

Risk of Natural Spread of *Hymenoscyphus fraxineus* with Environmental Niche Modelling and Ensemble Forecasting Technique

Elisa Dal Maso* and Lucio Montecchio

University of Padova, TeSAF Department, Viale dell'Università, Italy

Abstract

Ash dieback, caused by the ascomycete *Hymenoscyphus fraxineus*, is rapidly expanding over large geographical areas in Europe. A myriad of factors influence pest invasions and long-term establishment, i.e., species' life stage, the availability of suitable hosts and the suitability of the environment. This paper examines the principal environmental features that characterise naturally infected zones in order to forecast the potential distribution of the pathogen within the ranges of European ash species by means of Species Distribution Modelling and an ensemble forecasting technique. Furthermore, a network analysis permitted dispersal dynamics to be included in order to obtain realistic risk predictions for the natural spread. The multi-modelling procedure allowed the most endangered regions to be identified as the central and eastern Alps, Baltic States, southern Finland and the area encompassing Slovakia and southern Poland, whereas most marginal regions of the study area appeared less suitable for the natural establishment and spread of the disease. Statistical model predictions were highly correlated with abundant summer precipitation, high soil moisture and low air temperature. A novel approach to the ensemble forecasting technique in epidemiological modelling of plant pathogens is suggested as a tool to aid the survey of this infectious disease. Moreover, the final potential distribution maps may promote discussions about the control of the disease and the risks associated in the trade or movement of ash species.

Keywords: Ash dieback; *Hymenoscyphus pseudoalbidus*; *Chalara fraxinea*; *Fraxinus*; Species distribution models; Epidemiology

Abbreviations: GDD: Growing Degree Days; TSS: True Skill Statistic; AUC: Area Under The Curve; ROC: Receiver Operating Characteristic; GLM: Generalised Linear Model; LOG: Logistic Regression Model; SVM: Support Vector Machine Model; MLP: Multilayer Perceptron Artificial Neural Network; CHAID: Chi-Squared Automatic Interaction Detector Classification Tree; WA: Weighted Average

Introduction

Ash trees in Europe are threatened by a major disease caused by the ascomycete *Hymenoscyphus fraxineus* (T. Kowalski) Baral, Queloz, Hosoya, comb. nov. (basionym *Chalara fraxinea* T. Kowalski, synonym *Hymenoscyphus pseudoalbidus* Queloz et al.) [1-4], most likely introduced from East Asia [5]. The disease was first observed on *Fraxinus excelsior* L. in northeastern Poland in the 1990s [6], but the pathogen was identified as the primary causal agent of ash dieback in 2006 [1]. Symptoms were also observed in both European (*F. angustifolia* Vahl. and, only under artificial conditions, *F. ornus* L.) and introduced ash species (*F. nigra* Marsh., *F. pennsylvanica* Marsh., *F. americana* L. and *F. manschurica* Rupr.) [7-10].

Wind-dispersed ascospores, produced during the summer in apothecia on the previous year's leaf remnants in the litter, can penetrate and infect ash leaves via appressoria [10-12]. The symptoms that subsequently develop include necrotic leaf spots: wilting of leaves and young shoots; premature shedding of leaves; crown dieback; and necrotic bark lesions extending to the phloem, paratracheal parenchyma and parenchymatic rays below the bark [13,14].

At the present time, fully effective measures to control the disease are still lacking [15-18]. Due to its rapid spread in the majority of European countries [1,11,19], *H. fraxineus* was added to the EPPO Alert List in 2007 but was later deleted because sufficient alert was considered to have been given [20].

Predicting the spread of emerging infectious diseases is fundamental

for forecasting potential ecological consequences and designing control strategies [21-26], and mathematical models have long been widely used for agricultural and forest diseases [21,22,27-30], in particular, to predict the spread of parasites and pests [31-35] or the risk of infection in pest-free areas [36-38]. Among those extensively employed, Species Distribution Models (SDMs) can identify statistical or logical functions linking species' occurrences to a series of predictors, and project these relationships onto a geographical space, allowing range dynamics to be estimated and suitability maps defined [39]. In fact, the SDM approach is based on the concept of Grinnellian niche as a constraint to the potential distribution of species and can easily be implemented using ecological and evolutionary assumptions (i.e., selecting the most causal environmental predictors or determining a restricted set of competing models in multi-model inference) [40,41].

Some environmental characteristics connected to a pathogen's biology are known or hypothesised from field observations and laboratory experimental proofs. In terms of temperature requirements, *H. fraxineus* can be classified as a mesophile [42], considering that most isolates in pure culture show their maximal growth rate at approximately 20°C and cease to develop at approximately 30°C [43]. However, in ash tissues, the fungus exhibits less tolerance to heat [16]. On the other side, the pathogen is considered a cold-tolerant organism because of the ability of producing necroses during the winter and phialides at low

*Corresponding author: Elisa Dal Maso, University of Padova, TeSAF Department, Viale dell'Università, Italy, Tel: 0039-0498272887; Fax: 0039-0498272890, E-mail: elisa.dalmaso@studenti.unipd.it

Received September 19, 2014; Accepted November 09 2014; Published November 16, 2014

Citation: Dal Maso E, Montecchio L (2014) Risk of Natural Spread of *Hymenoscyphus fraxineus* with Environmental Niche Modelling and Ensemble Forecasting Technique. Forest Res 3: 131. doi:10.4172/21689776.1000131

Copyright: © 2014 Dal Maso E, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

temperature [43,44]. The asexual stage of the pathogen is most likely strongly associated with the pseudosclerotial plates that it produces on infected rachises [10] and that allow the fungus to overwinter [12,13]. The main hypothesis for subsequent fertilisation is proposed by Gross et al. [45] and supposes that conidiospores (spermatia), readily produced on the petiole in autumn, could be mediated by free water till the fusion with an ascogonium [13,45]. Ascospores of *H. fraxineus*, produced in the leaf litter by apothecia, are windborne and secure the dispersal and spread of the pathogen [45,46]. They are produced in abundance during several months in late spring and summer and are considered drought sensitive [10]. Furthermore, Husson et al. [47] found a positive correlation between soil moisture and the percentage of affected collar circumference caused by *H. fraxineus* in northeastern France. Additionally, Gross et al. [45] supposed that moist soil conditions could favour the survival of the pathogen on ash rachises in the litter and apothecia production. The discharge of spore has a peak in the morning [11], most likely to prevent spore desiccation and to facilitate germination [11,42]. Moreover, depending on altitude and related climatic conditions, the pathogen's apothecia first appear at the end of May, June or early July, subsequently with a different duration of dispersal [48]: in addition, the genetic intrapopulation variability of *H. fraxineus* is highly dependent on elevation, and, together, on the number of days with snow cover [49].

The artificial long-distance movement of infected ash commodities is known to have contributed to the spread of the disease [3,10], so that a Plant Health Order was introduced in Great Britain to restrict imports of ash plants and seeds to those originating in pest-free areas, despite the confirmed presence of the pathogen in a number of sites in the country [50], but little is known about the natural spread potential of *H. fraxineus* when considering habitat suitability. According to official reports [11,20], not all the distribution ranges of *F. excelsior* and *F. angustifolia* are affected yet. By means of an ensemble forecasting technique, resulting from a combination of nine distribution models, the main objective of this study was to examine the potential natural distribution of the parasite in European and neighbouring regions according to the geographical distribution of its hosts and to the main environmental features of the sites in which the natural presence of the disease was reported. Secondly, the natural spread of the pathogen was simulated, in order to consider the role of airborne spores in dispersal. In this perspective, nurseries and recent plantations that may be associated with the movement of infected plants for planting [3,10,51,52] were intentionally excluded.

Materials and Methods

Study area

The extension of the study area was based on the natural distribution maps of the three indigenous European ash species known to be susceptible to the parasite under natural conditions [8,53]. *F. excelsior* and *F. angustifolia* distribution maps were obtained from EUFORGEN and FRAXIGEN official databases with previous authorisation [54,55], imported into Quantum GIS software (QGIS Development Team) [56], and then converted into a single georeferenced shapefile corresponding to the study area. On the contrary, *F. ornus* was not included in the modelling because this species can develop limited necrotic lesions after artificial inoculation but appears to be resistant under field conditions [3].

Pathogen's presence

The greatest number of scientific reports of *H. fraxineus* were

collected in a data-set by means of a wide bibliographic study (used keywords "*Hymenoscyphus pseudoalbidus*", "*Chalara fraxinea*", "ash dieback", "dieback" or "decline of *Fraxinus excelsior*" or "European ash" or "common ash" or "*Fraxinus angustifolia*" or "narrow-leaved ash", and the combinations of these in all languages of the countries included in the study area: time frame for the research 01/06/2013-30/09/2013: publications had to be scientific papers referring to records of the disease that were spatially included in the study area: 27 final papers considered) [11,47,56-78] (authors, unpublished data). As the movement of asymptomatic, infected plants for planting is responsible for the artificial spread of the disease [3,10], reports in plantations younger than 3 years (the minimum time for confirmation that the site could be suitable for the pathogen) [79] and in nurseries were excluded from processing (16 records excluded). In this way, 252 sites with symptomatic ashes within the study area were considered (Figure 1).

This type of data can show patchy spatial coverage and some regions where the detected ash dieback had a greater recorded density than others, which was most likely derived from different sampling methods [80,81]. Moreover, these types of records are often closer to roads, rivers, coasts, towns and cities or concentrated in areas that are of more interest to collectors [74] than they would be if the survey effort were completely random [82-85]. To correct for this spatial bias, the resolution of the final study area was fixed in a 0.5° x 0.5° regular grid considering the spatial accuracy and precision of species records, according to Dungan et al. [86]. Presence points were then intersected with this grid, thereby reducing the number of presences to 177 patches containing at least one infected point [87].

Environmental variables

The predictor set included 12 environmental variables. For every predictor with a temporal scale, a subset January 1992 - December 2013 [11] was considered and monthly averages were computed in Raster Map Calculator in GRASS GIS (Grass Development Team, S. Michele all'Adige, Italy) [88]. The variables were selected for their relevance to the pathogen's biology and current main hypotheses on its life cycle, as reported in the introduction section. Precipitation, frequency of days with frost and monthly mean temperature maps were obtained from available Climatic Research Unit (CRU) time-series datasets [89-90], and the Growing Degree Days (GDD) computation was performed with a temperature threshold of -10°C [42,91]. Maximum, mean and minimum monthly temperature at a height of 2 m, surface temperature, soil moisture (0-10 cm depth) and runoff were obtained from NCEP/

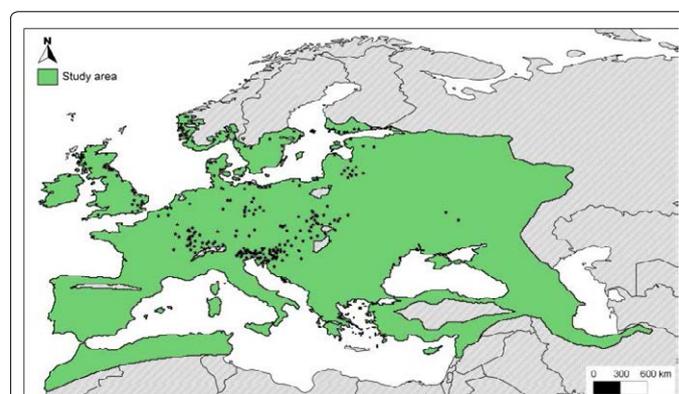


Figure 1: Study area and presences of *H. fraxineus* derived from natural infection. The area was obtained by merging the chorological maps of *F. excelsior* and *F. angustifolia*. Stars indicate the 252 localities where the presence of the pathogen was associated with a natural infection process.

NCAR Reanalysis 2 [92,93]. Snow cover and elevation maps were acquired from MODIS/Terra Snow Cover Monthly Dataset [94] and from SRTM 90 m Digital Elevation Data [95], respectively. Wind speed and direction were obtained from NCEP/NCAR Reanalysis 2 [92,93] but, after computing monthly averages in the considered temporal range, final maps showed no spatial pattern in the study area [56,96], also for the months more suitable for spores' dispersal [11]. Therefore, these predictors were discarded from further analyses.

To avoid multi-collinearity [97] and model over-fitting [98], the 122 environmental predictors (10 monthly predictors plus digital elevation data and GDD November-March, according to Table 1) were subjected to a collinearity control, based on the Pearson correlation between predictors [99] (Table 1). In this way and according to Peers et al. [100], when the correlation between two variables was statistically significant for $r > 0.85$ and $p < 0.0001$ (IBM SPSS Statistics software v. 22, International Business Machines Corp., New York, USA) [101], the most adequate predictor was selected using information about the fungal biology [22,102,103]. According to Merow et al. [104], this approach eliminates correlation and allows more parsimonious and interpretable models. Finally, the resulting maps were overlaid with the grid's study area while considering the average values in the centroids (the centre points of defined areas [105]).

Pre-processing of data

The modelling was directly trained in the whole study area as no other regions with similar environmental ranges, hosts and disease's presence have been detected to date [20], nor could a subplot represent all the considered climatic zones [106,107]. Moreover, taking into account the unavailability of absence data for the pathogen [108], background data (also referred as "pseudoabsences") were included in model construction [109-111]. Pseudoabsences were generated randomly, reducing the number of background points to three times the number of presences, according to Wisz and Guisan [112] and Barbet-Massin et al. [111]. The resulting data were then separated in three partitions in a split-sample approach (IBM SPSS Statistics software v. 22, International Business Machines Corp., New York, USA) [101,113]: training (65%, comprising 115 presences and 345 pseudoabsences) and validation (15%, with 27 presences and 80 pseudoabsences) sets were used in the construction and calibration of the individual models with the control of overfitting [114], respectively. The remaining data (test set, 20%, with 35 presences and 106 pseudoabsences) were used for comparing models [115].

Species distribution modelling procedures

In accordance with Elith et al. [109] and Guisan and Thuiller [41], more than one modelling algorithm, both classical and novel, was adopted. Methods were grouped on the basis of algorithm class into the following five categories: 1) Regression based models, 2) Classification Trees [116] and, within the machine learning community, 3) Artificial Neural Networks (ANN) [117], 4) Support Vector Machine (SVM) [118] and 5) Maxent [119,120]. The chosen regression-based models were backward stepwise logistic regression (with only main effect or with first order interactions: LOG) and a Generalised Linear Model (GLM), considering a binomial distribution, previously used extensively in species' distribution studies [121-127]. CHAID (Chi-squared Automatic Interaction Detector), belonging to the category "Classification tree", was chosen in order to take advantage of multiple splitting pathways for each grid's node [128-129] and two models were built, considering both boosting and bagging (bootstrap aggregation) procedures [130]. Boosting and bagging procedures were also

performed within the Artificial Neural Networks (ANN) Multilayer Perceptron category (MLP), which is considered more powerful than multiple regression models when modelling nonlinear relationships [121,131]. The last two machine learning algorithms used were the Support Vector Machine (SVM), recently introduced in a species distribution context [132,133], and Maxent (Maximum Entropy Model), estimating the target probability distribution by finding the probability distribution of maximum entropy [119-120,129]. Among the five categories of model construction, 1 to 4 were built in IBM SPSS Statistics software (v. 22, International Business Machines Corp., New York, USA) [101], and the SVM algorithm was implemented in LibSVM library (v. 3.17) [134], while the Maxent model was produced in Maxent software (v 3.3.3k) [119-120]. For each type of model, respective statistical parameters were calibrated in order to optimise the resulting sensitivity. Multiple runs (maximum = 50) for each model gave the distribution probability in each cell, which generated a final output with a mean predictive cell value ranging from 0 to 1.

Evaluation statistics

To evaluate the performances of the nine models, the predicted values were compared with the test set by means of contingency tables (also called a confusion matrix) [135] obtained with IBM SPSS Statistics software (v. 22, International Business Machines Corp., New York, USA) and considering the conventional threshold of 0.5 [136]: predicted relative probabilities ≥ 0.5 were classified as presence, whereas relative probabilities < 0.5 were classified as absence. The classical measures derived from the confusion matrix and calculated in Microsoft Excel (v. 2007, Microsoft Corporation, Redmond, USA) [137] were a) overall accuracy, b) specificity, c) sensitivity, d) Kohen's Kappa statistic and e) the True Skill Statistic (TSS) (Table 2) [136,138]. The Area Under the Curve (AUC) of the receiver operating characteristic (ROC) [139-142] was obtained in IBM SPSS Statistics software (v. 22, International Business Machines Corp., New York, USA), using the obtained measures as the probability that one score associated to a random presence site is higher than the probability of a random pseudoabsence site [143]. The F_{pb} index, specifically relying on presences and pseudoabsences [135], was then calculated from the contingency tables. In this way, the regions where the pathogen's establishment is possible but did not occur or was not yet detected, were not considered. For each category of model construction, only those that performed the highest F_{pb} measures were used to generate relative suitability surfaces in the study area and a qualitative comparison in Quantum GIS (QGIS Development Team) [56] was performed. A quantitative comparison was performed on the basis of the percentages of agreement among the predicted probabilities of selected models that

Chosen predictor	Monthly averages calculation	Variables correlated ($r > 0.85$, $p < 0.0001$) and discarded
Digital elevation data		
Mean temperature	*	Maximum, minimum and mean temperature at a height of 2 m; Skin temperature
Frost day frequency	*	Snow cover
Precipitation	*	
GDD November-March		Maximum, minimum and mean temperature at a height of 2 m; Skin temperature
Soil moisture (0-10 cm depth)	*	Runoff

Table 1: Shows the variables used for model building and the discarded predictors after the collinearity control based on the Pearson correlation (r). Stars indicate the selected predictors for which a monthly average was computed.

Measure	Formula
Overall accuracy	$\frac{TP + TN}{n}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Kappa statistic	$\frac{\left(\frac{TP + TN}{n}\right) - \frac{(TP + FP)(TP + FN) + (FN + TN)(TN + FP)}{n^2}}{1 - \frac{(TP + FP)(TP + FN) + (FN + TN)(TN + FP)}{n^2}}$
TSS	$Sensitivity + Specificity - 1$
F_{pb}	$\frac{2 \times TP}{TP + FN + FP}$

Table 2: Parameters used in the evaluation of individual models and the weighted average consensus model. n-Total Number of Cases; TN-True Negative; FP-False Positive; TP-True Positive; FN-False Negative [135].

were calculated on the whole dataset considering the conventional threshold of 0.5 between presence and pseudoabsence [136].

The weighted average consensus model and spatially realistic probability

Considering that the nine models gave partially different probability maps but performed very well in the comparison with the test set, rather than selecting just one as definitive, their prediction outputs were combined using a consensus modelling framework procedure [144-146]. Furthermore, this approach can enable more robust decision-making in the face of uncertainty, in particular in a conservation planning context [147]. Therefore, a Weighted Average (WA) was calculated on the previous evaluation of the selected modelling techniques but because pseudoabsences cannot be considered as confirmed *H. fraxineus* absences, instead of using conventional AUC values as weights [145,148], the new measure of F_{pb} was exploited:

$$WA_j = \frac{\sum_{ij} (F_{pb_i} \times P_{ij})}{\sum_i F_{pb_i}}$$

with P representing the predicted relative suitability [149] of the single model i for each grid cell j. The performance of the new WA consensus model was assessed with the same statistics and test set used for the individual models described above.

Because the natural spread of a pathogen is an intrinsically spatial process [150], the spatially explicit model for *H. fraxineus* in the study area was produced to identify possible suitable areas not reachable with a natural spread process, using the following procedure. According to a precautionary approach, the patches resulting as suitable for the pathogen or useful as transitional zones were selected in Quantum GIS (QGIS Development Team) [56] from the potential map obtained from the ensemble modelling technique [145,148]. This operation was made through a binary transformation and considered the threshold maximising the True Skill Statistic [111,151]. A script in R [152] was written *ad hoc* to generate a network [153-155] among neighbour polygons with a distance lower than 1.3° (approximately 120 km) and with a safety factor of one (two times the maximum spread distance indicated in the literature, given that the natural spread rate of *H.*

fraxineus can reach 60 km/year [10,156]). R code for the construction of the spatially explicit model is included as part of the Supplementary Information (Appendix S1: Map S2 for execution). The spread of the pathogen from the presence points in the network was then simulated in R (100 iterations over time). In this way, the final prediction excluded the areas potentially prone to natural spread in the WA consensus model, but not the regions gradually reachable from presence areas.

Relative importance of predictor variables

The contributions of each environmental variable to the construction of the models used in the WA consensus procedure obtained from the software were merged in a single relative importance value (OI, Overall Importance) to achieve a more readable result. This arrangement was achieved by computing a weighted average (Microsoft Excel v. 2007, Microsoft Corporation, Redmond, USA) [136] using the F_{pb} value associated with each individual model, similarly to the construction of the WA consensus model:

$$Overall\ Importance\ of\ Predictor_j = \frac{\sum_{i,j} (F_{pb_i} \times Importance\ of\ Predictor_{i,j})}{\sum_i F_{pb_i}}$$

where j represents the predictor and i the individual model.

Results

Extension of the study area

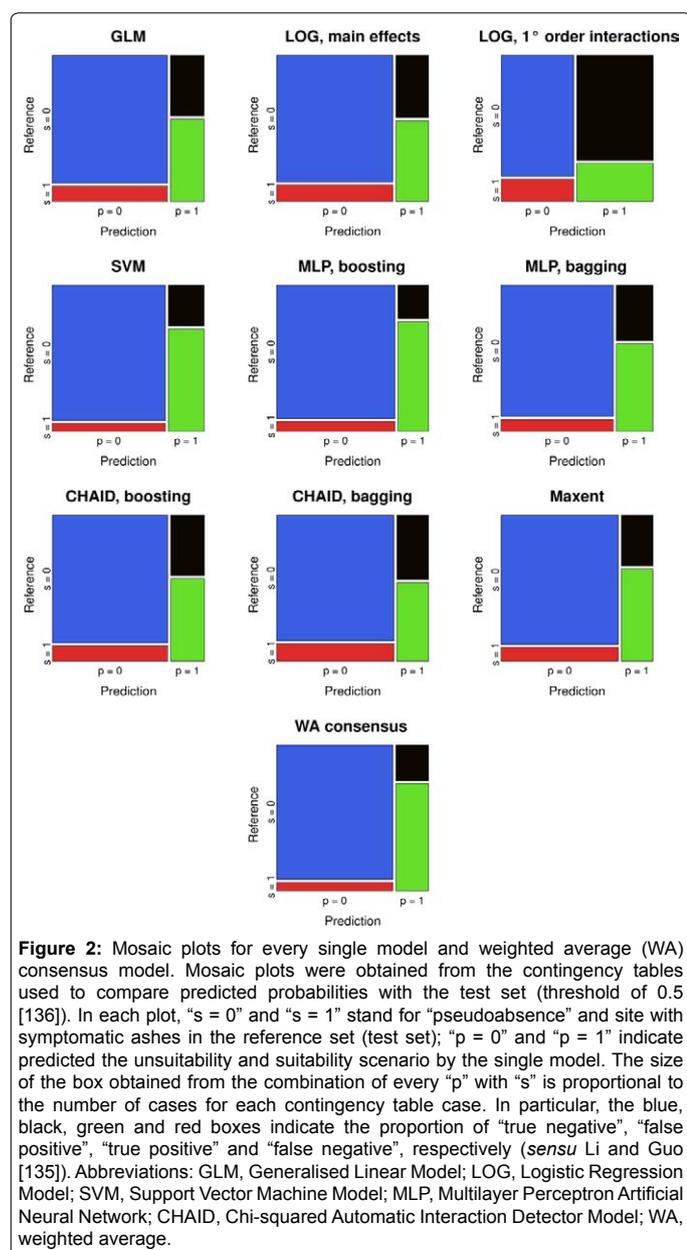
The shapefile corresponding to the study area included most of Europe plus neighbouring countries, from a northern limit in southern Scandinavia to some parts of north-western Africa and Anatolia, from Ireland and Portugal to western Russia and northern Iran (Figure 1).

Chosen predictors

The collinearity test based on the Pearson correlation allowed the number of predictors to be reduced from 122 to 50, as reported in Table 1. In particular, the maximum and minimum temperature at a height of 2 m, snow cover and runoff were discarded from further analyses. From the intersection of the maps with the grid, 4576 background samples were obtained and then reduced to 531 to allow model building.

Model fits and comparison

During model building, each type of algorithm was optimised and the best final set parameters are reported as part of the Supplementary Information in Appendix S3. The relative efficacy of the models on the test set was evaluated by comparing contingency tables (Figure 2, see Appendix S4 for a deepen explanation) and a series of parameters (Table 3). Among the models, SVM and MLP built with boosting or bagging procedures and Maxent achieved the highest measures of overall accuracy, sensitivity, Kappa statistic and TSS. In the comparison of the algorithms on the basis of the ROC curve (Figure 3), the SVM and the two MLP models were the best performing in predictive accuracy. This result was confirmed by the respective AUC values (AUC>0.9: Table 3). Considering specificity and F_{pb} , the values covered a greater range, indicating that LOG with first order interactions and CHAID bagging models performed significantly worse than the others in the same categories. As a result, the models selected for the WA consensus model for each category of construction on the basis of F_{pb} measures were GLM, SVM, CHAID built with boosting procedure, Artificial Neural Network MLP with boosting procedure, and Maxent. The WA consensus model often achieved higher performances on the basis of the evaluation parameters than the single models used for its construction.



The projections of the selected models in the QGIS software were visually different both in predicted extent and in the levels of the relative probabilities (Figure 4). In particular, the GLM forecast a wider potential area, with eastern extremes in the Moscow region (Figure 4A). In the SVM, the potential area was more restricted and had higher associated relative probabilities (Figure 4B). A similar result was obtained for the MLP, but with more irregular and fragmented areas (Figure 4C) in addition to the CHAID regression tree model (Figure 4D). The spatial pattern associated with the Maxent model was completely different and consisted of a smoother and larger potential distribution with a low relative suitability of presence, which also reached some southern zones in the study area (Figure 4E).

Although the models tended to differ in the magnitude of predicted relative probabilities, agreement was reached by all the models in highlighting the central, northern and eastern Alps, Baltic States, southern Finland and the zone including Slovakia and southern Poland

as more suitable areas for the pathogen as potential scenarios.

Considering the quantitative comparison among these models on the basis of predicted relative probabilities in the whole dataset after applying the 0.5 threshold, the percentage of agreement varied from 86.3% to 93.4%, whereas the accordance of the models with the WA consensus model achieved higher percentages (89.4% - 96.7%: Table 4).

WA consensus and spatially realistic models

The potential map of *H. fraxineus* drawn from the WA consensus model appeared to be an intermediate result in comparison with previous models (Figure 5A): the areas at major risk of spread ($p > 0.7$), such as the central and eastern Alps, Austria, Switzerland, eastern France, central Ukraine, southern and northern Poland, northeastern Germany, southern Sweden and Finland, central Denmark, southeastern England and the Baltic States, were confirmed and connected by areas of low-medium relative probabilities ($0.25 < p < 0.7$).

The final map obtained from the network analysis (threshold obtained for TSS maximisation 0.25) represented the spatially explicit distribution for *H. fraxineus* (Figure 5B), with the predicted distribution overlapping the greater part of the WA consensus map but with some patches, such as in the Pyrenees and Caucasus, which were not considered, being not gradually reachable from the potential area in Central Europe.

Environmental variables associated with the natural spread of *H. fraxineus*

Important variables in creating model fits were consistent in all models except for Maxent. Of the 50 predictors, the most important ones ($OI \geq 2$, $OI =$ Overall Importance) are reported in Table 5. Precipitation in July and August were the two most important variables, with OI values of 7.2 and 12.1, respectively. Precipitation in June and soil water content in August were also relatively strong predictors correlated with the occurrence of ash dieback ($OI = 5.7$ and 5.6), while temperature in July and August played a moderate role ($OI = 3.9$ and 3.2). In general, apart from elevation ($OI = 2$), the other predictors represented averages during the summer months. In particular, when compared to the average values of the whole study area, the presence of *H. fraxineus* was associated with a low mean temperature between June and September (16.6 °C), abundant summer precipitation (> 80 mm) and higher soil moisture content ($> 30\%$).

Discussion

Due to the rapid spread of *H. fraxineus* in Europe reported in recent years [10,20], this study was performed to provide a spatial prediction of the vulnerability of indigenous ash tree species in Europe, considering the distribution of hosts and the main environmental factors associated with naturally infected sites.

Among the nine algorithms compared, the Support Vector Machine (SVM), Artificial Neural Network Multilayer Perceptron (MLP) with boosting procedure and Maxent models achieved the highest measures of specificity, Kappa statistic, Area Under the Curve (AUC) and F_{pb} , demonstrating that they fit the test set better than the other models, which allows projections of observed patterns into independent situations and minimises over-fitting of data [157]. Sensitivity, an essential measure in models with presences and pseudoabsences, was significantly higher in the Generalised Linear Model (GLM), SVM and MLP Artificial Neural Network models. The generally higher performance of machine-learning methods, most likely due to peculiar advantages, such as robust parameter estimates, model structure

Model	Overall accuracy	Specificity	Sensitivity	Kappa statistic	TSS	AUC	F _{pb}
GLM	0.81	0.83	0.71	0.49	0.54	0.87	0.88
LOG, main effects	0.81	0.88	0.58	0.45	0.46	0.88	0.80
LOG, 1° order interactions	0.55	0.52	0.65	0.11	0.17	0.56	0.47
SVM	<i>0.89</i>	<i>0.91</i>	<i>0.81</i>	<i>0.69</i>	<i>0.72</i>	<i>0.92</i>	1.22
MLP, boosting	<i>0.90</i>	<i>0.94</i>	<i>0.74</i>	<i>0.69</i>	<i>0.68</i>	<i>0.92</i>	1.21
MLP, bagging	0.84	0.88	0.71	0.56	0.59	0.92	0.98
CHAID, boosting	0.81	0.88	0.67	0.49	0.54	0.85	0.88
CHAID, bagging	0.81	0.88	0.55	0.43	0.43	0.83	0.76
Maxent	<i>0.85</i>	<i>0.90</i>	0.65	0.55	0.55	0.90	0.95
WA consensus model	<i>0.90</i>	<i>0.93</i>	0.77	0.70	0.70	0.94	1.23

Table 3: Performances of the individual models and the weighted average consensus model.

Performances were computed on the test set considering overall accuracy, specificity, sensitivity, Kappa statistic, True Skill Statistic (TSS), the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) and F_{pb} (Table 2). The best four values for each parameter are italicised; the bold values indicate the best model for each category (according to those presented in the Materials and Methods section) on the basis of F_{pb} measures. Abbreviations: GLM: Generalised Linear Model; LOG: Logistic Regression Model; SVM: Support Vector Machine Model; MLP: Multilayer Perceptron Artificial Neural Network; CHAID: Chi-squared Automatic Interaction Detector Classification Tree; WA: Weighted Average.

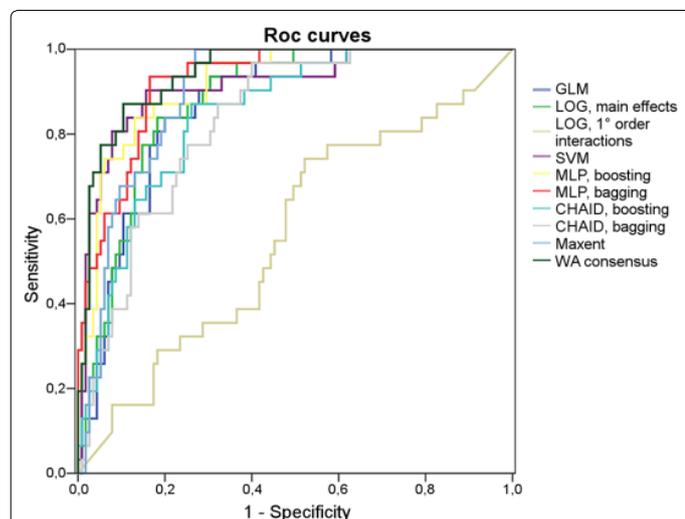


Figure 3: ROC curves for the individual models and for the WA consensus model. Sensitivity is plotted against the corresponding proportion of false positives (1-specificity), at various threshold settings. Abbreviations: GLM, Generalised Linear Model; LOG, Logistic Regression Model; SVM, Support Vector Machine Model; MLP, Multilayer Perceptron Artificial Neural Network; CHAID, Chi-squared Automatic Interaction Detector Classification Tree; WA, weighted average

learned from data and easy fitting of complex interactions, in spite of considering the use of pseudoabsences in models evaluation [110], was therefore confirmed [158].

Among the performance measures considered, prominence was given to the F_{pb} statistic. This accuracy assessment was recently proposed by Li and Guo [135] for presence-only modelling, to specifically consider presences and pseudoabsences instead of true absences in the confusion matrix [159]. Given that the performance of such models with regard to F_{pb} were quite robust, but that their predictive maps partially differed in the extension and magnitude of relative suitability of the pathogen's presence, in accordance with Stohlgren et al. [160], the consensus ensemble forecast was calculated as the weighted average of the best models, highlighting the areas of agreement among models as expected and thereby minimising the weakness of any given algorithm [144,147-148]. The resulting model, highlighting the areas suitable for the pathogen, generally outperformed any single algorithm based on the evaluation parameters (mainly Kappa statistic, AUC and F_{pb}) and suggested a potential distribution map with higher risk for the central and eastern Alps, Baltic States, southern Finland and Sweden,

Slovakia and southern Poland. This approach can enable more robust decision-making in the face of uncertainty [147], however, as suggested by Elith et al. [161], caution should be taken in selecting models for an ensemble forecast. An understanding of the data, single models and predictions should not be underestimated, especially in the case of a climate change context.

Species Distribution Models (SDMs) and ensemble forecasting lead to interesting conclusions on the ecological appropriateness of some areas to the potential pathogenic spread of *H. fraxineus* [40]. To take into account the dispersal limitation [162] and obtain a more realistic projection, a novel approach to the network analysis was implemented, which considered the potential map obtained from the ensemble modelling technique and the points where the disease presence can be considered as derived from a natural infection. In this way, most of the edging areas of the *F. excelsior* chorological map resulted as unsuitable for a natural spread in the final scenario (i.e., western Ireland, part of France and northern Spain, all the southern areas in central Italy,

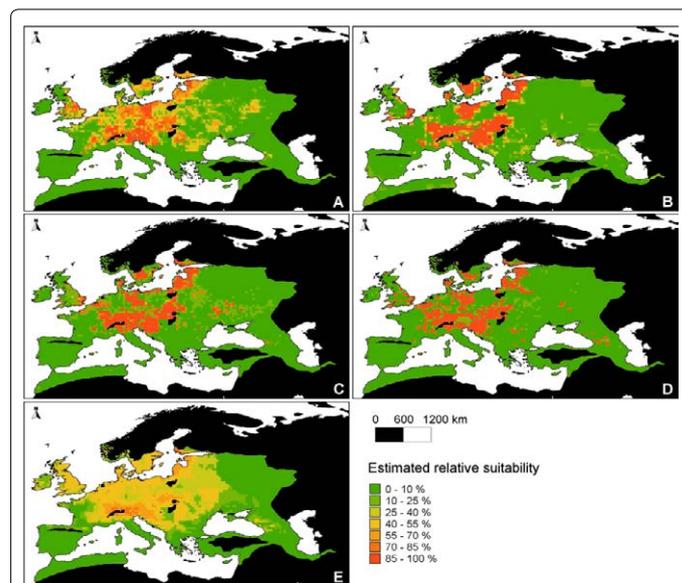
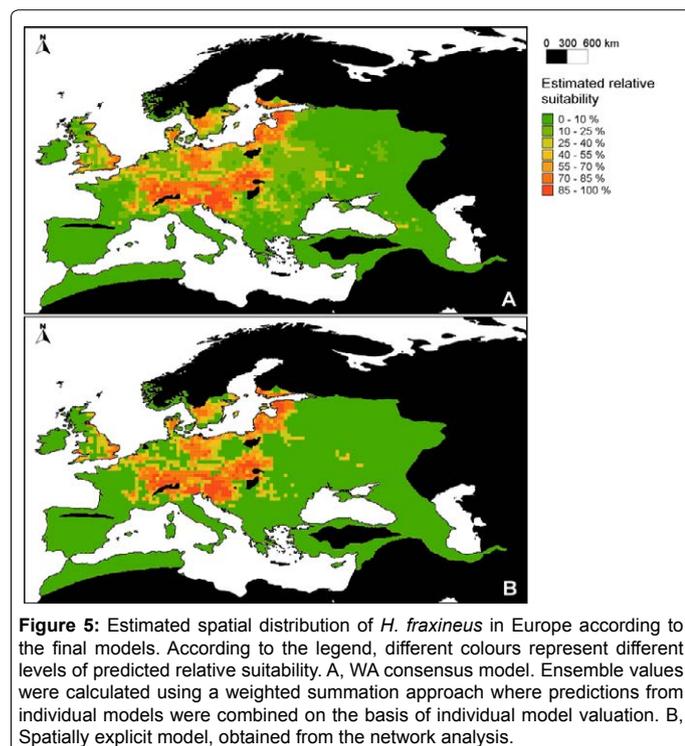


Figure 4: Estimated spatial distribution of *H. fraxineus* in Europe according to the individual models. According to the legend, different colours represent different levels of predicted relative suitability. A, Generalised Linear Model (GLM). B, Support Vector Machine (SVM). C, Artificial Neural Networks Multilayer Perceptron (MLP), with boosting building. D, Chi-squared Automatic Interaction Detector (CHAID) Classification Tree, with boosting procedure. E, Maxent.



the Balkans and northern Turkey and western areas in Russia, the Caucasus and Iran). The reported consensus ensemble forecast potential distribution map may therefore indicate the areas where the trade of ash species should be under particular control. In any event, caution should be taken in transferring predicted results from a wider scale to a more local scale [163,164]. The disease in Europe most likely originated from a single introduction event of the pathogen of at least two individuals with compatible mating types and was first observed in the early 1990s in Poland [165], although there is an hypothesis of the introduction of the pathogen together with the importation of *F. mandshurica* to Estonia [166]. The computed network analysis based on current presence points aimed to deliver realistic predictions, so the potential distribution in the case that the disease originated in others regions was not implemented.

The expected spread to *F. angustifolia* largely resulted as low, except for some areas in northern Italy, Slovenia, Croatia, Bosnia-Herzegovina and Romania, confirming the hypothesis of Gross et al. [10], who reported that the epidemic rate in Europe is slowing down, the sill of a sigmoid function of spatial growth against time has been reached. Similar studies on invasive pest modelling suggest that environmental conditions may serve as a constraint limiting the spread in respect to the hosts' distribution. For instance, Podger et al. [167] considered the establishment of *Phytophthora cinnamomi* in potential areas with annual mean temperature < 7.5°C and annual mean rainfall < 600 mm as unlikely. Moreover, Koch and Smith [168] estimated the potential spread of non-native *Xyleborus glabratus* in the southeastern U.S. and concluded that climatic conditions could prevent the spread from coastal plain to eastern interior forests.

As a result of the modelling, precipitation, soil moisture and air temperature were shown to be significantly more influential than other predictors in model building of the potential distribution of *H. fraxineus*. In particular, the species distribution appeared to be highly dependent on abundant rainfall and high soil moisture content in

the summer months, confirming the hypothesis of more intense ash dieback near water courses or in high soil moisture sites [169,170] and supporting the hypothesis of Gross et al. [45] about the importance of free water for the fertilisation of the *H. fraxineus* anamorph on petioles. Summer mean temperatures were also relevant for model construction of the pathogen niche, consistent with available information on the species [16]. In an ecological and biological interpretation, the temperature ranges highlighted by the models could be necessary for apothecia production, known to occur from May to October, with a peak in July, with a minimum temperature of 1.1°C and optimum temperature of 22°C [11,171]. Considering the low January mean temperature in areas where the species was present (-0.5°C), various studies indicate that the fungus can develop within the plants over the winter, causing necrosis [19], and the mycelium tolerates freezing at -20°C for at least two months and can even survive -70°C for at least one month [10]. In addition, conidial sporulation is favoured *in vitro* by temperatures between 5 and 15°C [43] and was observed in nature in autumn on ash rachises in the ground litter [53].

Modelling was directly trained on the entire study area in order to

Model	GLM	SVM	MLP, boosting	CHAID, boosting	Maxent	WA consensus model
GLM	-	88.4	88.8	86.3	87.9	89.4
SVM	88.4	-	93.4	91.7	89.9	96.3
MLP, boosting	88.8	93.4	-	91.6	90.0	96.7
CHAID, boosting	86.3	91.7	91.6	-	89.1	94.4
Maxent	87.9	89.9	90.0	89.1	-	91.5
WA consensus model	89.4	96.3	96.7	94.4	91.5	-

Table 4: Percentages of agreement in relative probabilities predicted by selected individual models and the weighted average consensus model on the whole dataset.

The table reports the agreement among relative probabilities predicted by the best models for each category chosen according to the F_{pb} measures reported in Table 3. Predicted relative probabilities are rounded to 0 (pseudoabsence) or 1 (presence) using the conventional threshold of 0.5 [136] before percentages' computation and comparison.

Environmental predictor	Overall importance	Mean value in presence dataset
Elevation	2.0	359.2 m
Mean temperature, January	2.3	-0.5 °C
Mean temperature, June	2.3	16.8 °C
Mean temperature, July	3.9	18.1 °C
Mean temperature, August	3.2	17.9 °C
Mean temperature, September	2.8	13.7 °C
Frost day frequency December	2.6	21.7 °C
GDD November-March	2.0	1743.1 °C
Precipitation, May	2.5	77.8 mm
Precipitation, June	5.7	91.4 mm
Precipitation, July	7.2	100.4 mm
Precipitation August	12.1	93.8 mm
Precipitation, September	2.0	82.1 mm
Soil moisture, March	2.2	0.35
Soil moisture, July	3.3	0.31
Soil moisture, August	5.6	0.30

Table 5: Overall Importance (OI) of environmental predictors included in model building.

Variables with OI>2 are shown; predictors with OI>6 are in bold, italicised text indicates the variables with 3<OI<6. The average value for each predictor considering the dataset of *H. fraxineus* presences are displayed in the right hand column.

Abbreviation: GDD: Growing Degree Days.

cover all the environmental gradients in the distributions of *F. excelsior* and *F. angustifolia* and to avoid underestimating climatic factors in delimiting species' distribution [106]. Moreover, attention must be paid in considering relative suitability predictions of the models, because of the possible lack of equilibrium, which is typical of an invasive pathogen not yet reaching its full potential distribution [161]. For this reason, further reports of ash dieback over time, including the probable original Asian distribution [5,172], could be easily taken into account to enlarge the boundaries of the Grinnellian niche (closer to equilibrium, according to Pulliam [173-175]). The availability of a wider time series, including data on ash dieback severity and host abundance, will allow the consideration of the spread dynamics of the disease in the context of different landscapes and in a climate change scenario [175-177]. More detailed mathematical analyses are in progress, to identify the specific high performance components in the machine learning models able to describe the biological and ecological complex interactions involved in the expression of the disease.

Acknowledgments

This work was financially supported by the Land Environment Resources and Health (L.E.R.H.) doctoral course (<http://intra.tesaf.unipd.it/school/lerh.asp>: University of Padova) and by the University of Padova ("ex-60% 2012"). The authors thank Dr. G. Frigimelica and D. Jurc for their kind cooperation, the FRAXIGEN project and the EUFORGEN programme for the license to use the ash distribution maps, and the members of the FPS COST Action FP1103 (Fraxback).

References

- Kowalski T (2006) *Chalara fraxinea* sp. nov. associated with dieback of ash (*Fraxinus excelsior*) in Poland. For Pathol 36: 264-270.
- Queloz V, Grünig CR, Berndt R, Kowalski T, Sieber TN, et al. (2010) Cryptic speciation in *Hymenoscyphus albidus*. For Pathol 41: 133-142.
- Pautasso M, Aas G, Queloz V, Holdenrieder O (2013) European ash (*Fraxinus excelsior*) dieback - A conservation biology challenge. Biol Conserv 158: 37-49.
- Baral HO, Queloz V, Hosoya T (2014) *Hymenoscyphus fraxineus*, the correct scientific name for the fungus causing ash dieback in Europe. IMA Fungus 5: 79-80.
- Zhao YJ, Hosoya T, Baral HO, Hosaka K, Kakishima M (2012) *Hymenoscyphus pseudoalbidus*, the correct name for *Lambertella albida* reported from Japan. Mycotaxon 122: 25-41.
- Przybyl K (2002) Fungi associated with necrotic apical parts of *Fraxinus excelsior* shoots. For Pathol 32: 387-394.
- Drenkhan R, Hanso M (2010) New host species for *Chalara fraxinea*. New Dis Rep 22: 16.
- Kirisits T, Matlakova M, Mottinger-Kroupa S, Halmschlager E, Lakatos F (2010) *Chalara fraxinea* associated with dieback of narrow-leaved ash (*Fraxinus angustifolia*). Plant Pathol 59: 411.
- Webber J, Hendry S (2012) Rapid assessment of the need for a detailed Pest Risk Analysis for *Chalara fraxinea*. Forest Research, Forestry Commission, UK.
- Gross A, Holdenrieder O, Pautasso M, Queloz V, Sieber TN (2014) *Hymenoscyphus pseudoalbidus*, the causal agent of European ash dieback. Mol Plant Pathol 15: 5-21.
- Timmermann V, Børja I, Hietala AM, Kirisits T, Solheim H (2011) Ash dieback: Pathogen spread and diurnal patterns of ascospore dispersal, with special emphasis on Norway. Bull OEPP 41: 14-20.
- Cleary MR, Daniel G, Stenlid J (2013) Light and scanning electron microscopy studies of the early infection stages of *Hymenoscyphus pseudoalbidus* on *Fraxinus excelsior*. Plant Pathol 62: 1294-1301.
- Schumacher J, Kehr R, Leonhard S (2009) Mycological and histological investigations of *Fraxinus excelsior* nursery saplings naturally infected by *Chalara fraxinea*. For Pathol 40: 419-429.
- Dal Maso E, Fanchin G, Mutto Accordi S, Scattolin L, Montecchio L (2012) Ultrastructural modifications in Common ash tissues colonised by *Chalara fraxinea*. Phytopathol Mediterr 51: 599-606.
- Cooke L, Fleming C, McCracken A (2013) DARD E&I project 12/3/S7: efficacy of biocides, disinfectants and other treatments to limit the spread of ash dieback caused by *Chalara fraxinea*. Agri-Food and Biosciences Institute.
- Hauptman T, Piškur B, de Groot M, Ogris N, Fertan M, et al. (2013) Temperature effect on *Chalara fraxinea*: heat treatment of saplings as a possible disease control method. For Pathol 43: 360-370.
- Dal Maso E, Cocking J, Montecchio L (2014) Efficacy tests on commercial fungicides against ash dieback in vitro and by trunk injection. Urban For Urban Gree: in press.
- Hauptman T, Aco Celar F, de Groot M, Jurc D (2014) Application of fungicides and urea for control of ash dieback. iForest: in press.
- Sansford CE (2013) Pest Risk Analysis for *Hymenoscyphus pseudoalbidus* for the UK and the Republic of Ireland. Forestry Commission, Bristol.
- EPPO (2014) *Hymenoscyphus pseudoalbidus* (CHAAFR). EPPO Global Database.
- Paul PA, Munkvold GP (2005) Regression and artificial neural network modeling for the prediction of gray leaf spot of maize. Phytopathology 95: 388-396.
- Dupin M, Reynaud P, Jarošik V, Baker R, Brunel S, et al. (2011) Effects of the training dataset characteristics on the performance of nine species distribution models: Application to *Diabrotica virgifera virgifera*. PLoS ONE 6: e20957.
- Meentemeyer RK, Cunniffe NJ, Cook AR, Filipe JAN, Hunter RD, et al. (2011) Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990-2030). Ecosphere 2: 1-24.
- Jönsson MT, Thor G (2012) Estimating coextinction risks from epidemic tree death: affiliate lichen communities among diseased host tree populations of *Fraxinus excelsior*. PLoS One 7: e45701.
- Santini A, Ghelardini L, De Pace C, Desprez-Loustau ML, Capretti P, et al. (2013) Biogeographical patterns and determinants of invasion by forest pathogens in Europe. New Phytol 197: 238-250.
- Löhmus A, Runnel K (2014) Ash dieback can rapidly eradicate isolated epiphyte populations in production forests: A case study. Biol Conserv 169: 185-188.
- Van Maanen A, Xu XM (2003) Modelling plant disease epidemics. Eur J Plant Pathol 109: 669-682.
- Meentemeyer R, Rizzo D, Mark W, Lotz E (2004) Mapping the risk of establishment and spread of sudden oak death in California. For Ecol Manage 200: 195-214.
- Bergot M, Cloppet E, Péramaud V, Déqué M, Marçais B, et al. (2004) Simulation of potential range expansion of oak disease caused by *Phytophthora cinnamomi* under climate change. Glob Chang Biol 10: 1539-1552.
- Auclair, Allan ND, Heilman WE, Brinkman B (2010) Predicting forest dieback in Maine, USA: a simple model based on soil frost and drought. Can J For Res 40: 687-702.
- Sturrock RN, Frankel SJ, Brown AV, Hennon PE, Kliejunas JT, et al. (2011) Climate change and forest diseases. Plant Pathol 60: 133-149.
- Venette RC, Cohen SD (2006) Potential climatic suitability for establishment of *Phytophthora ramorum* within the contiguous United States. For Ecol Manage 231: 18-26.
- Klopfenstein NB, Kim M-S, Hanna JW, Richardson BA, Lundquist JE (2009) Approaches to predicting potential impacts of climate change on forest disease: an example with *Armillaria* root disease. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
- Pukkala T, Mõykkynen T, Thor M, Rönnerberg J, Stenlid J (2005) Modeling infection and spread of *Heterobasidion annosum* in even-aged Fennoscandian conifer stands. Can J For Res 35: 74-84.
- BenDor TK, Metcalf SS, Fontenot LE, Sangunett B, Hannon B (2006) Modeling the spread of the Emerald Ash Borer. Ecol Modell 197: 221-236.
- Ganley RJ, Watt MS, Manning L, Iturriza E (2009) A global climatic risk assessment of pitch canker disease. Can J For Res 39: 2246-2256.
- Kelly M, Guo Q, Liu D, Shaari D (2007) Modeling the risk for a new invasive forest disease in the United States: an evaluation of five environmental niche models. Comput Environ Urban Syst 31: 689-710.
- Robinet C, Kehlenbeck H, Kriticos DJ, Baker RHA, Battisti A, et al. (2012) A suite of models to support the quantitative assessment of spread in Pest Risk Analysis.

39. Kamino LH, Stehmann JR, Amaral S, De Marco P Jr, Rangel TF, et al. (2012) Challenges and perspectives for species distribution modelling in the neotropics. *Biol Lett* 8: 324-326.
40. Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol* 78: 419-433.
41. Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8: 993-1009.
42. Hietala AM, Timmermann V, Børia I, Solheim H (2013) The invasive ash dieback pathogen *Hymenoscyphus pseudoalbidus* exerts maximal infection pressure prior to the onset of host leaf senescence. *Fungal Ecol* 6: 302-308.
43. Kowalski T, Bartnik C (2010) Morphological variation in colonies of *Chalara fraxinea* isolated from ash (*Fraxinus excelsior* L.) stems with symptoms of dieback and effects of temperature on colony growth and structure. *Acta Agrobot* 63: 99-106.
44. McKinney LV, Thomsen IM, Kjær ED, Nielsen LR (2012) Genetic resistance to *Hymenoscyphus pseudoalbidus* limits fungal growth and symptom occurrence in *Fraxinus excelsior*. *For Pathol* 42: 69-74.
45. Gross A, Zaffarano PL, Duo S, Grünig CR (2012) Reproductive mode and life cycle of the ash dieback pathogen *Hymenoscyphus pseudoalbidus*. *Fungal Genet Biol* 49: 977-986.
46. Bengtsson SBK, Vasaitis R, Kirisits T, Solheim H, Stenlid J (2012) Population structure of *Hymenoscyphus pseudoalbidus* and its genetic relationship to *Hymenoscyphus albidus*. *Fungal Ecol* 5: 147-153.
47. Husson C, Caël O, Grandjean JP, Nageleisen LM, Marçais B (2012) Occurrence of *Hymenoscyphus pseudoalbidus* on infected ash logs. *Plant Pathol* 61: 889-895.
48. Kräutler K, Kirisits T (2012) The ash dieback pathogen *Hymenoscyphus pseudoalbidus* is associated with leaf symptoms on ash species (*Fraxinus* spp.). *J Agr Ext Rural Dev* 4: 261-265.
49. Kraj W, Kowalski T (2013) Genetic variability of *Hymenoscyphus pseudoalbidus* on ash leaf rachises in leaf litter of forest stands in Poland. *J Phytopathol* 162: 218-227.
50. Elith J, Leathwick JR (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annu Rev Ecol Syst* 40: 677-697.
51. Harwood TD, Xu X, Pautasso M, Jeger MJ, Shawa MW (2009) Epidemiological risk assessment using linked network and grid based modelling: *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK. *Ecol Modell* 220: 3353-3361.
52. Great Britain Forestry Commissioners (2012) Plant Health (Forestry) (Amendment) Order 2012 (S.I. No. 2707 of 2012).
53. Kirisits T, Matlakova M, Mottinger-Kroupa S, Cech TL, Halmeschlager E (2009) The current situation of ash dieback caused by *Chalara fraxinea* in Austria. Proceedings of the conference of IUFRO working party. SDU Faculty of Forestry Journal A: 97-119.
54. EUFORGEN (2013) European Forest Genetic Resources Programme. Biodiversity International.
55. FRAXIGEN Research Project (2013) Ash for the future: defining European ash populations for conservation and regeneration. Supported by European Commission under the Fifth Framework Programme.
56. QGIS Development Team (2013) QGIS Geographic Information System. Open Source Geospatial Foundation Project.
57. Schumacher J, Wulf A, Leonhard S (2007) Erster Nachweis von *Chalara fraxinea* T. Kowalski sp. nov. in Deutschland – ein Verursacher neuartiger Schäden an Eschen. *Nachr Dtsch Pflanzenschutz* 59: 121-123.
58. Kirisits T (2008) Eschenpathogen *Chalara fraxinea* nun auch in Kärnten nachgewiesen. *Fortschutz Aktuell* 45: 28-30.
59. Bakys R, Vasaitis R, Barklund P, Thomsen IM, Stenlid J (2009) Occurrence and pathogenicity of fungi in necrotic and non-symptomatic shoots of declining common ash (*Fraxinus excelsior*) in Sweden. *Eur J For Res* 128: 51-60.
60. Kowalski T (2009) Expanse of *Chalara fraxinea* fungus in terms of ash dieback in Poland. *Sylvan* 153: 668-674.
61. EPPO (2010) *Chalara fraxinea* occurs in Lithuania.
62. Jankovský L, Holdenrieder O (2009) *Chalara fraxinea* – ash dieback in the Czech Republic. *Plant Prot Sci* 45: 74-78.
63. Kowalski T, Czekaj A (2010) Symptomy chorobowe i grzyby na zamierających jesionach (*Fraxinus excelsior* L.) w drzewostanach Nadleśnictwa Staszów. *Leśne Prace Badawcze* 71: 357-368.
64. Chandelier A, Delhaye N, Helson M (2011) First report of the ash dieback pathogen *Hymenoscyphus pseudoalbidus* (Anamorph *Chalara fraxinea*) on *Fraxinus excelsior* in Belgium. *Plant Dis* 95: 220.
65. Husson C, Scala B, Caël O, Frey P, Feau N, et al. (2011) *Chalara fraxinea* is an invasive pathogen in France. *Eur J Plant Pathol* 130: 311-324.
66. Kunca A (2011) Occurrence of Pest Agents in Slovak Forests in 2010 and Prognosis for 2011. Forest Research Institute, Zvolen.
67. Rytkönen A, Lilja A, Drenkhan R, Gaitnieks T, Hantula J (2011) First record of *Chalara fraxinea* in Finland and genetic variation among isolates sampled from Åland, mainland Finland, Estonia and Latvia. *For Pathol* 41: 169-174.
68. Witzel G, Metzler B (2011) Eschentriebsterben in Stangen- und Baumhölzern. *AFZ-Der Wald* 1: 24-27.
69. Baric L, Županic M, Pernek M, Diminic D (2012) First records of *Chalara fraxinea* in Croatia – a new agent of ash dieback (*Fraxinus* spp.) (Prvi nalazi patogene gljive *Chalara fraxinea* u Hrvatskoj – novog uzročnika odumiranja jasena (*Fraxinus* spp.)). *Šumarski List* 136: 461-468.
70. Baumann VM, Matschulla F, Helbig R (2012) Das Eschentriebsterben in Sachsen. *AFZ-DerWald* 3: 12-17.
71. Conedera M, Engesser R, Maresi G (2012) *Chalara fraxinea*: nuova minaccia per il bosco ticinese? *Agricoltura Ticinese* 144: 10.
72. Koltay A, Szabó I, Janik G (2012) *Chalara fraxinea* incidence in Hungarian ash (*Fraxinus excelsior*) forests. *Journal of Agricultural Extension and Rural Development* 4: 236-238.
73. Lenz H, Strasser L, Petercord R (2012) Eschentriebsterben begünstigt Auftreten sekundärer Schadorganismen. *Fortschutz Aktuell* 54: 26-28.
74. Stenström A, Bengtsson V, Finsberg C (2012) Askskottsjuka, ett nytt hot mot våra skyddsvärda träd. Länsstyrelsen i Västra Götalands län, naturvårdsenheten, Göteborg.
75. Županic M, Baric L, Pernek M, Diminic D (2012) Distribution of fungi *Chalara fraxinea* in Croatia (Rasprostranjenost gljive *Chalara fraxinea* u Hrvatskoj). *Radovi* 44: 125-134.
76. Bakys R, Vasaitis R, Skovsgaard JP (2013) Patterns and severity of crown dieback in young even-aged stands of European ash (*Fraxinus excelsior* L.) in relation to stand density, bud flushing phenotype, and season. *Plant Prot Sci* 49: 120-126.
77. Davydenko K, Vasaitis R, Stenlid J, Menkis A (2013) Fungi in foliage and shoots of *Fraxinus excelsior* in eastern Ukraine: a first report on *Hymenoscyphus pseudoalbidus*. *For Pathol* 43: 462-467.
78. Forestry Commission (2013) Map 2b: *Chalara fraxinea* - confirmed infection sites.
79. Varstvo gozdov Slovenije (2013) Jesenov ožig. Informacijsko središče za varstvo gozdov v Sloveniji.
80. Newbold T (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Prog Phys Geogr* 34: 3-22.
81. Loiselle BA, Jørgensen PM, Consiglio T, Jiménez I, Blake JG, et al. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J Biogeogr* 35: 105-116.
82. Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *J Biogeogr* 30: 1719-1727.
83. Hijmans RJ, Garrett KA, Huamán Z, Zhang DP, Schreuder M, et al. (2000) Assessing the geographic representativeness of Genebank collections: the case of Bolivian wild potatoes. *Conserv Biol* 14: 1755-1765.
84. Soberón JM, Llorente JB, Oñate L (2000) The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. *Biodivers Conserv* 9: 1441-1446.
85. Funk VA, Richardson KS (2002) Systematic data in biodiversity studies: use it or lose it. *Syst Biol* 51: 303-316.
86. Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, et al. (2002) A balanced view of scale in spatial statistical analysis. *Ecography* 25: 626-640.

87. Farina A (2001) Ecologia del paesaggio. In: Principi, metodi e applicazioni, UTET Libreria Srl, Torino.
88. GRASS Development Team (2013) Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation Project.
89. Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol 25: 693-712.
90. Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int J Climatol 34: 623-642.
91. Snyder RL (1985) Hand calculating degree-days. Agric For Meteorol 35: 353-358.
92. Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, et al. (1996) The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77: 437-470.
93. Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo JJ, et al. (2002) NCEP-DEO AMIP-II Reanalysis (R-2). Bulletin of the American Meteorological Society 83: 1631-1643.
94. National Operational Hydrologic Remote Sensing Center (2004) Snow Data Assimilation System (SNODAS) Data Products at NSIDC. National Snow and Ice Data Center, Boulder.
95. Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled SRTM for the Globe Version 4.
96. Fink D, Hochachka WM, Zuckerberg B, Winkler DW, Shaby B, et al. (2010) Spatiotemporal exploratory models for broad-scale survey data. Ecol Appl 20: 2131-2147.
97. Graham MH (2003) Confronting multicollinearity in ecological multiple regression. Ecology 84: 2809-2815.
98. Peterson AT, Papes M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30: 550-560.
99. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27-46.
100. Peers MJ, Thornton DH, Murray DL (2013) Evidence for large-scale effects of competition: niche displacement in Canada lynx and bobcat. Proc Biol Sci 280: 20132495.
101. IBM Corp. Released (2013) IBM SPSS Statistics for Windows, Version 22.0. International Business Machines Corporation, New York.
102. Jopp F, Reuter H, Breckling B (2011) An Introduction into Ecological Modelling for Students, Teachers & Scientists. In: Modelling Complex Ecological Dynamics. Springer-Verlag, Berlin Heidelberg.
103. Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, et al. (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. Ecography 36: 971-983.
104. Merow C, Smith MJ, Jr Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36: 1058-1069.
105. de Smith MJ, Goodchild MF, Longley PA (2007) A comprehensive guide to principles, techniques and software tools. In: Geospatial analysis, Troubador Publishing Ltd., Leicester, UK.
106. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, et al. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Modell 222: 1810-1819.
107. Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, et al. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecol Modell 263: 10-18.
108. Václavík T, Meentemeyer RK (2009) Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecol Modell 220: 3248-3258.
109. Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.
110. Cheffaoui RM, Lobo JM (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. Ecol Modell 210: 478-486.
111. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3: 327-338.
112. Wisz M, Guisan A (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol 9: 9.
113. Van Houwelingen JC, Le Cessie S (1990) Predictive value of statistical models. Stat Med 9: 1303-1325.
114. Sherriff A, Ott J, ALSPAC Study Team (2004) Artificial neural networks as statistical tools in epidemiological studies: analysis of risk factors for early infant wheeze. Paediatr Perinat Epidemiol 18: 456-463.
115. Nelles O (2001) From Classical Approaches to Neural Networks and Fuzzy Models. In: Nonlinear System Identification, Springer-Verlag, Berlin Heidelberg.
116. Rokach L, Maimon O (2008) Theory and applications. In: Data mining with decision trees World Scientific Publishing Co. Pte. Ltd, Singapore.
117. Priddy KL, Keller PE (2005) An introduction. In: Artificial Neural Networks, The Society of Photo-Optical Instrumentation Engineers, Washington.
118. Wang L (2005) Theory and applications. In: Support Vector Machines, Springer-Verlag, Berlin Heidelberg.
119. Phillips SJ, Dudik M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. Proceedings of the 21st International Conference on Machine Learning. ACM Press, New York.
120. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190: 231-259.
121. Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Modell 135: 147-186.
122. Rushton SP, Ormerod SJ, Kerby G (2004) New paradigms for modelling species distributions? J Appl Ecol 41: 193-200.
123. Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. J Biogeogr 31: 1555-1568.
124. Elith J, Graham CH (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32: 66-77.
125. Rupprecht F, Oldeland J, Finckh M (2011) Modelling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modelling approaches? J Veg Sci 22, Special feature: Ecoinformatics and Global Change: 647-659.
126. Zurell D, Grimm V, Rossmannith E, Zbinden N, Zimmermann NE (2011) Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps. Ecography 35: 590-603.
127. Smith JP, Hoffman JT (2001) Site and stand characteristics related to white pine blister rust in high-elevation forests of southern Idaho and western Wyoming. West N Am Nat 61: 409-416.
128. Scarnati L, Attorre F, De Sanctis M, Farcomeni A, Francesconi F, et al. (2009) A multiple approach for the evaluation of the spatial distribution and dynamics of a forest habitat: the case of Apennine beech forests with *Taxus baccata* and *Ilex aquifolium*. Biodivers Conserv 18: 3099-3113.
129. Clark JT, Fei S, Liang L, Rieske LK (2012) Mapping eastern hemlock: Comparing classification techniques to evaluate susceptibility of a fragmented and valued resource to an exotic invader, the hemlock woolly adelgid. For Ecol Manage 266: 216-222.
130. De'ath G (2007) Boosted trees for ecological modeling and prediction. Ecology 88: 243-251.
131. Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, et al. (2009) Using species distribution models to predict new occurrences for rare plants. Divers Distrib 15: 565-576.
132. Guo Q, Kelly M, Graham CH (2005) Support vector machines for predicting distribution of Sudden Oak Death in California. Ecol Modell 182: 75-90.
133. Drake JM, Randin C, Guisan A (2006) Modelling ecological niches with support vector machines. J Appl Ecol 43: 424-432.

134. Chih-Chung C, Chih-Jen L (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol 2*: 27:1-27:27.
135. Li W, Guo Q (2013) How to assess the prediction accuracy of species presence-absence models without absence data? *Ecography 36*: 788-799.
136. Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography 28*: 385-393.
137. Microsoft (2007) Microsoft Excel. Redmond, Washington.
138. Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: revalence, kappa and the true skill statistic (TSS). *J Appl Ecol 43*: 1223-1232.
139. Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *J Appl Ecol 38*: 921-931.
140. Brotons L, Thuiller W, Araújo MB, Hirzel AH (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography 27*: 437-448.
141. Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, et al. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Glob Chang Biol 11*: 2234-2250.
142. Peterson AT, Papes M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol Modell 213*: 63-72.
143. Elith J (2000) Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Ferson S, Burgman M (editors), *Quantitative Methods for Conservation Biology*, Springer New York, New York.
144. Araújo MB, Whittaker RJ, Ladle RJ, Erhard M (2005) Reducing uncertainty in projections of extinction risk from climate change. *Glob Ecol Biogeogr 14*: 529-538.
145. Marmion M, Hjort J, Thuiller W, Luoto M (2009) Statistical consensus methods for improving predictive geomorphology maps. *Comput Geosci 35*: 615-625.
146. Grenouillet G, Buisson L, Casajus N, Lek S (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography 34*: 9-17.
147. Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol 22*: 42-47.
148. Marmion M, Parviainen M, Luoto L, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. *Divers Distrib 15*: 59-69.
149. Phillips SJ, Elith J (2013) On estimating probability of presence from use-availability or presence-background data. *Ecology 94*: 1409-1419.
150. Bian L (2004) A conceptual framework for an individual-based spatially explicit epidemiological model. *Environ Plann B Plann Des 31*: 381-395.
151. Jimenez-Valverde A, Lobo JM (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecol 31*: 361-369.
152. Dean CB, Nielsen JD (2007) Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal 13*: 497-512.
153. Brooks CP, Antonovics J, Keitt TH (2008) Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. *Am Nat 172*: 149-159.
154. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, New York.
155. Firestone SM, Christley RM, Ward MP, Dhand NK (2012) Adding the spatial dimension to the social network analysis of an epidemic: investigation of the 2007 outbreak of equine influenza in Australia. *Prev Vet Med 106*: 123-135.
156. Solheim H, Timmermann V, Børja I, Hietala AM (2011) Yggdrasils helsestilstand—Åskeskuddsjuke er på frammarsj. *Skogeiaren 96*: 34-36.
157. Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *J Biogeogr 33*: 1677-1688.
158. Valle M, van Katwijk MM, de Jong DJ, Bouma TJ, Schipper AM, et al. (2013) Comparing the performance of species distribution models of *Zostera marina*: Implications for conservation. *J Sea Res 83*: 56-64.
159. Li W, Guo W (2014) A new accuracy assessment method for one-class remote sensing classification. *IEEE T Geosci Remote 52*: 4621-4632.
160. Stohlgren TJ, Ma P, Kumar S, Rocca M, Morisette JT, et al. (2010) Ensemble habitat mapping of invasive plant species. *Risk Anal 30*: 224-235.
161. Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods Ecol Evol 1*: 330-342.
162. Svenning JC, Skov F (2004) Limited filling of the potential range in European tree species. *Ecol Lett 7*: 565-573.
163. Boychuk D, Perera AH, Ter-Mikaelian MT, Martell DL, Chao L (2004) Modelling the effect of spatial scale and correlated fire disturbances on forest age distribution. *Ecol Modell 95*: 145-164.
164. Münzbergová Z (2004) Effect of spatial scale on factors limiting species distributions in dry grassland fragments. *J Ecol 92*: 854-867.
165. Gross A, Hosoya T, Queloz V (2014) Population structure of the invasive forest pathogen *Hymenoscyphus pseudoalbidus*. *Mol Ecol 23*: 2943-2960.
166. Drenkhan R, Sander H, Hanso M (2014) Introduction of Mandshurian ash (*Fraxinus mandshurica* Rupr.) to Estonia: is it related to the current epidemic on European ash (*F. excelsior* L.)? *Review. Eur J For Res 133*: 769-781.
167. Podger FD, Mummery DC, Palzer CR, Brown MJ (1990) Bioclimatic analysis of the distribution of damage to native plants in Tasmania by *Phytophthora cinnamomi*. *Aust J Ecol 15*: 281-289.
168. Koch FH, Smith WD (2008) Spatio-temporal analysis of *Xyleborus glabratus* (Coleoptera: Curculionidae [corrected] Scolytinae) invasion in eastern U.S. forests. *Environ Entomol 37*: 442-452.
169. Cech TL (2008) Esch disease in Lower Austria - new study results. *Forest Protection News 43*: 24-28.
170. Ogris N (2008) Ash blight, *Chalara fraxinea*. *News from forest protection no. 1*.
171. Ogris N (2010) Rachises as key to ash decline due to *Chalara fraxinea*. *EPPO Workshop on Chalara fraxinea*, June 30th–July 2nd, Oslo, NO.
172. Zheng H-D, Zhuang W-Y (2013) *Hymenoscyphus albidoides* sp. nov. and *H. pseudoalbidus* from China. *Mycol Progr 13*: 625-638.
173. Pulliam HR (2000) On the relationship between niche and distribution. *Ecol Lett 3*: 349-361.
174. Václavík T, Meentemeyer RK (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Divers Distrib 18*: 73-83.
175. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, et al. (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol Evol 19*: 535-544.
176. Rosenzweig C, Iglesias A, Yang XB, Epstein PR, Chivian E (2001) Climate change and extreme weather events; implications for food production, plant diseases, and pests. *Glob Change Hum Health 2*: 90-104.
177. Chakraborty S (2005) Potential impact of climate change on plant-pathogen interactions. *Australas Plant Pathol 34*: 443-448.