Using district-level occurrences in MaxEnt for predicting the invasion potential of an exotic insect pest in India

Sunil Kumar *, Jim Graham [1], Amanda M. West, Paul H. Evangelista

Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, United States

ABSTRACT

Insect pests are a major threat to agricultural biosecurity across the world, causing substantial economic losses. Majority of the species distribution modeling studies use precise coordinates (latitude/longitude) of species occurrences in MaxEnt (or maximum entropy model). However, lack of precise coordinates of insect pest occurrences at national/regional level is a common problem for many countries including India. This is because of the limited resources, lack of nationally coordinated surveys, and growers/farmers' privacy issues; district-level occurrences are commonly available (e.g., National Agricultural Pest Information System or NAPIS in the United States; http://pest.ceris.purdue.edu/). We demonstrated the use of MaxEnt to generate a preliminary, district-level map of the potential risk of invasion by an exotic cotton mealybug *Phenacoccus solenopsis* (Tinsley) (Hemiptera: Pseudococcidae) in India. District-level occurrence data were integrated with bioclimatic variables (values averaged within districts) using MaxEnt. The MaxEnt model performed better than random with an average test AUC value of 0.86 (±0.05). Our model predictions matched closely with the documented occurrence of *P. solenopsis* in all nine cotton growing states, and also predicted suitable habitats in other districts across India. The greatest threat of *P. solenopsis* infestations were predicted in most districts of Gujarat, Maharashtra, Andhra Pradesh, southwestern Punjab, northwestern Rajasthan, and western Haryana. Precipitation of coldest quarter, temperature annual range, and precipitation seasonality were the strongest predictors associated with *P. solenopsis* distribution. Precipitation of coldest quarter was negatively correlated with *P. solenopsis* occurrence. Mapping the potential distribution of invasive species is an iterative process, and our study is the first attempt to model national-level risk assessment of *P. solenopsis* in India. Our results can be used for selecting monitoring and surveillance sites and designing local, regional and national-level integrated pest management policies for cotton and other cultivated crops in India. The maps of potential pest distributions are urgently needed by agriculture managers and policymakers. Our approach can be used in other countries that lack precise coordinates of insect pest occurrences and generate a preliminary map of potential risk because it may be too late to wait for the precise coordinates of pest occurrences to generate a perfect map.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Invasive species are one of the major and most rapidly growing threats to agricultural biosecurity, livelihoods, human and animal health, forestry and biodiversity and result in huge economic losses (Davis, 2009; Pimentel, 2011). Growing trade and transportation along with other elements of globalization are facilitating invasions at an unprecedented rate (Levine and D'Antonio, 2003; Hulme, 2009). Mealybugs belong to one of the more common groups of small sap-sucking insects. They are considered a major agricultural pest on multiple continents causing serious problems (e.g. crop failure) when introduced to new geographic areas (Miller et al., 2002). Recent infestations of exotic cotton mealybug *Phenacoccus solenopsis* (Tinsley) in nine cotton growing states of India and several states in Pakistan have resulted in millions of dollars of damage to cotton crops and increased need for insecticides and other preventive measures (Aheer et al., 2009; Nagrare et al., 2009). The importance of cotton to India cannot be overstated; one-quarter of worldwide cotton acreage is planted there and over 60 million citizens depend on this crop for livelihood (Raju et al., 2008; NCIPM, 2009).

* Corresponding author. Address: 1499 Campus Delivery, Colorado State University, Fort Collins, CO 80523, United States. Tel.: +1 970 491 7056; fax: +1 970 491 1965.

E-mail address: sunil.kumar@colostate.edu (S. Kumar).

[1] Current address: Environmental Science and Management, Humboldt State University, Arcata, CA, United States.

P. solenopsis is native to the United States, and was first described by Tinsley (1898) in New Mexico, where it is found widespread on several ornamental and fruit crops. It is considered an exotic pest in Southeast Asia and other countries around the world including Argentina, Australia, Brazil, Chile, China, Ecuador, India and Pakistan (Vennila et al., 2010). The earliest *P. solenopsis* infestations in India were recorded in 2005 in Gujarat state (Jhala et al., 2008). It spread rapidly after it was first introduced, and was reported from all nine cotton growing states of India by 2008 (Nagrare et al., 2009). Wang et al. (2010) developed a global potential distribution map of *P. solenopsis* using average climate surface variables at 0.5° × 0.5° spatial resolution (approximately 55 km at the equator). Their map is too coarse to use for a national or regional scale planning and decision making. Detailed information on the potential habitat distribution of *P. solenopsis* in India remains unknown, and there is concern that future infestations may impact cotton and other cultivated crops such as okra, tomato, chili peppers, brinjal, and potato. Risk maps that show the potential distribution of *P. solenopsis* will be important management tools for early detection and monitoring, and integrated pest management planning (Macfadyen and Kriticos, 2012).

P. solenopsis is a polyphagus insect pest with a wide range of host plants including all four species of cotton and hundreds of other plant species. There are several natural enemies of mealybugs that control their populations including the parasitoid *Aenasius bambawalei* Hayat (Hymenoptera: Encyrtidae) (Prasad et al., 2011). *P. solenopsis* exhibited obligate sexual ovoviviparous reproduction (Prasad et al., 2012) and generally lays 500–600 eggs. It has a life cycle of 24–30 days and a female mealybug may produce 10–15 generations per year (Hanchinal et al., 2011). Initially the insect breeds on weeds such as parthenium (*Parthenium hysterophorus*), milkweed (*Asclepias* spp.), *Chenopodium* spp., and datura (*Datura alba*), and later migrates to cotton (*Gossypium hirsutum*) and other crops. Mealybug nymphs spread from infected to healthy plants via wind, irrigated water, rain, ants, and birds or by sticking/clinging to equipment, animals or people (Tanwar et al., 2007). Mealybugs can feed on all parts of a plant, but prefer actively growing leaf tissue, petioles, and leaf veins. They damage the plants by sucking sap from leaves, twigs, stems, roots and fruiting bodies. They inject toxic saliva into the plant parts causing chlorosis, stunting, deformation and death of plants (Tanwar et al., 2007).

Distribution and abundance of insects are highly influenced by climatic factors (temperature, moisture, humidity and their seasonal variations); especially the effects of temperature (Sutherst, 2000; Bale, 2002). Temperature and soil moisture may also interact to affect different developmental life stages of an insect. Temperature is one of the most influential environmental factors that affects distribution and abundance of different species of mealybugs (Amarasekare et al., 2008; Chong et al., 2003, 2008; Kim et al., 2008; Prasad et al., 2012). A detailed study examining the effects of temperature on the life cycle of *P. solenopsis* under laboratory conditions found lower development temperature thresholds for female and male at 11.7 °C and 10.1 °C, respectively. Female development was optimum at 32 °C and the upper temperature threshold for *P. solenopsis* development was around 39 °C (Prasad et al., 2012). Under laboratory conditions, relative humidity between 40% and 90% was found to be adequate for sustaining *P. solenopsis* populations (Vennila et al., 2010; Nagrare et al., 2011). Rainfall has been found to reduce the severity of *P. solenopsis* but it can also increase pest incidence because rainwater splashes act as a dispersal vector (Vennila et al., 2010).

Ecological niche models (ENM) and species distribution models (SDM) integrate species occurrence records with climatic and other environmental variables and generate maps of species potential or realized distribution (Bentlage et al., 2013). The distribution maps produced by ENM/SDM are used to design scientific surveys and manage insect pest infestations. These models can also identify environmental factors that limit a species' distribution. ENM/SDM approach is increasingly being used to map potential distributions of many species including insect pests (De Meyer et al., 2010; Wang et al., 2010; Evangelista et al., 2011; Parsa et al., 2012). In this study we used maximum entropy modeling (or MaxEnt) to predict the invasion potential of an exotic cotton mealybug, *P. solenopsis*. We hypothesized that an ENM/SDM will be able to predict the potential distribution of *P. solenopsis* using district-level occurrence data with high accuracy, and that climate factors alone would be good predictors. Our objectives were to: (1) generate a preliminary district-level map of the potential distribution of *P. solenopsis* in India, (2) quantify relative risk of invasion by *P. solenopsis* across all Indian states, and (3) identify bioclimatic factors associated with *P. solenopsis* distribution.

2. Materials and methods

2.1. Occurrence records and climate data

Geographic coordinates (i.e., latitude and longitude) of locations where a species was found present are typically used for ENM/SDM. For this study, precise locality coordinates for *P. solenopsis* were not available, so the district-level occurrence data published by Nagrare et al. (2009) were used (n = 42 records). Additional district-level occurrences (n = 11) for Karnataka, Tamil Nadu and Madhya Pradesh states were obtained from other published articles (e.g., Hanchinal et al., 2009, 2010, 2011) and Krishi Vigyan Kendra (Agricultural Science Center) Action Plan reports for different districts published in 2008, 2009, and 2010. We could not use GoogleEarth to generate approximate coordinates of *P. solenopsis* occurrence in cotton fields in the above districts because a GIS layer for cotton crop in India was unavailable. Therefore, a total of 53 district-level records (Fig. 1a) were used to generate a preliminary, district-level map of potential distribution for *P. solenopsis*, thus making use of the best available data. We did not use centroids of districts as surrogates of species occurrence points as some authors have done (e.g., Asian tiger mosquito, *Aedes albopictus* (Benedict et al., 2007; Medley, 2010). The centroid method may be acceptable if the target scale of prediction is global but may not be appropriate at national, state or finer scales; districts are not homogeneous, and some of them can be quite large. We calculated district-level averages of climatic variables in ArcMap (version 9.3, ESRI, Redlands, CA, USA) and used those as predictors. This is a relatively unconventional use of ENM/SDM, and the results may be useful for designing detailed surveys and making district-level state, regional or national pest management policies before more detailed, precise data for this species become available.

We obtained 19 bioclimatic data layers from the WorldClim dataset (Hijmans et al., 2005; http://www.worldclim.org/; Table 1) at ~1-km spatial resolution to represent current climatic conditions. The WorldClim dataset was generated using an interpolation technique using altitude and monthly temperature and precipitation records from 1950 to 2000. The 19 bioclimatic variables that define general trends, seasonality and extremes are considered biologically more meaningful than simple monthly or annual averages of temperature and precipitation in defining a species' ecophysiological tolerances (Nix, 1986; Kumar et al., 2009). We checked all bioclimatic variables for high cross-correlations using Pearson correlation coefficient ($r \geqslant 0.70$ or $\leqslant -0.70$). To reduce problems due to multicollinearity (Dormann et al., 2013) we included only one variable from a set of highly correlated variables (Appendix A). The decision to include or drop one of each set of highly correlated variables was made based on their potential biological relevance to *P. solenopsis* and their relative predictive power
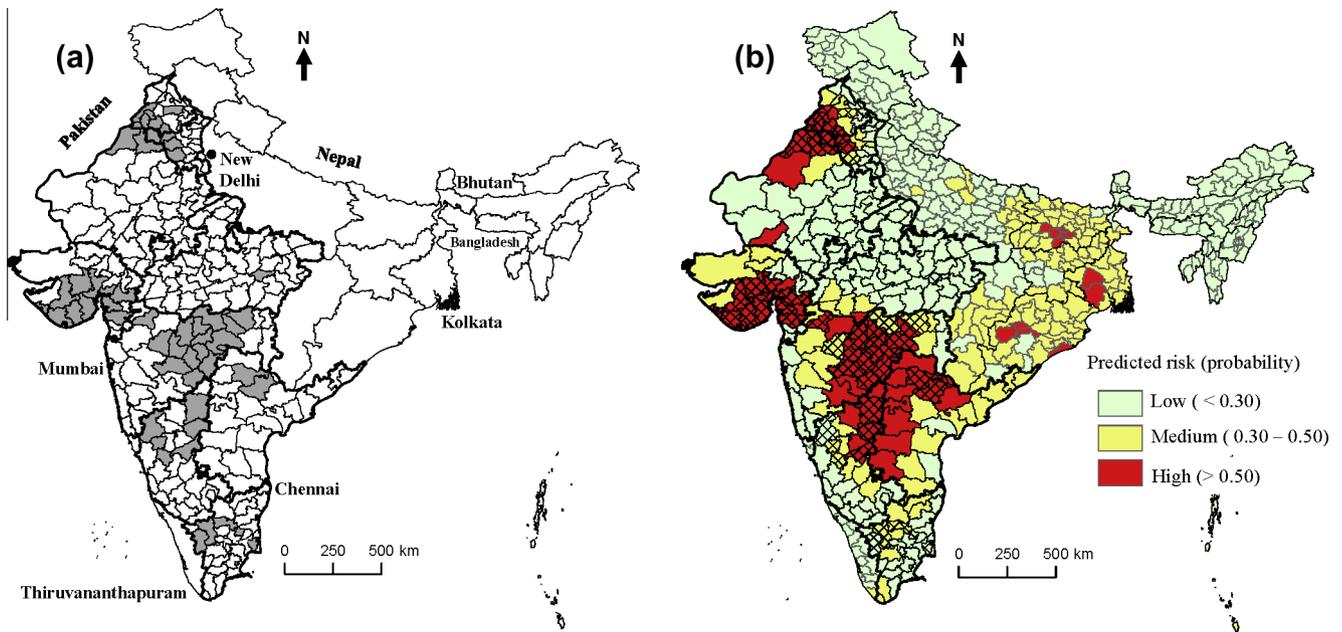
Fig. 1. (a) Current *P. solenopsis* occurrences (*n* = 53; grey shaded districts) in nine cotton growing states (represented by darker boundaries) in India, and (b) predicted potential risk of invasion by *P. solenopsis* in India; cross-hatched districts currently have *P. solenopsis*.

Table 1
Relative contribution of different bioclimatic variables to MaxEnt model for *P. solenopsis*. Percent contribution values are averages over 100 replicate runs. General statistics show the bioclimatic profile of *P. solenopsis* and were calculated based on 53 district-level *P. solenopsis* occurrence records.

Variable	Percent contribution	Mean	Standard deviation	Minimum	Maximum
Precipitation of coldest quarter (Bio19; mm)	30.8	41.2	64.7	1.2	358.7
Temperature annual range (Bio7; °C)	26.8	27.5	6.5	14.5	37.3
Precipitation seasonality (CV) (Bio15)	23.3	119.0	21.3	72.4	160.4
Precipitation of wettest month (Bio13; mm)	10.4	218.7	89.5	62.8	539.2
Annual mean temperature (Bio1; °C)	8.8	26.2	1.1	24.2	28.6

assessed based on training gain. Some variables were dropped because of their lower predictive power (i.e., percent contribution and jackknife training gain). The final model included only five bioclimatic variables (Table 1).

2.2. Modeling procedure

We used maximum entropy modeling or MaxEnt algorithm (version 3.3.3k; (Phillips et al., 2006) for quantifying relative risk of invasion and mapping the potential geographic distribution of *P. solenopsis* in India. We chose MaxEnt because: (1) it is a presence-only modeling algorithm (no absence data are needed), (2) it has performed relatively better than other modeling methods (Elith et al., 2006; Evangelista et al., 2008; Kumar et al., 2009), and (3) it is relatively robust to small sample sizes (Pearson et al., 2007; Kumar and Stohlgren, 2009). MaxEnt estimates the probability of presence of a species based on presence records and randomly generated background points by finding the maximum entropy distribution (Phillips et al., 2006). It uses a regularization parameter to control overfitting and can handle both categorical and continuous variables. MaxEnt uses five different features (linear, quadratic, product, threshold, and hinge) that constrain the geographical distribution of a species. The output from MaxEnt is an estimate of habitat suitability for a species that generally varies from 0 (lowest) to 1 (highest). We tested different settings in MaxEnt by varying regularization parameter, number of iterations and feature types. However, default settings in MaxEnt

yielded the best model for *P. solenopsis*. Our final model included only linear, quadratic and hinge feature types which could be due to the smaller sample size. Since we did not have precise coordinates of *P. solenopsis* occurrence, we used district-level occurrence as 'presence' locations in MaxEnt. We used 'samples with data' (SWD) format in MaxEnt using district level summaries of climatic variables and latitude and longitude of district centers as placeholders (MaxEnt does not use latitude/longitude information in model fitting when using SWD format). MaxEnt randomly selects 10,000 background points from the landscape. However, we restricted background point selection to nine cotton growing states and used 203 districts (excluding 53 presence districts) in MaxEnt. The resulting model using 53 presences and 203 background points from the nine states was then projected to all districts across India to identify potential new areas of invasion. Model predictions for all the districts were brought into a geographic information system (GIS) and maps were generated using ArcMap. Three arbitrary categories of risk of invasion by *P. solenopsis* were defined as low (<0.3), medium (0.3–0.5) and high (>0.5) based on predicted habitat suitability.

2.3. Model evaluation and validation

Area Under the ROC (receiver operating characteristic) curve or AUC (Swets, 1988) metric was used to evaluate the model performance. The AUC is a threshold-independent measure of a model's ability to discriminate presence from absence (or background). It

varies from 0.5 to 1; an AUC value of 0.5 shows that model predictions are not better than random, values <0.5 are worse than random, 0.5–0.7 signifies poor performance, 0.7–0.9 signifies reasonable/moderate performance, and >0.9 indicates high model performance (Peterson et al., 2011). MaxEnt calculates AUC values slightly differently than other traditional approaches. It calculates AUC by defining specificity using the predicted area and not true commission; see Phillips et al. (2006) for more details. Since we used the SWD format in MaxEnt we used 'PresenceAbsence' package in R, version 2.15.1 (R Development Core Team, 2012), to calculate AUC values.

Model validation was performed using 'subsampling' procedure in MaxEnt. Seventy percent of the *P. solenopsis* data were used for model calibration (training data: 38 districts) and the remaining 30% for model validation (test data: 15 districts). One hundred replicates were run and average AUC values for training and test datasets were calculated using 'PresenceAbsence' package (R Development Core Team, 2012). Percent variable contribution and jackknife procedures in MaxEnt were used to investigate relative importance of different bioclimatic predictors. Response curves were used to study the relationships between bioclimatic variables and the predicted probability of presence of *P. solenopsis*.

3. Results

3.1. Predicted current and potential distribution of P. solenopsis

The MaxEnt model predicted 91% of the currently documented *P. solenopsis* occurrences (i.e., 48 out of 53 districts) in high and medium risk categories with habitat suitability >0.30. The model predicted highly suitable areas for *P. solenopsis* in most of Gujarat (excluding a few districts in northeastern parts of the state), Maharashtra and Andhra Pradesh, southwestern Punjab, northwestern Rajasthan, and western Haryana (Fig. 1b; Table 2). Currently, the highest number of infested districts (15) is in Maharashtra state and our model also predicted highest risk in this state with 25 districts predicted under high and medium risk of invasion (Table 2). Gujarat state was predicted to have the second highest risk of infestation with 13 and 4 districts under high and medium risk, respectively.

The model predicted higher risk of *P. solenopsis* infestation in Karnataka and Andhra Pradesh where currently only a few districts show *P. solenopsis* presence (Fig. 1a and b; Table 2). Madhya Pradesh and Rajasthan were predicted under relatively lower risk compared to other cotton growing states (Fig. 1a and b; Table 2). The model predicted very low habitat suitability for *P. solenopsis* in Jabalpur district (probability 0.10) of Madhya Pradesh and lower suitability for Thiruvarur and Perambalur districts in Tamil Nadu (probabilities 0.22 and 0.24, respectively; Appendix A). These districts are currently infested with *P. solenopsis* yet the model

predicted them under a lower risk category. The model predicted low risk in all Northeastern states, Jammu and Kashmir, Himachal Pradesh, Uttarakhand and parts of Uttar Pradesh and Madhya Pradesh including coastal districts in Western Ghats in southern India (Fig. 1b). The model also predicted high risk of invasion for five districts in Bihar, four in Orissa, and two in West Bengal (Fig. 1b).

3.2. Model performance and influencing factors

MaxEnt predicted potential distribution of *P. solenopsis* with high accuracy for a generalist invader with an average test AUC value of 0.86 (±0.05) and an average training AUC value of 0.91 (±0.01). The final model included only five variables. Model predictions closely matched the documented occurrence of *P. solenopsis* in all nine cotton growing states and also showed potentially suitable districts in other states of India (Fig. 1b). Precipitation of coldest quarter (Bio19), temperature annual range (Bio7), and precipitation seasonality (Bio15) were the strongest predictors of *P. solenopsis* distribution with 30.8, 26.8, and 23.3 percent contributions, respectively (Table 1). Jackknife results also showed that temperature annual range had the highest predictive power (highest regularized training gain and AUC; Fig. 2a and b). Individual response curves for different bioclimatic variables (i.e. model created using only the corresponding variable) showed that the predicted probability of presence of *P. solenopsis* was positively correlated with temperature and negatively correlated with precipitation (Fig. 3). Probability of *P. solenopsis* presence increased up to 26.7 °C average annual temperature and decreased sharply after that (Fig. 3a); similar trends were observed for temperature annual range (Fig. 3b). Probability of *P. solenopsis* presence decreased with the increasing precipitation of coldest quarter (Bio19; Fig. 3c) but slowly increased with increasing precipitation seasonality and then sharply declined after a value of 120 (Fig. 3d).

4. Discussion

To our knowledge this is the first study to demonstrate MaxEnt's use for district-level species occurrences and predict invasion potential of an insect pest at landscape/regional level. All of the previous species distribution modeling studies using MaxEnt used precise coordinates of species occurrences. We showed that in the absence of precise coordinates district-level data can be used in MaxEnt to generate a *preliminary* map of potential distribution of a species. This map can be used to design future more detailed surveys and better planning for using limited funds and human resources. The maps should be updated as soon as the precise coordinates become available.

This study presents a preliminary map of potential distribution of *P. solenopsis* distribution in India using ecological niche modeling. MaxEnt model was highly successful in predicting currently

Table 2
Number of currently infested districts and the number of districts predicted under different categories of risk of invasion by *P. solenopsis* in nine cotton growing states in India. Risk categories were arbitrarily defined by assigning higher risk to districts with high predicted suitability.

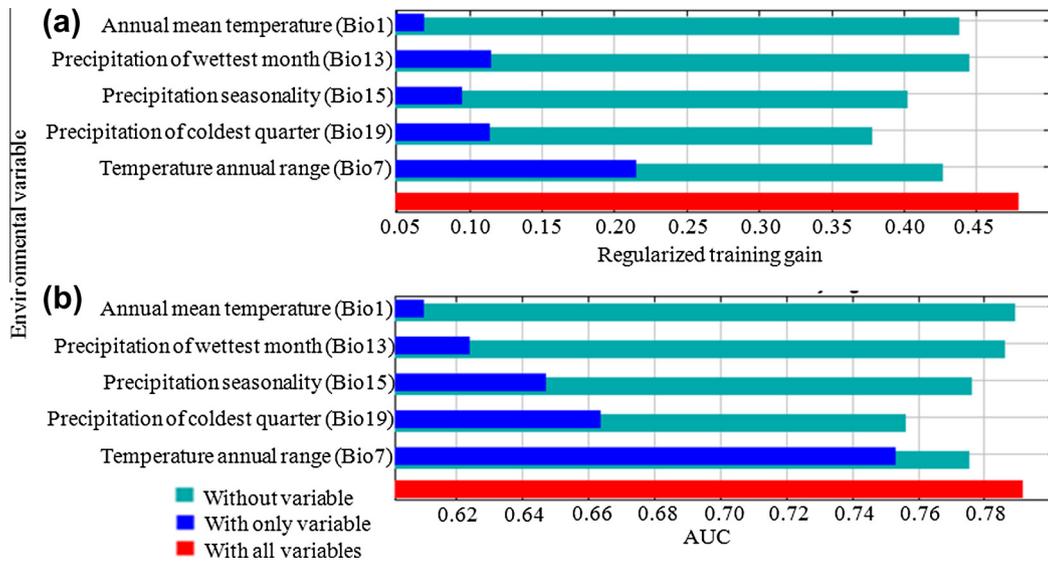| State	No. of districts with current cotton mealybug presence	No. of districts predicted under different risk categories (suitability range)		
		High (>0.5)	Medium (0.3–0.5)	Low (<0.3)
Maharashtra	15	15	10	9
Gujarat	11	13	4	8
Punjab	7	7	5	6
Karnataka	6	6	7	14
Tamil Nadu	5	–	9	21
Haryana	4	3	4	12
Andhra Pradesh	2	11	10	2
Rajasthan	2	4	1	27
Madhya Pradesh	1	1	4	43

Fig. 2. Relative importance of different environmental variables based on results of jackknife tests in MaxEnt. Graphics show variable contributions to (a) regularized training gain, and (b) AUC (area under the ROC curve). Values shown are averages over 100 replicate runs.
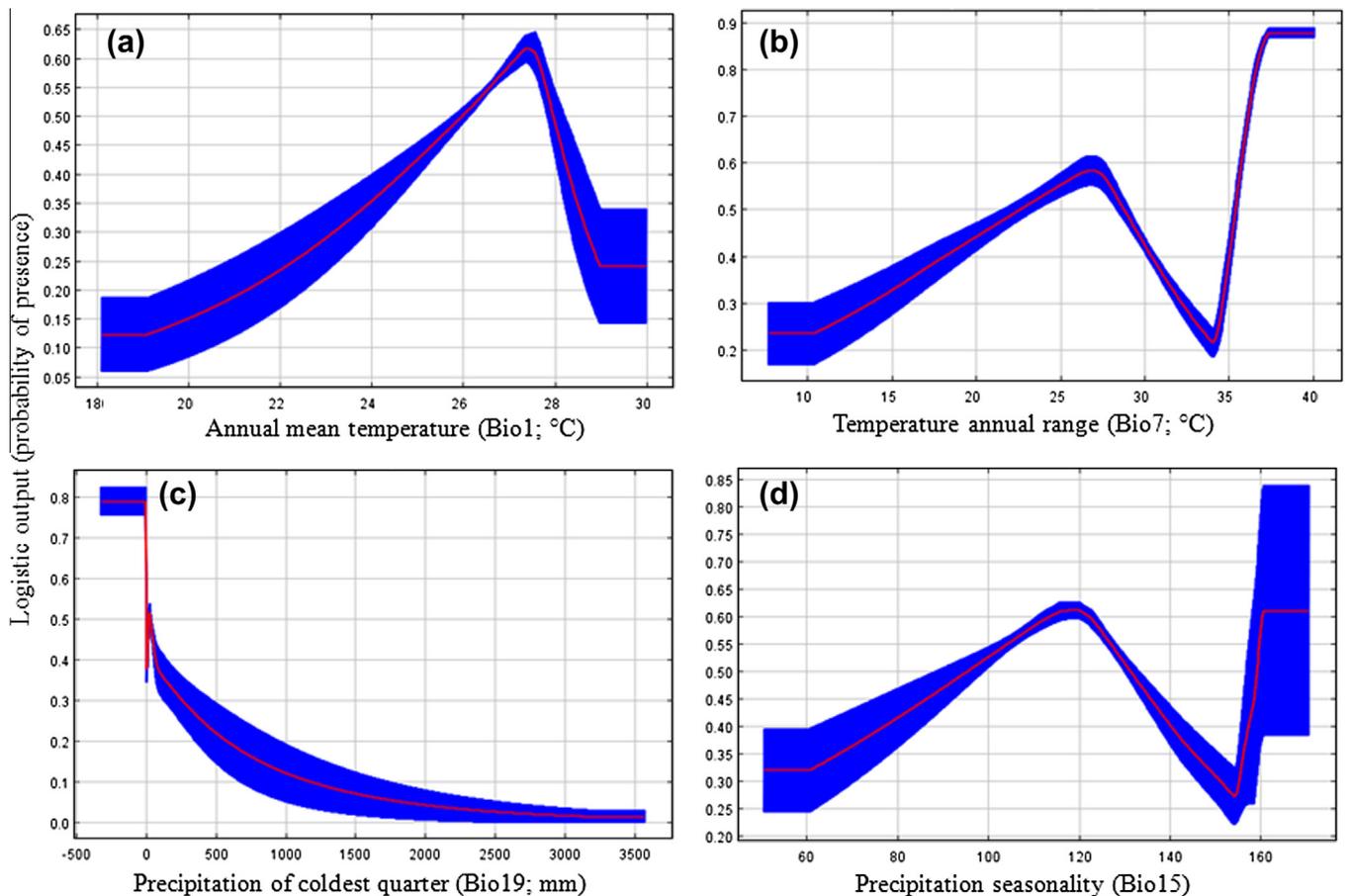

Fig. 3. Relationships between top environmental predictors and the probability of presence of *P. solenopsis* in India. (a) Mean annual temperature (Bio1, °C), (b) temperature annual range (Bio7, °C), (c) precipitation of coldest quarter (Bio19, mm), and (d) precipitation seasonality (Bio15). Red curves show the mean response and blue margins are ±1 SD calculated over 100 replicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

documented occurrences of *P. solenopsis*. The model also identified several potentially suitable areas for *P. solenopsis* across India, which indicates that *P. solenopsis* has not yet reached its full potential range. Precipitation and temperature variables were important

determinants of *P. solenopsis* distribution. Since the model produced using district-level species occurrence data is relatively coarse there is an urgent need to collect precise coordinates (latitude and longitude) of *P. solenopsis* occurrence locations to increase

the predictive accuracy of this model. Results from this study can be used for designing local, regional and national-level integrated pest management policies for cotton and other cultivated crops in India. For example, our results can be used to prioritize control efforts by first targeting highly suitable districts (Fig. 1b) and design future sampling strategies for early detection and rapid response. Our map also showed areas where *P. solenopsis* is most likely to invade which can be helpful in preventing its spread to non-infested but potentially vulnerable districts.

4.1. Potential threat of P. solenopsis to cotton and other cultivated crops in India

MaxEnt model performed excellent and accurately predicted majority of the currently infested districts under high and medium risk of *P. solenopsis* invasion (Fig. 1a and b). Out of nine cotton growing states in India, the model predicted significantly high risk of *P. solenopsis* infestations to cotton and other crops in Maharashtra, Gujarat, Andhra Pradesh and southern parts of Punjab and northern Rajasthan (Fig. 1b). The model predicted low risk in most of the Northern states including Jammu and Kashmir, Himachal Pradesh, Uttarakhand, and all Northeastern states (Fig. 1b). This could be because of the very low temperatures in the coldest months (<10 °C) in these areas that are not optimal for *P. solenopsis* growth (Prasad et al., 2012). The model predicted low risk for many districts in Rajasthan and Madhya Pradesh which is possibly due to extremely hot (average temperatures >40 °C) and dry conditions in these areas during summer (Fig. 1b). The model also predicted high risk of invasion for five districts in Bihar, four in Orissa, and two in West Bengal (Fig. 1b) which might be erroneous because *P. solenopsis* has not be reported from these areas. This error could be because of the model extrapolations beyond the training region (i.e. nine cotton growing states; Stohlgren et al., 2011). It is also possible that these areas are climatically suitable but do not have host plant species for *P. solenopsis*. For example, rice, wheat, mustard, lentil and pea are the major cultivated crops in the four predicted high risk districts in Bihar.

In a previous study, Wang et al. (2010) developed a potential distribution model for *P. solenopsis* at global level using the CLIMEX model. CLIMEX simulates different mechanisms that limit geographical distributions of a species (Sutherst, 2003). Inputs for the CLIMEX model include values of species specific ecophysiological variables such as lower and upper limits of moisture, temperature, and different stress indices (Sutherst and Maywald, 1985). This model was developed at $0.5° \times 0.5°$ spatial resolution and may not be suitable for state or regional level planning and policy making. The Wang et al. (2010) model predicted high to medium climate suitability of *P. solenopsis* for most parts of India except Northern most parts of the country and western Rajasthan. Our model predictions agreed with Wang et al. (2010) predictions in most of the cotton growing areas but differed significantly in other areas such as Northeastern states, central India and coastal areas (Fig. 1b). These differences could be because of the model specific assumptions, calibration settings, uncertainties and errors, and spatial resolution of climatic layers. For example, Wang et al. (2010) used only two locations from Northwestern India (Bathinda and Hisar) in CLIMEX model to analyze the climatic suitability for *P. solenopsis*. Differences in the two model results can also be due to the ways these models quantify the potential distribution of a species; for example, MaxEnt is a correlative model whereas CLIMEX is a mechanistic or process-based model. The mechanistic approach is based on direct measures of physiological variables and ignores biotic interactions, whereas the correlative approach is based on observations that already include effects of biotic interactions (e.g., presence of host plant species at insect pest occurrence locations) on distributions of species; both approaches have

strengths and weaknesses (Soberon and Peterson, 2005; Dormann et al., 2012). Our study produced a higher resolution district-level map of risk of invasion by *P. solenopsis* that is more useful than previous map for designing integrated pest management policies at state, regional, and national levels.

4.2. Effects of climatic factors on P. solenopsis distribution

Both precipitation and temperature variables were strongly correlated with *P. solenopsis* distribution (Fig. 2a and b). Precipitation of coldest quarter had strong negative influence on *P. solenopsis* whereas temperature had a positive effect (Fig. 3). Short term studies have shown reduction in *P. solenopsis* severity due to rainfall which is consistent with our results (Vennila et al., 2010). The model results matched closely with the recent laboratory studies and field observations of temperature effects on *P. solenopsis*. The models results showed that the probability of presence of *P. solenopsis* had a unimodal response to temperature (Fig. 3a). The probability of presence of *P. solenopsis* was lowest at 10 °C temperature and increased up to 27 °C (Fig. 3a and b) which is consistent with the recent findings of Prasad et al. (2012) who found 10.1 °C and 11.7 °C as the lower development temperature thresholds for male and female *P. solenopsis*, respectively. Their results also showed the optimum development at 32 °C and no development at 40 °C. Field observations on the distribution of *P. solenopsis* from Sindh state in Pakistan showed the common occurrence of *P. solenopsis* at sites with temperatures from 30.5 to 39.5 °C (Khuhro et al., 2012); this validates model predictions of steep relationship between the probability of occurrence for *P. solenopsis* and temperature from 35 to 40 °C (Fig. 3b; Sreedevi et al., 2013). Overall, these results show absence of *P. solenopsis* in colder environments with average temperatures <10 °C and very hot environments (i.e., temperatures >40 °C).

4.3. Caveats, future directions and management implications

ENM/SDM are not without limitations and assumptions, and are often affected by the quality of data and different decisions made during the model fitting. For example, they can be affected by sampling bias (Kadmon et al., 2004), sample size (Stockwell and Peterson, 2002), multicollinearity (Dormann et al., 2013), and spatial autocorrelation (Segurado et al., 2006). Performance of different ENM/SDM also depends on species characteristics, spatial resolution and extent of the study area, and choice of predictor variables (Evangelista et al., 2008; Guisan et al., 2007a,b). Regardless of these and other conceptual and practical problems (Soberon and Peterson, 2005; Menke et al., 2009) ENM/SDM are increasingly used to predict potential distributions of species of concern (Franklin, 2009; Elith et al., 2010; Peterson et al., 2011) because resource managers need accurate maps of species distribution and abundance for risk analysis. MaxEnt is robust to small sample sizes but it is affected by the way background data points are selected (Phillips, 2008; Phillips et al., 2009; VanDerWal et al., 2009). We dealt with some of these issues by dropping highly correlated variables to reduce multicollinearity effects and by training our model using background points from only nine cotton growing states. Extrapolations of niche models to novel environments might result in erroneous or uncertain predictions of potentially suitable areas for a species (Stohlgren et al., 2011). It is possible that our model predictions may have high uncertainty in areas that are far away from currently infested districts; for example, areas in extreme Northern and Northeastern states of India (Fig. 1b). The Multivariate Environmental Similarity Surface or MESS analysis implemented in MaxEnt (Elith et al., 2011) can provide insight into the novel environments and appropriate interpretation of model projections.

Future modeling work should include information on major host plant species distributions (e.g., cotton growing areas), land use land cover types, species specific phenology variables (e.g., lower and upper temperature development thresholds) (Buckley et al., 2011; Sambaraju et al., 2012), and remotely sensed variables (e.g., Normalized Difference Vegetation Index) with precise species occurrence records (Prabhakar et al., 2013). There is a temporal mismatch between the climatic data and *P. solenopsis* occurrence records because bioclimatic variables in our study were generated using temperature and precipitation records from 1950 to 2000 (Hijmans et al., 2005). If possible, future modeling efforts for *P. solenopsis* distribution should include a new set of bioclimatic variables generated using the most recent climatic data layers. Since temperature was one of the most important factors affecting *P. solenopsis* distribution it is highly likely that its phenology will shift as a result of climate change (Hodgson et al., 2011). Future work may include the assessment of effects of changing climatic conditions on the *P. solenopsis* distribution because potential future changes in temperature and precipitation might affect insect pests' population growth rates, increase the number of generations, extend the development season, and alter the crop-pest synchrony and interspecific interactions (Porter et al., 1991). A species niche is also shaped by its natural enemies, host plants, their interactions, and climatic conditions (Zhou et al., 2010). The climatic niche for *P. solenopsis* predicted in this study represents the niche space that could be utilized in the absence of competition and its natural enemies. MaxEnt model did not take into account the natural biotic mortality (e.g., due to parasitoids and predators), which governs the relative abundance of the species at any given location. A comprehensive pest management plan for *P. solenopsis* should consider the distribution of its natural enemies, host plants, and spatial and temporal variation in its phenology across India. Areas predicted under high and/or medium risk of *P. solenopsis* invasion that also have its host plant species (e.g. cotton plants in nine cotton growing states) should be targeted first.

Our study presents a novel approach to assessing the invasion potential of an insect pest using district-level occurrence data using niche models. Results provide insight into the climatic factors that affect *P. solenopsis* distribution in India. This is the first attempt to model national-level risk of invasion by *P. solenopsis* in India. Risk analysis of biological invasions is an iterative process and development of an initial model is recommended using the best available species occurrence data; further improvements should be made as more accurate data become available (Stohlgren and Schnase, 2006). The preliminary map developed in this study can be refined to regional scales by integrating more detailed species occurrence data collected using scientifically designed field surveys (Stohlgren, 2007) and higher resolution predictor variables. The predicted map can also serve as a testable alternative hypothesis for *P. solenopsis* occurrence in non-infested districts. Results will be useful for designing local, regional and national-level pest management policies for cotton and other cultivated crops in India. For example, results can be used for selecting monitoring and surveillance sites and planning the reduced applications of pesticides in cotton crops in districts that are predicted under low risk (Appendix B). The maps of potential pest distributions are urgently needed by agriculture managers and policymakers. Our approach can be used in other countries that lack precise coordinates of insect pest occurrences and generate a preliminary map of potential risk because it may be too late to wait for the precise coordinates of pest occurrences to generate a perfect map.

Acknowledgements

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2014.02.007.

References

Aheer, G.M., Shah, Z., Saeed, M., 2009. Seasonal history and biology of cotton mealy bug, *Phenacoccus solenopsis* Tinsley. J. Agric. Res. 47 (4), 423–431.

Amarasekare, K.G., Chong, J.H., Epsky, N.D., Mannion, C.M., 2008. Effect of temperature on the life history of the mealybug *Paracoccus marginatus* (Hemiptera: Pseudococcidae). J. Econ. Entomol. 101 (6), 1798–1804.

Bale, J.S., 2002. Insects and low temperatures: from molecular biology to distributions and abundance. Philos. Trans. Roy. Soc. Lond. Ser. B – Biol. Sci. 357 (1423), 849–861.

Benedict, M.Q., Levine, R.S., Hawley, W.A., Lounibos, L.P., 2007. Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. Vector-Borne Zoonotic Dis. 7 (1), 76–85.

Bentlage, B., Peterson, A.T., Barve, N., Cartwright, P., 2013. Plumbing the depths: extending ecological niche modelling and species distribution modelling in three dimensions. Glob. Ecol. Biogeogr. 22 (8), 952–961.

Buckley, L.B., Waaser, S.A., MacLean, H.J., Fox, R., 2011. Does including physiology improve species distribution model predictions of responses to recent climate change? Ecology 92 (12), 2214–2221.

Chong, J.H., Oetting, R.D., Van Iersel, M.W., 2003. Temperature effects on the development, survival, and reproduction of the Madeira mealybug, *Phenacoccus madeirensis* Green (Hemiptera: Pseudococcidae), on chrysanthemum. Ann. Entomol. Soc. Am. 96 (4), 539–543.

Chong, J.H., Roda, A.L., Mannion, C.M., 2008. Life history of the mealybug, *Maconellicoccus hirsutus* (Hemiptera: Pseudococcidae), at constant temperatures. Environ. Entomol. 37 (2), 323–332.

Davis, M.A., 2009. Invasion Biology. Oxford University Press, Oxford, UK, 288p.

De Meyer, M., Robertson, M.P., Mansell, M.W., Ekesi, S., Tsuruta, K., Mwaiko, W., Vayssieres, J.F., Peterson, A.T., 2010. Ecological niche and potential geographic distribution of the invasive fruit fly *Bactrocera invadens* (Diptera, Tephritidae). Bull. Entomol. Res. 100 (1), 35–48.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Romermann, C., Schroder, B., Singer, A., 2012. Correlation and process in species distribution models: bridging a dichotomy. J. Biogeogr. 39 (12), 2119–2131.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Marquez, J.R.G., Gruber, B., Lafourcade, B., Leitao, P.J., Munkemuller, T., McClean, C., Osborne, P.E., Reineking, B., Schroder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36 (1), 27–46.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129–151.

Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. Methods Ecol. Evol. 1 (4), 330–342.

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17 (1), 43–57.

Evangelista, P.H., Kumar, S., Stohlgren, T.J., Jarnevich, C.S., Crall, A.W., et al., 2008. Modelling invasion for a habitat generalist and a specialist plant species. Divers. Distrib. 14 (5), 808–817.

Evangelista, P.H., Kumar, S., Stohlgren, T.J., Young, N.E., 2011. Assessing forest vulnerability and the potential distribution of pine beetles under current and future climate scenarios in the Interior West of the US. For. Ecol. Manage. 262 (3), 307–316.

Franklin, J., 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, UK, 320p.

Guisan, A., Graham, C.H., Elith, J., Huettmann, F., the NCEAS Species Distribution Modelling Group, 2007a. Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib. 13 (3), 332–340.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S., Peterson, A.T., 2007b. What matters for predicting the occurrences of trees: techniques, data or species' characteristics? Ecol. Monogr. 77 (4), 615–630.

Hanchinal, S.G., Patil, B.V., Bheemanna, M., Hosamani, A.C., Sharanabasappa, S.G., 2009. Incidence of mealy bug on cotton in Tungbhadra project area. In: Proc. Dr. Leslie C. Coleman Memorial National Symposium Plant Protection December 04–06, 2008, University of Agricultural Sciences GKVK, Bangalore.

Hanchinal, S.G., Patil, B.V., Bheemanna, M., Hosamani, A.C., 2010. Population dynamics of mealybug, *Phenacoccus solenopsis* Tinsley and it's natural enemies on Bt cotton. Karnataka J. Agric. Sci. 23 (1), 137–139.

Hanchinal, S.G., Patil, B.V., Basavanagoud, K., Nagangoud, A., Biradar, D.P., Janagoudar, B.S., 2011. Incidence of invasive mealybug (*Phenacoccus solenopsis* Tinsley) on cotton. Karnataka J. Agric. Sci. 24 (2), 143–145.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25 (15), 1965–1978.

Hodgson, J.A., Thomas, C.D., Oliver, T.H., Anderson, B.J., Brereton, T.M., Crone, E.E., 2011. Predicting insect phenology across space and time. Glob. Change Biol. 17 (3), 1289–1300.

Hulme, P.E., 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. J. Appl. Ecol. 46 (1), 10–18.

Jhala, R.C., Bharpoda, T.M., Patel, M.G., 2008. *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae), the mealy bug species recorded first time on cotton and its alternate host plants in Gujarat, India. Uttar Pradesh J. Zool. 28 (3), 403–406.

Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol. Appl. 14 (2), 401–413.

Khuhro, S.N., Kalroo, A.M., Mahmood, R., 2012. Present status of mealy bug *Phenacoccus solenopsis* (Tinsley) on cotton and other plants in Sindh (Pakistan). World Cotton Research Conference-5, Mumbai 07–11 November 2011, pp. 268–271. <https://www.icac.org/meetings/wcrc/wcrc5/Proceedings.pdf> (accessed 24.09.13).

Kim, S.C., Song, J.H., Kim, D.S., 2008. Effect of temperature on the development and fecundity of the cryptic mealybug *Pseudococcus cryptus* in the laboratory. J. Asia Pacific Entomol. 11 (3), 149–153.

Kumar, S., Stohlgren, T.J., 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. J. Ecol. Natur. Environ. 1 (4), 94–98.

Kumar, S., Spaulding, S.A., Stohlgren, T.J., Hermann, K.A., Schmidt, T.S., et al., 2009. Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. Front. Ecol. Environ. 7 (8), 415–420.

Levine, J.M., D'Antonio, C.M., 2003. Forecasting biological invasions with increasing international trade. Conserv. Biol. 17 (1), 322–326.

Macfadyen, S., Kriticos, D.J., 2012. Modelling the geographical range of a species with variable life history. PLoS ONE 7, e40313.

Medley, K.A., 2010. Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. Glob. Ecol. Biogeogr. 19 (1), 122–133.

Menke, S.B., Holway, D.A., Fisher, R.N., Jetz, W., 2009. Characterizing and predicting species distributions across environments and scales: Argentine ant occurrences in the eye of the beholder. Glob. Ecol. Biogeogr. 18 (1), 50–63.

Miller, D.R., Miller, G.L., Watson, G.W., 2002. Invasive species of mealybugs (Hemiptera: Pseudococcidae) and their threat to US agriculture. Proc. Entomol. Soc. Wash. 104 (4), 825–836.

Nagrare, V.S., Kranthi, S., Biradar, V.K., Zade, N.N., Sangode, V., Kakde, G., Shukla, R.M., Shivare, D., Khadi, B.M., Kranthi, K.R., 2009. Widespread infestation of the exotic mealybug species, *Phenacoccus solenopsis* (Tinsley) (Hemiptera: Pseudococcidae), on cotton in India. Bull. Entomol. Res. 99 (5), 537–541.

Nagrare, V.S., Kranthi, S., Kumar, R., Jothi, B.D., Amutha, M., Deshmukh, A.J., Bisane, K.D., Kranthi, K.R., 2011. Compendium of cotton mealybugs. Nagpur, Maharashtra, India: Central Institute for Cotton Research. 42 p. <http://www.cicr.org.in/pdf/compendium_of_cotton_mealybugs.pdf> (accessed 24.09.13).

NCIPM, 2009. Revolution in Indian Cotton. National Center for Integrated Pest Management and Directorate of Cotton Development. Ministry of Agriculture, Department of Agriculture & Cooperation Government of India. <http://www.ncipm.org.in/NCIPMPDFs/Revolution_in_Indian_Cotton.pdf> (accessed 24.09.13).

Nix, H.A., 1986. A biogeophic analysis of Australian Elapid snakes. In: Longmore, R. (Ed.), Australian Flora and Fauna Series, 8. Australian Government Publishing Service, Canberra, Australia, pp. 4–15.

Parsa, S., Kondo, T., Winotai, A., 2012. The cassava mealybug (*Phenacoccus manihoti*) in Asia: First records, potential distribution, and an identification key. PLoS ONE 7, e47675.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Peterson, A.T., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34 (1), 102–117.

Peterson, A.T., Soberon, J., Pearson, R.G., Anderson, R.P., Martinez-Meyer, E., Nakamura, M., Araujo, M.B., 2011. Ecological Niches and Geographic Distributions. Princeton University Press, Princeton, N.J., 328p.

Phillips, S.J., 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007).. Ecography 31 (2), 272–278.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3–4), 231–259.

Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., 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., 2011. In: Pimentel, D. (Ed.), Biological Invasions Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species, second ed. CRC Press, Boca Raton, Florida, 463p.

Porter, J.H., Parry, M.L., Carter, T.R., 1991. The potential effects of climate-change on agricultural insect pests. Agric. For. Meteorol. 57 (1–3), 221–240.

Prabhakar, M., Prasad, Y.G., Vennila, S., Thirupathi, M., Sreedevi, G., Rao, G.R., Venkateswarlu, B., 2013. Hyperspectral indices for assessing damage by the solenopsis mealybug (Hemiptera: Pseudococcidae) in cotton. Comput. Electron. Agric. 97 (7), 61–70.

Prasad, Y.G., Prabhakar, M., Sreedevi, G., Thirupathi, M., 2011. Spatio-temporal dynamics of the parasitoid, *Aenasius bambawalei* Hayat (Hymenoptera: Encyrtidae) on mealybug, *Phenacoccus solenopsis* Tinsley in cotton based cropping systems and associated weed flora. J. Biol. Control 25 (3), 198–202.

Prasad, Y.G., Prabhakar, M., Sreedevi, G., Rao, G.R., Venkateswarlu, B., 2012. Effect of temperature on development, survival and reproduction of the mealybug, *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae) on cotton. Crop Protection 39 (9), 81–88.

R Development 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>.

Raju, A.R., Reddy, P.R., Khadi, B.M., Bharambe, B.M., 2008. Agro economic analysis of cotton based cropping systems. Indian J. Fertilizers 4 (5), 19–22.

Sambaraju, K.R., Carroll, A.L., Zhu, J., Stahl, K., Moore, R.D., Aukema, B.H., 2012. Climate change could alter the distribution of mountain pine beetle outbreaks in western Canada. Ecography 35 (3), 211–223.

Segurado, P., Araujo, M.B., Kunin, W.E., 2006. Consequences of spatial autocorrelation for niche-based models. J. Appl. Ecol. 43 (3), 433–444.

Soberon, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers. Inform. 2, 1–10.

Sreedevi, G., Prasad, Y.G., Prabhakar, M., Rao, G.R., Vennila, S., Venkateswarlu, B., 2013. Bioclimatic thresholds, thermal constants and survival of Mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae) in response to constant temperatures on Hibiscus. PLoS ONE 8 (9), e75636. http://dx.doi.org/10.1371/journal.pone.0075636.

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148 (1), 1–13.

Stohlgren, T.J., 2007. Measuring Plant Diversity: Lessons from the Field. Oxford University Press, Oxford, 408p.

Stohlgren, T.J., Schnase, J.L., 2006. Risk analysis for biological hazards: what we need to know about invasive species. Risk Anal. 26 (1), 163–173.

Stohlgren, T.J., Jarnevich, C.S., Esaias, W.E., Morisette, J.T., 2011. Bounding species distribution models. Curr. Zool. 57 (5), 642–647.

Sutherst, R.W., 2000. Climate change and invasive species: a conceptual framework. In: Mooney, H., Hobbs, R.J. (Eds.), Invasive Species in a Changing World. Island Press, Washington, DC, USA, pp. 211–240.

Sutherst, R.W., 2003. Prediction of species geographical ranges. J. Biogeogr. 30 (6), 805–816.

Sutherst, R.W., Maywald, G.F., 1985. A computerized system for matching climates in ecology. Agric. Ecosyst. Environ. 13 (3–4), 281–299.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240 (4857), 1285–1293.

Tanwar, R.K., Jeyakumar, P., Monga, D., 2007. Mealybugs and their management. Technical Bulletin 19, September, 2007. National Centre for Integrated Pest Management, LBS Building, Pusa Campus, New Delhi 110 012. <http://www.ncipm.org.in/mealybugs/Bulletin-Mealybugs%20(English).pdf> (accessed 24.09.13).

Tinsley, J.D., 1898. An ants'-nest coccid from New Mexico. Can. Entomol. 30 (2), 47–48.

VanDerWal, J., Shoo, L.P., Graham, C., William, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol. Model. 220 (4), 589–594.

Vennila, S., Ramamurthy, S.S., Deshmukh, A.J., Pinjarkar, D.B., Agarwal, M., Pagar, P.C., Prasad, Y.G., Prabhakar, M., Kranthi, K.R., Bambawale, O.M., 2010. A Treatise on Mealybugs of Central Indian Cotton Production System. Technical Bulletin No. 24, National Centre for Integrated Pest Management, New Delhi, p. 39.

Wang, Y.P., Watson, G.W., Zhang, R.Z., 2010. The potential distribution of an invasive mealybug *Phenacoccus solenopsis* and its threat to cotton in Asia. Agric. For. Entomol. 12 (4), 403–416.

Zhou, Z.S., Chen, Z.P., Xu, Z.F., 2010. Niches and interspecific competitive relationships of the parasitoids, *Microplitis prodeniae* and *Campoletis chlorideae*, of the Oriental leafworm moth, *Spodoptera litura*, in tobacco. J. Insect Sci. 10, 1–12.