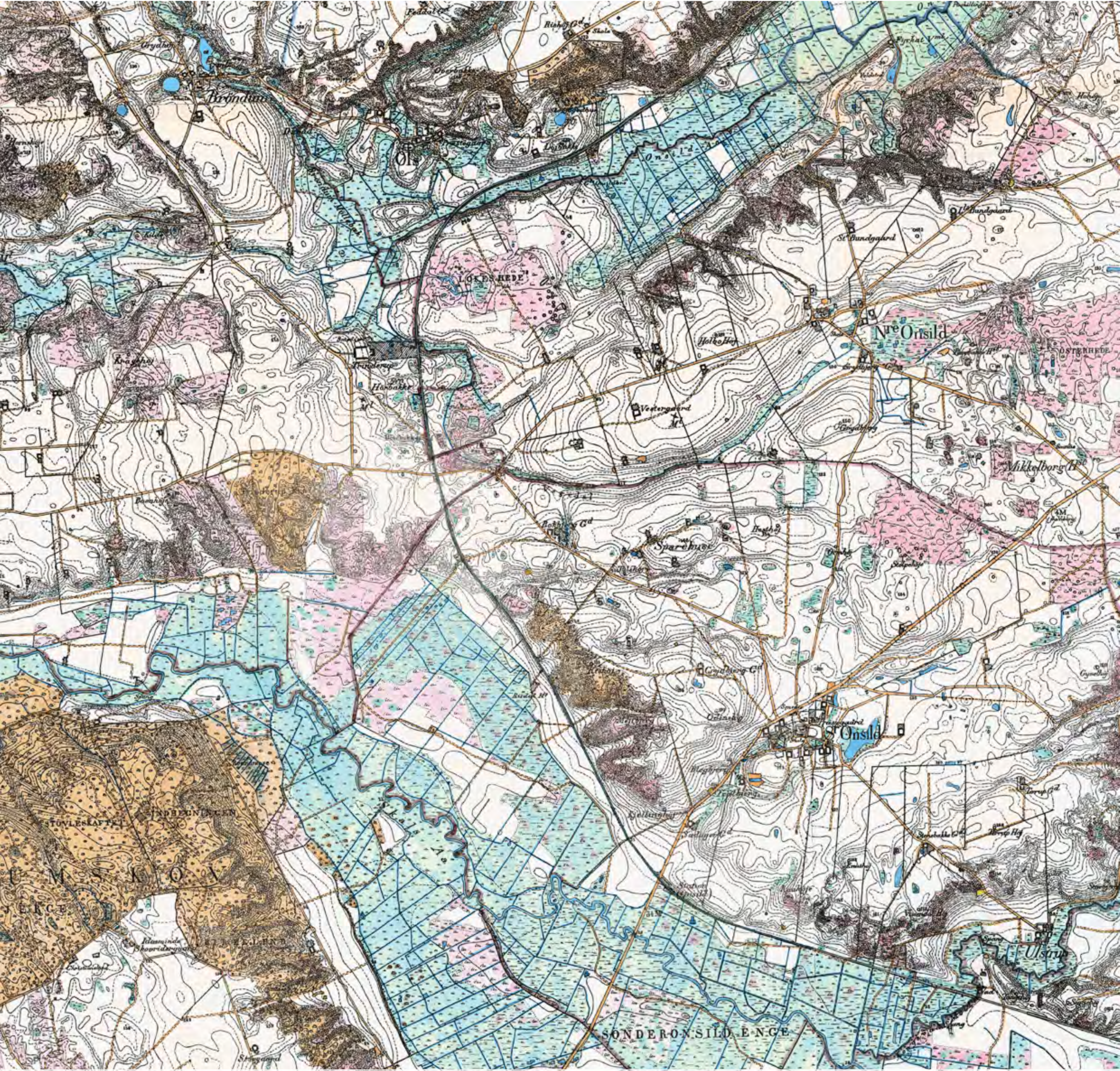


AUTOMATED PRODUCTION OF SPATIAL DATASETS FOR LAND CATEGORIES FROM HISTORICAL MAPS

Method development and results for a pilot study of Danish late-1800s topographical maps

Scientific Report from DCE – Danish Centre for Environment and Energy No. 389 2020

AARHUS UNIVERSITY
DCE - DANISH CENTRE FOR ENVIRONMENT AND ENERGY

[Blank page]

AUTOMATED PRODUCTION OF SPATIAL DATASETS FOR LAND CATEGORIES FROM HISTORICAL MAPS

Scientific Report from DCE – Danish Centre for Environment and Energy No. 389 2020

Gregor Levin[1]
Geoffrey Brian Groom[2]
Stig Roar Svenningsen[3]
Mads Linnet Perner[3]

[1] Aarhus University, Department of Environmental Science
[2] Aarhus University, Department of Bioscience,
[3] Royal Danish Library

AARHUS
UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Data sheet

Series title and no.: Scientific Report from DCE – Danish Centre for Environment and Energy No. 389

Title: Automated production of spatial datasets for land categories from historical maps
Subtitle: Method development and results for a pilot study of Danish late-1800s topographical maps

Authors: Gregor Levin[1], Geoffrey Brian Groom[2], Stig Roar Svenningsen[3], Mads Linnet Perner[3]
Institutions: [1]Aarhus University, Department of Environmental Science,
[2]Aarhus University, Department of Bioscience,
[3]Royal Danish Library

Publisher: Aarhus University, DCE – Danish Centre for Environment and Energy ©
URL: http://dce.au.dk/en

Year of publication: September 2020
Editing completed: August 2020

Referee: Peder Dam, Odense Bys Museer
Quality assurance, DCE: Jesper R. Fredshavn
Linguistic QA: Ann-Katrine Holme Christoffersen

External comments: This report is excepted, as it was written before the new guidelines regarding external comments were implemented.

Financial support: 15. Juni Fonden

Please cite as: Levin, G., Groom, G.B., Svenningsen, S.R. & Perner, M.L. 2020. Automated production of spatial datasets for land categories from historical maps. Method development and results for a pilot study of Danish late-1800s topographical maps. Aarhus University, DCE – Danish Centre for Environment and Energy, 121 pp. Scientific Report No. 389 http://dce2.au.dk/pub/SR389.pdf

Abstract: This report records the methods and the results of a pilot project aimed at automated production of machine-readable spatial datasets for land categories from Danish topographical maps from the late 1800s. The study was undertaken for two study areas in Jutland, covering around 300 km². Target land categories were: heath, sand dune, wetland, forest and water bodies. The automated geo-data production comprised a combination of object based image analysis, vector GIS, colour segmentation and machine learning processes. Results of an accuracy assessment indicate accuracies that for most categories are around 90 % or higher. A change assessment for the period from the late 1800s until today, revealed a dynamic characterised by decrease in open habitat types due to cultivation and afforestation. We conclude, that automated production of LULC category digital geo-data from historical maps offers a less time consuming and consequently more resource efficient alternative to traditional manual vectorisation

Keywords: Historical maps, topographical maps, Høje Målebordsblade, object based image processing, automated production, landscape change

Layout: Ann-Katrine Holme Christoffersen
Front page photo: Example of the "Høje Målebordsblade" map series, West of Hobro in eastern Jutland (Map from Agency for Data Supply and Efficiency (2019))

ISBN: 978-87-7156-511-9
ISSN (electronic): 2245-0203

Number of pages: 121

Internet version: The report is available in electronic format (pdf) at http://dce2.au.dk/pub/SR389.pdf

Contents

[Blank page]

Preface

This report records the methods and the results of a pilot project aimed at automated production of machine-readable spatial datasets for land categories from Danish topographical maps from the late 1800s. In 2019, 15. Juni Fonden financed this project with 500,000 DKK.

The project was conducted as a collaboration between the Department of Bioscience and the Department of Environmental Science at Aarhus University and the Special Collections at the Royal Danish Library.

Main authors of the sections in the report are:

- 1 Introduction: Gregor Levin
- 2 Description of the Danish topographical maps from the late 1800s: Stig Roar Svenningsen and Mads Linnet Perner
- 3 Automated production of spatial datasets of land categories: Geoffrey Brian Groom
- 4 Accuracy assessment: Stig Roar Svenningsen, Mads Linnet Perner and Gregor Levin
- 5 Assessment of land use and land cover changes: Gregor Levin
- 6 Discussion and conclusion (all)

In 2020, the map layers, which were generated in this project, will be made available via the website of Aarhus University.

Sammenfatning

Historiske kort indeholder værdifuld geografisk information omkring fortidens arealanvendelse og arealdække. Informationer om arealanvendelse og arealdække i historiske kort er imidlertid repræsenteret grafisk med en kombination af farver, signaturer og tekst. For at gøre den arealinformation, som er indeholdt i historiske kort anvendelig til kvantitative arealanalyser og analyser af arealændringer med moderne geografiske informationssystemer (GIS) er der behov for at konvertere denne arealinformation til digitale maskinlæsbare geografiske data (digitale geodata). Traditionelt er dette blevet gjort ved tids- og ressourcekrævende visuel tolkning og manuel digitalisering. Derfor er arealanalyser på baggrund af historiske kort oftest blevet udført for relativt små områder. Nylig udvikling inden for digital billedbehandling giver hidtil usete muligheder for mindre ressourcekrævende automatiseret produktion af digitale geodata fra historiske kortblade.

I 2019 har Aarhus Universitet og Det Kgl. Bibliotek gennemført et pilotstudie for at udvikle, teste og evaluere automatiseret produktion af digitale geodata fra de Høje Målebordsblade fra slutningen af 1800-tallet, som er det første landsdækkende topografiske kortværk for Danmark i stor målestok. Pilotstudiet er blevet gennemført for to undersøgelsesområder: Hirtshals i Nordjylland og Hobro i det centrale Jylland. Områderne dækker omkring 300 km² samt en stor variation af arealkategorier. Digitale geodata blev produceret for arealkategorierne: Hede, klit, vådområde, skov og vandoverflader. Den automatiserede produktion af digitale geodata blev baseret på en kombination af objektbaseret billedbehandling, farvesegmentering, vektor-GIS og maskinlæringsprocesser. For omkring 27.500 kontrolpunkter blev der udført en visuel kontrol af nøjagtigheden af den automatiserede metode. For de fleste arealkategorier viser resultaterne producentnøjagtigheder på omkring 95 % eller højere og brugernøjagtigheder på omkring 90 % eller højere. Pilotstudiet viste også, at især farvevariationer mellem forskellige kortblade kræver udvikling af relativt sofistikerede metoder for at opnå tilfredsstillende resultater. En sammenligning med nutidige arealkort viser arealændringer, som overordnet er karakteriseret ved en tilbagegang af lysåbne naturtyper, såsom heder, vådområder og klitter som en følge af opdyrkning og skovrejsning. Vi konkluderer at automatiseret produktion af digitale geodata fra historiske kortblade er mindre tidskrævende og derfor mere ressourceeffektiv and traditionel, manuel digitalisering.

Summary

Historical maps contain valuable, geographic information about past land use and land cover (LULC). However, historical maps are merely graphic raw material comprising colours, symbols, lines, and text labels that represent different aspects of LULC. In order to make these maps useful for assessment and quantification of LULC and LULC changes (LULCC) with modern geographical information systems (GIS), the LULC categories represented in the historical maps must be transformed into machine-readable spatial datasets. Traditionally, this production of LULC category digital geo-data from historical maps has been done by time and resource demanding visual interpretation of the maps and manual digitisation. Consequently, analyses of historical maps have typically been elaborated only for rather small areas. Recent developments within digital image processing have led to unprecedented possibilities for less time-consuming automated extraction of LULC categories from historical maps.

In 2019, Aarhus University and the Royal Danish Library conducted a pilot study to develop, test, and evaluate automated production of land category digital geo-data from Danish 1:20,000 topographical maps from the late 1800s (the Høje Målebordsblade map series). The study was undertaken for two areas: Hirtshals, in northern Jutland and Hobro in central Jutland, covering around 300 km² and a large variety of different Danish landscapes. Produced digital geo-data were for the land categories heath, dune sand, wetland, forest and water bodies. The automated geo-data production comprised a combination of object based image analysis, vector GIS, colour segmentation and machine learning processes. An accuracy assessment, based on visual interpretation of the applied map sheets for around 27,500 control points, was conducted. For most of the target categories, results indicated producers accuracies of around 95 % or higher and users accuracies of around 90 % or higher. The pilot study revealed that colour variations between map sheets is a significant factor in determining the level of method sophistication required to achieve satisfactory results. A comparison with contemporary maps revealed LULC changes, which are generally characterised by a decrease in open habitat types, such as heath, dunes and wetland due to cultivation and afforestation. We conclude, that automated production of LULC category digital geodata from historical maps offers a less time consuming and consequently more resource efficient alternative to traditional manual vectorisation.

1 Introduction

1.1 The relevance of maps

Often covering large areas over various periods of time, maps are unique documents containing information about places, human activities, and biophysical properties in the past, for which no or only limited alternative information sources exist (Chiang et al. 2014, Liu et al. 2019).

Maps are spatially explicit. In a map, categorisations of space, such as land use and land cover (LULC), are delineated as objects with a specific spatial location and extent. In contrast to tabular data, such as LULC statistics, information contained in maps can be applied for spatially explicit analyses of the presence, pattern and distribution of LULC categories. By spatially overlaying historical LULC information with contemporary LULC maps, spatially explicit changes in LULC can be explored, quantified and modelled. Applications include analyses of how changing drivers affect land use dynamics (e.g., Konkoly-Gyuró et al., 2019, Mikusiska, 2013, Svenningsen et al., 2019, Van den Berghe, 2019), and more specific explorations, such as of relations between land use dynamics and biodiversity (e.g., Käyhkö & Skånes, 2006), or estimation of historical carbon fluxes (e.g., Fuchs et al. 2016).

Although spatially explicit, historical maps are, just as any other kind of cartographic documents, not an objective representation and categorization of space. Rather, spatial categories in maps should be understood as a categorisation, which reflects the need for spatial information of the organisation or individual requesting production of the map. This also implies that definition of categories can change over time according to changing purposes of mapmaking, such as demonstrated by Svenningsen et al. (2015). Therefore, historical analyses of land use and land cover change (LULCC) must account for categorisation of space in historical maps by investigating the mapping practices of the mapmaker. Today, map production is mainly digital. In digital maps, geographic information is provided as spatial datasets (e.g., points, lines, and polygons in vector format or layers in raster format) which can be explored in modern analytical environments, such as geographic information systems (GIS).

However, particularly historical map documents are graphic raw material, comprising geographical information expressed in colours, symbols, lines and text labels. In order to enable the utilisation of the contained geographic information in a GIS, the geographic information needs to be recognised and extracted. In other words, geographic information in the form of historical map documents has to be “unlocked” through map processing.

Traditionally, extraction of geographic information from historical maps has been based on time- and resource consuming visual interpretation and manual delineation of geographic features. Analyses of historical maps have therefore typically only been elaborated for a limited number of map sheets and relatively small study areas (see e.g. Caspersen & Fritzbøger, 2002; Helkiær Jensen & Jensen, 1979; Kaim et al., 2016; Kienast, 1993; Konkoly-Gyuró et al., 2019, Kristensen et al., 2009; Odgaard & Rømer, 2009; Picuno et al., 2019; Svenningsen et al., 2015). There are, however, exceptions, such as Dam (2005), who, for all of Denmark, manually extracted geographical information from late

18th century maps. As demonstrated by Schaffer and Levin (2014), country-wide historical geographic datasets, enable analysis of patterns of landscape changes for different regions. Nevertheless, manual processing of historical maps for GIS layers not only implies generation of non-reproducible data that can suffer from high degrees of inaccuracy and introduced subjectivity, but also does not scale well for handling large map coverages. Therefore, country or region-wide information about historical LULC and LULC changes is, with a few exceptions (Dam & Jakobsen, 2008), generally limited to aggregated historical statistics (2008; Levin & Normander, 2008; Odgaard & Rømer, 2009).

1.2 Digital map processing

Digital map processing refers to computational procedures aimed at the automatic or semiautomatic production of digital spatial datasets of geographic features presented in the maps. The need for computational solutions with high degrees of automation for map processing becomes evident if one considers that millions of map documents have already been scanned and stored in digital archives (Chiang et al., 2014). For instance, in Denmark, a large proportion of available historical maps sheets have over recent years been scanned and made publicly available. This applies to all historical topographic maps from the archives of the Agency of Data Supply and Efficiency (SDFE). In addition, SDFE has also produced a few seamless nationwide and fully georeferenced raster datasets of key topographic maps series, which means that these maps series can be accessed and used in GIS.

Incorporating geographical information from historical maps into a GIS environment, where data can be used for spatial analyses and overlaid with other spatial data, creates unprecedented opportunities for multi-temporal and multi-contextual spatial analyses. Without digital maps processing solutions, large proportions of the spatial information in historical maps will remain inaccessible.

Digital map processing is a relatively young research area, which grew out of disciplines such as graphics recognition and document analysis (Chiang et al., 2014, Freeman & Pieroni, 1984, Liu et al., 2019). Due to various graphical quality issues and the general complexity of map content, extraction of information from maps is generally more difficult compared to processing of other data sources, such as written documents (Lladós et al., 2002). In maps, layers of geographic features, such as roads, contour lines, text labels, and LULC categories, often overlap with each other, increasing the degree of content complexity and colour mixing. Furthermore, due to inconsistency of applied symbols and colours, processing of manually drawn maps is even more challenging (Leyk & Boesh, 2009). General techniques for graphics recognition and document analysis can therefore not be directly applied to map processing.

In most historical maps, colours, symbols, lines and text items, represent explicit, overlapping layers of geographic information. Colours often represent thematic information of LULC categories, such as forest or water. Symbols, and to some degree text items describe different types of geographic characteristics. Lines can represent geographic features, such as roads or watercourses, boundaries of LULC categories, administrative boundaries or topographical characteristics, such as contour lines. A major objective of digital map processing is to spatially isolate these overlapping layers of geographic information from each other and to filter out those map elements, which obstruct the extraction of the geographical layer of interest. Digital map processing comprises two major methodological approaches, which are described below.

1.2.1 Recognition of map symbols, linear features and text labels

In maps, symbols, linear features and text labels describe different types of geographic characteristics. Symbols are sets of distributed graphic signs or shapes representing some feature or property (e.g., wetness)(Kent & Vujakovic, 2009). Symbol recognition uses knowledge of the semantics attributed to these symbols and then applies rules to extract them. Rules are mainly based on geometric characteristics, such as shape, orientation, texture and pattern. Symbol recognition has been applied to both recognise and extract layers of specific geographic features (e.g. Baily et al., 2011; le Riche, 2020; Liu et al., 2018; Pezeshk & Tutwiler, 2011; Uhl et al., 2018; Xu et al., 2016) and to filter out or remove features, such as text items, in order to improve the extraction of other layers of geographic features (e.g., Fuchs et al., 2013; Gobbi et al., 2019; Kim et al., 2014; Ostafin et al., 2017). Applied computation methods include Object Based Image Analysis (OBIA) (e.g., Gobbi et al., 2019; Liu et al., 2018) and graphics recognition (e.g., Chiang et al., 2016; Cordella & Vento, 2000). The process of extracting and recognising geographic features includes among others a variety of techniques that stem from the fields of document analysis and image processing and fall into three major categories. I: Template matching, where algorithms are developed to find objects that are similar in shape and size to an object given as a template (e.g., Laumer et al., 2020; Leyk & Boesch, 2009). II: Morphological operators, such as dilation, erosion and thinning are commonly used to modify image contents to facilitate the recognition and extraction of point, linear, and area features (e.g., Fuchs et al., 2013). III: Shape descriptors, which are commonly used to identify text and symbolic features, range from simple metrics (e.g., geometric measure of height, width, or area) (e.g., Kim et al., 2014) to computationally more complex descriptors based on the Fourier transform for estimating texture representations (e.g., Chiang & Knoblock, 2006) or the Hough transform for identifying imperfect linear objects (Dhar & Chanda, 2006).

1.2.2 Colour image segmentation

Many historical maps contain thematic information represented in predefined colours (e.g., blue for water, green for vegetation or red for elevation). Colour image segmentation (CIS) utilises these colours to group pixels of similar colours and extract these as vector or raster layers representing a specific geographic feature type (e.g., a LULC category). Utilised colour characteristics can include the red, green, and blue (RGB) colour space, alternative colour spaces such as hue, saturation, and intensity / lightness, or image grey scale transformations such as histogram normalisation. Studies applying CIS for digital processing of historical maps, have used supervised classification methods, where sample inputs (training data) are provided to guide the automated labelling of image pixels or objects to the trained classes (see e.g., Auffret et al., 2017; Baily et al., 2011; Fuchs et al., 2013; Leyk, 2010; Mariano et al., 2013; Ostafin et al., 2017) or unsupervised classification methods, where groupings of pixels or objects are formed without training data classification (see e.g., Godfrey & Eveleth, 2015). It is increasingly the case that applied approaches combine methods, in particular, pixel-based and object-based processing (e.g. Liu et al., 2018).

1.2.3 Digital historical map processing in land use/land cover studies

Existing studies applying digital map processing aimed at extracting and analysing land use and land cover (LULC) have been conducted for a large range of historical maps. Auffret et al. (2017) applied supervised image processing, based on RGB colour tones to arable land, forest, meadow, water and open land from Swedish economic maps from the period from 1859 to 1934 and from 1935 to 1978. Baily et al. (2011) applied supervised classification image processing techniques based on colour spaces, texture and symbol patterns to extract seven land use categories from maps from the First Land Utilisation Survey of Britain from the 1930s. Gobbi et al. (2019) carried out a semi-automatic digitisation of Italian heritage maps from 1859, 1926 and 1992. They applied object-based image analysis for filtering of undesirable feature, such as text, symbols and boundary lines and then applied colour segmentation to delineate LULC categories. Ostafin et al. (2017) created a forest cover mask from historical topographic maps from Poland and Switzerland. They applied supervised colour image segmentation and subsequent morphological techniques to filter out noise originating from undesired map elements but note that high densities of contour lines and shadings in mountainous areas highly impact accuracies. Liu et al. (2018) combined an old paper-based US civil war map and modern aerial photos to derive land-use history and landscape dynamics at fine scales for a region near Chancellorsville, USA, from 1867 to 2014. They apply a pixel based maximum likelihood classifier (MLC) and an object-based classification approach. Fuchs et al. (2015) applied a supervised maximum likelihood classification to elaborate a forest mask of central Europe based on a coarse scale land cover map from around 1900. Subsequently, they applied morphological operators to filter text items. Leyk & Boesh (2009) extracted forest areas from Swiss, 19th century topographic maps, applying a combination of symbol recognition techniques and morphological operators.

To estimate accuracies, results from digital map processing are usually compared to manual extraction of LULC categories. This builds on the assumption that manual extraction always is accurate and precise. However, manual extraction is subject to subjectivity – that is, it is difficult to defend a claim that the manual extraction interpretation and delineation of geographic information is always perfect or acceptable (Chiang et al., 2014). For instance, Ostafin et al. (2017) show that accuracies of manual extraction of forest areas vary from 91 to 99 %. In spite of this ambiguity regarding accuracy measures, studies applying digital map processing techniques to historical maps generally refer to achieved accuracies beyond 80 %. However, it is evident that accuracies are highly influenced by the complexity, quality and resolution of the utilised map as well as the number of LULC categories, which are extracted. For instance, for older, high resolution, manually drawn maps, attained accuracies are rather low, such as 53 % in the study by Ostafin et al. (2017) and 75 % in the study by Liu et al. (2018). In comparison, for higher quality and less complex 20th century maps (e.g., Auffret el al., 2017; Baily et al., 2011; Gobbi et al., 2019) generally much higher accuracies of more than 90 % are attained. However, for instance, Baily et al. (2011) also finds that while accuracies for large-extent LULC categories (meadow and arable) are very high, suggesting a 97 % correlation between the original map areas and the extracted data, other categories, such as agriculturally unproductive land, were more problematic with a correlation around 75 %. Leyk & Boesh (2009) and Fuchs et al. (2015) attained accuracies of around 90 %. However, both studies exclusively aimed at extracting forest areas, while other LULC categories were ignored.

In spite of variations in attained accuracies, a general finding is that digital map processing techniques are much more time and resource efficient than manual approaches. For example, for extraction of LULC categories from British land utilization maps from the 1930s, Baily (2007) concludes that for one map sheet, automated extraction of geographic layers takes 11.5 hours compared to 93 hours for manual extraction.

1.2.4 The Danish context

In Denmark, there is a unique opportunity to develop and apply digital map processing methods for automated extraction of LULC categories from historical maps. All publicly produced topographical maps are freely available as high-resolution scan digital raster data (Agency for Data Supply and Efficiency, 2019). An automated extraction of LULC categories from historical topographic maps will give the opportunity to generate countrywide map layers, representing the location and extent of different LULC categories at different points in time. Such map layers can be applied in spatially detailed assessments of LULC changes and of trajectories of LULC change. Being spatially explicit, the map layers can be applied in assessments at any spatial level, from a specific location or plot to different spatial delineations, such as parishes, municipalities, watersheds or the whole country.

2 Description of the Danish topographical maps from the late 1800s

The 1:20,000 topographical maps 'Høje målebordsblade' (HMB) represent the first large-scale topographic mapping of Danish territory. The HMB map series consists of approximately 1124 sheets, mostly 31.4 x 37.7 cm, at a scale of 1:20,000 surveyed between 1842 and 1899 and was developed by the Danish military as a response to a growing need for detailed topographical information about the landscape for military purposes. Although Denmark had been mapped by the Royal Society for Sciences and Letters at a scale of 1:120,000 during the second half of the 18th century, these maps were not sufficiently detailed for military use, thus creating a demand for a new military survey.

The HMB has a long and complex production history and was the culmination of a military mapping programme, initiated by the Quartermaster Department of the General Staff (QDGS). The wars against Britain in 1801 and from 1807 to 1814 had demonstrated an acute need for a detailed mapping of landscape elements and ground conditions. In the first half of the 19th century, the QGDS conducted several surveys on Zealand. In 1842, following a reorganization of the military survey organization by forming the Topographic Department of the General Staff, the HMB programme was commissioned. Although the military still was in a desperate need for the maps, the survey progressed slowly, due to both technological development in mapmaking and interruptions from the two wars in Schleswig (1848-1850 and 1864). After the war in 1864, the survey progressed more systematically, starting with the southern part of Jutland near the new border to Germany, and by 1899 the survey was completed in the hinterland of Copenhagen (Svenningsen, 2014; Svenningsen, 2016a). Figure 2.1 shows the year of survey of HMB map sheets between 1856 and 1899.

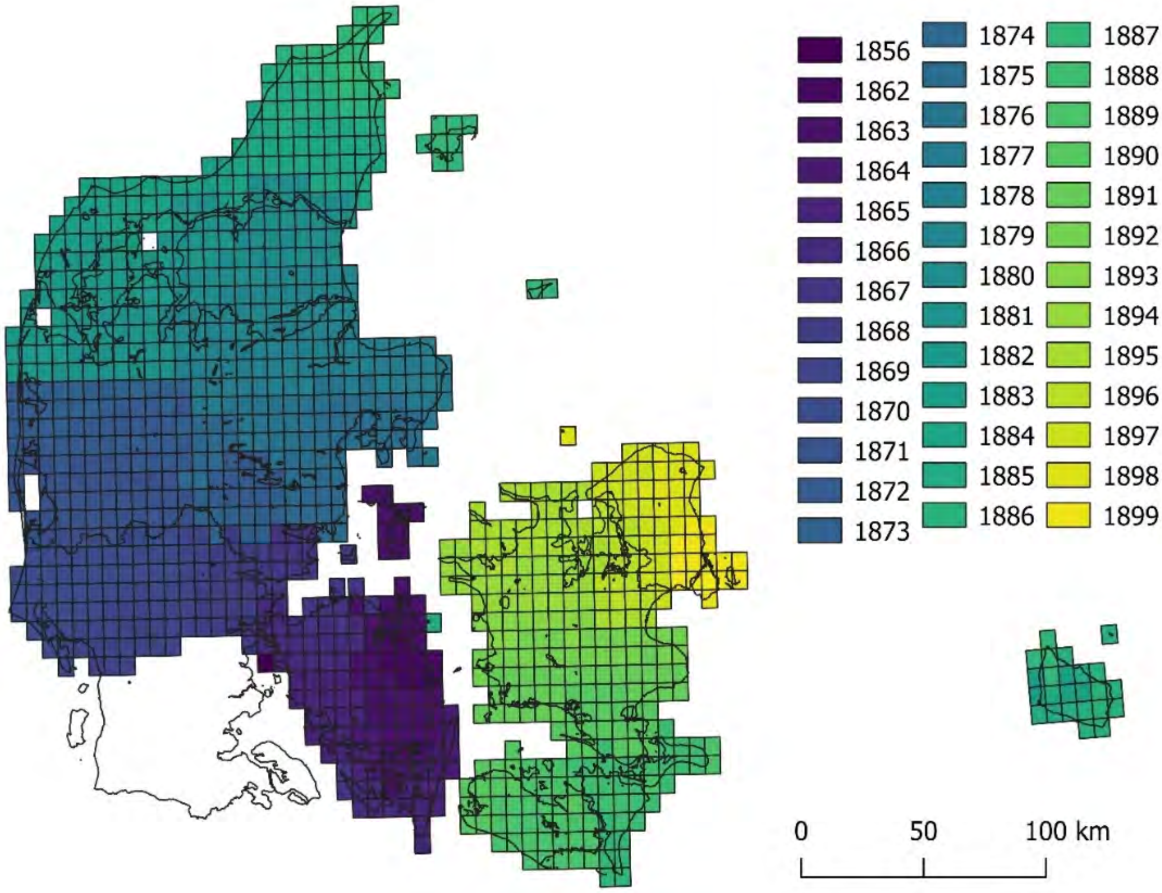


Figure 2.1 Year of survey of the HMB map sheets in scale 1:20,000 between 1856 and 1899.

2.1 Organisation of the survey

The HMB survey was done by military personnel from the Topographic Department of the General Staff in annual campaigns during the summer months, Michaelsen (2004). The surveys were organised in survey teams (målebrigade), who usually surveyed several parishes during one field campaign. A parish was normally subdivided in several small sections, so the paper could be handled on a surveying table (for an example of Klejtrup parish see Figure 2.2). Several surveyors were usually involved in the survey of one parish, e.g. in the case of Klejtrup two surveyors were involved, each handling two subsections. For a subsection, the surveyor would draw a manuscript map, where all measurements of spot heights and contour lines, as well as categorisations of land categories and landscape elements were recorded. Subsequently, another surveyor verified the manuscript map in the field during the campaign, before the results could be approved. An example of a manuscript 12b for Klejtrup parish can be seen in Figure 2.3.

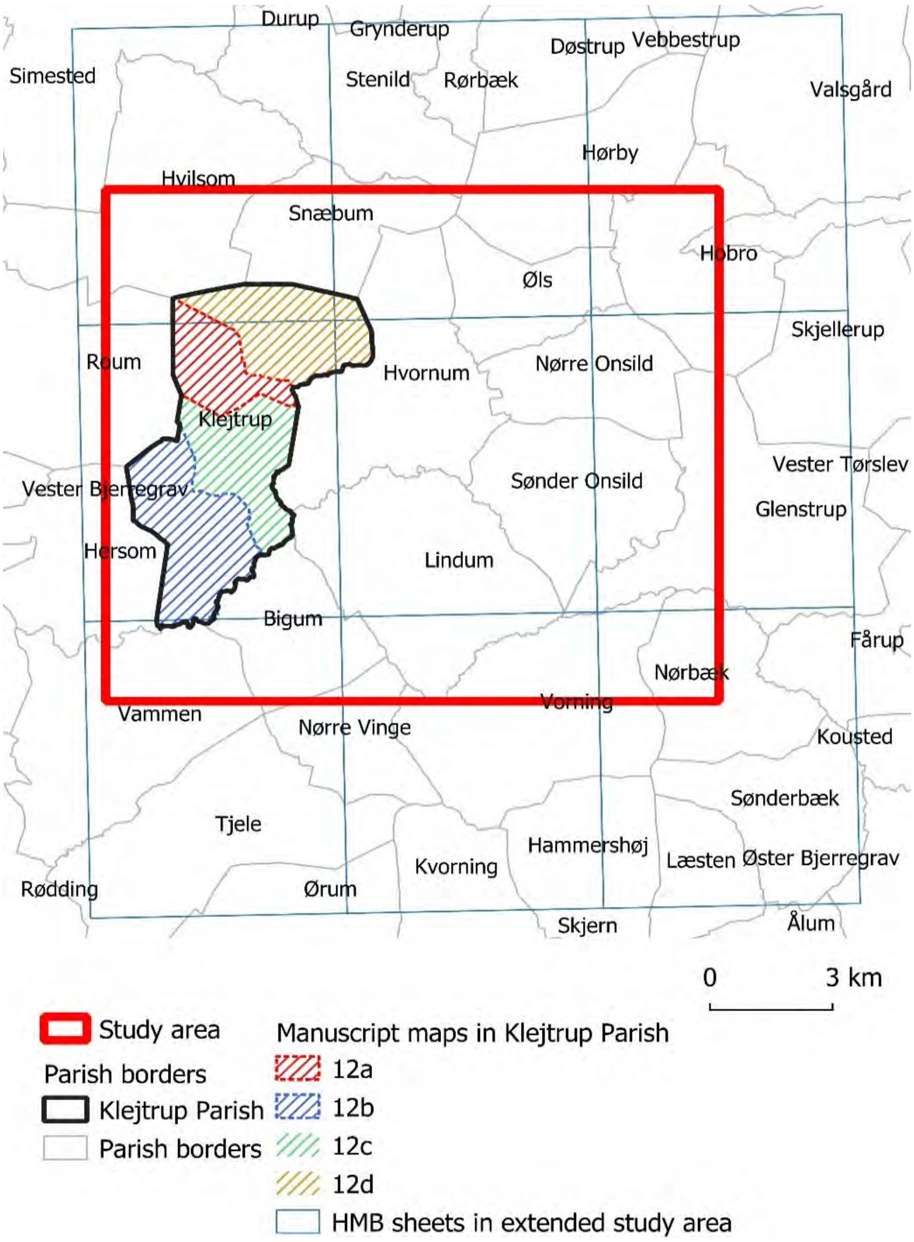


Figure 2.2 Illustration of the subdivision of parishes into survey sections for Klejtrup parish in the southern study area. Each subdivision corresponds to a manuscript map. This is also an illustrative example of the fact, that each HMB sheet was compiled from a number of manuscript maps from different surveyors.

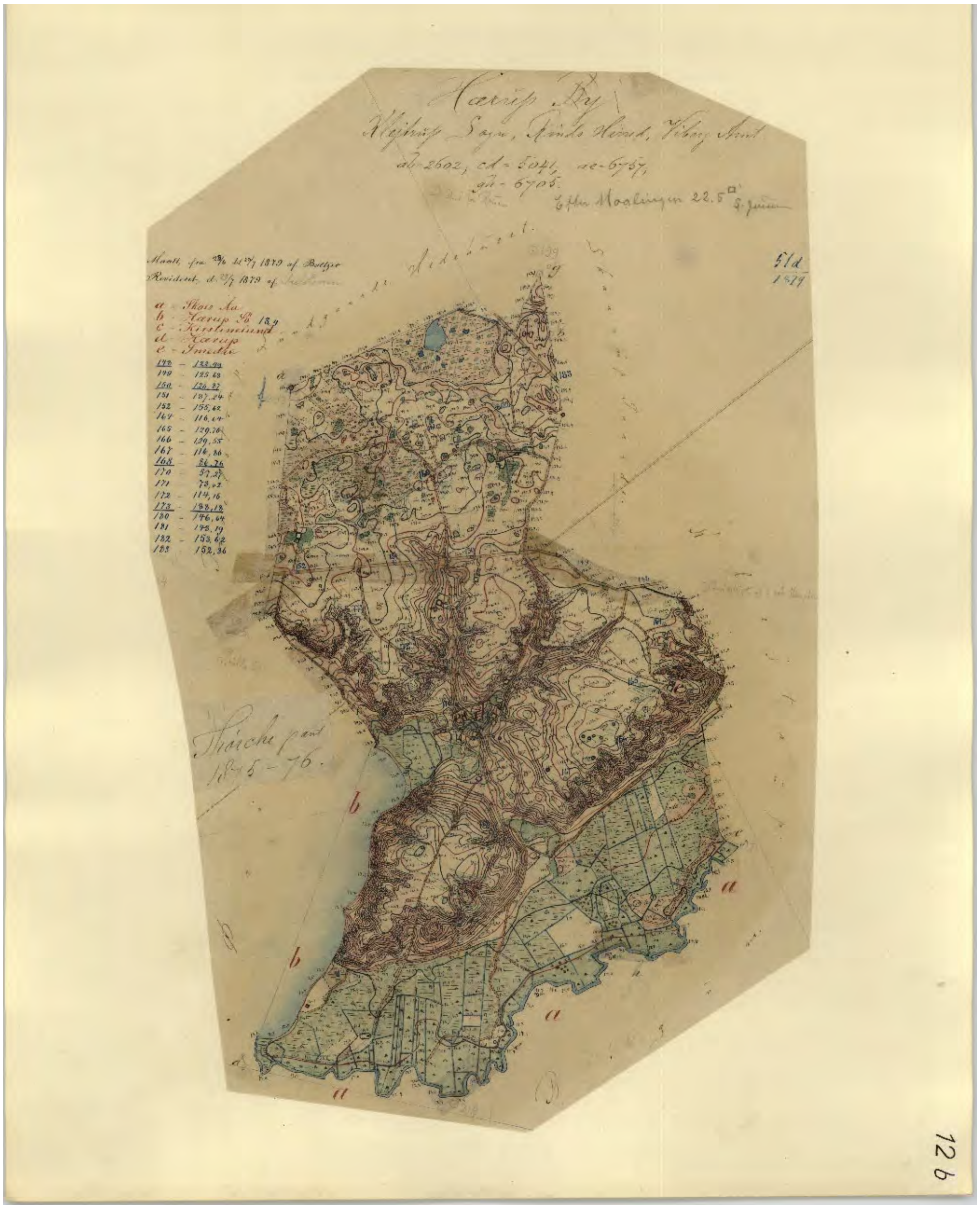


Figure 2.3 An example of a manuscript map from section 12b of Klejtrup Parish, surveyed between 23 June and 28 July 1879 by Baltzer and verified in the field 29 July 1879.

During the winter periods, the manuscript drafts were compiled into a full map sheet and prepared for printing. Because the survey progressed systematically parish by parish, using the cadastral maps as the geographic frame for the survey, a single printed map sheet can hold data collected in different years (Korsgaard, 2006, pp. 58–61). This means that the time of survey follows parish borders, rather than the individual map sheets. For one of the areas, which are applied in this study, Figure 2.4 shows the survey year for the parishes and the map sheet indexes. All map sheets within the validation area (red square) are based on surveys from 1879. However, some of the map sheets in the extended study area, outlined with blue are compiled based on surveys from other years, which means that data contained in one map sheet can have been collected across several years, as it is the case for sheet N21 Hvilsom. The vectorization of the HMB data, thus enables the possibility to specify the exact time of survey (year) for each feature in a parish, rather than just a period of up to three years for a complete map sheet.

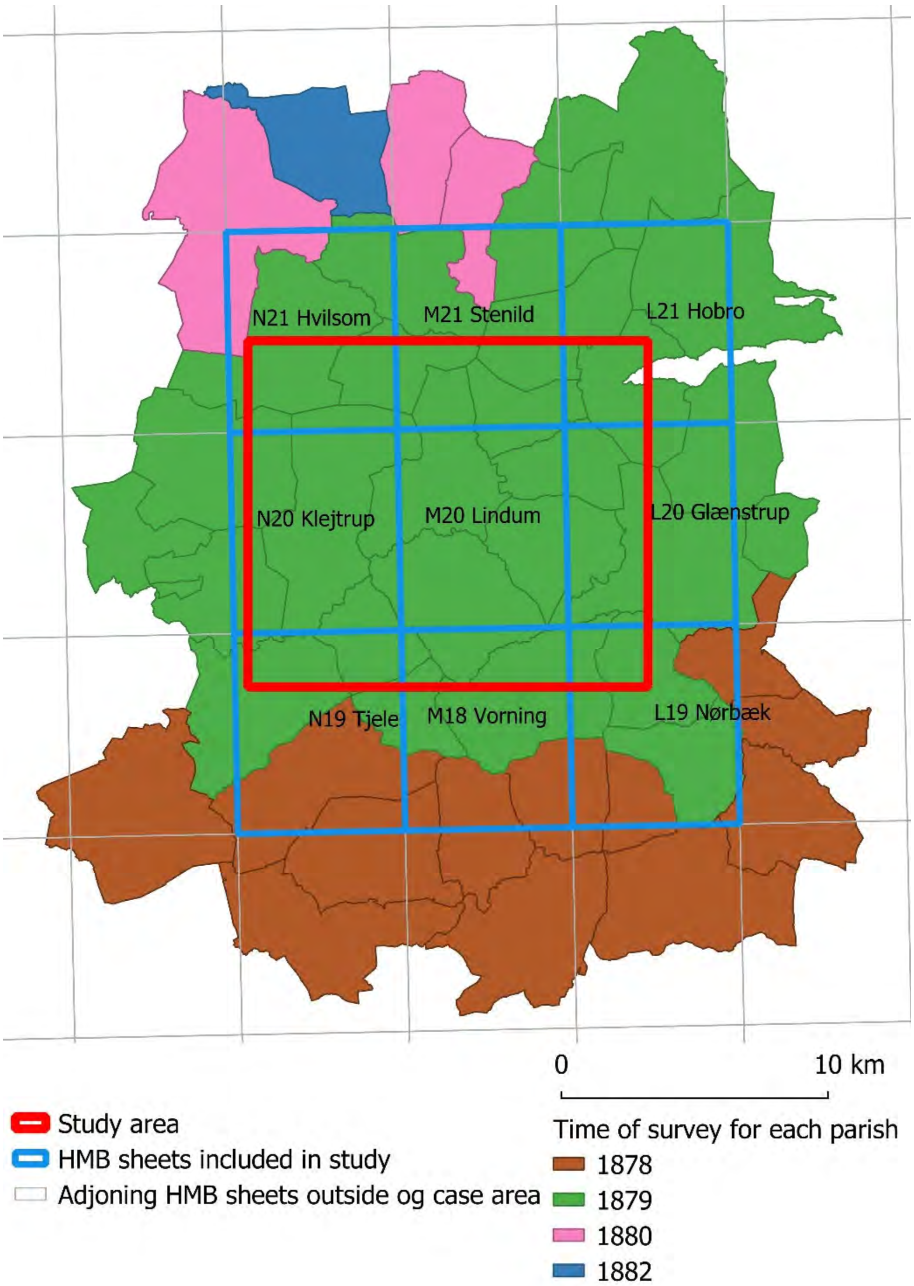


Figure 2.4 The figure shows, how the survey for the HMB sheets was done parish by parish. A HMB sheet typically included several parishes, which means that the time of survey can vary up to several years within one map sheet.

2.2 Original mapping instructions

The HMB maps are the primary source for detailed information about the landscape before the 20th century and have been used for a number of studies and administrative purposes. However, despite the importance of the maps as a source for information about the landscape of the past, the surveys and production behind the maps have only to a limited degree been subject to research. Although the military origin of the maps and the possible implications for landscape analysis have been recognized (Brandt, 2004; Münier, 2009), our understanding of the land classification system is still mainly limited to what is stated in the map legends (Korsgaard, 2006). This means that information about the categorisation and representation of land categories and landscape elements in the map is sparse and in principle limited to the signature tables published with the map series. Although some of the land categories shown in the legend seem to be directly comparable to modern LULC classification typologies, such as forest vegetation, recent work shows that classification of land categories has changed between different map types and over time (Svenningsen et al., 2015).

Consequently, in order to use data from the HMB in LULCC analyses, there is a need to analyse mapping practices and land classifications, which were used in the HMB maps. However, the amount of printed source material about the HMB maps is limited. In addition to the printed legend of the HMB, only a small booklet with guidelines for the surveyors provides some additional insight into the land classification and survey principles used in the production of the HMB map sheets (Generalstabens Topografiske Afdeling, 1877). Furthermore, the instructions in the booklet are very superficial and the majority of the text describes various administrative procedures for travel, accommodation, workhours as well as cleaning of instruments etc. The definition and description of land categories and landscape elements to be included in the survey take up just a single page along with an additional table in an appendix with a list of land categories. These land categories are generally only described with their name, such as 'meadow', without any further explanation. An exception is the land area type 'bog', where it is specified that this area type is characterised by peaty soil. However, the symbols to be used to represent land with peaty (organic) soil depends on the land use. For instance, bogs, which have been ploughed or used for grazing, are to be represented as meadow. The table in the appendix of the booklet of 1877 also list a few combinations of land area categories, such as combinations of meadow, bog and / or heath[1].

Based on a review of the land categories represented in the printed HMB sheets and from analysis of the manuscript maps, it is evident, that these land categories were combined in other ways than described in the instructions. Thus, the relative short description of land categories suggests that this categorisation of land was well known by the surveyors as well as by the military, who was the primary user of these maps. Therefore, in order to understand the categorisation of land by the military surveyors, there is a need to further explore the mapping practices of the Danish military in the 19th century.

As already mentioned, the first permanent military survey in Denmark was established when the QDGS was ordered to conduct topographic surveys following the wars against the Britain in 1801, and from 1807 to 1814. Although the QDGS commissioned several surveys and terrain reconnaissance in the first half of the 19th century, no coherent framework for military surveying and mapmaking existed at the time (Svenningsen, 2014; Svenningsen, 2016a). In 1830, the QDGS started to develop a concept for the production of military topographic maps, with a written survey instruction (Steinmann, 1832). Later followed a series of booklets written by the officer Oluf Nicolai Olsen, which served as a guideline for military topographic surveys and map production. The booklets were published between 1831 and 1836 and were later compiled into a single book. The first booklet, 'Topographisk Kaartsignaturer' from 1831, is the most important in relation to the definition and delineation of land categories and the related representational practices (Olsen, 1831). In the 1831

[1] The Danish word 'Lyng' is used for a land category in the original signature key and survey instructions for the HMB. This term be cannot directly translated to the English 'heather' which relates to the plant species Calluna vulgaris, as the 'lyng' land category probably covers a broader range of heathland, including other types of dwarf-shrubs and grasses. Consequently, we use the 'heath' for the land category 'lyng' in the HMB map.

publication, Olsen, in contrast to the later survey instructions, provided a classification system for landscape elements and land categories.

The categorization described by Olsen is focused on recording land categories and landscape elements, which had an influence on mobility, visibility, navigation as well as on force sustainment, which reflect the military purposes of the maps (Svenningsen, 2014; Svenningsen, 2016a). In this system, areas were classified according to two principles 1; Ground / soil condition and 2; classification of vegetation, which can be seen in Tables 2.1 and 2.2. The system allowed for combinations of different land categories, such as an area characterized with soft ground and with a vegetation cover of heather.

Table 2.1 Soil/ground classification by Olsen (see also appendix 1 and 2).

Ground classification	Definition (if stated in text)	Representational practices
Soft soil	Can be crossed by horsemen	Low intensity of green colour green and/or simple hatching.
	Can be crossed on foot	Medium Intensity of green colour and/or medium hatching.
	Can be crossed on foot only with difficulty (and/or assistance).	High intensity of green colour and/or dense hatching.
Hard soil	Arable land	Not to be represented
	Marsh land	Dark blue-grey colour.
	Heath	Light pink colour and/or symbol for heather.
	Hills	
Loose soil	Sand	Orange-brown colour and/or symbol for sand

Table 2.2 Vegetation types by Olsen (see also appendix 1 and 2).

Vegetation type	Definition (if stated in text)	Representational practices
Cereal/cropland		Not to be represented on the map
Permanent Grass	In meadows and such areas, which are in permanent use for grazing or haymaking.	Green colour and/or symbol for grass.
Broadleaf trees		Brown colour and/or symbol for deciduous trees.
Needleleaf trees		Brown colour and/or symbol for coniferous trees.
Scrubland		
Heath		Light pink colour and/or symbol for heather.
Plantations		
Gardens		Large gardens to be represented as forest

In addition to defining which land categories and which landscape elements were supposed to be surveyed and represented on military maps, an important part of the work by Olsen was to set the standard for symbols and colours to be used for military mapmaking. According to Olsen there was a need for a standardization, as already at that time several competing sets of symbols existed (Olsen, 1831, p. 1). Thus, land categories on large-scaled maps were supposed to be represented with a combination of symbols and colours (see appendix 1 for colours and appendix 2 for the symbols to be used). The use of a combination of symbols and colours to represent variation of land categories (combinations of soil type and vegetation cover), allowed for detailed cartographic representation of small changes in the landscape. As such, this representation practice suggests that Olsen drew on an earlier military cartographic tradition, where most often military maps were prepared as manuscript maps and then copied by hand. In this tradition, combinations of colour and symbols were used extensively to illustrate variations in the landscape (Medyńska-Gulij & Żuchowski, 2018). This mapping practice allowed for a detailed representation of gradual change between land categories such as increasing wetness in, for instance, a river valley.

Generally, the classification system setup by Olsen focuses on representing information related to land categories and landscape elements, which had either a positive or a negative influence on military mobility as well as visibility. Thus, areas in the Danish military mapping practices outlined by Olsen, basically seem to function as a friction matrix (see Table 2.3). An example of a frictionless area could be an area characterised by hard soil and with a vegetation cover of cereals on a totally flat surface. Such an area would have no negative or positive impact on the movement across the area. However, if a road traversed the area, this would improve the trafficability in the specific area, while a stream surrounded by wetland would limit the trafficability. The impacts on mobility and visibility are not necessarily related. An area could have a significant negative impact on mobility, but with no restriction on visibility. This is the case for a meadow with soft soil and a vegetation cover of grass. Such an area type would be difficult to cross but have a high visibility and low degree of concealment.

Table 2.3 Examples of relationship between land categories and mobility, visibility and concealment in the classification by Olsen (1831).

Area type	Impact in visibility	Impact on concealment	Impact on mobility
Forest (on hard soil)	-	++	-
Coniferous forest (on hard soil)	- -	+++	- -
Deciduous forest (on hard soil)	-	++	-
Heath (on hard soil)	+	-	-
Meadow (soft soil and grass vegetation)	++	- - -	- -
Bog (without specified vegetation cover)	+	- -	- - -
Dune	++	- -	-
Arable fields (hard soil and cereals)	+++	- - -	neutral
Hills (hard soil, no vegetation)	+/-	+/-	+/-
Scrubland (hard soil)	- - -	++	- -
Road	+/-	-	+++

In 1832, the QDGS issued additional instructions for the survey, which specified the use of the classification system and representational practices made by Olsen (Steinman, 1832). In relation to classification of land area type, the additional instruction stated that soil conditions e.g. hard and soft soil were to be classified independently of potential vegetation cover. However, presence of soft soil should always be represented in combination with vegetation cover. This creates some ambiguity in the representational practices related to soft soil and permanent grass, as both categories were to be represented with different green colours. Permanent grass, which is referred to as meadow (eng)[2], was to be represented with 'Kaisergrøn' and soft soil with mixture of

[2] Meadow (eng) is defined in the instructions as permanent grass, which can carry cross-country movement of infantry, cavalry and artillery units. Seasonal variation in softness in very humid years, or during spring and autumn is not to be considered: 'Enge. Hvorved forstaas Jordsmon, som altid ligger til Græs, og er meer og mindre sidliggende, men stedse har fast Grund, der kan passeres af Vaabenarter, betegnes blot med Græssignatur og ingen skravering. At sådanne Enge i almindelig fugtige Aar, For- og Efteraar ere meget bløde, kommer ikke i Betragtning ved Tegning'. Source: Steinmann 1832.

'Kaisergrøn' and 'Gummigult'. Figures 2.5 and 2.6 show how these guidelines were used in the field surveys conducted by the QDGS in the 1830s. In Figure 2.5, the variation in wetness is indicated by hatchings within the meadow around lake 'Bækkene'. The meadow is symbolised using a combination of green colour and symbols for grass. In Figure 2.6, the situation is the other way around. Here the green colours are used to indicate soft soil, while permanent grass is only marked with grass symbols along both sides of the border demarcating the area with soft soil. As such, this illustrates the use of combination of symbols and colours to represent gradual change in vegetation and soil conditions.

Figure 2.5 Extract of a manuscript map 'Section Arninge' surveyed in summer of 1841 by Pl.t L.E. von Staffeldt. Source: SDFE-file: GKD-073.

Figure 2.6 Extract of manuscript map 'Section Vestenskov' surveyed in summer of 1839 by Adjoint Premierlieutenant Andræ. Source: SDFE-file GKD-072.

In 1842, the topographic survey of the General Staff was reorganised, and a new Topographic Department of the General Staff was formed with O. N. Olsen as the leader. In the period after 1842, new survey and representational techniques were introduced and it was decided that the topographic maps would to be printed in a scale of 1:20,000 rather than 1:80,000. In addition, instead of hachures, contour lines were introduced to represent the relief. A few test maps were produced to test the use of contour lines for north-western Zealand and Møn. These test maps are illustrative examples of the development of the topographic maps and the development of the system by Olsen into what later became the survey instructions published in 1877. The test map from 1858 (Figure 2.7) still shows traces of the representational practices described in the guidelines by Olsen, but when studied in detail, small changes can be observed. Soft soil is still represented with dense hachures and areas that appear to be permanent grassland are marked with a green colour. Although some fuzzy borders exist between soft soil and grassland, the borders between grassland and hard soil (potentially agricultural land or commons) are now in some places marked with stippled lines. However, a comparison with the later version of the HMB sheet from 1895 (Figure 2.8), show the development in the representational practices, where all wet soils seem to be shown with a green colour and with symbols for the vegetation (grass / meadow and reeds). However, as late as in 1865 mapping of gradients in soil wetness can be observed in the manuscript maps analysis of land reclamation (Stenak, 2005, p. 151).

Figure 2.7 Extract of the test map 'Veirhöi' from 1846. Source: Royal Danish Library, Map Collection, shelf mark: KBK 1111,251,633,5-0-1846/1.

Figure 2.8 Extract of the HMB sheet e8 'Dragsholm' surveyed in 1895 and printed in 1897. Source: SDFE seamless HMB dataset.

2.3 Overview of land use/land cover types

The classification of land categories in the HMB maps is clearly related to the standards set by O. N. Olsen in the 1830s, but was modified according to the instructions from 1832 and 1877. Based on the legend and the survey instructions from 1877 a set of land categories can be identified (see Table 2.4). However, on the final printed versions of the maps several additional combinations of these land categories can be identified, as well as a number of land categories not described in the legend or in the survey instructions. The 1877 survey instructions specify that different land categories were to be separated with a line or stippled lines, such as indicated on the manuscript map and the corresponding printed maps sheet in Figure 2.9. Here, the area around the lake is shown as a combination soft soil, with heather and grass vegetation, and a stippled line marks the border to the area represented as pure heather. Despite this rule in the guidelines, several examples of gradual change within a land area type can be observed in the printed maps, which means that the representational practices in the maps did not follow the guidelines in all cases. The examples of such exceptions are numerous, and do particularly occur in complex landscapes with graduated transitions form one land area type to another, such as from costal dunes to heathland.

Table 2.4 Overview of land categories provided in the 1877 guidelines for surveyors.

Number in original table	Name of area type in table (Danish)	Name of area type in table (English)	Land area category	Interpreted definition	In map legend 1902
30.	Eng.	Meadow	Soil condition and vegetation	Wet soil with permanent grass cover	+
31.	Mose.	Bog[3]	Soil condition	Wet peat soil, with natural vegetation	+
32.	Lyng.	Heath	Vegetation	Heath cover on hard soil.	+
33.	Klit.	Dune	Soil condition	Loose sand, no or sparse vegetation	+
34.	Sø, Vandhul, Dam.	Lake/pond	Ground condition	Open water.	+
35.	Skov af Løvtræer.	Forest of broadleaf trees	Vegetation	Broadleaf tree cover, high density (forest)	+
36.	Skov af Naaletræer.	Forest of needleleaf tree	Vegetation	Needleleaf tree cover, high density (forest)	+
37.	Krat af Løvtræer.	Thicket of broadleaf trees	Vegetation	Broadleaf tree cover, low density	-
38.	Krat af Naaletræer.	Thicket of needleleaf tree	Vegetation	Needleleaf tree cover, low density	-
39.	Plantage.	Plantation	Vegetation	Tree cover, planted	-
40.	30 i Forbindelse med 31.	30 in connection with 31.	Soil/ground condition and vegetation	Permanent grass vegetation in combination with natural vegetation on peat soil.	-
41.	31 i Forbindelse med 32.	31 in connection with 32.	Soil/ground condition and vegetation	Natural vegetation in combination with heath on peat soil	-
42.	30 i Forbindelse med 31 og 32.	30 in connection with 31. And 32.	Soil/ground condition and vegetation	Combination of semi-permanent grass cover with natural vegetation and heath on peat soil.	-
43.	Skrænt	Slope	Ground condition	Slope, hard soil.	+

[3] Definition in text: Bog are areas with peat soil, always to be hatched, except when ploughed or used for grazing. In such cases, they are to be shown as meadow.

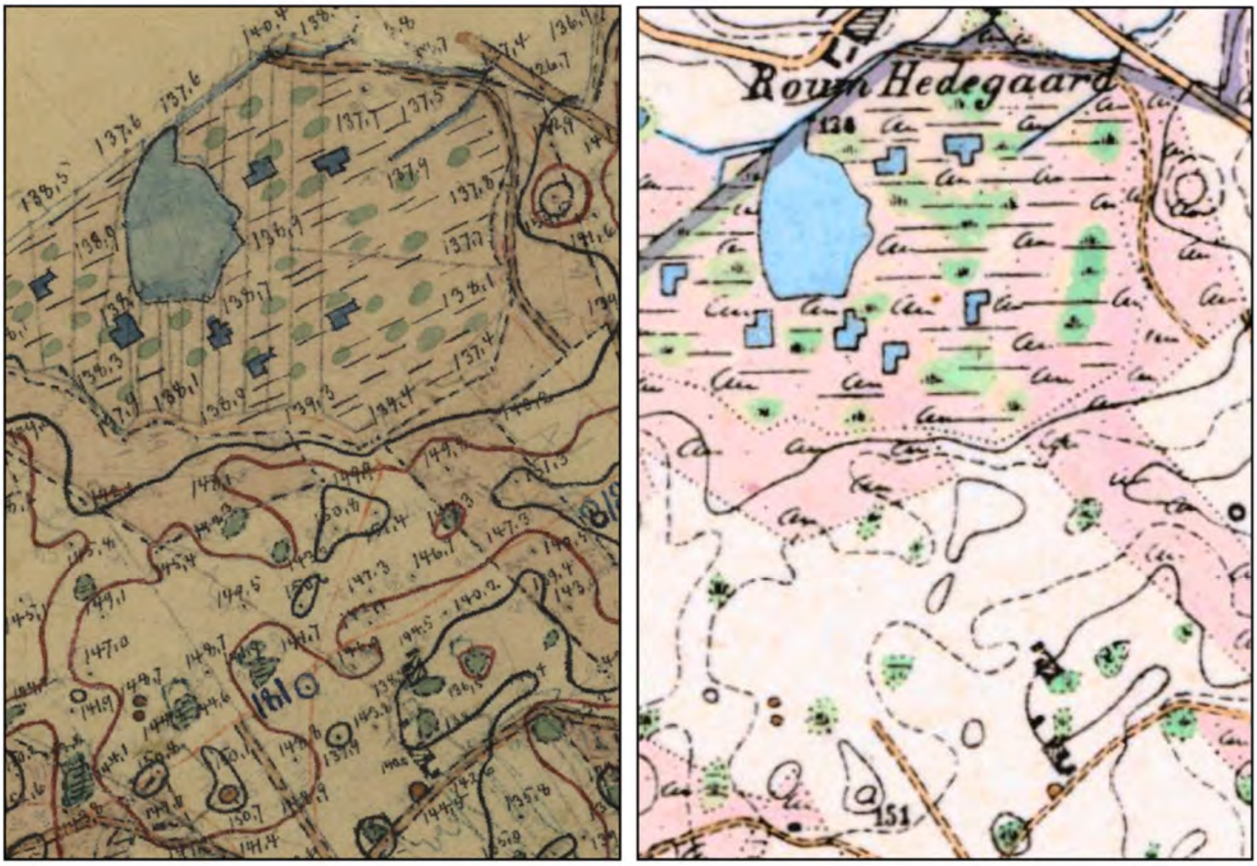


Figure 2.9 A comparison between an extract of survey sheet 12b Hærup By, Klejstrup Sogn, surveyed between 23/6 – 28/7 1879 by Baltzer, verified in the field 29th July 1879. The final version of the HMB sheet N20 Klejtrup, drawn by L. Jensen and published in 1880. Here the heath bog (combination of heath and bog signatures are separated from (dry) heath, which shows how different land areas types were delineated by stippled lines.

In summary, it is our conclusion that the land categories used in the HMB maps seem to be based on an adjusted and simplified version of Olsen's classification system. The detailed classification of ground / soil condition and vegetation seems to have been replaced by a more coarse classification of vegetation and soil / ground conditions, which have an impact on military mobility and visibility. This is the case with the fine-graded classification of soft soil, which was replaced by a single category of especially soft ground ("Sid bund"). In addition, permanent grassland as a vegetation category seems to have disappeared, and grassland is now only represented by the meadow category (grass signatures and green colour). However, the green colour is also used for bogs and marshland, suggesting that the green colour is applied to soft soil and is no longer used for representation of grassland. This is supported by the fact that bogs (which are defined as areas characterized by peat soil) are to be represented as meadow if used for grazing or ploughed. This addition makes sense, if the meadow symbols designates a land use / vegetation cover, and the green colour indicates soft soil and / or increased wetness.

However, several additional types of land area categories seem to be included in the maps and the corresponding legends. Thus, based on a review of the representational practices in the HMB maps and corresponding legends, a number of 'basic' symbols and colours can be identified. Table 2.4 presents the 'basic' form of the different signatures and colours used to represent land area categories in the HMB maps. The different signatures and colours can be combined, to represent several mixed categories, which is also indicated in the list of land area categories presented in Table 2.5, but several additional combinations have been identified. Although several combinations of land area categories can be identified, it is possible to group these combinations into following seven overarching land area categories, which are characterised by their own distinct representational practices:

- Water: Represented with blue colour (both on full-coloured version and three coloured version), and designates areas with permanent open water surfaces.
- Wetland: Represented with a range of different symbols (meadow, bog) and a green colour, and designates areas characterized with soft soil (often organic), and with potential seasonal flooding.
- Dune sand: Represented with a pattern of dots and yellow colour (for sand related to gravel pits), this class designates areas characterized with loose soil in the form of sand.
- Heath: Represented with symbol for heath and a rose colour, and designates areas covered with heather or similar dwarf shrub vegetation.
- Forest: Represented with symbols for broadleaf and needleleaf trees and thicket as well as a brown colour, and designates areas with tree cover.
- Steeply sloping ground: Represented with a symbol for slope (cross-lines between the height isolines), and designates areas with steeply sloping ground.
- Rock: Represented with symbol for rock (dense pattern of double shading), and designate areas with rocks (mainly at Bornholm).
- Built-up / urban: Represented with symbol for gardens / urban area (dense oblique lines), and designates urban areas and gardens.

Table 2.5 Overview of the relation between signatures as well as colours and the different land area categories.

List of signatures and colours	Combinations of colour and signatures.	Overarching land area category
Meadow	Soft soil (green), heath (light pink)	Wetland
Dune sand	Soft soil (green), heath (light pink), sand pit (yellow)	Dune
Bog	Soft soil (green), heath (light pink)	Wetland
Reeds	Water (blue)	Water bodies
Heath	Soft soil (green), heath (light pink)	Heath
Broadleaf trees	Soft soil (green), heath (light pink), forest (brown)	Forest
Needleleaf trees	Soft soil (green), heath (light pink), forest (brown)	Forest
Thicket of broadleaf trees		Forest
Thicket of needleleaf trees		Forest
Slope	Forest (brown), heath (light pink)	Slope
Soft soil	Soft soil (green)	Wetland
Rocks		Rocks
Peat extraction	Water (blue)	Water bodies
Marsh land	Soft soil (green)	Wetland
Buildings, urban areas and gardens		Build-up/urban

3 Automated production of spatial datasets of land categories

3.1 Selection of study areas

Our criteria for the selection of study areas were to (a) enable the use of the results of the developed information production method as an indication of their more general applicability and (b) enable LULCC analyses of the resulting land category mappings for a meaningful set of parish extents. The first of these objectives implied choice of study areas that between them included extensive areas of all five target land categories and source map data with differences typical of the range found across the full HMB map series.

The characteristics of the two study areas that we selected are summarised in Table 3.1. Figure 3.1 shows the location of the areas. The Hobro study area, located in northeast Jutland, was formed of the bounding-box of the parishes of Klejtrup, Nørre Onsild, Sønder Onsild, Lindum, Øls and Hvornum in eastern Jutland. This study area of 196 km² includes extensive areas of forest, wetland and heath. The Hirtshals study area in northern Jutland was formed of the bounding-box of the parishes of Hirtshals, Horne and Asdal. This study area of 79 km² includes extensive areas of heather, wetland and dune. Because the HMB sheets do not follow parish boundaries, nine HMB sheets were needed to completely cover the selected parishes in the Hobro study area, while six HMB sheet were needed for the Hirtshals study area. Therefore, data was extracted for all of these HMB sheets, thus creating extended study areas, covering 425 km2 and 13 parishes for Hobro study area and 170 km2 and seven parishes for the Hirtshals study area. Both study areas are covered by HMB source data of colouration and symbol form variation, meaning that the symbols, e.g. for lyng vary, in form (curvature, slant) and how distinct they are (some are hardly visible as the full form).

Table 3.1 Overview of selection criteria and key parameters for the two study areas.

Selection criteria for study areas	Hobro study area	Hirtshals study area
Land categories included in the study area	Water, Wetland, Heath and Forest	Water, Wetland, Dune, Heath and Forest
Number of parishes included in the study area	5	3
Number of parishes in extended study area	7	13
Size of study area	196 km^2	79 km^2
Size of extended study area	425 km^2	170 km^2
Number of HMB sheets included	6	9

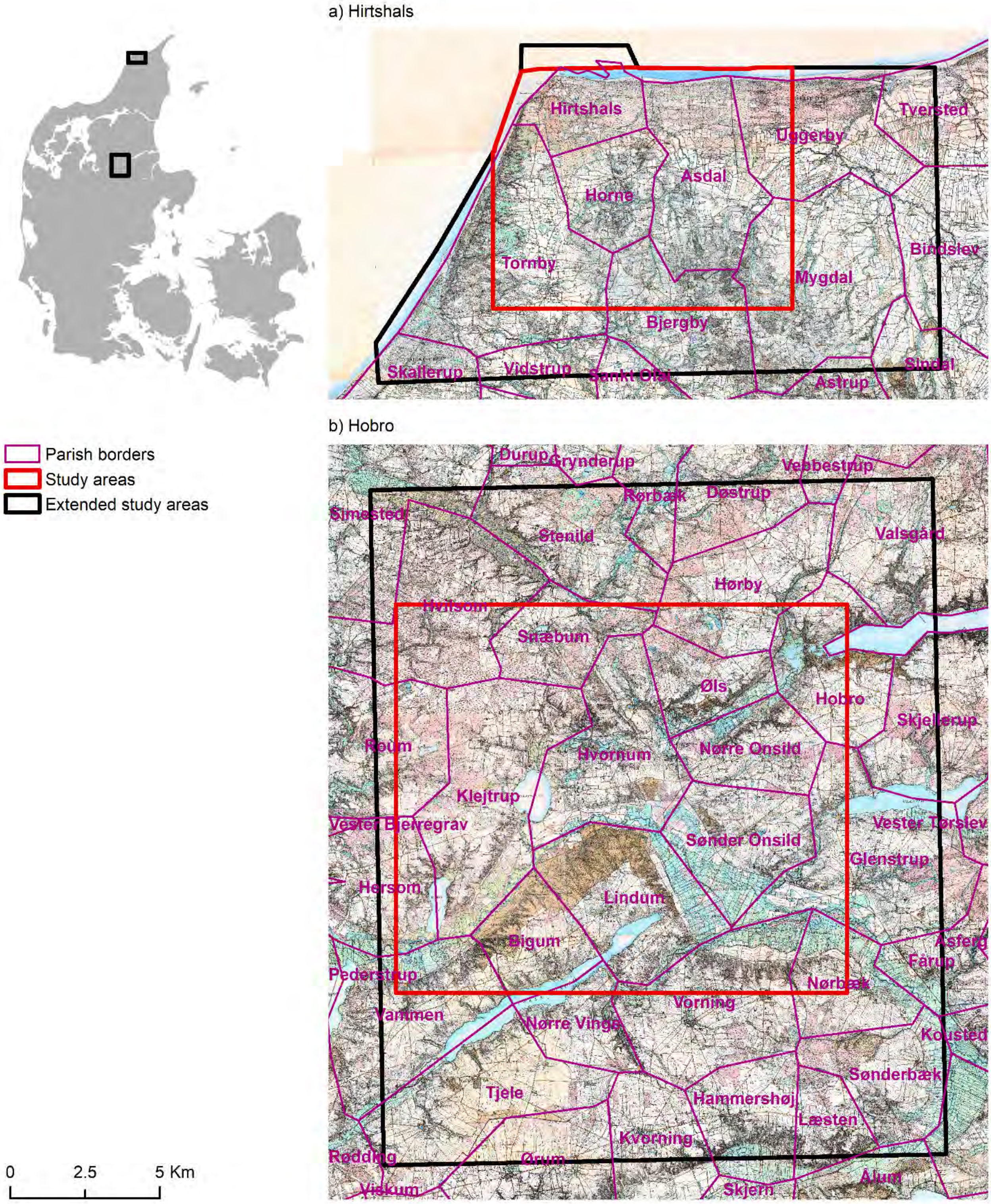


Figure 3.1 Overview of the study areas and extended study areas.

3.2 HMB sources

There are two forms of HMB source material that this work has had available and has used. One is from the digital archive of the Danish national mapping agency (Styrelsen for Dataforsyning og Effektivisering, SDFE / Agency for data Supply and Efficiency), of scanning's made of one set of HMB paper maps. These source data are publically available via the SDFE "Kortforsyningen" web portal (https://download.kortforsyningen.dk/ accessed May 2019) (Agency for Data Supply and Efficiency, 2019) as georeferenced

(ETRS_1989_UTM_Zone_32N, resampled to a pixel size of 2 m) image data that have been mosaicked and then cut into a set of 20 x 20 km tiles as .tif (8-bit unsigned, LZW lossless compressed) files (hereafter, referred to as the "SD" image set). In this project, the actual SD archive source material used was tile files previously acquired by Aarhus University and organised as an ESRI ArcGis national raster mosaic dataset. The other source is the archive of HMB paper sheets held by the Danish Royal Library (Royal Danish Library (KB)) and are original paper sheets, including their margins.

The KB archive paper map sheets and the map sheets that were used for the SD images are not entirely the same raw material. Firstly, only approximately two-thirds of the sheets in the KB collection are full-colour (i.e. tint-coloured, by hand, post printing) with the remainder (mainly eastern Denmark) being just 3-colour (black, blue, brown), whereas the entire set of SD image files are full colour. Where maps of both sources are full-colour, as is the case for both of the study areas of this work, there are places where the local application of colour differs, indicating that the SD and the KB represent two, independently hand-coloured, sets of the paper maps. Differences between the data of the two sources present some local interpretation challenges but, besides of that, the situation represents one of beneficial complementarity for the automated digitisation.

Use of HMB map items from the KB archive required scanning and geo-referencing, which was applied to just the map sheets needed in order to cover the project's study areas. Scanning of the KB sourced map sheets was done using a ScannTECH 600i-fb, with the standard settings of that scanner for colour artwork. A digitisation resolution of 600 dpi was used, which is appropriate for 1:20,000 scale source material (Tobler, 1988). The resulting .tif file sizes were mainly between 320 and 330 MB, but with some maps sheets, giving files of 430 – 440 MB.

Both of the project's study areas are within parts of Denmark for which the KB source HMB are available in the full-colour form rather than the three-colour form (as is the case for most of Zealand and about half of Fyn). It was therefore possible in the automated information production methods that were developed and applied for this project, to compensate for localised poorer map source quality in the SDFE sourced HMB maps with full-colour source HMB maps held by the KB.

3.3 Geo-referencing and geographic structuring of the mapping

Geo-referencing of the .tif files of the scanned KB map sheets was made using the ESRI ArcGis 10.6 geo-referencing tools. Tie-points were set between the not geo-referenced .tif file and the geo-referenced raster mosaic of the SD images. Notwithstanding the colour differences noted above and differences related to the different applied scanning resolutions (coarser in the case of the SD images), the source (KB) and the reference (SD) images are essentially the same (i.e. one common piece of cartography), so finding good tie-points was relatively straight-forward. Typically between 30 and 80 tie-points were applied with a spline fitting model. Resampling of the KB sourced image data was made using nearest neighbour resampling (in order to not form pixel values not present in the original scan image) to a pixel size of 1 x 1 m. A pixel size of 1 x 1 m was applied as that matches the cartographic scale of the HMB maps (1:20,000) and the applied scanning resolution.

The geo-referenced KB files included the map sheet borders. The border parts were clipped-away, using the ArcGis raster clipping tool and a handmade vector polygon (ESRI shape format) of the non-border portion of each sheet. The same polygon was then used to clip the corresponding portion out of the SD sourced raster mosaic, also resampling (nearest neighbour) that to 1 x 1 m pixels. However, the two post-clip image products do not have exactly the same geographic coverage : the SD clip is from a national mosaic but the KB clip coverage is that of just the source paper map sheet extent, scanned, and slightly rotated by its georeferencing and thereby leaving wedges without useful data (Figure 3.2). Simple raster processing, using ArcGis Reclassify and Raster Calculate tools were applied to form an additional raster (8-bit 1 band .tif) with a pixel value 0 for the part with both KB and SD HMB map data and a pixel value of 100 for the parts comprising the KB side wedges. That additional raster was used as a no-data image in the subsequent processing for automated production of the target category geo-data.

Figure 3.2 Illustration for a map-sheet (H341) of the difference (left) between the KB geo-coded and map margin clipped image data extent, with the raster extent out-lined in black and the part of that raster with actual KB data outlined in red, and (centre) the data extent of a corresponding clip of the SD mosaic, that was resolved in the automated processing using an image that defined as no-data the non-overlapping extent (right).

One HMB map sheet extent, with pixels of 1 x 1 m is typically more than 40 million pixels (i.e. larger than 6000 x 7000). As part of the image processing for production of the target category map information, many additional full size image layers are made, and at several points in the process a set of single pixel objects for the full image is required. Even with 32 GB RAM, the processing of a full-size map sheet can give PC memory problems that cause the processing to end abruptly without completion (i.e. “crash”). The work was therefore structured as a partitioning to smaller sub-sections for each map sheet. Typically, these “map-sheet-parts” (MSPs) were the northwest (NW). northeast (NE), southwest (SW), and southeast (SE) quadrants, defined based on the standard Danish 1 km grid (“kvadratnet”), with 1 km overlap, E-W and N-S, between sub-sections (Figure 3.3). The smaller MSPs have a size of approximately 12 million pixels and the larger a size of approximately 19 million pixels. Production of the no-data images (described above) was made at the level of the MSPs. In the cases of several of the map sheets in the Hirtshals study area, an alternative sub-division was used (Figure 3.4):

- For map sheets H158, H324 and H185, they being coastal with relatively small land extents, no sub-division was applied.
- For map sheets H323 and H184 MSPs SW, SE and North (NN) were used, without overlap between NN and either SW or SE, on account of a marked

spatial colour variation in these sheets (cf. the later discussion in this Section 3.5.1).

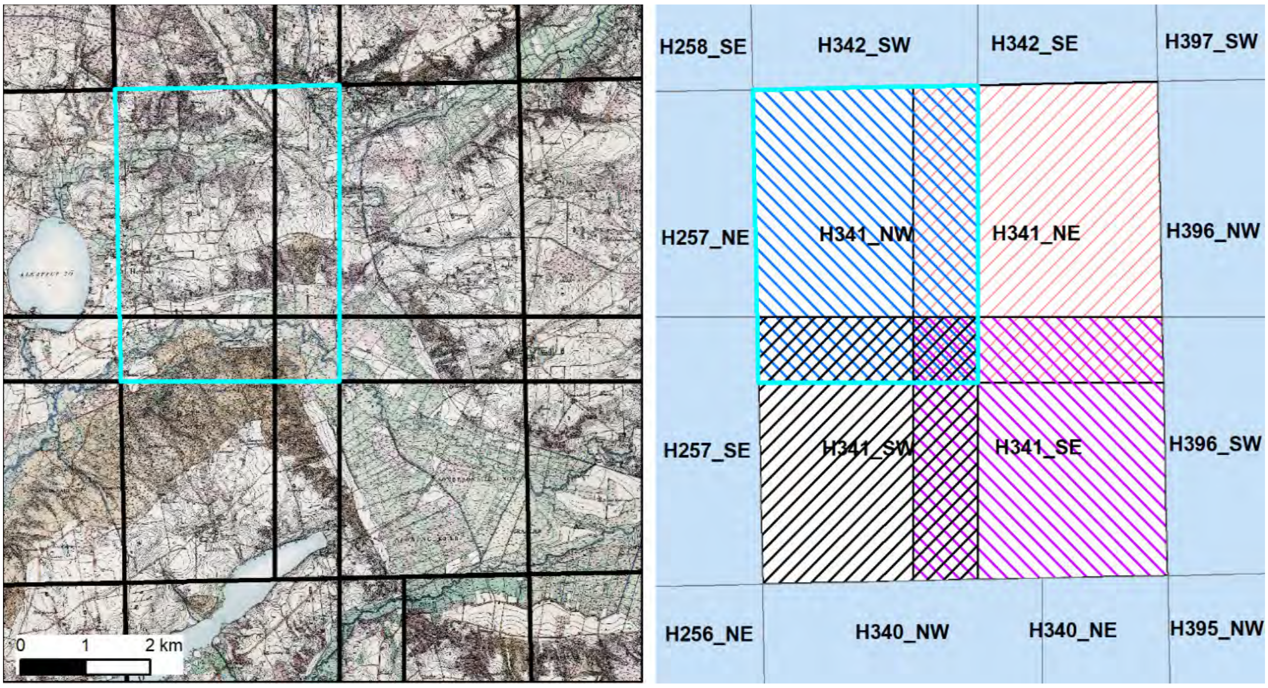


Figure 3.3 Illustration of a standard set of NW, NE, SW,SE map-sheet parts, here shown (right) for the extent represented by the KB archive map sheet H341, with the map-sheet part H341_NW selected (cyan).

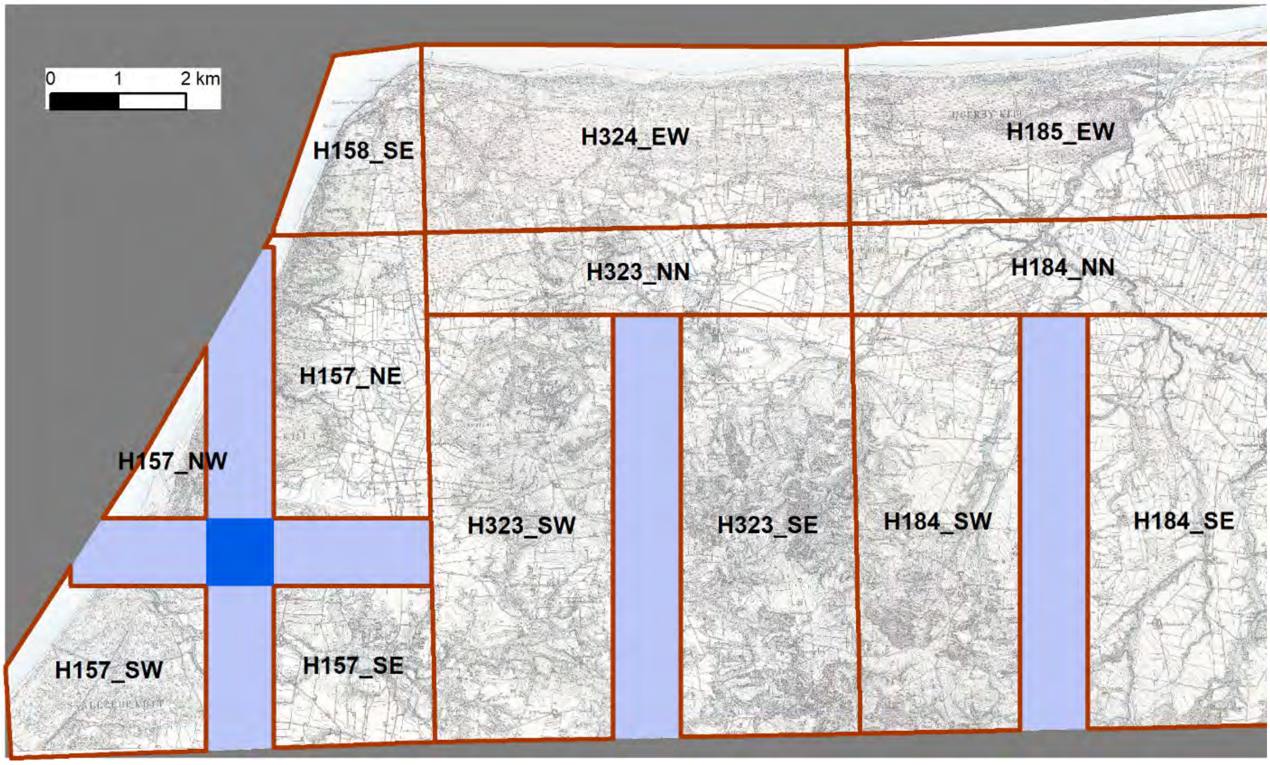


Figure 3.4 The map sheet parts applied for the Hirtshals study area, with 2-way overlaps shown in light blue and a 4-way overlap shown in dark blue.

The presence of map sheet part overlaps are significant in that it resulted in there being either two or four independently generated mappings of target categories within the overlap zones. Ideally, there should not be target category mapping differences related to map sheet part, so checks made of the multiple mappings in the overlap zones served as an additional form of, within-the-workflow, control of the quality of the automated mappings method.

3.4 Automated production of HMB land category digital spatial datasets

Mappings of the target land categories are made from the HMB data via a sequence of processing operations applied to the image data within a single software environment that is purposeful to object based image analyses (OBIA). The OBIA operation sequence for the five target land categories comprises many thousands of individual operations that are organised as distinct modules and code blocks. In the OBIA operations, the part associated with the no-data pixel value in the no-data mask image is effectively ignored.

The overall workflow (Figure 3.5) includes the data geo-registration and geographic structuring described above followed by the automated digitisation of the target categories, plus:

Beforehand, that the quality of the SD and KB source data is visually assessed to check for the possible need for non-standard partitioning of the map sheet and / or application of non-standard workflow mapping modules, and,

Afterwards, that each of the produced target category mappings are visually assessed against the source SD and KB data for the overall quality of the automated mappings.

Typically, these visual assessments should each involve approximately a few minutes. If a need for a non-standard partition of the map sheet is determined, then between 15 and 30 minutes additional time is typically required for data preparation. If a need for a non-standard workflow mapping module is determined, the time required for development and testing of an alternative mapping module can be between 5 minutes and over an hour. Typically, that comprises a need for just use of alternative derived image layers. The likely benefit of that and the choice of derived image layer can be assessed, visually, by interrupting the workflow following the production of the derived image layers. The time required arising from visual assessment of the resulting target category mappings is dependent upon the applied level of punctiliousness, i.e. the applied success criteria, since it is likely that the resulting target category distribution will not match the source HMB legend category distribution 100 %; furthermore, ambiguities between the KB and SD data will give rise to an element of subjectivity in the assessment. However, should a target category mapping be deemed to be unsatisfactory, additional time, quite possibly more than one hour, will be required to determine the spatial pattern of the failures and to develop and test a more successful non-standard workflow module.

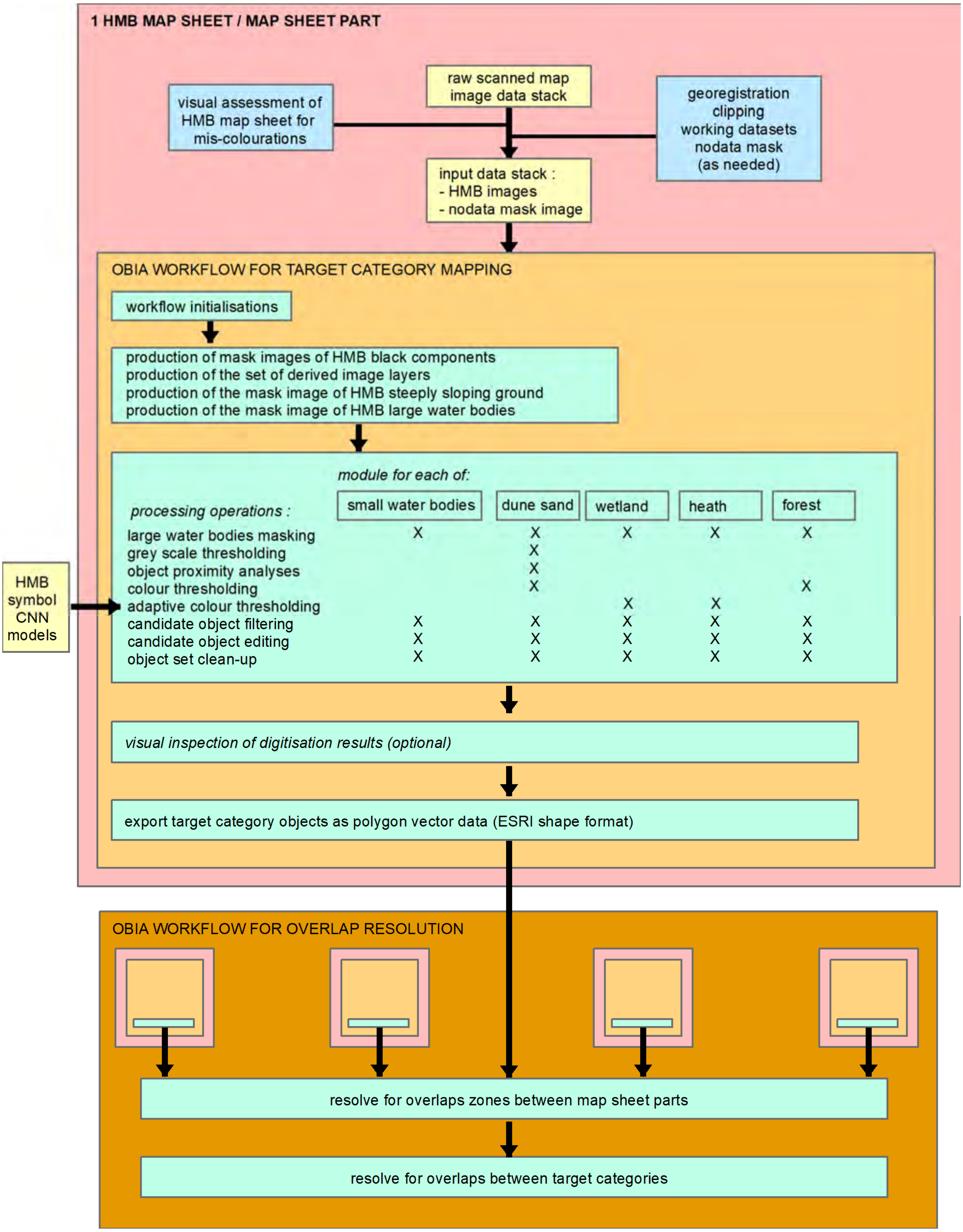


Figure 3.5 Overview of the automated digitisation OBIA workflow.

3.4.1 OBIA operations

Within the applied OBIA software environment a single data processing operation comprises:

a. Selection of the type of operation (e.g. segmentation, object label assignment, object merging, etc.),

b. Setting of the domain that the operation should be applied over (e.g. image pixels, or objects that meet the conditions of a set of filters), and
c. Setting of operation specific parameters (e.g. the label to apply upon objects that pass the filter defined in the domain setting).

The choice of an operation to use, the domain setting and the parameters settings have in this work been set based on expert knowledge, i.e. empirically from the image data, using a heuristic (trial and error) approach. The only exception to that has been certain points within the sequence, where variables derived by one part of the sequence are automatically applied later on in the sequence as domain and parameter settings, the principle example of that being the automated setting of segmentation thresholds via an automated mapping of target category symbols (described below). The applied solutions do not use methods that are commonly used for image data classification, such as parametric or nonparametric statistical models with or without classifier training, apart from the use of training data for the convolutional neural network model based prediction mapping of HMB map symbols.

Image object orientated mapping for a target category, in general, can be simplified to a process involving a mix of two basic forms of operations: (a) operations that either create or modify the shape of objects and (b) operations that change a label or a variable value that an object has. Typically, an object-based image analysis solution, for a target category, comprises an alternating sequence of type (a) and type (b) operations (Figure 3.6). The general rules of good OBIA practice are, that it is often beneficial to break the mapping of a target category down into sub-modules for mapping of subsets of target objects (i.e. "divide-and-rule"), and to focus initially on the sub-parts of the target category that are most clearly distinguished from the non- target category parts of the image. Many temporary subsets of objects, with re-cycled temporary labels and variables, are generated, and subsequently discarded, as part of the process.

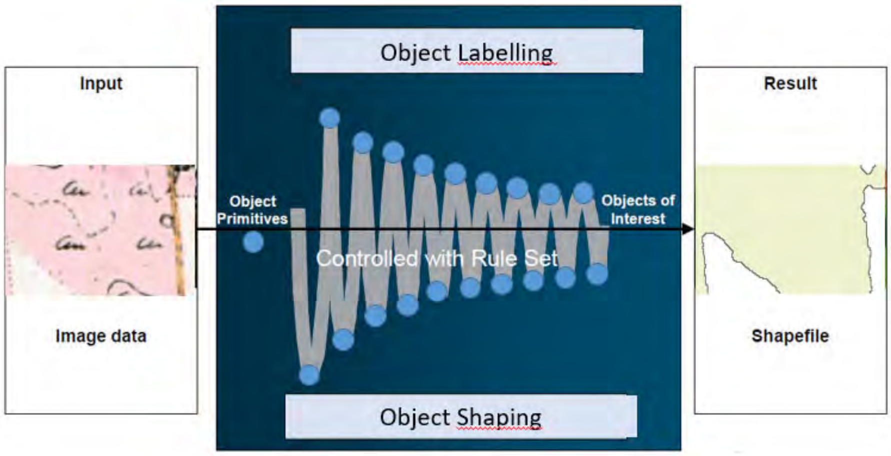


Figure 3.6 Typical OBIA flow between object forming and object labelling operations. (modified after original, © Trimble Inc.).

The full OBIA operations sequence also included numerous standard programming structures (e.g. Do-Loops, If-Then-Else blocks) that manage the running of modules and sub-modules. A singular sequence of operations was applied for the majority of the map sheet parts. For some map sheet parts, for some target categories, variations of the standard workflow were applied to address non-standard aspects of the source data, as noted earlier.

It is not practical or particularly meaningful to present the full processing sequence, operation-by-operation, as there are many thousands of individual operations, and since a small number of generic OBIA operations were applied repeatedly throughout the full sequence, albeit with customisations for the context they were applied to. These most frequently used OBIA operations are each described in their generic form later.

Within the sequence, two key generic modules were applied prior to the target category specific modules (see Figure Susie). One of these was for production of a set of derived image layers. The other was for production of sets of objects that map key content components of the HMB maps. These two generic modules provide vital enhanced flexibility and efficiency to the overall processing sequence. They too are described in detail later.

The mapping of each target category was made as an independent module, meaning that it was possible that the overall processing sequence produced a mapping to more than a single target category for the same piece of land. This way of working acknowledged the reality of the HMB maps, with the location of target categories shown as both colouration and symbols. Thereby situations exist where there is at best ambiguity as to the intended interpretation of the surveyed land characteristic. This way of working also respected that two forms of the HMB maps are utilised in this work, being that from SDFE and the sheets held by KB. Whilst line work and symbols are largely consistent (i.e. it is believed that both the SD and the KB forms originate from a common set of blank-white prints), the patterns of colouration, such as in areas with a mosaic of wetland and heath, do in several locations differ between the SD and the KB data. The generation of independent mappings of each target class therefore enabled the generation of mixed categories extents, i.e. extents where there is marked target category ambiguity. The mixed category mappings generated in this work do not represent a result of a spatial simplification. The final mappings of pure and mixed categories, also taking account of mapping differences in areas with map sheet part overlaps, was done by a second OBIA workflow that took as its inputs the mappings produced by the target category mapping modules of the first OBIA workflow. That second workflow is described later.

3.5 HMB data challenges for the automated land category mapping

The OBIA solutions for the mapping of the target categories have had to be developed such that they work in spite of numerous challenges associated with specific characteristics of the HMB map paper sheets, i.e. ways in which the HMB data detract from a “perfect” situation. The principle such issues are:

3.5.1 Variation in colouration

Variation in colouration is present between the SD scanned image data and the KB held map sheets, between sheets, and within sheets; this is illustrated for the target category wetland in Figure 3.7. Colour variations are probably related to the use of several individuals for the hand colouring, the colour vagaries of the paint media, and the different external conditions (e.g. temperature, humidity, sunlight) that different sheets / sheet-parts have been subject to since. The provenance of the SD data with respect to the source and the current location(s) of the paper sheets that were scanned, and the applied scanner, is unknown. This makes it difficult to more fully investigate possible causes of colour variations in the SD data.

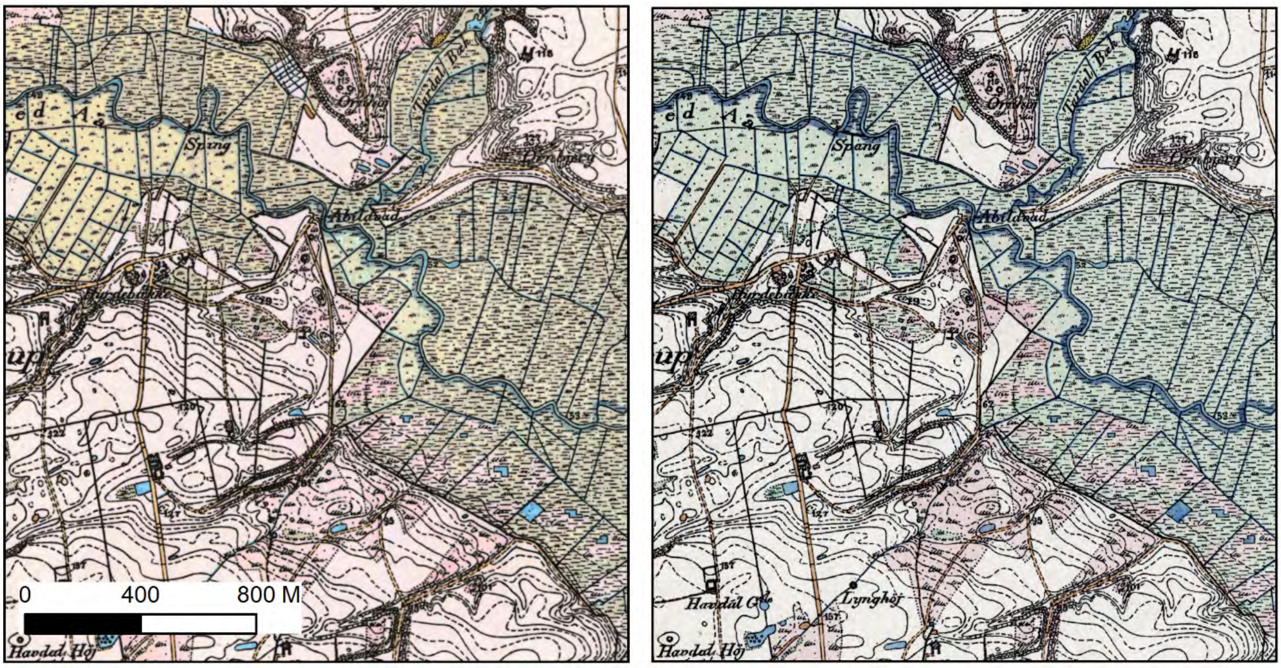


Figure 3.7 The SD image data (left) and the KB image data (right) for a part of the KB archive map sheet H258, showing a case of within map sheet colour variation for the “wetland” category in just the SD data. Note also the variation in the colouration of the background, both within the SD sheet and between the SD and KB sheets.

It is possible to speculate that differential degrees of exposure to sunlight may be one significant cause of SD data colour variations that occur single map sheet extents (i.e. as determined from KB archive), since some parts of the SD data have approximately rectilinear extents with variation present also in the non-coloured “background” parts of the maps (Figure Jane). That this project undertook the work in terms of the individual KB map sheets was a response to the complexity of colour variation in the SD data, with, in most cases, the colouration within a single KB sheet being consistent in the KB data, and also in the SD data. Colour variations within a working extent have a marked negative affect on the possibility to apply operations, such as image colour thresholding with constants as the threshold values. It was therefore, as noted above, necessary in some cases to work with sub-parts of the KB map sheets other than the nominal NE-NW-SE-SW divisions, in order to have working parts with relatively consistent colouration patterns, as for example for H323 and H124 with a smaller northern sub-part without overlap to the SW and SE parts (see Figure 3.4). Image data enhancements such as histogram equalisation / normalisation provided a means for assessing the severity of colour variations (Figure 3.8) and determining the requirements for using more customised within map-sheet sub-divisions.

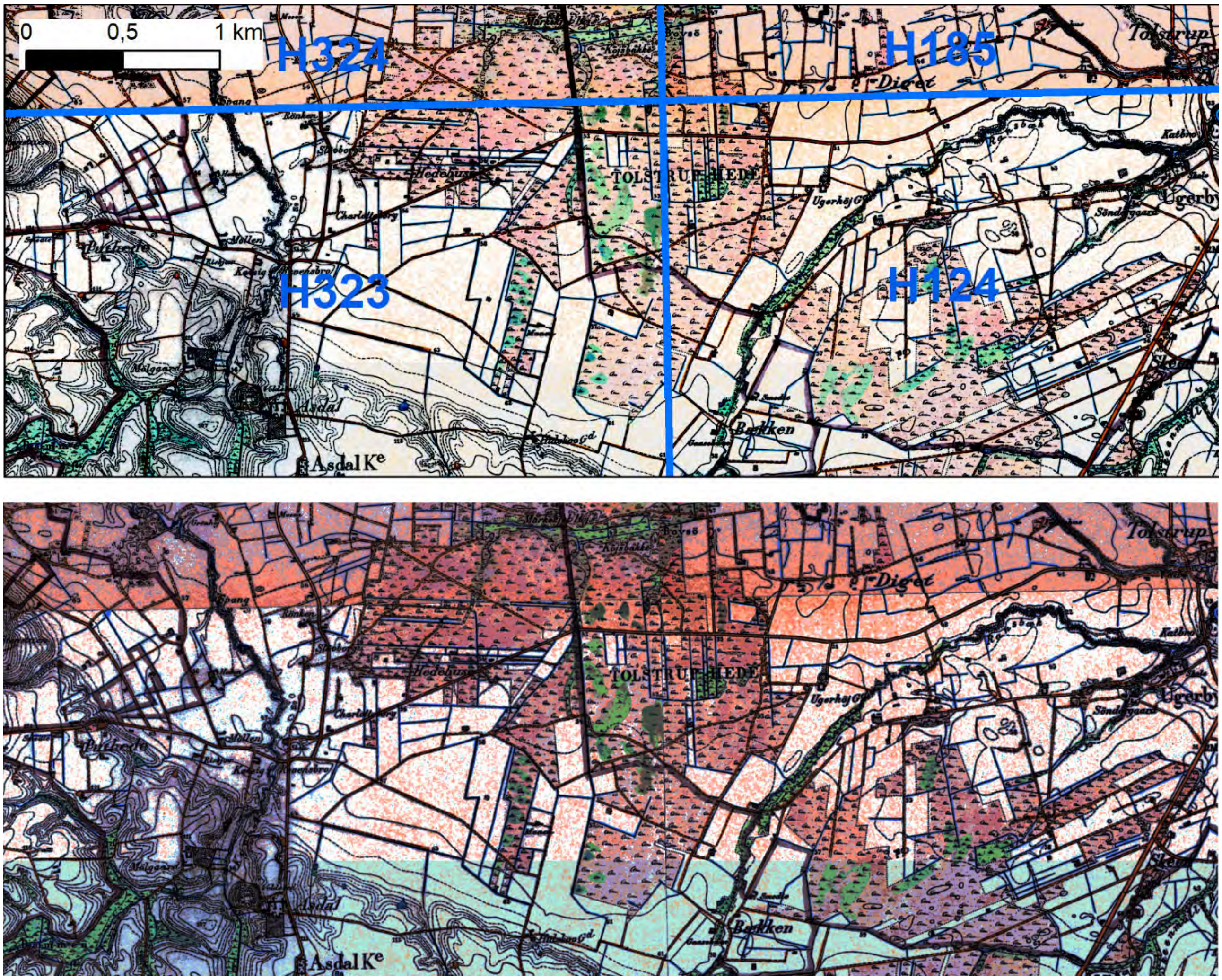


Figure 3.8 Illustration of a rectilinearly distributed background (i.e. matrix) discolouration (light orange speckle and blotches) in parts of the SD data. Upper: the SD data with a 1 standard deviation stretch overlaid with the edges and shelf mark numbers from the KB archive: the discolouration of the background parts is faintly visible. Lower: the SD with a histogram equalisation stretch applied: the colour variation between map sheets is enhanced (e.g. between the parts corresponding to the KB archive sheets H323 and H324) but also the colour variation within sheets, such as within the parts corresponding to the KB archive sheets H323 and H184.

3.5.2 Variations in symbol form and quality

HMB map symbols for the target category heath and the target category wetland (i.e. one of the several specific symbol forms used in the context of "mose, eng, marsk") vary with respect to their exact form and the fidelity of the marks they comprise (Figure 3.9). This is also so for other HMB legend symbols, such as the two forms of symbol used for needleleaf and broadleaf forest; (the symbols for forest were not used in this work, but in further method development they could well be applied in a similar way as the symbols for heath and wetland). Modern AI numerical analysis methods, such as convolutional neural network (CNN) models, are able to operate effectively in spite of marked target variability, providing that the variability is covered by the training samples that are used to form the CNN models. As already noted, it is believed that the SD and the KB data originate from a common original cartography, and no difference in the placement of symbols has been noted to indicate otherwise. However, the general fidelity of the symbol marks is less sharp in the SD data than the KB data, since the former originate in raster data with coarser cell size, and possibly, but not confirmable, a lower fidelity scanning. Therefore, CNN models were formed with samples gathered just from the KB data, and for the symbol mapping the CNN models were applied upon just the KB data. Mappings of the heath and the wetland symbols were used as an assistance to the mappings of heath and wetland that was, as is described later, based in use of the colouration of areas of these categories (Groom et al., 2020).

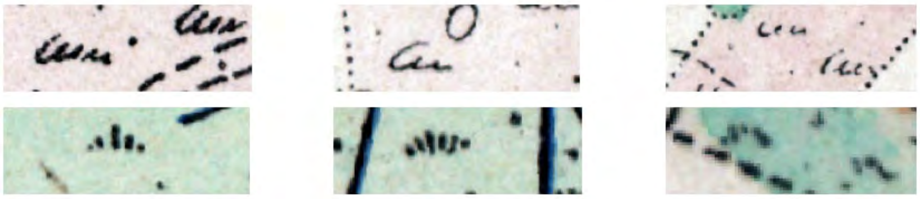


Figure 3.9 Examples of the variation in the Heath (top) and Wetland (bottom) symbol forms in the HMB data of the test areas.

3.5.3 Overlap of target category HMB colouration with black HMB content

Overlap of target category HMB colouration with black text, line-work or symbol marks is a further complicating factor. (For work with the 3-colour HMB map data, this complication would also include the brown height isolines.) To form target category objects that fully cover the interpreted extent of a category, the parts with included black HMB elements have needed to be isolated and filled-in using mixtures of object domain constrained object reshaping, object re-labelling, object merging, etc. operations (Figure 3.10).

Figure 3.10 Illustration of text (place names) overlapping a map theme (water body) colouration (left) and the associated completed mapping of the water bodies (right).

3.5.4 Administrative boundaries marked in the HMB maps

Narrow bands in shades of brown and blue that represent boundaries between administrative units (Figure 3.11) after the colouration of the target categories are a further, albeit rather localised, complicating factor. Attempts were made to use the historical DIGDAG data (Digitalt atlas over Danmarks historisk-administrative geografi, 2019) to isolate these but significant local differences between each of those datasets and the boundaries as they are expressed on the HMB maps have precluded use of the DIGDAG data in that way. In many cases, the administrative boundaries comprise long straight parts, making their manual digitisation possible, but significant other parts are convoluted, as for example where a border follows a river course. Attempts were made to automate mapping of the boundaries in the HMB maps, but as they were included in the HMB maps such that they do not fully obscure the underlying map detail, the problem is complex and developing a full solution was not possible with the available resources. Having a full digitisation of the borders shown in the HMB maps would be a major assistance for increasing the quality of the mapping of the target categories.

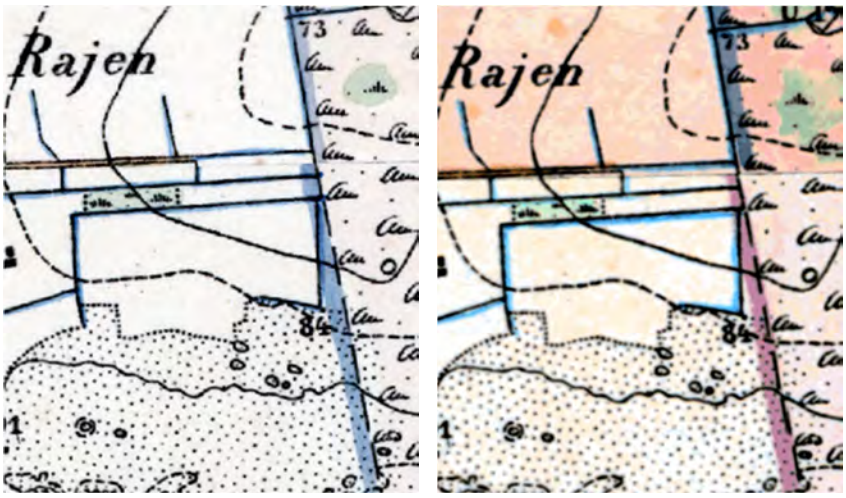


Figure 3.11 Illustration of the representation of administrative boundaries in the HMB maps, with use of different colouration between the KB (left) and SD (right) versions of the map image data.

3.5.5 Forest and road colouration similarity

In both the SD and the KB data, in full colour forms, there is high colour similarity between the colouration applied for forest areas and the colouration applied for roads (Figure 3.12). This represented a particular challenge for avoidance of the false-positive mapping of roads as the forest target class, which was resolved by making an initial mapping of "not forest" that included roads, as is described in Section 3.6.5.

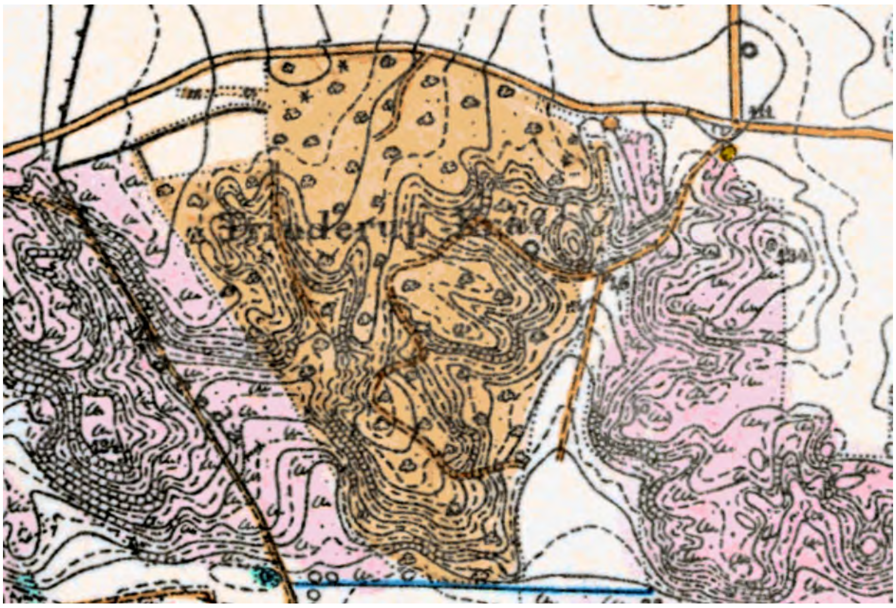


Figure 3.12 Illustration, from map sheet H341, SD data, of the colouration similarity of forest and roads.

There are also locations where the SD and the KB data express a quite different reality via their use of colouration (Figure 3.13), possibly because line work items have been over-looked or misinterpreted during the hand-colouration process. These cases can never be mapped correctly in relation to both the SD and KB data, and complicate the process of accuracy assessment. Fortunately, they are not numerous.

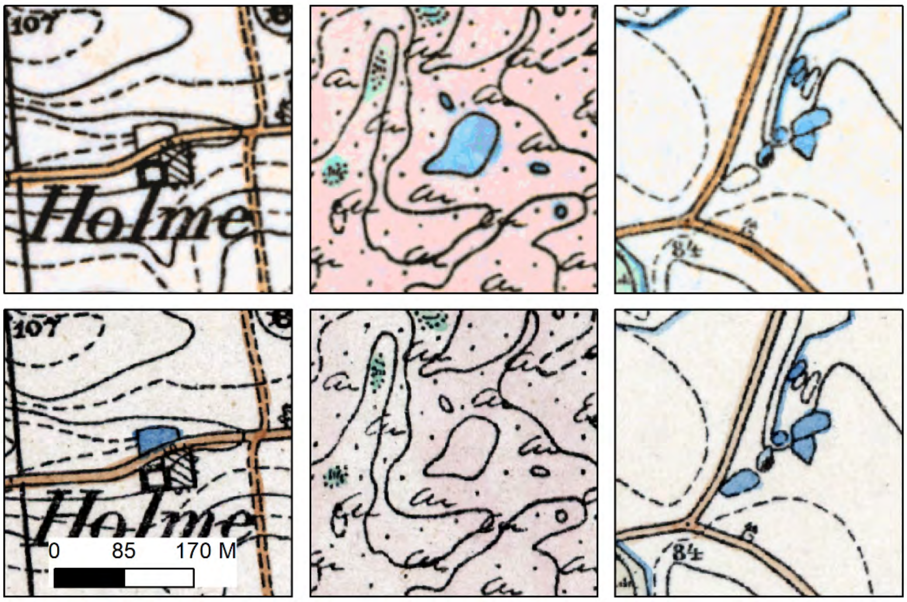


Figure 3.13 Three examples of map differences for water bodies between the SD and the KB HMB source material, related to subjectivity in the hand-colouration. Up-per row: SD; lower row: KB. Left and centre examples are from sheet H342, right example is from sheet H184. Same scale in each example.

3.6 Land category mapping methodology: detailed description

The detailed description of the methodology that follows is structured as:

- Description of applied set of derived image layers
- Description of prior mapping of key HMB components
- Description of the key generic object based image analysis
- Description of the use of convolutional neural networks
- Description of the key aspects of the mapping modules for each of the five target categories
- Description of integrations for map sheet part overlaps and between target category mappings

3.6.1 Derived image layers

Production of a set of derived image layers was a key initial generic module for subsequent flexible and effective implementation of the target category specific mapping modules. Derived image layers are OBIA processing sequence temporary raster layers that each represent an enhancement of specific tonal (i.e. black-white) or colouration properties of one or more of the six (i.e. 3 SD R-G-B + 3 KB R-G-B) source image layers. As described earlier, the scannings made of the paper HMB map sheets produced digital images that each comprise three 8-bit unsigned integer image layers (also often called "bands" or "channels"), which represent, for each pixel, the density of red, green and blue in the analogue source material, i.e. the paper maps. So, for example, a light blue colouration used for an area of open water results in image data that, for the water body pixels has a low density (low pixel value) for the red image layer, a higher density (higher pixel value) for the green image layer and a higher still density for the blue image layer. Thus, the image is basically a set of pixels with red-green-blue (RGB) value triplets. Base pixel values can be changed in many ways to give new images that enhance the usability of the image data for an applied purpose, such as mapping the target categories.

The derived image layer set used in this work comprises layers resulting from combinations of kernel based spatial domain filters, image pixel value statistical normalisations, and application of arithmetic expressions involving the image layers. The derived image layers are a major component of the applied mapping solution since they represent enhanced image data for specific target categories (Figure 3.14) and thereby enabled application of simple image layer thresholding rules to form initial sets of relevant objects. (The six source R-B-G image layers were not themselves used in the target category specific mapping solutions.) The full set of derived image layers, and their roles in the mapping modules are presented in Table 3.2.

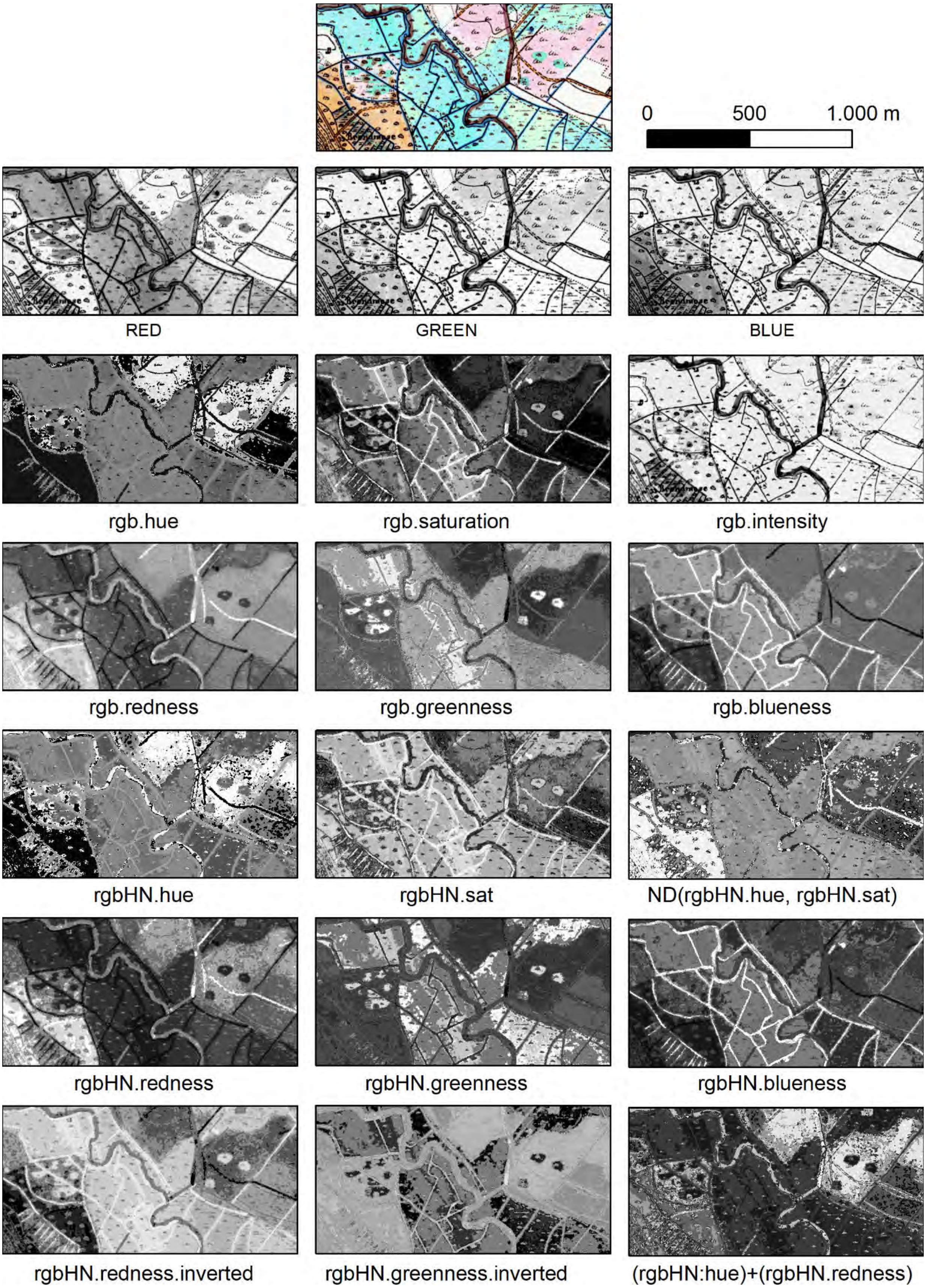


Figure 3.14 Illustration of the ability of derived image layers to enhance different colour components of the HMB maps. The example used is a part of the map sheet part H341_NE, with the SD data. All parts are displayed with application of a 3.0 standard deviation stretch based on the statistics of the dis-played data. HN: histogram normalised. ND: normalised difference. Note the lower degree of differences between HMB categories in the raw red, green and blue image layers (2nd row from top) compared to the derived image layer (rows 3 – 7).

Table 3.2 Key to the standard set of derived image layers and their uses in the automated digitisation, including also the initial raw image layers (row-2). Unless explicitly noted as either “sd” or “kb”, each item was derived for both the SD and the KB image data. Under Uses, both the SD and the KB versions were used as indicated, unless noted otherwise. “dilip” means that an image layer is used as an input for another derived image layer. (The names shown here only approximate to derived image layer names used in Figure 3.14.)

Type of transformation	Applied function/arithmetic expression	Name	Uses in mask or target category mapping modules
None (raw image data)		rR, rG, rB	
Kernel based spatial domain filtering	See text above	fR, fG, fB	
HIS	Hue(rR,rG,rB)	rRGBhue	
HIS	Saturation (rR,rG,rB)	rRGBsat	
HSI	Intensity (rR,rG,rB)	rRGBint	black parts mask large water bodies mask (kb)
HSI	Hue(fRf,fG,fB)	fRGBhue	large water bodies mask (sd) steeply sloping ground mask (sd) heath forest
HSI	Saturation (fR,fG,fB)	fRGBsat	large water bodies mask (kb) wetland (sd) heath forest
HSI	Intensity (fR,fG,fB)	fRGBint	
Histogram Normalisation	Histogram Normalisation (rR)	rRhn	dilip (sd,kb)
Histogram Normalisation	Histogram Normalisation (rG)	rGhn	dilip (sd,kb)
Histogram Normalisation	Histogram Normalisation (rB)	rBhn	dilip (sd,kb)
Histogram Normalisation	Histogram Normalisation (fR)	fRhn	dilip (sd,kb)
Histogram Normalisation	Histogram Normalisation (fG)	fGhn	dilip (sd,kb)
Histogram Normalisation	Histogram Normalisation (fB)	fBhn	dilip (sd,kb)
Arithmetic	rRhn/(rGhn+rBhn)	rhnRness	
Arithmetic	rGhn/(rRhn+rBhn)	rhnGness	
Arithmetic	rBhn/(rRhn+rGhn)	rhnBness	
Arithmetic	fRhn/(fGhn+fBhn)	fhnRness	Large water bodies mask (sd) Small water bodies (kb) Dunes (sd) Heath
Arithmetic	fGhn/(fRhn+fBhn)	fhnGness	Wetland (kb) Dunes (kb) Heath
Arithmetic	fBhn/(fRhn+fGhn)	fhnBness	Small water bodies Heath
HSI	Hue(rhnRness, rhnGness, rhnBness)	rhnHue	
HSI	Saturation(rhnRness, rhnGness, rhnBness)	rhnSat	
HIS	Hue(fhnRness, fhnGness, fhnBness)	fhnHue	Heath (sd)
HSI	Saturation(fhnRness, fhnGness, fhnBness)	fhnSat	Dune (kb) Heath
Normalised Difference	(fhnHue-fhnSat)/(fhnHue+fhnSat)	fhnNDhs	Heath Forest
Arithmetic	1-fhnRness	fhnRnessINV	Wetland (sd)
Arithmetic	1-fhnGness	fhnGnessINV	Heath (kb)
Arithmetic	fhnHue + fhnRness	DILsum1	Heath (kb)
Arithmetic	fhnRness – fhnGness	DILdiff1	Heath (kb)
Arithmetic	fhnBness – fhnGness	DILdiff2	Heath (kb)
Arithmetic	DILdiff1 + DILdiff2	DIFdiffsum1	Heath

In more detail, the three forms, as named above, of pixel change applied in the production of the derived image layers have been:

a) Kernel based spatial domain filtering: applied for every pixel, this replaces the value for a central pixel within a moving window of pixels with a value that is a weighted mathematical function of the window set of pixels. This processing was applied to (1) reduce the degree of local, pixel-to-pixel variation in the HMB colouration and (2) reduce the intensity of fine black lines and other black markings in the HMB maps, in order to enhance the possibilities of using the colouration to fully map the target categories (Figure 3.15). The applied filters were first a maximum value filter with a 3x3 pixel kernel, followed by a median value filter with a 5x5 pixel kernel.

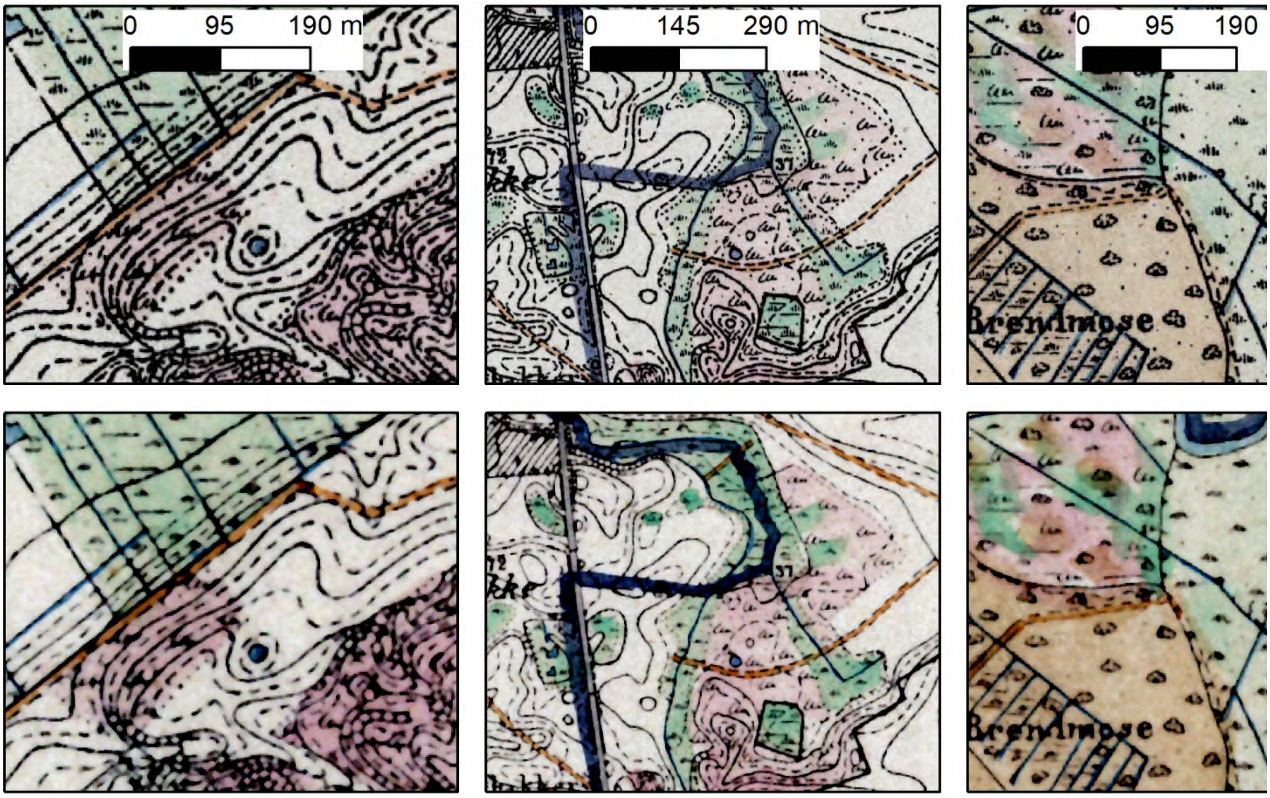


Figure 3.15 Illustration of the effect of the kernel-based spatial domain filtering on the image data. The example used is a part of the map sheet part H341_NE, with the KB data. All parts are displayed with application of a 3.0 standard deviation stretch based on the statistics of the whole image. Top row: unfiltered image data. Bottom row: 3x3 kernel maximum value followed by 5x5 kernel median value filtered image data.

b) Pixel based changes such that each pixel gets a new value based on the relationship of its value, for a specific image layer, to the statistics, for that same image layer, of all the pixels. This has been applied here in the form of histogram normalisation (Figure 3.16) of the source image data in its raw and its spatially filtered forms. The histogram normalised image layers have then been applied for the pixel based arithmetic function changes described below.

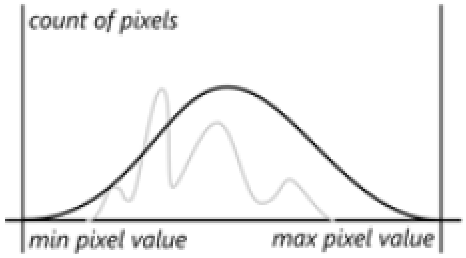


Figure 3.16 Sample of histogram after histogram normalization: raw image data frequency distribution in grey, histogram normalised frequency histogram in black. (Source: Trimble eCognition Developer Reference Book Version 9.5.1, Revision 1.0, December 2019).

c) Pixel-based change such that each pixel gets a new value based on an arithmetic function of its value in a set of input image layers. This type of pixel change has comprised conversion to Hue-Saturation-Intensity (HSI) pixel values, conversion to Redness, Greenness and Blueness pixel values and various arithmetic other operations applied to combinations of these to make enhancements for specific mapping purposes. HSI is a so-called alternative colour space to RGB (Fortner and Meyer, 1997). Redness, Greenness and Blueness are ratios of one colour band to the sum of a set of the other two colour bands, e.g. Redness = Red value / (Green value + Blue value).

The histogram normalisation (i.e. (b)) and the arithmetic function processing (i.e. (c)) were applied upon both the original SD and KB R-G-B image data layers and the spatial domain filtered SD and KB R-G-B image data layers. Furthermore, the arithmetic function processing (i.e. (c)) was applied upon both original filtered and unfiltered R-G-B image data layers and the histogram normalised filtered and unfiltered R-G-B image data layers.

3.6.2 Prior mapping of key HMB content

The second initial generic element was production of sets of objects that map key content components of the HMB maps. These mappings assisted the target category modules.

One such prior mapping was for the black elements in the HMB maps. Many HMB map items use black: text, symbols, height isolines, buildings, curtilages (diagonal line pattern in the full colour printed maps), dune sand areas (evenly spaced dots), ditches (black line along one side), the edges of roads and the edges of target categories either as solid lines (water bodies) or dot-lines (heath, wetland, forest, dune sand). Having all black parts mapped prior to the target category mapping modules was significant for control of the growth of mappings of small water bodies, and for in-fill within otherwise mapped larger areas of heath, wetland, forest and dune sand. Black element mapping was made using the Intensity derived image layer of the SD and the KB and minimum(SD,KB) original (i.e. unfiltered, un-normalised) R-G-B image data. For each of these three inputs, a black mapping was made using a pair of pixel value threshold values.

Another prior mapping was that of off-coast water and larger lakes. This was prescribed since these elements in the HMB maps often have highly variable colourations (e.g. lakes are often hand tinted in blue only close to their shores). Furthermore, there is often marked bi-modality in the overall size range of “blue” water body extents in the HMB maps (i.e. some very large water bodies plus many very small water bodies), which presented challenges for mapping of all water bodies within a single module.

Prior mapping was also made for steeply sloping ground. Forest or heath on steeply sloping ground are characterised by colouration that is broken-up into small areas of colour by the cross lines between the height isolines that are used in the HMB maps to indicate steeply sloping ground. The forest and heath mapping modules each therefore included separate sub-modules for mapping of the categories on flatter ground and on steeply sloping ground. Similarly, dune sand with marked dune development are most effectively detected as steeply sloping ground with proximity to dune sand areas that are otherwise mapped by the dune sand mapping module. It was therefore beneficial to have all steeply sloping ground mapped prior to the target category modules.

Object based image analysis processing steps were developed and applied to get initial mappings of these three components. The mappings of each component (Figure 3.17) were applied in the target category modules as binary image layers in order that they may be utilised there without interfering with the object sets of the target category specific modules.

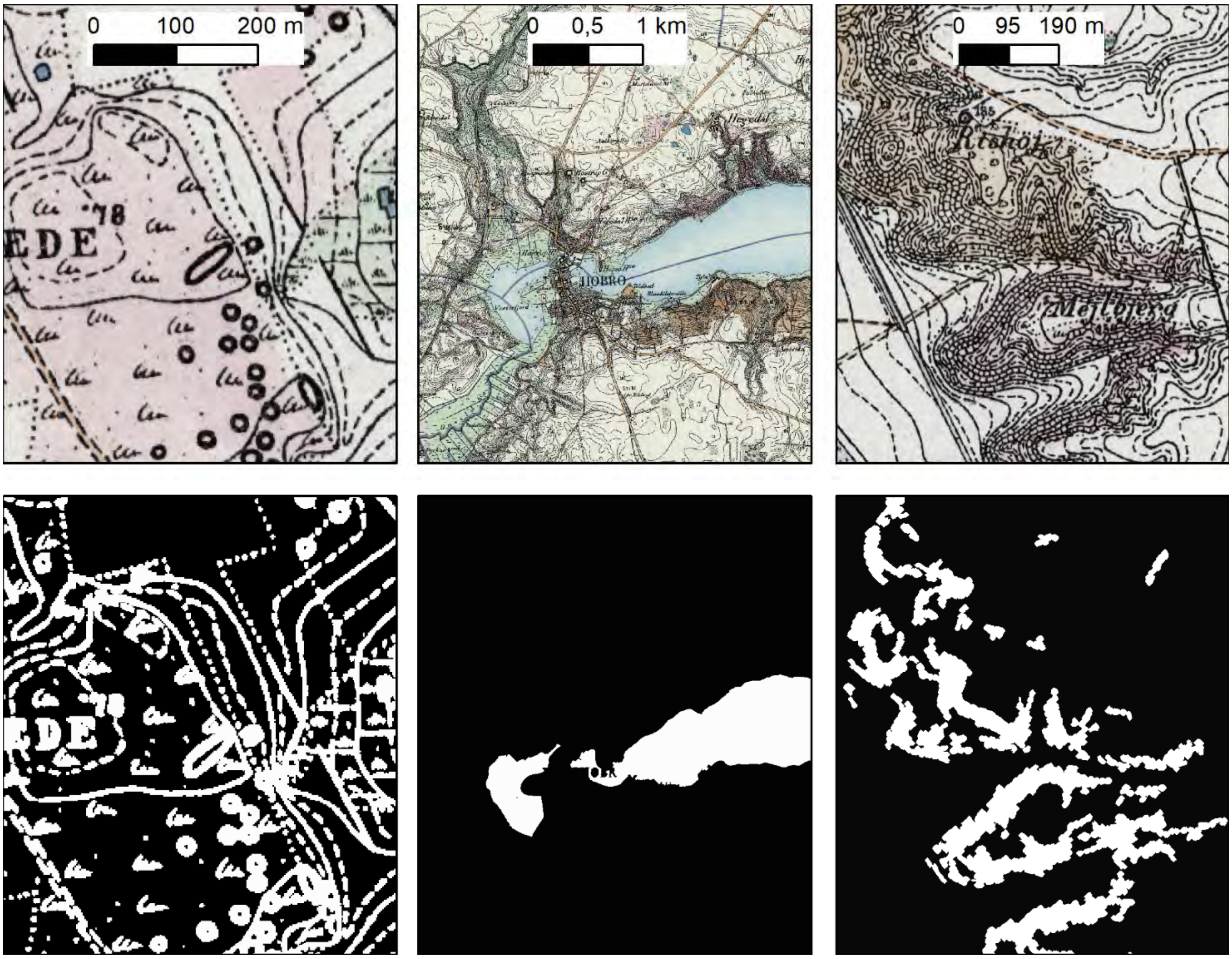


Figure 3.17 Examples of the prior mappings. Right: black elements, example used is a part of the map sheet part H341_NE. Centre: large water bodies, example used is a part of the map sheet part H397_SW. Right: steeply sloping ground; example used is a part of the map sheet part H341_NE. Top row: the KB image data, displayed with application of a 3.0 standard deviation stretch based on the statistics of the whole image. Lower row: pixel value 0 or 100 prior mappings.

3.6.3 The key generic OBIA operations

A relatively small set of OBIA operations was applied repeatedly, with different domain and parameter settings, dependent upon the applied context. These are described below, but only in simplified terms since the applied OBIA environment (Trimble eCognition v. 9.51) provides high levels of flexibility and versatility in the application of each of its operations, and full description of which is not meaningful here (see https://docs.ecognition.com/v9.5.0/ (last accessed 6thMay 2020) for full description of these possibilities.) The most often used OBIA operations were:

- Multi-threshold segmentation: For a domain, which can be either the entire set of image pixels (i.e. no existing objects are present or are retained) or an existing object set/subset, form new objects based on a threshold value(s) of the pixels in one specific image layer, and label the object sets formed from image data with pixel values either side of the threshold value(s) as required.
- Chessboard segmentation: For a domain, which can be either the entire set of image pixels (i.e. no existing objects are present or are retained) or an existing object set/subset, form new square objects of a specified size.
- Assign: Assign operation relabel objects of an object set, with the set being specified in terms of, among other possibilities, object class labels and conditions such as thresholds on object size or object mean pixel value in an image layer or the degree to which an object is a neighbour to specific other objects
- Merging : Merging of sets of objects defined by domain settings (class label, conditions, etc.) that have adjacency
- Pixel-based object re-sizing: this can involve growth, coating (to objects of a new label) or shrinkage (to objects of a new label) of domain defined sets of objects; via its parameter settings these operations can also be controlled in terms of pixels with specific category label, specific image layer value ranges and in terms of desired degree of object geometry smoothing to remove undesirable protrusions and indentations.

3.6.4 Applied method and use of convolutional neural networks

As well as with colouration, the HMB maps carry HMB legend category representation by placement of specific category related map symbols. Two distinctive HMB symbols are (a) between four and seven approximately vertical (map viewed N-S) black line strokes, of increasing height towards the centre of the symbol (see Figure Bill) used in the HMB maps as a symbol of land with difficulty to cross due to wetness, i.e. approximating to the project's target category "wetland", and (b) large right-slanted "C" curves with the base of that part continued to the right as a sequence of smaller right-slanted "C" curves of decreasing height (Figure Bill) that is used in the HMB maps as a symbol for heath ("lyng" – see alse footnote on page 18). Having the distribution of these two symbols in the HMB maps automatically detected was an assistance in the mapping of the full extents of the associated target categories, based on their colourations. As noted above, the colourations vary. Detected symbols provided a means of determining the appropriate threshold, as a variable, to apply to map a category in a specific map sheet (Groom et al. 2020). The resulting target category mappings were far superior to ones made using constants for the thresholds (Figure 3.18). Convolution neural networks (CNN) is a part of deep learning (which is a part of machine learning and in turn artificial intelligence – Figure 3.19) that is well-able to identify the parts of images that match the pattern of image data presented as training examples, particularly where the trained pattern comprises distinctive line work such as small shapes. HMB symbol pattern sampling and CNN modelling was made as one-off operations in a separate workflow, using the same OBIA software environment. The CNN symbol models were then applied for symbol detection and use within the target category mapping modules (Figure 3.20).

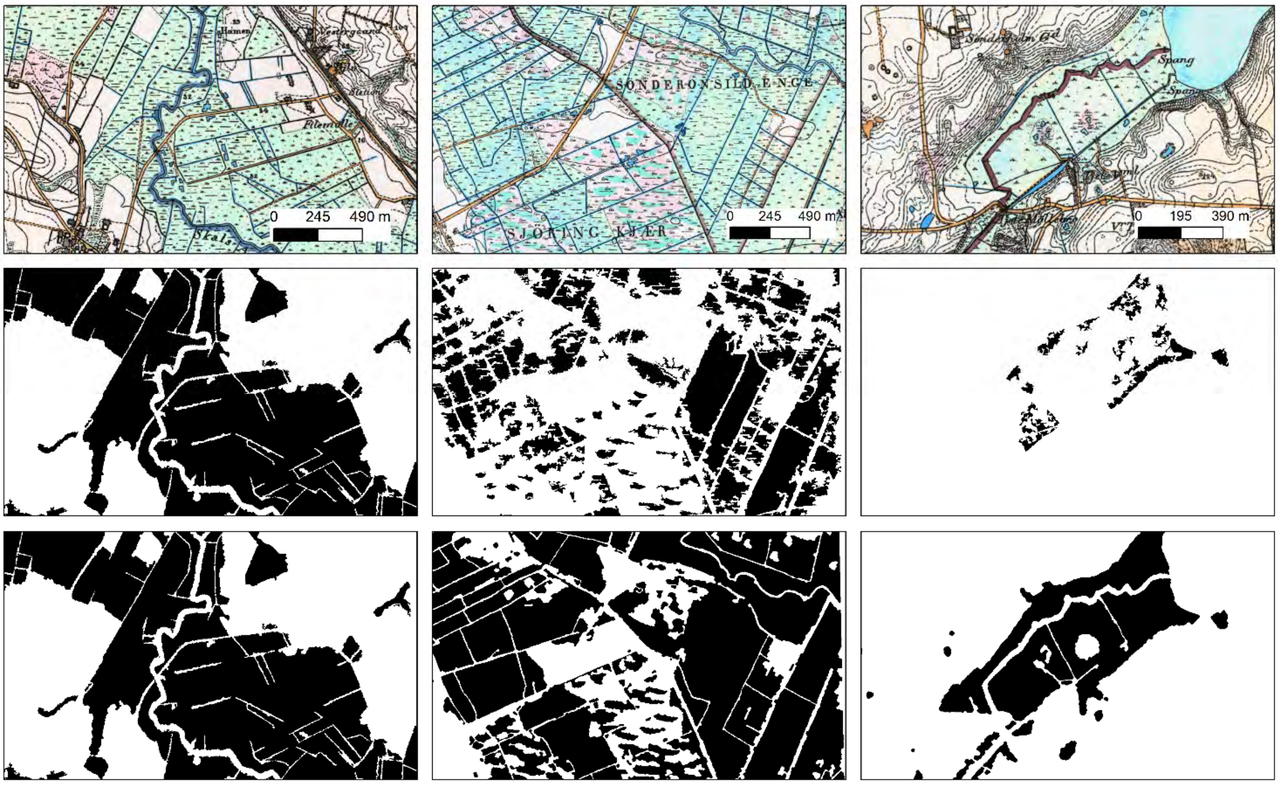


Figure 3.18 Result mapping of the target class wetland for three HMB map sheets (top row) with a constant threshold values for all three map sheets (middle row) and variable thresholds set via use of CNN symbol mappings (bottom row).

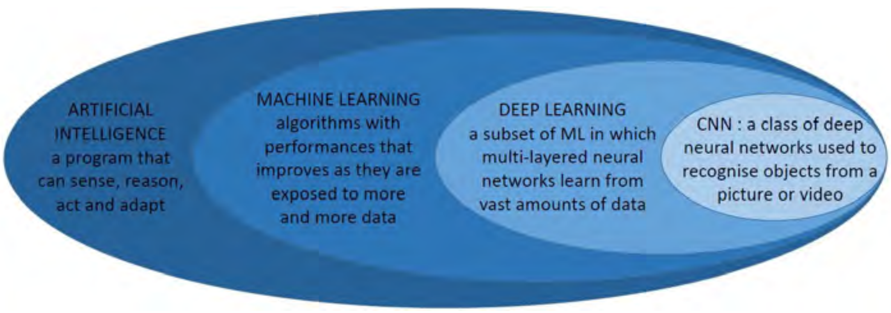


Figure 3.19 Relationship of convolutional neural networks (CNN) to deep learning, machine learning and artificial intelligence.

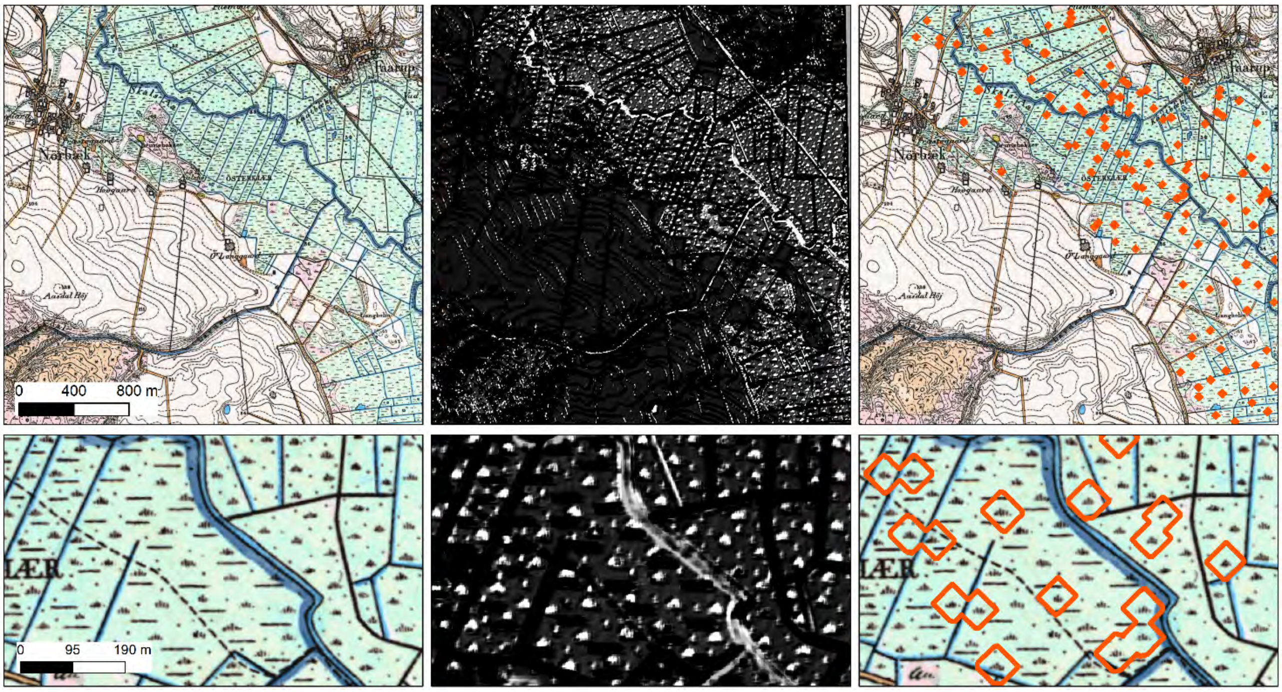


Figure 3.20 Wetland symbol raw CNN "heatmap" (centre) and symbol heatmap hotspot objects, grown and filtered to zone objects (right) to use for threshold variable setting.

3.6.5 Key aspects of the mapping modules for each of the five target categories

Apart from the mapping of forest and dune sand, the solutions for the mapping of the target categories all followed a similar pattern:

- Identify either one or two derived image layers that well isolate the areas of the target category. For many HMB map sheets parts, for a specific target category, the same derived image layers could be taken as a good discriminator for a category (see Table 3.2). The degree to which that was likely to be the case was assessed in advance by visual inspection of the HMB sheets in their R-G-B form. If that indicated that a sheet was anomalous, more detailed visual inspection of the different derived image layers was made prior to running of the target category specific modules, to determine which layers could best discriminate the category.
- Image layer thresholding on the selected derived image layer, to derive an initial set of candidate objects for the target category. For some categories (e.g. for the smaller water bodies) the colouration is sufficiently consistent in all HMB map sheets, with a corresponding consistent derived image layer that represents marked discriminating power, to enable use of a constant value as the applied threshold. For heath and wetland, the colouration is, as noted above, highly variable between map sheets, which required use of a variable threshold value, with the value being derived map-sheet-part independently via use of the CNN mapped symbols associated with the category.
- Refinement of the candidate object set in terms of false positive object exclusion, false negative object inclusion and object re-shaping. This includes the inclusion into the set of category objects, of text and symbols the within the extent of a larger category candidate object, as well as exclusion of smaller initial category objects.

The final set of target category objects, also for forest and dune sand, were then exported by the OBIA workflow as polygons in the ESRI ArcGis shape format. Key aspects of the mapping of each of the target categories is described in the next paragraphs:

Water bodies

Mapping of larger water bodies, inland and coastal, was done as one of the initial stages, as described above. Smaller inland water bodies are present, in the full-colour, as well as the 3-colour HMB maps, as a mid-blue colouration, with black linework defining their extent. The combination of the solid blue colouration and the black linework is a highly distinctive signature for smaller water bodies in the HMB maps. Similar blue colourations are used to mark ditches, but with black linework to just one side. The target category "water bodies" of this work does not include ditches.

The basic method for the mapping of small inland water bodies began with thresholding on a black components image (derived as described above) followed by thresholding over the remainder of the image on a derived image layer that is a strong discriminator of the blue colouration (typically, one of the hnBness image layers). A selection was then made of objects from the second thresholding that have a high relative border to the objects from the first thresholding. This provided a set of candidate objects that could be taken as a final mapping of many of the small water bodies and at least a partial mapping of all of the remainder. Typically, there were also some false positives, in particular some ditches with additional black linework adjacent to the blue. Object filters were applied to remove the false positive small water body candidate objects. Completion of the mapping of the partially mapped small water bodies was made by pixel-based object resizing, accruing pixels of the unclassified objects irrespective of the blueness, but with growth limited by the objects of the black image components (Figure 3.21).

Rivers represent a particular form of water body for the automated mapping process. Many larger rivers are distinct and are detected too, but many rivers were also used at the time of the HMB survey and cartography as administrative boundaries and have, as noted above, the normal colouration for water obscured by the colourations (variously, pink, brown, blue) that have been applied in the HMB maps for administrative boundaries. This resulted in these parts of even wider rivers not being consistently mapped as water bodies by the automated process. Therefore, the mapping of rivers, as opposed to all other forms of water bodies, is incomplete. For some map sheets, the markedly elongated morphology of water body objects associated with rivers enabled isolation of river water bodies, which were then exported to a specific Rivers set of output vector polygons.

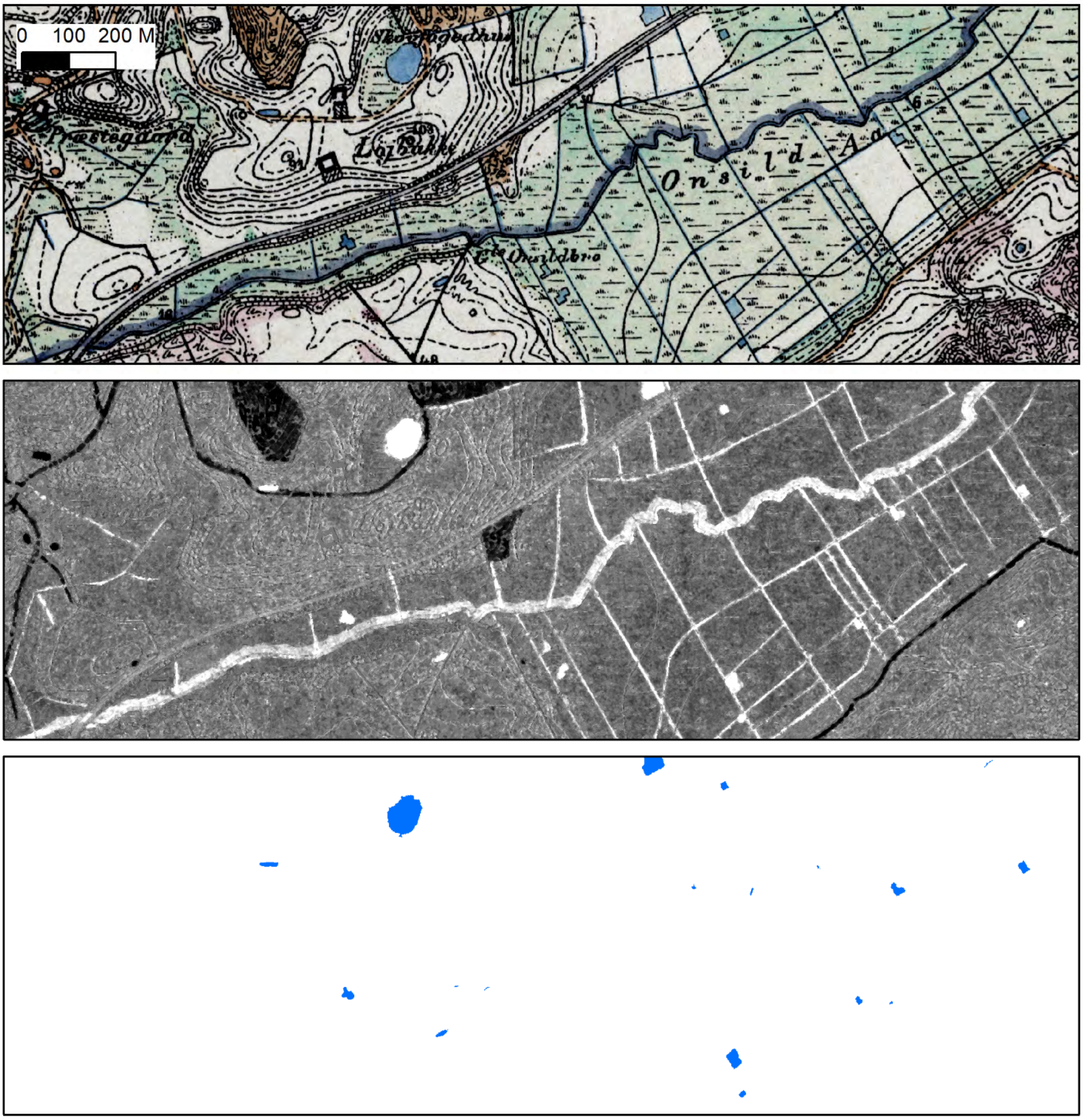


Figure 3.21 Illustration of the mapping of small water bodies. Top: part of HMB sheet H341, KB image data with a 2.0 standard deviation stretch. Middle: spatial domain filtered KB R-G-B, histogram normalised, expression of relative blueness (i.e. kbF.normHis.Bness), with a 3.0 standard deviation stretch. Bottom: result geo-data for small water bodies (blue).

Wetland

In almost all HMB full-colour map sheets, wetlands are highly distinctive in terms of a green or blue-green colouration. However, as already noted, there is also considerable variability in this colouration that disrupts the simplicity of making colour thresholding with a constant. This was resolved via use of image segmentation threshold values that were set adaptively for each map sheet part via use of the CNN mapping of the HMB symbol for "eng" / meadow (Groom et al. 2020) (see Figures 3.18 and 3.20), and independent thresholding on two derived image layers, being one from the SD data and one from the KB data. Routine object set filtering and object geometry editing procedures were then applied upon the initial candidate objects to make the final mapping (Figure 3.22). The distinctive and unique colour used for wetland in the full colour HMB maps precluded occurrence of false positive mapping cases, apart from between wetland and water bodies. Cases of that error were resolved by having parts mapped as small water bodies take priority over parts mapped to wetland, as a part of the second (finalisation) workflow.

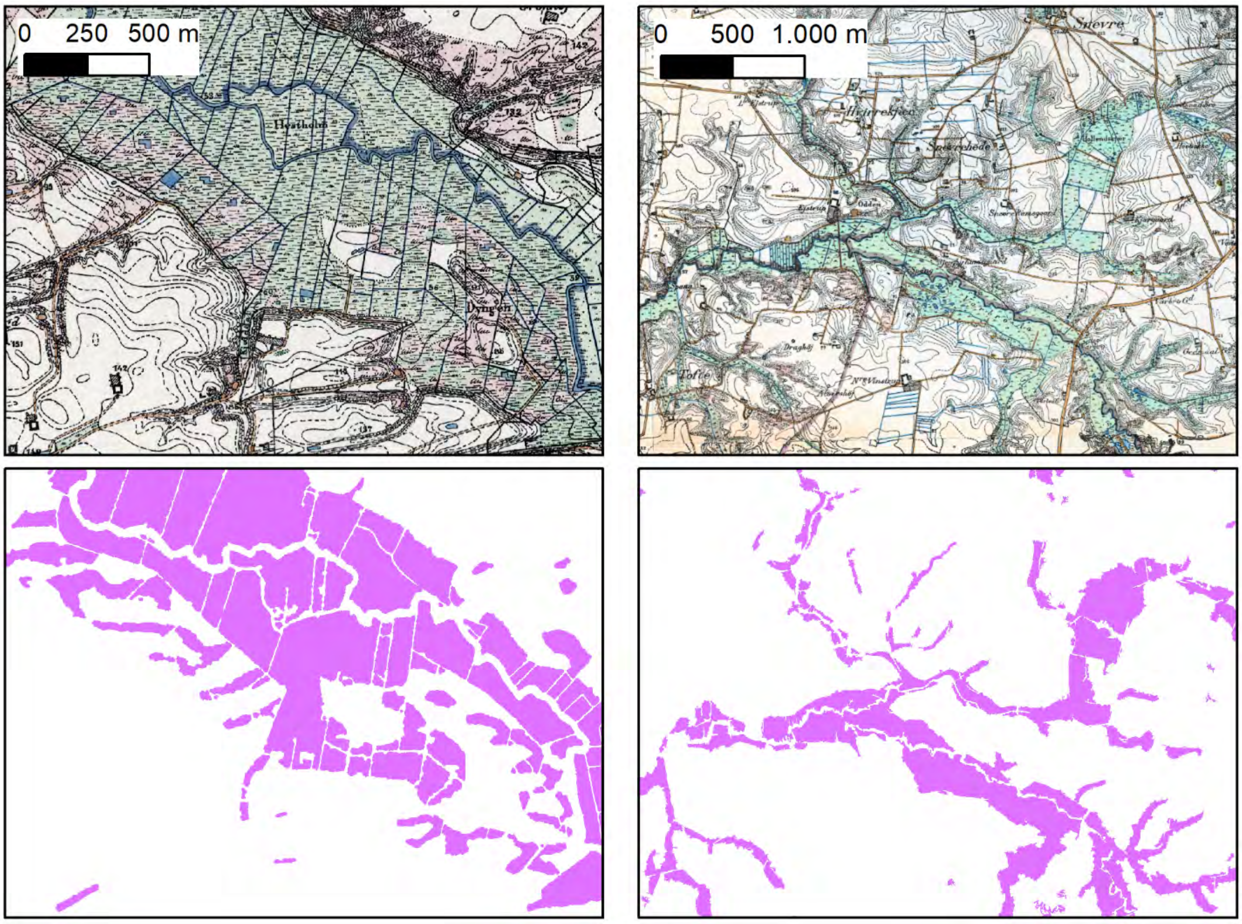


Figure 3.22 Examples of the HMB maps (top) and the final mapping of the target category wetland (bottom, lilac). Left: a part of HMB map sheet H258 (KB data shown). Right: a part of HMB map sheet H323 (SD data shown).

Heath

In almost all HMB full-colour map sheets, heath is highly distinctive in terms of a light pink colouration. However, there is also considerable variability in this colouration that disrupts the simplicity of making colour thresholding with a constant. This was resolved via use of image segmentation threshold values that were set adaptively for each map sheet part via use of the CNN mapping of the HMB heath symbol (Groom et al. 2020), and independent thresholding on two derived image layers, being one from the SD data and one from the KB data.

An additional initial process was applied to make focused mapping of heath on steeply sloping ground. Heath occurs often in the HMB maps in conjunction with steeply sloping ground. As steeply sloping ground is presented in the HMB maps with cross lines between the height isolines, the associated pink heath colourations are small (typically less than 200 pixels) and would otherwise be lost by object size filtering. To reduce the possibility of this type of false negative error, a modulation of the image layer threshold was applied upon objects formed from the image mask of steeply sloping ground made in the prior stages. Each of the resulting small candidate heath objects was then grown across the remaining steeply sloping ground objects to form larger, coalesced heath objects (Figure 3.23). Routine object set and object form editing procedures were then applied upon the initial candidate objects to make the final mapping.

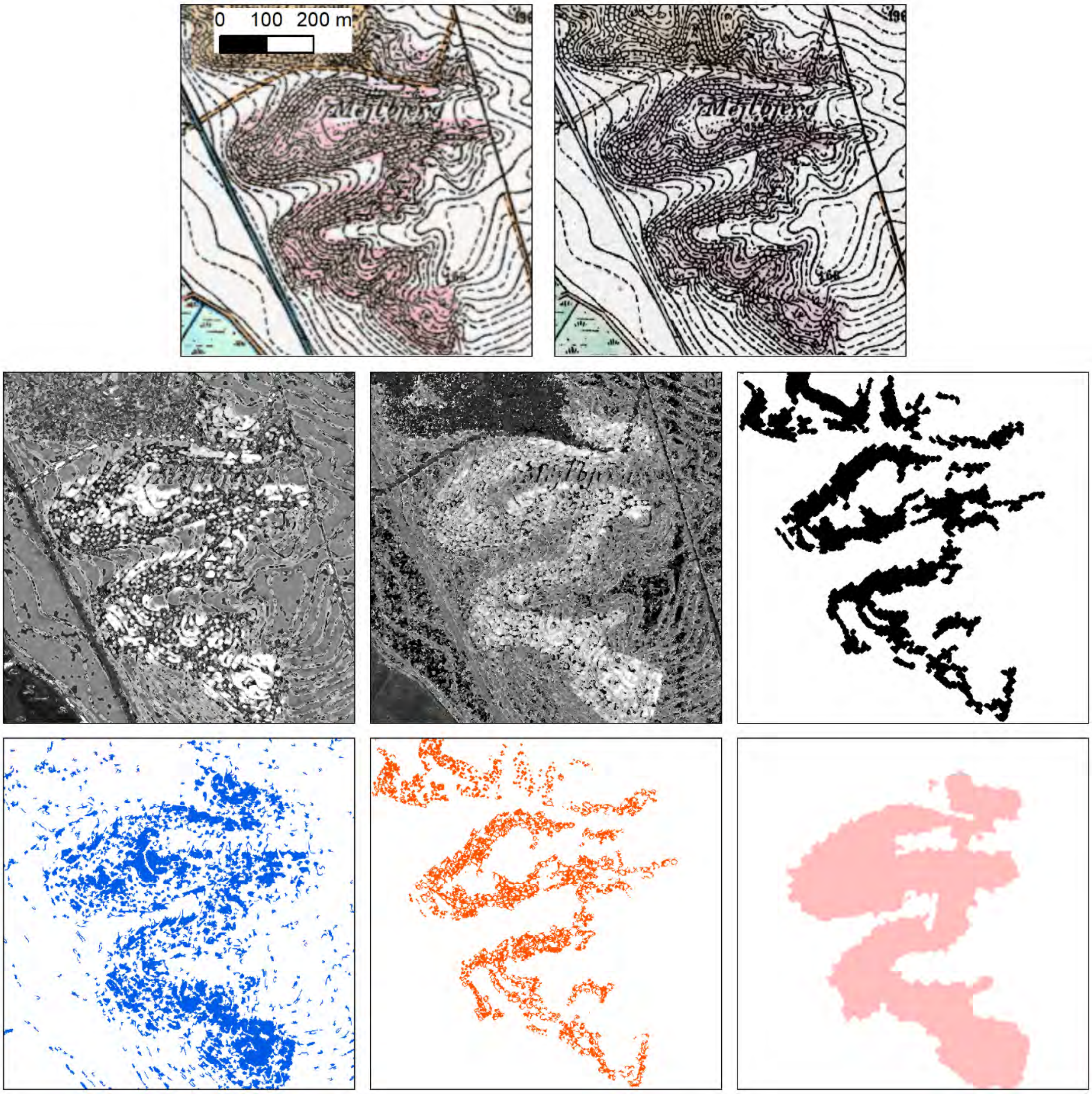


Figure 3.23 Illustration of the mapping of heath on level ground and on steeply sloping ground. Top row: the SD and the KB HMB image data. Middle row, left to right: the applied derived image layers from the SD and the KB raw HMB image data, the derived mask image layer of steeply sloping ground. Bottom row: left, the initial mapping of heath with adaptive thresholding on the SD and the KB derived image layers; centre, the initial heath mapping for steeply sloping ground; right, the final heath mapping. The example is a part of the HMB map sheet H341.

Forest

HMB full-colour map colouration based mapping of forests is complicated by the similarity, in the HMB full colour maps, of the colouration used for forests and that used for roads, in both cases, being orange or light brown. It is further complicated by the widespread occurrence of forests on steeply sloping ground. The applied solution for forest mapping was a reversed form of the standard procedure. Instead of initially mapping for forest, initial objects were made that represent all parts of the map sheet apart from forests and roads. These objects were then grown by a sufficient extent to include roads away from forests. The remaining parts, being just forest plus roads in forests, were then grown to compensate for the earlier growth of non-forest objects, parts on steeply sloping ground were included via use of the image layer of steeply sloping ground, and the forest objects were edited to infill for black line-work text and symbols and height contours (Figure 3.24).

For many of the HMB map sheets, the forest is limited to a few, rather regularly shaped areas, and it was more effective to manually digitise shape format polygons for forest in these map sheets.

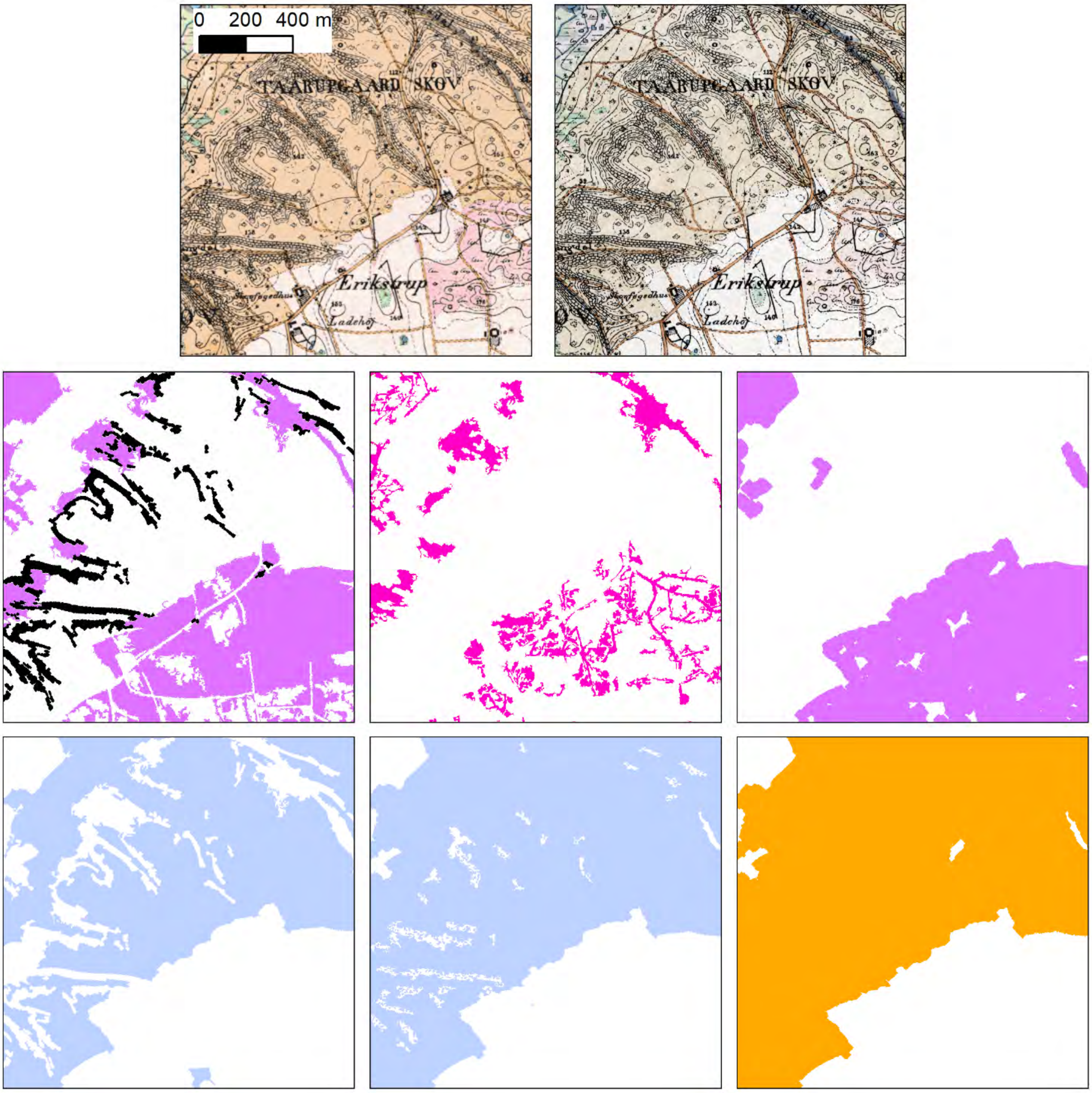


Figure 3.24 Illustration of the mapping of forest. Top row: the SD and the KB HMB image data. Middle row, left: steeply sloping ground (black) and general non-forest matrix (lilac); centre: general matrix parts with forest-like colouration; right: growth of remaining large general matrix objects over forest-like objects. Bot-tom row: left: large unclassified objects (blue); centre, inclusion of enclosed steeply sloping ground and remnant forest-like objects; right, the final forest mapping. The example is a part of the HMB map sheet H257.

Dune sand

Dune sand areas are not in the HMB maps represented by any distinctive colouration, but rather by a spatial patterning of dark dots. The dots are of similar sizes and have an approximately regular spacing; these are the HMB signature features that formed the basis for automated mapping of dune sand areas. Initially, all dark parts in an HSI colour-space intensity image layer were mapped to a set of objects based on a crude threshold value. This object set included text, symbols, height contours, etc. as well as the dune sand dots. Object size thresholding was applied to remove most of the non- dune sand objects. For the remaining dune sand candidate objects, objects with distance to the closest other object of over 20 m were removed. Remaining sand dune dot candidate objects were coated with other objects to a distance of 20 m and for the coating object a floating point raster was made of the distance of each pixel to a dot candidate object. The distance raster was subjected to a threshold based segmentation (value < 11) with the resulting parts merged with the dot candidate objects. The initial candidate object set of the dune sand area includes false-positives. Sets of clear-cut false-positive cases were filtered for via fixed threshold segmentation using other derived image layers (Figure 3.25).

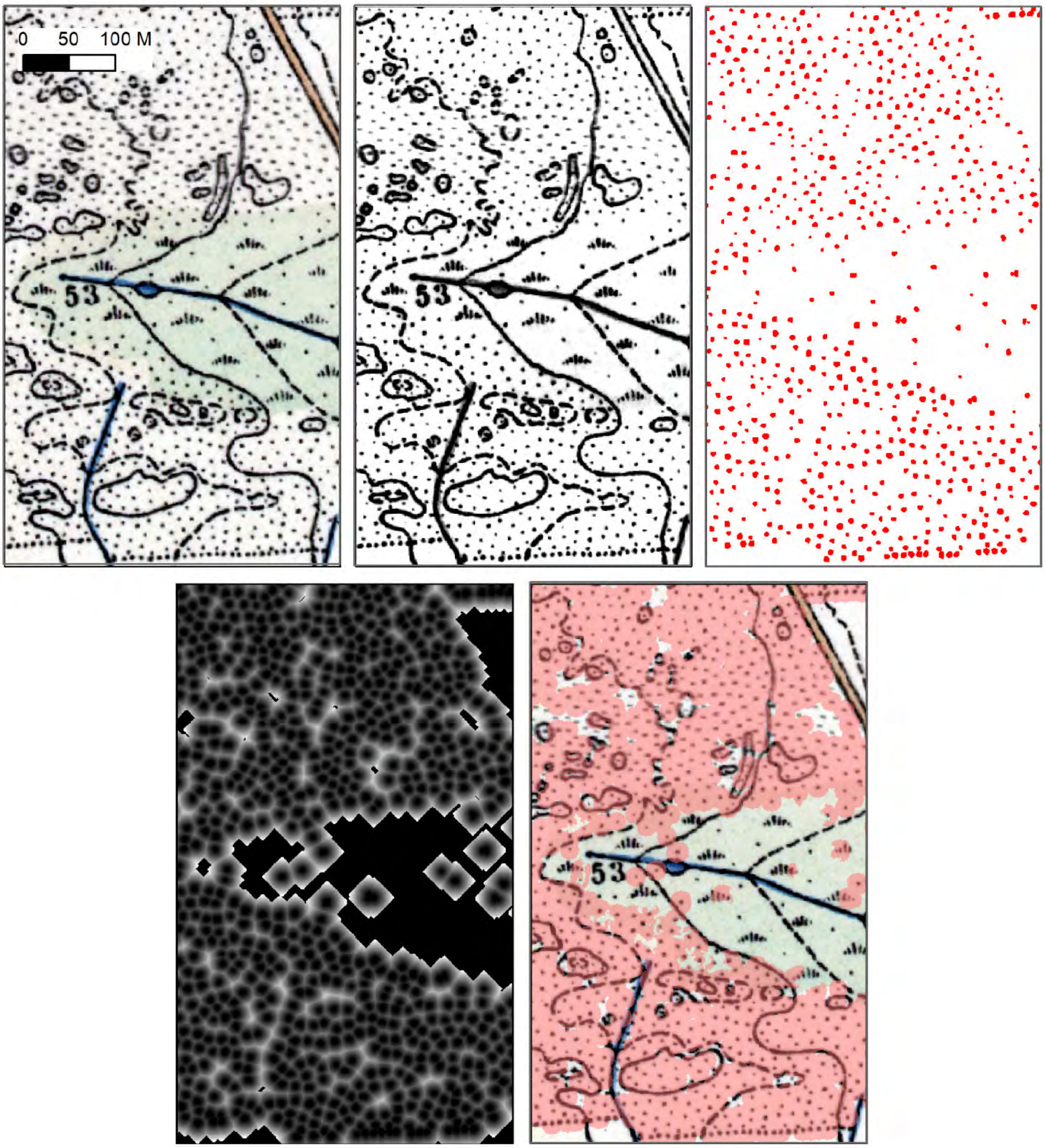


Figure 3.25 Illustration of the mapping of dune sand. Top row, left: an example sand dune area from HMB sheet H157; centre: the KB R-G-B intensity derived image; right: the initial set of dune sand dot candidate objects (red). Bottom row, left: distance map to dune sand dot candidate objects; right: The threshold segmented distance map objects, merged with the dot candidate objects, and filtered for clear-cut false-positive parts (N.B. this is not the finalised dune sand mapping), with the HMB KB image data as the backdrop.

In areas with marked sand dune development, the dune sand dots are obscured in the HMB maps by the HMB maps use of height isolines and cross-lines between the isolines to represent steeply sloping ground. Sand dunes were therefore false-negative cases with respect to the set of dune sand objects. In order to include sand dunes in the dune sand mapping, use was made of the mask image layer of the HMB represented steeply sloping ground. A coating was made of the dune sand objects into the mask image and smaller unclassified objects that lie within were merged (Figure 3.26).

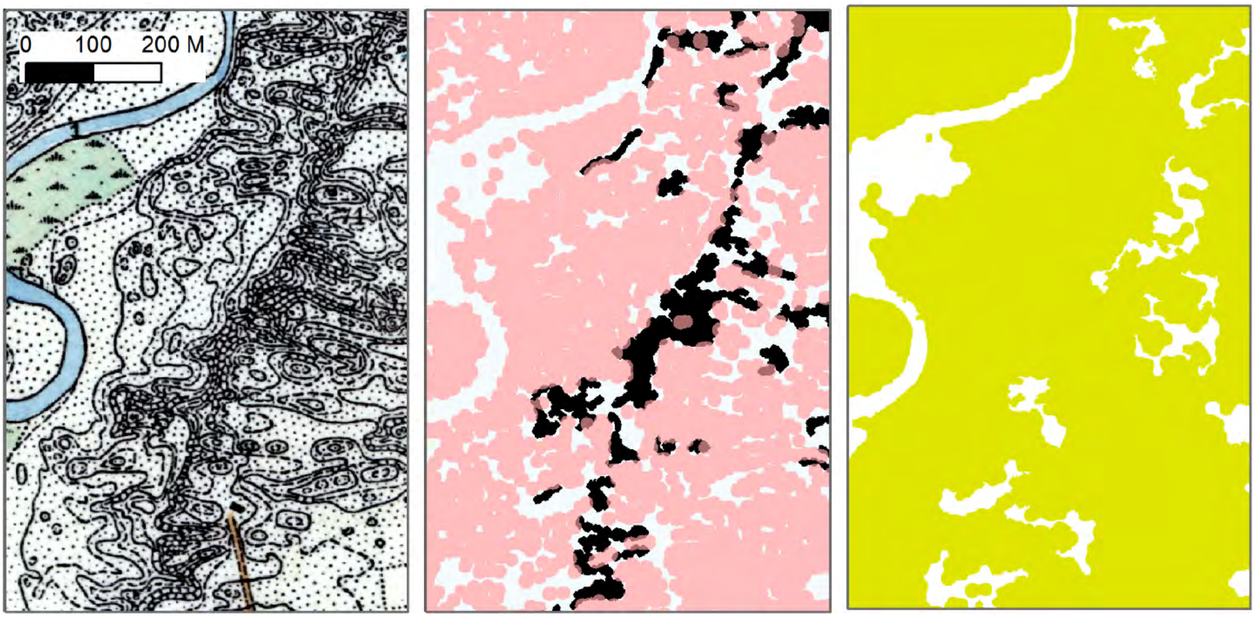


Figure 3.26 A part of HMB map sheet H157 with dune sand and probable sand dunes within (left). Centre: the dune sand objects (pink), with the mask image of steeply sloping ground (black). Right: the final dune sand, including sand dune objects.

3.6.6 Integrations for map sheet part overlaps and between target category mappings

The mappings made independently for each of the parts (typically four) of a HMB map sheet extent resulted in multiple mappings for the 1 km wide overlap zones between the parts. Furthermore, that mappings were made independently for each of the five target categories, possibilities existed for the same piece of land to have been mapped as more than just a single category, related to ambiguities in the HMB maps. Mapping overlaps of map sheet parts and categories were resolved by a second OBIA workflow.

Overlaps for each category, between map sheets parts were resolved first. This involved a visual assessment of the similarity / dissimilarity of overlap zone mappings from the respective map sheet parts, with the OBIA environment configured to enable clear visualisation of the separate mappings (Figure 3.27). A decision was made for each overlap zone (and each no-overlap zone) as to whether the final made for that category in the overlap zone should be as a logical AND or a logical OR, or just one of the inputs (applied rarely, if there were clear false-positive errors in just one of the inputs). The decisions were then applied via a nine-element vector of decisions to form for each category the final mapping for the full map sheet.

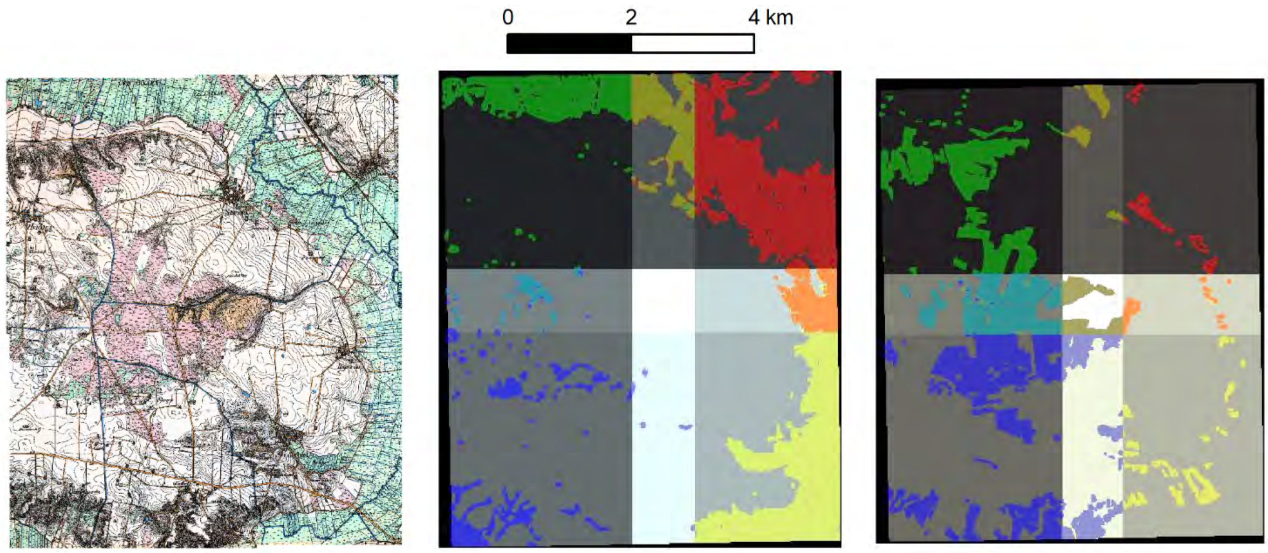


Figure 3.27 Production of the full HMB map sheet mappings from the independent map-pings of the map sheet parts. Example illustrated is HMB map sheet H395 (left, SD data). Assessment views based on transparent display of map sheet part mapping masks with different colours for each part, for wetland (centre) and heathland (right), with the four no-overlap zones and the five overlap zones of the map sheet as a grey tone backdrop. In these two cases, there are high levels of agreement between the map sheet part mappings.

The second component of this second OBIA workflow was the processing for overlaps in the mappings of the five target classes. This was resolved via binary sequence mask additions between the mappings of the five categories, i.e. an image layer with pure water body parts as value 2, pure forest parts as value 4, pure heath parts as value 8, etc. and thereby water-forest mix parts as value 6, water-heath mix parts as value 10, etc. A fixed set of rules for resolving mixture parts were then applied, e.g. that "a heath-water mix always goes as water", "a heath-wetland mix results in a mixed category mapping". Spatial context considerations are included in some of the rules, relating to object size and relative border to other objects of specific categories. Further contextual fixed rule-based processing is finally applied to clean up both pure and mixed category objects for smaller objects.

4 Accuracy assessment

4.1 Creation of validation layer

In order to validate the automatically extracted vector data, we aimed to build a test layer classified by visual interpretation of the HMB maps sheets. In this process, we chose a layer of points rather than polygons for two reasons. Firstly, points are faster to classify than areas. Secondly, classifications of polygons are at risk of being less accurate, due to difference in shape and size of land categories across the maps. Points are also more suitable units of digitisation in many cases such as for areas consisting of mixed-classes with fuzzy borders, as shown in Figure 4.1. In these cases, it would be difficult to draw the border between the land categories. Instead, with points, we are able to classify each object based on a qualitative evaluation of the signatures around it. A manual digitisation of the target categories within our two study areas would have enabled us to compare a manual approach and the automated approach in terms of achieved classification accuracies, delineation of categories and required time and resources. However, a manual digitisation was beyond the resources available to this project.

The validation layer was created as a regular point grid with an equidistance of 100 meters, resulting in a total of 7,868 validation points for the Hirtshals study area and 19,630 points for the Hobro area. The distance of 100 meters was considered to be small enough to capture the variation of different land categories while still being manageable to interpret within the resources of this project. Each point was classified visually based on the colours and signatures present up to 50 meters from the point. The applied land categories are summarised in Table 4.1. As illustrated in Figure 4.1, a point can be classified either as a pure category, such as heath or wetland or as a combination of several categories, such as heath and wetland or as heath and forest. For both study areas, Figure 4.2 shows the full classification of the validation points into the five target categories for our project. Apart from the land category column, the layer contains another column in which points that are located directly on a black line, such as a height isolines or a delineation line, were flagged (see Figure 4.3).

Table 4.1 Breakdown of basic land categories applied for the point validation layer.

Land category in the HMB map sheets	Definition	Code	Land category applied for the validation
Open land	White matrix on map sheet	0	Other
Meadow	Area characterized with meadow signatures and green colour	A	Wetland
Heath	Area marked with heath symbols and ‘purple’ colour	B	Heath
Bog	Area marked with bog symbols and ‘green colour’	C	Wetland
Dune sand	Area marked with sand symbols	D	Dune sand
Water	Area coloured with blue	E	Water body
Forest	Area marked with symbols for trees (both coniferous and deciduous) and brown colour	F	Forest
Marsh	Area marked with marsh signature and green colour	G	Water body
Road	Area marked with road symbol and ‘brown colour’	H	Other
Buildings	Area marked with black symbols for buildings	I	Other
Peat cuts	Area marked with symbols for peat cutting and blue colour	J	Other
Reed	Area marked with symbol for reed	K	Water body
Sand pits	Area marked with yellow colour, sometimes also sand signatures	L	Other
Clay pits	Area marked with brown colour and no other signatures	M	Other
Cemetery	Area marked with symbol (cross) for cemetery	N	Other

Figure 4.1 The maps show the classification of validation points into different land categories and combination of categories. Each validation point was classified based on a visual interpretation of symbols and colours within half of the distances to the next point.

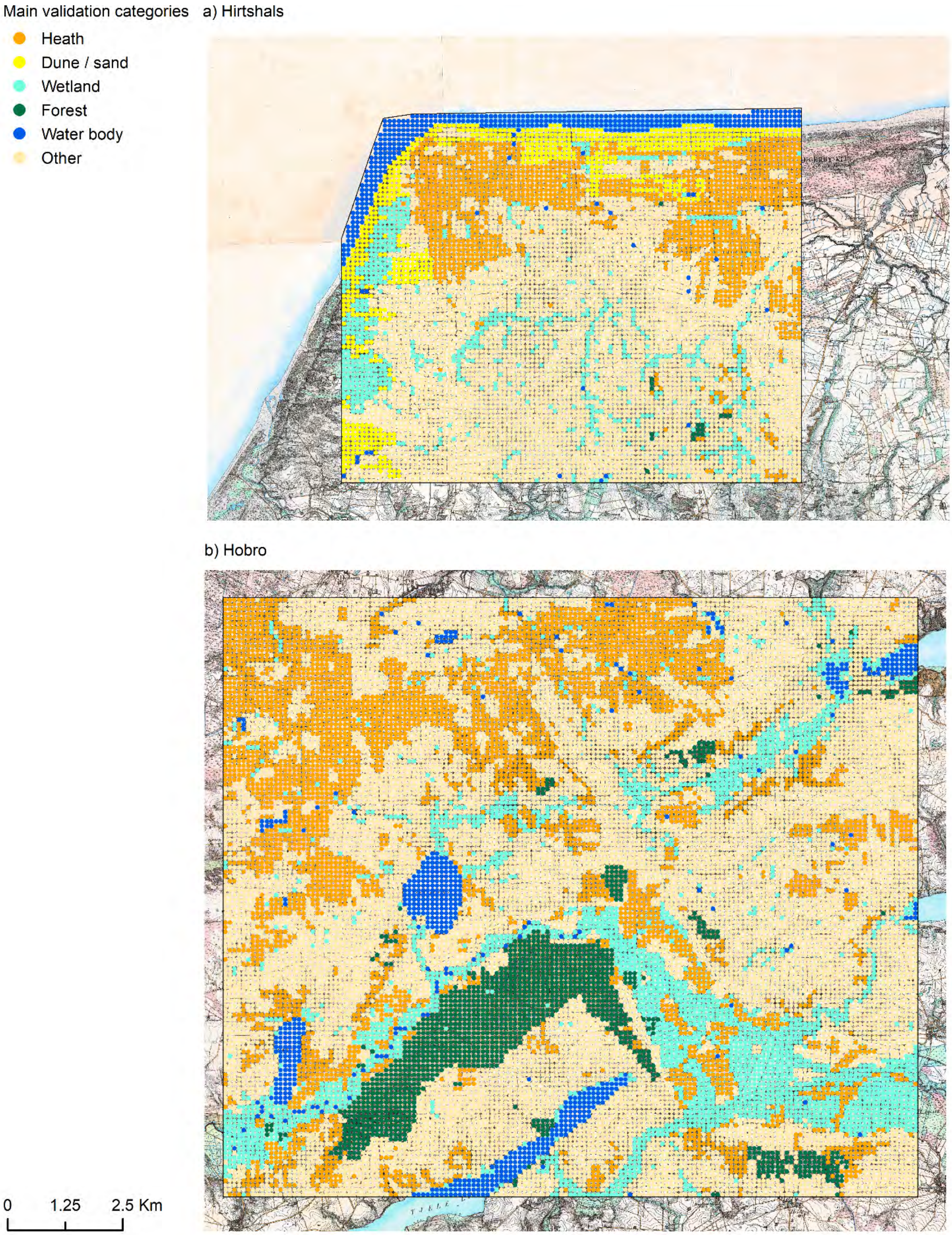


Figure 4.2 Overview of the distribution of main validation categories for the two study areas.

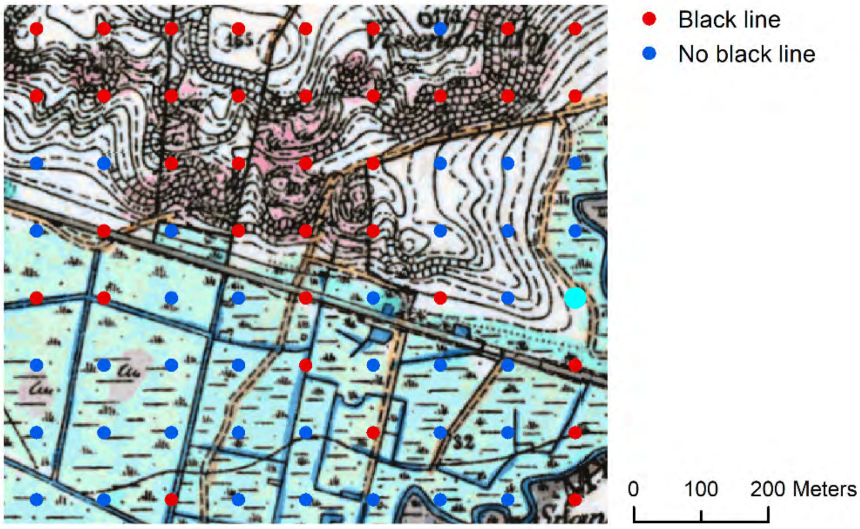


Figure 4.3 Illustration of the registration of validation points located directly on black lines.

4.2 Results of the accuracy assessment

Table 4.2 contains the results of the accuracy assessment for all combinations of land categories. Accuracy has been assessed for both false positives (points, assigned to a specific category in the extracted data but not by the validation dataset) and false negatives (points, assigned to a specific category by the validation dataset but not in the extracted data from the HMB maps). Overall accuracy, which is the percentage of points where categories extracted from HMB maps agree with the categories from the validation dataset, is 87.5 %. The results also show that errors, both in terms of false positive and false negative are generally lowest for pure categories, where the point in both the validation dataset and the extracted categories from HMB maps, has only been assigned to one category. The result also show that black elements in the HMB map sheets, such as text elements or height isolines, can affect obtained accuracies. Therefore, we removed all points overlapping with black elements from the sample, corresponding to 10,297 of 27,498 points (Table 4.3). Removing points overlapping with black elements result in higher category specific accuracies, both in terms of false negatives and false positives and in terms of a higher overall accuracy, raising that to 90.7 %.

Table 4.2 Results of the accuracy assessment for all individual land categories. Points, which were correctly extracted, are marked with grey colour. Overall accuracy: 87.5 %.

	Validation layer														
Extracted layers	Heath	Heath and Dune sand	Heath and Wetland	Dune sand	Wetland	Wetland and Dune sand	Wetland and Water bodies	Forest	Forest and Heath	Forest and Wetland	Water bodies	Other	Total	Producer's accuracy (%)	False positive (%)
Heath	4,175	40	339	2	17	0	0	2	70	0	5	195	4,845	86.2	13.8
Heath and Wetland	8	0	63	0	9	0	0	0	0	0	0	1	81	77.8	22.2
Dune sand	3	34	1	615	3	41	0	0	0	0	7	17	721	85.3	14.7
Wetland	23	1	366	1	2,411	28	3	3	0	31	9	193	3,069	78.6	21.4
Wetland and Dune sand	0	0	0	3	17	62	0	0	0	0	0	2	84	73.8	26.2
Forest	4	0	0	0	2	0	0	1,122	19	12	0	19	1,178	95.2	4.8
Forest and Heath	4	0	0	0	0	0	0	1	2	0	0	0	7	0.0	100.0
Forest and Wetland	0	0	1	0	2	0	0	3	0	2	0	0	8	25.0	75.0
Forest and Water bodies	0	0	0	0	0	0	0	0	0	0	1	0	1	0.0	100.0
Water bodies	1	0	1	9	13	0	0	0	0	0	1,021	6	1,051	97.1	2.9
Other	804	40	175	128	558	16	0	58	13	10	53	14,598	16,453	88.7	11.3
Total	5,022	115	946	758	3,032	147	3	1,189	104	55	1,096	15,031	27,498		
User's accuracy	83.1	0.0	6.7	81.1	79.5	42.2	0.0	94.4	1.9	3.6	93.2	97.1			
False negative	16.9	100.0	93.3	18.9	20.5	57.8	100.0	5.6	98.1	96.4	6.8	2.9			

Table 4.3 Results of the accuracy assessment for all individual land categories, where points, overlapping with black elements in the map sheets are removed. Points, which were correctly extracted, are marked with grey colour. Overall accuracy: 90.7 %.

	Validation layer													
Extracted layers	Heath	Heath and Dune sand	Heath and Wetland	Dune sand	Wetland	Wetland and Dune sand	Wetland and Water bodies	Forest	Forest and Heath	Forest and Wetland	Water bodies	Other	Total	Producer's accuracy (%)	False positive (%)
Heath	2,760	29	305	1	9	0	0	1	24	0	1	60	3,190	86.5	13.5
Heath and Wetland	5	0	60	0	7	0	0	0	0	0	0	0	72	83.3	16.7
Dune sand	2	20	1	388	2	16	0	0	0	0	1	8	438	88.6	11.4
Wetland	13	1	323	1	1,870	14	2	1	0	31	4	77	2,337	80.0	20.0
Wetland and Dune sand	0	0	0	2	15	47	0	0	0	0	0	2	66	71.2	28.8
Forest	0	0	0	0	1	0	0	540	2	6	0	4	553	97.6	2.4
Forest and Heath	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0	100.0
Forest and Wetland	0	0	1	0	2	0	0	2	0	2	0	0	7	28.6	71.4
Forest and Water bodies	0	0	0	0	0	0	0	0	0	0	1	0	1	0.0	100.0
Water bodies	0	0	0	4	4	0	0	0	0	0	988	2	998	99.0	1.0
Other	232	17	120	42	142	6	0	7	1	8	19	8,945	9,539	93.8	6.2
Total	3,012	67	810	438	2,052	83	2	551	27	47	1,014	9,098	17,201		
User's accuracy	91.6	0.0	7.4	88.6	91.1	56.6	0.0	98.0	0.0	4.3	97.4	98.3			
False negative	8.4	100.0	92.6	11.4	8.9	43.4	100.0	2.0	100.0	95.7	2.6	1.7			

Comparing both pure and combined categories obfuscates accuracy assessment where one of the combined categories has in fact been mapped correctly, e.g. a point, which has been extracted as both heath and wetland and in the validation dataset is categorised as wetland is labelled as incorrectly mapped, although the wetland category is mapped correctly. Therefore, we also conducted an assessment, where we included combined categories. In this assessment, at least one of the categories extracted from the HMB maps should be present in the validation dataset, to count as a correctly extracted category. Results are shown in Table 4.3. The overall achieved accuracy is 91.3 %. Furthermore, for all categories, false negative values are around 20 % or lower and false positive values are around 10 % or lower. Then, after also removing points overlapping with black elements (Table 4.4), overall accuracy increased to 95.4 %. For all categories, false negative values are around 10 % or lower and false positive values are around 6 % or lower.

Table 4.3 Results of the accuracy assessment, where combined categories are included. I.e., at least one of the categories extracted from the HMB maps should be present in the validation dataset, to count as a correctly extracted category. Overall accuracy: 91.3 %.

	Validation layer								
Extracted layers	Heath	Dune sand	Wetland	Forest	Water bodies	Other	Total	Producer's accuracy (%)	False positive (%)
Heath	4,695	2	17	2	5	195	4,916	95.5	4.5
Dune sand	3	693	4	0	7	17	724	95.7	4.3
Wetland	23	2	2,930	3	9	196	3,163	92.6	7.4
Forest	4	0	2	1,165	0	19	1,190	97.9	2.1
Water bodies	1	9	14	0	1,022	6	1,052	97.1	2.9
Other	844	128	749	81	53	14,598	16,453	88.7	11.3
Total	5,570	834	3,716	1,251	1,096	15,031	27,498		
User's accuracy (%)	84.3	83.1	78.8	93.1	93.2	97.1			
False negative (%)	15.7	16.9	21.2	6.9	6.8	2.9			

Table 4.4 Results of the accuracy assessment, where combined categories are included. I.e., at least one of the categories extracted from the HMB maps should be present in the validation dataset, to count as a correctly extracted category. Points overlapping with black elements in the map sheets are removed. Overall accuracy: 95.4 %..

	Validation layer								
Extracted layers	Heath	Dune sand	Wetland	Forest	Water bodies	Other	Total	Producer's accuracy (%)	False positive (%)
Heath	3,183	1	9	1	1	60	3,255	97.8	2.2
Dune sand	2	426	3	0	1	8	440	96.8	3.2
Wetland	13	2	2,312	1	4	79	2,411	95.9	4.1
Forest	0	0	1	552	0	4	557	99.1	0.9
Water bodies	0	4	4	0	989	2	999	99.0	1.0
Other	249	42	268	16	19	8,945	9,539	93.8	6.2
Total	3,447	475	2,597	570	1,014	9,098	17,201		
User's accuracy (%)	92.3	89.7	89.0	96.8	97.5	98.3			
False negative (%)	7.7	10.3	11.0	3.2	2.5	1.7			

For the six target categories, false positive values range from 2.9 to 4.5 %, when including black elements and from 1.0 – 4.1 % when excluding black elements. This indicates that our model does not significantly overestimate the area of the target categories. False negative values were lowest for water bodies (6.8 % with black lines; 2.5 % without black lines) and slightly higher for forest (6.9 % with black lines; 3.3 % without black lines). False negative values are markedly higher for heath (15.7 % with black lines; 7.7 % without black lines) for dune / sand (16.9 % with black lines; 10.3 % without black lines) and for wetland (21.2 % with black lines; 11.0 %). These higher false negative values indicate that our model was less successful at consistently recognising these categories, but also that these categories often occur in complex combinations of other categories and black elements. This is especially the case with the heath and the dune sand categories, with occurrences often located on steeply sloping ground and / or dune area with many small hills, and consequently many black lines and signatures for steeply sloping ground. For the categories heath, wetland and dune sand, the maps in Figure 4.4 shows the validation points grouped into positives (matches), false negative and false positive. For all three categories, the spatial distributions of false negatives show localized, semi-clustered patterns. This suggests special local factors leading to repeated occurrence of an error within a local area, such as narrow extents of a category (e.g. for wetland in a narrow-bottomed valley), markedly cut-up by text and linework. The few false positive events show dispersed patterns adjacent to positive events, suggesting small delineation discrepancies in the produced mapping. Vast extents of the occurrences of each class are free of false negatives and false positives.

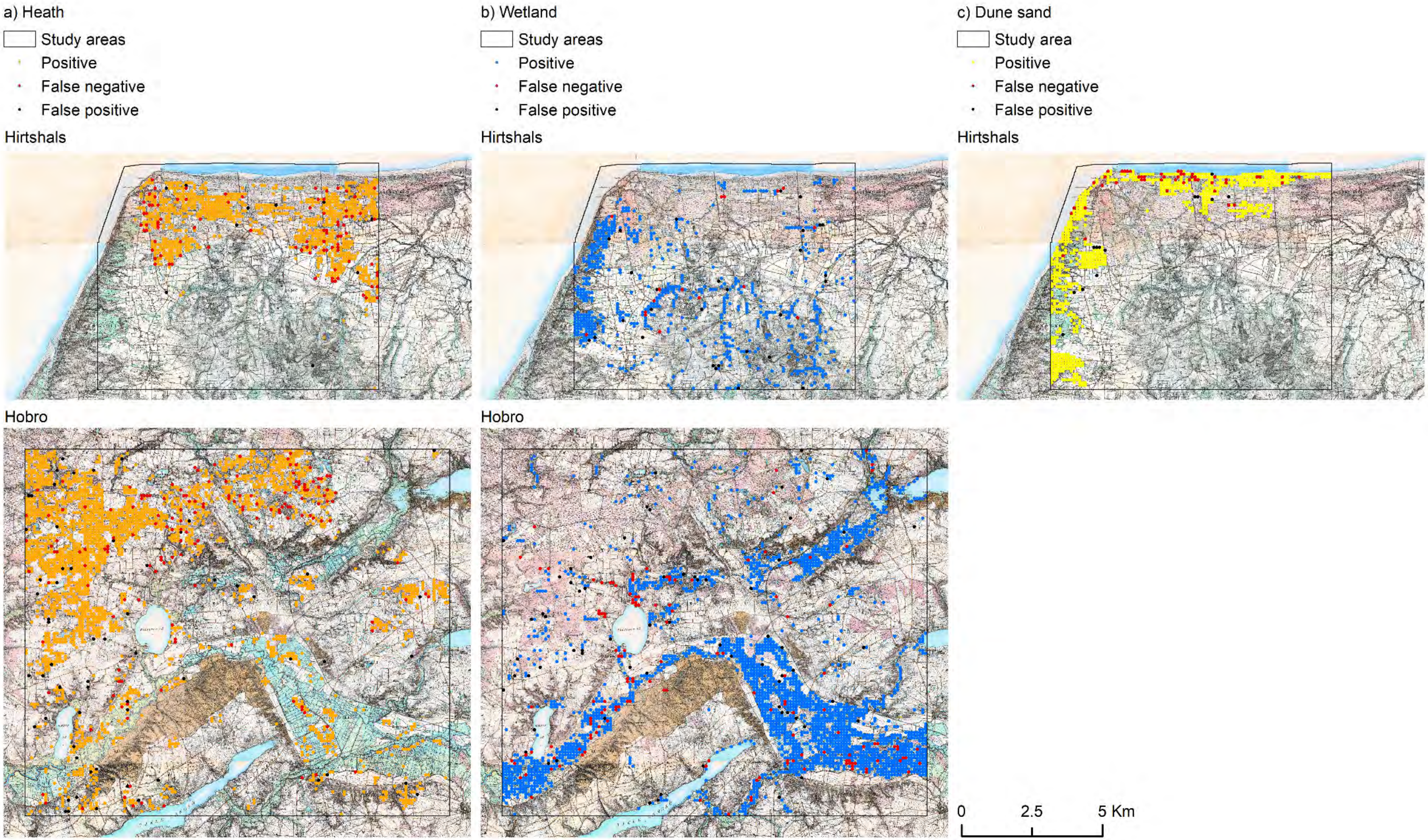


Figure 4.4 Validation points grouped into positives (matches), false negative and false positive for both study areas and for the categories heath (a), wetland (b) and dune sand (c).

5 Assessment of land use and land cover changes

In order to assess land use and land cover (LULC) changes, we combined the layers of land categories produced from the HMB maps sheets with contemporary LULC information. Contemporary LULC information was derived from Basemap03 (Levin, 2019). Basemap03 is a countrywide map of LULC, which is based on overlaying and spatially adjusting available up-to-date geographical layers, containing land use and land cover information into a raster map with a cell-size of 10x10 meters. LULC information contained in Basemap03 corresponds to the year 2018.

5.1 Pre-processing

5.1.1 Aggregation of contemporary land categories

Basemap03 contains a total of 677 individual LULC categories, which we aggregated into nine major LULC categories (Table 5.1), including the five target categories, which were extracted from the HMB map. In the HMB maps, dry grassland and agriculture do not exist as land categories, and built / infrastructure was not extracted from the HMB maps as part of this project. We included these categories, as they can provide important insight into LULCC.

Table 5.1 Aggregated land categories for 2018.

Land category	Description	Sources
Heath	Uncultivated areas, dominated by vegetation of heather, gorse and coarse grasses	Registration of protected habitat types; Management plans of state forests; Management plans of Defence sites; Natura 2000 habitat types; topographical database (FOT)
Dry grassland	Uncultivated areas, dominated by grass and herb vegetation on dry and/or calcareous soils	Registration of protected habitat types
Dune sand	Uncultivated areas, dominated by dune sand and coastal sand. Includes partly vegetated dune areas	Management plans of state forests; Management plans of Defence sites; Natura 2000 habitat types; topographical database (FOT)
Wetland	Uncultivated areas, dominated by grass and herb vegetation on permanently or seasonally water-logged soils	Registration of protected habitat types; Management plans of state forests; Management plans of Defence sites; Natura 2000 habitat types
Forest	Areas, dominated by woody vegetation. Including temporary un-stocked areas. Excludes Christmas trees, energy forest and orchards	Management plans of state forests; Management plans of Defence sites; Natura 2000 habitat types; topographical database (FOT), Agricultural census data
Water bodies	Water bodies, including lakes, ponds, streams, sea. Includes water bodies covered by reed vegetation	Registration of protected habitat types; Management plans of state forests; Management plans of Defence sites; Natura 2000 habitat types; topographical database (FOT)
Agriculture	Agricultural land. Includes cropland, Christmas trees, energy forest, orchards and permanent grasslands, which are not registered as habitat types	Management plans of state forests; Management plans of Defence sites; topographical database (FOT), Agricultural census data
Built/infrastructure	Built up land, infrastructure and associated land, such as gardens, recreational and sport facilities, road and rail verges	Management plans of state forests; Management plans of Defence sites; topographical database (FOT), cadastre map
Unmapped	Areas without any land use or land cover information	

5.1.2 Adjustment to current land categories

A direct overlay between the land categories produced from the HMB map sheet and the contemporary land categories would result in inclusion of a sizable accumulation of areas of apparent change, that are the result of small discrepancies between the delineations of land categories in the two datasets, rather than representing actual changes. Therefore, for the change assessment, we spatially adjusted the land categories produced from the HMB maps sheets with the contemporary land categories derived from Basemap03. In a first step, the vector layers of land categories produced from the HMB maps were converted to raster format with a cell size of 10x10 meters, corresponding to the spatial support of Basemap03.

The layers produced from the HMB map sheets were produced independently for each target category. As a consequence, for some locations, the produced layers overlap, resulting in combined categories. For some locations, these category overlaps cases are actually present in the HMB maps. There are, e.g. locations, where the dotted signature for dune sand overlaps with the bluish-green colour representing wetlands. However, in most cases, these overlaps between categories are a consequence of small and narrow discrepancies between the delineations. Therefore, and in order to avoid too much complexity in the change assessment, in a second step, we eliminated these overlaps. This was done by, first overlaying the overlapping categories with the contemporary land categories. If an area according to the HMB map sheets was mapped as e.g. both wetland and dune sand, and the contemporary land category was dune sand, the combined category was assigned to dune sand. Next, for locations without agreement between one of the combined categories and contemporary categories, cells with combined categories were eliminated by assigning them to the closest adjacent pure category with agreement to one of the combined categories. In order to reduce the bias from small and narrow discrepancies between the delineation of the two layers, in a third step, we spatially adjusted the land categories from the HMB maps sheets to the contemporary land categories. The applied method is described in Figure 5.1. Table 5.2 summarises the area of 1880 land categories before and after spatial adjustment to the 2018 land categories. In the adjusted map, combined categories (e.g. heath and wetland) have been eliminated. Furthermore, after adjustment, the area of other (not mapped) categories is reduced by 27 Km^2 (~7%), while areas of heath, dune sand, wetland and forest are enlarged.

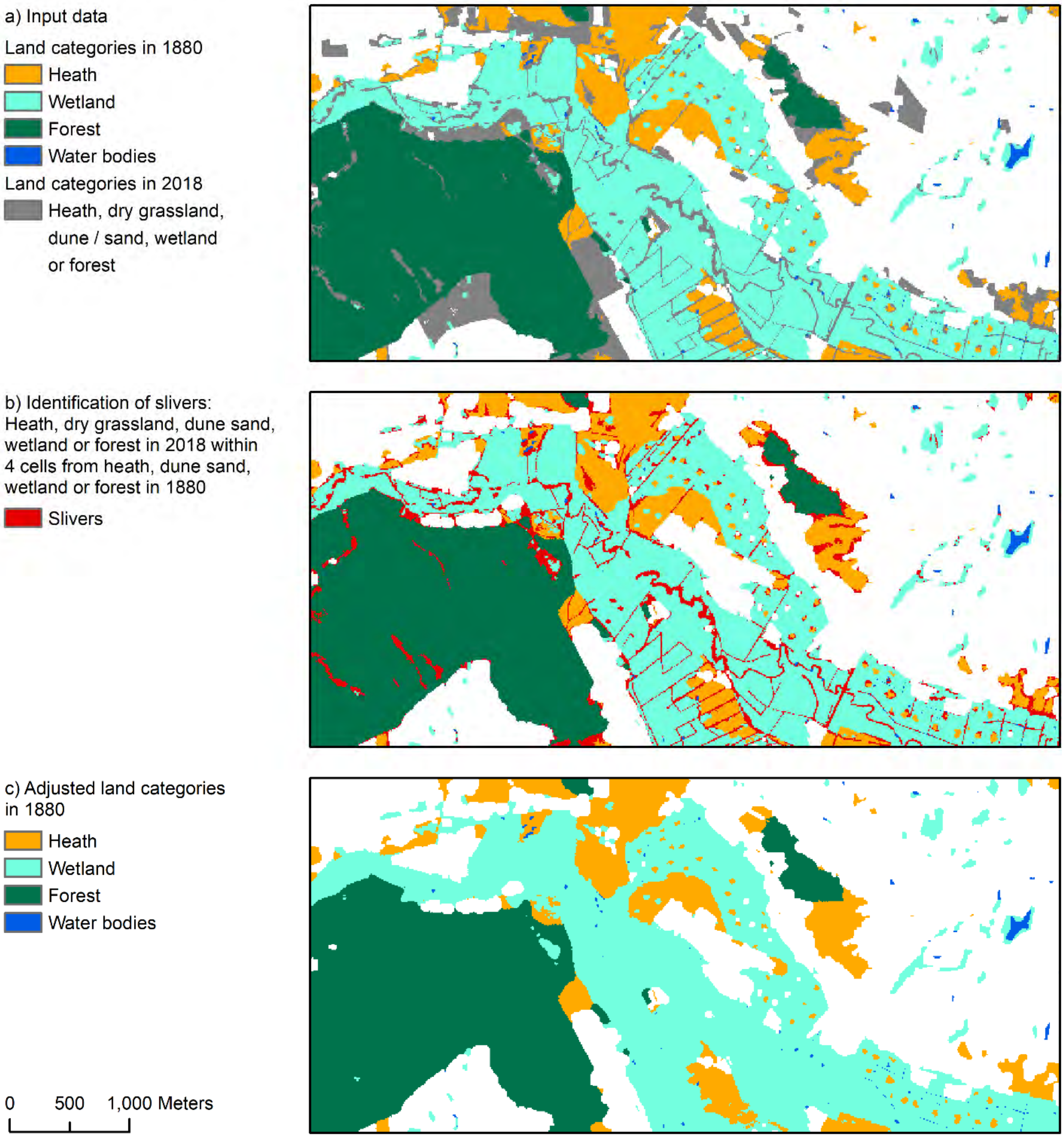


Figure 5.1 Illustration of applied methodology for adjustment of land categories produced from the HMB maps to contemporary land categories for 2018. Land categories from the HMB map sheets were overlaid with contemporary land categories, representing heath, dry grassland, dune sand, wetland or forest (a). Next, so-called “slivers” were identified. These were defined as cells which in the contemporary land category layer are mapped as heath, dry grassland, dune sand, wetland or forest, and which are located within at least 4 cells (40 meters) from any of the target categories from the HMB map sheets (b). Finally, we eliminated these slivers by assigning cells to the closest adjacent land category from the HMB map sheets (c).

Table 5.2 Area of 1880 land categories before and after adjustment to 2018 land categories.

	Before adjustment to 2018 land categories	After adjustment to 2018 land categories	Difference non-adjusted and adjusted land categories	
1880 land categories	Area (Km²)	Area (Km²)	Area (Km²)	Proportion (%)
Heath	93,8	106,2	12,4	13,3
Heath and Wetland	1,2	-	-	-
Dune sand	17,7	20,9	3,2	17,9
Dune sand and Wetland	1,0	-	-	-
Wetland	58,8	71,7	12,9	22,0
Forest	14,7	15,9	1,3	8,5
Forest and Heath	0,2	-	-	-
Forest and Wetland	0,1	-	-	-
Forest and Water	0,0	-	-	-
Water	24,6	24,2	-0,4	-1,5
Other	383,2	356,2	-27,0	-7,1
Total	595,2	595,2		

The HMB map sheets, which we applied for the two study areas, are based on surveys conducted in the period between 1879 and 1887, the Hobro area between 1879 and 1880 and the Hirtshals area between 1885 and 1887. However, for sake of clarity, in the remaining sections we refer to land categories from the HMB map sheets as land categories for 1880 and contemporary land categories as land categories for 2018.

5.2 Results

5.2.1 Accumulated changes

For the two study areas, Figures 5.2 and 5.3 show the mapped land categories for the Hirtshals and the Hobro area for 1880 and for 2018. The area and proportion of land categories are summarised in Table 5.3 and 5.4. In both areas in 1880, open, non-forested habitat types made up around one third of the total area. By 2018, the area proportion of these categories declined to less than 20 % in the Hirtshals area and to less than 10 % in the Hobro area. Meanwhile, proportions of forest increased significantly from 0.5 to 13.0 % in the Hirtshals area and from 3.5 to 12.3 % in the Hobro area. In both study areas, the proportion of water bodies increased slightly. In the HMB map sheets, agriculture, does not exist as a land category and we did not produce layers for built / infrastructure. For 1880, these categories are therefore part of the “other” category. In the Hirtshals area, the total proportion of other did only change insignificantly, while in the Hobro area, it increased from around 60 % in 1880 to around 76 % in 2018. In both study areas there were almost no changes from forest in 1880 to other categories in 2018.

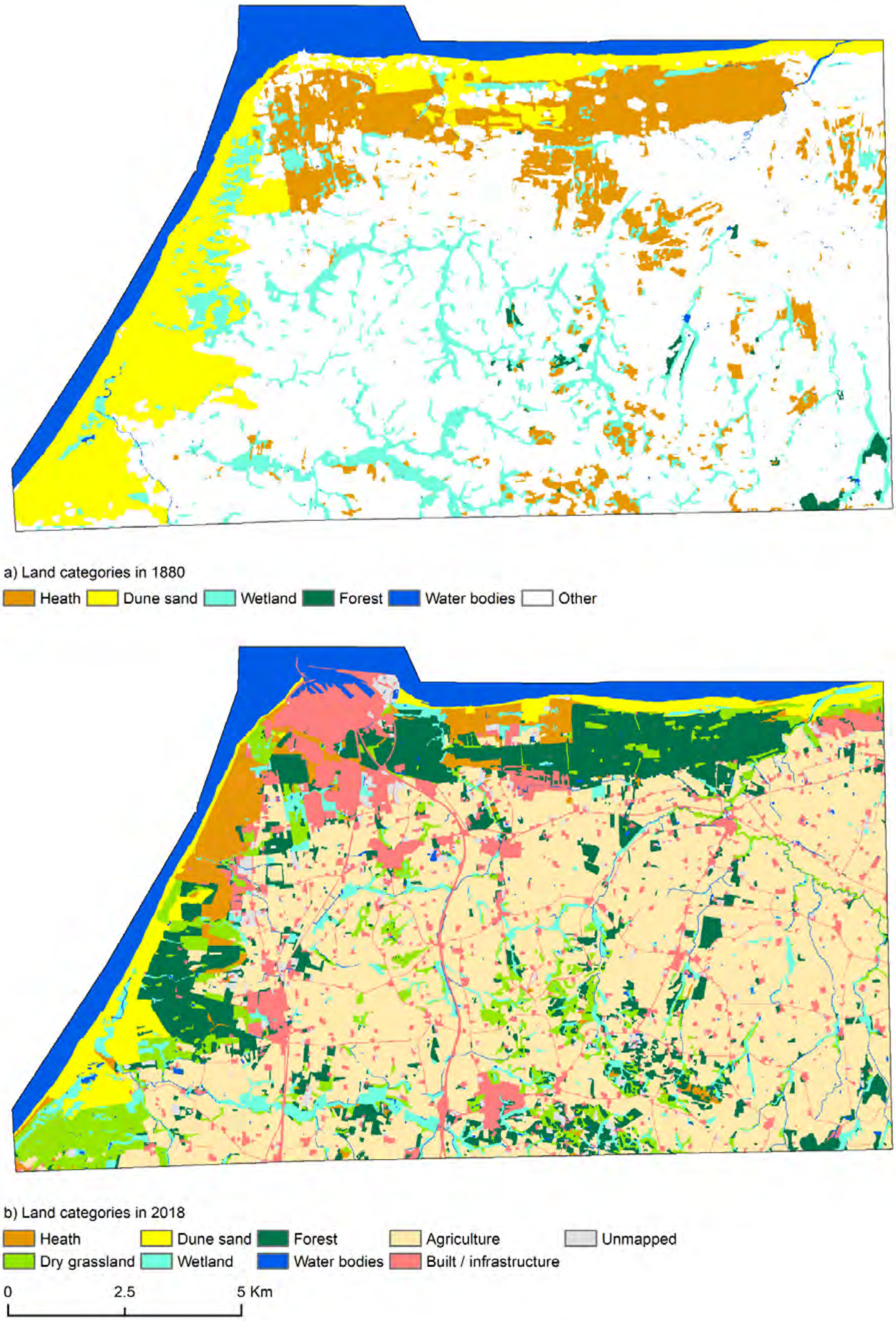


Figure 5.2 Land categories in the Hirtshals area in 1880 and in 2018.

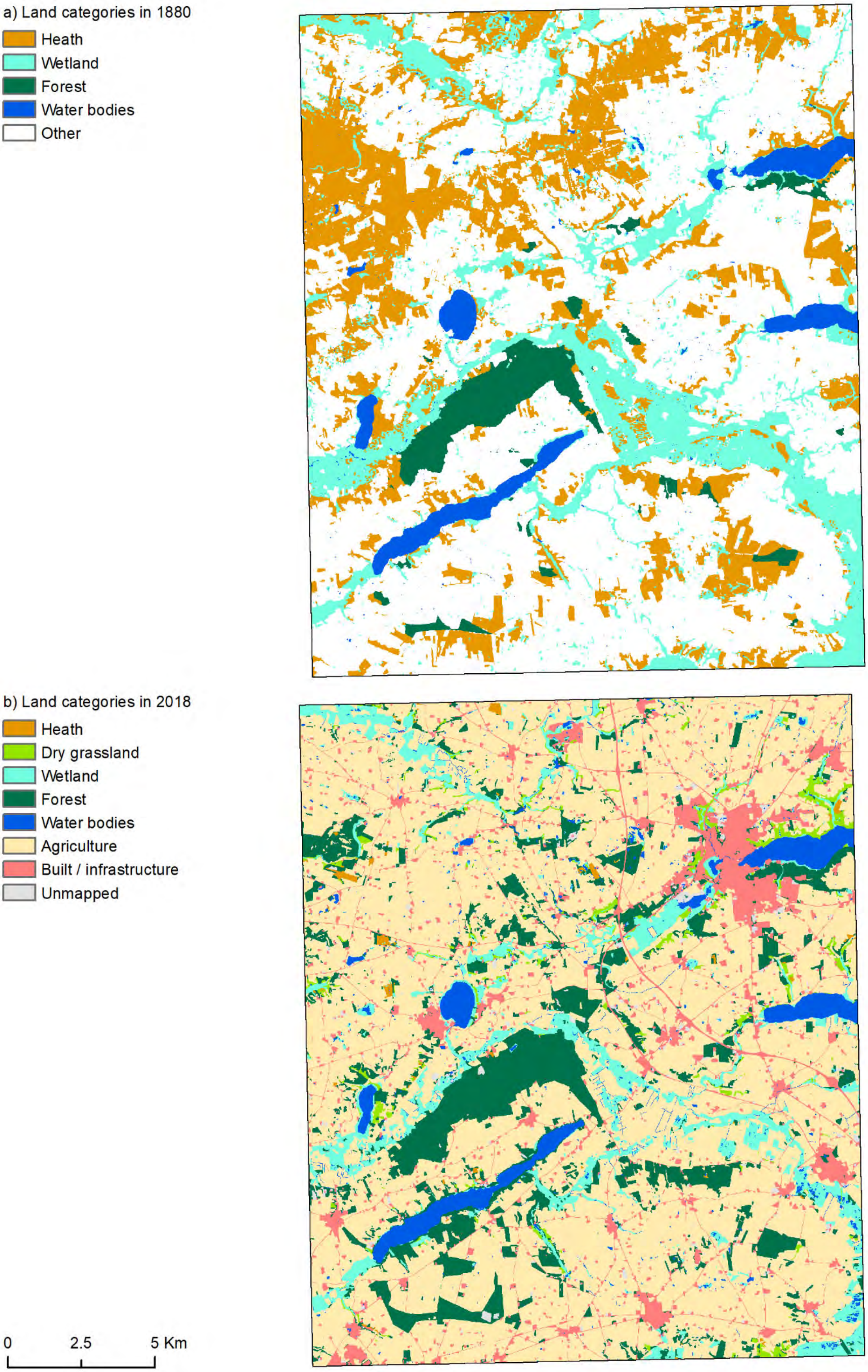


Figure 5.3 Land categories in the Hobro area in 1880 and in 2018.

Table 5.3 Area and proportion of land categories in 1880 and 2018 in the Hirtshals area.

		1880	2018	1880	2018	1880-2018	
Overall land categories	Specific land categories	Area (km²)	Area (km²)	Proportion of total area (%)	Proportion of total area (%)	Area (km²)	Proportion of 1880 (%)
Open, non-forested habitat types	Heath	19.2	5.7	11.3	3.4	-13.4	-70.1
	Dry grassland	-	11.6	-	6.8	-	-
	Dune sand	20.9	6.5	12.3	3.8	-14.4	-68.9
	Wetland	14.0	7.5	8.2	4.4	-6.5	-46.4
	Total	54.1	31.4	31.8	18.5	-22.7	-42.0
Forest	Forest	0.9	22.1	0.5	13.0	21.3	2,480.8
Water bodies	Water bodies	10.4	12.4	6.1	7.3	2.0	18.7
Other	Agriculture	-	83.3	-	49.0	-	-
	Built/infrastructure	-	17.3	-	10.2	-	-
	Unmapped	-	3.5	-	2.1	-	-
	Other	104.6	-	61.6	-	-	-
	Total	104.6	104.1	61.6	61.2	-0.5	-0.5
Total		170.0	170.0	100.0	100.0		

Table 5.4 Area and proportion of land categories in 1880 and 2018 in the Hobro area.

		1879	2018	1880	2018	1880-2018	
Overall land categories	Specific land categories	Area (km²)	Area (km²)	Proportion of total area (%)	Proportion of total area (%)	Area (km²)	Proportion of 1880 (%)
Open, non-forested habitat types	Heath	87.1	1.3	20.5	0.3	-85.7	-98.5
	Dry grassland	-	6.5	-	1.5	-	-
	Wetland	57.7	27.4	13.6	6.4	-30.3	-52.5
	Total	144.8	35.2	34.1	8.3	-109.6	-75.7
Forest	Forest	15.1	51.5	3.5	12.1	36.4	241.9
Water bodies	Water bodies	13.8	15.5	3.2	3.6	1.7	12.3
Other	Agriculture	-	280.9	-	66.1	-	-
	Built/infrastructure	-	37.0	-	8.7	-	-
	Unmapped	-	5.2	-	1.2	-	-
	Other	251.6	-	59.2	-	-	-
	Total	251.6	323.0	59.2	76.0	71.4	28.4
Total	Total	425.2	425.2	100.0	100.0		

5.2.2 Spatially explicit LULC dynamics

The accumulated LULC changes can conceal spatially explicit dynamics between the different land categories. I.e., while the total area and proportion of a land category might not have changed significantly, the geographic distribution of the land category might have changed, resulting in more significant, spatially explicit dynamics. Table 5.5 and 5.6 contain change matrixes depicting spatially explicit changes between land categories for the Hirtshals area and the Hobro area. From the tables it can be seen that e.g. in the Hirtshals area just 0.9 km² were mapped as heath in both 1880 and in 2018 (Table 5.5a). Only 4.8 % of the area mapped as heath in 1880 was also heath in 2018, while 45.2 % had changed to forest and 24.8 % to agriculture (Table 5.5b). The table also shows that 16.0 % of the total area that was heath in 2018 was also heath in 1886, while 55.6 % was dune sand and 17.5 % was wetland. In the following sections, we describe and discuss changes between specific land categories in more detail.

Table 5.5 Land use/land cover change matrix for the Hirtshals area.

a) Area (Km²)	2018									
1880	Heath	Dry grassland	Dune sand	Wetland	Forest	Water bodies	Agriculture	Built/ infrastructure	Unmapped	Total
Heath	0.9	1.5	0.2	0.4	8.7	0.1	4.8	2.3	0.3	19.2
Dune sand	3.2	4.0	5.2	1.0	4.5	1.5	0.4	1.0	0.1	20.9
Wetland	1.0	1.2	0.2	4.4	1.1	0.4	4.8	0.7	0.2	14.0
Forest	0.0	0.1	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.9
Water bodies	0.0	0.0	0.2	0.1	0.0	9.3	0.1	0.5	0.2	10.4
Other	0.6	4.8	0.7	1.6	7.1	1.0	73.2	12.8	2.7	104.6
Total	5.7	11.6	6.5	7.5	22.1	12.4	83.3	17.3	3.5	170.0
b) Proportion of 1880 (%)	2018									
1880	Heath	Dry grassland	Dune sand	Wetland	Forest	Water bodies	Agriculture	Built/ infrastructure	Unmapped	Total
Heath	4.8	7.9	1.3	2.2	45.2	0.4	24.8	11.9	1.4	100.0
Dune sand	15.2	18.9	24.8	4.7	21.7	7.4	1.9	4.7	0.6	100.0
Wetland	7.2	8.8	1.2	31.3	7.8	3.2	34.5	4.8	1.3	100.0
Forest	1.2	7.8	0.0	3.0	80.3	1.9	3.6	1.8	0.4	100.0
Water bodies	0.1	0.3	1.7	0.9	0.2	89.3	0.7	4.9	1.9	100.0
Other	0.6	4.6	0.7	1.5	6.8	1.0	70.0	12.2	2.6	100.0
Total	3.4	6.8	3.8	4.4	13.0	7.3	49.0	10.2	2.1	100.0
c) Proportion of 2018 (%)	2018									
1880	Heath	Dry grassland	Dune sand	Wetland	Forest	Water bodies	Agriculture	Built/ infrastructure	Unmapped	Total
Heath	16.0	13.1	3.8	5.7	39.2	0.6	5.7	13.2	7.7	11.3
Dune sand	55.6	34.0	79.9	13.1	20.5	12.4	0.5	5.7	3.7	12.3
Wetland	17.5	10.5	2.6	58.4	4.9	3.6	5.8	3.9	5.0	8.2
Forest	0.2	0.6	0.0	0.3	3.1	0.1	0.0	0.1	0.1	0.5
Water bodies	0.1	0.3	2.8	1.2	0.1	75.2	0.1	3.0	5.5	6.1
Other	10.7	41.5	10.9	21.3	32.2	8.0	87.9	74.0	77.9	61.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 5.6 Land use/land cover change matrix for the Hobro area.

a) Area (Km²)	2018								
1880	Heath	Dry grassland	Wetland	Forest	Water bodies	Agriculture	Built/ infrastructure	Unmapped	Total
Heath	1.0	3.8	1.2	20.1	0.4	56.2	3.4	1.0	87.1
Wetland	0.1	0.8	24.1	2.5	2.6	25.0	1.9	0.6	57.7
Forest	0.0	0.0	0.0	14.6	0.0	0.2	0.1	0.1	15.1
Water bodies	0.0	0.0	0.7	0.6	11.7	0.5	0.2	0.0	13.8
Other	0.2	1.9	1.3	13.7	0.7	199.0	31.4	3.4	251.6
Total	1.3	6.5	27.4	51.5	15.5	280.9	37.0	5.2	425.2
b) Proportion of 1880 (%)	2018								
1880	Heath	Dry grassland	Wetland	Forest	Water bodies	Agriculture	Built/ infrastructure	Unmapped	Total
Heath	1.1	4.4	1.4	23.1	0.5	64.5	3.9	1.2	100.0
Wetland	0.2	1.4	41.8	4.3	4.5	43.4	3.3	1.1	100.0
Forest	0.3	0.1	0.2	97.0	0.2	1.3	0.5	0.4	100.0
Water bodies	0.0	0.1	5.4	4.2	84.9	3.5	1.6	0.3	100.0
Other	0.1	0.7	0.5	5.5	0.3	79.1	12.5	1.4	100.0
Total	0.3	1.5	6.4	12.1	3.6	66.1	8.7	1.2	100.0
Proportion of 2018 (%)	2018								
c) 1880	Heath	Dry grassland	Wetland	Forest	Water bodies	Agriculture	Built/ infrastructure	Unmapped	Total
Heath	74.4	58.5	4.4	39.0	2.6	20.0	9.2	19.6	20.5
Wetland	10.5	12.2	88.0	4.9	16.9	8.9	5.2	12.1	13.6
Forest	3.3	0.3	0.1	28.4	0.2	0.1	0.2	1.2	3.5
Water bodies	0.2	0.2	2.7	1.1	75.6	0.2	0.6	0.9	3.2
Other	11.7	28.8	4.7	26.7	4.7	70.8	84.8	66.2	59.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Changes from open, non-forested habitat types

For open, non-forested habitat types (including heath, dune sand and wetland), which were mapped in 1880, the diagrams in Figure 5.4 show the proportion of the different land categories mapped in 2018. The changes are also shown in the maps in Figure 5.5. The figures and maps illustrate overall LULC dynamics, characterised by a decline in open, non-forested habitats, mainly to the gain of afforestation, agriculture and built / infrastructure. There are, however, significant differences between the two study areas. While, in Hirtshals around 43 % of open, non-forested habitat types in 1880 were also non-forested habitat types in 2018, in Hobro, this proportion is only around 22 %. In comparison, in the Hirtshals area, some 18 % had, by 2018 changed to agriculture, while this proportion is more than 56 % in the Hobro area. In Hirtshals, around 27 % had by 2018 changed to forest, compared to around 16 % in Hobro. A larger proportion had changed to build / infrastructure in Hirtshals (~7%) compared to Hobro (~4%). Finally, also a larger proportion had changed to water bodies in Hirtshals (~4%) compared to Hobro (~2%).

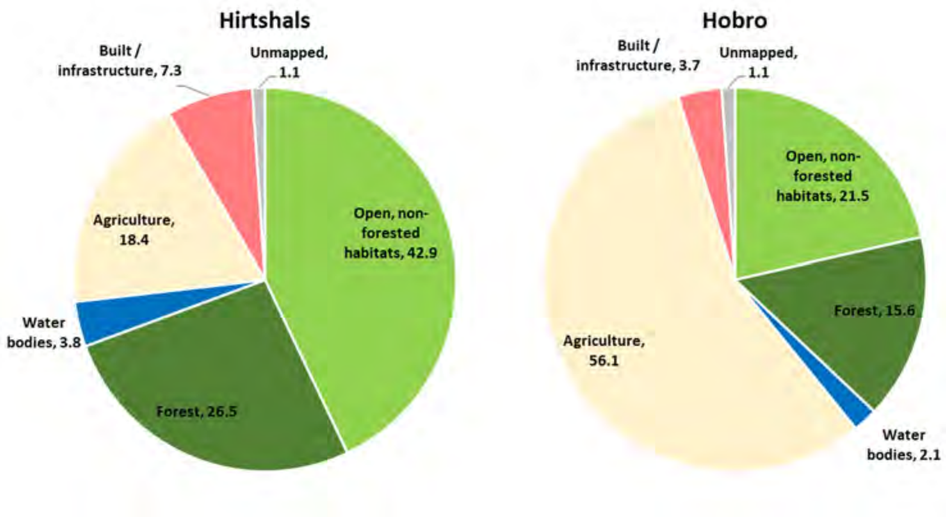


Figure 5.4 2018 land category proportions of 1880 open, non-forested habitat.

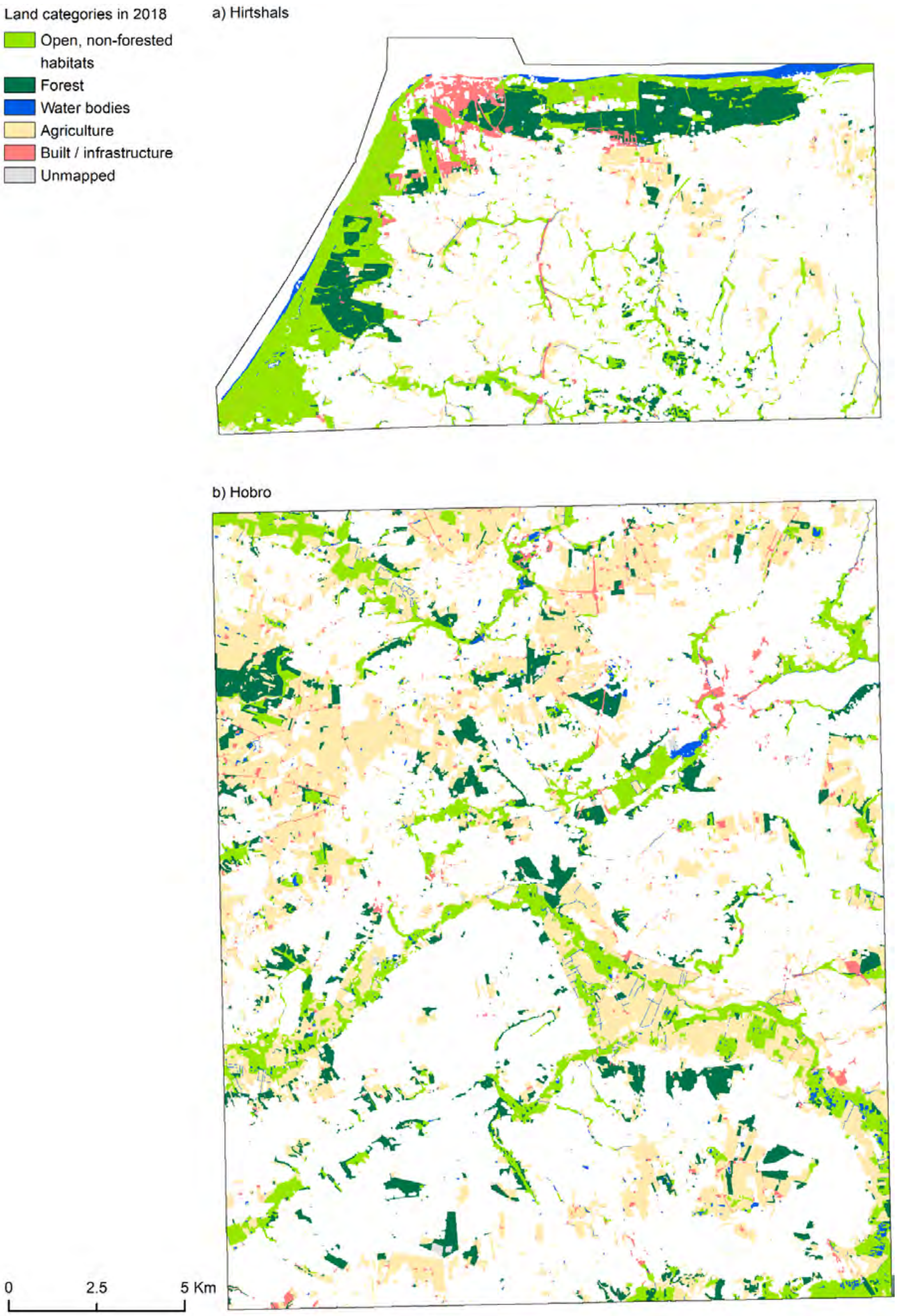


Figure 5.5 2018 land categories of 1880 open, non-forested habitat.

Changes from heath

For heath mapped in 1880, the diagrams in Figure 5.6 show the proportion of the different land categories in 2018. The changes are also shown in the maps in Figure 5.7. The figures and maps illustrate overall LULC dynamics, characterised by a decline in heath, mainly to the gain of afforestation, agriculture and, particularly in the Hirtshals area, built / infrastructure. There are significant differences between the two study areas. While, in Hirtshals, by 2018 the largest proportion had changed to forest (> 45%), in Hobro, the largest proportion had changed agriculture (> 56%). Furthermore, in both areas, a considerable proportion of heath was by 2018 mapped as dry grassland (~ 8% in Hirtshals; ~4% in Hobro), which was not included as a land category in the HMB survey. The map in Figure 5.8 shows an example from the north-eastern part of the Hobro area, which in 1880 was mapped as heath and in 2018 as dry grassland. As can be seen, these areas are located on rather steep slopes and are in 2018 characterised by a mixture of grass, shrubs and tree cover. This can indicate two things: First, that the heat category in the HMB maps relates to vegetation types, which to some degree are convergent with vegetation types characterising dry grassland in contemporary habitat mapping. Second, that parts of 1880 heath areas by 2018 have changed to dry grasslands, characterised by mixtures of grass, shrub and tree cover.

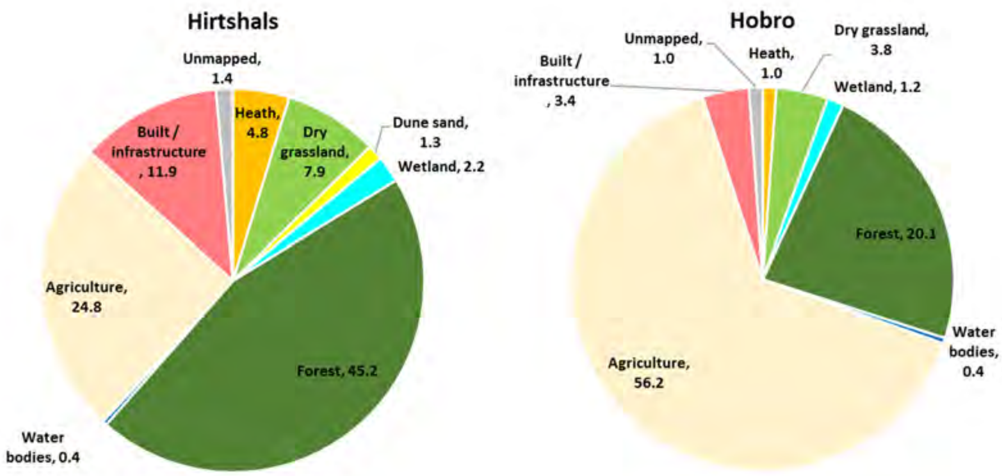


Figure 5.6 2018 land category proportions of 1880 heath.

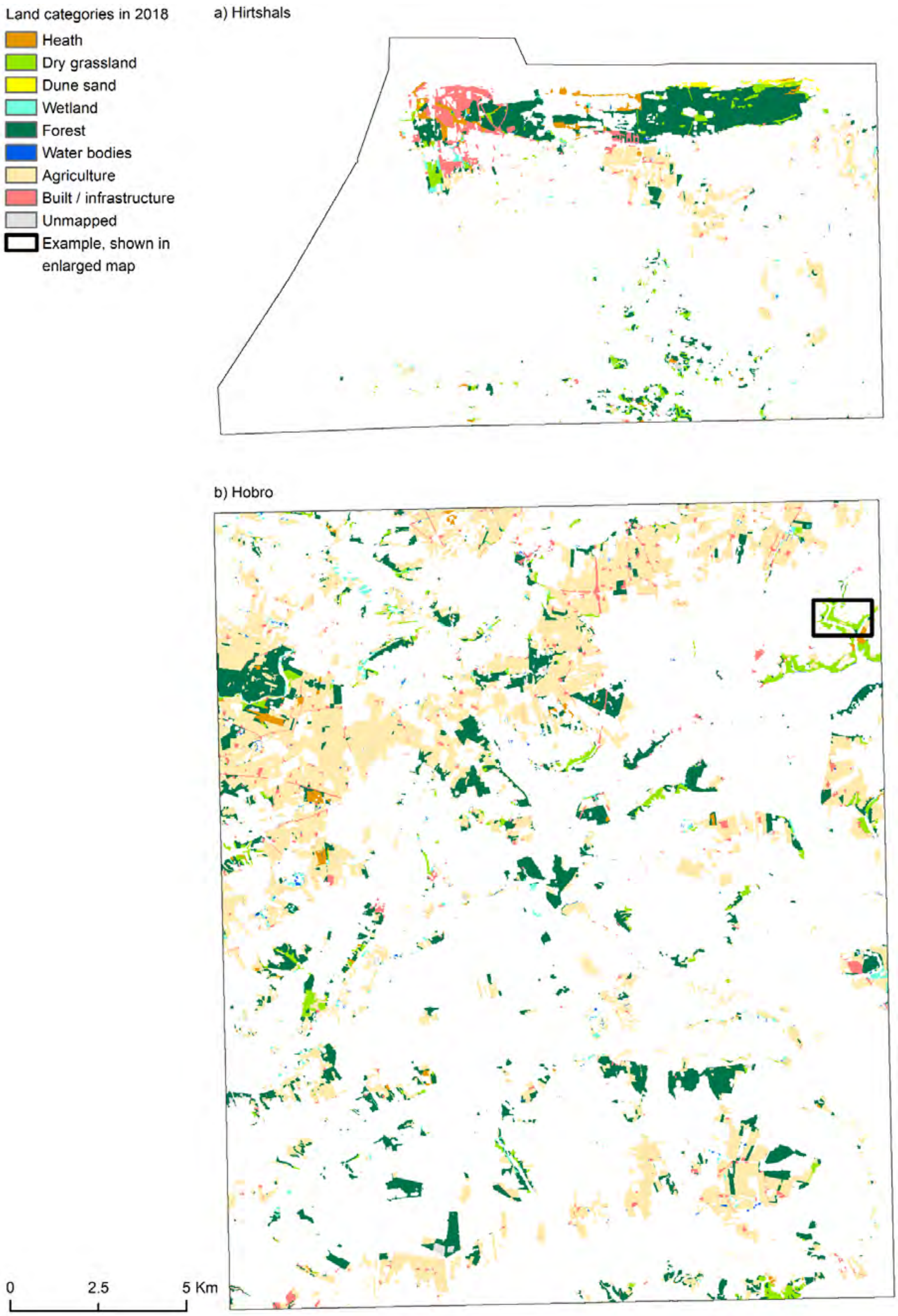


Figure 5.7 2018 land categories of 1880 heath.

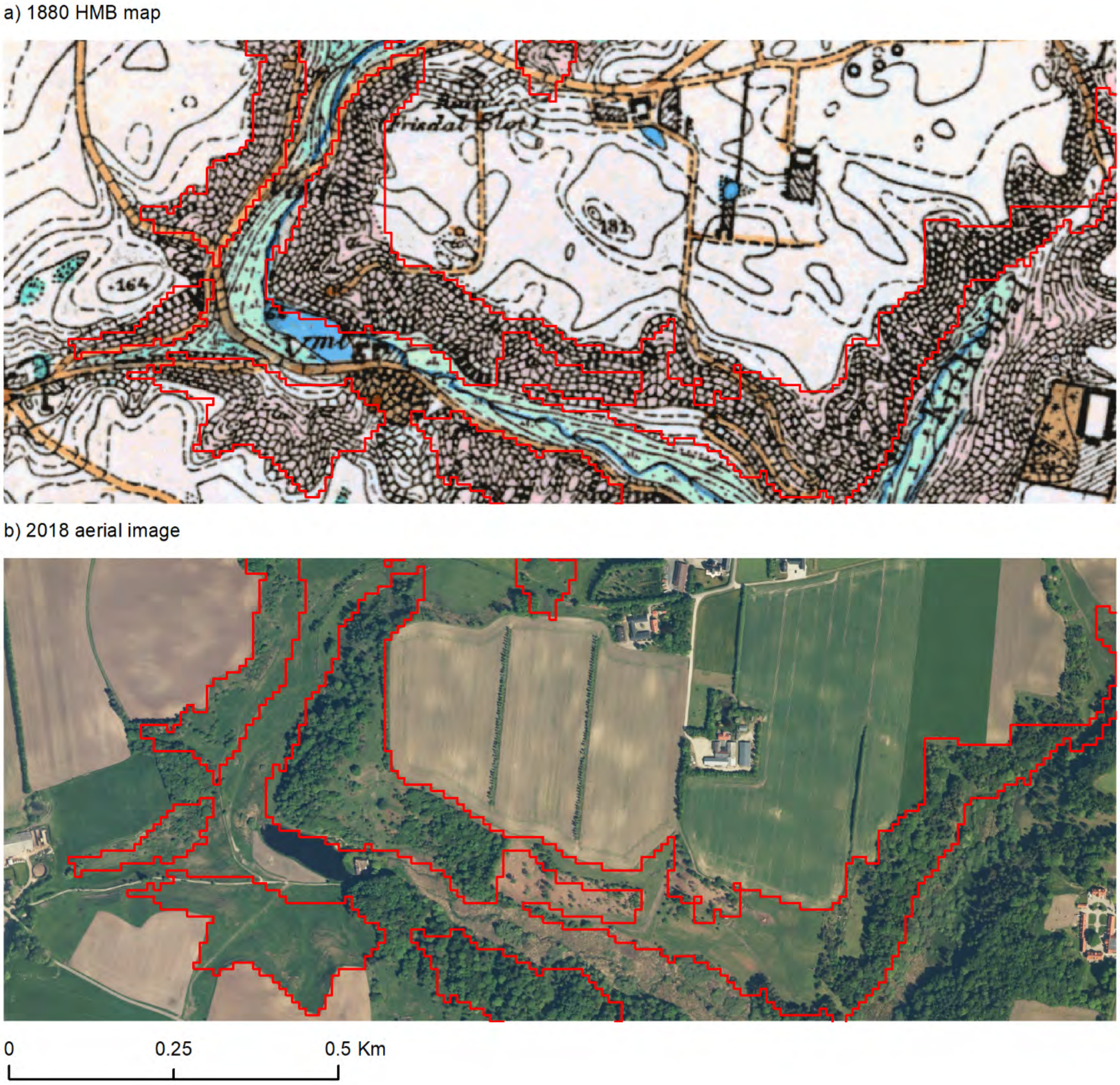


Figure 5.8 Example from the Hobro area, mapped as heath in 1880 and as dry grassland in 2018.

Changes from Dune sand

For dune sand that was mapped in 1880 in the Hirtshals area, the diagram in Figure 5.9 shows the proportion of the different land categories in 2018. The changes are also shown in the map in Figure 5.10. Overall, changes are characterised by a decline of the area of dune sand to the gain of forest (~ 22%), dry grassland (~ 19%), heath (> 15%) and water bodies (> 7%). Around 25 % of the 1880 dune sand area was also mapped as dune sand in 2018. Subdividing the dune sand category from Basemap03 shows that only 1.5 % of 1880 dune sand area was in 2018 mapped as so called shifting dunes, characterised by active sand drift while 6.2 % was mapped as sand surfaces (primarily along the sea shore). Around 17 % was in 2018 mapped as dunes with vegetation cover. The map in Figure 5.11 shows an example from the south-western part of the Hirtshals area, which in 1880 was mapped as dune sand and in 2018 as dry grassland and Figure 5.12 shows an example from the north-western part of the Hirtshals area, which in 1880 was mapped as dune sand and in 2018 as heath. In both examples, these areas are characterised by the sand dune signature on the HMB map sheet. On inspection of the 2018 aerial images, it appears that they are dominated by grass and herb vegetation with some scattered trees and bushes. Figure 5.13 shows an area, which was mapped as dune sand in both 1880 and in 2018. The subdivision of the dune sand category illustrates that in 2018, a large area in the eastern part in 2018 was mapped as dune with vegetation cover and the area along the shoreline was mapped as sand surface. Only a narrow area to the east of the shore was in 2018 mapped as shifting dunes. The overall numbers and the examples indicate a development, where particularly afforestation and measurements for the prevention of sand drift, such as planting of lyme grass and other grass and herb species led to a substantial decline in the area of shifting dunes, characterised by active erosion and deposition of sand. However, the change from dune sand to grass and herb dominated vegetation cover can also be subscribed to the fluidity between the mapping concepts of the HMB maps and the more recent with a shift from mapping of the land for mainly military interests to a broader purpose base, including nature conservation.

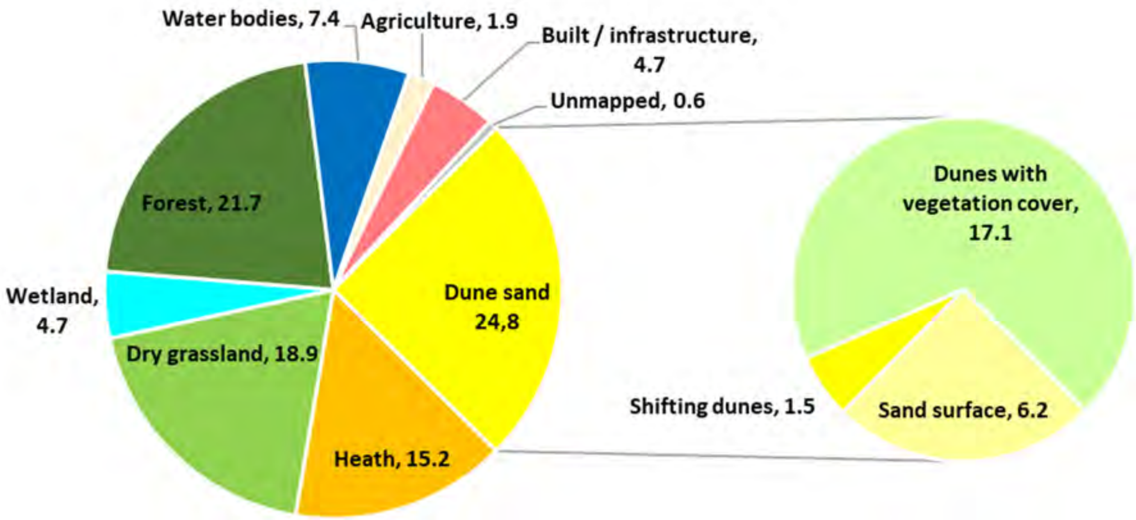


Figure 5.9 2018 land category proportions of 1880 dune sand in the Hirtshals area.

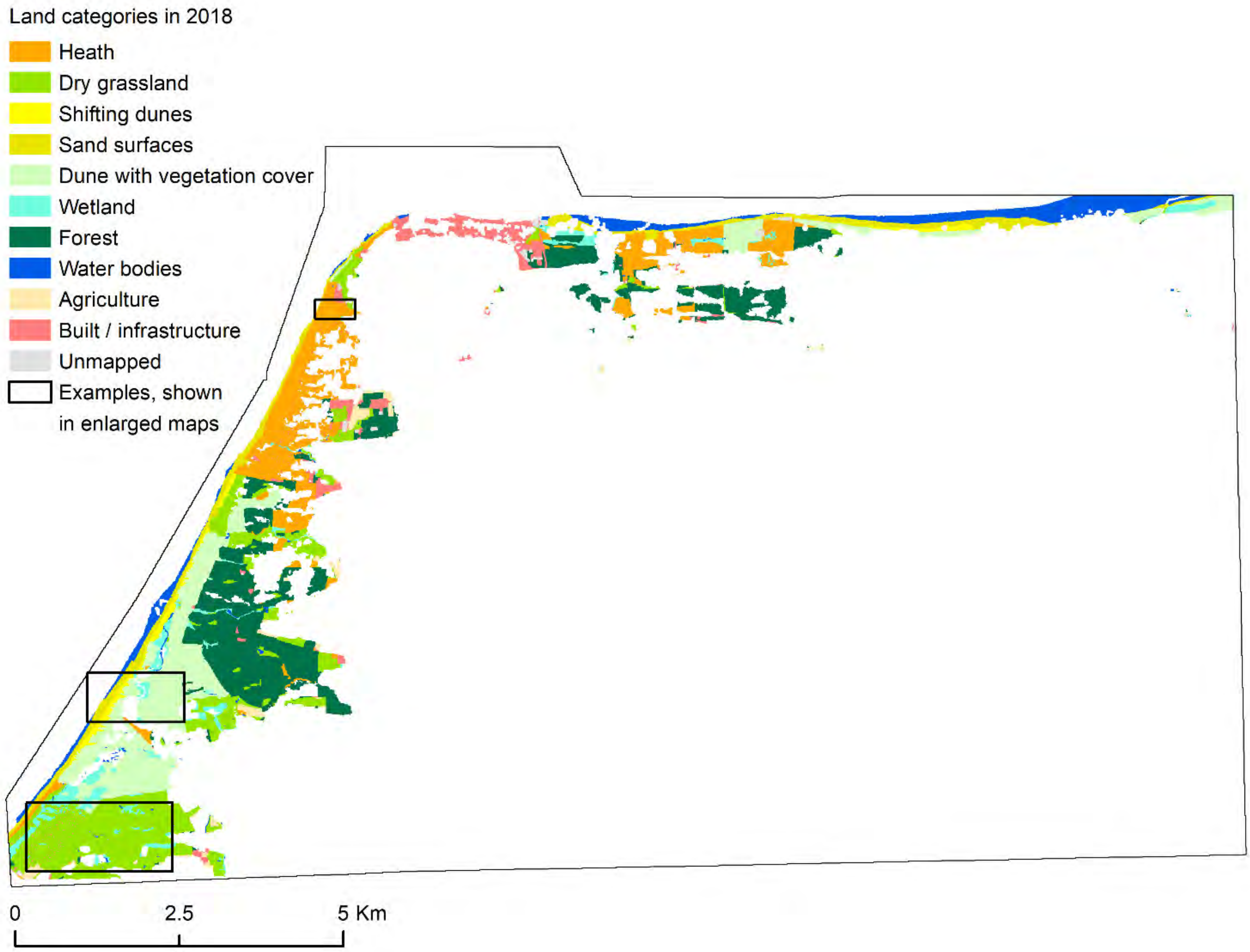


Figure 5.10 2018 land categories of 1880 dune sand in 1880 in the Hirtshals area.

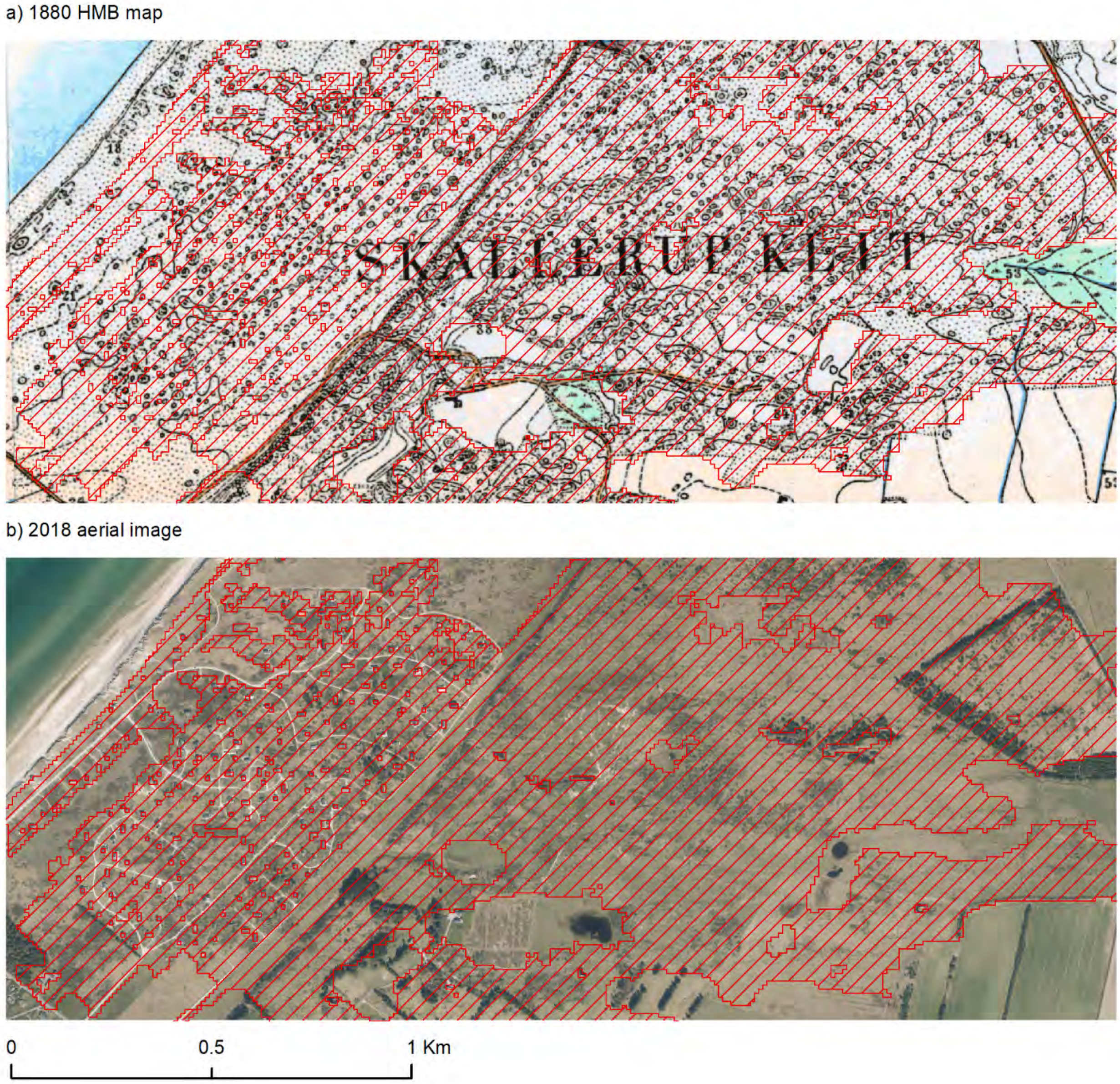


Figure 5.11 Example from the Hirtshals area, mapped as dune sand in 1880 and as dry grassland in 2018.

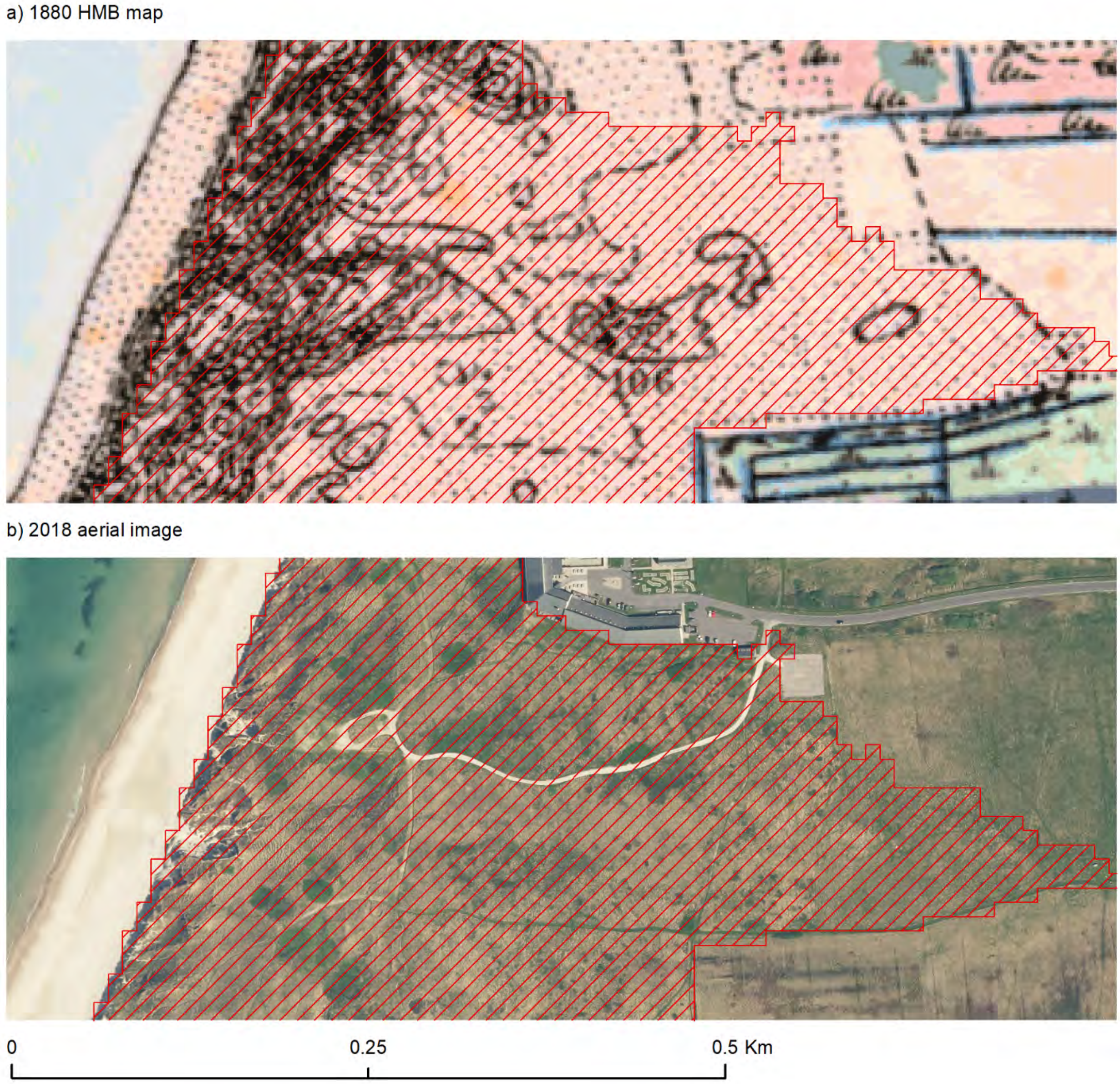


Figure 5.12 Example from the Hirtshals area, mapped as dune sand in 1880 and as heath in 2018.

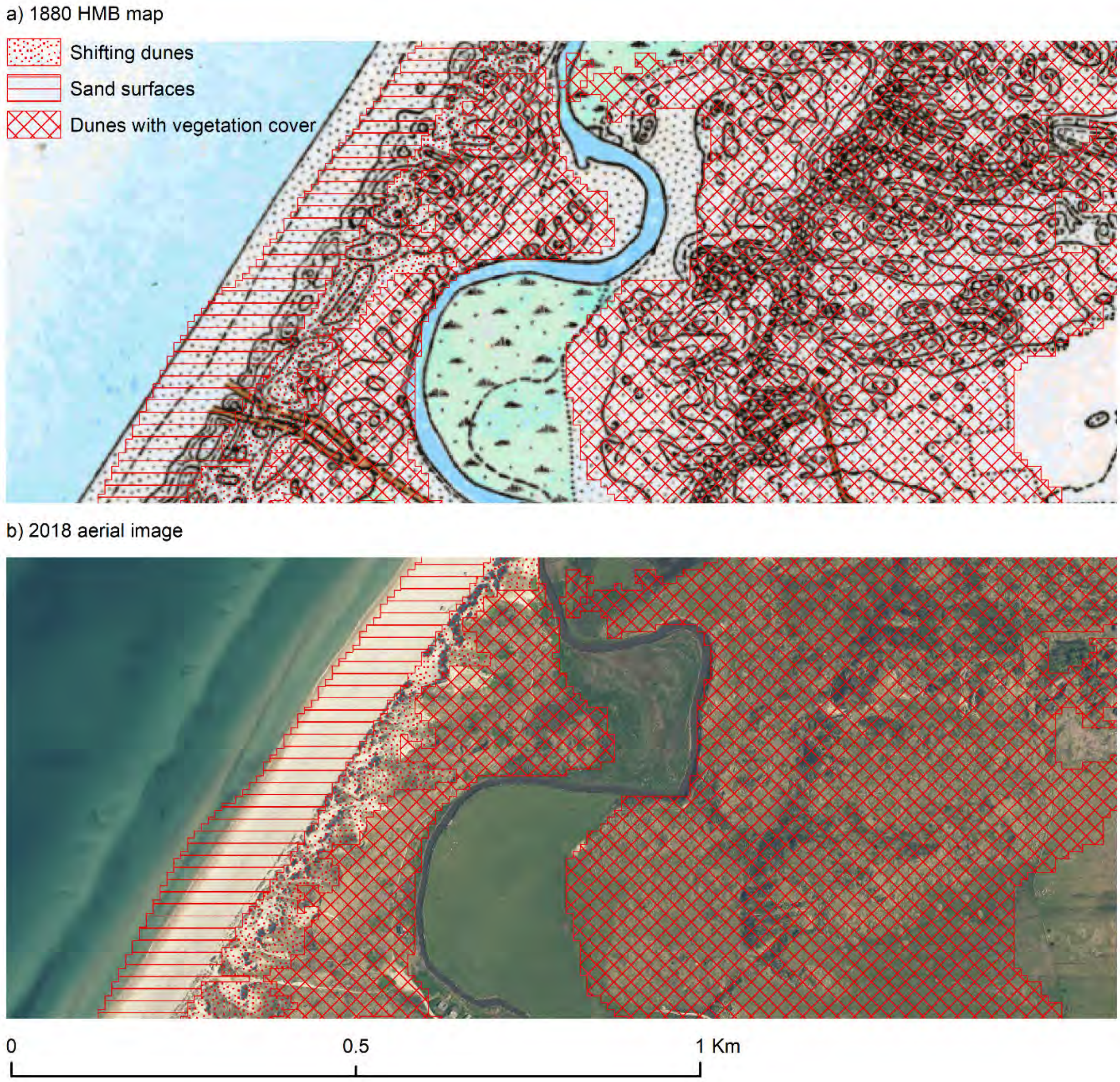


Figure 5.13 Example from the Hirtshals area, mapped as dune sand in 1880 and in 2018.

Changes from wetland

For wetland that was mapped in 1880, the diagrams in Figure 5.14 show the proportion of the different land categories mapped in 2018. The changes are also shown in the maps in Figure 5.15. The figures and maps illustrate overall LULC dynamics, characterised by a decline in wetland, mainly to the gain of agriculture (~ 35% in the Hirtshals area; ~ 44% in the Hobro area) and some afforestation (~ 8% in the Hirtshals area; ~ 4% in the Hobro area). In the Hirtshals area, a considerable proportion of the 1880 wetland had in 2018 changed to dry grassland (~ 9%) or to heath (~ 7%). Figure 5.16 shows an example from the western part of the Hirtshals area, which in 1880 was mapped as wetland and in 2018 as dry grassland. On the 2018 aerial image, several ditches can be seen, which inspection of the HMB maps reveals as not having been mapped for 1880. This indicates more recent, manmade drainage. Figure 5.17 shows an example from the western part of the Hirtshals area, which in 1880 was mapped as wetland and in 2018 as heath. Possible explanations are deposition of sand from the dune sand areas to the east and / or drainage of the area in relation to establishment of holiday houses, which can be seen on the 2018 aerial image.

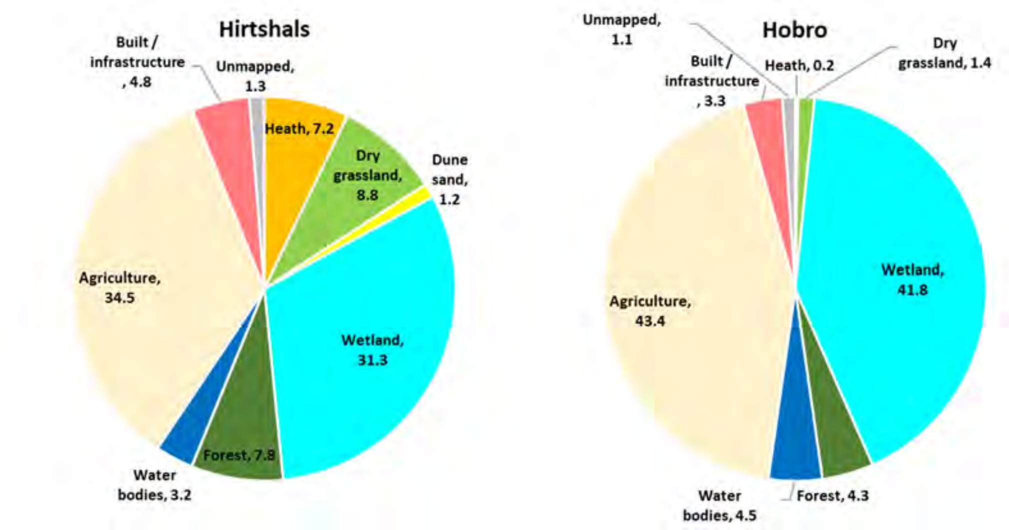


Figure 5.14 2018 land category proportions of 1880 wetland.

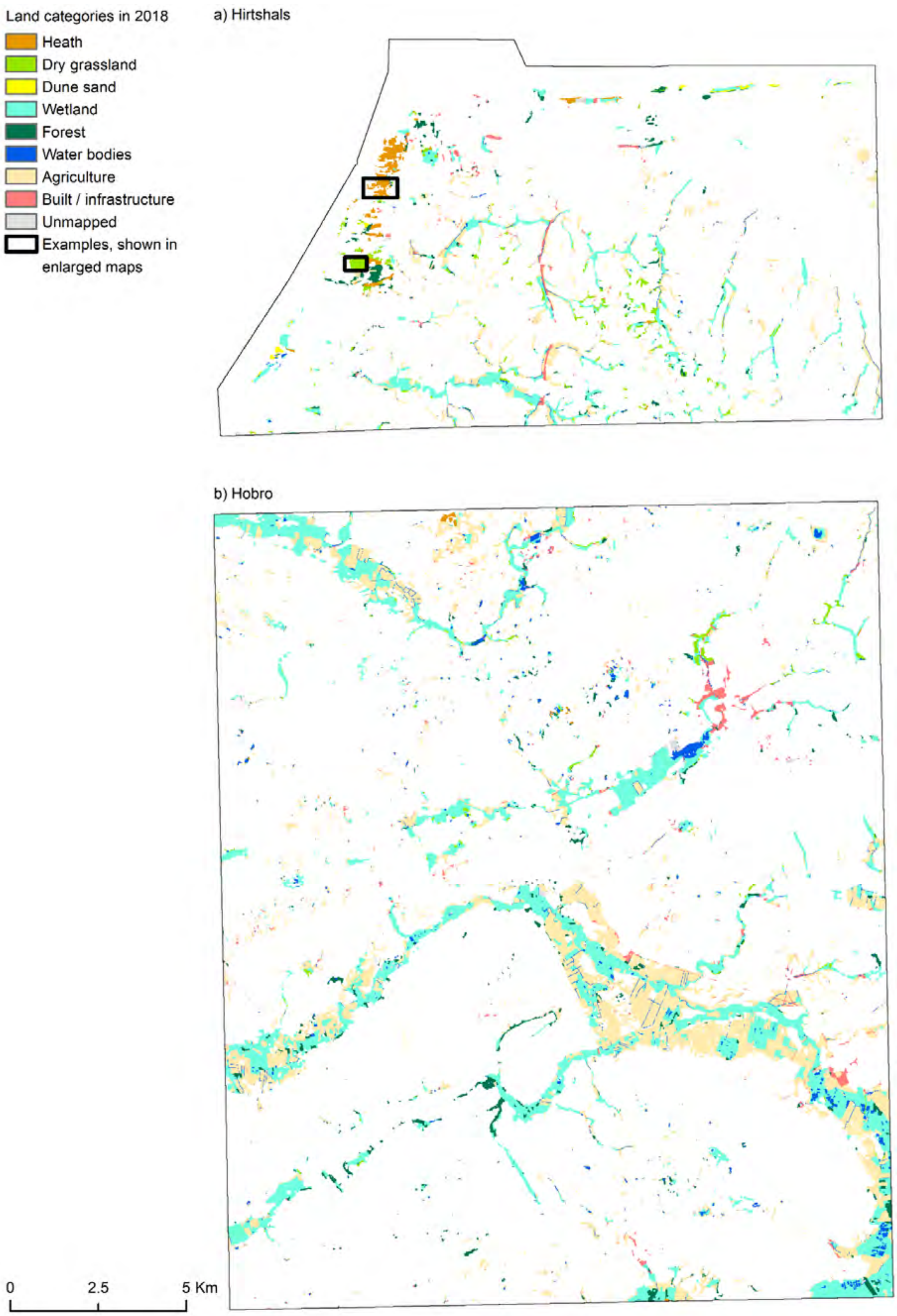


Figure 5.15 2018 land categories of 1880 wetland.

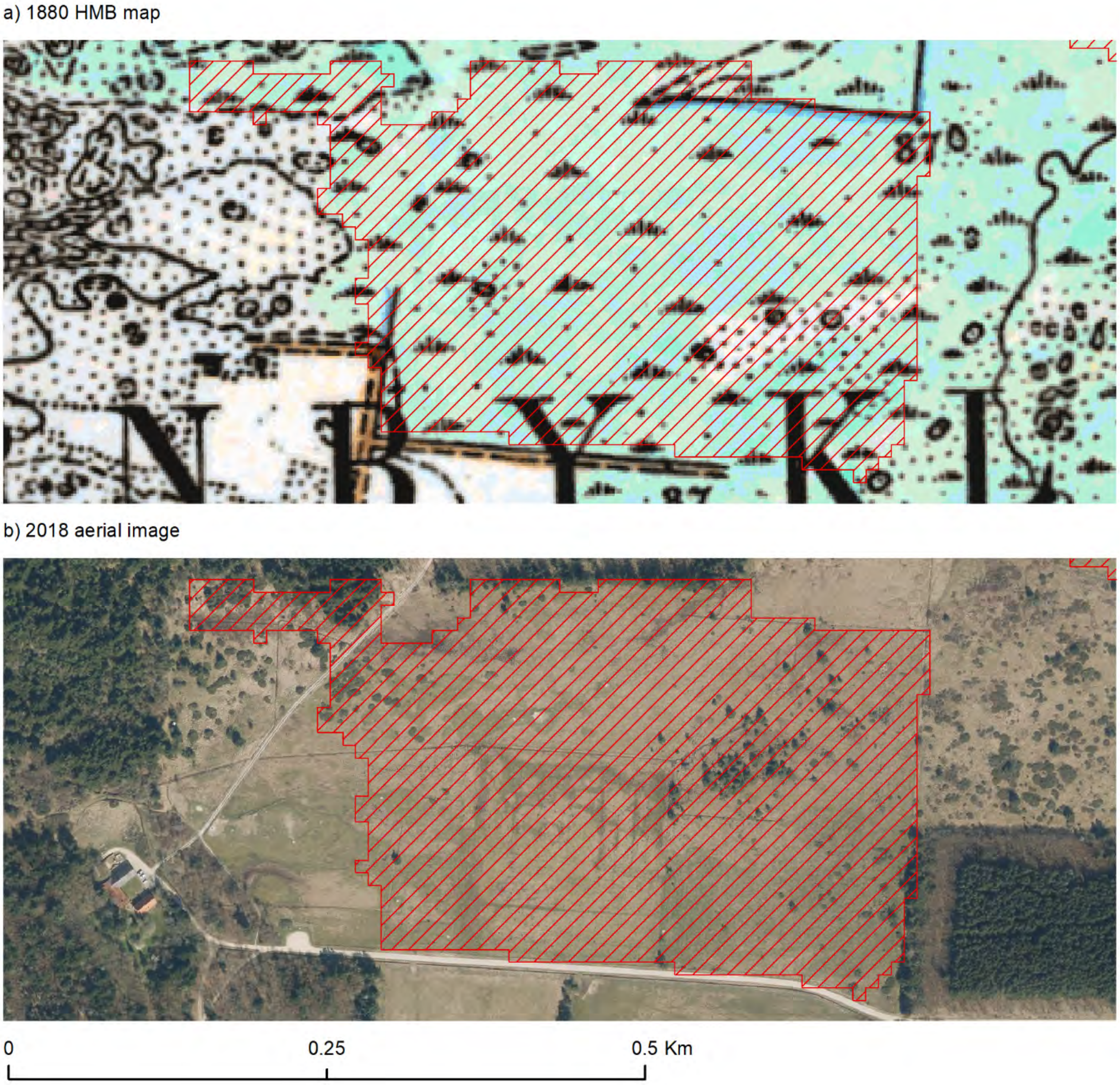


Figure 5.16 Example from the Hirtshals area, mapped as wetland in 1880 and as dry grassland in 2018.

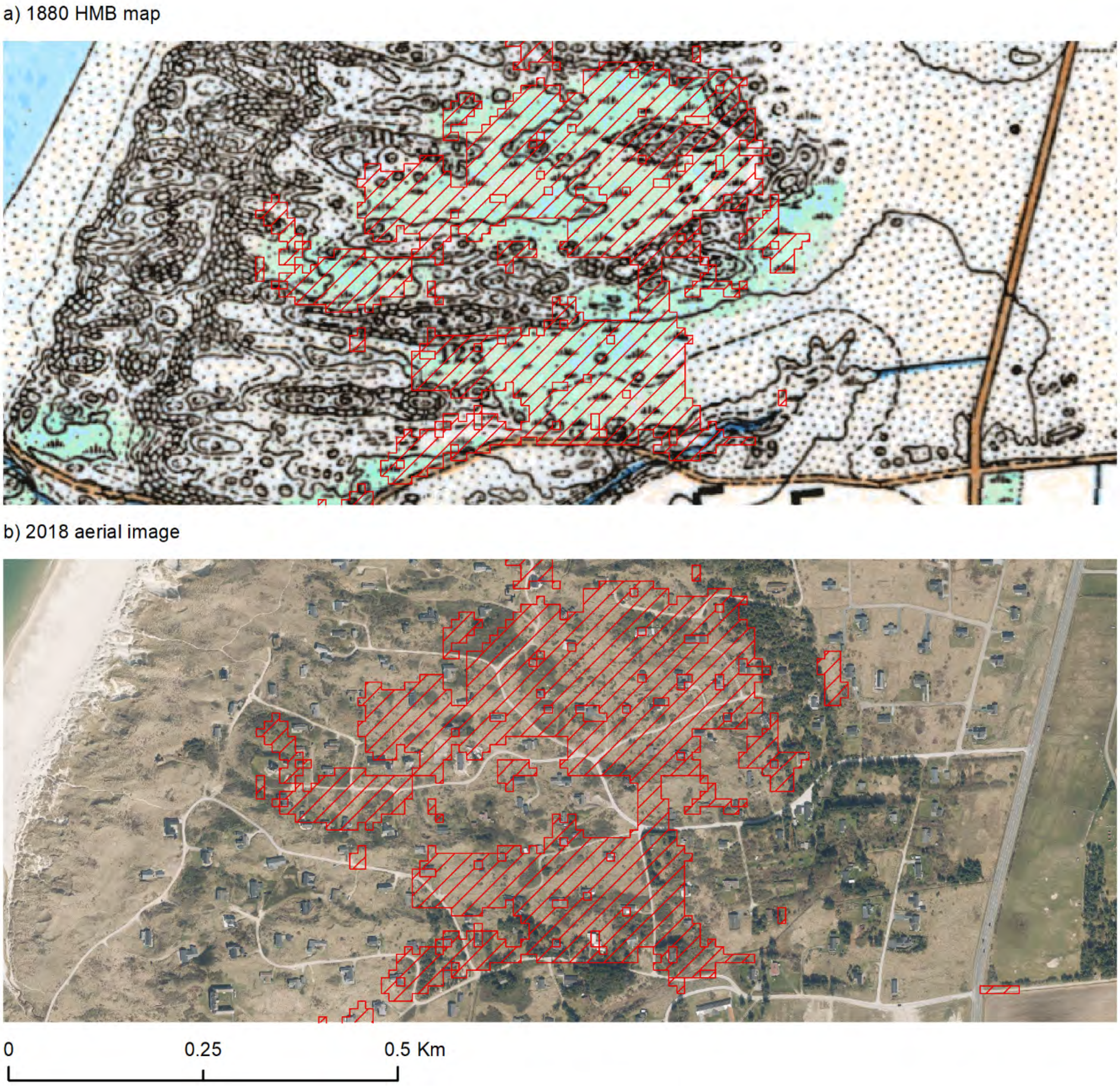


Figure 5.17 Example from the Hirtshals area, mapped as wetland in 1880 and as heath in 2018.

Changes from water bodies

For water bodies that were mapped in 1880, the diagrams in Figure 5.18 show the proportion of the different land categories mapped in 2018. The changes are also shown in the maps in Figure 5.19. For both areas, the figures and maps show that the major part of water bodies in 1880, in 2018 was still water. However, In the Hirtshals area, some had in 2018 changed to built / infrastructure (~ 5%) and in the Hobro area, small proportions had changed to forest (~ 4%) and to wetland (~ 5%). Figure 5.20 shows and example from the Hirtshals area, where larger areas of sea were reclaimed for the construction of the Hirtshals harbour. Figure 5.21 shows the Klejtrup Lake in the Hobro area, where large areas along the lakeshore have changed from open water in 1880 to wetland in 2018. Figure 5.22 shows an example from the Tjele Lake in the Hobro area, where, by 2018, large areas along the lakeshore had changed to forest. A subdivision of the forest category from Basemap03 elucidates that these areas are also categorised as alluvial forest. Consequently, both examples illustrate a natural process of sedimentation and consequent change from open water surfaces to permanently or periodically wet habitat types.

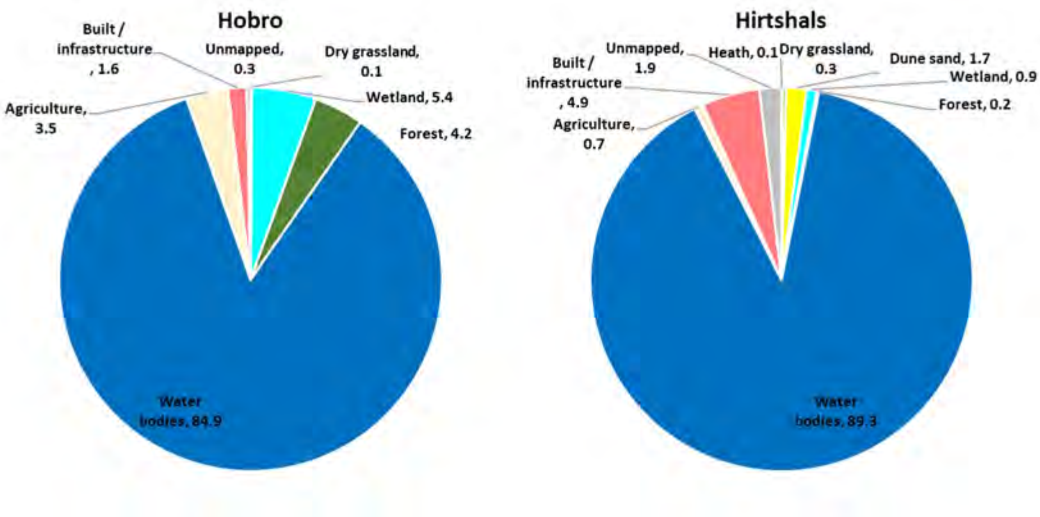


Figure 5.18 2018 land category proportions of 1880 water bodies.

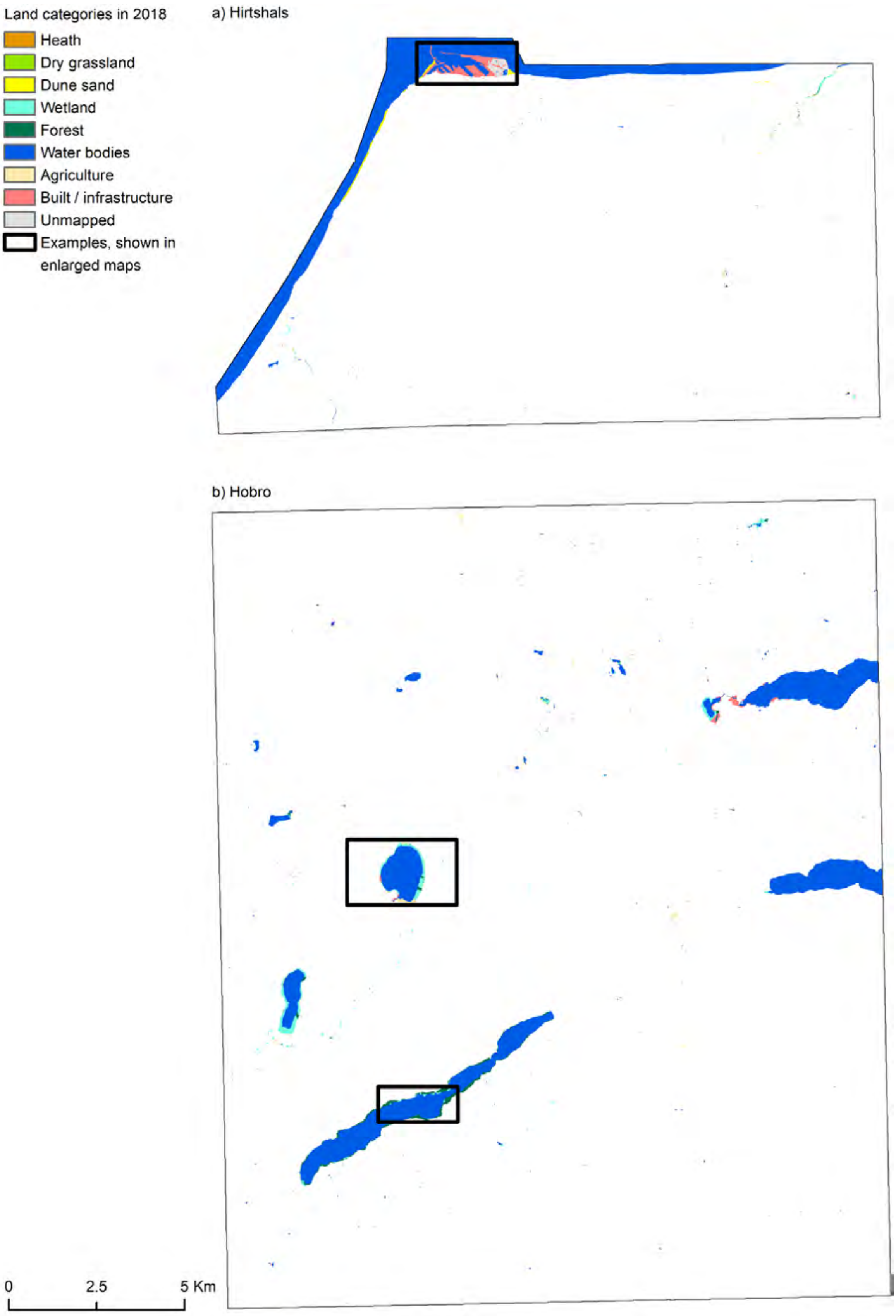


Figure 5.19 2018 land categories of 1880 water bodies.

Figure 5.20 Example from the Hirtshals area, mapped as water body in 1880 and as built/infrastructure in 2018.

a) 1880 HMB map

b) 2018 aerial image

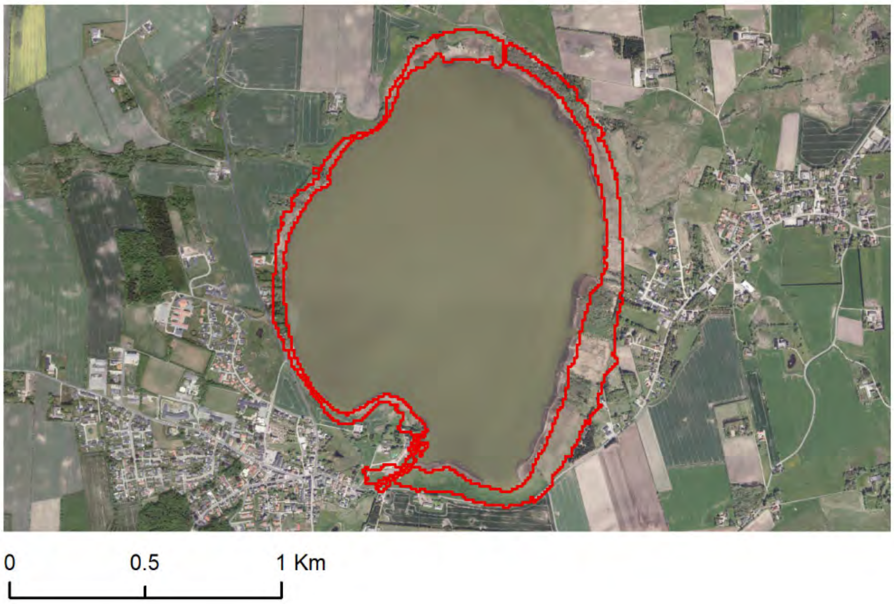


Figure 5.21 Example from the Hobro area, mapped as water body in 1880 and as wetland in 2018.

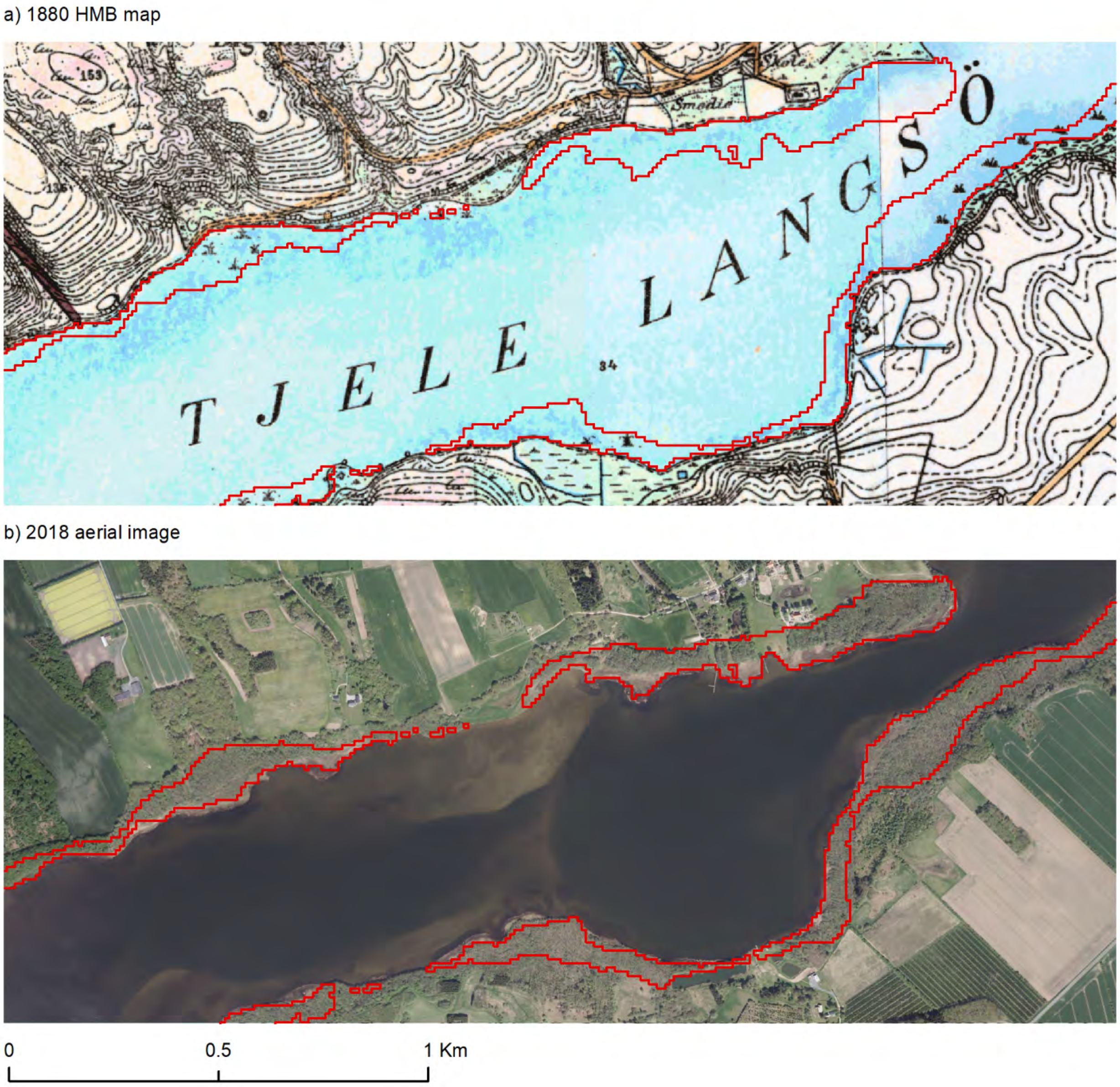


Figure 5.22 Example from the Hobro area, mapped as water body in 1880 and as forest in 2018.

Changes to agriculture

For agricultural land in 2018, the diagrams in Figure 5.23 show the proportion of the different land categories mapped in 1880. The changes are also shown in the maps in Figure 5.24. The figures and maps illustrate overall LULC dynamics, characterised by an increase in agricultural land at the expense of heath and wetland. Agricultural land was not included as a land category in the HMB survey but is a subset of the "other" category. It can thus be assumed that a large proportion of the area, which in 2018 was mapped as agricultural land and which was not mapped in the HMB maps, in 1880 were used for agriculture. In Hirtshals, the "other" category makes up around 88 % of 2018 agricultural land in Hobro around 70 %. This indicates a more significant increase in agricultural land in the Hobro area than in the Hirtshals area.

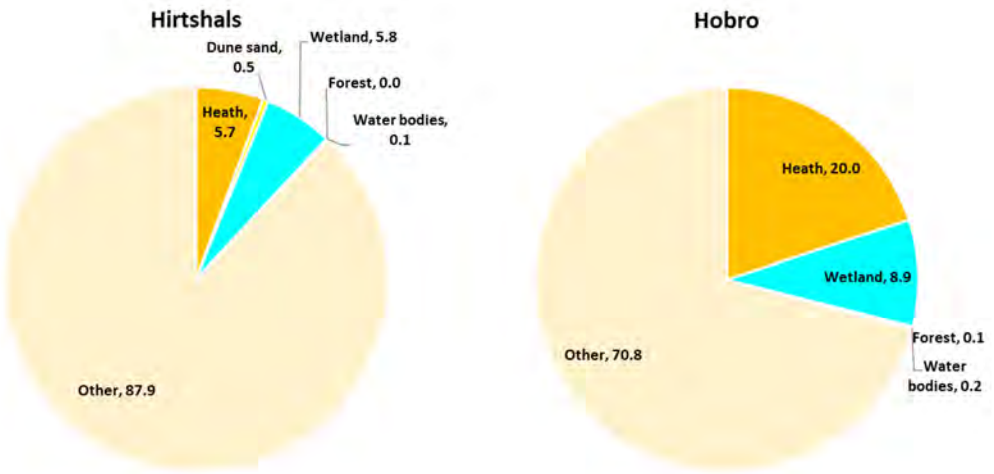


Figure 5.23 1880 land category proportions of 2018 agriculture.

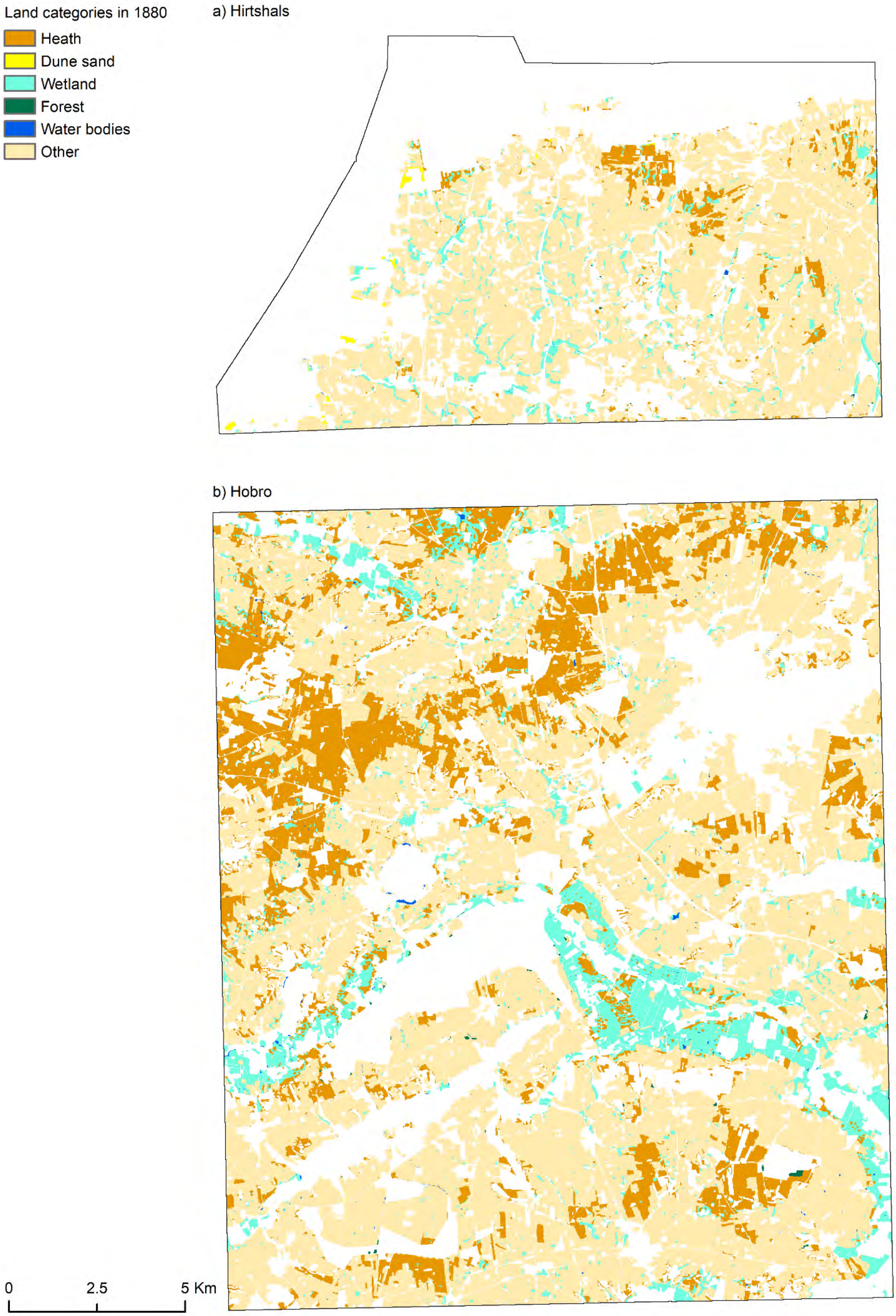


Figure 5.24 1880 land categories of 2018 agriculture.

Changes to forest

For forest in 2018, the diagrams in Figure 5.25 show the proportion of the different land categories mapped in 1880. The changes are also shown in the maps in Figure 5.26. The figures and maps illustrate overall LULC dynamics, characterised by an increase of forest the expense of open, non-forested habitat types and other land, which probably mainly comprises agricultural land. While in the Hobro area, about 28 % of the 2018 forest area was also forest in 1880, this is only the case for around 3 % in the Hirtshals area. This indicates a higher rate of afforestation in the Hirtshals area, compared to the Hobro area.

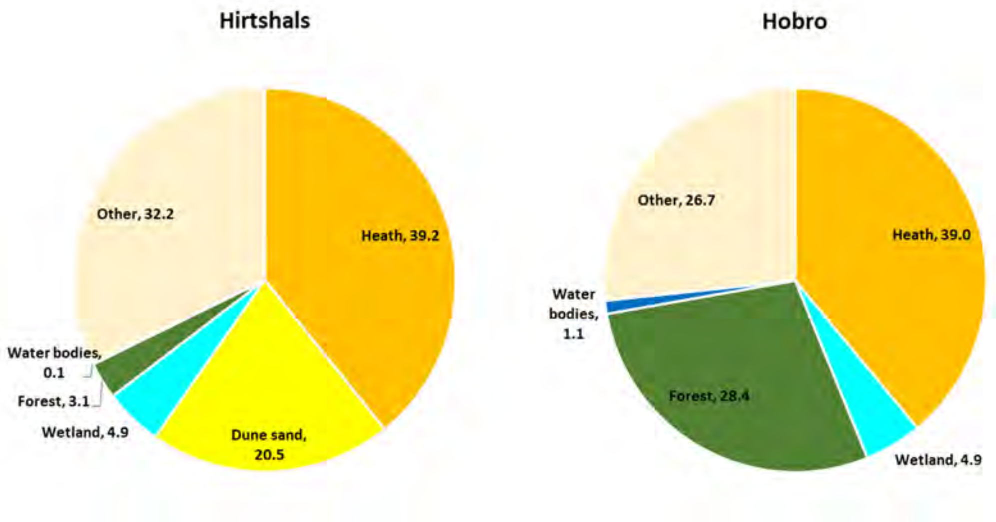


Figure 5.25 1880 land category proportions of 2018 forest.

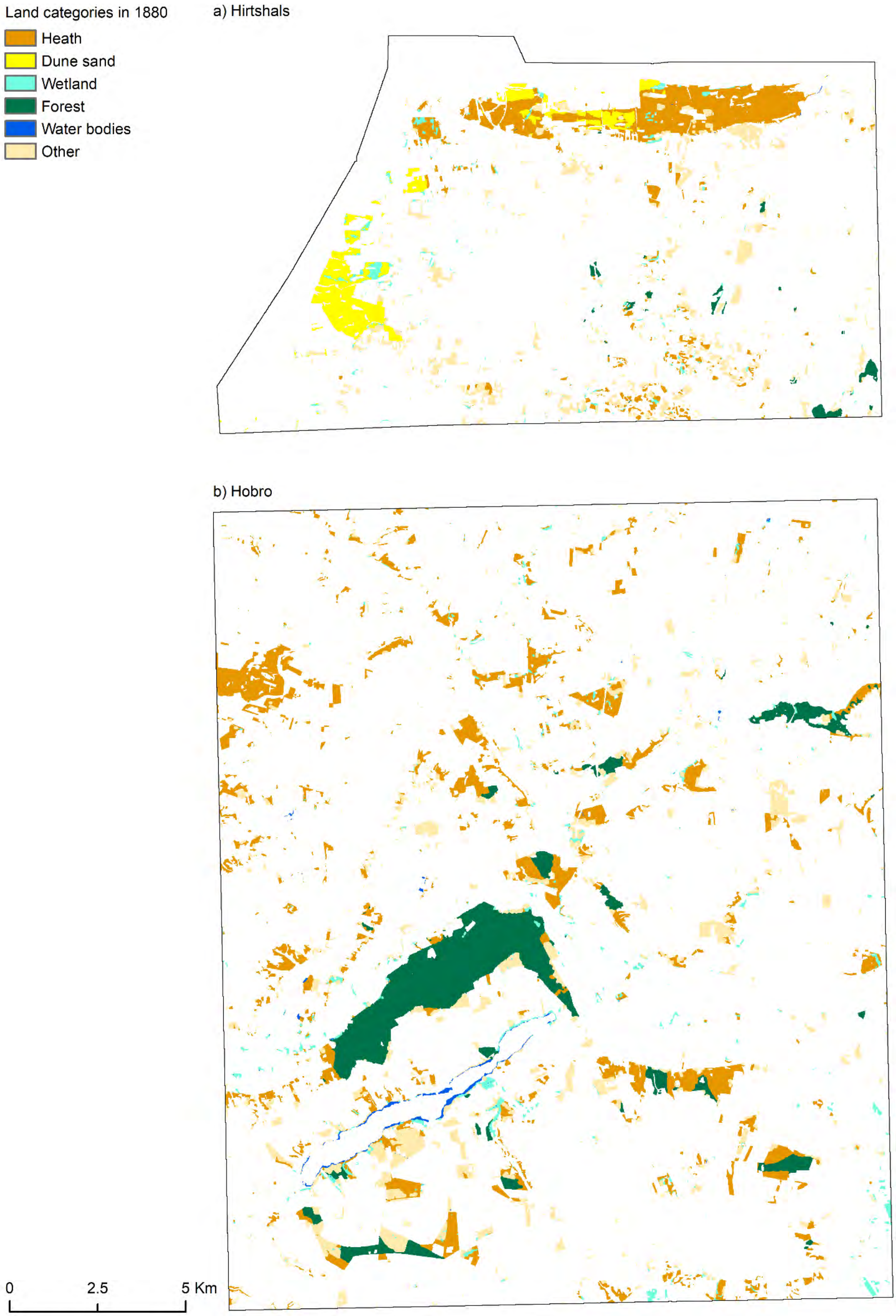


Figure 5.26 1880 land categories of 2018 forest.

Change to open, non-forested habitat types

Overall LULC dynamics in the two study areas are characterised by a decline of open, non-forested habitat types. However, the change assessment also revealed areas, which in 2018 are mapped as open, non-forested habitat types, which were not included in the mapping, based on the HMB map sheets. Table 5.7 summarises the area of the different habitat types in 2018, which were not extracted from the HMB map sheets. These areas are also shown on the maps in Figure 5.27. The proportion is highest for dry grassland (~ 43% in the Hirtshals areas; ~ 29% in the Hobro area. Figure 5.28 shows an example from the central part of the Hirtshals area, where there are no signatures indicating the presence in 1880 of forest, wetland, heath or dune sand, while from the 2018 aerial image, the area appears dominated by grass and herb vegetation. In this case, the lack of the area being present in the HMB map as a mapped category is assumed to be a consequence of the dry grassland category not having been mapped as part of the HMB survey. Figure 5.29 shows an example from the north-eastern part of the Hirtshals study area, where a large belt of dunes is mapped in 2018 but is not included in the HMB map. This might be due to the 2018 dune sand category also including vegetated dunes. It might also be the consequence of sand drift from the coastline depositing sand inland. Finally, Figure 5.30 shows an example from the eastern part of the Hobro area, where a narrow band of wetland was mapped in 2018 but is missing in the wetland layer derived from the HMB map sheets. Here, our methodology for extracting land categories from the HMB maps was unable to detect the HMB signature, possibly due to the indistinct colouration and small colour extent in the HMB maps.

Table 5.7 Area and proportion of open, non-forested habitat types, which were only mapped in 2018.

	Hirtshals			Hobro		
	Total	Not included in HMB map		Total	Not included in HMB map	
	Area (km²)	Area (km²)	Proportion (%)	Area (km²)	Area (km²)	Proportion (%)
Heath	5.7	0.6	10.7	1.3	0.2	11.7
Dry grassland	11.6	4.8	41.5	6.5	1.9	28.8
Dune sand	6.5	0.7	10.9	-	-	-
Wetland	7.5	1.6	21.3	27.4	1.3	4.7

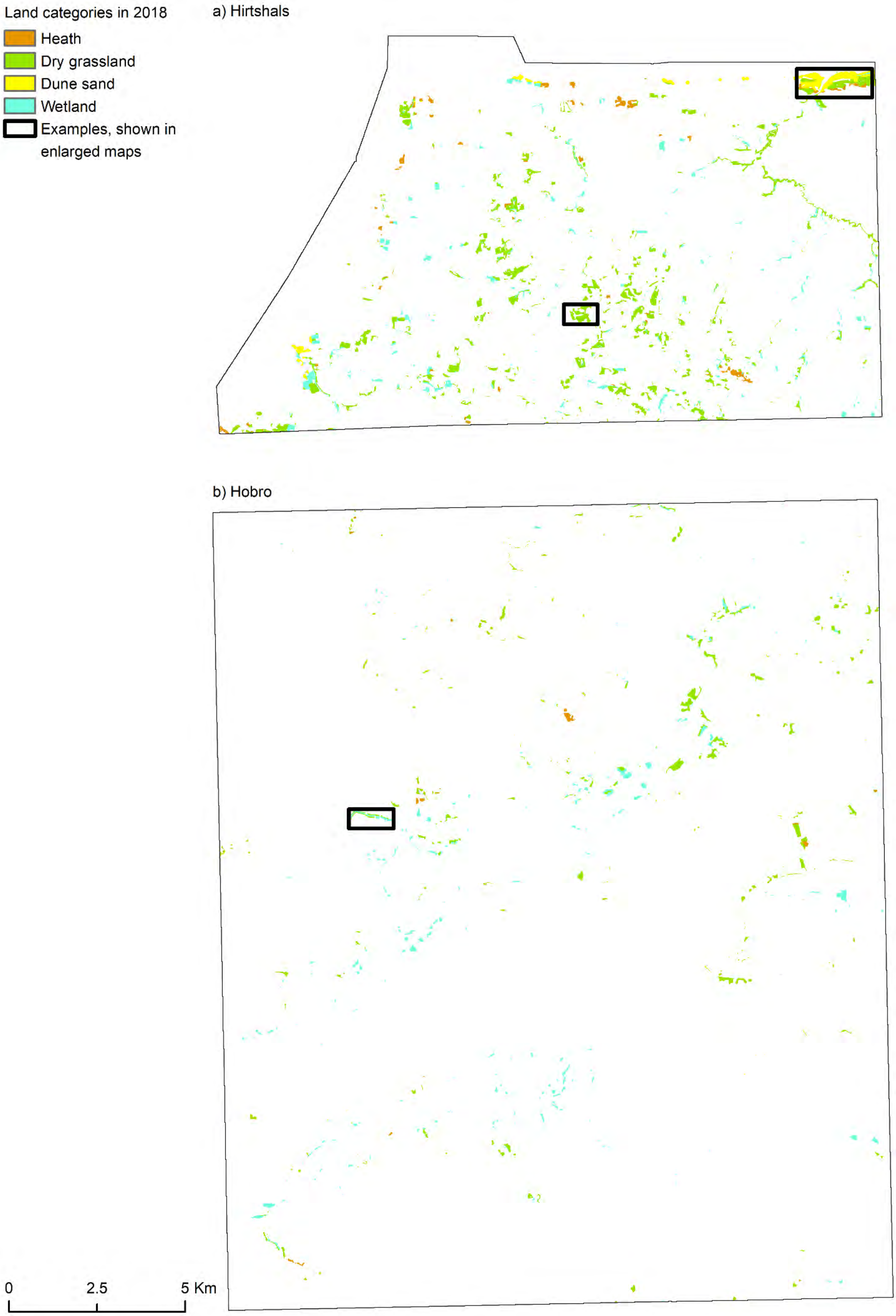
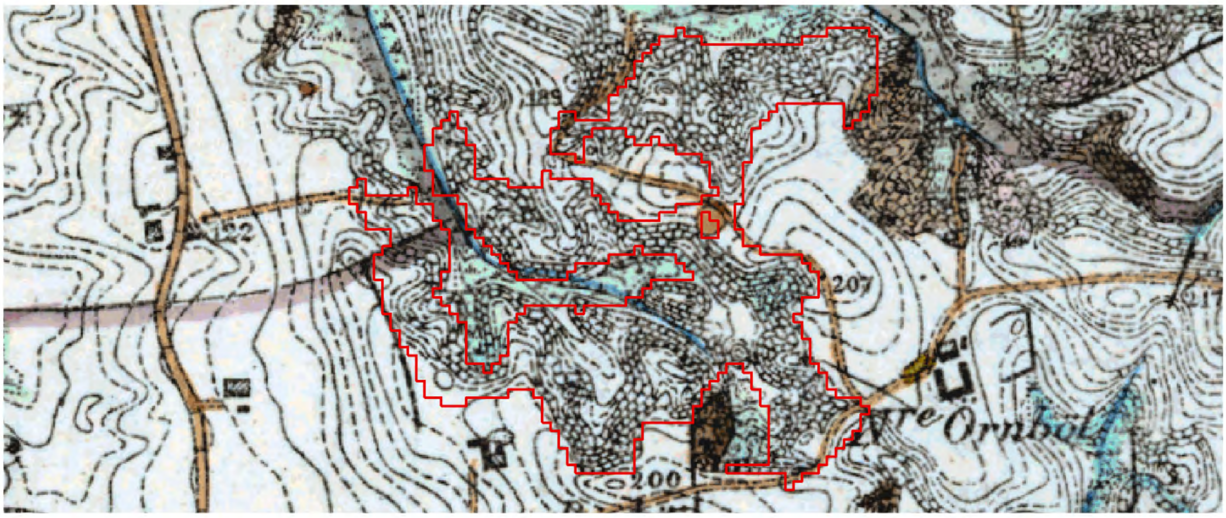


Figure 5.27 Open, non-forested habitat types, which were only mapped in 2018.

a) 1880 HMB map



b) 2018 aerial image

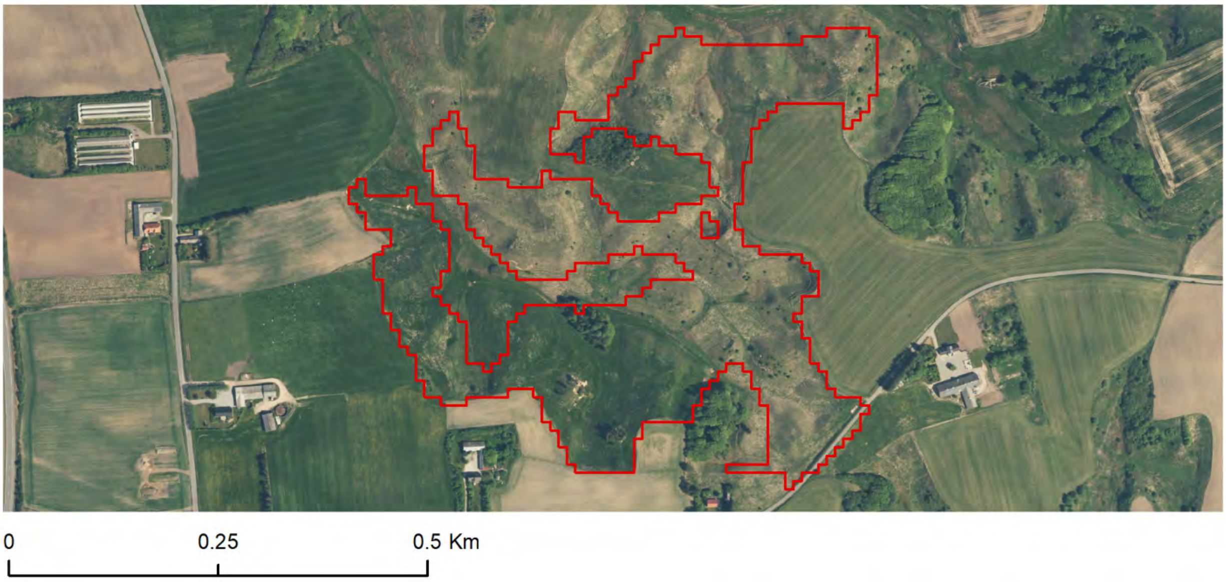


Figure 5.28 Example from the Hirtshals area, mapped as dry grassland in 2018 and not mapped in 1880.

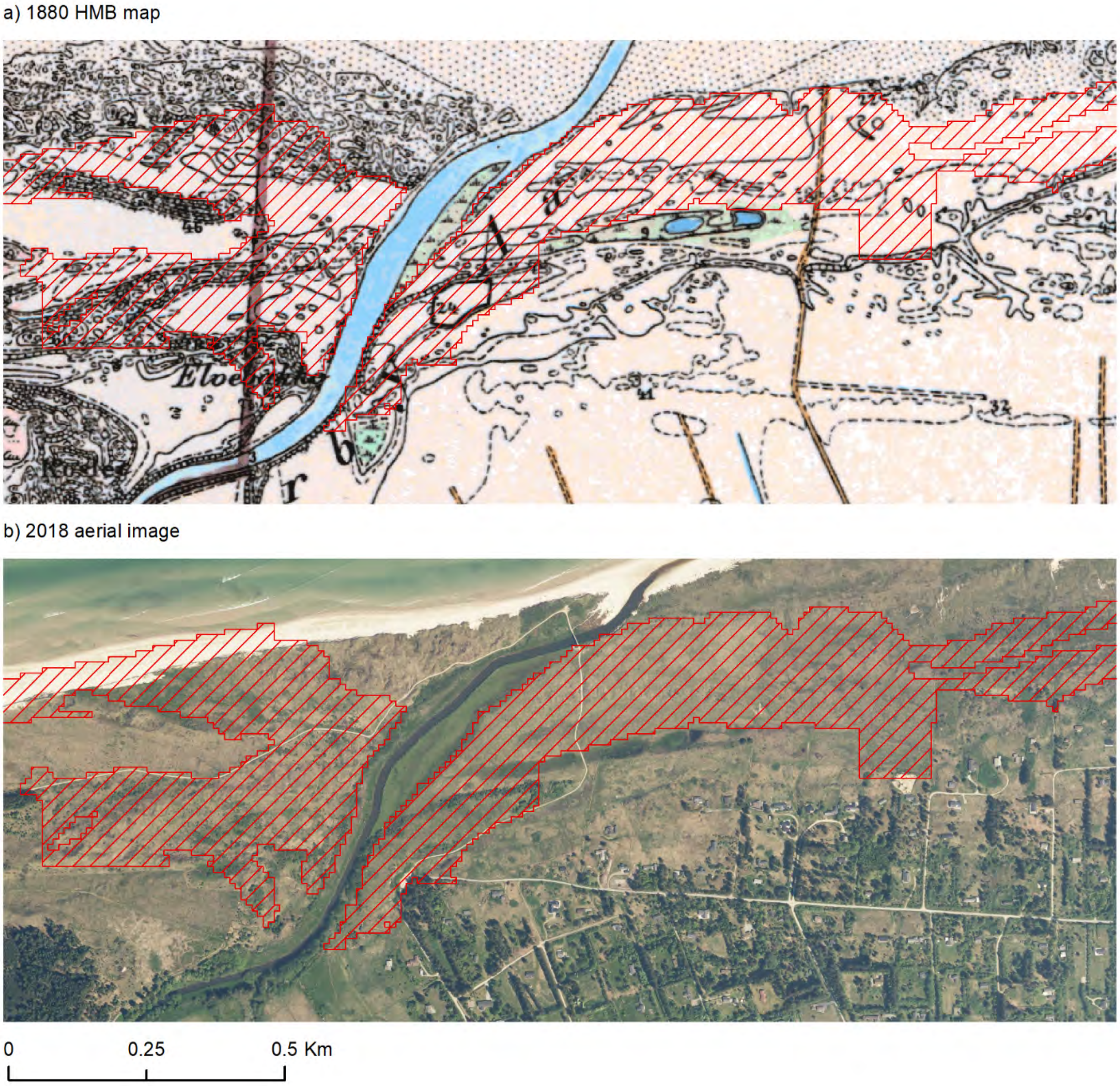
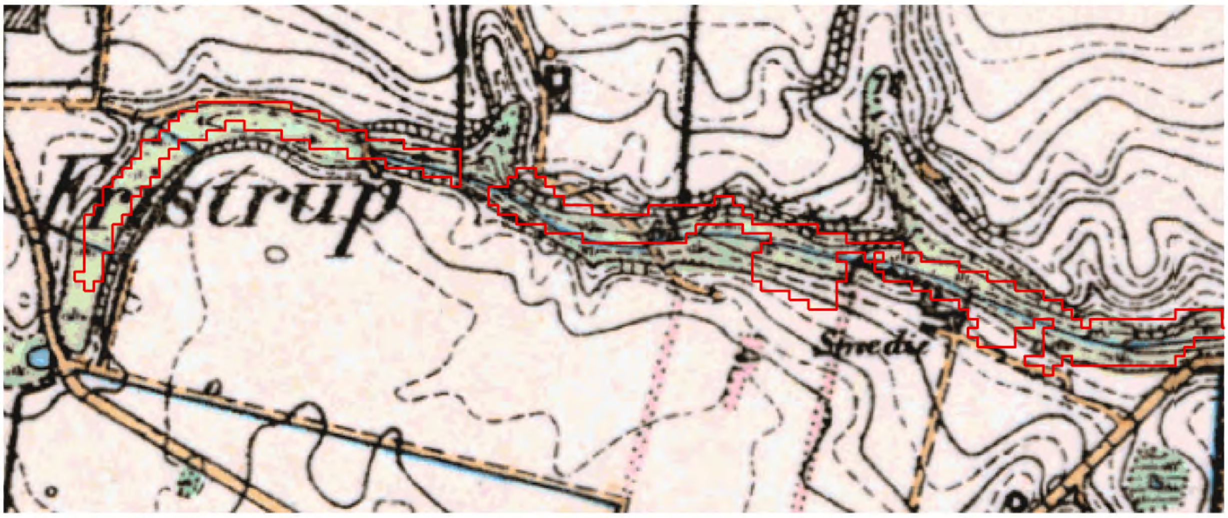


Figure 5.29 Example from the Hirtshals area, mapped as dune sand in 2018 and not mapped in 1880.

a) 1880 HMB map



b) 2018 aerial image

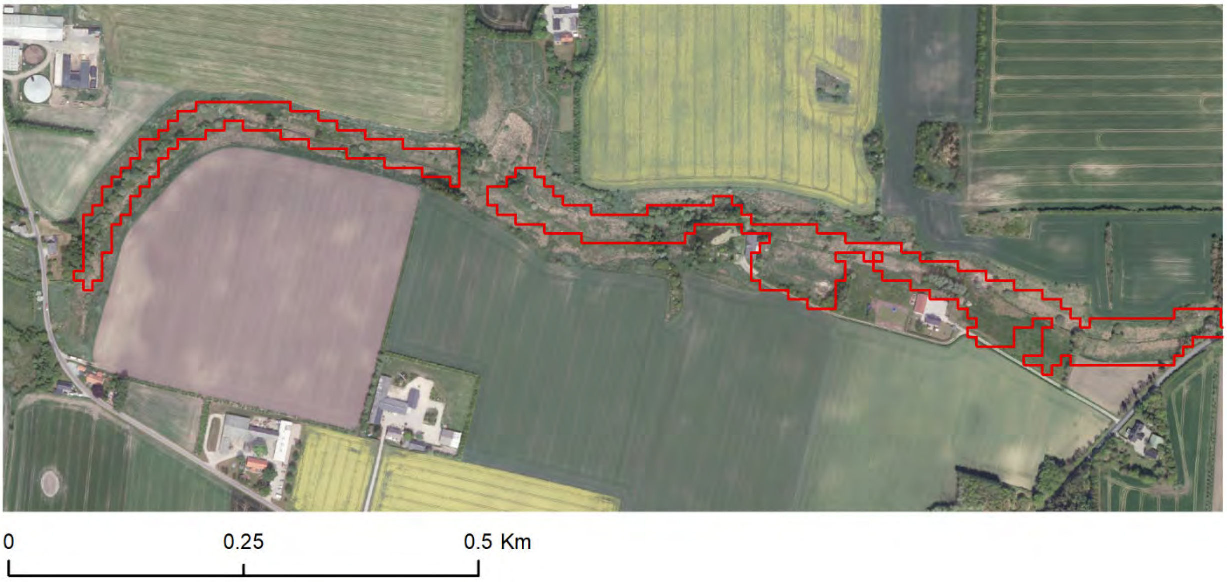


Figure 5.30 Example from the Hobro area, mapped as wetland in 2018, which was not extracted from the HMB map with our model.

Changes to water bodies

For water bodies in 2018, the diagrams in Figure 5.31 show the proportion of the different land categories mapped in 1880. The changes are also shown in the maps in Figure 5.32. For both study areas, the figures and maps show that a proportion of around 25 % of water bodies, mapped in 2018 were not mapped as water bodies in 1880. In the Hirtshals area about 12 % was in 1880 mapped as Dune sand. Figure 5.33 shows an example in the north-eastern part of the Hirtshals area, where a large area of dune sand by 2018 had changed to sea. The example illustrates a process of coastal erosion. Figure 5.34 shows an example from the Hobro area, where a lake has been established on a former wetland area.

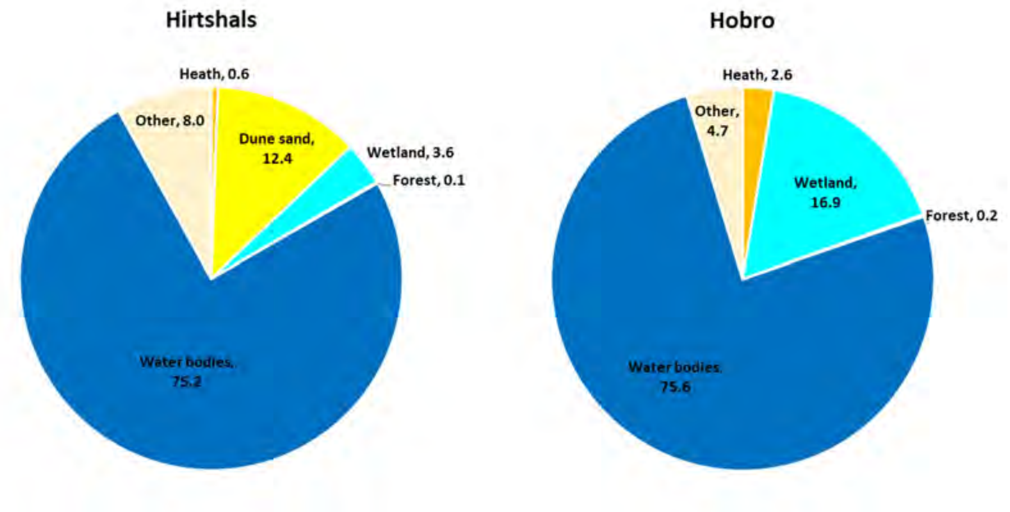


Figure 5.31 1880 land category proportions of 2018 water bodies.

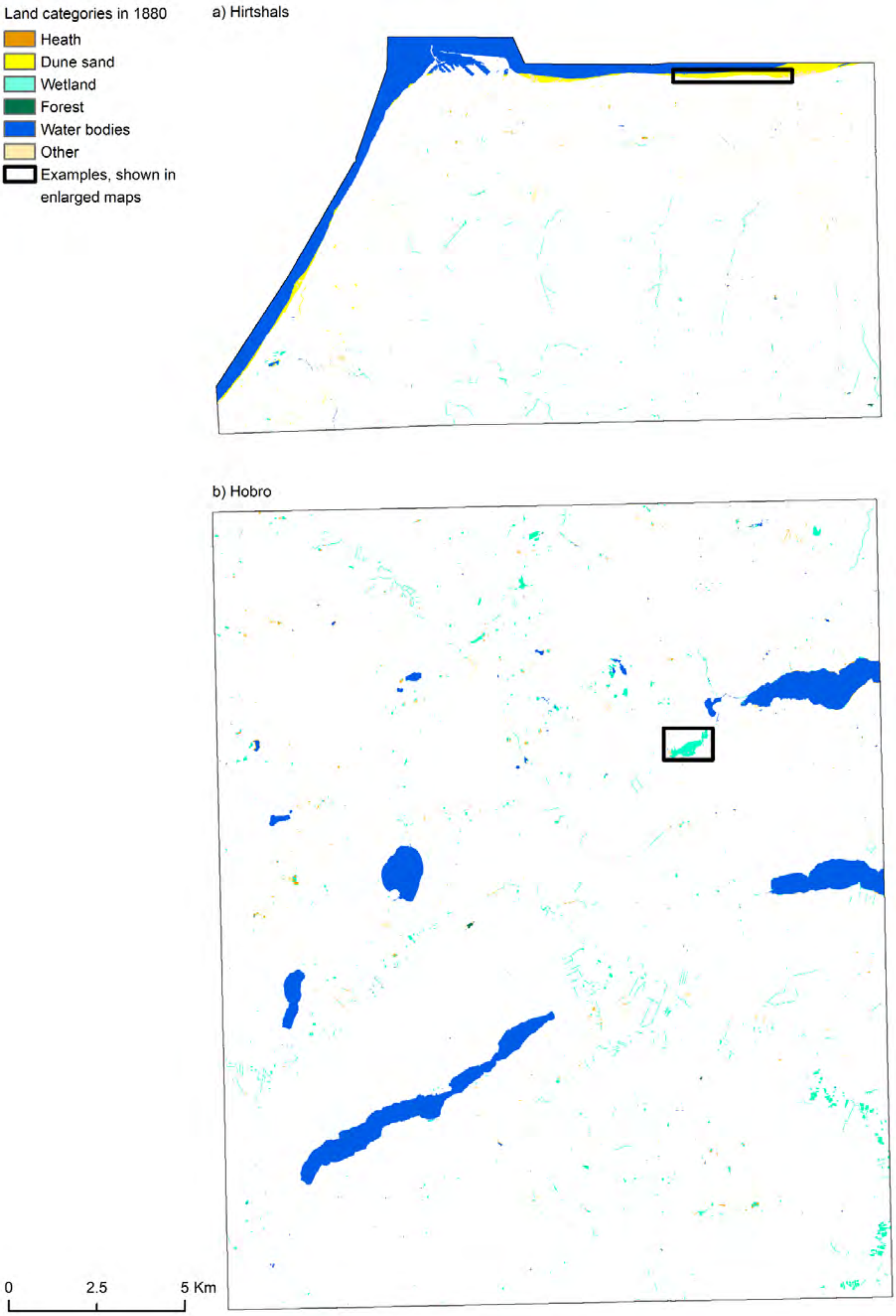
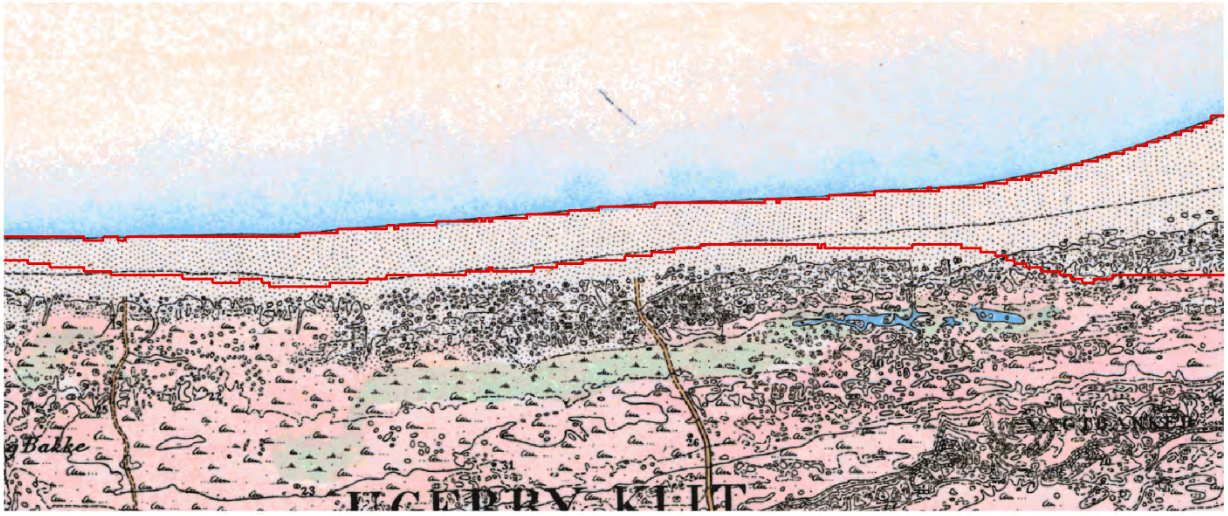


Figure 5.32 1880 land categories of 2018 water bodies.

a) 1880 HMB map



b) 2018 aerial image

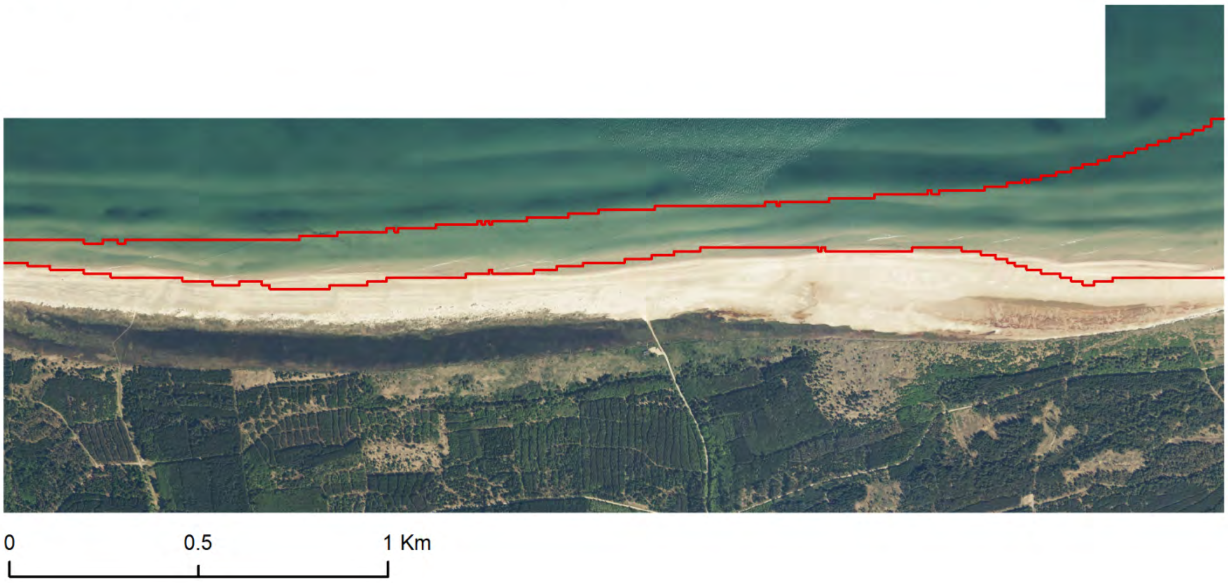


Figure 5.33 Example from the Hirtshals area, mapped as dune sand in 1880 and as water body in 2018.

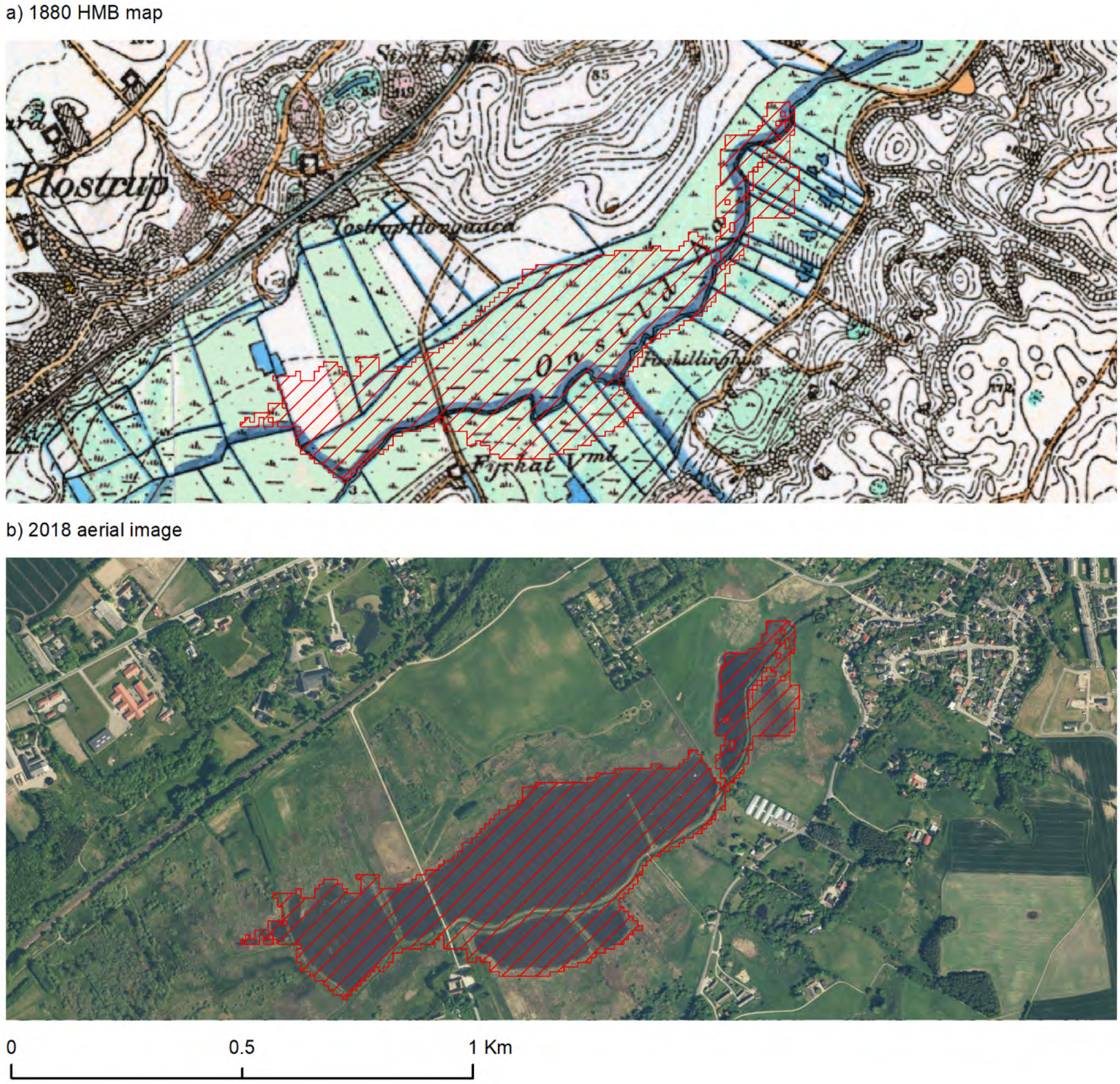


Figure 5.34 Example from the Hobro area, mapped as wetland in 1880 and as water body in 2018.

5.3 Discussion

For both study areas, the results of our analysis reveal overall LULC dynamics, which for the period from 1880 to 2018 are characterised by a decrease in open, non-forested land categories, such as heath, dune sand and wetlands at the gain of afforestation and an increasing area of agricultural land. Our results also point at differences between the two areas. While in the Hirtshals area, the loss of open, non-forested land categories is mainly the consequence of afforestation, in the Hobro area, it is the consequence of a change to agricultural land use. For the Hirtshals area, our results indicate changes from dune sand to land categories dominated by grass and herb vegetation. This change could be the consequence of efforts to prevent sand drift. However, most probably, part of this change must be subscribed to the differences between the mapping concepts of the HMB maps and more recent mapping with a shift from mapping of the land for mainly military interests to a broader purpose base, specifically including nature conservation and more generally inventory motives (Svenningsen, 2016b; Svenningsen et al., 2015). This illustrates the need for thorough studies of categorisation in historical as well as contemporary maps.

Nevertheless, the two study areas show clear evidence of a general intensification of the landscape with change from open non-forested land categories to agriculture and forest, which are in line with comparable studies. However, the results also reveal some major differences in LULC dynamics between the two areas. In the Hobro area, open non-forested habitat types generally changed to agriculture, while in the Hirtshals area these categories generally changed to forest. This difference probably relates to soil fertility e.g. the potential for conversion of open non-forested categories to agriculture and the planting of forest to mitigate sand drifting. Yet, while these differences might not be surprising, such dynamics have not yet been investigated based on spatial data as countrywide studies are based on statistics. In addition, the few spatial specific studies of LULC dynamics in Denmark have primarily been conducted for parts of central Jutland (Caspersen & Fritzbøger, 2002; Jensen & Jensen, 1977; Jensen & Jensen, 1979; Kristensen et al., 2009; Münier, 2009) and a few studies have been conducted for eastern and western Jutland, Fyn and Zealand (Münier, 2009; Primdahl et al., 2010; Stenak, 2005). This shows, a potential for investigating local and regional LULC dynamics based on a countrywide dataset of HMB data. Such a countrywide dataset could also provide comprehensive machine-readable spatially specific information about persistent open non-forested categories, which are of importance for contemporary nature and landscape management. However, additional analyses might be needed to ensure that a site has been unchanged throughout the period.

6 Discussion and conclusion

6.1 Perspectives for extraction of additional land categories and landscape elements

In this project, we have focused on the automated production of digital geodata for the land categories heath, dune sand, wetland, forest and water bodies. We choose these categories in order to be able to map dynamics between major habitat types. Due to limited resources, we decided not to include categories, such as buildings and associated land or roads. Since urbanisation is one of the most important historical and ongoing landscape changes (Levin & Normander, 2008), future method development should also include urban land categories and roads. Furthermore, we also did not include line elements, such as streams and ditches, which would be relevant to future method development.

6.2 Perspectives for a countrywide extraction of land categories for the late 1800s

As it has been for the HMB map sheet sections in the test areas, it should be that the same rule-set set of procedures can be applied to map the same target categories for other parts of the HMB coverage. The test areas were selected to include sizable extents of the target categories, in a range of landscape contexts (e.g. both off and on steeply sloping ground) and to include a good representation of the range of category colourisations observed more widely, in SD data, across Denmark.

It would have to be recognised that the HMB map sheet category colourations do vary, in particular for wetland and heath. Therefore, initial visual inspection and assessment of each map sheet, in terms of the similarity of its category colourations to the standard colourations is recommended. In some cases this inspection will need to include close visual inspection of the derived image layers that are used by the standard routines in order to assess that those layer(s) are appropriate to use for a mapping. If the inspection indicates that they have poor discriminatory power, one of the alternative derived image layers will need to be chosen instead. In extreme cases, an alternative processing block for a category may need to be chosen, and possibly modified.

In the test areas, it has in general been beneficial to use derived image layers based on the KB archive HMB map sheets. That has been related either to the more consistent colouration of the KB maps and / or the finer resolution of their scanning. However, in some cases in this study, the colouration of the KB sheets has been markedly pale (e.g. for heath), requiring alternative use of derived image layers based on the SD image data. For much of Zeeland, and parts of Funen, the KB map sheets are only 3-colour, which markedly limits their use there in the same ways as they have been used for our study areas. Including another study area on Zeeland or Funen would have increased the general applicability of our study for the whole country. But this was beyond our resources. We have, however, noted that the SD image data of Zeeland and Funen are of good quality throughout, being more colour consistent than the SD image data of the test areas. It is therefore expected that the same procedures as applied for the test areas can be applied where the KB 3-colour maps are absent, albeit with some small changes to the rule set. The 3-colour

HMB maps, with just water bodies using a blue print or hand-tint component, represent a potentially strong basis for the mapping of all water associated HMB map features, including ditches.

The benefit for this project of complementarity between the SD and the KB HMB image data was noted earlier. Similar complementarity possibilities may be fewer for parts with just the SD set of full-colour HMB data, but as the SD data of eastern Denmark are in general of good quality, it is expected that good quality automated geo-data can be produced also from a single source data set for those parts. The possibility of KB having a full set of full-colour sheets is being investigated. An additional source would be the “Mønsterblade” from SDFE (digital) and National Archive (physical). The SDFE seamless dataset was made from Mønsterblade (Generalstabens Topografiske Afdeling, 1910), but a new scanning of the Mønsterblade would provide higher resolution data.

6.3 The potential to apply automated extraction to other historical topographical maps

Following completion of the HMB map series surveys in 1899, survey and cartography of a new series, the Lave Målebordsblade (LMB) began in 1901, continuing through to 1933. The map sheets of the LMB series are each 1½ times as broad as the HMB sheets and each sheet covers 71 km2, rather than the 47 km2 of the HMB map sheets. The HMB and LMB survey protocols were similar, but the LMB cartography was only ever published as three-colour prints, with blue used for the sea, lakes, rivers, etc, brown for height isolines and black for all other lines, marks and text. The LMB were not published with hand-tinting as was applied to the HMB sheets.

The full colour colouration of the HMB sheets has been the most significant aspect of those maps to enable automated digitisation of four of the five target categories. In the three-colour LMB maps (as for the three-colour HMB maps sheets of Funen and Zeeland), water bodies alone could in principle be automatically digitised with similar colour segmentation processing as has been applied in this study. Dune sand is marked in the same way in the LMB maps as in the HMB, but is distinguished from beaches, which are characterised by a different dot density. For other categories, such as the other three categories digitised in this study, processing would need to make far greater use of the associated LMB legend symbols for the categories. That would imply there being a greater reliance on the associated CNN or other machine learning methods to make the symbols with very low levels of both false-positives and false-negative. In general, the graphical quality of the LMB maps is high, with legend symbols that are well defined and uniform. Except for situations where the symbols are obscured by other black map components that should be possible by use of more symbol training samples and model parameterisations. Since height isolines are in brown, finding symbols of forest and heath on steeply sloping ground should also be possible. If all symbols of a legend type can be detected with high accuracies, automated definition of the areal unit of a category could be made by bounded growth. Boundaries to growth would need to include the presence of symbols of other types, natural limits such as rivers and roads, and LMB map sheet indications of legend item limits such as dotted lines.

Automated digitisation of categories in the LMB maps could also be done with semantic and geographic integration to the products of the automated digitisation of the HMB maps. For example, that areas mapped to a category in the HMB map would receive a weighting that would increase their interpretation to the same category in automated digitisation of the LMB maps. There would need to be sound semantic grounds for such weighting, such as that areas surveyed as forests in the early years of the LMB surveys would have probably have been forest, and been surveyed as such, also in the later years of the HMB surveys.

The production history of the LMB is however, more complex compared to the HMB, as the LMB sheets was produced in several revised versions between 1901 and 1970s. In some urban areas, more than eight different versions have been published, and in some versions, only the road network was updated. This means that maps sheets to be included in such a project should be selected carefully, in order to include sheets were the land categories and other landscape features have been updated.

Whereas the coverage of the HMB map series is the extent of Denmark between 1864 and 1920, the LMB map series covers the full modern extent of Denmark. For the decades when the HMB maps were surveyed, the remainder of the modern state of Denmark was mapped by the Preussiske Maaleboldsblade (PMB) map series. The published printed editions of the PMB series use just blue, for sea, lakes, rivers, etc., and automated digitisation of these components should be possible with the same forms of processing as used for the HMB maps. All other lines, marks and texts are in black. Legend item symbols are similar, but not identical, to those of the HMB and LMB map series. The graphical quality of the PMB maps is high, so symbol CNN model combined with OBIA methods should enable digitisation of legend category areal units. As the PMB maps were made under commission of the Prussian authorities that administered southern Jutland at that time, the PMB maps have much in common with other late 1800s maps of parts of modern Germany and several other European lands. Thus, methods for effective automated digitization of the PMB maps could also be applied far more widely.

References

Agency for Data Supply and Efficiency, 2019: Høje Målebordsblade. Copenhagen, Agency for Data Supply and Efficiency. Assessed May 19, 2019 at https://www.kortforsyningen.dk/

Auffret, A.G., Kimberley, A., Plue, J., Skånes, H., Jakobsson, S., Waldén, E., Wennbom, M., Wood, H., Bullock, J.M., Cousins, S.A.O., Gartz, M., Hooftman, D.A.P & Tränk, L., 2017: HistMapR: Rapid digitization of historical land-use maps in R. Methods in Ecology and Evolution 8, 1453–1457.

Baily, B., 2007: The extraction of digital vector data from historic land use maps of Great Britain using image processing techniques. e-Perimetron, 2, 209-223.

Baily, B., Riley, M., Aucott, P., Southall, H., 2011: Extracting digital data from the First Land Utilisation Survey of Great Britain – Methods, issues and potential. Applied Geography 31, 959-968.

Brandt, J., 2004: Anvendelse af topografiske kort til monitering af biotoper. Geoforum Perspektiv, 5, 53–61.

Caspersen, O.H., Fritzbøger, B., 2002: Long Term Landscape Dynamics - a 300-Years, Case Study from Denmark. Danish Journal of Geography, Special Issue 3, 13–27.

Chiang, Y.-Y., Knoblock, C.B., 2006: Classification of line and character pixels on raster maps using discrete cosine transformation coefficients and support vector machine. In Proceedings of the 18th International Conference on Pattern Recognition. 1034–1037.

Chiang, Y.-Y., Leyk, S., Nazari, N.H., Moghaddam, S., Tan, T.X., 2016: Assessing the impact of graphical quality on automatic text recognition in digital maps. Computers & Geosciences 93, 21–35.

Cordella, L.P., Vento, M., 2000: Symbol and shape recognition. In Proceedings of the 3rd International Workshop on Graphics Recognition, Recent Advances (GREC'99). Springer-Verlag, London, UK, 167–182.

Dam, P., 2005: Videnskabernes Selskabs kort 1768-1805 - En introduktion til såvel analog som digital brug. HisKIS' Digitale Årsskrift 2003-2005, 36–44.

Dam, P., Jakobsen, J.G.G., 2008: Historisk-Geografisk Atlas. Copenhagen, Geografforlaget, Det Kongelige Danske Geografiske Selskab.

Dhar, D.B., Chanda, B., 2006: Extraction and recognition of geographical features from paper maps. International Journal of Document Analysis and Recognition 8, 4, 232–245.

Digitalt atlas over Danmarks historisk-administrative geografi, 2019: Sognegrænser, Digitalt atlas over Danmarks historisk-administrative geografi. Assessed March 8, 2019 at http://www.digdag.dk/

Fortner, B., T.E.Meyer, 1997: Number by colors: a guide to using color to understand technical data. New York, Springer Science+Business Media.

Freeman H, Pieroni, G.G., 1982: Map Data Processing. London, Academic Press.

Fuchs, R., Schulp, C.J.E., Hengeveld, G.M., Verburg, P.H., Clevers, J.G.P.W., Schelhaas, M-J., Herold, M., 2016: Assessing the influence of historic net and gross land changes on the carbon fluxes of Europe. Global Change Biology 22, 2526–2539

Fuchs, R., Verburg, P.H., Clevers, b, J.G.P.W., Herold, M., 2015: The potential of old maps and encyclopaedias for reconstructing historic European land cover/use change. Applied Geography 59, 43-55.

Generalstabens Topografiske Afdeling, 1877: Ledetraad i Detailmaalingen. trykt hos C. L. E. Hansen.

Generalstabens Topografiske Afdeling, 1910: Farvemønsterblade, Danmark. Rigsarkivet: Generalstabens Topografiske Afdeling, Farvemønsterblade, Danmark (Officielt), pk 1-16.

Gobbi, S., Ciolli, M., La Porta, N., Rocchini, D., Tattoni, C., Zatelli, P., 2019: New Tools for the Classification and Filtering of Historical Maps. ISPRS Int. J. Geo-Inf., 8, 455-.

Godfrey, B., Eveleth, H., 2015: An Adaptable Approach for Generating Vector Features from Scanned Historical Thematic Maps Using Image Enhancement and Remote Sensing Techniques in a Geographic Information System. Journal of Map & Geography Libraries, 11, 18-36.

Groom, G., Levin, G., Svenningsen, S., Perner, M.L., 2020: Historical Maps - Machine learning helps us over the map vectorisation crux. Proceedings of the ICA Workshop on Automatic Vectorisation of Historical Maps, pp. 89-98. Budapest, 2020.
http://lazarus.elte.hu/avhm/AVHM_Proceedings.pdf#page=89

Helkiær Jensen, R., Jensen K.M., 1979: Kulturlandskabet i Borris og Sdr. Felding — En kortbladsanalyse af et Vestjysk landbrugssamfund og en dokumentation for dets udvikling. Danish Journal of Geography 78–79, 61–99.

Jensen, K.M., Jensen, R.H., 1977: En landbrugsgeografisk analyse af Gadbjerg og Givskud sogne. Danish Journal of Geography 76, 35–51.

Kaim, D., Kozak, J., Kolecka, N., Ziólkowska, E., Ostafin, K., Ostapowicz, K., Gimmi, U., Munteanu, C., Radeloff, V.C., 2016: Broad scale forest cover reconstruction from historical topographic maps. Applied Geography 67, 39-48.

Käyhkö, N. & Skånes, H., 2006: Change trajectories and key biotopes—Assessing landscape dynamics and sustainability. Landscape and Urban Planning 75, 300–321.

Kent, A.J., & Vujakovic, P., 2009: Stylistic Diversity in European State 1 : 50 000 Topographic Maps. Cartographic Journal 46, 179–213.

Kienast, F., 1993: Analysis of historic landscape patterns with a geographical information system – a methodological outline. Landsc. Ecol 8, 103–118.

Kim, N.W., Lee, J., Lee, H., Seo, J., 2014: Accurate segmentation of land regions in historical cadastral maps. J. Vis. Commun. Image R. 25, 1262–1274.

Konkoly-Gyuró, E., Balázs, P., Tirászi, Á., 2019: Transdisciplinary approach of transboundary landscape studies: a case study of an Austro-Hungarian transboundary landscape, Danish Journal of Geography 1, 52-68.

Korsgaard, P., 2006: Kort som kilde: En håndbog om historiske kort og deres anvendelse. Dansk Historisk Fællesråd: Sammenslutningen af Lokalarkiver.

Kristensen, S.B.P., Reenberg, A., Peña, J.J.D., 2009: Exploring Local Rural Lands-cape Changes in Denmark: A Human-Environmental Timeline Perspective. Danish Journal of Geography 109, 47–67.

Laumer, D., H. Gümgümcü, M.Heitzler, L.Hurni, 2020: A Semi-automatic Label Digitization Workflow for the Siegfried map. In, International Workshop on Automatic Vectorisation of Historical Maps - 13 March 2020 - ELTE, Budapest https://doi.org/10.21862/avhm2020.02

le Riche, M., 2020: Identifying Building Footprints in Historic Map Data using OpenCv and PostGIS. In, International Workshop on Automatic Vectorisation of Historical Maps - 13 March 2020 - ELTE, Budapest
https://doi.org/10.21862/avhm2020.02

Levin, G., 2019: Basemap03. Technical documentation of the method for elaboration of a land-use and land-cover map for Denmark. Aarhus University, DCE – Danish Centre for Environment and Energy, 86 pp. Technical Report No. 159.

Levin, G., Normander, B., 2008: Arealanvendelse i Danmark siden slutningen af 1800-tallet. Aarhus University, National Environmental Research Institute, (Technical Report from NERI – National Environmental Research Institute; No. 682), 46 pp.

Leyk, S., 2010: Segmentation of colour layers in historical maps based on hierarchical colour sampling. In Graphics Recognition. Achievements, Challenges, and Evolution, Ogier, J.-M., Liu, W., Lladós, J., Eds., Lecture Notes in Computer Science, Vol. 6020. Springer, Berlin, 231–241.

Leyk, S., Boesch, R., 2009: Extracting Composite Cartographic Area Features in Low-Quality Maps, Cartography and Geographic Information Science 36, 71-79.

Liu, D., Toman, E., Fuller, Z., Chen, G., Londo, A., Zhang, X., Zhao, K., 2018: Integration of historical map and aerial imagery to characterize long-term land-use change and landscape dynamics: An object-based analysis via Random Forests. Ecological Indicators 95, 596-605.

Liu, T., Xu, P., Zhang, S., 2019: A review of recent advances in scanned topographic map processing. Neurocomputing 328, 5–87.

Lladós, J, Valveny, E., Sanchez, G. Marti, E., 2002: Symbol recognition: Current advances and perspectives. In Graphics Recognition Algorithms and Applications, D. Blostein and Y.-B. Kwon, Eds., Lecture Notes in Computer Science, Vol. 2390. Springer, Berlin, 104–128.

Marciano, R.J., Allen, R.C., Hou, C.Y., Lach, P.R., 2013: Big Historical Data Feature Extraction, Journal of Map & Geography Libraries 9, 69-80.

Medyńska-Gulij, B., Żuchowski, T., 2018: European topography in eighteenth-century manuscript maps (Wyd. 1). Bogucki Wydawn. Naukowe.

Michaelsen, P., 2004: Målebordsblade og 4 cm kort. Bordmåling, rekognoscering og instrukser for arbejdet. Geoforum Perspektiv 3, 5.

Mikusiska, A., Zawadzka, B., Samojlik, T., Jedrzejewska, B.J., Mikusinsj, G., 2013: Quantifying landscape change during the last two centuries in Białowieza Primeval Forest. Applied Vegetation Science 16, 217–226.

Münier, B. 2009: Landskabets udvikling siden midten af 1800-tallet. In B. Odgaard & J. R. Rømer (Eds.), Danske landbrugslandskaber gennem 2000 år: fra digevoldinger til støtteordninger (pp. 81–98). Aarhus Universitetsforlag.

Nilsson, B., Wiese, M.B., Tougaard, L., Højberg, A., Troldborg, L., Greve, M.H., Greve, M.B., Tind, S.L., 2014: Forstudium til vurdering af udbredelse og påvirkning af grundvandsafhængige terrestriske økosystemer i det udrænede og drænede landskab. Rapport er udarbejdet af GEUS og Aarhus Universitet for Naturstyrelsen.

Odgaard, B. & Rømer, J.R., 2009: Danske landbrugslandskaber gennem 2000 år: Fra digevoldninger til støtteordninger. Århus Universitetsforlag.

Olsen, O.N., 1831: Topographisk Kaartsignaturer. den kongelige danske Generalqvarteermester-Stab.

Ostafin, K., Iwanowski, M., Kozak, J., Cacko, A., Gimmi, U., Kaim, D., Psomas, A., Ginzler, C. & Ostapowicz, K., 2017: Forest cover mask from historical topographic maps based on image processing. Geoscience Data Journal 4, 29-39.

Pezeshk, A. & Tutwiler, R.L., 2011: Automatic Feature Extraction and Text Recognition From Scanned Topographic Maps. IEEE Transactions on Geoscience and Remote Sensing 49, 5047-5063.

Picuno, P., Cillis, G. & Statuto, D., 2019: Investigating the time evolution of a rural landscape: How historical maps may provide environmental information when processed using a GIS. Ecological Engineering 139, 105580.

Primdahl, J., Kristensen, L., Busck, A.G., & Vejre, H., 2010: Functional and structural changes of agricultural landscapes: How Changes are conceived by local farmers in two Danish rural communities. Landscape Research 35, 633–653.

Schaffer, G., Levin, N., 2014: Mapping Human Induced Landscape Changes in Israel Between the end of the 19Th Century and the Beginning of the 21Th Century. Journal of Landscape Ecology, 7, 110–145.

Steinmann, 1832: Instruction til den militaire Landmaaling ved Generalquartermester-Staben (Rigsarkivet: 0201-022 Generalstabens Taktiske Afdeling, Krigsføringsdepotet A. Sagsakter 1780-1943. VI Topografi Pk. 144, læg 1c Instruction til den militaire Landmaaling ved Generalquartermester-staben; Tillæg til Instruktion og betingelse af 23de Febr. 1830 for de Officerer af Armeen, som attraae at bruges ansættes til militair Landmaaling ved Generalquartermester- Staben, 1832). Den Kongelige Generalquartermester-Stab.

Stenak, M., 2005: De inddæmmede landskaber: en historisk geografi. Landbohistorisk Selskab.

Svenningsen, S.R., 2014: Krigen og landskabet – 250 års dansk militærgeografi. Fra Krig Og Fred -Dansk Militærhistorisk Kommissions Tidsskrift 1, 11–58.

Svenningsen, S.R., 2016a: Mapping the Nation for War: The Landscape in Danish Military Cartography 1800–2000. Imago Mundi 68, 196–201.

Svenningsen, S.R., 2016b: The emergence of an environmental cartography in Denmark. The Cartographic Journal 53, 101-113.

Svenningsen, S.R., Levin, G. Perner, M.L., 2019: Military land use and the impact on landscape: A study of land use history on Danish Defence sites. Land Use Policy 84, 114-126.

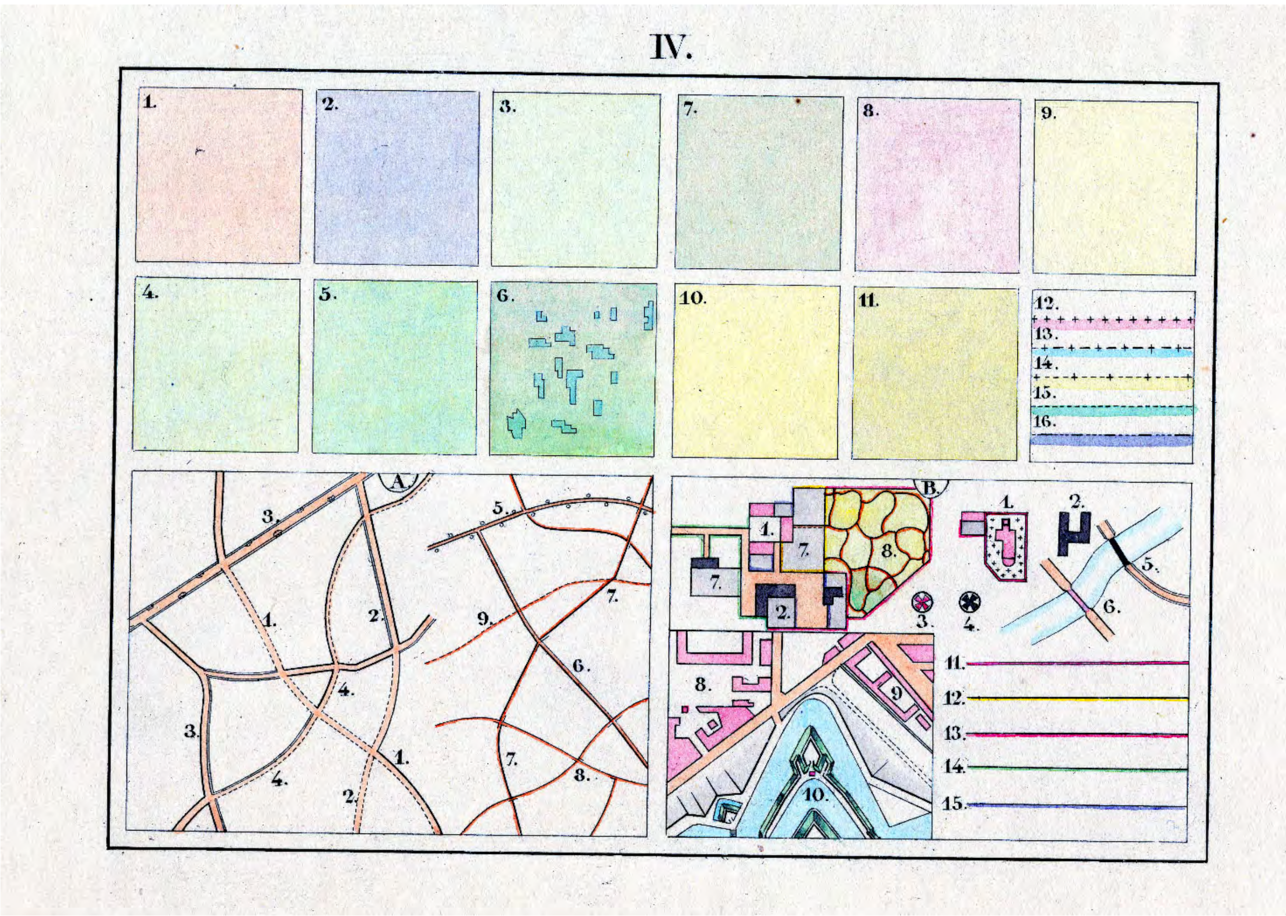
Svenningsen, S.R., Levin, G., Jepsen, M.R., 2015: Decrease in Danish semi-natural grassland – a social construct or a real-world change? Danish Journal of Geography 115, 157-166.

Uhl, J.H., Leyk, S., Chiang, Y.-Y., Duan, W., Knoblock, C.A., 2018: Spatialising uncertainty in image segmentation using weakly supervised convolutional neural networks: a case study from historical map processing. IET Image Processing 12, 2084-2091.

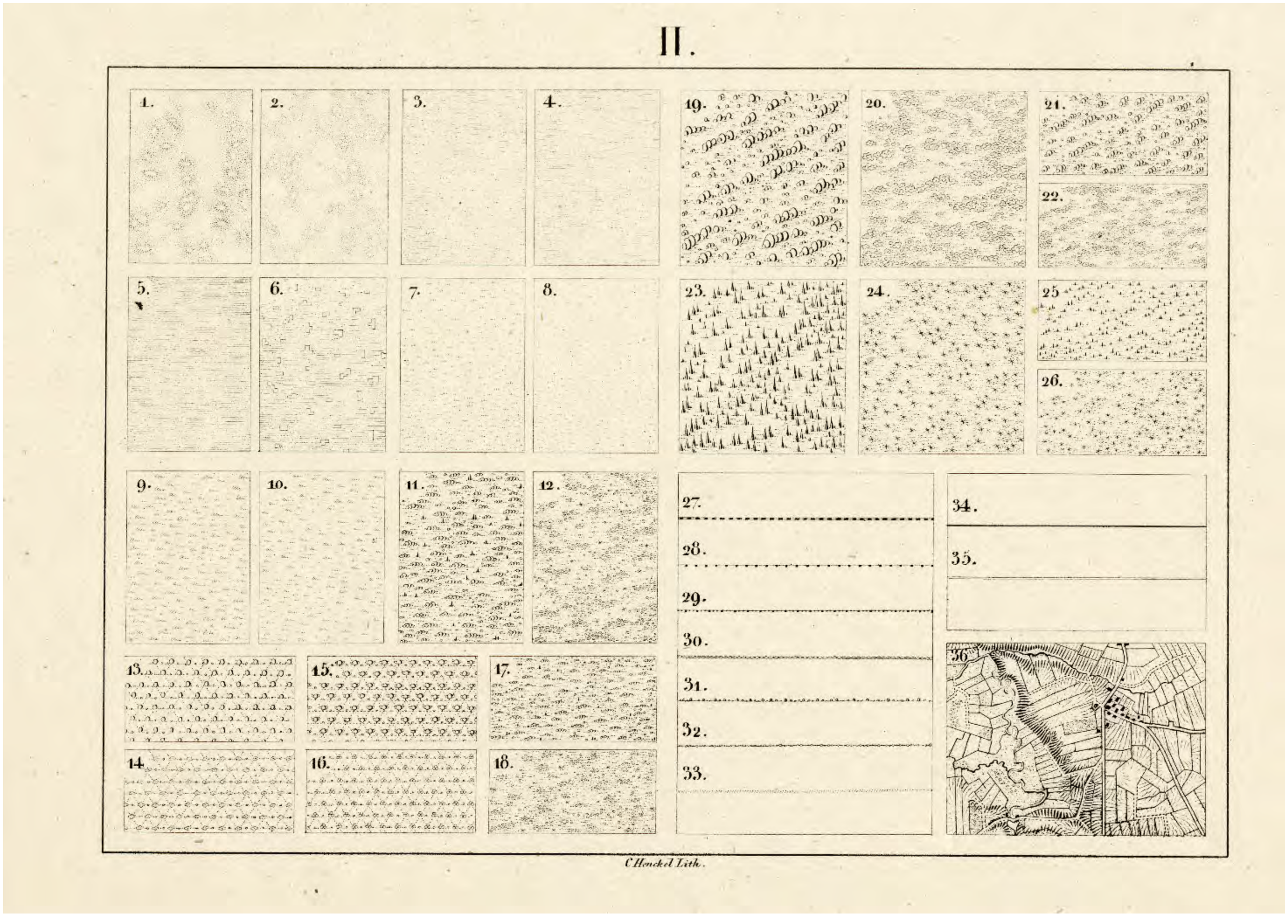
Van den Berghe, H., Gheyle, N., Note, W., Stichelbaut, B., Van Meirvenne, M., Bourgeois, J., Van Eetvelde, V., 2019: Revealing the preservation of First World War shell hole landscapes based on a landscape change study and LiDAR, Danish Journal of Geography 1, 38-51.

Xu, B., Chen. J., Yao, M., 2016: Identification of Contour Lines from Average-Quality Scanned Topographic Maps. Mathematical Problems in Engineering 2016, 14 p.

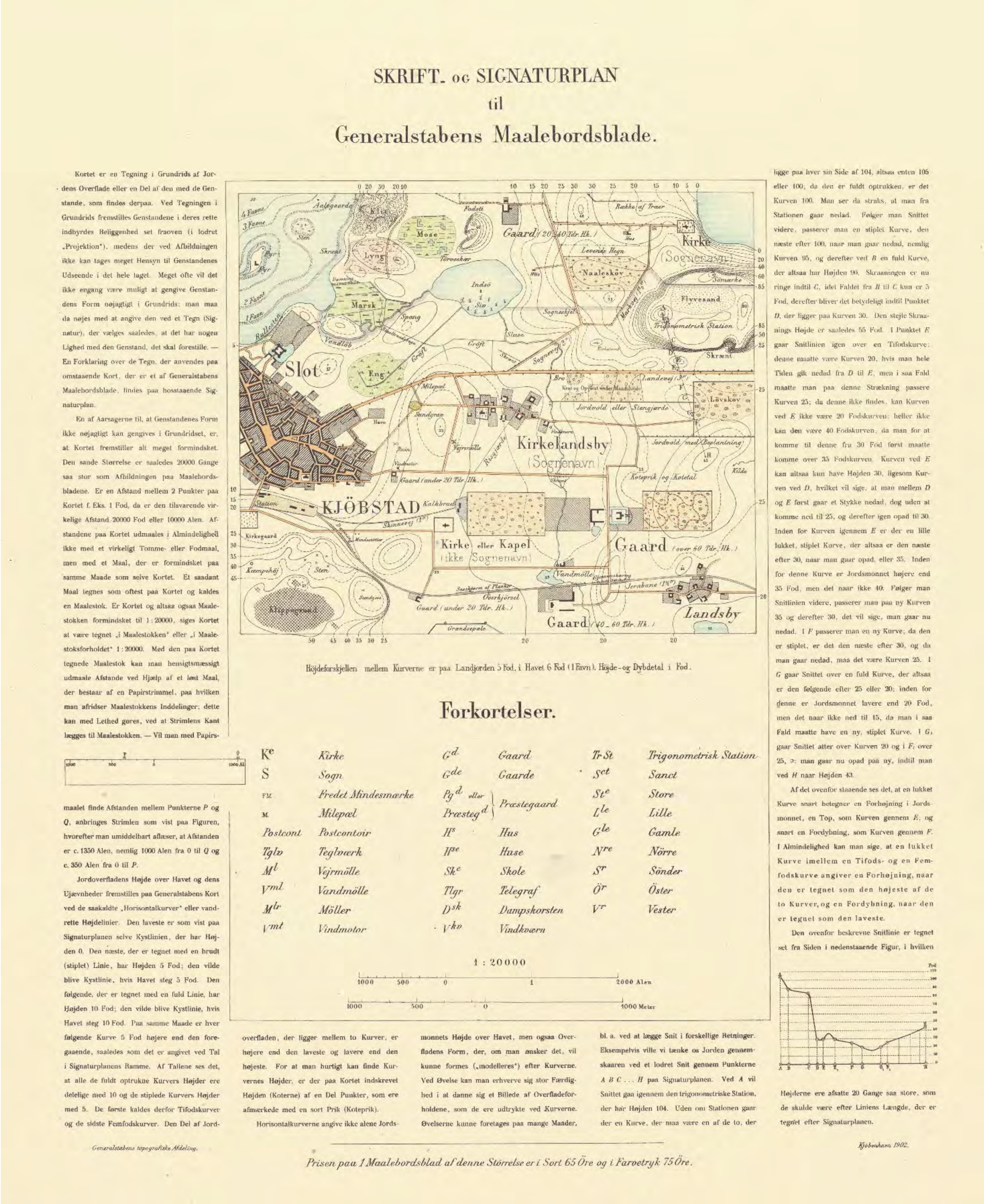
Appendix



Appendix 1 Colour for categories by O.N. Olsen (1831). The colours plate to be used on large-scale topographic maps. Colours adjusted from original, as the original colours were fainted. Numbers on corresponds to the following categories 1: Sand, 2: Marsh, 3: soft soil (can be crossed by horsemen), 4: Soft soil (can be crossed by men on foot), 5: soft soil (can only be cross with difficulties or with assistance), 6: peat cuts; 7: grass, 8: heather, 9: scrubland, 10: Deciduous trees, 11: Coniferous trees.



Appendix.2 Appendices Symbols for categories by O.N. Olsen (1831). Figure AS2. Signatures for large-scale topographic maps. 1: Sand, 2: sand, 3: soft soil can be crossed by horsemen, 4: soft soil, can be crossed on foot, 5: Soft soil, can only be crossed with difficulties or assistance, 6: peat cuts, 7: grass, 8: grass, 9: heather, 10: heather, 11: scrubland (perspective), 12: scrubland (plan), 13: Plantation, 14: plantation, 15: orchard, 16: orchard, 17: scrubs, 18: scrubs, 19: Deciduous trees, 20: Deciduous trees, 21: Deciduous trees, 22: Deciduous trees, 23: Coniferous trees, 24: Coniferous trees, 25: Coniferous trees, 26: Coniferous trees.



SKRIFT- og SIGNATURPLAN
til
Generalstabens Maalebordsblade.

Kortet er en Tegning i Grundrids af Jordens Overflade eller en Del af den med de Genstande, som findes derpaa. Ved Tegningen i Grundrids fremstilles Genstandene i deres rette indbyrdes Beliggenhed set fraoven (i lodret „Projektion"), medens der ved Afbildningen ikke kan tages meget Hensyn til Genstandenes Udseende i det hele taget. Meget ofte vil det ikke engang være muligt at gengive Genstandens Form nøjagtigt i Grundrids; man maa da nøjes med at angive den ved et Tegn (Signatur), der vælges saaledes, at det har nogen Lighed med den Genstand, det skal forestille. — En Forklaring over de Tegn, der anvendes paa omstaaende Kort, der er et af Generalstabens Maalebordsblade, findes paa hosstaaende Signaturplan.

En af Aarsagerne til, at Genstandenes Form ikke nøjagtigt kan gengives i Grundridset, er, at Kortet fremstiller alt meget formindsket. Den sande Størrelse er saaledes 20000 Gange saa stor som Afbildningen paa Maalebordsbladene. Er en Afstand mellem 2 Punkter paa Kortet f. Eks. 1 Fod, da er den tilsvarende virkelige Afstand 20000 Fod eller 10000 Alen. Afstandene paa Kortet udmaales i Almindelighed ikke med et virkeligt Tomme- eller Fodmaal, men med et Maal, der er formindsket paa samme Maade som selve Kortet. Et saadant Maal tegnes som oftest paa Kortet og kaldes en Maalestok. Er Kortet og altsaa ogsaa Maalestokken formindsket til 1 : 20000, siges Kortet at være tegnet „i Maalestokken" eller „i Maalestoksforholdet" 1 : 20000. Med den paa Kortet tegnede Maalestok kan man hensigtsmæssigt udmaale Afstande ved Hjælp af et løst Maal, der bestaar af en Papirstrimmel, paa hvilken man afridser Maalestokkens Inddelinger; dette kan med Lethed gøres, ved at Strimlens Kant lægges til Maalestokken. — Vil man med Papirsmaalet finde Afstanden mellem Punkterne *P* og *Q*, anbringes Strimlen som vist paa Figuren, hvorefter man umiddelbart aflæser, at Afstanden er c. 1350 Alen, nemlig 1000 Alen fra 0 til *Q* og c. 350 Alen fra 0 til *P*.

Jordoverfladens Højde over Havet og dens Ujævnheder fremstilles paa Generalstabens Kort ved de saakaldte „Horisontalkurver" eller vandrette Højdelinier. Den laveste er som vist paa Signaturplanen selve Kystlinien, der har Højden 0. Den næste, der er tegnet med en brudt (stiplet) Linie, har Højden 5 Fod; den vilde blive Kystlinie, hvis Havet steg 5 Fod. Den følgende, der er tegnet med en fuld Linie, har Højden 10 Fod; den vilde blive Kystlinie, hvis Havet steg 10 Fod. Paa samme Maade er hver følgende Kurve 5 Fod højere end den foregaaende, saaledes som det er angivet ved Tal i Signaturplanens Ramme. Af Tallene ses det, at alle de fuldt optrukne Kurvers Højder ere delelige med 10 og de stiplede Kurvers Højder med 5. De første kaldes derfor Tifodskurver og de sidste Femfodskurver. Den Del af Jordoverfladen, der ligger mellem to Kurver, er højere end den laveste og lavere end den højeste. For at man hurtigt kan finde Kurvernes Højder, er der paa Kortet indskrevet Højden (Koterne) af en Del Punkter, som ere afmærkede med en sort Prik (Koteprik).

Horisontalkurverne angive ikke alene Jordsmonnets Højde over Havet, men ogsaa Overfladens Form, der, om man ønsker det, vil kunne formes („modelleres") efter Kurverne. Ved Øvelse kan man erhverve sig stor Færdighed i at danne sig et Billede af Overfladeforholdene, som de ere udtrykte ved Kurverne. Øvelserne kunne foretages paa mange Maader, bl. a. ved at lægge Snit i forskellige Retninger. Eksempelvis ville vi tænke os Jorden gennemskaaren ved et lodret Snit gennem Punkterne *A B C . . . H* paa Signaturplanen. Ved *A* vil Snittet gaa igennem den trigonometriske Station, der har Højden 104. Uden om Stationen gaar der en Kurve, der maa være en af de to, der ligge paa hver sin Side af 104, altsaa enten 105 eller 100; da den er fuldt optrukken, er det Kurven 100. Man ser da straks, at man fra Stationen gaar nedad. Følger man Snittet videre, passerer man en stiplet Kurve, den næste efter 100, naar man gaar nedad, nemlig Kurven 95, og derefter ved *B* en fuld Kurve, der altsaa har Højden 90. Skraaningen er nu ringe indtil *C*, idet Faldet fra *B* til *C* kun er 5 Fod, derefter bliver det betydeligt indtil Punktet *D*, der ligger paa Kurven 30. Den stejle Skraanings Højde er saaledes 55 Fod. I Punktet *E* gaar Snitlinien igen over en Tifodskurve; denne maatte være Kurven 20, hvis man hele Tiden gik nedad fra *D* til *E*, men i saa Fald maatte man paa denne Strækning passere Kurven 25; da denne ikke findes, kan Kurven ved *E* ikke være 20 Fodskurven; heller ikke kan den være 40 Fodskurven, da man for at komme til denne fra 30 Fod først maatte komme over 35 Fodskurven. Kurven ved *E* kan altsaa kun have Højden 30, ligesom Kurven ved *D*, hvilket vil sige, at man mellem *D* og *E* først gaar et Stykke nedad, dog uden at komme ned til 25, og derefter igen opad til 30. Inden for Kurven igennem *E* er der en lille lukket, stiplet Kurve, der altsaa er den næste efter 30, naar man gaar opad, eller 35. Inden for denne Kurve er Jordsmonnet højere end 35 Fod, men det naar ikke 40. Følger man Snitlinien videre, passerer man paa ny Kurven 35 og derefter 30, det vil sige, man gaar nu nedad. I *F* passerer man en ny Kurve; da den er stiplet, er det den næste efter 30, og da man gaar nedad, maa det være Kurven 25. I *G* gaar Snittet over en fuld Kurve, der altsaa er den følgende efter 25 eller 20; inden for denne er Jordsmonnet lavere end 20 Fod, men det naar ikke ned til 15, da man i saa Fald maatte have en ny, stiplet Kurve. I G_1 gaar Snittet atter over Kurven 20 og i F_1 over 25, ɔ: man gaar nu opad paa ny, indtil man ved *H* naar Højden 43.

Af det ovenfor staaende ses det, at en lukket Kurve snart betegner en Forhøjning i Jordsmonnet, en Top, som Kurven gennem *E*, og snart en Fordybning, som Kurven gennem *F*. I Almindelighed kan man sige, at en lukket Kurve imellem en Tifods- og en Femfodskurve angiver en Forhøjning, naar den er tegnet som den højeste af de to Kurver, og en Fordybning, naar den er tegnet som den laveste.

Den ovenfor beskrevne Snitlinie er tegnet set fra Siden i nedenstaaende Figur, i hvilken Højderne ere afsatte 20 Gange saa store, som de skulde være efter Liniens Længde, der er tegnet efter Signaturplanen.

Højdeforskjellen mellem Kurverne er paa Landjorden 5 Fod, i Havet 6 Fod (1 Favn). Højde- og Dybdetal i Fod.

Forkortelser.

Ke	Kirke	Gd	Gaard	Tr St	Trigonometrisk Station
S	Sogn	Gde	Gaarde	Sct	Sanct
FM	Fredet Mindesmærke	Pgd eller Præstegd	Præstegaard	Ste	Store
M	Milepæl			Lle	Lille
Postcont	Postcontoir	Hs	Hus	Gle	Gamle
Tglv	Teglværk	Hse	Huse	Nre	Nörre
Ml	Vejrmölle	Ske	Skole	Sr	Sönder
Vml	Vandmölle	Tlgr	Telegraf	Ör	Öster
Mlr	Möller	Dsk	Dampskorsten	Vr	Vester
Vmt	Vindmotor	Vkv	Vindkværn		

1 : 20000

1000 500 0 1 2000 Alen

1000 500 0 1000 Meter

Generalstabens topografiske Afdeling.

Kjøbenhavn 1902.

Prisen paa 1 Maalebordsblad af denne Störrelse er i Sort 65 Öre og i Farvetryk 75 Öre.

Appendix 3 Legend (eventyrkort) for the HMB maps, version from 1902. Legend with illustration of the use signature table for the HMB maps from 1902.

AUTOMATED PRODUCTION OF SPATIAL DATASETS FOR LAND CATEGORIES FROM HISTORICAL MAPS

Method development and results for a pilot study of Danish late-1800s topographical maps

This report records the methods and the results of a pilot project aimed at automated production of machine-readable spatial datasets for land categories from Danish topographical maps from the late 1800s. The study was undertaken for two study areas in Jutland, covering around 300 km². Target land categories were: heath, sand dune, wetland, forest and water bodies. The automated geo-data production comprised a combination of object based image analysis, vector GIS, colour segmentation and machine learning processes. Results of an accuracy assessment indicate accuracies that for most categories are around 90 % or higher. A change assessment for the period from the late 1800s until today, revealed a dynamic characterised by decrease in open habitat types due to cultivation and afforestation. We conclude, that automated production of LULC category digital geo-data from historical maps offers a less time consuming and consequently more resource efficient alternative to traditional manual vectorisation.

ISBN: 978-87-7156-511-9
ISSN: 2245-0203