Assessing the suitable cultivation areas for Fibraurea tinctoria in Vietnam using the Maxent and spatial autoregressive model (SAR)

Tuyet T.A. Truong (truongthianhtuyet@tuaf.edu.vn)
Hiroshima University
Cuong Dang Nguyen
Thai Nguyen University of Agriculture and Forestry
Nikki Heherson A. Dagamac
University of Santo Tomas
Thong Van Vu
Thai Nguyen University of Agriculture and Forestry
Tetsuro Hosaka
Hiroshima University

Research Article

Keywords: cultivation, medicinal plant, palmatine, species distribution, spatial autocorrelation, SAR

Posted Date: March 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1430234/v1

Abstract

Cultivating medical plants is an efficient conservation strategy to reduce stress from the loss of wild populations and meet the increasing demand for plant-based medicine. Identifying cultivation areas for the medicinal plant should consider the suitable environmental conditions for both plant growth and the accumulation of bioactive constituents. *Fibraurea tinctoria* Lour., a perennial woody climber, is widely used as a traditional medicine in Southeast Asia. However, due to overexploitation, the species has been suffering a rapid population decline in Vietnam. To promote better conservation and cultivation of *F.tinctoria*, the paper predicted potentially suitable cultivation areas using the Maxent model and built a spatial autoregressive model to identify the correlation between environmental factors and the active ingredient content in *F. tinctoria*. The results showed that *F. tinctoria* has a wide distribution range from the north to the south of Vietnam, excluding the Mekong, Red River Delta, and Central Highland. The precipitation in the driest season, soil types, and broadleaf cover highly influence the species' distribution. Although altitude does not contribute much to the distribution, it highly correlates to the palmatine content. The mean diurnal range (bio2), isothermality (bio3), and precipitation seasonality also influence the palmatine content. The northwest region of Vietnam is the most suitable cultivated area with the highest palmatine. The study provides helpful information for the government and pharmaceutical companies to identify suitable areas for the growth of species and ensure the quality of chemical content; thereby human and material resources in developing medicinal cultivation areas in Vietnam will be prioritized allocated appropriately.

1. Introduction

Fibraurea tinctoria Lour. (Menispermaceae family) is an original plant in China, Indonesia, Malaysia, Thailand, and Vietnam (Wu et al. 1962). *F. tinctoria* is a perennial yellow woody climber which can climb up to 40m long with a stem diameter of 5 cm and is often distributed in lowland forests with an elevation less than 1200m (Al-Saikhan 2020). The species has been widely used in traditional medicines (Perry and Metzger 1980; Niwat et al. 2005). Its stem bark is utilized to cure wounds and inflammation (Al-Saikhan 2020), food poisoning and paralyze, dysentery, analgesic, antipyretic, antidote (Galappathie et al. 2014), diarrhea, and hepatitis (Niwat et al. 2005). In Vietnam, *F. tinctoria* is distributed in secondary forests from the northern mountainous provinces to the Central provinces with elevations less than 1,000m above sea level. In the past, *F. tinctoria* has large populations in the natural forest but is considered as a by-product of forest products and less paid attention. Hence, overexploitation for a long time led to the dramatic decline of the species (Pham and Nguyen 2015). This plant has been listed in the Red Book of Vietnam since 1996 (Ministry of Environmental Science and Technology, 1996), and belongs to group IIA which needs to be protected (Decree 06/2019/ND-CP).

Strategic science and technology research programs to develop the pharmaceutical chemical industry to the year 2020 in Vietnam clearly stated the goal to study and develop the raw material area of *F. tinctoria* to extract 1,000kg palmatine hydrochloride/year (Decision Decree No. 61/2007/QD/TTg). Recently, *F. tinctoria* has been experimented and planted successfully in many places (Vu et al. 2017; Pham, 2014), however, whether the plants meet the relevant chemical composition standards has not been paid attention which can reduce cultivation efficiency and waste of human and material resources. Moreover, the quality of chemical metabolites in herbs is influenced by a multitude of factors, such as the germplasm, plant collection time, and environmental factors. Above all, research has demonstrated that variation in the quality of medical plants is mainly due to environmental factors (Guo et al. 2013; Ncube et al. 2012).

Ecological niche models (ENM) are statistically robust representations that have been widely used for many years to estimate the potential geographic distribution of any species by using (i) known geographical occurrences and (ii) projected environmental factors (Guisan and Zimmermann, 2000; Pearson, 2010). These models simply employ the relationship between the two aforementioned components to determine the suitable habitats where the populations of that particular species can thrive (Almadrones-Reyes and Dagamac 2018). Hence, such models have been used as a promising tool for conservation and understanding species persistence especially to localities whose records are considered limited (Lu et al. 2012; Peterson 2006, Phillips et al. 2004). With fewer data requirements, single algorithms for species distribution are preferred to be employed effectively (Kaky et al. 2020) for poorly studied taxa (Kearney and Porter, 2009) or data-poor regions (Truong et al. 2017). Such regression or correlational algorithms has been used widely in predicting the potential geographical distribution of many species including those plant species that have important medicinal values like *Justicia* sp. (Yang et al. 2013), *Rosa* sp. (Abdelaal et al. 2019), *Carthamus* sp. (Wei et al. 2018), etc.

To assess the correlation between medicinal components with environmental factors, linear regression analysis using Ordinary Least Square (OLS) method has been used in some studies (e.g. Xu et al. 2020; Yuan et al. 2020). However, in fact, real data, especially ecological data often exhibit spatial patterns (Beale et al. 2010; Ver Hoef et al. 2018). Hence, without considering the spatial autocorrelation, this non-spatial regression method can lead to low precision and high error rates (Beale et al. 2010). Spatial autoregessive (SAR) model is specifically designed to model spatially autocorrelated data based on neighborhood relationships which can improve the performance of spatial data (Beale et al. 2010; Pace and Gilley 1997; Ver Hoef et al. 2018).

Intending to develop medicinal plant cultivation for conservation and to improve the quality of traditional medicine in Vietnam, this study, therefore, aimed to (i) determine ecological factors affecting the distribution of *F. tinctoria* and accumulation of palmatine content in *F. tinctoria* and (ii) identify suitable areas in Vietnam where *F. tinctoria* can be successfully cultivated. To achieve these objectives, the study built an environmental suitability map of the habitat for *F. tinctoria* based on the Maxent and SAR model to assess the correlation between the chemical content and environmental variables. Finally, the predicted areas with suitable environments for the cultivation of *F. tinctoria* with high content of palmatine were identified.

2. Material And Methods

2.1. Species occurrence record and chemical information

A total of 177 localities of F. tinctoria in Vietnam were used as the occurrence data. These occurrences were collected from field observations from June 2017- December 2017 in six regions of Vietnam (North East, North West, Red River Delta, North Central Coast, Central Highlands, South East). Observations were designed based on line-transect sampling method with a distance of 3km x 3km in study areas. All samples were collected in wild populations to avoid the human interference.

The latitude and longitude coordinates of these localities were identified by a global positioning system (GPS). The geographical clusters of localities can lead to the overfitting of the species distribution models towards environmental biases and inflation of model performance values (Boria et al. 2014). Thus, we used the spThin package in R to deal with problems associated with spatial sampling biases. This dataset contains 177 verified occurrence records that were spatially thinned, using a thinning distance of 10 km (Radosavljevic and Anderson 2014). Finally, a total of 43 occurrence records of *F. tinctoria* has remained for modeling (Fig. 1)

The palmatine content of 177 *F. tinctoria* germplasm resources, which were collected from the six regions of Vietnam, were extracted. The fresh root samples, ranging between 300g and 500g, were recorded were kept in a plastic bag and weighted. The plastic bags were labeled and sealed. *F. tinctoria* samples were analyzed and extracted Palmatine content by applying high-performance thin-layer chromatography (HPLC) as described by Peng L *et al*. and Thin-layer Chromatography Scanner (Urbain and Simões-Pires 2006).

2.2. Environmental parameters

Bioclimatic variables are biologically meaningful for defining the environmental niche of a species. Hence, they have been broadly utilized for species distribution modeling (Cruz-Cárdenas et al. 2014; Booth et al. 2017; William et al. 200) In the study, we used 19 bioclimatic variables from the WorldClim (Version 2), which is obtained from measurements recorded during the period 1970–2000 from climate stations worldwide (Fick and Hijmans, 2017) downloaded from http://www.worldclim.org. Elevation data at 30-arc-second is also downloaded from Worldclim. In addition, the Land Cover layer map is derived from Copernicus Global Land Operations https://land.copernicus.eu. The Dynamic Land Cover map at 100 m spatial resolution (CGLS-LC100) was downloaded by using Google Earth Engine. The product provides proportional estimates for vegetation/ground cover for the land cover types. The Land Cover maps (v3.0.1) are provided for the year 2019 over the entire Globe, derived from the PROBA-V 100 m through the use of a Sentinel time-series. 30 arc-second raster soil database from Harmonized World Soil Database (Nachtergaele et al. 2009) was also extracted and included for the species distribution modeling.

Table 1
The selected climate variables for modelling the habitat suitability distribution of *F. tinctoria* in Vietnam

Data source	Environmental variables	Abbreviation	Units	Type of data
WorldClim	Mean Diurnal Range (Mean of monthly (max temp - min temp))	Bio2	°C × 10%	continuous
	Isothermality (bio2/bio7)	Bio3	°C × 10	continuous
	Mean Temperature of Wettest Quarter	Bio 8	°C	continuous
	Annual Precipitation	Bio 12	mm	continuous
	Precipitation of Driest Month	Bio 14	mm	continuous
	Precipitation Seasonality (CV)	Bio 15	%	continuous
	Precipitation of Warmest Quarter	Bio 18	mm	continuous
	Precipitation of Coldest Quarter	Bio 19	mm	continuous
	Elevation	alt		continuous
Harmonized World Soil Database	Soil types	soil		categorical
Land cover	Land cover	LCVN		categorical

In total, 22 environmental layers were clipped and resampled to the same geographic boundaries and cell size (at 30-arc-second spatial resolution for Vietnam) in R software (R Core Team 2020). To avoid multiple collinearity of these variables and overfitting of the Maxent model, the environmental variables were filtered by Variance Inflation Factor (VIF) (Fois et al. 2015). VIF values were calculated by the *vifcor* and *vifstep* function of usdm R package (Naimi et al. 2014). Since VIF values greater than 10 is a signal that the model has a collinearity problem, only value less than 10 were remained. Finally, 11 environmental variables (Table 1) were chosen to run the model.

2.3. Modelling Habitat Suitability of Species

To model the habitat suitability of species, we used the maximum entropy species distribution model (Maxent, Version 3.3.3k) (Phillips 2013). Maxent has been widely used since it requires presence-only data and provides good results even with very small sample sizes (Papes and Gaubert, 2007; Hernandez et al. 2008).

2.3.1. Model optimization

Current studies pointed out that setting default parameters may make overfitting models whose results are difficult to interpret (Radosavljevic and Anderson, 2014; Warren et al. 2014; Warren and Seifert, 2011). An R program package ENMeval was developed to select the optimal model parameters by regulating regularization multiplier (RM) and feature combination (FC) (Muscarella et al. 2014). Different parameter combinations are analyzed to find out the combination with the lowest complexity for modelling (Zhao et al. 2021).

The study used ENMeval package in R (R Core Team, 2020) to optimize the model parameters Muscarella et al. (2014). RM parameter was set from 0.5–4 with 0.5 intervals (Elith et al. 2011; Guevara et al. 2018). Five features were set including linear (L), quadratic (Q), hinge (H), and product (P) and threshold (T) (Phillips et al. 2006). We selected 5 feature combinations: L, LQ, H, LQH, and LQHP. To quantify the degree of overfitting, we used four metrics: (1) The Akaike information criterion correction (delta.AICc) was used to evaluate the fit and complexity of the model (Akaike, 1973; Burnham and Anderson, 2004); (2) the difference between training and testing AUC (avg.diff.AUC), (3)10% training omission rate (or.10p.avg) (Pearson et al. 2010; Warren and Seifert 2011), and (4) AUC train and test average (Warren and Seifert 2011). The parameter combination with the minimum delta.AICc value, low difference between training and testing AUC, low 10% training omission rate and high AUC had been chosen as the optimal parameters to build the model.

2.3.2. Model simulation

To reduce the potential effects of spatial autocorrelation in the model building, we ran the model 10 times, with a random 30% of the presence points withheld each time period, and averaged the results. The selected output grid format was “logistic”, in which the pixel values range from 0 to 1. The contribution of each variable to the habitat model of *F. tinctoria* was calculated using the software's built-in jack-knife test. The response curves of each variable were calculated showing the quantitative relationship between the environmental variables and the logistic probability of habitat suitability. The logistic probability of a presence being higher than 0.5 indicated that the value range of its corresponding ecological factor was suitable for the growth of *F. tinctoria*. However, when the logistic probability of presence was lower than 0.1, the corresponding range was not suitable for growth.

2.3.3. Model evaluation

To evaluate the predictive performance of the model, we calculated the AUC (area under the receiver operating characteristic curve) values. Models with AUC values from 0.7 to 0.9 were considered to have a good fit. The suitable habitat for *F. tinctoria* was predicted. The results (in the range from 0 to 1) indicated different degrees of habitat suitability. Although changes in nature do not occur in discrete intervals, for conceptualization and easy interpretation of the results of the Maxent model, the predictive results were regrouped and plotted into four levels based on Jenks' natural breaks: unsuitable habitat (0–0.1), low suitable habitat (0.1–0.3), moderate suitable habitat (0.3–0.5), high suitable habitat (0.5–1). This criterion has been widely used in species distribution models (Arabameri et al. 2020; Li et al. 2019; Zhao et al. 2021). Also, to assess the importance of environmental variables of *F. tinctoria* contribution rate, permutation importance and Jackknife test were used.

2.4. Determining the environmentally suitable cultivation areas

To assess the correlation between the content of the chemical components and environmental variables, SAR model was used. Before spatial modeling is used, the data was tested for spatial dependencies with Moran's I method. This method used to test autocorrelation between locations of palmatine samples. The hypotheses of Moran's I test including H0: $\rho = 0$ (there is no spatial autocorrelation), H1: $\rho \neq 0$ (there is spatial autocorrelation). If Ho is rejected, it means there is a spatial dependency among observations. Then the spatial lag autoregressive model is used to demonstrate the spatial relationship between independent variables (X) and dependent variable (Y). The location effects on the data are embodied by weights.

$$y = \rho W y + X\beta + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2 I)$, I is identity matrix, y is a vector of observations (% palmatine content) on the dependent environmental variable (nx1), X is a matrix of observations (k) on the independent variables nxk, ρ is spatial autoregressive coefficient and $|\rho| < 1$. Moran's I test and spatial lag autoregressive model was performed in R with spdep package (R Core Team, 2020).

The optimal content of palmatine in root and stem of *F. tinctoria* according to Bich et al. (2003) is 2–3%. Hence, 2% of palmatine content was set as the threshold level to identify the suitable areas for cultivation. The areas with environmental suitability for the cultivation of *F.tinctoria* were visualized in ArcGIS v. 10.3 using the Maxent results and the previously set threshold level for palmatine production.

3. Results And Discussion

3.1. Model optimization and accuracy evaluation

Results from ENMeval package showed that RM=3 and FC= LQH has low AICc value with delta.AICc close to 0. Also, this combination set had both small values of AUC.DIFF and OR10, while higher average AUC value (Figure 2). This indicated that the model under this parameter set had a low degree of overfitting. Hence, we selected the combination setting RM = 3 and FC = LQH as the optimal model parameters to predict the potentially suitable habitat of *F. tinctoria*. Under this parameter setting, the average AUC value of 10 times of repeated training set was 0.939± 0.017 and the average AUC value of the test set was 0.906± 0.036, which now indicates high model accuracy for *F. tinctoria* prediction.

3.2. *Suitable habitat of F. tinctoria*

The prediction of *F. tinctoria* distribution in Fig. 5 (a) shows that *F. tinctoria* has a wide range of suitable habitats in Vietnam. This wide distribution range obtained from the Maxent result aligns with the documentation of previous studies (Pham and Nguyen, 2015; Vu et al. 2017). The total area of potential high suitable habitat estimated from the suitable habitat map is 10473992.8 ha, accounting for 32% of Vietnam's territory. The vast majority of these areas mainly lie in North and Central Vietnam. In addition, habitats that are not suitable for the distribution of the species include the following: (i) Mekong River Delta, (ii) Central Highland, and (iii) a part of Red River Delta.

Generally, the habitat to achieve the palmatine content is similar to the potentially suitable habitat for the distribution of *F.tinctoria* predicted by Maxent modelling. The only difference is a 63912.5-hectare decline lying in some points in Quang Tri, Thua Thien Hue, Lam Dong, Dong Nai (Fig. 5b, 5c). This means the species can still occur in those areas but the palmatine content is not high. Hence, if decisions on developing medicinal plant areas are made based on the growth and distribution index only, we would probably get biased results that the highly suitable regions for medicinal plants may not be the optimum cultivation regions with high quality of medicinal materials.

3.3. Evaluation of important climatic variables

The results of percent contribution and permutation importance showed that land cover, and precipitation of driest month (bio14) and soil are the most important variables to the distribution of *F. tinctoria* with cumulative contribution rate and permutation importance value reached 95.1%., and 89.9%, respectively (Table 2). Jacknife results also indicated that land cover, and precipitation of driest month (bio14) and soil contain more effective information than others.

Table 2
Percent contribution and permutation importance of variables

Variable	Percent contribution	Permutation importance
landcover	53.4	38.5
bio14	33.5	44.9
soil	8.2	6.5
bio2	1.9	4.9
bio19	0.7	0.3
bio3	0.6	1.2
bio18	0.6	1.4
bio8	0.4	0.6
bio12	0.3	0.6
altVn	0.3	0.6
bio15	0.1	0.5

The response curves (Fig. 4) of environmental variables indicated that *F. tinctoria* mostly prefers to grow in areas where the precipitation of driest month (bio14) is 20mm or higher. Although being a vine with a deep and efficient root system which allows them to avoid drought stress and be more competitive than trees in the dry season, the rainfall is still crucial to the development of those species. The height growth of lianas doubles in wet season than dry season (Schnitzer 2005). Remaining photosynthetically active during seasonal droughts (Opler et al. 1991;Kalácska et al. 2005) is a trait of lianas which may require the minimum precipitation for the driest season to ensure for photosynthetic activities of the plant.

F. tinctoria is also suitable in closed evergreen, broadleaf forest with tree canopy > 70% (Fig. 4). In terms of soil type, *F. tinctoria* has been influenced by Orthic Acrisols, Lithosols, and Chromic Luvisols, which are generally acidic, low level of fertility, highly occur in upland areas (FAO 1988) and typical for tropical forests soils (JICA 2002). *F. tinctoria* seems also not suitable with wetland, highly clay soils like Gleysols, Thionic Fluvisols and Rhodic Ferralsols which are mainly found in Mekong River Delta, Central Highland and Red River Delta. Those areas also have low rainfall in the driest season (less than 20mm during the December to April), low broadleaf cover (high area of cultivated and managed vegetation/cropland). A number of studies on the factors that affect liana abundance and community structure in tropical forests also showed that the rainfall, soil properties, disturbance and host trees were one of the most influencing factors to the development of liana (Gianoli 2015; Hogan et al. 2017; Schnitzer 2005; Schnitzer and Bongers 2011). Disturbances like deforestation, cultivation which destroy the broadleaf trees and change the water regime, structure of soil can lead to the decline of *F. tinctoria* distribution range. In the literature review of Ali et al. (2016) on medicinal climbers' distribution and conservation, they stated that the continuous extraction of wild species resulted in the shift and decline of their distribution during the last two decades due to the substantial loss of their habitats .

Table 3. Soil types and land cover used in the prediction of suitable habitat for *Fibraurea tinctoria* in Vietnam. The bold texts are land/soil types that have high contribution to the distribution of. *F. Tinctoria*

Code	Landcover	Soil type
1	Shrubs	Ferric Acrisols
2	**Herbaceous vegetation**	Ferric Acrisols
3	Cultivated and managed vegetation/agriculture (cropland)	Ferric Acrisols
4	Urban / built up	Gleyic Acrisols
5	Bare / sparse vegetation	Gleyic Acrisols
6	Permanent water bodies	**Orthic Acrisols**
7	Herbaceous wetland	Orthic Acrisols
8	Closed forest, evergreen needle lea	Eutric Gleysols
9	**Closed forest, evergreen, broad leaf**	Eutric Gleysols
10	Closed forest, deciduous broad leaf	**Lithosols**
11	Closed forest, mixed	Lithosols
12	Closed forest, unknown	Calcaric Fluvisols
13		Eutric Fluvisols
14	Open forest, evergreen broad leaf	Thionic Fluvisols
15		**Chromic Luvisols**
16		Eutric Regosols
17	Open forest, unknown	Pellic Vertisols
18	Open sea	Acrid Ferralsols
19		Orthic Ferralsols
20		Rhodic Ferralsols
21		Dystric Gleysols

3.4. Correlation between chemical constituents and environmental variables

The results from Moran' I test with Z statistics test is 6.0135, Moran's I is 0.2910 and p-value = 9.075e-10 showed that there are spatial dependences on palmatine content. Therefore, a spatial model was used to identify the correlation between palmatine content and environmental variables. The result of spatial model using Spatial Autoregressive (SAR) is shown in Table 4.

Table 4
Summary of spatial autoregressive (SAR) model

Variables	Estimate	Std. Error	z value	Pr(>|t|)
(Intercept)	15.2710	4.3044	3.5478	0.0004
Orthic Acrisols	1.2466	0.6312	1.9754	0.0482
Altitude	0.0015	0.0003	4.6919	0.0000
Bio 2	-0.1212	0.0520	-2.3292	0.0198
Bio 3	-0.0089	0.0020	-4.4599	0.0000
Bio 15	-0.0576	0.0209	-2.7578	0.0058

The results indicated that palmatine content is highly correlated to the Orthic Acrisols, altitude, the temperature oscillation (bio2 and bio3) and the Precipitation Seasonality (bio15). While the higher altitude will yield positive influence on the palmatine content, the increase of monthly temperature oscillation, high isothermal activity, and precipitation seasonality tend to have negative impacts on the palmatine content (Table 4).

Lam Dong and Dong Nai where the isothermality was recorded highest (over 500) had the lowest average of Palmatine (Table 5) and unsuitable for cultivation (Fig. 5b,c). Also, the palmatine is only around threshold (2–3%) in provinces such as Kontum, Lai Chau, Ha Tinh, Thua Thien Hue, Lam Dong where the average of altitude is below 200m or over 1000m, together with high bio 2 (> 59) The average of palmetin is highest in provinces that have high altitude (400-1000m) such as provinces in the Northwest (Dien Bien, Lao Cai, Son La, Cao Bang) and Dac Nong. Pham and Nguyen (2015) also pointed out that in Vietnam, *F. tinctoria* grows best in the elevation of less than 1000m. Data from Phillips and Miller 2002 also found that, the liana species grow best in lowland tropical forests with had a mean annual rainfall ≥ 500 mm and was ≤ 1,000 m in elevation. Many studies tested the chemical content among various medicinal plants at different elevations and also found that the altitudinal gradient influence not only species distribution but also the chemical content (Lakey and Dorji 2016; Sun et al. 2016). This can be explained by the positive correlation between elevation to soil moisture, pH, nitrogen, phosphorus, and exchangeable calcium and magnesium (Moser et al. 2009). Those factors were proved to have a significant positive correlation with the palmatine content in plants (Zhang et al. 2011).

Table 5
Palmatine content by provinces and environmental factors. The bold texts are provinces with high average palmatine content

Provinces	Average of Palmetin	Average of bio3	Average of bio15	Average of bio2	Average of altVn
Backan	3.96	314.71	84.11	52.30	316.20
Binhphuoc	2.93	503.99	78.75	59.50	244.00
Caobang	**4.58**	**312.66**	**73.40**	**59.49**	**647.58**
Daclak	2.46	480.69	73.34	61.00	652.67
Dacnong	**5.79**	**469.98**	**73.51**	**61.00**	**441.50**
Dienbien	**5.22**	**376.47**	**79.70**	**55.15**	**757.46**
Dongnai	2.60	506.16	76.26	59.09	144.33
Gialai	4.29	440.48	72.01	59.50	801.50
Hagiang	4.03	318.16	71.49	60.33	132.17
Hatinh	2.61	238.94	67.99	60.00	31.67
Kontum	2.49	456.86	76.65	60.00	1170.00
Laichau	2.89	367.18	76.59	60.40	1199.00
Lamdong	0.53	503.70	84.83	61.70	634.20
Laocai	**4.52**	**341.39**	**81.41**	**59.20**	**420.20**
Phutho	3.48	330.19	88.98	49.97	384.00
Quangbinh	4.31	332.46	75.51	55.49	30.50
Quangtri	2.85	358.50	76.27	58.47	48.40
Sonla	**5.10**	**345.17**	**65.39**	**61.00**	**429.00**
Thainguyen	**4.74**	**318.13**	**70.84**	**57.25**	**237.08**
Thanhhoa	3.81	315.15	77.14	62.67	90.83
Thuathienhue	2.88	343.57	80.80	52.53	183.87
Tuyenquang	3.25	315.74	71.73	56.75	69.25
Vinhphuc	3.62	338.65	96.68	41.00	1101.33
Yenbai	3.60	341.95	95.79	41.00	1011.88

Our results aligned with others studies (Guo et al. 2013; Ncube et al. 2012) that showed environmental factors affect not only plant growth and reproduction, but also the content of active ingredients that are related to environmental stresses. Using the spatial autoregressive model, the relationship between palmatine content and environmental factors was established for this study, identifying now the ultimate climate regions in Vietnam where the species can be successfully cultivated.

Conclusion

F. tinctoria with high palmitin content in roots and stems have been widely used as traditional herbal plants and tend to be developed in the pharmaceutical chemical industry. The species has a wide distribution range in Vietnam. However, to cultivate efficiently the species with high palmitin content, Dac Nong and northwest provinces such as Dien Bien, Son La would be the best suitable cultivation area. Red River and Mekong delta river and Central Highland are unsuitable for the distribution of the species. It is also noted that in provinces such as Thua Thien Hue, Dong Nai, Lam Dong where the species may occur, the cultivation of *F. tinctoria* in this area may not be efficient due to the low palmatine content. Through analysing environmental factors, it seems that while precipitation in the driest season, broadleaf forest cover, and Orthic Acrisols soil have the most contribution to the distribution of the species, the altitude and isothermality are the additional factors that shape the palmatine content in *F. tinctoria.* Since environment factors control the geospatial distribution characteristics and are closely correlated with the overall quality of the herbal drug, comprehensive quality control requires measuring both the chemical constitutes and the chemical bioactivities under specific growth conditions. Disturbances and future warming may cause the planting regions suitable for *F. tinctoria* to continue to expand northward. However, if the expansion leads to the shift of species to over 1000m altitude, the decline of the distribution and chemical content can happen. Hence, special attention should be paid to the influence of extreme climatic conditions on the distribution and active ingredients. Targeted cultivation based on the geospatial distribution characteristics of active ingredients in *F. tinctoria s*hould be taken seriously.

Declarations

Funding

The study was conducted with the financial support of the Ministry of Science and Technology in Vietnam and Hiroshima Univesity.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Tuyet TA Truong prepared input data, performed models and interpreted results, wrote manuscript and acted as corresponding author. Cuong Dang Nguyen collected data in the field, prepared data in Arcgis, edited the manuscript, and acted also the corresponding author. Thong Van Vu collected data in the field.

Tetsuro Hosaka provided guidance for SAR model, and edited the manuscript. Nikki Dagamac contributed to editing the manuscript.

References

1. Abdelaal, M., Fois, M., Fenu, G., & Bacchetta, G. (2019) Using MaxEnt modelling to predict the potential distribution of the endemic plant Rosa arabica Crép. in Egypt. Ecological informatics 50: 68–75.
2. Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*. Springer, New York, NY, pp 199–213
3. Almadrones-Reyes, K. J., & Dagamac, N. H. A. (2018) Predicting local habitat suitability in changing climate scenarios: Applying species distribution modelling for Diderma hemisphaericum. Curr. Res. Environ. Appl. Mycol 8:492–500.
4. Arabameri, A., Asadi Nalivan, O., Saha, S., Roy, J., Pradhan, B., Tiefenbacher, J. P., & Thi Ngo, P. T. (2020) Novel ensemble approaches of machine learning techniques in modelling the gully erosion susceptibility. Remote Sensing 12(11):1890.
5. Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010) Regression analysis of spatial data. Ecology letters 13(2): 246–264.
6. Booth, T. H. (2017) Assessing species climatic requirements beyond the realized niche: some lessons mainly from tree species distribution modelling. Climatic Change 145(3):259–271.
7. Burnham, K. P., & Anderson, D. R. (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research 33(2): 261–304.
8. Cruz-Cárdenas, G., López-Mata, L., Villaseñor, J. L., & Ortiz, E. (2014) Potential species distribution modelling and the use of principal component analysis as predictor variables. Revista Mexicana de Biodiversidad 85(1): 189–199.
9. Chesworth, W. (Ed.). (2007) *Encyclopedia of soil science*. Springer Science & Business Media.
10. Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011) A statistical explanation of MaxEnt
11. FAO (1988) Report of the mission on Agricultural Sector Review
12. Fick, S.E. and R.J. Hijmans (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302–4315.
13. Fois, M., Fenu, G., Lombrana, A. C., Cogoni, D., & Bacchetta, G. (2015) A practical method to speed up the discovery of unknown populations using Species Distribution Models. Journal for Nature Conservation 24:42–48.
14. Guevara, L., Gerstner, B. E., Kass, J. M., & Anderson, R. P. (2018) Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. Global change biology 24(4): 1511–1522.

15. Hogan JA, Mayorquin S, Rice K, Thompson J, Zimmerman JK, Brokaw N. (2017) Liana dynamics reflect land-use history and hurricane response in a Puerto Rican forest. J Trop Ecol. 33:155–164.
16. JICA-Japan International Cooperation Agency (2002) Technical Manual Series on the Feasibility Study on the Forest Management Plan in the Central Highland in Socialist Republic of Viet Nam. Japan Overseas Forestry Consultants Association (JOFCA). Pasco Corporation, Viet Nam.
17. Kaky, E., Nolan, V., Alatawi, A., & Gilbert, F. (2020) A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. Ecological Informatics 60:101150.
18. Li, H. Q., Liu, X. H., Wang, J. H., Xing, L. G., & Fu, Y. Y. (2019) Maxent modelling for predicting climate change effects on the potential planting area of tuber mustard in China. The Journal of Agricultural Science 157(5):375–381.
19. Liu, R. M., Zhao, S. X., & Zhu, R. H. (1981) A note on the identification of fibralactone from the root of Fibraurea recisa Pierre (author's transl). Yao xue xue bao = Acta pharmaceutica Sinica 16(6):479–480.
20. Nachtergaele, F., Velthuizen, H. V., Verelst, L., & Wiberg, D. (2009) Harmonized world soil database (hwsd). *Food and Agriculture Organization of the United Nations, Rome*.
21. Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014) ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in ecology and evolution 5(11):1198–1205.
22. Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., and Toxopeus, A.G. (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37 (2): 191–203.
23. Opler, P. A., H. G. Baker, and G. W. Frankie. 1991. Seasonality of climbers: a review and example from Costa Rican dry forest in F. E. Putz and H. A. Mooney, eds. The biology of vines. Cambridge University Press, Cambridge, pp 377–391
24. Pace, R. K., & Gilley, O. W. (1997) Using the spatial configuration of the data to improve estimation. The Journal of Real Estate Finance and Economics 14(3):333–340.
25. Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Townsend Peterson, A. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 34(1):02–117.
26. Peng L, et al. (2014) Spaceflight environment-induced variation in root yield and active constituents of *Salvia miltiorrhiza*. Planta Medica. 80:1029–1035. doi: 10.1055/s-0034-1382594.
27. Perry, L.M. and Metzger, J. (1980) Medicinal Plants of East and Southeast Asia. MIT Press, Cambridge, pp. 269.
28. Phillips, O. L., and J. S. Miller (2002) Global patterns of plant diversity: Alwyn H. Gentry’s forest transect data set. Missouri Botanical Garden, St. Louis
29. Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006) Maximum entropy modelling of species geographic distributions. Ecological Modelling 190(3–4): 231–259.

30. Phạm Hữu Hạnh (2014) Nghiên cứu một số đặc điểm sinh học và kỹ thuật gây trồng cây Hoàng Đằng (Fibraurea tinctoria Lour) tại huyện Hoành Bồ, tỉnh Quảng Ninh.
31. Pham Huu Hanh, Nguyen Huy Son (2015) Effects of fertilizer and light cover on growth of *Fibraurea tinctoria* Lour at the stage of nursery. Vietnam Journal of Forest Science 3:3889–3896.
32. R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
33. Radosavljevic A., R.P. Anderson (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation, J. Biogeogr. 41 (4):629–643
34. Schnitzer SA. (2005) A mechanistic explanation for global patterns of liana abundance and distribution. Am Nat. 166:262–276. doi:10.1086/431250.
35. Schnitzer SA, Bongers F. (2011) Increasing liana abundance and biomass in tropical forests: emerging patterns and putative mechanisms. Ecol Lett. 14:397–406. doi:10.1111/j.1461-0248.2011.01590.x.
36. Tekwa, I. J., Shehu, H. E., & Maunde, S. M. (2011) Soil nutrient status and productivity potentials of lithosols in Mubi Area, Northeastern Nigeria. Agriculture and Biology Journal of North America 2(6):887–896.
37. Thông, V. V., Thúy, P. T., Vy, V. P. T., & Nguyên, V. T. (2017) Nghiên cứu kỹ thuật nhân giống vô tính cây hoàng đằng (Fibraurea tinctoria Lour). TNU Journal of Science and Technology 171(11):201–206.
38. Urbain, A., & Simões-Pires, C. A. (2006) Thin-layer chromatography of plants, with chemical and biological detection methods. Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, 1–22.
39. Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M. J. (2018) Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs 88(1): 36–59.
40. Warren, D.L.; Wright, A.N.; Seifert, S.N. and Shaffer, H.B. (2014) **Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern**. Divers. Distrib., 20:334–343
41. Warren, D.L. and Seifert, S.N (2011) **Ecological niche modelling in Maxent: the importance of model complexity and the performance of model selection criteria**, Ecol. Appl., 21:335–342
42. Wei, B., Wang, R., Hou, K., Wang, X., & Wu, W. (2018) Predicting the current and future cultivation regions of *Carthamus tinctorius* L. using MaxEnt model under climate change in China. Global Ecology and Conservation 16:e00477.
43. Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009) Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions 15(4): 565–576.
44. Wu, M. C., Su, C. W., Chang, L. Y., & Lee, C. K. (1962). A pharmacognostical study of chinese drug huang-teng. The roots and stems of *Fibraurea tinctoria* Lour. Yao xue xue bao = Acta pharmaceutica Sinica 9:233–241.

45. Xu, N., Meng, F., Zhou, G., Li, Y., Wang, B., & Lu, H. (2020) Assessing the suitable cultivation areas for *Scutellaria baicalensis* in China using the Maxent model and multiple linear regression. Biochemical Systematics and Ecology 90:104052.
46. Yang, X. Q., Kushwaha, S. P. S., Saran, S., Xu, J., & Roy, P. S. (2013) Maxent modelling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. Ecological engineering 51:83–87.
47. Yu, T. E. (Ed.). (1997) *Chemistry of variable charge soils*. Oxford University Press.
48. Yuan, Y., Tang, X., Jia, Z., Li, C., Ma, J., & Zhang, J. (2020) The effects of ecological factors on the main medicinal components of Dendrobium officinale under different cultivation modes. Forests 11(1):94.
49. Zhang J, Jin H, Li W, Wang Y, Zhang J. (2011) Correlation analysis soil nutrition as sell as yield and active compounds of Coptis teeta. Zhongguo Zhong Yao Za Zhi 36(19):2629–31. Chinese. PMID: 22242419.
50. Zhao, Y., Deng, X., Xiang, W., Chen, L., & Ouyang, S. (2021) Predicting potential suitable habitats of Chinese fir under current and future climatic scenarios based on Maxent model. Ecological Informatics 64:101393.

Figures

Figure 1

Occurrence records of *F. tinctoria* in Vietnam

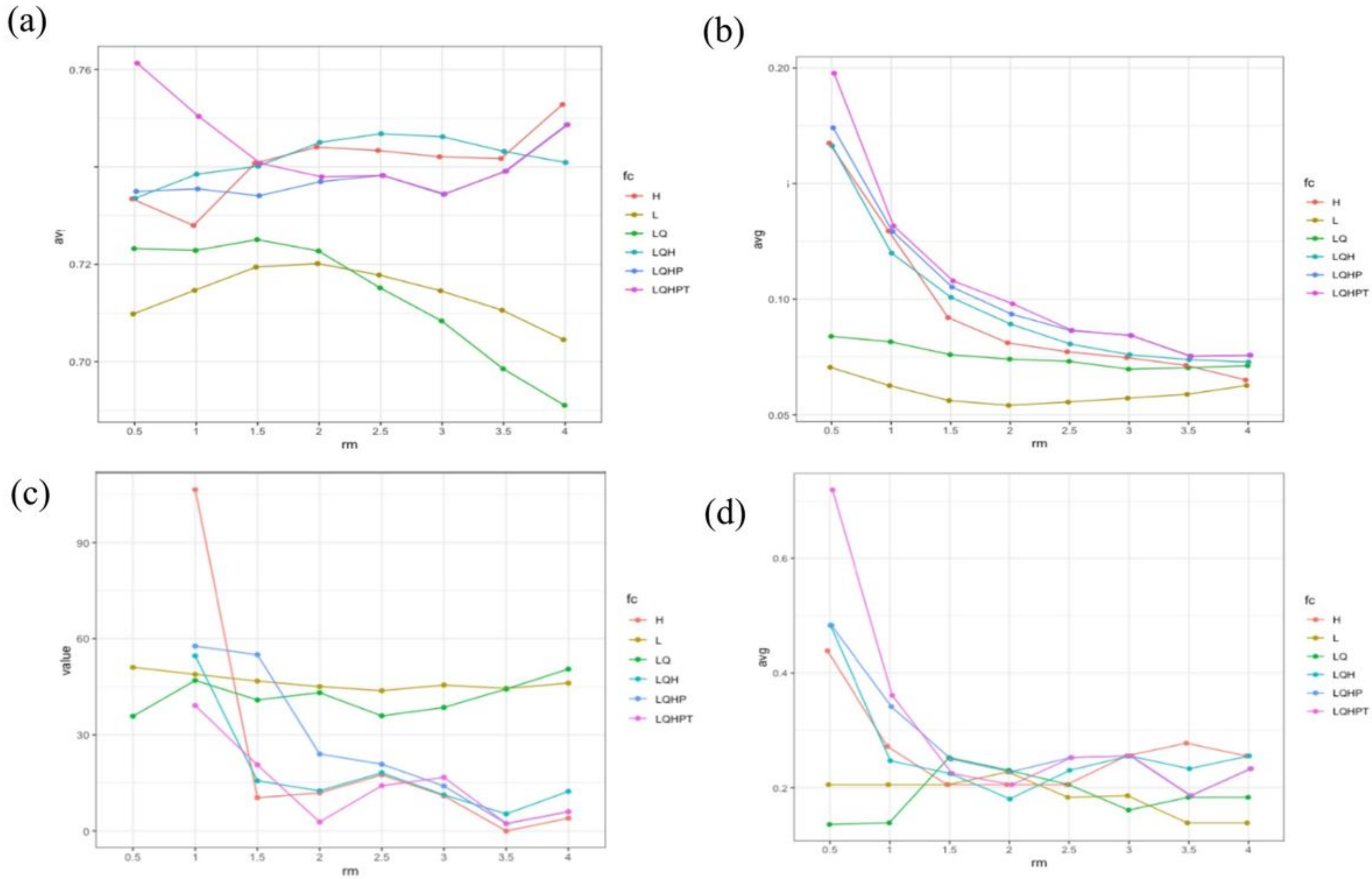

Figure 2

(a) Delta.AICc, (b) AUC.DIFF and mean of all validation, (c) AUC test average, (d) 10percent omission rates for *F. tinctoria* resulting from Maxent models under different parameter combinations. Legends denote different feature classes (L = linear, Q = quadratic, H = hinge, P = product and T = threshold).

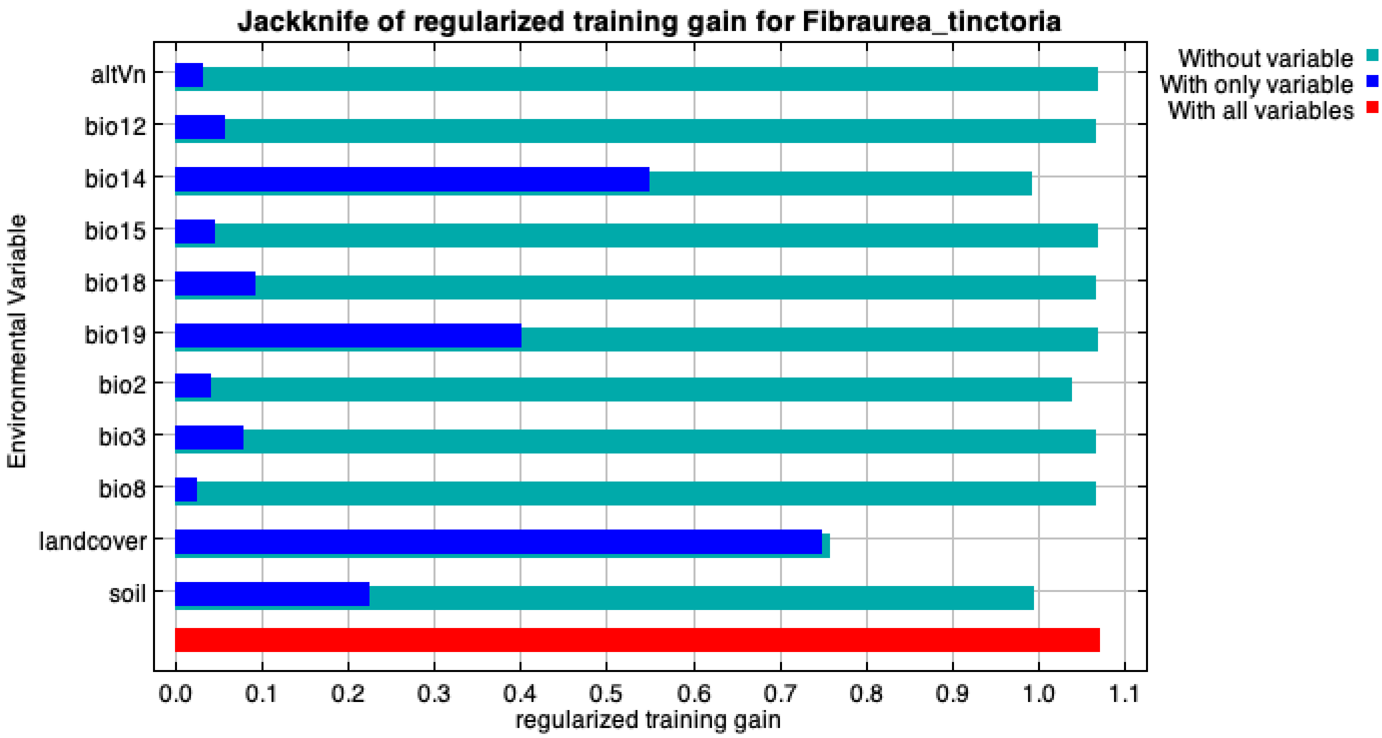

Figure 3

Jackknife of regularized training gain for *Fibraurea tinctoria*

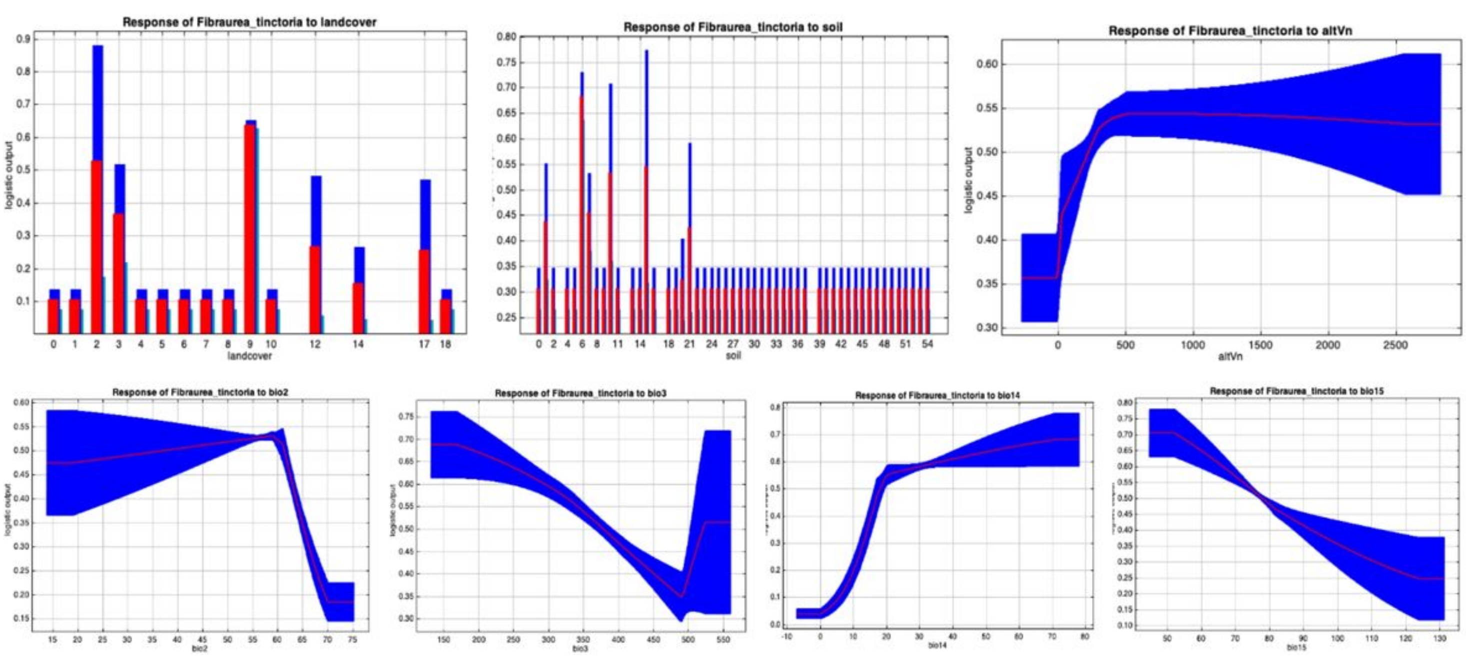

Figure 4

Marginal response curves of *Fibraurea tinctoria* for variables. The red curve in each plot is average response curve and the blue is standard deviation across all 10 partition runs.

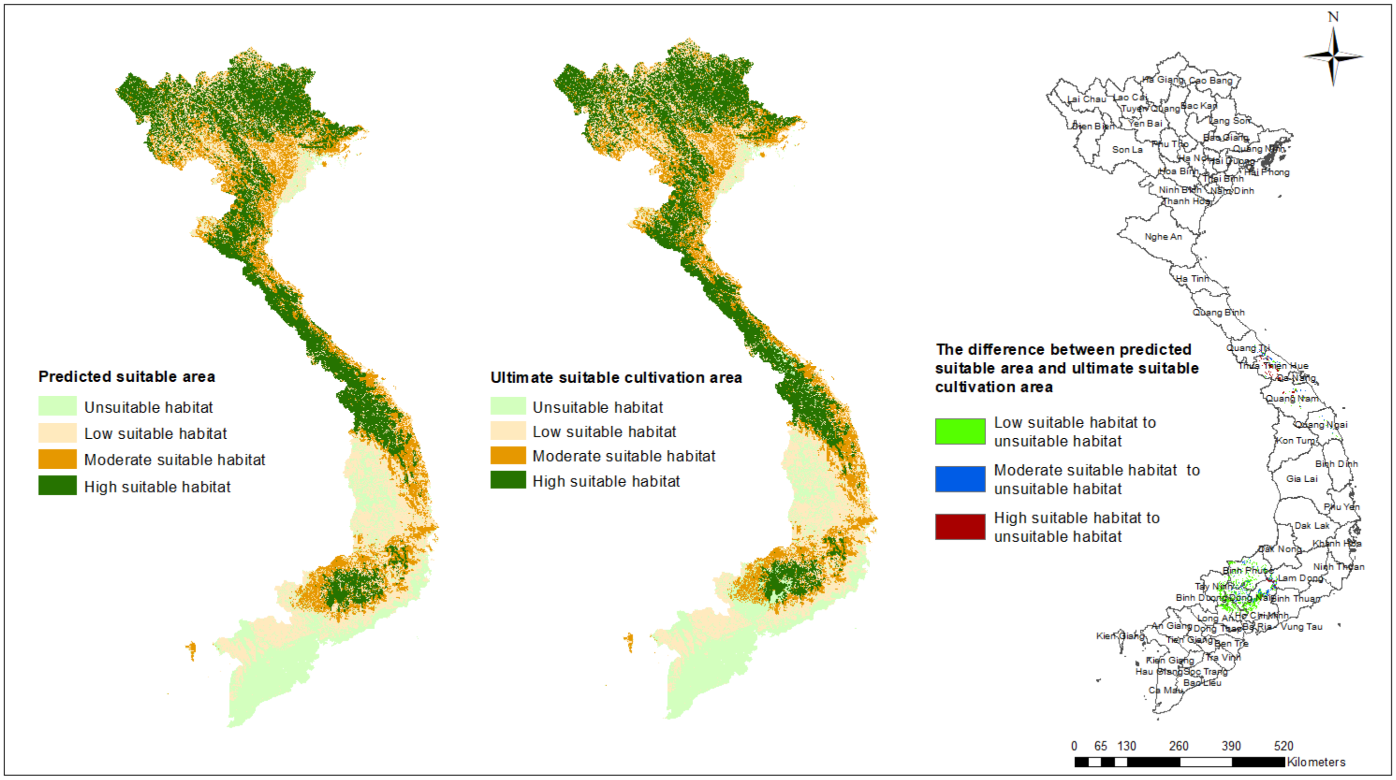

Figure 5

Predicted suitable habitat of *Fibraurea tinctoria*

(a) Predicted suitable habitat of *Fibraurea tinctoria* using Maxent

(b) The ultimate climate cultivation areas of *Fibraurea tinctoria* with minimum 2% Palmatine content

(c) The difference between predicted suitable habitat and ultimate suitable cultivation area of *Fibraurea tinctoria*