Biol Invasions (2018) 20:3527–3544
https://doi.org/10.1007/s10530-018-1793-1

ORIGINAL PAPER

Impact of biased sampling effort and spatial uncertainty of locations on models of plant invasion patterns in Croatia

Andreja Radović · Stefan Schindler · David Rossiter · Toni Nikolić

Received: 14 July 2015 / Accepted: 25 June 2018 / Published online: 9 July 2018

Abstract Very frequently biological databases are used for analysing distribution of different taxa. These databases are usually the result of variable sampling effort and location uncertainty. The aim of this study was to test the influence of geographically biased sampling effort and spatial uncertainty of locations on models of species richness. For this purpose, we assessed the pattern of invasive alien plants in Croatia using the Flora Croatica Database. The procedure applied in testing of the sensitivity of models consisted of sample area sectioning into coherent ecological classes (hereinafter Gower classes). The quadrants were then ranked based on sampling effort per class. This resulted in creation of models using varying numbers of quadrants whose performance was tested with independent validation points. From this the best fitting model was determined, as well as a threshold of sampling effort. The data from quadrants with sampling effort below the threshold were considered too unreliable for modelling. Further, spatial uncertainty was simulated by adding a random term to each location and re-running the models using the simulated locations. Biased sampling effort and spatial uncertainty of locations had similar effects on model performance in terms of the magnitude of the affected area, as in both cases 7% of the quadrants showed statistically significant deviations in alien plant species richness. The model using only on the

---

A. Radović (✉)
Department of Applied Geoinformatics and Spatial Planning, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague 165 21, Czech Republic
e-mail: radovic@fzp.czu.cz;
andreja.radovic100@gmail.com

S. Schindler
Environment Agency Austria, Spittelauer Lände 5, 1090 Vienna, Austria
e-mail: stefan.schindler@umweltbundesamt.at

S. Schindler
Division of Conservation Biology, Vegetation and Landscape Ecology, University of Vienna, Rennweg 14, 1030 Vienna, Austria

D. Rossiter
New York State College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14850, USA

D. Rossiter
Faculty of Geo-Information Science and Earth Observation, Formerly of University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

T. Nikolić
Marulićev trg 20/II, 10000 Zagreb, Croatia
e-mail: toni.nikolic@biol.pmf.hr

quandrants with the highest 35% quantile sampling effort best balanced the sampling effort per quadrant and overall geographical coverage. It predicted a mean number of 3.2 invasive alien plant species per quadrant for the Alpine region, 5.2 for the Continental, 6.1 for the Mediterranean and 5.3 for the Pannonian region of Croatia. Thus, the observational databases can be considered as a reliable source for species richness models and, most likely, for other types of species distribution models, given that their limitations are accounted for in the data selection process. In order to obtain precise estimates of species richness it is required to sample the whole range of ecological conditions of the study area.

Keywords Biodiversity databases · Balkans · Data quality · Regression kriging · Spatial analysis

Introduction

Reliability and precision of predictions by species distribution modelling depend on three main groups of factors: (1) the nature and quality of the original input data for model specification (e.g., models of presence, presence/absence, or density); (2) the predictive power of the environmental variables; and (3) the modelling techniques used (Araújo and Guisan 2006; Ferrier et al. 2002; Guisan and Zimmermann 2000; Lobo 2008b). Biological databases are often used in analysing distribution of different taxa. However, usually they are characterised by gaps in both, taxonomic and geographic spaces, the so-called "Wallacean shortfall" (Chefaoui et al. 2005; Bini et al. 2006; Hortal and Lobo 2006; Rocchini et al. 2011). They suffer from oversampling of biologically interesting areas and other extraordinarily attractive areas (Dennis and Thomas 2000; Zaniewski et al. 2002; Fourcade et al. 2014). Problems in using information from such biological databases in preparing spatial ecological models are numerous. The use of all available distributional information on species in a given area, such as the incorporation of species records from unsystematic samplings and data observations with uncertain coordinates of location, e.g., toponyms (Rocchini et al. 2011), frequently results in a biased picture of species distribution and biodiversity. For instance, this may be the case due to the uneven distribution of sampling effort (Lobo 2008a; Philips et al. 2009; Fourcade et al. 2014). Several solutions have been proposed to overcome the problem of data deficiency (Hortal and Lobo 2006). These solutions range from models predicting single species distribution (Guisan and Zimmermann 2000; Fourcade et al. 2014) via synecological models predicting assemblages of species (Hortal and Lobo 2006; Guisan and Rahbek 2011) to models predicting entire communities (Ferrier and Guisan 2006). To account for uneven sampling effort some authors suggest the use of species accumulation curves (Hortal and Lobo 2005), while others suggest nonparametric techniques such as the Chao index (Chao and Tsung-Jen 2004). However, since maps of any biodiversity surrogate can be interpolated from even a few well-characterized sites (Hortal and Lobo 2005), biological databases can be a valuable source of information, if their reliability is carefully assessed (Robertson et al. 2010; Santos et al. 2010). A comprehensive sampling effort assessment is required in order to reveal areas that are sufficiently covered and to treat their inventories as reliable information on biodiversity (Garcillán et al. 2003; Hortal and Lobo 2005; Fourcade et al. 2014). In those cases when direct information on sampling effort is missing, an indirect measure, such as the total number of records, may be used as a sampling-effort surrogate for distinguishing localities that are likely well-surveyed and that should therefore be used for modelling (Lobo 2008a; Philips et al. 2009).

Invasive alien species, both purposively and accidentally introduced, are recognized as one of the major drivers of global biodiversity change (Sala et al. 2000; Genovesi and Shine 2003; Bellard et al. 2013; Tittensor et al. 2014). They present new challenges in managing natural ecosystems (Thuiller et al. 2007) since they have the potential to establish viable populations with high growth rates. Such populations eliminate parts of the native biota (Rejmanek 1999). Further, they modify the disturbance regime (Brooks et al. 2004), and transform ecosystem structures and functioning (Dukes and Money 1999). As a consequence, they adversely affect local diversity, economy, ecosystem services (Pimentel et al. 2005) and the health of humans and wildlife (Mack et al. 2000; Taramarcaz et al. 2005; Scalera et al. 2012; Simberloff 2013; Schindler et al. 2015; Roy et al. 2016). For these reasons, knowledge of spatial occurrence patterns of invasive alien species, often obtained by predictions based on distribution modelling, is essential for decision-making. However, for invasive alien species, due to the dynamics of their invasion it is particularly difficult to differentiate poorly sampled areas from truly absent areas (Crall et al. 2013; Philips et al. 2009). As the initial conditions strongly affect the performance of methods for correcting sampling bias there is a need to develop systematic (step-by-step) guidelines (Fourcade et al. 2014). A process of iterative modelling is suggested in particular for invasive alien species (Stohlgren and Schnase 2006).

In this study, we have modelled the species richness of invasive alien plants in Croatia. The study primarily aims to test the sensitivity of model performance and outcome while accounting for two major constraints in observational data: geographically biased sampling efforts and spatial uncertainty of locations. Thus, our study delivers new insights because it differs from similar studies by (1) dealing with sampling bias and spatial uncertainty, (2) focusing on invasive alien plant species (but see e.g. Crall et al. 2013), and by (3) modelling species richness instead of distributions of single species. A novelty of the approach is the use of several nested datasets along a gradient from weak comprehensive spatial coverage but low sampling bias to good spatial coverage but strongly biased sampling effort per location. This approach includes the assessment of a threshold for input data that represent the optimal trade-off between these two constraints. Further, our objectives included testing the performance of our approach by assessing the error in modelling results caused by biased sampling efforts, by comparing models based on the input data set of the optimal trade-off with those based on the full original data sets. We also tested the effects of spatial uncertainty of locations by comparing model alternatives that do and do not account for this limitation of the input data. Secondary aims of this study were to present, for applied purposes, a model of the invasive alien plant species richness of Croatia at 25 km^2 quadrants, provide certain insight about the robustness of this model, and the from the model an assessment of the biogeography of Croatian invasive alien plant species richness. These three secondary aims are all crucial for effective control and management measures.

Materials and methods

The study area: Croatia

Croatia is at the intersection of Central Europe, Southern Europe and the Mediterranean. Despite its relatively small surface, four out of eleven European biogeographic regions are found there: Alpine, Continental, Mediterranean and Pannonian. As a refuge for diverse plant species during the last glacial period, today, Croatia is extraordinarily rich in vascular flora (Nikolić et al. 2013) and, as such, it has been recognised as a regional biodiversity hotspot (Griffiths 2004; Fady and Cyrille 2010; Hewitt 2011; Zachos and Habel 2011). The invasive alien flora in Croatia has a similar taxonomical composition to that reported in other European countries (Nikolić et al. 2014). Recently, a checklist of invasive alien plant species was prepared (Boršić et al. 2008), including their potential impact on native flora (Mitić et al. 2008). Their spatial distribution has been assessed (Nikolić et al. 2013). However, some Croatian botanists have observed inadequate/poor knowledge of flora in some parts of Croatia (Nikolic et al. 1998, 2014).

The Croatian flora database

For this study and the methodological approach developed within (Fig. 1), we used the Flora Croatica Database (hereinafter: FCD, http://hirc.botanic.hr/fcd/) that contains the most currently-available information on Croatian flora. The FCD was compiled from diverse studies and research projects with diverse objectives, sampling intensities, spatial extents, and temporal coverages. Further, it contains the information extracted from published literature, as well as from herbarium collections. For this study, we extracted all 11,823 occurrences of alien plant species from the FCD (Fig. 2). The biogeographic regions were clearly subjected to different sampling efforts. The mean number of observation records per 25 km^2 quadrant was the greatest in the Alpine (n = 400) and Mediterranean (n = 377) region, yet significantly lower in the Continental (n = 230) and Pannonian (n = 104) region of Croatia. These numbers, scaled by the total number of taxons ever recorded in each region, as a proxy for vegetation diversity of the zones, show similar disproportion (Alpine region 3.3; Continental 2.8, Mediterranean 2.5 and Pannonian 1.5).

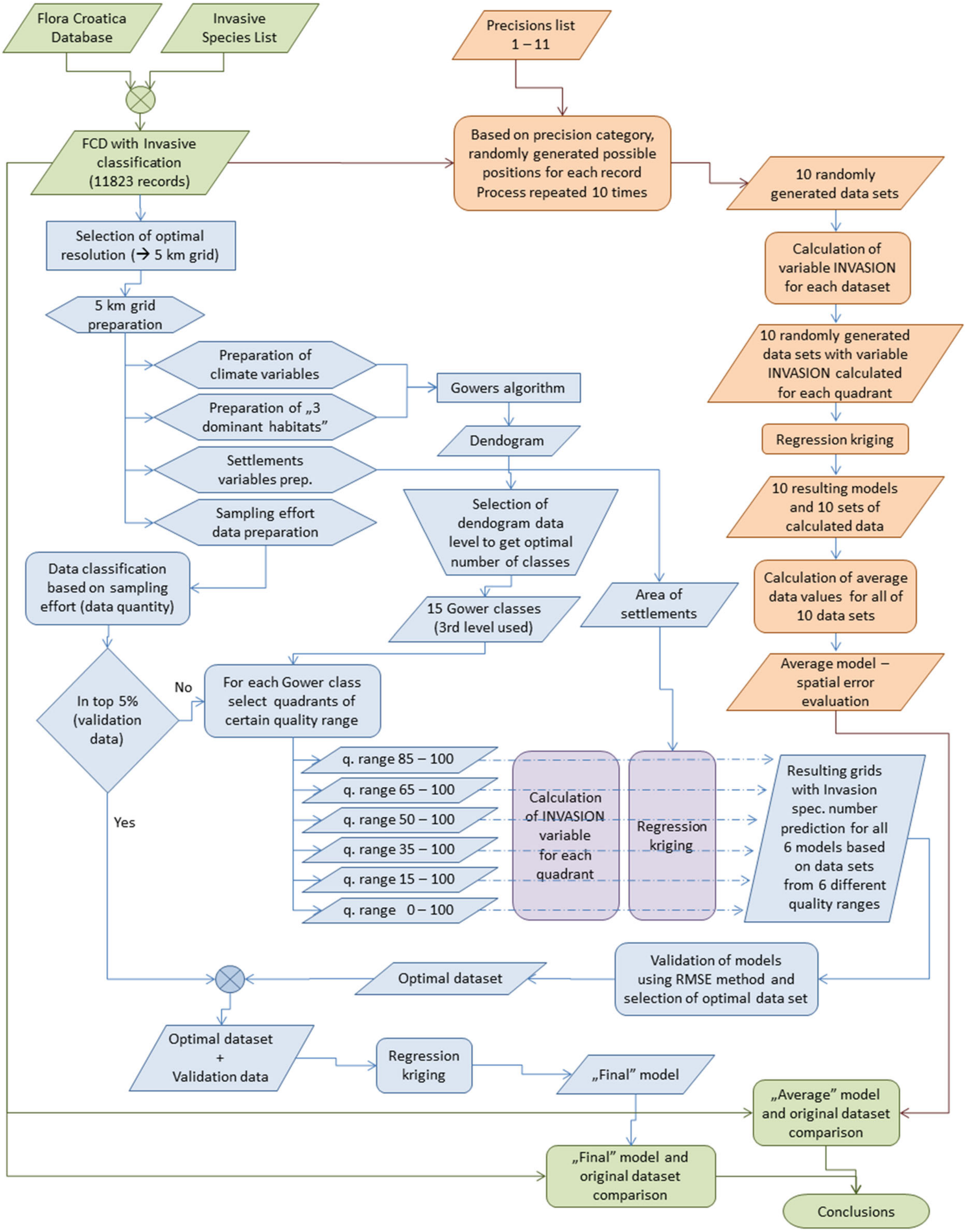


Fig. 1 Detailed scheme of the methodological steps conducted in this study

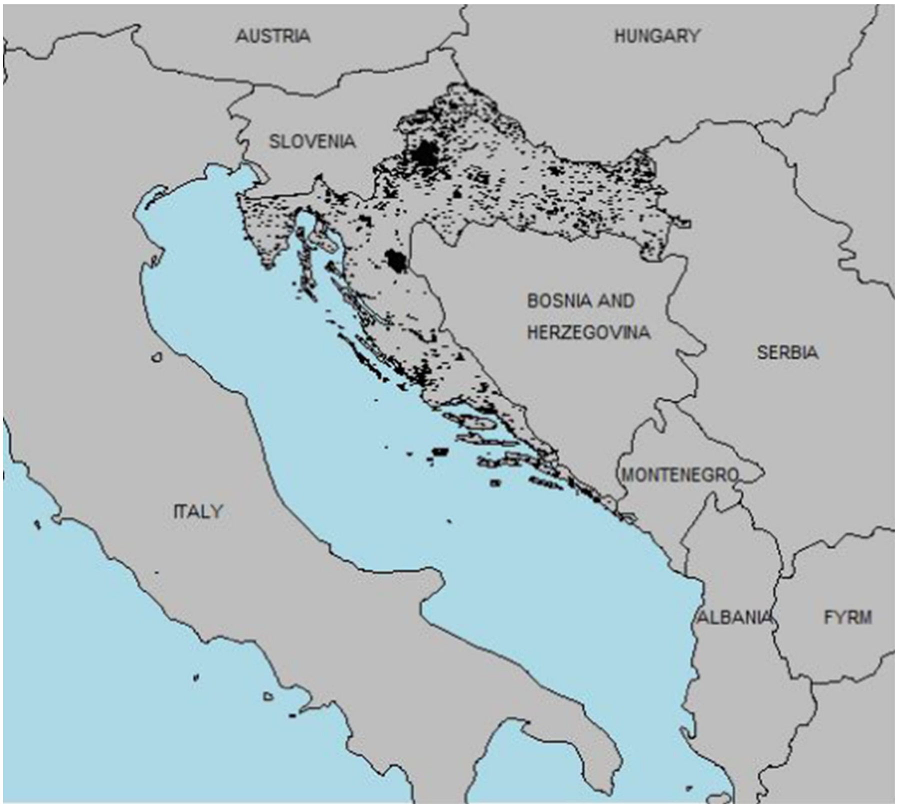


Fig. 2 Point locations of invasive alien plant species extracted from the Flora Croatica Database

The oldest data records in the database were obtained from herbarium collections. In total, such records make $< 6\%$ of total data. Amongst those, the oldest recorded data (before 1900) account for less than 1% of the overall occurence data in FCD. The most data (52%) are gathered through field observations; these were collected over the past 20 years. The information on occurrences from the literature (42%) originate mainly from the period after 1950. Therefore, approximately 94% of the data used in this study were collected after the appearance of most invasive alien plant species in Croatia. Each FCD data record has information on the spatial uncertainty of the location (Table 1). A portion of these observations were georeferenced only by toponyms (e.g. nearby settlements) or by a Central European MTB grid coordinate (German "Messtischblätter", i.e., $6' \times 10'$ map sheets) (Nikolic et al. 1998), or their divisions by 4 ($3' \times 5'$), 16 ($1.5' \times 2.5'$) or 64 ($0.75' \times 1.25'$). The spatial uncertainty of each FCD location was coded, with the code ranging from 0 (present somewhere in Croatia) to 11 (the observation was geo-referenced using a GPS receiver to the spatial precision of the device). Thus, the reported spatial uncertainty ranged from 5 m to 500 km (Table 1). The point shapefile of invasive alien species records used in this modelling exercise had variable spatial uncertainty ranging from category 2–11, with 93.8%, having greater spatial precision than our mapping resolution of 5 km. For this reason, we conducted the work with 5 km × 5 km (i.e., 25 km^2) quadrants, after transferring the original species data and predictor variables (see below for details on predictors) to this spatial resolution. Point information on the presence of invasive alien plant species was, therefore, used to compute the variable "invasive alien plant species richness", i.e., the number of alien plant species detected per 25 km^2 quadrant. In order to test the effect of the spatial uncertainty of positions, we used regression kriging (see below) to interpolate over the mapping grid, each of the ten interpolations after preparing a separate set of x–y simulated coordinates with associated uncertainties as specified in Table 1. Transformation of point data to grid data of the desired resolution was done by assigning the unique species IDs to quadrants by the overlap method from "sp" package in R

Table 1 Spatial uncertainty of locations (in radius distance [m]) of records from Flora Croatica Database (FCD)

Spatial uncertainty code FCD	Spatial uncertainty [m]	Data references by MTB or UTM	Numbers of observations in point data set
0	500,000		0
1	200,000		73
2	5643.3	MTB, UTM 10 × 10	664
3	2821.7	$\mathrm{MTB}^{1/4}$	533
4	1784.6	$\mathrm{MTB}^{1/16}$	547
5	1261.9		2931
6	564.4		2113
7	100		1705
8	50		2
9	25		259
10	5		30
11	< 5		2964

Uncertainties in FCD that were given in terms of areas (km^2) were transformed to linear measures (m)

(function 'over') and then aggregating the number of unique invasive alien species IDs with the function 'aggregate' in R.

Predictor variables for modelling

As a surrogate for sampling effort, we used the total number of records in the FCD per spatial unit as variable "effort". We assume that this variable is an appropriate surrogate, because it is likely that an entire target group of species, observed by similar methods, will share similar bias (Philips et al. 2009).

Habitat, climate, landscape heterogeneity and human disturbance mainly affect plant invasions (Woodward 1987; Hunter and Yonzon 1993; McKinney 2001; Araújo 2003; Pino et al. 2005; Chytry et al. 2008). Thus, to describe Croatia by its main ecological characteristics, we used information on habitats and climate. Croatian State Institute for Nature Protection provided the spatial database of habitat types (SINP 2009). For each 10 km^2 we extracted information on its three main habitat classes. We downloaded nineteen bioclimatic variables from the Worldclim database (http://www.worldclim.org) (Hijmans et al. 2005). Since they were highly correlated, we summarised them by using principal component analysis to obtain uncorrelated variables suitable for ordinary least squares (OLS) regression. Based on the eigenvalue criterion, we used the first three standardized principal components for further analyses. Since the resulting set of variables was mixed (three continuous numerical principal components and three categorical variables describing three main habitat classes), we used the Gower algorithm (Gower 1971; Kaufman and Rousseeuw 1990), the only classification algorithm that can deal with such mixed variable structure, to calculate the similarity of quadrants. We decided to divide Croatia into 15 classes (hereinafter called "Gower classes"; "Appendix B-a") because this number of classes was obtained with the Gower algorithm dendrogram at the third level of separation. This classification resulted in coherent geographic clusters (see "Appendix A-b").

As a surrogate for human disturbance for each quadrant, we used the area of settlements, obtained from Oikon Ltd, Institute of Applied Ecology in the form of a GIS-shapefile, which was then gridded and up-scaled to fit the model resolution.

The biogeographic classification of Croatia into Mediterranean, Alpine, Continental, and Pannonian ("Appendix A"), was prepared by the Croatian State Institute for Nature Protection (EEA 1998). Recently, a new classification of the Croatian territory into only 3 biogeographic regions was proposed by the European Environment Agency (EEA, http://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe) by merging the Continental and the Pannonian area. However, it was our preference to keep the old classification due to the particularities of the Pannonian part in terms of climate, biogeography and sampling effort.

All environmental variables were first prepared as gridded data with a high resolution of 50 m and then aggregated to 25 km^2 resolution for final modelling by summing for the total area of settlements, by calculating mean values for the climatological variables.

Analysing a threshold in sampling effort to derive an optimal set for modelling

Since the sampling effort was extremely unevenly distributed, particularly in large areas of East and Central Croatia, certain ecological conditions were not well covered. This was the reason for using the Gower classes as a tool and ranking all quadrants per Gower class according to their respective sampling effort. The first step in the following analysis was to identify the best-sampled quadrants (5%, arbitrary threshold) from all over Croatia. These quadrants were not used for modelling. They were kept apart for validation. Moreover, we assumed that we have knowledge of the total pool of invasive alien plant species for these quadrants. Then we prepared for each Gower class six nested sets of quadrants corresponding to six quantile ranges of sampling effort (85–100, 65–100, 50–100, 35–100, 15–100 and 0–100) of the total number of records in the quadrant and Gower class (Table 2, "Appendix C"). These datasets are referred to hereinafter as D85, D65, D50, D35, D15, D0, respectively, and constitute a gradient related to different intensities of the sampling bias ranging from "huge gaps not considered for model calibration" (i.e. D85) to "the majority of considered quadrants are very poorly sampled" (i.e. D0).

Model preparation and validation

To predict the number of invasive alien plant species we used Regression Kriging technique (RK) (Hengl et al. 2007), i.e. a regression on environmental predictors and interpolation of the model residuals by simple kriging (Pebesma 2006). We chose RK because it has proven to be useful in ecological applications (e.g. Liebhold 1993; Hengl 2009) and also because RK performed well in the methods comparison studies (Wang et al. 2012; Meng 2014). We used ordinary least squares regression (OLS) rather than generalized least squares (GLS) for the regression portion of RK. The regression coefficients found by OLS are unbiased estimates of those found by GLS (Draper 1998, p. 223–4). The variance of the coefficients differs, but this is no problem, because we do not use the variance in the model to compute the regression residuals in the kriging portion of RK but only the coefficients.

Table 2 Root mean squared error (RMSE) at validation points for different data subsets (Do–D85)

Dataset	Quadrants used (Percentiles)	RMSE
D85	85–100	1425.01
D65	65–100	652.65
D50	50–100	660.36
D35	35–100	708.64
D15	15–100	707.67
D0	0–100	705.87

Best model at validation points is D65, i.e. the model based on the quadrants with sampling effort greater than the 65 percentile per Gower class

We used the Gower class and the area of settlements of each quadrant as an overall description of habitat and climate for the regression part of the models, because previous studies described connections between habitat composition, climate and human presence on the invasion level (Hunter and Yonzon 1993; McKinney 2001; Araújo 2003; Chytry et al. 2008).

We compared the six prepared datasets (i.e. D85, D65, D50, etc.) according to their predictive accuracy as evaluated by the root mean squared error of prediction (RMSE) (Hengl 2009). RMSE is the square root of the average squared difference between observed and predicted values at the independent validation points. We selected the dataset that provided the most accurate model (hereinafter: "the best dataset") and computed the linear model and its residuals. These we interpolated using the 'krige' function in the 'gstat' R package using fitted variogram models of estimates from empirical variograms of the regression residuals fitted with the 'fit.variogram' function of the 'gstat' R package (Pebesma 2004). We then summed the interpolated residuals and regression predictions at each cell to give a final model prediction (hereafter "best model"). This method has already proved its usefulness in creating models from incomplete inventories (Hortal and Lobo 2011).

Finally, we merged the best dataset and validation quadrants and then repeated the regression kriging procedure for calibration of a new model. We did this since, for the applied purpose of our model, it is of advantage to include the best-sampled areas of the country. The resulting, "final model" could not be validated because we included our independent validation quadrants for its fitting. However, we examined this model to identify areas with large per-grid cell difference between this model and the full model D0. To understand underlying data constraints and biogeographic patterns, we aggregated by summing them for Gower classes, sampling effort zones and biogeographic regions in SAGA GIS and using 'RSAGA' R package (Brenning 2008). We calculated the significance of the obtained differences in predictions using the 'SigDiff' function of the 'SDMTools' R package (VanDerWal et al. 2012) on the zonal statistics for biogeographic regions, sampling effort zones and Gower classes. Significant differences were calculated first at Gower class level to account for the different inherent plant diversity in the class. Then they were merged together in one resulting grid. For the purpose of multiple general linear hypotheses testing we used the 'glht' function of the 'multcomp' R package (Hothorn et al. 2008) to detect the effects of, and interactions between, our predictors.

Analysing how spatial uncertainty of locations influences the modelling results

To assess the influence of spatial uncertainty for the locations of the input data records, we simulated possible coordinates of each observation of our dataset "D65 plus validation points" by using the 'sample' R function. We added independent uniformly distributed random values, ranging from minus to plus one unit of reported uncertainty for each observation (Table 1), on the values of each pair of X, Y coordinates. We repeated the regression kriging procedure as explained in the previous section on all ten realizations of simulated coordinates. We applied this modification of our final model from the previous section to predict the number of invasive alien plant species, and averaged the ten obtained maps for each 25 km^2 quadrant. The modelling results were compared with the final model from the previous section by assessing the magnitude of prediction differences of the models, again using the 'SigDiff' function of the 'SDMTools' R package (VanDerWal et al. 2012). Finally, we tested how significant differences were distributed biogeographically and across the sampling effort zones using analysis of variance, 'aov' function of the 'stats' R package. The research effort zone grid was split according to the quantile statistics into 4 zones of research intensity giving code 1 to the zone of lowest research intensity, up to 1st quartile effort, code 2 for 1st quantile to median, code 3 for mean to 3rd quantile and finaly code 4 for those quadrants having effort above 3rd quartile.

The script for the analysis was carried out in the R environment for statistical computing (R Core Team 2012) using diverse packages ('rgdal', 'rxlsx', 'sp', and 'maptools') for data manipulation and aggregation, and also in the SAGA "System for Automated Geoscientific Analyses" GIS version 2.0.8 (http://www.saga-gis.org/en/index.html). R functions and scripts were executed within SAGA by means of the 'rsaga' package connecting SAGA and R.

Results

Geographically biased sampling effort and trade off with data coverage

Model predictions at validation points were quite inaccurate if models were built using only the top 15% quantile sampling effort quadrants (D85) per Gower class (Table 2). The best dataset was the one built with the top 35% quantile sampling effort (D65; Table 2). That dataset provided best-balanced sampling effort per quadrant and overall geographic coverage. Adding quadrants with lower sampling effort (i.e., D50 to D0) gradually lowered the accuracy. The best model, i.e. the one derived from D65, predicted zero to 38 invasive alien species per quadrant.

The mean predicted number of alien plant species differed across biogeographic regions with the Alpine having 1.5 (range 0–15) species, the Continental 4.2 (range 0–38) species, the Mediterranean 4.6 (range 0–36) species and the Pannonian 4.8 (range 0–26) species. In the process of regression kriging we detected autocorrelation in residuals of the regression model and fitted a pentaspheric variogram model. The semivariance reached a definite sill at approximately

26 km separation, with a nugget parameter of 17.6 species and structural sill of 8.1 species. Thus, the nugget is a large proportion (68%) of the total sill, showing moderate spatial dependence of the model residuals.

Comparison of final model versus full model

Differences in predictions of the 'final model' (D65+ validation points) and the 'full model' (D0) ranged between − 15 and + 16 invasive alien plant species (Fig. 3a) and were clustered throughout the country (Fig. 3a). Significant differences were detected in 6.9% of quadrants (Fig. 3b) after scaling the number of predicted invasive alien plant species with mean species richness per Gower class per quadrant. The scaling was necessary due to extreme differences of species richness per Gower class. On average, the final model predicted higher levels of invasive alien plant species richness than the full model.

The model performance varied across regions ($p < 0.0001$) and sampling effort zones ($p < 0.0001$), also, the interaction between these two factors was statistically significant ($p = 0.012$). We obtained greater differences in model predictions for the zone sampled with lowest intensity (Table 3). The greatest absolute differences in predicted species richness between final and full model were found in the Continental region, while in the Alpine region the absolute differences were the lowest ("Appendix D-a"). In all bio-geographical regions, final and full model had the greatest discrepancy at areas with low richness of invasive alien plant species; while they provided similar results when the invasive alien plant species richness was high (see "Appendix D-b"). In all biogeographic regions, except the Mediterranean, there were more quadrants with significantly higher than significantly lower values in the final model when compared to the full model ("Appendix D-c"). Moreover, differences per Gower class showed that the final model by default produced higher values. Differences that are more significant were observed in the cases where the final model had higher values than where values were lower ("Appendix E"). The same pattern emerged when pooling quadrants across sampling effort zones ("Appendix E").

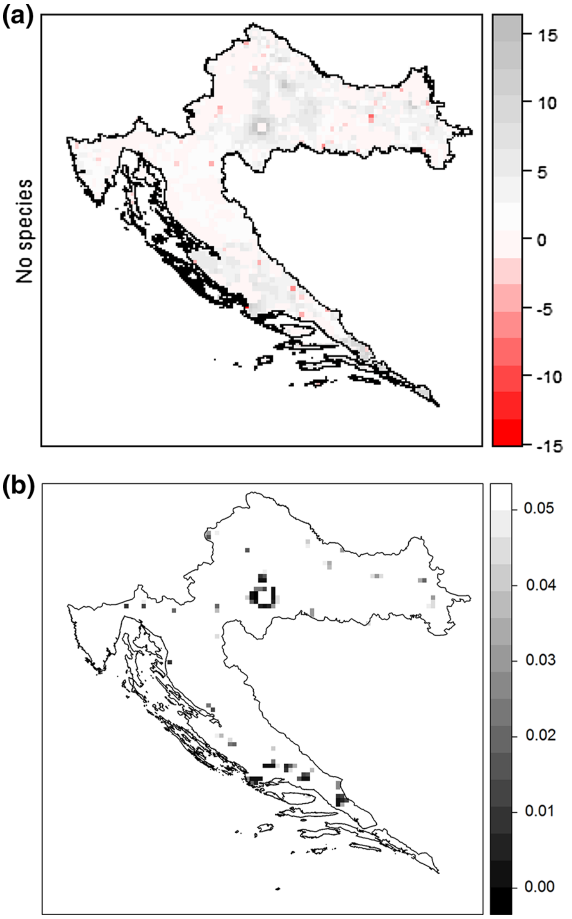


Fig. 3 **a** Differences in predictions of alien plant species richness between final model (D65 plus validation points) and full model D0; **b** significance of differences in predictions after scaling with mean number of overall species per Gower class; all grid cells with p value greater or equal 0.05 are plotted white

Impact of spatial uncertainty of locations

We detected significantly different predictions of invasive alien plant species richness in 7% of the quadrants when comparing our 'final dataset' of the last section to the one that included random error in coordinates according to the reported spatial uncertainty. The residual variogram from regression on environmental predictors revealed similar ranges than those from the final model (19.9 vs. 26.2 km) ("Appendix B-b"). When incorporating spatial uncertainty into the modelling procedure, the significance of detected differences did not differ across biogeographic regions ($p = 0.209$), while the effect of sampling effort was strong ($p < 0.001$). The highest

Table 3 Results of testing effect of sampling effort on statistical significance of detected differences in (a) final model (D65 + validation points) versus full model (D0) and (b) final model with consideration of spatial uncertainty of the locations versus final model without consideration of spatial uncertainty

	(a) Final model versus full model			(b) Final model plus precision uncertainty versus final model		
Sampling effort zones	Estimate	t-value	Significance (*p* value)	Estimate	t-value	Significance (*p* value)
2 versus 1	− 0.02	− 1.89	ns	0.01	0.68	ns
3 versus 1	− 0.05	− 4.15	< 0.001	− 0.01	− 0.45	ns
4 versus 1	− 0.04	− 3.25	< 0.01	− 0.09	− 5.38	< 0.001
3 versus 2	− 0.03	− 2.25	ns	− 0.02	− 1.12	ns
4 versus 2	− 0.02	− 1.35	ns	− 0.10	− 6.03	< 0.001
4 versus 3	0.01	0.90	ns	− 0.08	− 4.92	< 0.001

Significance testing was performed by using the 'SigDiff' function of the 'SDMTools' R package (VanDerWal et al. 2012). Coding for sampling effort zones: lowest sampling effort—code 1; highest sampling effort—code 4

sampling effort zone had a significantly lower amount of detected differences between the final model considering uncertainty and the final model ignoring uncertainty of locations than any other sampling effort class (Table 3).

The final model, considering spatial uncertainty of locations, gave slightly higher predictions of alien plant species richness across bio-geographical regions, compared to the final model ignoring spatial uncertainty. The Alpine region had a mean of 3.2 (range 0–18); the Continental a mean of 5.2 (range 0–42), the Mediterranean a mean of 6.1 (range 0–29), and the Pannonian a mean of 5.3 (range 0–18) species. The overall pattern among regions remained the same. This model (Fig. 4) we consider as the best possible prediction of invasive alien flora given the available information and, as such, it should be a valuable tool for decision-making.

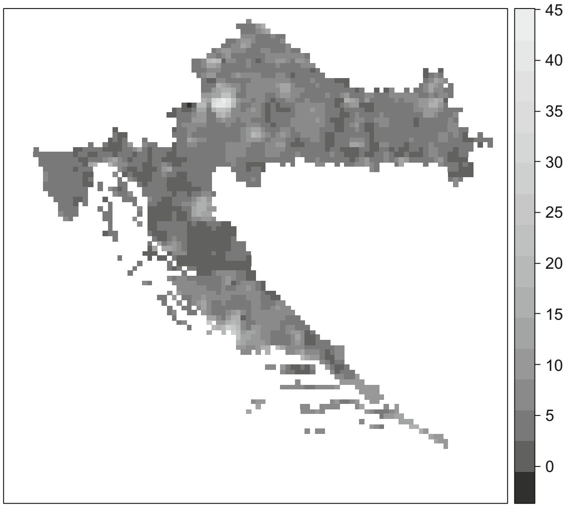


Fig. 4 Invasive alien plant species predictions using model Final model plus precision uncertainty obtained controlling both, geographically biased sampling bias and spatial uncertainty source of error

Discussion

Impact of geographically biased sampling effort and spatial uncertainty of locations

We assessed the threshold representing the optimal trade-off between data coverage and unbiased sampling effort and tested the influence of geographically biased sampling effort and spatial uncertainty of locations on models of invasive alien plant species richness for Croatia. Our results show that these two types of data limitations significantly affected the modelling predictions, with surprisingly similar effects on model performance. Statistically significant differences in predictions of 'final' (D65 plus validation points) versus 'full' (D0) models were detected in 7% of the country. Similarly, the significant differences in predictions of the 'final model' and the model incorporating the spatial uncertainty of locations occurred in 7% of the country. Adding more quadrants of lower sampling effort into models made them less precise. The model based on the 15% best sampled quadrants only was the one with worst performance. This is most likely due to large, poorly-sampled regions of the country, which were not covered at all in this set, as well as too low number of the quadrants included for modelling purposes, e.g. for reliable variogram fitting to detect spatial autocorrelation. Due to differences in approaches and outcome measures, it is difficult to compare these results with those obtained by other researchers. However, our results match those of Fourcade et al. (2014) who assessed in their comprehensive study on different sampling biases and correction methods that in the aspect of the differences between unbiased and biased models often remained moderate, despite significant deviations depending on target species and bias type. Our results also match those of Crall et al. (2013), who showed that stratifying the sampling area yields in better results of habitat suitability models for invasive alien species.

The pattern of statistically significant differences in predictions of final' versus 'full' model (cf. Appendices A, B, E) differed across sampling effort zones, showing statistically the largest discrepancy on predictions at the least-sampled areas (e.g. Gower class 6). In the Alpine region, a comparably high proportion of quadrants showed statistically significant differences despite small differences in absolute invasive alien plant species richness between the two models ("Appendix E-a"). However, high absolute differences without statistically significance were frequent for the quadrants with high average invasive alien plant species richness, rather poor sampling effort and over-predictions by the full model (Gower class 8, 11, 13). There were no significant differences observed between the final and full model in the areas of highest sampling efforts (e.g. Gower classes 1, 3, 4).

Taking into account spatial uncertainty resulting in, on average, higher estimates for species richness than when ignoring it. This implies, vice versa, that ignoring spatial uncertainty can lead to underestimations of the numbers of invasive alien plants and the size of their distributional ranges. The spatial uncertainty of the locations also implied limitations on the spatial resolution of the models. At 25 km^2 resolution, 6.3% of our point information of presence of invasive alien species had greater uncertainty than our prediction resolution. For making predictions at 1 km^2, this value would increase to 40% (Radović, unpublised). The pattern of differences among the final model considering and ignoring spatial uncertainty of locations, again, reveal that most significant differences occur in the least sampled areas of the country (cf. Table 3).

Invasion pattern in Croatia

Although based on observational data with several limitations, our models revealed clear patterns of invasive alien plant invasion for Croatia. The level of invasion significantly differs across the four biogeographic regions. The highest mean and maximum species richness was predicted for Pannonia, followed by the Mediterranean. The Mediterranean is considered particularly endangered by alien invasive species, due to its sensitivity to most drivers of global change, with biotic exchange indicated as one of the main threatening factors (Sala et al. 2000). For the Croatian Alpine region, our results show a low level of invasion by invasive alien plant species. Likely, the reason for this is relatively low human impact on the region (Pauchard et al. 2009). These major biogeographic patterns remained stable across the different kind of data sets and modelling approaches applied in this study.

Methodological solutions applied in this study

We are suggesting the described procedure as a solution when modelling species richness or species distribution based on observational databases with biased sampling effort and relevant spatial uncertainty of locations. The thresholds representing optimal trade-offs between data coverage and homogenous sampling effort can be obtained according to the model performances at the validation points. After selecting the optimum threshold, the validation points can be added back to the prediction set in order to produce a final model, which we consider an optimal output for applied purposes. Testing the performance of the final model should be conducted by sampling across different ecological conditions, but this approach is time consuming and expensive. Our approach was to use the RMSE of model residuals at the validation points, i.e. the 5% best sampled quadrants in Croatia, assuming that species lists at these quadrants are complete.

The environmental predictors used in our regression model were Gower classes based on habitat and climate and the area of settlements in each quadrant. These variables have shown to be linked with the level of invasion and are probably the most relevant in this context. However, additionally, other variables could be used in such modelling efforts, for instance, landscape heterogeneity and other landscape patterns having strong relations to species richness (Allouche et al. 2012; Tschantke et al. 2012; Schindler et al. 2013; Stein et al. 2014). Further, we could improve the sectioning by using more than three dominant habitats for the classification for better incorporation of habitat complexity. Incorporating imperfect detection of the species in the field could also strongly improve the final modelling outcome (Kery and Schmid 2004; Kéry et al. 2010). We are aware that our approach is mainly applicable in dealing with sampling bias and spatial uncertainty in the areas with rich data, because the statistical procedures used in this approach are hardly possible with small sample sizes. In particular, variogram estimation becomes increasingly unreliable as the number of points becomes smaller and their density lower. However, given the current trends in biodiversity data availability (e.g. Tittensor et al. 2014), we are confident that data availability will further increase for many areas, leaving data heterogeneity as principal challenge.

Conclusions and recommendations

We developed and assessed a new approach for modelling species richness using observational data collected with different sampling efforts due to a focus on areas attractive to naturalists and researchers. A clear-cut threshold for sampling effort per quadrant would leave large areas of specific habitat and climate conditions completely unrepresented in the dataset used for model preparation. Thus, we developed an approach where thresholds in sampling effort were applied for each of 15 homogenised parts defined according to three dominant habitat classes and climate conditions.

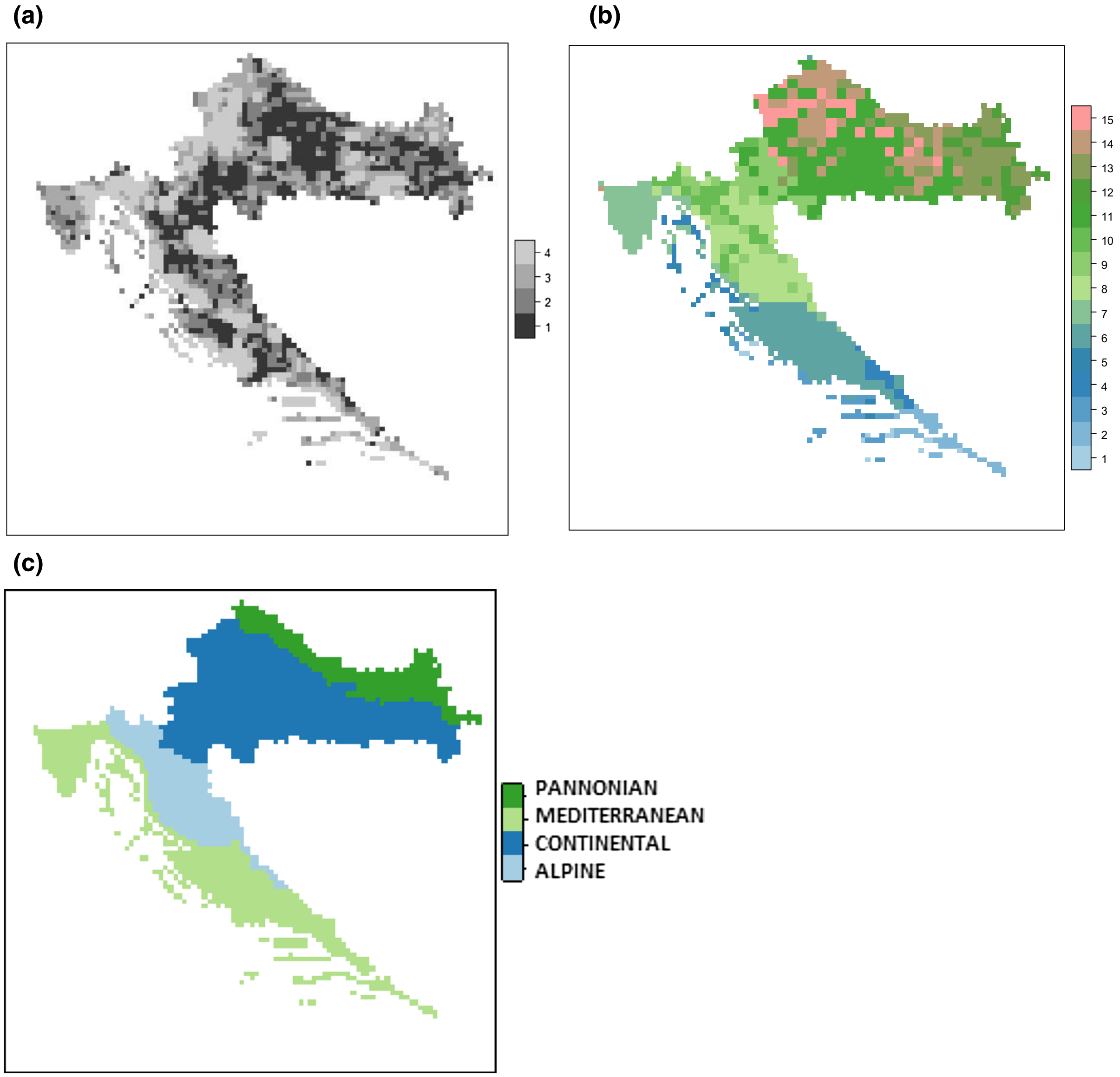
Our results clearly show that observational databases are a valuable source for ecological models that can lead to robust results (cf. Hortal and Lobo 2005). However, our approach also demonstrates that the usefulness of observational databases would strongly benefit from standardized sampling effort in a whole range of ecological conditions. Ignoring biased sampling efforts and uncertainty of locations could lead to significantly different predictions and to underestimations of numbers and distributions of invasive alien plant species. The threshold obtained according to the model performance at the validation points certainly depend on the input data, the predictive power of the environmental variables and the modelling techniques used and may well vary in other studies and regions. Furthermore, we are suggesting that the procedure described in this study be used whenever using the data from observational databases in model preparation. Such results can be complemented by the level of uncertainty of the predictions by testing the sensitivity of the results for different methodological options. The spatial uncertainty of the locations caused limitations for the model resolution. The data with higher spatial uncertainty of locations than grid resolution should be excluded. However, we expect that the problem of spatial uncertainty will improve, because more and more observers meanwhile report precisely georeferenced data.

Furthermore, our results lead to conclusion that that differential sampling effort is strongly related to ecological conditions, which, again, is a constraint for obtaining data qualities that serve the modelling purposes. It is, therefore, crucial to set up sampling and monitoring schemes that are stratified across ecological conditions. Croatian botanists have already observed the inadequate knowledge of flora in some parts of the country (Nikolić et al. 2014). This should change in a way that at least species checklists should become available for MTB, or even better, MTB64 quadrants (approximately 1 km^2). The use of modelling techniques can give the probability of species occurrence according to the known relationship among species and environmental variables. This information should be used to assess potential species assemblages across Croatia. As such, it should fill the gap in knowledge that exists now due to lack of records for some parts of Croatia's vegetation/floristic composition (Nikolić et al. 2014). This situation is not specific for vegetation data: databases for other taxa, like birds, suffer from even more extreme biases.

Acknowledgements This study was prepared under Grant 119-1191193-1227 of the Croatian Ministry of Science, Education and Sports. We would like to thank to two referees for improving previous versions of the manuscript.

Appendix A

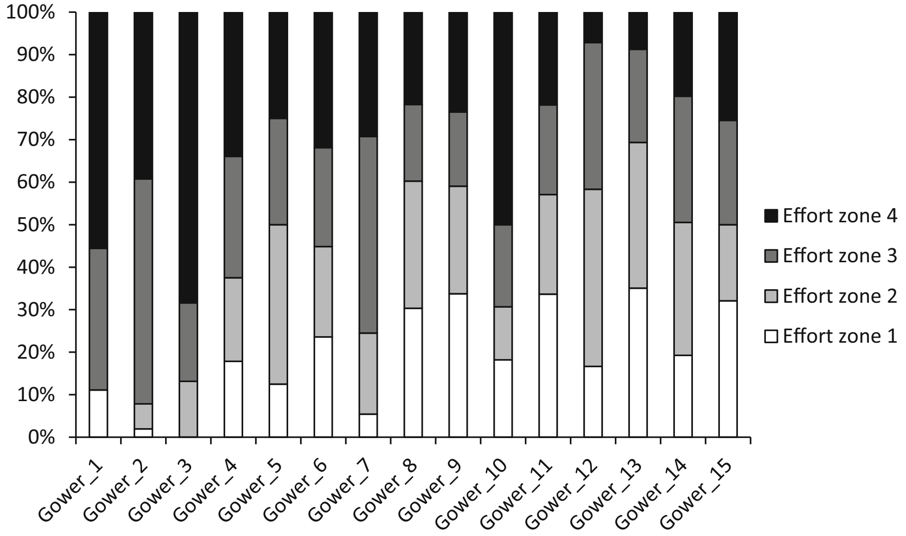
Raster data sets used to model alien plant species richness: (a) total number of floristic records in Flora Croatica Database (FCD)—transferred to 4 zones according to summary statistics; (b) Gower classes determined by means of clustering on similarity matrix obtained via Gower algorithm using information on three dominant habitat classes and climate variables; (c) biogeographic regions of Croatia according to State Institute of Nature Protection (EEA 1998).

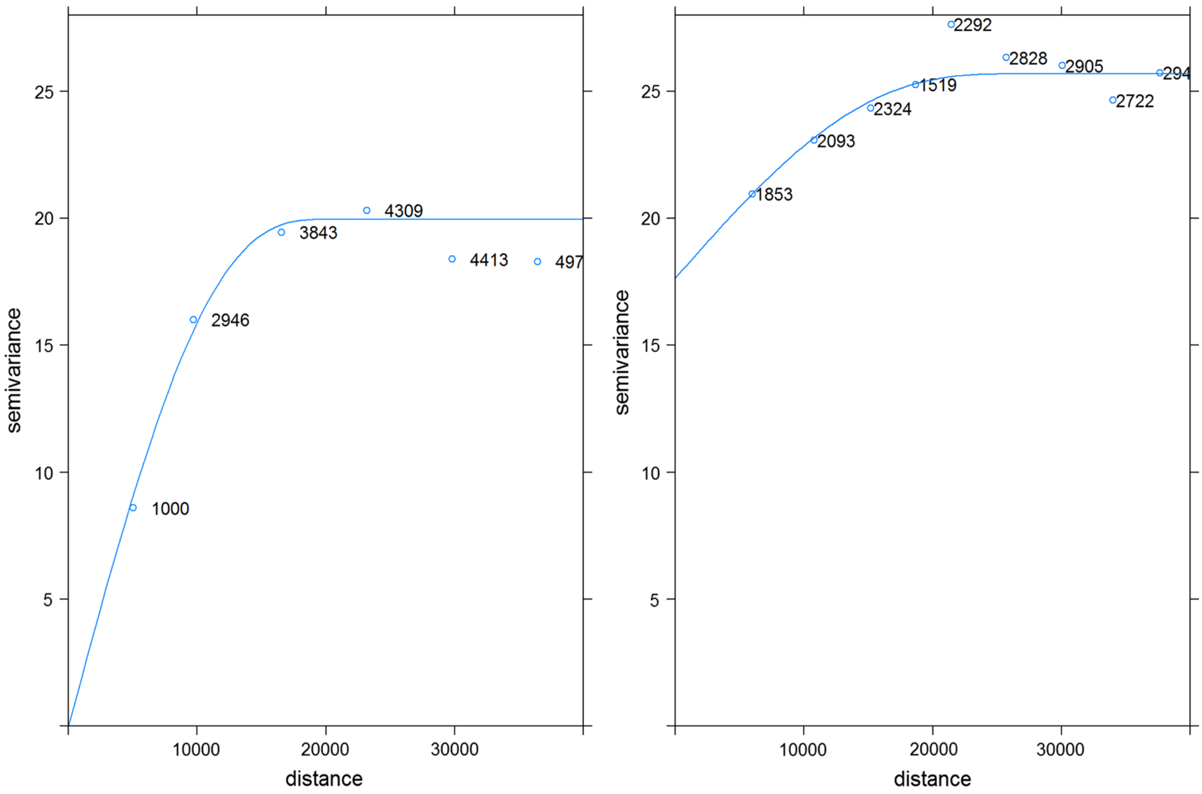


Appendix B

(a) Distribution of grids across sampling effort zones (1-white: lowest effort; 4-black: highest effort) per Gower class.

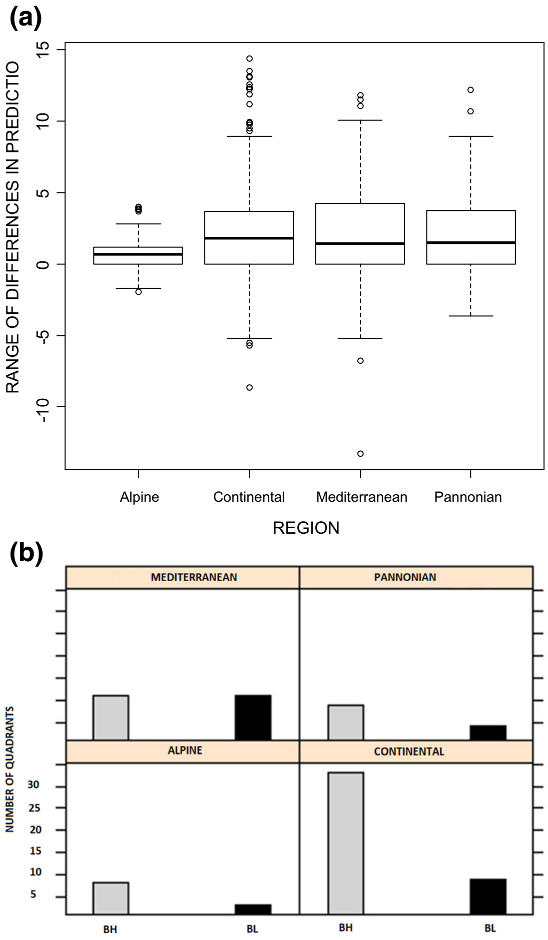
(b) Residual variograms from regressions on environmental predictors from the model that includes random errors in coordinates according to the reported spatial uncertainty (left panel) and the final model prepared with D65 plus validation points (right panel).





Appendix C

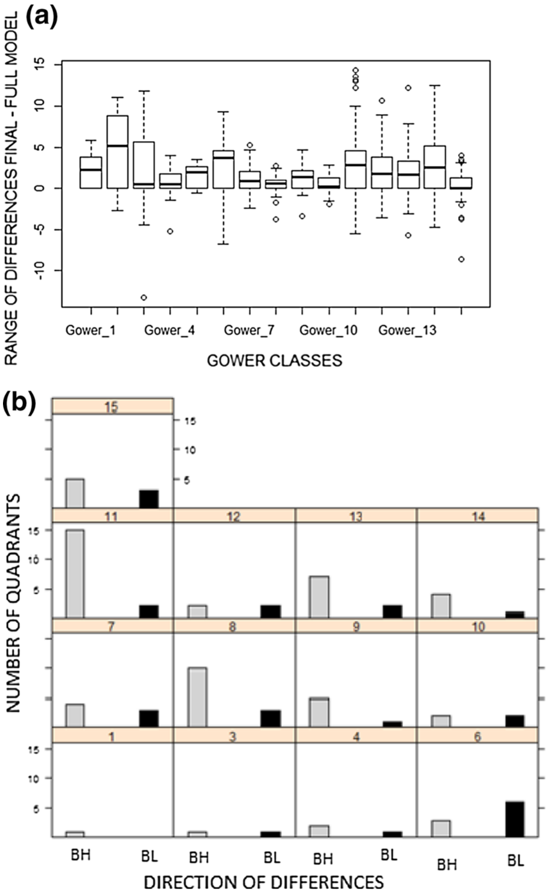
Invasive alien plant species richness predictions of final and full model: (a) differences of predictions across biogeographic regions; figure shows how final and full model predict differently regarding biogeographic regions—parallel box plots of differences in predicted invasive alien plant species richness; (b) number of quadrants with significantly different predictions of invasive alien plant species richness per biogeographic region (significantly higher (BH) and significantly lower (BL) at 0.05 level of significance).



Appendix D

Invasive alien plant species richness predictions of final and full model: (a) absolute differences of predictions across Gower classes; (b) Number of quadrants with significantly lower (BL) and significantly higher (BH) predictions of final model per region at 0.05 level per Gower classes (only Gower classes with detected significant differences are presented in plot); (c) Distribution of grids across sampling effort zones (1-white: lowest effort; 4-black: highest effort) per Gower class; number of quadrants per biogeographic region used in each dataset. For the datasets D85, D65, D50, D35, D15, and D0, quantile sampling effort was referring to each Gower class, not to the entire biogeographic region.

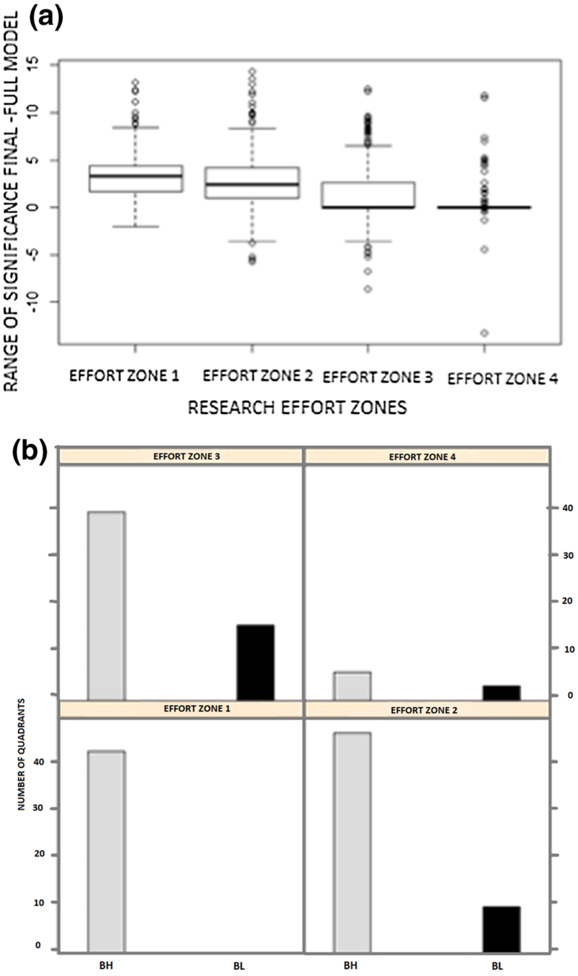
In these figures we presented how our models (final and full model) predicts differently regarding Gower classes (a) revealing those habitat/climate classes where differences are most pronounced as Gower classes 2, 3, 6, 11 and 14. Figure b present the direction of significant differences in predictions. Only in Gower class 6 the final model predicted more often significantly lower than significantly higher species richness than the full model.



(c)

Dataset	Alpine	Conti-nental	Mediter-ranean	Pannonian
D85	34	89	67	25
D65 (= best model)	95	265	194	92
D50	143	412	290	128
D35	193	540	381	179
D15	266	728	501	229
D0	310	920	636	278
Validation (best 5% across Country)	33	35	45	0
TOTAL	343	955	681	278
D65 + validation points (= final model)	147	349	291	104

Appendix E



Invasive alien plant species richness predictions of final and full model across research effort classes: (a) absolute differences of predictions across sampling effort zones revealing that final models predicted higher numbers and that the range of prediction differences stayed constant over all four research effort zones; (b) species richness predicted by final vs full model; (c) Number of quadrants with significantly lower (BL) and significantly higher (BH) predictions of final model at 0.05 level per sampling effort zones (coding for sampling effort zones: Highest sampling effort (> 3rd quartile)—zone 4 up to lowest sampling effort (< 1st quartile)—zone 1.

References

Allouche O, Kalyuzhny M, Moreno-Rueda G, Pizarro M, Kadmon R (2012) Area-heterogeneity trade-off and the diversity of ecological communities. Proc Natl Acad Sci (USA) 109:17495–17500

Araújo MB (2003) The coincidence of people and biodiversity in Europe. Glob Ecol Biogeogr 12(1):5–12

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33:1677–1688

Bellard C, Thuiller W, Leroy B, Genovesi P, Bakkenes M, Courchamp F (2013) Will climate change promote future invasions? Glob Change Biol 19(12):3740–3748

Bini LM, Diniz-Filho JAF, Rangel TF, Bastos RP, Pinto MP (2006) Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. Divers Distrib 12:475–482

Boršić I, Milović M, Dujmović I, Bogdanović S, Cigić P, Rešetnik I, Nikolić T, Mitić B (2008) Preliminary checklist of invasive alien plant species(ias) in Croatia. Natura Croatica 17:55–71

Brenning A (2008) Statistical geocomputing combining R and SAGA: the example of landslide susceptibility analysis with generalized additive models. In: Boehner J, Blaschke T, Montanarella L (eds) SAGA—seconds out (= Hamburger Beitraege zur Physischen Geographie und Landschaftsoekologie vol 19, pp 23–32

Brooks ML, Dantonio CM, Richardson DM, Grace JB, Keeley JE, DiTomaso JM, Hobbs RJ, Pellant M, Pyke D (2004) Effects of invasive alien plants on fire regimes. Bioscience 54(7):677–688

Chao A, Tsung-Jen S (2004) Nonparametric prediction in species sampling. J Agric Environ Stat 9:253–269

Chefaoui RM, Hortal J, Lobo JM (2005) Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. Biol Conserv 122:327–338

Chytry M, Maskell LC, Pyšek P, Vila M, Font X, Smart SM (2008) Habitat invasions by alien plants: a quantitative comparison among Mediterranean, subcontinental and oceanic regions of Europe. J Appl Ecol 448–458

Crall AW, Jarnevich CS, Panke B, Young N, Renz M, Morisette J (2013) Using habitat suitability models to target invasive plant species surveys. Ecol Appl 23(1):60–72

Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. J Insect Conserv 4:73–77

Draper NR, Smith H (1998) Applied regression analysis, 3rd edn. Wiley-Interscience, New York

Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? Trends Ecol Evol 14(4):135–139

EEA (1998) http://www.eea.europa.eu/data-and-maps/figures/biogeographical-regions-in-europe-1998, downloaded 12.10.2012

Fady B, Cyrille C (2010) Macroecological patterns of species and genetic diversity in vascular plants of the Mediterranean basin. Divers Distrib 16:53–64

Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. J Appl Ecol 43:393–404

Ferrier S, Watson G, Pearce J, Drielsma M (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast new south wales, Species-level modelling. Biodivers Conserv 11:2275–2307

Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PLoS ONE 9(5):97122

Garcillán PP, Ezcurra E, Riemann H (2003) Distribution and species richness of woody dryland legumes in baja california, mexico. J Veg Sci. https://doi.org/10.1016/j.tree.2013.01.014

Genovesi P, Shine C (2003) European strategy on invasive alien species: convention on the conservation of European wildlife and habitats (Bern convention). Council of Europe

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857–874

Griffiths HI, Kryštufek B, Reed JM (eds) (2004) Balkan biodiversity. Pattern and process in the European hotspot. Kluwer Academic Publishers, Dordrecht, p 377

Guisan A, Rahbek C (2011) Sesam—a new framework integrating macroecological and species distribution models for predicting spatio temporal patterns of species assemblages. J Biogeogr. https://doi.org/10.1111/j.1365-2699.2011.02550.x

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Hengl T (2009) A practical guide to geostatistical mapping. LULU 291 pp

Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: from equations to case studies. Comput Geosci 33:1301–1315

Hewitt GM (2011) Mediterranean peninsulas. The evolution of hotspots. In: Zachos F, Habel CJ (eds) Biodiversity hotspots. Distribution and protection of conservation priority areas. Springer, Berlin, pp 123–147

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25(15):1965–1978

Hortal J, Lobo JM (2005) An ED-based protocol for optimal sampling of biodiversity. Biodivers Conserv 14:2913–2947

Hortal J, Lobo JM (2006) Towards synecological framework for systematic conservation planning. Biodivers Inf 16–45

Hortal J, Lobo JM (2011) Can species richness patterns be interpolated from a limited number of well-known areas? Mapping diversity using GLM and kriging. Nat Conserv 9:200–207

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. Biomet J 50:346–363

Hunter ML, Yonzon (1993) Altitudinal distributions of birds, mammals, people, forests and parks in Nepal. Conserv Biol 7(2):420–423

Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley, New York

Kery M, Schmid H (2004) Monitoring programs need to take into account imperfect species detectability. Basic Appl Ecol 5:65–73

Kéry M, Gardner B, Monnerat C (2010) Predicting species distributions from checklist data using site-occupancy models. J Biogeogr 37:1851–1862

Liebhold AM, Rossi RE, Kemp WP (1993) Geostatistics and geographic information systems in applied insect ecology. Annu Rev Entomol 38:303–327

Lobo JM (2008a) Database records as a surrogate for sampling effort provide higher species richness estimations. Biodivers Conserv 17(4):873–881

Lobo JM (2008b) More complex distribution models or more representative data? Biodivers Inf 5:14–19

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz F (2000) Biotic invasions: causes, epidemiology, global consequences and control. Ecol Appl 10:689–710

Mckinney ML (2001) Effects of human population, area and time on non-native plant and fish diversity of the us. Biol Conserv 243–252

Meng Q (2014) Regression kriging versus geographically weighted regression for spatial interpolation. Int J Adv Remote Sens GIS 3(1):606–615

Mitić B, Boršić I, Dujmović I, Bogdanović S, Milović M, Cigić P, Rešetnik I, Nikolić T (2008) Alien flora of Croatia:

proposals for standards in terminology, criteria and related database. Natura Croatica 17(2):73–90

Nikolic T, Bukovec D, Šopf J, Jelaska SD (1998) Mapping of Croatian flora—possibilities and standards. Natura Croatica 7:1–62

Nikolić T, Mitić B, Milašinović B, Jelaska SD (2013) Invasive alien plants in Croatia as a threat to biodiversity of south-eastern Europe: distributional patterns and range size. CR Biol 336:109–121

Nikolić T, Mitić B, Ruščić M, Milašinović B (2014) Diversity, knowledge and spatial distribution of the vascular flora of Croatia. Plant Biosyst 148(4):591–601

Pauchard A, Kueffer K, Dietz H, Daehler CC, Alexander J, Edvards PJ, Arevalo JR, Cavieres LA, Guisan A, Haider S, Jakobs G, Mcdougall K, Millar CI, Naylor BJ, Parks CG, Rew LJ, Seipel T (2009) Ain't no mountain high enough: plant invasions reaching new elevations. Front Ecol Environ 7(9):479–486

Pebesma EJ (2004) Multivariable geostatistics in s: the gstat package. Comput Geosci 30(7):683–691

Pebesma EJ (2006) The role of external variables and GIS databases in geostatistical analysis. Trans GIS 10:615–632

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19(1):181–197

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol Econ 52:273–288

Pino J, Font X, Carbó J, Jové M, Pallarès L (2005) Large-scale correlates of alien plant invasion in Catalonia (NE of Spain). Biol Conserv 122:339–350

Rejmanek M (1999) Invasive plant species and invasible ecosystems. In: Sandlund OT, Schei PJ, Viken A (eds) Invasive species and bio-diversity anagement. Kluwer Academic Publishers, Dordrecht, pp 79–102

Robertson MP, Cumming GS, Erasmus BFN (2010) Getting the most out of atlas data. Divers Distrib 16:363–375

Rocchini D, Hortal J, Lengyel S, Lobo JM, Jimenez-Valverde A, Bacaro G, Chiarucci A (2011) Uncertainty in species distribution mapping and the need for maps of ignorance. Prog Phys Geogr 35:211–226

Roy H (2016) Invasive species: control wildlife pathogens too. Nature 530:281

R Core Team (2012) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.r-project.org/

Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Hanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global biodiversity scenarios for the year 2100. Science 287(5459):1770–1774

Santos AMC, Jones OR, Quicke DLJ, Hortal J (2010) Assessing the reliability of biodiversity databases: identifying evenly inventoried island parasitoid faunas (Hymenoptera: Ichneumonoidea) worldwide. Insect Conserv Divers 3:72–82. https://doi.org/10.1111/j.1752-4598.2010.00079.x

Scalera R, Genovesi P, Essl F, Rabitsch W (2012) The impacts of invasive alien species in Europe. European Environment Agency, Copenhagen

Schindler S, von Wehrden H, Poirazidis K, Wrbka T, Kati V (2013) Multiscale performance of landscape metrics as indicators of species richness of plants, insects and vertebrates. Ecol Ind 31:41–48

Schindler S, Staska B, Adam M, Rabitsch W, Essl F (2015) Alien species and public health impacts in Europe: a literature review. NeoBiota 27:1–23

Simberloff D (2013) Biological invasions: What's worth fighting and what can be won? Ecol Eng 65:112–121

SINP (2009) National habitat classification scheme. [Nacionalna klasifikacija staništa] (in Croatian)

Stein A, Gerstner K, Kreft H (2014) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecol Lett 17:866–880

Stohlgren TJ, Schnase JL (2006) Risk analysis for biological hazards: what we need to know about invasive species. Risk Anal 26:163–173

Taramarcaz P, Lambelet C, Clot B, Keimer C, Hauser C (2005) Ragweed (Ambrosia) progression and its health risks: Will Switzerland resist this invasion? Swiss Med Week 135:538–548

Thuiller W, Richardson MD, Midgley GF (2007) Will climate change promote alien plant invasions? In: Nentwig W (ed) Ecological studies biological invasions. Springer, Berlin

Tittensor DP, Walpole M, Hill SLL, Boyce DG, Britten GL, Burgess ND, Butchart SHM, Leadley PW, Regan EC, Alkemade R, Baumung R, Bellard C, Bouwman L, Bowles-Newark NJ, Chenery AM, Cheung WWL, Christensen V, Cooper HD, Crowther AR, Dixon MJR, Galli A, Gaveau V, Gregory RD, Gutierrez NL, Hirsch TL, Höft R, Januchowski-Hartley SR, Karmann M, Krug CB, Leverington FJ, Loh J, Kutsch Lojenga R, Malsch K, Marques A, Morgan DHW, Mumby PJ, Newbold T, Noonan-Mooney K, Pagad SN, Parks BC, Pereira HM, Robertson T, Rondinini C, Santini L, Scharlemann JPW, Schindler S, Sumaila UR, The SLS, van Kolck J, Visconti P, Ye Y (2014) A mid -term analysis of progress towards international biodiversity targets. Science 346(6206):241–244

Tscharntke T, Tylianakis JM, Rand TA, Didham RK, Fahrig L, Batary P, Bengtsson J, Clough Y, Crist TO, Dormann CF, Ewers RM, Frund J, Holt RD, Holzschuh A, Klein AM, Kleijn D, Kremen C, Landis DA (2012) Landscape moderation of biodiversity patterns and processes— eight hypotheses. Biol Rev 87:661–685

Vanderwal J, Falconi L, Januchowski S, Shoo L, Storlie C (2012) Sdmtools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises. R package version 1.1-13. http://cran.r-project.org/package=sdmtools

Wang K, Zhang C, Li W (2012) Comparison of geographically weighted regression and regression kriging for estimating the spatial distribution of soil organic matter. GISci Remote Sens 49(6):915–932

Woodward FI (1987) Climate and plant distribution. Cambridge University Press, Cambridge

Zachos FE, Habel CJ (eds) (2011) Biodiversity hotspots. Distribution and protection of conservation Priority areas. Springer, Berlin, p 546

Zaniewski AE, Lehmann A, Overton JM (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol Model 157:261–280