EFITA

Digitizing Agriculture

Conference Proceedings

12th EFITA-HAICTA-WCCA CONGRESS 2019

Edited by

D. Bochtis, M. Lampridi, A. C. Tagarakis, C. Achillas

2019

EFITA-HAICTA-WCCA CONGRESS

A Conference supported by CIGR VII and INFITA

Rhodes, Greece | June 27 - 29

EFITA

EUROPEAN FEDERATION FOR INFORMATION TECHNOLOGY IN AGRICULTURE, FOOD AND THE ENVIRONMENT

Digitizing Agriculture

Conference Proceedings

Edited by

Dionysis Bochtis, Maria Lampridi, Aristotelis Christos Tagarakis, Charisios Achillas

ISBN 978-618-84798-0-7

EFITA CIGR 1930 HAICTA INFITA EurAgEng

iBO Institute for Bio-economy & Agri-technology

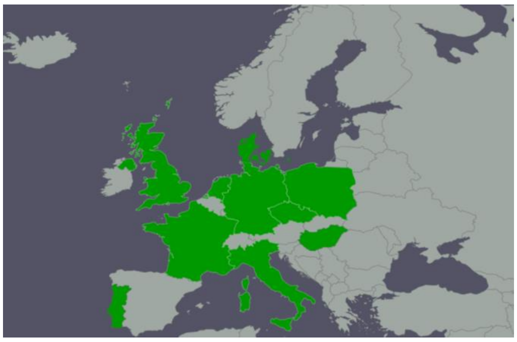
INTERNATIONAL HELLENIC UNIVERSITY

Past Conferences

"European Federation for Information Technology in Agriculture, Food and the Environment"

The first EFITA was held in 1997.

The conference facilitates participants form over 25 countries worldwide dealing with knowledge sharing and thinking on the future of ICT technologies within the agri-food and bio-resource sectors. EFITA conferences include ICT demonstrations, digital poster presentations, plenary sessions, breakout sessions in various topical areas (e.g. remote sensing, robotics, farm management information systems, data mining and analysis, agri-food and biomass chains, as well as ICT-agri education and professional issues), special sessions, and technical tours to exceptional paradigms of ICT applications in real-life primary production.



1997 - Copenhagen (DK) - Iver Thysen
1999 – Bonn (DE) - Gerhard Schiefer
2001 – Montpellier (FR) - Jérôme Steffe and Francis Sévila
2003 – Debrecen (HU) - Miklos Herdon
2005 – Vila Real (PT) - José Boaventura
2007 – Wageningen (NL) - Caroline Parker
2009 – Torino (IT) - Sjaak Wolfert
2011 – Prague (CZ) - Pavel Simek
2013 – Torino (IT) - Remigio Berruto
2015 – Pozman (PL) - Jerzy Weres
2017 – Montpellier (FR) - Jean-Pierre Chanet, Olivier Naud, Bruno Tisseyre
2019 – Rhodes (GR) - Dionysis Bochtis

EFITA 2019 Organizers

Institutes

Institute for Bio-economy
and Agri-technology
Center for Research and Technology Hellas

Universities

International Hellenic University
Department of Supply Chain Management

Organizations

European Federation for Information
Technology in Agriculture, Food and the
Environment

International Network for Information
Techiology in Agriculture.

European Society of Agricultural Engineers

Hellenic Association for Information and
Communication Technologies in Agriculture,
Food and Environment

Conference Chairman

Dionysis Bochtis
Ibo/CERTH

Organizing Committee

Chairman

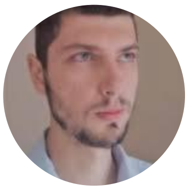
Dr. Dimitrios Aidonis
Assistant Professor
International Hellenic University

Members

Dr.Charisios Achillas
Assistant Professor
International Hellenic University

Dr. Christos Keramidas
Assistant Professor
International Hellenic University

Dimitrios Katikaridis
Technical Scientific Staff
Institute for Bio-economy and Agri-technology
iBO/CERTH



Vasileios Moysiadis
Technical Scientific Staff
Institute for Bio-economy and Agri-technology
iBO/CERTH

Dr. Alexandra Michailidou
Collaborating Researcher
Aristotle University of Thessaloniki

Soumela Chatziantoniou
Technical Staff
International Hellenic University

Elias Tsourapas
IT Manager
HAICTA

International Scientific Committee

Michael Clasen
President
EFITA Board

Yiannis Ampatzidis
Assistant Professor
University of Florida

Simon Pearson
Director
Lincoln Institute of Agrifood

Fedro Zazueta
President iAABE
Secretary General CIGR

Pål Johan From
Professor
NMBU

Bui Thi Nga
Associate Professor
Vietnam National University

Spyros Fountas
Associate Professor
Agricultural University of Athens

Lampros Stergioulas
Professor
University of Surrey

Daoliang Li
Professor
China Agricultural University

Jean-Paul Douzals
Researcher
IRSTEA

Thomas Bartzanas
Research Director
iBO

Fatima Baptista
Professor
Universidade de Evora

Emilio Gil
Associate Professor
Universitat Politecnica De Catalunya

Gerrit Polder
Senior Researcher
Wageningen University

Patrizia Busato
CIGR VII Chair
University of Turin

Till Riedel
Lab Lead
KIT/TECO

Tamme Van Der Wal
Partner
AeroVision B.V

Monica Varga
Research Fellow
Kaposvar University

Michel Berducat
Deputy Director
IRSTEA

Panos Pardalos
Professor
University of Florida

Seishi Ninomiya
Professor
University of Tokyo

Jean-Pierre Chanet
Researcher
IRSTEA

Kostas Arvanitis
Professor
Agricultural University of Athens

Markus Gandorfer
GIL President
Bavarian State Research Center

Krzysztof Koszela
Associate Professor
Poznan University of Life Sciences

Bournaris Thomas
Associate Professor
Aristotle University of Thessaloniki

Jose Aguilar Pereira
Professor
University of Costa Rica

Andreas Meyer Aurich
GIL Vice-President
ATB

Koutsou Stavriani
Professor
Alexander Technological Institute

Frits Van Evert
Senior Researcher
Wageningen University

Dimitrios Moshou
Professor
Aristotle University of Thessaloniki

Ragkos Athanasios
Researcher
Demeter

Laszlo Varallyai
Assistant Professor
University of Debrecen

Maro Vlachopoulou
Professor
University of Macedonia

Jiannong XIN
Associate Director
University of Florida

Michael Nørremark
Senior Advisor
University of Aarhus

Paolo Balsari
Full Professor DiSAFA
University of Turin

Ian Houseman
Secretary General EFITA
EFITA

Jurgen Vangeyte
Scientific Director
ILVO

Parisses Constantinos
Professor
Western Macedonia University

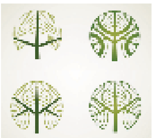
Conference Topics

Sensors
Wireless sensor networks
Remote Sensing and GIS applications
Bio-sensors
Physical and Chemical sensors
Optical sensors

Decision
Modelling and Simulation
Prediction models
Multi-Agent systems
Planning tools
Environmental ICT management systems
Farm management systems (FMIS)
Decision Support Systems

Action
On-line farm services
Web applications
Cloud computing applications
Monitoring Robots, Action Robots
Machine embedded ICT tools

Data
Big data management
Data mining
Data visualization
Data and Knowledge Management
Metadata and data standards
Ontologies for agriculture
Knowledge bases and Knowledge repository services
Web of Data and Open Data
Image processing



Cross-cutting Themes
Social Networking in agriculture
e-agribusiness
ICT and business
Rural economies and ICT policies for rural development
Traceability tools
Human-Computer Interaction
Open topic

Preface

As humanity is faced with a number of issues that threaten its well-being and preservation, such as the production of adequate and safe food for the rapidly increasing population and the protection of the deteriorating environment, the need for immediate measures is imperative. Technology continuous to evolve in a fast pace, penetrating all aspects of human activity, automizing and facilitating tasks and in most cases contributing towards the increase of production and the improvement of labor conditions. In the evolving field of agriculture, the advent of information and communication technologies aims to contribute towards the increase of production while lowering the environmental impact. Additionally, the automation of agricultural tasks, upgrades the farmers' working environment, making it an attractive occupation as it can involve a variety of different specialties.

The introduction of applied sciences in Agriculture has turned the research interest towards new and innovative topics related to primary production such as: Robotics and automation, Machine-embedded ICT tools, Internet of Things (IoT) in agri-food production and agri-food chains, Remote sensing and GIS applications, AI applications in agriculture, Decision support systems for agriculture, ICT applications for precision farming, Traceability and agri-food chains systems and Big data and data mining for agricultural information systems. The scientific advancements in the field of Agri-technology are being promoted under the theme of the **12th EFITA International Conference - "Digitizing Agriculture".**

The **EFITA Conference,** with presence from 1997 and participants from 25 countries, is an International Conference dedicated to the state-of-the-art and future use of ICT in the agri-food and bio-resources sectors. The conference is supported by the **European Federation for Information Technology in Agriculture, Food and the Environment** (EFITA) with the collaboration of the **World Congress on Computers in Agriculture** (WCCA) and the **Hellenic Association for Information and Communication Technologies in Agriculture, Food and Environment** (HAICTA), and the **International Commission of Agriculture and Biosystems Engineering** (CIGR) **Section VII** (Information Technology).

These proceedings contain selected peer reviewed research papers accepted for publication at the **12th EFITA-HAICTA-WCCA Congress** which was held in **Rhodes Island, Greece**, on **June 27-29, 2019**. Through this preface we take the opportunity to thank the authors and EFITA 2019 participants for their contribution to these proceedings. We would also like to thank the reviewers, members of the Scientific Committee, for ensuring the Proceedings quality as well as the invited keynote speakers Prof. George Vellidis, and Prof. Raj Khosla for their inspired lectures.

The Editors,

Prof. Dionysis Bochtis	**Ph.D.C. Maria Lampridi**	**Dr. Aristotelis-Christos Tagarakis**	**Dr. Charisios Achillas**
iBO/CERTH	iBO/CERTH	iBO/CERTH	IHU

Contents

Thematic Area: Sensors

Thematic Area: Data

Thematic Area: Decision

Thematic Area: Action

Thematic Area: Cross-Cutting Themes

NOVEL PROXIMAL AND REMOTE SENSING APPROACHES FOR DERIVING VEGETATION INDICES: A CASE STUDY COMPARING PLANT-O-METER AND SENTINEL-2 DATA

Miloš Pandžić[1], Aristotelis C. Tagarakis[1, 2], Vasa Radonić[1], Oskar Marko[1], Goran Kitić[1], Marko Panić[1], Nataša Ljubičić[1]

[1]University of Novi Sad, BioSense Institute, Serbia

[2]Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Greece

milos.pandzic@biosense.rs, a.tagarakis@certh.gr, vasarad@biosense.rs, oskar.marko@biosense.rs, gkitic@biosense.rs, panic@biosense.rs, natasa.ljubicic@biosense.rs

ABSTRACT

With an increasing interest of the agricultural community in precision agriculture, this paper aims to compare two novel sensing approaches for crop monitoring. The recently developed multispectral proximal sensor named Plant-O-Meter and Sentinel-2 satellite, which carries a multispectral optical instrument, are two sensors suitable for agricultural applications. Each of them has pros and cons regarding spatial, spectral and temporal resolutions and their complementary use will surely bring added value compared to information retrieved by a single sensor. In order to correctly address the problem of data fusion, compatibility studies between the two sensors are necessary. In this study, a maize field was sensed on several dates in 2018 growing season using both sensors. Numerous vegetation indices based on different spectral channel combinations were calculated and the results were compared using linear regression analysis. First results showed good positive correlations between the indices obtained by the two sensors.

Keywords*:* crop monitoring, proximal sensing, Sentinel-2, vegetation indices, correlation

1. INTRODUCTION

Recent advances in technology provided an unprecedented opportunity for further development of precision agriculture that has been practiced commercially since 1990's (Mulla, 2013). Both remote and in-field sensors are used for monitoring plant deficiency for nutrients and water, plant health status and soil condition (Lee et al., 2010) and the development of low-cost sensors, as well as the liberalization of data access by data providers such as the European Space Agency (ESA) and NASA, have paved the way for acquisition of vast amounts of sensor data. However, compatibility studies between datasets acquired by different sensors are necessary prior to any kind of data fusion in practice.

Proximal sensing or ground-based remote sensing is performed by sensors at a relatively short distance from the object of interest. Hand-held devices or sensors mounted on tractors and other vehicles are usually referred to as proximal sensors. Their limitation is the small area coverage (Jackson, 1986), but they also have significant advantages, such as high spatial resolution and independent choice of the time of acquisition. Another advantage is that their measurements are not compromised by cloudiness and are ideal for practical applications such as on-the-go variable rate fertilization (Shanahan et al., 2008). Over the years, various different optical proximal sensors found practical applications, such as SPAD meter (Konica Minolta Inc., Osaka, Japan), Hydro N-sensor (Yara International ASA, Oslo, Norway), GreenSeeker (Tribmle Inc., CA, USA), Crop Circle (Holland Scientific, NE, USA), CropScan (Next Instruments, Sydney, Australia), etc.

On the other hand, satellite remote sensing has been used in agriculture since 1970's when the first Landsat satellite was launched. Over the period of nearly half a century, the resolution of satellite images, as well as the revisit frequency, increased dramatically (Mulla, 2013). However, a big drawback of this sensing approach has historically been the high price of satellite imagery. ESA and NASA changed their policies in the last decade and made certain satellite imagery available to general public at no cost (Woodcock et al., 2008; Aschbacher and Milagro-Pérez, 2012). Landsat's 40-year long archive is now freely available together with on-going Landsat missions and state-of-the-art Earth observation program Copernicus operated by ESA on behalf of European Commission. These led to an increased interest of the agricultural community toward satellite remote sensing in the last decade.

Although proximal and remote sensing were extensively studied for assessing crop dynamics (Corti et al., 2018), direct inter-comparison between satellite remote sensing and proximal sensors with respect to crop monitoring has rarely been discussed. Bausch and Khosla (2010) compared QuickBird satellite-derived indices with ground-based Exotech radiometer-derived indices and found good correlation, with highest agreement in green normalized difference vegetation index normalized for reference area (NGNDVI). Caturegli et al. (2015) tested ground-based multispectral measurements (using Licor spectroradiometer and GreenSeeker) and GeoEye-1 satellite images for estimating nitrogen status of turfgrasses. Comparing NDVI values acquired from these instruments, the highest Pearson correlation coefficient was found between GreenSeeker and satellite derived NDVI ($r \approx 1$). Yang et al. (2008) found substantial linear correlation ($r > 0.7$) between NDVI measured from Formosat-2 satellite images and ground portable spectroradiometer GER-2600. Wagner and Hank (2013) revealed Pearson correlation coefficient of 0.85 between RapidEye and YARA-N sensor-derived Red Edge Inflection Point (REIP). Within this study, the necessary modification was made in RapidEye measurements using YARA-N sensor-based model, so that the REIP could be calculated. Bu et al. (2017) confirmed that yields of sugar beet root, spring wheat, corn and sunflower can be predicted with GreenSeeker, Crop Circle and RapidEye red and red-edge imagery.

The use of vegetation indices (VIs) is of great importance in monitoring crop dynamics and predicting the yield. Hence, it is essential to quantify the level of similarity between different sensor measurements prior to data fusion. In this paper, various VIs derived from measurements made with a recently developed, active, multispectral proximal sensor named Plant-O-Meter (POM), were compared to VIs derived from Sentinel-2 optical satellite imagery. Although the spatial resolution of POM is higher than Sentinel-2's and more detailed information can be obtained, the latter would be more suitable for covering larger agricultural areas. In this regard, POM measurements could serve as ground-truth or they could be used for on-the-go in-field applications. Nevertheless, both sensors represent modern active optical instruments that are likely to find broader use in the near future.

2. Materials and Methods

The present study was carried out during the 2018 growing season on a commercial field located in Begeč in Serbia (45° 14' 32.712" N and 19° 36' 21.486" E), whose size was 6 ha. The field was sown with "Exxupery" hybrid (R.A.G.T. Semences, France) of maize (Zea mays L.) on 15 April 2018. Seeding was done in 300 m long rows, at the plant distance of 0.2 m within rows and 0.7 m between rows. A total of 300 kg ha^{-1} of 15:15:15 NPK fertilizer was applied at planting.

In-field reflectance measurements were made using POM sensor, recently developed by BioSense Institute (Republic of Serbia). This proximal sensor is connected to Android-operated devices through a user-friendly application and has the ability to record georeferenced point measurements and map the canopy properties of a field crop, using the internal GPS of the Android device. It records the data in four different spectral bands, namely blue (465 nm), green (535 nm), red (630 nm) and near-infrared (850 nm). Every tenth row of the experimental field was scanned by walking along the rows, holding the sensor directly on top of the crop row with the scanning footprint perpendicular to the row direction. The measuring frequency was 1 Hz, which roughly corresponded to 1 m distance between the POM record points along the row. POM measurements were performed at four different dates and were carried out in the following stages of maize development: 6-leaf growth stage (V6), beginning of tasseling (VT), silking (R1) and at the end of blister stage (R2), (Table 1).

Table 1. Corresponding acquisition dates for POM and Sentinel-2 and development stage of maize.

POM date	Sentinel-2 date	Crop development stage
01.06.2018	30.05.2018	6-leaf (V6)
21.06.2018	24.06.2018	Tassel (VT)
04.07.2018	14.07.2018	Silking (R1)
26.07.2018	29.07.2018	Blister (R2)

Sentinel-2 is a constellation of two identical satellites and the joint revisit time of A and B satellites is 5 days at the equator. Each carries an optical multispectral instrument that provides images in 13 spectral bands with spatial resolutions of either 10, 20 or 60 m (European Space Agency, 2015). Bands used in the experiment are blue (490 nm), green (560 nm), red (665 nm) and NIR (842 nm) bands with a 10 m resolution and the narrow NIR (865 nm) band with a 20 m resolution. With respect to POM measurement dates, corresponding cloud-free satellite images were downloaded and processed. Atmospherically corrected images were downloaded from the official Copernicus Open Access Hub (https://scihub.copernicus.eu/) and processed with official Sentinel-2 Toolbox (SNAP) software and QGIS. Acquisition dates for Sentinel-2 images are given in Table 1.

Since the narrow NIR band of Sentinel-2 images was only available at a 20 m resolution, all images were resampled using the nearest neighbor method. Thus, the blue, green, red and NIR bands from Sentinel-2 images were down-sampled from 10 m to 20 m resolution.

Due to the higher resolution of POM measurements compared to Sentinel-2 images, i.e. several POM measurements points fell within a single Sentinel-2 image pixel (Fig. 1), all POM measurements inside a Sentinel-2 pixel were averaged. By employing this, there was only one corresponding value per POM spectral band for a single image pixel. Hence, 1-1 mapping between measurements of the two sensors was achieved. Using different spectral band combinations, various indices were calculated (Table 2).

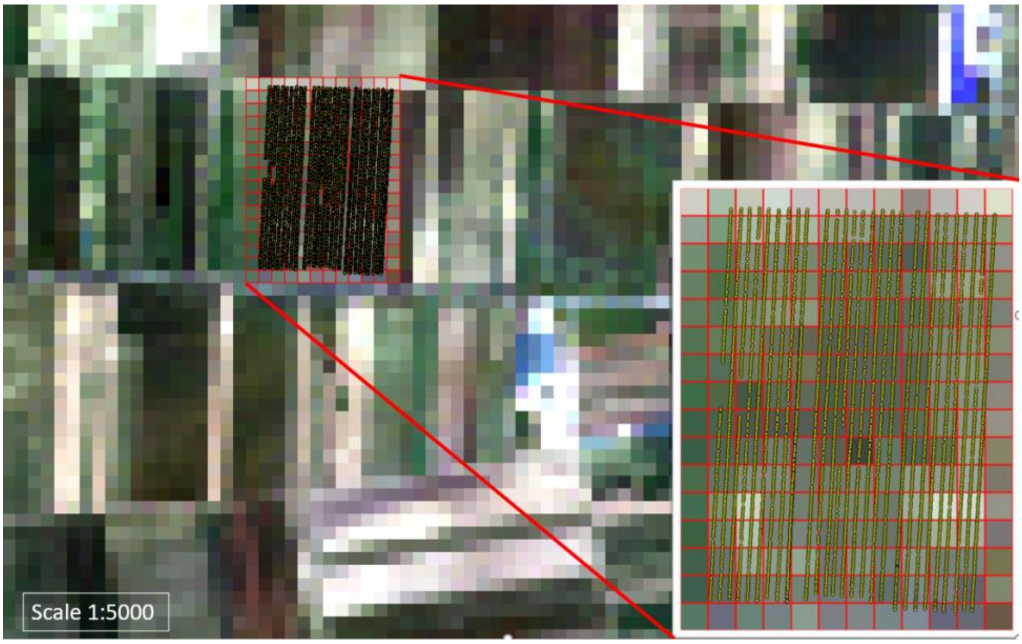


Figure 1. Sentinel-2 image of the experimental field in Begeč at 20 m resolution where yellow dots represent POM measurement points.

Pixels that were known to be outliers were manually excluded from further analysis. Those were either border pixels, contaminated by the features outside the field, or pixels contaminated by other objects located inside the parcel.

3. Results and Discussion

The analysis of the Sentinel-2 image acquired on 24 June 2018 provided poor results due to the significant effect of a layer of clouds over the experimental field. Therefore, this date was excluded from the analysis. This is a good example of the constraints of the use of optical satellite images as they highly depend on the weather (Mulla, 2013).

Table 2. Coefficient of determination (r^2) and Root Mean Square Error (RMSE) form the regression between indices calculated from Sentinel-2, using the wide range NIR band, and POM.

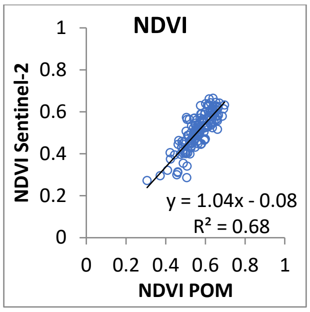
date	01-06-2018		04-07-2018		26-07-2018	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
NDVI	0.680	0.075	0.162	0.093	0.036	0.093
SR	0.612	0.905	0.147	4.283	0.045	4.616
IPVI	0.668	0.045	0.162	0.047	0.036	0.046
NDVIg	0.616	0.058	0.102	0.209	0.008	0.210
NDVIb	0.652	0.103	0.050	0.096	0.000	0.134
SIPI	0.546	0.527	0.059	0.266	0.000	0.267
EVI	0.325	0.182	0.002	0.719	0.000	1.327
GSAVI	0.614	0.091	0.105	0.319	0.008	0.318
GOSAVI	0.615	0.058	0.103	0.210	0.008	0.210
GCI	0.574	0.500	0.087	4.625	0.028	4.611
NLI	0.478	0.014	0.124	0.006	0.060	0.004
TDVI	0.672	0.115	0.167	0.091	0.034	0.089
WDRVI	0.648	0.120	0.156	0.226	0.041	0.230
GRNDVI	0.659	0.062	0.177	0.236	0.022	0.239
GBNDVI	0.676	0.061	0.097	0.244	0.008	0.274
RBNDVI	0.700	0.131	0.143	0.152	0.010	0.184
PNDVI	0.686	0.082	0.155	0.259	0.017	0.288
Average	**0.580**	**0.179**	**0.111**	**0.677**	**0.020**	**0.736**

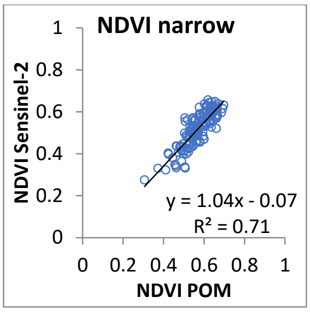
Table 3. Coefficient of determination (r^2) and Root Mean Square Error (RMSE) form the regression between indices calculated from Sentinel-2, using the narrow range NIR band, and POM.

date	01-06-2018		04-07-2018		26-07-2018	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
NDVI	0.710	0.069	0.162	0.100	0.045	0.103
SR	0.644	0.863	0.135	4.791	0.058	5.615
IPVI	0.696	0.042	0.162	0.050	0.045	0.052
NDVIg	0.630	0.059	0.106	0.217	0.012	0.224
NDVIb	0.676	0.099	0.044	0.101	0.002	0.141
SIPI	0.579	0.522	0.057	0.265	0.001	0.264
EVI	0.344	0.176	0.003	0.715	0.000	1.312
GSAVI	0.627	0.093	0.108	0.331	0.012	0.339
GOSAVI	0.629	0.060	0.107	0.218	0.012	0.225
GCI	0.595	0.515	0.094	5.015	0.032	5.349
NLI	0.493	0.014	0.116	0.006	0.097	0.004
TDVI	0.702	0.106	0.167	0.096	0.043	0.098
WDRVI	0.677	0.114	0.153	0.243	0.052	0.261
GRNDVI	0.683	0.057	0.178	0.248	0.029	0.260
GBNDVI	0.696	0.056	0.095	0.255	0.017	0.293
RBNDVI	0.728	0.124	0.136	0.162	0.022	0.200
PNDVI	0.710	0.076	0.151	0.274	0.028	0.313
Average	**0.603**	**0.175**	**0.110**	**0.733**	**0.028**	**0.842**

The linear regression analysis provided an insight of which indices calculated using POM are in better agreement with the same ones calculated using Sentinel-2 satellite images. The differences were mainly due to the deviations in the operating wavelengths for the two sensors and in the different sensitivity of each sensor at different bands. Sentinel-2 provides two measurements in the NIR channel: wide (785 – 900 nm) and narrow (855 – 875 nm) range. According to the statistical analysis, using the narrow range NIR in calculations of Sentinel-2 indices provided better correlation to the POM indices (Tables 1, 2; Figure 2). This was expected since the measuring range for the two bands, narrow NIR band of Sentinel-2 and NIR band of POM, is much closer.

In general, good positive correlations were obtained for most of the indices measured by the two sensors at V6 growth stage of maize (01-06-2018; Tables 2 and 3). This is an indication that POM has a high potential for providing reliable measurements of the canopy reflectance and plant status during maize growing stages, and it can serve as a good alternative to the satellite sensors, having the benefits that the active proximal sensors offer: high spatial resolution, flexibility in the measurement timing and independence from cloudiness, as given by (Bu et al., 2017).





a **b**

Figure 2. Linear regression between the NDVI calculated from POM measurements and Sentinel-2 satellite images using wide (a) and narrow (b) bands, at V6 maize growth stage.

POM measurements at the V6 growth stage showed good correlation with Sentinel-2 results, mainly due to the uniformity of the color of the canopy across the field. Concerning the NDVI, which is the most widely used vegetation index (Tagarakis and Ketterings, 2017; Hatfield et al., 2008), the linear regression showed significant correlation for the narrow band (r^2= 0.71, RMSE = 0.069; Table 3, Figure 2) showing a 1:1 relationship; the slope of the linear model was almost 1 and the constant approached 0 (Figure 2). After tasselling, the measurements showed considerably lower correlation between the two sensors explained by the mixture of colors after the tassels appear, and the different shades of the canopy from green to yellow as the plants approach maturity. Due to the large difference in the spatial resolution of the measurements of the two sensors in the study, this random mixture of colors affected the results of each sensor differently.

4. Conclusions

Ground based proximal sensing provides comparable results to the indices calculated from Sentinel-2 satellite images. Cloudiness is an important limiting factor of satellite remote sensing. POM active proximal sensor can be an alternative to satellite images, as it provides comparable measurements at a high spatial resolution, independent of weather and illumination conditions. The plant development stage plays an important role in the agreement between the indices derived by POM and Sentinel-2 due to the large difference in spatial resolution of the measurements.

Acknowledgement

This work was supported by Ministry of Education, Science and Technological Development of Republic of Serbia through project III 44006 - Development of new ICT, based on advanced mathematical methods, with application in medicine, telecommunications, power systems, protection of national heritage and education and by Provincial Secretariat for Higher Education and Scientific Research of Vojvodina through project Sensor technologies for integrated monitoring of agricultural production.

References

Aschbacher, J. and Milagro-Pérez, M. P. (2012) 'The European Earth monitoring (GMES) programme: Status and perspectives', Remote Sensing of Environment, 120, pp. 3-8. doi: 10.1016/j.rse.2011.08.028.

Bausch, W.C. and Khosla, R. (2010) 'QuickBird satellite versus ground-based multi-spectral data for estimating nitrogen status of irrigated maize', Precision Agriculture, 11, pp. 274-290. doi: 10.1007/s11119-009-9133-1.

Bu, H., Sharma, L. K., Denton, A. and Franzen, D. W. (2017) 'Comparison of Satellite Imagery and Ground-Based Active Optical Sensors as Yield Predictors in Sugar Beet, Spring Wheat, Corn, and Sunflower', Agronomy Journal, 109, pp. 299–308. doi: 10.2134/agronj2016.03.0150.

Caturegli, L., Casucci, M., Lulli, F., Grossi, N., Gaetani, M., Magni, S., Bonari, E. and Volterrani, M. (2015) 'GeoEye-1 satellite versus ground-based multispectral data for estimating nitrogen status of turfgrasses', International Journal of Remote Sensing, 36, pp. 2238–2251. doi: 10.1080/01431161.2015.1035409.

Corti, M., Cavalli, D., Cabassi, G., Gallina, P. M. and Bechini, L. (2018) 'Does remote and proximal optical sensing successfully estimate maize variables? A review', European Journal of Agronomy, 99, pp. 37-50. doi: 10.1016/j.eja.2018.06.008.

European Space Agency (2015) 'Sentinel-2 User Handbook', Issue 1, pp. 1-64. https://sentinel.esa.int/documents/247904/685211/sentinel-2_user_handbook

Hatfield, J.L., Gitelson, A.A, Schepers, J.S. and Walthall C.L. (2008) 'Application of Spectral Remote Sensing for Agronomic Decisions', Agronomy Journal, 100, pp. 117–131. doi: 10.2134/agronj2006.0370c.

Jackson, R. D. (1986) 'Remote Sensing of Biotic and Abiotic Plant Stress', Annual review of Phytopathology, 24, pp. 265-287. doi: 10.1146/annurev.py.24.090186.001405.

Lee, W.S., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D. and Li, C. (2010) 'Sensing technologies for precision specialty crop production', Computers and Electronics in Agriculture, 74, pp. 2–33. doi: 10.1016/j.compag.2010.08.005.

Mulla, D. J. (2013) 'Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps', Biosystems Engineering, Special Issue: Sensing in Agriculture, pp. 358–371. doi: 10.1016/j.biosystemseng.2012.08.009.

Shanahan, J.F., Kitchen, N.R., Raun, W.R. and Schepers, J.S. (2008) 'Responsive in-season nitrogen management for cereals', Computers and Electronics in Agriculture, 61, pp. 51-62. doi: 10.1016/j.compag.2007.06.006.

Tagarakis, A. C. and Ketterings, Q. M. (2017) 'In-Season Estimation of Corn Yield Potential Using Proximal Sensing', Agronomy Journal, 109, pp. 1323-1330. doi: 10.2134/agronj2016.12.0732.

Wagner, P. and Hank, K. (2013) 'Suitability of aerial and satellite data for calculation of site-specific nitrogen fertilisation compared to ground based sensor data', Precision Agriculture, 14, pp. 135–150. doi: 10.1007/s11119-012-9278-1.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A. and Wynne, R. (2008) 'Free access to Landsat imagery', Science, 320, pp. 1011. doi: 10.1126/science.320.5879.1011a.

Yang, C.-M., Liu, C.-C. and Wang, Y.-W. (2008) 'Using Formosat-2 Satellite Data to Estimate Leaf Area Index of Rice Crop', Journal of Photogrammetry and Remote Sensing, 13, pp. 253-260.

USE OF A 3D IMAGING DEVICE TO MODEL THE COMPLETE SHAPE OF DAIRY CATTLE AND MEASURE NEW MORPHOLOGICAL PHENOTYPES

C. Allain[1], A. Caillot[2], L. Depuille[1], P. Faverdin[2], J. M. Delouard[3], L. Delattre[3], T. Luginbuhl[3], J. Lassalas[2], Y. Le Cozler[2]

[1]Institut de l'Elevage, Monvoisin, 35652 Le Rheu, France

[2]PEGASE, Agrocampus Ouest, INRA, 35590 Saint-Gilles, France

[3]3DOUEST, 5 Rue de Broglie, 22 300 Lannion

clement.allain@idele.fr, anais.caillot@inra.fr, laurence.depuille@idele.fr, philippe.faverdin@inra.fr, delouard@3douest.com, delattre@3douest.com, luginbuhl@3douest.com, jacques.lassalas@inra.fr, yannick.lecozler@agrocampus-ouest.fr

ABSTRACT

Monitoring of body weight variation, body condition and/or morphological changes allows optimal management of animal health, production and reproduction performance. However, due to implementation difficulties (handling, time consumption, investments), this type of monitoring is not very common within commercial farms. The development of three-dimensional imaging technologies is an interesting solution to meet these needs. The purpose of this study was to develop, test and validate a device (Morpho3D) offering the possibility of recording and analysing complete 3D forms of dairy cattle. To evaluate the performance of this tool, manual measurements were performed on 30 Holstein dairy cows: height at withers (HG), chest circumference (TP), chest depth (PP), hip width (LH), buttock width (LF) and ischium width (LI). They were compared to those estimated by the Morpho3D device. Correlations coefficients between Morpho3D measurements and manual measurements were 0.89 for PP, 0.80 for LH, 0.78 for TP, 0.76 for LF, 0.63 for LI and 0.62 for HG. For the Morpho3D system, the repeatability standard deviation ranged from 0.34 to 1.89 (coefficient of variation (CV) from 0.26 to 9.81) and the reproducibility standard deviation ranged from 0.55 to 5.87 (CV from 0.94 to 7.34). These values are close to those obtained with manual measurements. This new device offers the possibility of measuring new phenotypes such as the total volume of the animal or the body surface and thus offers new opportunities for new researches and studies.

Keywords*:* 3D imaging, body measurement, dairy cow, precision breeding, sensors

1. INTRODUCTION

Monitoring the evolution of the morphology of dairy cattle allows to adapt feeding, reproduction and general management for optimal operation of the farm. Currently, with the exception of weight, most measurements are done manually (tape, measuring rod) or visually (Heinrichs and Hardgrove, 1987). These measurements, which are time-consuming, are sources of stress and accidents for farmers and animals and rarely available on the farm. Developing precise, automatic and easy-to-use tools to overcome these problems is therefore of interest. Imaging techniques offer interesting alternatives to manual measurements and / or costly methods (Pezzulo et al, 2018). 2D imaging approaches, used in pork with some success (Marchant et al, 1993, Schofield et al, 1998), do not allow approaching the third dimension. In addition, distortion problems, the calibration procedure, the need for multiple cameras and finally the complexity of the 3D reconstruction models have reduced their use. The 3D imaging technologies have thus been used successfully to analyse the Body Condition Score (BCS) of dairy cattle (Fischer et al, 2015, Kuzuhara et al, 2015). Negretti et al (2008), Buranakarl et al (2012), Guo et al (2017) and Pezzuolo et al (2018) have also developed 3D image technologies with the aim of obtaining a 3D image of the whole animal, but many problems remain. Pezzuolo et al (2018), used low-cost portable equipment based on the Microsoft Kinect v1 sensor, and concluded that their method still requires a lot of engineering to enable the automatic collection and extraction of data satisfactorily. A new device (called "Morpho3D") has been developed to easily capture the complete shape of cows and measure their morphological traits. Measurements obtained with this device were compared with values collected directly from live animals. To validate the method, repeatability and reproducibility were also analysed.

2. MATERIAL AND METHODS

2.1 Device

The device tested is installed in the dairy experimental farm of INRA-UMR PEGASE located in Le Rheu (France). The system comprises of a total of 5 camera sensors, each in combination with a laser projector emitting at 650 nm to limit the risks to humans and animals. The pairs ‘camera-laser’ are installed on a mobile portico (Figure 1), set at 0.40 and 1.77 m above the ground on both sides and a fifth in the middle of the top of the portico. The portico moves at an average speed of 0.5 m s^{-1} from back to front and returns to its original position at an average speed of 0.3 m s^{-1}. 80 photos per second and per camera are recorded only during the first way of the portico (Le Cozler et al. 2019a). To secure the entire device, four stainless steel cables mark the movement of the cows. Animals can also be blocked by a self-locking head door if necessary.

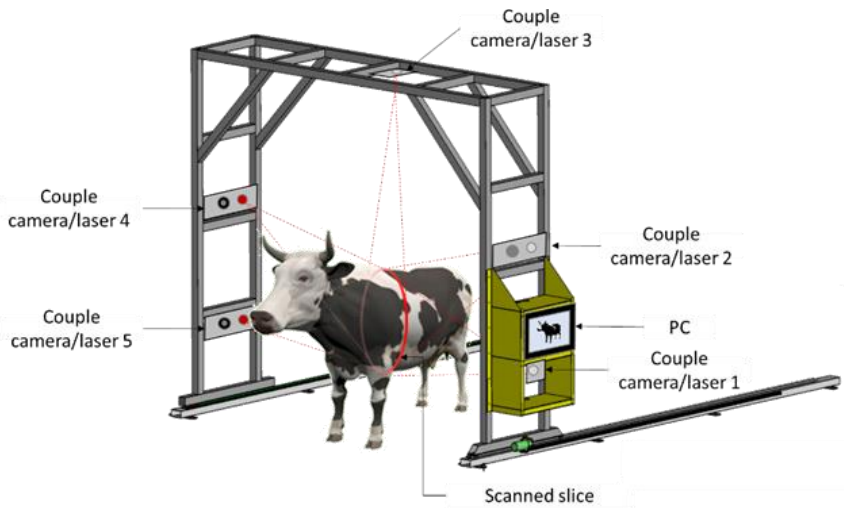


Figure 1. The Morpho3D scanning system

The images of the laser bands projected on the animal are captured by their corresponding camera and sent to a computer to reconstruct the 3D information. The images of each camera are first processed to build point clouds and the complete 3D reconstruction of the animal is performed by recording and merging the multiple views of the 3D data from the point clouds of the 5 camera-laser pairs (Figure 2). A distance threshold is set to ignore points too far from the camera and not belonging to the animal. An example of the living animal process is available at https://vimeo.com/219370900.

Free software for processing and editing 3D triangular meshes was then used to clean the data (Meshlab, Cignoni et al, 2008). A Poisson surface reconstruction algorithm was applied to construct a triangulated mesh and perform shape smoothing (Kazhdan and Hoppe, 2013). The different stages of the treatment are shown in figure 3.

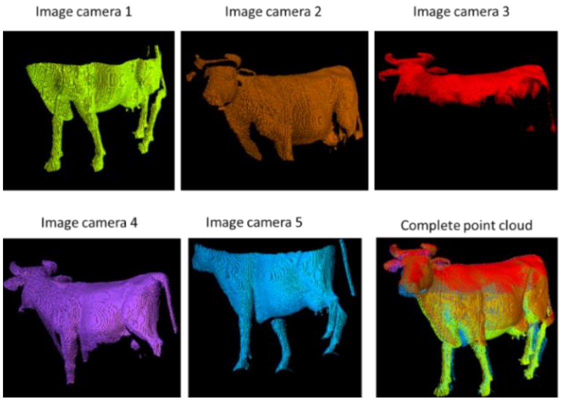


Figure 2. Point cloud reconstruction

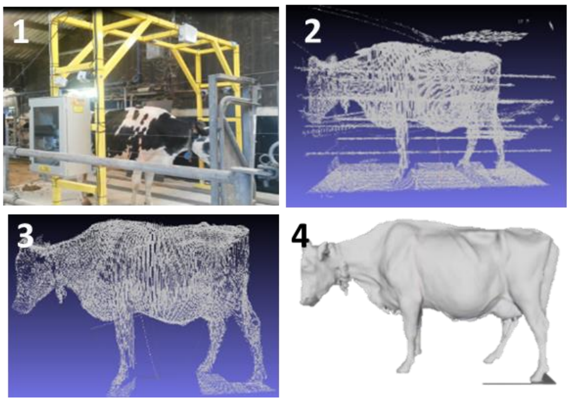


Figure 3. From Acquisition to Final Data:
Image 1: Data Acquisition. Image 2: Raw cloud. Image 3: Cloud after cleaning. Image 4: final image after normalization and Poisson reconstruction.

2.2 Animals and measures

Data were collected between May and June 2017 on 30 Holstein dairy cows, aged 3.0 (+ 1.2) years on average, producing 25.5 (+ 3.6) kg of milk per day. Manual measurements were made directly on cows and this same measurements were estimated on 3D images from Morpho3D device. For both methods, 6 of these 30 cows underwent a series of repeated measurements of the same indicators (6 times each), in order to estimate the reproducibility of the methods. For the estimation of repeatability, a plastic cow model was used. The indicators measured on live animals and images included wither height (WH), heart girth (HG), chest depth (CD), hip width (HW), thirl width (TW) and ischium width (IW). For manual measurements, a tape and a measuring rod were used. In the reconstructed images, measurements were made using dedicated software (Metrux2α®, 3D Ouest).

2.3 Data analysis

Repeatability and reproducibility of the two methods were evaluated. Repeatability makes it possible to evaluate the error generated when estimating an indicator several times on the same sample with the same methodology, in the same environment, over a short period of time. It was estimated by making measurements 6 times the same day, from the same 3D scan of the same animal (plastic model cow). Reproducibility evaluates the same error but under varying environmental conditions. It was estimated by scanning 6 cows, 6 times each the same day, and making the measurements once on each 3D scan. The 3D variations were corrected to account for the effect of animals in extracting ANOVA model residues. The coefficients of variation for repeatability (CVr) and reproducibility (CVR) were evaluated as $CVr = (\sigma r / \mu r) * 100$ and $CVR = (\sigma R / \mu R) * 100$, where σr and σR are respectively the standard deviations of the corrected 3D measurement for the repeatability and reproducibility datasets and μr and μR are respectively the average 3D measurements of the repeatability and reproducibility data. Similarly, the repeatability and reproducibility of manual measurements were estimated by correcting the variability of the measures for the effect of cows and operators (2 operators performed the same measures). The Anova 1 model then included "cow identity" as a factor in case of repeatability, and the Anova 2 model includes "cow identity" and "expert identity" as explanatory factors of the measure for reproducibility. The correlation analysis between the 3D image measurements and the reference values was performed using the statistical software R, (R Core Team

2013. R Foundation for Statistical Computing, Vienna, Austria) and the analyses concerning repeatability and reproducibility were carried out with the SAS software (SAS institute, 2016).

3. RESULTS AND DISCUSSION

Comparison between the two methods shows that most manual measurements have values lower than those obtained from 3D images (Table 1). The highest difference was observed for the ischium (difference of 11.2%), while the lowest was noted for the withers height (1.3%).

Table 1. Comparison of measurements manually performed on 30 cows or made from 3D images

Measure	Manual (cm)	Morpho3D (cm)	Error rate (%)	*p*- value
Heart girth (HG)	207.5	221.5	6.3	< 0.0001
Chest depth (CD)	79.4	83.8	5.3	< 0.0001
Wither height (WH)	146.9	148.8	1.3	< 0.003
Hip width (HW)	55.5	54.4	2.0	< 0.02
Thirl width (TW)	51.9	54.4	4.6	< 0.008
Ischium width (IW)	17.4	19.6	11.2	< 0.02

The correlation between the two types of measurements is also high (Table 2). The highest values were observed for chest depth (0.89) and the lowest values for ischium width (0.63). The prominent bones at the hips certainly explain the small differences observed between manual measurements and Morpho3D, as noted by Pezzulo et al (2018). On the contrary, the prominent bones are less visible for the ischium, which may explain the meager performance at this level. For some measurements (Heart girth or chest depth), an overestimation exists because in some cases, the position of the front leg on the image did not allow a satisfactory access. The correlation values between the two approaches are generally lower than those reported by Buranarkal et al (2012) and Pezzulo et al (2018). Buranarkal et al (2012) performed their measurements under laboratory conditions and used visual markings stuck on animals, unusable under commercial conditions. Pezzulo et al (2018) performed their analyses on mean values.

Table 2. Coefficient of correlation and P-value, between manual measurements and those obtained on 3D images

Measure	Coefficient of correlation	*p*- value*
Heart girth (HG)	0.78	< 0.001
Chest depth (CD)	0.89	< 0.001
Wither height (WH)	0.62	< 0.001
Hip width (HW)	0.80	< 0.001
Thirl width (TW)	0.76	< 0.01
Ischium width (IW)	0.63	< 0.01

** test of Student*

The repeatability and reproducibility values are quite similar between methods (Table 3). For data from 3D images, the σr ranged from 0.34 to 1.89 (CV 0.26 to 9.81) and σR from 0.55 to 5.87 (CV from 0.94 to 7.34).). Using manual measurements, the σr ranged from 0.21 to 1.32 (CV 0.11 to 10.30) and σR ranged from 0.49 to 1.19 (CV 0.42 to 4.46). According to Fischer et al (2015), measurement methods with repeatability and reproducibility CVs below 4% can be considered as interesting methods, which is the case in this study. Many authors stress the important effect of the animal's position on the fluctuations of the measurements made and the importance of selecting, often manually, the best images to limit undesirable variations (Kmet et al. 2000; Stajnko et al. 2008). Fischer et al (2015) also showed the impact of animal's position in the work done to estimate BCS by 3D imaging. But only a few authors went so far as to qualify (repeatability, reproducibility) the method tested.

Another important point that can change the quality of the images and therefore ultimately the repeatability and reproducibility of measurements, is the environment (Tscharke and Banhazi. 2013). Indeed, most technologies are sensitive to daylight and are conducted in controlled light conditions. Similarly, the control of animal movements to obtain exploitable images is a crucial point.

An estimate of the time spent per method to obtain all the values of the indicators used was carried out and corresponded to 2.5 and 14 minutes for the manual and automatic systems. In the second case, the acquisition is fast (6 s on average) but the time needed to analyse and get the final results was about 14 min. It is clearly possible to reduce this time in the future through the optimization of models and equations

Table 3. Repeatability and reproducibility of body measurements obtained directly from animals (Manual) or 3D images (Morpho3D). A plastic cow model was used for the repeatability study and 6 cows were used for the reproducibility study.

CVr and CVR are respectively the coefficients of variation for repeatability (CVr) and reproducibility (CVR), σr and σR the standard deviations of the corrected 3D measurement for the repeatability and reproducibility datasets and μr and μR the average 3D measurement of the repeatability and reproducibility data.

Measure		Repeatability			Reproducibility		
		μr (cm)	σr	CVr (%)	μR (cm)	σR	CVR (%)
Heart girth (HG)	Manual	194.2	0.21	0.11	204.2	0.86	0.42
	Morpho3D	195.8	1.89	0.97	221.1	5.87	2.63
Chest depth (CD)	Manual	75.1	0.42	0.56	79.1	0.49	0.62
	Morpho3D	76.5	0.44	0.58	84.4	0.92	1.09
Wither height (WH)	Manual	129.1	1.04	0.80	148.9	1.07	0.72
	Morpho3D	131.1	0.34	0.26	148.6	2.12	1.42
Hip width (HW)	Manual	39.8	0.35	0.88	55.5	1.01	1.82
	Morpho3D	39.9	0.67	1.68	58.6	0.55	0.94
Thirl width (TW)	Manual	50.9	0.36	0.71	50.8	1.19	1.82
	Morpho3D	52.6	0.34	0.64	55.5	1.82	3.28
Ischium width (IW)	Manual	12.8	1.32	10.30	17.3	0.77	4.46
	Morpho3D	17.5	1.78	9.81	15.4	1.13	7.34

4. Conclusion

This new technology is very promising. Despite a longer time of obtaining the final result, the 'animal handling' part is very short, limiting the risk of accidents for humans and animals, which is interesting for other productions where the handling of animals is more delicate (suckler cattle for example). The automation of the various phases of the acquisition process (cleaning, reconstruction and automatic measurement) is a major concern. The development of a new technology based on a 'one shot' may be a solution to solve the problem of moving animals. The possibility of obtaining a 3D image of the whole animal makes it possible to consider many valuations: automatic BCS, automated morphological score for animal selection, estimation of body weight, measurement of the surface and / or the volume of the animal (Le Cozler et al. 2019b).

5. Acknowledgements

The authors thank all those involved in the project, especially the technicians from the experimental station who took great care of the animals. The Morpho3D project is supported by the CASDAR National Fund, an Incentive Credit from the INRA Phase Department and the ANR DEFFILAIT Program.

References

Buranakarl, C., Indramangala, J., Koobkaew, K., Sanghuayphrai, N., Sanpote, J., Tanprasert, C., Phatrapornnant, T., Sukhumavasi, W., Nampimoon, P., 2012. Estimation of body weight and body surface area in swamp buffaloes using visual image analysis. Journal of Buffalo Science. 1, 13–20.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008. MeshLab: an open-source mesh processing tool. In: Sixth Eurographics Italian Chapter Conference, pp. 129–136.

Fischer, A., Luginbuhl, T., Delattre, L., Delouard, J.M., Faverdin, P., 2015. Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows. Journal of Dairy Science. 98, 4465–4476.

Guo, H., Ma, X., Ma, Q., Wang, K., Su, W., Zhu, D., 2017. LSSA_CAU: an interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows and pigs. Computers and Electronics in Agriculture. 138, 60–68.

Heinrichs, A.J., Hargrove, G.L., 1987. Standards of weight and height for Holstein heifers. Journal of Dairy Science. 70, 653–660.

Kmet, J., Sakowski, T., Huba, J., Peskovicova, D., Chrenek, J., Polak, P., 2000. Application of video image analysis in the slaughter value estimation of live Simmental bulls. Archiv Tierzucht Dummerstorf 43, 411–416.

Kuzuhara, Y., Kawamura, K., Yoshitoshi, R., Tamaki, T., Sugai, S., Ikegami, M., Kurokawa, Y., Obitsu, T., Okita, M., Sugino, T., Yasuda, T., 2015. A preliminarily study for predicting body weight and milk properties in lactating Holstein cows using a three dimensional camera system. Computers and Electronics in Agriculture. 111, 186–193.

Le Cozler, Y., Allain, C., Caillot, A., Delouard, J. M., Delattre, L., Luginbuhl, T., & Faverdin, P., 2019a, High-precision scanning system for complete 3D cow body shape imaging and analysis of morphological traits. Computers and electronics in agriculture, 157, 447-453.

Le Cozler, Y., Allain, C., Xavier, C., Depuille, L., Caillot, A., Delouard, J. M., ... & Faverdin, P., 2019b), Volume and surface area of Holstein dairy cows calculated from complete 3D shapes acquired using a high-precision scanning system: Interest for body weight estimation. Computers and Electronics in Agriculture, 165, 104977.

Marchant, J.A., Schofield, C.P., 1993. Extending the snake image processing algorithm for outlining pigs in scenes. Computers and Electronics in Agriculture. 8, 261–275.

Negretti, P., Bianconi, G., Bartocci, S., Terramoccia, S., Verna, M., 2008. Determination of live weight and body condition score in lactating Mediterranean buffalo by Visual Image Analysis. Livestock Science. 113, 1–7.

Pezzuolo, A., Guarino, M., Sartori, L., Marinello, F., 2018a. A feasibility study on the use of a structured light depth-camera for three-dimensional body measurements of dairy cows in free-stall barns. Sensors 18, 673. https://doi.org/10.3390/s18020673.

Schofield, C.P., Marchant, J.A., White, R.P., Brandl, N., Wilson, M., 1999. Monitoring of pig growth using prototype imaging system. Journal of Agricultural Engineering Research. 72 (3), 205–210.

Stajnko, D., Brus, M., Hocevar, M., 2008. Estimation of bull live weight through thermographically measured body dimensions. Computers and Electronics in Agriculture. 61, 223–240.

Tscharke, M., Banhazi, T.M., 2013. Review of methods to determine weight and size of livestock from images. Australian Journal of Multi-Disciplinary Engineering. 10, 1–17.

TOWARDS 5G REQUIREMENTS: PERFORMANCE EVALUATION OF A SIMULATED WSN USING SDN TECHNOLOGY

José Olimpio R. Batista Jr., Gustavo M. Mostaço, Roberto F. Silva, Graça Bressan, Carlos E. Cugnasca, Moacyr Martucci Jr.

Department of Computer Engineering and Digital Systems,

Escola Politécnica da Universidade de São Paulo (USP), Brazil

olimpio.rodrigues@usp.br, gmostaco@usp.br, roberto.fray.silva@gmail.com, graca.bressan@usp.br, carlos.cugnasca@usp.br, mmartucc@usp.br

ABSTRACT

The 5G, Fifth Generation of Mobile Networks, currently in its final development stage, promises to innovate the Internet of Things (IoT) ecosystem. It has the potential to aid in problem solving and improve the quality of existing and future services and applications. Some of the main applications include Wireless Sensor Networks (WSN), which may benefit from its very high speed and low latency in communications. Many services and applications related to WSNs are limited due to low speed and high latency connections. Some of its uses in agriculture range from fixed sensors networks for gathering weather data for irrigation control, to mobile WSNs with nodes attached to animals in the field, collecting health and productivity data, among many others. In this paper, we simulated an ad-hoc network with and without Software-Defined Networking (SDN) technology, to verify the average latency and packet delivery rate in conditions to support 5G requirements. To do so, COOJA and it-SDN were used as WSN simulators. It was observed that the use of SDN resulted in similar packet loss rate (1%) and in a considerably lower latency (at least 47%) compared to the other protocols evaluated.

Keywords*:* IoT, Wireless Sensor Networks, 5G; SDN, Autonomous Data Communication Networks

1. INTRODUCTION

The 5G technologies can be applied to many different domains, especially when high speed and low latency are essential. This will improve considerably the adoption of Internet of Things (IoT) technologies (Zanella et al., 2014). This has the potential to change many areas, such as cities, health, industry, and farm production, among many others. 5G is more than just an improvement on current telecommunication technologies Parvez et al. (2018), presenting specific improvements and implementation requirements in comparison to previous technologies.

There are several requirements for implementing 5G technologies in farming operations and processes, but most of the literature focuses mainly on theoretical aspects. Furthermore, as most domains have specific requirements, it is important, when studying the implementation of technologies in a specific domain, to understand what are the adequate requirements and which routing protocol would be the most beneficial to use.

The main objective of our research is to evaluate different routing strategies for an application of Wireless Sensor Network (WSN) on animal production. We used simulation to evaluate the Collection

Tree Protocol (CTP), the IPv6 Routing Protocol for Low-Power and Lossy Networks (RPL), and the Software-Defined Networking (SDN) strategies on a cattle weighting and water consumption scenario, for cattle in pasture, fulfilling 5G and domain requirements.

The main research question we explored is: 'can SDN improve Packet Delivery Rate (PDR) and latency on a WSN application for animal production?' It is important to mention that the same methodology used in this paper can be applied for research in other domains, considering their requirements and adapting the simulation model to better fulfill them.

2. Methodology

The methodology used in this paper has three main steps:

1. Requirements identification for implementing WSN for cattle weighing on the farm;
2. Requirements identification for implementing 5G in animal production operations;
3. Simulation of the CTP and RPL protocols and the SDN technology, using software COOJA (COntiki Os JAva), a WSN simulator that allows for the simulation of real hardware platforms (Mehmood, 2017).

Table 1 illustrates the parameters of the simulation and the versions of the software used. These simulations will allow the estimation of the latency and PDR of the WSN, with and without SDN, considering different routing strategies.

The protocols used were the following:

- CTP, which is capable of routing data, calculating and maintaining the routes for one or more sinks, focusing on building minimum cost trees. In this case, the sink acts as a root;
- RPL, which builds an acyclic graph, oriented towards a specific destiny, to route the packages.

Table 1. Parameters used in the simulations and software versions

Variable	Attribution	Variable	Attribution
Metrics	Latency and Delivery Rate	MAC Layer	CSMA
Topology	"Grade", Top Left Sink	RDC	NullRDC
TX Range/INT Range	50m/60m	CCR	128Hz
Sucess Rate TX/RX	99%/99%	Radio-Propagation Model	UDGM Distance Loss
Time of each Simulation	12 minutes	PHY Layer	IEEE 802.15.4
Speed of each Simulation	100%	COOJA	v1.8
Nodes Types	Tmote Sky; Statics	Ubuntu	v18.04
Number of Source Nodes	7 or 6 (SDN)	Qt Creator	v5.8
Number of Sink Nodes	1	Instant Contiki	v3.0
Protocols' Utilized	CTP, RPL	it-SDN (Alves et al., 2017)	v0.3

3. Requirements for 5G technologies

According to Yousaf et al., 2017, 5G networks represent the next major phase of the telecommunications industry in terms of voice and data communication. Its basic requirements for operation include the entire telecommunications infrastructure (cabling, radio base stations, etc.) and the regulation of frequencies that will be used.

The three main logical requirements that shall characterize a 5G network are based on its ability to support:

1. Enhanced Mobile BroadBand (eMBB) - a scenario in which a high transmission rate and high traffic flow capacity are the main requirements;
2. Massive Machine Type Communications (mMTC) - a scenario that is related to IoT, in which a high density of devices and high energy efficiency are key issues;
3. Ultra Reliable Low Latency Communications (uRLLC) - a scenario in which very low latency (Parvez et al., 2017) and high reliability are the main requirements.

In terms of agribusinesses, we may consider the scenarios of (eMBB, uRLLC) and (uRLLC, mMTC). These entail 5G networks to provide increased peak bit rates at Gbps per user, higher spectrum efficiency, better coverage, and support for a massively increased number of diverse connectable devices. In addition, 5G systems must be cost-efficient, reliable, flexibly deployable, elastic, agile and above all, programmable. The 5G mobile network system is thus going to be multi-tiered and slices need to be deployed and managed at each level, resulting in not only a complex architecture, but also posing big challenges in terms of 5G network, sliced infrastructure, and traffic management. In this regard some of the principal factors are:

1. Seamless and flexible management of physical and virtualized resources across the three tiers;
2. Agile end-to-end service orchestration for each respective service vertical, where each vertical may have multiple service instances;
3. Enabling end-to-end connectivity services to each service instance, which is also programmable.

Considering the above challenges, two key technologies have been developed in order to cater for the scalability, flexibility, agility, and programming requirements of 5G mobile networks: NFV (Network Function Virtualization) and SDN (Yousaf et al., 2017).

4. Software-defined networking - SDN

To provide a flexible and scalable architecture for handling network congestion and complexity in edge-cloud environments, a modern network technology, SDN, has emerged. Within an SDN environment, a single controller or group of controllers may provide data routing control plan services for a greater number of nodes, thus allowing a system-wide view of network resources. The SDN enables data to dynamically be routed on a flow-by-stream basis, using source and destination information, then adapting to possible topology changes, providing better speeds and latency, eliminating potential bottlenecks architectures (Kaur et al., 2018).

The OpenFlow, a protocol that provides an abstraction layer from the physical network to the control element, allows the network to be configured or manipulated through software that works harmonically with SDN technology. Latency is considered to be minimized with a peer-to-peer or distributed SDN controller infrastructure because it shares the communication load of the controller.

Traditional centralized cloud computing has a global view of the network but it is not suitable for applications that require low latency, real-time operation, and high-quality data for Artificial Intelligence services. Edge computing has as its main goal to extend the functions of cloud computing to the edges of the network. Because of its proximity to end users and its decentralized deployment, edge computing can support quality applications and services that present requirements such as real-time execution, low latency, and high mobility with location recognition. This makes it more suitable for applications that generally do not have enough computing and storage resources. By taking these aspects into account, this paper points to a possible new cloud and edge-processing framework, allowing real-time data analysis on IoT networks with SDN.

Communication networks are currently undergoing a major evolutionary change in order to be capable to flexibly serve the needs and requirements of massive numbers of connected users and devices to enable the functioning of the new set of envisioned applications and services in an agile and programmable way. Key terms in that context are IoT, virtualization, software, and cloud-native. In order to be able to maintain and run these networks over 5G slices, SDN technology is widely considered as one key enabler in network architecture (Yousaf et al., 2017).

5. Requirements of WSN for cattle weighing in the farm

The world demand for animal-based food products is predicted to increase by 70% by 2050. Meeting this demand in a way that has a minimal impact on the environment and that improves the quality and performance of livestock farming will require the implementation of advanced technologies.

The use of IoT technologies (ITU-T, 2012) aims at providing full coverage of the processes related to livestock production by collecting and transmitting data along the entire agroecosystem. It should be present on each participant of a livestock chain, bringing and collecting information about their processes, increasing the possibilities for control and improvement on the efficiency of their tasks.

Real-time monitoring of livestock indexes is of special interest to farmers. However, it can become difficult to be evaluated in large herds, especially at grass-fed cattle. For example, beef cattle weighing, in this raising method, is done only a few times during the life-cycle of the animals, due to the difficulty on gathering the animal, or the entire herd together, in a single point, where there is a weighing scale.

Some research developed automated camera-based systems to monitor behavioral activities, diseases, oestrus, as well as individual animal mass (Norton, Berckmans, 2018; Condotta et al., 2018). Although, this technology predominantly is applied for in-house raising methods, in which the animals are confined to a building, making it easier to obtain quality images, or even handle the animals to a weighing scale more frequently.

For that, an automatic weighing system specially designed to open field, grass-fed cattle may bring the aforementioned essential information for the farmer. Using a weighing scale placed close to a cattle drinking through will enable the collection of both the animal’s weight and identification tag on a daily basis, meeting the requirements for an automated weighing system on a smart farming environment. Additionally, it becomes easier to monitor water consumption by the herd.

6. Results

Figure 1 illustrates the topology that was used as a base for our simulations. Only the behavior of Mesh 1 was simulated.

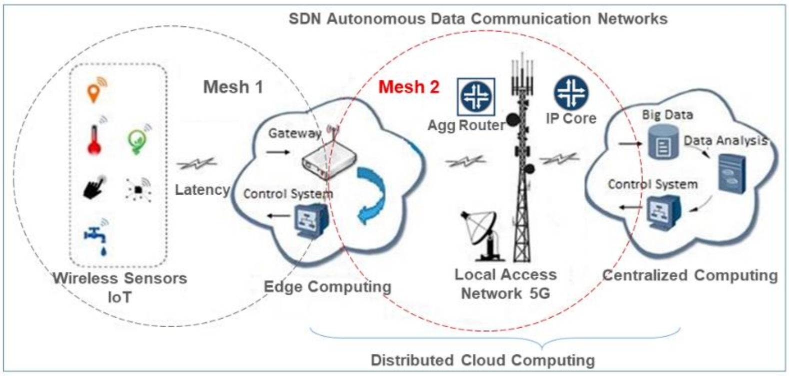


Figure 1. Basic topology of the computational structure and the autonomous network for implementing IoT on 5G. Source: Adapted from Parvez et al., 2017

The topologies used for the simulation on COOJA are illustrated in Figure 2. Figure 3 shows the results obtained under the three simulated scenarios, two of them without SDN (varying the routing protocols) and the last one with SDN.

CTP and RPL protocols, which are capable of generating and maintaining a data routing structure obtained PDRs of 98% and 99%, respectively. With the use of SDN technology, a higher PDR (99%) and a lower latency (56 ms) were obtained compared to CTP and RPL without SDN (132 and 107 ms, respectively), probably due to the centralized control and dynamic behavior of the WSN network.

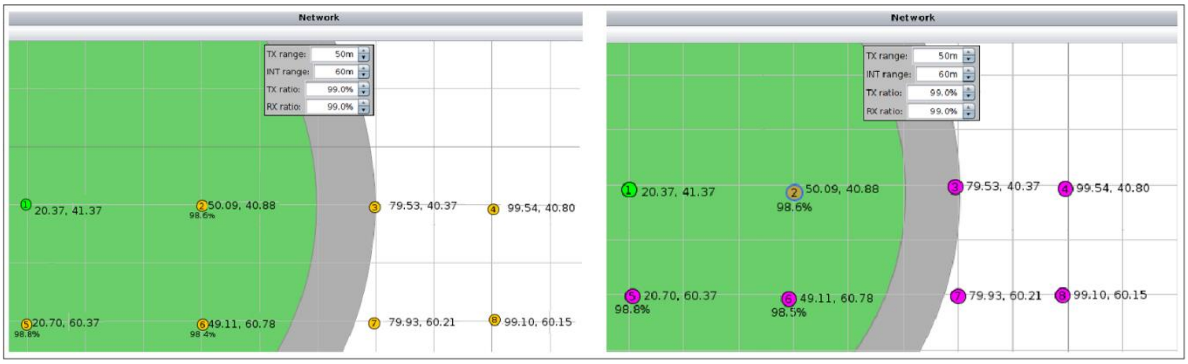


Figure 2. Simulated topologies. a) On the left, the topology without SDN, with node 1 as the sink node; b) on the right, the topology with SDN, with node 1 as the master and node 2 as the sink.

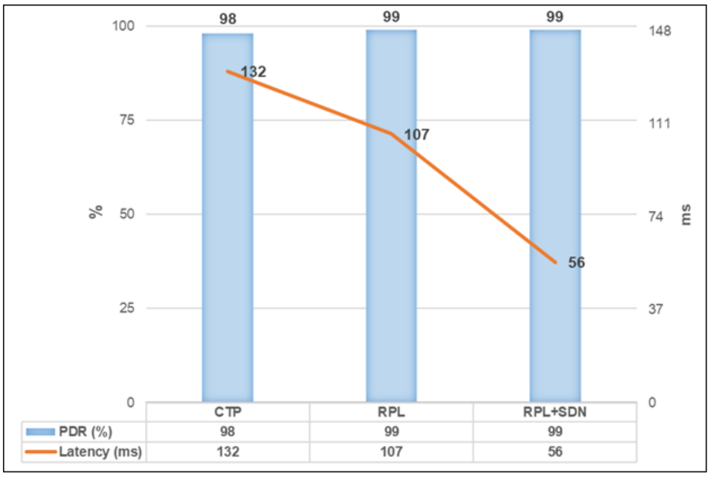


Figure 3. PDR and latency obtained from the 3 different scenarios

7. Discussion

It becomes clear the performance improvement of the RPL protocol over CTP and, especially, with the use of SDN in conjunction with RPL. For networks with mobile nodes, the connectivity should be superior with SDN. Although, we suggest the development of new simulations with mobile nodes.

Data communication networks tend to be autonomous, towards the best 5G performances, as well as the autonomous vehicles of the future, capable of being driven by themselves, without any human intervention (Sandano, 2018). Further, those networks also tend to be worked with Artificial Intelligence under advanced analysis. This requires a coherent combination of intelligence with human control and supervision, software-controlled automated operational processes, and underlying programmable infrastructure so that they can adapt, self-configure, monitor, repair and, optimize by constantly evaluating changes in its own automatic reallocation of resources.

These programmable and autonomous networks will use machine learning and optimization capabilities to adapt dynamically to the service demands and traffic patterns. In this way, software control, will enable the creation and deployment of automated services in autonomous networks, opening new research opportunities for SDN and autonomous networks.

8. Conclusions

There are currently several technical constraints in the models of data communications networks regarding computational rates, cache sizes, communication bandwidths and latencies in front of future Internet traffic volume, which is expected to more than double within the next two years. The IoT networks still suffer from the limitation and capacities of the devices for memory and internal processing, communication protocols, etc. Given the advances in nanoelectronics, there may be a rapid progression in the number of "things" connected to the Internet. On the other hand, the standardization of IoT device data will continue to be a limiting factor for transmission and analysis.

In the simulation done, we achieved a 47% reduction in latency with SDN. For lower latencies, it is necessary to orchestrate cloud computing via software control in autonomous and intelligent networks. The autonomy of the network will not only bring better latency but also other technical-commercial benefits to users and providers. The 5G network orchestration needs to be better studied and standardized due to demands related to low latency, integrality, accuracy, and precision.

Acknowledgments

This work was supported by Itaú Unibanco S.A. through the Itaú Scholarship Program, at the Centro de Ciência de Dados (C²D), Universidade de São Paulo, Brazil, and by the National Council for Scientific and Technological Development (CNPq).

References

Alves, R.C., Oliveira, D.A., Segura, G.N., Margi, C.B. (2017) 'IT-SDN: Installation Guide (for Linux 64 bits)' - March.

Condotta, I.C.F.S. et al. (2018) 'Using an artificial neural network to predict pig mass from depth images'. In: 10th International Livestock Environment Symposium (ILES X). American Society of Agricultural and Biological Engineers, pp. 1.

ITU-T (2012) 'Y.2060 - Overview of the Internet of Things. Recommendation', ITU-T. https://doi.org/10.1021/ic00245a007.

Kaur, K. et al. (2018) 'Edge Computing in the Industrial Internet of Things Environment: Software-Defined-Networks-Based Edge-Cloud Interplay', Communications Magazine, 56(2), pp. 44-51. ISSN 1558-1896.

Mehmood, T. (2017) 'COOJA Network Simulator: Exploring the Infinite Possible Ways to Compute the Performance Metrics of IOT Based Smart Devices to Understand the Working of IOT Based Compression and Routing Protocols', Dept. of Electrical Engineering, SEECS, NUST Islamabad.

Norton, T., Berckmans, D. (2018) 'Precision Livestock Farming: the Future of Livestock Welfare Monitoring and Management?'. Animal Welfare in a Changing World, pp. 130

Parvez, I. et al. (2018) 'A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions', IEEE Communications Surveys & Tutorials, 20(4), pp. 3098-3130. doi: 10.1109/COMST.2018.2841349.

Sandano, I. (2018) 'The Self-Driven Network, A trajetória rumo à automação e seus estágios´, São Paulo, Brasil.

Yousaf, F. Z. et al. (2017) 'NFV and SDN-Key technology enablers for 5G networks', IEEE Journal on Selected Areas in Communications, 35(11), pp. 2468–2478. doi: 10.1109/JSAC.2017.2760418.

Zanella et al. (2014) 'Internet of Things for Smart Cities´, IEEE Internet of Things,' 1(1), pp. 22-32. ISSN: 2327-4662.

Preliminary results of multispectral camera mounted on unmanned aerial vehicle for soil properties estimation and mapping

Theodora Angelopoulou[1,2], Nikolaos Tziolas[2,3], Athanasios Balafoutis[1], Georgios Zalidis[2,3], Dionysis Bochtis[1]

[1]Centre for Research and Technology – Hellas (CERTH), Institute for Bio-economy and Agri-technology (iBO), 57001 Thermi, Thessaloniki, Greece

[2]Laboratory of Remote Sensing, Spectroscopy and GIS, Department of Agriculture, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece

[3]Interbalkan Environment Center (i-BEC), 18 Loutron Str., 57200 Lagadas, Greece

d.angelopoulou@certh.gr, ntziolas@auth.gr, a.balafoutis@certh.gr, zalidis@agro.auth.gr, d.bochtis@certh.gr

Abstract

Soil quality continuously deteriorates due to extensive agricultural practices, risking food security. Thus, soil quality sustainability is vital to extend agricultural land productivity potential. However, soil properties estimation entails time consuming, laborious and expensive procedures. Smart agriculture schemes include novel and potentially low cost in situ observations such as unmanned aerial vehicles (UAVs) that are rapidly maturing and becoming viable alternatives to costlier traditional solutions for digital soil mapping. The objective of this research was to evaluate the capabilities of multispectral imagery (400-810 nm) predictive ability for soil properties estimation acquired in bare soil conditions in a 6-ha experimental field in Rizomilos, Thessaly, Greece. A comparative analysis was performed with laboratory spectral measurements of 18 soil samples (0–30 cm) collected from the same field covering the complete VNIR-SWIR range (400-2500 nm). The soil samples were also determined by wet chemistry methods to calibrate the developed prediction models. Considering the imagery data values, the laboratory spectral signatures and the produced spectral indices as input features, a support vector machine for regression algorithm (SVR) was used for model calibration. Laboratory soil spectroscopy resulted in R^2= 0.58 while UAV application R^2= 0.48.

Keywords*:* UAV, soil organic matter, multispectral, spectral indices**.**

1. Introduction

Soil as part of the natural environment, which is non-renewable to a great extent, is considered to be one of the most important natural resources that performs several ecological and non-ecological functions (Blum, 2005). Soil organic matted (SOM) is considered a significant soil parameter which affects soil fertility, sustainability of agricultural systems, and crop productivity. Therefore, there is an increasing concern about the SOM levels, specifically towards climate change adaptation and mitigation (Muñoz-Rojas *et al.*, 2017). The Greek soils in particular have shown low OM content ranging approximately form 1.0 to 1.5% (Tsadilas *et al.*, 2005). Consequently, there is a great interest in selecting the proper approach to achieve sustainable land use management to maintain or even increase SOM levels. However, conventional soil chemical analysis is time consuming and costly to provide the spatial variability information needed (Viscarra Rossel *et al.*, 2006). Therefore, other techniques are evaluated as alternative and cost-effective methods. Soil reflectance spectroscopy has been used in various domains from laboratory conditions to proximal and remote sensing applications. It has attracted much interest in soil science due to its advantages over conventional methods i.e. minimal to no sample preparation, simultaneous measurement of many constituents, large number of samples measured within a day and no chemicals requirement leading to a safer working environment (McBratney *et al.*, 2006).

The need to implement new technologies in agriculture for monitoring purposes that would allow site specific sustainable management practices, has led to the use of sensors mounted on UAVs for in situ observations to achieve inexpensive, rapid and high-resolution soil digital mapping. Their use has increased due to the high spatial resolution they provide, and the flexibility in the data acquisition timing (Angelopoulou *et al.*, 2019). Recent studies proposed the use of spectral indices, as derived by the data from different bands, in order to enhance the information instead of utilizing only the simple reflectance values (Gholizadeh *et al.*, 2018). To our knowledge there are limited studies for SOM estimation with the use of UAVs in real field conditions (Aldana-Jague *et al.*, 2016).

In this study we aim to compare the performance of laboratory soil spectroscopy to multispectral imagery acquired in bare soil conditions mounted on a UAV for SOM estimation. We also retrieved several spectral indices to assess their contribution in SOM estimation and to improve its prediction.

2. MATERIALS AND METHODS

2.1 Study area

The study area was a 6-ha field located in Rizomilos, Volos in central Greece (Figure 1 Study area mapFigure 1). The field measurements were performed at bare soil, seed bed level conditions to minimize the effects of various surface conditions i.e. soil roughness and the existence of plants or plant residues.

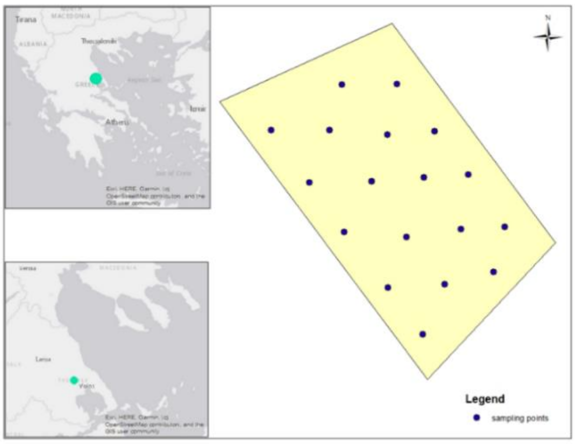


Figure 1 Study area map

2.2 Soil sampling and chemical soil analysis

In total 18 topsoil (0-30 cm) soil samples were collected (Figure 1) with a Dutch auger following the zigzag sampling pattern which is determined to obtain the best coverage on smaller areas. Soil samples were analyzed for soil texture, soil moisture and SOM content. To determine the percentage of sand, silt, and clay, the inorganic fraction of soil was measured by particle size analysis (Bouyoucos method analysis) and the total organic carbon was determined by the Walkley-Black method. The conversion to SOM was performed using the following equation:

Organic matter (%) = Total organic carbon (%) x 1.724 (Wiley, 1906)

2.3 Multispectral measurements

For this study the Parrot Sequoia+ multispectral camera (Parrot SA, Paris, France) was used. The camera has four separate sensors which cover the Green (530-570nm), Red (640-680nm), Red-edge (730-740nm) and NIR (770-810nm) regions. It was mounted on the eBee platform (https://www.sensefly.com/), and the flight altitude was 50 m. Photogrammetric processing was performed in the RAW imagery data. In this context, the PiX4D photogrammetric software (PiX4D SA, 2019) was fed with the aerial imagery data and their corresponding orientation metadata. We employed the structure from motion algorithm to find sparse point clouds, generate dense point clouds, build mesh and produce the high resolution orthomosaics of each band.

2.4 Laboratory spectral measurements

For the laboratory spectral measurements, the protocol developed by Ben Dor, Ong and Lau, (2015) was followed. The spectroradiometer used in the study was the PSR +3500 model of Spectral Evolution Company with a spectral range between 350-2500 nm. The procedure was as follows: (i) white reference measurement for instrument calibration, (ii) spectral measurement of the two internal standards for the standardization of soil spectral signatures, and (iii) soil sample measurement. The aforementioned calibration and standardization procedure was repeated for every five soil samples.

2.5 Spectral indices retrieval

Combining the spectral bands of the multispectral camera and the respective bands from the laboratory measurements, 14 spectral indices were calculated aiming at improving the prediction capability. The selected spectral indices were normalized difference vegetation index (NDVI), Green normalized difference vegetation index (GNDVI), transformed vegetation index (TVI), soil-adjusted vegetation index (SAVI), (1/green-1/red edge)*NIR (CRI), normalized difference red edge index (NDRE), NIR/red (SR), NIR/red edge (SR_E), red edge/NIR (CSM), (NIR/red edge)-1 (CI), optimized soil adjusted vegetation index (OSAVI), renormalized difference vegetation index (RDVI), weighted difference vegetation index (WDVI) and bright related index (BI2). The aforementioned indices were selected due to their sensitivity to changes in SOM content and bright related indices which are sensitive to soil texture as indicated in the literature (Levin et al., 2007; Gholizadeh *et al.*, 2018).

2.5 Regression and geostatistical analyses

An SVR algorithm with radial basis kernel was utilized for model calibration, due to its ability to generalize unseen data. Considering the low number of samples a leave one out cross validation was selected (Soriano-Disla *et al.*, 2014). Model evaluation was performed using the coefficient of determination (R^2), the root mean square error of prediction (RMSE) and the ratio of performance to interquartile distance (RPIQ). The same procedure was followed for both laboratory and UAV spectral measurements. The interpolation procedure followed for mapping SOM from laboratory spectral measurements was the inverse distance weighting (IDW).

3. Results and Discussion

3.1 Physicochemical analysis of soil samples

Soil texture was mainly characterized as clay loam to sandy clay loam since clay content ranged from 20.1 % to 38.1 % and sand content from 22.3 % to 45.6 %. The SOM content was found to range between 3.74 – 8.81 % which is considered adequate for most cultivations.

3.2 Soil organic matter prediction and mapping using spectroscopic data

The estimation of SOM using data from laboratory spectral measurements with the combined use of the spectral indices provided relatively good results with the SVR algorithm (C=32, sigma=0.1218). The statistical accuracy regarding R^2= 0.58, RMSE = 0.81 % and RPIQ = 1.18 indicated potential distinction between high and low values (Figure 2a). Compared to other studies (Nawar *et al.*, 2016), the results show lower accuracy which could be attributed to the small dataset and the small range and low variability of the SOM values.

The use of the SVR (C=16, sigma=0.084) algorithm to predict SOM from the multispectral data showed lower coefficient of determination (R^2 = 0.48) and ratio of performance to interquartile distance (RPIQ = 1.58) compared to the laboratory measurements. However, the RMSE was lower (RMSE = 0.74 %). This could lead to the inference of relatively similar results between the two approaches (Figure 2b). A related study from Aldana-Jague *et al.*, (2016) showed better accuracy, but that was a long term experiment conducted in Rothamsted, UK using data of high variability in soil organic carbon content. In addition, the study area was significantly larger. Overall, the most predominant spectral range was at red band. This band appears to be associated with SOM (Tsakiridis *et al.*, 2019). Among all spectral indices, SAVI provided the strongest correlations with SOM. This is consistent with previous studies (Gholizadeh *et al.*, 2018).

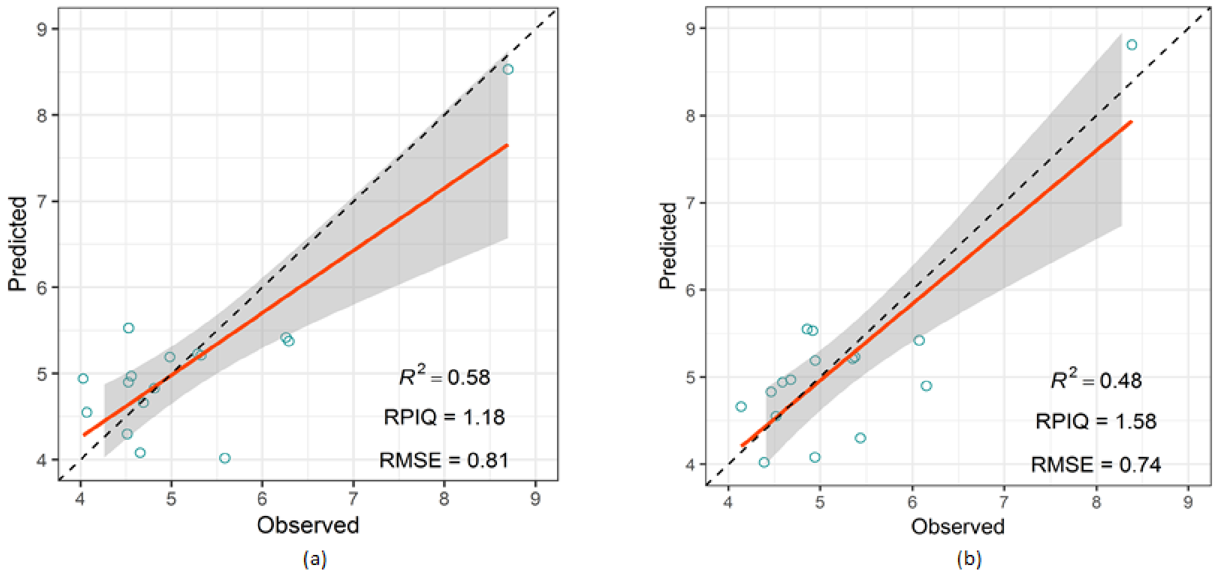


Figure 2. Scatter plots of chemically measured SOM vs. predicted values of SOM from laboratory (a) and UAS measurements (b) based on SVM regression coefficients

Figure 3 shows the generated SOM maps from laboratory (a) and UAV spectral measurements (b). The spatial distribution of SOM derived by the IDW method has slightly underestimated higher values of SOM. The UAV method overestimated the predicted values when SOM was high and underestimated when SOM content was low. This can also be seen in the scatter plot (Figure 2b). Comparing the two images both approaches gave similar results as it is shown in the upper part of the field, presenting the potential of using UAS for SOM estimation.

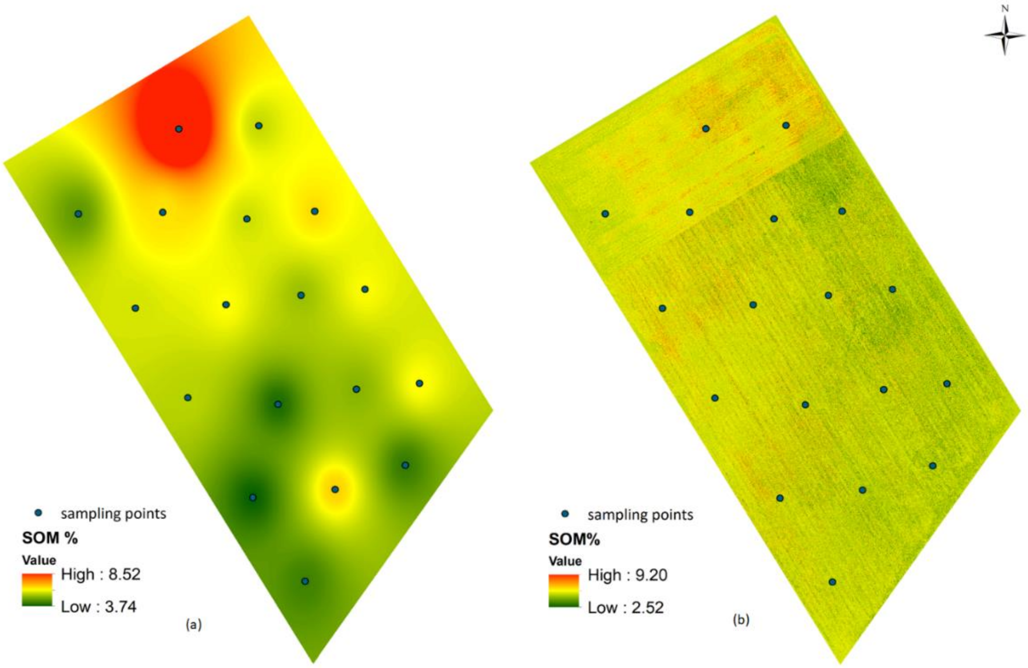


Figure 3. Maps of SOM generated from laboratory (a) and UAS measurements (b)

4. Conclusions

According to the preliminary results both laboratory and UAV spectral measurements showed adequate and similar performance for SOM estimation. Considering the spatial distribution derived from both instruments and modelling approaches the data trend was similar. The data acquisition procedure plays a very important role in the accuracy of soil property's estimation under real field conditions, and its practicality depends on appropriate radiometric calibration and other corrective tasks that are necessary in the pre-flight and post-flight steps (i.e. radiometric and geometric corrections to maintain the scientific rigor of the results). Although multispectral imaging lacks in spectral information compared to laboratory soil spectroscopy, it could provide data with higher spatial resolution and less effort. In this context, we can conclude that UAVs enable the provision of vast spatial coverage for soil parameters estimation paving the way for constant mapping and monitoring soil campaigns. To achieve better results it is imperative to perform experiments in regions with various agricultural management tehniques and different levels of OM showing. The alignment of multispectral aerial data with laboratory and proximal sensing data is the field to which science has turned its attention in order to exploit the large soil spectral libraries that have been created around the world.

Acknowledgement

The work was supported by the project "Research Synergy to address major challenges in the nexus: energy-environment-agricultural production (Food, Water, Materials)"—NEXUS, funded by the Greek Secretariat for Research and Technology (GSRT)—Pr. No. MIS 5002496.

References

Aldana-Jague, E., Heckrath, G., Macdonald, A., van Wesemael, B., Van Oost, K . (2016) 'UAS-based soil carbon mapping using VIS-NIR (480-1000 nm) multi-spectral imaging: Potential and limitations', Geoderma, 275, pp. 55–66. doi: 10.1016/j.geoderma.2016.04.012.

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D. (2019) 'Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review', Remote Sensing, 11(6), p. 676. doi: 10.3390/rs11060676.

Blum, W. E. H. (2005) 'Functions of Soil for Society and the Environment', Reviews in Environmental Science and Bio/Technology. Kluwer Academic Publishers, 4(3), pp. 75–79. doi: 10.1007/s11157-005-2236-x.

Ben Dor, E., Ong, C. and Lau, I. C. (2015) 'Reflectance measurements of soils in the laboratory: Standards and protocols', Geoderma, 245–246. doi: 10.1016/j.geoderma.2015.01.002.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L. (2018) 'Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging', Remote Sensing of Environment, 218(September), pp. 89–103. doi: 10.1016/j.rse.2018.09.015.

Levin, N., Kidron, G. J. and Ben-Dor, E. (2007) 'Surface properties of stabilizing coastal dunes: combining spectral and field analyses', Sedimentology. John Wiley & Sons, Ltd (10.1111), 54(4), pp. 771–788. doi: 10.1111/j.1365-3091.2007.00859.x.

McBratney, A. B., Minasny, B. and Viscarra Rossel, R. (2006) 'Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis', Geoderma, 136(1), pp. 272–278. doi: 10.1016/j.geoderma.2006.03.051.

Muñoz-Rojas, M., Abd-Elmabod, S.K., Zavala, L.M., De la Rosa, D., Jordán, A. (2017) 'Climate change impacts on soil organic carbon stocks of Mediterranean agricultural areas: A case study in Northern Egypt', Agriculture, Ecosystems & Environment. Elsevier, 238, pp. 142–152. doi: 10.1016/J.AGEE.2016.09.001.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M. (2016) 'Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy', Soil and Tillage Research. Elsevier, 155, pp. 510–522. doi: 10.1016/J.STILL.2015.07.021.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M.J. (2014) 'The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties', Applied Spectroscopy Reviews. Taylor & Francis, 49(2), pp. 139–186. doi: 10.1080/05704928.2013.811081.

Tsadilas, C. D., Mitsios, I. K. and Golia, E. (2005) 'Influence of Biosolids Application on Some Soil Physical Properties', Communications in Soil Science and Plant Analysis. Taylor & Francis Group , 36(4–6), pp. 709–716. doi: 10.1081/CSS-200043350.

Tsakiridis, N.L., Theocharis, J.B., Ben-Dor, E., Zalidis, G.C. (2019) 'Using interpretable fuzzy rule-based models for the estimation of soil organic carbon from VNIR/SWIR spectra and soil texture', Chemometrics and Intelligent Laboratory Systems. Elsevier, 189, pp. 39–55. doi: 10.1016/J.CHEMOLAB.2019.03.011.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O. (2006) 'Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties', Geoderma, 131(1), pp. 59–75. doi: 10.1016/j.geoderma.2005.03.007.

Wiley, H. W. (1906) Principles and practice of agricultural analysis; a manual for the study of soils, fertilizers, and agricultural products; for the use of analysists, teachers, and students of agricultural chemistry. 2d ed., re. Easton, Pa. : The Chemical publishing co. Available at: https://www.biodiversitylibrary.org/item/75931.

IN-FIELD TESTING OF NEW LOW-COST MULTISPECTRAL SENSOR FOR ASSESSING MAIZE YIELD POTENTIAL

Aristotelis C. Tagarakis[1,2], Marko Kostić[3], Natasa Ljubičić[1], Bojana Ivošević[1], Goran Kitić[1], Miloš Pandžić[1]

[1] University of Novi Sad, BioSense Institute, Serbia

[2]Institute for Bio-economy and Agri-technology - iBO, Center for Research and Technology Hellas - CERTH, Greece

[3]University of Novi Sad, Faculty of Agriculture, Serbia

a.tagarakis@certh.gr, natasa.ljubicic@biosense.rs, gkitic@biosense.rs, gkitic@biosense.rs, marko.kostic@polj.uns.ac.rs

ABSTRACT

Active proximal sensing has been increasingly used to provide information about canopy properties in a large range of crops. In this study a low cost, active multispectral optical device named Plant-O-Meter (POM) was tested in real conditions at two experimental fields comparing it with the GreenSeeker handheld device. Treatments included five nitrogen (N) fertilisation rates applied during sowing. Maize was scanned between V5 to V8 growth stages. The results showed that measuring with the POM sensor within this growth stage window can provide good estimation of end-of-season yield, comparable to the GreenSeeker. This indicates that Plant-O-Meter exhibits strong potential for accurate plant canopy measurements and for real time variable rate fertilisation applications in maize.

Keywords*:* proximal sensors, multispectral, crop sensing, yield estimation

1. INTRODUCTION

Optical sensors have become very popular for nitrogen status diagnostic. Simple handling make them suitable for a wide range of applications on different crops, and instant data can be collected following a non-destructive sampling method (Magney et al., 2016; Kostić et al, 2016). The results of numerous studies on optical sensors offer a vast number of models which could be used in the prediction of maize maturity, yield potential, plant health estimation, etc. (Zecha et al., 2018). Spectral analysis of reflected waves from plant canopy is valuable in the recognition of spectral "fingerprints", which help identify some biotic or abiotic processes that are otherwise undetectable by humans or machines. Active multispectral sensors show great potential for rapid spatial assessment of nitrogen (N) status of growing plants in early season providing adequate precondition for optimization of nitrogen management (Bean et al., 2018). In addition, such sensors are not being compromised by differences in illumination conditions, such as cloudiness, thus are ideal for on-the-go variable rate N applications (Solari et al., 2008). Therefore, active proximal sensing has been increasingly used in agriculture for assessing crop status and growth, and has proved to be promising approach for end-of-season yield estimation in a range of crops (Tagarakis and Ketterings, 2017). Active sensors emit and measure the reflectance of specific spectra of light, typically in the visible and the near-infrared, from the plant canopy providing a range of vegetation indices such as the Normalized Difference Vegetation Index (NDVI; Rouse et al., 1973). The NDVI is the most widely used index for deriving yield estimates (Hatfield et al., 2008), but other indices have also been used. It has also been related to nitrogen status, chlorophyll content, biomass, and leaf area, at micro and macro scale (Solari et al., 2008; Wang et al., 2016).

Yield estimation from mid-season spectral canopy measurements is of particular importance since it is the first step in the development of an algorithm for real-time variable rate N applications (Moges et al., 2007). The timing of sensing, in terms of growth stage, greatly impacts the accuracy of yield predictions from sensor data (Raun et al., 2005). Previous studies in maize defined V7 – V8 (7 – 8 fully developed leaves) as the growth stages that provide the highest accuracy of end-of-season yield estimation and V6 as the stage with the highest variability in NDVI measurements (Tagarakis and Ketterings, 2017; Raun et al., 2005), important elements for maximizing the benefit of variable rate fertilisation.

Raun et al. (2002) introduced the In Season Estimated Yield (INSEY) (Eq 1,2) as an approach that normalizes NDVI measurements across time and various environmental conditions (Teal et al., 2006), accounting for the growing conditions from planting to sensing and providing an estimate of the N uptake per day (Lukina et al., 2001) and the biomass produced per day (Raun et al., 2005).

$$INSEY_{DAP} = NDVI / DAP \quad (1)$$

Where: $INSEY_{DAP}$ is the In Season Estimated Yield, NDVI is the Normalized Vegetation Index, DAP is the number of Days After Planting for days with GDD > 0, and GDD is defined as the growing degree days.

$$INSEY_{GDD} = NDVI / GDD \quad (2)$$

Where: $INSEY_{GDD}$ is the In Season Estimated Yield, NDVI is the Normalized Vegetation Index, and GDD is defined as the cumulative growing degree days from planting to sensing.

Teal et al. (2006) developed models to predict maize grain yield based on NDVI, $INSEY_{GDD}$, and $INSEY_{DAP}$ showing similarly good results (R^2 ranged from 0.73–0.77). In recent study conducted by Rogers (2016) good correlation was achieved between combined optical sensor readings, collected during two growing seasons at V6 to V8 growth stage, and final yield (R^2>0.68). Similarly good results were reported by Tagarakis and Ketterings (2017) who defined V7 as the growth stage that provides the most accurate estimation of end-of-season yield for grain maize in New York (R^2 = 0.78).

Most active proximal sensors vary in central wavelengths or bandwidths for calculating NDVI (Kim et al., 2010; Yao et al., 2013). Therefore, this study has been conducted to evaluate the performance of the recently developed active multispectral proximal sensor Plant-O-Meter in real field conditions, and compare it with the GreenSeeker handheld which is a widely accepted sensor. The main objectives of this study were to: (1) define the relationship between the NDVI measurements derived from the two sensors, (2) determine the specific growth stage at which the sensors provide more reliable end-of-season yield estimation under the specific climatic conditions in Vojvodina region, and (3) define the ability of Plant-O-Meter to estimate end-of-season yield from mid-season canopy measurements as compared with the GreenSeeker handheld sensor device.

2. MATERIALS AND METHODS

Field trials and experimental design

The present study was carried out at two experimental fields in Bajmok and Ravno Selo, located respectively in northern and central Vojvodina region in Serbia. Three maize hybrids (Zea mays L.) of different maturity classes and length of vegetation period were sown, namely P9537 (FAO 340), P9911 (FAO 450) and P0412 (FAO 530). The fields were sown at 70 cm inter-row spacing and 20 cm spacing between plants in the row. The study included five different N treatments (0, 50, 100, 150 and 200 kg N ha^{-1}) applied pre-plant by incorporating granular urea (46% N). The experiment was conducted following randomized complete block design (RCBD) with three replications. Each plot contained four rows of maize; the central part, 6m long, of the two middle rows was measured while the side rows served as guard rows.

Sensor measurements and sensor description

Two active proximal sensors were used to measure NDVI at V5, V6 and V8 growth stages of maize; the GreenSeeker handheld (Trimble Inc., CA, USA) and a recently developed active multispectral optical sensor named Plant-O-Meter (POM) (BioSense Institute, Serbia).

GreenSeeker is an active hand-held sensor, which emits light and measures the reflectance at 660 nm (R) and 770 nm (NIR) calculating the NDVI (Tremblay et al., 2009). In-field reflectance measurements were taken by holding the GreenSeeker sensor approximately 60 cm above the crop canopy, with the sensing footprint perpendicular to the row direction, manually recording four average measurements from the measuring area in each plot.

Plant-O-Meter senses at four bands (Blue, Green, Red and Infrared) and provides the reflectance separately for each band providing the ability to calculate more than 20 different indices. It connects to any Android device and uses its processing and storing capacity for data logging and processing. In addition, it uses the device’s GNSS antenna to georeference the measurements. In-field reflectance measurements were taken by holding the Plant-O-Meter approximately 60 cm above the crop canopy, with the sensing footprint perpendicular to the row direction and scanning the whole length of the two middle rows in continuous mode at frequency of 1Hz. The central part, (6 m) of each measured row (2 middle rows per plot), was selected after processing the data using GIS software (QGIS Development Team, 2018). The NDVI measurements with both instruments were made close to noon, between 11:00 a.m. and 1:00 p.m.

Harvest

At the stage of full maturity, the plants from the central part, 6 m long, of the two middle rows in each plot were hand-harvested by manually picking all developed ears and collecting in pre-labelled bags. The gross weight of each plot was measured and the content of each bag was shelled to calculate the net grain weight. A GAC® 2500-INTL Grain Analysis Computer (Dickey-John, IL, U.S.A.) was used to measure grain moisture content (MC) and the final yield was normalized at 14% MC.

Data analysis

In order to analyze the sensors’ data, the measurements were transformed to $INSEY_{DAP}$ dividing the NDVI by the days after planting as recommended by previous studies (Tagarakis and Ketterings, 2017). Using INSEY adjusted the sensor measurements to the specific growing conditions of each field from planting until sensing. Regression analysis was used to define the relationship between the GreenSekeer and Plant-O-Meter NDVI measurements. In addition, linear regression models were used for the relationships between the INSEY and end-of-season yield for each growth stage.

3. RESULTS AND DISCUSSION

Regression analysis between INSEY and end-of-season yield showed that for both sensors, good yield estimation can be achieved when scanning at V6 – V8 growth stages. This finding is in agreement with the results by Tagarakis and Keterings (2017) who defined V6 as the earliest growth stage for accurate yield estimations. The most accurate estimation of yield for both sensors was achieved at V8 growth stage consistent with the findings of previous studies (Teal et al., 2006). The coefficient of determination (R^2) was 0.8 and 0.75 for Plant-O-Meter and GreenSeeker respectively (Figure 1). In general, Plant-O-Meter seemed to provide better estimation of end-of-season yield for all three maize growth stages. However, this needs to be further investigated in various environmental and climatic conditions and at more detailed temporal resolution.

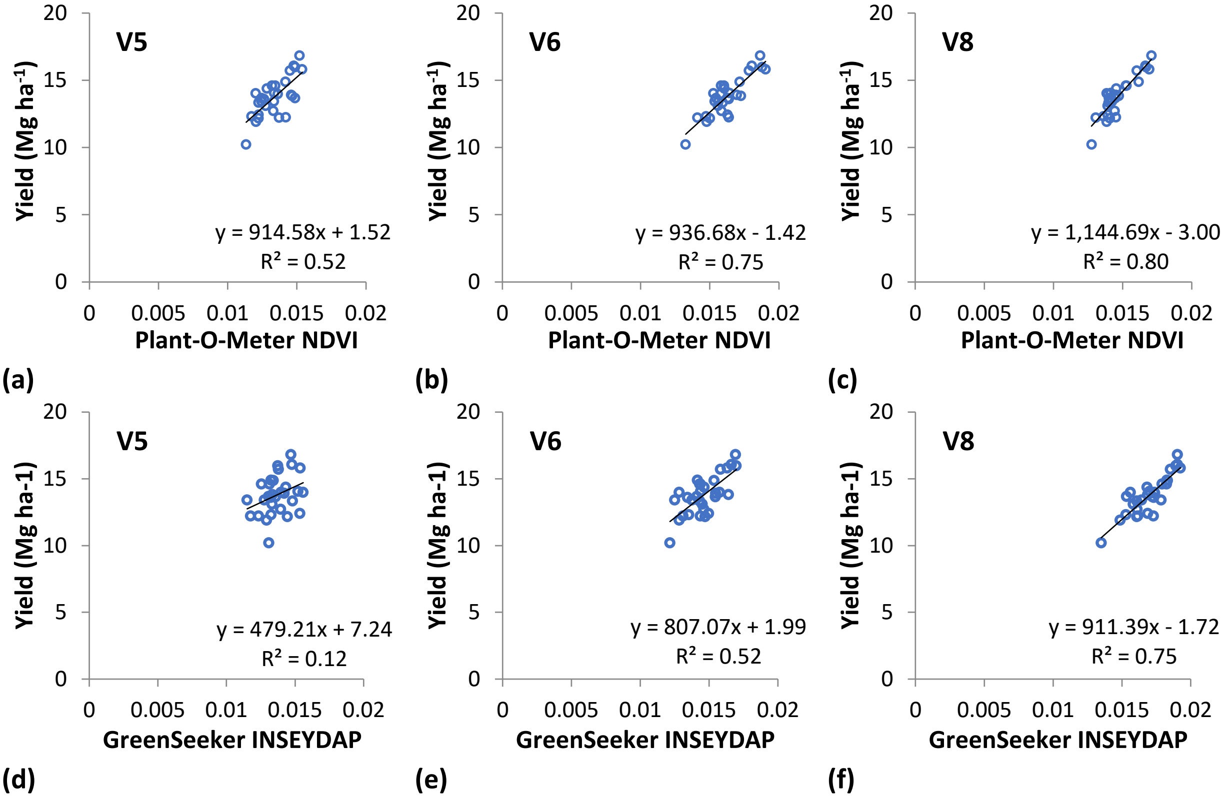


Figure 1. Relationship between end-of-season yield and In Season Estimated Yield (INSEYDAP) measured using the Plant-O-Meter at V5 (a), V6 (b) and V8 (c) maize growth stages and using the GreenSeeker at V5 (d), V6 (e) and V8 (f) maize growth stages.

Regression analysis was used for the comparison between the Plant-O-Meter and GreenSeeker sensors. Despite the fact that both Plant-O-Meter and GreenSeeker sensors measure canopy reflectance at different wavelengths, the results confirmed a strong relationship between the NDVI measurements from the two sensors (R^2 = 0.89), on overall basis (Figure 2). This result suggests that the Plant-O-Meter sensor shows good potential to be used for on-the-go variable rate applications as it performs similarly to the GreenSeeker in real field conditions.

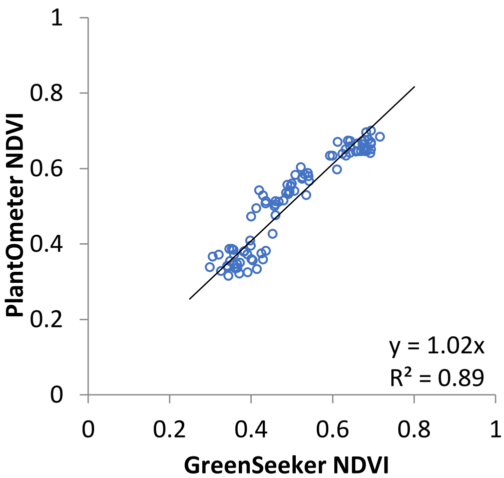


Figure 2. Relationship between NDVI measured using the Plant-O-Meter (y axis) and NDVI measured using the GreenSeeker (x axis)

This was the first trial for in-field testing of the Plant-O-Meter multispectral active proximal sensor. The results revealed its potential to estimate end-of-season yield form mid-season canopy measurements which is the first step in the development of an algorithm for variable rate nitrogen applications (Moges et al., 2007).

5. CONCLUSIONS

Based on the present findings reliable end-of-season yield estimation is attained measuring the mid-season NDVI between V7 and V8 stage. The overall results indicated that NDVI obtained using Green-Seeker was very similar to the NDVI measured by the Plant-O-Meter. In addition, both sensors provided good estimation of end-of-season yield at V8 growth stage of maize crop. The Plant-O-Meter provided slightly better estimation of end-of-season yield especially for the measurements performed earlier in the season (V5 and V6 stages). This result indicates that Plant-O-Meter exhibits good potential for accurate plant canopy measurements and for real time variable rate fertilisation applications in maize. Considering that the optimal stages for measuring NDVI depend on the environmental conditions, further studies in more diverse conditions are needed to test and evaluate Plant-O-Meter's performance. The final equation for end-of-season yield estimation using mid-season crop canopy measurements need to be finalized completing the first step for the development of an algorithm for real time VRN application. Furthermore, the low cost and the ease of use of the Plant-O-Meter sensor are expected to make it a reliable and affordable solution for small and medium size farmers in order to apply precision agriculture.

ACKNOWLEDGEMENTS

This study was supported by the project “Development of the device for measurement and mapping of nitrogen as the most important parameter in sustainable agriculture”, contract no. 114-451-2794/2016-03, funded by Provincial Secretariat for Higher Education and Scientific Research of Autonomous Province of Vojvodina, Republic of Serbia and by the project “Improvement of the quality of tractors and mobile systems with the aim of increasing competitiveness and preserving soil and environment”, contract no. TR-31046, funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

Bean, G. M., Kitchen, N. R., Camberato, J. J., Ferguson, R. B., Fernandez, F. G., Franzen, D. W., Laboski, C. A. M., Nafziger, E. D., Sawyer, J. E., Scharf, P. C., Schepers, J. and Shanahan, J. S. (2018) ‘Active-optical reflectance sensing corn algorithms evaluated over the United States Midwest corn belt’, Agronomy Journal, 110, pp. 2552–2565.

Hatfield, J.L., Gitelson, A.A., Schepers, J.S. and Walthall, C.L.(2008) ‘Application of spectral remote sensing for agronomic decisions’, Agronomy Journal, 100, pp. 117–131. doi: 10.2134/agronj2006.0370c.

Kim, Y., Huete, A., Miura, T. and Jiang, Z. (2010) ‘Spectral compatibility of vegetation indices across sensors: band decomposition analysis with Hyperion data’, Journal of Applied Remote Sensing 4(1) 043520. doi: 10.1117/1.3400635.

Kostić, M., Rakić, D., Savin, L., Dedović, N. and Simikić, M. (2016) ‘Application of an original soil tillage resistance sensor in spatial prediction of selected soil properties’, Computers and Electronics in Agriculture, 127, pp. 615-624. doi: 10.1016/j.compag.2016.07.027.

Lukina, E. V., Freeman, K. W., Wynn, K. J., Thomason, W. E., Mullen, R. W., Stone, M. L., Solie, J. B., Klatt, A. R., Johnson, G. V., Elliott, R. L. and Raun, W. R. (2001) ‘Nitrogen fertiliyation optimization algotithm based on in-season estimates of yield and plant nitrogen uptake', Journal of Plant Nutrition, 24(6), pp. 885-898. doi: 10.1081/PLN-100103780.

Magney, S. T., Eitel J. U. H., Huggins D. R., Vierling L. A. (2016) 'Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality', Agricultural and Forest Meteorology, 217, pp. 46–60. doi: 10.1016/j.agrformet.2015.11.009.

Moges, S. M., Girma, K., Teal, R. K., Freeman, K. W., Zhang, H. and Arnall, D. B. (2007) 'In-season estimation of grain sorghum yield potential using a hand-held optical sensor', Arch. of Agron. and Soil Sci., 53(6), pp. 617–628. doi: 10.1080/03650340701597251.

Raun, W.R., Solie, J.B., Johnson, G.V., Stone, M.L., Mullen, R.W. and Freeman, K.W. (2002) 'Improving nitrogen use efficiency in cereal grain production with optical sensing and variable rate application', Agron. J., 94, pp. 815–820. doi: 10.2134/agronj2002.8150.

Raun, W. R., Solie, J.B., Martin, K.L., Freeman, K.W., Stone, M. L., Johnson, G.V. and Mullen, R.W. (2005) 'Growth stage, development, and spatial variability in corn evaluated using optical sensor readings', J. Plant Nutr., 28, pp. 173-182. doi: 10.1081/PLN-200042277.

Rogers, N. G. (2016) 'Sensor Based Nitrogen Management for Corn Production in Coastal Plain Soils', All Theses. 2579.

Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. (1973) 'Monitoring vegetation systems in the Great Plains with ERTS', NASA. Goddard Space Flight Center 3d ERTS-1 Symp., 1, pp. 309–317.

Solari, F., Shanahan, J., Ferguson, R. B., Schepers, J. S. and Gitelson, A. A. (2008) 'Active sensor reflectance measurements to corn nitrogen status and yield potential', Agronomy Journal, 100, pp. 571–579. doi: 10.2134/agronj2007.0244.

Tagarakis A. C. and Ketterings Q. M. (2017) 'In-season estimation of corn yield potential using proximal sensing', Agronomy Journal, 109(4), pp. 1323-1330. doi: 10.2134/agronj2016.12.0732.

Teal, R. K., Tubana, B., Girma, K., Freeman, K. W., Arnall, D. B., Walsh, O. and Raun, W. R. (2006) In-season prediction of corn grain yield potential using normalized difference vegetation index', Agron. J., 98, pp. 1488–1494. doi: 10.2134/agronj2006.0103

Tremblay N., Wang Z., Ma, B. L., Belec, C. and Vigneault, P. (2009) 'A comparison of crop data measured by two commercial sensors for variable-rate nitrogen application', Precision Agriculture, 10, pp. 145-161. doi: 10.1007/s11119-008-9080-2

Wang, R., Cherkauer, K. A. and Bowling, L. C. (2016) 'Corn Response to Climate Stress Detected with Satellite-Based NDVI Time Series', Remote Sensing, 8(4), pp. 269. doi: 10.3390/rs8040269.

Yao, X., Yao, X., Jia, W., Tian, Y., Ni, J., Cao, W. and Zhu, Y. (2013) 'Comparison and intercalibration of vegetation indices from different sensors for monitoring above-ground plant nitrogen uptake in winter wheat', Sensors, 13(3), pp. 3109-3130. doi: 10.3390/s130303109.

Zecha, C. W., Peteinatos, G. G., Link, J. and Claupein, W. (2018) 'Utilisation of ground and airborne optical sensors for nitrogen level identification and yield prediction in wheat', Agriculture, 8(6) pp. 79. doi: 0.3390/agriculture8060079.

OPTIMA - OPTIMISED INTEGRATED PEST MANAGEMENT FOR PRECISE DETECTION AND CONTROL OF PLANT DISEASES IN PERENNIAL CROPS AND OPEN-FIELD VEGETABLES

Athanasios Balafoutis[1], Nikos Mylonas[2], Spyros Fountas[2], Dimitris Tsitsigiannis[2], Paolo Balsari[3], Massimo Pugliese[3], Emilio Gil[4], David Nuyttens[5], Gerrit Polder[6], Fausto Freire[7], Jose Paulo Sousa[7], Mathilde Briande[8], Valerie Le Clerc[8], Jean-Paul Douzals[9], Amedeo Caffini[10], Lars Berger[11], Zisis Tsiropoulos[12], Daniele Eberle[13], Sarah Bellalou[14], Andreas Thierfelder[15]

[1]Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Volos, Greece

[2]Agricultural University of Athens, Athens, Greece

[3]University of Torino, Torino, Italy

[4]Universidad Politecnica de Cataluna, Barcelona, Spain

[5]Instituut voor Landbouw-, Visserij- en Voedingsonderzoek, Merelbeke, Belgium

[6]Wageningen University & Research, Wageningen, the Netherlands

[7]Universidade de Coimbra, Coimbra, Portugal

[8]Institute for life, food and horticultural sciences and landscaping – AGROCAMPUS OUEST, Rennes, France

[9]National Research Institute of Science and Technology for Environment and Agriculture - IRSTEA, France

[10]CAFFINI, Verona, Italy

[11]Pulverizadores FEDE, Valencia, Spain

[12]AGENSO, Athens, Greece

[13]Terre da Vino, Barolo, Italy

[14]INVENIO, France

[15]European Crop Protection Association - ECPA, Brussels, Belgium

a.balafoutis@certh.gr, nmylonas@aua.gr, paolo.balsari@unito.it, david.nuyttens@ilvo.vlaanderen.be, emilio.gil@upc.edu, gerrit.polder@wur.nl, fausto.freire@del.uc.pt, valerie.leclerc@agrocampus-ouest.fr, jean-paul.douzals@irstea.fr, amedeo.caffini@caffini.com, lberger@fedepulverizadores.com, tsiropoulos@agenso.gr, eberle@terredavino.it, s.bellalou@invenio-fl.fr, andreas.thierfelder@ecpa.eu

ABSTRACT

OPTIMA is an H2020 research project that will develop an environmentally friendly IPM framework for vineyards, apple orchards and carrots by providing a holistic integrated approach which includes all critical aspects related to integrated disease management, such as i) use of novel biological Plant Protection Products, ii) disease prediction models, iii) spectral early disease detection systems and iv) precision spraying techniques. It will contribute significantly to the reduction of the European agriculture reliance on chemical Plant Protection Products resulting in reduced use of agrochemicals, lower residues and reduced impacts on human health.

Keywords: IPM, biological plant protection products, disease prediction models, spectral disease detection, precision spraying

1. INTRODUCTION

As the global population approaches 9 billion by 2050 (UN, 2015), the UN Food and Agriculture Organization (FAO) expects that demand for agricultural outputs will increase by 60% compared with the annual average from 2005 through 2007, representing an increase of approximately 1% per year. Furthermore, the total crop yield reduction caused by all crop pest species (estimated to be around 67,000 —including plant pathogens, weeds, invertebrates and some vertebrate species) reaches 40% (Oerke et al., 1994). Therefore, global food security is undermined by pests alongside other constraints, such as inclement weather, poor soils and farmers' limited access to technical knowledge (Chandler et al., 2011). Ways to make crop protection more sustainable are required and Integrated Pest Management (IPM) is promoted as the best way forward, shown also by its central position in the 2009/128/EC Sustainable Use Directive on Plant Protection Products (PPPs) (EU, 2009).

IPM could make a difference in this effort, as it emphasizes on the growth of a healthy crop with the least possible disruption to agro-ecosystems, and encourages natural mechanisms for pest management (Lamichhane et al., 2016). IPM is a systems approach that combines different crop protection practices and subsequent integration of appropriate measures that discourage the development of populations of harmful organisms and keep the use of plant protection products and other forms of intervention to levels that are economically and ecologically justified and reduce or minimize risks to human health and the environment (Flint and van den Bosch, 1981).

Synthetic PPPs have been common practice in industrialised countries for pest management since the Green Revolution and together with optimised crop varieties, intensive mechanization, irrigation and crop nutrition through rigorous fertilization, they increased agricultural yields significantly (70% in Europe and 100% in the USA (Lamichhane et al., 2016)). Ideally, PPPs are used to exterminate the targeted pests. Unfortunately, the application of PPPs worldwide is being executed with limited consideration to the dosage rate, optimum number of applications, timing and frequency resulting in rampant use of these agrochemicals, under the axiom: "if little is good, much more will be better" (Wasim Aktar et al., 2009), having as result the contamination of natural resources by PPPs, including soil, water, turf and all vegetation types. On the contrary, Bio-PPPs are a particular group of crop protection tools that can substitute or support synthetic PPPs and are ideal for IPM schemes. Bio-PPPs are defined as a mass-produced agent manufactured from a living microorganism or a natural product and sold for the control of plant pests (EU, 2009). That said, Bio-PPPs show attractive properties, such as their selectivity, their little or no toxic residue production, and the significantly lower development costs in comparison to synthetic PPPs (Hajek, 2004).

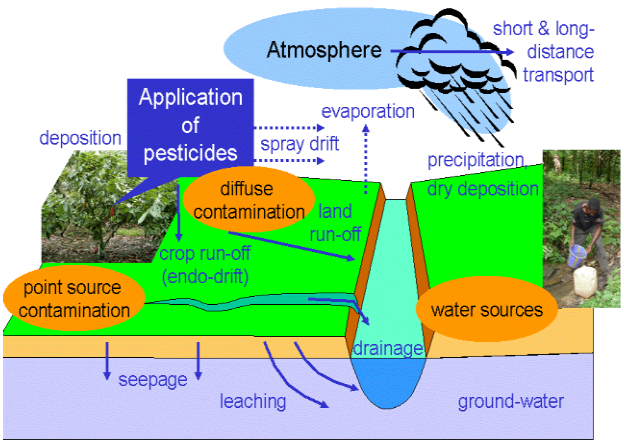


Figure 1: Routes of environmental contamination from PPPs. Source: Wikipedia

PPPs are mainly applied, especially in conventional agriculture, using hydraulic and hydro-pneumatic sprayers. The principle of operation is to convert a PPP formulation that in most cases is diluted in water (or another liquid carrier) into droplets that will be sprayed upon the canopy of the selected crop to spread the chemical compound. Unfortunately, dose-transfer to the biological target (i.e. the pest) through PPP spraying has high inefficiencies and significant amounts of the active ingredient end up elsewhere in the environment (Graham-Bryce, 1977) contaminating natural resources (water, soil, air). There are numerous routes of environmental contamination from PPPs (Figure 1). Contamination can be either point source which is mainly related to the handling of PPPs on a farm during cleaning, filling, remnant liquid management, transport and storage (TOPPS, 2008a) or diffuse source which is mainly related to run-off from field after application, discharge from drainage and off target deposition of spray due to wind (spray drift) (TOPPS, 2008b). Therefore, diffuse sources can be mainly reduced through optimization of spraying technology. Novel spraying equipment targeting to a precise application spraying process uses only the approved and advised amount of PPPs reducing therewith both wastes – such as remnants in the sprayer - and risks, yet ensuring optimised biological efficacy with the greatest input cost effectiveness.

The overall objective of OPTIMA is to develop an environmentally friendly Integrated Pest Management (IPM) framework for use-cases in orchards, vineyards and open-field vegetables by providing a holistic approach which includes the major elements related to integrated disease management: (i) combined use of bio-PPPs and synthetic PPPs, (ii) DSS for disease prediction, (iii) spectral disease detection systems and (iv) precision spraying techniques.

OPTIMA advanced IPM framework will consist of 4 main pillars (Prediction, Detection, Selection and Application), and will focus on plant diseases that annually damage high-value crops (fruits and vegetables) and demand high amounts of fungicides to be applied in numerous spraying applications. OPTIMA will work on apple scab, grape downy mildew and Alternaria leaf blight of carrot, based on the importance of the aforementioned diseases and crops for the European agriculture.

To accomplish OPTIMA's vision, the specific project objectives are:

- Objective 1: Optimize plant disease prediction models and develop advanced early disease detection methods
- Objective 2: Evaluate and screen biological and synthetic PPPs and assess plant and pathogen resistance mechanisms for successful disease control
- Objective 3: Enhance and develop innovative precision spraying technologies
- Objective 4: Test and evaluate the proposed new IPM elements under field conditions
- Objective 5: Assess health, environmental and socioeconomic impacts and risks of the proposed IPM system

2. Methodology

To achieve all specific objectives of OPTIMA and deliver a fully functional IPM system, eight Work Packages (WP) will undertake the required actions, as shown in Figure 2:

1. **WP1** focuses on assessing end-users' requirements following the co-creation approach, while participating in the design and evaluation of the developed products, tools and strategies throughout the project life-time. End-users will also evaluate the features of the demonstrated IPM system in the pilot areas.
2. **WP2** focuses on the development of a DSS for the prediction of grape downy mildew, apple scab and Alternaria leaf blight of carrots based on agro-climatic, pathogen biological cycle algorithms and users' testimonies and on the development of a portable advanced detection system for in-field localization and monitoring of the selected diseases using state of the art machine learning;
3. **WP3** deals with the development of novel sustainable IPM strategies, to be applied as preventive or therapeutic control methods for the selected plant diseases both at experimental and commercial scale. Several biological and chemical intervention strategies will be tested experimentally to expand the range of available tools and the best performing options will be optimised for the optimum application strategy to be pilot testing. The focused diseases will be

Alternaria leaf blight in carrots, grape downy mildew in vineyards and apple scab in apple orchards.

4. **WP4** brings together technological advancement to primarily define optimal spray configuration and parameters for the different crop-disease combinations and reduce spray drift while applying PPPs. These findings together with different variable rate application technologies will be used to develop three smart sprayers for each disease/crop combination.
5. **WP5** is the phase of OPTIMA project that will receive all the components of the IPM system and evaluate them separately and as a whole. The DSS and the disease detection system will be tested in three selected pilot areas, while the efficacy of selected synthetic and bio-PPPs will be evaluated as well. The smart sprayers for each crop will be tested to quantify the improvements on deposition, coverage and drift reduction in comparison to conventional sprayers, at commercial fields. Finally, all components of the IPM system as a whole will be tested in-field and compared with conventional crop protection methods.
6. **WP6** focuses on performing an extended Life-Cycle Assessment (LCA) combined with Human and Environmental Risk Assessment (HERA) using a Multi-Criteria Decision Analysis (MCDA) to identify and quantify the human health, environmental and socio-economic impacts and risks of treatments of the selected crops, comparing the conventional with the proposed IPM crop protection system
7. **WP7** covers the dissemination, communication and exploitation strategy for achieving high visibility of the OPTIMA project results and enable fast exploitation of the developed products.
8. Finally, **WP8** is the project management aiming to achieve smooth evolution of the project regarding technical coordination and monitoring, efficient communication, financial and administrative issues, and appropriate data management.

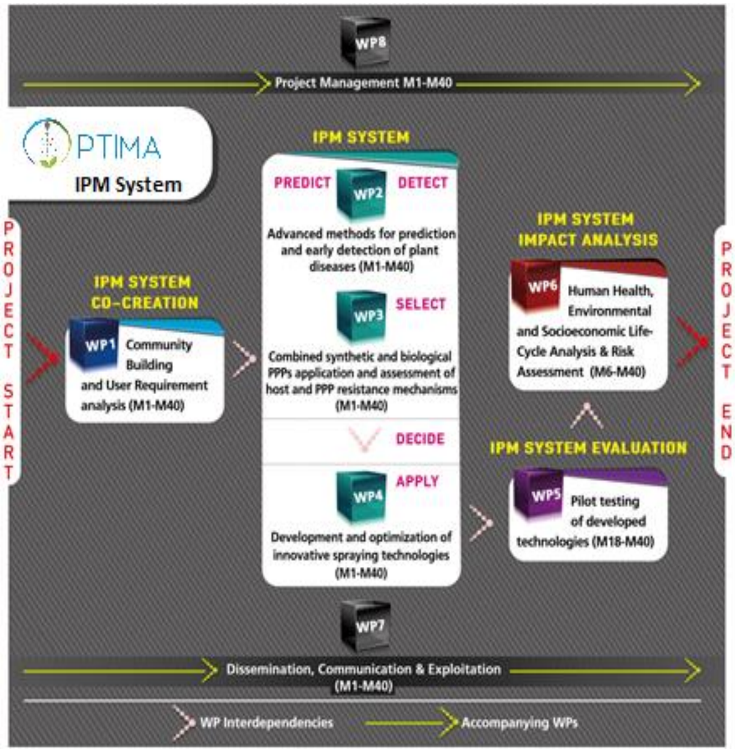


Figure 2: Overall structure of the OPTIMA project

3. Expected Results

OPTIMA will:

1. **optimise already existing disease prediction models** using agro-climatic and biological algorithms; users' testimonies; and geostatistical methods, which will be used for the development of a Decision Support System (DSS).
2. **develop advanced spectral detection systems** for in-field localisation and monitoring of the selected diseases in the use-case crops. OPTIMA's detection system will be used as standalone system providing information at site-specific level that can be used to export the variable rate spraying prescription maps; or as embedded to the three developed in the course of the project smart sprayers for the three crop types, where the image detection results will be transferred directly to the control unit of the sprayer in real-time to adjust the spraying quantity on site.
3. **evaluate a collection of 10 different bio-PPP agents** per host of commercial (market-ready) and under development bio-PPPs (from partner's culture collections) together with selected new generation synthetic fungicides against apple scab, downy mildew in vineyards and Alternaria leaf blight in carrots to identify the most efficient combinations to efficiently control the diseases.
4. **develop three smart prototype sprayers** for carrots, vineyards and apple orchards actuating different nozzle types, sprayer settings and adopting variable rate application control based on optimal selection of spray parameters, canopy and disease characteristics, together with the integration of innovative drift reducing technologies in order to minimize losses to the environment.
5. **follow the co-creation process** to take advantage of the end-users' experience in its functions. After the technical validation and evaluation of the proposed IPM system and according to the numeric assessment of the benefits that will be obtained from the system evaluation in terms of farm economics, environmental and human health protection and social aspects, the pragmatic depiction of the impacts of such system will become clear.
6. **pilot test the OPTIMA developed IPM approach** in three selected use case areas. Trials of the OPTIMA IPM system and its individual components (DSS, disease detection system, combinations of PPPs and smart sprayers) will be carried out to evaluate the improvements and benefits when compared with the normal practices applied in the specific areas. Field experiments will be conducted in close collaboration with the three farmers' cooperatives that participate in the project and will follow the field experiments on a daily basis. Primarily, the developed standalone early disease detection system will be evaluated and optimised in field assessment by correlation with conventional field scouting techniques. At the same time, a data collection procedure will be arranged to quantify the savings on PPPs, water and time. The prediction model will be evaluated in two growing seasons (year 1 and 2) and tested with the other IPM system components in year 3.
7. **field-evaluate the selected synthetic and bio-PPPs combination** using the developed three smart sprayers for each crop. Depending on the crop characteristics, and following the EPPO guidelines for every single crop, a complete evaluation strategy of disease control will be conducted. Good spraying practices and the most accurate working parameters developed by OPTIMA will be demonstrated and compared with conventional practices. Benefits in terms of less PPP amount and less contamination will be measured and explained to the end-users for the purposes of system evaluation.
8. **test and evaluate the three smart sprayers** developed in field conditions to define their ability to generate a safe, efficient and improved spray application process. Over the established experimental plots, field trials will be conducted in order to evaluate the spray quality distribution (deposition on leaves, coverage, penetration, etc.), the risk of drift, the potential reduction of PPP loses, and the capability of the three developed sprayers to adapt the spray distribution to the canopy characteristics. During the field trials a complete canopy characterization using different electronic sensors (ultrasonic, LiDAR, etc.) will be performed to collect the necessary information about the intended target. Once the canopy will be characterized, the smart sprayers will be ready to arrange the Variable Rate Application process

that will be executed either based on previously generated canopy and risk maps, or on the on-going measurements provided by the on-board sensors.

9. **Evaluate and quantify the disease incidence severity** of the three selected diseases at the end of the experimental session
10. **Conduct an extended Life-Cycle Assessment** (LCA) of the complete OPTIMA IPM framework, and a full Human and Environmental Risk Assessment (HERA), together with Social LCA and the economic viability to address life-cycle costs together with scenario and sensitivity analysis.
11. **Conduct a Multi-Criteria Decision Analysis** (MCDA) to fully assess the proposed crop protection systems for apple scab, grape downy mildew and Alternaria leaf blight in carrots based on the complementary use of LCA and HERA.

REFERENCES

UN. (2015) http://www.un.org/en/development/desa/news/ population /2015-report.html, accessed on April 25th, 2019

Oerke E.C., Dehne H-W., Schonbeck F., Weber A. (1994). Crop production and crop protection: estimated losses in major food and cash crops. Elsevier doi: 10.1016/C2009-0-00683-7

Chandler, D., Bailey A.S., Tatchell G.M., Davidson G., Greaves J., Grant W.P. (2011). The development, regulation and use of biopesticides for integrated pest management, Phil. Trans. R. Soc. B, 366 (1573):1987-1998. doi: 10.1098/rstb.2010.0390

EU. (2009) http://eur-lex.europa.eu/LexUriServ/ LexUriServ do?uri=OJ:L:2009:309:0071:0086:en:PDF

Lamichhane J.R., Dachbrodt-Saaydeh S., Kudsk P., Messéan A. (2016). Toward a Reduced Reliance on Conventional Pesticides in European Agriculture, Plant Disease, 100 (1), 10-24. doi: 10.1094/PDIS-05-15-0574-FE

Flint, M.L. & van den Bosch, R. (1981). Introduction to integrated pest management. New York: Plenum Press. doi: 10.1017/S0014479700013715

Wasim Aktar, Dwaipayan Sengupta, Ashim Chowdhury (2009). Impact of pesticides use in agriculture: their benefits and hazards, Interdiscip Toxicol. 2(1): 1–12. doi: 10.2478/v10102-009-0001-7

Hajek, A. (2004). Natural enemies: an introduction to biological control. Cornell University, New York Cambridge University Press, doi: 10.1017/CBO9780511811838

Graham-Bryce, I.J. (1977) Crop Protection: A Consideration of the Effectiveness and Disadvantages of Current Methods and of the Scope for Improvement, Philosophical Transactions Royal Society London, Series B, Biological Sciences, 281 (980): 163-179. doi: 10.1098/rstb.1977.0131

TOPPS. (2008a) http://www.topps-life.org/uploads/8/0/0/3/8003583/_topps_best_management_practices_findoc_070613.pdf

TOPPS (2008b) http://www.topps-life.org/uploads/8/0/0/3/8003583/ansicht_drift_book_englisch.pdf

DEEP LEARNING BASED PLANT PART DETECTION IN GREENHOUSE SETTINGS

Manya Afonso[1], Ruud Barth[2] and Aneesh Chauhan[3]

[1]Biometris, Wageningen University and Research, The Netherlands

[2]Greenhouse Horticulture, Wageningen University and Research, The Netherlands

[3]Food and Bio-based Research, Wageningen University and Research, The Netherlands

manya.afonso@wur.nl, ruud.barth@wur.nl, aneesh.chahan@wur.nl

ABSTRACT

Precision agriculture challenges such as automatic harvesting, phenotyping, and yield prediction require precise detection of plant parts such as the fruits, leaves or stems. Deep learning has emerged as the state-of-the-art technology for image segmentation and object detection in several domains, notably in self-driving vehicles and medical imaging. In recent years, deep learning methods are being increasingly adopted in vision-based applications for precision agriculture. In previous work, methods were investigated to segment the image for plant parts. However, such an approach did not yield object instances. In this work, we applied the state-of-the-art deep learning object detector, MaskRCNN, to the problem of detecting fruit and other plant parts, in the sweet pepper (capsicum annuum) plant. An extensive study was carried out where we investigated different transfer learning schemes, different convolutional neural network architectures, and varying numbers of training images. Experimentally, we found that MaskRCNN trained with the synthetic data and fine-tuned with very few empirical images is able to detect more than 95% of the sweet pepper fruit. It was also found that training on the synthetic data and then fine-tuning over a few empirical images led to a better performance in the detection of fruit, over training only on the limited set of empirical images. Furthermore, results show that the best model could successfully generalize to different imaging conditions. This work is a necessary step for applying deep learning in high-throughput robotics and phenotyping approaches and will open up many opportunities for smart farming and more efficient use of resources. Currently, training deep learning models is dependent on the knowledge and expertise of the scientists involved. The insights gained from this work should lead to more automatic training protocols, allowing widespread use in very different applications.

Keywords: Computer Vision, Robotics, Deep Learning, Synthetic Data, Instance Segmentation.

1. INTRODUCTION

Automation and robotic tasks in agriculture such as harvesting, pruning or localized spraying require detailed and accurate localization of plant parts from images (Bac et. al., 2014). Traditionally, computer vision solutions to detect these parts were based on hand-crafted features such as shape and texture (Kapach et. al., 2012). However due to the large variability between different varieties of the crop, specimens of the same crop and variety, and in the imaging conditions (Barth, 2018; Bac, 2015), methods from the domain of machine learning such as convolutional neural networks, are being increasingly used due to their ability to cope with large amounts of variation in training data. Deep neural networks (Lecun et. al., 2015) have emerged as the state-of-the-art in computer vision, from image level classification (Krizhevsky et. al., 2012) to pixel-wise or semantic segmentation (Long et. al., 2015). In many applications including agriculture, object detection rather than semantic segmentation is required to discern individual instances in order to determine manipulation or application points for each object. FastRCNN (Girshick et. al., 2015) is an object detector which uses region based convolutional neural networks (R-CNN) combined with the selective search method to detect region proposals. Another fast and popular deep learning object detector is the You Only Look Once (YOLO) (Redmon et. al., 2016) which applies a single neural network to the full image which divides the image into regions and predicts bounding boxes and probabilities for each region, unlike Fast-RCNN and its variants which apply the model to an image at multiple locations and scales. These methods provide a classification with a bounding-box localization of instances of the object of interest. A variant of FastRCNN has been used to detect sweet peppers and other fruit from color and near infrared images (Sa et. al., 2016). A recent method, Mask-RCNN, extends FastRCNN by a branch for predicting an object mask in parallel with the existing branch for bounding box recognition (He et. al., 2017). In our previous work, it was shown that plant parts could be successfully semantically segmented in the image, based on synthetic data bootstrapping (Barth et. al., 2018) and fine-tuning with a small empirical dataset (Barth et. al., 2017).

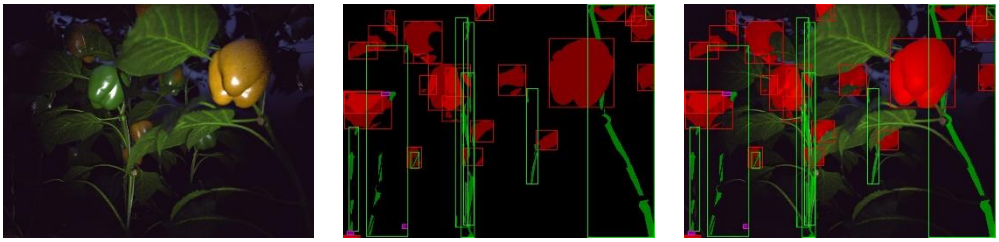
In this paper we build upon previous pixel-wise segmentation results, by using Mask-RCNN to obtain instance detection of plant parts. This paper also investigates different transfer learning techniques, with and without training on the synthetic or empirical dataset. Experimental results show that under the best training configuration (architecture and transfer learning scheme), almost 100% of the fruits, 75% of the stems, and 80% of the peduncles can be detected, with few false positives. An improvement over the Intersection over union (IOU) metric was obtained compared to previous pixel-wise segmentation. Thus, the use of MaskRCNN, resulted in a practically useful method for sweet pepper part detection with state-of-the-art performance.

2. MATERIALS AND METHODS

2.1 Dataset

The image dataset consists of the first 200 synthetic images out of the dataset from (Barth et. Al., 2017 and Barth, 2018), modelled to approximate the empirical set visually and 50 empirical (real) images of a sweet-pepper crop in a commercial high-tech greenhouse. The dataset is available online. Because the ground truth of this dataset was on a per-pixel level and did not contain instances, the empirical dataset was manually relabeled for unique objects of the classes fruit, stem, and peduncle using the LabelMe tool (Russell et. al., 2008). For the synthetic images, the instance wise ground truth was obtained by re-rendering each object alone in the 3D scene, thus producing the complete shape of each object. To deal with occlusions, a post processing step was used to include only the class pixels for each instance that were also present in the original ground truth. The 200 synthetic images were used only for training. The first 30 images from the empirical dataset were used for training or fine-tuning while image numbers 31 to 50 are used for testing. An alternative empirical dataset of 20 images, that used different illumination and camera, was also acquired and labeled with the aim to verify the trained model's generalizability and robustness to new conditions in other datasets.

An example image from the synthetic dataset and its ground truth are shown in Figure 1. It can be seen that the color image is quite realistic, compared to empirical ones in Figures 2(a) and 3(a).



(a) (b) (c)

Figure 1: Example of synthetic image (left), of its corresponding ground truth instance labels (middle), and of the GT overlaid (right).

2.2 Deep Learning Software Setup and algorithms

Detectron, Facebook AI Research's software with the implementation of Mask R-CNN, was used to train convolutional neural network models with the image datasets from the previous section to perform instance segmentation of plant parts. It was installed on a workstation with an NVIDIA GeForce GTX 1080 Ti 11GB GPU with Ubuntu LTS 16.04 64-bit supported by CUDA 9.0.

Mask-RCNN uses a convolutional neural network architecture such as ResNet (He et. al., 2016) as the backbone, which extracts the feature maps. The region proposal network (RPN) applies a sliding window over this feature map and calculates region proposals, which are then pooled with the feature maps and over each, a classifier is applied resulting in a bounding box prediction corresponding to an instance of the particular class. The scheme until this point is the same as FasterRCNN. An additional convolutional network is applied on the aligned region and feature map to obtain a mask for each bounding box.

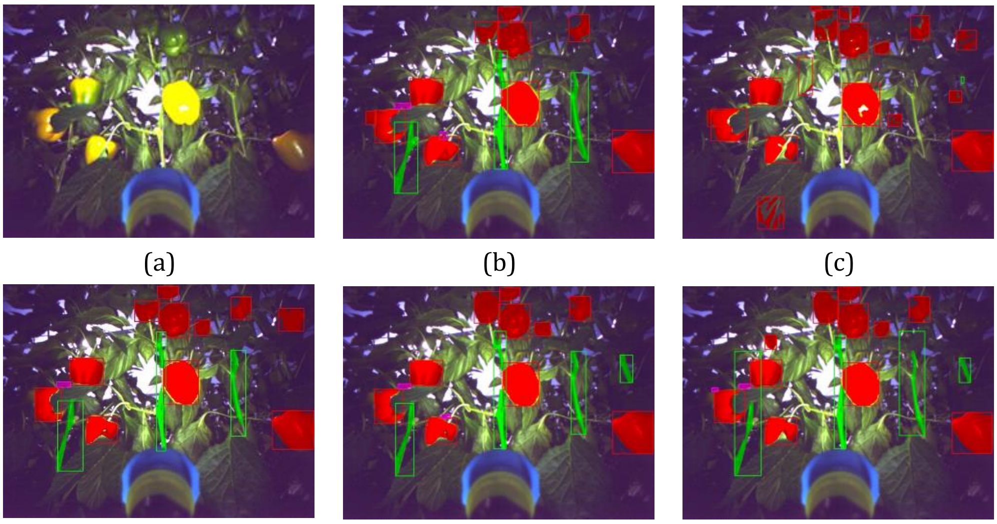
The following backbone CNN architectures were varied: the 50 and 101-layer ResNets,and ResNext (Xie et. al., 2017) configurations ResNext 101-32x8d (101 layers, cardinality 32, bottleneck width 8) and ResNext 101-64x4d (101 layers, cardinality 64, bottleneck width 4). For these architectures, models pre-trained on the ImageNet1k dataset were downloaded from the Detectron repository. Since this paper addresses the problem of sweet-pepper plant part detection, models pre-trained on ImageNet-1K offer an advantage since one of the classes in this dataset is 'bell-pepper'.

Different transfer learning schemes, training only on the synthetic dataset or only on the empirical dataset, or training on the synthetic set followed by fine-tuning on the empirical were also studied.

3. RESULTS

For the detected objects, true positives, false positives, and false negatives were determined using a criterion that a 25% in overlap of the segmented instance and the ground truth masks should result in a positive detection thereof. This level of overlap was chosen to take into account the fact that some fruit and stems may have disjoint sections due to occlusions, not all of which get detected as part of the same instance. From these measures, the precision and recall were calculated for each plant part class. For comparison with earlier pixel-wise segmentation, the Jaccard Index similarity coefficient or intersection-over-union (IOU) (Csurka et. al., 2013) was calculated by combining all pixels for a particular class and image and comparing against the pixel-wise ground truth.

Figures 2 and 3 present results of segmentation obtained with different transfer learning schemes and architectures, and Table 1 summarizes the figures of merit for the different configurations.



(a) (b) (c)

(d) (e) (f)

Figure 2: Detection on empirical image no. 36: (a) RGB image; (b) RGB with overlaid GT; detections with: (c) R101 trained on synthetic set; (d) R101 trained only on emp. images 1-30; (e) R101 and (f) X101 64x4d, both trained on synthetic and fine-tuned on emp. images 1-30.

Table 1: Summary of detection statistics using Mask-RCNN with different backbone architectures and training schemes. The 3 best values for each metric over the empirical test set are in bold.

Network		**Fruit**			**Stem**			**Peduncle**	
	P	**R**	**IoU**	**P**	**R**	**IoU**	**P**	**R**	**IoU**
Tested on: emp 31-50									
Pixelwise (VGG)			0.76			0.41			0.39
Experiment 1, Trained on: synth 1-200, Tested on: emp 31-50									
ResNet101	0.56	0.92	0.56	0.11	0.04	0.01	0.73	0.3	0.08
ResNext101 32x8d	0.43	**0.99**	0.54	0.17	0.14	0.09	0.56	0.24	0.09
ResNext101 64x4d	0.41	0.95	0.45	0.4	0.21	0.1	0.7	0.4	0.16
Experiment 2, Trained on: synth 1-200, Finetuned: emp 1-30, Tested on: emp 31-50									
ResNet101	**0.9**	0.96	**0.78**	0.92	0.59	**0.39**	**0.85**	0.71	**0.41**
ResNext101 32x8d	0.87	**0.97**	0.77	**1**	0.56	**0.37**	0.78	**0.84**	**0.46**
ResNext101 64x4d	0.87	0.96	**0.78**	**0.98**	**0.76**	**0.47**	0.77	**0.82**	**0.48**
Experiment 3, Trained on: emp 1-30, Tested on: emp 31-50									
ResNet101	0.87	**0.97**	0.76	0.97	0.56	0.33	0.79	0.62	0.36
ResNext101 32x8d	0.87	0.94	0.74	**1**	**0.67**	0.36	0.82	0.65	0.34
Tested on: alt. emp 1-20									
ResNet101 train syn 1-200	0.21	0.6	0.32	0	0	0.02	0.06	0.08	0.02
ResNet101 train emp 1-30	0.77	0.71	0.72	0.77	0.5	0.16	0.64	0.6	0.2
R101 train syn 1-200 ft emp 1-30	0.67	1	0.69	0.7	0.46	0.17	0.33	0.46	0.07

(a) (b) (c)

Figure 3: Detection results alternative emp. image no. 11: (a) RGB image; (b) RGB with overlaid GT; (c) detection with R101 trained on synthetic images and fine-tuned on emp. images 1-30.

4. Discussion

From the results in Table 1, it can be seen that the best transfer learning scheme was training on the synthetic dataset, followed by fine-tuning on the empirical dataset. This scheme was found to perform better than training only on the empirical dataset. For the detection of the fruits, the best architecture was found to be ResNet101, whereas ResNext101 64x4d was the best for the detection of the stems. The two did not vary much in the detection of the peduncles. With this combination of transfer learning, architectures, and the full empirical training dataset, we obtained for the fruits, a precision of above 0.90 and a recall of close to 1.0, for the stems, a precision of close to 1.0 and a recall above 0.75, and for the peduncles a precision and recall of close to 0.80.

Overall, the intersection over union values for each class were higher than the respective ones reported in previous work on pixel-wise segmentation. Thus, state-of-the-art performance was achieved in detection of these plant parts using Mask-RCNN whilst determining the optimal transfer learning strategy, network architecture and dataset size.

5. Conclusions

In this work, deep learning instance detection was applied to the problem of detecting pepper plant parts. Experimental results show that this approach works well for the detection of fruit, which is useful for practical applications such as harvesting and yield estimation. Our results also show that the synthetic pepper dataset is useful for initial training, to detect peppers in real life images, thus reducing the need for a large number of manually annotated images.

In the experiments, the best numerical results were obtained for the detection of the fruit. The detection of other plant parts such as stems and peduncles is a harder problem likely due to the similarity in color to leaves, twigs, and other unlabeled parts. The segmentation of plant parts could be useful for other plants grown in greenhouses and similar environments, for example tomatoes regarding yield estimation or stem thickness as an important trait for phenotyping.

References

Bac, C. W. (2015). Improving obstacle awareness for robotic harvesting of sweet-pepper. PhD thesis, Wageningen University and Research.

Bac, C. W., van Henten, E. J., Hemming, J., and Edan, Y. (2014). Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. Journal of Field Robotics, 31(6):888–911.

Barth, R. (2018). Vision principles for harvest robotics : sowing artificial intelligence in agriculture. PhD thesis, Wageningen University and Research.

Barth, R., IJsselmuiden, J., Hemming, J., and E.J. Van Henten (2017). Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation. Computers and Electronics in Agriculture.

Barth, R., IJsselmuiden, J., Hemming, J., and E.J. Van Henten (2018). Data synthesis methods for semantic segmentation in agriculture: A capsicum annuum dataset. Computers and Electronics in Agriculture, 144:284 – 296.

Gabriela Csurka, Diane Larlus, F. P. (2013). What is a good evaluation measure for semantic segmentation? In Proceedings of the British Machine Vision Conference. BMVA Press.

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., and He, K. (2018). Detectron. https://github.com/facebookresearch/detectron.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In Computer Vision (ICCV), 2017 IEEE International Conference on, pages 2980–2988. IEEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.

Kapach, K., Barnea, E., Mairon, R., Edan, Y., and Ben-Shahar, O. (2012). Computer vision for fruit harvesting robots – state of the art and challenges ahead. Int. J. Comput. Vision Robot., 3(1/2):4–34.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105.

LeCun, Y., Bengio, Y., and Hinton, G. (2015), Deep learning, nature, 521(7553):436.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 779–788.

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). Labelme: A database and web-based tool for image annotation. International Journal of Computer Vision, 77(1):157–173.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. Sensors (Basel), 16(8):1222. 27527168[pmid].

Uijlings, J. R., Van De Sande, K. E., Gevers, T., and Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104(2):154–171.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 5987– 5995. IEEE.

Exploring OWA Operators For Aggregating Fuzzy Cognitive Maps Constructed by Experts/Stakeholders in Agriculture

Konstantinos Papageorgiou[1], Elpiniki Papageorgiou[2,3], Asmaa Mourhir[4] and George Stamoulis[1]

[1] University of Thessaly, Department of Computer Science, 35100 Lamia, Greece

[2] University of Thessaly, Faculty of Technology, 41110 Larisa, Greece

[3] Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Greece

[4]Computer Science Department, School of Science and Engineering, Al Akhawayn University in Ifrane, Morocco

konpapageorgiou@uth.gr, elpinikipapageorgiou@uth.gr, a.mourhir@aui.ma

Abstract

Fuzzy cognitive maps are graph-based models mostly constructed in a participatory setting either by experts to given domains or by a large number of stakeholders in various scientific areas of interest. Usually each expert or stakeholder designs individually a FCM based on his/her knowledge or opinion for the specific application domain. For scenario analysis and decision-making purposes, an overall FCM for the specific problem needs to be constructed, aggregating all individual FCMs designed by the experts and/or stakeholders. The average aggregation method for weighted interconnections among concepts is the most common method in FCM modeling, which is quite simple, regardless the inherent uncertainty induced by the different experts' or stakeholders' opinions. The aim of this research work is two-fold: (i) to propose an alternative aggregation method based on learning Ordered Weighted Average (OWA) operators in aggregating FCM weights, assigned by many experts and/or stakeholders, and (ii) to present a new software tool for FCM aggregation, called FCM-OWA, leveraging the different FCM aggregation methods. A precision farming problem, considering apple yield prediction, is used to show the applicability and usefulness of the proposed methodology in modeling. The results after comparing OWA operators with the traditional average aggregation method imply that the proposed approach is really challenging on modeling experts' knowledge in agricultural domain.

Keywords*:* fuzzy cognitive maps, knowledge aggregation, OWA, participatory approach, decision making.

1. Introduction

A Fuzzy cognitive map (FCM) is a flexible and innovative technique which combines neural networks and graph theory, for modelling human knowledge in decision-making processes. Introduced by Kosko in 1986, FCMs are simple models that can easily incorporate human knowledge and adapt to a given domain (Stach, 2010), so they have gained considerable research interests and have been extensively applied to many research fields and application areas (Papageorgiou & Salmeron, 2013). Designing a FCM for modelling a given system often includes the aggregation of knowledge from a variety of sources. Typically, aggregated input models are developed by multiple experts from the application domain (Stach, 2010), aiming at improving the reliability of the final model which is less susceptible to potentially erroneous beliefs of a single expert or knowledge discrepancies among participants. There are a couple of techniques for combining multiple FCMs into a single collective model commonly used in a wide range of real-life problems (Mezei & Sarlin, 2016), which are namely the weighted average, and the OWA method introduced by Yager (1988).

In a traditional FCM, considering the weighted average method, the weights are fuzzy values in the interval [0, 1] and the collective FCM model is constructed by averaging values for a given interconnection (Gray et al., 2014). This approach is also particularly useful when the causal relationships are described using linguistic terms. The rules for a given interconnection are aggregated using fuzzy operators, and the overall output is elaborated by using the weighted average of each rule's output (Papageorgiou & Stylios, 2008).

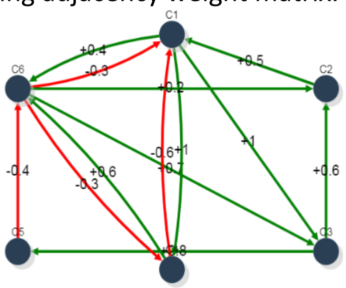
From the reviewed literature, few research papers were found regarding the application of OWA operators in aggregating individual FCMs. Among them, Zhenbang and Lihua (2007) introduced OWA operators, as numeric aggregation operators, into the conventional FCMs. They also discussed the issues related to determining weights for the OWA aggregation under different conditions. The aim of this work is to develop an alternative FCM aggregation method by learning OWA operators' weights. A precision farming problem with apple yield prediction is used to show the applicability and usefulness of the proposed methodology to support decision making.

This paper is structured as follows. Section 2 describes the theoretical background about the OWA operators along with the proposed algorithm for learning OWA operator weights for aggregation purposes. Section 3 presents the FCM aggregation tool. The results of the examined apple yield case study are presented and discussed in Section 4. Section 5 summarizes the study's findings and conclusions.

2. METHODOLOGY

2.1 Fuzzy Cognitive Maps Overview

FCMs, introduced by Kosko (1986), allow representing knowledge in the form of a directed graph whose nodes denote the main factors of the analyzed problem, and links represent the causal relationships between concepts. Each of FCM's link is associated with a weight value w_{ij} which varies from -1 to 1, that describes the strength of the corresponding relation between two concepts C_i and C_j. There are three different types of possible causalities between every pair of concepts C_i and C_j: (i) $W_{ij} > 0$, which designates a positive causality, (ii) $W_{ij} < 0$, which designates a negative causality, and (iii) $W_{ij} = 0$, which designates no causality. Figure 1 shows an example of a Fuzzy Cognitive Map with its corresponding adjacency weight matrix.



$$W = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0.4 \\ 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0.8 & 0 \\ -0.6 & 0 & 0 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & -0.4 \\ -0.3 & 0.2 & 0.7 & -0.3 & 0 & 0 \end{bmatrix}$$

Figure 1. Fuzzy cognitive map (left) and the correspondent weight adjacency matrix (right), showing the positive and negative causal influences.

Typically, a FCM of n concepts could be represented mathematically by a $n \times n$ weight matrix (W). By feeding the fuzzy cognitive map with an initial stimulus state vector X(t) (state vector at time (t)), it can model the evolution of a scenario over time by evolving forward and letting concepts interact with one another. Each subsequent value of the concept state $X^{(t+1)}$ can be computed as previous state $X^{(t)}$ and weight matrix multiplication, according to Eq. (1).

$$X_i^{(\kappa+1)} = f\left(\sum_{j=1, j\neq i}^{n} w_{ji} \times X_j^{\kappa}\right) \tag{1}$$

In the average aggregation method, all the numerical weighted interconnections suggested by various experts, are summed and then divided by the number of experts, thus producing an average numerical weight. An example of this average method is presented in (Stach, 2010).

2.2 Main aspects of ordered weighted averaging (OWA) operators

An OWA operator of dimension n is a mapping: $f: R^n \rightarrow R$, that has an associated weighting vector W, $W = [w_1\ w_2\ \dots\ w_n]^T$, such that $\sum_i w_i = 1$; $w_i \in [0, 1]$ and $f(a_1, \dots, a_n) = \sum_{j=1}^{n} w_j\ b_j$, where b_j is the j[th] largest element of the collection of the aggregated objects $a_1, a_2\ \dots, a_n$. The function value $f(a_1, \dots, a_n)$ determines the aggregated value of arguments $a_1, a_2\ \dots, a_n$ (Filev & Yager, 1994).

A fundamental aspect of the OWA operator is the re-ordering step, in particular, an argument a_i is not associated with a particular weight w_i but rather a weight w_i is associated with a particular ordered position i of the arguments.

It can be easily shown that the OWA operators are aggregation operators, satisfying the commutativity, monotonicity and idempotency properties and that they are bounded by the Max and Min operators (Yager, 1993), for OWA operators

$$\operatorname*{Min}_i a_i \leq f(a_1, \dots, a_n) \leq \operatorname*{Max}_i a_i$$

Since this class of operators runs between the $Max\ (or)$ and the $Min\ (and)$, Yager (1988) introduced a measure to characterize the type of aggregation being performed for a specific value of the weighting vector. This measure called the $orness\ measure$ of the aggregation is defined as shown by (2).

$$Orness(W) = \frac{1}{n-1} \sum_{i=1}^{n} (n-1) w_i. \tag{2}$$

As suggested by Yager (1988), this measure which lies in the unit interval, characterizes the degree to which the aggregation is like an $or\ (Max)$ operation. It can be shown that

$$orness([1\ 0 \dots 0]^T) = 1, \qquad orness([0\ 0 \dots 1]^T) = 0, \qquad orness\left(\left[\frac{1}{n}\ \frac{1}{n} \dots \frac{1}{n}\right]^T\right) = 0.5.$$

Therefore the Max, Min and arithmetic mean operators can be regarded as OWA operators with a degree of orness, respectively, 1, 0 and 0.5. A second measure introduced by Yager (1988) was the dispersion or entropy associated with a weighting vector:

$$Disp(W) = \sum_{i=1}^{n} w_i\ \ln w_i. \tag{3}$$

This was suggested for use in calculating how much of the information in the arguments is used during an aggregation based on W.

2.3 Calculating OWA weights from experts/stakeholders' opinions

Based on the previously suggested method by Yager (1988) for obtaining the weights associated with the OWA aggregation when observed data on the arguments have been provided, this research study presents an algorithm which can be used for aggregating weights assigned by experts/stakeholders' opinion in designing FCMs. The proposed algorithm learns the weights associated with a particular use of the OWA operator from a group of experts and stakeholders for the specific scientific domain. The steps below, describe the procedure to obtain the learning OWA weights (Fig.2).

We consider Expert opinions as argument values ($a_{k1}, a_{k2} \dots, a_{kn}$), and sample as each edge-weight (n) of data.

Step 1: Generate slightly different parameters ρ for each argument which represents the optimism of the decision maker, 0≤ρ≤1.

Step 2: Calculate the aggregated values for each sample using the Hurwics method according to which, the aggregated value d obtained from a tuple of n arguments, $a_1, a_2 \dots, a_n$, is defined as a weighted average of the Max and Min values of that tuple.

$$\rho \mathrm{Max}\ a_i + (1-\rho)\mathrm{Min} a_i = d$$

Step 3: Reorder the objects ($a_{k1}, a_{k2} \dots, a_{kn}$).

Step 4: Calculate the current estimate of the aggregated values dk

$$\hat{d}_k = b_{k1}w_1 + b_{k2}w_2 + \cdots + b_{kn}w_n$$

with initial values of the OWA weights $w_1 = 1/n$.

Step 5: Calculate the total $\hat{d}_k$, d_k, b_{ki} for each i. The parameters λ_i determine the weights of OWA and are updated with the propagation of the error $\hat{d}_k - d_k$ between the current estimated aggregated value and the actual aggregated value (Filev & Yager, 1994).

Step 6: Calculate the current estimates of the λ_i

$$\lambda_i(l+1) = \lambda_i(l) - \beta w_i(l)\left(b_{ki} - \hat{d}_k\right)\left(\hat{d}_k - d_k\right)$$

with initial values $\lambda_i(0) = 0$, i = (1, n), and a learning rate of β = 0.35.

Step 7: Use λ_i, $i = (1, n)$, to provide a current estimate of the weights

$$w_i = \frac{e^{\lambda_i(l)}}{\sum_{j=1}^{n} e^{\lambda_i(l)}},\ i = (1, n)$$

Step 8: Update w_i and $\hat{d}_k$ at each iteration until the estimates for all the λ_i converge to, that is Δ = |$\lambda(l+1)$ – $\lambda(l)$| are small.

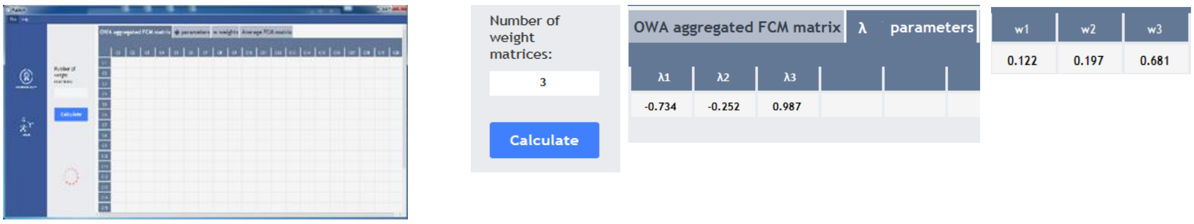
Figure 2. The algorithm for learning the OWA weights

In what follows, an explanatory paradigm from precision farming devoted to apple yield prediction will better illustrate the proposed FCM construction approach by experts (see section 3).

2.4 FCM-OWA Tool Description

The proposed algorithm based on learning OWA operator weights for aggregating FCM models was implemented with the FCM-OWA, a new software tool developed in java programming language. An overview snapshot of the user interface, where the main menu with the “import” and “export” buttons and the “OWA aggregation” function are present, can be seen in Fig. 3a.

The user needs first to define the number of input weight matrices as shown in Fig. 3b, and next to insert an excel file with the weight matrices, one in each sheet, defined by each expert. Once the user clicks on “Calculate”, the OWA aggregated FCM matrix appears in the first tab, while in the following 3 tabs, the lamda parameters, the w weights and the average FCM matrix are displayed (Fig.3). Additionally, users are given the option to export the OWA aggregated FCM matrix in an excel file.



(a) The FCM-OWA tool (b)Number of matrices (c) lamda parameters tab (d) w weights tab

Figure 3. Screenshots of (a) The FCM-OWA tool for aggregation (b)The number of weight matrices option, (c) “lamda parameters” tab and (d) “w weights” tab

3. Results

As a case study we selected a precision farming problem of apple yield prediction previously published in Papageorgiou et al. (2013). Specifically, a number of experts designed 17 individual FCM models that were collected and aggregated to provide support to policy making. The FCM models which were designed and developed to represent experts’ knowledge for yield prediction and crop management, consist of 9 concepts and 13 weighted connections among them. The developed FCM model comprises nodes linked by directed edges, where the nodes represent the main soil factors affecting yield (such as soil texture (clay (C7) and sand content (C8)), soil electrical conductivity (EC)- (C1), calcium (Ca)-

(C2), potassium (K)- (C3), organic matter (OM)- (C4), phosphorus (P)- (C5), and zinc (Zn)- (C6) contents), and the directed edges show the cause-effect (weighted) relationships between the soil properties and yield. In this work, we consider only 3 out of 17 expert models to demonstrate the implementation of our proposed aggregation methodology. The expert opinions are considered as argument values $a_{k1}, a_{k2} \dots, a_{kn}$ and the weight between two concepts is shown in Table 1.

The aggregated values were calculated using various values for parameter ρ within ρ (0.01≤ ρ≤0.2). For example, ρ=0.153; 0.131; 0.181; 0.075; 0.055. These values have been randomly selected by the algorithm. Using min and max values of ρ, the aggregated value of weight was calculated as follows.

$$0.153(0.58) + (1 - 0.153)(0.43) = 0.45$$

We initialized $\lambda_i(0) = 0$, i = (1, n), β = 0.35 και $w_1 = w_2 = w_3$=0.33. The estimated values of λ_i after 108 iterations were: $\lambda_1 = 0.63,\ \lambda_2 = -0.19,\ \lambda_3 = 0.82$

The following OWA weights were calculated considering the above λ_i:

$$w_1 = 0.146, \qquad w_2 = 0.227, \qquad w_3 = 0.626$$

Applying the OWA tool, all the weight matrices of 3 experts of precision farming were uploaded and their OWA weights were calculated (see Table 1). Furthermore, the same tool calculates the average values of weights (benchmark method Wb) through the Kosko's aggregation method. We followed the same process for parameter ρ for the following ranges: 0.3<ρ<0.5 and 0.5<ρ<0.7.

Table 1. Aggregated (benchmark) weights from 3 experts' opinions and weights produced by learning OWA operators (Wowa)

	Experts' opinions			**Aggregated value**	**Weights produced by learning OWA operators (Wowa)**			**ΔW-Deviation Wb-Wowa**		
Weight	Exp. 1	Exp. 2	Exp. 3	Wb Average	Weight-1 0.01<ρ<0.2	Weight-2 0.3<ρ<0.5	Weight-3 0.5<ρ<0.7	ΔW1	ΔW2	ΔW3
C1-C9	0.43	0.50	0.58	0.45	0.47	0.53	0.58	0.02	0.08	0.13
C2-C1	0.57	0.60	0.68	0.58	0.33	0.40	0.45	0.07	-0.18	0.13
C2-C9	0.57	0.60	0.68	0.58	0.59	0.64	0.68	0.01	0.06	0.10
C3-C1	−0.30	-0.35	-0.25	-0.32	-0.32	-0.28	-0.25	0.00	0.04	0.07

Table 1 gathers the calculated values for OWA aggregated weights for some exemplary interrelationships among FCM concepts, as well as the deviations between the benchmark weight Wb (average method) and the Wowa, weight produced by learning OWA operators. Also, considering the ρ values, it is presented that there are significant deviations for higher values of ρ (ie. for weights C2-C1 and C6-C9) in the overall FCM model. For most of the weighted interconnections above, the deviations are relatively small.

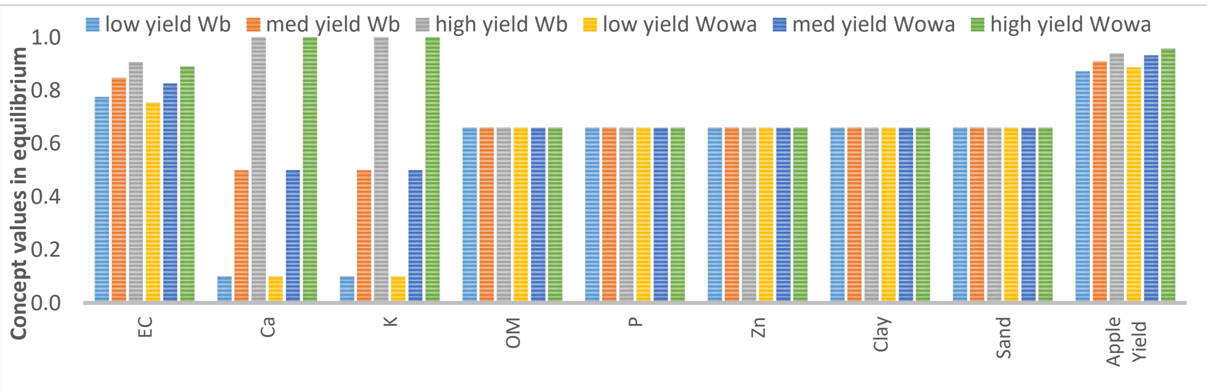


Figure 4. Preliminary results of three scenarios of low, medium and high yield

Preliminary simulations with the aggregated FCM produced by OWA have been accomplished and some illustrative results are given in Fig. 4. Three scenarios of low, medium and high yield were examined. In all cases, the two key concepts affecting apple yield, Ca and K, are modified to generate perturbation in the yield. For example, in the case of high yield, Ca and K have a significant contribution on decision making. The simulation results show deviations between Wb and Wowa for the three examined scenarios of low, medium and high apple yield in the equilibrium. Similar trends are observed overall for both FCMs.

4. Discussion & Conclusions

In this work, an OWA-based FCM aggregation method by learning OWA operator weights from data was introduced and applied to a precision farming case study for apple yield prediction. The first evident outcome that makes this research study important regards the relationships' strengths calculated with the proposed approach. It came up that each weighted relation has a value similar to the benchmark weight when ρ takes small values. Specifically, the absolute difference between them is equal to 0.06 on average, which means that the weight obtained by OWA is slightly greater than the benchmark one. In this preliminary work, it is proved that there is a consistency between the values of the weights came from the OWA method and the benchmark weight, justifying the usefulness of the presented methodology as well as the easiness of use of the new software tool proposed in this study. As future work, we will focus on a more thorough investigation of this aggregation method by performing extended scenario analysis on both OWA aggregating FCMs and benchmark aggregation method, and further comparing the results in the basis of policy making in agriculture.

References

Axelrod, R. (1976). *Structure of Decision*. Princeton, USA: Princeton University Press.

Filev, D. and Yager, R.R. (1998). On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*. 94, 157-169.

Gray, S.A., Zanre, E. and Gray, S.R.J. (2014). Fuzzy Cognitive Maps as Representations of Mental Models and Group Beliefs, In: Papageorgiou, E.I. (Ed.), Fuzzy Cognitive Maps for Applied Sciences and Engineering. Springer Berlin Heidelberg, pp. 29-48.

Kosko, B. (1986). Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies.* 24, 65–75.

Leyva-Vαzquez, M., Pιrez-Teruel, K. and John, R. (2014). A Model for Enterprise Architecture Scenario Analysis Based on Fuzzy Cognitive Maps and OWA Operators, IEEE conference.

Mezei, J. and Sarlin, P. (2016). Aggregating expert knowledge for the measurement of systemic risk. *Decision Support Systems.* 88: 38–50.

O'Hagan, M. (1988). Aggregating template rule antecedents in real-time expert systems with fuzzy set logic, *Proc. 22nd Ann. IEEE* Conf. on Signals, Systems and Computers, Pacific Grove, CA. 681-689.

Papageorgiou, E. and Stylios, C. (2008). *Fuzzy cognitive maps. Handbook of granular computing*. pp. 755-774. Chichester, England: John Wiley and Son Ltd, Publication Atrium.

Papageorgiou, E.I., Aggelopoulou, K.D., Gemtos, T.A. and Nanos, G.D. (2013). Yield prediction in apples using Fuzzy Cognitive Map learning approach. *Computers & Electronics in Agriculture*. 91, 19-29.

Sharif, A.M. and Irani, Z. (2006). Exploring Fuzzy Cognitive Mapping for IS Evaluation. *European Journal of Operational Research*, vol. 173, pp.1175-1187.

Stach, W. (2010). Learning and Aggregation of Fuzzy Cognitive Maps - An Evolutionary Approach. PhD Dissertation, University of Alberta.

Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics.* 18, 183–190.

Yager, R.R. (1993). Families of OWA operators, Fuzzy Sets and Systems. 59, 125-148.

Zhenbang, Lv. & Lihua, Zhou. (2007). Advanced Fuzzy Cognitive Maps Based on OWA Aggregation. International journal of computational cognition (http://www.ijcc.us), vol. 5, no. 2.

GRAPEVINE YIELD PREDICTION USING IMAGE ANALYSIS - IMPROVING THE ESTIMATION OF NON-VISIBLE BUNCHES

Gonçalo Victorino[1,2], Guilherme Maia[1], José Queiroz[1], Ricardo Braga[1], Jorge Marques[2], José Santos-Victor[2], Carlos Lopes[1]

[1]LEAF, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal

[2]ISR, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

gvictorino@isa.ulisboa.pt

ABSTRACT

Yield forecast is an issue of utmost importance for the entire grape and wine sectors. There are several methods for vineyard yield estimation. The ones based on estimating yield components are the most commonly used in commercial vineyards. Those methods are generally destructive and very labor intensive and can provide inaccurate results as they are based on the assessment of a small sample of bunches. Recently, several attempts have been made to apply image analysis technologies for bunch and/or berries recognition in digital images. Nonetheless, the effectiveness of image analysis in predicting yield is strongly dependent of grape bunch visibility, which is dependent on canopy density at fruiting zone and on bunch number, density and dimensions. In this work data on bunch occlusion obtained in a field experiment is presented. This work is set-up in the frame of a research project aimed at the development of an unmanned ground vehicle to scout vineyards for non-intrusive estimation of canopy features and grape yield. The objective is to evaluate the use of explanatory variables to estimate the fraction of non-visible bunches (bunches occluded by leaves). In the future, this estimation can potentially improve the accuracy of a computer vision algorithm used by the robot to estimate total yield.

In two vineyard plots with Encruzado (white) and Syrah (red) varieties, several canopy segments of 1 meter length were photographed with a RGB camera and a blue background, close to harvest date. Out of these images, canopy gaps (porosity) and bunches' region of interest (ROI) files were computed in order to estimate the corresponding projected area. Vines were then defoliated at fruiting zone, in two steps and new images were obtained before each step.

Overall the area of bunches occluded by leaves achieved mean values between 67% and 73%, with Syrah presenting the larger variation. A polynomial regression was fitted between canopy porosity (independent variable) and percentage of bunches not occluded by leaves which showed significant R^2 values of 0.83 and 0.82 for the Encruzado and Syrah varieties, respectively.

Our results show that the fraction of non-visible bunches can be estimated indirectly using canopy porosity as explanatory variable, a trait that can be automatically obtained in the future using a laser range finder deployed on the mobile platform.

Keywords: Precision viticulture, bunch occlusion, image analysis, robot, yield estimation

1. INTRODUCTION

Accurate yield estimation is extremely useful for the management of any crop. In viticulture this information can help vineyard managers in several aspects such as defining staff and machinery needed for harvest, future fertilization planning or anticipation of cellar needs (e.g. allocating tank space for wine making). If done early enough, yield forecasting can bring advantages towards planning bunch thinning needs (in order to prevent excessive yield and consequent poor wine quality), planning purchases and/or grape sales, establishing grape prices and managing wine stocks and grape and wine market, programming investments and developing marketing strategies (Dunn and Martin, 2004). However, yield estimations early in the season carry higher risk of lower accuracy due to unpredictable negative effects of biotic and abiotic factors, which may affect the final number and size of berries. These factors induce a high yield variability, both spatial and seasonal (Bramley and Hamilton, 2004; Victorino *et al.*, 2017) and therefore, yield predictions need to be considered for every season.

Traditional methods for yield estimation are based on counting sampled yield components which can be done all along the growing cycle. The veraison stage is one of the most used phenological stages to apply these methods as it is early enough for crop and winery managers to adapt their plans if needed, while also close enough to harvest to not jeopardize the estimation's accuracy. Estimations based on manual counting of yield components are simple enough to be accessible to anyone, and are still the most common practice today. However these methods are very labor-intensive and have low accuracy due to the difficulties related to sampling a large amount of vineyard area (Dunn and Martin, 2004).

To overcome the above-mentioned limitations, several methods have been developed. While some of them remain at a research level (Tarara *et al.*, 2014; Fraga and Santos, 2017) others are already in use by the industry, as is the case of the aeropalynological forecast models (Besselat, 1987; Cunha *et al.*, 2016). However, the aeropalynological methods are used mainly at a regional scale and are not recommended to be used by an individual winegrower, as the pollen grains transported by the wind can come from a highly unpredictable range of places and distances, not being site specific.

Recently, a big research effort has been done regarding the use of sensor-based technologies to address the overall yield estimation challenge. Nevertheless, so far, no commercial imaging-systems are available for grapevine yield estimation (Taylor *et al.*, 2018) but many recent new approaches are being developed.

Sensor-based methodologies encompass several challenges, being image data analysis one of the main ones. With the advancement of machine learning and its recent migration to agriculture applied technologies, new image processing algorithms have been developed bringing the possibility of analyzing great amounts of images in a short period of time. With such technology, pictures taken from the vineyard with, for example, a mobile platform, can today be automatically inspected, a task that would otherwise take many hours if done by a person.

One of the first attempts to use image analysis for yield estimation in viticulture was done by Dunn and Martin (2004). Since then, many new approaches that included machine learning algorithms were developed and are thoroughly reviewed up until 2017 by Seng *et al.* (2018). However, being such a trending technology, several recent works have been developed in the meantime for vineyard field conditions using machine learning. Such algorithms attempt to automatically extract information regarding a desired yield component in the collected image. This technology serves as an upgraded way to count yield components for yield estimation in automatic systems that collect extensive amounts of image data.

Neural networks (a particular type of machine learning algorithm) have been used for flower detection by Liu *et al.* (2018), for bunch detection by Milella *et al.* (2019), and for single berry detection by Aquino *et al.* (2018). Other models have been successfully used for this purpose, such as Boolean models (Millan *et al.*, 2018) and Random Forest Classifiers (Riggio *et al.,* 2018) for single

berry detection, as well as Support Vector Machines (Pérez-Zavala *et al.*, 2018) for bunch segmentation.
In several of these cases, images were collected using an on-the-go platform, some at night time with artificial lighting. This type of lighting prevents variability on light conditions, caused by different sunshine intensity and orientation, while also removes background noise from vines behind the targeted ones.
All of the previous algorithms had successful results at detecting yield components in grapevine images. However all of them are dependent of leaf removal, as vegetation in normal conditions covers a great percentage of the grape bunches (Nuske *et al.*, 2014). However, as leaf removal is not a generalized practice, yield estimation methods should rely on non-manipulated canopies, where part of the grape bunches are occluded by leaves. The fraction of bunches occluded by leaves is dependent of canopy porosity (fraction of gaps in the fruiting zone; Smart and Robinson, 1991). Compared to a dense canopy, a sparse one will have more gaps enabling more bunch visibility. Canopy porosity is most commonly measured using the *Point Quadrat* method, adapted to grapevines by Smart (1987). However, more recently it has been estimated using image analysis (De Bei *et al.*, 2016; Diago *et al.*, 2016), which would be most adequate when using sensor-based technology for yield estimation.
The present work explores the possibility of using grapevine canopy porosity as an explanatory variable to estimate grape bunches occluded by leaves, to be applied on yield estimation methods based on image analysis. Preliminary results from the 2018 season are shown.

2. MATERIAL AND METHODS

2.1. Site description and climatic conditions

The experiment was carried out during the 2018 season in two adult experimental vineyards plots located at Tapada da Ajuda, Lisboa (38°42'24,61" N; 9°11'05,53" W). In the first vineyard plot, grapevines of the white variety Encruzado were planted in 2006 with a spacing of 2.5m between and 1.0m within row, on a N-S row orientation. The vines are spur pruned on a unilateral Royat cordon and trained to a vertical shoot positioning trellis system with two pairs of movable wires. The second vineyard plot, grapevines of the red variety *Syrah*, were planted in 1998 with a plant density of 3,333 plants/ha (spacing of 2.5m between and 1.2m in row) and a N-S row orientation. The vines are spur pruned on a bilateral Royat cordon and trained to a vertical shoot positioning trellis system with two pairs of movable wires.

2.2. Image acquisition

Lateral grapevine field images were captured with a commercial camera (Nikon D5200) at the end of August 2018 when plants reached the BBCH phenological stage 85 (Lorenz *et al.*, 1995). The camera, configured in auto mode, was mounted on a tripod located approximately 2 m away from the row axis and 1 m above the ground. Images were collected with a blue background in a total of six 1 m segments (corresponding approximately to one vine) at three to five manual defoliation steps (Fig. 1). A total of 30 and 21 images were analyzed for *Encruzado* and *Syrah*, respectively.
Images of non-defoliated vines were collected. Out of these images, canopy gaps and bunches' *region of interest* (ROI) files were computed in order to estimate the projected area of these parameters. Grapevines were then defoliated, at fruiting zone, in two steps in order to create different canopy porosity levels. First, half the leaves were taken off (half defoliation) then, the remaining ones (full defoliation). Between each defoliation moment new images were obtained using the same methodology described above. Images were collected close to 11:00 a.m. from the eastern side of the canopy.

Figure 1. Images collected between defoliation steps on the variety *Encruzado*. A) non-defoliated vine. B) half-defoliated vine. C) fully-defoliated vine.

2.3. Image analysis & estimation of parameters

Images were analyzed using the ImageJ software (v1.52e, National Institutes of Health, EUA). The original images were cropped to include only the fruiting zone, which reaches from the cordon up to approximately 40 cm above the cordon. For each defoliation step, grape bunch projected area was outlined in order to estimate the corresponding number of pixels.

The percentage of visible bunches (% VB) was then calculated (Eq. 1) by dividing the visible bunch pixels (VBpx) by the total bunch pixels (TBpx), computed from the fully-defoliated vine.

$$\%\mathrm{VB} = \frac{\mathrm{VBpx}}{\mathrm{TBpx}} \times 100 \qquad \text{(Eq. 1)}$$

In the same image, canopy gaps were classified with the blue background (Fig. 1) using the Hue-Saturation-Brightness (HSB) representation of the image. The Brightness channel was ignored to maintain uniform levels of brightness across all images. The Hue and Saturation histograms were tuned in order to classify only the blue background.

The percentage of gaps (porosity; % Por) was finally calculated as the number of blue pixels classified (Gaps), divided by the total number of pixels in the image (Totalpx; Eq. 2).

$$\%\mathrm{Por} = \frac{\mathrm{Gaps}}{\mathrm{Totalpx}} \times 100 \qquad \text{(Eq. 2)}$$

A polynomial regression was then fitted using %Por as independent variable to estimate %VB. All data analysis was performed using SAS®.

3. Results & discussion

Both varieties present a similar number of bunches but large differences on bunch weight and yield with Syrah showing the lower values (Table 1).

Table 1. Summary results for yield and yield components, porosity and bunch occlusion by leaves, accessed one week before the harvest, per variety. Mean ± standard error. Data for non-defoliated vines.

Variables	*Encruzado*	*Syrah*
Number of bunches (bunches/m)	22.0 ± 1.4	20.0 ± 0.6
Average bunch weight (g)	200.3 ± 16.4	119.7 ± 8.1
Yield (kg/m)	4.4 ± 0.4	1.9 ± 0.3
Yield (ton/ha)	17.7 ± 1.5	6.3 ± 1.1
Porosity (%)	22.5 ± 3.3	16.8 ± 2.0
Bunches occluded by leaves (%)	66.7 ± 4.5	72.8 ± 6.2

In non-defoliated vines, leaves occluded up to 66.7% and 72.8% for *Encruzado* and *Syrah* varieties, respectively, showing the importance of this feature regarding bunch visibility (Table 1).

A high variation of this occlusion appears to be explained by canopy porosity as it was previously proposed for this work, showing significant relationship (R^2=0.78) in a simple linear regression analysis, for the *Encruzado* variety. To improve this relationship a third order polynomial regression was fitted with a $R^2 = 0.83$ (eq. 5; Fig. 4).

$$y = -0.0016x^3 + 0.1503x^2 - 2.2786x + 38.475 \qquad \text{(eq. 5)}$$

As for the *Syrah* variety, a simple linear regression analysis also showed significant results (R^2=0.80) which was again improved by a polynomial regression of third order that yielded a $R^2 = 0.82$ (eq. 6).

$$y = -0.0009x^3 + 0.0805x^2 - 0.3281x + 20.071 \qquad \text{(eq. 6)}$$

Porosity never reaches values above 80% even after full defoliation, because other organs remain present (mostly branches and bunches).

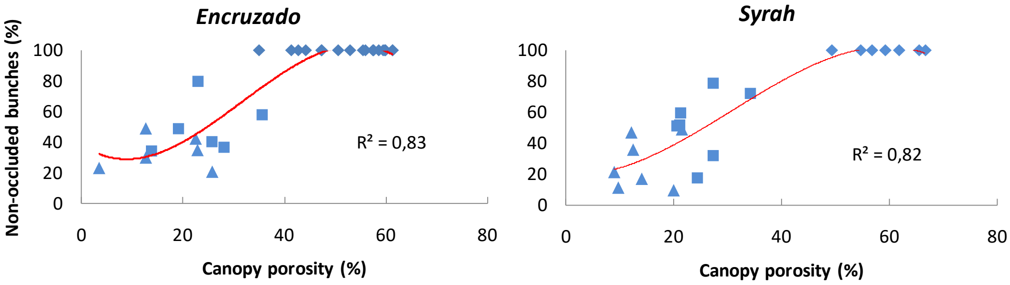


Figure 2. Polynomial regressions between non-occluded bunches and canopy porosity for the varieties *Encruzado* (n=30) and *Syrah* (n=21). Triangles, squares and diamonds represent vines that are non-defoliated, half-defoliated and fully-defoliated, respectively.

Results indicate that it is possible to estimate yield on non-defoliated vines based on canopy porosity. This information goes against what was previously stated by Aquino *et al.* (2018) claiming that the randomness of fruit exposure makes accurate yield predictions unsuitable.

4. Conclusions

A suitable way to explore yield estimation based on image analysis without recurring to invasive methods such as defoliation is explored in this work. It was firstly observed that leaves were the main vine organ occluding grape bunches and that canopy porosity has an impact on bunch exposure.

The significant relationships obtained for both varieties between canopy porosity and % exposed bunches indicates that canopy porosity is a reliable predictor for the fraction of visible bunches detected on lateral vine images taken in field conditions. Therefore, by segmenting visible bunch pixels and estimating canopy porosity it might be possible to indirectly estimate the portion of non-visible bunches.

Further research is ongoing focusing on the increase of the number of seasons and vines analyzed. Additionally, different levels of manipulated porosity are being explored on separated sets of vines in order to simulate a more broaden spectrum of field conditions. Furthermore, it was observed that a significant part of bunch occlusion is caused by neighboring bunches, a factor that is currently also being evaluated. Finally, work is ongoing to explore the prediction of occluded bunches and its viability for yield estimation models.

References

Aquino, A. *et al.* (2018) 'Automated early yield prediction in vineyards from on-the-go image acquisition', *Computers and Electronics in Agriculture*. doi: 10.1016/j.compag.2017.11.026.

De Bei, R. *et al.* (2016) 'Viticanopy: A free computer app to estimate canopy vigor and porosity for grapevine', *Sensors (Switzerland)*. doi: 10.3390/s16040585.

Besselat, B. (1987) 'Les prévisions de récolte en viticulture', 1985(Tableau 1), pp. 1–12.

Bramley, R. G. V and Hamilton, R. P. (2004) 'Understanding variability in winegrape production systems 2. Within vineyard variation in quality over several vintages', *Australian Journal Of Grape And Wine Research*, 10(1), pp. 32–45. doi: 10.1111/j.1755-0238.2004.tb00006.x.

Cunha, M., Ribeiro, H. and Abreu, I. (2016) 'Pollen-based predictive modelling of wine production: Application to an arid region', *European Journal of Agronomy*. Elsevier B.V., 73, pp. 42–54. doi: 10.1016/j.eja.2015.10.008.

Diago, M. P. *et al.* (2016) 'Assessment of vineyard canopy porosity using machine vision', *American Journal of Enology and Viticulture*. doi: 10.5344/ajev.2015.15037.

Dunn, G. M. and Martin, S. R. (2004) 'Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest', *Australian Journal of Grape and Wine Research*. doi: 10.1111/j.1755-0238.2004.tb00022.x.

Fraga, H. and Santos, J. A. (2017) 'Daily prediction of seasonal grapevine production in the Douro wine region based on favourable meteorological conditions', *Australian Journal of Grape and Wine Research*. doi: 10.1111/ajgw.12278.

Liu, S. *et al.* (2018) 'A robust automated flower estimation system for grape vines', *Biosystems Engineering*. doi: 10.1016/j.biosystemseng.2018.05.009.

Lorenz, D. H. *et al.* (1995) 'Growth Stages of the Grapevine: Phenological growth stages of the grapevine (Vitis vinifera L. ssp. vinifera)—Codes and descriptions according to the extended BBCH scale', *Australian Journal of Grape and Wine Research*, 1(2), pp. 100–103. doi: 10.1111/j.1755-0238.1995.tb00085.x.

Milella, A. *et al.* (2019) 'In-field high throughput grapevine phenotyping with a consumer-grade depth camera', *Computers and Electronics in Agriculture*. Elsevier, 156(November 2018), pp. 293–306. doi: 10.1016/j.compag.2018.11.026.

Millan, B. *et al.* (2018) 'On-the-Go Grapevine Yield Estimation Using Image Analysis and Boolean Model'. doi: 10.1155/2018/9634752.

Nuske, S. *et al.* (2014) 'Automated visual yield estimation in vineyards', in *Journal of Field Robotics*. doi: 10.1002/rob.21541.

Pérez-Zavala, R. *et al.* (2018) 'A pattern recognition strategy for visual grape bunch detection in vineyards', *Computers and Electronics in Agriculture*. Elsevier, 151(September 2017), pp. 136–149. doi: 10.1016/j.compag.2018.05.019.

Riggio, G., Fantuzzi, C. and Secchi, C. (2018) 'A Low-Cost Navigation Strategy for Yield Estimation in Vineyards', *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2200–2205. doi: 10.1109/ICRA.2018.8462839.

Seng, K. P. *et al.* (2018) 'Computer vision and machine learning for viticulture technology', *IEEE Access*. IEEE, 6, pp. 67494–67510. doi: 10.1109/ACCESS.2018.2875862.

Smart, R. E. (1987) 'Influence of light on composition and quality of grapes', *Acta Horticulturae*, 206, pp. 37–48.

Smart, R.E. and Robinson, M. (1991) 'Sunlight into Wine: A Handbook for Winegrape Canopy Management'. Winetitles, Adelaide.

Tarara, J. M. *et al.* (2014) 'Use of cordon wire tension for static and dynamic prediction of grapevine yield', *American Journal of Enology and Viticulture*, 65(4), pp. 443–452. doi: 10.5344/ajev.2014.14021.

Taylor, J. A. *et al.* (2018) 'Considerations on spatial crop load mapping', *Australian Journal of Grape and Wine Research*. doi: 10.1111/ajgw.12378.

Victorino, G., Braga, R. and Lopes, C. M. (2017) 'The effect of topography on the spatial variability of grapevine vegetative and reproductive components', pp. 510–516.

A DATAHUB FOR SEMANTIC INTEROPERABILITY IN DATA-DRIVEN INTEGRATED GREENHOUSE SYSTEMS

Jack Verhoosel[1], Barry Nouwt[1], Roos Bakker[1], Athanasios Sapounas[2] and Bart Slager[2]

[1]TNO, Department of Data Science, Soesterberg, The Netherlands

[2]TNO, Construction Innovation Centre, Delft, The Netherlands

jack.verhoosel@tno.nl, barry.nouwt@tno.nl, roos.bakker@tno.nl, athanasios.sapounas@tno.nl, bart.slager@tno.nl

ABSTRACT

This paper deals with the challenge of semantic alignment of different data sources in the horticultural sector. In this sector, greenhouses are used to grow vegetables and plants and the main goal for a greenhouse grower is to control the climate such that crop is optimally cultivated against the lowest cost. Combining available data sources to extract trends and patterns via data analysis, it is important to better support growing decisions. A Common Greenhouse Ontology (CGO) has been developed and used in a Datahub to make data sources accessible via Resource Description Framework (RDF) and a SPARQL interface on top of an Apache Jena Fuseki triplestore. The Datahub was applied in a trial use case in which three data sources where made accessible for a linear regression component that derived patterns between nutrients used and crop growth. One of the lessons learned is that the use of a common ontology very well supports the aligned use of data in analysis and thus better supports decision making.

Keywords*:* semantic alignment, ontologies, greenhouse, data analysis, decision support

1. INTRODUCTION

In the horticultural sector, greenhouses are used to grow vegetables and plants. The main goal for a greenhouse grower is to control the climate such that crop is optimally cultivated against the lowest cost. Sufficient expertise of the grower about the crop in relation to the greenhouse climate is an important prerequisite for achieving this goal. On the other hand, a greenhouse climate computer is one of the most important systems to support the grower in his decisions about defining the climate and growing strategy. However, there are also other data sources around the greenhouse that can be used to further support growers, such as weather, configuration and performance of greenhouse systems, crop growth, yield figures, fertigation strategies and labour planning. The challenge is to semantically align the data from different data sources, such that it can be jointly used to extract trends and patterns via data analysis to better support decisions taken by the grower or even to be used for automatic guidance of control systems. In a national project DDINGS (Data-Driven INtegrated Greenhouse Systems), a datahub was developed with a Common Greenhouse Ontology (CGO) as a solution to this challenge together with main greenhouse construction companies and equipment suppliers. First, some related work on horticultural ontologies is described. The remainder of the paper describes the Datahub, CGO and some application results.

1.1 Related Work on Horticultural Ontologies

In agriculture, several efforts have been made to develop ontologies for the domain. A few webportals provide search engines in a larger set of agricultural ontologies, such as AgroPortal (http://agroportal.lirmm.fr) and GODAN (https://vest.agrisemantics.org). However, they do not contain an ontology that describes the concepts in and around a greenhouse. In addition, (Rehman, 2015) gives a brief overview. AGROVOC and the Advanced Ontology Service (AOS) project was proposed by the Food and Agriculture Organization of the United Nations (FAO) for the development of agricultural ontologies based on their multilingual thesaurus as described by (Soergel et al., 2004). AGROVOC is a multilingual agricultural thesaurus and contains over 32,000 concepts in 27 languages. It comes close to an ontology and is the largest available agricultural thesaurus that is still being maintained. Smaller and older examples of ontologies are the PLANTS ontology (Goumopoulos et al., 2004), OntoCrop (Maliappis, 2009), Crop-Pest Ontology (Beck et al., 2005), Irrigation Ontology (Cornejo et al., 2005), AgriOnto (Xie et al., 2008), ONTAgri (Rehman and Shaikh, 2011) and the Crop Research Ontology (http://www.cropontology.org/). In (Roussey et al., 2013) and (Amarger et al., 2014) a small crop production ontology is described as well as an approach to use ontology design patterns for combining different ontologies into one. Most of these ontologies are either out-of-date, not maintained anymore, only partly available or simply not covering the specific greenhouse domain that is in scope. Therefore, we developed our own Common Greenhouse Ontology (CGO) to support the integration of data in and around the greenhouse. Where possible, the CGO reuses these existing ontologies that describe part of the domain.

2. The DDINGS Datahub

The DDINGS datahub connects and combines various data sources and enables data analysis for better decision support of the grower. In order to show the feasibility of our approach and the datahub, data from a trial experiment with different fertilizer recipes for a small plant crop was used. Data on water and leaf chemical analysis and crop growth parameters were measured. These data sources were mapped to the CGO and made available via our datahub. In addition, a linear regression algorithm was used to find the relation between a specific nutrient in the water and leaf chemical analysis and the crop growth.

2.1 The Datahub Concept

The type of customers of greenhouse construction company is shifting from growers to investors. Investors demand guarantees on production in a greenhouse build with certain construction concepts. Therefore, greenhouse construction companies need to shift from providing a turnkey greenhouse construction company to a service provider that supports the grower in its activities. As a result, growers, investors and construction companies want remote monitoring of the greenhouse and its connected devices and equipment. They want insight in the performance of the greenhouse in terms of yield related to growing strategies and greenhouse configuration. They need to tackle anomalies in the climate or crop in the greenhouse by combining data about growing strategies and greenhouse configuration. Moreover, they want to optimize greenhouse construction concepts using intelligence from available data with new ICT analysis techniques, like deep learning.

A lot of data is already available in separate data sources that can help to deal with these challenges. However, a Datahub is needed to connect and combine these data sources (see Figure 1). Our Datahub:

1. Saves development time/money by connecting a data source ***once*** and reuse it for ***multiple*** applications.
2. Aligns ***semantics*** of different data sources to get unambiguous meaning of information.
3. Provides a ***standardized*** interface to speed up application development on top of it.
4. Offers ***storage*** capabilities for long-term historic data capturing and comparison.

5. Enables ***data analysis*** over historic and real-time data of multiple combined data sources.
6. Secures data access via ***authentication*** and ***mandates*** for data users.

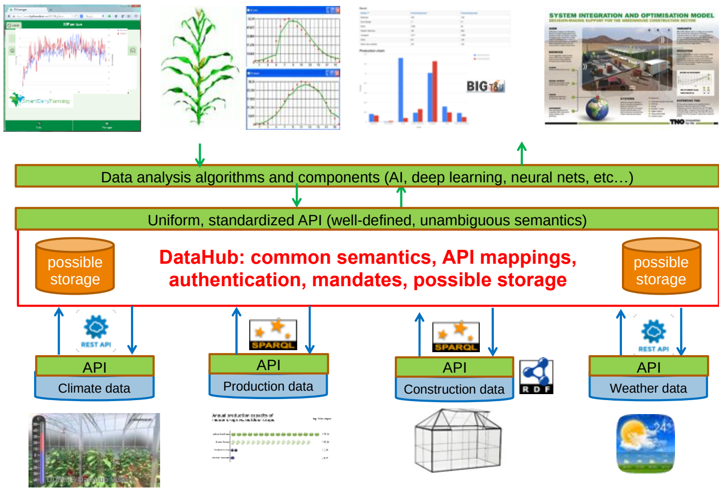


Figure 1. Sketch of the DDINGS Datahub and its context and functionality

The datahub connects relevant data sources and uses a CGO expressed in the Ontology Web Language (OWL) for semantic alignment. The CGO is used to make data accessible in generic, common terms that describe the concepts in and around the greenhouse, such as the construction and topology, the systems present, the crop, the growing system, the fertigation and nutrients strategies, possible diseases, labour planning etc. The CGO is represented using linked data technology based on RDFand is stored in an Apache Jena Fuseki triplestore that forms the main component of our Datahub. The Datahub provides a SPARQL interface that can be used to query the CGO and its concepts and relations. It contains a mapping from the CGO to the specific terms in the data sources made accessible. Finally, the datahub provides data analysis components on top of this interface, to support the grower in making better decisions using machine learning algorithms, such as linear regression, clustering and decision trees.

2.3 The Common Greenhouse Ontology

A greenhouse is a concept which in itself contains many other concepts: crops, technical systems, insects, and more. For each of these concepts one could think up an ontology, therefore existing ontologies were reused. The CGO serves two purposes: the first is to give a truthful representation of the different elements in a greenhouse, the second to support the integration of data such that our datahub can easily access the data.

Different kinds of data need to be able to integrate with the CGO: for example, the thickness of the stem of a flower, the color of a leaf, or the nutrition level of a water sample. One thing that these data types have in common is that they are somehow generated through an observation by some sensor. The semantic sensor network (SSN) ontology was used to describe sensors and their observations (Compton et al., 2012). An observation of some feature of interest, which has an observable property, is made by a sensor that implements a certain procedure to get a result. For example, when the color of a leaf is observed, the leaf is a *Feature of interest*, the color is the *Observable property*, the observer can be a human or a technical *Sensor*, who uses a color scale as a *Procedure* to determine the color. The *Result* of the observation can in this case be expressed in a numerical value and some color code. To express the results of observations, the ontology of units of measure (OM) by (Rijgersberg et al., 2013) was used. This ontology contains a class *Measure*, which is a combination of a numerical value and a measuring unit and can be seen as a subclass of the result class of the SSN ontology.

On top of these ontologies, several classes were added to describe characteristics of a greenhouse, such as the Greenhouse class itself, a *Plant* class which denotes plants and its subclass *Crop*, and

Flower. Another class that was added is the *Part* class, which requires some explanation. A lot of the data that needs to be expressed in the ontology is about a part of a whole, for instance the leaf or the stem of a flower. Also other parts of wholes in the greenhouse domain such as construction parts or system parts need to be modelled. Therefore, it is useful to gather these parts under one class instead of making many different classes. The work of (Rector et al., 2005) was used as an inspiration for the part/whole relations that were defined. In relation to the SSN ontology, the parts are possible features of interest for observations.

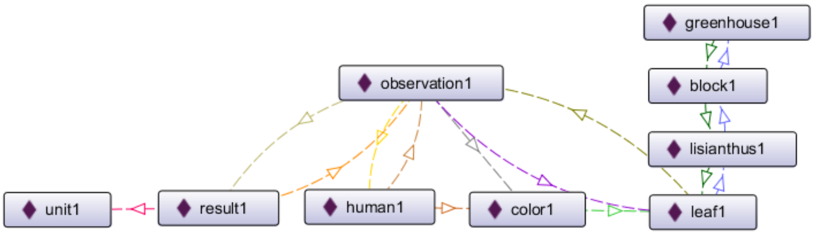


Figure 2. A High-Level Snapshot of an Instantiation of the Common Greenhouse Ontology

Data can be expressed in our CGO concepts and it is shown in Figure 2 how the color of a leaf fits in the ontology through individuals. The area on the right-hand side shows individuals in the CGO that have a part-whole relation with each other. This relation is shown with blue and green arrows. Greenhouse1 has a part (has_part) that is named block1, block1 has a part lisianthus1, which is a flower, and this flower has a part leaf1: the leaf of the flower. This leaf can then be used as a feature of interest for a certain observation: observation1. The observation is made of some property, color1, by a sensor, human1, and gives a result, result1. This result is expressed by a numerical value and a unit as defined by the OM ontology.

3. Application

3.1 Use Case Analysis

In order to show the feasibility of our Datahub, data was used from a half-year trial experiment. The experiment focused on 6 different fertilizer recipes for Lisianthus, a small plant crop. During three consecutive crop rounds of 6-8 weeks, data related to (1) water chemical analysis, (2) leaf chemical analysis and (3) crop growth parameters were measured in CSV format. This data was transformed to RDF using the LODRefine tool (https://sourceforge.net/projects/lodrefine/) and an RDF skeleton mapped to the CGO and stored it as 3 different data sources into an Apache Jena Fuseki triplestore. As shown in Figure 1, the Datahub mediates between the data sources at the bottom and the applications at the top that give earned value to the users. Data analysis of the datasets collected was done during the three consecutive crop rounds. The resulting application uses the Datahub to retrieve measurements of plant characteristics and analyse and visualize these to the user. A linear regression analysis component was used to find the relation between a specific nutrient in the water and leaf chemical analysis versus the crop growth.

Data analysis was started with consulting the growers about what plant characteristics are most important for them and what is considered to be an optimal plant with respect to, for example, stem thickness, number and colour of leaves and plant length. The result of this consultation was that in general growers prefer fast growing larger plants. Also, higher branch weight, greener leaves and faster blooming is important. The latter because it reduces the heating costs considerably. Based on this consultation, 1) analysis was done on which of the nutrient recipes produced the largest plants and 2) find the influence of the level of natrium on the growth rate. The following Python packages were used: *Numpy*, a fundamental package for scientific computing, *Pandas*, a library providing high-performance, easy-to-use data structures and data analysis tools, *Matplotlib*, a Python 2D plotting library which produces publication quality figures and *Sklearn*, a simple and efficient tool for data mining and data analysis.

During the pre-processing phase, the measurements of the crop were collected and normalized in such a way that it supported the two analysis questions mentioned above. Each row of the resulting table captures per measurement on a particular plant the crop round, the nutritional recipe, the supposed level of natrium in this recipe, the length on that particular date, the age of the plant and its average growth per day since the previous measurement. For the first phase of our analysis, only the measurements from the last week of the crop round were taken to determine the final length of the plant and correlate these with the different recipes. For the second phase, the age of the plant and the average growth per day were taken and correlated with the level of natrium in their nutritional recipe.

3.2 Results

From the first analysis that tried to determine which of the 6 different nutritional recipes produced the largest plants, correlations were found between the height of the plants and the recipe they received. The second phase of the analysis produced Figure 3.

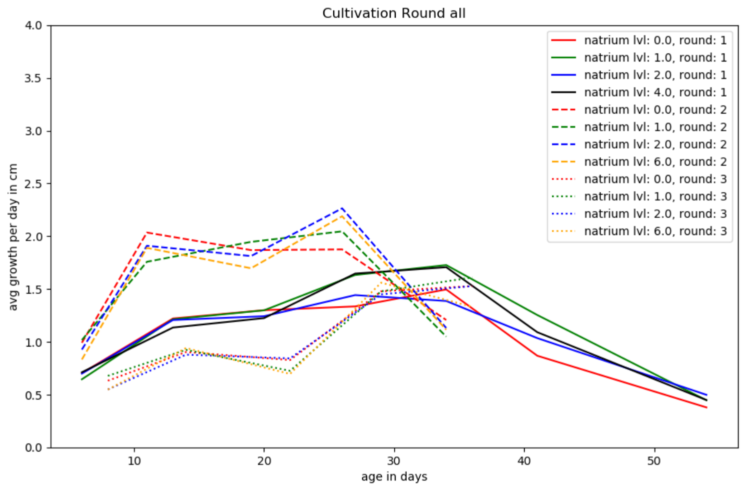


Figure 3. Data Analysis Results for the Nutrient Use case

The age, growth rate and the level of natrium the plants received are correlated. The horizontal axis represents the age in days of the plants, the vertical axis represents the average growth per day of the plant, the color of the lines represents the level of natrium in their recipe and the dash-ness of the lines represent the crop round in which the plants took part. Although higher levels of natrium did not inhibit the growth rate of the plants, as was expected, the figure very clearly shows the influence of the seasons to the growth pattern of the plants. These results have been visualised via the front-end of the datahub.

4. Conclusions

Lessons learned show that making data sources linkable and combining them in a Datahub with ontologies is feasible. A common ontology enables semantic alignment of the data in different sources towards its users. As an important result, analysis components can use this semantically aligned data unambiguously. Nevertheless, pre-processing of data towards an analysis algorithm remains an important and cumbersome task. Since it is often unclear upfront which combination of features will yield the best results, the preprocessing phase should provide a way to choose these features flexibly and this puts requirements on the technology used during preprocessing. Our conclusion is that the combination of RDF/OWL and SPARQL gives the flexibility that is necessary for a proper data analysis, while still providing normalized and cleaned data. Another conclusion can be drawn on the effects of keeping data at its source on data analysis. This yields a different approach as opposed to centralizing data for analysis. For instance, a reasoning component that orchestrates the data analysis process can be used to retrieve distributed data on request during the analysis without storing the data first in a central data source. Our Datahub is currently being extended with a few use cases that shows this behavior. Finally, a difference in modelling approaches was encountered between the domain expert and the ontology designer, as they tend to look differently at the concepts to be modelled in the ontology. For instance, a gap existed between model technical terms like 'Part' or 'Sensor' (to also include human observators) and domain terminology who rejects some of the model-technical terms. Our solution was to introduce a 'model-technical term' flag that was put on concepts in the ontology that were necessary from a modelling perspective but not well recognized by domain experts. Currently, our Datahub is also being extended with other external data sources with different API's.

References

Rehman A.U. (2015) 'Smart Agriculture: An Approach Towards Better Agriculture Management', OMICS group eBooks, https://www.esciencecentral.org/ebooks/smart-agriculture-an-approach-towards-better-agriculture-management/pdf/agricultural-ontologies.pdf, Foster City, USA.

Soergel D, Lauser B, Liang A, Fisseha F, Keizer J, et al. (2004) 'Reengineering Thesauri for New Applications: the AGROVOC Example', Journal of Digital Information 4: 1-23.

Goumopoulos C, Christopoulou E, Drossos N, Kameas A (2004) 'The Plants system: enabling mixed societies of communicating plants and artefacts', Ambient Intelligence p: 184-195.

Maliappis M.T. (2009) 'Using Agricultural Ontologies', Metadata and Semantics p: 493-498.

Beck H.W., Kim S., Hagan D. (2005) 'A Crop-Pest Ontology for Extension Publications', in 2005 EFITA/WCCA Joint Congress on IT in Agriculture, Vila Real, Portugal p: 1169-1176.

Cornejo C, Beck H.W., Haman D.Z., Zazueta F.S. (2005) 'Development and Application of an Irrigation Ontology', in Joint conference, 5th Conference of the European Federation for Information Technology in Agriculture, Food and Environment, 3rd World Congress of Computers in Agriculture and Natural Resources, Vila Real, Portugal.

Xie N., Wang W, Yang Y (2008) 'Ontology-based Agricultural Knowledge Acquisition and Application', Computer and Computing Technologies in Agriculture 1: 349-357.

Rehman A.U., Shaikh Z.A. (2011) 'ONTAgri: Scalable Service Oriented Agriculture Ontology for Precision Farming', in 2011 International Conference on Agricultural and Biosystems Engineering (ICABE 2011), Advances in Biomedical Engineering p: 411-413.

Roussey C., Chanet J-P., Cellier V., Amarger F. (2013) 'Agronomic Taxon', in Proceedings of the Second International Workshop on Open Data, BNF Paris, France.

Amarger F., Chanet J-P., Haemmerle O., Hernandez N., Roussey C. (2014) 'SKOS Sources Transformations for Ontology Engineering: Agronomical Taxonomy Use case', in 2014 Proceedings of Metadata and Semantics Research Conference (MTSR 2014), Karlsruhe, Germany, Springer, Communications in Computer and Information Science, Volume 478, pp 314-328.

Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., ... & Huang, V. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Web semantics: science, services and agents on the World Wide Web*, *17*, 25-32.

Rijgersberg, H., Van Assem, M., & Top, J. (2013). Ontology of units of measure and related concepts. *Semantic Web*, *4*(1), 3-13.

Rector, A., Welty, C., Noy, N., & Wallace, E. (2005). Simple part-whole relations in OWL Ontologies.

EARTH OBSERVATION DATA AND SPATIAL DATA SETS ANALYSIS

Pavel Simek, Jan Jarolimek, Eva Kanska, Michal Stoces, Jiri Vanek, Jan Pavlik and Alexandr Vasilenko

Department of Information Technologies, Faculty of Economics and Management, Czech University of Life Sciences Prague

simek@pef.czu.cz, jarolimek@pef.czu.cz, kanska@pef.czu.cz, stoces@pef.czu.cz, vanek@pef.czu.cz, pavlikjan@pef.czu.cz, vasilenko@pef.czu.cz

ABSTRACT

Aerial or satellite imagery allows for non-destructive remote sensing and monitoring in agriculture and related fields. The advantage of data from Sentinel missions is their availability and regular acquisition (approximately every 3 days in the Czech Republic). For effective remote sensing, there are often some limitations resulting from the properties of the acquired data – large data transmission (especially in the case of mobile use), resolution, cloudiness, sunlight reflections, etc. The research was focused on data sets from SENTINEL-2 Level-2A and higher, which are already preprocessed, especially in cartographic projection (UTM / WGS1984), creating tiles 100 x 100 km with atmospheric corrections with subsequent possibility to calculate vegetation indices, especially NDVI, GNDVI, LAI, EVI, RENDVI and MSI. Due to the size of the data, the usable datasets (area of interest and without clouds) are divided into a 5x5 km tile network grid, calculated vegetation indices and saved in datastorage. The main goal is to propose a solution for the calculation of indexes from smaller datasets, to design a prototype and to subsequently verify the solution on a pilot.

Keywords*:* Spatial data, Earth observation, remote sensing, Sentinel

1. INTRODUCTION

Aerial or satellite imagery enables non-destructive remote sensing and monitoring in agriculture and related fields, such as monitoring the condition of agricultural and forest land, water in the landscape, rural development, and offering operational use for modeling agri-technical interventions (Fülöp et al., 2015). Remote sensing of optical and radar data can help map out crop types and estimate biophysical parameters, especially with the availability of an unprecedented amount of free SENTINEL satellite data within the Copernicus program (Veloso et al., 2017). Especially the SENTINEL-2 multispectral imaging is of great importance to agriculture (Wang et al., 2018).

The use of SENTINEL and Landsat satellites can be divided into use in crop and livestock production. However, these two directions are very closely linked. Images can be used for modeling and calculating spectral indexes, and they can just as well be used for visualization in the form of color synthesis, whether in the visible spectrum or in other wavelength ranges. Longer mapping of land use / cover can also be automated, for example in the field of monitoring forested areas (Szostak et al., 2019). Monitoring of vegetation cover in the winter months is also of great importance (Denize et al., 2019). Before processing, satellite images must be pre-processed using atmospheric corrections, resampling, and spatial cropping (Rodriguez-Ramirez et al., 2019). Also, many projects that tried to integrate various data of various formats were implemented recently (Řezník et al., 2015).

A suitably chosen color synthesis may indicate plant stress or soil erosion by a different color shade than is normal for healthy and prosperous growth. The most recently usable indices in plant production include, in particular, the Normalized Difference Vegetation Index - NDVI, Enhanced Vegetation Index - EVI (Halabuk et al., 2015), but also the Red Edge Normalized Difference Vegetation Index - RENDVI (Deng et al., 2018), Green Normalized Difference Vegetation Index - GNDVI, Moisture Stress Index – MSI or Leaf Area Index - LAI.

The original images of SENTINEL missions may be too large or even unusable to process and calculate the indexes (e.g. due to clouds). The main goal is to propose a solution for the calculation of indexes from smaller datasets, to design a prototype and subsequently to verify the solution on a pilot.

2. Methodology

The basic scientific method of analysis and synthesis was used to tackle the research problem. Sentinel satellites are made up of several types of satellites, usually discharged in pairs to provide faster data recovery. Currently, the satellites of these missions are two SENTINEL-1 satellites designed for radar mapping of the Earth's surface and the detection of information available through their sensors, two SENTINEL-2 satellites that provide image data across multiple spectral bands, two SENTINEL-3 satellites and one SENTINEL-5P satellite.

The research was focused mainly on data from the SENTINEL-2 mission. To ensure potential use of SENTINEL-2 data in time series, data must be geometrically registered and radiometrically corrected. Both procedures are the basic operation in Level-1 data level processing. Level 2 includes atmospheric corrections. Level-3 processing creates space-time synthesis from “Bottom of Atmosphere” images of SENTINEL-2 Level-2A. One important step is resampling images into a cartographic projection (UTM / WGS1984), creating 100 x 100 km tiles, and calculating cloud masks, land surface and water surfaces (Level-1C).

A SENTINEL-2 product refers to a directory folder that contains a collection of information. It includes: a manifest.safe file which holds the general product information in XML, a preview image in JPEG2000 format, subfolders for measurement datasets including image data (granules/tiles) in GML-JPEG2000 format, subfolders for datastrip level information, a subfolder with auxiliary data (e.g. International Earth Rotation & Reference Systems (IERS) bulletin) and HTML previews.

The Level-2 product is also in SAFE format, which groups together several types of file: metadata file (XML file), preview image (JPEG2000 with GML geo-location), tiles files with BOA reflectances image data file (GML / JPEG2000) for each tile, datastrip files, auxiliary data and ancillary data (Ground Image Processing Parameters (GIPPs)) (ESA).

Default dataset name from the SENTINEL-2 mission looks like:

```
MMM_MSIL1C_YYYYMMDDHHMMSS_Nxxyy_ROOO_Txxxxx_<Product Discriminator>.SAFE
```

where:

- MMM: is the mission ID (S2A/S2B)
- MSIL1C: denotes the Level-1C product level
- YYYYMMDDHHMMSS: the data sensing start time
- Nxxyy: is the Processing Baseline number (e.g. N0204)
- ROOO: is the Relative Orbit number (R001 - R143)
- Txxxxx: is the Tile Number
- SAFE: Product Format (Standard Archive Format for Europe) (ESA)

The processing of remote sensing image data and the interpretation of image information involves several sequential digital image analysis procedures. Lillesand et al. (2008) describes several basic steps in digital image processing; for the purposes of this research, the following are the most important:

- Image Preprocessing - targets image correction for the most accurate representation of the original scene. Includes correction of geometric, radiometric and atmospheric distortion and image noise. Preprocessing procedures depend on the characteristics of the sensor used to capture the image.
- Image Enhancement - represents methods of effective image representation for its further processing or visual interpretation. This creates a new image with increased information content. It involves manipulation of the color component of the image or contrast, or, for example, the use of spectral bands to highlight vegetation in the form of vegetation indices, etc.

For the pilot Image Enhancement software for downloading selected images of SENTINEL-2 Level-2 mission and the subsequent calculation of indexes were written in C ++. For the processing itself a high performance computer with two processors AMD EPYC 7601 32 core, 2.2 GHz and 1.4 TB RAM, OS Windows 10 was used.

3. Results

Sentinel mission data are effectively used in the agrarian sector. Each data set represents the image data of a surface sub-area of the Earth. However, it is necessary to reckon with a large production of data, especially during their continuous renewal. The data are available through the Sentinel organizations' web presentations. The basic idea is slightly different, given by local customs and view of open data, but the data are in accessible form that allows automatic download and subsequent machine processing. All datasets are available via https://scihub.copernicus.eu/.

The current state of calculations within existing datasets assumes work with selected sectors recorded by the SENTINEL-2 scan tool. The tile, which is then available at the Science Open Hub, has the dimensions of 100 x 100 km, which requires considerable computing. The way to avoid this is to divide the tile into sub-units (making a grid of smaller squares). Due to the optimum ratio of computational complexity and size of the displayed area, the tiles should have a size of 5 x 5 km.

Each tile has even partial addressing - the columns are marked with letters and lines with numbers as a grid. Thus, the grid will have the origin A1 (the first column marked as A and the first row marked as 1) and the end T20 (the last 20th column marked as T and the last 20th row marked as 20). For addressing was used ASCII characters. This makes it possible to use the current API, which is based on string search in the name of data sets. Before that, however, it is still necessary to perform a cloud analysis to ascertain whether the image is usable or not, and the source data is recalculated to a pixel resolution of 10 m. Subsequently, the tiles that are meaningful for the calculation of vegetation indices are selected, e.g. without clouds, correct location (in the Czech Republic), and so on.

Figure 1. The procedure of calculating indexes (starting from Open Access data download)

Accessing the Copernicus repository at national level is often accomplished with the same software as the ESA Open Access Hub. When using a national repository, it is not possible to influence selections and work with data other than displaying currently published datasets. These can be displayed based on the selected satellite, selected tool (sensor), time stamp and polygon (location of location within data sets).

Among the most widely used spectral indices in crop production are: NDVI, GNDVI, LAI, EVI, RENDVI and MSI. For index calculations, new datasets are uploaded to the repository that respect the following naming rules:

Original (default) dataset name:

`S2B_MSIL2A_20190330T101029_N0211_R022_T33UVR_20190330T144328`

New dataset name:

`T1_S2B_MSIL2A_20190330T101029_N0211_R022_T33UVR_20190330T144328`

where the identification of the grid tile of the dataset was added - here the square T1 - the twentieth column and the first row

After calculating the index, its name is reflected in the dataset naming, e.g.

`NDVI_T1_S2B_MSIL2A_20190330T101029_N0211_R022_T33UVR_20190330T144328`

As can be seen from the example, the name reflects that it is NDVI from T1 tile (5x5 km) inside the 33UVR tile (100x100 km).

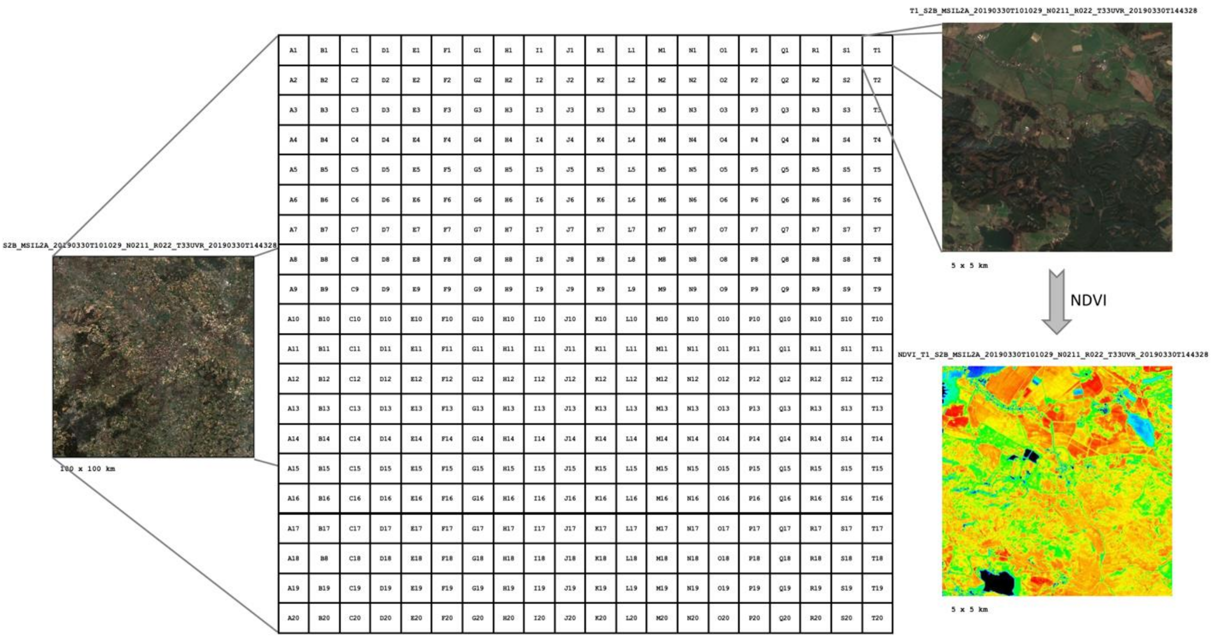


Figure 2. Layout of 100x100 km frame divided into 5x5 km tiles

4. Discussion

The advantage of data from SENTINEL-2 is their availability and regular acquisition (approximately every 3 days in the Czech Republic). To work with remote sensing data, there are often some limitations that result from the properties of the acquired data – large data transmission (especially in the case of mobile use), resolution, cloudiness, sunlight reflections, etc. For the analysis of cloudiness, it is necessary to create a cloudiness mask from SENTINEL-2 data. This is already provided at level 1C, and is subject to validation (Baetens et al., 2019) but even then, cloud classification can be difficult (Lie et al., 2019).

Figure 3. Example of cloud mask cirrus, source: ESA, 2019.

Another source for the calculation of vegetation indices can be data sets from the Landsat-8 program, or it is possible to combine both data sets, which can be beneficial, for example, in longer term forest monitoring (Arekhi et al., 2019) or landscape coverage (Carrasco et al., 2019).

The big limit is the addition of indices requiring heavy calculations that cannot be realized shortly after the release of the new dataset. These are mainly indices that require resampling. Thanks to the current cloud computing capabilities, this limit can be successfully overcome using?? applications for processing remote sensing images significantly reduce processing time (Defourny et al., 2019).

Images for processing and indices computing can also be captured with UAV. In this case, the images are at a much higher resolution, but at a much higher price. However, for certain crops, SENTINEL-2 images produce some distortion (Khaliq et al., 2019).

5. Conclusions

The advantage of data sets from the SENTINEL-2 mission of the Copernicus program is their availability, regularity of new images and their preprocessing into cartographic projection (UTM / WGS1984), creation of tiles 100 x 100 km, calculation of cloud mask, earth surface and water surfaces and atmospheric corrections (Level-2A). The disadvantage can be its spatial resolution, size of files and especially cloudiness.

The research was focused on the possibility of automated SENTINEL-2 data usage, with subsequent possibility of calculation of vegetation indices and other indices, especially NDVI, GNDVI, LAI, EVI, RENDVI and MSI. Due to the size of the data, the usable datasets (the area of interest) are divided into a 5x5 km grid network labeled A1 - T20 (400 tiles from each 100x100 km grid square). Before the calculation of the vegetative indices, the area to be examined will be selected and the cloud will be analyzed (whether the calculation from the given tile makes sense). Subsequently, from much smaller tiles, the desired vegetation indexes are calculated (smaller tiles are faster and easier to process). The whole project is in the phase of the proposed prototype and preparations are currently underway to test a functional prototype with a focus on data and computing complexity.

Acknowledgements

This article was created with the support of the Internal Grant Agency (IGA) of FEM CULS in Prague, no. 20181010 – „Identification of limitations in processing of remote sensing data and in their efficient use“.

References

Arekhi, M., Goksel, C., Sanli, F. B. and Senel, G. (2019) ‘Comparative Evaluation of the Spectral and Spatial Consistency of SENTINEL-2 and Landsat-8 OLI Data for Igneada Longos Forest’, ISPRS International Journal of Geo-information, vol. 8, issue 2, article no. 56. doi: 10.3390/ijgi8020056.

Baetens, L., Desjardins, C. and Hagolle, O. (2019) ‘Validation of Copernicus SENTINEL-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure’, Remote Sensing, vol. 11, issue 4, article no. 433, EISSN 2072-4292. doi: 10.3390/rs11040433.

Carrasco, L., O'Neil, A. W., Morton, R. D. and Rowland, C. S. (2019) ‘Evaluating Combinations of Temporally Aggregated Sentinel-1, SENTINEL-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine’, Remote Sensing, vol. 11, issue 3, article no. 288. doi: 10.3390/rs11030288.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L,. Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Begue, A., Dejoux, J. F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P. S., Traore, S. S. and Koetz, B. (2019) ‘Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world’, Remote Sensing of Environment, vol. 221, pp. 551-568. doi: 10.1016/j.rse.2018.11.007.

Denize, J., Hubert-Moy, L., Betbeder, J., Corgne, S., Baudry, J. and Pottier, E. (2019) 'Evaluation of Using Sentinel-1 and-2 Time-Series to Identify Winter Land Use in Agricultural Landscapes', Remote Sensing, vol. 11, issue 1, article no. 37. doi: 10.3390/rs11010037.

Deng, L., Mao, ZH., Li, XJ., Hu, ZW., Duan, FZ and Yan, YN. (2018) 'UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras', ISPRS Journal of Photogrammetr and Remote Sensing, vol. 146, pp. 124-136. doi: 10.1016/j.isprsjprs.2018.09.008.

European Space Agency. 'Data Formats' [on-line] Available at https://sentinels.copernicus.eu/web/sentinel/user-guides/SENTINEL-2-msi/data-formats.

European Space Agency. 'Naming Convention' [on-line] Available at https://sentinels.copernicus.eu/web/sentinel/user-guides/SENTINEL-2-msi/naming-convention.

Fülöp, G., Sárközi, E. and Szigeti, I. S. (2015) 'Applying earth observation data for detection and estimation of woody biomass volumes of agricultural areas accessible for energy production', Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment, HAICTA 2015. CEUR Workshop Proceedings, vol. 1498, pp 64-67. ISSN 16130073.

Halabuk, A., Mojses, M., Halabuk, M and David, S. (2015) 'Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series', Remote Sensing, vol. 7, issue 5, pp. 6107-6132. doi: 10.3390/rs70506107.

Lillesand, T. M., Kiefer, R. W. andChipman, J. W. (2008) 'Remote sensing and image interpretation', 6th issue, Hoboken, NJ: John Wiley and Sons. ISBN-10: 0470052457, ISBN-13: 978-0470052457.

Liu, C-C., Zhang, Y-C., Chen, P-Y., Lai, C-C., Chen, Y-H,, Cheng, J-H. and Ko, M-H. (2019) 'Clouds Classification from SENTINEL-2 Imagery with Deep Residual Learning and Semantic Image Segmentation', Remote Sensing, vol. 11, issue 2, article no. 119. doi: 10.3390/rs11020119.

Khaliq, A., Comba, L., Biglia, A., Aimonino, D. R., Chiaberge, M. and Gay, P. (2019) 'Comparison of Satellite and UAV-Based Multispectral Imagery for Vineyard Variability Assessment', Remote Sensing, vol. 11, issue 4, article no 436. EISSN 2072-4292. doi: 10.3390/rs11040436.

Reznik, T., Lukas, V., Charvat, K., Horakova, S and Charvat jr., K. (2015) 'Towards Farm-Oriented Open Data in Europe: the Scope and Pilots of the European Project "FOODIE"', Agris on-line Papers in Economics and Informatics, vol. 8, issue 1, pp. 51-58. ISSN 1804-1930.

Rodriguez-Ramirez, R., Sanchez, M. G., Rivera-Caicedo, J. P., Fajardo-Delgado, D and Avila-George, H. (2019) 'Automating an Image Processing Chain of the SENTINEL-2 Satellite', Proceedings of the Trends and Applications in Software Engineering, Books series: Advances in Intelligent Systems and Computing, vol. 865, pp. 216-334. doi: 10.1007/978-3-030-01171-0_20.

Szostak, M., Knapik, K., Likus-Cieslik, J., Wezyk, P and Pietrzykowski, M. (2019) 'Monitoring the spatial range of forested areas in the reclaimed sites using SENTINEL-2 images', Sylwan, vol. 163, issue 1, pp 55-61. ISSN 0039-7660.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F. and Ceschia, E. (2017) 'Understanding the temporal behavior of crops using Sentinel-1 and SENTINEL-2-like data for agricultural applications', Remote Sensing of Environment, vol. 199, pp 415-426. ISSN 00344257. DOI 10.1016/j.rse.2017.07.015.

Wang, B., Jia, K., Liang, SL., Xie, XH., Wei, XQ., Zhao, X., Yao, YJ and Zhang, XT. (2018) 'Assessment of SENTINEL-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover', Remote Sensing, vol. 10, issue 12, article no. 1927. doi: 10.3390/rs10121927.

IMAGE EXTRACTION BASED ON DEPTH INFORMATION FOR CALF BODY WEIGHT ESTIMATION

Naoki Fukuda[1], Takenao Ohkawa[1] , Chikara Ohta[2] , Kenji Oyama[3], Yumi Takaki[1] and Ryo Nishide[4]

[1]Graduate School of System Informatics, Kobe University, Japan

[2]Graduate School of Science, Technology and Innovation, Kobe University, Japan

[3]Food Resources Education and Research Center, Kobe University, Japan

[4]The Center for Data Science Education and Research, Shiga University, Japan

i_l_i123@cs25.scitec.kobe-u.ac.jp, ohkawa@kobe-u.ac.jp, ohta@port.kobe-u.ac.jp, yumi@people.kobe-u.ac.jp, oyama@kobe-u.ac.jp, ryo-nishide@biwako.shiga-u.ac.jp

ABSTRACT

This paper aims to facilitate body weight estimation by using calf's images taken in a loose barn. In our method, the procedures from image extraction of calves to the resultant body weight estimation are automated. The images with a single calf are used for body weight estimation. Most of the images are unusable, as several or none of the calves are included in the images, otherwise, the body parts are not properly extracted due to the calf's posture. In this paper, we propose a method to select only the images appropriate for body weight estimation. First, the information such as the calf's posture, body information and the angle of the calf to the camera are obtained. Then, this information is examined based on a certain threshold to extract only the appropriate images. Depth images are used because they are less affected by the surrounding environments and are considered useful for extracting calf area. The calf area is extracted by using background subtraction with a depth image. The images which meet all criteria are chosen as appropriate images for body weight estimation. Efficiency of the proposed automated method and manual work are compared by MAPE (Mean Absolute Percentage Error) of estimated calf weight. The MAPE by using manually-selected images was 10.34% and that by using proposed method was 13.94%, which yields the difference of 3.6%. From this result, we confirmed that the proposed method for automatically selecting appropriate images for body weight estimation can fairly perform as well as manual selection and can be effective to reduce human effort.

Keywords*:* Cattle, Calf, Weight estimation, Image processing, Depth camera**.**

1. INTRODUCTION

In general, the calving interval is often used to examine the fertility of beef cattle. Some of the cattle, however, cannot produce healthy and growing calves effectively nor cannot foster calves appropriately. In order to examine the fertility of beef cattle, maternal abilities are also indispensable other than calving interval, and thus, the calf's weight has been used as an indicator of maternal ability for the past decades (Meyer, 1992). Measuring the calf's weight manually, however, burdens farmers. Various studies have been conducted on methods for body weight estimation (Song et al., 2018, Ogata et al., 2011).

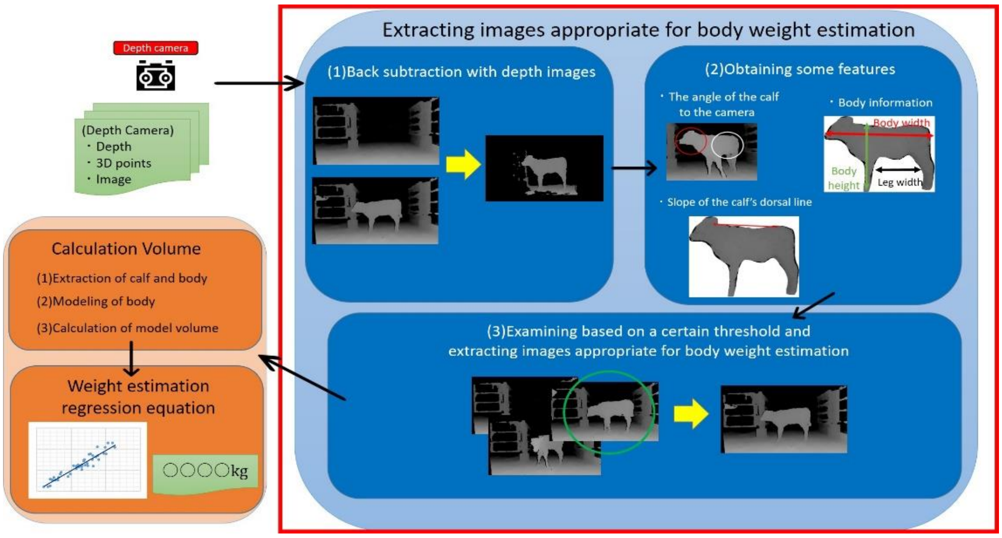


Figure 1. Flow of body weight estimation

We aim to facilitate body weight estimation by proposing a method for body weight estimation by using calf's images taken in a loose barn. In our method, the procedures from image extraction of calves to the resultant body weight estimation are automated. The images with a single calf are used for body weight estimation. Most of the images are unusable, as several or none of the calves are included in the image, otherwise, the body parts are not properly extracted correctly due to the calf's posture. The proposal method utilizes only the images appropriate for body weight estimation by obtaining the information such as the calf's posture, body information and the angle of the calf to the camera and examining them based on a certain threshold.

The rest of this paper is organized as follows. In Section 2, we introduce the flow of body weight estimation and propose the method to select only the images appropriate for body weight estimation. In Section 3, we conduct an experiment applying the proposed method. In Section 4, the effectiveness of the proposed method is evaluated based on the experimental results. Section 5 concludes the paper with some future research directions.

2. METHODOLOGY

2.1 Overview of Image Selection

Fig. 1 shows the flow of body weight estimation system by using calf's images. Body weight estimation consists of three parts; (1) Extracting of calf and body, (2) Modeling of body, and (3) Calculation of model volume (Yamashita et al., 2018, Nishide et al.,2018). Calf's body width, body height and leg width for extracting of calf and body are used. The authors assume that the calf looks sideways to the camera, and the calf's spine is on a perpendicular line to the camera (Left picture of Fig. 2) by using the fact that calf's body is symmetrical around its spine. This line is used for modeling the body, however, the modeling is affected in case that calf's position is oblique to the camera (e.g. Right picture of Fig. 2). For these reasons, only the images appropriate for body weight estimation are selected by using calf's body information and the angle of the calf to the camera (Red frame of Fig. 1).

In a proposed method, the calf area is initially extracted by using background subtraction with depth images. Second, the angle of the calf's position to the camera, calf's body information and the slope of the calf's dorsal line from the root of neck to the hip by using the calf area are obtained. Finally, the images which meet all criteria are chosen as appropriate images for body weight estimation.

2.2 Extracting Calf's Area with Background Subtraction

In this paper, background subtraction is used for extracting a calf, and depth images are used because they are less affected by the surrounding environments and are considered useful for extracting calf area (Fernandez-Sanchez et al., 2013). By using background subtraction, the differences between depth values of calf's input image and a background image are obtained, and calf's area are extracted by comparing with a certain threshold. Fig. 3 shows an example image obtained with background subtraction.

2.3 Calf's Body Information

Horizontal direction is defined as x axis and the vertical direction as y axis as shown in Fig. 4. The longest part of a calf's body in x direction is the part from tip of nose to buttocks defined as *cow_width.* The longest part of a calf's body in y direction is the part from backbone to phalange defined as *cow_height.* When taking pictures of calves, calves are not always standing still but they are sometimes opening their legs or walking. In that case, a ratio between legs width and body width or body height change more drastically than usual. Therefore, we detect both legs and define the differences of both x coordinates as *leg_width*.

2.4 The Angle of Calf Image Shot from Camera

The images used for analysis are taken from the side of calf, however, calf do not always appear perpendicularly to the camera as shown in right picture of Fig. 2. Therefore, oblique calves shot from the camera was detected by using differences between depth values around head and buttocks as shown in Fig. 5.

2.5 The Slope of Calf's Dorsal Line

Calves sometimes raise their buttocks or put their head down as show in Fig. 6. In such cases, calf's body length or body height changes drastically, as the calf's dorsal line from the root of neck to buttocks becomes larger than usual. Therefore, in the proposed method, we also calculate the slope of the calf's dorsal line.

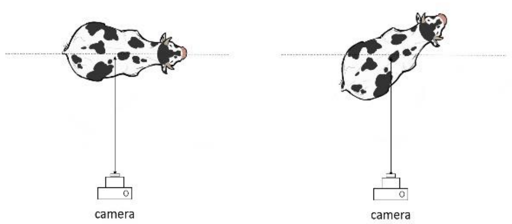


Figure 2. The angle of the calf to the camera

Figure 3. Image obtained with background subtraction

Figure 4. Definition of coordinate system

Figure 5. Example of a calf tilt to the camera

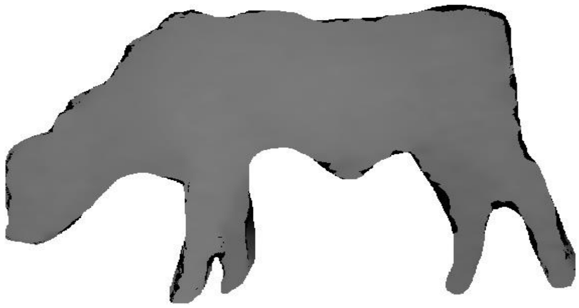


Figure 6. Example of calf lowering its head

3. EXPERIMENT

3.1 Data Set

We conducted experiments on black calves bred at the Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University, from newborn to the ones weighing under 100 kg. The camera used for taking pictures is RealSense Depth Camera D415[1] manufactured by Intel. The camera was set to shoot toward the place where only calves walk. When taking pictures, calves were conducted to the front of camera manually by the staff. Data to be used are images of 14 calves in total, which were taken on February 4, 8 and 11, 2019. The number of images is about 40,000. They are all taken within these three days, and include a lot of unstable ones which none of the calves are included in. Only the images appropriate for body weight estimation are selected by using our proposed method. Later, we calculate correlation coefficient and linear regression equation between measured body weight and estimated body weight. When we obtain multiple images of the same calf in same day, we average the result of estimated volume. In addition, we conduct leave-one-out cross validation using obtained data points and calculate estimated weight $\mathrm{W_i^{estimate}}$ from estimated volume V_i. Finally, we calculate error e_i between $\mathrm{W}_i^{\mathrm{estimate}}$ and measured weight $\mathrm{W}_i^{\mathrm{actually}}$, and show the average of error (Mean Absolute Percentage Error MAPE).

$$e_i = \frac{\left|W_i^{estimate} - W_i^{actually}\right|}{W_i^{actually}} \times 100 \qquad (3.1)$$

In the experiment, images are selected automatically by using our proposed method and body volume is estimated by using the selected images. Only circle fitting is used when modeling calf body (Yamashita et al., 2018, Nishide et al., 2018). We also estimate body volume using manually-selected images, and compare MAPE when using manually-selected images and the one when using auto-selected images.

3.2 Estimation Result of Volume and Linear Regression Equation

As a result of selecting images from 40,000 by using a proposed method, we obtained 62 usable images. Left graph of Fig. 7 shows the result of body volume estimation using circle fitting. The horizontal axis of the graph is the volume of the obtained model and the vertical axis is the measured weight, as mentioned above.

As a result of selecting images to 40000 ones manually, we obtained 49 ones. Right of Fig. 7 shows the graph of result of body volume estimation using circle fitting. The horizontal axis of the graph is the volume of the obtained model and the vertical axis is the measured weight, as mentioned above.

We conduct leave-one-out cross validation with estimated body volume and measured weight, and take the average of errors for MAPE. We calculate that in each graph. Table 1 shows obtained MAPE and Correlation Coefficient from each graph.

---

[1]https://www.mouser.com/pdfdocs/Intel_D400_Series_Datasheet.pdf

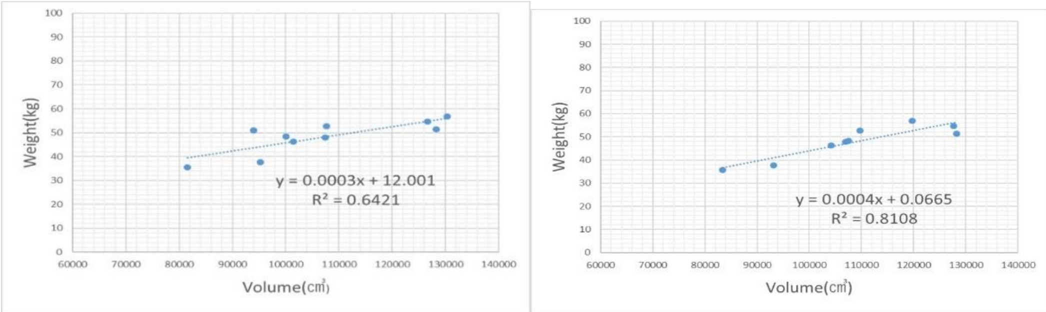


Figure 7. Result of volume calculation (Left: Auto-selected, Right: Manually-selected)

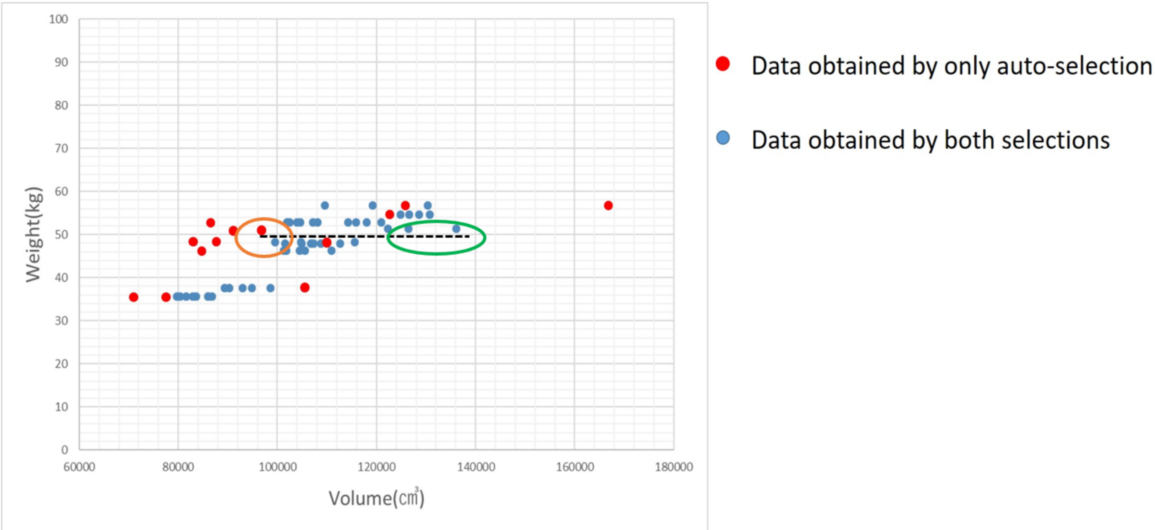


Figure 8. Result of non-averaged volume calculation

Table 1. Mean Absolute Percentage Error(MAPE) and Correlation Coefficient

	MAPE(%)	Correlation Coefficient
Auto-selected	13.94	0.6421
Manually-selected	10.34	0.8108

4. DISCUSSION

Table 1 shows that MAPE obtained when using auto-selected images is greater than the one obtained when using manually-selected images by a slight difference of 3.6%. However, because manually-selected images are selected by manpower, that method can be considered applicable in an ideal environment. Avoiding the effort of manual selection, the proposed method can be a possible solution for automatically selecting images appropriate for body weight estimation. Besides, all 49 manually-selected images are contained in 62 auto-selected images. Therefore, the proposed method has been performed well. However, inappropriate images were extracted by using the proposed method. Hence, inappropriate images must be eliminated. Moreover, the reason why the proposed method did not perform well is mentioned as follows Fig 8. shows all non-averaged data, obtained by both methods. Red points are obtained by only auto-selection and blue ones are obtained by both selections, auto-selection and manually-selection. If the value of weight is equal, then those points refer to the same calf. There are some points which have smaller volume compared with other points with same amount of weight. For example, points circled in orange or green, and a drawn broken line, have same value of weight. However, points circled in orange have smaller volume compared with points circled in green. It is considered that some images which the calf was oblique to the camera were extracted incorrectly. Therefore, width was detected to be shorter than usual, and calf was detected to have a smaller body than usual.

5. Conclusions

In this study, we proposed a method for selecting the images appropriate for body weight estimation by taking calf's pictures with a depth camera and obtaining information from calf's area. Besides, we compared the result of body weight estimation with images obtained by using the proposed method and with the ones obtained manually and verified the precision of the proposed method. As a result, we could at least show the effectiveness of the proposed method.

In the experiment, we could select images appropriate for body weight estimation, but on the other hand, a few images unsuitable for body weight estimation were selected. Thus, we should get rid of the images unsuitable for body weight estimation by using machine learning with auto-selected images or using outlier detection. We plan to further the research on body weight estimation practically by developing the proposed method, so that we don't have to force the calves to stand still in front of the camera.

Acknowledgement

This work was supported by the subsidized projects of Japan Livestock Technology Association, the Council for Information Utilization of Beef Cattle Improvement, the Ministry of Agriculture, Forestry and Fisheries, and JST CREST Grant Number JPMJCR1682, Japan. Note that any implications in this work are not their official opinions.

References

Meyer, K. (1992) 'Variance components due to direct and maternal effects for growth traits of Australian beef cattle', Livestock Production Science, Vol.31, pp.179-204.

Song, X., Bokkers, E. A. M., Tol, P.P.J. van der., Groot Koerkamp, P. W. G. and Mourik, S. van. (2018) 'Automated body weight prediction of dairy cows using 3-dimensional vision', Journal of Daily Science, Volume 101, Issue 5, May, pp.4448-4459.

Ogata, M., Nakahashi, Y., Shibuya, R., and Kuchida, K. (2011) 'Possibility and its precision of body weight prediction by image analysis method for Japanese Black calves', Research Bulletin Obihiro University, 32, 14–19.

Yamashita, A., Ohkawa, T., Oyama, K., Ohta, C., Nishide, R. and Honda, T. (2018) 'Calf Weight Estimation with Stereo Camera Using Three-Dimensional Successive Cylindrical Model', Journal of the Institute of Industrial Applications Engineers, Vol.6, pp.39-46.

Nishide, R., Yamashita, A., Takaki, Y., Ohta, C., Oyama, K. and Ohkawa, T. (2018) 'Calf Robust Weight Estimation Using 3D Contiguous Cylindrical Model and Directional Orientation from Stereo Images', SoICT '18, December 6-7, Da Nang City, Viet Nam.

Fernandez-Sanchez, E. J., Diaz, J., and Ros, E. (2013) 'Background Subtraction Based on Color and Depth Using Active Sensors', Sensors, 13, 8895–8915.

BIG DATA MANAGEMENT TOOLS FOR SMART FARMING DATA

Michal Stočes, Pavel Šimek, Edita Šilerová, Jan Masner and Jan Jarolímek
Department of Information Technologies of Faculty of Economics and Management at Czech University of Life Sciences Prague, Czech Republic
stoces@pef.czu.cz, simek@pef.czu.cz, silerova@pef.czu.cz , masner@pef.czu.cz, jarolimek@pef.czu.cz

ABSTRACT

This paper deals with the issue of role of middleware in the process of data transformation, aggregation, storage and analysis. Data used in the concept of precision agriculture is very diverse. Not only in terms of sources but also formats. Farms use small sensor data as well as large files from satellite systems for various analyzes. Selected middleware services will be evaluated using a multiple-criteria decision analysis.

In agriculture, the collected data need to be continuously analyzed and worked with. Data processing procedures, especially with regard to Big Data, are not yet properly addressed in agriculture. As the volume of data collected grows, the demands for efficient storage grows as well. It is necessary to deal with this issue. The farmer's data sources can be divided between the data acquired by the farm from its own internal, private data source and data obtained externally. External data can be used from public open data databases or purchased.

On the basis of a multiple-criteria decision analysis of the weighted sum method, key criteria, selected weights for the use of a small start-up company, ThingsBoard was released as a compromise option with the highest value of 0.9. The second was DeviceHive with 0.833 and Mainflux, WSo2 IoT and Thinger.io.

Keywords: middleware, internet of things, big data, sensor data, data transformation, data aggregation, data store, data analytics

1. INTRODUCTION

The paper focuses on the issue of middleware as a tool for managing data from smart agriculture. The paper continues on previous authors research in the field of IoT, Big data and smart agriculture. (Stočes et al., 2018) and (Stočes et al., 2018).

1.1 Internet of Things and Middleware

Atzori et al. (2013) divided the Internet of Things paradigm into three groups in his work. The first group, such as Internet-based Vision, which includes web stuff, Internet 0 and IP (Internet Protocol address) for smart devices, is just connecting the device to the network. The second group is a semantically oriented vision that can be summed up as knowledge of working with information. This group includes data analysis and thus the middleware discussed in this work. The intersection of these two groups is then smart connected middleware. The last group is a device-oriented vision that includes the devices themselves. The intersection of all three groups is then the Internet of Things. The whole scheme is shown in Figure 1 below (Dastjerdi et al., 2016).

Middleware is a layer or set of sub-layers of software between different levels of an application. The aim of middleware is to hide details of various technologies, protocols, network environments (Stočes et al., 2018), duplicate data, as a higher goal is to shield the programmer from problems that are not directly relevant to its scope and mission-specific development (Jarolímek et al., 2019). In addition, middleware masks the heterogeneity of computer architectures, operating systems, programming languages, and network technologies to facilitate application programming and management (Stočes et al., 2019). The reasons for using middleware in IoT are much from the security discussed in the later chapter, as well as its abstraction from the devices themselves (Kumarage et al., 2016). The middleware discussed in this work is not an abstraction of the devices themselves, their I / O systems, or the abstraction of networking (Cimino et al., 2019).

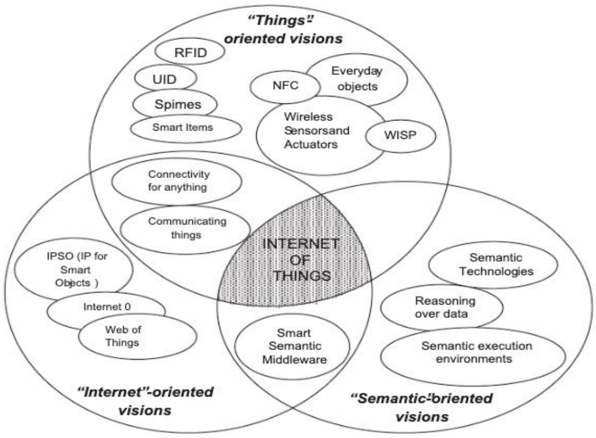


Figure 1. Internet of things (Atzori et al., 2010)

1.2. Middleware architectural designs

Middleware can be divided (Razzaque et al., 2016) into seven groups according to their design: event driven, service oriented, virtual machine driven, agent driven, tuple space paradigm, database oriented, and application specific. Some middleware is a combination of several groups that take advantage of selected groups. The three most used and basic ones from which others can be derived will be described. Event-based middleware, where middleware plays an active role in the data flow and sends the data itself to the application layer. Service-oriented middleware where it is passive and the data provides and exposes the API (Application Programming Interface). The last described pattern will be a virtual machine-controlled middleware in which the middleware is passive in terms of application layer and data only provides, but active against the physical layer in which it deploys virtual machines (Liangzhao et al., 2004.).

2. Methodology

2.1 Middleware evaluation

Middleware services were selected, individual solutions for specific use cases were evaluated. The main case is a small farm that wants to penetrate the smart farming industry and collect data from them (10 data user). They are interested in further developing this project and have the capacity of developers to develop their own solutions. Middleware services will be evaluated using a multiple-criteria decision analysis (MCDA) (Fiala et al., 1994), for which points will first be determined on a scale from 1 to 10, where 10 is the best value. Subsequently, according to the use case, suitable scales will be selected for literary research, the chosen use case as well as a number of experts dealing with the issue of the Internet of Things at the Czech University of Life Sciences Prague and the Farms dealing with smart agriculture. After the evaluation, a suitable compromise option will be chosen, or the development options based on the results for individual use cases will be described. Based on the survey, the following 8 evaluation criteria have been selected by the experts: Licensing, Documentation & Community, Development, User & Device Management, Format Support, Database, Data Distribution, Visualization (da Cruz et al., 2018).

2.1 Selected Middleware

The main selection criterion for the selection of middleware services was their availability and the preferred option was open source, the possibility of testing in a time limited trial version or limited demo functionality - the program was available for trial without credit card insertion. This excluded options like AWS - Amazon Web Services with its IoT Core, Microsoft Azure - IoT Hub and the ability to create its own middleware service using Elastic's ELK file - ElasticSearch for search, database and analysis, LogStash for logging and data collection and Kiban for visualization. (Šimek, et al., 2019) These solutions are very robust and are not needed for small business use. Middleware services were also selected according to current developments, the latest update could not be older than a few months and at the same time being developed by a company or group, not by an individual, this requirement will exclude a small, unsearchable middleware that will be many. These small projects can bring value, but in the case of using a small business that wants to continue its long-term development, and the product is actively used, the likelihood of closing a product that is covered by an individual is too high (Somani et al., 2019).

Five middleware services were selected for the final evaluation: Thinger.io, WSo2, IoT Mainflux DeviceHiv, ThingsBoard. All selected middleware are available online.

3.RESULTS

For the weighted sum method, scores were assigned for each criterion of each middleware service (Table 1.) All points for the weighted sum method are listed in the table below (Table 2.).

Table 1. Weighted sum method I.part.

Criteria	**Middleware**				
	Thinger.io	WSo2 IoT	Mainflux	DeviceHive	ThingsBoard
Licensing	3	6	10	10	10
Documentation & Community	7	3	6	10	10
Development	6	3	6	8	10
User & Device Management	7	2	7	8	10
Format Support	3	10	8	8	3
Data Distribution	4	4	8	7	10
Visualization	5	7	8	8	9
Database	3	3	7	7	9

The next step after filling the weighted table is to calculate the total benefit of the variant. This is calculated by the scalar product of the values in the row with the values selected by the balance. The result is a table of total benefits below (Table 3.)

Table 2. Weighted sum method part II.

Criteria	**Criterium weight**	Middleware				
		Thinger.io	WSo2 IoT	Mainflux	DeviceHive	ThingsBoard
Licensing	**0.15**	0	0.43	1	1	1
Documentation & Community	**0.2**	0.57	0	0.43	1	1
Development	**0.2**	0.43	0	0.43	0.71	1
User & Device Management	**0.05**	0.63	0	0.63	0.75	1
Format Support	**0.15**	0	1	0.71	0.71	0
Data Distribution	**0.1**	0	0	0.67	0.5	1
Visualization	**0.15**	0	0.5	0.75	0.75	1
Database	**0.05**	0	0	0.67	0.67	1

Table 3. Result of multi-criteria analysis of variants.

Middleware	Thinger.io	WSo2 IoT	Mainflux	DeviceHive	ThingsBoard
Overall benefit of the variants	0.231	0.289	0.672	0.833	0.9
Order	5.	4.	3.	2.	1.

On the basis of a multiple-criteria decision analysis of the weighted sum method, key criteria, selected weights for the use of a small start-up company, ThingsBoard was released as a compromise option with the highest value of 0.9. The second was DeviceHive with 0.833 and Mainflux, WSo2 IoT and Thinger.io.

4. Discussion

The multiple-criteria decision analysis is based on the best-rated compromise variant of the middleware service ThingsBoard. If a small start-up company is taken into account, this service is very interesting for it. Initially, the company does not have the money to buy a robust ready-made solution, and for this free middleware they would have the opportunity to free up their 10 users. The suitability of ThingsBoard would be enhanced by the user interface, which is much more accessible to non-technical users and offers suitable presentation materials for potential clients. The second proposed option could be to implement a custom solution that is being developed by a large multinational corporation with library support in 4 major languages. Both variants have great documentation with a detailed description of all key aspects, this fact supports better support on GitHub, on which ThingsBoard is actively responding. ThingsBoard would be particularly useful when there is a need for rapid prototyping of a solution, both to demonstrate an idea, a feasibility test, or to validate a concept. Fast prototyping is enabled by the graphical scripting framework Rule Engine.

In the case of recommendations for medium to large farms, a different situation needs to be analyzed. Large companies always prefer to develop a customized system tailored to their specific requirements, a second DeviceHive compromise option would be appropriate for this example. By covering the programming languages with their libraries - Python, Java, JavaScript, Go - they cover most of the common market in finding experts.

Other options had a greater gap in the multiple-criteria decision analysis than the first two variants among themselves, which was due to the prevailing lack of evaluation in the core criteria, Documentation and Community and Development. The last two variants of WSo2 IoT and Thinger.io had poorly processed visualization and limited connectivity to external databases.

Comparing the results of multiple-criteria decision analysis with other sources and articles is not possible because it cannot be found through a common search for similar comparisons. Online are to find the introduction of individual middleware services, but not their comparison. There were lots of resources on the Internet that compare middleware services in terms of their architectures and implementations.

5. Conclusions

The aim of this work was to evaluate middleware services used for aggregation, transformation and distribution of data obtained from smart agriculture sensor networks. In the first part of paper, the key requirements for middleware services resulting from the literary research were analyzed. Each chosen criterion was presented, described on the basis of what knowledge the method and procedure of the evaluation follows and defined. The extent to which middleware has been addressed as a topic of work has been precisely defined and the case of using a small start-up company for which criteria and their weights have been described. The documentation and method of developing a particular middleware service were collectively the most important criteria as key parameters for middleware solution development.

In the next chapter, the middleware services were selected according to the basic criteria: available online, open source, it is possible to try them in a time limited trial version or demo without the need to insert a credit card. Middleware services also had to be actively supported by developers and the latest update could not be older than a few months. Based on these initial requirements, 5 services were selected - Thinger.io with UI development, WSo2 IoT with multi-format support and UI development environment, Mainflux custom solution implementation service, DeviceHive with detailed documentation, and ThingsBoard middleware that offered custom UI scripting via Rule Engine framework. The services were presented and further described, their functionality, architecture and other characteristics according to the selected criteria.

In the last part, selected middleware services were evaluated using multi-criteria analysis and suitable compromise variants applied to individual cases of use were evaluated. The results are as follows ThingsBoard with the highest value of 0.9. The second was DeviceHive with 0.833 and Mainflux, WSo2 IoT and Thinger.io. For the case of using a small start-up company, ThingsBoard appeared to be a suitable compromise option, which had the most benefit in DeviceCritical analysis and DeviceHive. ThingsBoard is suitable for rapid prototyping with the UI Rule Engine framework, which can be implemented with its own implementation for this robust tool, offering asset management, user and group management as an Asset Management feature. DeviceHive is suitable for both a small farm and larger companies that want to work on their own solution with the support of an open source middleware service that is being developed by a multinational company. DeviceHive has its own library in multiple languages - Python, JavaScript, Java and Go. These results and findings can be followed by a real project implementation, which will collect data from the Internet of Things devices and help in selecting the appropriate middleware service.

ACKNOWLEDGMENTS

The results and knowledge included herein have been obtained owing to support from the following institutional grants. Internal grant agency of the Faculty of Economics and Management, Czech University of Life Sciences in Prague, grant no. 2019B0009 – "Life Sciences 4.0".

REFERENCES

Atzori, L., Iera, A. and Morabito, G. (2010). 'The Internet of Things: A Survey'. Computer networks, 54(15), pp. 2787-2805. doi: 10.1016/j.comnet.2010.05.010

Cimino, L.D., de Resende, J.E.E., Silva, L.H.M. et al. (2019) 'A middleware solution for integrating and exploring IoT and HPC capabilities', Software: Practice and Experience, 49(4), pp. doi: 10.1002/spe.2630

Dastjerdi, A. V. and Buyya.,R. (2016). 'Fog Computing: Helping the Internet of Things Realize Its Potential'. Computer, 49(8), pp. 112-116. doi: 10.1109/MC.2016.245

da Cruz, M. A. A., Rodrigues, J. J. P. C., Al-Muhtadi,J., Korotaev,V. V. and de Albuquerque, V. H. C. (2018) 'A Reference Model for Internet of Things Middleware' IEEE Internet of Things Journal 5(2), pp.871-883. doi:10.1109/JIOT.2018.2796561

Fiala, P., Jablonský, J. and Maňas, M. (1994) 'Vícekriteriální rozhodování' Vysoká škola ekonomická v Praze, ISBN 80-7079-748-7

Jarolímek, J., Pavlík, J., Kholova, J. and Ronanki, S. (2019) 'Data Pre-processing for Agricultural Simulations', AGRIS on-line Papers in Economics and Informatics, 11(1), pp. 49-53. doi: 10.7160/aol.2019.110105.

Kumarage, H., Khalil, I., Alabdulatif, A., Tari, Z. and Yi, X. (2016) 'Secure Data Analytics for Cloud-Integrated Internet of Things Applications' IEEE Cloud Computing, 3(2), pp.46-56. doi:10.1109/MCC.2016.30.

Liangzhao Zeng, B. Benatallah, Ngu, A. H. H., Dumas, M., Kalagnanam, J. and Chang., H. (2004) 'QoS-aware middleware for Web services composition' IEEE Transactions on Software Engineering 30(5), pp. 311-327. doi: 10.1109/TSE.2004.11

Razzaque, M. A., Milojevic-Jevric, M., Palade, A., and Clarke, S. (2016) 'Middleware for Internet of Things: A Survey' IEEE Internet of Things Journal, 3(1), pp. 70-95. doi: 10.1109/JIOT.2015.2498900

Šimek P., Vaněk J., Stočes M., Jarolímek J. and Pavlík J. (2017) 'Mobile accessibility expense analysis of the agrarian WWW portal' Agricultural Economics, 63, pp. 197-203. doi: 10.17221/313/2015-AGRICECON

Somani, G., Zhao, X. Srirama, S. N.and Buyya, R. (2019) 'Integration of Cloud, Internet of Things, and Big Data Analytics' Software: Practice and Experience, 49(4), pp 561-564. doi: 10.1002/spe.2664.

Stočes, M., Masner, J., Jarolímek, J., Šimek, P. (2018) 'Internet of Things and Big Data Processing in Agriculture.' In AFITA/WCCA 2018 Proceedings - Research Frontiers in Precision Agriculture 24.10.2018, Powai, Mumbai, pp. 49-51.

Stočes, M., Šilerová, E., Vaněk, J., Jarolimek, J. and Šimek, P. (2018) 'Possibilities of using open data in sugar & ugar beet sector' Listy cukrovarnické a řepařské, 134(3), pp. 117-121. ISSN: 1210-3306.

Stočes, M. – Masner, J. – Kánsaká, E. – Jarolímek, J. (2018) 'Processing og Big Data in Internet of Things and Precision Agriculture'.International Scientific Conference Agrarian Perspectives XXVII. Food Safety – Food Security. pp. 353-358. ISSN 2464-4781

PROPOSAL OF AN ARCHITECTURE FOR DATA INTEGRATION AT AGRICULTURAL SUPPLY CHAINS, CONSIDERING THE IMPLEMENTATION OF IOT, NOSQL AND BLOCKCHAIN TECHNOLOGIES

Roberto F. Silva, Gustavo M. Mostaço, Fernando Xavier, Antonio Mauro Saraiva, Carlos E. Cugnasca

Department of Computer Engineering and Digital Systems,

Escola Politécnica da Universidade de São Paulo (USP), Brazil

roberto.fray.silva@gmail.com, gmostaco@usp.br, fxavier@usp.br, saraiva@usp.br, carlos.cugnasca@usp.br

ABSTRACT

Agricultural supply chains produce a huge amount of data related to traceability, production processes, environmental monitoring, among others. These are very important for the decision-making processes of the different supply chain stakeholders. With the implementation of technologies related to the Internet of things (IoT), the quantity and variety of data generated will increase even further. In order to extract useful information in an efficient way, it will be important to consider many aspects related to data management, such as security, processing, storage, transfer, etc. In this paper, we gather the requirements to implement IoT on agricultural SCs and propose an architecture that uses local databases to store raw and confidential data, a NoSQL database on the cloud to aggregate data that is important for decision-making by different stakeholders, and a blockchain that aggregates and safely stores the information for two main users. They are (i) consumers, considering aspects that guarantee product quality, together with what is relevant for them to choose between different products or brands; and (ii) government, which is related to data used for inspections, quality control, customs processes, and certifications. We conclude this paper by presenting at which level each of the functional, non-functional and domain-specific requirements will be fulfilled, and its main advantages in comparison with other architectures.

Keywords*:* agri-food, supply chains, IoT, big data management, blockchain

1. INTRODUCTION

A supply chain (SC) can be defined as a group of agents that are responsible, directly or indirectly, for supplying the demand of customers (Chopra, Meindl, 2013). Agricultural SCs deal with agents and products related to the agri-food domain, and they normally involve links such as production (at farms), warehousing (of raw materials and end products), processing (at industries), distribution centers, wholesalers, retailers and consumers.

An SC generates a considerable amount of heterogeneous and non-standardized data from different sources, which are processed by different software at the agents' systems and stored on different databases (Hribernik et al., 2010; Corella, Rosale, Simarro, 2013). In the case of agricultural SCs, these are related to three main systems: (i) traceability and product identification; (ii) environmental monitoring; and (iii) process monitoring. These will be further explored in Section 2.

The different stakeholders have different needs in terms of information availability, ease of use, confidentiality, among others. With the implementation of technologies related to the Internet of Things (IoT), these needs tend to increase. Nevertheless, new technical problems tend to rise, mainly related to the lack of interoperability and the generation and processing of huge amounts of data, demanding innovative ways to manage and organize the data flows.

In this sense, the objective of this research is to propose an architecture that uses local databases to store raw and confidential data in a NoSQL database, stored on the cloud, to aggregate data that is important for decision-making by different stakeholders, and a blockchain that aggregates and safely stores the information that is relevant for consumers and government officials.

2. DATA IN AGRICULTURAL SUPPLY CHAINS

There are three main types of systems that generate data on agricultural SCs (Silva et al., 2015; Pang et al., 2015; Verdouw et al., 2013; Chopra, Meindl, 2013; Verdouw et al., 2016):

1. **Traceability and product identification** → related to data generated during activities such as product and batches identification, tracking and tracing. The main technologies involved are barcodes, QR codes and radio-frequency identification (RFID). Data capturing can be automatic or semi-automatic. Product location is also part of this system;
2. **Environmental monitoring** → related to data generated while monitoring environmental variables that affect the product's quality. Ranges from the production at the farm, through processing at the industry, transportation or warehousing, all the way to the wholesalers and retailers shelves. The main environmental variables monitored for agricultural SCs are temperature, relative humidity, light, and pressure;
3. **Process monitoring** → related to data generated while monitoring production processes, mainly in the industrial environment. These are highly heterogeneous and can vary considerably among agents in the SC and even inside the same link. Most of the data generated on this system are not open for other agents in the SC and can contribute little for their decision-making processes. Nevertheless, some data, such as which ingredients and raw materials used, should be made public, as well as the transformation processes.

System 1, related to traceability, is very important for agricultural SCs and food products, as it is crucial to identify product contaminations, batches with problems, and the location of products in the SC, among others. Traceability can be defined as the tracking and tracing of product batches throughout the SC, allowing for the identification of critical quality control points (Juran, Godfrey, 2000).

As several agents adopt different technologies to allow product traceability throughout the SC, a series of interoperability problems can be generated. One way to reduce the occurrence of these problems is to utilize a common and widely accepted paradigm that describes how each technology should interact on the SC. One such paradigm is the IoT, which will be described in the next section.

3. RELATED WORKS

According to Atzori, Iera and Morabito (2010), IoT is a paradigm in which technologies related to product traceability, identification and monitoring - such as RFID, Wireless Sensor Networks (WSN), Global Positioning System (GPS), among others - interact with each other. This can be used, for example, to remotely monitor the weather at a vineyard, while connecting grapes harvesting data in each plot to an RFID tag, achieving traceability from the farm link to the end consumer. Several kinds of research use this paradigm for the generation, collection, processing, and storage of data from agricultural SCs (Gubbi et al., 2013; Pang et al., 2015; Corella, Rosale, Simarro, 2013).

Lack of interoperability is a very important problem for the implementation of the IoT paradigm. Currently, there is no universally accepted standard for this. According to Harris, Wang and Wang

(2015), two main factors cause problems on the adoption of IoT technologies on SCs: (i) lack of interoperability between existing systems and (ii) long implementation periods associated with ICTs. To the best of our knowledge, most architectures, such as the IoT-A (Carrez, 2013), do not fulfill all the requirements from agricultural SCs.

To address the problem of excessive volume, variety, and velocity of data generation, the NoSQL database was proposed. Its main advantages over traditional database models are a fast read and write operations, easy expansion, support for mass data storage and low cost. Moreover, the use of NoSQL databases is driven by other performance reasons, such as the avoidance of unneeded complexity, high throughput and horizontal scalability.

While traditional databases are mostly based on the relational model, NoSQL databases use other data models, such as key-value or column-oriented documents, each one presenting its related advantages and disadvantages. The study of Kamilaris, Kartakoullis and Prenafeta-Boldú (2017) reviews the Big Data technologies used in agriculture, with the use of cloud computing, machine learning and NoSQL databases in several studies. As highlighted by these studies, the increased availability of data collection devices represents both an opportunity and a challenge to turn this massive volume of heterogeneous data into useful information that could be addressed by the NoSQL model. Nevertheless, it is also important to make relevant information accessible to the agents in the supply chain. Among these, consumers and the Government demand data that could be easily accessible and trustworthy. The blockchain technology is an option to address this need.

A blockchain can be described as a sequence of information blocks that are connected with each other using cryptographic hashes. Each block contains information in its header, data from the present transaction, besides the unique ID of the previous block (Zheng et al., 2017). The header information is used to ensure data consistency within the chain by a cryptographic hash. In the blockchain, there are mechanisms for authorization and validation of the read and write operations of the blocks, ensuring security in the transactions (Hackius, Petersen, 2017; Tian, 2016).

Using a blockchain, data control is completely decentralized, bringing more transparency, without losing control of issues such as access security, and data integrity (Hackius, Petersen, 2017). In addition, with fewer intermediary elements on the transactions, there may also be performance gains. Decentralization, security and integrity are labeled as fundamental advantages of blockchain technology in relation to other technologies, due to the use of other concepts already established in Computer Science, such as distributed computing, cryptography and fault tolerance mechanisms.

4. Methodology

The methodology used in this research is based on two steps:

1. Identification of the main functional, non-functional and domain-specific requirements, through an extensive and in-depth literature review using the Scopus database;
2. Proposal of an architecture to address the requirements identified on the first step, considering real-world scenarios, as well as the increase of the IoT paradigm adoption, without standardization of all its technologies.

5. Requirements Identification

An in-depth literature review was conducted to identify the main requirements for data integration at agricultural SCs, considering the IoT paradigm. A series of proposals and frameworks for the implementation of IoT on SCs were analyzed. For means of this research, it was discarded models that: (i) considered only specific links in the SC; (ii) did not consider interoperability problems; (iii) did not consider the requirements of the different agents in the SC; and (iv) did not consider product quality monitoring.

The main requirements were identified and separated into three groups: functional (6 requirements), non-functional (4 requirements), and domain-specific (7 requirements). Functional requirements are related to the existence of services, functionalities, techniques, and tools to manage different hardware and software resources and aspects such as energy management. Non-functional requirements are related to performance, quality of service level, security, availability, and trustworthiness. Domain-specific requirements are related to implementation in a particular domain. Those requirements are presented in Table 1.

6. PROPOSED ARCHITECTURE

The proposed architecture is illustrated in Figure 1. It is composed by:

- Data generation and collection at each SC agent. Data should be standardized, if possible;
- Raw data storage at local databases. Each agent should store data that is not yet suited for decision-making on the SC level. Nevertheless, the agents should, as much as possible, treat their data to eliminate errors, outliers, null values, etc.;
- Transmission of non-confidential local data from the three systems to the cloud. The agent should, if possible, label the data to help on fusion and processing activities. If not, unsupervised machine learning methods, such as k-means could be used in the cloud to create labels for these data. In the cloud, two processes will happen: (i) data processing and fusion using a middleware; and (ii) storage of processed data in a NoSQL database;
- Transmission of only relevant data for consumers and the Government to the blockchain, which is openly available. This would contain data related to product batches origin (System 1), and the processes and ingredients that were used during its transformation (System 3)

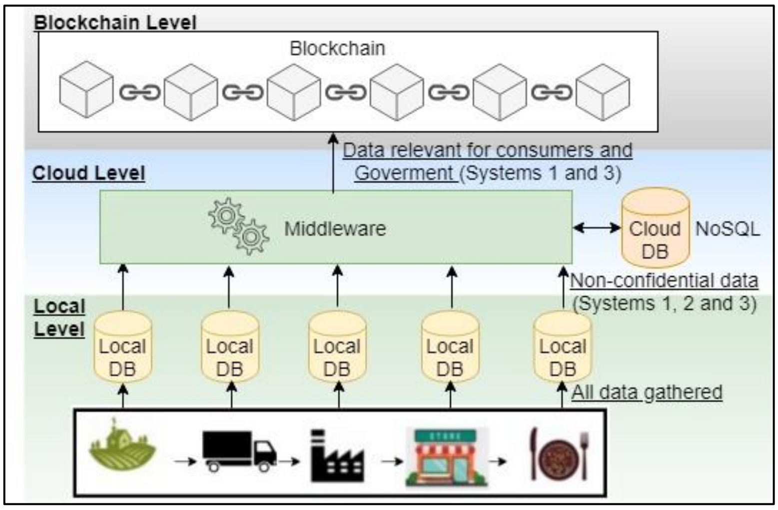


Figure 1. Proposed architecture for data integration at agricultural SCs

Table 1 contains the descriptions of how the proposed architecture addresses the functional, non-functional and domain-specific requirements previously identified. The functional and domain-specific requirements will be fulfilled at the local, cloud and blockchain levels, while the non-functional requirements will be fulfilled at the cloud and blockchain levels.

Table 1 - Proposed architecture and fulfillment of functional requirements

Requirement*	**How it will be fulfilled in the proposed architecture**
F1. Automatic identification	**1. Local level:** a) Standardizing the agent system as much as possible

F2. Resources management **F3.** Data management **F4.** Events management **F5.** Interoperability **F6.** Context identification	**2. Cloud level:** a) Functionalities developed on top of the IoT-A framework; b) Middleware for processing, NoSQL for data storage **3. Blockchain level:** a) Consumers and Government visualizations
NF1. Autonomy **NF2.** Middleware scalability **NF3.** Resilience **NF4.** Privacy and security	**1. Cloud level:** a) Functionalities developed on top of the IoT-A framework; b) Middleware that uses parallel computing **2. Blockchain level:** a) Fault tolerance and authorization
D1. Technologies coexisting **D2.** Creation of virtual SCs **D3.** Distributed traceability **D4.** Virtual object quality control **D5.** Heterogeneous data **D6.** Authorization, reliability **D7.** Information fusion	**1. Local level:** a) Standardizing the agent system as much as possible; b) Guaranteeing that the agent follows traceability rules **2. Cloud level:** a) Functionalities developed on top of the IoT-A framework; b) Middleware for data processing; c) Authorization levels **3. Blockchain level:** a) Fault tolerance and authorization

*Groups are: functional (F), non-functional (NF) and domain-specific (D).

Unlike the model proposed by Tian (2016), in proposed architecture, the data does not need to follow standards to be used for decision-making. The use of a middleware on the cloud will allow both a faster implementation (as the agents will not need to adopt standards before they participate in the system), but it will also allow for coexisting technologies. In addition, to the best of our knowledge, the proposed architecture takes the lead on clearly identifying the systems that will generate most of the data in an agricultural SC.

7. Conclusions

With the implementation of IoT technologies, agricultural SCs will generate a huge amount of heterogeneous data. The proposed architecture can be considered more adequate in this context, in which huge amounts of data becomes a challenge to extract useful information. In order to store, process and present relevant information for consumers and Government officials, the requirements were identified and an architecture was presented for implementing IoT on agricultural SCs. This was encompassed by the use of local databases at the agents to store confidential data and the use of a central cloud database with NoSQL technology. They are connected by a middleware, capable of parallel processing and that receives important data for the SC, and the use of a blockchain to encrypt and present relevant data.

The main difficulty on developing this architecture was the lack of open real-world data for the requirements identification, as companies rarely share their data. Overcoming this limitation demanded a thorough literature review. Nevertheless, several points may be developed on future research, such as a more thorough description of each level in the architecture, a description of how it can be implemented together with the IoT-A framework, and the development of the middleware that will allow all the data to be processed and stored on the NoSQL database.

Acknowledgments

This work was supported by Itaú Unibanco S.A. through the Itaú Scholarship Program, at the Centro de Ciência de Dados (C²D), Universidade de São Paulo, Brazil, by the National Council for Scientific and Technological Development (CNPq), and also by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

REFERENCES

Atzori, L., Iera, A., Morabito, G. (2010) 'The internet of things: A survey', Computer Networks, 54(15), pp.2787-2805.

Carrez, F., Bauer, M., Boussard, M., Bui, N. (2013) 'Final architectural reference model for the IoT v3. 0', EC FP7 IoT-A Deliverable, 1.

Chopra, S., Meindl, P. (2013) 'Supply chain management: Strategy, planning, and operation', 5th ed. New Jersey, USA: Pearson Education, 528pp.

Corella, V.P., Rosalen, R.C., Simarro, D.M. (2013) 'SCIF-IRIS framework: a framework to facilitate interoperability in supply chains', International Journal of Computer Integrated Manufacturing, 26(1-2), pp.67-86.

Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M. (2013) 'Internet of Things (IoT): A vision, architectural elements, and future directions', Future Generation Computer Systems, 29(7), pp.1645-1660.

Hackius, N., Petersen, M. (2017) 'Blockchain in logistics and supply chain: trick or treat?', In: Proceedings of the Hamburg International Conference of Logistics (HICL), pp. 3-18.

Hribernik, K.A., Warden, T., Thoben, K.D., Herzog, O. (2010) 'An internet of things for transport logistics–an approach to connecting the information and material flows in autonomous cooperating logistics processes', In: Proceedings of the 12th international MITIP conference on Information Technology & Innovation Processes of the Enterprises, pp. 54-67.

Juran, J. M., Godfrey, A. B (2000) 'Juran's Quality Handbook', 5th ed. USA: McGraw-Hill Publisher, 1872pp.

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X. (2017) 'A review on the practice of big data analysis in agriculture', Computers and Electronics in Agriculture, 143, pp.23-37.

Pang, Z., Chen, Q., Han, W., Zheng, L. (2015) 'Value-centric design of the internet-of-things solution for food supply chain: value creation, sensor portfolio and information fusion', Information Systems Frontiers, 17(2), pp.289-319.

Silva, R.F., Praça, I., Yoshizaki, H., Cugnasca, C.E. (2015) 'Proposal of a traceability model for the raw Brazilian sugar supply chain using RFID and WSN', In: Production and Operations Management Society, POMS 26th Annual Conference, Washington DC, USA.

Tian, F. (2016) 'An agri-food supply chain traceability system for China based on RFID & blockchain technology', In: 13th international conference on service systems and service management (ICSSSM), pp. 1-6.

Verdouw, C.N., Vucic, N., Sundmaeker, H., Beulens, A. (2013) 'Future internet as a driver for virtualization, connectivity and intelligence of agri-food supply chain networks'. International Journal on Food System Dynamics, 4(4), pp.261-272.

Verdouw, C.N., Wolfert, J., Beulens, A.J.M., Rialland, A. (2016) 'Virtualization of food supply chains with the internet of things'. Journal of Food Engineering, 176, pp.128-136.

Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H. (2017) 'An overview of blockchain technology: architecture, consensus, and future trends', In: 2017 IEEE International Congress on Big Data (BigData Congress), pp. 557-564.

Feature Selection and Grouping of Cultivation Environment Data to Extract High/Low Yield Inhibition Factor of Soybeans

Katsuhiro Nagata[1], Midori Namba[1], Seiichi Ozawa[1], Yuya Chonan[2], Satoshi Hayashi[2], Takuji Nakamura[2], Hiroyuki Tsuji[2], Noriyuki Murakami[2], Ryo Nishide[3] and Takenao Ohkawa[1]

[1]Kobe University, Japan

[2] Hokkaido Agricultural Research Center, NARO, Japan

[3] The Center for Data Science Education and Research, Shiga University, Japan

nagata@cs25.scitec.kobe-u.ac.jp, namba@cs25.scitec.kobe-u.ac.jp, ozawasei@kobe-u.ac.jp, ohkawa@kobe-u.ac.jp, y.chonan@affrc.go.jp, hayashis@affrc.go.jp, takuwan@affrc.go.jp, tuzihiro@affrc.go.jp, noriyuki@affrc.go.jp, ryo-nishide@biwako.shiga-u.ac.jp

Abstract

This work aims to extract high or low yield factors by analyzing soybean cultivation data and its cultivation environment. In this study, methods for using soybean data and investigating the cause of yield affected by the cultivation environment are proposed. As the soybean is affected by surrounding environment at each growing stage, the cultivation environment is also examined at each corresponding stage by dividing environment data. Qualitative values are generated at each stage, and the plants' condition for each stage is simply expressed because the slight changes of the environment do not affect much the growth of soybean. Then, similar cultivation environments at each growing stage are grouped by clustering. In the grouping, features of cultivation environment are selected to eliminate groups that have the few number of soybean fields whose environments are similar, and features with low possibility to be classified as either the high and low yield factors are removed. Thus, a distinctive group that is inclined toward either high yield or low yield of soybean cultivation has been identified. The cultivation environment that may affect the yield of soybeans has been revealed.

Keywords*:* soybean, feature selection, clustering.

1. Introduction

In recent years, there is a high demand in agricultural efficiency due to the aging society and a lack of successor in Japan. To solve this problem, data mining using past data is increasingly used (Harel, D. et al, 2014). Not many methods have been established to improve the amount of yield and quality of soybeans (Japan Agricultural Development and Extension Association (JADEA), 2012), which is the target of our work. Previous studies used LCM (Uno, T. et. Al, 2004), or tried to discover a frequent pattern from soybean cultivation data (Umejima, K. et. al, 2016, Namba, M. et. al, 2017). However, these studies did not investigate the dual effects of cultivation environments at multiple stages.

The yield is considered to be affected by the dual effects of cultivation environments at multiple stages. Therefore, the authors have studied a method using decision tree (Buchanan, G. , B. et al, 1978). Decision tree is easy to interpret visibly and can represent the dual effects of cultivation environments at multiple stages. However, decision tree makes factors more complicated to grasp because of its hierarchical structure. In this paper, we propose a method using clustering and feature selection. We do not intend to deepen the tree by using clustering, Ward's hierarchical method (Ward, Jr. , H. , J. , 1963), and feature selection.

2. METHODOLOGY

2.1 Consideration of soybean growth by stage division and qualitative value conversion

It is impractical to manage the daily growth of a crops because it burdens the farmers. In addition, since soybeans are known to be affected by the surrounding environment depending on the stage, the influence is to be examined by dividing the cultivation environment data by each stage or the corresponding period. In this study, we focus on the fact that the growth of soybean depends on the temperature before the flowering stage and the number of days thereafter (National Agriculture and Food Research Organization (NARO), 2003). The growth of soybean is divided into 8 stages by accumulated temperature before flowering beginning, and by 10 days after flowering beginning into 5 days. The accumulated temperature is a value obtained by adding the mean temperature every day from the sowing date, and the next stage begins from the next day when the value exceeds a reference value. Then, let the typical value for every cultivation environment element in each stage be a value in the stage (Figure 1).

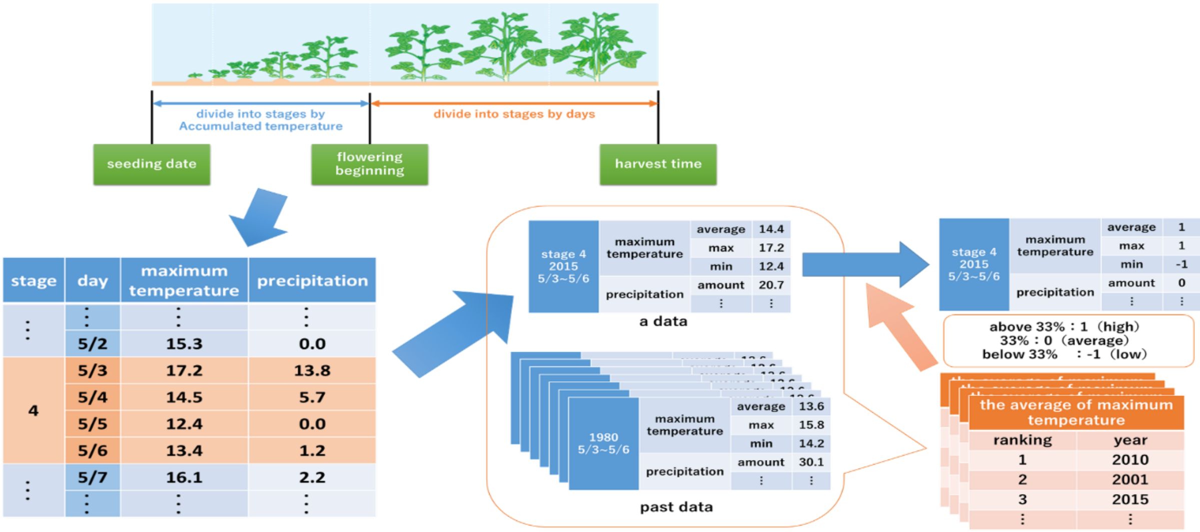


Figure 1. Dividing stage and making qualitative value (Syngenta Japan, "bits of knowledge about soybeans", https://www.syngenta.co.jp/cp/columns/hatasaku_blank_01?TB_iframe=true&width=740&height=800, Accessed 26 April 2019 (in Japanese))

Soybeans and other crops are not affected by small environmental changes. Therefore, the change of the environment is roughly grasped by converting the representative value of each element acquired for each stage to a qualitative value. As shown in Figure 1, each element is ranked by comparing its value with values in the same period in the last 30 years. According to the ranking, the elements are classified by the following three conditions; ranked in the top 33% as "high", bottom 33% as "low", and the others as "average". These three conditions are represented as {1, -1, 0} respectively.

2.2 Creation of field groups with similar cultivation environment by clustering using feature selection

There are many attributes of the data after being qualitatively valued. As the value varies depending on the field, too many complex yielding factors will be obtained if this data is used. Complex factors are difficult to use in actual soybean cultivation and are not useful. Therefore, we make groups of fields with similar cultivation environment by using clustering (Ward, Jr. , H. , J. , 1963), which enables to classify given data without external criteria. Ward's method is used in this work. The clustering was performed between fields and not within the fields. For example, clustering of Figure 2 shows the degrees of similarity are calculated for seven fields, and clusters are formed in terms of similarity. As a result, a total of five clusters are created.

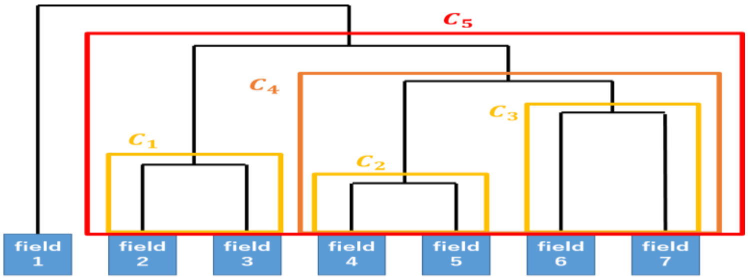


Figure 2. Example of clustering

Among the clusters created in this way, a cluster with a small number of fields is not useful because there are few fields that apply to the high or low yield factors in the end, and the factor will be unreliable. So, in this research, we will try not to create such a cluster so that the results are not influenced by feature selection. We select the elements used to create clusters with more than the number of differences between high and low yielding fields (we defined this number as 'dif') and remove elements that have never been used. Thereafter, clustering is performed again using only the selected element.

2.3 Extraction of high/low yield factors by decision trees

The data after clustering is only considered as influential at each stage. Since the crop grows with time, its yield is considered to be affected by the dual effects of cultivation environments at multiple stages. So, it is necessary to take into consideration the influence between stages in the whole period, not only at each stage. Thus, we use the decision tree (Buchanan, G. , B. et al, 1978) to find the factor from the information of fields to the clusters. For example, Figure 3 shows that fields, which do not belong to cluster 1 in stage 3 and belong to cluster 2 in stage 10, become high yielding fields. When the condition belonging to cluster 1 and cluster 2 in Table 1, a high yield factor "The average of maximum temperature is high in the stage 3 and the average of precipitation is low in the stage 10." can be finally obtained.

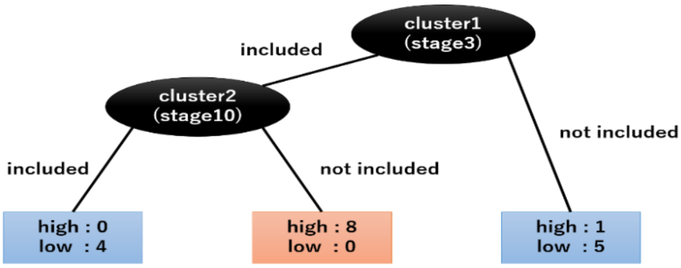


Figure 3. Sample of decision tree

Table 1. Example of the condition belonging to clusters

Cluster	The condition belonging to clusters
Cluster1	The average of maximum temperature is high in stage 3
Cluster2	The average of precipitation is low in stage 10

3. Experiment

In order to verify the effects of clustering, experiments have been conducted for the method using “Without clustering method” described in Section 2.3 and the proposed method called as “With clustering method” here in comparison. The outline of the field data are shown in Table 2. The total of 42 types of data are used in six fields from the soybean cultivation areas in Hokkaido. The yield of each field is ranked, and the experiments have been performed with the top 21 fields as high-yielding and the bottom 21 fields as low-yielding fields. We acquired the cultivation environment for every field from NARO mesh agriculture meteorological data (Table 3). In this experiment, dif in Section 2.3 is defined as

$$dif = 6$$

The Ward method is used for clustering, and the C5.0 algorithm is used for the decision tree.

Table 2. Field data

The number of fields	42
Location	Iwamizawa, Naganuma, Hitsujigaoka, Memuro, Bibai, Asahikawa
Year	2004 – 2017
Seeding date	12 May – 6 June
Cultivar	Yukihomare, Toyomusume, Toyoharuka, Yukishizuka

Table 3. environmental data acquired from NARO

The elements	Mean temperature	Maximum temperature	Minimum temperature	The amount of solar radiation	Precipitation	Sunshine duration
Location	Iwamizawa, Naganuma, Hitsujigaoka, Memuro, Bibai, Asahikawa					
Year	2004 – 2017					
The period of data acquisition	1 May – 30 November					
Qualitative value	Average, Maximum, Minimum, Range	Average, Maximum, Minimum	Average, Maximum, Minimum	Sum, Maximum, Minimum, Range	Sum, Maximum, Minimum	Sum, Maximum, Minimum, Range

We finally discuss the results of experimental methods with the experts (i.e. the cultivators of soybean in Hokkaido) to examine whether the results are matched with their experience.

4. Results

The results obtained by “With clustering method” and “Without clustering method” are summarized in Figures 4 and 5. In “With clustering method”, we obtained the following:

- If the range of the mean temperature in stage 3 is high and the total value of the solar radiation amount is below the average, then the yield is high.
- A field that does not fulfill all the conditions shown in Figure 5 has low yields.

In “Without clustering method”, we obtained the following:

- If the total value of the amount of solar radiation in stage 9 is low, the yield will be low.

- Even if the above conditions are not met, low yields may occur when the maximum of sunshine duration in stage 8 is high.
- If the amount of solar radiation in stage 9 is above the average, and the maximum of the sunshine hours in stage 8 is below the average, the yield is low.

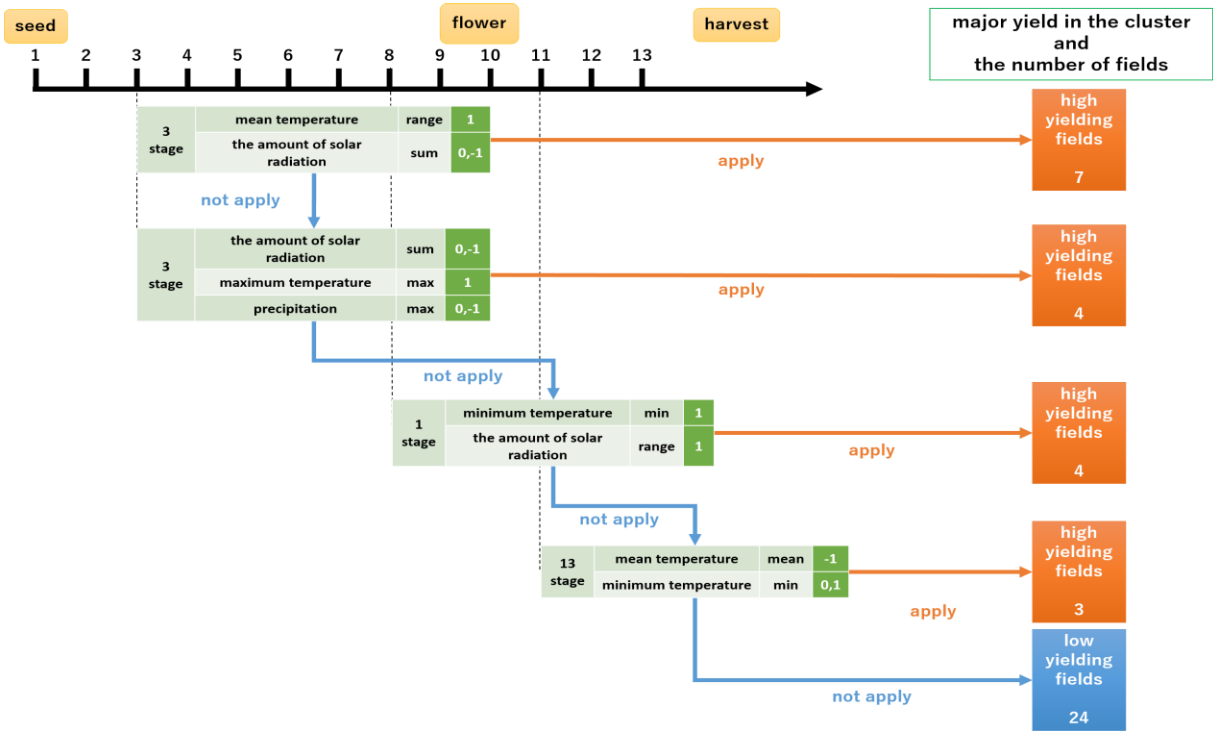


Figure 4. Decision tree of "With clustering method"

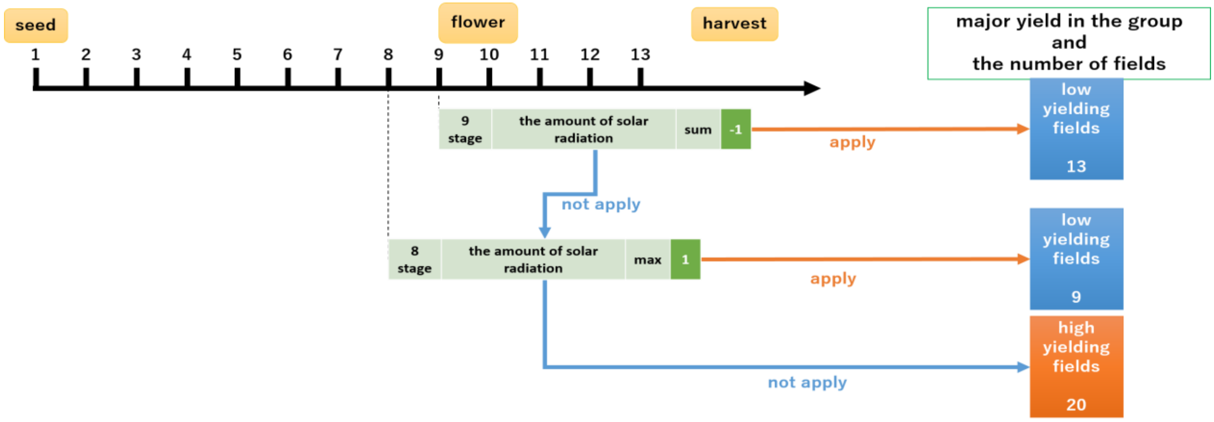


Figure 5. Decision tree of "Without clustering method"

5. Discussion

First of all, as the number of nodes of decision tree is fewer in "Without clustering method", it is easy for the "Without clustering method" to interpret. Looking at each factor, the factor of the first condition in stage 9 is matched with the experts' knowledge that high amount of solar radiation around the beginning of flowering leads to high yield. On the other hand, the second factor is not matched with the experts' knowledge.

Next, we analyze the results of "With clustering method". The individual conditions of "With clustering method" are all matched with the experts' opinions. Moreover, the condition in stage 3 is appeared. Usually, this stage is considered to be less related to growth. We may discover a new knowledge that the environment of stage 3 is much important for soybean growth.

6. Conclusions

In this study, a method is developed to obtain more reliable high or low yield factors by clustering and using feature selection for the factors that influence the yield of soybeans. In order to confirm the effectiveness of “With clustering method”, the results were compared with the practice and knowledge of experts in actual soybean cultivation. The overall result of “With clustering method” is relatively correct, and is useful information for future soybean cultivation.

The points of the study will be summarized and referred for future works. First, soybean growth does not always depend on the elements of Table 3. Other elements of soybean environment must be considered. Second, we experimented with only 42 data of fields. It is necessary to experiment with additional data to get more reliable results. Third, cross validation needs to be conducted to prove the reliability of our method.

Acknowledgement

This study was partially supported by Ministry of Agriculture, Forestry and Fisheries: Development of diagnostic methods and countermeasure techniques for overcoming high-yield inhibitory factors.

References

Harel, K. , Fadida, H. , Slepoy, A. and Shilo, K. (2014) ‘The Effect of Mean Daily Tem-perature and Relative Humidity on Pollen, Fruit Set and Yield of Tomato Grown in Commercial Protected Cultivation’, Agronomy, Vol. 4, Issue. 1, pp. 167-177.

Japan Agricultural Development and Extension Association (JADEA) (2012) ‘Soybean Making for Improvement of Yield and Quality and Stable Production Q & A’, https://www.jadea.org/houkokusho/daizu/documents/daizu-kaitei.pdf, Accessed 26 April 2019 (in Japanese).

Uno, T. , Asao, T. , Uchida, Y. , and Arimura, H. (2004) ‘An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases’, Lecture Notes in Articial Intelligence, pp. 16-31.

Umejima, K. , Arimitsu, F. , Ozawa, S. , Murakami, N. , Tsuji, H. , and Ohkawa, T. (2016) ‘Optimal Pattern Mining from Time-Series Cultivation Data of Soybeans for Knowledge Discovery’, Proceedings of Joint Workshop on Time Series Analytics and Collaborative Agents Research & Development (in conjunction with the 29th Australasian Joint Conference on Articial Intelligence), pp.19-24.

Namba, M. , Umejima, K. , Nishide, R. , Ohkawa, T. , Ozawa, S. , Murakami, N. and Tsuji, H. (2016) ‘Optimal Pattern Discovery based on Cultivation Data for Elucidation of High Yield Inhibition Factor of Soybean’, Proceedings of the 5th IIAE International Conference on Intelligent Systems and Image Processing, pp.209-216.

Hira, M. , Z. and Gillies, F. , D. (2015), ‘A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data’, Advances in Bioinformatics Volume 2015, Article ID 198363, pp. 13.

National Agriculture and Food Research Organization(NARO) (2003) ‘Cultivation technology after planting utilizing soybean "Yukihomare"’, http://www.naro.affrc.go.jp/project/results/laboratory/harc/2003/cryo03-03.html, Accessed 15 April 2019 (in Japanese).

Ward, Jr. , H. , J. (1963), ‘Hierarchical Grouping to Optimize an Objective Function’, Journal of the American Statistical Association, Vol. 58, pp. 236-244.

Buchanan, G. , B. and Mitchell, M. , T. (1978), ‘Model-directed learning of production rules’, Pattern-Directed Inference Systems, Academic Press, New York.

MULTIPASS : MANAGING THE CONSENTS OF ACCESS TO FARM DATA IN A CHAIN OF TRUST TO MAKE NEW SERVICES EMERGE FOR FARMERS

Bruno Lauga[1], Béatrice Balvay[2], Laurent Topart[3], Juliette Leclaire[4], Anthony Clenet[5], François Brun[6], François Pinet[7], Catherine Roussey[7], Mehdi Sine[1]

[1]ARVALIS-Institut du végétal, agricultural technical institute for arable crops, France

[2]IDELE, breeding technical institute, France

[3]ORANGE, telecommunications company, France

[4]FIEA, France Informatique Elevage et Agriculture, France

[5]SMAG, farm software provider, France

[6]ACTA, Agricultural Technical Coordination Association, France

[7]Université Clermont Auvergne, IRSTEA, UR TSCF, Clermont-Ferrand, France

b.lauga@arvalis.fr, beatrice.balvay@idele.fr, laurent.topart@orange.com, jleclaire@fiea.fr, aclenet@smag-group.com, francois.brun@acta.asso.fr, francois.pinet@irstea.fr

ABSTRACT

With the emergence of digital technologies, farms become a relevant source of data to meet the challenges of multi-performance agriculture. Beyond the services provided, access to farmers' data depends on a clear understanding of their use, which must be done in a transparent way. Several codes of conduct at a national or international level push for a voluntary commitment to respect some good practices in the use of agricultural data. To provide a tool and answer farmer's questions on the control of their data and the transparency of the data processing, the partners of the MULTIPASS project, have imagined an interoperable ecosystem of farmer consents management, protecting farmers from no consented uses of their data.

Farmers' expectations of such an ecosystem have been expressed during workshops. They want to better identify existing data flows, including actors, data processes, and data clusters. Based on the farmers' expectations, the MULTIPASS project stakeholders have proposed the architecture of an ecosystem integrating two consent management tools as "pilots". This ecosystem should take in charge the interoperability between each consent management tools or with future tools.

This solution is based on a shared typology of data and data processes as well as on the specifications of the consent message content. All these elements should be easily accessible to meet the interoperability need of the ecosystem. It is also based on a router, which provides unified access to consent management tools (using API). In particular, it provides the farmer (beneficiary) with an exhaustive view of his/her consents (which can be distributed on several consent management systems), meeting farmers' expectations for transparency. It is also the point where a data provider can check whether the consent required to provide data exists, without needing to know which consent management system is concerned.

In this project, the stakeholders want to demonstrate to agricultural professional organizations the benefits and feasibility of a consent management ecosystem. By strengthening the confidence of farmers to share data, the project will allow the emergence of new knowledge and new services.

Keywords*:* farm data, data management, consent, transparency, chain of trust.

1. Introduction

Farmers are engaged in a progress for sustainable and productive agriculture. With the emergence of digital technologies, farms become a relevant source of data to meet the challenges of multi-performance agriculture. These data are the basis of the decision-making process. There is a data-driven agriculture based on the data transfer within the farm. These data also make it possible to create new knowledge or tools that improve the precision and relevance of agricultural operations in order to increase yields without negative impact on the environment.

Beyond the services provided, access to farmers' data depends on a clear understanding of their use, which must be done in a transparent way (Brun et al., 2016). This is a real concern for both farmers, who cannot control the uses of their data, and also data providers who have difficulties in determining the access and reuse permissions they can provide on the farmers' data they host. Access rights must be properly managed, as well as the farmer's consent for the uses of her/his data. The conditions related to this consent must be easily accessible and modifiable by the farmer.

It is this chain of trust that the MULTIPASS project wants to implement. Through this project, the partners want to make available to farmers and data producers an interoperable farmers' consent management ecosystem, protecting data exchanges improving confidence to share their data with other organizations.

2. Challenges for a Consent Management Ecosystem

2.1 Towards a widespread use of consents

Consents are the adherence of one party to the request made by another. In the case of personal data, consent is one of the 6 legal bases provided by the EU General Data Protection Regulation (GDPR, 2016) which authorizes the implementation of data processing. The law does not require the systematic collection of consents before processing personal data, because other legal bases can be invoked to process these data such as a mission of general interest or a contractual commitment (CNIL, 2018). Nevertheless, consents will enable the management of agricultural data exchanges not specified in the contracts.

To authorize an agricultural data processing and to reinforce the transparency of these uses, the farmer must be able to express her/his consent as shown in Figure 1.

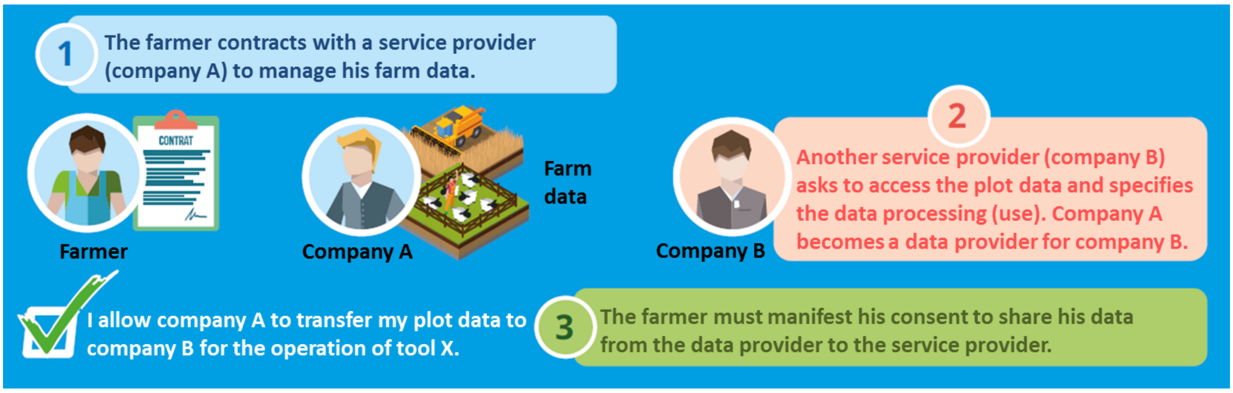


Figure 1. Example of a consent use for data exchanges

2.3 Build a chain of trust

Multipass project goal was to build an ecosystem of stakeholders to manage consents and to create the engagement rules of these actors. We defined the typology of stakeholders presented in Table 1 involved in any farm data exchanges and consents management.

Table 1. Typology of actors in a consent management ecosystem

Term	Definition
Right holder	The person who has the rights on the data. The consent of this person is needed to exchange data. In the MULTIPASS project, she/he is a farmer or breeder.
Delegatee	The right holder has delegated to a person or an organization (i.e., a delegatee) the right to give consents on her/his behalf.
Consent manager	The manager in charge of a consent management system.
Service provider	The organization that sells service to farmers and that needs an access to data. It is the beneficiary of the consent.
Data provider	The manager of the service (database) in charge of providing the data to the service provider.
Consent recorder	The organization that registers consents in the consent management system.

In this chain of trust, each actor has a responsibility and must satisfy good practices related to the use of agricultural data and consents.

2.3 Respecting good practices

Jurists seem to think that, in the absence of a specific legal regulation, the control of agricultural data is ensured only by contracts with the farmer (Douville, 2019). The control will not come from the law but from a voluntary commitment made by the parties to respect some good practices in data uses. The French DataAgri code of conduct (FNSEA, 2018) leaded by the “Fédération Nationale des Syndicats d'Exploitants Agricoles” (FNSEA) and the “Jeunes Agriculteurs” (JA), and the European CODE OF CONDUCT (EU code of conduct, 2018) clearly goes in this direction.

In this context, farmers’ expectations of such an ecosystem have been expressed in various workshops. Farmers regret that so far, they had not been consulted much when the service providers processed their data. They expressed a need for transparency and want to better identify existing flows, including stakeholders, data uses and associated data categories. Based on farmers' expectations, the MULTIPASS project stakeholders have proposed one architecture of an ecosystem integrating two consent management tools as "pilots" and the conditions for their interoperability with each other or with future tools.

3. Implementation of the Multipass Ecosystem

3.1 Proposed architecture

Consent management solutions dedicated to agriculture already exist. These systems are often designed for particular needs. These different consent management systems can be freely chosen by the ecosystem stakeholders. Consents are stored in the consent management systems with the only constraint to register the information expected in the MULTIPASS ecosystem interfaces.

The main tool defined in the MULTIPASS ecosystem is a router that guarantees the interoperability of the different consent management systems. It allows a unified access to consents to provide a list of them (by right holders, service providers, etc.) or to verify the existence of consents before data exchanges. For this, it knows and can query the various consent management systems which will have interfaces (APIs) similar to those of the MULTIPASS router.

In particular, it provides the right holder with an exhaustive view of her/his consents (which can be distributed across several consent management systems), meeting farmers' transparency needs. The router also allows a data provider to check if the consent required for a data exchange exists, without needing to know in which consent management system it is managed. There is also a traceability of these controls. The use case diagram presented in Figure 2 shows the expected roles of each of the actors as well as the functional scope of the MULTIPASS router.

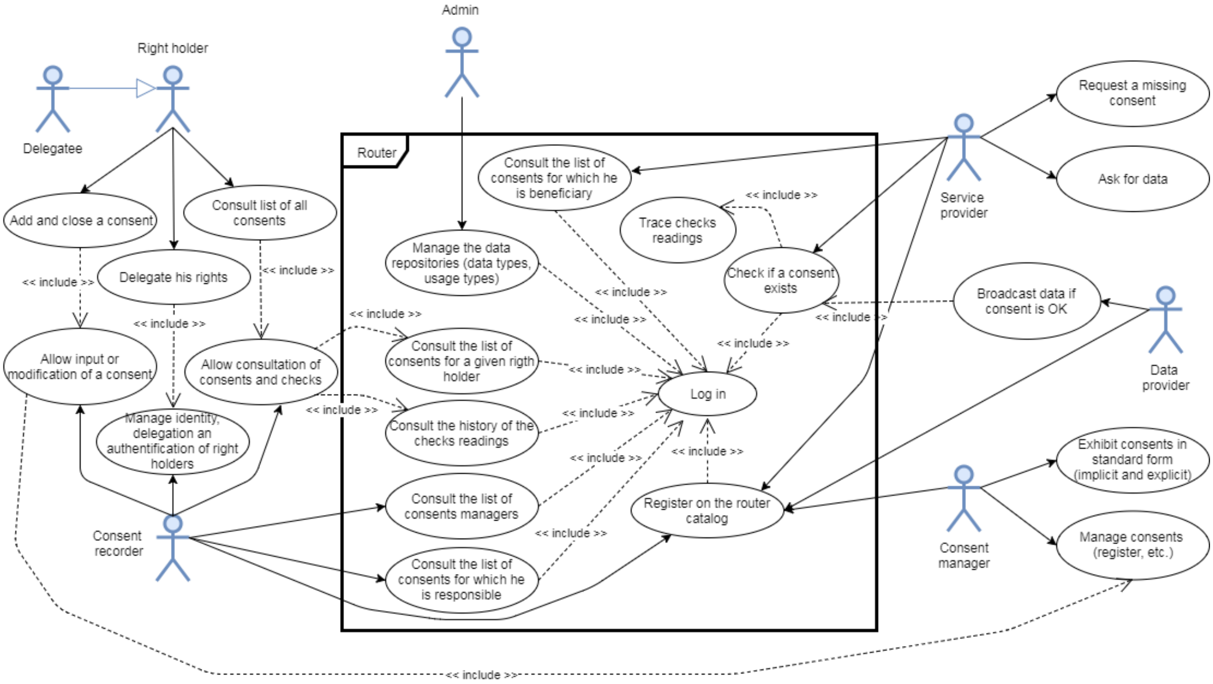


Figure 2. Use case diagram of MULTIPASS router

All actors other than right holders must register on the router before they can use it. Their registration is validated by a router's administrator. A right holder (or her/his delegatee) does not interact with the router. It is the role of the consent recorder to allow the input or the modification of consents. Only the right holder can see her/his consents once she/he is authenticated. The consent recorder cannot see them. This security is especially needed when the consent recorder is also a service provider (it must not see if the farmer works with its competitors). For this, either it will have made a contractual commitment in its contract with the farmer, or it will be committed by adhering to a charter or it will be obliged by GDPR in the case of personal data.

The management of the data repositories is the responsibility of a router's administrator.

3.2 Technical architecture

The router has a Java REST API that exposes business and administrative services. As consents are by nature sensitive data that must be secured, HTTPS is used for the exchanges. The OAuth protocol is used for authentication. A signature mechanism guarantees the API that the token issued during authentication process has been generated by the system. The passwords of the different users are stored in a SSHA hashed form in an LDAP server. A Java human-machine interface allows system administrators to manage the different users and data repositories (data categories and uses) of the router.

A reverse proxy "HA Proxy" is used to secure the application upstream. This system will also be used for load balancing between different downstream application servers. The PostgreSQL database that registers actors, data repositories and logs could eventually be transferred into an elastic stack.

3.3 Conditions for ecosystem interoperability

The router is an important part of the ecosystem interoperability. It is based on the main concept of consent. Consents are not managed in the database of the router, but only in the interfaces. The identification of the companies (farm, data or service provider) is done by the French SIRET identifier but the system allows the use of another identifier.

Table 2. Description of the concept of consent

SIRET number of the farm exploitation (data producer)	**WHO: actors of the data exchange**
SIRET number of the service provider (beneficiary)	
SIRET number of the data provider	
SIRET number of the consent recorder	
Data categories	**WHAT: What is the data exchange about**
Use case (codes)	
Use case description (free wording)	
Consent beginning	**Scope of consent**
Consent end	
Restrictions on consent (and data): (optional free wording)	
Anonymisation	**Constraints**
Contract (explicit, implicit ...) : If yes, contract reference or terms of use	
Reversibility of the consent (not possible if based on a contract)	

Ontologies are one of the possible solutions for solving data interoperability issues. The word ontology covers a large number of different data sources ranging from thesauri to schemas shared on the Web through semantic Web technologies (Roussey et al., 2011). In the MULTIPASS project, the partners studied different agricultural data exchange schemes, and in particular GIEA ("Gestion des Informations de l'Exploitation Agricole" – a model for Farm Information Management), a model created in France for data sharing (Pinet et al., 2009). These schemes propose a vocabulary dedicated to agriculture, but too complex and not suitable for the uses in the context of consent management. The definition of consents will be associated with a typology of data and a typology of uses that remains to be defined. We recommend that these lists will be organized (hierarchies of category) and shared on the Web to meet the interoperability need of the ecosystem.

4. Discussion

A Blockchain could constitute the ecosystem on its own, but the challenge at this stage is to explore its promises in terms of trust decentralization. For this, in the second phase of the project, two consent management tools will be compared within use cases. The first one is based on a trusted third party (France Génétique Elevage, 2016) and the second one will be based on Blockchain technology.

MULTIPASS does not have the ability to interfere with consent management systems. They have to verify that the person who registers a consent is the one for whom the consent is given. It is therefore recommended to clearly identify the users with the creation of identity providers for agriculture, as there are elsewhere (French administration, Google or Facebook). Finally, it is the responsibility of the consent manager to ensure the legal value of the consents collected. The participants of the MULTIPASS workshop held on Sept 27th, 2018 (bringing together socio-economic partners of the farmer) highlighted the overlap in the regulatory bases of contracts and consents. There may be a risk of contradiction between a consent and a pre-existing contract.

5. Conclusions

The soft law (codes of conduct) provides a framework for agricultural data exchange, putting forward consent management. The project aims to demonstrate to agricultural professional organizations the benefits and feasibility of a consent management ecosystem through limited but concrete use cases in France. The Blockchain technology will be evaluated to explore its promises in terms of trust decentralization. The router designed by the partners will implement a proof of concept for interoperability between existing and future consent management systems.

It provides a solution ("data passport") to farmers for the control on their data and on the transparency of the data uses, in a chain of trust. Now, the governance of the ecosystem should open to all agricultural actors. New knowledge and new services are expected as the confidence of farmers to share their data should be strengthened.

Acknowledgments

MULTIPASS (2018-2020) is funded by the French Ministry of Agriculture (in CASDAR program) and its partners.

References

Douville, T. (2019) ‘Contrat et données agricoles’, Droit rural, n° 469, 7p.

CNIL (2018), ‘Conformité RGPD : comment recueillir le consentement des personnes ?’ https://www.cnil.fr/fr/conformite-rgpd-comment-recueillir-le-consentement-des-personnes

GDPR (2016) ‘General Data Protection Regulation EU 2016/679, April 27th, 2016.

FNSEA and JA (2018) ‘DataAgri Charte sur l’utilisation des données agricoles’, 12p.

EU code of conduct (2018) ‘EU Code of conduct on agricultural data sharing by contractual agreement’, 20p.

Roussey, C., Chanet, J-P., Soulignac, V. and Bernard, S. (2011) ‘Les ontologies en agriculture’, journal ‘Ingénierie des systèmes d’information (ISI)’, special number ‘Systèmes d'informations pour l'environnement’ Vol 16, n°3, pp55-84.

Brun, F., Siné, M., Gallot, S., Lauga, B., Colinet, J., Cimino, M., Haezebrouck, T-P., Besnard, J., (2016) ‘ACTA - Les Instituts Techniques Agricoles: L’accès aux données pour la Recherche et l’Innovation en Agriculture. Position des Instituts Techniques Agricoles’.

FRANCE GENETIQUE ELEVAGE (2016) ‘Le consentement des éleveurs pour l’accès à leurs données - Comment le gérer ? Comment le faire respecter ?’, note proposed by the Information Systems Strategy Commission.

Pinet, F., Roussey, C., Brun, T., Vigier, F (2009) ‘The use of UML as a tool for the formalisation of standards and the design of ontologies in agriculture’. Chapter in: Advances in Modeling Agricultural Systems, Springer, p.131-147.

DETECTING OF APPROACHED INTERACTION WITH CATTLE IN ESTRUS BASED ON COMMUNITY TRANSITION AND CATTLE DISTANCE

Shunta Fukumoto[1], Ryo Nishide[2], Yumi Takaki[1], Chikara Ohta[3], Kenji Oyama[4] and Takenao Ohkawa[1]

[1] Graduate School of System Informatics, Kobe University, Japan

[2] The Center for Data Science Education and Research, Shiga University, Japan

[3] Graduate School of Science, Technology and Innovation, Kobe University, Japan

[4] Food Resources Education and Research Center, Kobe University, Japan

s.fuku5@cs25.scitec.kobe-u.ac.jp, yumi@people.kobe-u.ac.jp, ohkawa@kobe-u.ac.jp, ryo-nishide@biwako.shiga-u.ac.jp, ohta@port.kobe-u.ac.jp, oyama@kobe-u.ac.jp

ABSTRACT

In stock breeding of beef cattle, it is essential to efficiently produce calves to maintain stable management. For this purpose, most farmers conduct artificial insemination, which should be performed after half a day of estrus detection. Therefore, in order to perform artificial insemination successfully, it is necessary to detect cattle in estrus precisely. Generally, searching the estrus is performed visually. Sometimes, a pedometer or a temperature sensor is used to grasp cattle's condition. These methods, however, do not consider the fact that the cattle live in a community with other cattle. Taking such cattle's sociality into account may enable to grasp the change in cattle's condition and detect cattle in estrus more accurately. This study focuses on the sociality of grazing cattle and aims to detect estrus by grasping the change of social behavior. Cattle in estrus tend to be approached by other cattle continuously for several hours. The behavior of approaching or being approached (stated as approaching-approached behavior from here on) is quantified by using the position information. In order to minimize the influence of momentum in quantification, we propose a method focusing on the direction of the cattle during movement. Moreover, we employed weight based on community history and distance between two cattle in order to reduce noise due to unintended approaching-approached behavior. Furthermore, we performed anomaly detection with a state-space model to detect cattle in estrus based on the quantification of the approaching-approached behavior in real-time. We verified the effectivity of the quantification for estrus detection with performed artificial inseminations. As a result, the precision was 0.579, the recall was 0.733 and F-measure was 0.647, and thus, confirmed the effectivity of our method.

Keywords: Cattle, Cow, Estrus detection, Heat detection, Anomaly detection, GPS.

1. INTRODUCTION

In stock breeding of beef cattle, there is an obstacle to efficiently produce calves by performing artificial insemination (AI). It is necessary to find cattle in estrus in order to perform AI successfully. Currently, livestock workers who have knowledge about the change in cattle's condition and behavior take an observation of cattle. Therefore, as the area of a pasture and the number of cattle become larger, it is more difficult to find out the cattle in estrus. In order to reduce their burden in human resources, in this study, an attempt is made to detect cattle in estrus by focusing on the approaching-approached behavior. The approaching-approached behavior is observed frequently in cattle in estrus because it tries to avoid other cattle that tend to approach. In the previous year, we attempted to quantify the approaching-approached behavior when in estrus (Fukumoto et al., 2018). This method focused on the moving direction of cattle which were the target for estrus detection. However, there are several drawbacks in the method. First, the method did not consider the situation in more than three cattle and thus, the behaviors of more than two cattle are ignored. Second, the unintended approaches are likely to affect the quantification, and thus may cause errors for estrus detection. In order to deal with such problems, in this paper, after quantification of the approaching-approached behavior between two cows, we try to get rid of the unintended approaches from the quantification by assigning a weight for each cow based on community transition and distance. After that, we verify the effectivity through an experiment. In Section 2, we describe the method of the approaching-approached behavior, Section 3 shows an experimental result to evaluate method and finally, Section 4 summarizes our work with future tasks.

2. METHODOLOGY

2.1 Extraction of Interaction Information Related to Cattle in Estrus

Cattle are animal with sociality and express particular interactions among them when in estrus. For example, smelling their buttocks each other and standing are symbolic actions of that (Roelofs at al., 2005). During these interactions, abnormally approaching is frequently observed. The left side of Figure 1 shows the appearance of abnormally approaching toward a cow in estrus by another cow. Cattle in estrus are more likely to generate other cattle's interest and continue to be approached by any other cattle because their vulva becomes swollen and hyperemic. As a result of that, interactions such as standing are caused. Abnormally approaching during estrus lasts for a few hours and such an interaction is less likely occurred except cattle in estrus. Therefore, it is said that this behavior is a good indicator of cattle in estrus (Roelofs at al., 2005). That is why it is probably useful for estrus detection to quantify the approaching-approached behavior and detect the state of being abnormally approached with high accuracy.

We put GPS (Global Positioning System) receiver around their neck to acquire position information of each cow (stated as the GPS collar from here on). The overview of a GPS collar is shown on the right side of Figure 1.

Figure 1. Approaching-approached behavior (left) and GPS collar (right) *Note.* The source of the pictures: (Fukumoto et al., 2018)

The GPS collar is made by putting a GPS receiver and a USB battery (ARKNAV Co., Taiwan) in a waterproof box. The position information is recorded into the text file as a log at daily intervals and it contains latitude, longitude, velocity on the ground and so on. From the time series data of position information, we can calculate the distance and angle between two positions.

Cattle form some groups to live and the interactions are likely to be observed inside those groups. Therefore, we express the unit of those groups as communities and focus on the approaching-approached behavior inside the communities. In order to determine the community, we use Louvain algorithm based on the total time for which the distance between two cattle are less than 10 m (Blondel et al., 2008).

2.2 Quantification of approaching-approached behavior

A method of the quantification of the approaching-approached behavior is nearly the same as the previous year's (Fukumoto et al., 2018). This method focuses on the angle between a moving vector $V_{c_i}^t$ and a direction vector D_{c_i,c_j}^t. The cow c_i is the target for estrus detection. $V_{c_i}^t$ has the moving direction. D_{c_i,c_j}^t has the direction from cow c_i to any other cow c_j. On the other hand, the method in this paper focuses on the angle between two vectors shown in Figure 2.

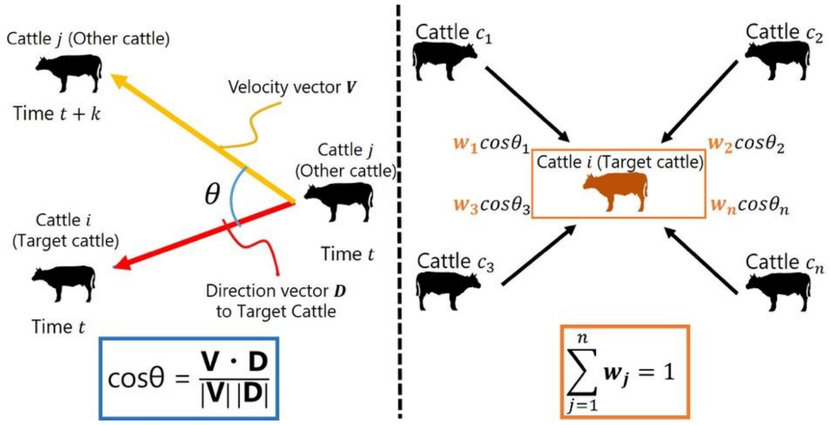


Figure 2. Concept of the proposed method

One is a moving vector $V_{c_j}^t$ of any other cow c_j, and the other is a direction vector D_{c_j,c_i}^t which has the direction from any other cow c_j to a target cow c_i. The approached index between two cows S_{c_i,c_j}^t at time t is calculated by

$$\textbf{Approached index between two cows} : \boldsymbol{S_{c_i,c_j}^t} = \frac{\boldsymbol{V_{c_j}^t} \cdot \boldsymbol{D_{c_j,c_i}^t}}{\left\|\boldsymbol{V_{c_j}^t}\right\| \left\|\boldsymbol{D_{c_j,c_i}^t}\right\|}. \quad (\mathbf{1})$$

The approached index between two cows S_{c_i,c_j}^t is expected to become nearly 1 when a target cow is approached by any other cow and become -1 when a target cow is approaching toward any other cow.

2.3 Employing weight based on cattle distance and community transition

During grazing, unintended approaching-approached behavior can occur accidentally because of feeding behavior and so on. Considering such things, we make an attempt to quantify the approaching-approached behavior among more than three cattle. At first, we decided to assign weights towards other cattle to get rid of the effect of the cattle's behavior which has nothing to do with the target cattle's behavior. These weights are assigned to only cattle which belong to the same community as the target cattle. The weights are determined by considering two factors as below.

The first is past community records. Most of the approaching-approached behavior in estrus are performed between friendly cows. These cows are more likely to travel with each other and thus, more likely to belong to the same community. Therefore, it is possible to assign more weights toward friendly cows to get rid of unintended approaching-approached behavior. The weights based on community transition is determined by the function of the number n_{c_j} that the cow c_j became the same community as the target cattle in the past X times.

The second is the distance to the target cattle. The distance at the approaching-approached behavior in estrus is smaller than the one when feeding and resting. Therefore, the approaching-approached behavior is more likely to occur with the nearer cow especially even in the same community. The weights based on cattle distance is determined by the function of the mean distance d_{c_j} between the target cow and any other cow c_j of every community generation interval Y.

Each function adopted to the determination of the weight is shown in Figure 3.

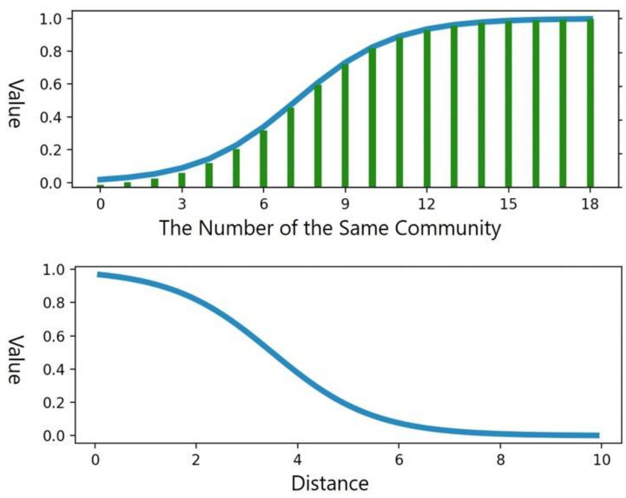


Figure 3. Functions used to determine the weights

These functions are generated based on sigmoid function. The weights $\boldsymbol{w_{c_j}^t}$ of cow $\boldsymbol{c_j}$ are calculated by

$$\textbf{Approached index}: \ \boldsymbol{w_{c_j}^t = \lambda \frac{w_{n_{c_j}}^t}{\sum_{c_j \in C} w_{n_{c_j}}^t} + (1-\lambda) \frac{w_{d_{c_j}}^t}{\sum_{c_j \in C} w_{d_{c_j}}^t}}, \quad (2)$$

where $w_{n_{c_j}}^t$ are the weights based on the community transition, $w_{d_{c_j}}^t$ are the weights based on distance to the target cattle and λ is the coefficient which indicates the rate of two weights.

The approached index $S_{c_i}^t$ at time t is calculated by

$$\boldsymbol{S_{c_i}^t = \frac{\sum_{c_j \in C} w_{c_j}^t \times S_{c_i,c_j}^t}{\sum_{c_j \in C} w_{c_j}^t}}, \quad (3)$$

where S_{c_i,c_j}^t is the approached index between 2 cows and $w_{c_j}^t$ is the weight. This expression adopts the form of weighted mean and thus the sum of all weights become 1, which are expected to prevent community size from affecting values. The rest of the process adopts from the previous year's (S, Fukumoto et. al, 2018). In short, we calculate the mean $y_{c_i}^t$ for T from time t for the purpose of detecting the state that the target cattle are abnormally approached by any other cow. The mean $y_{c_i}^t$ is calculated by

$$\boldsymbol{y_{c_i}^t = \frac{k}{T} \sum_{l=0}^{T/k} S_{c_i}^{t+k \times l}}, \quad (4)$$

where k is the interval of GPS acquisition. We express this $y_{c_i}^t$ as the approached value in this paper. The range of the value becomes $-1 \leq y_{c_i}^t \leq 1$ as same as the cosine similarity.

3. EXPERIMENT RESULT AND EVALUATION

3.1 Experimental setting

We put GPS collars on the cattle which have been bred at the Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University in order to examine the possibility of estrus detection. The position of each cow was measured for eight months from May 1, 2018, to December 31, 2018. The measurement was conducted for thirty-six cattle during this period. The detection interval for location information by GPS was 5 seconds. Communities are generated in every $Y = 10$ min. The weights based on community transition are determined by the community of latest $X = 18$ times. We adopted three kinds of the coefficient $\lambda = 0.0, 0.5, 1.0$. The interval of calculating approached values is $T = 180$ min and we calculated the approached value every 10 min. We adopted the local-level model which is a kind of the state-space model for anomaly detection (Durbin et al., 2012). We evaluate the proposed method by $\mathrm{Precision}$, Recall, $\mathrm{F\text{-}measure}$. The detector examines whether the cattle are in estrus or not every half a day. We compare the case when pedometers are used in conjunction with the proposed method with the case when pedometer are used solely. The pedometers have the same functions as previous year's (Fukumoto et al., 2018).

3.2 Result and Discussion

Table 1 shows each $\mathrm{Precision}$, Recall, $\mathrm{F\text{-}measure}$ of the case when pedometers are used solely and are used with the approached value (no weight, $\lambda = 0.0, 0.5, 1.0$).

Table 1. Evaluation scores of each method

Method	Precision	Recall	F-measure
Pedometer only	0.378	0.933	0.538
Pedometer+Approached value (no weight)	0.529	0.600	0.563
Pedometer+Approached value ($\boldsymbol{\lambda = 0.0}$)	0.579	0.733	0.647
Pedometer+Approached value ($\boldsymbol{\lambda = 0.5}$)	0.524	0.733	0.611
Pedometer+Approached value ($\boldsymbol{\lambda = 1.0}$)	0.526	0.667	0.589

During the data acquisition period, there were thirty-seven days on which the artificial insemination was performed while both GPS data and the pedometer data were obtained. Among those days, pregnancy was confirmed fifteen times during that period. The increase in the number of steps occurred fourteen times out of the fifteen. This fact confirmed the improvement of $\mathrm{F\text{-}measure}$ due to the proposed method. Moreover, the approached value with weight has higher $\mathrm{F\text{-}measure}$ than that without weight every time. Especially, the weight based on the community transition has the highest $\mathrm{Precision}$, Recall, $\mathrm{F\text{-}measure}$. Figure 4 shows the result of the pedometer, the approached value without weights and the approached value with weights. We performed the artificial insemination to the cow on 24th, October and later on, the cow became pregnant . This indicates the cow was in estrus during the period framed in red. In addition, the artificial insemination was also carried out in the morning of 29th, August and 17th, September. There are sharp increases in the approached value during both of these days. In such cases, even though the artificial insemination failed, this peak may also resemble estrus because these periods corresponds to that of cow's estrus. Thus, the positive effect of quantification could be confirmed. Particularly, the weights based on community transition are working well and eliminating unintended approaching-approached behavior to some extent.

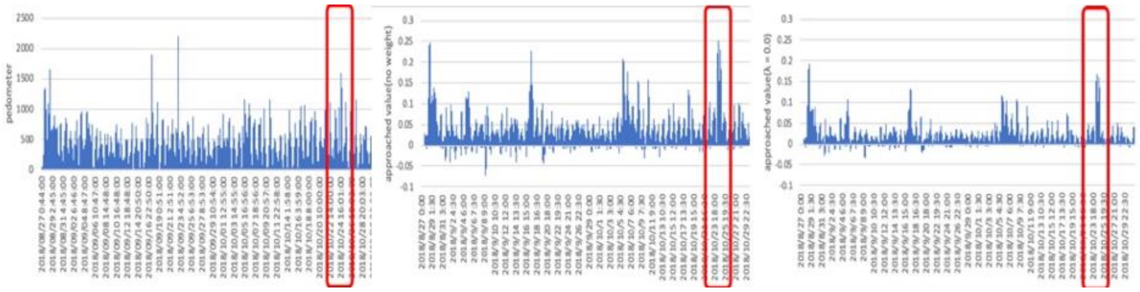


Figure 4. Result of cattle 20295 from pedometer (left), approached value with no weight (middle) and approached value with weight (right)

The weights based on cattle distance are also feasible but its effectivity is inferior to that of the weights based on community transition possibly due to positioning error. The communities were generated by clustering based on the time when the cattle distance became less than 10 meter, which is the reason the weights based on community transition are robust to temporarily positioning error. On the other hand, the weights based on cattle distance were generated by considering only the distance itself and thus, they are vulnerable to the positioning error. Therefore, we conceive that the approached value ($\lambda = 0.0$), which adopts the weights based on the community transition, is the most effective method.

4. Conclusions

In this paper, we focused on the approaching-approached behavior and applied the knowledge that the cattle in estrus are more likely to be approached. In quantification, we added weights to consider cattle's community and distance. When our method was combined with the conventional method using the pedometer, F-measure rose by 2.5% and we could grasp the estrus cycle more precisely. There is a certain effect of employing weight, which raised F-measure by another 8.4%. However, we assume that our method is much effective because we did not examine some missed estrus. This is because in this work, our method was used together with the data of pedometer and the implicit estrus has been observed which the pedometer itself cannot detect. Thus, it is necessary to verify the usability of the proposed method by conducting artificial insemination for the implicit estrus of cows detected from only the approached value, which cannot be seen from the pedometer.

Acknowledgement

This work was supported by JST CREST Grant Number JPMJCR1682.

References

Roelofs, J, B., Eerdenburg F, J, C, M, van., Soede, N, M. and Kemp, B. (2005) 'Various behavioral signs of estrous and their relationship with time of ovulation in dairy cattle', Theriogenology, 63, 5, pp. 1366–1377.

Blondel, V, D., Guillaume, J, L., Lambiotte R. and Lefebvre, E. (2008) 'Fast Unfolding of Communities in Large Networks', Journal of Statistical Mechanics, 2008, pp. 10008.

Fukumoto, S., Nishide, R., Takaki, Y., Ohta, C., Oyama, K. and Ohkawa, T. (2018) 'Quantifying the Approaching Behaviors for Interactions in Detecting Estrus of Breeding Cattle', SoICT, pp. 235-242. doi: 10.1145/3287921.3287944.

Durbin, J. and Koopman, S, J. (2012) 'Time Series Analysis by State Space Methods: Second Edition', 38, Oxford University Press, Oxford.

BLOCKCHAIN: A TOOL FOR SUPPLY CHAIN CERTIFICATION

Crescenzio Gallo[1], Nicola Faccilongo[2], Francesco Conto'[2] and Nino Adamashvili[2]

1Department of Medical and Surgical Sciences, University of Foggia, Italy

2Department of Economics, University of Foggia, Italy

crescenzio.gallo@unifg.it, nicola.faccilongo@unifg.it, francesco.conto@unifg.it, nino.adamashvili@unifg.it

ABSTRACT

Blockchain technology allows to develop product certification processes, offering greater guarantees on the history of food, from the collection of raw materials to the flow between operators in the supply chain and up to the final consumer. This new technology can therefore represent a strategic element for the agro-food supply chains (and not only). On one hand, in order to defend production from climatic variability, prompt action to manage production factors can be taken, as well as to contain costs and minimize production and environmental risks. Additionally, to guarantee brand safety and protect specific territorial features from illegal competition with counterfeit products. In this work authors introduce the Blockchain technology and its usage and implications for the supply chain certification. Initially its main aspects are illustrated together with its historical origins; in the next sections the specific issues related to supply chain and the certification of products are examined, and then we propose an adoption scheme of blockchain technology for fostering a sound and effective certification of the production in the entire supply chain.

Keywords: Blockchain, Traceability, Trustworthiness, Supply chain, Certification

1. INTRODUCTION

Blockchain innovation is surrounded by the fast-growing industrial ambience. The hype about this technology has to do with its effectiveness for achieving and maintaining integrity in distributed peer-to-peer systems, with the capacity of reshaping whole existing industries by disintermediation. In this system, each node maintains its own copy and, following the blockchain's consensus algorithm, all nodes reach one consistent version of the shared ledger.

The blockchain technology through decentralization, trustworthiness and collective maintenance ensures a reliable register of well-systemized information (Karikari et al., 2019). Taking into consideration all these characteristics of blockchain, it is a good foundation for industries to use in order to simplify product tracking along supply chain, certification process and establishment of new relationships.

In fact, currently the attention is focusing on the food traceability since there is an increasing need of certification of origin and quality of product. Thanks to Blockchain technology, all the players in the supply chain would no longer need to use "paper documents" or rely on central or third-party entities for the certification of the various information and documents produced during the different stages of the supply chain.

2. BACKGROUND ANALYSIS

2.1. The Blockchain Technology

The basic idea behind the blockchain technology was born in 1991, when Huber and Stornetta (1991) in their work described how to sign documents digitally in a way to be easily shown that none of the documents signed in the collection had been modified. This system was first used for digital currency in 2008 by an anonymous programmer Satoshi Nakamoto.

Blockchain (Prathyusha et al., 2018) is a peer-to-peer distributed database of transactions (ledger) secured by cryptography. Transactions are registered in a continuously growing list (chain) of records (blocks) which are append-only (it is possible to write a new block at the end of the structure only) and immutable (or at least very hard to change). The blockchain can be updated only via agreement among peers (consensus). For obtaining identical results when aggregating transactions, it is mandatory to preserve the order in which transaction data are added to the history. Any transaction not being part of that history cannot be trusted on.

In order to maintain integrity, only those transactions that are formally/semantically correct and authorized, get added to the blockchain. Before adding a new block into the blockchain, it has to be validated through the so-called hash mechanism. A hash function transforms any data into a sequence of fixed length figures (characters and/or bits), regardless of the size of the input data. Cryptographic hash functions create a (quick) digital digest for any data in deterministic form (same input => same output), with pseudorandom behaviour (a little change in input data turns into a big change in the output digest) and one-way usage (no reverse-engineering; it is not possible to get the original data from the output digest). If the data had been changed, both, cryptographic hash value and data location, would be different and the hash reference would become invalid.

2.1.1 Blockchain data structure and immutability

The first block of the chain is called the "Genesis block". Each block stores the following information:

- **Index**: the position of the block in the chain (the genesis block has index 0).
- **Timestamp**: the time when the block was created (used for keeping the blockchain in the correct order).
- **Hash**: a numeric value that uniquely identifies the block's data (the digital fingerprint of data) and has the following properties: Fixed length (typically 256 bits); Easy to compute; Not reversible (it is not possible to get the original data from hash); If data changes, hash changes (in particular, small change in data leads to big change in hash).
- **Previous hash**: the hash reference of the previous block in the chain.
- **Data**: the data (transactions) stored in the current block. Changing the data will change the hash, so it become invalid. Subsequent blocks will also be invalid, leading to a cascading invalidation of blocks in the chain.
- **Nonce**: the number of iterations needed to find a valid hash, i.e. a hash with a required predefined number of leading zeroes, called *difficulty* (3 in the example shown in fig. 1).

Figure 1. A hash value (with difficulty 3)

The hash of the current block comes from the combination of its index, timestamp, data and nonce and the previous block's hash. Mining is the process of finding a valid hash for the block. A new block being added to the blockchain needs to meet the following requirements:

- Block index = latest block index + 1;
- Block previous hash = latest block hash;

- Block hash meets difficulty requirement (no. of leading zeroes);
- Block hash is correctly calculated.

Different peers on the network are simultaneously trying to add blocks to the blockchain. For this reason, new blocks need to be validated before becoming part of the chain. Peers ask each other to find who has the most up-to-date blockchain version.

If a block is modified, it and its subsequent blocks become invalid and are rejected by the peers on the network. Earlier blocks will be harder to alter because there are more subsequent blocks to re-mine. The only way to mutate a block would be to mine the block again, and all the blocks after. Since new blocks are always being added, it’s nearly impossible to mutate the blockchain.

2.2. Food supply chain: the risk of fraud and tracking production through Blockchains

The interest in the concept of the Supply Chain has steadily increased since the 1980s, when companies from different sectors have verified the advantages resulting from the construction of collaborative relationships within and outside their own organization. According to Van der Vorst *et al.* (2007), the supply chain is a sequence of processes (decision-making and executive) and flows of materials, information and money, which occur at different stages of the journey of products and services from the point of production to the point of consumption and cross the borders between the organizations involved. La Londe & Masters (1994) proposed and Lambert et al. (1998) follow a slightly different logic, more focused on the role of active subjects in the supply chain. It identifies the Supply Chain as a set of enterprises or identities through that brings products or services to the market.

In today's constantly transforming world, agriculture management as almost all other fields is becoming more complex. A more frequent need for intermediaries causes the supply chains to get longer (Mylrea & Gourisetti, 2018). Hence, the number of documentation and their copies for all involved parties increases that makes it challenging to understand the origin of products (Mei & Dinwoodie, 2005). Despite the fact that agri-food supply chains are already digitalized (cloud computing, artificial intelligence, internet of things), there are a lot of remarkable inefficiencies in the farming operations, distribution and selling. Globally, the cost of food fraud is US$40 billion every year (PWC, 2017). From a process point of view, companies still suffer with a lack of integration and organization of services, especially regarded to certification processes. The necessity of these companies concerns the simplification of the control and certification procedures and the reduction in the hours/person employed for this activity. A further critical element that can be found within the organic supply chain is represented by the problem deriving from the conflict of interest generated by the controlled-controller relationship that has led over the last few years to a loss of reputation in consumers related to the certification system and the organic supply chain as a whole.

So, traceability becomes the crucial factor in agri-food supply chain, in terms of the ability to trace and follow the history of final product in the supply chain, and possessing necessary information on all stages of production process, warehousing, distribution and trade (Aung & Chang, 2014). It will minimize possible human error reducing the steps and procedures by directing the company and the certifier towards a process automation. The blockchain, in this sense, is opening a world of opportunities that will offer a huge sustainable competitive advantage.

IBM demonstrated the use of blockchain as a fresh food tracking system for Walmart (Hackett, 2017), tracing the movements of each individual product from harvesting to packaging, from cold storage to sorting centers. In Italy, Barilla has adopted a similar process (Morabito, 2017; Petek & Zajec, 2018) to follow the growth of basil plants used for pesto. It starts with sowing and continues with delivery to haulers up to the factory where the basil is transformed into pesto. Everything is under control and not a single batch can go unnoticed by the company. The aim is to strengthen the image of quality of the raw material along the entire chain and the anti-counterfeiting control. The interest in Blockchain technology is stronger in the case of Protected Geographical Indication (PGI), Protected Designation of

Origin (PDO) and organic products since the detailed information about steps passed by the product gets crucial importance.

3. SUPPLY CHAIN TRACKING SYSTEM BY BLOCKCHAIN TECHNOLOGY

Carrying out a comparative analysis of supply chain efficiency ex ante and ex post the IT technology introduction, La Sala et al. (2017) argue the importance of efficient use and sharing of information and introduce the scheme of the IT system adopted (fig. 2). The scheme shows that the interaction between different stakeholders (supplier companies, farmers, processing companies, distribution company, retailer stores) of the industry is much more simplified using advanced technologies.

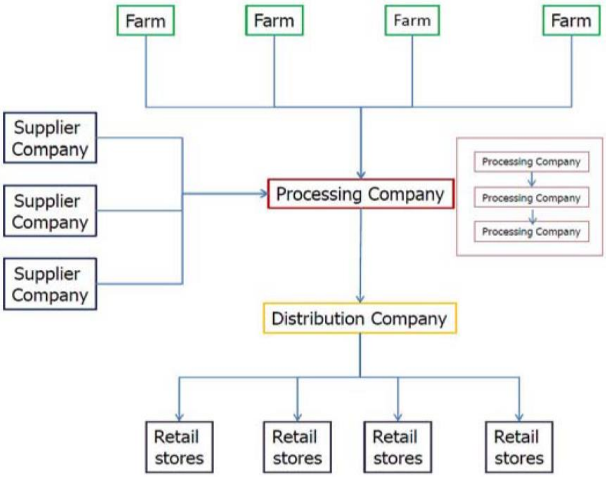


Figure 2. Scheme of the IT system adopted

Started from this scheme, authors propose the model where authenticity of all information is proved by Blockchain technology. It is supposed, that all supplier companies and farmers use Blockchain for registering the transactions. So, all the activities carried out by each stakeholder are registered on the blockchain: after the activity has been done, the information is declared, then proved and protected by hush. Thus, the information provided by supplier companies and farmers are trustful as falsification is almost impossible. Subsequently, processing companies get trustworthy information about input such as the geographical indications, weather conditions, soil management, seeds' nutrients etc. Then similarly they register techniques and technologies of processing and add new block to a chain (or make new Genesis block for specific product as it is shown on the fug. 3). Next, distribution company describes the transportation conditions, safety, temperature, vehicle, delivery and adds new block. After that, retail stores register some other information like delivery details, storage, safety and additional block is constructed. Finally, as consumers as regulatory authorities can track all steps passed by specific product simply by scanning its QR code.

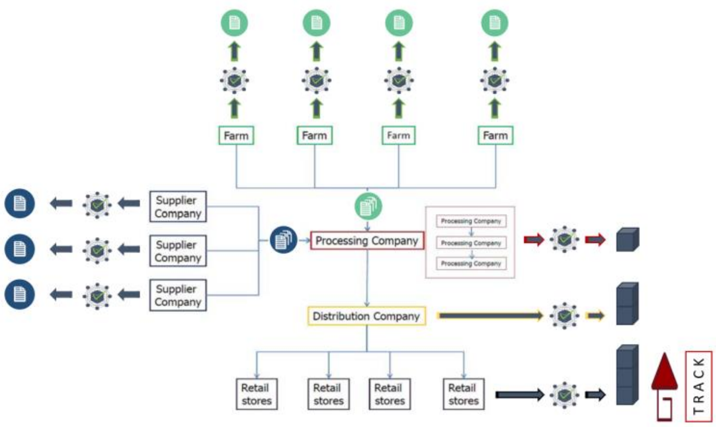


Figure 3. Supply chain tracking system by Blockchain technology

4. Discussion

In the current agri-food market, there is the increasing need to certify the origin and quality of products using a new traceability model capable of setting a quantitative and qualitative model of information shared along the supply chain. In this way, the problems related to food safety would be solved, at the same time ensuring the maximum transparency for final consumers and regulatory authorities. This requirement is even more important in the organic farming sector where the control and certification of all stages of the supply chain become important to guarantee compliance with the certification schemes that the correct agronomic and processing practices related to organic farming are done.

Blockchain technology would allow agri-food information to be shared in a reliable and secure environment, also guaranteeing its immutability. All the players in the supply chain would no longer need to use "paper documents" or rely on central or third-party entities for the certification of the various information and documents produced during the different stages of the supply chain. Before buying a product, consumers will be able to verify all the data and consult the certified documentation: not only the origin, but also, for example, if the frozen food has been transported safely at the right temperature.

5. Conclusion

Blockchain technology improves the monitoring of production process and reduces the work intensity necessary to guarantee the quality and certification processes. A further effect on the production processes and transformation is giving the ability to all the players in the supply chain to participate in a shared manner in the construction of quality processes. The implementation of the technology and related management model guarantee the nodes mapping of the supply chain where the gup takes place. Through this mapping it will be possible to provide the data for better control and for improvement of the processes. The monitoring will allow to identify any inefficiencies in the supply chain and the consequent intervention for their minimization. There will be an intensification of information exchanges between operators and consumers which will determine an increase in awareness of the final consumer about the organic product and the value of the supply chain processes.

The limitation of this paper is its descriptive character and the future research plans of the authors is to examine the blockchain technology in practice.

References

Aung, M. M., & Chang, Y. S. (2014). Traceability in a food supply chain: Safety and quality perspectives. *Food Control 39*, 172-184.

Hackett, R. (2017). Walmart and 9 Food Giants Team Up on IBM Blockchain Plans. *Fortune.*

Karikari, A., Zhu, L. & Dara, R. 2019, "Blockchain: The next step in the development of the Internet of Things", *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018*, pp. 341.

Katz, J., Menezes, A. J., Van Oorschot, P. C., & Vanstone, S. A. (1996). *Handbook of applied cryptography*. CRC press.

Ku, R. S. R. (2017). The creative destruction of copyright: Napster and the new economics of digital technology. In *Copyright Law* (pp. 207-268). Routledge.

La Londe, B. J., & Masters, J. M. (1994). Emerging logistics strategies: blueprints for the next century. *International journal of physical distribution & logistics management*, 24(7), 35-47.

Lambert, D. M., Stock, J. R., & Ellram, L. M. (1998). *Fundamentals of logistics management*. McGraw-Hill/Irwin.

La Sala, P., Faccilongo N., & Fiore M. (2017). Integrated management of the PGI 'Matera' Bread chain. World Review of Entrepreneurship, Management and Sust. Development, 13 (5/6), 665-683.

Mei, Z. & Dinwoodie, J., (2005). Electronic shipping documentation in China's international supply chains. *Supply Chain Management. Vol 10, 3*, 198-205

Morabito, V. (2017). Business Innovation Through Blockchain. *Cham: Springer International Publishing*.

Mylrea, M. & Gourisetti, S.N.G., (2018). Blockchain for Supply Chain Cybersecurity, Optimization and Compliance. *Proceedings - Resilience Week 2018, RWS 2018*. Article number 8473517, 70-76

Nakamoto, Satoshi. (2008). *Bitcoin: a peer-to-peer electronic cash system*. https://bitcoin.org/bitcoin.pdf.

Petek, I., & Zajec, N. (2018, April). Collaborative Intelligence and decentralized business community building–Potentials in Food/Nutrition sector. In *2018 IEEE 34th International Conference on Data Engineering Workshops (ICDEW)*. IEEE.

Prathyusha, T., Kavya, M., & Akshita, P. S. L. (2018). Block Chain Technology. http:://academicscience.co.in

PWC. (2017). Food fraud vulnerability assessment and mitigation. *Retrieved from Are you doing enough to prevent food fraud?: https://www.pwccn.com/en/migration/pdf/fsis-food-fraud-nov2016.pdf*;

Van der Vorst, J. G., Da Silva, C., & Trienekens, J. H. (2007). *Agro-industrial supply chain management: concepts and applications.* FAO.

DEFINITION OF REFERENCE MODELS FOR FUNCTIONAL PARAMETERS AND PRICE FOR COMBINE HARVESTERS

Tatevik Yezekyan[1], Giannantonio Armentano[2], Samuele Trestini[1], Luigi Sartori[1] and Francesco Marinello[1]

[1]Department of Land, Environment, Agriculture and Forestry, University of Padova, Via dell'Università 16, 35020 Legnaro, Italy

[2]Edizioni L'Informatore Agrario srl, Via Bencivenga - Biondani, 16, 37133 Verona, Italy

tatevik.yezekyan@phd.unipd.it, g.armentano@informatoreagrario.it, samuele.trestini@unipd.it, luigi.sartori@unipd.it, francesco.marinello@unipd.it

ABSTRACT

Operational and functional parameters of agricultural machines have essential importance and direct influence on farm fleet definition or optimisation (both for tractors and implements), machinery planning and management. Decision support systems and models have been developed in the past mainly analysing and quantifying farm costs while information lack existing correlation with functional parameters such as weight, power, efficiency, etc. Conversely, such parameters play an important role, not only with direct and indirect costs but also with agronomic and environmental performances. Such aspects are highly essential and to some extent critical for capital intensive machines such as combine harvesters.

In the current research the functional parameters of combine harvesters have been analysed (power, weight and tank capacity) providing linear models for the variables which exhibit the highest predictive potential. Highest correlation exhibited power of the machine in a relation with price (r = 0.91) and tank capacity (r = 0.90), which allows to perform forecast analyses related to the evaluation and prediction of costs and performances, thus contributing to the optimisation of the fleet selection process and investments.

Keywords*:* combine harvester, functional parameter, reference model, fleet management, statistical analyses.

1. INTRODUCTION

Management of agricultural systems, machinery and operations rely on complete understanding of operational functions, parameters, and processes. Coordination of planning, decision making and application of determined solutions are based on specifications of the farm and crop with the purpose to meet operation performance requirements and expected profit (Peart & Shoup, 2004). Agriculture, like every business, is supposed to be profitable. An economic evaluation of facilities and justification of preferences related to the investments based on the actual needs create the logical background for adjustment of the best profitable management approach.

Agriculture management is based on composite interaction of available data and resources. Identification of optimal synergy is very complex due to the machinery system combination with agronomic, biological and climatic features (Kaspar et al., 2003; Søgaard & Sørensen, 2004). Complete understanding and consideration of each parameter within the boundaries of the actual needs, their interaction effect on the other parameters and the outcome of the operation might increase the potential of performance, ensure sustainable production and profit.

Continuously advancements in the technology on the farm, in processing, and in agribusiness lead to the development of farm management tools and systems for ensuring successful performance of agricultural operations and machines (Kortenbruck et al., 2016; Peart & Shoup, 2004; Vougioukas et al., 2011). At the same time, intensification of farm practices, saturation of machinery construction, specifications and features lead to more complex decision-making phase, machinery selection process and adjustment of innovative solutions. The adoption of management strategies for enhancement of the efficiencies applied in mechanisation system lead to the modification of cost structure and price policy, impacting decision-making regarding fleet accomplishment and machinery. To predict those impacts and to forecast outcomes of decision alternatives modern agricultural management challenges to a problem-solving framework utilising mathematical models and computer software packages, simulations, spread sheets, linear and non-linear programming, scheduling routines, etc. (Bulgakov et al., 2015; Camarena et al., 2004; Kaspar et al., 2003; Pezzuolo et al., 2014; Rodias et al., 2017; Sopegno et al., 2016). Eventually, all these models and methods are called to simplify and support farms with the most sensitive and requested issues related to the fleet, farm organisation and overall crop production, however, diversity and complexity of available systems not always meet the needs and expectations of real farms (Sopegno et al., 2016). However, machinery management approaches and application diversity available in the market lack the possible correlation of technical parameters of the machines and their economic prerequisites, relations and consequences. Additionally, most of the methods are limited to the research level or applied in the restricted number of experimental farms. While, identification of parameters' performance criteria and their relation to the operation, and evaluation of correlations might create a fundamental base for justification of relatively high investments, operation cost definition and environmental protection. Optimisation of costs and maximisation of efficiencies have primary importance for modern farm management, and it is more obvious and critical in the case of harvesting operation with large capital-intensive investment on the harvesting machines.

Current research is aimed to analyse functional parameters (power, weight and tank capacity) of combine harvesters, to define the impacts of technical and design parameters, to provide linear regression models for the variables which exhibit the highest predictive potential, thus contributing to the optimisation of the fleet selection process, evaluation and prediction of costs and performances.

2. MATERIALS AND METHODS

Combine harvesters are complex agricultural machines intended for crop harvesting operation and designed for three crop processing modules: threshing, separation and cleaning processes (De Baerdemaeker & Saeys, 2013; Hermann, 2018). The threshing system consists of a transverse (conventional) or longitudinal rotor (axial) with threshing elements to separate the grain kernels from the heads, stems or pods. The conventional harvesting machines have a transverse threshing rotor and a straw walker for separation, the axial combination has one or two parallel rotors with threshing elements in front and separation in the rear end, and the hybrid construction of the combines has a transverse threshing rotor and a longitudinal separation rotor. In the modern combine harvesters, the threshing and separation systems are designed as one coherent system. The performance of combine harvester is characterised by grain loss and damage, cleanliness and straw quality. Enhancement of combine harvester's performance is considered as a complex optimisation issue due to a large number of parameters and interdependencies, which may have nonlinear correlations and be even conflicting (Hermann, 2018).

The current study is based on the analyses of 120 models of combine harvesters available in the modern agricultural machinery market intended to harvest a different kind of grain crops. Dataset consists of the variation of combines with conventional, axial and hybrid threshing systems equipped for hillside harvesting with and without a self-levelling control system. For a definition of the most correlated functional parameters of harvesting combines and their influence on the performance of the machine and price, data were analysed with the application of Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) software. Power, weight, tank capacity and list price of the machines, according to the market, were investigated as main variables for applied statistical analyses (Table 1).

Table 1. The range for minimum and maximum values of the considered variables according to the dataset

Variable	Value
Power, kW	110 - 480
Tank capacity, L	4200 - 14500
Weight, kg	7600 - 19500
Price VAT excl., k€	130 - 595
Weight power index, kg kW^{-1}	35.7 – 87.0

Dependencies between considered variables were studied with the application of linear regression analyses. The relevance of the models was quantified by means of correlation studies and dependences modelled according to the linear characteristic equation:

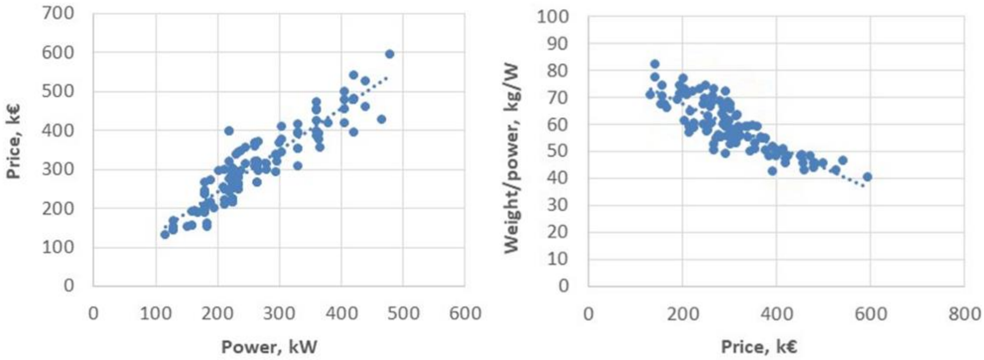
$$y = m_1 \cdot x + q \qquad (1)$$

where *x* and *y* represent respectively independent and dependent variables, *mi* is the slope (or linear coefficient) related to the *i*-th independent variables, *q* is the intercept between *y* and *x* variables.

According to the linear regression analysis data reported regarding Pearson correlation coefficient *r*, slope *m* and intercept *q* of the linear models. The simplicity of linear models allows their wide application and consideration by interested parties (farmers, farm management applications and agencies, authorities, etc.) for justification of machines parameters' requirements and forecasting. The liner models considered as not large models that provide a high degree of detail and precision, nevertheless, the models are based on the defined functional parameters and represent robust and realistic output, provide sufficient details for decision making and selection optimisation.

3. RESULTS

Data collected for the 120 models were statistically analysed as discussed above. A graphical representation as reported in Figure 1 already gives evidence of existing correlations between variables, in general acceptably approximable through linear trends. In particular, high linearity can be recognised between power, weight and price, as also evidenced by the following analyses.



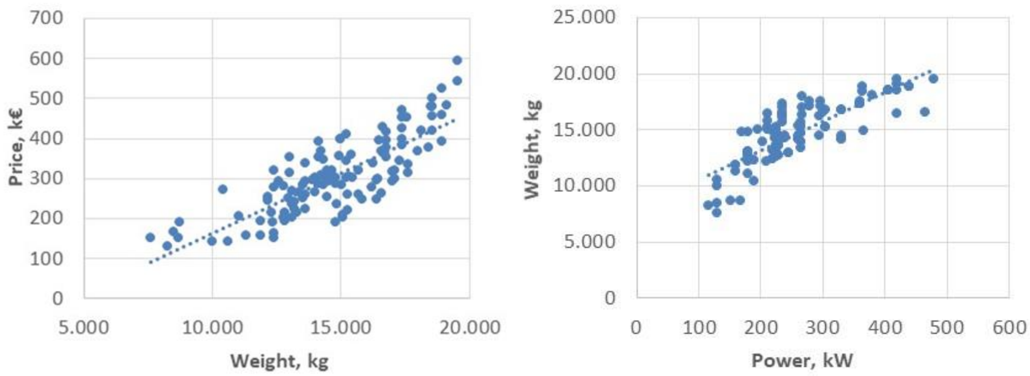


Figure 1. Graphical representation of most interesting correlation between studied variables

The linear modelling and correlation analysis for combine harvesters exhibited relatively high Pearson coefficients between all considered variables. The highest correlation was found between price and power (r = 0.91), as well as between power and tank capacity (r = 0.90). Slightly lower values could be recognised between weight and tank capacity (r = 0.8), and between price and weight (r = 0.81), as reported in Table 2.

Compared to previously published linear regression analyses performed on seeding machines (Tatevik et al., 2018) and sprayers (Yezekyan et al., 2018), combine harvesters models exhibit in general higher correlation between performance parameters. This can be probably explained as a consequence of the lower number of companies that compete in the market: such condition most probably reduces competitiveness and increases homogenization of performances and constructive principles.

Table 2. Correlation matrix of Pearson coefficient *r* for functional parameters of combine harvesters

	Tank capacity, L	Weight, kg	Power, kW	Price VAT excl.,€
Tank capacity, L	1	0.8023	0.9059	0.8220
Weight, kg		1	0.8212	0.8084
Power, kW			1	0.9102
Price VAT excl.,€				1

Linear models were estimated for all the variables' combinations according to the linear regression analysis. Results are summarised in Table 3. For combine harvesters, it can be noticed how for each additional tonne of the machine, a power supply of 26 kW has to be counted, while for the same weight, a volume of tank capacity of about 0.66 cubic meter has to be taken into account. With regard to needed investment, an average price of 1070 € has to be considered for each kW of power. The same table reports coefficients of determination R^2 and standard errors for the same models. Dealing with linear models, coefficients of determination are clearly in agreement with Pearson coefficients. Standard errors are in general relatively low, with higher uncertainty levels especially in the case of linear regressions based on weight. However, in particular, in the case of high power/high weight machinery, the models provide good forecasting results.

Table 3. Correlation matrix of Pearson coefficient *r* for functional parameters of combine harvesters

	Price	ε	R^2
Power	$P = 0.0008 \cdot Pr + 23.11$	33.9	0.8285
Weight	$M = 0.0216 \cdot Pr + 8120$	1510	0.6535
Capacity	$C = 0.0181 \cdot Pr + 3686$	1203	0.6757
	Power	**ε**	**R^2**
Weight	$M = 25.85 \cdot P + 8023.$	1467	0.6744
Capacity	$C = 23.39 \cdot P + 3128$	895	0.8207
Price	$Pr = 1070 \cdot P + 27970$	39500	0.8285
	Weight	**ε**	**R^2**
Power	$P = 0.0261 \cdot M - 124.3$	46.5	0.6744
Capacity	$C = 0.6579 \cdot M - 485.57$	1256	0.6436
Price	$Pr = 30.19 \cdot M - 138590$	56390	0.6535
	Capacity	**ε**	**R^2**
Weight	$M = 0.9782 \cdot C + 5740$	1531	0.6436
Power	$P = 0.0351 \cdot C - 62.92$	34.5	0.8207
Price	$Pr = 37.43 \cdot C - 38270$	54540	0.6757

For different groups of machines: *C*: tank capacity (L); *M*: weight (kg); *P*: power (kW); *L*: working width (m); *Pr*: estimated price (€)

4. Discussion and conclusions

Many farm management solutions provide the transition of collected information into knowledge. However, there is an actual strategic issue in the model building, to develop generalised and sufficient predictive model with simple and robust forecasting abilities. Modelling approach needs to meet actual needs raised from required interests and meet the expectations of applicants. From the current point of view, developed linear models for combine harvesters allow making simplified calculations of main technical requirements related to the machine power, weight and capacity. The latest allows having a better overview of the farm organisation, operation management and performance prediction, and to have a better arrangement of work, timeliness and independence in scheduling individual operations. According to the performed analyses the highest impact on the price formation exhibited the power of the machine, thus initial forecast of the combine harvester investment needs to consider 30 k€ and increasing by 1100 euro for each kW of the harvesting equipment. Additionally, considering the corresponding correlation of capacity of 0.35 m^3 and weight of 26 kg per each kW.

The linear construction of the models here reported allows further implementation and combination with farm management systems and decision-making approaches as initial data for functional parameters definition and price prediction. Furthermore, price forecasting and definition of initial investment might allow the parties involved arrive at the economically practical decision of operation management definition related to the machine purchase, rent or leasing, which is crucial in the case of self-propelled harvesting equipment, which requires huge investment and has comparatively very high prices in the market.

References

Bulgakov, V., Adamchuk, V., Arak, M., & Olt, J. (2015). Mathematical Modelling of the Process of Renewal of the Fleet of Combine Harvesters. Agriculture and Agricultural Science Procedia. https://doi.org/10.1016/j.aaspro.2015.12.027

Camarena, E. A., Gracia, C., & Cabrera Sixto, J. M. (2004). A Mixed Integer Linear Programming Machinery Selection Model for Multifarm Systems. Biosystems Engineering, 87(2), 145–154. https://doi.org/10.1016/j.biosystemseng.2003.10.003

De Baerdemaeker, J., & Saeys, W. (2013). Advanced control of combine harvesters. IFAC Proceedings Volumes (IFAC-PapersOnline). https://doi.org/10.3182/20130828-2-SF-3019.00069

Hermann, D. (2018). Optimisation of Combine Harvesters using Model-based Control. In DTU Elektro.

Kaspar, T. C., Colvin, T. S., Jaynes, D. B., Karlen, D. L., James, D. E., Meek, D. W., … Butler, H. (2003). Relationship between six years of corn yields and terrain attributes. Precision Agriculture. https://doi.org/10.1023/A:1021867123125

Kortenbruck, D., Sapounas, A. A., Griepentrog, H. W., Paraforos, D. S., Ziogas, V., Vassiliadis, V., & Stamkopoulos, K. (2016). A Farm Management Information System Using Future Internet Technologies. IFAC-PapersOnLine. https://doi.org/10.1016/j.ifacol.2016.10.060

Peart, R. M., & Shoup, W. D. (2004). Agricultural Systems Management Ontimizing Efficiency and Performance. Marcel Dekker Inc., New York, pp. 280

Pezzuolo, A., Basso, B., Marinello, F., & Sartori, L. (2014). Using SALUS model for medium and long term simulations of energy efficiency in different tillage systems. Applied Mathematical Sciences. https://doi.org/10.12988/ams.2014.46447

Rodias, E., Berruto, R., Bochtis, D., Busato, P., & Sopegno, A. (2017). A computational tool for comparative energy cost analysis of multiple-crop production systems. Energies. https://doi.org/10.3390/en10070831

Søgaard, H. T., & Sørensen, C. G. (2004). A model for optimal selection of machinery sizes within the farm machinery system. Biosystems Engineering, 89(1), 13–28. https://doi.org/10.1016/j.biosystemseng.2004.05.004

Sopegno, A., Busato, P., Berruto, R., & Romanelli, T. L. (2016). A cost prediction model for machine operation in multi-field production systems. Scientia Agricola. https://doi.org/10.1590/0103-9016-2015-0304

Sørensen, C. A. G., Milan, M., Bochtis, D., Tieppo, R. C., & Romanelli, T. L. (2018). Modeling cost and energy demand in agricultural machinery fleets for soybean and maize cultivated using a no-tillage system. Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2018.11.032

Vougioukas, S. G., Bochtis, D. D., Sørensen, C. G., Suomi, P., & Pesonen, L. (2011). Functional requirements for a future farm management information system. Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2011.02.005

Yezekyan, T., Marinello, F., Armentano, G., & Sartori, L. (2018). Analysis of cost and performances of agricultural machinery: Reference model for sprayers. Agronomy Research. https://doi.org/10.15159/AR.18.049

Yezekyan, Tatevik, Marinello, F., Armentano, G., Trestini, S., & Sartori, L. (2018). Definition of Reference Models for Power, Weight, Working Width, and Price for Seeding Machines. Agriculture. https://doi.org/10.3390/agriculture8120186

MEASUREMENT OF INCOME RISK AS BENCHMARKING TOOL FOR DAIRY FARMS

Gašper Petelin [1] and Jaka Žgajnar [2]

[1] Datalab Tehnologije d.d., Hajdrihova ulica 28c, SI-1000 Ljubljana, Slovenia

[2] University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Groblje 3, SI-1230, Domžale, Slovenia

gasper.petelin@datalab.eu, jaka.zgajnar@bf.uni-lj.si

ABSTRACT

With increasing risks in agriculture, their management is becoming increasingly topical. Climate changes, price fluctuations and changes of institutional measures increase variability of farms' income, which can have a negative effect. Effective risk management strategy can help farmers protecting themselves from such events and help stabilize their income. Given that we are talking about various factors that influence the risks, a multistep approach is required. First, layers of risks should be considered separately. Normal risks are usually managed on the farm with different costs strategies by farmers (OECD, 2011) while market and catastrophic risks are managed by the private sector or the state. Additionally, an effective risk measurement tool should be established. On top of that benchmarking could be an effective approach for increasing farm efficiency. The present paper therefore presents a conceptual approach to use the farm management information system PANTHEON Farming as a risk measurement and benchmarking tool. It is presented on the case of Slovenian dairy farms. First goal of such tool is to use farm level data, created through bookkeeping and accounting, to measure volatility of individual financial parameters. In such a manner, farmers would get a detailed overview of the income variability on his/her farm as well as an insight to what actually causes such variability. In the next step, tool would enable user to compare their results to benchmarks or to similar results of other farms. Farms that use PANTHEON Farming could share information about risk measurement and some additional details like production type, production intensity, location, etc. Such insight would enable farmers to compare (benchmark) their results and see, if their variability in some way stands out from benchmarks of sector or from farms in same location, intensity of production, feed usage, etc. These results would also indicate to the farmer some additional information like which type of production, feed or animal breed could help reduce income variability in similar condition. An important step of risk measurement is also time frame, selected for analysis, which can have a significant impact on results. Therefore, developed tool would enable farmer to choose either he wants to review only last year, last three years or Olympic average results which would properly recalculate the variability of each parameter. Such functionality as well as whole benchmarking tool could help farmers with their business. For farmers to be able to cope with increasing risks, adequate information should be provided as well as individual risk sources should be identified. This could also increase production optimization and risk management on farm level and could therefore increase resilience of farms to exposure risk. Policy makers could also benefit on the long run, as risk management on farm level would increase and could on the one hand minimize the need for policy interference with risk management and on the other minimize the exposure of individual farms to catastrophic risks.

Keywords*:* Benchmarking, Income risk, Dairy farm, Risk measurement***.***

1. Introduction

Due to its close connection with nature, agriculture has always been a risky business (Chapman, et al., 2007). Factors like weather (Tangermann, 2011) or disease outbreaks (Aimin, 2010; Meuwissen et al. 2001) already present important drivers of farm uncertainties with additional increase in recent period. Additional elements like price risks of products (Antón, et al., 2011) as well as outputs (Aimin, 2010; Schaffnit-Chatterjee, 2010) pose another major threat for individual farm. Institutional risks must also be considered. Changes in legislation (Berg and Kramer, 2008), late payment of subsidies or environmental restrictions (Agrosynergie, 2011; Schils et al., 2007) can also make an important contribution to the income variability. Additional smaller, but not at all negligible risk elements combine factors like exchange rate and farm workforce (Hardaker, et al., 2004).

However, to be able to effectively tackle on farm risks, they first have to be identified and measured. According to Meuwissen et al. (2001), livestock producer are seeing more importance in production, price and institutional risks and hence our focus is on these groups. Additionally, on farm risk strategies should consider risk layers. Farmers are only able to manage normal risks for themselves whereas other risks, like market or catastrophic, are usually covered by other stakeholders (OECD, 2011). The main problem with this fact is that, according to Meuwissen et al. (2008), there is no clear dividing line between layers, which makes it difficult to determine when to include external tools to on farm risk management. To overcome this issue risk management tools or different researches use 15% of income drop as a threshold value for normal risks Tangermann (2011).

Within layers of normal risks that farmer can and should manage for himself (Tangermann, 2011; Lipińska, 2016), some additional characteristics have to be considered. Systemic risks (like price or institutional risks) are very hard to manage on farm level (Girdžiute, 2012; Matthews, 2010). Nevertheless, even if the farm recognizes increased exposure to such risks, they usually affect the wider area and indirectly affect functioning of such farm. Additionally, connections between risk groups should also be considered as ignoring such connections could jeopardize the relevance of the model (Lien, 2003). Positive correlation usually increases variability and, according to El Benni in Finger (2014), it usually appears in regional markets without significant impact to world markets (like Slovenia). Negative correlation is on the other hand desirable to some extent as, according to Benni in Finger (2014), it can decrease variability. Still, no matter how we consider each group of risks, it should be clear that the characteristics of the individual farm have the greatest impact on income variability.

Measured parameters, in our case variability of financial parameters, do not tell much by themselves if we do not have some values to compare them with. There are many methods of data comparison (Ronan and Cleary, 2000; Torun et al., 2018) with benchmarking being one of them. Method is simply described as comparison of ones performance with the performance of other engaged in similar activity (European Commission, 2017). Concept uses systematic approach to business improvement where best practice is sought and implemented to improve desired process (Torun, et al., 2018). Since its first efforts in the 1960, it was formally incorporated into the corporation-wide improvement efforts in 1983 (Ronan and Cleary, 2000). Nowadays, benchmarking is a very well-known concept also in agriculture. Calculating and comparing technological parameters to benchmarks or experiences of other farmers to improve performance, so called horizontal benchmarking (Franks and Haverty, 2005), is already implemented in many agricultural sectors (European Commission, 2017), especially sectors like granivore sector, arable sector and dairy production.

Despite benchmarking for the financial analysis has been identified as a tool that helps farmers identify excessive costs and inefficiencies that can solidify and stabilize the financial structure of farms (Franks and Haverty, 2005) its application is very diverse. Although it is a widespread practice in some environments, like Australia (Wilson, et al., 2005), its spread in the EU is only around 30% - 40% (European Commission, 2017). There are several reasons for this situation. One of the most important is insufficient knowledge of the benchmarking method. There is often said that benchmarking is a tool for large firms and not for small farmers. The reason for this is that key performance indicators (KPIs) should be defined by experts and that only large farms can find acceptable benchmarking partners, which could also be associated with trust issues regarding sharing of financial data (Franks and Haverty, 2005). On top of that, data collecting still represent a huge burden as over 40% of data in the EU are still sent in a non-electronic way. However, we are seeing some improvement in this area. With adoption of ICT, farmers are getting more relevant data easier, which enables better insight in an individual farm (European Commission, 2017).

Hence our motivation lies within the fact that financial data in the EU are not sufficiently used to analyze the actual state of the farm. To improve this, our analysis contains an approach that measures income risk from financial data and uses them in a benchmarking tool to compare data with other farms or defined benchmarks. The remainder of this paper is thus structured as follows: section 2 presents methodology on how the income risk measurement model was established and unified as well as how data are obtained. Section 3 represents the results of benchmarking between selected farm samples, which are further discussed in section 4. Section 5 at the end draws conclusions and describes some possibilities for future development in this area.

2. METHODOLOGY

To be able to effectively benchmark income variability, we first had to establish income risk measurement model. The model was developed from bookkeeping and accounting data of an individual farm. Standardized chart of accounts, which enables simplified implementation (Argilés and Slof, 2001) that are used for accounting, have been recognized as a useful tool to measure and analyze risks (Antón, et al., 2011). The construction of such module was made possible by the adopted chart of accounts, established by Slovenian accounting and agriculture experts (MKGP, 2014). Detailed chart of account is based on Farm Accountancy Data Network (FADN) accounting methodology and combines FADN methodology with accounting (MKGP, 2012). As a result, we can get detailed financial structure of individual farm that is comparable to FADN results.

Primary data source in our analysis are bookkeeping data from PANTHEON Farming (Datalab, 2019). As farm management information system (FMIS) software enables recording of farm events (animal birth, insemination, treatment, etc.) as well as bookkeeping (received/issued invoices, stock management, personnel records, etc.). How data are obtained and prepared was thoroughly presented by Petelin and Žgajnar (2016). On the basis of this approach we then developed a model to measure and analyze risks. In this part we used combination of chart of account and FADN methodology to prepare a hierarchical structure of variables (Figure 1). This approach and combination of different methodologies would theoretically enable farmer to compare their bookkeeping data with FADN results. Variables were then grouped and summarized to calculate revenues, variable costs, fixed costs and income.

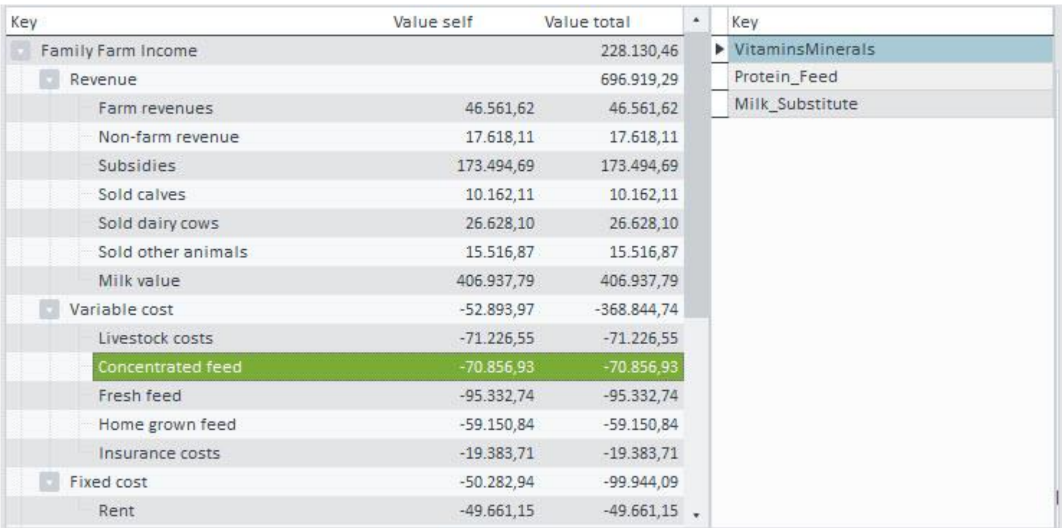
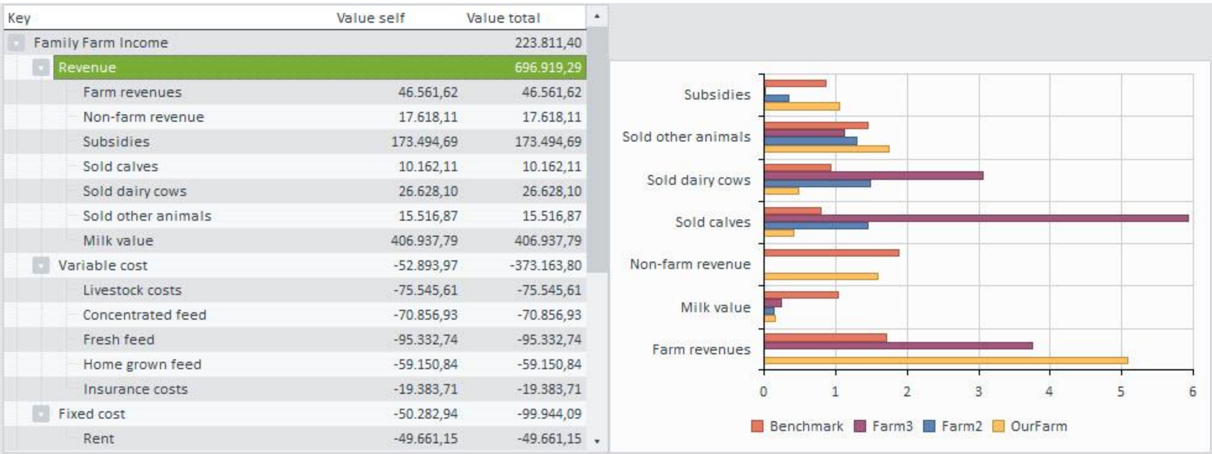


Figure 1: Hierarchical structure of income risk measurement variables (Datalab, 2019)

The last step of model construction was adapting the existing model for benchmarking. First, we calculated parameters that will serve to measure risk. As it has been previously indicated (Hardaker, et al., 2004), coefficient of variation (CV) is an appropriate measure of risks as its relative value enables comparison of different values. Benchmarking on the level of CV also brings another important advantage. As it is calculated from actual values as a relative measure, it does not disclosure any actual financial information. This is why we calculate CVs for each variable from historical data from PANTHEON Farming and share them through web service to other PANTHEON Farming users. The final result of this approach is a graphical comparison of CVs for the whole hierarchical structure of variables. Results of such analysis are described in the next chapter.

3. RESULTS

This section presents results of our analysis. Based on previously established risk measurement model, we managed to develop an approach that would enable comparison of CV from different farms. Figure 2, therefore, shows benchmarking of financial parameters. The left section of the picture shows the hierarchical structure of income that is calculated from selected variables with calculated result of selected farm. In our case, we are looking ad fixed structure of income that enables comparison of results from different farms. As was previously already described, we constructed variables based on chart of accounts and FADN variable structure.



Key	Value self	Value total
Family Farm Income		223.811,40
Revenue		696.919,29
Farm revenues	46.561,62	46.561,62
Non-farm revenue	17.618,11	17.618,11
Subsidies	173.494,69	173.494,69
Sold calves	10.162,11	10.162,11
Sold dairy cows	26.628,10	26.628,10
Sold other animals	15.516,87	15.516,87
Milk value	406.937,79	406.937,79
Variable cost	-52.893,97	-373.163,80
Livestock costs	-75.545,61	-75.545,61
Concentrated feed	-70.856,93	-70.856,93
Fresh feed	-95.332,74	-95.332,74
Home grown feed	-59.150,84	-59.150,84
Insurance costs	-19.383,71	-19.383,71
Fixed cost	-50.282,94	-99.944,09
Rent	-49.661,15	-49.661,15

Figure 2: Benchmarking of CV from different farms (Datalab, 2019)

The right section of Figure 2 shows the result of benchmarking for revenue from sale of animals. Selected parameter (Revenue) on the right is further divided on variables subsidies, sold other animals, sold dairy cows, sold calves, non-farm revenue, milk value and farm revenues. Diagram shows the CV for a farm that is using PANTHEON Farming (OurFarm), two individual farms (Farm2 and Farm3) as well as benchmark sample (Benchmark). For latest variable, we selected data from FADN sample and includes all dairy farms in Slovenia. Although we cannot draw any firm conclusions, developed approach provides an insight into the analysis that the approach enables. On the one hand we can see that variability of sale of calves and dairy cows is very high on Farm3. This indicates that flow of animals (arrivals and departures) on farm is very diverse in a sense that different animal groups are bred on farm. If farm would wand to better control its risks, it should focus more on milk production and bred animals mainly for that purpose. However, milk value shows significantly low variability. In comparison with other farms, OurFarm shows significantly lower variability of sale of calves, but a bit higher variability of sale of dairy cows. As we are seeing also with other farms, some are expressing high variability for sale of calves (Farm3) and some farm revenues (Farm3 and OurFarm) which indicates on diversity of selected sample.

4. Discussion

As has been repeatedly established before, risk management is becoming ever more topical. To be able to manage risks, they must be measured, analyzed and presented in a comprehensible manner. Since farmers do not favor the use of computer equipment (of do not have time to use it) and usually sees is as appropriate for larger farm enterprises (European Commission, 2017) the latter is probably the most important. FMIS in connection with enterprise resource planning (ERP) could prove as a useful tool for such measurement and analysis since it includes not only technical information of individual farm, but also financials. Benchmarking of these results is an additional step of risk management. To be able to follow income variability of farms from a similar environment and adopt useful practices for risk management could improve the resilience of smaller farms. Since data structure is compatible with other data sources like FADN methodology, benchmarks could be established on external data sources and included in analysis, which would enable farmers to compare their results with expected values within their sector. Nevertheless, presented results just indicate the possibility of using such a tool and require further research and development in this field.

5. Conclusions

The developed approach shows possibilities in how to use FMIS to analyze risks on individual farm and to benchmark those risks either with other farms or with benchmarks, established by public organizations or other stakeholders. This would provide benefit for all stakeholders. Farmers itself would get clear insight in risk structure in their farm which is a first step into risk management. Other stakeholders like expert organizations or state could also benefit from it on a long run as tool could provide depth insight into the structure of risk either in specific sector, location, etc., which could help to develop risk management tools.

Nevertheless, it is clear that presented contribution is only the first step towards an effective risk benchmarking tool. As stated before, characteristic of individual farm are very important in risk management hence such tool should share more information regarding specifics of individual farm. Data series should be clearly marked, possibly with option to exclude individual year, which would enable to either exclude individual bad year or analyze only specific period. Most important tool should provide long data series for risks, which would provide relevant analysis of risk. Therefore such tools should be developed and put in use rather sooner than later.

REFERENCES

Agrosynergie, 2011. *Evaluation of income effects of direct support,* Brussels: Agrosynergie.

Aimin, H., 2010. Uncertainty, Risk Aversion and Risk Management in Agriculture. *Agriculture and Agricultural Science Procedia,* Volume 1, pp. 152-156.

Antón, J., Kimura, S. and Matini, R., 2011. *Risk management in Agriculture in Canada.* 40 ured. s.l.:OECD Food, Agriculture and Fisheries Working Papers.

Argilés, J. and Slof, J., 2001. New Opportunities for Farm Accounting. *European Accounting Review,* 10(2), pp. 1-37.

Berg, E. and Kramer, J., 2008. Policy options for risk management. V: *Income stabilization in European agriculture.* Wageningen: Wageningen Academic Publishers, pp. 143-167.

Chapman, D., Malcolm, I., Neal, M. and Cullen, B., 2007. *Risk and uncertainty in diary production systems: Research concepts, tools and prospects.* Melbourne, Australasian dairy science symposium, pp. 476-490.

Datalab, 2019. *PANTHEON Farming.* www.datalab.ch (21.04.2016).

El Benni, N. and Finger, R., 2014. Where is the risk? Price, yield and cost risk in Swiss crop production. *Review of Agricultural and Environmental Studies,* 95(03), pp. 299-326.

European Commission, 2017. *EIP-AGRI Focus Group on Benchmarking: Final report,* Brussels: EIP-AGRI.

Franks, J. and Haverty, M., 2005. Benchmarking farm enterprises. *Journal of farm management,* 12(3), pp. 143-158.

Girdžiute, L., 2012. Risks in agriculture and opportunities of their integrated evaluation. *Procedia - Social and Behavioral Sciences,* Volume 62, pp. 783-790.

Hardaker, J., Huirne, R., Anderson, J. and Lien, G., 2004. *Coping with risk in agriculture.* 2nd ed. Oxfordshire: CABI Publishing.

Lipińska, I., 2016. Managing the risk in agriculture production: the role of government. *European Countryside,* Volume 2, pp. 86-97.

Matthews, A., 2010. *Perspectives on addressing market instability and income risk for farmers,* Dublin: Trinity College.

Meuwissen, M., Huirne, R. and Hardaker, J., 2001. Risk and risk management: an empirical Analysis of Dutch livestock farmers. *Livestock Production Science,* Volume 69, pp. 43-53.

Meuwissen, M., van Asseldonk, M. and Huirne, R., 2008. *Income stabilization in European agriculture; design and economic impact of risk management tools.* First Edition ed. Wageningen: Wageningen Academic Publishers.

MKGP, 2012. *Navodilo za knjiženje poslovnih dogodkov v kmetijski dejavnosti z vzpostavitvijo povezave med FADN in davčnim knjigovodstvom in temu ustrezen kontni načrt.* http://www.mkgp.gov.si/si/delovna_podrocja/kmetijstvo/knjigovodstvo_na_kmetijah/

MKGP, 2014. *Priporočen Kontni Načrt.* http://www.mkgp.gov.si/fileadmin/mkgp.gov.si/pageuploads/podrocja/Kmetijstvo/Knjigovodstvo/PEKN.pdf

OECD, 2011. *Managing Risk in Agriculture: Policy Assessment and Design,* Pariz: OECD Publishing.

Petelin, G. and Žgajnar, J., 2016. *Tools for income risk management in agriculture.* Ljubljana, Društvo agrarnih ekonomistov Slovenije.

Ronan, G. and Cleary, G., 2000. Best Practices Benchmarking in Australian Agriculture: Issues and Challenges. *Agribusiness Perspectives Papers 2000,* Volume 39.

Schaffnit-Chatterjee, C., 2010. *Risk management in agriculture: Towards market soultions in the EU,* Frankfurt am Main: Deutsche Bank Research.

Schils, R. et al., 2007. DairyWise, A Whole-Farm Dairy Model. *Journal of Dairy Science,* Volume 90, pp. 5334-5346.

Tangermann, S., 2011. *Risk Management in Agriculture and the Future of the EU's Common Agricultural Policy.* 1 ed. Geneva: International Centre for Trade And Sustainable Development.

Torun, M. et al., 2018. Assessing business incubation: A review on benchmarking. *International Journa of Innovation Studies,* pp. 1-10.

Wilson, R., Charry, A. and Kemp, D., 2005. Performance indicators and bechmarking in Australian agriculture: synthesis and perspectives. *Extention Farming Systems,* 1(1), pp. 45-58.

NUMERICAL DETERMINATION OF AIR EXCHANGE RATE OF AND INSIDE A NATURALLY VENTILATED BARN DEPENDING ON INCOMING WIND ANGLE AND BARN'S LENGTH/WIDTH

Moustapha Doumbia[1], Sabrina Hempel[1], David Janke[1] and Thomas Amon[2]

[1]Leibniz Institute for Agricultural Engineering and Bioeconomy, Department of Engineering for Livestockmanagement, Potsdam, Germany

[2]Institut of Animal Hygiene and Environmental Health, Department of Veterinary Medicine, Freie Universität Berlin, Germany

mdoumbia@atb-potsdam.de, SHempel@atb-potsdam.de, DJanke@atb-potsdam.de, tamon@atb-potsdam.de

ABSTRACT

The air exchange rate is an important parameter in order to evaluate the gas emission of naturally ventilated barns. At the same time, understanding the flow inside such barns helps to investigate the comfort of the animals inside. Yet the air exchange rate is still difficult to assess mainly because of fluctuating influence stimuli such as wind (speed and angle), temperature, barn geometry etc. The main objective of the BELUVA project, financed by the German Research Foundation, is to systematically address those influences. The present study has been carried out in order to investigate the impact of incoming wind angle on the air exchange rate of animal occupied zones of barns with different length/width ratios.

A numerical model with quadrilateral and polyhedral grid cells (hybrid mesh) was set up in ANSYS Workbench. The numerical domain was divided into blocks for better meshing controls and quality. The influence of the roughness height was studied as well to reduce the numerical error related to the non-horizontal homogeneity of the velocity profile downstream.

Going from this numerical model, the air exchange rate of the whole barn and of the animal occupied zones inside has been determined and compared for different cases. The cases distinct themselves by three different incoming wind angles (0 °, 45 ° and 90 °), three different barn's length (L) / barn's width (W) (L/W=2,3,4) ratio and three different velocity magnitudes (1, 3 and 5 m s^{-1}).

The results show that the influence of the velocity incident angle on the air exchange rate of the overall barn and the animal occupied zones inside can be classified in types at least for the 0 ° and 90 ° incident angles. This, depending on the *L/W* ratio, gives important information about the height of the local in air exchange rate.

Keywords*:* air exchange rate, CFD simulation, meshing strategy, animal occupied zone.

1. INTRODUCTION

The implementation of computational fluid dynamics (CFD) in the study of barns has been growing in the recent years (Rong et al., 2016). The purposes, however, are totally different. For example, Wu et al., 2012 used CFD technics to evaluate air exchange rate measurement methods. Bjerg et al., 2013 and Fiedler et al., 2014 analyzed the gas emission from barns while Norton et al., 2008 investigated the influence of openings on the animal thermal comfort.

But still a lot has to be done to fully comprehend the influence of the different factors on the flow outside and inside a barn. Indeed, there are geometrical parameters such as the length/width (*L/W*) ratio of a barn that has yet to be taken into account. A barn design is relatively simple compared to other buildings. This simplicity allows the farmers to build the barn depending on the number of animals they possess, which results in a high variety of *L/W* ratio.

The paper should be considered as a second part of the paper published in Ciosta conference (Doumbia et al., prediction of the local air exchange rate in animal occupied zones of a naturally ventilated barn). While the latter paper focuses on the CFD validation and results analysis, this paper deals mainly with the numerical pre-studies aspect and preparation that was undergone to achieve reliable results.

2. NUMERICAL SIMULATION

All the numerical investigations were done in wind tunnel of scale (1:100).

2.1 Methods Barn details: Domain size and dimension

In figure 1 the dimensions of the computational domain are presented as well as the mesh refinement box with H = 114 mm (barn's height) and W = 342 mm (barn's width) as reference lengths. The dimensions were chosen to be large enough so that the walls do not impact the main flow. For the same reasons the side walls were used as the velocity inlets, see Figure 2 for the full list boundary conditions. The animal occupied zones had the following dimensions: H_AOZ = 30 mm, L_AOZ = L/4, W_AOZ = W/4.

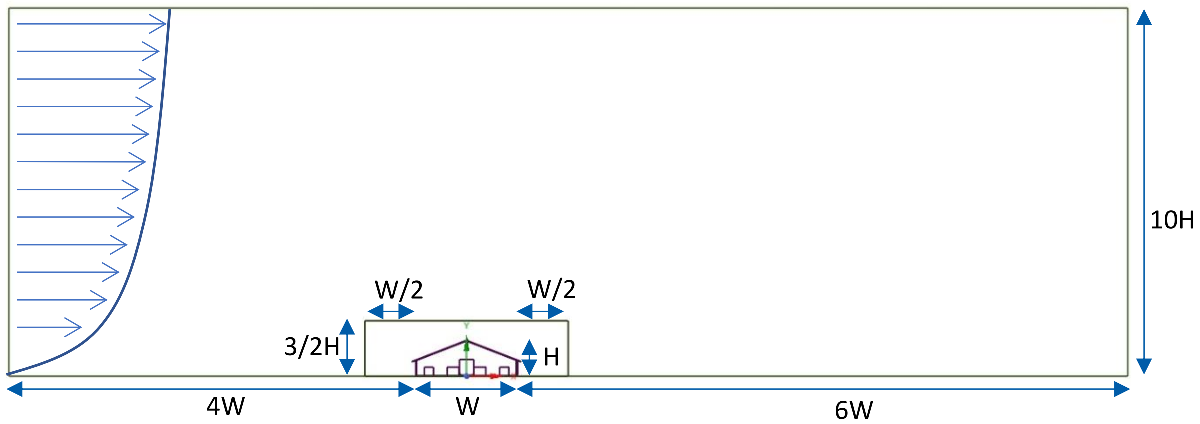


Figure 1. Dimensions of the numerical domain

Along with the velocity profile at the domain inlet, the profile for *k* (the turbulent kinetic energy) and *ε* (the turbulent dissipation rate) were implemented as well. The parameters *k* and *ε* were obtained from the wind tunnel measurement data of the turbulent intensity I using the following equations described in ANSYS Fluent User's Guide:

$$k = \frac{3}{2}\left(U_{avg}I\right)^2$$

$$\epsilon = \frac{k^{3/2}}{l}$$

where U_{avg} and l are the average velocity and the turbulent length scale respectively.

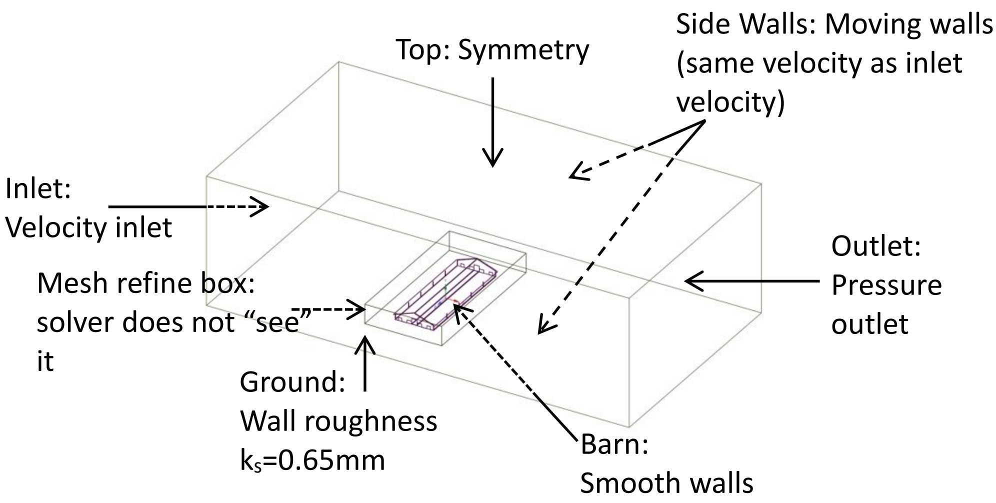


Figure 2. Boundary conditions

For the studied cases the k-ε realizable turbulence model was selected as recommended by Fluent (Lanfrit et al., 2005) for external dynamic flows. The pressure-based solver was preferred over the density-based since the flow is supposed to be incompressible, steady, with gas transport and isotherm (no energy source). In addition, the coupled scheme was selected since, from our experience and Fluent theory guide, it offers better performance and a more "robust and efficient single-phase implementation for steady flows" over the segregated schemes.

2.2 Meshing strategy

The computational domain was divided into blocks to achieve a better mesh control (cf: figure 3). The resulting hybrid mesh is a combination of tetrahedral cells (barn's block only) and quadrilateral cells (all other blocks), which puts the effort of mesh quality improvement only to the barn's box.

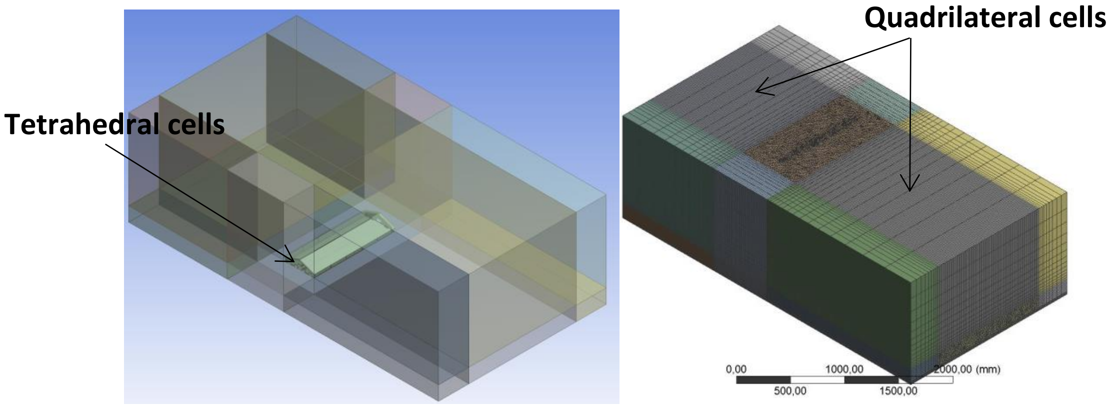


Figure 3. Computational domain divided into blocks

Tetrahedral mesh was chosen for the barn's box since this type of mesh is able to capture complex geometry details like sharp angles and small segments. Information about the cells number and quality for each *L/W* ratio obtained from Ansys Meshing are listed in table 1. Referring to Ansys meshing user guide, mesh quality metrics like the skewness, which evaluates how close to ideal a cell is, is important since it affects the simulation accuracy and stability. Based on table 2, the average skewness of the domain for all *L/W* ratios is evaluated as almost excellent.

Moreover, a Ansys Fluent technique, consisting of converting tetrahedral cells to polyhedral, was used to further reduce the total cell number up 40 % and thus reduce computational time.

Table 1. Cells number and mesh average skewness for each *L/W* ratio

L/W ratio	2	3	4
Cells number (in million)	12.09	17.18	21.66
Average skewness	0.256	0.254	0.253

Table 2. Skewness mesh metrics spectrum

Cell quality	Degenerate	Bad	Poor	Fair	Good	Excellent	Equilateral
Value of skewness	1	1 < - 0.9	0.9 – 0.75	0.75 – 0.5	0.5 – 0.25	0.25 - > 0	0

2.3 Roughness height k_s

One of the main difficulties in modeling flow in atmospheric boundary layer (ABL) (1:200 m in real scale) is to maintain the horizontal uniformity of the velocity profile in the flow direction. And one of the important parameters impacting the uniformity is the roughness height k_s of the numerical domain, which is supposed to simulate the obstacles of terrain surrounding the object under investigation (Bocken and al, 2008). In order to find the right k_s corresponding to the wind tunnel setup and arrangement and at the same time minimizing the non-uniformity of the velocity profile, a pre-study was done in an empty domain. In detail the equation given in Durbin et al., 2001 and Blocken et al, 2007 was solved to obtain the mathematical k_s:

$$U(y) = U_{ref} \times \frac{\ln\left(\frac{y}{y_0}\right)}{\ln\left(\frac{y_{ref}}{y_0}\right)}$$

In Ansys Fluent the roughness is defined as:

$$k_s = \frac{9.793\, y_0}{C_s}$$

Where *U(y)* is the wind velocity at the height *y*, U_{ref} a reference velocity at a reference height y_{ref} and C_S a roughness constant with a standard value of 0.5 (Blocken et al, 2007). Then giving velocity profile from the measurement data at the inlet of the domain, simulations were performed with different k_s values. The relative discrepancies of the velocity at different distances from the inlet compared with the inlet velocity were evaluated and are shown in figure 4. In the domain of interest (the first 4 m > 11 *W*), the errors were the smallest for ks = 0.00065 m.

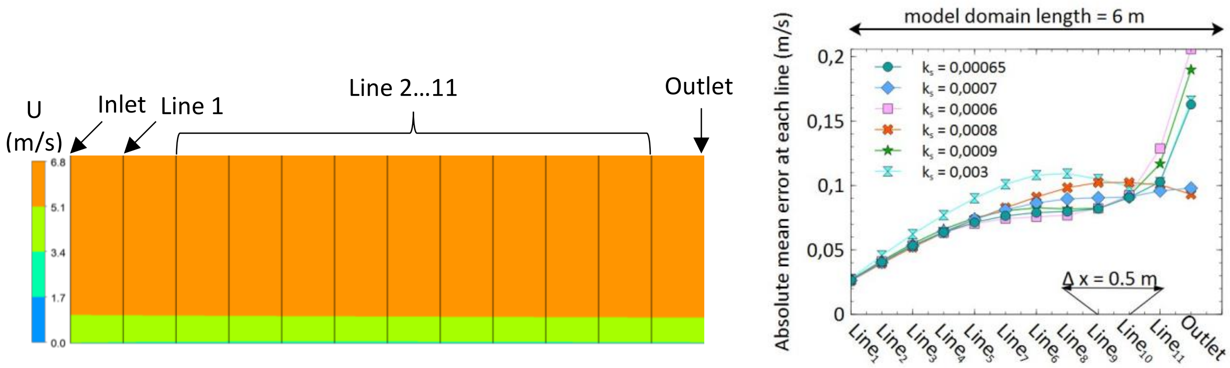


Figure 4. Relative velocity discrepancies along the domain compared to inlet velocity

3. RESULTS AND DISCUSSION

After validation of the numerical model with the wind tunnel measurements, the impact of *L/W* (= 2, 3, 4), velocity magnitude at the barn's inlet U (= 1, 3, 5 m s^{-1}) and incident velocity angle (0 °, 45 °, 90 °) on the air exchange rate of the whole barn (AER_{Barn}) and the animal occupied zone (AER_{AOZ}) was investigated. The simulations were done with Reynolds number Re > 80000 to ensure fully developed turbulence. The AER for the individual AOZs and the barn are presented in Figure 5 in s^{-1} (s^{-1} instead of h^{-1} since the calculations has been done in scale dimension 1:100).

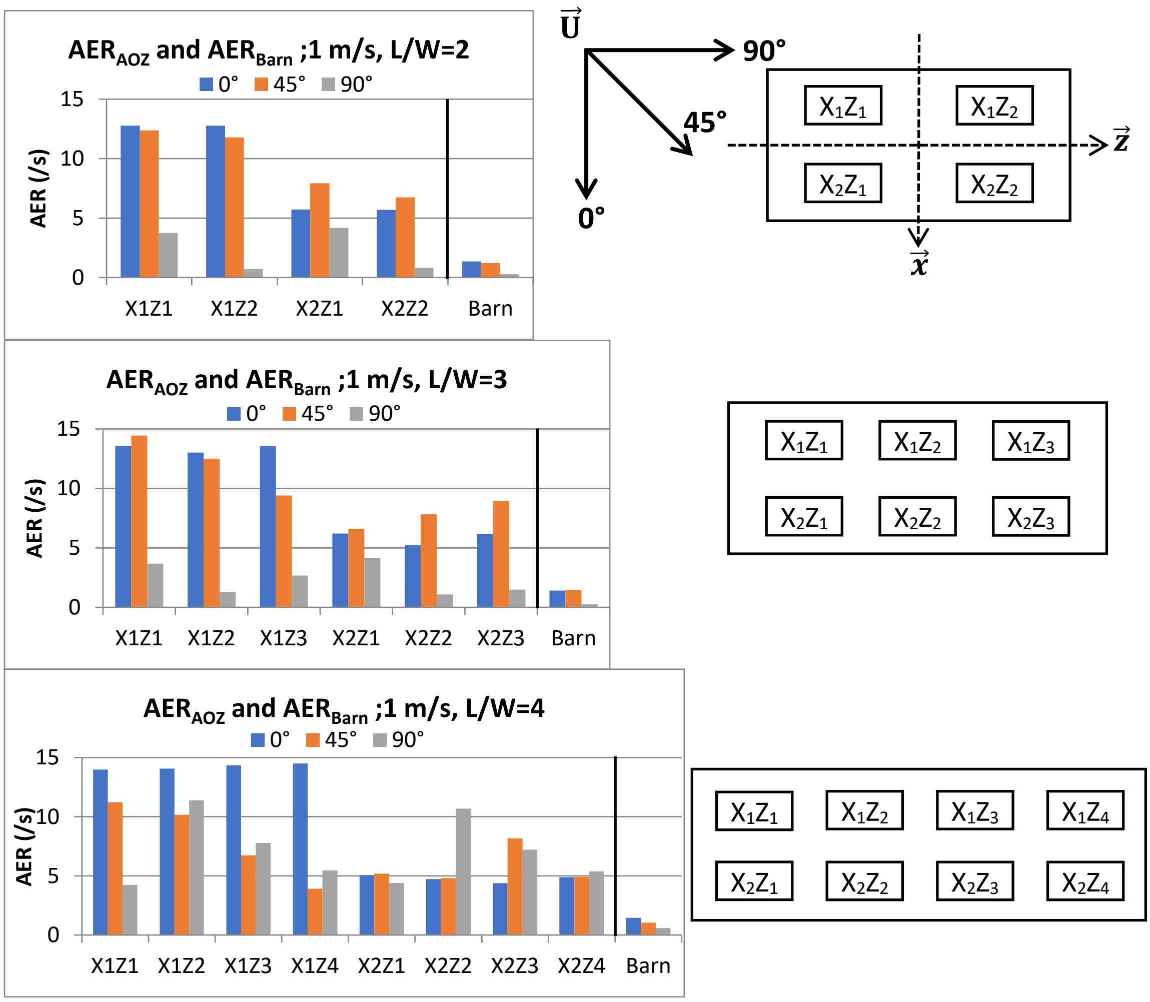


Figure 5. AER_{Barn} and AER_{AOZ}s for the 1 m s^{-1} case; Right: AOZs positions

From figure 5 the main findings can be summarized as follows:

- The AER_{Barn} alone is unable to provide valuable information about the changes inside.
- Each velocity incident case demonstrates a distinct pattern independent of the *L/W* ratio.
- For 0 ° incident angle, the AER_{AOZ} of the second line (X_2Z_b, b=1..*L/W*) is around half of the AER_{AOZ} of the first line (X_1Z_b, b=1..*L/W*).
- For 90 ° incident angle, considering the Z direction as symmetric axis, the AER_{AOZ} of one side (AER_{X1Zi}) have almost the same values as the one of the other side (AER_{X2Zi}).
- The 45 ° incident angle is the most unpredictable case, in the first line (X_1Z_i) values are decreasing from Z_i to Z_{i+1}. In the second line (X_2Z_i) the pattern changes with the *L/W* ratio.

The same can be observed for the other tested velocity magnitudes as well.

5. Conclusions

This paper focuses on the numerical modelling preparation for the study of AER in barns depending on various parameters. A meshing strategy by blocks was used in order to have better control and reduced cells number. The roughness k_s = 0.00065 m was found to minimize the numerical error in the ABL. Finally, it was shown that the influence of the velocity incident angle on the air exchange rate of the animal occupied zones inside the barn cannot be ignored. It can be classified in typical pattern at least for the 0 ° and 90 ° incident angles. This allows obtaining important information on the height of the local air exchange rate depending on the position inside the barn.

Further work will take into account the convection and gas flow in a real sized barn in the pursuance of a better comprehension of AER dependency.

References

Rong, L. (2001) 'Mechanisms of natural ventilation in livestock buildings: Perspectives on past achievements and future challenges', Biosystems engineering. doi: 10.1016/j.biosystemseng.2016.09.004.

Wu, W. (2012) 'Evaluation of methods for determining air exchange rate in a naturally ventilated dairy cattle building with large openings using computational fluid dynamics (CFD), Atmospheric Environment, 63, pp. 179 to 188. doi: 10.1016/j.atmosenv.2012.09.042.

Bjerg, B. (2013) 'Modelling of ammonia emissions from naturally ventilated livestock buildings. Part 3: CFD modelling', Biosystems engineering, 116, pp. 259 to 275. doi: 10.1016/j.biosystemseng.2013.06.012.

Norton, T. (2008) 'Assessing the ventilation effectiveness of naturally ventilated livestock buildings under wind dominated conditions using computational fluid dynamics', Biosystems engineering, 103, pp. 78 to 99. doi: 10.1016/j.biosystemseng.2009.02.007.

Fiedler, T. (2008) 'Spatial distribution of air flow and CO2 concentration in a naturally ventilated dairy building', Environmental Engineering and Management Journal, 13, 9, 2193-2200.

Blocken, B. (2017) 'CFD simulation of the atmospheric boundary layer: wall function problems', Atmospheric Environment. doi: 10.1016/j.atmosenv.2006.08.019.

Lanfrit, M. (2005) 'Best practice guidelines for handling Automotive External Aerodynamics with FLUENT'.

ANSYS Fluent User's Guide, version 18.1 (2017)

ANSYS Fluent Theory Guide, version 18.1 (2017)

DRY ABOVE GROUND BIOMASS FOR A SOYBEAN CROP USING AN EMPIRICAL MODEL IN GREECE

Christos Vamvakoulas, Stavros Alexandris and Ioannis Argyrokastritis

Lab. of Agricultural Hydraulics, Department of Natural Resources Development & Agricultural Engineering, 75 Iera Odos str. 118 55 Athens, Greece

chrisvamva@aua.gr, stalex@aua.gr, jarg@aua.gr

ABSTRACT

A new empirical equation for the estimation of daily Dry Above Ground Biomass ($\mathrm{D\text{-}AGB}$) for a hybrid of soybean (*Glycine max L.*) is proposed. This equation requires data for three crop dependent parameters, Leaf Area Index, plant height and cumulative crop evapotranspiration. Bilinear surface regression analysis is used in order to estimate the factors entering the empirical model. For the calibration of the proposed model, yield data from a well-watered soybean crop for the year 2015, in the experimental field (0.1 ha) of the Agricultural University of Athens, are used as a reference. Verification of the validity of the model was obtained by using data from 2014 cultivation period for well-watered soybean cultivation (100% of crop evapotranspiration water treatment), as well as, data from three irrigation treatments (75%, 50%, 25% of crop evapotranspiration) for two cultivation periods (2014-2015). The proposed method for the estimation of $\mathrm{D\text{-}AGB}$ may be proved as a useful tool for estimations without using destructive sampling.

Keywords*:* Dry Above Ground Biomass, Soybean, Empirical models, Bilinear regression analysis.

1. INTRODUCTION

Nowadays agronomists and irrigation experts use crop productivity models for the simulation and prediction of dry above ground biomass ($\mathrm{D\text{-}AGB}$). Complex crop growth models (CGM) require a large number of input parameters, usually not available from ideal sites, which leads to significant and systematic cumulative errors in determining crop yield and above ground biomass.

In this study, it is attempted to present a simple model using geostatistical methods in a simple form. Generally empirical equations are a tool for local estimations of attributes without many parameters as inputs. In the past algorithms for the creation of such equations have been used for the estimation of reference evapotranspiration (ET_o) (Alexandris and Kerkides, 2003; Alexandris et. al, 2006) and for the estimation of crop evapotranspiration (Poss et. al, 2004). Also, empirical bilinear regression equations have been used for the prediction of human and rat tissue (Meulenberg and Vijverberg, 2000), while multiple regression analysis has been applied for un-mixing of surface temperature data in an urban environment (Wicki and Parlow, 2017).

In this paper, a model for the daily estimation of $\mathrm{D\text{-}AGB}$ for a soybean hybrid (PR91M10) in central Greece was formulated. The model has been parameterized by experimental observations on the soybean crop. Also, the model is examined for water stressed and non-water stressed plants, under field conditions. The final equation obtained, is based on Leaf Area Index (LAI), plant height ($\mathrm{h_c}$) and cumulative crop evapotranspiration ($\mathrm{cumET_c}$).

2. METHODOLOGY

The experiment was performed in the experimental field of the Agricultural University of Athens in Aliartos plain (38° 23' 40'' N, 23° 05' 08'' E and 95 m altitude), during 2014 and 2015 cultivation periods. Data from an experiment with four irrigation treatments (100%, 75%, 50%, 25% of crop evapotranspiration respectively) in a randomized complete block design, with four replications, were used. Daily grass reference evapotranspiration and crop evapotranspiration were estimated by using the Penman-Monteith equation and crop coefficients as it is suggested by Allen et al (1998). The plot size of each irrigation treatment was 3 m × 12 m and the spacing between each main plot was 3 m in order to minimize water movement among treatments. The experimental plots were 3 m × 6 m and consisted of 5 rows with 0.75 m apart. PR91M10 is a highly-productive variety of the early maturity group (00). Seeds were hand-planted, using a seeding depth of about 3 cm, on 30 May 2014 (Julian day, JD: 150) and on 31 May 2015 (Julian day, JD: 151), respectively. Treatment plots consisted of 5 rows planted, 75 cm apart, with 4–5 cm row spacing and the sowing density was 33 seeds m^{-2}.Irrigation scheduling was based on the daily water balance calculation and on results obtained using the computer model ISAREG (Pereira et al, 2003), which utilized data collected during consecutive cultivation periods from 2011 to 2015.

Rainfall during the 2014 and 2015 cultivation period was 46.1 mm and 176.7 mm respectively. The ground water table was at 1.2 m depth for both cultivation years. Irrigation was applied to provide 100%, 75%, 50% and 25% of the crop evapotranspiration needs.

A surface drip irrigation system was used for irrigation. A 16 mm diameter polyethylene pipe with inline pressure compensating drippers at 33cm intervals was placed on one side of each soybean row. The average discharge of emitters was 4.4 l/h at 0.1 MPa.

Periodically, every 7 days approximately plant height ($\mathrm{h_c}$, cm) was measured and destructive sampling was performed by collecting 3 plants from the 3 interior rows of each plot, for Leaf Area Index (LAI) and Dry Above Ground Biomass ($\mathrm{D\text{-}AGB}$, ton/ha) estimation. Sampling was performed at 25, 34, 41, 48, 54, 60, 66 and 75 days after planting (DAP) for 2014 cultivation period and at 24, 33, 40, 47, 53, 59, 65 and 75 (DAP) for the 2015 cultivation period respectively.

The parameterization of the model was done for the 2015 cultivation year data, because precipitation was higher than that of 2014 giving better environmental conditions for the non-water stressed plants (100% treatment). The model represents the simulation curve of the $\mathrm{D\text{-}AGB}$ for the first 75 days of the growing period. The last 20 days of the maturity stage are not included in the simulation curve. Surface regression analysis was used to establish the new model to simulate daily ($\mathrm{D\text{-}AGB}$). The empirical model was derived by surface polynomial regression using the three crop dependent parameters, measured values of Leaf Area Index (LAI), plant height ($\mathrm{h_c}$) and cumulative crop evapotranspiration ($\mathrm{cumET_c}$), in a general form $\mathrm{D\text{-}AGB} = f(\mathrm{LAI}, \mathrm{h_c}, \mathrm{cumET_c})$. It utilizes four unknown parameters (k_0, k_1, k_2, k_3) which are determined in a three stage approach. Experimental lines for the $\mathrm{D\text{-}AGB}$ obtained from the destructive sampling, were used as standard values. Calculated $\mathrm{D\text{-}AGB}$ values are then regressed against mean daily values of pairs of LAI and $\mathrm{h_c}$ (first stage) and LAI and $\mathrm{cumET_c}$ (second stage) in a bilinear equation of the form:

$$z = f(x, y) = k_0 + k_1 \cdot y + k_2 \cdot x + k_3 \cdot x \cdot y$$

x, y denoting daily values of either LAI and $\mathrm{h_c}$ (cm), in the first stage of investigation, or LAI and $\mathrm{cumET_c}$ (mm/time), in the second stage, z standing for $\mathrm{D\text{-}AGB}$ (ton/ha). As expected, the first and second stages end up with the estimation of two sets of four parameters a_i, b_i $(i = 1,...4)$ as shown in the equations (1) and (2) below:

$$C_1 = a_1 + a_2 \cdot \mathrm{h_c} + a_3 \cdot \mathrm{LAI} + a_4 \cdot \mathrm{LAI} \cdot \mathrm{h_c} \tag{1}$$

Where $a_1 = -0.143, \quad a_2 = 0.095, \quad a_3 = -6.33, \quad a_4 = 0.058$.

$$C_2 = b_1 + b_2 \cdot \mathrm{cumET}_c + b_3 \cdot \mathrm{LAI} + b_4 \cdot \mathrm{LAI} \cdot \mathrm{cumET_c} \tag{2}$$

Where $b_1 = -0.115, \quad b_2 = 0.0066, \quad b_3 = -2.4, \quad b_4 = 0.0129$.

In the above equations C_1 and C_2 represent Dry Above Ground Biomass ($\mathrm{D\text{-}AGB}$) in ton/ha. Tables 1 and 2 show the cross-correlation/covariance of the factors entering in the first and second stage of regression respectively.

The $\mathrm{D\text{-}AGB}$ values are now regressed against the results obtained from the previous stages shown as C_1 and C_2 bilinear expressions (stage 3). This last regression ends up with the estimation of four parameters m_i $(i = 1,...4)$ and the final working formula for $\mathrm{D\text{-}AGB}$ on a daily basis is given by the following Eq.(3) in an implicit form, since C_1 and C_2 are functions of the attributes LAI, $\mathrm{h_c}$ and $\mathrm{cumET_c}$:

$$\mathrm{D\text{-}AGB}(\mathrm{ton/ha}) = m_1 + m_2 \cdot C_2 + m_3 \cdot C_1 + m_4 \cdot C_1 \cdot C_2 \tag{3}$$

Where $m_1 = 0.0082, \quad m_2 = 1.11, \quad m_3 = -0.12, \quad m_4 = 0.0032$.

In Fig.1a the iso-lines of ($\mathrm{D\text{-}AGB}$) derived from Eq.(1) as a function of LAI, plant height ($\mathrm{h_c}$) and $\mathrm{D\text{-}AGB}$ measurements, through curve interpolation lines respectively in a daily basis, are presented. Similarly, Fig.2b shows the results of the second stage $\mathrm{D\text{-}AGB}$=*f*(LAI, $\mathrm{cumET_c}$). Higher sensitivity showed the LAI-$\mathrm{h_c}$ correlation, than the one of LAI-$\mathrm{cumET_c}$ for the $\mathrm{D\text{-}AGB}$ factor.

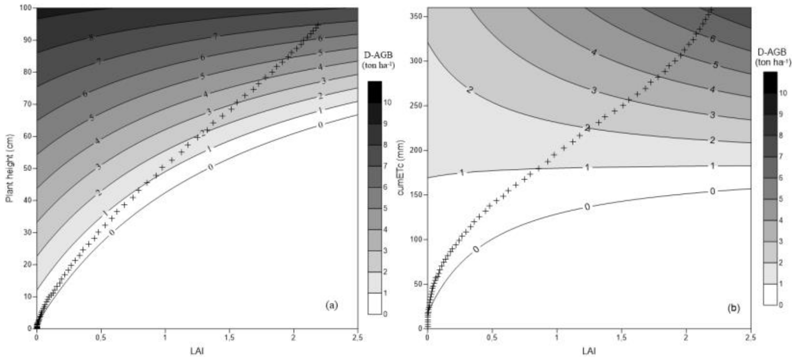


Figure 1. The iso-lines of ($\mathrm{D\text{-}AGB}$) derived from Eq.(1) and Eq.(2)

Table 1. Cross-correlation/covariance between LAI, $\mathrm{h_c}$, $\mathrm{D\text{-}AGB}$ from the first stage of regression.

	LAI	$\mathrm{h_c}$	$\mathrm{D\text{-}AGB}$
Variable correlation			
LAI	1.000	0.996	0.957
$\mathrm{h_c}$	0.996	1.000	0.935
$\mathrm{D\text{-}AGB}$	0.957	0.935	1.000
Variable covariance			
LAI	0.614	25.637	1.613
$\mathrm{h_c}$	25.637	1080.251	66.108
$\mathrm{D\text{-}AGB}$	1.613	66.108	4.626

Table 2. Cross-correlation/covariance between LAI, $\mathrm{cumET_c}$, $\mathrm{D\text{-}AGB}$ from the second stage of regression.

	LAI	$\mathrm{cumET_c}$	D-AGB
Variable correlation			
LAI	1.000	0.987	0.957
$\mathrm{cumET_c}$	0.987	1.000	0.929
D-AGB	0.957	0.929	1.000
Variable covariance			
LAI	0.614	86.497	1.613
$\mathrm{cumET_c}$	86.497	12505.448	223.44
D-AGB	1.613	223.44	4.626

It is obvious from the Tables 1 and 2 that the strongest correlation exists between LAI and $\mathrm{h_c}$ (0.996) and that all three attributes LAI, $\mathrm{h_c}$ and $\mathrm{cumET_c}$ are also strongly correlated to $\mathrm{D\text{-}AGB}$ (all correlation coefficients are above 0.92, see tables 1 and 2).

A statistical analysis was further performed in order to provide quantitative indices to our estimates. For this purpose the following statistical indices were estimated (Fox, 1981; Willmott, 1982): (i) Mean bias error (MBE), (ii) Variance of the distribution of differences s_d^2 which expresses the variability of $(P-O)$ distribution about MBE, (iii) Root mean square error (RMSE), (iv) Mean absolute error (MAE), (v) Index of agreement, d, (Willmott, 1982), where n is the number of cases. *O* denotes the experimental values of $\mathrm{D\text{-}AGB}$ measured during the 2014-2015 cultivation periods for all irrigation treatments (I_{100}, I_{75}, I_{50} and I_{25}). P denotes the simulated values as these are estimated by the proposed methodology. All the above mentioned relevant statistical indices are provided in Table 3.

Table 3. Summary statistics of daily Dry Above ground Biomass ($\mathrm{D\text{-}AGB}$) tested against the reference method.

Treatments	Slope	MBE	RMSE	MAE	s_d^2	*d*	R^2
2015,(N=75)							
I_{75}	1.113	0.239	0.450	0.262	0.640	0.998	0.986
I_{50}	1.073	0.315	0.498	0.371	1.010	0.996	0.966
I_{25}	1.347	0.446	0.678	0.490	1.990	0.988	0.978
2014,(N=75)							
I_{100}	1.213	0.226	0.380	0.245	0.539	0.997	0.992
I_{75}	1.211	0.393	0.579	0.414	1.522	0.994	0.974
I_{50}	1.008	0.211	0.378	0.321	0.485	0.998	0.965

3. Results And Discussion

The regression equations between daily simulated $\mathrm{D\text{-}AGB}$ values against the experimental and the cross-correlation coefficient (R^2) are shown in table 3 for the 2015 and 2014 cultivation periods respectively.

Fig.2 presents the development of $\mathrm{D\text{-}AGB}$, both measured and simulated, during cultivation period 2015 expressed in days after planting (DAP). As it is depicted in Fig.2a the simulated and experimental curve interpolated lines are almost coincided for the 100% treatment because the model has been calibrated for this treatment and cultivation period. Fig.2b, 2c, 2d show the 75%, 50% and 25% treatments for the 2015 cultivation period respectively. It is obvious that the predictions by the model for the 75%, 50% and 25% treatments, give results very close to the measurements for the 2015 cultivation period.

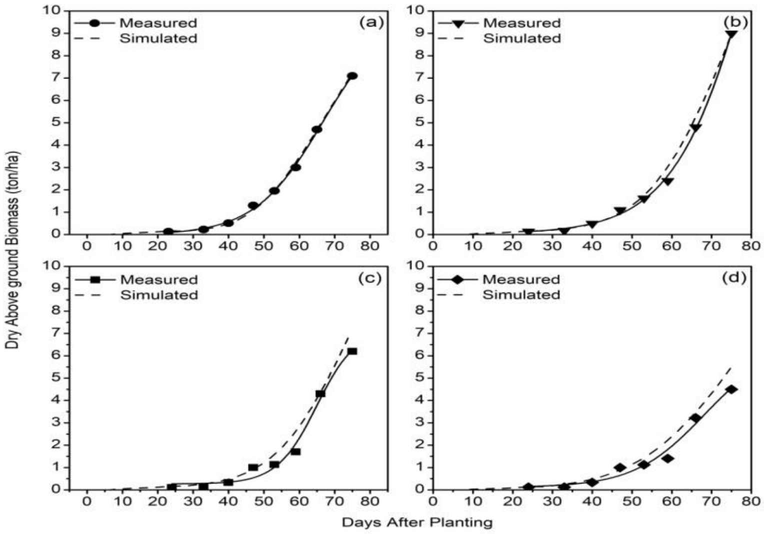


Figure 2. The relationship between Dry Above Ground Biomass (D-AGB), (ton/ha) and days after planting (D-AGB) for 2015 cultivation period and PR91M10 hybrid. The a, b, c, d parts depict the 100%, 75%, 50% and 25% of ETc water treatments respectively.

In Fig.3 the development of D-AGB both measured and simulated curve interpolated lines during cultivation period 2014 expressed in days after planting (DAP) are presented. For 2014 cultivation period all four figures Fig.3a, 3b, 3c and 3d are used for verification purpose. Fig.3a, 3b, 3c and 3d show the 100%, 75%, 50% and 25% treatments respectively. From Fig.3a at 100% treatment can be assumed that measured and simulated curve interpolated lines were very close and at DAP 75 the model predicted D-AGB 4.951 ton/ha, while experimental D-AGB for the DAP 75 for the non water stressed soybean was 4.385 ton/ha. Similarly the 75%, 50% and 25% water treatments were perfect fitted till 55 DAP approximately for the 2014 cultivation period. However, the response of the plant to the water stress mechanism is a fairly complex process involving both biophysical and biochemical functions that could differentiate predictions of the experimental observations. This induces the differences after DAP 55 for the water stressed treatments in 2014 cultivation period (Fig.3b,3c,3d) and for the 2015 water stressed treatments (Fig.2b,2c,2d).

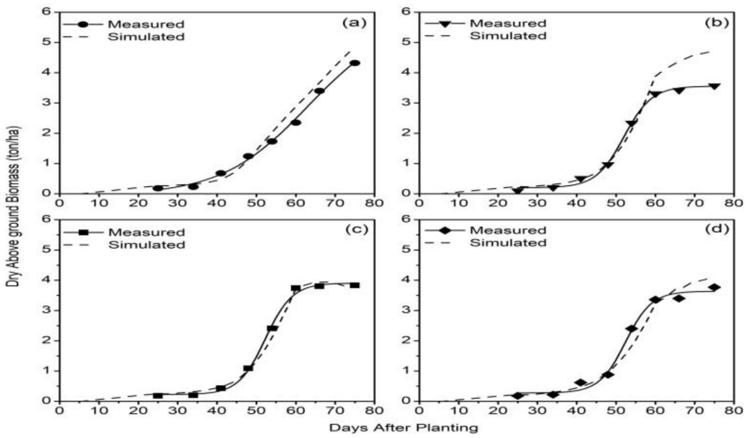


Figure 3. The relationship between Dry Above Ground Biomass (D-AGB), (ton/ha) and days after planting (DAP) for 2014 cultivation period and PR91M10 hybrid. The a, b, c, d parts depict the 100%, 75%, 50% and 25% of ETc water treatments respectively.

4. Conclusion

For the first time an already existing empirical methodology for the prediction of reference evapotranspiration (ET_0) coupled with crop geometrical characteristics (LAI, h_c) and $cumET_c$ as inputs, was used in order to predict daily $D\text{-}AGB$. The statistical analysis showed very satisfactory adjustment of the experimental and simulated values especially for the non-water stressed treatments of the 2014 cultivation period.
Further experimentation for different regions and a wider range of $D\text{-}AGB$ values is needed in order to verify the goodness of fit, for the parameters used in the methodology, in different climate regimes and for more cultivation species. An important advantage of the methodology that has been followed, in addition to the use of three readily measured fundamental parameters ($cumET_c$, LAI, h_c), is that the model can easily be calibrated (different coefficients) for any crop and in any climatic environment.
However, it could be set a more complex algorithm using more environmental attributes of the soil-plant-atmosphere system, which might be adjusted better to the simulated values into the experimental ones, especially for the plants under non water stress or under irrigation deficit.

5. Acknowledgements

This research was supported by the Alexander S. Onassis Foundation.

References

Allen R.G., L.S. Pereira, D. Raes, M. Smith (1998). Crop Evapotranspiration, FAO Irrigation and Drainage paper No 56, Rome, Italy,*pp.* 327.

Alexandris, S., & Kerkides, P. (2003). New empirical formula for hourly estimations of reference evapotranspiration. Agric. Water Manage. 60, 181–198.

Alexandris, S., P., Kerkides, A., Liakatas (2006). Daily reference evapotranspiration estimates by the ''Copais'' approach. Agric. Water Manage. 82, 371-386.

Fox, M.S. (1981). An organizational view of distributed systems. IEEE Trans. Syst. Man Cybernetics 11, 70–80.

Meulenberg, C. J. W., & M.Vijverberg, H. P. (2000). Empirical relations predicting human and rat tissue: air partition coefficients of volatile organic compounds. Toxicol. Appl. Pharmacol. 165, 206–216.

Pereira, L.S., P.R., Teodoro, P.N., Rodrigues, J.L.,Teixeira 2003. Irrigation scheduling simulation: the model ISAREG. In: Rossi, G., Cancelliere, A., Pereira, L.S., Oweis, T., Shatanawi, M., Zairi, A. (Eds.) Tools for Drought Mitigation in Mediterranean Regions. Kluwer, Dordrecht, *pp*. 161-180.

Poss, J.A., W.B., Russell, P.J., Shouse, R.S., Austin, S.R., Grattan, C.M., Grieve, J.H., Lieth, L., Zeng (2004). A volumetric lysimeter system (VLS): an alternative to weighing lysimeters for plant-water relations studies. Comput. Electron. Agric. 43, 55–68.

Wicki, A., & Parlow, E. (2017). Attribution of local climate zones using a multitemporal land use/land cover classification scheme. J. Appl. Remote Sens., 11, 026001.

Willmott, C.J. (1982). Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 63, 1309–1313.

IDENTIFICATION OF REQUIREMENTS FOR IMPLEMENTING IOT ON THE SMART BEEF VALUE CHAIN

Gustavo Marques Mostaço, Roberto Fray Silva, Carlos Eduardo Cugnasca
Department of Computer Engineering and Digital Systems,
Escola Politécnica da Universidade de São Paulo (USP), Brazil
gmostaco@usp.br, roberto.fray.silva@gmail.com, carlos.cugnasca@usp.br

ABSTRACT

Smart services for livestock value chains can improve both productivity and sustainability by providing valuable information for the decision-making and management systems. By blending Internet of Things (IoT) technologies with Big Data analytics, it is possible to increase information and data exchange among supply chain agents, gain predictive insights in farming and marketing operations, drive real-time operational decisions and increase the efficiency of processes. This work provides an overview of the beef cattle VC comprising its stages, stakeholders, processes, and the informational flow. It also identifies the main requirements for implementing IoT on the beef cattle value chain, which are organized in 9 main services and 31 sub-services or activities. It sets the ground requirements for the future development of a framework for Smart Beef Cattle Services. These can also be used for developing a more general Smart Livestock Farming framework. Farm management research may benefit from the resulting requirements, using it to build services and architectures for a panorama of further automation and autonomous operations of the farms and agricultural supply chains.

Keywords: IoT, farm management systems, Smart Livestock Farming, agribusiness, value chain.

1. INTRODUCTION

Value chains (VCs) can be defined as a series of interconnected processes or services, going through production, storage, processing, transportation and marketing of a specific product or commodity, adding value as it moves through the VC (Wolfert et al., 2017). Those commodities can be marketed for domestic or international markets and involve a complex series of stakeholders and products.

Smart services for the Beef Cattle farming refers to modern Information and Communication Technologies (ICT) applied to this livestock VC. They have the potential to deliver a more productive and sustainable production by integrating processes of the Precision Livestock Farming (PLF), Management Information Systems (MIS) and agricultural automation in order to provide better decision-making and more effective exploitation of operations (Banhazi et al., 2012). Additionally, it is possible to better comply with food-quality standards and sustainability labels to achieve higher value markets, increasing the value of their products.

The use of Internet of Things (IoT) technologies (ITU-T, 2012) aims at providing full coverage of the processes related to this livestock's VC. This can be done by collecting, transmitting, analyzing, and storing data from the entire agroecosystem. One example in the beef cattle VC is a system that registers and controls inputs purchase and consumption throughout the production processes, encompassing breeding, growing, finishing, slaughtering, meat processing, transportation and logistics, until its products reach the retailers and wholesalers, and finally the consumers' table.

Aspects of the IoT paradigm are vital to the Smart Beef Cattle VC concept. By using IoT, one is able to run the workflow among farmer, service provider, logistics provider, market and consumer synchronously all together. That means to establish contact with each participant of the VC, bringing data and collecting information about their processes, increasing the possibilities for controlling and improving the efficiency of their tasks (ITU-T, 2012). On the other hand, without applying IoT concepts, beef cattle farming may continue to work but will not be able to achieve higher performance levels and mitigate quality and environmental problems timely. Therefore, this work aims at setting the ground requirements for the beef cattle VC to operate under the IoT percepts.

2. Methodology

The methodology used in this paper can be divided into 2 steps:

1. **VC description:** this step aimed at describing the main stages, stakeholders, processes, and the informational flow of the beef cattle VC, based on a thorough literature review;
2. **Requirements and services identification**: we identified the main requirements and services needed to manage the information flow in the beef cattle VC, considering the use of IoT technologies and based on both periodic meetings with experts from International Telecommunication Union - Standardization Sector (ITU-T) and an in-depth literature review.

3. Results

This section contains the main results of our research. It is divided into two parts: 3.1, providing an overview of the beef cattle VC; and 3.2, containing the requirements and services identified.

3.1. Overview of the beef cattle value chain

Different stages and perspectives are present in the beef VC, such as quality control, efficiency control, certification, sustainability, logistics, business economics, marketing channels, and informational flows (CEMA, 2017). The commodities inside such chains can be marketed for domestic (national, intra or inter-state), or international markets, involving a complex series of stakeholders and products. They can respond to ICT changes at the same time that are subjected to local, national and international laws and requirements. One main concern with the present work is to address the domestic facet of those VCs, leaving the international issues for future research.

The main stakeholders involved in the beef cattle VC are: Inputs and Services Suppliers; Farmers; Processing and logistics agents (transport, slaughterhouses, industries); Market agents (warehouses, distributors, traders and retailers); and Consumers. Figure 1 shows an overview of a generic beef cattle VC. It represents the main stages and steps during planning, production, processing, transportation and sale of animal products.

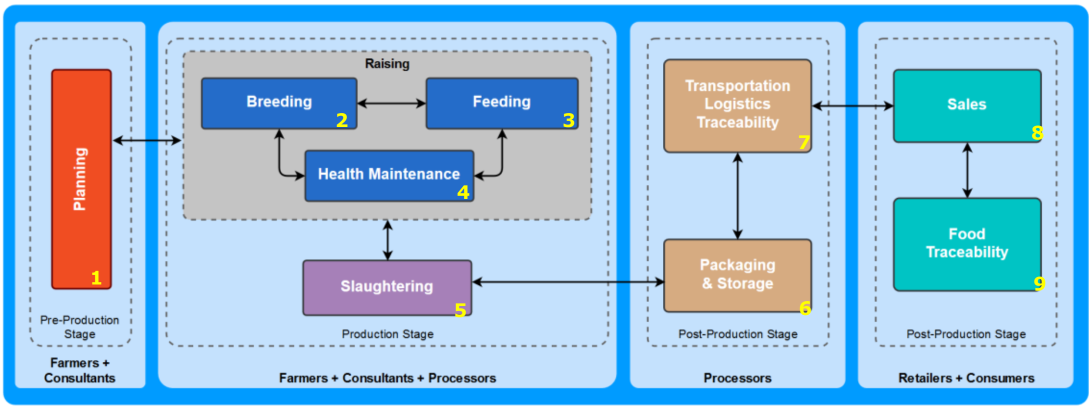


Figure 1. Overview of the beef cattle value chain

The **Planning** phase (number 1 in Figure 1) is the main component of the Pre-Production Stage. At this phase, the resources and raw materials are allocated according to the activities, workforce and production capacity. It includes such things as land, buildings, equipment, supplies and processes, as well as laws and regulations (environmental and product quality) that affect the business. In this step, the IoT requirements are brought up and the system capabilities are applied to meet them.

Breeding or Reproduction (number 2 in Figure 1) is the first step inside the Production Stage. It refers to herd increase through animal reproduction inside the farm or by artificial insemination or embryo transfer techniques. In this step, the main IoT requirement is the animal identification and breeding record, which is a demand for consumers of higher quality products.

Feeding (number 3 in Figure 1) comprises one of the most resource consuming activities in the Production Stage. It considers mainly animal feeding and water consumption and affects directly on the final product results and resource use efficiency. For ruminants, quality grass consumption is a major concern, but in many climates or rearing methods, hay, silage, as well as energy and protein-rich foods can be added to the diet (Coleman, Moore, 2003). In this step, it is important to record details about food components, the quality and quantity of ingested food by the animals and their weight gain. The most common IoT requirements for this step are Animal identification, Productivity monitoring, Health management, and Animal welfare.

Health maintenance (number 4 in Figure 1) is directly connected to the raising activities. Together with animal welfare, implies great concerns among consumers, due to their impact on product quality and safety. These not only represent a possible threat to human health but can also lead to direct production losses and resource use inefficiency. These concerns can commonly include how animals are kept and treated inside the farm, how they are transported and how they are slaughtered (HFAC, 2018; WSPA, 2011).

Slaughtering (number 5 in Figure 1) considers the reception of the living animals at the slaughterhouse and encompasses the activities related to killing them and processing the carcasses, resulting in the final products that are sold to the market. Due to sanitary concerns, animal slaughtering is only allowed in specific and quality controlled facilities. Some countries may require that these establishments ensure that the animals are handled and slaughtered humanely (Grandin, Nami, 2017). In this step, the most important IoT requirements are Animal identification, Climate control, and Animal welfare.

Logistics (throughout the whole VC) is present in great part of the beef cattle VC. It is important to provide the necessary amount of raw materials and inputs at the correct time during production, as well as it is responsible for providing a constant flow of animals for the slaughterhouse, to maintain its optimal working flow. Incorrect animal transportation and handling can result in product quality reduction, and in severe cases can lead to death (Sheridan et al., 1991). The most important logistics requirements are Animal identification, Climate control, Health management, and Animal welfare.

Traceability or product tracking and tracing (throughout the whole VC) is the capability to follow the path of a specified production batch throughout the VC, as it moves from one organization to the next, allowing the identification of critical quality control points. In the case of food contamination, it helps identifying the product batches that may have been contaminated (Dabbene et al., 2014). In general, food businesses engaged in the wholesale supply, production or food import must have a well-defined system, including production records covering, clearly described in a written document, to ensure that a fast and efficient recall is possible and timely (so called internal traceability). This information should be readily accessible for both Governmental agencies, and the end consumer, to identify what batches/product units must be recalled and their location in the VC (Verdouw et al., 2018). It should also be accessible for other companies in the chain, so they can maintain a description of the whole path of the product in the chain (so called external traceability). Recording can be done through means of barcodes, QR codes, RFID tags, and other tracking media.

3.2. Requirements and services identification

The identified requirements for the smart beef cattle VC services are presented in Table 1.

Table 1. Smart Beef Cattle VC main services and sub-services or activities

Services	Sub-Services or Activities	Steps related
Animal monitoring	Animal identification	ALL
	Animal tracking	2, 3, 4
Animal health maintenance	Prevention and control of diseases	4
	Increasing productive lifetime	2, 3, 4
	Increasing disease resistance	2, 4
	Animal welfare	2, 3, 4, 5
Nutrition and Productivity control	Forage quality improvement	3, 4
	Dietary improvements and substitutes	3, 4
	Feed supplements	3, 4
	Digestibility control (rumen microbiome)	3, 4
	Precision feeding	3, 4
Breeding and Genetic Improvement	Animal selection	2, 3, 4
	Genomic selection	2, 3, 4, 5
	Increasing performance on low-quality feed	3
	Low-methane production	2, 3, 4
Pasture Management	Pasture quality and quantity management*	3
	Carbon sequestration increase*	1, 3
	Integrated and Mixed systems*	1, 3
Waste Handling and Treatment	Manure collection*	4
	Biogas digesters efficiency management	1
	Gas emission control*	1
	Fertilizer production	1, 3
Climate control	Climate sensing and control for living animals	4, 5
	Final product temperature monitoring	6, 7, 8, 9
Logistics Management	Distribution recording	7, 8, 9
	Transport environment control	7
	Market analysis alerts	1, 8
Traceability Management	Production recording	1, 2, 3, 4
	Processing control	5, 6
	Product tracing	ALL
	Recall management	ALL

*Possible sustainability and environmental impacts

Animal identification is one of the key information along the beef cattle VC, since this register accompanies the animal throughout its life inside the farm and should remain attached to all products derived from it. Becomes essential to allow product traceability. The requirement is to use any kind of electronic identification (internal, external or subcutaneous) allowing easy and instantaneous identification of the animal, using a proper device or antenna. For example, RFID tags can be used on the skin and/or earrings.

Productivity monitoring is mainly applied to the Raising set of tasks on the Production Stage, which demands the management of feeding and the main processes inside the farm, so that the animals may experience optimal weight gain (based on feed conversion ratio and energy expenditure). The requirements in this category are related to the correct association of important production parameters with each animals' ID, such as weight and food consumption over time (in order to obtain feed conversion ratio), and movement (to estimate energy expenditure).

Health management is very important to guarantee product quality and safety. As indicated before, this category can be responsible for productivity losses or even possible threats to human health. Animal health monitoring can be improved with the use of internal and external sensors, capable of sensing and communicating important parameters to the server or the cloud, at a time interval that is appropriated for decision-making. Vital information such as the control of vaccinations, exams, pests, diseases, bruises, and other health-related records should be registered on the management system and at the animal's electronic identification device.

Climate control such as the ambient temperature and relative humidity can affect considerably the metabolism of farm animals, leading to thermal stress. This results in efficiency loss, increasing the animals' energy use. Other environmental factors may also influence the animals' thermal sensation and heat dissipation, such as air velocity and radiation (Babinszky et al., 2011). These factors can also hasten the final products quality reduction during transport, at processing units or at distributors and retailers. Regarding this issue, climate control can be improved with the help of environment sensors, capable of sensing and communicating important parameters at an appropriated time interval, and actuators capable of changing important factors in a timely manner.

Animal welfare is related to the reduction of animal suffering. When the animal is considered to suffer for any reason its efficiency in converting feed into muscle decreases, leading to direct losses for the farmer. This issue is interrelated to the climate control and health monitoring categories. At this point, systems capable of detecting abnormal behavior should be adopted, through GPS monitoring (for extensive farming) or image recognition (for confinements or intensive animal farming), in order to mitigate the issues related to this category.

Information and Communication Security regards the proposal of IoT devices integration on this VC, which is to seamlessly connect all stakeholders, providing the flow of important, strategical and sometimes crucial information. For this to happen, it is crucial to guarantee the continuity, integrity and security of communications between the stakeholders and the services here identified. Some principles that should be met are confidentiality, integrity, availability, authentication, lightweight solutions, heterogeneity, policies and key management systems (Mahmoud et al., 2015). Confidentiality and integrity are mandatory when dealing with livestock tracking, due to the sensitive nature of location data, especially in real-time solutions. It is important to mention that all the communication between users and the devices on animals, machines or places should be supported in real-time by the IoT infrastructure.

4. Discussions

On the beef cattle VC, multiple technologies will coexist for the different stages of data collection, processing and storage. This is especially evident in the Farm stage, due to its heterogeneous nature in comparison to the other VC links.

Since our main focus is to promote an overview and identify the main requirements and services of the beef cattle VC, the proposal of a framework considering those aspects, as well as the implementation aspects, must be further researched and discussed with both Farm administrators and stakeholders from other stages of the VC. In this sense, we believe that our work is an initial step towards this discussion, as it provides a subsidy for the development of a framework for Smart Beef Cattle Services based on IoT, as well as sets the ground rules for the development of similar research for other species or rearing methods. Therefore, it can be expected that farm management and operations research will greatly benefit from the resulting panorama and requirements.

Future work is related to the validation with different VC agents, the development of a framework for the beef cattle VC to operate under the IoT paradigm, as well as the extension to other livestock VCs, creating a Smart Livestock Farming infrastructure, and then, perhaps, expand to other agro-industrial VCs.

5. Conclusions

This work provides an initial step towards the creation of a Smart Beef Cattle value chain. It provides an overview of a typical beef production chain, specifies the stakeholders involved, its processes and the informational flow. It also identifies the main requirements for implementing IoT on the beef cattle value chain, which are organized in 9 main services and 31 sub-services or activities, allowing the value chain to function optimally using IoT technologies.

This work is an initial step towards a broader discussion, as it provides a subsidy for the development of a framework for Smart Beef Cattle Services based on IoT, as well as sets the ground rules for the development of similar research for other species or rearing methods, possibly scaling it up to a Smart Livestock Farming system.

In this sense, farm management research may benefit from the resulting requirements, using it to build services and architectures for a panorama of further automation and autonomous operations of the farms and supply chains.

Acknowledgments

This work was supported by the National Council for Scientific and Technological Development (CNPq), and also by the Itaú Unibanco S.A. through the Itaú Scholarship Program, at the Centro de Ciência de Dados (C²D), Universidade de São Paulo, Brazil.

References

Babinszky, L., Halas, V. and Verstegen, M.W.A. (2011) 'Impacts of climate change on animal production and quality of animal food products', InTech.

Banhazi, T. M. *et al.* (2012) 'Precision livestock farming: an international review of scientific and commercial aspects. International Journal of Agricultural and Biological Engineering, 5(3), 1-9.ITU-T, 2012

CEMA. (2017) 'CEMA - European Agricultural Machinery - Innovative Livestock Technologies: Making Livestock Farming More Animal-Friendly, Sustainable & Competitive.'

Coleman, S. W. and Moore, J. E. (2003) ''Feed quality and animal performance'. Field Crops Research, 84(1-2), 17-29.

Dabbene, F., Gay, P. and Tortia, C. (2014) 'Traceability issues in food supply chain management: A review'. Biosystems engineering, 120, 65-80.

Grandin, T. and NAMI. (2017) 'Recommended Animal Handling Guidelines & Audit Guide: A Systematic Approach to Animal Welfare'.

HFAC. (2018) 'Animal Care Standards: Beef Cattle'. Encyclopedia of Reproduction.

ITU-T. (2012) 'Y.2060 - Overview of the Internet of Things'. Recommendation ITU-T.

Mahmoud, R., Yousuf, T., Aloul, F. and Zualkernan, I. (2015) 'Internet of things (IoT) security: Current status, challenges and prospective measures', In: Internet Technology and Secured Transactions (ICITST), 2015 10th International Conference for. IEEE, pp. 336-341.

Sheridan, J. J. *et al.* (1991) 'Guidelines for slaughtering, meat cutting and further processing'. FAO.

Verdouw, C. N. *et al.* (2018) 'A reference architecture for IoT-based logistic information systems in agri-food supply chains'. Enterprise information systems, 12(7), 755-779.

Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M. J. (2017) 'Big data in smart farming – a review'. Agricultural Systems, 153, 69-80.

WSPA. (2011) 'Universal Declaration of Animal Welfare (UDAW) - GlobalAnimalLaw.Org'.

COMPARISON OF THE K-MEANS AND SELF-ORGANIZING MAPS TECHNIQUES TO LABEL AGRICULTURAL SUPPLY CHAIN DATA

Roberto F. Silva, Gustavo M. Mostaço, Fernando Xavier, Antonio Mauro Saraiva, Carlos E. Cugnasca

Department of Computer Engineering and Digital Systems,

Escola Politécnica da Universidade de São Paulo (USP), Brazil

roberto.fray.silva@gmail.com, gmostaco@usp.br, fxavier@usp.br, saraiva@usp.br, carlos.cugnasca@usp.br

ABSTRACT

The data produced in agricultural supply chains may be divided into three separated systems: (i) product identification and traceability, related to identifying production batches and places where the product has passed on the supply chain; (ii) environmental monitoring, considering mainly the temperature and relative humidity in storage and transportation; and (iii) processes, related to the data describing production processes and inputs used. Systems (i) and (ii) produce mainly structured data, while system (iii) produces non-structured data, and these are present in all agents in the supply chain. Data labeling on the different systems is an important step towards improving supply chain coordination and decision making related to traceability, production, and certification, among others. Nevertheless, it is a labor-intensive task, whose adoption is discouraged in data management activities. The main objective of this paper was to contribute to the reduction of interoperability problems by applying two clustering algorithms to label non-standardized data from agricultural supply chains. First, the data were clustered using k-means++ and self-organizing maps, with different model parameters. Then, a series of inferences were made to evaluate if the labels were correctly assigned, based on the characteristics of the data on each of the three systems. Lastly, a series of recommendations to improve the results of the models are discussed.

Keywords: agri-food, supply chains, unsupervised learning

1. INTRODUCTION

A supply chain (SC) can be described as a group of companies that are responsible for fulfilling the demands of a consumer segment (Chopra, Meindl, 2013). In the case of agricultural SCs, the products involved are agricultural products, such as grains, fruits, flowers, animal products, among others. An SC is divided into links or stages, which are groups of companies that are characterized by doing specific types of activities, such as farms, industries, logistics service providers, retailers, among others (Chopra, Meindl, 2013). Each company in an SC is termed an agent.

SCs generate a vast amount of heterogeneous data from different sources, processed on different systems and stored on different databases and formats (Corella, Rosale, Simarro, 2013). Generally, this data is not standardized among agents in the same link, and it can be processed and stored on different solutions, resulting in interoperability problems.

The data in an agricultural SC can be divided into three main systems: (i) product identification and traceability, related to identifying production batches and places where the product has passed among the SC; (ii) environmental monitoring, considering mainly the temperature and relative humidity during the warehousing and transportation activities; and (iii) processes, related to the description of production processes and inputs used. Systems (i) and (ii) produce mainly structured data, while system (iii) produces mainly non-structured data (Pang et al., 2015; Verdouw et al., 2013; Chopra, Meindl, 2013; Verdouw et al., 2016).

As the Internet of Things (IoT) technologies become widely adopted in the SCs, the quantity of data that is generated and its associated problems tend to increase. Lack of interoperability, long periods of implementation and lack of communication are some of the problems of adopting this paradigm without proper standardization and coordination among agents (Harris, Wang, Wang, 2015).

Interoperability is a significant problem, as most SCs will probably continue to have coexisting technologies in terms of data generation, storage, and processing, resulting in data on different formats and resolutions. The complexity and time needed to standardize the data generated by all agents are prohibitive. Researches by Park and Song (2015) and Pang et al. (2015) consider the implementation of middlewares for the reduction of interoperability problems. Nevertheless, most of the research is theoretical and does not consider practical applications.

The main objective of this paper is to contribute to the reduction of interoperability problems by applying clustering algorithms to label non-standardized data from agricultural SCs. It is unpractical to manually label this data due to its volume and variety. Yet, this would lead to improvements in SC coordination and decision making related to traceability, production, and certification, among others. Our main research question is: “Can the K-means++ and self-organizing maps (SOM) models help on agricultural SC data labeling?” This is evaluated in light of the results from the models’ implementations.

To the best of our knowledge, this is the first attempt to address this problem using unsupervised machine learning techniques. Two methods are implemented and studied: K-means++ and SOM. Logical inferences are developed to evaluate and improve the results. In addition, a framework for automatic data labeling will be briefly introduced.

The k-means is a clustering algorithm that was invented more than 50 years ago and is still used due to its good results (Jain, 2010). It basically creates points that will be used to calculate distances and form clusters, through a series of iterations. Nevertheless, it suffers from a problem due to the method used for its initialization, leading to local optima. To reduce the impact of this problem, the k-means++ algorithm was proposed (Arthur, Vassilvitskii, 2007). The k-means++ algorithm has a considerably lower processing time than using multiple random initializations.

The SOM algorithm is a neural network model for unsupervised machine learning and was created in 1982 (Kohonen, 1982). According to Kind and Brunner (2013), SOMs can be used to project n-dimensional data into a 2D map. It preserves the data topology, allowing the identification of hidden patterns that may not be observed in a k-means implementation, especially on datasets with a considerable amount of features. It can also be used for cluster analysis and labeling generation. Nevertheless, it is not nearly as used as the k-means algorithm.

2. METHODOLOGY

The methodology adopted for this work is divided into five main steps:

1. **Data collection and processing** using Python on the Jupyter IDE and LibreOffice Calc. It is composed of the following activities: **(i) data collection; (ii) data preprocessing** (unique identification of rows, separating target column from the data, dealing with missing data with zeros and interpolations, and .csv exporting); **(iii) data processing** (eliminate errors and outliers); **(iv) data fusion** (identify common features on the datasets, merge them, solve missing data problems and select which final features will enter the model); and **(v) data normalization** (using the MinMax scaler on Scikit-Learn package);

2. **Implementation of the K-means++ algorithm** using Scikit-Learn, considering seven implementations: K-means++, PCA with two features and five variations of k-means with random initialization (10, 20, 30, 40 and 50);
3. **Implementation of the SOM algorithm** using MiniSom[1], with the triangle neighborhood function and several combinations of the following parameters: learning rate (0.2, 0.3 and 0.5), sigma (3, 4, 5 and 6), and number of rounds (100, 500 and 1000);
4. **Definition and implementation of logical inferences** using Python in the Jupyter IDE, based on statistical analysis of the features and an in-depth literature review;
5. **Comparison of the implemented algorithms** using Scikit-Learn, divided into two parts: (I) unsupervised part, using silhouette score, homogeneity and processing time; and (ii) supervised part, using a confusion matrix and a classification report with the target labels.

Due to the lack of practical researches on clustering of agricultural SC data, and of suitable datasets that could be readily used in our research, some simplifying assumptions were made:

1. The predicted labels are related to the three main systems described in Section 1;
2. The considered SC is composed of four links: farm, industry, transportation, and retailer;
3. The dataset used for clustering was formed by the fusion of seven open datasets, described in Table 1. Each one represents partial data generated by different links in the SC as well as the most common information architecture present on it. This selection was based on a thorough search for open databases on Kaggle[2], Datahub[3] and Github[4]. Datasets features were analyzed during the data fusion activities;
4. The final dataset represents the situation of an agricultural SC, in which agents send their data to a common database in the cloud, without standardization.

Table 1. Open datasets selected to form the final dataset

Dataset number	SC Link and System	Description	Total number of features
01	Farm - 2	Environmental data collected[5]	9 (6 common, 3 unique)
02	Farm - 3	Herbicide application per plot[6]	10 (5 common, 5 unique)
03	Farm - 3	Plant growth measurements per plot[6]	12 (5 common, 7 unique)
04	Farm - 3	Productivity per plot[6]	6 (5 common, 1 unique)
05	Transportation - 1	Transportation from farm to industry[7]	9 (6 common, 3 unique)
06	Industry - 2	Environmental data at the industry[8]	15 (8 common, 7 unique)
07	Industry - 3	Operational data at the industry[9]	12 (5 common, 7 unique)

3. RESULTS AND DISCUSSION

Among the several implementations of k-means tested, the k-means++ and k-means random 50 times presented the best results. In relation to the SOMs implementations, the best results were achieved

1 https://github.com/JustGlowing/minisom
2 https://www.kaggle.com
3 https://datahub.io
4 https://github.com
5 INMET - Data from Rio Verde, GO - http://www.inmet.gov.br
6 Scribner et al (2003); Silva et al (2012); Omer et al (2015); Agridat - https://github.com/kwstat/agridat
7 Cashew Truck Arrivals - https://www.kaggle.com/extralime/cashew-truck-arrivals
8 Wsn-indfeat-dataset - https://github.com/apanouso/wsn-indfeat-dataset
9 Vega shrink-wrapper degradation - https://www.kaggle.com/inIT-OWL/vega-shrinkwrapper-runtofailure-data

with the following parameters: 3 (sigma), 0.2 (learning rate), and 500 (rounds). Table 2 contains the unsupervised quality metrics for the above mentioned methods.

From the analysis of these results, we can conclude that: (i) k-means++ was the best method due to a higher homogeneity score compared to the SOM model, it has a similar silhouette score compared to the k-means random 50 times, but presents significantly lower run time when compared to it; and (ii) the SOM model had a considerably worse silhouette score, and a much higher run time compared to k-means methods.

Some of the main reasons that may explain these results are: (i) the data imbalance among labels (System 1 contains 671 data points, while System 2 contains 20.527, and System 3 contains 15.153); (ii) the considerably small size of the dataset, especially for a neural network method; and (iii) the heterogeneity of data itself. Nevertheless, these results can work as a baseline for other implementations of unsupervised learning on the agricultural SCs domain.

Table 2. Model implementation results

Model	Processing time (s)	Homogeneity	Silhouette
K-means random 50 times	2.68	0.640	0.472
K-means++	0.41	0.640	0.426
SOM	52.50	0.569	0.012

After implementing the models and analyzing the unsupervised machine learning metrics, it was implemented a confusion matrix and a classification report, comparing the predicted clusters with the real labels. Table 3 contains the classification report for the k-means++ model and Table 4 contains the classification report for the SOM model. It is possible to observe that k-means++ can still be considered as the best model to cluster our dataset, presenting a considerably high precision, recall and F1-score across clusters 2 and 3. Nevertheless, it did not correctly classify cluster 1, missing all values. We believe that this is due to the data imbalance and the heterogeneity of this cluster. For the SOM model, although it manages to have a high recall for cluster 1, it achieves that by associating many more data points to this cluster than it really has (about 10 times more).

Table 3. Classification report for the k-means++ model

Quality metric	Cluster 1	Cluster 2	Cluster 3	Weighted average
Precision	0.00	0.86	1.00	0.90
Recall	0.00	1.00	0.81	0.90
F1-Score	0.00	0.93	0.90	0.90
Data points associated	250	23813	12288	-
Real number of data points	671	20527	15153	-

As a conclusion of both the unsupervised and supervised metrics analysis, the k-means++ model performed better than the SOM model. Nevertheless, this is an initial analysis of this domain, as stated before. More experiments with different datasets, are needed to validate this hypothesis. Based on the dataset that was created during our analyses, it is possible for other researchers to further improve on our findings.

Table 4. Classification report for the SOM model

Quality metric	Cluster 1	Cluster 2	Cluster 3	Weighted average
Precision	0.10	1.00	1.00	0.98

Recall	1.00	0.85	0.80	0.84
F1-Score	0.18	0.92	0.89	0.90
Data points associated	6589	18261	11501	-
Real number of data points	671	20527	15153	-

Based on our model implementation, we devised an initial framework for automatic data labeling for agricultural SCs. It is composed of the following steps:

1. **Data generation**, with standardization as the main target;
2. **Data collection**, with automatic row and dataset unique identification. This should, ideally, be done locally, if possible. However, as most agricultural SCs suffer from a lack of coordination, it could be executed automatically in the cloud;
3. **Data processing**, generating a new dataset without outliers, which are eliminated using logical inferences based on the features mean and standard deviation. Part of this process can be automatized. It is essential to maintain the original dataset, as some food quality problems are directly related to extreme variations on temperature, RH, gases, among others;
4. **Data fusion**, considering both the identification and selection of common features and the percentage of null cells in each feature. We believe this step can present some automation problems, as it would involve a considerable amount of inferences, which may not be clear before analyzing the datasets. More research is needed on this topic;
5. **Data normalization**, using the MinMax or standard scalers;
6. **Application of the labels on the original datasets**. This process can be automated, as the individual observations are uniquely identified. Future work is needed in order to identify if it is possible to extend the predicted labels to the observations that were considered initially as outliers and if any value is added to the SC by doing so.

4. CONCLUSIONS

In this paper, a dataset, representative of the data in an agricultural SC, was built and two models were implemented for clustering analysis to generate labels for these data: k-means++ and SOM. It was observed that the k-means++ performed better for both unsupervised and supervised machine learning metrics, while also having a considerably lower processing time. The results can be considered satisfactory, and a framework was proposed for automatic data labeling on this domain, becoming an important contribution of this work towards improving supply chain coordination and decision making in the agri-food sector.

Some limitations were observed, related to (i) the lack of open datasets that could be used to evaluate agricultural SCs, especially the ones containing data from the whole SC; and (ii) the lack of a framework to analyze heterogeneous, non-standardized data generated throughout agricultural SCs. Further work is needed to improve and enlarge the dataset. The implementation of other models, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Expectation-Maximization (EM) is also intended, as well as dealing with the current cluster imbalance in our data. The development of a common dataset, that can be used for evaluating data labeling in this domain, should be done in cooperation with partners from the different links of an agricultural SC.

ACKNOWLEDGMENTS

This work was supported by Itaú Unibanco S.A. through the Itaú Scholarship Program, at the Centro de Ciência de Dados (C²D), Universidade de São Paulo, Brazil, by the National Council for Scientific and

Technological Development (CNPq), and also by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

REFERENCES

Arthur, D., Vassilvitskii, S. (2007) 'K-means++: The advantages of careful seeding', In: Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027-1035.

Carrasco Kind, M., Brunner, R.J. (2014) 'SOMz: photometric redshift PDFs with self-organizing maps and random atlas', Monthly Notices of the Royal Astronomical Society, 438(4), pp.3409-3421.

Chopra, S., Meindl, P. (2013) 'Supply chain management: Strategy, planning, and operation', 5th ed. New Jersey, USA: Pearson Education, 528pp.

Corella, V.P., Rosalen, R.C., Simarro, D.M. (2013) 'SCIF-IRIS framework: a framework to facilitate interoperability in supply chains'. International Journal of Computer Integrated Manufacturing, 26(1-2), pp.67-86.

Harris, I., Wang, Y., Wang, H. (2015) 'ICT in multimodal transport and technological trends: unleashing potential for the future', International Journal of Production Economics, 159, pp.88-103.

Jain, A.K. (2010) 'Data clustering: 50 years beyond K-means', Pattern Recognition Letters, 31(8), pp.651-666.

Kohonen, T. (1982) 'Self-organized formation of topologically correct feature maps', Biological Cybernetics, 43(1), pp.59-69.

Omer, S.O., Abdalla, A.W.H., Mohammed, M.H., Singh, M. (2015) 'Bayesian estimation of genotype-by-environment interaction in sorghum variety trials'. Communications in Biometry and Crop Science, 10, pp.82-95.

Pang, Z., Chen, Q., Han, W., Zheng, L. (2015) 'Value-centric design of the internet-of-things solution for food supply chain: value creation, sensor portfolio and information fusion', Information Systems Frontiers, 17(2), pp.289-319.

Park, S., Song, J. (2015) 'Self-adaptive middleware framework for Internet of Things', In: 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE), pp. 81-82.

Scribner, E.A., Battaglin, W.A., Dietze, J.E., Thurman, E.M. (2003) 'Reconnaissance data for glyphosate, other selected herbicides, their degradation products, and antibiotics in 51 streams in nine Midwestern States', No. 2003-217.

Silva, A.M.D., Degrande, P.E., Suekane, R., Fernandes, M.G., Zeviani, W.M. (2012) 'Impacto de diferentes níveis de desfolha artificial nos estádios fenológicos do algodoeiro', Revista de Ciências Agrárias, 35(1), pp.163-172.

Verdouw, C.N., Vucic, N., Sundmaeker, H., Beulens, A. (2013) 'Future internet as a driver for virtualization, connectivity and intelligence of agri-food supply chain networks'. International Journal on Food System Dynamics, 4(4), pp.261-272.

Verdouw, C.N., Wolfert, J., Beulens, A.J.M., Rialland, A. (2016) 'Virtualization of food supply chains with the internet of things'. Journal of Food Engineering, 176, pp.128-136.

4CROP WEB APPLICATION FOR MONITORING RAIN-FED CROPS IN THE SAHEL

Tiziana De Filippis, Patrizio Vignaroli, Leandro Rocchi, Elena Rapisardi
National Research Council - Institute of Biometeorology, Florence, Italy
t.de.filippis@ibimet.cnr.it, p.vignaroli@ibimet.cnr.it, l.rocchi@ibimet.cnr.it, e.rapisardi@ibimet.cnr.it

ABSTRACT

Agro-geoinformation is the key information in the agricultural decision making and policy formulation process, especially in the countries where food security mainly depends on rain-fed crops production. It's the case of Sudano-Sahel zone where scarce economic resources hamper regular monitoring of crops development; a context that requires new approaches to detect crops risk zones during the agricultural season. The advances of Earth observation and sensing technologies, as well as geoprocessing web tools, enable new opportunities and challenges in applying agro-geoinformation to crop monitoring and assessment.

This paper presents the "4Crop" web application, an open source and interoperable solution for agricultural drought risk identification and forecasting in the Sahelian countries. The goal is the development of an operational tool that balances the lack of sufficient and timely acquisition of ground data using meteorological satellite open data sets. The whole web geoprocessing is based on the Crop Risk Zone (CRZ) model. The model performs a soil water balance to evaluate the satisfaction of crop water requirements in each phenological stage of the growing period. The model also provides a qualitative evaluation of the expected crop yields compared with the potential one, taking into account both water stress intensity and the phenological stage of crops.

The 4Crop web application currently running on Niger and Mali, and the outputs aim to identify installation and phenological phases of the main rain-fed crops (millet, sorghum, groundnut, cowpea) and to create crop risk zones images for each selected country.

The goal is to support Sudano-Sahel Early Warning Systems and any other local users in decision making and foster drought risk reduction and climate change resilience.

The proposed approach aims to encourage the integration and sharing of interoperable and open source solutions and thus contribute to the setting-up of distributed climate services in developing countries.

Keywords: agro-geoinformation, web application, geoprocessing, open source, climate services, Niger, Mali.

1. INTRODUCTION

Food security is still one of the major concerns that Sahelian populations have to face. In the Sahel, agriculture is primarily based on rainfed crops and it is often structurally inadequate to face the climatic variability. In general, low rainfall during the growing season can lead to crop yields decrease and, sometimes, to food crises (Sultan et al., 2005).

Crop yields may suffer significantly from either a late onset or early cessation of the rainy season, as well as from high frequency of damaging dry spells (Mugalavai et al., 2008). Therefore, the choice of the sowing date is of paramount importance for farmers. The ability to estimate effectively the onset of the season and potentially dangerous dry spells becomes therefore vital for planning rainfed agriculture practices aiming to minimize risks and maximize yields. Field observations in Mali, since 1983, show that when farmers use agro-meteorological information to plan the sowing date and to choose the varieties to be used, yields are higher compared to traditional choices (Hellmuth et al., 2007). As a consequence, reliable prediction of rainfall onset and duration is effective to reduce risk related to sowing date, planting practice and choice of varieties (Stewart, 1991).

In this context, advices to farmers are a fundamental component of prevention allowing a better adaptation of traditional crop calendar to climatic variability. Past experiences, in fact, show that agro-meteorological information can play a key role supporting food security, reducing the vulnerability of farmers, strengthening the rural production systems (Kleschenko et al., 2004) and that appropriated advices and agrometeorological information contributes in increasing crop productions.

In the Sudano-Sahel zone where scarce economic resources hamper regular monitoring of crops, the development of new approaches to detect crop risk zones during the agricultural season are required. The advances of Earth observation and sensing technologies, as well as geoprocessing web tools, enable new opportunities and challenges in applying agro-geoinformation to crop monitoring and assessment. As a matter of fact, in this Region web-based geoprocessing has not been yet fully explored for agricultural applications, whilst stand-alone applications and software are still widely used. The weak points of stand-alone solutions are time and effort required to install and manage the set-up, including the collection of geospatial data from a variety of sources, and pre-processing and analysing the data on local machines (Zhao et al., 2012). Indeed LDCs (Least Developed Countries) stand-alone applications, without continuous user support and updates, often makes the analysis and the application of geospatial data very expensive and time consuming.

The unavailability of timely meteorological data and the scarcity of funds for hardware and software maintenance, do not ensure near real-time drought monitoring on regular basis by National Early Warning Systems (EWSs) for food security. Moreover, the information and knowledge from geospatial data cannot be shared and integrated across organizations and communities.

The present work is focused on the development of a multipurpose, integrated crop monitoring web application, called 4Crop, which allows to monitor the impact of the drought stress on the main rain-fed agricultural production systems in the Sahel region.

2. Methodology

The web application, targeting Niger and Mali National Meteorological Services (NMSs), was implemented on a coherent Open Source web-based Spatial Data Infrastructure to treat all input and output data in an interoperable, platform-independent and uniform way.

The 4Crop client/server architecture has been implemented using open source tools and OGC (Open Geospatial Consortium) standards in order to guarantee the web application sustainability and the implementation of customized geospatial functions required by the end users.

The whole web geoprocessing is based on the Crop Risk Zone (CRZ) model (Vignaroli et al., 2016) that performs a soil water balance to evaluate the satisfaction of crop water requirements in each phenological stage during the growing period. The model is initialized by a rain threshold (10, 15 or 20 mm); this threshold depends on the crop calendar and the varieties traditionally used by farmers in different agro-ecological zones (Bacci et al., 2009). To best adapt simulation to the real behaviour of various cropping systems, the CRZ model allows users to customize some parameters: crop types and varieties, sowing conditions (rain threshold and start of simulation period) and extent of analysis area. So far the model has been tested on the following four crops: pearl millet (85 days and 130 days), cowpea (75 days), groundnut (100 days and 140 days), sorghum (110 days). The model data input can be summarized as follows:

- Gridded daily Cumulated Rainfall Estimate Images;
- Gridded daily Cumulated Precipitation Forecast (0 - 240 h);

- Gridded average daily potential evapotranspiration (PET) (from MOD16 Global Terrestrial Evapotranspiration Data Set);
- Average start of growing season (computed on the last 10 years);
- Average end of growing season (computed on the last 10 years);
- Gridded soil water storage capacity (from FAO Harmonized World Soil Database);
- Phenological phase lengths and crop coefficient Kc (FAO, 1998) for each simulated crop.

Due to the lack of a dense weather station network in Africa and the availability and consistency of long-term rainfall data for the Sahel Region, open satellite-derived datasets have been used to provide the CRZ model with the input data required for analysis. NCEP/NOAA Global Forecast System (GFS) is the reference data source for precipitation forecast images at 0.25° resolution; the Climate Prediction Center (CPC) Rainfall Estimator supplies daily Rainfall Estimates (RFE) at 0.1° resolution and EUMETSAT Earth Observation Portal makes available historical series of Multi-Sensor Precipitation Estimate (MPE) at 3 km resolution. For such reason, in the system architecture, chains for automatic data downloading (figure 1) ensure a continuous update of each data set, while specific procedures and services have been built up and integrated into the 4Crop environment to handle CRZ model input data flow.

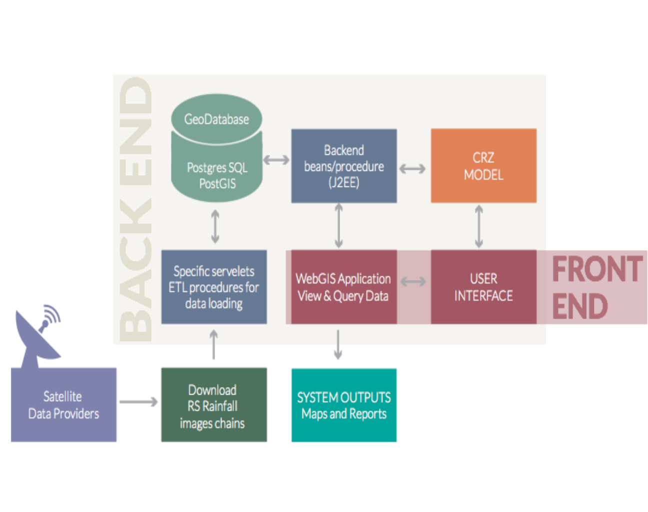


Figure 1. 4Crop System Architecture

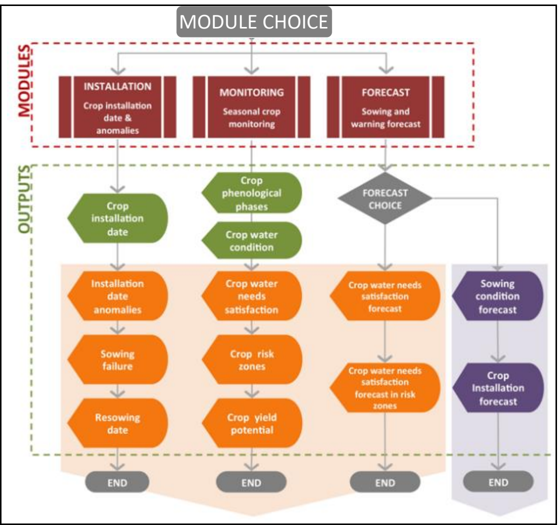


Figure 2. CRZ model flow chart

The CRZ model is composed of three modules of analysis (figure 2). The first two modules use rainfall estimate images as input data and operate in diagnostic mode. The third module employs precipitation forecast images and it works in predictive mode.

- **Installation module** provides an overview of the dates of successful seeding, showing areas where sowing failures may have occurred due to water stress. Zones where a crop was installed later than normal - because of the late onset of the rainy season or due to a first sowing failure - are also highlighted.
- **Monitoring module** performs a diagnostic of crop condition after its installation. The algorithm assesses the water requirements satisfaction and shows the areas where a water stress happened. The model also provides an estimate of the potential crop yield as a consequence of the water stress intensity and the phenological stage in which the stress occurred.
- **Forecast module** is composed of two sub-routines. The first, Sowing Forecast, provides outputs on the occurrence of favourable conditions for sowing and the subsequent crop installation. The second, called Warning Forecast, performs a prognosis on the possible occurrence of a water stress situation for crops already sown crops.

The UI (User Interface) has been implemented to allow users in initializing of analysis modules and to manage model outputs (viewing and saving or download data).

2.1 Open Source geoprocessing tools for CRZ model implementation

The CRZ model was developed using PL/pgSQL - SQL Procedural Language for PostgreSQL database system and PostGIS library built-in PostgreSQL. Each CRZ module is composed of a main PL/pgSQL function, performing initialization processes, and an iteration of functions for crop simulation processes. The CRZ modules work on input vector and raster data stored previously in the GeoDataBase. For example, in the “Installation module” the initialization processes set the parameters defined by the user (e.g. crop type, season length, country name) and extract the input data from GeoDB (e.g. daily RFE, daily PET images, season end and average sowing date images). All raster input images are clipped with the country’s boundaries, so time and resource consuming are optimized for the following processing phases. Within the installation module the iteration of functions generates module outputs (e.g. crop installation, sowing failures, etc.). The ST_MapAlgebra, as call-back function, performs pixel-by-pixel operations over raster images defined by the CRZ model algorithm.
At the end of the iteration cycles of each module, the main PL/PgSQL function stores the results in the GeoDB with all metadata information related to the model run.
Finally a JAX-WS, using the PostGIS predefined functions, publishes classified output images on 4Crop web Interface.

2.2 “4Crop” WebGIS application user requirements

During web applications development phase user requirements were defined through a User Consultation Process (UCP) involving the technical staff of Niger and Mali National Meteorological Services. The operators/users were interviewed in order to understand their specific needs in terms of usage, information products (maps and reports), and also to assess the usability needs in view of their previous experience with the CRZ `model` stand-alone software. The interviews allowed to better focus on user requirements, particularly for UI development. The UI prototype was shared with the users to obtain their feedback and further suggestions. In order to avoid any language barrier, which could prevent a wider use of the web application, 4Crop is available in French, the official language of the target countries (figures 3).
4Crop system has been conceived as a multipurpose tool (figure 4) in order to meet the needs of different categories of stakeholders, from farmers to political decision makers.

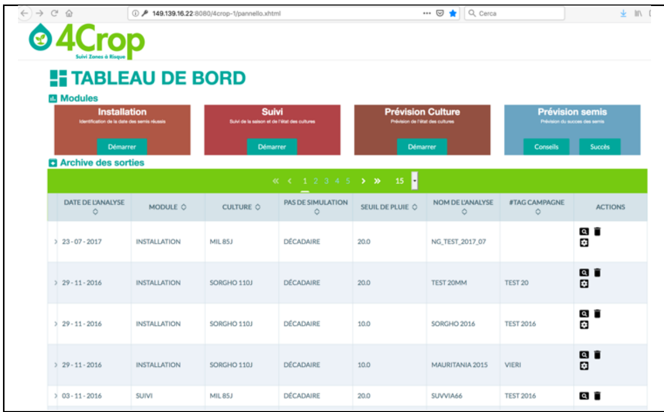


Figure 3. 4CROP User Interface

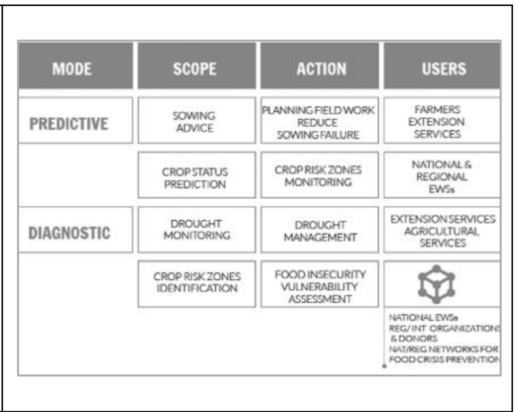


Figure 4. Multipurpose Tools

3. RESULTS

At present 4Crop is directly available by defined users’ profiles (personnel staff of NMSs of Niger and Mali) throughout a login procedure (http://149.139.16.22:8080/4crop-1/).
4Crop aims to localize potential crop risk zones, allowing an early evaluation of potentially involved production systems and population. Indeed, 4Crop aims to intercept agricultural drought phenomena

starting from the crops installation, detecting seeding delays and failures, and analysing the crop water balance during the cropping season. The model can simulate most widespread crops and varieties in the Sahel: pearl millet, sorghum, cowpea and groundnut.
The models have an overall tendency to postpone the sowing dates with an overall accuracy within a range between +/- 10 days.

The model results can be affected by rainfall estimated images used as input. RFE images which initialize 4Crop model, despite the lower spatial resolution, seem to better represent the rainfall patterns compared to the Meteosat Second Generation (MSG) images. This behaviour is more evident in the agricultural areas with a late installation of the monsoon circulation and a short duration of the rainy season.

Regarding the estimated potential yields, validation was performed on yield of millet and sorghum with respect to the data of the agricultural statistics of Mali at third administrative level (Cercles). The validation results have shown that the degree of accuracy of the estimates improves when the model is applied on homogeneous agro-ecological zones, rather than on the whole agricultural area of the country. Tests performed on sorghum using two different starting dates of sowing, early for the southern agricultural zone and delayed for the north, have marked values of overall accuracy significantly higher (0.75%) than those obtained with a single one (0.50%). At the same time, the double simulation has allowed to reduce the number of false negatives with an important increase of all the accuracy indexes.

The results of output validation process demonstrate that the 4Crop model performances in simulating the real field situations mainly depend from choice of the initialization parameters.

4Crop has been conceived to work over extended areas (until entire Sahel region) where there are different climatic and agronomic conditions. For this reason, it would be appropriate to perform analysis on areas which have homogeneous characteristics as regards the production systems (type of crops and most used varieties, the growing season length, rainy season period installation, etc.).

4. Discussion

A further element to consider in order to ensure an appropriate exploitation of 4Crop outputs is the understanding of the potential and limitations of the system. The validation process has shown how the analysed outcomes are influenced by the characteristics of the images of rainfall estimate. The setting of the start date of growing cycle and the choice of the useful rainy threshold useful for sowing are also essential for a right simulation of crops behaviour. The evaluation activities during the agricultural season necessarily require the involvement and participation of the final beneficiaries, in particular the NMSs. Their knowledge of different production systems and of the strategies adopted by farmers is fundamental for a rigorous evaluation of products and outcomes provided by 4Crop. Finally, Open Source solutions adopted in developing 4Crop System can make an important contribution to capacity building of local Institutions, which are the main actors in planning and implementation of prevention and response policies to potential food crises. Indeed, the 4Crop approach is meant to encourage the integration and sharing of interoperable and open source solutions among national and international stakeholders.

5. Conclusions

In developing countries, the development of climate services has been often hampered by the lack of strong national institutions, particularly weather services. Consequently, the setting up of interinstitutional alliances and collaborations involving research, national and international institutions are showing beginning results.

In this sense, 4Crop perspective is to contribute to the setting-up of distributed climate services, allowing stakeholders at different level to access and share information products through web services and standard protocols. Accessibility and availability, re-use and re-distribution of scientific products are recognized to be the prerequisite to build climate services worldwide based on the paradigm of Open Data Policies that should encourage the use of a trans disciplinary approach and collaboration to build products and services tailored on users' needs.

The 4Crop study case in Mali and Niger, aims to evaluate impacts due to drought stress during the whole crop growth cycle, providing farmers with information in order to implement appropriate and timely response strategies that minimize risk exposure to food security (Vignaroli et al., 2016).

The Spatial Data Infrastructure is designed to manage further agrometeorological products and services derived from advances in research in applied meteorology and climatology. The use of open source tools and standardized interoperable web services ensure sustainability in the development and deployment of web applications with geo-referenced data and customized territorial analysis that could be connected to other interoperable climate services. Indeed, the development of climate services that provide products tailored to different users implies the capability of using multiple data sources and mastering of competencies, which are not always available in national meteorological services.

Moreover, the whole infrastructure framework code is open source and can be shared to foster cooperation among software and interface designers, experts, practitioners, and researchers.

ACKNOWLEDGEMENTS

This work was supported by GFDRR - The Challenge Fund "Building Resilience to Drought in the Sahel by Early Risk Identification and Advices" Project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Bacci M., Di Vecchia A., Genesio L., Tarchiani V., Vignaroli P. (2009a). Identification et Suivi des Zones à Risque agro-météorologique au Sahel. In Changements Globaux et Dèveloppement Durable en Afrique,volume 3. Rome, Aracne Editrice. ISBN 978-88- 548- 2980- 0.

Hellmuth, M.E., Moorhead, A., Thomson, M.C., and Williams, J. (eds) 2007. Climate Risk Management in Africa: Learning from Practice. International Research Institute for Climate and Society (IRI), Columbia University, New York, USA.

Kleschenko A., Grom L., Ndiaye M., Stefanski R., 2004. The impact of agrometeorological applications for sustainable management of farming, forestry and livestock systems. Report CagM WMO/TD No. 1175, Geneva: World Meteorological Organization.

Mugalavai E.M., Kipkorir E.C., Raes D., Rao M.S., 2008. Analysis of rainfall onset, cessation and length of growing season for western Kenya. Agricultural and forest meteorology 148: 1123-1135.

Stewart, J.I., 1991. Principles and performance of response farming. In: Ford, W., Muchow, R.C., Bellamy, Z.A. (Eds.), Climatic Risk in Crop Production. Model and Management for the Semi-arid Tropics and Sub-tropics. CAB International, Wallingford.

Sultan B., Baron C., Dingkuhn M., Sarr B., Janicot S., 2005. La variabilité climatique en Afrique de l'Ouest aux échelles saisonnière et intra-saisonnière. II : applications à la sensibilité des rendements agricoles au Sahel. Sécheresse, 16(1):23-33.

Vignaroli, P., Rocchi, L., De Filippis, T., Tarchiani, V., Bacci, M., Toscano, P., Pasqui, M. and Rapisardi, E.: The Crop Risk Zones Monitoring System for resilience to drought in the Sahel. Geophysical Research Abstracts, Vol. 18, EGU2016-16616-3, 2016. EGU General Assembly, 2016.

Zhao, P., Foerster, T., & Yue, P. (2012). The geoprocessing web. Computers & Geosciences, 47, 3-12.

WATER USE EFFICIENCY (WUE) IN A HIGH DENSITY OLIVE GROVE: VARIETY, PLANTING DENSITY AND IRRIGATION EFFECTS

Athanasios Gertsis[1]and Gregory Yannakis[1]

[1]Perrotis College-American Farm School, Thessaloniki, Greece

agerts@afs.edu.gr,gyiann@perrotis.edu.gr

ABSTRACT

The new production systems in olives, with high (HD) and super high planting (SHD) densities, adapted fully to mechanical harvesting with row harvesters, represent a great challenge for all olive production countries. The available information is scarce and very localized, in terms of irrigation amount needed by the systems around the world. Great variations also exist, indicating that new approaches must be adapted to minimize water resources and increase the water use efficiency (WUE) in a tree species, which is a historical and cultural trademark in the Mediterranean areas and is rapidly expanding in all over the world. This study presents data from along-term high-density olive grove in Greece, a unique study where all major agronomic inputs are evaluated (varieties, planting densities, irrigation and ground and foliar fertilization rates). Results in the past years revealed that there are many opportunities to increase irrigation water use efficiency (WUE) and achieve sustainable production. The study follows a holistic approach and precision agriculture technologies are applied to increase the overall sustainability value of these systems. The WUE of the super high densities were higher than the lower densities, in both varieties used (Koroneiki and Arbequina), while Koroneiki was higher than Arbequina in WUE and production. These trends were also shown in the past years, under different climate conditions, while 2018 year was adverse for olive production in Greece and in most Mediterranean areas.

Keywords*:*olive, high density, Water Use Efficiency (WUE), Koroneiki, Arbequina.

1. INTRODUCTION

The use of water in agriculture (both in crop and animal production systems) represents a very significant portion of the total deposits in Earth and is under strong monitoring globally in the last decades. Optimal use of irrigation water will result in a higher Water use efficiency (WUE). In a generic agronomic approach, the crop WUE is meant to be the ratio of total water used by crops (irrigation+rainfall) over the total economic yield achieved at field conditions (Kijne et al., 2003; Molden et al., 2010a). Olive production systems can be classified in two main groups, based on the final type of product: table olive production which requires significant amount of irrigation water, while olive oil production systems can be either rainfed or irrigated. Recently the new high-density linear systems require irrigation to produce annually and sustainably. However, there are controversial information on the literature about the amount as well as the WUE of these systems. The complexity of the irrigation scheduling is increased by the important changes in the climate and the new olive cropping systems that emerged over the last 25 years, and by the current wide diversity of production systems inputs and cultural practices (tree density, irrigation water availability and use, fertilization practices, etc.).Approaches to schedule irrigation are based on developing specific growth stage ET coefficients and specific ET models (Ortaz et al., 2016; Subedi and Chavez, 2015), or stem water potential (SWP) periodic measurements.

In the context of this study, the agronomic approach will be used, since the emphasis is on how much total water was provided to the field (total of rainfall and irrigation) and how much economic product was output. This is a simple but more realistic approach and easily measured even by farmers. Also, the irrigation timing was mainly based on "crop available water" sensed by the plant in the form of soil water potential (Ψs).

Increases in water use efficiency are commonly cited as a response mechanism of plants to moderate to severe soil water deficits, and has been the focus of many programs that seek to increase crop tolerance of drought e.g. project AZORT (http://www.cespevi.it/azort/azort.html). However, there is some question as to the benefit of increased water-use efficiency of plants in agricultural systems, as the processes of increased yield production and decreased transpiration water loss (that is, the main driver of increases in water-use efficiency) are fundamentally opposed (Bacon, 2004).If there existed a situation where water deficit induced lower transpiration rates without simultaneously decreasing photosynthetic rates and biomass production, then water-use efficiency would be both greatly improved and a desired trait in crop production.

This study is a follow up of the past years (Gertsis et al., 2017) and differentiates in the new approach with the use of water marks sensors to monitor the soil water potential (Ψs,) rather than the soil volumetric water content (VWC) or stem water potential (SWP) used in few studies reported. It is not a labor intensive and continuous monitoring is technically feasible.

2. METHODOLOGY

The Educational-Research-Demonstration olive grove (N40° 34' 13.42" and E22° 59' 12.25") was established in late 2011 in Perrotis College, American Farm School of Thessaloniki, Greece (Fig. 1). It aims in applications of precision agriculture methodologies (Gertsis et al., 2013) and in long-term evaluation of the new linear production systems using high planting densities and adapted to fully mechanical harvesting. The grove Includes the following main treatments: two varieties (Koroneiki and Arbequina), three planting densities acronym Super-High Density (SHD), High density (HD) and Medium (MD) density at 1670, 1000 and 500 trees/ha, correspondingly, two irrigation levels (Conventional irrigation- CIand Deficit irrigation -DI) and two fertilization levels (conventional fertilization and 50% less). A number of trees from each treatment were harvested manually. CI and DI were calculated based on ET and soil water potential (Ψs) monitored by watermark soil moisture sensors (Fig. 2) installed in 6 lines and at three depths (20-40-60 cm). The critical value of SWP (Soil Water Potential) for deficit irrigation was set at 120 mbars (0.012MPa), a reference set value based on the typical root depth system of the linear olive systems. It was observed that the majority of the active root system is within the reported depth of 0-60 cm, due to the drip irrigation system used, the high planting density and the tree height, usually not exceeding 2.5 m to facilitate mechanical harvesting with straddle type harvesters.

The total amount of water used as irrigation and was rainfall 484 mm for the season to harvesting (Figure 3). It was monitored by a nearby installed automated weather station (Davis Vantage Pro, http://www.davisnet.com/solution/vantage-vue/) in the growing season ending to before harvesting, and was used to estimate WUE (total water used/total yiled) The average yield for each planting density across all irrigation treatments values and varieties will be presented.

Figure 1. The Educational-Research-Demonstration olive grove in Perrotis College, American Farm School of Thessaloniki, Greece

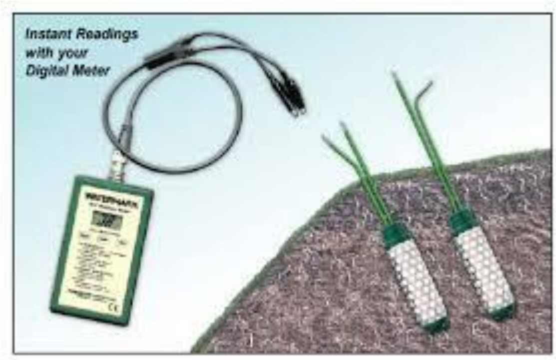


Figure 2. The soil sensors (watermarks) measuring soil water potential at 3 depths (20-40-60 cm)

3. Results

The results from the yield and the WUE for each planting density and the 2018 growing season are shown on Table 1.

Table 1. Yield, irrigation, and Agronomic Water Use Efficiencies (WUE) for the three planting densities (SHD-HD-MD) in 2018 season

Irrigation treatment	Variety	Average yield/tree (kg)			Total irrigation (mm)		
		SHD	HD	MEDIUM	SHD	HD	MD
Full	KORONEIKI	3.22	3.12	4.12	72.65	43.50	21.75
Deficit	KORONEIKI	3.15	2.95	4.09	44.59	26.70	13.35
Full	ARBEQUINA	1.80	1.95	2.01	72.65	43.50	21.75
Deficit	ARBEQUINA	1.67	1.72	1.98	44.59	26.70	13.35

AVERAGE YIELD (Kg/ha) adjusted			WUE(kg m^{-3})		
SHD	HD	MD	SHD	HD	MD

6990.6	4056.0	2678.0	1.26	0.77	0.53
6838.6	3835.0	2658.5	1.29	0.75	0.53
3907.8	2535.0	1306.5	0.70	0.48	0.26
3625.5	2236.0	1287.0	0.69	0.44	0.26

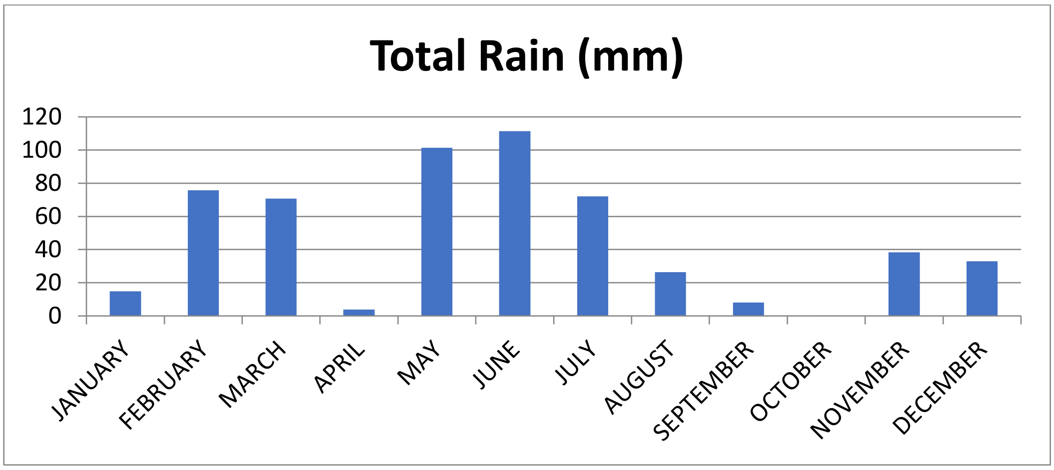


Figure 3. Rainfall amount and distribution in 2018 (Thessaloniki, Greece)

4. Discussion

Table 1 presents the calculated WUE for all three planting densities as an average across irrigation treatments and varieties for the 2018 growing season. The WUE trend of the high density systems shown is in close agreement but lower, when adjusted, with other conventional olive production systems (Hijazi et al., 2014 and Nuberg and Yunsua, 2003). Since all the trees are relatively young in age (< 7-8 years), the WUE was affected by the planting density and so, the SHD system produced the highest WUE as compared with the HD and the MD density. It must be mentioned that this year (2018) was considered one of the worst in terms of production in Greece, for both table olives and olive oil. This was due to unusual climate conditions (heavy rainfall in the flowering period (April-May) and in summer (May to early August. In addition it was followed by a continuous drought period from early August to October and the high damages caused by the insect *Bactrocera* (Dacus) *oleae* and the fungus *Colletotrichum*, and mainly by C. *gloeosporioides*, C. *acutatum* και C. *clavatum*. Therefore, the data presented herein are indicative of a comparatively very low yielding olive production season. The olive oil quality reported in another study, followed the same trends with yield and WUE among all treatments and in general was low.

Hijazi et al. (2014) reported from an olive grove in Syria, WUE values ranging from 0.4 to 2.1 kg $m^{-3}$for a range of irrigation systems. Average production of fruits was 8.53 tons/ha or 48 kg/tree and a planting density of 180 trees/ha. Using drip irrigation, WUE increased from 1.3 kg/m^3 to 2.36 kg/m^3 compared with surface irrigation. These values are for conventional planting systems and older trees (18 years old) and are comparable to the high-density systems of the presented study under drip irrigation, considering approximately similar tree age and yields/ha. Nuberg and Yunusa (2003) reported for Australia the WUE for two years ranged from 0.4 to 2.1 kg/m^3.

The SHD olive systems have an advantage during the first 4-10 years, due to higher population per unit area and, therefore, more efficient use of water and fertilizer inputs; however, the yield of the SHD systems was reported to level-off or slightly decreasing after 8-10 years. In this later period, the HD systems may have an overall advantage in WUE. This is a speculation not yet reported for Greece.

Comparing the CI and DI irrigation treatments, it was shown that the deficit irrigation has a very similar WUE to the conventional while Koroneiki variety had a much higher WUE compared to Arbequina, under the same irrigation amounts provided to both varieties. The performance of Koroneiki, was consistent in previous years, under more optimum climatic conditions (Gertsis et al., 2017).

5. Conclusions

It appears that olive trees in super high density (SHD) systems utilized better the total water in the field, per unit area, than the lower density systems and resulted in a higher WUE. These results followed the same trends shown in previous years (Gertsis et al., 2017) although this year was an exceptionally low production situation. Koroneiki out-yielded Arbequina in a variety of climate conditions. An important conclusion, verified again in this year, is that the WUE of the low irrigation levels are very competitive and can result in significant savings of irrigation and other production inputs for the new olive production systems. This long-term study will produce in the forthcoming years, additional data to further evaluate the interaction of WUE and the Total Fertilizer Use Efficiency (TFUE), in an effort to increase the WUE of the systems evaluated and provide more practical information to olive farmers for optimal management of high density olive systems in Greece and worldwide.

Acknowledgments

The authors express their gratitude to Anastasopoulos Nurseries, Greece (www.anastasopoulos-nurseries.com), for providing the plant materials, their harvesting equipment and their valuable know-how on these olive production systems.

References

Bacon, M. 2004. Water Use Efficiency in Plant Biology. Oxford: Blackwell Publishing Ltd.

Gertsis, A., D. Fountas, I. Arpasanu and M. Michaloudis. 2013. Precision Agriculture applications in a high density olive grove adapted for mechanical harvesting in Greece. Procedia Technology 8 (2013) 152-156

Gertsis, A., K. Zoukidis and A. Mavridis. 2017. Evaluation of the water footprint and water use efficiency in a high density olive (*Olea europea* L.) grove system. *European Water* Special Issue III-59. Paper No. 49. EWRA2017 - 10th Word Congress on Water Resources and Environment). http://www.ewra.net/ew/issue_59.htm

Hijazi, A., M. Doghoze, N. Jouni, V. Nangia, M. Karrou, and T. Oweis. 2014. Water requirement and water-use efficiency for olive trees under different irrigation systems. 7th International Conf. on Water Resources in the Mediterranean Basin, Oct 10-12, Marrakech, Morocco.

Kijne, J.W., Barker, R., and Molden, D. (eds) 2003. Water Productivity in Agriculture: Limits and Opportunities for Improvement. Comprehensive Assessment of Water Management in Agriculture Series 1. CAB International, Wallingford, UK in association with International Water Management Institute (IWMI), Colombo.

Molden D, Oweis T, Steduto P , Bindraban P, Hanjra MA , Kijne J. 2010a. Improving agricultural water productivity: Between optimism and caution. Agricultural Water Management 97: 528–535

Nuberg, I. and I. Yunusa. 2003. Olive water use and yield – monitoring the relationship. RIRDC Publication No 03/048 RIRDC Project No UA-47A

Orgaz, F., Testi, L., Villalobos, F.J., Fereres, E. 2006. Water requirements of olive orchards II: determination of crop coefficients for irrigation scheduling. Irrigation Science, 24:77-84.

Subedi, Abhinaya and José L. Chávez. 2015. Crop Evapotranspiration (ET) Estimation Models: A Review and Discussion of the Applicability and Limitations of ET Methods. Journal of Agricultural Science; Vol. 7, No. 6; 2015.

.

ASSESSING THE EFFICIENCY OF ARABLE CROPS PRODUCTION IN A CROSS-NATIONAL CONTEXT

Spyros Niavis[1], Leonidas Sotirios Kyrgiakos[2], Christina Kleisiari[2] and George Vlontzos[2]

[1]Department of Planning and Regional Development, University of Thessaly, Greece

[2.]Department of Agriculture, Crop Production and Rural Environment, University of Thessaly, Greece

spniavis@uth.gr, lkyrgiakos@uth.gr, chkleisiari@uth.gr, gvlontzos@uth.gr

ABSTRACT

According to FAO, improving the efficiency of resources use for the production of agricultural products is a key prerequisite for achieving global agricultural sustainability. Thus, it is necessary to incorporate performance evaluations in order to assess sustainability from an academic standpoint in the context of supporting international agricultural policy such as the Common Agricultural Policy (CAP). While efficiency is a major issue, there is a marked lack of bibliographic references focusing on agricultural efficiency issues on a global scale. However, much of the relevant international agricultural literature focuses on issues that concern individual countries or groups of countries with common characteristics. In the light of the above, this research seeks to evaluate the efficiency of national agricultural sectors in the production of arable crops internationally, a fact that will be achieved through Data Envelopment Analysis (DEA).

Keywords*:* agricultural efficiency**,** arable crops, sustainability, Data Envelopment Analysis

1. INTRODUCTION

As United Nations estimate, current human population is 7,7 billion people and it will be increased to 9,7 billion until 2050 (United Nations, 2019), meaning that resource management is essential to ensure food security under sustainable development principles. It should be notified that population models predict an average of 12.9% of urban population rise, leading to the creation of territories with extremely high needs of energy and food in a very small scale. In this context, agricultural sector should deal with the new challenges, maximizing production output using the least needed resources as inputs. Considering also climate change consequences, efficiency of arable crops should be examined in a global scale perspective to achieve production of adequate amounts of food and feed, while on the same time environmental impacts should be minimized.

Cereals production is associated with almost 50% of total cultivated area globally (14 billion acres), covering human and animal nutritional needs, while being an essential ingredient for the production of a wide range of products (FAOSTAT, 2019). Food security, high energy value and adequate protein concentration are the primer reasons for which cereals hold leading positions in all national agricultural sector, both in developed and developing countries. Cereals can be found all around the globe apart from hot and very wet regions. Maize contributes 45% of global cereal production, playing a significant role both for human nutrition and animal industry as feed. Moreover, a series of products can be made out of maize and its by-products affecting multiple sectors. Maize, sorghum and rice grown in countries with warmer climate. Asia holds 92% of global rice production and maize feeds the largest amount of Americans. Wheat, barley, rye and oats can be concerned as colder climate cereals. Wheat is mainly produced in Asia and Europe while barley is mainly cultivated to Russia, Canada and Germany due to its ability of producing adequate amounts of yield in a wide range of soil and climatic conditions (Gustafson, Raskina, Ma, & Nevo, 2009). Rising interest for leguminous crops have been observed lately, due to their ability to produce high protein seeds, to increase soil fertility through nitrogen assimilation and being a great alternative as green manure instead of the conventional ones (Plaza-Bonilla, Nolot, Raffaillac, & Justes, 2017).

Land availability, labour cost and adaptation of new technologies can be critical factors for achieving high efficiency scores in national level. In order to face increasing needs for food due to rising human population and climate change consequences, a more efficient use of current resources is required. For example, Zhang *et al.* (2015) present a scenario of a greater distribution of nitrogen resources on a global scale depending on each country needs, for increased nitrogen efficiency use ratios and food security. Moutinho *et al.* (2018) have investigated European national agricultural sectors from 2005-2012 using DEA and SFA methodology in order highlight differences on resources use and environmental impact between different countries. Through this analysis Cyprus, Denmark and France achieved the highest efficiency scores thus Belgium and Bulgaria had the lowest scores. It should be notified that in this survey no climate data have been taken into consideration. Wang *et al.*, (2019) have also identify efficiency differences in China's provinces, concluding that the average grain production could be 25% more efficient for the reference period 2000-2010. A global scale analysis, examining economic and regional characteristics such as temperature, rainfall and slope, resulted in a high yield gap of 46%, 50%, and 46% for wheat, maize, and rice accordingly, revealing the need for an agricultural production management in a global context (Neumann *et al.,* 2010).

According to previous remarks, different regions have been specialized in different types of crops, producing bigger amounts of products and gaining a competitive advantage among others in global markets. Aim of this paper is to assess efficiency levels of different agricultural sectors around the globe, while the efficiency differences between various types of crops. In order to achieve the above-mentioned goal, the basic research question of the present is the following:

1. Is there any efficiency deficit on arable crops farming among the different areas of the world?

2. Methodology

In order to run a DEA model we suppose that we have *n* national agriculture sectors expressed as NAS_j $(j = 1,2,\ldots,n)$, and $x_{ij} > 0$ is the amount of input i used in NAS_j and $y_{rj} > 0$ is the amount of output r of NAS_j. Then, the efficiency score for the NAS_i is extracted by solving the following model (Zhu, 2014)

$$\begin{aligned} &\varphi^* = \max \varphi \\ &\text{s.t.} \sum_{j=1}^{n} x_{ij}\lambda_j \le x_{io} \quad i = 1,2,\ldots,m \\ &\sum_{j=1}^{n} y_{rj}\lambda_j \ge \varphi y_{ro} \quad r = 1,2,\ldots,s \\ &\lambda_j \ge 0 \qquad \mathrm{j} = 1,2,\ldots,n \end{aligned} \tag{1}$$

Where φ^* is the decision variable which represents the relative technical efficiency of NAS_0 and λ_j the weights that NASj places on NAS_0 to construct its efficiency reference set. For a fully efficient NAS the φ^* score is 1, whilst inefficient NASs acquire a score above 1. Analysis incorporates various models which correspond to different types of arable crops which are presented in Table 1. Model 1 considers the total aggregated production of arable crops whilst the Models 2-4 focus on specific types of crops. For all the models the output is the total produced output volume expressed in thousand tonnes whilst the input is the total cultivated area expressed in hectares.

Table 1. The Data Envelopment Analysis models for the different types of arable crops

DEA Model	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Arable Crops Considered	Total Arable Crops (Cereals, Coarse Grains, Fiber and Roots, Pulses, Vegetables	Cereals	Coarse Grains	Pulses
Output	Production in thousand Tonnes			
Input	Area in Hectares			

Efficiency assessment is conducted in a two-step approach. In the first step, the globe is divided in 17 sub-regions and respective efficiency evaluations are conducted among the countries of each region by applying all the four models (the first stage DEA analysis is omitted for North America and Oceania due to the small sample of countries). For each region the two most efficient countries under the four models are selected and then included in the final global sample of countries. Then, in the second step all four models are applied on the global sample in order to extract efficiency evaluations and respective rankings of the considered countries. Figure 1 presents the methodological context of this paper explaining the two stages of the analysis. First stage contains the assessment of efficiency of national agricultural sectors (NAS) both in regional and global scale, giving them a representative ranking. For this purpose, Data Envelopment Analysis has been used in order to classify different national agricultural sectors using an output maximization approach (output-oriented).

3. RESULTS

3.1 Results of regional and global efficiency assessment

In Table 2 the basic descriptive statistics of the variables of the global model are presented. The area is expressed in thousand hectares (kha) and the production in thousand tonnes (kt). The mean cultivated area for arable cops is estimated at 68,293kha and the output at 467,632kt. The coefficient of variation (CV) values of the two variables signify that the output present larger variations than the cultivated land among the countries. The larger variability of output is also testified by the larger max-min difference. The lowest variability of the variables is found for the pulses as the CV coefficients are smaller than these of the other crops.

Table 2. The descriptive statistics of the variables of the DEA global models. Source FAOSTAT (2019)

	Variables							
Statistics	**Total Arable Crops**		**Cereals**		**Coarse**		**Pulses**	
	Area	Production	Area	Production	Area	Production	Area	Production
Mean	68,293	467,632	32,210	142,475	25,327	139,194	3,992	5,590
St.dev	206,728	1,813,154	87,578	479,845	78,205	578,600	8,696	13,592
CV%	303%	388%	272%	337%	309%	416%	218%	243%
Min	58	363	6	14	6	23	3	4
Max	1,444,043	13,198,639	597,164	3,558,932	435,997	386,2154	44,713	65,079

In addition, Table 3 illustrates the results of the four models. In more detail, the first column shows the average yield values for the 17 regions, followed by descriptive statistics of the efficiency scores, as well as the three best performing countries and the worst performing countries for different types of arable crops. Individual results of all countries can be shared by the creators upon request, as they are not presented here due to the difficulty caused by the limited size of the paper. Concerning the different types of arable crops, the results show that there are considerable variations in the efficiency of the regions. The average yield score for the total production of arable crops (6.495) is higher than the average scores of all the four models, indicating that it has the highest levels of inefficiency. On the contrary, cereal production has the lowest inefficiency with a yield score ranging to 2,932. Finally, the Coefficient of Variation proves that coarse grain production has the largest variability of efficiency, while the lowest variability belongs to pulses′ production.

Table 3. The Data Envelopment Analysis results per region and the three best and worst performing country per each model

Average Efficiency per Region	Total Arable Crops	Rank	Cereals	Rank	Coarse Grain	Rank	Pulses	Rank
Central Europe	1.985	1	1.032	1	1.158	1	1.283	1
West Asia	2.378	2	4.147	15	3.757	11	2.692	4
East Asia	2.861	3	1.448	2	2.943	9	3.540	7
South Europe	3.755	4	2.366	6	1.910	2	2.786	5
South America	3.926	5	1.622	4	2.436	5	4.115	9
North Europe	4.883	6	1.555	3	2.084	4	2.028	2
North Africa	5.418	7	3.689	13	4.434	13	3.981	8
Central Asia	5.464	8	3.069	12	2.772	7	2.383	3
Southeast Asia	6.117	9	2.025	5	2.574	6	5.353	11
North America	6.560	10	2.467	8	1.958	3	4.274	10
Central Africa	6.817	11	7.005	16	7.321	14	6.115	14
South Asia	7.341	12	2.451	7	2.903	8	5.365	12
Oceania	7.983	13	3.025	11	2.982	10	3.240	6
Central America	10.379	14	3.965	14	3.793	12	5.432	13
South Africa	10.469	15	2.819	9	9.350	16	7.482	16
West Africa	13.090	16	3.024	10	8.280	15	7.173	15

Statistics

	Total Arable Crops	Cereals	Coarse Grain	Pulses
Grand Total	6.495	2.932	4.106	4.392
Max	32.263	14.904	26.459	12.729
Coefiicient of Variation	0.784	0.778	1.025	0.564
Most Efficient	Netherlands Jordan Belgium	Ireland Belgium Netherlands	Belgium Mexico Israel	Ireland Belgium Netherlands
Lest Efficient	Mali Namibia Mauritania	Jordan Trinidad and Tobago Congo	Congo Mauritania Namibia	Ghana Zambia Mauritania

Central Europe, represented by the countries of Belgium, the Netherlands and Ireland, is the region with the highest efficiency in the four models. In more detail, Ireland is the best player in the cereal and legume models, Belgium in the coarse grains model and the Netherlands in total arable crops.

With exception of Models 1 and 3, the countries in the region are represented in the best performing team in all models. In more detail, Jordan appears to be the second most effective country in the sample in the first model, while the latter model, Belgium, Mexico and Israel appear to be the leading countries. Central Europe is followed by countries in Western and Eastern Asia, South Europe and South America. As far as West Asia is concerned, the efficient production of legumes and other vegetables is the reason for its high yields, while cereals and coarse cereals are less efficient. East Asia's high position is due to efficient cereal production, while South Europe's competitive advantage in the production of coarse grains is responsible for the region's high ranking.

In contrast, South and West Africa, but also Central America, are the least effective regions of the genus that are proven to have low yields in terms of the overall arable crop model, with rather low yields for individual crops as well. In addition, countries in both West and South Africa, with the aim of focusing on cereal production, have the potential to improve their overall competitiveness. Mauritania and Namibia are considered the weakest countries as they are most often included in the worst performing categories.

The revealed variability of the scores and rankings of the various regions signify their strengths and weaknesses regarding their ability to remain competitive in the production of the various types of arable crops. Therefore, regions which are rather inefficient in a type of crop could be strongly competitive in the production of other crops. Illustrating, Central Asia is very weak in cereal production as it ranks 12th whilst possessing a strong advantage in pulses productions signified by the fact that it is found to be the third most efficient region of the world. The same stands true for the countries of Oceania which also present a competitive advantage in pulses production. In addition, North America seem to present modest performance in all types of crops with excepting the coarse grain for which is found to be the third most efficient region. Finally, for the least developed countries of South and West Africa it is apparent that their competitive advantage could be based on the cereals production for which its relevant position is far better than the respective position under the other types of crops.

5. Conclusions

In this paper, total agricultural production along with cereals, coarse grains and pulses production of agricultural sectors around the globe have been examined with the use of four different models in order to highlight efficiency differences in both national and continent level. Countries of Central Europe have achieved the highest efficiency scores regarding arable crops production. Average scores of different NAS have resulted in no significant differences of efficiency in continent level, highlighting the fact that developing countries should focus on arable crop production in order to gain an advantage in the global market. Moreover, developing countries should intensify their efforts on arable crops production in order to use land more efficiently for cultivation of suitable cultivars depending on their climatic conditions. Increased food security in national and worldwide level, higher revenues for local food networks and greater resistance of food supply chain to climate change can be achieved from the intensification of arable crops production, under the circular economy context.

References

FAOSTAT. (2019). FAOSTAT. Retrieved February 27, 2019, from 2019 website: http://www.fao.org/faostat/en/#data

Gustafson, P., Raskina, O., Ma, X., & Nevo, E. (2009). Wheat Evolution, Domestication, and Improvement. In *Wheat Science and Trade*. https://doi.org/10.1002/9780813818832.ch1

Moutinho, V., Madaleno, M., Macedo, P., Robaina, M., & Marques, C. (2018). Efficiency in the European agricultural sector: environment and resources. *Environmental Science and Pollution Research*, *25*(18), 17927–17941. https://doi.org/10.1007/s11356-018-2041-z

Neumann, K., Verburg, P. H., Stehfest, E., & Müller, C. (2010). The yield gap of global grain production: A spatial analysis. *Agricultural Systems*, *103*(5), 316–326. https://doi.org/10.1016/j.agsy.2010.02.004

Plaza-Bonilla, D., Nolot, J. M., Raffaillac, D., & Justes, E. (2017). Innovative cropping systems to reduce N inputs and maintain wheat yields by inserting grain legumes and cover crops in southwestern France. *European Journal of Agronomy*, *82*, 331–341. https://doi.org/10.1016/j.eja.2016.05.010

United Nations. (2019). World Population Prospects. Retrieved February 26, 2019, from https://population.un.org/wpp/

Wang, P., Deng, X., & Jiang, S. (2019). Global warming, grain production and its efficiency: Case study of major grain production region. *Ecological Indicators*, *105*, 563–570. https://doi.org/10.1016/j.ecolind.2018.05.022

World Economic Forum. (2018). The Global Competitiveness Report 2017-2018. Retrieved February 23 2019, from http://reports.weforum.org/global-competitiveness-index-2017-2018/#topic=data

Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., & Shen, Y. (2015). Managing nitrogen for sustainable development. *Nature*, *528*(7580), 51–59. https://doi.org/10.1038/nature15743

Zhu, J. (2014). Quantitative Models for Performance Evaluation and Benchmarking. In *Operations Research*. https://doi.org/10.1007/978-3-319-06647-9

Economic Analysis of Unmanned Ground Vehicle Use in Conventional Agricultural Operations

Maria G. Lampridi[1,2], Dimitrios Kateris[1], Spyridon Tziakas[1] and Dionysis Bochtis[1]

[11]Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Greece

[2]Faculty of Agriculture, Forestry and Natural Environment, School of Agriculture, Aristotle University of Thessaloniki, Thessaloniki, Greece

m.lampridi@certh.gr, d.kateris@certh.gr, s.tziakas@certh.gr, d.bochtis@certh.gr

Abstract

With the rapidly developing technology of robotic vehicles and smart farming systems, conventional agricultural practices are evolving towards a new era of automation that promises to increase their efficiency and effectiveness contributing towards the need for increased production with lower economic and environmental costs. However, such technology is relatively new and most of the resulting products and services are used in an experimental stage. The scientific community and the industry mainly focus on the advancement of technology with the introduction of new smart farming products and services for which an extensive economic feasibility analysis has not yet been carried out. To that end, the aim of the paper is to perform an economic feasibility assessment of replacing conventional agricultural machinery and human labor with "smart" farming systems. The methodology used adopts the principles of conventional agricultural machinery cost calculation adjusting them to the use of Unmanned Ground Vehicles (UGV). On this basis, the cost of performing a conventional agricultural operation with the use of a robotic vehicle is estimated for a variety of different production scenarios. The scenarios are distinguished on the basis of the cultivation size and the application of different operation management schemes, as for example different charging times and the use of multiple vehicles to avoid the dead times caused by charging. The results highlight the effect of operation management in the overall efficiency of such systems which eventually affects the operation duration and the resulting cost, despite the fact that there are still many factors that need to be further investigated for the accurate cost estimation, e.g. repair and maintenance cost, salvage value.

Keywords*:* UGV, economic analysis, precision farming, robotic system

1. Introduction

The employment of autonomous robotic systems is connected to the rapid evolution of technology (Bochtis et al., 2014) and the turning towards an online world were all operations are automated (Bochtis *et al.*, 2015; Ampatzidis et al., 2017). As a result, all the aspects of production and services are being robotized in order to satisfy increased demand more efficiently (Qureshi and Syed, 2014; Marinoudi *et al.*, 2019). The swift towards robotization is also evident in the agricultural sector since the need for increased efficiency of agricultural practices with a reduced environmental burden (Toledo *et al.*, 2014) drives the evolution of technology towards integrated "smart" farming systems that replace conventional agricultural machinery and practices (Bochtis *et al.*, 2011). Precision agriculture techniques that employ robotic agricultural systems have been introduced towards the reduction of inputs considering real field condition and needs (Bongiovanni and Lowenberg-Deboer, 2004). However, as it is the case with all the newly introduced technologies it is important to assess their feasibility to support their adoption and acceptance by their wider sector (Moradi et al., 2018). Also, in the context of a wider sustainability assessment it is important to determine the economic feasibility of agricultural robotic systems, and examine the costs and benefits arising from their use.

In the agricultural sector, Pedersen et al. (2006) performed an economic feasibility analysis to compare three conventional agricultural systems and the respective autonomous systems. Their work concluded that the robotic systems were more economically feasible than the respective conventional in all the cases tested. Toledo et al. (2014) assessed an electricity powered prototype mechanical weed control system to determine the energy costs of its operation concluding that the evolution of autonomous vehicle technology could further reduce the energy costs. Lampridi et al. (2019) attempted to propose a methodology for the economic analysis of the employment of robotic systems in arable farming adopting the methodology of conventional machinery use cost estimation (Lampridi et al., 2019). The methodology was implemented for the estimation of the cost of light soil cultivation.

The present paper attempts to examine the feasibility of robotic system's employment for a variety of different operational schemes concerning the robotic system's functionality. The robotic system considered for the purposes of this study is an autonomous UGV performing light soil cultivation (Grimstad and From, 2017). The application of the system is examined in the small and the large scale and it is compared to the relevant conventional application.

2. METHODOLOGY

In order to estimate the cost of robot employment in light soil cultivation the model presented in Figure 1 was followed. The model along with the respective equations is described in detail in Lampridi et al. (2019). In accordance with the calculation of the cost of performing a conventional agricultural operation, the total cost comprises of the Ownership and the Operational cost. The Ownership includes the capital and depreciation costs of the machinery as well as the housing and insurance costs. These costs can also be characterized as fixed costs since they are not related to the operation performed but they are determined by the machinery purchase cost and economic parameters such as the interest rate and the inflation.

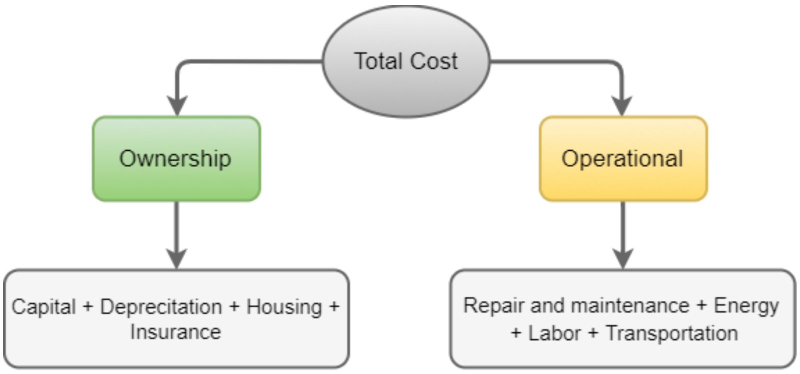


Figure 1. Cost Estimation Model

On the other hand, the operation cost is determined by the nature of the operation performed. It consists of the energy, the labor the transportation and the repair and maintenance cost. In the case of conventional machinery for the calculation of the repair and maintenance cost the ASABE standards are used (American Society of Agricultural and Biological Engineers, 2011). However, in the case of a robotic system such standards cannot apply. To that end, the approach of Bubeck et al. (2016) is followed. According to this approach the repair and maintenance cost of the robotic system is considered as a percentage of the respective conventional one.

3. Case Study

The aforementioned methodology was used to estimate the cost of performing light soil cultivation with the use of an unmanned ground vehicle in different operational schemes. The results are also compared with those of performing the operation with the use of a conventional system in order to maintain a base of reference. The system is examined in two different area scales. The Small Scale concerns an area of 10ha while the large scale an area of 100ha. In the case of the robotic system the same UGV is used in both small- and large-scale applications while in the case of conventional system different tractors are used. Table 1 presents the input parameters that were used for the cost estimation of the basic scenarios. In more detail a 40kw and an 80kw tractor caring a mechanical row-crop cultivator (C-shank) was selected for the small-scale and the large-scale scenarios respectively. In all the scenarios examined the interest rate is 9%, the inflation is 4% while the economic life of the machinery is 15 years.

Table 1. Input parameters for the economic model

		Robotic	Conventional Small-scale	Conventional Large-scale
Field operation parameters	**Working width (m)**	1.2	2.6	6.0
	Speed (km h^{-1})	4	8	8
Investment and ownership parameters	**Purchase price (€)**	50,000	40,000	130,000
	Implement purchase price (€)	1,000	3,000	6,000
	Salvage value (%)	10.9 [1,2]	10	10
	Housing coefficient	2.25	0.75	0.75
	Insurance coefficient	0.75	0.25	0.25
Machinery parameters	**Repair and Maintenance (R&M) factors**	-	0.003	0.003
		-	2	2
	Implement R&M factors	0.17	0.17	0.17
		2.2	2.2	2.2
	Machine power (KW)	3.4	40	80
	Energy cost (€ KWh^{-1})	0.145	0.496	0.496
	Labour Cost (€ h^{-1})	7.5[1]	15	15

[1] (Lampridi *et al.*, 2019) [2] (Propfe *et al.*, 2012)

For the robotic system basic scenario, the following assumptions were made. The UGV has a total power of 3.4kW and carries a 48V and 70Ah Lithium-Ion Battery. Considering that for prolonging the life-span of the battery it will discharge at a maximum 80% of its capacity before recharging and that the vehicle does not operate always on its maximum power, the UGV has an autonomous operation of maximum 1.34 hours. Additionally, the charging duration is calculated at 3.7 hours, considering a 15 A hourly charging rate. Regarding the assembly and disassembly of the machinery a total of 2 hours is considered. The UGV is equipped with a mechanical row-crop cultivator. It should also be stated that the UGV has the possibility to carry a second battery without any adjustments, doubling the operation time. A worker is considered for the observation of the equipment and the charging and battery replacement processes. The hourly wage is considered as the half of the respective conventional.

For the purposed of the present study the base line scenarios were compared to five alternative scenarios concerning the operational functionality of the robot and both in small- and large-scale applications. More specifically alternative scenarios that increase the efficiency of the robotic system were examined. These alternative scenarios include:

Scenario 1. Installation of a second battery of the same capacity in order to double the operation duration.

Scenario 2. Operation of the UGV's in pairs caring one battery. In this scenario for each UGV operating there is another one charging, waiting to replace the operating one. This scenario reduces the charging idle time from 3.7h to 1.02h.

Scenario 3. Operation of the UGV's in pairs caring two batteries. This scenario is the same as scenario 2 but doubles the operation time of the robot.

Scenario 4. Replacement of the battery in the operating UGV caring one battery. In this scenario the robot caries one battery which is replaced as soon as it discharges in order to avoid the idle charging time. However also in this scenario an idle time of 0.1h is considered for the replacement of the battery.

Scenario 5. Replacement of the battery in the operating UGV caring two batteries. The last scenario is the same as scenario 4 but with the use of two batteries in the UGV. In that case the idle time for replacing the batteries is increased in 0.2h.

4. RESULTS

Figure 2 presents the results for the Small-Scale scenario. Considering the basic scenarios, the robotic system is 5.3 times more expensive than the conventional one. This is attributed to the low efficiency of the robotic system (approximately 25.6%) caused by the small operating time and long charging time. The total operation cost is 99.97 €·ha^{-1} for the robotic system compared to the 18.58 €·ha^{-1} of the conventional one. However, when installing a second battery (Scenario 1) the total cost is reduced to 64.37 €·ha^{-1} and the ratio to 3.46 times, while the efficiency increases to 39.3%. The system's efficiency is further increased (52.18%) when introducing robot operation in pairs with one battery per UGV (Scenario 2), however due to big investment costs the total cost of the system remains 3.50 times higher than the conventional one. The efficiency of the robotic system reaches up to 66% in Scenario 3 with the corresponding cost further reducing to 51.65 €·ha^{-1}. In the last scenarios (4 and 5) the efficiency reaches up to 82.67% and 83.3% respectively, which are very similar to the efficiency of conventional systems. The respective cots are approximately 1.7 times higher than the conventional (31.17 €·ha^{-1} for scenario 4 and 30.92 €·ha^{-1} for scenario 5).

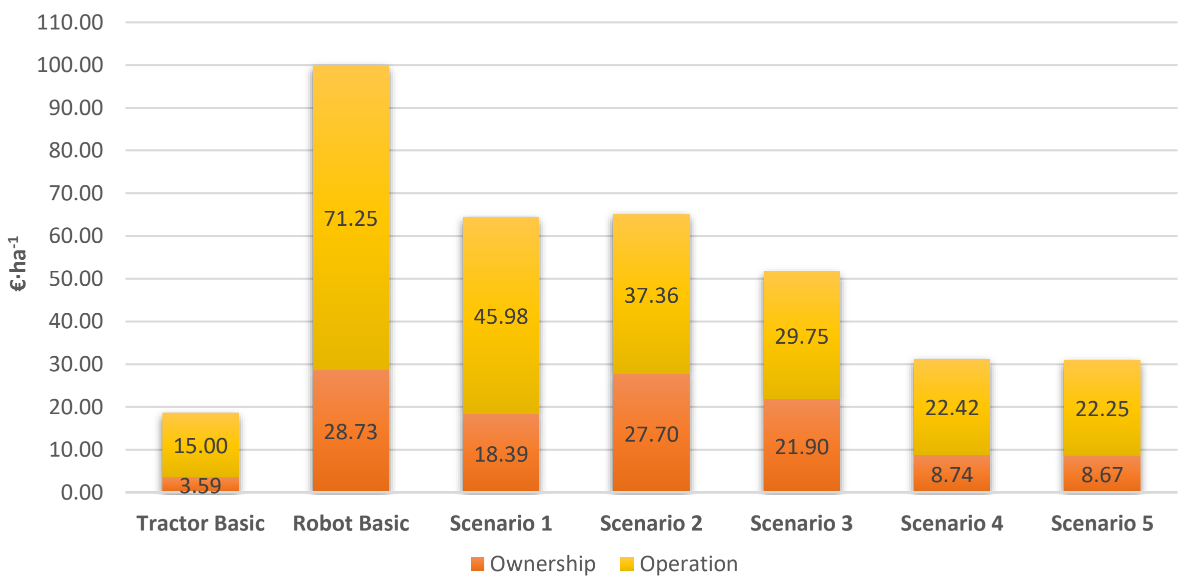


Figure 2. Total Cost – Small Scale Scenario (10ha)

Figure 3 demonstrated the results for the Large-scale scenario. In principle the robotic system is 3.48 times more expensive than the conventional one requiring 7 units to perform the operation in the given time frame, however, this ratio is reduced to 2.40 times when installing a second battery to the UGV (Scenario 1) (the vehicles required are reduced to five). In Scenario 2, eight vehicles are required to perform the operation (considering that four are operating and four are waiting to replace them) demonstrating that the high cost (42.74 €·ha^{-1}) of these robotic systems is a determining factor of their feasibility. Respectively in scenario 3 the cost is reduced to 34.64 €·ha^{-1} due to the increased efficiency of the system (50.8% in Scenario 2 and 67.1% in scenario 3). As was the case for the small-scale application, scenarios 4 and 5 demonstrate the highest efficiency (78.3% and 84.5% respectively) while their cost is only 1.49 and 1.38 times higher than the cost of the conventional system indicating its potential.

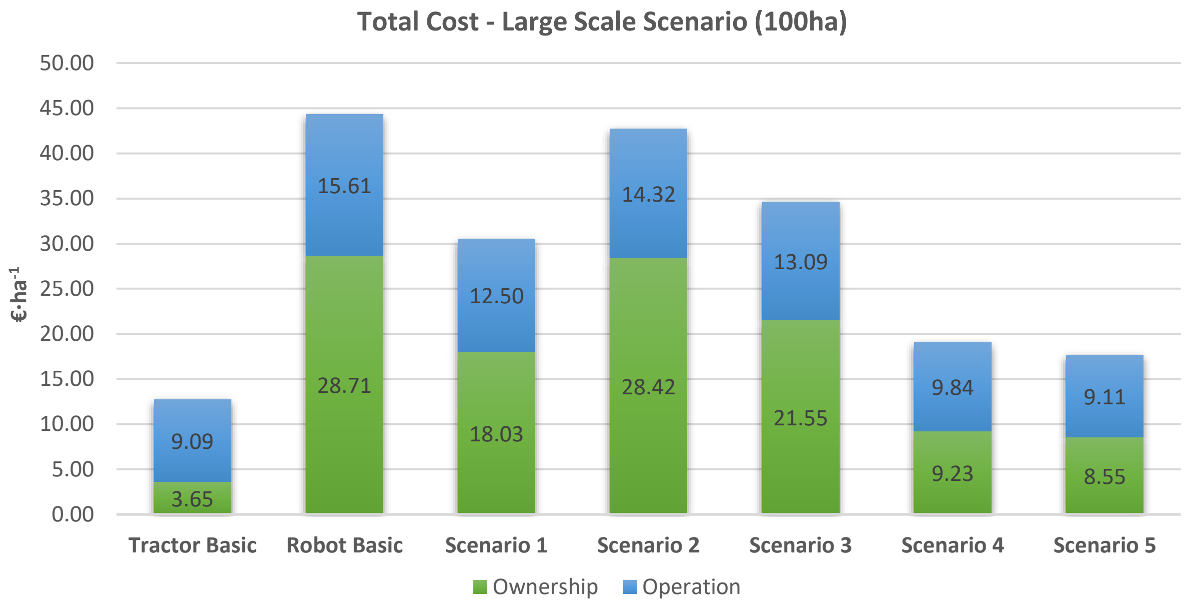


Figure 3. Total Cost – Large Scale Scenario (100ha)

5. Conclusions

The present paper attempted to step a little further towards the examination of the feasibility of adopting autonomous vehicles in conventional agricultural operations. According to the scenarios presented there is a promising potential however, the cost of these systems, which are still not massively produced, is high making them unprofitable compared to the conventional ones. Additionally, there are still areas for further investigation for the accurate cost estimation, as for example with the repair and maintenance cost or the salvage value. Nevertheless, several technical improvements can lead to the increase of their efficiency and their eventual release from the need of an operator thus further reducing operational costs.

Acknowledgements

The work was supported by the project "Research Synergy to address major challenges in the nexus: energy-environment-agricultural production (Food, Water, Materials)" - NEXUS, funded by the Greek Secretariat for Research and Technology (GSRT) – Pr. No. MIS 5002496.

References

American Society of Agricultural and Biological Engineers (2011) 'ASAE D497.7 MAR2011 Agricultural Machinery Management Data', ASABE. doi: 10.1126/science.85.2210.446-a.

Ampatzidis, Y., De Bellis, L. and Luvisi, A. (2017) 'iPathology: Robotic Applications and Management of Plants and Plant Diseases', Sustainability. doi: 10.3390/su9061010.

Bochtis, D., Griepentrog, H. W., Vougioukas, S., Busato, P., Berruto, R. and Zhou, K. (2015) 'Route planning for orchard operations', Computers and Electronics in Agriculture. doi: 10.1016/j.compag.2014.12.024.

Bochtis, D., Sørensen, C. G., Jørgensen, R. N., Nørremark, M., Hameed, I. A. and Swain, K. C. (2011) 'Robotic weed monitoring', Acta Agriculturae Scandinavica Section B: Soil and Plant Science. doi: 10.1080/09064711003796428.

Bochtis, D. D., Sørensen, C. G. C. and Busato, P. (2014) 'Advances in agricultural machinery management: A review', Biosystems Engineering. doi: 10.1016/j.biosystemseng.2014.07.012.

Bongiovanni, R. and Lowenberg-Deboer, J. (2004) 'Precision agriculture and sustainability', Precision Agriculture. doi: 10.1023/B:PRAG.0000040806.39604.aa.

Bubeck, S., Tomaschek, J., & Fahl, U. (2016). 'Perspectives of electric mobility: Total cost of ownership of electric vehicles in Germany'. Transport Policy, 50, pp. 63–77. doi: 10.1016/j.tranpol.2016.05.012

Grimstad, L. and From, P. (2017) 'The Thorvald II Agricultural Robotic System', Robotics, 6(4), p. 24. doi: 10.3390/robotics6040024.

Lampridi, M. G., Kateris, D., Vasileiadis, G., Marinoudi, V., Pearson, S., Sørensen, C. G., Balafoutis, A. and Bochtis, D. (2019) 'A Case-Based Economic Assessment of Robotics Employment in Precision Arable Farming', Agronomy, 9(4), p. 175. doi: 10.3390/agronomy9040175.

Marinoudi, V., Sørensen, C. G., Pearson, S. and Bochtis, D. (2019) 'Robotics and labour in agriculture. A context consideration', Biosystems Engineering. doi: 10.1016/j.biosystemseng.2019.06.013.

Moradi, Morteza, Moradi, Mohammad and Bayat, F. (2018) 'On robot acceptance and adoption: A case study', in 2018 Artificial Intelligence and Robotics, IRANOPEN 2018 and the 8th Conference on Artificial Intelligence and Robotics. Qazvin, Iran, 10, April, 2018: IEEE, pp. 21–25. doi: 10.1109/RIOS.2018.8406626.

Pedersen, S. M., Fountas, S., Have, H. and Blackmore, B. S. (2006) 'Agricultural robots - System analysis and economic feasibility', Precision Agriculture. doi: 10.1007/s11119-006-9014-9.

Propfe, B., Redelbach, M., Santini, D. J. and Friedrich, H. (2012) 'Cost Analysis of Plug-in Hybrid Electric Vehicles including Maintenance & Repair Costs and Resale Values Implementing Agreement on Hybrid and Electric Vehicles', Proceedings of 26th International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium, 5, p. 6862. doi: 10.3390/wevj5040886.

Qureshi, M. O. and Syed, R. S. (2014) 'The impact of robotics on employment and motivation of employees in the service sector, with special reference to health care', Safety and Health at Work. doi: 10.1016/j.shaw.2014.07.003.

Toledo, O. M., Steward, B., Steward, B., Gai, J. and Tang, L. (2014) 'Techno-economic analysis of future precision field robots', in American Society of Agricultural and Biological Engineers Annual International Meeting 2014, ASABE 2014. Montreal, Quebec Canada,13-15, July, 2014.

INTELLIGENT CONVERSATIONAL AGENT INTEGRATION TO A SOCIAL MEDIA PLATFORM FOR CONTROLLING IOT DEVICES IN SMART AGRICULTURE FACILITIES

Eleni Symeonaki[1, 2], Konstantinos Arvanitis[1], Panagiotis Papageorgas[3] and Dimitrios Piromalis[2]

[1]Department of Natural Resources Management and Agricultural Engineering, Agricultural University of Athens, Greece

[2]Department of Industrial Design and Production Engineering, University of West Attica, Greece

[3] Department of Electrical and Electronic Engineering, University of West Attica, Greece

esimeon@uniwa.gr, piromali@uniwa.gr, karvan@aua.gr, ppapag@uniwa.gr

ABSTRACT

The issue of establishing interaction methods among users, applications and systems involved in Smart Agriculture through interfaces which are simple and friendly in end-usage is considered to be essential for achieving the maximum possible penetration of the IoT technologies in this sector, for the benefit of sustainability. Herewith, in this paper an attempt is made to encounter this issue through the involvement of intelligent conversational agents in controlling IoT devices applied to Smart Agriculture facilities, by introducing the idea of developing a chatbot system which is integrated to a messenger application of a popular social media platform in natural language environment. This solution is considered to provide an efficient, effective and user-friendly mean of interaction between the end-users and the IoT devices deployed in agriculture facilities.

Keywords: smart agriculture, Internet of Things, intelligent conversational agents, chatbots, natural language processing.

1. INTRODUCTION

Sustainability is considered nowadays as one of the most imperative targets to be achieved globally in order to cope with the imminent climate change related challenges. This target leads to the stronger involvement of innovative technologies in agricultural facilities in the context of Smart Agriculture (Lipper et al., 2014), enabling accordingly the active participation of the stakeholders so as to improve productivity by maximizing the efficiency inputs and minimizing their environmental impacts.

The technology of the Internet of Things (IoT), which is continuously evolving and maturing, is considered to be a valuable asset in the development of Smart Agriculture through the extensive use of intelligent remote-controlled production equipment such as Wireless Sensor Networks (WSNs) and mobile embedded systems. The most novel solutions regarding IoT agricultural applications tend to adopt ubiquitous interconnectivity methods along with cost-effective cloud services granted by smart mobile devices (Dlodlo and Kalezhi, 2015). In this context, several applications specially designed to run on smartphones, tablets and other mobile devices have been introduced up to present with the purpose to establish interaction methods among the agricultural IoT objects (physical and artificial) accessing and transacting information via a highly distributed public network such as the Internet. Mobile applications specially addressing to the IoT for Smart Agriculture have been lately presented

(Costopoulou, 2016), offering the opportunity to increase yields through modernized production methods with respect to the environment, contributing in this way to the global sustainable growth. However, the adoption and usage of such innovative technology practices in agriculture facilities is still rather limited and fragmentary since, as several studies indicate, only a rather small proportion of agricultural stakeholders takes advantage of the opportunities offered by the consolidation of the IoT with smart mobile devices. This seems to be due to the fact that resistance to change remains an obstacle in agriculture and familiarity to the Information and Communication Technologies (ICT) features continues to be a challenge in rural areas. Since social networking has been recorded as the second largest traffic volume contributor worldwide, with an average share of over 15% of total mobile data traffic, the integration of agricultural mobile applications to social media messaging platforms could be the key for overcoming the barriers of the IoT technologies penetration in agriculture facilities (Lathiya, 2015).

This work attempts to introduce a user-friendly, efficient and secure framework for controlling IoT agricultural devices in natural language dialogues, through the deployment of an Intelligent Conversational Agent (chatbot) and its integration to an instant messaging application of a popular social media platform. On this ground, the conceptual framework focusing on the features of intelligent conversational agents and the benefits of employing chatbot systems as interfaces for IoT agricultural applications are reviewed in brief. Thereafter the architecture, the operating features as well as an overview of a chatbot system capable of controlling a group of IoT devices through the interaction via instant messaging in a popular social media platform is described. Finally, the conclusions deriving from this attempt are presented and some reference on the ongoing research work is made.

2. Conceptual Framework Overview

Spoken dialogue technology refers to the turn-by-turn interaction between humans and intelligent systems in terms of natural language communication ranging from only a small set of words (such as the digits 0–9 and the words yes/no) to large vocabulary dialogues (Dybkjær et al., 2004). At present, due to the progress in language processing and dialogue modelling, there is a broad variety of systems that deploy spoken dialogue technology methods ranging from simple question-answering models which can answer a single question at a time, to sophisticated dialogue systems, which allow extended conversational interaction between end-users and devices (Braun et al., 2017).

Conversational User Interfaces (CUI) are software dialogue systems which facilitate any average user to interact with any device, anywhere and at any time, without the need of special skills or training, by involving a variety of written or oral natural communication forms in order to simulate actual conversations in the end-users' native languages rather than in specific command-line syntax (Schnelle-Walka et al., 2016). Advanced CUIs support the situated language understanding of any probably ambiguous, insufficient or partial multimodal inputs and the deriving of completely correlated outputs, through the implementation of Artificial Intelligence (AI) techniques, using Natural Language Processing (NLP) and Natural Language Understanding (NLU) programs (McTear et al., 2016).

The recently growing need to enhance the extensive access to web services and online information through intelligent, effective, dynamic, flexible, multimodal and user friendly means of Human-Machine Interaction (HMI), through the integration of Machine Learning (ML) techniques (Schnelle-Walka et al., 2016; Stoyanchev et al. 2016) and natural language understanding functions in various services (such as localized search, dialogue management, remote control, the Internet of Things, etc.), are among the most important reasons for the recent impetus of conversational user interfaces in the form of conversational agents (Ferrara et al., 2016).

Chatbots (also known as Chatterbots or simply as bots) are the most prominent conversational agents of current conversational user interfaces. The term "Chatbot" refers to an interactive software dialogue system which enables real-time communication with the end-users by simulating and reproducing turn-by-turn intelligent conversations in natural language via textual (textbots) or even auditory methods (voicebots). A chatbox system consists of three main modules (Braun et al., 2017): Request Interpretation, Response Retrieval and Message Generation. In the context of Request Interpretation, a “request” is not necessarily a question, but can also be any user input, while equally a “response” to this input could be any output statement. The Message Generation follows the classical Natural Language Generation (NLG) pipeline (Reiter and Dale, 2000).

Chatbot systems constitute a technological trend which strongly coordinates with the current IoT concept providing a highly effective and user-friendly interface solution (Kar and Haldar, 2016). From this prospect, there are some significant advantages in employing chatbots as an IoT interface rather than conventional applications for different platforms and versions. Some of the reasons why chatbots are such an appropriate interface solution in the field of IoT are as follows:

- Interaction in natural language as they are capable of creating triggered rules for IoT smart devices so as to activate any action requested by the end-user. Natural language processing algorithms using artificial intelligence features are responsible for unpacking the intent and pass any required instructions to the IoT gateway for processing. Moreover, the chatbot system is empowered through artificial intelligence learning techniques.
- De-parameterized environment of interaction as chatbots do not require the systematic input of all parameters in order to complete a request.
- End-users are not obliged to learn the operation of different IoT applications as they can query IoT networks of devices simply by using their native language, without having to know any interface sequence or command structure. Additionally, end-users do not have to separately download applications as these may be centrally and directly accessible through existing chat clients.
- Refining of end-user requests for subsequent interactions and control reducing in this way the problems of information abstraction which present in conventional interface solutions.
- No demand for constant application updates neither for any maintenance of older versions at the back-end of different operating systems versions or mobile platforms.
- End-users of IoT applications are offered a natural, pleasant and simple interaction environment, which operates on any messenger service platform whether this is mobile, in-app or via web chat.

3. METHODOLOGY

For the materialization of the project, data acquired by a Wireless Sensors and Actuators Network (WSAN) (Piromalis and Arvanitis, 2016) deployed in an agriculture facility were used. These raw data were transmitted via a gateway, using the LoRaWAN communication protocol, to a context-aware middleware cloud component where they were centrally processed and managed. The contextual information deriving from the middleware cloud is then made available to end-users through an AI-based chatbot, which is integrated to an instant messaging application of a popular social media platform (as communication user interface), for providing them with context-aware services related to the monitoring as well as the control of the agriculture facility of interest in question–answer sessions conducted in their native languages.

The AI-based chatbot system was chosen to be integrated to the Facebook Messenger application since it encompasses high penetration to end-users (more than 1.3 billion users worldwide) along with several features, among which open official Application Programming Interfaces (APIs) for chatbot development, WebView objects for increased control over the user interface and advanced configuration options, unimpeded file sharing (text, audio, image, video, etc.) as well as analytics and feedback. What is more the Facebook Messenger application employs Natural Language Processing and Machine Learning algorithms which allow the understanding and extraction of information (entities) out of instant messaging dialogues carried out as natural language conversations.

Although Facebook comes up with 'Wit.ai' as its own chatbot development platform, in order to grant flexible further integration of the chatbot to other popular messenger applications which provide official APIs for this purpose (i.e. Twitter, Viber, Slack, Line, Telegram, etc.), the 'Dialogflow' platform offered by Google [https://dialogflow.com] was selected for its development. 'Dialogflow' (formerly known as Api.ai) is an open-source platform which enables the deployment of chatbots through several features similarly to 'Wit.ai' and on top of that it involves an in-line code editor allowing the performance of various tasks straight from the console. The architecture of the proposed solution is presented in Figure 1.

Figure 1. Chatbot System Architecture.

The basic flow of conversation in 'Dialogflow' involves the aspects of: a) the input given by the end-user, b) the given input parsed by the agent and c) the response returned to the end-user by the agent. In this context, for defining the flow of the conversation, intents that map the end-user input to responses were created in the agent, whereas in each one of these intents, examples of user expressions triggering the intent, actions extracted from each expression and the ways of response, were properly defined. Intents consist of four main components allowing the mapping of the end-user's input to responses: a) intent name, b) training phrases, c) actions and parameters (defining the relativity of information in the expressions) and d) responses. 'Dialogflow' performs the matching of user expressions to intents using the training phrases defined and the significant values, words or phrases specified within them. Figure 2 presents an example of how 'Dialogflow' performs a successful matching of a user's expression to an intent.

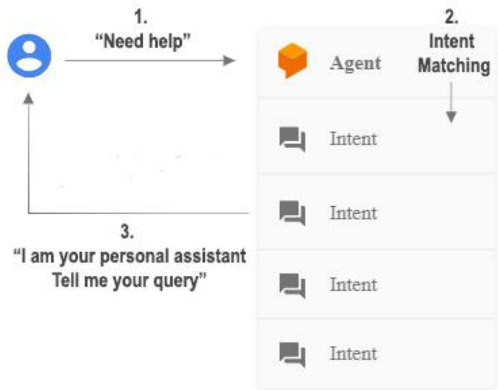


Figure 2. Indicative conversational flow when a successful matching is performed.

Moreover the 'Node.js' platform [https://nodejs.org] was used in addition to the 'Dialogflow', for providing the development of the chatbot with a server-side JavaScript run-time environment so as to create a simple webserver with two webhook (also known as HTTP push API or web callback) endpoints, the first for the initial Facebook verification and the second for being responsible for any other messages from the Facebook Messenger. Finally, in order to integrate this solution into the cloud and ensure its constant execution over the internet, the cloud computing platform services from Microsoft Azure platform [https://azure.microsoft.com] were used for the deployment of the chatbot.

4. Implementation and Results

Based on the methodology previously described, an AI-based chatbot system was developed and integrated to the Facebook Messenger application for monitoring as well as for controlling a WSAN deployed in an agriculture facility. The functionalities developed for the chatbot system, concerning its adaption to different forms of dialogue which achieve the same intents, are introduced below. It has to be noticed that although the dialogues presented in this section are in English, the chatbot system can be easily adapted to several natural languages as it is supported by appropriate NLP and ML algorithms. In Figure 3 some indicative results of the conversations and operations performed in the chatbot system are presented.

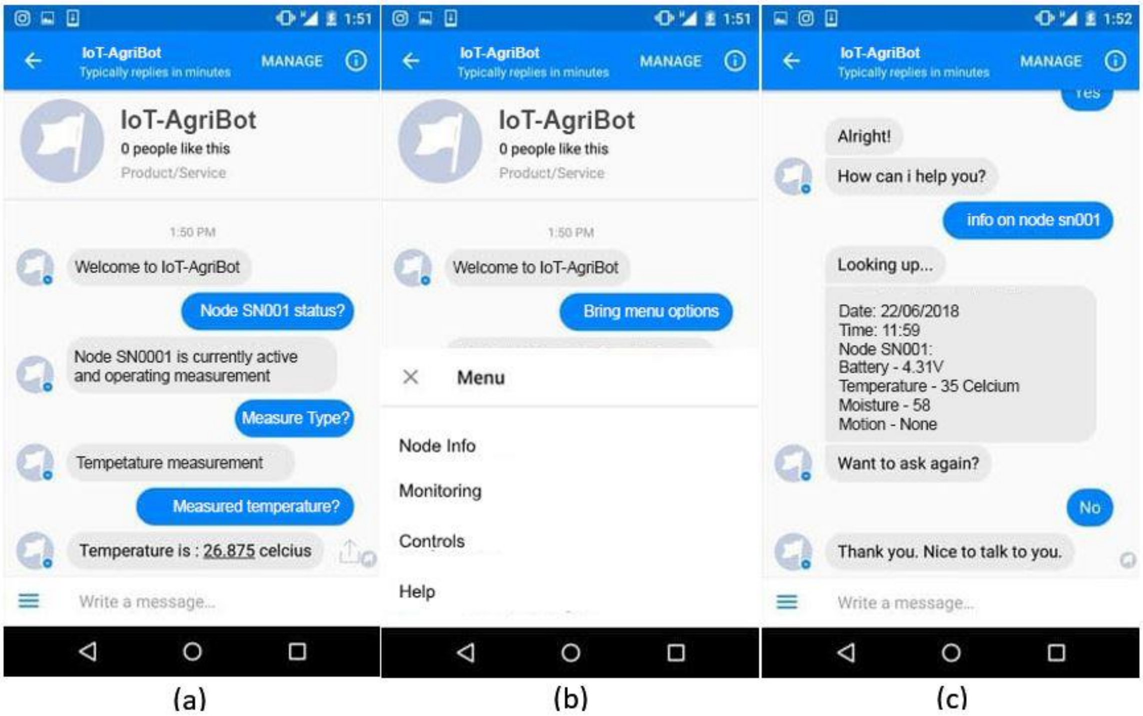


Figure 3. Indicative results of conversations and operations in the chatbot system.

Intent 1: Greeting; a greeting interaction is important in the conversation in order to make the system's usage more accessible and friendly.

Intent 2: Menu; the options menu is offered in order to provide the end-user with a more direct interaction interface with the chatbot and increase its usability.

Intent 3: Help; a help session was developed in the conversation in order to encounter any difficulties of end-users to understand the chatbot functionalities or in case any operation details have to be confirmed. In this context, the chatbot is capable of detecting the users' probable problems or doubts and offer them the help required.

Intent 4: **Monitoring and Control**; this intent concerns the establishment of interaction with the defined entities (IoT devices of the agriculture facility) offering monitoring information about the cultivation and control actions of the equipment.

5. Conclusions and Further Research

The integration of a chatbot system, as an intelligent conversational agent, to the instant messaging application of a popular social media platform is considered to assist the automation of agriculture in a great extent through achieving the maximum possible penetration of the IoT technologies in this sector. Given the observed aspects in this research, it was possible to achieve this objective, presenting a promising solution which provides an efficient, effective and user-friendly mean of interaction between the end-users and the IoT devices deployed in agriculture facilities, based on the Facebook messenger application and the cognitive services of the 'Dialoglow' platform and 'Node.js' technology. As future work, the chatbot system is planned to be also integrated to other popular social media platforms and tests will be performed in order to obtain results regarding its performance and ascertain its efficient operation.

REFERENCES

Braun, D., Mendez, A.H, Matthes, F. and Langen, M. (2017) 'Evaluating natural language understanding services for conversational question answering systems', in proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, Association for Computational Linguistics, Saarbrücken, Germany, pp.174–185

Costopoulou, C., Ntaliani, M. and Karetsos S. (2016) 'Studying Mobile Apps for Agriculture', IOSR Journal of Mobile Computing & Application, 3(6), pp. 44-99

Dlodlo, N. and Kalezhi, J. (2015) 'The internet of things in agriculture for sustainable rural development', in Proceedings of the International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Windhoek, Namibia, pp. 13–18

Dybkjær, L., Bernsen, N.O. and Minker, W. (2004) 'Evaluation and usability of multimodal spoken language dialogue systems' Speech Communication, 43(1-2), pp. 33–54, doi: 10.1016/j.specom.2004.02.001

Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. (2016) 'The Rise of Social Bots', Communications of the ACM, 59(7), pp.96–104, doi: 10.1145/2818717

Kar, R. and Haldar, R. (2016) 'Applying Chatbots to the Internet of Things: Opportunities and Architectural Elements', International Journal of Advanced Computer Science and Applications (IJACSA), 7(11), pp. 147-154, doi:http://dx.doi.org/10.14569/ IJACSA.2016.071119

Lathiya, A., Rathod, A., and Choudhary, K. (2015), 'Role of social media in agriculture' International Journal of Commerce and Business Management, 8, pp. 268-273, doi: 10.15740/HAS/IJCBM/8.2/268-273.

Lipper, L. et al. (2014) 'Climate-Smart Agriculture for Food Security', Nature Climate Change Journal, 4, pp.1068 -1072, doi: https://doi.org/10.1038/nclimate2437

McTear, M., Callejas, Z. and Griol, D. (2016) 'Implementing spoken language understanding', The conversational interface: Talking to Smart Devices, Springer International Publishing, Switzerland, pp. 187–208, doi:10.1007/978-3-319-32967-3_9

Piromalis, D. and Arvanitis, K. (2016) 'SensoTube: A scalable hardware design architecture for wireless sensors and actuators networks nodes in the agricultural domain', Sensors, 16(8), p. 1227, doi: 10.3390/s16081227

Reiter, E. and Dale, R. (2000) 'Building Natural Language Generation Systems', Cambridge University Press, Cambridge, UK, ISBN:0-521-62036-8

Schnelle-Walka, D., Radomski, S, Milde B., Biemann, C. and Muhlhauser, M. (2016) 'NLU vs. Dialog Management: To whom am I speaking?', in Proceedings of the Joint Workshop on Smart Connected and Wearable Things (SCWT'2016), Sonoma, California, USA, doi: https://doi.org/10.13140/RG.2.1.1928.4247

Stoyanchev, S., Lison, P. and Bangalore, S. (2016) 'Rapid prototyping of form driven dialogue Systems using an open-source framework', in Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, Los Angeles, USA, pp. 216–219

PATH PLANNING FOR AGRICULTURAL VEHICLES BASED ON GENERATED PERCEPTION MAPS

Dimitrios Katikaridis[1,2], Vasileios Moysiadis[1,2], Giorgos Vasileiadis[2], Damianos Kalaitzidis[2], Panagiotis Papazisis[2], Dimitrios Vrakas[1], Dionysis Bochtis[2]

[1]School of Informatics, Aristotle university of Thessaloniki, Greece

[2]Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Greece

d.katikaridis@certh.gr, v.moisiadis@certh.gr, g.vasileiadis@certh.gr, damianos.kalaitzidis@hotmail.com, 1panagiotis60@outlook.com, dvrakas@csd.auth.gr, d.bochtis@certh.gr

ABSTRACT

Agricultural field operations, such as spraying, harvesting and seeding, can potentially be automated and executed either by conventional machinery equipped with automation systems, or by autonomous machines. An automated operation must be planned and organized in a manner that, on one hand reduces risk and failure potential and, on the other hand, optimizes productivity and efficiency. However, the diversity of the natural out-door environment and the huge amount of diversified in type data required to picture the operational environment, comprise the hardest challenges for the deployment of fully automated agricultural operations.

In the context of this paper an algorithmic approach was developed aiming at solving one of the various problems encountered in the autonomous agricultural operations. Specifically, the problem addressed is the navigation in the semi-structured environment of orchards. The navigation process consists of seeking a valid path connecting two predefined points in the field enabling the vehicle to travel between them. This is also a core functional component for robotic vehicles in agricultural operations.

The developed software receives as input pre-processed data, a geotagged depiction of an orchard farm, which is obtained by an unmanned aerial vehicle. The pre-processing formats the coordinates which define the field's tracks. Based on this data, the software creates a grid-based map related to the accessible areas, utilized by a graph-based algorithm that produces the topological path planning solution. Subsequently, the solution is translated as a sequence of coordinates which define the produced optimal path.

The software was executed, and its functionality validated in routing applications in an orchard using an autonomous farming vehicle.

Keywords: path planning, UGV, routing, graph based

1. INTRODUCTION

In precision agriculture, an automated operation should be programmed and scheduled resulting an increased productivity and efficiency while risk factors are being avoided. In the other hand, dynamically changed environments, alongside with the required data that describes the respective field, consists in the most serious challenge that must be achieved in order to accomplish a fully automated and unmanned in-field navigation. The exploitation of such technologies requires the review of the traditional techniques that takes place into the field. Oksanen (2007) introduced two path planning algorithms, one is separating a complicated field area to smaller and simpler areas while the other one is using prediction methods. Sørensen et al. (2004) are using a method which compiles a combinational optimization on the planning patterns, based on combined field's, vehicle's and implement's characteristics. In many cases, new solutions provided a more efficient result comparing with the traditional ones (Hameed et al., 2011). As suggested by Bochtis et al. (2012) an optimized vehicle path planning offers a plenty of advantages such as reduced required initial data for an operation, cleaner and healthier production method, decreased environmental impact and a better product maintenance. At recent years, the implementation and optimization of unmanned automated vehicles have triggered the interest about the robotics and path planning (Keicher and Seufert, 2000) . Research on this area has focused on solving problems such as minimizing vehicle maneuvers in the field by reducing its complexity and fragmenting fields into smaller areas to create simple patterns (Jin and Tang, 2006). Still in research, Cariou et al. (2010) is studying to find optimal maneuver patterns when changing a crop line. In this paper, the approach of the problem in relation to the speed and the maneuvers of the vehicle is examined. Then there are various kinds of changes when turning the vehicle. A research which is also remarkable is Jensen et al. (2012) in which it is planned to design a path of unit supporters within a field (within the crop lines, on the headland) as well as outside the field (rural roads, field entrances). At work Ali et al. (2009) a complete approach is developed in which a continuous coverage of crop lines is presented. While the study focuses mainly on harvesting, it can also be applied to feed applications as it covers the case of moving off-road vehicles for landing purposes. The Han (2019) approach is also remarkable. In this work, a path-finding implementation has been carried out in three dimensions. This problem has more challenges than finding a path in the two-dimensional world. In this approach, a method has been developed where a subset of obstacles translates into terms of significance within the path.

In the context of this paper an algorithmic approach was developed aiming at solving one of the various problems encountered in the autonomous agricultural operations. The developed software receives as input pre-processed data, a geotagged depiction of an orchard farm, which is obtained by an unmanned aerial vehicle. Based on this data, the software creates a grid-based map related to the accessible areas, utilized by a graph-based algorithm that produces the topological path planning solution. Subsequently, the solution is translated as a sequence of coordinates which define the produced optimal path.

The software was executed, and its functionality validated in routing applications in an orchard using an autonomous farming vehicle.

2. Methodology

2.1 Analysis and assumptions

In this work, linear crops are examined, with the common feature of the impossibility of moving vehicles perpendicular to the lines. Given this limitation and the inherent structure of the system it is possible to create a topological grid that fully describes the field. To solve the problem, it was necessary to make some assumptions. Initially, it was assumed that there would be no mechanical failure or malfunction in the vehicles. Then, in the context of optimization, the motion of the main and secondary vehicles was calculated simultaneously and cooperatively. In addition, dynamic obstacle avoidance is not part of the path design as this operation is performed using lower level hardware and sensors. Finally, the algorithm can be applied to either convex or non-convex fields, provided that the crop lines are parallel curves.

2.2 Review

The field is presented as a topological grid where there are nodes in it. Each node can represent the following states, Reserved point / obstacle, Free point, Starting point of the path and Endpoint of the path. Each node is associated with four (4) movements within the space: Up, Down, Right, Left.

Also, for every adjacent node to which traffic is allowed, there is a direct connection (**Figure** 1). Required input data for the algorithm is the number of cultivation lines. Under these circumstances, the problem goes back to the problem of finding the shortest path in a grid that joins the original to the final node. In the case where more than one path is found, the elongated distance of each one is calculated, and ultimately it is chosen with the lowest cost.

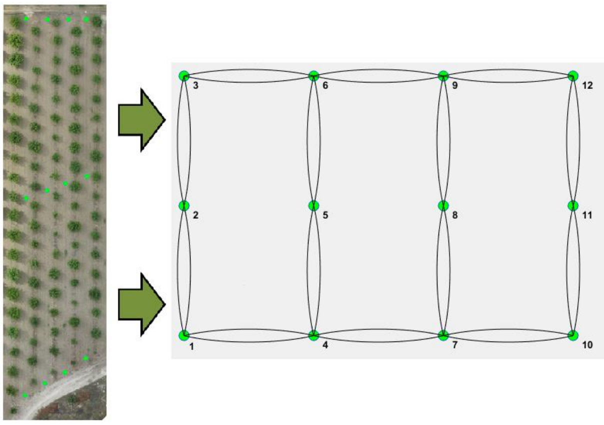


Figure 1. Map to topological grid

2.3 Representative topological grid of the field

A field can be represented as a topological grid contains points that can be characterized by a state (Reserved point / obstacle, free point, starting point, end point).

Reserved point / obstacle: Any actual obstacle or still working vehicle. Considered to be immobilized

Free point: Any point of the topological grid where the vehicle can reach.

Starting point: The starting point of the path
End point: The endpoint of the path

3. IMPLEMENTATION

The implementation of the solution is divided into the following steps: The software reads and edits the data from the UAV. A grid-based map is produced using that data and, in addition, the optimal path is created using the algorithmic approach that is presented is this paper. Furthermore, a KML (Keyhole Markable Language) which describes the produced path, is created. Finally, the KML file is imported to an unmanned ground vehicle in order to be parsed and execute the respective operation (**Figure** 2).

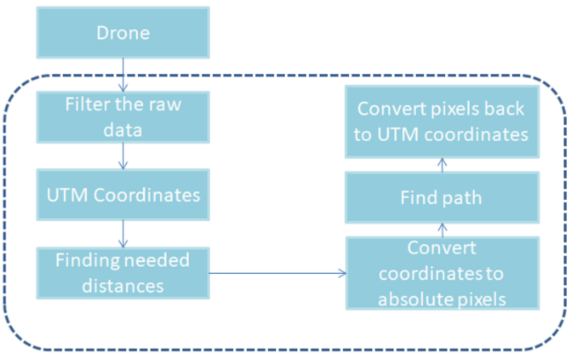


Figure 2. Raw data as algorithm's input

The input data consists of UTM (Universal Transverse Mercator) coordinates that form the crop lines as well as the corresponding pixels of the coordinates on the images (**Figure** 3).

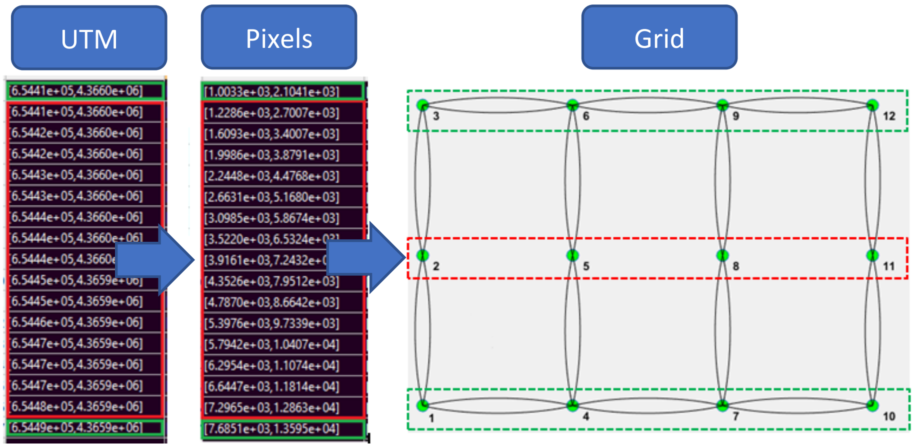


Figure 3. UTM coordinates to topological map

A routine is responsible for the integrity of the data while it cleans out any faulty coordinate or empty entity. After that, 3 class objects are created describing the grid-based map and the lines of the field. The obstacle declaration is done by the user, stating the location of the entity (actual obstacle, parked machinery or implement, working labor). Considering the existing topology map, the nodes belonging to the path are colored in blue while obstacle is colored in red. When all nodes are designed, a dashed blue line joins these points (Figure 4) producing two (2) results. The graph of the nodes and the planned path in the real field.

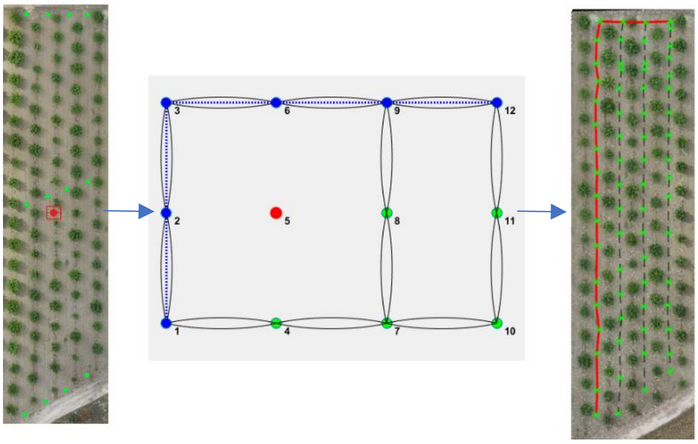


Figure 4. The produced result

4. CASE STUDIES

The algorithm was tested in two (2) different usage scenarios presented below (**Figure** 5 and **Figure** 6). In all cases, the scope was the same field. Differentiation lies in the different points that are placed as the starting point and the end point. The results produced by the software are as follows: Stamping the path to the image of the mosaic, mapping the path to the topological map, create a graph describing the path based on the pointers of its points and create a KML file and present the path using the Google Earth software.

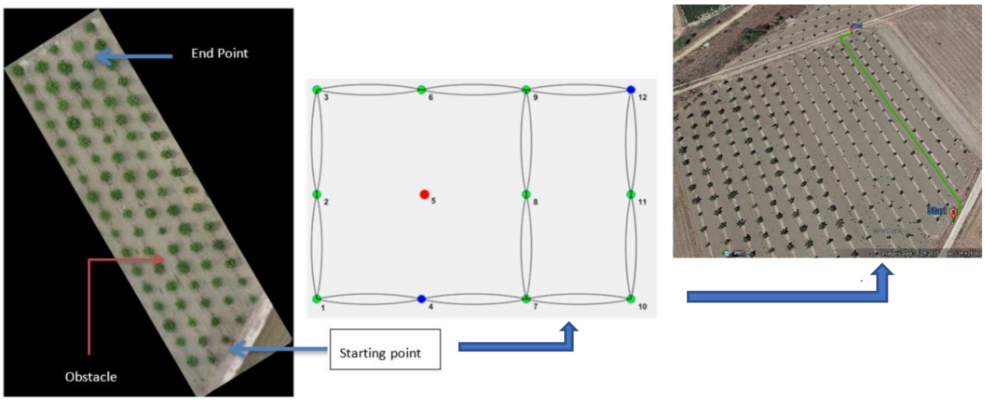


Figure 5. Usecase 1 with 1 obstacle

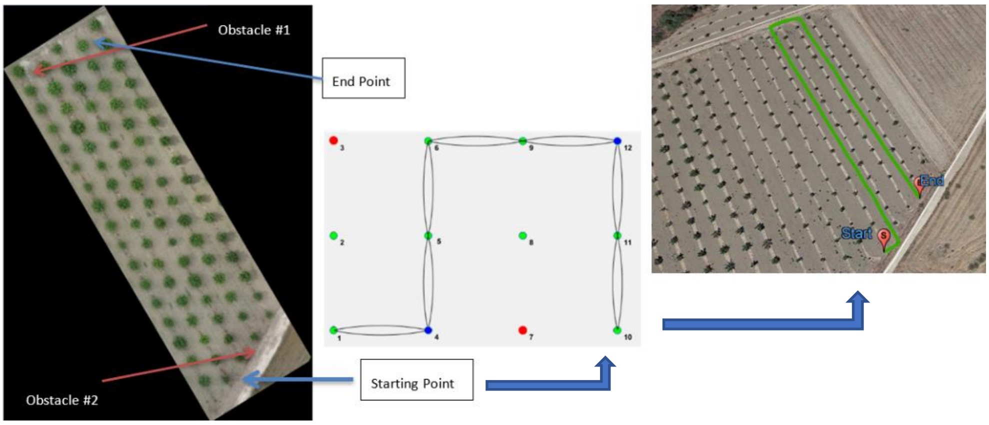


Figure 6. Usecase 2 with 2 obstacles

The produced KML files created by the software can be used for various purposes with different media. Some forms of use include tracing a path with a handheld GPS receiver or importing the file into a robotic vehicle. In this paper, a robotic ground vehicle was used to trace the path that is produced by the software (**Figure** 7). The robotic vehicle fitted a GPS receiver, parses the KML file, and finally, using the integrated computer, the UGV is moving corresponding to the KML's coordinates.

Figure 7. UGV traces the generated path

5. Conclusions and future steps

This work solves a part of the problems that exist when trying to navigate autonomously within the field. The next steps that will be studied increase the expectations for further development of this sector. One goal consists of the software ability to produce within the KML file new entities that will represent the obstacles. Further options and arrangements will be added for the vehicle morphology and obstacles. Appropriate sensors will then be attached to the stand-alone vehicle in order to dynamically avoid obstacles. A laser sensor and a. A depth camera are suitable for applications requiring object recognition when moving the vehicle into space (Fu et al., 2015). Finally, the objective is the interaction of the UAV with the UGV in real time will allow it to solve even more complicated and complicated problems (Sivaneri and Gross, 2018).

Acknowledgements

The work was supported by the project "Research Synergy to address major challenges in the nexus: energy-environment-agricultural production (Food, Water, Materials)" - NEXUS, funded by the Greek Secretariat for Research and Technology (GSRT) – Pr. No. MIS 5002496.

References

Ali, O., Verlinden, B. and Van Oudheusden, D. (2009) ‘Infield logistics planning for crop-harvesting operations’, Engineering Optimization. Taylor & Francis , 41(2), pp. 183–197. doi: 10.1080/03052150802406540.

Bochtis, D. D., Sørensen, C. G. and Green, O. (2012) ‘A DSS for planning of soil-sensitive field operations’, Decision Support Systems. North-Holland, 53(1), pp. 66–75. doi: 10.1016/J.DSS.2011.12.005.

Cariou, C., Lenain, R., Thuilot, B., Humbert, T. and Berducat, M. (2010) ‘Maneuvers automation for agricultural vehicle in headland’, p. 10 p. Available at: https://hal.archives-ouvertes.fr/hal-00523407/ (Accessed: 29 January 2019).

Fu, H., Xu, D., Lin, S. and Liu, J. (2015) ‘Object-based RGBD image co-segmentation with mutex constraint’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 4428–4436. doi: 10.1109/CVPR.2015.7299072.

Han, J. (2019) ‘An efficient approach to 3D path planning’, Information Sciences. Elsevier, 478, pp. 318–330. doi: 10.1016/J.INS.2018.11.045.

I. A. Hameed, I. A., D. D. Bochtis, D. D. and C. G. Sorensen, C. G. (2011) ‘Driving Angle and Track Sequence Optimization for Operational Path Planning Using Genetic Algorithms’, Applied Engineering in Agriculture. American Society of Agricultural and Biological Engineers, 27(6), pp. 1077–1086. doi: 10.13031/2013.40615.

Jensen, M. A. F., Bochtis, D., Sørensen, C. G., Blas, M. R. and Lykkegaard, K. L. (2012) ‘In-field and inter-field path planning for agricultural transport units’, Computers & Industrial Engineering. Pergamon, 63(4), pp. 1054–1061. doi: 10.1016/J.CIE.2012.07.004.

Jin, J. and Tang L. (2006) ‘Optimal path planning for arable farming’, elibrary.asabe.org. Available at: https://elibrary.asabe.org/abstract.asp?aid=20643 (Accessed: 29 January 2019).

Keicher, R. and Seufert, H. (2000) ‘Automatic guidance for agricultural vehicles in Europe’, Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S0168169999000629 (Accessed: 29 January 2019).

Oksanen, T. (2007) *Path planning algorithms for agricultural field machines*. Available at: https://aaltodoc.aalto.fi/handle/123456789/2976 (Accessed: 29 January 2019).

Sivaneri, V. O. and Gross, J. N. (2018) ‘Flight-testing of a cooperative UGV-to-UAV strategy for improved positioning in challenging GNSS environments’, Aerospace Science and Technology. Elsevier Masson, 82–83, pp. 575–582. doi: 10.1016/J.AST.2018.09.035.

Sørensen, C., Bak, T. and Jørgensen, R. N. (2004) ‘Mission planner for agricultural robotics’, Proc. AgEng, (2004/9/12), pp. 894–895. doi: 10.1.1.93.4994.

INTEGRATED INFORMATION SYSTEM FOR ROBOTICS APPLICATIONS IN AGRICULTURE

Ioannis Menexes[1], Vasileios Kolorizos[1], Christos Arvanitis[1], Georgios Banias[1], Dimitrios Kateris[1] and Dionysis Bochtis[1]

[1]Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Greece

i.menexes@certh.gr, v.kolorizos@certh.gr, c.arvanitis@certh.gr, g.banias@certh.gr, d.kateris@certh.gr, d.bochtis@certh.gr

ABSTRACT

In the information era, focusing on the Agriculture modernization and introducing it to the new digital world, the most important elements of which are the online systems and cloud services, should be highly prioritized. Unfortunately it is not considered as vital when compared to the rest of the industry. Practices based on empirical observations and non-optimized methods for tasks, such as weed spraying, fertilizing, yield prediction or plant health monitoring, lead to inefficient resource use and provide undesired results with both economic and environmental negative impact. These are problems that could be alleviated with the help of the latest technological achievements. The developed information system presented is the framework that integrates all the different autonomously operating subsystems, ensuring bidirectional communication among them. State of the art Unmanned ground vehicles are combined with advanced hardware equipment and enable them to navigate autonomously inside fields. Laser-based sensors, that use the LIDAR technology, and Global Navigation Satellite System compatible devices, which provide centimeter-level accuracy, ensure that the robots are fully aware of their surroundings and location in the real world. Simultaneously, unmanned aerial vehicles fly above the work area and collect information that are input to the developed information system. Using the ground robots in the same framework with the unmanned aerial systems creates a network that consistently and reliably feeds the information system with data. This data is analyzed and stored on the cloud, in order to be compiled into applicable information and for future use.

Keywords*:* robotics, agriculture, sensors, UGV, UAV

1. INTRODUCTION

The world as it used to be known, has changed and evolved into an environment of digital information and data, structured within the framework of the internet. Dynamic as it has become, the need to upgrade the nature of its elements, from static to dynamic, is now of utmost importance than ever before. However, although this process has been in progress for quite a while now, not all sectors have received the appropriate attention in order to smoothly transition into the information era. Unfortunately, agriculture belongs to this list of sectors that have been overlooked and thus, have not managed to keep up with the latest technological advancements. As a result, many methods and practices that are used in agriculture are still tied to empirical observations and conclusions that lack efficiency and optimization. Nevertheless, there is great potential for improvements towards that direction, if the power of the latest technological advancements is harnessed effectively and methodically.

Such advances enable the integration of technologies tightly connected to mobile and wireless broadband and cloud computing, to Knowledge based Agriculture (Morgan and Murdoch, 2000) and manage to accomplish the automation of human labor requiring tasks, significantly dropping the final cost of such jobs. Consequently, resource usage is reduced while efficiency in yield production is increased achieving the goal of a much desired paradox balance. This sustainability defined goal, aiming at higher yield demands is thereby accomplished, positively affecting multiple aspects of the bio-economy industry, among which the most important are the environmental and economic ones (Von Wirén-Lehr, 2001). However, the ever growing population applying pressure towards production quantity, in conjunction with the inhibitory factors of soil compaction (Hamza and Anderson, 2005) and climate instability (Mueller et al., 2012), put extra weight on the demands. To make matters worse, a 20 years depth analysis which shows a 0.489% increase in calorific demands per capita annually, according to the Food and Agriculture Organization (FAO) of the United Nations, translated into a total of 10.25% in two decades, becomes a significant variable in the formula that indicates the urgency for further increase in supply to compensate for this demand.

Information Technology provides the infrastructure to create the much needed environment for Information driven Agriculture. Information systems aggregating and processing valuable agriculture related information serve purposes of improving the level of the intelligent management and decision of agricultural production (Sørensen et al., 2010; Yan-E, 2011). Input data to the information system regarding soil quality, field and cultivation state, plant health as well as a rich list of other precious metrics can be reliably and efficiently obtained with the use of automated robotic subsystems (Sorensen et al., 2010) combined with sophisticated hardware (sensors, cameras etc.) (Adamchuk et al., 2004). State of the art ground and aerial vehicles, equipped with equally advanced hardware, can be introduced to Information Agriculture through integrated information systems (Sørensen et al., 2011).

In this paper, an integrated information system collecting data via both Unmanned Ground Vehicles (UGV) and Unmanned Aerial Vehicles (UAV) is proposed. It aims at modernizing and automating some of the aforementioned sectors of agriculture and expects to tackle problems the source of which emanates from methods based on traditional practices and lack of centralized aggregated data.

2. METHODOLOGY

The framework that the UGV and UAV subsystems are integrated into is the presented developed information system. The information system's back-end is supported by the Amazon Web Services (AWS) cloud computing (Amazon, 2016) which offers a suitable platform to accommodate the needs of the information system's infrastructure (Figure 1). High reliability and great up-down scalability potential as well as the "serverless application" feature were the three decisive reasons why the AWS cloud computing was chosen to be the application host. The "serverless application" model removes the "server manual maintenance" variable out of the "human resources management" equation, accelerating the development speed of the application by focusing human labor directly to the front-end side of the application.

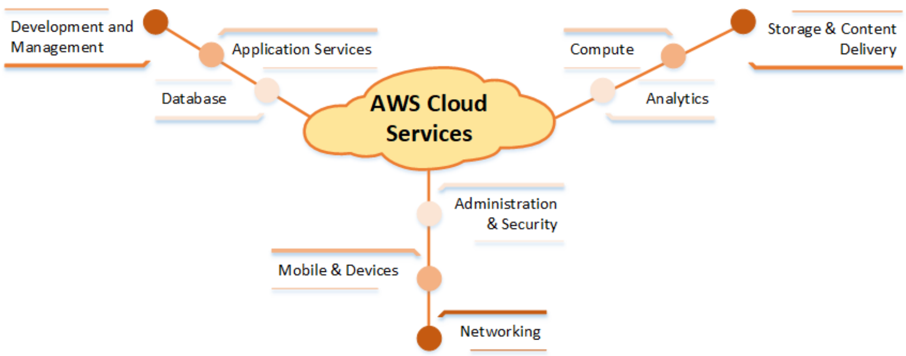


Figure 1. AWS cloud Infrastructure Services

For the front-end system development, it was necessary to use a set of different tools. ReactStrap (Hernandez, Burrell and Sharp, 2018), BootStrap (Otto et al., 2019), HTML and CSS is the combination that gives life to the user-side application by providing the Graphic User Interface (GUI) through which the information system can be accessed by its users. The data exchange between the application's front-end and back-end side is realized with the help of React language (Facebook Open Source, 2017). React is a JavaScript library, suitable for creating user interfaces, which is commonly used for developing webpages and web applications. BootStrap is a webpage style-tool that enables the interface to be easily transferred to a great list of portable devices without breaking the style format of the interface itself. ReactStrap is a tool that enables developers to use BootStrap objects alongside with React, easing the portability of the developed applications to numerous devices such as smartphones, computers and tablets.

UGVs are operated by Robot Operating System (ROS) (Quigley et al., 2009), which is a Linux based framework for writing and developing robot software. ROS has a modular character and a network structure. It consists of a central unit called Rosmaster and a network of nodes. This network is operated by the Rosmaster and has no limitations as for the number of nodes it can accommodate. These nodes are independent from one another, each serving different purposes, executing tasks of different priorities, in parallel. Nodes communicate with each other and all of them are coordinated by the Rosmaster (Figure 2). ROS is commonly used in the industry as a framework to operate robots, or fleets of robots, and it is so popular due to its Open Source nature as well as the large collection of software libraries and tools it offers that can be combined into various applications. Additionally, ROS comes with a large variety of implemented software drivers and eases the process of connecting industry, state of the art, equipment to the robot. Such equipment can be hardware devices i.e. GNSS receivers, laser-scan sensors, RGB/depth cameras, Inertial Measurement Units (IMU) etc. all used to set robots aware of their surroundings and give them real-world perception (Thrun, 2008).

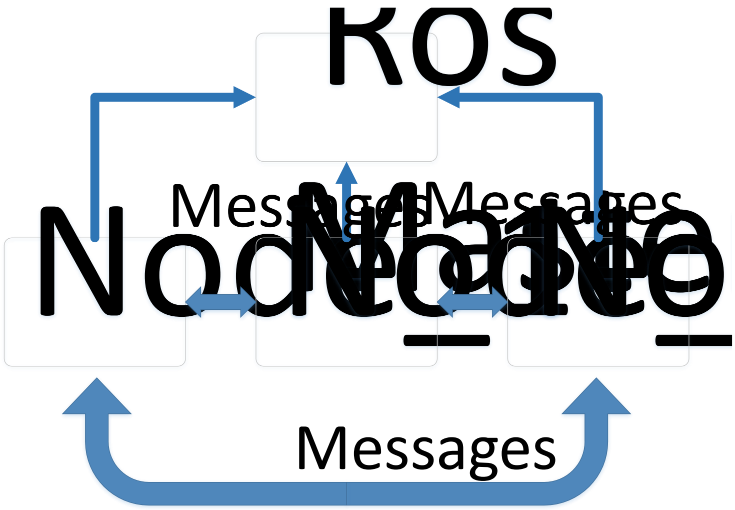


Figure 2. ROS network structure

UGVs equipped with such hardware operate autonomously inside the framework that the developed information system provides (Sørensen and Bochtis, 2010). Tasks can be assigned to the robots and will be executed according to the schedule, as set by the user. The information system constantly logs the current state of each robot, providing feedback that is used to keep track of the robot's general status. Data about robot's location, speed, battery level and health, task progress and uptime are live in order to keep the user updated and able to monitor the robot's current state. UGVs' autonomous navigation is assisted by UAVs that operate simultaneously above the working area, feeding the information system with real-time data about the working area (Ravankar et al., 2018).

The UAVs fly above the field, within which the UGVs operate, and collect photographic material that is used to determine the presence of terrestrial obstacles. These data feeds the information system from where the user can see the live aerial view of the respective field. Route plans are generated by analyzing the incoming data and UGVs are updated accordingly. UAVs assisted navigation sends alerts to the UGVs that indicate the existence of an obstacle at a specific area inside the field. By utilizing the sensing mounted hardware, UGVs can cross-examine the validity of these alerts, update their route plan and avoid areas inside the field that are blocked by obstacles.

As a result, important amount of data is aggregated in the information system, stored on the cloud and later compiled into applicable information. Photographic material acquired by UGVs, operating along their planned route, is suitable for use in numerous applications (**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**).

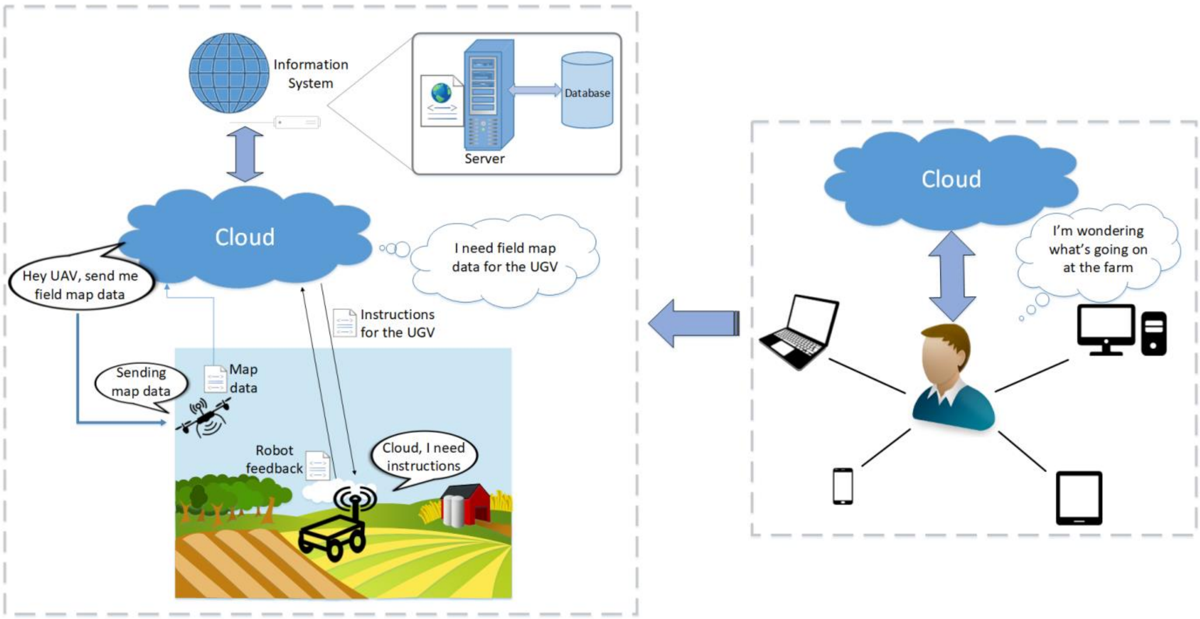


Figure 3. Interaction of information system's entities with the end user

Plant health monitoring becomes a trivial task (Liakos et al., 2018) with reduced human resource cost and greatly increased accuracy (Khirade and Patil, 2015). This leads to better and more reliable yield prediction, which in turn guarantees higher accuracy in economic estimations for the information system users. Weed spraying can now be scheduled for specific areas inside fields where the problem appears to be more intense, contributing in resource waste reduction and therefore, higher efficiency and smaller environmental negative impact (McFadyen et al., 2011).

3. CONCLUSIONS

The presented information system integrates unmanned ground and aerial vehicles in order to support the functions of a fleet of independent data collectors with further possibilities of generation-assignment and transmission of tasks to be executed. The automated interaction of UGVs and UAVs with the information system, targets at creating a sustainable network of constant data flow through the information system entities. Yield prediction is going to be supported by large amounts of analyzed photographic material and will no longer be subject to uncertain estimations. Plant health monitoring aims at informing the user for potential in-field disease-related hazards as well as reduce the error margin for diagnosis of common plant diseases. Finally, methodical tracking of fields for weed growth is expected to provide guidance for precision-spraying with the use of weed growth-zones information carrying heatmaps, generated by post-analyzing cloud stored data. Here, a brief concept of the system has been presented. Parts of the presented information system are already being implemented and the development status is in constant progress. The results derived from the so far use of the information system are greatly encouraging.

ACKNOWLEDGEMENTS

The work was supported by the project "Research Synergy to address major challenges in the nexus: energy-environment-agricultural production (Food, Water, Materials)" - NEXUS, funded by the Greek Secretariat for Research and Technology (GSRT) – Pr. No. MIS 5002496.

REFERENCES

Adamchuk, V. I., Hummel, J. W., Morgan, M. T. and Upadhyaya, S. K. (2004) 'On-the-go soil sensors for precision agriculture', Computers and Electronics in Agriculture, 44(1), pp. 71–91. doi: 10.1016/j.compag.2004.03.002.

Amazon (2016) Amazon Web Services (AWS)- Cloud Computing Services, Amazon Web Services, Inc. Available at: https://aws.amazon.com/ (Accessed: 13 July 2019).

Facebook Open Source (2017) JavaScript library for building user interfaces, Facebook Inc. Available at: https://reactjs.org/ (Accessed: 12 July 2019).

Hamza, M. A. and Anderson, W. K. (2005) 'Soil compaction in cropping systems: A review of the nature, causes and possible solutions', Soil and Tillage Research. doi: 10.1016/j.still.2004.08.009.

Hernandez, E., Burrell, C. and Sharp, E. (2018) ReactStrap - React Bootstrap 4 components. Available at: https://reactstrap.github.io/ (Accessed: 12 July 2019).

Khirade, S. D. and Patil, A. B. (2015) 'Plant Disease Detection Using Image Processing', in 2015 International Conference on Computing Communication Control and Automation. IEEE, pp. 768–771. doi: 10.1109/ICCUBEA.2015.153.

Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D. (2018) 'Machine Learning in Agriculture: A Review', Sensors. Multidisciplinary Digital Publishing Institute, 18(8), p. 2674. doi: 10.3390/s18082674.

McFadyen, R., Adkins, S., Walker, S., Lemerle, D., South Wales Leslie Weston, N., South Wales Sandy Lloyd, N., Australia, W., Panetta, D., Shane Campbell, Q., Wayne Vogler, Q., Simon Brooks, Q., Wu, H. and South Wales, N. (2011) 23rd Asian-Pacific Weed Science Society Conference The Sebel Cairns. Available at: http://www.apwss.org/documents/23rd APWSS Conference Proceedings Vol 2.pdf#page=108.

Morgan, K. and Murdoch, J. (2000) 'Organic vs. conventional agriculture: Knowledge, power and innovation in the food chain', Geoforum. doi: 10.1016/S0016-7185(99)00029-9.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N. and Foley, J. A. (2012) 'Closing yield gaps through nutrient and water management', Nature. doi: 10.1038/nature11420.

Otto, M., Thornton, J., Rebert, C., Lauke, P., Mazovetskiy, G., Galante, A. and Cuppens, M. (2019) Bootstrap - The most popular HTML, CSS, and JS library in the world. Available at: https://getbootstrap.com/ (Accessed: 12 July 2019).

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R. and Ng, A. (2009) 'ROS: an open-source Robot Operating System', Icra, 3(Figure 1), p. 5. doi: http://www.willowgarage.com/papers/ros-open-source-robot-operating-system.

Ravankar, Ankit, Ravankar, Abhijeet, Kobayashi, Y. and Emaru, T. (2018) 'Autonomous Mapping and Exploration with Unmanned Aerial Vehicles Using Low Cost Sensors', Proceedings, 4(1), p. 44. doi: 10.3390/ecsa-5-05753.

Sorensen, C. G., Pesonen, L., Fountas, S., Suomi, P., Bochtis, D., Bildsøe, P. and Pedersen, S. M. (2010) 'A user-centric approach for information modelling in arable farming', Computers and Electronics in Agriculture. Elsevier, 73(1), pp. 44–55. doi: 10.1016/J.COMPAG.2010.04.003.

Sørensen, C. G., Fountas, S., Nash, E., Pesonen, L., Bochtis, D., Pedersen, S. M., Basso, B. and Blackmore, S. B. (2010) 'Conceptual model of a future farm management information system', Computers and Electronics in Agriculture. Elsevier, 72(1), pp. 37–47. doi: 10.1016/J.COMPAG.2010.02.003.

Sørensen, C. G., Pesonen, L., Bochtis, D. D., Vougioukas, S. G. and Suomi, P. (2011) 'Functional requirements for a future farm management information system', Computers and Electronics in Agriculture. Elsevier, 76(2), pp. 266–276. doi: 10.1016/J.COMPAG.2011.02.005.

Sørensen, C. G. and Bochtis, D. D. (2010) 'Conceptual model of fleet management in agriculture', Biosystems Engineering. Academic Press, 105(1), pp. 41–50. doi: 10.1016/J.BIOSYSTEMSENG.2009.09.009.

Thrun, S. (2008) 'Simultaneous localization and mapping', Springer Tracts in Advanced Robotics. doi: 10.1007/978-3-540-75388-9_3.

Von Wirén-Lehr, S. (2001) 'Sustainability in agriculture - An evaluation of principal goal-oriented concepts to close the gap between theory and practice', Agriculture, Ecosystems and Environment. doi: 10.1016/S0167-8809(00)00197-3.

Yan-E, D. (2011) 'Design of intelligent agriculture management information system based on IoT', in Proceedings - 4th International Conference on Intelligent Computation Technology and Automation, ICICTA 2011. IEEE, pp. 1045–1049. doi: 10.1109/ICICTA.2011.262.

MAPPING AGRICULTURAL AREAS USING AN AUTOMATED UNMANNED AERIAL VEHICLE

Vasileios Moysiadis[1], Dimitrios Katikaridis[1], Ioannis Menexes[1], Dimitrios Vrakas[2], Aristotelis C. Tagarakis[1], Dionysis Bochtis[1]

[1]Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Greece

[2]Aristotle University of Thessaloniki

v.moisiadis@certh.gr, d.katikaridis@certh.gr, i.menexes@certh.gr, dvrakas@csd.auth.gr, a.tagarakis@certh.gr, d.bochtis@certh.gr

ABSTRACT

Mapping procedure consist an essential operation towards fully automated navigation within any operational environment. This paper focuses on the development of a system, as a first approach, for the one-way communication between an unmanned aerial vehicle (UAV) and an unmanned ground vehicle (UGV) for the detection and registration of operational environment entities and the extraction of the geographical coordinates as a basis information for the subsequent automated navigation of the ground vehicle. The case study took place in a commercial orchard installed with walnut trees located in Central Greece. As a mandatory scenario for the tree's identification, was characterized the hypothesis of distinct, arranged circle like trees' vegetation within the field. The developed system consist an inaugural process towards fully automated operations in agricultural fields. In addition, it contributes to the two-way communication between autonomous vehicles (UAV-UGV) focusing on collaborative predictions and actions for both pre-planned and real-time planning processes.

Keywords: UAV, UGV, Precision Agriculture, Mapping, trees identification, path extraction.

1. INTRODUCTION

The rapid development of technological advances, in recent decades, led to the implementation of innovative systems into production processes. Supervisory control systems have been developed and implemented in various application areas (Colomina and Molina, 2014; Quintin et al., 2017). With the incorporation of the above in agricultural production, the concepts of «precision agriculture» and «agriculture 4.0» have been introduced (Weltzien, 2016; Ozdogan, Gacar and Aktas, 2017; Rose and Chilvers, 2018). In contrast to traditional agriculture, precision agriculture constitutes a comprehensive concept which aims at better management of the fields applying the right amount at the right place and time. The use of autonomous robotic systems is intended to assist in processing operations which are vital for the growing season (Pedersen et al., 2006; Mousazadeh, 2013; Bechar and Vigneault, 2016). To further explore, the use unmanned aerial vehicles (UAV) and unmanned ground vehicles (UGV) is documented to be a promising approach with a main objective of mapping field and crop properties. In tandem with computational efficiency and state of the art sensors, the implementation of the high-resolution cameras in UAV systems facilitates the incorporation of pattern recognition algorithms.

Previous studies on the identification of crop rows used pattern recognition methods. Initially the mosaic was transformed into gray-scale, to prevent the shading effect, and then the Hough transformation was implemented (Jiang *et al.*, 2016; Zhang *et al.,* 2008). Numerous methodologies were developed incorporating machine learning algorithms. Yang *et al.*, (2009) developed a framework which utilizes the Adaboost algorithm (Freund and Schapire, 1997) to identify trees from aerial images. Guerrero *et al.*, (2012) followed a different approach for the trees identification problem with the use of Support Vector Machine algorithm (SVM) combined with Otsu method (Otsu, 1979) to transform RGB grayscale images in order to distinguish the areas of interest (crop rows).

More recently, the potential benefits of the point cloud data exploitation were analyzed. (Torres-Sánchez et al., 2018) proved that the use of point cloud data potentially offers satisfactory results. Taking advantage of the altitudinal variations in an orchard in conjunction with OBIA (Object Based Image Analysis) algorithm (Blaschke, 2010) produced satisfactory results in recognizing trees from aerial images.

This paper focuses on the development of an autonomous system that uses the rapid mapping capabilities of unmanned aerial vehicle (UAV) for mapping agricultural large-scale areas and defines the path for the navigation of unmanned ground vehicles (UGV). In particular, the aim of this work is to identify and export areas (with accurate geographical coordinates) that denote collision free navigation for UGV within an agricultural environment (orchard). The challenge in this methodology is to correctly identify the trees in the outline of the field to properly export the required paths that the UGV should follow.

2. Methodology

2.1 Experimental Procedure

The experiment took place in 2018 in a commercial walnut orchard located in Rizomilos, Magnesia, Central Greece. The orchard was flat and there were no significant geomorphological variations. A DJI S1000+ octacopter (SZ DJI Technology Co., Ltd., Shenzhen, China) equipped with Pixhawk 2.1 autopilot (Proficnc®), Here+ GPS (Proficnc®) and Sony Cyber-shot RX100 III digital camera (SONY, Minato, Tokyo) was utilized to execute automated flights over the orchard. Initial test flights occurred in order to ensure the quality of the data and the proper execution of the experimental procedure. After the fine tuning, the landscape front lap and portrait side lap were set to 75% and the relative altitude was set to 40m providing ground resolution of 0.9 cm/px.

The unmanned aerial system (UAS) took pictures from the field using the survey technique with the QGroundControl. According to this technique, the spatial placement of specific distance points (based on the overlapping percentage) took place across the field. The collection of data required for the development of the algorithms was carried out at two different stages of the walnuts development, during August (walnut filling stage) when the canopy was fully developed and November (post-harvest) when the leaves were brown and the trees were defoliating.

2.2 Data Preprocessing

As a first step to the pre-processing procedure is the image geotagging. To that end, a linux shell script algorithm was developed to extract the UAV's position for each point of interest from the UAV's log file. Subsequently, all the geographical coordinates were assigned to the aggregated images. The Pix4Dmapper software suite was utilized for creating two-dimensional and three-dimensional field representation models.

The essential function of the algorithm was to identify and export points of interest within the field that denote the free movement of a UGV. For this purpose, a georeferenced ortho-mosaic was required. As a result, the geographical coordinates of both the orchard trees and the aforementioned areas were exported. In addition, a two-dimensional gray scale map was produced, which contained the reference points of a relative coordinate system for defining the UGV distance range. The produced map contributes to the autonomous navigation of the UGV using the Robot Operating System (ROS).

The proposed algorithm consisted of three distinct steps; a) mosaic's color transformation, b) trees Identification and c) path extraction.

In orchard fields, the emergence of vegetation of non-cultivated plants (weeds etc.) is quite common. This results in the occurrence of noise in the data. To address the phenomenon, a mosaic RGB transformation function was developed to the following color spaces to eliminate this noise and distinguish the trees from the weeds and the shadow effect. Three color transofmations were examined; HSV, YCbCr, and CIE L * a * b *. The HSV transformation provided better results compared to the other two transformations. In particular, using this transofmation the shading phenomenon and vegetation areas are distinguishable by the color code assigned to them. The YCbCr transformation provided better results for the images acquired in August, however, the results for the November images were quite poor. Therefore the HSV was selected for further analysis.

As a next step, a proper color filter was selected in order to disentangle the cultivated trees. Two basic color categories dominate the orchards' transformation in the color space. The areas of interest were isolated from the rest of the orchard. This process was performed on a function that takes an RGB image as an input, converts it into HSV, isolates the areas of interest, and applies the areas of lower interest to the original photograph.

The final step of mosaic's color transformation was to delete the noise and transform the mosaic into grayscale. A function was modified and incorporated for the elimination of noise and gray scale transformation leading to the final mosaic.

The Circular Hough Transform (CHT) technique was used to identify the orchard trees. A function was developed to calculate the CHT, taking into account previous work (Illingworth and Kittler, 1987; Manzanera et al., 2016). The shape of the final mosaic consists of grayscale pixels and non-compact circles (circle-like points). To identify circles with various R values, the iterative method was used to determine the radius of each potential cycle. To ensure proper identification of trees in environments with increased noise, a function was developed that recognizes the noise in the identified clusters based on their area. The area of the clusters followed normal distribution, with the area of the actual trees belonging to the interval [μ-σ, μ + σ] where μ is the mean value of the areas and σ is their standard deviation. Therefore, any area smaller than the domain was defined as noise and was excluded (Figure 1). For the images taken in November, the noise level was considerably higher due to the trees defoliation and the decolorization of the leaves.

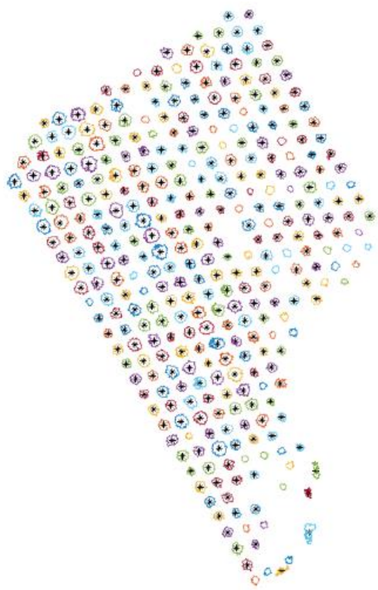


Figure 1. Identified trees, location and canopy shape, after the application of the noise depletion function in the ortho-mosaic.

Subsequent to the coordinates' transformation was the creation of virtual trees. Hence, a function was developed that takes into account the trees' arrangement in order to create trees in places where trees should exist (Figure 2).

For the path extraction, the combination of virtual and identified trees was used. To define the path, the mean distance between the tree rows was used. To properly classify the identified trees into their rows, an alignment was required.

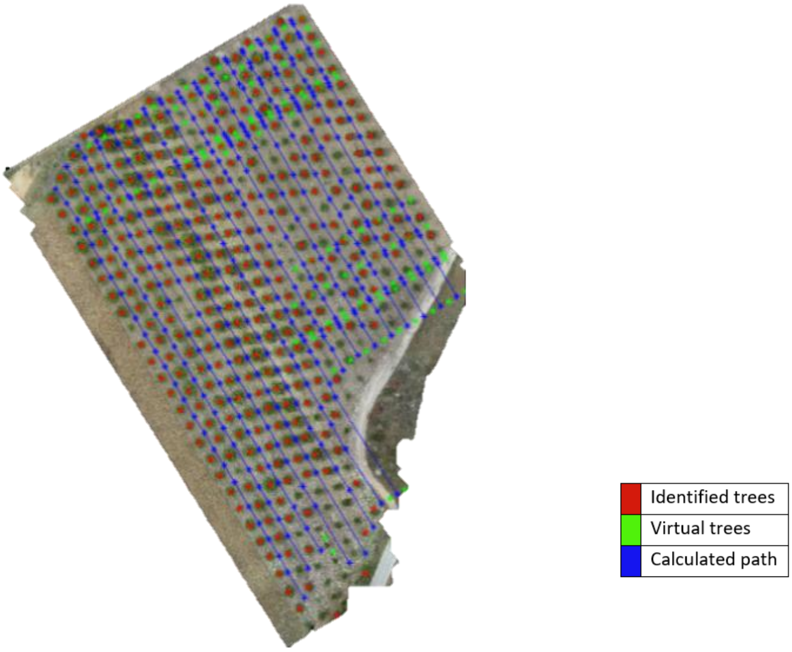


Figure 2. Path extraction using the ortho-mosaic from the walnut orchard in the study.

3. VALIDATION AND RESULTS

The algorithm was validated using images from areas of the field with increased noise due to the occurrence of developed weeds in a significant proportion of the orchard's surface. The system performed well recognizing the vast majority of the trees (Figure 3).

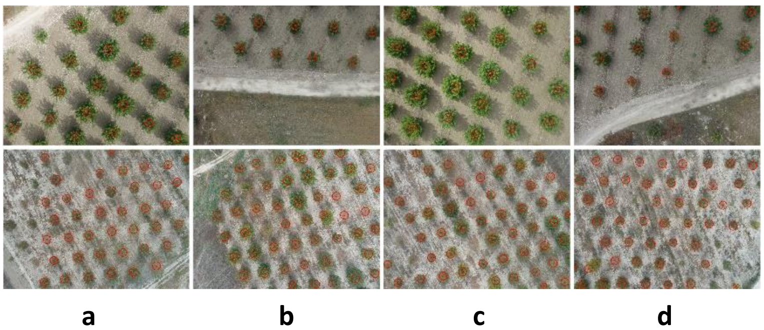


Figure 3. Trees identification validation for the aerial images acquired in August (top) and November (bottom) in four selected locations within the orchard with increased noise.

The actual location of the trees was manually recorded for ground truthing. The results from the developed image analysis procedure were compared with the manually recorded tree locations. The tree recognition rates were satisfactory, varying from 88.8% to 100% for the images taken in August. The recognition rate for the November data set was considerably lower, ranging from 78.4 to 94.1%, due to the partial defoliation of the trees, the color change of the leaves (turning brown) and the increased noise (Table 1).

Table 1. Accuracy of trees recognition using the developed image analysis algorithm for images taken in August and November.

ID	August	November
a	90.9	78.43
b	100	94.11
c	88.8	89.83
d	100	82.75

The path generated form the developed methodology for image analysis was used by the UGV to navigate in the orchard. The UGV followed the path without facing any issues and managed to avoid all the obstacles in real field environment. This was an indication that the UAV-UGV synergy is feasible and the two platforms can exchange data and information efficiently. The study is on-going and further tests in real field conditions will take place to confirm this outcome.

5. Conclusions

The nature of the operational environment in orchards is governed by unpredictable factors that introduce noise and contribute to the alteration of the requested data. The weather conditions combined with the cultivation practices and the constantly changing environment within the orchards during the growing season, bring changes in the tree phenology and the morphology of the soil surface. all these factors introduce different levels of noise in the images acquired form the orchard depending on the acquisition timing.

The development of a pattern recognition algorithm produced satisfactory results while minimizing computational costs since it required relatively small datasets. In addition, it allowed easy modification of the algorithm for various case studies. A major disadvantage of this method was the response to images with a lot of noise.

UAVs consist a rapid means for field scouting and monitoring making them an excellent tool for draw alternative routes for obstacle avoidance by the UGVs in the field. This type of UAV-UGV synergy is promising and may increase the efficiency of autonomous vehicles for in-field operations.

Acknowledgments

The work was supported by the project "Research Synergy to address major challenges in the nexus: energy-environment-agricultural production (Food, Water, Materials)" - NEXUS, funded by the Greek Secretariat for Research and Technology (GSRT) – Pr. No. MIS 5002496.

References

Bechar, A. and Vigneault, C. (2016) 'Agricultural robots for field operations: Concepts and components', Biosystems Engineering. Academic Press, 149, pp. 94–111. doi: 10.1016/J.BIOSYSTEMSENG.2016.06.014.

Blaschke, T. (2010) 'Object based image analysis for remote sensing', ISPRS Journal of Photogrammetry and Remote Sensing. Elsevier, 65(1), pp. 2–16. doi: 10.1016/J.ISPRSJPRS.2009.06.004.

Colomina, I. and Molina, P. (2014) 'Unmanned aerial systems for photogrammetry and remote sensing: A review', ISPRS Journal of Photogrammetry and Remote Sensing. Elsevier, 92, pp. 79–97. doi: 10.1016/J.ISPRSJPRS.2014.02.013.

Freund, Y. and Schapire, R. E. (1997) 'A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting', Journal of Computer and System Sciences. Academic Press, 55(1), pp. 119–139. doi: 10.1006/JCSS.1997.1504.

Guerrero, J. M., Pajares, G., Montalvo, M., Romeo, J. and Guijarro, M. (2012) 'Support Vector Machines for crop/weeds identification in maize fields', Expert Systems with Applications. Pergamon, 39(12), pp. 11149–11155. doi: 10.1016/J.ESWA.2012.03.040.

H. Zhang, H., B. Chen, B. and L. Zhang, L. (2008) 'Detection Algorithm for Crop Multi-Centerlines Based on Machine Vision', Transactions of the ASABE. American Society of Agricultural and Biological Engineers, 51(3), pp. 1089–1097. doi: 10.13031/2013.24511.

Illingworth, J. and Kittler, J. (1987) 'The Adaptive Hough Transform', IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9(5), pp. 690–698. doi: 10.1109/TPAMI.1987.4767964.

Jiang, G., Wang, X., Wang, Z. and Liu, H. (2016) 'Wheat rows detection at the early growth stage based on Hough transform and vanishing point', Computers and Electronics in Agriculture. Elsevier, 123, pp. 211–223. doi: 10.1016/J.COMPAG.2016.02.002.

Manzanera, A., Nguyen, T. P. and Xu, X. (2016) 'Line and circle detection using dense one-to-one Hough transforms on greyscale images', EURASIP Journal on Image and Video Processing. Springer International Publishing, 2016(1), p. 46. doi: 10.1186/s13640-016-0149-y.

Mousazadeh, H. (2013) 'A technical review on navigation systems of agricultural autonomous off-road vehicles', Journal of Terramechanics. Pergamon, 50(3), pp. 211–232. doi: 10.1016/J.JTERRA.2013.03.004.

Otsu, N. (1979) 'A Threshold Selection Method from Gray-Level Histograms', IEEE Transactions on Systems, Man, and Cybernetics, 9(1), pp. 62–66. doi: 10.1109/TSMC.1979.4310076.

Ozdogan, B., Gacar, A. and Aktas, H. (2017) 'DIGITAL AGRICULTURE PRACTICES IN THE CONTEXT OF AGRICULTURE 4.0', Journal of Economics, Finance and Accounting-JEFA, 4(2), pp. 184–191. doi: 10.17261/Pressacademia.2017.448.

Pedersen, S. M., Fountas, S., Have, H. and Blackmore, B. S. (2006) 'Agricultural robots—system analysis and economic feasibility', Precision Agriculture. Kluwer Academic Publishers-Plenum Publishers, 7(4), pp. 295–308. doi: 10.1007/s11119-006-9014-9.

Quintin, F., Iovino, S., Savvaris, A. and Tsourdos, A. (2017) 'Use of Co-operative UAVs to Support/Augment UGV Situational Awareness and/or Inter-Vehicle Communications', IFAC-PapersOnLine. Elsevier, 50(1), pp. 8037–8044. doi: 10.1016/J.IFACOL.2017.08.1229.

Rose, D. C. and Chilvers, J. (2018) 'Agriculture 4.0: Broadening Responsible Innovation in an Era of Smart Farming', Frontiers in Sustainable Food Systems. Frontiers, 2, p. 87. doi: 10.3389/fsufs.2018.00087.

Torres-Sánchez, J., de Castro, A. I., Peña, J. M., Jiménez-Brenes, F. M., Arquero, O., Lovera, M. and López-Granados, F. (2018) 'Mapping the 3D structure of almond trees using UAV acquired photogrammetric point clouds and object-based image analysis', Biosystems Engineering. Academic Press, 176, pp. 172–184. doi: 10.1016/J.BIOSYSTEMSENG.2018.10.018.

Weltzien, C. (2016) 'Digital agriculture-or why agriculture 4.0 still offers only modest returns'. doi: 10.15150/lt.2015.3123.

Yang, L., Wu, X., Praun, E. and Ma, X. (2009) 'Tree detection from aerial imagery', in Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 131–137.

A CONCEPT NOTE ON A ROBOTIC SYSTEM FOR ORCHARDS ESTABLISHMENT

Giorgos Vasileiadis[1,2], Dimitrios Katikaridis[1], Vasileios Moysiadis[1], Dimitrios Kateris[1] and Dionysis Bochtis[1]

[1]Institute for Bio-economy and Agri-technology, Centre for Research and Technology Hellas, Greece

[2]Faculty of Agriculture, Forestry and Natural Environment, School of Agriculture, Aristotle University of Thessaloniki, Greece

g.vasileiadis@certh.gr, d.katikaridis@certh.gr, v.moisiadis@certh.gr, d.kateris@certh.gr, d.bochtis@certh.gr

ABSTRACT

In the dawn of the Agriculture 4.0 era orchard establishment remains a manual and empirically executed task even though it affects the performance of the agricultural holding throughout its service life. The typical workflow involves a significant amount of manual labour, requiring teams of workers operating within a narrow window of availability. The narrow application time window is a major hindrance that is intensified by the decline of workforce availability in the developed countries. The work presented aims to alleviate these factors by introducing contemporary technological advancements. Furthermore, these features will be integrated in a fully autonomous framework. The compatibility and seamless integration of subsystems is insured by software, hardware and communication protocol choices. The system will be designed on available and widely used technologies, taking advantage of the open source community and contributing back to it the applied developed solutions. A central pillar of development will be the Robotic Operating System (ROS), ensuring compatibility and expandability of software and hardware components. Ubiquitous sensors can be used to control the system and add an increased level of unmanned operational capacity. On the mechanical design aspect of the system the framework for product development will be based on the functional design paradigm and will be grounded by the Quality Function Deployment framework. The resulting system is going to be designed to fit an array of Unmanned Ground Vehicles (UGV) and standardized conventional machinery. The operation is divided in stages and developed subsystems are going to be part of a modular architecture. Therefore, the system is flexible and adaptable to the diversified needs of spatially dispersed locations. Positive results are expected both on the pure operational performance, due to the automated nature of the system and on the future operations planning, due to the precise georeferenced planting maps.

Keywords*:* orchard, establishment, robot, UGV, planting

1. INTRODUCTION

Agriculture has supported world growth since the begging of time. To achieve this, cultivating techniques, cultivars and tools evolved and fueled further advances in the agricultural sector and influenced heavily all other sectors. As world population growth is expected to increase substantially (Gerland et al., 2014) and exert pressure on agri-food production, environmental change impacts negatively yield capacity and land availability, sustainability isn't achievable with conventional agricultural practices. Precision agriculture (PA) offers a framework of resource handling that promises rationalized and optimized use of resources (McBratney et al., 2005). This enables farming with reduction of ineffective inputs preserving resources, e.g. fresh water, and simultaneously increases yield capacities and product quality (Blackmore, Godwin and Fountas, 2003). To implement a PA framework use of tools like Information and Communication Technology (ICT), automatization and robotic vehicles can be used to achieve the accuracy level of operations prescribed by PA (Sørensen et al., 2010).

Robots, ICT and automation in agriculture are expanding their field of applicability rapidly, ranging from retro-fitting existing conventional equipment with auto-steering systems and Real-Time Kinematic (RTK) Global Navigation Satellite System (GNSS) to disruptive development of fully robotic platforms capable of executing tasks in the field. Blackmore et al. (2009) suggested a classification scheme for task-oriented robots in arable farming. Numerous robotic systems have been introduced in recent years, some of which aimed specifically for orchard cultivation. Especially for surveying work that can be executed both by UGVs and Unmanned Aerial Vehicles (UAVs) with the use of relatively inexpensive sensors (Pravakar et al., 2015). Several other developed systems are deployed for agricultural tasks (Bergerman et al., 2013; Ball et al., 2015) also in a very active field of development and research. The need for fully automized, planned, controlled, measurable and documented operations is strong, even though farmers tend to perceive them as an inconvenience rather than solution. To abate this perception, systems need to be able to operate autonomously and with a comprehensive to the farmer workflow.

Establishment is a corner stone operation for a multiyear orchard cultivation and can benefit significantly from the introduction of ICT and automatization. Furthermore, the ability to alleviate consequences of mistakes in this step is notably limited, if not impossible, underlining the need for documented and controlled operations. In this work we expand the integration of Unmanned Ground Vehicles (UGV) in orchard cultivation through the proposed robotic system that is to be developed to serve the needs of orchard establishment.

2. System Requirements

Orchard establishment is a labor-intensive task that usually involves numerous workers. The farm manager has to decide on planting pattern and parameters depending on cultivar and field. After this preparatory stage, typical workflow includes staking the field, hole drilling and transplanting the young tree.

In order to operate in this environment, the proposed system dictates the use of a carrying vehicle. A critical requirement is the ICT compliant platform. This is a prerequisite for a fully automated and remote-controlled operation of both the vehicle and the proposed earth drill. Furthermore, the vehicle needs to carry the equipment and provide a rigid support structure to fulfill the mission.

Even though GNSS boards are an integrated feature for almost all modern machinery, the way that this module interfaces with the rest of the system is a system requirement for this concept. Direct access to localization data through popular protocols is needed for the mission planner to be able to utilize it and provide the positioning accuracy needed for the operation. Accuracy needs to be at centimeter level in order to achieve an increased level of operation.

Autonomy is another requirement that needs to be considered by designing a system that is both energy and operational efficient. Field operations demand high levels of autonomy in order for systems to be considered as economically feasible and in many cases to utilize the limited time window available for the operation.

Drilling depth is a function that influences design of the physical system heavily. Apart from the physical dimensions of the auger, the energy consumed is also dependent on the type of soil that is to be processed. Hence, various soil types need to be considered from easy to process loamy soils to compacted clay soils.

3. System Design

Evaluating the above requirements, the architecture proposed is analyzed in the following paragraphs. The resulting system is a compilation of elements that aim to simplify construction as well as utilize technologies that are at a high readiness level and therefore applicable out of the shelf. The proposed system is illustrated in **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**.

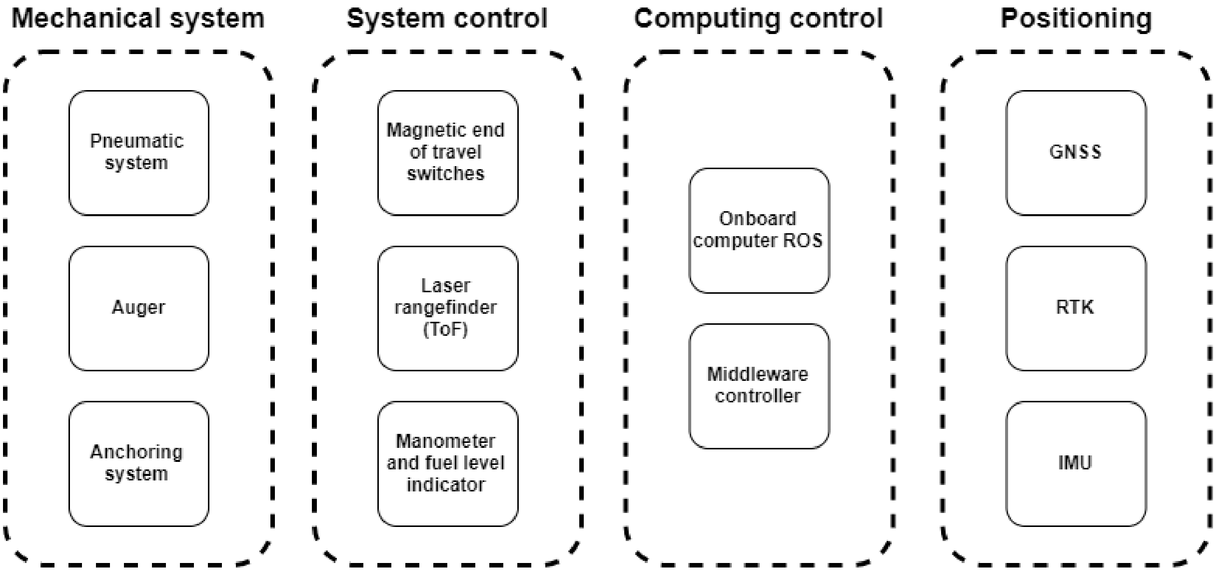


Figure 1. System Architecture

3.1 Architecture

The proposed architecture is based on the Thorvald autonomous robotic vehicle. This platform has physical features that simplify the design and integration process significantly, namely the hollow in the middle frame and its ability to adjust the width. The first enables the design of a compact and rigid system with a footprint similar to that of a UGV without a fully functional implement engaged. Additionally, the smaller footprint allows turns in less space, easier transport to and from the field. The latter is important in future scaling of operations to multi-agent systems executing the task simultaneously. Having an adjustable platform from the conceptualization stage is that the implement can be optimally designed with less restrictions.

As mentioned above, the system needs to navigate and localize with high precision in open fields. The technological solutions for this task are a GNSS board with RTK corrections to increase accuracy. Furthermore, an Inertial Movement Unit (IMU) is used to monitor heading and contribute to accuracy of movement.

The core task is going to be executed using a modified petrol-powered auger. The use of a petrol-powered auger offers great advantages in terms of autonomy levels. The Internal Combustion (IC) engine is going to provide sufficient power to the auger and ensure that the electrically powered UGV isn’t going to be deprived of power resources needed for movement to application points. The second power source is also going to be used to drive the air compressor module that is needed to alter the height of the implement. This will increase its efficiency as the IC will essentially be engaged in either drilling or air compressing work, reducing significantly idle time.

To achieve stability the system is going to exploit the kinematic features of the platform and when the auger is in operation the wheels will turn 45° degrees to immobilize the vehicle and counteract to some extend the torque generated by the soil displacement. To further ensure stability during the operation, a retractable dedicated system is designed to anchor the robot during drilling and be disengaged when not needed.

The physical components need to be coordinated and monitored. This role is served by the connecting framework for all these functions, the Robotic Operational System (ROS) that controls the UGV, the GNSS and communicates with the middleware controller that connects the sensors and monitors the implement. This can be a single board computer or a programmable microcontroller that can read incoming signals, process them at a basic level onboard and output serially the status of the system.

3.2 Mechanical system

In more detail the auger is going to be dimensioned according to experimental data that are going to be collected to determine the optimal design parameters for these operational conditions. Flute angle, nose angle, number of flutes and other parameters are to be determined according to the force and consequently the torque that they are going to exert on the structure. A preliminary version of this assembly can be seen in **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** and Figure 3.

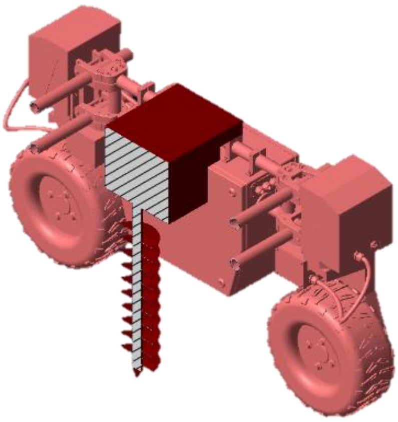


Figure 2. Section view of auger on UGV

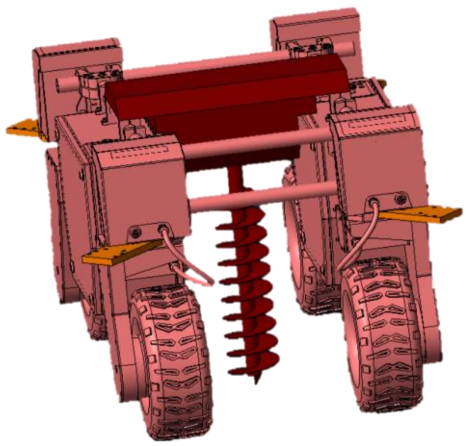


Figure 3. Auger integration

The abovementioned procedure is going to determine the number of anchor points that the system needs to achieve stability during operation. The anchoring system is going to be retractable and utilize a preloaded subframe that aids in traction between the system and the soil.

The operation of the system relies on components being retracted and deployed as needed. The pneumatic system providing this functionality is going to use compressed air linear actuators. The system will use energy generated by the IC engine that drives the auger. This contributes to system autonomy notably since compressed air is one of the most energy consuming features of the concept.

3.3 System control

The moving subsystems need to be controlled and monitored by the middleware. Robustness is a primary factor to be considered, since the operating environment is adverse. Furthermore, the components need to be as simple and globally available as possible in order to be serviceable and replaceable. In this light, magnetic switches for reaching end of travel control are chosen. As the operating principle isn't relying on contact some misalignment and contamination levels are allowed.

To accurately measure travel position of the auger, an optical sensor is chosen that is based on time of flight measurements of a laser diode. Since this is also a non-contact sensor with no moving parts, it can be considered robust enough for operation in the field.

The compressed air tank manometer is also going to be monitored, as well as the fuel level for the IC engine using a simple fuel tank float switch.

3.4 Computing control

The controlling sensors described in Section 3.3 are to be connected to a programmable micro controller, that will be embedded in the implement. This primary control level is going to aid in reporting operational status and monitor the systems task progress. The need to have a middleware controller is also dictated by the requirement for the system to be transferable and easily integrated to other vehicles. Furthermore, lower level control can be dealt with internally, standardizing the equipment's features and procedures and allowing 3rd party integration or development.

The main computing resource is the onboard computer of the UGV that handles primarily all the functions of the vehicle. Since the UGV is running ROS the route planning of the operation is handled by ROS planners and communication with the middleware can be achieved by serial protocol. Consequently, there is a need to develop software that passes commands to the implement and vice versa, interprets the data transmitted by the middleware.

4. Discussion

The proposed concept is designed to automatize a labor-intensive procedure. To achieve this various component are utilized to assemble a novel implement that can be integrated in autonomous robotic vehicles. The system relies on two power sources, the batteries in the UGV and the petrol tank of the IC used for the auger and the air-compressor. Future design iterations should focus on a fully electrical version. In particular, the experimental procedure of defining the optimal auger geometrical parameters will also provide valuable input as for the real work and power requirements. These can be used to calculate a feasible autonomy level that fulfils the operational requirements of orchard establishing. Last, even though the system is task-specific designed, minor modifications can enhance its functionality to similar field work. e.g. soil sampling could also be executed using the proposed system, automatizing a tedious and expensive procedure, enabling farmers to include valuable soil analysis data in their decision processes.

References

Ball, D., Ross, P., English, A., Patten, T., Upcroft, B., Fitch, R., Sukkarieh, S., Wyeth, G. and Corke, P. (2015) 'Robotics for Sustainable Broad-Acre Agriculture', in. Springer, Cham, pp. 439–453. doi: 10.1007/978-3-319-07488-7_30.

Bergerman, M., Billingsley, J., Reid, J. F. and Deere, J. (2013) 'IEEE Robotics and Automation Society Technical Committee on Agricultural Robotics and Automation [TC Spotlight]'. doi: 10.1109/MRA.2013.2255513.

Blackmore, B. S., Fountas, S., Gemtos, T. A. and Griepentrog, H. W. (2009) 'A specification for an autonomous crop production mechanization system', in Acta Horticulturae, pp. 201–216. doi: 10.17660/ActaHortic.2009.824.23.

Blackmore, S., Godwin, R. J. and Fountas, S. (2003) 'The analysis of spatial and temporal trends in yield map data over six years', Biosystems Engineering. Academic Press, 84(4), pp. 455–466. doi: 10.1016/S1537-5110(03)00038-2.

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B. K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G. K. and Wilmoth, J. (2014) 'World population stabilization unlikely this century', Science. American Association for the Advancement of Science, 346(6206), pp. 234–237. doi: 10.1126/science.1257469.

McBratney, A., Whelan, B., Ancev, T. and Bouma, J. (2005) 'Future Directions of Precision Agriculture', Precision Agriculture. Kluwer Academic Publishers, 6(1), pp. 7–23. doi: 10.1007/s11119-005-0681-8.

Pravakar, R., Stefas, N., Peng, C., Bayram, H., Tokekar, P. and Isler, V. (2015) Robotic Surveying of Apple Orchards. doi: 10.13140/RG.2.2.28507.00806.

Sørensen, C. G., Fountas, S., Nash, E., Pesonen, L., Bochtis, D., Pedersen, S. M., Basso, B. and Blackmore, S. B. (2010) 'Conceptual model of a future farm management information system', Computers and Electronics in Agriculture, 72(1), pp. 37–47. doi: 10.1016/j.compag.2010.02.003.

FARM-TO-FORK TRACEABILITY: BLOCKCHAIN MEETS AGRI-FOOD SUPPLY CHAIN

Magdalena Stefanova[1], Michail Salampasis[2]

[1] Sofia University, Faculty of Mathematics and Informatics, Department of Computing Systems

[2] International Hellenic University, Department of Informatics & Electronics Engineering

stefanova.magdalena@gmail.com, mikesalampasis@gmail.com

ABSTRACT

The transfer of goods along the agri-food supply chain currently lacks sufficient visibility. Starting from providers of raw materials to farmers, to food processors, to distributors, to retail outlets and finally to the consumer the ability to trace the history, application or location of any food substance and entity has become an important priority. Consumer food decisions are not only affected by taste and price. Consumers are also concerned about transparency, traceability and social impact. They also expect all participants in the agri-food supply chain to have effective practices in place that allow for the rapid identification, location, and withdrawal of food lots when problems are suspected or confirmed. On top of consumer expectations, there are new regulations and standards that require an improved evaluation of foreign suppliers, fully documented processes and an assessment of all vulnerabilities in the supply chain. These social, economic and legislative demands as well as the fact that most existing traceability solutions are usually proprietary IT systems, make it difficult to develop a scalable, universal and cost-effective traceability system. Emerging technologies, such as blockchain could transform supply chain traceability as we know it and bring more transparency through the value chain, creating value to stakeholders. From a technology perspective, the proposed solution leverages blockchain to keep track of the flow of physical goods. This paper introduces FoodBlock, a theoretical, 'farm-to-fork' traceability solution, which integrates Hyperledger blockchain, mobile apps and GS1 identification standards. We have created a framework for building a minimal viable prototype by using existing technologies.

Keywords: Blockchain, Hyperledger, Food Traceability, Smart Contracts, Smart Agriculture

1. INTRODUCTION

Due to the opaque nature of the agri-food supply chain, regulators, as well as customers have been unable to determine with accuracy the food quality and safety. The endpoints of the food supply-chain - farmer and consumer are pressured to trust a system without having visibility into the processes of the intermediaries. Consumers, despite their growing concerns, are still in the dark on which products are sustainable and have verified high quality. Today, the vast majority of traditional logistic information systems in Agriculture and Food (Agri-Food) supply chains merely track and store orders and deliveries, without providing effects such as transparency, traceability and auditability (Stefanova, 2019).

Traceability is a general term referring to the completeness of information about every step in a process chain (Jansen-Vullers et al., 2003). From a food supply chain perspective, a more rigorous definition defines traceability 'as the ability to trace and follow a food, feed, food-producing animal or substance intended to be, or expected to be incorporated into a food or feed through all stages of production, processing and distribution' (European Community Regulation, 187/2002). In the food industry traceability has been a focal point for several years (Bechini, et al., 2005). The needs for food traceability can be summarized as food safety, quality verifications, and legal and market requirements (Wang & Li, 2006). A traceability system can be useful in the cases of recalls of affected products, improved quality controls, increased consumer confidence in food products and production optimization (Salampasis, et al., 2012).

Currently, there is a large variety of traceability mechanisms used in food supply chains. Some older traceability schemes are paper based while more recent are obviously IT based. Solutions based on IT are usually proprietary systems which are developed ad-hoc from food businesses to respond to particular demands on food traceability originating from EU legislation or demands from large retailers. These systems typically respond to needs within a single operator in the supply chain. The diversity of the systems used among different operators makes their integration difficult. Also, current traceability systems address traceability of food products in such a way which usually does not allow an integrated and transparent retrieval of traceability information along the whole supply chain (Cao et al, 2010). Each link in the supply chain requires keeping information about preceding and succeeding links in local databases, without making it possible to get a complete view of traceability recorded information.

Without a doubt there is a significant technology development which call towards the development of new generation traceability systems. This new technology push is Blockchain (Underwood, 2016). Blockchain can transform supply chain traceability as we know it and could bring more transparency through the value chain, creating value to stakeholders both upstream and downstream. In this paper we discuss a theoretical framework and we provide practical insights on how Blockchain could be used for traceability solutions in a farm-to-fork setting.

2. PROPOSED SOLUTION

The food supply chain is one of the most complex and fragmented of all supply chains. The production is found all over the world, both on land and in water, making many producers and intermediaries difficult to identify and track. This brings uncertainty to all stakeholders in the production chain. Examples of problems that have been difficult to solve with current technologies include preventing fraud and counterfeiting. These issues can have serious domino-effects on public health and the environment and increase financial costs of unnecessary recalls of food products.

Using blockchain to keep track of the entire food supply chain, we have a technology and an infrastructure for secure and trustworthy data storage for all traceability data involved that can also be presented to the end consumer. Tracing and controlling food are possible without the blockchain, however, the implementation of Blockchain technology brings three major inherit benefits to the table:

- The data cannot be manipulated.
- The supply chain can keep traceability and control secure without all participants disclosing all their customers and suppliers' data to a central party. The level of privacy to enforce can be decided by the participants in the system/network.
- The blockchain creates trust in low cost IT solutions. You can use email, Word, mobile phones, and still be sure data is accurate. This allows for example rural farmers and independent truck drivers to seemingly integrate with a larger traceability system.

2.1. Key criteria for food-traceability blockchain

Blockchain networks can come in many variations. The first step before implementing a specific solution is to map the needs and key criteria that a blockchain should meet in order to serve the food-traceability domain. Below we summarize the key requirements of a traceability application. The design and development of our framework is mainly inspired and driven from these requirements.

SPEED AND LATENCY. The food-traceability blockchain is tasked with mapping the digital world to the physical one, which means transactions should be reflected in the chain with the same speed, which is usually required to exchange the food physically. Consequently, the time interval needs to be short enough for convenience and relevance. An interval between 30 seconds and 1 minute is optimal.

REPUTATION AND GOVERNANCE. Considering the latest scandals in the crypto space (The Chain Media, 2018) and the relative conservatism of food chain participants to technology crazes (Tamirat, et al., 2017), a network with clear reputation and preferably already successfully realized industry prototypes should be selected. Businesses need sufficient governance to run blockchain effectively over the long-term.

SCALABILITY. It should be anticipated that the number of participants in a food-traceability blockchain could grow from a handful of enthusiasts to millions of daily users. Still, the blockchain could be limited to a certain agricultural sector, even bounded to a given country, but the capacity of the system to grow and manage increased demand should be there, because its value is directly proportional to its size, also known as network effect.

TYPE OF BLOCKCHAIN. There are three core types of blockchains: public, permissioned and private. Firstly, public networks are large and decentralized, anyone can participate within them at any level. They tend to be more secure and immutable then private or permissioned networks. However, they are often slower, more expensive to use and have limited storage capacity. Within the food-traceability use case, the participants' roles and relative number, except for that of end consumers, are known in advance. To some extent, the participants can be regarded as trusted parties, who demand fast and easy-to-use blockchain network. As a result, the inherent properties of public blockchains are not suitable for the food transparency scenario. Secondly, private networks are shared between trusted parties and may not be viewable to the public. They're very fast and may have no latency. However, the food-traceability ledger must be readable for the general public so that transparency is guaranteed. Finally, permissioned networks are viewable to the public, but participation is controlled (Investopedia, 2018). Their properties such as scalability, low latency and lower cost to build applications on top of them, make them a more optimal choice for the food-traceability scenario.

2.2 Why hyperledger?

There are possibly many suitable blockchain networks for food-traceability, but the investment by many top-players in Hyperledger as well as the resulting quality and robustness, make it a preferable choice compared to volatile networks. Hyperledger (Hyperledger, 2018) is under the guardianship of the Linux Foundation. Hyperledger has more than 230 organizations as members—from Airbus to VMware. The platform offers the benefits of high reputation as well as flexibility and robustness thanks to the successful open-source model. Its projects Hyperledger Fabric or Sawtooth meet all the requirements for food- traceability blockchain listed above and add additional advantages, such as no vendor lock-in, lower total cost of ownership, access to source code, out-of-the box templates, etc.

2.3 Technology

The proposed solution leverages blockchain to enable the traceability of agri-food products. Given the task of providing supply chain system that would connect many small farms and the performance challenges of public blockchains, the need for expensive infrastructure is minimized by employing a permissioned blockchain such as the one provided by the Hyperledger effort. Since the roles of the participants in the agri-food supply chain are known and can be mapped, a permissioned blockchain model is chosen for increased security. This means a person needs to meet certain requirements to perform certain actions on the blockchain. For example, a retailer can verify food status only of lots, which are already known to be purchased by him, so that he will not interfere with the food status of his competitor's inventory. Alternatively, a registered vegetable producer is not allowed to produce data about milk products, etc.

PROTOTYPE: SMART CONTRACTS ON HYPERLEDGER COMPOSER

Hyperledger Composer is a set of tools that can be used to build a blockchain business network on top of Hyperledger Fabric. Hyperledger Composer is best used for the creation of development versions and proof-of-concepts. It makes it simple and fast to create smart contracts, since it is built with the widespread JavaScipt language and supports lots of modern tools such as node.js and popular editors. An example smart contract, which is called chaincode in the Hyperledger world, is a check for expiry date of eggs –egg batches, which were recorded by farmers 27 days ago or more, are automatically rejected for processing on the blockchain. Hyperledger Composer enables easy alignment between business requirements and technical development. However, after the PoC phase, going for Hyperledger Fabric native development is advised.

FOOD TRACEABILITY ON HYPERLEDGER FABRIC

Hyperledger Fabric serves as a decentralized ledger or more simply -distributed database. Hyperledger Fabric is a permissioned blockchain network that can be set up and owned by the organizations forming a consortium. In the agrifood case that could be the farming associations, big producers and retailers, interested in providing transparency and traceability to the end customers. These organizations are called members. Each member is responsible to setup their network of peers. Peers would include raw material providers, farmers, distributors and processors. All peers should be configured with appropriate cryptographic material including certificates ensuring their identity. Peers can use different clients to create transaction invocation requests. A client can be any specific mobile app, web application or organization portal. Client applications use either Hyperledger Fabric SDK or REST web service to interact with the Hyperledger Fabric network. All the peers maintain the ledger they are subscribed to; hence we have Distributed Ledger Technology (DLT). But unlike most other blockchain networks, in Hyperledger Fabric peers have different roles – Endorser peer, Anchor peer, Orderer peer and General peer (Figure 1).

ENDORSER PEER

If a peer is marked as Endorser Peer, it is responsible for receiving the transaction invocation requests. Upon receiving such a request, the Endorser peer:

- Validates the transaction by checking certificate details and the roles of the requester
- Executes the Chaincode and simulates the outcome of the transaction

The Endorser peer does not update the ledger. It can only approve or disapprove a transaction. It is the only type of node that executes the Chaincode, which is why it is not necessary to install the Chaincode in every node of the network which greatly helps scalability.

In the agri-food case, each actor should have at least one Endorser peer so it can guarantee to the end customers a certain set of rules (Chaincode) has been followed during the process.

ANCHOR PEER

The Anchor peer or a cluster of Anchor peers is configured at the time of channel/ledger configuration. It receives updates from the Orderer peer and broadcasts to all other peers in the network. Anchor peers should be discoverable at any time by all other peers. It should reside in a cloud or data center that guarantees high availability and fast response rates with enough bandwidth.

ORDERER PEER

It is the central communication channel for the whole Hyperledger Fabric network. It is responsible for the consistent state across the network. The Orderer peer creates the new blocks, adds them to the blockchain and delivers them to all peers through the Anchor peers. Orderer peer is built on top of a message- oriented architecture. During development a solo instance is possible, creating a single point of failure for the whole system. A production-ready Hyperledger Fabric network should use a more reliable solution such as Kafka or any other high availability messaging service.

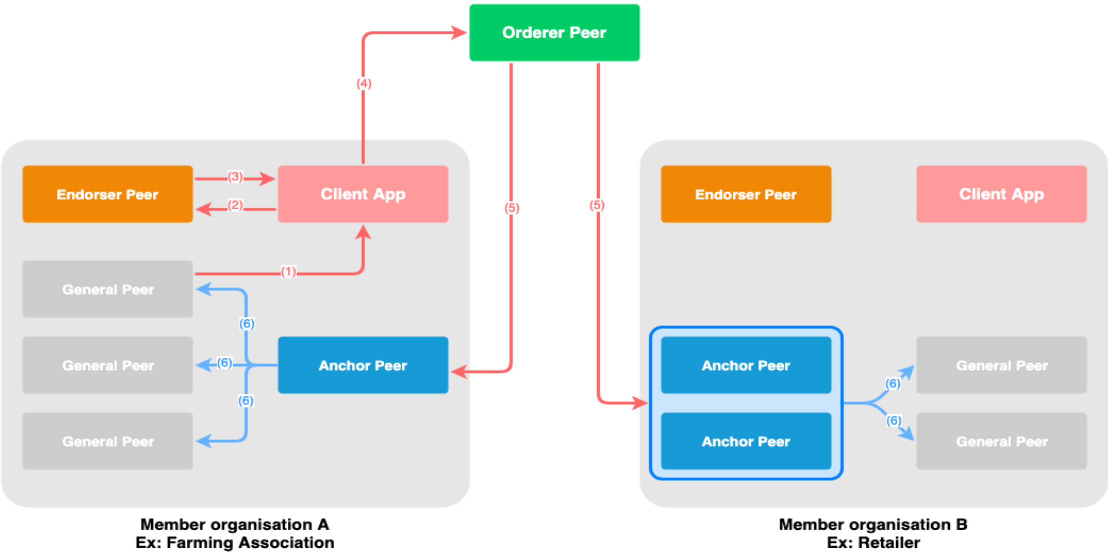


Figure 1. Architectural diagram of peers

BUSINESS LOGIC FLOW (FIGURE 1)

- Some general peer initiates a transaction e.g. a farm delivers goods to a processor, or a distributor arrives at the retailer's warehouse. The transaction is initiated using the Client App (Web App, Mobile App etc.)
- The Client App sends the transaction to the Endorser Peer of the corresponding Member Organization. The Endorser Peer applies the Chaincode/Smart Contracts business logic and either approves or disapproves the transaction. Chaincode logic can be organization specific e.g. different retailers can enforce different rules for freshness.
- After that the result is sent back to the Client App
- In case the transaction is approved, the Client App sends it to the Orderer Peer, and it gets added to the blockchain.
- The Orderer Peer sends the new version of the blockchain to all Anchor Peers, including the ones owned by all other Member Organizations (each Member Organization has at least one).
- All General Peers get access to the new version of the blockchain, containing the latest transactions via the Anchor Peer of the Member Organization that added them to the Hyperledger Fabric Network.

3. CONCLUSIONS

As showcased thanks to the described sample solution, implementing blockchain network for agri-food traceability is feasible. Leveraging an established platform such as Hyperledger, which is advancing through constant contributions by the blockchain community and trusted industry players, solves the inherent challenges for public blockchains and proves the initiative viable. Consumers have clearly articulated how highly they value proven food quality and transparency, which makes food traceability endeavors desirable for the end consumers. Existing technological solutions didn't manage to deliver on the promise for food transparency and even though blockchain is still perceived as complex, farmers and consumers can be presented with simple blockchain-based tools, which they can easily adopt.

We believe that we have already demonstrated the feasibility of building such a system since most of key characteristics of the proposed traceability application framework have already been applied in developing blockchain applications in other domains and environments.

We wish to further develop the FoodBlock framework in order to improve it with a large set of core and other services that will make the development of traceability applications easier and in adherence to the paradigm of Blockchain computing. In conclusion then we feel that in this paper we illustrated an architecture, a business model and basic processes which will enable members of the food supply chain to design and develop the traceability systems that meet many of the requirements identified and without being put under excessive economic costs.

REFERENCES

Liang, W., Jing, C. & Fan, J. (2015). Modeling and Implementation of Cattle/Beef Supply Chain Traceability Using a Distributed RFID-Based Framework in China.

Bechini, A. et al. (2008). Patterns and technologies for enabling supply chain traceability through collaborative information e-business. Information and Software Technology, 50, pp.342-359.

Cao W., Zheng L., Zhu H., Wu P. (2010) General Framework for Animal Food Safety Traceability Using GS1 and RFID. In: Li D., Zhao C. (eds) Computer and Computing Technologies in Agriculture III. CCTA 2009. IFIP Advances in Information and Communication Technology, vol 317. Springer, Berlin, Heidelberg

Hyperledger, An Introduction (2018). [Online] Available at: https://www.hyperledger.org/ [Accessed April 2019].

Investopedia: Public, Private, Permissioned Blockchains Compared. (2018) [Online] Available at: https://www.investopedia.com/news/public-private-permissioned-blockchains- compared/

[Accessed 2019].

Jansen-Vullers, M., Dorp, C. A. V. & A.J.M Beulens, (2003). Managing traceability information in manufacture. International Journal of Information Management, Volume 23, pp. 395-413.

Provenance, 2016. From Shore to Plate- Tracking Tuna on the Blockchain. [Online] Available at: https://www.provenance.org/tracking-tuna-on-the-blockchain Response Media, 2017. 2017 Transparency Study, s.l.: Response Media.

Salampasis, M., Tektonidis, D. & Kalogianni, E. (2012). TraceALL: a semantic web framework for food traceability systems. Journal of Systems and Information Technology, 14(4), pp. 302- 317.

Stefanova, M. (2019). Precision Poultry Farming: Software Architecture Framework and Online Zootechnical Diary for Monitoring and Collaborating on Hens' Health. Information and Communication Technologies in Modern Agricultural Development, pp. 191-205.

Tamirat, T. W., Pedersen, S. & Lind, K. M. (2017). Farm and operator characteristics affecting adoption of precision agriculture in Denmark and Germany. Acta Agriculturae Scandinavica, Section B - Soil & Plant Science 68(4), pp. 1-9.

The Chain Media, (2018). 10 Cryptocurrency Scandals You Need to Know About. [Online] Available at: https://thechain.media/10-cryptocurrency-scandals-you-need-to-know-about [Accessed 2019].

Underwood Sarah. (2016). Blockchain beyond bitcoin. Commun. ACM 59, 11 (October 2016), 15-17. DOI: https://doi.org/10.1145/2994581

Walker, G. (2015). Difficulty. [Online] Available at: https://learnmeabitcoin.com/guide/difficulty [Accessed March 2019].

Wang, X. & Li, Dong (2006). Value Added on Food Traceability: a Supply Chain Manage-ment Approach. IEEE 2006 International Conference on Service Operations and Logistics and Informatics, pp.493-498.

Opportunities, risks and obstacles to the implementation of digitisation technologies in German agriculture

Jana Munz, Nicola Gindele and Reiner Doluschitz
Institute of Farm Management, Agricultural Computer Science and Enterprise Management
Department of Agricultural Science, Hohenheim University, Germany
Jana.Munz@uni-hohenheim.de

ABSTRACT

Given that digitisation in agriculture is currently one of the most important ongoing developments in the agri-food sector in Germany, this paper presents the opportunities, risks and obstacles to implementing digitisation technologies on the German farm enterprise level. Based on an empirical survey with a response rate of 8.4 % carried out at the beginning of 2018, 329 questionnaires on this topic were evaluated and analysed. A factor analysis was used to conceptualise 20 variables regarding the individual opportunities, risks and obstacles to the implementation of digitisation technologies into 6 factors. It revealed positive effects of the use of digital systems, such as various economic and ecological advantages, as well as improved operational coordination. Risks of a socio-economic and financial nature could also be integrated. A lack of both knowledge and infrastructure continue to be major obstacles on the road to agriculture 4.0. Consequently, there is still a need for action in many areas to facilitate the implementation of digital systems in agriculture and to improve the functionality of the systems currently in use.

Keywords: Digitisation, Agriculture, Agribusiness

1. Introduction

Digital progress, one of the most important global transformation processes, is also predominant in German agriculture (Rohleder and Krüsken 2016). The use of digitisation technologies aims to increase the productivity and efficiency of agricultural production thereby contributing to sustainable development (El Bilali and Allahyari 2018). Basic technologies, such as positioning systems and the use of sensors, can launch a new era of agriculture 4.0. They permit the precise positioning of agricultural machinery in the field and optimise the use of plant protection products and pesticides in a site-specific manner (Mintert et al. 2016; Weltzien and Gebbers 2016). Radio-frequency identification technologies and sensors are used for individual animal recognition, the measurement of health parameters or to control stable technology (Weltzien and Gebbers 2016; Doluschitz and Spilke 2002). Farm Management Information Systems (FMIS) can be used to support farm data management by linking and automatically processing information from public sources and farm operating data (Griepentrog 2011; Wolfert et al. 2017). Despite far-reaching explanations of the opportunities that digitisation can open up in agriculture, the current state of digitisation in agricultural enterprises in Germany should be kept in mind. The goal of a cross-corporate data hub to link, collect and document data barrier-free along the value chain and to analyse the data fully automatically, is still a long way off (Möller and Sonnen

2016). The aim of this work is to identify the opportunities, risks and obstacles to implementing digitisation technologies in German agriculture in order to derive suitable recommendations for action on the adoption of digital systems.

2. METHODOLOGY

Given the explorative nature of the research aim, a written quantitative online survey was conducted. The sample consists of freely accessible e-mail addresses of agricultural training enterprises and registered cooperatives throughout Germany. The survey was carried out in two phases between the beginning of January and mid-February 2018. A total of 4,731 subjects could be contacted, of which 800 e-mails were undeliverable. The response rate to the survey was 8.4 %. The low rate is probably due to the fact that the farm enterprises receive a large number of surveys and, therefore, not all of them can be motivated to participate. Due to the selection of the sample (online portals, cooperative register), the responding farms in the sample reflect significantly larger farm structures than the average data provided by the German Farmers' Association (sample: Ø 470 ha UAA[1], median: 160 ha UAA / German Farmers' Association 2018: Ø 60.5 ha UAA) (Deutscher Bauernverband 2018). The company´s legal form of joint partnerships and legal entities are more strongly represented among the farmers in this survey. From a demographic point of view, it can also be seen that the educational level of the sampled farmers is well above the German average. These results can be explained by the fact that only agricultural training enterprises could be interviewed when selecting addresses. These farmers have a higher standard of education and it is assumed that they mainly represent full-time farm operating managers and they farm a larger agricultural area.

Overall, the results of this survey cannot be regarded as representative of German agriculture as a whole. Nevertheless, farms above the growth threshold of 100 ha are represented at an above-average rate (70.4 %). Against the backdrop of the ongoing structural change in agriculture, it can be assumed that farms will continue to grow (Stockinger 2009). Consequently, the current above-average size of the farms in the sample will, in future, be representative of German agriculture.

In the following, the structure discovering procedure of explorative factor analysis is used. The aim of this procedure is to reduce data by combining several variables into a smaller number of common factors. The method used to extract the factors is principal component analysis. In order to assess data quality, the Kaiser-Meyer-Olkin criterion (KMO) and Bartlett's significance were selected as the test parameters (Bühl 2014). Using factor analysis, 20 items were reduced to six extracted factors. The KMO measurement is 0.787 and can, therefore, be rated as "quite good" or "almost deserving" (Kaiser and Rice 1974). The significance of 0.000 underlines the suitability of the selected variables.

Further calculations were made with the six factors formed. Firstly, mean value comparisons were calculated, whereby the individual factors as dependent variables were compared with the size of the enterprise and the educational level. In addition, correlation calculations were performed between the six factors and the variable *estimate of the current significance of digitisation*. In addition, correlation calculations were performed with variables that include the *significance of individual digital systems* (IT systems, automation, data management, intelligent machines, sensors, airborne systems).

[1] Utilised agricultural area (UAA)

3. RESULTS

Table 1: Result of factor analysis, sum of declared total variance= 63.28 %; KMO=0.787; rotation method: Varimax; scale from 1 (very large) to 6 (very small); source: Own survey

Factors	Factor loading
Factor 1: Economic and ecological advantages (*0.863)	
Improved animal welfare	0.836
Early detection of animal diseases	0.785
Increased resource efficiency	0.752
Cost savings	0.747
Positive effects on the environment	0.740
Time savings	0.685
Factor 2: Improved cross-corporate coordination (*0.703)	
Compliance with the obligation to provide evidence for official controls	0.877
Improved transparency	0.757
Factor 3: Socio-economic risks (*0.638)	
Job losses	0.790
Existential risk for small and medium-sized enterprises	0.735
Overly complex topic	0.574
Data protection	0.423
Fear of increased controls	0.375
Factor 4: Financial risk (*0.535)	
Investment risk	0.830
Uncertainty whether high investments will generate a return	0.592
Factor 5: Educational deficits (*0.682)	
Lack of practice-oriented application examples	0.810
Concerns about the functionality and reliability of technology	0.693
Lack of digital competence	0.659
Factor 6: Lack of infrastructure (*0.558)	
Lack of broadband expansion	0.791
Lack of technology or software compatibility	0.743

*Cronbach's α

On the basis of a factor analysis of the variables of the topic areas of opportunities, risks and obstacles to implementing digitisation technologies, six areas could be identified that are of particular relevance to the subjects. Table 1 depicts an overview of the results of the factor analysis.

Factor 1 (economic and ecological advantages) is loaded by six elementary variables that are classified as opportunities in connection with the implementation of digital systems in farm enterprises: positive effects on the environment, cost savings, increased resource efficiency, early detection of animal diseases and improved animal welfare. Additional calculations with this factor confirmed that farms with a higher area endowment classify the economic and ecological advantages of using digital technologies as significantly greater (H-test, p=0.030). It could also be confirmed that graduates of universities or universities of applied sciences attribute higher value to the advantages than master craftsmen or technicians (U-test, p=0.025). In addition, the Spearman correlation produced highly significant correlations between the importance of all listed digital systems and the assessed economic and environmental benefits of using the technologies. The positive correlations are listed in the following: data acquisition, management and analysis (r=0.410, p=0.000), Intelligent agricultural machinery (r=0.382, p=0.000), automation (r=0.341, p=0.000), sensors (r=0.311, p=0.000), airborne systems (r=0.296, p=0.000) and IT systems (r=0.244, p=0.000). From these results it can be concluded that the economic and environmental benefits of digital technologies are deemed to be substantial as soon as they are implemented on farms. Overall, it can be deduced that farms with larger utilised agricultural areas are increasingly using digitisation technologies and that they clearly benefit both economically and ecologically. By means of corresponding economies of scale, which occur when more land is farmed, large enterprises derive greater economic benefit and consequently rate the implementation of digital systems as more positive. The results also show that skilled workers are needed in order to benefit from digitisation. In addition, improved cross-corporate coordination, which represents factor 2 and loads on two elementary variables (improved transparency, compliance with the obligation to provide evidence for official controls), is seen as an opportunity to use digitisation technologies on farms. The Spearman correlation provided highly significant results regarding the relationship between factor two and the importance of data management systems (r=0.261, p=0.000) and IT systems (r=0.263, p=0.000). Therefore, it can be concluded that farmers using digital systems to support their data management see this as a clear improvement in cross-corporate coordination.

Factor 3 reflects the socio-economic risks that are classified by the subjects through the implementation of digital systems. Here, the variables fear of increased controls, data protection, overly complex topic, existential risk for small and medium-sized enterprises and job losses play a particularly important role. In addition, elementary variables (uncertainty whether high investments will generate a return; investment risk), which are related to financial concerns about the digitisation of agricultural enterprises, load on factor 4 (financial risk). This factor represents a statistically significant negative correlation with the importance of digitisation (r=-0.140, p=0.020). This illustrates that the profitability of digital systems is questioned by non-users and can be cited as a proven reason why farmers do not invest in the purchase of digital systems.

Obstacles to the implementation of a digitisation strategy, which farmers have rated as particularly relevant, are reflected in factor 5 (educational deficits). It summarises the elementary variables *lack of digital competence*, *concerns about the functionality and reliability of technology* and a *lack of practice-oriented application examples*. These barriers can be attributed to a lack of education about the functions of digital technologies. In addition, elementary variables (lack of compatibility of technology or software, lack of broadband expansion), which indicate a lack of infrastructure, are summarised in factor 6. This factor correlates significantly and positively with the use of the following digital systems: data management (r=0.145, p=0.015), intelligent agricultural machines (r=0.256, p=0.000), automation (r=0.120, p=0.045) and sensors (r=0.194, p=0.001). These results prove that although individual digital systems are already established on farms and play a major role, the problem of the lack of infrastructure is ongoing. Therefore, it can be concluded that the lack of compatibility of technology or software and the lack of broadband expansion not only constitute an obstacle to the establishment of digital technologies, but also pose a problem for users of extensive digital systems.

4. DISCUSSION AND CONCLUSIONS

On the basis of the factor analysis carried out in connection with the topic areas of opportunities, risks and obstacles to implementing digitisation technologies, six areas could be identified which are of particular relevance to the subjects (KMO=0.787; p=0.000). Various economic and ecological advantages (Cronbach's α=0.863) and improved cross-corporate coordination (Cronbach's α=0.703) through the use of digital technologies were seen as major opportunities. The various risks, on the other hand, could be defined as socio-economic (Cronbach's α=0.638) and financial risk factors (Cronbach's α=0.535). The educational deficits (Cronbach's α=0.682) and a lack of infrastructure (Cronbach's α=0.558) were seen as obstacles. As the size of the farm enterprise and the level of education increased, the economic and ecological benefits were seen as greater (H-test, p=0.030; U-test, p=0.025). This confirms that economies of scale enhance the benefits of using digital technologies and that skilled workers need to be available to take advantage of digitisation. A profitability calculation by Lopotz (2013) confirms that the profitable use of digital systems for site-specific farming is very dependent on the homogeneity of the location, the size of the agricultural area used and the management capabilities. This conclusion underlines the results obtained which indicate that the extent of the use of technical systems depends on the educational level and the size of the farm enterprise (Lopotz 2013). It has also been confirmed that when digitisation technologies are implemented on farms, the benefits are proven in practice. What is particularly worth noting is that the use of IT systems that simplify data management plays a particularly important role in improving cross-corporate coordination (r=0.263, p=0.000). It also has been stressed that the lack of compatibility of technology and software (factor loading: r=0.743), and the lack of broadband expansion (factor loading: r=0.791) constitute problems for users of digital systems. Holster et al. (2012) confirm that these issues constitute a major challenge to achieving the goal of networking and exchanging data. Thus a lack of infrastructure on the way to agriculture 4.0 is still a major obstacle (Holster et al. 2012).

Positive effects can be demonstrated through the use of digital systems. However, there is still a need for action to facilitate the implementation of digital systems in agriculture and to improve the functionality of the systems currently in use.

Countries are required to digitise area data and make it available free of charge. Similarly, specific instructions for agricultural input should be digitised and made available in order to automatically check compliance with legal or voluntary standards (Deutscher Bauernverband 2016). Uniform data standards along the entire value chain should be introduced and incentives should be given to offer exclusively manufacturer independent applications and software which is compatible across platforms. Individual initiatives, such as standardised interfaces between tractors, tools and computers (ISOBUS), have not yet been sufficiently accepted in practice. Alternatively, research must advance solutions for universally compatible data integration and networking using semantic technologies. Furthermore, there should be clear legal regulations regarding data exploitation rights and data protection.

In order to overcome the educational deficits, knowledge about modern, digitised technology should be imparted both in agricultural training and in university studies. Training also has an important role to play in facilitating the use of digital systems and reducing inhibitions and prejudices in farmers who are on average older.

One starting point for an economic use of digital systems in small-scale agriculture is the creation of inter-company cooperation (e.g. machinery sharing, cooperatives). Qualified services from contractors are another option. More cost-effective applications, such as simple agricultural apps or farm management information systems, could potentially help to organise business processes more efficiently and save time by facilitating documentation.

References

Bühl, Achim (Ed.) (2014) 'SPSS 22. Einführung in die moderne Datenanalyse', 14th edition Pearson Deutschland: Pearson.

Deutscher Bauernverband (2016) 'Landwirtschaft 4.0-Chancen und Handlungsbedarf', Positionspapier des Präsidiums des Deutschen Bauernverbandes. Available online at http://media.repro-mayr.de/34/661134.pdf.

Deutscher Bauernverband (2018) 'Situationsbericht 17/18. 3. Agrarstruktur'. Available online at http://www.bauernverband.de/33-betriebe-und-betriebsgroessen-803628, last accessed on 29.03.2018.

Doluschitz, Reiner; Spilke, Joachim (2002) ‘Agrarinformatik’, Stuttgart (Hohenheim): Ulmer (UTB für Wissenschaft Agrarwissenschaften, 2230).

El Bilali, Hamid; Allahyari, Mohammad Sadegh (2018) ‘Transition towards sustainability in agriculture and food systems: Role of information and communication technologies’, Information Processing in Agriculture, 5 (4), pp. 456–464. doi: 10.1016/j.inpa.2018.06.006.

Griepentrog, Hans W. (2011) ‘Smart Farming: Praxisreife Lösungen, Erfordernisse und Techniken für morgen’, DLG Pressedienst 2011. Available online at http://presse.dlg.org/pdf/dlg.org/1/4937.

Holster, Henri; Horakova, Sarka; Ipema, Bert; Fusai, Benedicte; Gannerini, Ginfranco; Martini, Daniel; Shalloo, Laurence; Schmid, Otto (2012) ‘Current situation on data exchange in agriculture in the EU27 & Switzerland. Final Report’. Available online at http://library.wur.nl/WebQuery/wurpubs/fulltext/206268.

Kaiser, H.; Rice, J. (1974) ‘Little Jiffy, Mark IV’ Educational and Psychological Measurement’ (34), pp. 111–117. Available online at http://journals.sagepub.com/doi/pdf/10.1177/001316447403400115.

Lopotz, Harald (2013) ‘Precision Farming: Rechnen sich die Investitionen?’, Landwirtschaftskammer Nordrhein-Westfalen, 2013. Available online at http://www.duesse.de/rueckblick/pdf/2013-06-19-rentabilitaet-pf.pdf, last accessed on 28.04.2018.

Mintert, J.; Widmar, D.; Langemeier, M.; Boehlje, M.; Erickson, B. (2016) ‘The challenges of precision agriculture: is big data the answer?’, Paper presented at the Southern Agricultural Economics Association (SAEA) Annaul Meeting; San Antonio, Texas; February 6-9, 2016. Available online at https://ageconsearch.umn.edu/bitstream/230057/2/THE%20CHALLENGES%20OF%20PRECISION%20AGRICULTURE_manuscript_SAEA_2016.pdf.

Möller, Jens; Sonnen, Johannes (Ed.) (2016) ‘Datenmanagement in Landwirtschaft und Landtechnik. Intelligente Systeme- Stand der Technik und neue Möglichkeiten’, Ruckelshausen, A., et al. (Ed.): Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn, pp. 133-136. Available online at https://dl.gi.de/bitstream/handle/20.500.12116/788/133.pdf?sequence=1.

Rohleder, Bernhard; Krüsken, Bernhard (2016) ‘Digitalisierung in der Landwirtschaft’, Bitkom; Deutscher Bauernverband. Berlin, 2016. Available online at https://www.bitkom-research.de/WebRoot/Store19/Shops/63742557/5819/BD75/5F7A/C381/3D6E/C0A8/2BBA/AC38/Digitalisierung_in_der_Landwirtschaft.pdf.

Stockinger, C. (2009) ‘Was entscheidet über die Entwicklungsfähigkeit eines Betriebes? Chancen und Anforderungen an Familienbetriebe’, Deutsche Landwirtschafts-Gesellschaft (Ed.) (2009): Landwirtschaft 2020. Herausforderungen, Strategien, Verantwortung. Frankfurt a. M, pp. 101–1017.

Weltzien, Cornelia; Gebbers, Robin (2016) ‘Aktueller Stand der Technik im Bereich der Sensoren für Precision Agriculture’, Arno Ruckelshausen et. al (Ed.): GI-Edition Band 253 - Informatik in der Land-, Forst- und Ernährungswirtschaft, 36. GIL-Jahrestagung in February 2016, Osnabrück. Bonn: Köllen. Available online at http://www.gil-net.de/Publikationen/28_217.pdf.

Wolfert, Sjaak; Ge, Lan; Verduow, Cor; Bogaardt, M. (2017) ‘Big Data in Smart Farming- A review’, Agricultural Systems, 153, pp. 69–80. doi. https://doi.org/10.1016/j.agsy.2017.01.023

OPTIONS FOR AUTOMATIC IDENTIFICATION OF USER ACTIVITIES IN USABILITY TESTING

Jan Masner, Jiří Vaněk, Jan Pavlík, Eva Kánská, Edita Šilerová
Department of Information Technologies, Czech University of Life Sciences Prague, Czech Republic
masner@pef.czu.cz, vanek@pef.czu.cz, pavlikjan@pef.czu.cz, kanska@pef.czu.cz, silerova@pef.czu.cz

ABSTRACT

When testing usability of applications, it is often needed to analyze behavior of users in terms of identifying their activities. The activity may be that the user is working on the assignment without problems, is searching for something, is absolutely lost in user interface, is filling a form, is studying the manual, etc. Identifying of the activities is usually done by tagging a video and audio record of the testing, optionally together with visualization of eye movements (eye tracking). It is a very time-consuming work for the usability experts. When testing in a specialized laboratory, we can obtain data from various measurements. Besides audio-visual record, data from eye-tracking, click tracking and keyboard tracking can be analyzed. Moreover, we can engage biometrical data such as pulse, skin temperature, humidity, hand movements, etc. The research question of the paper is whether it is possible to analyze all the data and develop methods and algorithms to automatically identify the user activities. We have performed several experiments in the specialized laboratory for usability testing. The paper describes the research design, experiments, methods of data processing and analysis. Finally, first conclusions and findings are discussed, as well as forthcoming research. The proposed research can be generalized to usability of any products. For example, using glasses for eye tracking and smartwatches, it is possible to conduct usability research of agricultural machinery.

Keywords*:* usability, UX, software, video-tagging, human-computer interaction

1. INTRODUCTION

In general, usability can be defined as the ease of use and learnability of a human-made object. In software engineering, the usability is defined by ISO 9241-11 as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (IDF, 2019). Usability plays an important role in selling any products, whether it is software or equipment for precision farming (Jarolimek et al., 2019). In computer sciences, usability is often extended by using a broader term User Experience (UX) (Benda et al., 2017).

Software products are becoming increasingly complex and it is necessary to consult their development with target group. Therefore, testing usability with end-users has become very important. The testing which utilizes end-users as test subjects can be divided into two basic techniques, testing remotely and laboratory testing (Rubin and Chisnell, 2008). Distance testing requires that the users first start a support application that records their actions (Atterer et al., 2006). Moreover, distance testing tends to run into unexpected problems that complicate the process (Kaikkonen et al., 2005). The second mentioned method appears as preferable for many cases. For laboratory testing, users are located centrally in one place. Hardware and software testing requirements are thus secured by equipment in the test lab.

The complexity of the gathered data is constantly rising as new testing techniques are being implemented. A visual form, such as audiovisual records, is a typical style of output when performing a usability testing in laboratory. Analysis of such data is highly time-consuming (Prasse, 1990). In contrast of quantitative data, it is not yet possible to algorithmically analyze the audiovisual outputs or transform them into such form. Therefore, researchers have to perform a thorough review and analysis of the recordings. When testing higher amount of users and tasks, the time consumption of the analysis is astronomical (Nørgaard et al., 2006).

When testing usability of applications, it is often needed to analyze users' behavior in terms of identifying their activities. The activity may be that the user is working on the assignment without problems, is searching for something, is absolutely lost in UI, is filling form, is studying the manual, etc. The outputs then can be easily visualized using various technics such as pie chart. The pie chart has a significant predictive value and can be easily understood by management workers. The time consumption of the user activities is important, for example, in decision making about the usage of the tested software (Harel et al., 2008). The Department of Information Technologies at CULS in Prague used this type of output to support various usability research studies for the Czech Ministry of Agriculture and the Ministry of the Environment. This type of analysis usually needs to employ a log of the desired user activities. This can be achieved by the pre-analysis of the user activities during the test followed by detailed tagging of the testing video recording (Wynn and Still, 2011). However, this type of analysis has to be done manually by the usability experts. It is a very time-consuming activity. Some usability test may last dozens of minutes. Many parts of the video recording need to be reviewed repeatedly. One of the possible solutions to this critical problem is some kind of automatic processing of the video recordings, and auto-tagging of the user activities (Petropavlovskiy and Nefedova, 2017).

Recently, there have been a lot of new advanced methods and technologies for usability testing in a laboratory. The gathered data are highly valuable for understanding users' behavior (Çakar et al., 2017). Besides the audiovisual recordings, it is possible to track the user input (e.g. click tracking and keyboard tracking). It is possible to visualize and analyze the user behavior using eye tracking equipment (Hild et al., 2012) which can provide additional valuable insights (Wang et al., 2019). Moreover, it is possible to get various biometrical data. These data can enrich the overall picture of users' experience such as stress or emotions (Mirza-Babaei et al., 2011). In regard to such data volume and its complexity, there is a problem with the processing and transformation into a necessary form for testing hypotheses and drawing conclusions. Manual processing of such data is very difficult and often not suitable for real scenarios.

The objective of the paper is to introduce a methodology for the research to find methods and algorithms for automatic data processing. The output should be a correct identification of the user's activities during usability testing. The research is currently conducted at the Department of Information Technologies at CULS Prague.

2. METHODOLOGY

Currently, self-developed software is used to analyze the video record and tag the user activities, named VEA-UX. It supports definition of custom activities, option to save the project for later work, and dividing tagger's workload into sections. The sections mostly serve as checkpoints for the tasks which users have to perform during the testing. The application works on a slightly different principle than other similar software. The events are marked when they conclude. This approach helps to save time needed for the analysis. It avoids excessive rewinds to the beginning of the activity. The software also supports backward editing of any tagged events and sections. The capabilities of the software are illustrated in a Figure 1. Despite the fact, that the VEA-UX application can save time for the analysis, it is still time-consuming. Nevertheless, the VEA-UX will be used later for evaluation.

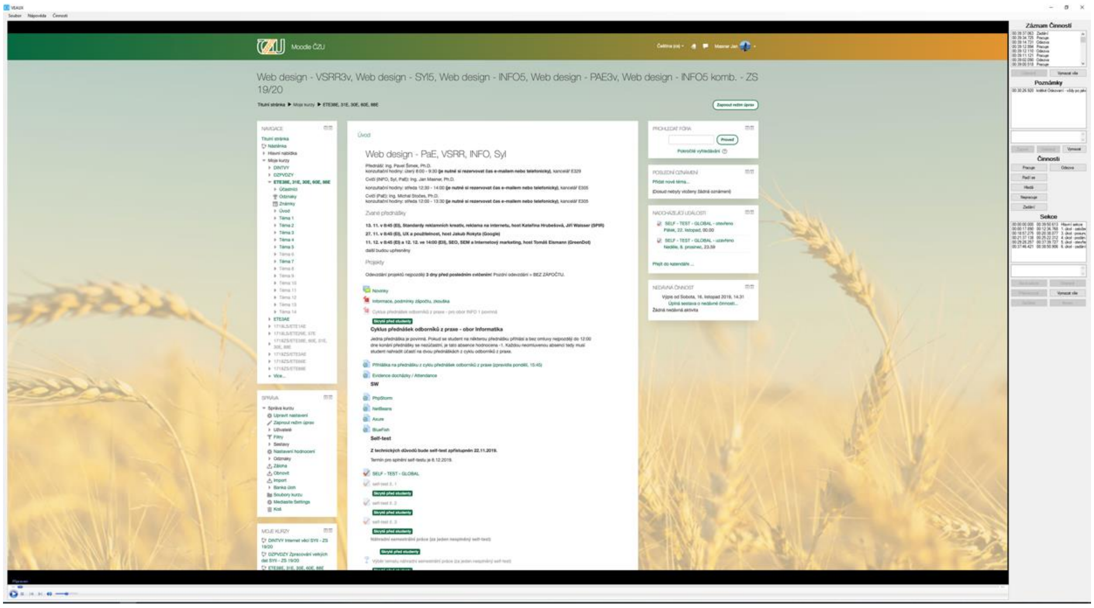


Figure 1. VEA-UX software used for tagging user activities from a video recording

Experiments

The proposed methodology consists of several steps. First of all, some experiments were carried out. At the Faculty of Economics and Management, specialized laboratory for usability testing (HUBRU – Human Behavior Research Unit) was used. The laboratory is equipped with computers, devices for eye tracking, and advanced recording equipment.

For the first set of experiments, testing environment in Moodle was created. It consisted of a copy of the existing course taught by the Department of Information Technologies. The testing scenario consisted of 5 tasks, ranging from very easy to difficult to accomplish. The tasks for users were as follows:

1. Find and write down at what time the Lectures are held.
2. Find out who attends the course with you. Choose one of your fellow students and send him a message.
3. Hide or show some item from a course.
4. Create a new topic in the course. Add to the topic some file material.
5. Change overall score (grade) of a student named Michael White to 200 points.

Eight users were tested to get the first results. The test group contained users with no experiences with Moodle as well as students with some knowledge about its functions, and experienced teachers. To gather the data for later analysis, the Tobii Pro X2 (60Hz version) eye tracker together with click and keyboard tracking, and Huawei Watch 2 were used. A very important setup which needs to be done before such testing, is to synchronize the times across devices. The synchronization is necessary to properly merge the data from different sources. We also placed a smartwatch on users' dominant hand (which they use to control the mouse). This setup optimizes the use of accelerometer and gyroscope data.

Data preparation

The eye tracking data together with click and keyboard tracking data were gathered through a Tobii Pro Studio software which operates the eye tracker device. The timestamp was retrieved from the computer time. Data from smartwatch are retrieved via the application SensorCap. It allows to synchronize the time with a custom NTP (Network Time Protocol) server and easily export sensor data in CSV format. The same NTP server has to be set on the computer with eye tracking. Selected sample data are shown on Table 1 – eye tracking, and in Table 2 – smartwatch.

Table 1. Sample data from eye tracking (selected columns)

Recording Timestamp	Local Timestamp	GazeEvent Type	GazeEvent Duration	GazePoint. Index	…
23	15:18:42.210	Fixation	400	1	…
39	15:18:42.226	Fixation	400	2	…
57	15:18:42.244	Fixation	400	3	…
73	15:18:42.260	Fixation	400	4	…
90	15:18:42.277	Fixation	400	5	…

Table 2. Sample data from smartwatch - heartrate

eventTimeNanos	deviceTimeMillis	ntpTimeMillis	sensorAccuracy	HTR
7464907545779	1548771430329	1548771430044	3	91.0
7465931257028	1548771431353	1548771431068	3	91.0
7466929089111	1548771432351	1548771432066	0	92.0
7467999088174	1548771433421	1548771433136	0	93.0
7468908427913	1548771434330	1548771434045	0	93.0

The data shown in Table 1 and Table 2 need to be joined together. Additionally, there are more smartwatch data sets. Firstly, we used data from accelerometer sensor. The data have similar structure as shown in Table 2. To sum up, we needed to merge three data sets – eye tracking with keyboard and click streams, heart rate data, and accelerometer data. The problem is that the granularity is different – the values are taken at different periods. Using Jupyter notebook and Python programming language we developed an algorithm to merge the datasets. Using the smallest granularity according to Local Timestamp in Table 1, the data were resampled into milliseconds. Subsequently, all three datasets can be merged and grouped.

3. Further Research

Data Analysis

Firstly, the exploratory analysis will be carried out. Correlation analysis will be used to determine whether there are variables that can be omitted. Visualization also helps in understanding the data. We use mostly Tableau. For the later model development, we plan to use machine learning. Therefore, we will firstly engage statistical classification using artificial neural networks. The TIBCO Statistica software can be easily used without further programming.

During the analysis, the input data sets will need to be adjusted. Several factors need to be considered. The input variables are the most significant for the result. Other variables can be computed, for example delayed values can help to recognize longer activities.

The data, which has been gathered, can serve as a good basis for further research. Very important factor is to properly prepare the testing environment as well as the tasks for the users. Additional tests with users will be done according to the data analysis. We have other smartwatches available. It is also possible to use dedicated equipment for biometric feedback. It can measure additional values such as skin temperature, skin conductance, motility, and advanced pulse data.

4. Conclusions

When finished, the proposed research and its output methods can help to save time during usability research studies. The methods and algorithms will be able to automatically identify the problem parts of the tested software. This can help, for example, in backwards analysis of the video records to identify the problem parts. Usability experts can then focus only on the most important aspects of the tested product. In some cases, it is required to analyze the effectivity of using the product. We can identify, when users spend too much time by searching, navigating, or by other activities which are not productive. The results of such usability studies can be delivered in a radically shorter time.

The proposed research can be generalized to usability of other then software products. For the tracking of eye movements, it is possible to use special glasses such as *Tobii Pro Glasses*. The user does not have to be at a fixed position in front of the computer or other device. For example, farmers can wear the glasses when operating a tractor, harvester, etc. Together with the use of smartwatches, users can wear the equipment for a longer period of time and in a real, natural environment. The testing would not be affected by any interference or stress caused by the unnatural environment in a laboratory. To analyze testing of such length, the use of traditional video analysis methods would be time-consuming, economically inefficient, and therefore ill-advised.

Acknowledgments

The results and knowledge included herein have been obtained owing to support from the following institutional grants. Internal grant agency of the Faculty of Economics and Management, Czech University of Life Sciences in Prague, grant no. 2019B0009 – “Life Sciences 4.0“.

References

Atterer, R., Wnuk, M., Schmidt, A., 2006. Knowing the user’s every move. https://doi.org/10.1145/1135777.1135811

Benda, P., Šimek, P., Masner, J., Vanek, J., 2017. Analysis of eAGRI web portal ergonomics and presentation of information in terms of the general public. Agris On-line Pap. Econ. Informatics 9. https://doi.org/10.7160/aol.2017.090401

Çakar, T., Rızvanoğlu, K., Öztürk, Ö., Çelik, D.Z., Gürvardar, İ., 2017. The use of neurometric and biometric research methods in understanding the user experience during product search of first-time buyers in E-commerce, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-58634-2_26

Harel, A., Kenett, R.S., Ruggeri, F., 2008. Decision support for user interface design: Usability diagnosis by time analysis of the user activity, in: Proceedings - International Computer Software and Applications Conference. https://doi.org/10.1109/COMPSAC.2008.61

Hild, J., Peinsipp-Byma, E., Klaus, E., 2012. Improving usability for video analysis using gaze-based interaction, in: Full Motion Video (FMV) Workflows and Technologies for Intelligence, Surveillance, and Reconnaissance (ISR) and Situational Awareness. https://doi.org/10.1117/12.919042

IDF, 2019. Usability [WWW Document]. URL https://www.interaction-design.org/literature/topics/usability

Jarolimek, J., Masner, J., Vanek, J., Pankova, L., 2019. Assessing Benefits of Precision Farming

Technologies in Sugar Beet Production. List. Cukrov. A Repar. 135, 57–63.

Kaikkonen, A., Kallio, T., Kankainen, A., Cankar, M., Kekalainen, A., 2005. Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. J. USABILITY Stud.

Mirza-Babaei, P., Long, S., Foley, E., McAllister, G., 2011. Understanding the contribution of biometrics to games user research. Proc. DiGRA.

Nørgaard, M., K, H., K, 2006. What do usability evaluators do in practice? An explorative study of think-aloud testing. In: Proceedings of the. DIS 2006 Conf. Des. Interact. Syst. ACM Press. New York. https://doi.org/10.1145/1142405.1142439

Petropavlovskiy, M., Nefedova, O., 2017. AUTOMATIZATION OF COLLECTION AND USE OF THE INFORMATION FROM OFFICIAL WEBSITES OF EDUCATIONAL ORGANIZATIONS. IJAEDU- Int. E-Journal Adv. Educ. https://doi.org/10.18768/ijaedu.336265

Prasse, M.J., 1990. Video analysis method. An integrated approach to usability assessment, in: Proceedings of the Human Factors Society. https://doi.org/10.1177/154193129003400436

Rubin, J., Chisnell, D., 2008. Handbook of Usability Testing, Indianapolis, IN: Wiley Pub. https://doi.org/10.1007/s13398-014-0173-7.2

Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., Fieldman, A., 2019. Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data. Int. J. Hum. Comput. Interact. https://doi.org/10.1080/10447318.2018.1464776

Wynn, J., Still, J.D., 2011. Motivating change and reducing cost with the discount video data analysis technique, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-21708-1_37

SPATIAL ANALYSIS OF CROP DIVERSITY IN HUNGARY BEFORE AND AFTER INTRODUCTION OF GREENING

Márta Gaál[1], András Molnár[2]

[1] Research Institute of Agricultural Economics, National Agricultural Research and Innovation Centre, Budapest, Hungary

[2] Takarek Group, Budapest, Hungary

gaal.marta@aki.naik.hu, andras.molnar.aki@gmail.com

ABSTRACT

Due to the introduction of greening measures during the latest reform of the Common Agricultural Policy (CAP), the assessment of crop diversity has increasingly been a focus of interest. Better understanding of the factors that may influence the outcomes of diversity calculation could help the further evaluation of the regulations. This study investigates the spatial pattern of crop diversity in different scales and the changes over time in Hungary. Regional differences in crop diversity can be demonstrated at all scales examined, but the patterns are not exactly the same. Some increase in diversity can be revealed after the introduction of greening, but there are no further increasing tendencies.

Keywords*:* crop diversity, Shannon index, spatial resolution, agricultural policy

1. INTRODUCTION

Many authors have studied the advantages of diversifying agriculture crops and the negative effects of having a monoculture system. A farm growing only few crops is very sensitive to be affected by the emergence of a pathogen, pest invasion, and weather damages, or economic and market risks as well. Crop diversity can be realized across time by crop rotation, while diversifying the portfolio of cultivated crops maintains diversity in space (Weigel et al., 2018). Diversity has a positive and significant effect on crop production, especially when rainfall is low (Donfouet et al., 2017), and can potentially contribute to offer ecological services (Montelone et al., 2018). On the other hand, the negative effects of the monoculture on soil quality and on ecological conditions are also well known. However, excessive diversification could lead to adverse fragmentation of resource capacities, which may prevent the development of an optimum production level and size.

The Common Agricultural Policy (CAP) also puts emphasis on efficient agro-environmental actions for sustainable development. The 2013 CAP reform introduced a payment for a compulsory set of ‘greening measures’, one of which is crop diversification. According to this, farms having 10 to 30 hectares have to grow at least two crops, with the main crop representing a maximum of 75% of the arable area. Farms over 30 hectares are required to have at least three crops, the main crop covering at most 75% of the land and the two main crops together at most 95% (EU, 2013). Based on the European Commission (EC, 2017) report 75% of arable land in the EU was affected by this measure in 2016, while more than 90% in Hungary.

Agriculture is a traditionally important sector in the Hungarian economy, as the country has favorable soil and climatic conditions for several crops. However, the combined area of wheat and maize, accounts for about 45% of the arable land. The present study aims to compare crop diversity in Hungary before and after introduction of greening, and to explore whether the regulation had a significant impact on the cropping pattern. The effect of the spatial resolution of the analysis was also examined.

2. METHODOLOGY

2.1 Database

The Hungarian Agricultural Land Parcel Identification System (Hungarian acronym: MePAR), which is mandatory for the administration of agricultural subsidies, is based on physical blocks (more than 300,000). The size of the blocks varies on a relatively wide range, while the average size of the eligible area is around 25 hectares. A physical block may contain several agricultural parcels. When farmers submit subsidy claims they indicate their parcels and crops within the blocks, but the country-wide parcels' geometry is available only for the Hungarian Paying Agency.

In a former study (Gaál et al., 2018) crop diversity at species-level was analyzed based on the administrative data of 2010-2014. For data protection reasons the agency transformed the parcels' data to a 1x1 km (100 hectares) grid, therefore the obtained database contained for each grid the crop codes and the area by crops. The total number of the grid cells was 94,293.

Recent calculations were based on the same administrative database for 2014-2018, but attribute data (block id, crop code and area) were obtained at parcel level, which allowed the calculations at block level. Additionally, the blocks could also be aggregated at higher administration levels, from which the settlements (LAU 2) were applied. There were 3,154 settlements with an average size of almost 3,000 hectares.

2.2 Methods

Many diversity indices for calculating richness and evenness are frequently applied in ecology and landscape metrics, like the Shannon diversity index (SDI), the Simpson diversity index, and the Berger-Parker index. The Shannon diversity index has been applied as an indicator of cropland diversity in several studies (e.g. Aguilar et al., 2015; Donfouet et al., 2017; Montelone et al., 2018; Weigel et al., 2018). SDI can be calculated according to the following equation:

$$SDI = -\sum_{i=1}^{N} p_i * lnp_i$$

where N is the number of land cover types and p_i corresponds to the proportional abundance of the i^{th} type. The index can vary between 0 to infinity in theory; a low value represents low diversity while a high value evidences high diversity. Calculations were done block and settlement levels, and each crop code was counted as separate crop type.

Optimized hot spot analysis was implemented to detect spatial clustering of Shannon diversity values. This tool identifies statistically significant spatial clusters of high values (hot spots) and low values (cold spots) using the Getis-Ord Gi* statistic.

Additionally, statistical analyses were carried out to test the significant differences among the diversity values. Due to the non-normal distribution, Kruskal-Wallis and Mann-Whitney test were applied to reveal the overall differences among the years. Pearson's chi-squared statistics and Kendall's tau-b were also applied to check the differences in data distribution.

Basic calculations were done in a PostgreSQL database. Spatial clustering (optimized hot spot) and the representation were carried out with ArcGIS Desktop software.

3. Results

3.1 Crop diversity at different scales

Considering the whole country, the total number of crop types varied between 314 and 400 per year. In a greater area there is a higher probability to cultivate more crops, therefore the crop diversity values depend on the size of the units examined. At block level the maximum number of the crop types was 27, with an average value of 2.5. Using the 1x1 km grid, the maximum number of crop types was 33, while the average was 6. The settlements cover much greater area, therefore at this level the maximum value was 133, while the average 23. Settlement level analysis of course hides diversity at lower levels of aggregation, especially the occurrence of the zero diversity values (only one crop) has reduced. Therefore, greater level of aggregation (blocks –> 1x1 km grid –> settlements) indicated a shift to the higher diversity values. However, adjusting the symbol categories to the given data, some clear regional patterns can be observed at any scale (Figure 1).

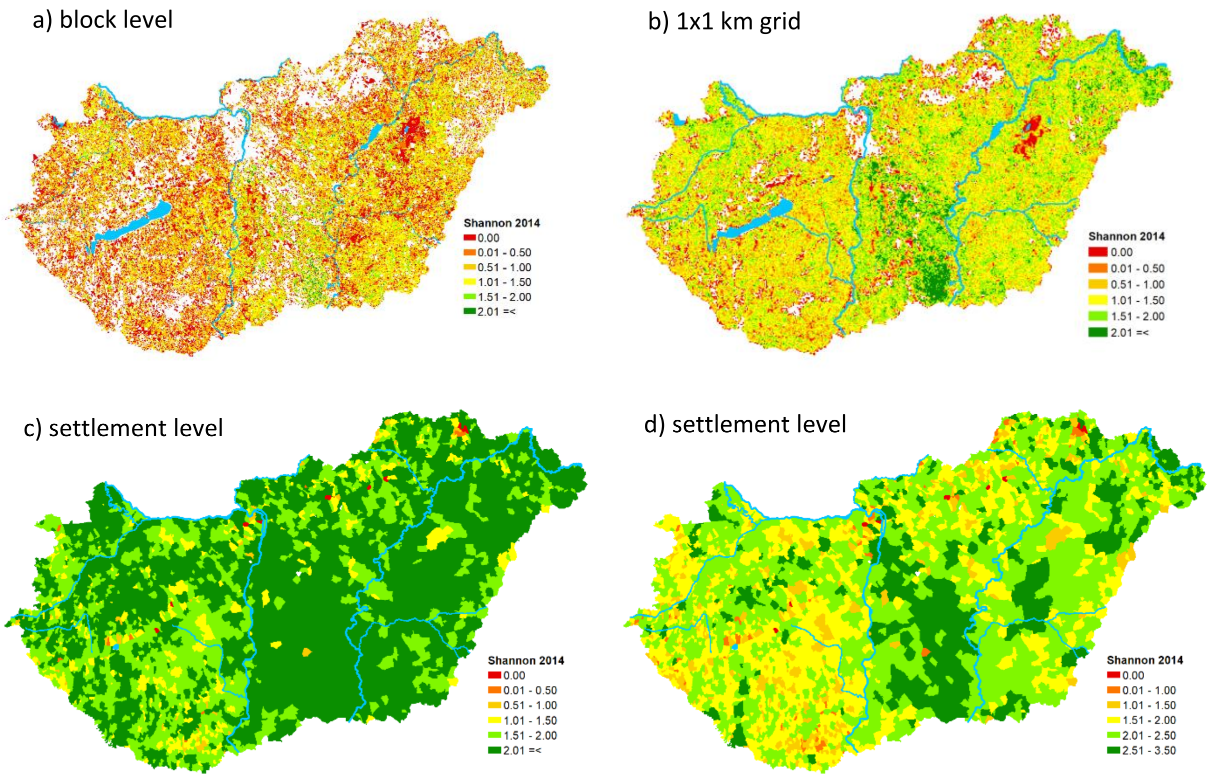


Figure 1. Scale dependency of crop diversity results
(a, b, and c with the same categories, d adjusted to the settlement level result)

3.3 Changes over time

The Shannon diversity index was calculated for each year from 2014 to 2018, for the physical blocks and for the settlements, as well. National average diversity values were calculated in two ways. First, as simple averages of the values obtained for the blocks/settlements, and secondly due to the different sizes of crop area in the blocks/settlements, area based weighted averages were also calculated (Table 1).

Except from the unweighted averages for blocks, there was an overall increase in crop diversity from 2014 to 2016, and until 2018 it remained almost stable. Based on the statistical analysis this increase proved to be significant for unweighted values at settlement level and weighted averages at block level. For weighted values at settlement level the significant difference between 2014 and 2015 could be kept only at 90% probability (p=0.075). The unweighted and weighted averages for the blocks show opposite tendencies over the years. Probably due to the increasing number of blocks (from 217,274 to 257,622 in total), and consequently the increasing number of the smaller blocks with one crop type (SDI = 0). In 2014, 87,736 blocks were found with mono-cropping, were the main crop types were permanent grassland (15,329 blocks, 17.5%), maize (14,856, 16.9%), and wheat (11,435, 13.0%). In 2018, 114,405 blocks were covered only by one crop type, which is about 30% more. The number of the blocks with wheat had a slight increase (plus 1,222 blocks), while the blocks with maize had decreased (with 1,018 less). The main changes are related to permanent grasslands, as the number of the blocks became more than twice (34,863), while the total grassland area has not changed. So, these were only administrative changes, which refer to the more precise delineation of grassland areas according to the maintenance of permanent grassland measure.

Table 1. Statistics of the Shannon diversity on national level

Calculation		2014	2015	2016	2017	2018	Chi-square	*p*- value
Blocks	average	0.53	0.54	0.51	0.50	0.49		
	median	0.51 *a*	0.51 *b*	0.44 *c*	0.40 *d*	0.38 *e*	1351.6	0.0001
Blocks, weighted	average	0.72	0.77	0.77	0.76	0.76		
	median	0.69 *a*	0.78 *b*	0.78 *c*	0.77 *d*	0.77 *e*	2088.6	0.0001
Settlements	average	1.96	2.04	2.06	2.04	2.03		
	median	1.98 *a*	2.07 *b*	2.09 *b*	2.08 *b*	2.08 *b*	106.2	0.0001
Settlements, weighted	average	2.13	2.21	2.23	2.20	2.20		
	median	2.16	2.24	2.25	2.22	2.22	4.054	0.3988

* different letters indicate the statistically different years

In addition to averages, the distribution of the diversity values is also important. The distributions were analyzed according to the categories presented in Figure 1a for block level and for settlement level Figure 1d. At both level significant correlations (p=0.000) were found between the categories and the years. In the case of settlements (Table 2) zero diversity values were hardly found and had a decreasing tendency. The remaining two settlements had only permanent grasslands. A decreasing tendency can be observed in the diversity category of 1.01-1.5, while a continuous increase in the category of 2.01-2.05, both in the number of settlements and the percentage of area affected.

Table 2. Distribution of the Shannon diversity values on settlement level

SDI categories	2014		2015		2016		2017		2018	
	count	area	count	area	count	area	count	area	count	area
0	9	0.0%	10	0.0%	6	0.0%	7	0.0%	2	0.0%
0.01-1.00	58	0.4%	51	0.2%	56	0.2%	66	0.3%	70	0.3%
1.01-1.50	346	5.6%	220	3.0%	215	2.9%	215	2.9%	217	2.9%
1.51-2.00	1199	29.5%	1065	23.3%	1004	21.9%	1018	23.5%	1042	23.9%
2.01-2.50	1252	47.8%	1449	51.7%	1486	52.9%	1479	52.9%	1488	55.1%
2.51-3.50	281	16.8%	355	21.9%	383	22.0%	365	20.4%	330	17.8%

At block level a weak negative correlation (Kendall's tau-b = -0.0269) can be observed between the diversity categories and the years examined, which can be a consequence of the increasing number of blocks.

3.2 Spatial distribution

Optimized hot spot analysis was implemented to identify statistically significant spatial clusters of high (hot spots) and low diversity values (cold spots). Hungarian geographical meso-regions were also used to understand the patterns. Due to the temporal variability, the patterns were slightly different over the years. Figure 2 shows the year 2018 as an example.

At block level, high diversity areas were found at Körös-Maros interfluve (1.13), in Hajdúság (1.11), Bácska Plain (1.3), and at the Small Plain (2.1), which are high-quality arable land with many crops. Additionally, the farm structure – more parcels in the blocks – can contribute to the high diversity between the Danube and Tisza (1.1-1.2), and at the Upper-Tisza region (1.6). Low diversity is typical for unfavorable soil conditions, and for the hilly-mountainous parts of the country.

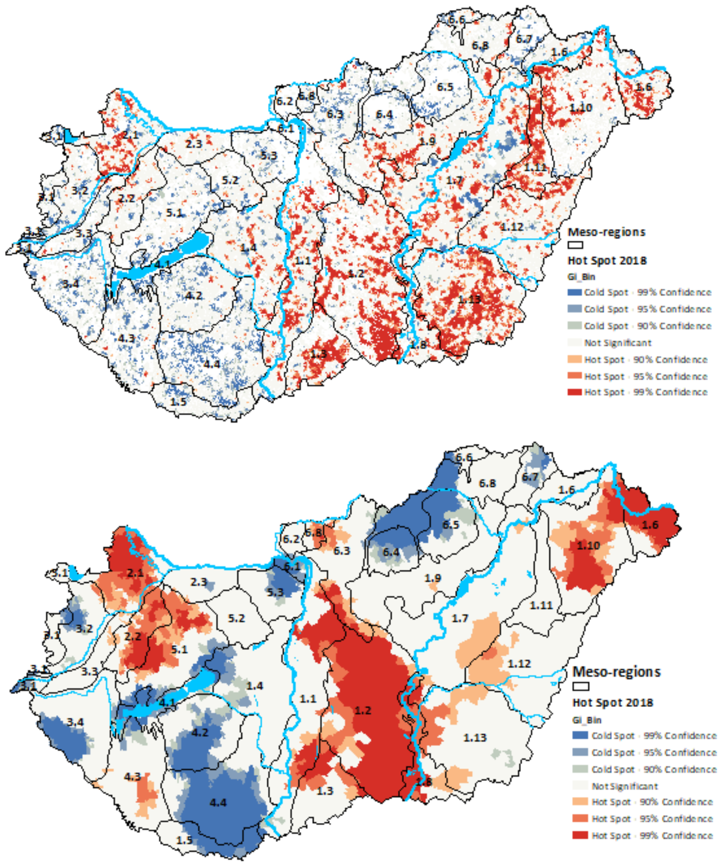


Figure 2. Crop diversity hot spot analysis results
(red areas are significant clusters of high diversity, while blue areas are clusters of low values)

Similar patterns can be recognized also at settlement level, e.g. the high diversity areas at meso-regions 1.2, 1.6, and 2.1. However, due to the aggregation, hot spot areas can appear at settlement level, when the blocks have few crops, but all together several crops are grown at the given area, e.g. in Nyírség (1.10). In contrast, high diversity at block level does not imply high diversity at settlement level. This may be typical to areas with greater blocks, like at Körös-Maros interfluve (1.13).

Cold spot areas were found in the Balaton area (4.1), Mecsek Mountains and Tolna hills (4.4), and Mátra and Bükk Mountains (6.4-6.5), which are typical wine regions.

4. DISCUSSION

Instead of the clear trends, diversity values showed both spatial and temporal variabilities over the years. The interpretation of the changes in diversity has difficulties, as the calculated results depend on several factors. Spatial diversity pattern could be partly explained by topographic, climatic, and soil factors, but farm structure has also an important effect. At the same time, the choice of the spatial unit is crucial during the analysis. However, some low/high-diversity areas can be identified at all scales examined. The diversity values were significantly higher in 2015 than in 2014, which might indicate the effectiveness of the introduction of greening. However, this increase was observed only at 65% of the settlements, and there are no settlements where the increase in diversity would be continuous since then.

In the presented results all crop types were counted as separate for the diversification measure, however, some of them like species of *Brassicaceae*, *Solanaceae*, and *Cucurbitaceae* should be aggregated according to the regulations. Therefore, the calculated diversity values could be an overestimate in some cases.

An analysis based on FADN data of ten countries (EC, 2017) showed that the crop diversification measure has resulted in an increase in the diversity of cultivated crops over 0.8% of the arable area. Compared to this, the obtained results indicate more favorable changes in Hungary.

5. CONCLUSIONS

Crop diversity has an important role in sustainable agro-ecosystems. Due to the changes in the physical block system, a higher aggregation level should be used for further monitoring. 1x1 km grid can provide an effective aggregation level, but parcels' geometry is needed for the calculation, which is not publicly available. Settlements level can also provide adequate information of the spatial differences and the changes over time.

REFERENCES

Aguilar, J., Gramig, G.G., Hendrickson, J.R., Archer, D.W., Forcella, F., Liebig, M.A. (2015) 'Crop species diversity changes in the United States: 1978–2012', PLoS ONE 10(8): e0136580. https://doi.org/10.1371/journal.pone.0136580

Donfouet, H.P.P., Barczak, A., Détang-Dessendre, C., Maigné, C. (2017) 'Crop Production and Crop Diversity in France: A Spatial Analysis', Ecological Economics 134 pp. 29-39. http://dx.doi.org/10.1016/j.ecolecon.2016.11.016

EC (2017) 'Evaluation study of the payment for agricultural practices beneficial for the climate and the environment', https://ec.europa.eu/agriculture/sites/agriculture/files/fullrep_en.pdf

EU (2013) 'Regulation (EU) No 1307/2013 of the European Parliament and of the Council of 17 December 2013', https://eur-lex.europa.eu/eli/reg/2013/1307/oj

Gaál, M., Blanco Mojica, L.F., Zubor-Nemes, A. (2018) 'Spatio-temporal analysis of crop diversity in Hungary', Proceedings of 60th Georgikon Scientific Conference, Keszthely. pp. 105-110.

Monteleone, M., Cammerini, A.R.B., Libutti, A. (2018) 'Agricultural "greening" and cropland diversification trends: Potential contribution of agroenergy crops in Capitanata (South Italy)', Land Use Policy 70: 591-600. http://dx.doi.org/10.1016/j.landusepol.2017.10.038

Weigel, R., Koellner, T., Poppenborg, P., Bogner, C. (2018) 'Crop diversity and stability of revenue on farms in Central Europe: An analysis of big data from a comprehensive agricultural census in Bavaria', PLoS ONE 13(11): e0207454. https://doi.org/10.1371/journal.pone.0207454

USE OF ELECTRONIC IDENTIFICATION AND NEW TECHNOLOGIES ON EUROPEAN SHEEP FARMS

Gautier Jean-Marc[1], Morgan Davies Claire [1,2], Keady Tim W. J.[1,3], Bohan Alan [1,3], Lagriffoul Gilles[1], Ocak Sezen[4], Beltrán De Heredia Ignacia[5], Carta Antonello[6], Gavojdian Dinu[7], Rivallant Pauline[8], Francois Dominique[8]

[1]Institut de l'Elevage, Toulouse, France

[2]Scotland's Rural College (SRUC), Hill & Mountain Research Centre, Kirkton, Scotland, UK

[3]TEAGASC, Animal and Grassland Research and Innovation Centre, Athenry, Ireland

[4]TOGEN, Gaziantep, Turkey

[5]Neiker Tecnalia, Vitoria-Gasteiz, Spain

[6]AGRIS-Sardegna - Research Unit: Genetics and Biotechnology, Olmedo, Italy

[7]Banat's University of Agricultural Sciences and Veterinary Medicine "King Michael I of Romania", Timisoara, Romania

[8]INRA Occitanie-UMR GenPhySE, Toulouse, France

jean-marc.gautier@idele.fr, claire.morgan-davies@sruc.ac.uk, Tim.Keady@teagasc.ie, alan.bohan@teagasc.ie, gilles.lagriffoul@inra.fr, sezenocak1@gmail.com, ibeltran@neiker.eus, acarta@agrisricerca.it, gavojdian_dinu@animalsci-tm.ro, pauline.rivallant@etudiants.purpan.fr, dominique.francois@inra.fr

ABSTRACT

Electronic identification of small ruminants is mandatory since 2010 in Europe. Associated with a context of widespread use of "connected" tools, the availability of solutions using new technologies to manage livestock and decrease workload, should become relevant for farmers. A survey was undertaken in the seven main EU sheep countries (France, Ireland, UK, Spain, Italy, Romania and Hungary) and Turkey (under two European projects: SheepNet and iSAGE) to determine the use of electronic identification (EID) associated technologies and barriers to the uptake of precision livestock farming (PLF) technologies. A total of 1,148 responses were collected and analysed. Sixty four percent of the respondents believe that EID and PLF are an opportunity for better flock/animal management but only 34% of them use it. This survey also highlighted the type of technologies used and the main motivations and barriers for PLF uptake. To date, in the sheep sector, new technologies are mainly related to drafting, animal location, concentrate feed management and performance testing. This is the first study undertaken at EU level targeting the sheep sector. It identifies the main gaps to tackle and proposes some pathways in order to foster the use of new PLF technologies.

Keywords*:* Sheep, precision livestock farming, Electronic Identification, technologies, Europe***.***

1. INTRODUCTION

Precision Livestock Farming (PLF) can be defined as "the 'sensor-based' individual animal approach" (Halachmi et al., 2019), using the principles and technology of process engineering. Precision Livestock Farming has been widely adopted in the management of high-value animals e.g. dairy cattle (Carpentier et al., 2018). Many commercial companies have already developed, or are in the process of developing, PLF applications for intensive farming such as pig (Banhazi et al. 2007), and poultry (Fernandez et al., 2018) production. However, PLF has not yet been applied to animal species considered to have a lower economic value or interest, as is the case in small ruminants such sheep and goats, or in extensive management systems. This is despite the production efficiencies and welfare advantages that may be achieved by applying PLF in these systems.

Sheep systems are vital for the rural economy and society, especially where the climatic and topographical conditions are challenging. However, despite their crucial role for the economy, rural fabric, biodiversity and cultural heritage, the increasing lack of farm labour in these systems is a major issue (Morgan-Davies et al., 2017). Farmers are confronted with increasing pressures to care for a larger number of animals per labour unit to have an economically viable business. This will become more acute in future years. Precision Livestock Farming technologies could alleviate some of the issues e.g. enable remote tracking or behaviour monitoring. Since 2010, all sheep in EU members states are now equipped with EID tags or bolus. The legislation should further pave the way for use of PLF technologies in livestock management. There has been recent research undertaken in sheep systems to enable PLF be included in sheep systems e.g. weight crate for feeding management (Morgan-Davies et al., 2018). However, despite these developments, PLF technology uptake is a major issue (Pierpaoli et al., 2013). In sheep systems, barriers for adoption have been reported (Ruiz-Garcia & Lunedai, 2011), but it is still unclear as to how wide the issue is, and what are the motivations behind farmers' use (or lack of use) of PLF technologies. This paper presents the results of a survey undertaken in the main EU sheep producing countries on sheep farmers' use of PLF technologies and their motivations and barriers for using such technologies.

2. Methodology

Three surveys, based on a pool of common questions, were undertaken to determine the use of the different technologic devices on sheep farms. The first survey was undertaken in the UK by SRUC (Scotland's RUral College) in 2015 and 2016. The second survey was undertaken in early 2018 in France by IDELE through the EU H2020 iSAGE project (https://www.isage.eu). The third survey was undertaken across the seven main EU sheep countries (France, Ireland, UK, Spain, Italy, Romania and Hungary) and Turkey in 2018 through the EU H2020 SheepNet project (www.sheepnet.network).

The SRUC survey was conducted via farmers' interviews during British livestock events. There were 11 questions relating to their farming activities and to their use of PLF tools. Fifty-four answers were collected, of which 26 from the sheep sector.

The iSAGE survey was realised with an online internet questionnaire, on a voluntary basis. Sheep and goat sectors were targeted. There were 29 questions dealing with the general description of the farm, the use of electronic identification (EID reader type, reader trademark, EID connected tool, flock management software, valorisation), opinion on EID (opportunity, constraints and limitations), information about the farmer (age, sex, education level, member to a breeding society). The internet link to the questionnaire was sent to the farmers through different networks: farmer's trade union, levy board organization, technical organization and newsletters, breeding organizations etc. A total of 1,035 surveys were collected for sheep and goats. Only 578 complete answers were retained, of which 471 came from sheep farmers (62 % in dairy and 38 % in meat).

SheepNet used the same survey as used by iSAGE, in the countries involved in SheepNet (France, Ireland, UK, Spain, Italy, Romania, Hungary and Turkey). The iSAGE questionnaire was translated in the different languages of the SheepNet countries and sent to farmers of the different countries through different means of communication. The SheepNet survey was completed on a voluntary basis. A total of 1,315 surveys were collected and 651 were retained. Eliminated answers were mostly from very small flocks or were incomplete.

A total of 1,148 responses across 8 countries were retained and analysed. The data analysis was done with Excel and was mostly a descriptive analysis.

3. Results

Almost 75% of the 1,148 completed surveys came from France (489) and Ireland (350). The remaining surveys originated from United-Kingdom (95), Turkey (92), Spain (60), Italy (40), Romania (18) and Hungary (4). Type of production and flock size were used to describe farm profile.

Three types of production were identified: meat ewes, dairy ewes and multi-purpose ewes, accounting for 61%, 26% and 13% of the sample, respectively. Flock size was classified into six categories: two extremes (less than 49 ewes and more than 1000 ewes) and four centrals (50 to 999 ewes) (Figure 1 a.). Overall, 31% of farms had a flock size between 50 and 199 ewes, 22% between 200 and 349 ewes and 17% between 350 and 499 and 17% had 500 and 999 ewes. The two extreme categories both encompassed 6% of the farms. When looking at flock size and type of production, it showed that meat and multi-purpose ewe farms were generally smaller than dairy ewes (Figure 1 b.), respectively between 50 and 199 ewes vs. between 350 and 999 ewes.

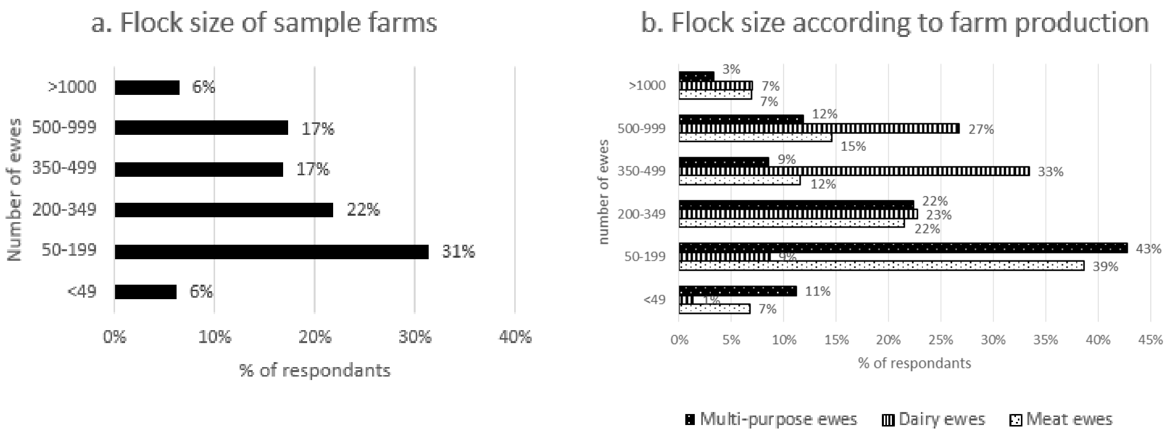


Figure 1. Sample flock size distribution: a. Overall sample, b. By type of production

An important point of the survey was that the majority of farmers (64%) agreed that sheep EID was an opportunity for sheep farming. This vision was shared by all farmers regardless of flock size. The lowest number of positive responses was from the smallest flocks (less than 199 ewes), which had a percentage of “Yes” around 50%. Conversely, 80% of the largest flocks (more than 500 ewes) thought that EID was an opportunity.

Thirty eight percent of European farmers were equipped with tools that utilised the benefits of EID (Figure 2). Four equipped farming profiles were defined: farmers equipped with an EID reader combined with farm management software (17% of farmers); farmers equipped with both an EID reader and equipment using EID (e.g. EID weigh crate, EID automatic feeder…; 14% of farmers); farmers with only an EID reader (4% of farmers); farmers with only digital equipment (3% of farmers). Equipment level changed with the type of production. Dairy flocks were the most equipped (62%), multi-purpose ewes flocks were the least equipped (16%) and meat flocks were intermediate (33%). Level of PLF farm equipment differed by country. Turkey, Ireland and Spain had more than 75% of non-equipped sheep farmers. France and United Kingdom had respectively 55% and 37% of non-equipped farmers. Romania, Spain and Italy had the best equipped farmers with less than 17% of non-equipped farmers.

There was a positive relationship between flock size and level of PLF equipment (Figure 3). The greater the number of ewes the more PLF equipment was available on farm. Less than 10% of farms with less than 199 ewes had equipment that used EID. Conversely, more than 75% of the farms with more than 500 ewes were equipped with EID PLF equipment. Farmers’ age did not influence the level of equipment on farm.

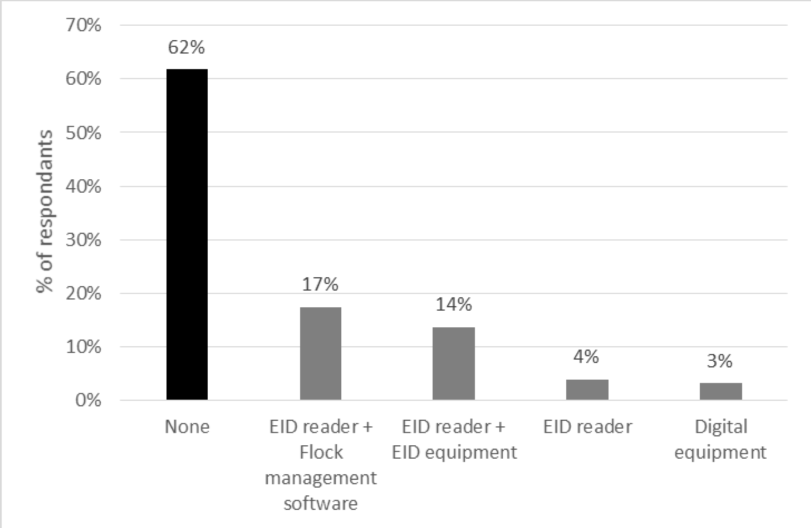


Figure 2. Repartition of equipment level of the farms surveyed

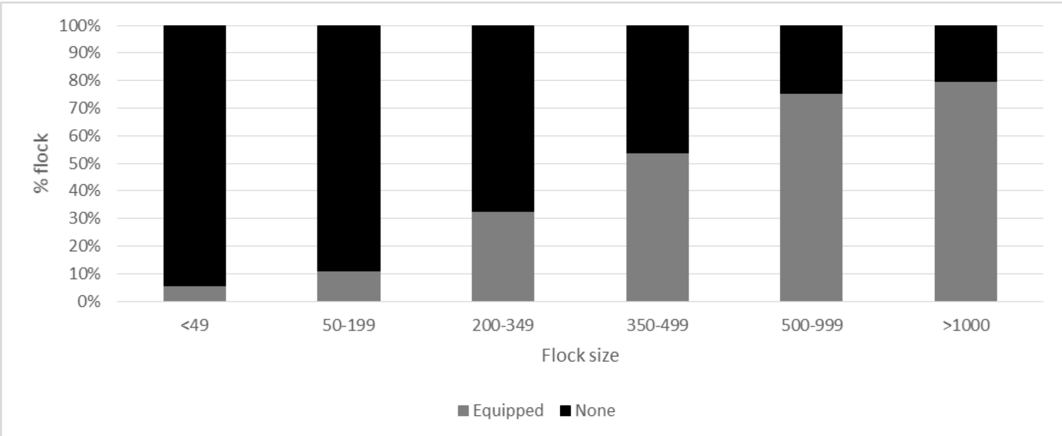


Figure 3. Equipment level according to flock size

Two major uses of EID were identified: recording animal movement (between 60% and 80% of farmers, depending on type of production) and sorting/drafting sheep into management groups (45% to 55% for meat ewes and dairy ewes). Electronic identification was used for automatic recording of animal weights in meat ewe production (40%) and for recording pedigree information for breeding in multi-purpose ewe production (almost 40%).

Precise flock management with the ease of recording flock data for management purposes, time saving and labour simplification were the three main motivations for PLF equipment. Modernisation, data sharing and welfare improvement did not appear to be triggers in the investment in PLF equipment. The main reasons for not using EID were rank by farmers per order of importance. The first one is the cost of equipment, second the flock size and third the accessibility to the equipment.

4. Discussion

This survey was undertaken in the seven main EU sheep producing countries and Turkey and thus is representative of the variation in flock size and the many different sheep production systems practiced in the EU. This survey shows global trends that can help to better understand motivations and constraints to the uptake of digital technologies.

Since 2010, EID for small ruminants has been mandatory in the EU. The aim of the EU regulation (n° 21/2004) was to ensure better traceability of all small ruminants during their lifetime. In practice, when a farmer sells animals, he has to complete a transportation document with all the official animal numbers. In the early 2010s, most farmers saw EID only as an EU regulation constraint with an extra cost (Holtz, 2015 - unpublished). This survey shows that almost 10 years after the enforcement of this EU regulation, mentalities seem to have changed, as 60% of the respondents saw EID as an opportunity. Since 2010, the democratisation of the smartphone technologies and the increase of digital tools might have helped farmers to better grasp the potential of digitalisation on their farm. This, somewhat, more positive farmers' vision on EID could have an impact on the use of new management equipment and software and on development's orientations for sheep farming. But despite a favourable context, and a positive farmers' vision, equipment level is still low and only 38% of the European farmers surveyed were equipped with tools enabling the benefits of EID. The level of PLF equipment on farm increased with flock size. Seventy five percent of farms with more than 500 ewes had PLF equipment that utilised EID.

The equipment rate of European farmers surveyed differs between dairy and meat production systems as there was a higher level of PLF equipment on dairy farms (62% vs 33% on dairy and meat farms, respectively). This difference could be partially explained by a higher flock size on dairy farms (34% and 22% of dairy and meat farmers had more than 500 ewes, respectively) and a higher income form dairy relative to meat production. Furthermore sheep dairy farmers are probably more aware of technologies used in the (PLF strong) dairy cow industry. In terms of equipment, 87% of equipped farms have an EID reader either linked or not linked with EID PLF equipment. Forty percent of the readers used were a stick reader, probably due to its relatively low price (between 700 to 900€). Forty nine percent of farmers equipped with an EID reader combined it with flock management software. Only 17% of the respondents had EID PLF equipment. The equipment differed between dairy and meat production farms. For dairy production, auto-feeders in the milking parlour represented 58% of the equipment, followed by milk counters (30%). For meat production, weigh crates represented 74% of the equipment on farms. The main obstacle to the equipment of farms remained the high cost of EID readers/equipment. Flock size ranked second and could be related to the cost/benefit of the investment. The accessibility of the material, lack of support and communication were considered the third, fourth and fifth most important barriers, respectively, to the uptake of EID PLF equipment.

The use of EID by farmers is still very limited and is heavily dependent on farm system and type of equipment on the farm. To date, EID use remains mainly for the recording of animal movements as well as sorting sheep into management groups. This could be explained by the regulations and needs to document animal movements. Conversely, the recording of ultrasound results, health data and the monitoring of mating, animal performance and ancestry data are still undervalued. Finally, the use of EID is still strongly related to its first (mandatory) purpose, to ensure traceability.

5. Conclusions

This study helps to better understand motivations and constraints for digital technologies uptake by EU sheep producers Sheep production is in a favourable position to develop PLF (with widespread use of EID and expansion of new digital technologies such as smartphone), as 60% of the farm surveyed considered EID as an opportunity. Yet, only 38% of farms were equipped with tools that maximise the benefits of EID. Overall, dairy farms were better equipped than meat farms and some differences appear between countries. The level of equipment on farms depends on flock size, with nearly 75% of farms with more than 500 ewes having some EID equipment. However, cost of technology is still the largest barrier to it uptake on sheep farms. The benefits of EID remain mostly limited to the management of animal movements, which is mandatory. To promote a better use of EID equipment, research on the approach on cost/benefit of investments should be carried out as well as a better communication on the possible benefits.

Acknowledgement

All farmers who have participated to the different surveys for this study.

References

Banhazi, T., Dunn, M., Cook, P., Black, J., Durack, M. and Johnnson, I. (2007) 'Development of precision livestock farming (PLF) technologies for the Australian pig industry', In 3rd European Precision Livestock farming Conference, 1, pp 219-228.

Carpentier, L., Berckmans, D., Youssef, A., Berckmans, D., Van Waterschoot, T., Johnston, D., Ferguson, N., Earley, B., Fontana, I., Tullo, E., Guarino, M., Vranken, E., and Norton, T. (2018) 'Automatic cough detection for bovine respiratory disease in a calf house', Journal of Agricultural Engineering Research, 173, pp 45-56.

Fernandez, A. P., Norton, T., Tullo, E., Van Hertem, T., Youssef, A., Exadaktylos, V., Vranken, E., Guarina, M. and Berckmans, D. (2018) 'Real-time monitoring of broiler flock's welfare status using camera-based technology', Journal of Agricultural Engineering Research, 173, pp 103-114.

Halachmi, I., Guarina, M., Bewley, J., and Pastell, M. (2019) 'Smart animal agriculture: Application of real-time sensors to improve animal well-being and production', Annual Review of Animal Biosciences, 7; *In press*.

Holtz, J., Gautier, J.M., Lefrileux, Y. and Caramelle-Holtz, E. (2016). 'Use of electronic identification and new technologies in French goat farms'. IGC, Antalia, 26-28 septembre 2016.

Morgan-Davies, C., Wilson, R., and Waterhouse, A. (2017) 'Impacts of farmers' management styles on income and labour under alternative extensive land use scenarios', Agricultural Systems, 155, pp 168-178.

Morgan-Davies, C., Lambe, N., Wishart, H., Waterhouse, A., Kenyon, F., McBean, D., and McCracken, D. (2018) 'Impacts of using a precision livestock system targeted approach in mountain sheep flocks', Livestock Science, 208, pp 67-76.

Pierpaoli, E., Carli, G., Pignatti, E., and Canavari, M. (2013) 'Drivers of precision agriculture technologies adoption: A literature review', Procedia Technology, 8, pp 61–69.

Ruiz-Garcia, L. and Lunadei, L. (2011) 'The role of RFID in agriculture: applications, limitations and challenges', Computers and Electronics in Agriculture, 79, pp 42–50.

A SERIOUS VIDEO GAME FOR SMART FARMING TECHNOLOGIES

Spyros Fountas[1], Zisis Tsiropoulos[2], Panagiotis Stamatelopoulos[2], Evangelos Anastasiou[1], Tim Hutzenlaub[3], Mladen Radišić[4], Vladan Minic[4], Patrick Rau[3]

[1] Agricultural University of Athens, Iera Odos 75, Athens, Greece;

[2] AGENSO, Markou Botsari 47, Athens, Greece;

[3] Mad about Pandas, Kottbusser Damm 73, 10967 Berlin, Germany;

[4] Inosens, Vojvode Misica 13 21000, Novi Sad, Serbia

sfountas@aua.gr

ABSTRACT

GATES is a near-to-market (Technology Readiness Level 7) serious game-based training platform that, through the use of a range of gaming technologies (3D scenarios, interactive storytelling, modelling and data), will train professionals and other stakeholders in the value chain in the use of Smart Farming Technology. It is a cross-platform (Desktop/Mobile/Web) available for Android and Windows featuring online and offline synchronized modes. GATES features learning and behavioural triggers for stimulating players' engagement, creativity and collaborative behaviours, while caring for gender and environmental issues related to the use of Smart Farming Technologies. In the current version, the player can choose between five crops (wheat, corn, potato, apples and vineyards) and can select one of multiple field operations, such as irrigation, fertilization, spraying and harvesting. The outcome of this serious game is a set of economic and environmental output of the selection of specific choices by the users. The gaming platform will be marketed as a white-label app, and will function either as a stand-alone or as a complementary tool to traditional training methods, covering a wide range of agricultural settings for the needs of different professionals in the Smart Farming Technologies value chain.

Keywords: smart farming, precision agriculture, gamification, H2020

1. INTRODUCTION

Smart Farming Technologies (SFT) are related to more efficient application of inputs (seeds, fertilisers, chemicals, water, fuel and labour), increased work speed and comfort and enhanced flexibility on-farm. Europe plays a dominant role in the global production of SFT, with nearly 30% of the global production output of agricultural machinery in 2014 having been produced by companies within the European Union. The European agricultural machinery industry includes 4,500 manufacturers (with over 95% SMEs) employs 250,000 people and in 2015 produced a value of 25 billion Euro (CEMA, 2014). SFT can contribute to the wider goal of meeting the increasing demand for food, feed, and raw materials while ensuring the sustainability of primary production; however, the fast pace of development of these technologies, their complexity and the lack of expertise of end-users have hindered its adoption (JRC, 2014).

The increasing pace of development of SFT over recent years, the costly equipment needed for its implementation and the wide range of existing solutions for addressing the increasing complexity of agricultural systems in a context of climate change, indicates the need for innovative and integrated training approaches in order to allow end-users to fully tap the benefits of SFT. Traditional training approaches (seminars, self-training material, technical simulators and on-line courses) have proved to be useful for training professionals on the benefits of SFT. On the other hand, they are not always scalable and cost-effective. In addition, the adoption of smart farming has not reached high level numbers, as it was expected due to many reasons that have been addressed by many studies (i.e. Fountas et al., 2015; Barnes et al., 2019).

Gamification is an approach for training professionals and has demonstrated potential to improve many aspects of how businesses provide training to staff and communicate and engage with consumers, boosting knowledge and loyalty among customers, channel partners and other stakeholders. Game-based learning has been demonstrated to be of great efficacy to comprehensively absorb and practice complex concepts and procedures under simulated working conditions, particularly in fields where workers have to learn how to use expensive and/or technologically complex equipment, such as SFT. Game-based learning increases participant's interest while making the training process more enjoyable, memorable and effective based on an engaging experience and a safe environment for failure.

Learning through play is a well-known concept. Educational computer games, underpinned by instructional goals and appropriate application of game mechanics, have considerable potential for training adults through their ability to engage, motivate and influence the behaviour of learners. It is the attributes of games, such as compelling storylines, attainable challenges, rewards, recognition, controlled and a safe environment to fail and try again that make them so powerful for learning (Groff, et al., 2010; Klopfer et al, 2009). Well-designed serious (educational) games make learning fun, challenging and rewarding. In terms of suitability for learning, there are significant parallels between many of the learning theories applied to the design of e-learning courses and the principles of game design.

The objective of this serious game was to develop a serious game-based training platform that, through the use of a range of gaming technologies, will train professionals and other stakeholders in the value chain in the use of Smart Farming Technology. GATES game aimed to develop a serious gaming cross-platform (Desktop/Mobile/Web) available for Android, iOS and Windows featuring online and offline synchronized modes.

2. Methodology

The development of the serious game followed the minimum viable product or Minimum Viable Game (MVG) approach, where three iterations were created. This general agile methodology that was used for the development of the MVG is a mix of Scrum and classical sequential project management ("waterfall"), which allows for an iterative process with the addition of having a more feasible control of the workload. Scrum is an iterative process, where the ability to respond to requirement volatility and unpredictable changes in an efficient manner is a key principle. An Iteration (or sprint) is a defined time frame, where a product increment is achieved. Advantages of this agile approach are among others, the flexibility in the development process, a high level of engagement and communication between the team members and good transparency of the team's velocity through burn-down charts. The three MVGs were evaluated by end-users (farmers, students, farm machinery professionals) in Spain, Serbia and Greece following a semi-structured questionnaire. For the development of the GATES game, the Unity3D Engine was chosen. Unity3D Engine provides a full - fledged engine for creating games but also an easy way of deploying the created game on several different platforms (web, android, iOS).

3. Results

The GATES training platform is made as a single player game with online capabilities. At the main menu, the player is able to start the game immediately, either creating a user's profile or without registering so as to be logged in online automatically. On demand, the user is able to register the account and associate a user name and password. By doing that, the player is able to save and resume the game, whenever they want. The “Main Story” game mode is split into stages, with each one of the stages focusing on increasing player’s awareness about the existing SFT and/or benefits deriving from their application. Each stage starts with a mini tutorial informing the player about the stage tasks, while the tutorial character Trektor (Figure 1) informs players during the stage progress about the effects of each of their actions in form of info tips.

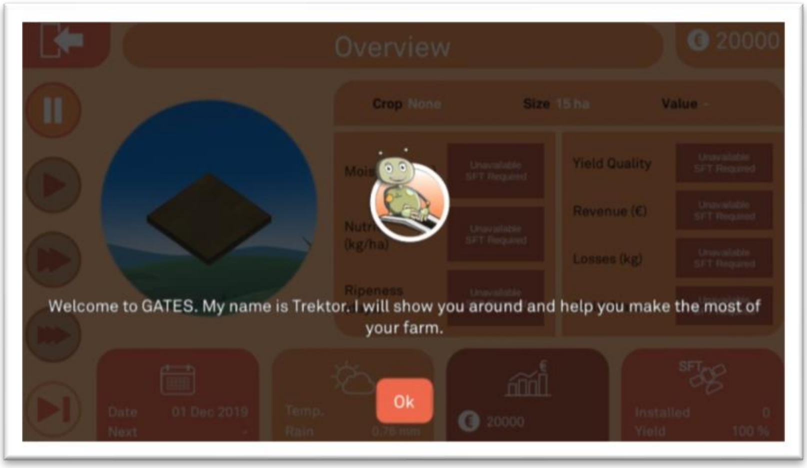


Figure 1. Welcome page with the Trektor instructor

GATES consists of a number of modules that are important to run the game. **The Farming Module** is the core of the GATES serious gaming platform. All the game modes use the farming module as their main playing environment, which contains all the modelling and algorithms needed for the GATES game mechanics. Through it, a representation/simulation of the real world is done, by running the necessary models for defining the negative or positive effects of each type of in-game interaction (e.g usage of a smart farming service by a farmer), or event (e.g. rain). The Tutorial module is an interactive in-game tutorial which helps players to perform required actions, resulting in better retention. GATES consists of an initial tutorial in which in the first minutes, users will become familiar with the User Interface and will complete their first tasks by purchasing SFT equipment/services (e.g. Variable Rate Seeder) and by applying it in the field (first seeding operation). **The Statistics module** illustrates complex information in a simple manner and allows the game to showcase the authentic effects of SFT in a comprehensive and precise way. This module explains the actions taken by the user by visualizing their effects, allowing players to investigate and analyze their actions and to change their gameplay in order to improve their performance and reach their goals. Finally, the Logger module is used by the GATES platform for improving the game itself and for providing better training results to the users.

In addition, the **Library module** is an in-game educational content, which provides to the players the basic knowledge needed for completing game tasks. The online wiki library provides extensive information on each aspect related with SFT, not only for helping players completing the various game tasks but also for providing an educational tool able to help them in deepening their knowledge on SFTs. The wiki page material is available in the online form, showing documents, presentations, videos, images, diagrams, and links. **The Scenario Creator module** allows users to create custom scenarios defining winning conditions, the SFTs usage, country, area, amount of money, etc (Figure 2). The level of the scenario creator customization ability will vary. Users with the lowest fee will be able to create simple scenarios and to invite friends to play, while users with the highest fee, will be able to create tutorial character info tips, upload 3D models, upload education material, create puzzles and create interactive guides. Moreover, users with the highest fee will have the option to evaluate the invited players’ performances in their created scenarios through the Logger Module. This will help them to find the weaknesses and the strengths of their invited users with respect to the scenario tasks and SFT displayed. The “Scenario Creator Module” will be available to Premium users according to the game Business Model.

Figure 2. Story scenarios selection

Moreover, the **SFT 3D Interactive module** is using GATES 3D engine. This module will simulate the real usage of certain SFTs in the 3D environment and will be used for teaching and preparing players to operate SFT equipment. The players will be able to see and interact with SFT interfaces and they will be able to create interactive guides (using hints and instructions) and puzzle development (Figure 3). For example, puzzles will start with very simple tasks (e.g. turning the machine on and starting it with default settings) with the difficulty to be increased gradually (e.g. player must make finer adjustments over the course of the game, optimizing its usage. Moreover, through this module, a virtual showroom with 3D-Models of various SFT equipment will be developed. Finally, **Social Module** has its own pages on Facebook and LinkedIn Social Networks, which is connected to the game through and Application Programming Interfaces (APIs). Besides regular activities that can be done on such networks they will also be used for triggering competition and engagement behaviors: e.g. when a player will achieve a learning goal and earn the "badge" for the particular SFT, a post will appear on the social networks mentioning her/his progress.

The GATES game environment is user friendly and easy to be played by the end-users with limited training and explanatory time and presents the main information that a person working with agriculture needs (Figure 4). These features are related to soil moisture, maturity level and nutrition needs. The outcome values of the game are yield, quality, field status and revenue. On the bottom of the main screen, can be seen the cultivation calendar, the current weather information from the already installed weather information from previous years, the total revenue, as well as the number of SFTs already installed in the game. In the picture window, the different operations are illustrated, such as soil cultivation, seeding, spraying, irrigation, fertilization or harvesting.

Figure 3. Farm machinery equipment with SFT systems in 3D environment

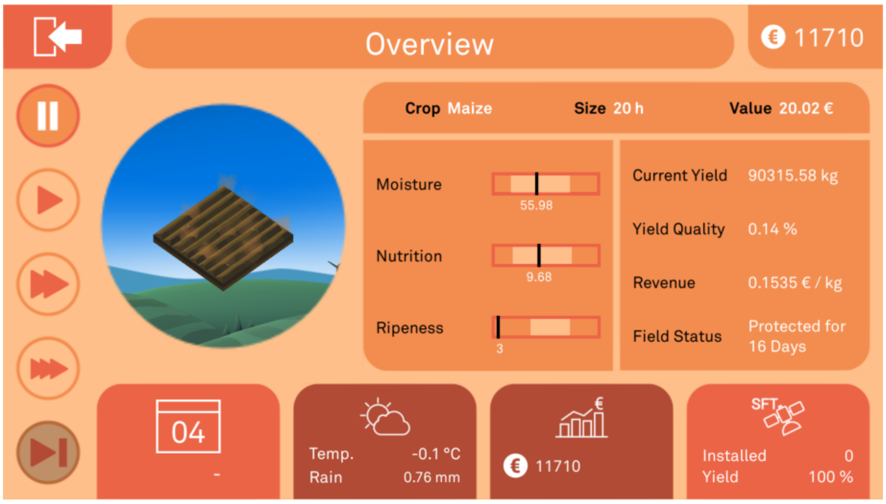


Figure 4. GATES main overview functions

In the evaluation stage of the project, in total 750 users participated in the three stages of validated learning cycles, but the number of users included in the workshops, presentations and direct contact was much higher. From those who fulfilled the surveys, this is the distribution by type: (Farmers: 253; Students: 297; farm machinery specialists: 204). The satisfaction level was increased from MVG1 to MVG3 and among all modules the 3D interactive module was perceived as the most useful and entertaining to play with.

4. Discussion and conclusions

The main value proposition of GATES is that it provides a serious game-based training platform, making use of different gaming technologies, in order to train professionals across the agricultural value chain on the use of Smart Farming Technology, thus allowing deploying its full economic and environmental potential in European agriculture. As for its main benefits, GATES offers a convenient online way for farmers, agricultural engineering students, machinery manufacturers, AgTech businesses, SFT consultants, farmer cooperatives, etc., to obtain knowledge related to the latest developments in smart farming technologies. Thus, this concept is effectively turning GATES into a repository of information regarding SFTs and a one-stop-shop for interactive learning about their positive impacts on agricultural production. GATES offers its platform as an appealing and fast interface needed to browse through a variety of SFT solutions and find the most suiting one.

A large number of end-users, who have tested the game commented that language selection is very important, as most farmers are not used in working with English terminology for their farming businesses and they would like to see an instructor/tutor in most scenarios helping them to choose the most appropriate SFTs for their fields. The lean start-up methodology with the three iterative cycles was an extremely important process for the future success of the whole GATES concept, which connected the developers with the testing communities counting several hundreds of users. The feedback collected and the process itself presents valuable points of direction where the final version of the game should lead to.

5. ACKNOWLEDGEMENTS

Results incorporated in this work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732358.

REFERENCES

Barnes, A.P., Soto, I. Eory, V., Beck, B., Balafoutis, A., Sánchez, B., Vangeyte, J., Fountas, S., van der Wal, T. Gómez-Barbero, M., 2019. Exploring the adoption of precision agricultural technologies: A cross regional study of EU farmers. Land Use Policy, 80, 163-174.

CEMA, 2014. European machinery sales. http://www.cema-agri.org/publication/european-agricultural-machinery-sales-decline-5-2014

Fountas, S., Blackmore, S., Ess, D., Hawkins, S., Blumhoff, G., Lowenberg-Deboer, J., Sorensen, C. G., 2005.Farmer Experience with Precision Agriculture in Denmark and the US Eastern Corn Belt. Precision Agriculture, 6, 121-141.

Groff, J., Howells, C., Cranmer, S., 2010. The impact of console games in the classroom: Evidence from schools in Scotland. Futurelab, UK

JRC. Eds. Pablo J. Zarco-Tejada et al. 2014. Precision Agriculture – an opportunity for EU farmers, Potential support with the CAP 2014-2020.

Klopfer, E., Osterweil, S., Salen, K., 2009. Moving learning games forward. Cambridge, MA: The Education Arcade.

INNOVATIVE AGRIBUSINESS IN GREECE

Avraam Mavridis[1,2*], Maro Vlahopoulou[2], Athanasios Gertsis[1]

[1]Department of Agroenvironmental Management Systems, Perrotis College, Greece & Department of Applied Informatics, University of Macedonia, Greece

[2]Department of Applied Informatics, University of Macedonia, Greece

amavri@afs.edu.gr, mavla@uom.edu.gr, agertsis@afs.edu.gr

ABSTRACT

The heterogeneous framework of activities in the agricultural sector requires efficient, dynamic and holistic approaches in fair and equitable ways of management, cultivation practices and education, research, trade and policy strategies. Such demanding, interdisciplinary applications can be operated extensively today, more than ever before through Exploitation of Informatics. Such activities require a secure, yet powerful framework of hardware and software facilities in order to unlock the potentials of innovative agribusinesses for all agricultural products' value chain parties and end-users, while aiming to support and develop benefits of agro-environmental sustainability. Latest accomplishments in Earth Observation technologies, Geoinformatics, AI (Artificial Intelligence), IoT (Internet of Things), Smart Farming Sensors, Precision Agriculture, Digital Agriculture, Cloud Services and Modular Agricultural Robotics constitute a fruitful region of continuous development. Current approaches worldwide are trying to develop new agricultural curriculum developing the educational background of the scientist who will support such activities: the data agronomist. Additionally, new digital skills are required to be developed by all actors participating in this framework: the farmers, the agronomist, the researcher, the academics, the consumers, the entrepreneurs and the policy makers. This paper will approach and present the level of exploitation and incorporation of different Information Technologies in the Agricultural Sector of Greece towards the development of innovative agribusiness opportunities, regardless of the gender, age, or/and educational background. Such innovative agribusiness potentials will be based on current available data repositories and web-based networking, while providing elements of further advancements in the near future, so as to combat restraints of economic and environmental crisis in the country.

Keywords*:* logistics Earth Observation technologies, Geoinformatics, Artificial Intelligence, Modular Agricultural Robotics, Smart Farming Sensors, Digital Agriculture

1. INTRODUCTION

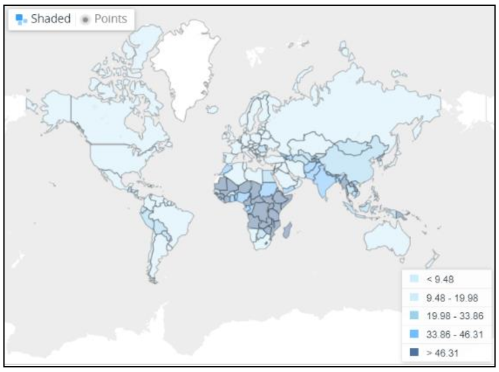
Agriculture, as we know it today, evolved through initiatives and activities at individual, collective and institutional level worldwide, starting 11,000 years ago, and it is continuously evolving, providing benefits to all sectors of economy with great impact to natural and socio-economic frameworks. The Agricultural framework has a very important role in the economy of countries at global scale, and is the main source of food, income and employment to their populations. At the same time, it is facing continuous economic, technological and regulatory pressures and changes, considering the fact that at the beginning of 2015, Agriculture accounted for one-third of global Gross Domestic Product (GDP). As worldwide population is estimated to increase over 9.5 billion people by 2050, there is a augmented need to increase the agricultural activities in order to produce 50% more food, feed and biofuel than it was needed in 2012 (Table 1), so as to meet the demand (Alexandratos et al., 2012).

Table 1. Increase in agricultural production to match projected demand, 2005–2050 (percent)

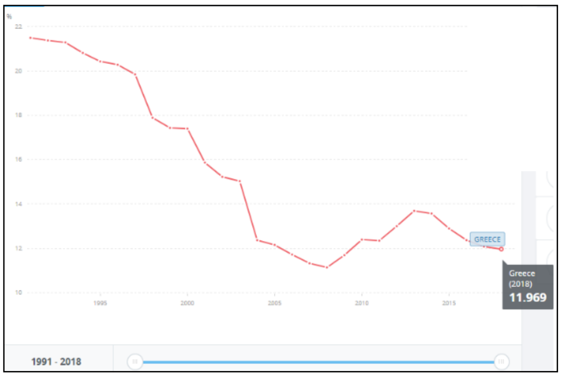
	2005/07	2050	2005/07 2012	2013-2050
World				
As projected in AT2050	100	159.6	14.8	44.8
With updated population projections	100	163.4	14.8	48.6

Source: FAO 2017, based on UN 2015, and Alexandratos and Bruinsma, 2012

Businesses related with Agriculture (in short, Agribusinesses) were of worth of more than 3.5 trillion US dollars (World Bank, 2019) at the beginning of 2017 globally and Agriculture was occupying over one-third (1/3) of total land area (FAO, 2018). Additionally, labor characteristics in agriculture are confusing as they constitute a mixed operational framework which it may be formal and informal, characterized by high levels of multi-activities and seasonality and also, it can be self-employed or unpaid, and/or supplied by family members. Under these perspectives and, as income grows and new technologies are introduced, labor in agriculture is continually reducing, occupying in 2018 more than 1 billion people (28.3% of the world's total employment, from 43.77% in 1991), with significant variations between developing and developed countries (Fig. 1), that stands also for the agricultural sector in Greece (Fig. 2) (World Bank, 2019).



1



2

Figure 1. Classification of employment in agriculture worldwide in 2018 (% of total employment)

Figure 2. Fluctuation of employment in agriculture in Greece (% of total employment) from 1991 up to 2018, with the comprehending percentage (11.969%) in 2018 (Source: World Bank, 2019)

Agriculture provides public goods, nutrition and labor worldwide, but it also contributes in severely degradation of agro-environmental resources and ecosystem services, as well as of humans' health. An additional point of interest is the fact that Agriculture both contributes to climate change and is affected by climate change into a framework of increasing scarcity of natural resources (such as arable land and water). However, the overall agricultural sector should be approached as a multi- and inter-operational framework, where "industry convergence" is applied with sectors that exist beyond the traditional fiber or food industries, which are based on raw products established by agricultural activities (such as bioenergy, bio-based plastic sector and pharmaceutical industry) (Boehlje et al, 2011), enhancing final demand for agricultural products in the medium and long-term.

In Greece, Agriculture is a major sector for the country's economy, comprising 2.9 per cent of GDP and 14 per cent of employment. European Union (EU) has an average of 1.2 per cent and 5 per cent, respectively. In Greek economy, in terms of contribution towards the GDP, we have the following sectors (and they can all be related with agricultural activities) (EU-MERCI project, 2018):

i) **Primary**: agriculture, 3.4%, 2) **Secondary**: industry, 20.8%, 3) **Tertiary**: services, 75.8%.

In addition, according to the future directives of the United Nations' (UN) Sustainable Development Goals (UN, 2015), supported and adopted by EU (European Commission, 2016), in front of the increasing scarcity of natural resources (such as arable land and water) and other socioeconomic burdens focused within the 17 SDGs, the agribusiness sector needs to change, pushing the agricultural-related productivity and activities through "industry convergence" to a new level.

New frontiers of agribusinesses should be connected with sustainable economic development, which is very much based on technological progress and innovation. As innovation-based industries create - in most cases - greater added value in an economy, they create products, services and metadata of extra value that are generally more efficient to manage, to produce and supply, being more exportable and offer more well-paid jobs and welfare throughout the horizontal and vertical agricultural-related chains and trade channels.

Future agricultural-related activities in Greece should incorporate innovation within, developing improvements and opportunities in the overall agricultural framework, focused on successful attainment of UN's SDGs, which are fundamental to achieving food security, poverty alleviation and welfare, as well as conservation of environmental resources and sustainable development.

2. Materials and Methods

As approaches to knowledge exchange, learning and innovation in agriculture and related activities are rapidly evolving, establishing a clear knowledge about current operation conditions and agricultural-related activities of innovative agribusinesses in Greece requires a "top-down / bottom-up" approach. Such consideration should be elaborated to all specific agricultural activities/practices, as well as towards horizontal and vertical agricultural-related chains and connections of industrial convergence sectors with Agriculture in Greece. According to FAO (2019): "*Agricultural innovation is the process whereby individuals or organizations bring new or existing products, processes or ways of organization into use for the first time in a specific context, to increase effectiveness, competitiveness and resilience with the goal of solving a problem*".

Under these perspectives, the following process will be followed in order to approach the overall current and future framework of the Innovative Agribusinesses in Greece (Fig.3): 1) Analysis of current Types of Agriculture and its Sectors in Greece, 2) Understanding of current and future trends of Innovation in agricultural-related businesses in Greece, 3) Presentation of best practices and approaches of Innovative Agribusinesses of EU, EU Member States and of organizations in Greece.

2.1 Analyzing current Types of Agriculture and its Sectors in Greece

Agriculture is related with many activities within the farm, as well as with other operational sectors of the industry which are widespread in different operational categories worldwide, but without uniformity throughout. One approach supports the fact that Agriculture is differentiated according to human intervention and mode of study of the surrounding nature. However, different patterns of agriculture which have been developed through centuries and especially during the last decades, can be classified into other types of agriculture through many and different ways. Some of the major criteria to be used for this scope include (Chandra, 2011):

i. Scale, ii)Type of crop, iii) Livestock combinations, iv) Intensity, v) Means of distribution of farm produce, vi) Level of mechanization/technologies used,

Agriculture sector is mainly divided into the following four sub-sectors were innovative technologies can be applied throughout the value chain:

a) Crops, b) Livestock, c) Fishery (Aquaculture), d) Forestry (Silviculture).

Agricultural sector covers the activities related to: a) Growing crops, fruits & vegetables, b) Harvesting & Threshing, c) Growing of trees & logging, d) Fishing & Breeding, e) Rearing of animals and poultry, f) Production of milk, eggs, etc.

Another approach (Bourbos et al., 1996) divides agriculture (overall agricultural Area: 3,254,078.9 hectares (HSA, 2019)) into three main categories:

a) **Natural**, b) **Conventional** (known also as Intensive, or Agrochemical), and c) **Sustainable**, including: i) Integrative (with Precision, Smart and Digital Agriculture), ii) Ecological (with Traditional, Organic, Biodynamic, Organic-biologic and Precision Organic Agriculture (Mavridis, 2008)). At early 2018, the total organic area (the sum of the "area under conversion" and the "certified area") in Greece was 410,140 ha.

The country covers an area of 131,621 km², of which 82.2% is rural, supporting the living of the 44.1% of the population (10,816,286 citizens). Over half of the country's 723,010 agricultural holdings are less than 2 hectares. Small and fragmented land parcels constitute one of the main characteristics of Greek agriculture, while the lack of skilled workers (only 3.5% of all farm managers have agricultural training) is a problem that innovation approaches should overcome.

Innovative Agribusinesses are a key challenge for such agricultural-related activities, as they could be based on the establishment of an increasing, skilled workforce with younger farmers (currently only 12.6% of farm managers are less than 35 years old), who are more receptive on new technologies.

2.2 Innovation in agricultural-related businesses in Greece

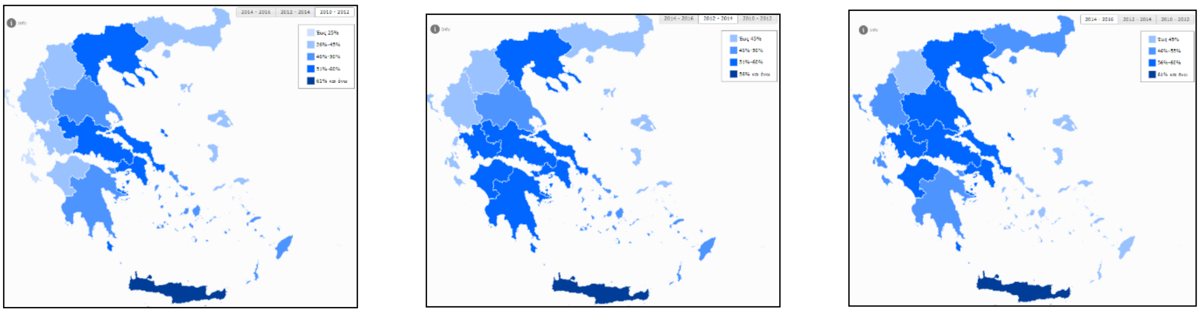
Digitization of agricultural activities in plant and animal production can improve in many ways the working conditions and welfare for farmers and their families, while reducing the environmental impact of agricultural practices, providing useful information to the end-user (e.g., the consumer). These technologies can offer significant benefits on aspects of remote measurement of soil conditions, on provision of better water/irrigation management and crop monitoring in Precision Agriculture. In Precision Livestock Farming (PLF) (Figures 3a, 3b & 3c), they can provide significant benefits for better nourishment and health of animals, by monitoring heat stress and their activity in the field through GPS (Global Positioning System) and better quality and quantity of milk and meat.

3a 3b 3c

Figure 3 (a, b & c). DigitAnimal (https://digitanimal.com) enterprise can select and analyze data throughout the day with collars (3a), showing cows' readiness for insemination, and (3b) monitor daily activity from the office/home, or through a smartphone (3c) through GPS, in real time.

Further exploitation of selected data (spatial and non-spatial), which can be provided in real-time and continuously transferred to a mainframe, can give to farmers/entrepreneurs/end-users better understanding of the variability of soil elements into their farm, or the development of crop patterns, or variations of animal's activity through the day, so as to apply more efficient practices. Towards that scope, Smart Farming Technologies constitute an operational sector of increased capabilities and augmented synergies. Such technologies in agriculture include: i) Satellite Monitoring, ii) Soil/Plant/Weather Sensors, iii) Mobile devices with specific farm apps, iv) Smart Zone Seeding technologies, v) Autonomous - Modular Robotics, vi) Weather Modelling, vii) Smart Micro Irrigation, viii) Fertilizer Modelling, ix) "On-the-spot" and "On-the-fly" spraying systems, x) Internet of Things, xi) UAVs/UAS (Unmanned Aerial Vehicles/ Unmanned Aerial Systems).

We have to consider the fact that all these technologies could be related with the industry and the society in a developing way, advancing new operational schemes and opportunities. However, innovation in Greek enterprises, in the last years, has shown significant spatial variation (Fig. 4):



4a **4b** **4c**

Figure 4 (a, b & c). Differentiation in adoption of innovation activities by Greek enterprises through the years: 2010-2012 (4a), 2012-2014 (4b) and 2014-2016 (4c) (Source: NISRT, 2019).

2.3 Best practices and approaches of Innovative Agribusinesses

Among numerous forces at play across the innovation in agribusiness value chain, certain practices/examples discriminate among others in adopting and sharing innovation in Agriculture:

i) At EU scale: Pioneering activities are promoted by EU through the European Innovation Partnership for Agricultural productivity and Sustainability (**EIP-AGRI**) (https://ec.europa.eu/eip/agriculture/en/about), contributing to the European Union's strategy 'Europe 2020' for smart, sustainable and inclusive growth. Among them is a catalogue with innovative projects presented in EIP-AGRI events. Among them we can distinguish the project **SMART-AKIS** (https://www.smart-akis.com), a European Network mainstreaming Smart Farming Technologies among the European farmer community and bridging the gap between practitioners and research on the identification and delivery of new Smart Farming solutions. Also, the project **INNOVARURALE** (https://www.innovarurale.it), the Italian Portal of knowledge and innovation in the agricultural and food system

ii) At National (Greek) scale: a) A startup enterprise (**CENTAUR**) (https://centaur.ag/) solutions enable end-users to maintain and trace pristine product condition across the supply chain -- from harvesting, to storage, to milling, to retail, b) The Aristoleo® Test Kit (https://aristoleo.com/) measures the health promoting phenolic compounds oleocanthal (anti-inflammatory) and oleacein (antioxidant), the most widely researched phenols in olive oil. This "lab in a vial" is calibrated by NMR (Nuclear Magnetic Resonance), being quick and inexpensive for olive oil producers to analyze their batches of EVOO (Extra Virgin Olive Oil).

3. Discussion and Conclusions

Innovation in Agriculture-related businesses in Greece can become a significant asset for the future of Greek Agriculture, in connection with the international operational framework. Aiming at this goal, innovative agribusinesses should be transformative and adaptive towards future trends and challenges transforming the Agribusiness Industry such as: 1. Sustained low commodity prices, 2. Improving crop/animal yields, 3. Increasing focus on sustainability, 4. Growing investment in Agricultural Technology, 4. Transformation of the production methods and tools, 5. Adaptation of IoT (Internet of Things) requirements: connectivity, robustness and legacy technologies, 6.Dealing with legacy technology, 7. Transparency of productions, 8. Promotion of services, 9. Operation within connected ecosystems, 10. Consolidation of agribusiness subsectors, 11. Take under consideration water limitations and labor shortages, 12.Potential changes to established global trade agreements

REFERENCES

Alexandratos, N. & Bruinsma, J. (2012). 'World agriculture towards 2030/2050: the 2012 revision'. ESA Working Paper No. 12–03. Rome, FAO (Food and Agriculture Organization).

Boehlje, M., Roucan-Kane, M., & Bröring, S. (2011). 'Future agribusiness challenges: Strategic uncertainty, innovation and structural change'. International Food and Agribusiness Management Review,14(5), 53-82.

Bourbos, B.A., Skountridakis, M.T., (1996)."Conventional and Sustainable Agriculture, Development, Capabilities and Restrictions". Proceedings of Greek Conference on "Organic Agriculture Problems and Perspectives", February 28-29, 1996. Chania, Crete. Greece.

Chandra, D. (2018). 'Types of Agriculture'. (Website: https://owlcation.com/agriculture/Types-of-Agriculture).

European Commission (2016). 'Next steps for a sustainable European future: European Union action for sustainability'. (Website: https://ec.europa.eu/europeaid/sites/devco/files/swd-key-european-actions-2030-agenda-sdgs-390-20161122_en.pdf).

EU-MERCI project, (2018). 'D4.2 - Picture of efficiency projects implemented by the Industry sector-by-sector and process-by-process / Analysis of the industrial sectors in different Countries: Greece'. HORIZON 2020 Project Nr. 693845, (Website: http://www.eumerci-portal.eu/documents/20182/38527/3+-+Greece.pdf).

Mavridis A., (2008). 'Contribution of Geographic Information Systems in Spatial Planning of Organic Agriculture in Combination with the Sustainable Development of the Natural Environment'. Dissertation Theses for the Doctorate of Philosophy, Laboratory of Geodesy and Geomatics, School of Civil Engineering, Faculty of Engineering, Aristotle University of Thessaloniki.

FAO. (2017). 'The future of food and agriculture – Trends and challenges'. Rome. ISBN 978-92-5-109551-5 (Website: http://www.fao.org/3/a-i6583e.pdf).

FAO. (2018). 'WORLD FOOD AND AGRICULTURE – STATISTICAL POCKETBOOK 2018'. Rome. 254 pp. ISBN 978-92-5-131012-0 (Website: http://www.fao.org/3/CA1796EN/ca1796en.pdf).

HAS (Hellenic Statistics Authority). (2019). 'Annual Agricultural Statistical Survey (Final Results)'. (Website: http://www.statistics.gr/en/statistics/-/publication/SPG06/).

National Information System for Research and Technology (NISRT), (2019). 'Innovation in Greek Enterprises over time'. (Website: http://metrics.ekt.gr/el/node/391).

United Nations, Department of Economic and Social Affairs, Population Division (2017). 'World Population Prospects: The 2017 Revision, Key Findings and Advance Tables'. Working Paper No. ESA/P/WP/248.

United Nations, General Assembly (2015). 'A/RES/70/1 - Transforming our world: the 2030 Agenda for Sustainable Development'. (Website: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E).

World Bank national accounts data, and OECD National Accounts data files (2019). Website: https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS

ISBN 978-618-84798-0-7