

Selecting Biologically Meaningful Environmental Dimensions of Low Discrepancy among Ranges to Predict Potential Distribution of Bean Plataspid Invasion

Gengping Zhu^{1,2*}, Matthew J. Petersen³, Wenjun Bu²

1 College of Environmental Science and Engineering, Nankai University, Tianjin, China, **2** College of Life Sciences, Nankai University, Tianjin, China, **3** Department of Entomology, Cornell University, Geneva, New York, United States of America

Abstract

Background: The Bean plataspid (*Megacopta cribraria*) (Hemiptera: Pentatomidae), native to Asia, is becoming an invasive species in North America; its potential spread to soybean producing areas in the US is of great concern. Ecological niche modelling (ENM) has been used increasingly in predicting invasive species' potential distribution; however, poor niche model transferability was sometimes reported, leading to the artifactual conclusion of niche differentiation during species' invasion.

Methodology/Principals: We aim to improve the geographical transferability of ENM via environmental variable selection to predict the potential distribution of Bean plataspid invasion. Sixteen environmental dimensions between native and introduced Bean plataspid populations were compared, and classified into two datasets with different degrees of discrepancy by the interquartile range (IQR) overlap in boxplot. Niche models based on these two datasets were compared in native model prediction and invading model projection. Classical niche model approaches (i.e., model calibrated on native range and transferred outside) were used to anticipate the potential distribution of Bean plataspid invasion.

Conclusions/Significance: Niche models based on the two datasets showed little difference in native model predictions; however, when projecting onto the introduced area, models based on the environmental datasets showing low discrepancy among ranges recovered good model transferability in predicting the newly established population of Bean plataspid in the US. Recommendations were made for selecting biologically meaningful environmental dimensions of low discrepancy among ranges to improve niche model transferability among these geographically separated areas. Outside of its native range, areas with invasion potential include the southeastern US in North America, southwestern Europe, southeastern South America, southern Africa, and the eastern coastal Australia.

Citation: Zhu G, Petersen MJ, Bu W (2012) Selecting Biologically Meaningful Environmental Dimensions of Low Discrepancy among Ranges to Predict Potential Distribution of Bean Plataspid Invasion. PLoS ONE 7(9): e46247. doi:10.1371/journal.pone.0046247

Editor: Brock Fenton, University of Western Ontario, Canada

Received: May 18, 2012; **Accepted:** August 29, 2012; **Published:** September 25, 2012

Copyright: © 2012 Zhu 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.

Funding: This research was supported by the China Postdoctoral Science Foundation (2012M510744), the National Education Project in Basic Science for Special Subjects (Insect Systematics, no. J0930005) and National Natural Science Foundation of China (no. 31071959). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gengpingzhu@hotmail.com

Introduction

Ecological niche modelling (ENM) has been used increasingly in predicting invasive species' potential distribution [1–4], and other aspects of ecology and evolution [5,6]. Based on integrating known occurrence with environmental variables, ENM seeks to characterize environmental conditions that are suitable for the species, and then to identify where suitable environments are distributed spatially [7]. The assumptions under which ENMs work are the equilibrium between species' distributions and ecological requirements, dispersal without limitation and niche conservatism [8]. The ecological niche of a species here can be defined as the set of environmental conditions under which it is able to maintain populations without immigrational subsidy [9,10].

The classical ENM approaches for invasive species prediction are to calibrate niche models in species' native range and then transfer models to identify areas of potential invasion. Niche model

transferability here refers to the capability of species' occurrence prediction from niche models projected onto novel areas. Several studies reported poor niche model transferability in predicting species' invasion (e.g., [11–14]), leading to the conclusion of niche differentiation during species' invasion, which violates one of the key assumptions of ENM (i.e., niche conservatism). However, other studies have demonstrated that such differences might be caused by sampling records [15,16], environmental datasets [4,17,18], or the method of ENM analysis [19,20]. Since the conclusions of niche differentiation can be due, in part, to poor niche model transferability, we hypothesized that if environmental dimensions were selected that reduced discrepancy between native and introduced populations, then niche model transferability might be improved among these areas. Some studies suggested niche model transferability could be improved by selecting fewer environmental dimensions for niche model calibration [4,17,21,22]. However, fewer environmental dimensions also

means fewer constraint on niche characterization (i.e., ENM calibration), which might result in higher commission error (predicted presence in areas of actual absence) (Personal communication with Dr. Peterson, Kansas University), although it indeed shows lower omission error (predicted absence in areas of actual presence).

Many true bugs (Hemiptera: Heteroptera) have extended their distributions remarkably in the last century, and some of them have increased their pest status after introduction beyond the native range [4,23–26]. Native to Asia, the Bean plataspid *Megacopta cribraria* (F.), also known as lablab bug or globular stink bug, is the first species of the family Plataspidae to be introduced into the Western Hemisphere [24]. The establishment was confirmed by specimens collected in northeastern Georgia, and by November 2010, it had been reported in Georgia, North Carolina, South Carolina, Tennessee, and Alabama [26]. The newly established US population has attracted much attention due to the damage caused by the species in its native area. In Asia, Bean plataspid is a serious pest of soybeans [27] where it damages young leaves, stems, and newly developed pods [26]. In China, the pest has caused soybean crop losses of 30–50% percent [28]. More importantly, recently it was found infesting soybean production areas in Georgia and South Carolina [29]. The only recent detections may indicate that Bean plataspid is still in early stages of invasion following introduction. Although the scope of the bugs' status as a US crop pest has not yet been determined, their potential for spread to large soybean producing areas in the US is of great concern [29].

In this study, we aim to improve niche model transferability by selecting biological meaningful environmental dimensions with low discrepancy among ranges (i.e., native and introduced) to predict the potential geographic distribution of Bean plataspid invasion. The environmental dimensions occupied by native and introduced populations were firstly compared, and classified into high and low discrepancy datasets by the interquartile range (IQR) overlap in boxplot. Principal component analysis was used to further compare the classified datasets in reduced dimensions. Classical niche model approaches (i.e., model based on native range and transferring to introduced areas) were then used to compare niche model transferability based on these two environmental datasets. The preferred models were chosen to anticipate the potential distribution of Bean plataspid across North America and the world. Model transferability responding to niche space comparison, and implications for variable selection to improve the niche model transferability were discussed in this study.

Materials and Methods

Occurrence data

Species occurrence localities were assembled from the literatures and from specimens deposited in the Institute of Entomology at Nankai University. Localities with ambiguous or unclear descriptions were excluded. All geographically recoverable localities were georeferenced in Google Maps, Gazetteer of China [30] or BioGeomancer (<http://bg.berkeley.edu/latest/>). Invasive records in the U.S. were largely county-level [31] and converted to points by digitizing the centroid of infested counties in Arc GIS 10 (ArcGIS, Environmental Systems Research Institute, Redlands, CA, USA). This method has been used in previous studies, with models based on random points not differing qualitatively from models based on such centre points [4,12,14]. A total of 166 and 132 occurrence records were prepared for native and introduced populations respectively.

Montandon described a species closely related to *M. cribraria*, i.e., *Megacopta punctatissima* (as *Coptosoma punctatissimum*, [32]), before later finding specimens that were intermediate between *M. cribraria* and *M. punctatissima* [33]. In a revision of the family Plataspidae from China, Yang considered *M. punctatissima* to be conspecific with *M. cribraria* [34]. Hosokawa et al. reported that the two species were capable of interbreeding and their offspring were found to reproduce successfully [35]. Herein, we treated *M. punctatissima* to be a junior synonym of *M. cribraria* and considered *M. cribraria* as the taxonomic entity and utilized occurrence records accordingly.

The native 166 occurrence points varied in spatial density due to variable sampling intensity over geography. These data might inflate measures of accuracy for presence-only niche models (e.g., Maxent or GARP, see below) [36]. As a result, and to avoid overemphasizing heavily on sampled areas, we selected points for model calibration using a subsampling regime to reduce sampling bias and spatial autocorrelation. Following Nuñez and Medley [37], we generated models using all available occurrence points and measured spatial autocorrelation among model pseudo-residuals ($1 - \text{probability of occurrence generated by model}$) by calculating Moran's *I* at multiple distance classes using SAM v4.0 [38]. Significance was determined using permutation tests. A minimum distance of 150 km was detected, we therefore created a grid with cell dimensions of $1.5 \times 1.5^\circ$ and selected the occurrence point that close to the centroid of each grid cell. This procedure reduced the number of native occurrences to 89 points used for model calibration, leaving the remaining points used for model testing. The procedure greatly reduced sampling bias and spatial autocorrelation, resulting in evenly distributed occurrence points across space [37].

Environmental variables

Environmental variables summarizing aspects of climate, topography and habitat were prepared to represent ecologically important factors known to impact the biological prosperity of Bean plataspid [39–41] (Table 1). Climate variables represented by bioclimatic factors of temperature and precipitation were derived from WorldClim [42], and of sunshine from CliMond [43]. Highly correlative variables were not included in the analysis. Topography variable represented by elevation was also obtained from the WorldClim database. Habitat variables were represented by the Normalized difference vegetation index (NDVI) derived from <http://edit.csic.es/>, re-calculated as the average of values for 12 months. All dimensions were set at a spatial resolution of 2.5 arc-min for analysis.

Direct comparison and Principal Component Analysis (PCA)

Raw environmental data was extracted from environmental rasters at species' occurrence records using ArcGIS 10, and compared in boxplot between native and introduced populations. Boxplot gives a good sense of environmental data distribution (median, minimum, maximum, and the first and third quartiles), that indicate the extent to which the data lies near the median, or near the extremes [44]. Here, the interquartile range (IQR) overlap between native and introduced populations was used to classify the environmental dimensions. The interquartile IQR is a measure of statistical dispersion, it is equal to the difference between the third (Q_3) and first (Q_1) quartiles (i.e., $IQR = Q_3 - Q_1$), which is represented by column in boxplot. Unlike total range, the IQR is a robust statistic, having a breakdown point of 25%, and is thus often preferred to the total range [44]. All boxplots were prepared in Sigmaplot 11.0 (Systat

Table 1. Principal components analysis (PCA) of environmental variables associated with occurrence of Bean plataspid.

Dataset I		Factor Loading		
Variable	Description	PC-1	PC-2	PC-3
BIO2	Mean diurnal temperature range	-0.70	0.54	0.12
BIO6	Minimum temperature of coldest month	0.79	0.01	0.57
BIO7	Temperature annual range	-0.83	0.14	-0.49
BIO13	Precipitation of wettest month	0.79	0.03	-0.15
BIO14	Precipitation of driest month	-0.77	-0.18	0.46
BIO15	Precipitation seasonality	0.87	0.24	-0.40
BIO21	Highest weekly radiation	-0.23	0.90	0.16
BIO22	Lowest weekly radiation	0.87	0.37	0.09
Eigenvalue		4.58	1.34	1.00
Percentage variance		57.31	16.80	12.48
Cumulative percentage variance		57.31	74.11	86.59
Dataset II		Factor Loading		
Variable	Description	PC-1	PC-2	PC-3
BIO1	Annual mean temperature	0.91	-0.12	-0.29
BIO3	Isothermality	0.75	0.45	0.42
BIO4	Temperature seasonality	-0.84	-0.45	0.22
BIO5	Maximum temperature of warmest month	0.58	-0.72	0.13
BIO12	Annual precipitation	0.10	0.39	-0.78
BIO20	Annual mean radiation	0.66	-0.10	0.26
DEM	Elevation	-0.16	0.82	0.03
NDVI	Normalized difference vegetation index	-0.06	0.54	0.50
Eigenvalue		2.88	2.04	1.24
Percentage variance		36.02	25.54	15.47
Cumulative percentage variance		36.02	61.56	77.04

Eigenvalues for the most important variables (>0.8) are in bold.
doi:10.1371/journal.pone.0046247.t001

Software Inc, Chicago, IL, USA), then classified into two datasets by IQRs (or boxplot columns) overlaps between the two populations. Of which, Dataset I with IQRs were not overlapped (i.e., BIO 2, 6, 7, 13, 14, 15, 21 and 22), Dataset II with IQRs were more or less overlapped (i.e., BIO 1, 3, 4, 5, 12, 20, DEM and NDVI) (Table 1, Figure 1).

Therefore, the Dataset I represented environmental dimensions of high discrepancy among the two ranges, and the Dataset II represented dimensions of low discrepancy. A principle component analysis (PCA) was then used to visualize the species niche space in reduced dimensions. Two correlation matrixes were prepared for the Dataset I and II to perform the PCA in SPSS 19 (IBM SPSS Statistics, Chicago, IL, USA). To facilitate visualization, occurrence records were grouped for native and introduced populations respectively. The PCA provided a representation of the bug's niche space across both ranges. The spread of points representing native and invasive populations in the PCA were compared [12,14]. If the species' niche is retained, invasive occurrences should cluster within the cloud of native range occurrences; separation of the two ranges in PCA space signifies a potential divergence from niche conservation, in a strict sense [22,45].

Two variable sets comparison

All models were developed using maximum entropy algorithm implemented in Maxent software (version 3.3.3k) [46–49]. In

exploring areas of potential invasion, another algorithm was used (i.e., GARP, see below). Maximum entropy is a machine-learning technique that predicts species distributions using detailed environmental variables associated with species occurrence. It follows the principle of maximum entropy and spreads out probability as uniformly as possible, but subject to the caveat that they must match empirical information such as known presence. Analysis was run on default program conditions (Logistic output, default convergence threshold (10^{-5}) and maximum number of iterations (500)). The logistic output with suitability values ranging from 0 (unsuitable habitat) to 1 (optimal habitat) gives an estimate of probability of presence, assuming that the sampling design is such that typical presence localities have probability of presence of about 0.5 [46,47]. A jack-knife procedure was used to evaluate the relative importance of each predictor variable and the ability to correctly predict new ranges in the model [50].

Niche models were calibrated against native range environmental data rasters clipped to appropriate size defined by a bounding box containing all known native range occurrences (i.e., the area defined as the geographic space available to the species). Constructed models using the above two environmental datasets were then transferred onto the US (not include Hawaii and Alaska) respectively. Maxent also calculates a Multivariate Environmental Suitability Surface (MESS) map indicating areas where environmental variables occur outside the range of values in the training

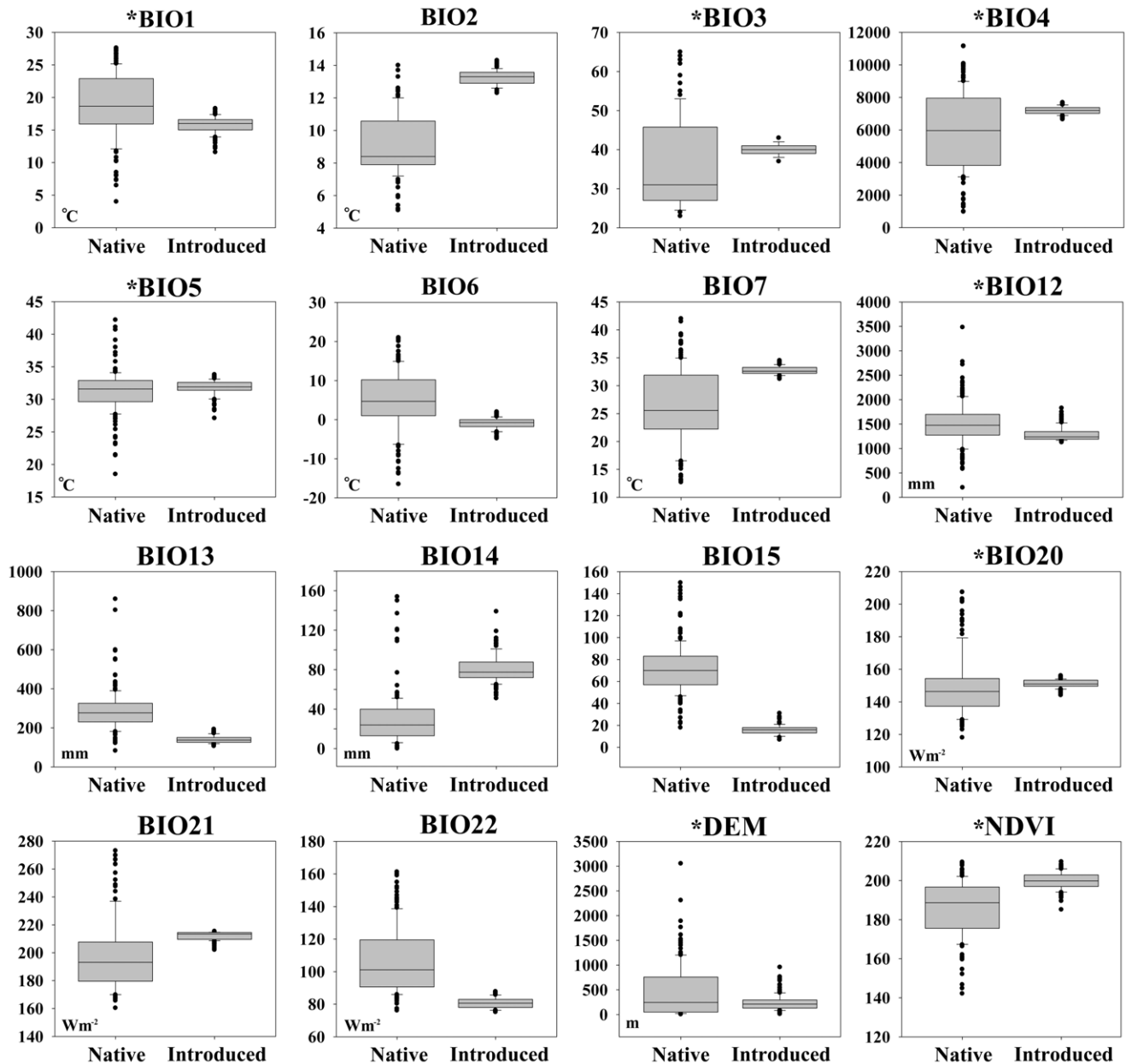


Figure 1. Direct comparison of Bean plataspid occurrence-associated variables between native and introduced distributional areas. The Dataset *I* (variables without asterisk) with columns (i.e., IQR) were not overlapped between the two areas representing variables of high discrepancy. The Dataset *II* (variables with asterisk) with columns were more or less overlapped representing variables of low discrepancy. doi:10.1371/journal.pone.0046247.g001

region, ENM suitability projections in these regions are unreliable, and should be treated cautiously [51]. MESS of native dimensions in contrast to the US (for Dataset *I*) and the world (for Dataset *II*) were prepared.

Area Under Curve (AUC) of the receiver operating characteristic (ROC) plot and omission rate were used for model evaluation. AUC weights the omission error and commission error equally, it is a threshold-independent measure that juxtaposes correct and incorrect predictions over the spectrum of threshold. AUC values range from 0 to 1, where 1 is a perfect fit. Useful models produce AUC values of 0.7–0.9, and models with ‘good discriminating ability’ produce AUC values above 0.9 [52]. Both AUC and omission rate were used for native model evaluation. Success of

models transferred to the US to correctly capture Bean plataspid occurrences was tested using omission rate. Omission rate assesses prediction error calculated by the proportion of test points that were not predicted at a particular threshold. We plotted omission rate across the threshold spectrum of Maxent’s logistic output values, specifically, we calculated omission rate at the increasing rate of 0.05 degrees against the total 1.0 logistic output.

Exploring areas of potential invasion

To explore areas of potential invasion globally, the Dataset *II* was used (Table 1). We calibrated models based on native range, and transferred their predictions onto the other continents. Considering that the record in the US does not characterize the

actual distribution, and the sample bias in native Asia, we used 89 occurrences of the reduced native sample for model calibration. Although Maxent has appeared superior to GARP in some previous studies [49], carefully assessments of model quality showed no significant differences between the two [53]. Recent studies suggested using multiple algorithms to infer a consensus estimate of niche dimensions [5,50,54–56]. Hence, we further used the Genetic Algorithm for Rule-set Prediction (GARP, [57]) to explore areas of potential invasion (Text S1). For model evaluation, we calculated binary omission rate of the remaining occurrence (including native and introduced records) at the threshold of the 10th percentile training presence, which assumed a grid cell was suitable if its suitability score was greater than the 10th percentile of training points. The 10th percentile threshold is highly conservative in estimating species' presence and has been more commonly used [2,58,59].

Results

Direct comparisons

Figure 1 summarized the sixteen environmental dimensions and their ranges among native and introduced Bean plataspid populations. PCA of environmental variables associated with Bean plataspid occurrence revealed reduced dimensions that accounted for the observed distribution (Table 1; Figure 2). The first three components were significant and explained 86.59% and 77.04% of the variance in Dataset *I* and *II*, respectively (Table 1). In Dataset *I*, the first component (PC-1) was associated with precipitation seasonality, lowest weekly radiation, and temperature annual range, the second (PC-2) was closely associated with highest weekly radiation, while the third (PC-3) was less clearly associated with single dimension. In Dataset *II*, the first component was associated with annual mean temperature and temperature seasonality, the second was associated with elevation while the third was less clearly associated with single dimension (Table 1). The niche space occupied by the introduced population departed from that occupied by the native population in the reduced dimensions in Dataset *I*, while in Dataset *II*, the orientation of introduced occurrences all fell more closely within those of native population (Figure 2). Comparing these two environmental datasets, one that showed niche difference between native and introduced populations (i.e., Dataset *I*), the other (i.e., Dataset *II*) did not.

Environmental datasets and model comparisons

In native predictions, models based on the two environmental datasets all showed good model performance compared to random predictions (Dataset *I*: AUC = 0.876; Dataset *II*: AUC = 0.867; Figure 3), and models trained on the two datasets showed little difference in omission rate across Maxent thresholds (Figure 4). However, when transferred onto the US, models based on Dataset *II* showed lower omission rates at the thresholds of 0.2 to 0.75, comparing to that based on Dataset *I*. At the thresholds of 0.25 to 0.5, the omission rates rose significantly in Dataset *I*, but stayed stable at a low level in Dataset *II* (Figure 4), suggesting good model transferability in Dataset *II*. Areas of potential invasion identified by Dataset *I* in the US include extensive areas of the lower Midwest and Southeast, Florida was also identified as suitable in Dataset *I*. In Dataset *II* projected model, the areas identified include most of the infested counties in Georgia, North Carolina and South Carolina, the other states including Mississippi, Alabama, Tennessee and Virginia, and the coastal areas along the Atlantic and Gulf of Mexico (Figure 3). All of these areas with

environmental variables were fell in that of the training data (Figure S1, S2).

Areas of potential invasion

Niche models based on the reduced 89 native points yield zero omission (Maxent) and 1.43% omission (GARP) of the independent test points (total 209 points), suggesting good model performance. Comparing to Maxent, the projection of GARP is a little conservative in Europe in contrast to other continents (Figure 5, Figure S3). Outside of native-range areas, high suitable areas identified by both modelling algorithms include the southeastern US in North America, southwestern Europe, southeastern South America, southern Africa, and the eastern coastal Australia (Figure 5, Figure S3). All of these areas with environmental variables were fell in that of the training data (Figure S2). Attention should be paid to quarantine and inspection when engaging in interchanges with the south Asia in these areas.

Discussion

Niche difference

Across its native range, the biology of Bean plataspid has been insufficiently studied to adequately understand the ecological constraints of the species [39–41,60]. A species' ecological niche can be characterized through either mechanistic or correlative approaches (i.e., ENM), with the former identifying the physiological determinants of a species' tolerance to environmental conditions, and the later identifying the ecological niche by associating known species' environmental tolerances derived from the conditions found at actual occurrences [7,18]. Correlative methods are made increasingly possible through the availability of species' occurrence data and geographic information system (GIS) data [6,7], which are particularly useful for poorly known species, such as Bean plataspid. In PCA, the annual mean temperature, temperature seasonality, temperature annual range, the highest/lowest weekly radiation, and elevation showed their significance in explaining the bug's distribution. In ENMs, the annual precipitation and radiation, temperature seasonality, the maximum temperature of the warmest month, and precipitation of the driest month appeared to be significant in model calibration.

However, care should be taken in interpreting their results, as ENM approaches were based on observations that already include effects of biotic interactions on species distributions [61], and thus recovered the realized or potential niche [20,62]. This portion of the fundamental niche, limited by biotic interaction or dispersal limitations, is unlikely to capture the full ecological tolerances of the species. In fact, the fundamental niche is rarely fully displayed in geographic space [20,22,63]. The spectrum of host plants available to Bean plataspid is limited due to an obligate relationship with a bacterial endosymbiont which allows it to feed on soybean and other legumes [26,35,64]. This association might influence the bug's potential for ecological expansion across the invasive range, reducing the occupancy of the range predicted by native range occurrences. It is unlikely that the current invasive range is in a distributional equilibrium, as many areas identified as suitable were not inhabited in the US. Therefore, our observed niche difference can be considered as the realized niche difference, the observed niche space in Dataset *I* and *II* can be considered as the realized niche manifested in two ecological dimensions (i.e., Dataset *I* and *II*).

The ENM here is to characterize the realized niche [20,62], whereas the PCA is to characterize the realized niche in specific ecological dimensions. In fact, the observed niche difference might tend to happen in the fine resolution environmental dimensions.

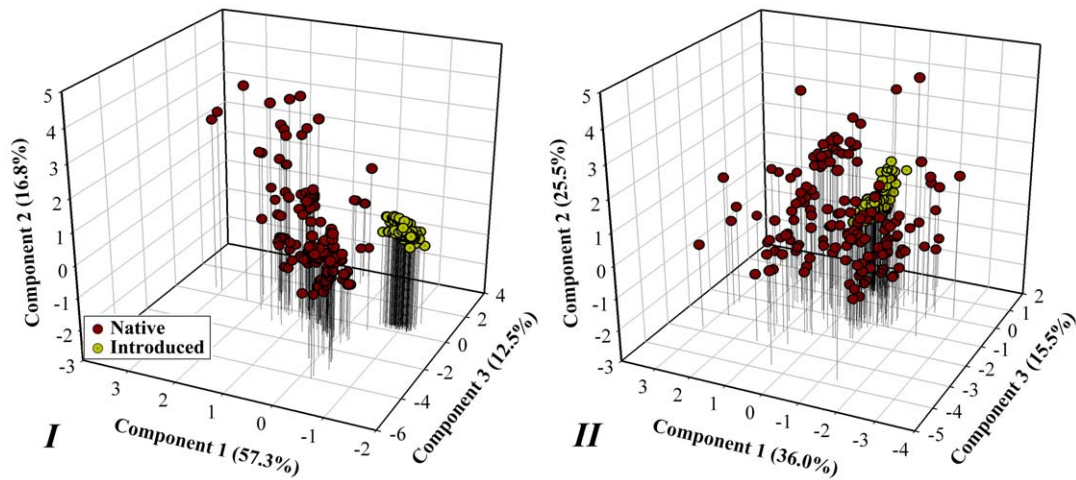


Figure 2. Principal component analysis of Dataset I (left) and II (right) associated with occurrences of Bean plataspid. Symbols represent Bean plataspid occurrences in native and introduced areas. doi:10.1371/journal.pone.0046247.g002

At small spatial extents, the fundamental niche might express differently because of heterogeneity of the landscape. While at broad spatial scale, the species interactions are weak, diffuse, or non-specific, it will be unusual to see such niche difference [61]. The observed “conserved” or “relaxed” variables [21] might not be simply because of their biological meaningful to the species (i.e., actually restricting species geographical range), but also because whether they were released by the fundamental niche in the heterogeneous landscape in which the species occur. In ecological dimensions, the realized niche difference might be displayed in some environmental axes, but not others [12]. The latter portion dimensions were also biological meaningful and might be more useful for niche model calibration. Actually, the other factors (e.g., sampling bias or environmental data quality) might also play a role in the observed niche difference, although they are not the topic here.

Environmental dimension and niche model transferability

Environmental variable selection is very important for niche model calibration. Apart from considering the biological factors that may restrict species’ distributions, the resolution, extent of

study range, and correlation among variables have to be taken into consideration [3,4]. In particular, comparing the ecological envelopes occupied by native and introduced populations offers useful information for variable selection prior to the modelling [4,12,21]. Actually, environmental dimensions had been compared using PCA and metrics summarizing differences between niches [11,14,45]. Herein, we found that after direct comparisons in boxplot, the environmental dimensions could be better classified into groups before further comparisons were made in the reduced dimensions of PCA. In the PCA, the realized niche in ecological dimensions of Dataset II is less discrepancy than that of Dataset I between native and introduced populations. Model projected onto the US based on Dataset II showed good model transferability in predicting the invasive occurrences (Figure 4), the niche model transferability here was improved via the PCA selecting environmental variables (i.e., Dataset II).

Model transfer outside the native range into the non-analog conditions is a challenge for niche modelling algorithms [65–67]. Many model protocols have been proposed to improve niche model transferability [56,68,69]. Putting the realized niche in ecological dimensions, the realized niches between native and introduced populations might be the same in some dimensions but

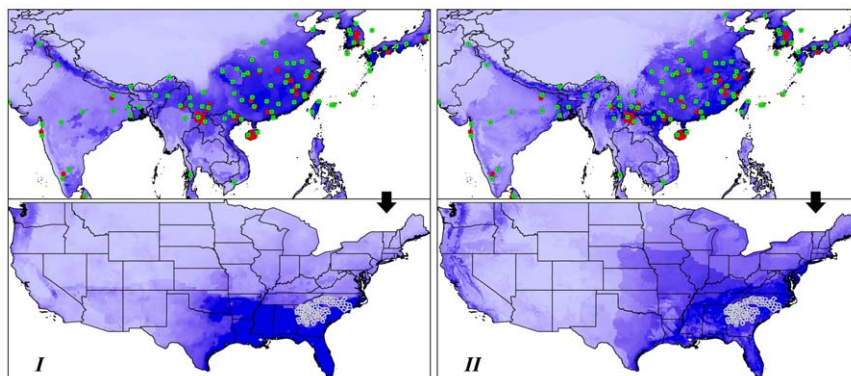


Figure 3. Niche model calibrated on native range and projected onto the US based on Dataset I (left) and II (right). Dark blue color represents high suitability, light blue indicates low suitability. Green dots indicate the occurrence points used for model calibration, red dots indicate the model testing point, white grids in the US indicate the infested counties. doi:10.1371/journal.pone.0046247.g003

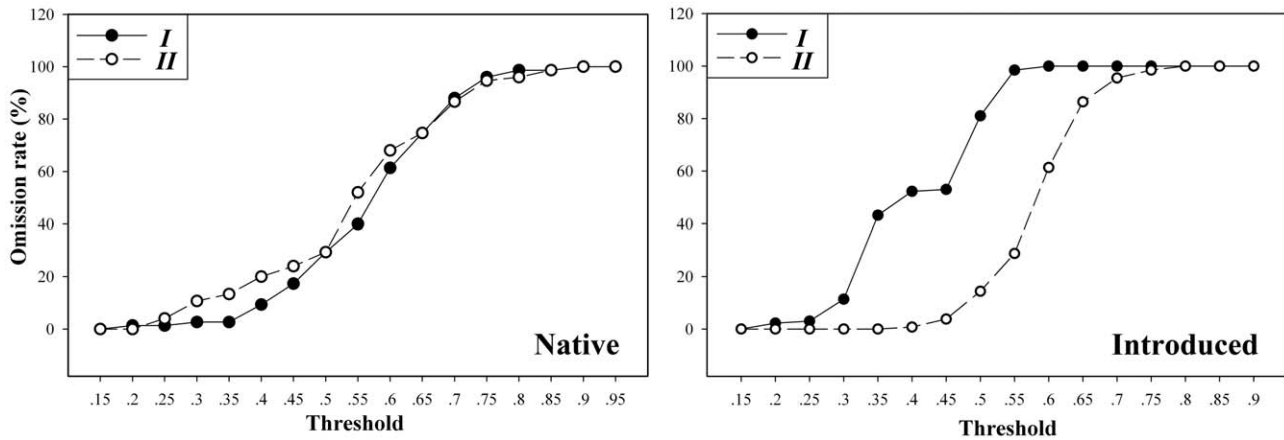


Figure 4. Omission rates among niche models based on Dataset I and II. Omission rates were plotted in native Asia models and their transferring in the US across the threshold spectrum of Maxent. doi:10.1371/journal.pone.0046247.g004

different in other dimensions [12,21]. Niche space comparison using PCA offers useful information on environmental dimensions of low or high discrepancy among ranges prior to the modelling. Our proposal was to select environmental dimensions with low discrepancy to improve niche model transferability among these distributional areas. When transferring niche models beyond the native area onto a larger area (comparing to model calibrating area) or globally, the environmental dimensions with low discrepancy and reduced dimensionality might be preferable (e.g., [4]). Our emphasis of selecting environmental dimensions of low discrepancy to improve niche model transferability by no means diminishes the importance of physiological relevance (e.g., [18]), or the effect of correlativeness of the variables (e.g., [3]).

Areas of potential invasion

High suitable areas were identified by Dataset II based model in the US, including the infested states (i.e., Georgia, North Carolina and South Carolina) and large portions of the surrounding states (i.e., Alabama, Tennessee, Mississippi, Louisiana, and Virginia). Other regions, such as Florida and Louisiana, were not supported by Dataset II but by Dataset I. Considering the bug's propensity to fly, land on, and get inside vehicles, inspection and quarantine should be considered for these states when engaging commercial activity with the infested states. Currently, the U.S. Department of Agriculture's (USDA) Animal and Plant Health Inspection Service (APHIS) is conducting surveys for the Bean plataspid to determine

the extent of infestation in Georgia, North Carolina and South Carolina [31]. Attention should also be paid to the areas of high-suitability that surrounding these infested states (Figure 3), and to areas predicted around the world (Figure 5, S3). This should be especially true for the developed areas where intensive trade activity and commercial interchange might facilitate new invasions.

While large areas of the US are proposed to be suitable to the Bean plataspid, these areas represent niche space without consideration of potential biotic interactions or the species dispersal ability. Many additional factors may influence the successful establishment of a non-indigenous species into a novel community, including existing species richness, competitors, predators, food availability, human footprint and climate similarity compared with the source ecosystem [21]. Although the area predicted as suitable for a species does not necessarily mean it will establish populations there, it does offer useful information for determining areas of potential invasion and spread.

Conclusion

Sixteen environmental dimensions occupied by native and introduced Bean plataspid populations were compared and classified into two datasets with different degrees of discrepancy among the two ranges. Niche models based on the dataset showing low discrepancy recovered good model transferability in the

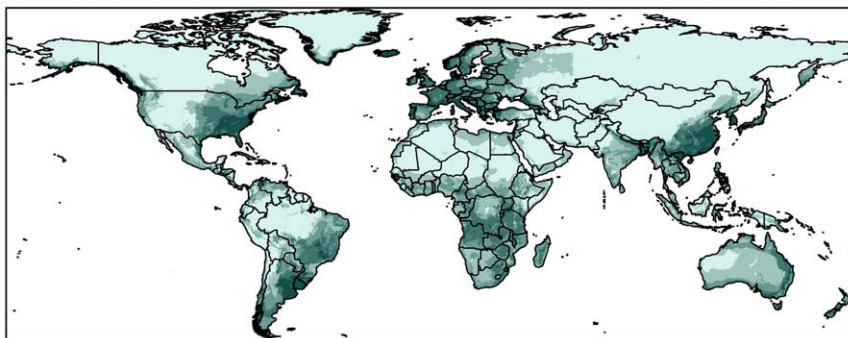


Figure 5. Niche models based on Dataset II and transferred worldwide using Maxent. Dark green color represents high suitability, light green indicates low suitability. doi:10.1371/journal.pone.0046247.g005

introduced areas, with low omission in recovering species occurrences. Recommendations were made for selecting biological meaningful environmental dimensions of low discrepancy among populations to improve niche model transferability among these geographically separated areas. When transferring niche model onto a larger area or globally, environmental dimensions with low discrepancy and reduced dimensionality were proposed. Outside of its native range, areas with invasion potential include the southeastern US in North America, southwestern Europe, southeastern South America, southern Africa, and the eastern coastal Australia. In the US, the states surrounding the current infested states (i.e., Georgia, North Carolina and South Carolina), including Alabama, Tennessee, Mississippi, Virginia and Florida should be monitored carefully as a result.

Supporting Information

Figure S1 MESS map for Dataset I in model comparison. Areas in red indicate one or more environmental variables outside the range present in the training data. (TIF)

Figure S2 MESS map for Dataset II when transferred the model worldwide. Areas in red indicate one or more

environmental variables outside the range present in the training data. (TIF)

Figure S3 Niche models based on Dataset II and transferred worldwide using GARP. Dark green color represents high suitability, light green indicates low suitability. (TIF)

Text S1 GARP protocol in exploring area of potential invasion. (DOCX)

Acknowledgments

We would like to thank Dr. Dávid Rédei (Hungarian Natural History Museum, Hungary) for the assistance in occurrence records collection and validation, and two anonymous reviewers for their helpful comments on an earlier version of this manuscript. Special thanks to Prof. A. Townsend Peterson (University of Kansas, USA) for valuable comments on model protocols and linguistic writing.

Author Contributions

Conceived and designed the experiments: GZ MJP WB. Performed the experiments: GZ. Analyzed the data: GZ WB. Contributed reagents/materials/analysis tools: GZ WB. Wrote the paper: GZ MJP.

References

- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78: 419–433.
- Liu X, Guo Z, Ke Z, Wang S, Li Y (2011) Increasing potential risk of a global aquatic invader in Europe in contrast to other continents under future climate change. *PLoS ONE* 6: e18429.
- Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, et al. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions* 13: 2785–2797.
- Zhu G, Bu W, Gao Y, Liu G (2012) Potential geographic distribution of Brown Marmorated Stink Bug invasion (*Halyomorpha halys*). *PLoS ONE* 7: e31246.
- Araújo MB, Whittaker RJ, Ladle RJ, Erhard M (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography* 14: 529–538.
- McCormack JE, Zellmer AJ, Knowles LL (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation? Insights from tests with niche models. *Evolution* 64: 1231–1244.
- Pearson RG (2007) Species' distribution modeling for conservation educators and practitioners. Synthesis. American Museum of Natural History. URL: <http://ncep.amnh.org>. Accessed 2011 October 10.
- Peterson AT, Soberón J, Sánchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science* 285: 1265–1267.
- Grinnell J (1917) The niche-relationships of the California thrasher. *Auk* 34: 427–433.
- Grinnell J (1924) Geography and evolution. *Ecology* 5: 225–229.
- Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson AT, et al. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10: 701–709.
- Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography* 16: 24–33.
- Gallagher RV, Beaumont LJ, Hughes L, Leishman MR (2010) Evidence for climatic niche and biome shifts between native and novel ranges in plant species introduced to Australia. *Journal of Ecology* 98: 790–799.
- Medley KA (2010) Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. *Global Ecology and Biogeography* 19: 122–133.
- Menke SB, Holway DA, Fisher RN, Jetz W (2009) Characterizing and predicting species distributions across environments and scales: Argentine ant occurrences in the eye of the beholder. *Global Ecology and Biogeography* 18: 50–63.
- Feeley KJ, Silman MR (2011) Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and Distributions* 17: 1132–1140.
- Peterson AT, Nakazawa Y (2008) Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography* 17: 135–144.
- Rödger D, Schmidtlein S, Veith M, Lötters S (2009) Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? *PLoS ONE* 4: e7843.
- Mandle L, Warren DL, Hoffmann MH, Peterson AT, Schmitt J, et al. (2010) Conclusions about niche expansion in introduced *Impatiens walleriana* populations depend on method of analysis. *PLoS ONE* 5: e15297.
- Soberón J, Peterson AT (2011) Ecological niche shifts and environmental space anisotropy: a cautionary note. *Revista Mexicana de Biodiversidad* 82: 1348–1355.
- Rödger D, Lötters S (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography* 18: 674–687.
- Peterson AT (2011) Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography* 38: 817–828.
- Rabitsch W (2008) Alien true bugs of Europe (Insecta: Hemiptera: Heteroptera). *Zootaxa* 1827: 1–44.
- Eger JEJ, Ames LM, Suiter DR, Jenkins TM, Rider DA, et al. (2010) Occurrence of the Old World bug *Megacopta cribraria* (F.) (Heteroptera: Plataspidae) in Georgia: A serious home invader and potential legume pest. *Insecta Mundi* 0121: 1–11.
- Fent M, Kment P (2011) First record of the invasive western conifer seed bug *Leptoglossus occidentalis* (Heteroptera: Coreidae) in Turkey. *North-Western Journal of Zoology* 7: 72–80.
- Jenkins TM, Eaton TD (2011) Population genetic baseline of the first plataspid stink bug symbiosis (Hemiptera: Heteroptera: Plataspidae) reported in North America. *Insects* 2: 264–272.
- Kikkuchi A, Kobayashi H (2010) Effect of injury by adult *Megacopta punctatissima* (Montandon) (Hemiptera: Plataspidae) on the growth of soybean during the vegetative stage of growth. *Japanese Journal of Applied Entomology and Zoology* 54: 37–43.
- USDA (2010) Invasive insect (Bean plataspid) poses risk to soybean crops and infests homes in southeastern states. U.S. Dept. of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine Factsheet. URL: <http://www.aphis.usda.gov/publications/>. Accessed 2011 October 10.
- Suiter DR, Ames LM, Eger JEJ, Gardner WA (2010) *Megacopta cribraria* as a nuisance pest. URL: <http://www.caes.uga.edu/publications/>. Accessed 2011 October 10.
- Gazetteer of China (1997) Sinomaps Press, Beijing, China.
- APHIS (2004) NAPIS PestTracker: A public website of the NAPIS/CAPS database. URL: <http://www.ceris.purdue.edu/napis/>. Accessed 2011 October 10.
- Montandon AL (1896) Plataspidae. Nouvelle série études et descriptions. *Annales de la Société Entomologique de Belgique* 40: 86–134.
- Montandon A L (1897) Les Plataspides du Muséum d'histoire naturelle de Paris. *Annales de la Société Entomologique de France* 1896: 436–464.
- Yang WI (1934) Revision of Chinese Plataspidae. *Bulletin of the Fan Memorial Institute of Biology* 5: 137–236.
- Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T (2006) Strict host-symbiont co-speciation and reductive genome evolution in insect gut bacteria. *PLoS Biology* 10: 1841–1851.

36. Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* 36: 2290–2299.
37. Nuñez MA, Medley KA (2011) Pine invasions: climate predicts invasion success; something else predicts failure. *Diversity and Distributions* 17: 703–713.
38. Rangel TF, Diniz-Filho JAF, Bini LM (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography* 15: 321–327.
39. Tayutivutikul J, Yano K (1990) Biology of insects associated with the kudzu plant, *Pueraria lobata* 2. *Megacopta punctissimum* (Hemiptera: Plataspidae). *Japanese Journal of Entomology* 58: 533–539.
40. Li YH, Pan ZS, Zhang JP, Li WS (2001) Observation on the biological property of *Megacopta cribraria* (F.). *Plant Protection Technology and Extension* 21: 11–12.
41. Chen Q, Wang JL, Guo SJ, Bai HX, Zhuo XN (2009) Study on the biological property of *Megacopta cribraria* (F.). *Journal of Henan Agricultural Science* 4: 88–90.
42. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
43. Kriticos DJ, Webber BL, Leriche A, Ota N, Macadam I, et al. (2011) CliMond: global high resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution* 3: 53–64.
44. Upton G, Cook I (1996) *Understanding Statistics*. Oxford University Press. p. 55.
45. Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62: 2868–2883.
46. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
47. Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.
48. Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, et al. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43–57.
49. Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, et al. (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29: 129–151.
50. Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102–117.
51. Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1: 330–342.
52. Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293.
53. 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.
54. Pearson RG, Thuiller W, Araujo MB, Martinez-Meyer E, Brotons L, et al. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography* 33: 1704–1711.
55. Waltari E, Hijmans RJ, Peterson AT, Nyári AS, Perkins SL, et al. (2007) Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* 2: e563.
56. Marmion M, Parviainen M, Luoto M, Risto K, Heikkinen RK, et al. (2009) Evaluation of consensus methods in predictive species distribution modeling. *Diversity and Distributions* 15: 59–69.
57. Stockwell DRB, Peters DP (1999) The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Systems* 13: 143–158.
58. Brito JC, Santos X, Pleguezuelos JM, Sillero N (2008) Inferring evolutionary scenarios with geostatistics and geographical information systems for the viperid snakes *Vipera latastei* and *Vipera monticola*. *Biological Journal of the Linnean Society* 95: 790–806.
59. Ficetola GF, Thuiller W, Padoa-Schioppa E (2009) From introduction to the establishment of alien species: bioclimatic differences between presence and reproduction localities in the slider turtle. *Diversity and Distributions* 15: 108–116.
60. Zhang SM (1985) *Economic insect fauna of China*, Fascicle 31, Hemiptera (1). Science press, Beijing, China.
61. Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2: 1–10.
62. Sillero N (2011) What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling* 222: 1343–1346.
63. Godsoe W (2010) Regional variation exaggerates ecological divergence in niche models. *Systematic Biology* 59: 298–306.
64. Fukatsu T, Hosokawa T (2002) Capsule-transmitted gut symbiotic bacterium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. *Applied and Environmental Microbiology* 68: 389–396.
65. Fitzpatrick MC, Hargrove WW (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation* 18: 2255–2261.
66. Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35: 276–288.
67. Zurell D, Elith J, Schröder B (2012) Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions*, doi: 10.1111/j.1472-4642.2012.00887.x.
68. Araujo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42–47.
69. Beaumont IJ, Gallagher RV, Thuiller W, Downey PO, Leishman MR, et al. (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions* 15: 409–420.