



## Predicting the potential distribution of *Carthamus lanatus* L. (Saffron thistle) using Maximum Entropy model in India

B. Sarath Babu<sup>1</sup>, Sheikh M. Sultan<sup>2</sup>, K. Rameash<sup>3</sup> and N. Sivaraj<sup>4</sup>

<sup>1,3,4</sup>ICAR-National Bureau of Plant Genetic Resources, Regional Station, Hyderabad 500 030, Telangana State.

<sup>2</sup>ICAR-National Bureau of Plant Genetic Resources, Regional Station, Srinagar 190 005, Jammu and Kashmir State.

### Abstract

*Carthamus lanatus* L. is closely related to safflower and originated from Mediterranean Basin. Wild species especially the wild relatives of crop plants are invaluable source or resistance to several biotic and abiotic stresses, yield, nutritional quality, adaptation and genetic diversity. In India, the species distribution is restricted to Himachal Pradesh and Jammu & Kashmir. Ecological niche modelling has been analyzed using Maximum Entropy method for present and future climatic conditions. The generated MaxEnt model indicated that parts of Gilgit, Chilas, Baramula, Srinagar, Kupwara, Kargil, Doda, Leh (Ladakh), Udhampur of Jammu & Kashmir and Chamba and Lahul&Spiti districts of Himachal Pradesh are the highest probability regions for the distribution of *C. lanatus*.

**Keywords:** Saffron Thistle, *Carthamus Lanatus*, DIVA-GIS, MaxEnt modelling

### I. INTRODUCTION

*Carthamus lanatus* L., is a closely related wild and weedy species of cultivated oilseed crop Safflower which is quite widely distributed in the Kashmir valley. It is also known as *wooly distaff thistle*, *false star thistle*, *wooly safflower* and *wooly star thistle* and belongs to the family Asteraceae of plant kingdom. A native plant species of Mediterranean region, Southern Europe and Central & West Asia which has spread to many temperate regions world including USA, Argentina, Chile, New Zealand, Australia, Greece, Morocco, Portugal, South Africa, Turkey, Uruguay and many other countries including temperate regions of India[1]. Wild species especially the wild relatives of crop plants are invaluable source or resistance to several biotic and abiotic stresses, yield, nutritional quality, adaptation and genetic diversity. However their utilization in the crop improvement programmes depends largely on their availability as well as their crossability relationship with the cultivated types.

*Carthamus lanatus* L. is a problematic highly prickly plant commonly occurring on dry *karewa* agricultural lands throughout Kashmir. The seeds germinate in May producing a rosette with spiny leaves. The flowering stems develop in June flowering occurs in July and seeds mature in August. The stems of young plants are usually covered with loose wooly or cobwebby hairs. The plant is erect; often much branched annual species grows upto 1-1.5 m tall. Stems are usually single in the basal part, branching in upper part, sparsely septate-hairy, cobwebby and glandular, striate. Leaves alternate, sessile, triangular and once-pinnate- deeply narrow lobed with a prominent strong spine at tips. Rosette leaves are narrow about 20 cm long, stem clasping. Stem leaves shorter, rigid, stem clasping and armed with stout spines. Flowers are solitary, sessile, yellow, egg shaped and surrounded by spiny bracts, borne at the ends of the branches. Seeds are brownish-grey, 3 mm long with a four-angled base, sometimes with a pappus of stiff bristles. Each flower capitulum produces

10-16 seeds. Saffron thistle has an unbranched tap root system. Seed buried in the soil may survive for over eight to ten years but hardly ever germinates at depths below 5 cm. Most seed germinates within 3 years of release but the seed has complex dormancy characteristics.

## II. CHARACTERISTICS OF THE STUDY AREA

Kashmir valley is the central part of North Western Indian Himalayan state of Jammu & Kashmir extending between 32° 17' N to 37° 05' N and 72° 31' E to 80° 20' E. Topographically Kashmir valley with a moderate climate is deep elliptical bowl-shaped valley bounded by lofty mountains of Pir Panjal in the South and Southwest and the Great Himalayan range in the north and east with 64% of total area being mountainous. It has an average annual precipitation of 660mm with average temperature around 13°C and the altitude ranging between 1500-2500 asl. The Physiographic of target districts are highly unequal. The high hills and the elevated plateaus of alluvial and locus trine materials are present in the study area. These are locally known as *Karewas* (Wudars). The *karewas* are divided from each other, sometimes cut into strips by valleys from 30-92 m (100 to 300 feet) in depth, occasionally they are surrounded altogether by lower ground by more generally they connect on to some of the mountains that bound the ravine. The *karewas* of study area are famous for the cultivation of saffron, apples and almonds. Besides, major crops grown in these districts are Rice, Wheat, Maize, Vegetables and fruits, oil seeds and saffron.

## III. MATERIALS AND METHODS

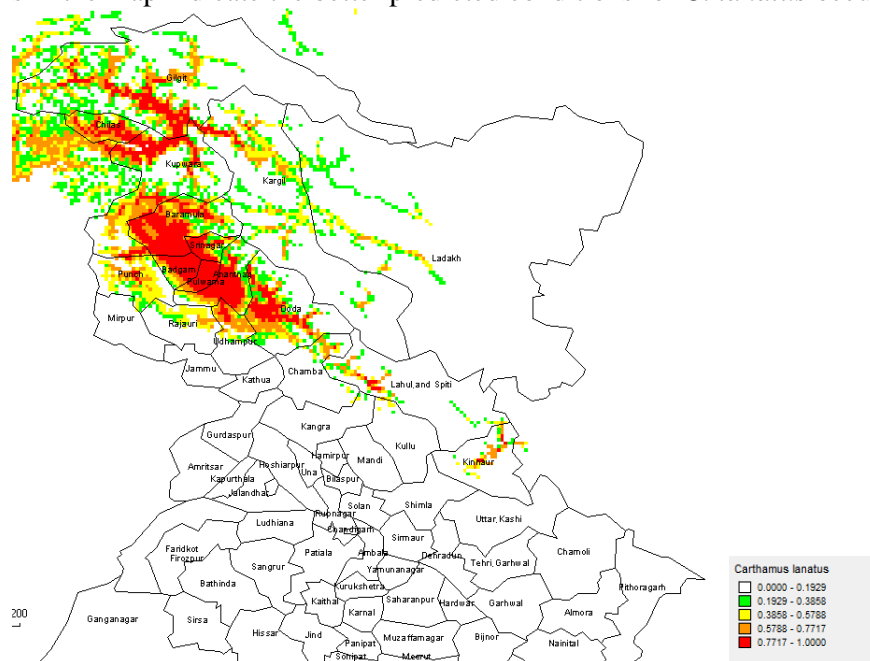
Crop presence data of wild safflower was generated by conducting two exploration surveys while collecting 24 samples from parts of Jammu and Kashmir by the National Bureau of Plant Genetic Resources, Regional Station, Srinagar during 2013. Fortunately, the crop presence data (geo-reference data) of *Carthamus lanatus* collected through exploration surveys provides a reliable and sound basis for the analysis using Maxent software. Bioclimatic variables (BC) are often used in ecological niche modeling and they represent annual trends, seasonality and extreme or limiting environmental factors. We used Bioclim For the current climate (baseline) of India we used monthly data from the WorldClim database [2] sourced from global weather stations publicly and freely available and downloadable from [www.worldclim.org](http://www.worldclim.org). WorldClim was developed using ~47,000 weather stations with monthly information on precipitation, ~23,000 stations with mean temperature data and ~13,000 locations with diurnal temperature range data, passed through a quality checking algorithm and then used to develop a continuous climate surface using a thin plate spline algorithm, with elevation, latitude and longitude as independent variables [2]. The data represent long term (1950-2000) monthly means of maximum, minimum, mean temperatures and total rainfall, at 30 arc-seconds (~1 km at the equator) for every land area of the globe. WorldClim provides interpolated global climate surfaces using latitude, longitude and elevation as independent variables and represents long term (1950-2000) monthly means of maximum, minimum, mean temperatures and total rainfall. The following settings were used during the MaxEnt run: presence records (24); 10024 points used to determine the Maxent distribution (background points and presence points). Environmental layers used (all continuous) (Bioclim\_2.5 arc min): bio1, bio10, bio11, bio12, bio13, bio14, bio15, bio16, bio17, bio18, bio19, bio2, bio3, bio4, bio5, bio6, bio7, bio8 and bio9. Regularization values: linear/quadratic:0.714, categorical:0.429, threshold:1.880, hinge:0.5. MaxEnt software version 3.3.3k downloaded from [www.cs.princeton.edu/~schapire/](http://www.cs.princeton.edu/~schapire/) MaxEnt was used for Saffron thistle species probability habitat modelling. Climate data for Future conditions were obtained from <http://www.worldclim.org> that was downscaled global climate model (GCM) data from CMIP5 (IPPC Fifth Assessment). The ACCESS1-0 climate model developed by CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia) in collaboration with BOM (Bureau of Meteorology, Australia) was used for the analysis at a special resolution of 2.5-minute (of a longitude/latitude degree) that is about 4.5 km at the equator. The greenhouse gas scenario was kept as RCP 4.5 (representative concentration pathway) for the year 2050 (average for 2041-2060). The RCPs are consistent with a wide range of possible changes in future anthropogenic (i.e., human)

greenhouse gas (GHG) emissions. RCP 4.5 assumes that global annual GHG emissions (measured in CO<sub>2</sub>-equivalents) peak around 2040, then with emissions declining substantially thereafter. Maximum entropy (MaxEnt) is a niche modelling method that has been developed involving species distribution information based only on known presences and is a general-purpose method for making predictions or inferences from incomplete information.

#### IV. RESULTS AND DISCUSSION

Species distribution modelling techniques produce maps of the potential distribution of species [3]. MaxEnt [4,5] is a predictive biogeography programme that uses a maximum entropy algorithm to match known locality points for a species to potential localities, based on their environmental characteristics. It is a useful technique as it does not require absence records to build a predictive model. We have used this technique to identify suitable locations to grow red banana in the changed climate regime.

Figure 1 represents the MaxEnt modelling probability matching locations where the wild safflower species distributed in India based on the presence points with current climatic scenario. The high probability value range of 0.77 to 1.0 for likely potential regions for distribution of *C. lanatus* occurs in parts of Gilgit, Chilas, Baramula, Srinagar, Kupwara, Kargil, Doda, Leh (Ladakh), Udhampur of Jammu & Kashmir and Chamba and Lahul&Spiti districts of Himachal Pradesh. Warmer colours in the map indicate the better predicted conditions for *C. lanatus* occurrence.



**Fig.1. MaxEnt model generated for *Carthamus lanatus* (current climate grids)**

Maximum Entropy (MaxEnt) is a niche modelling approach that has been developed linking species distribution information built only on identified presences and is a general-purpose method for making predictions or inferences from incomplete information. MaxEnt can take the environmental conditions at occurrence locations and produce a probability distribution that can then be used to assess every other location for its likely occurrence. The result is a map of the probability of conditions being favourable to occurrence. The basis for this model is the general notion that knowledge about environmental conditions at locations where *Carthamus lanatus* is successfully present should provide a basis for summarizing species growth parameters throughout the region. We have used MaxEnt niche modelling method for predicting potential distribution information based on known *C. lanatus* presences in India. The known presences of this important wild species are drawn from Jammu and Kashmir. Some common thresholds and corresponding omission rates are

provided in Table 1. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes  $6 * \text{training omission rate} + .04 * \text{cumulative threshold} + 1.6 * \text{fractional predicted area}$ .

**Table 1. Some common thresholds and corresponding omission rates for the MaxEnt Model on wild safflower (*Carthamus lanatus*)**

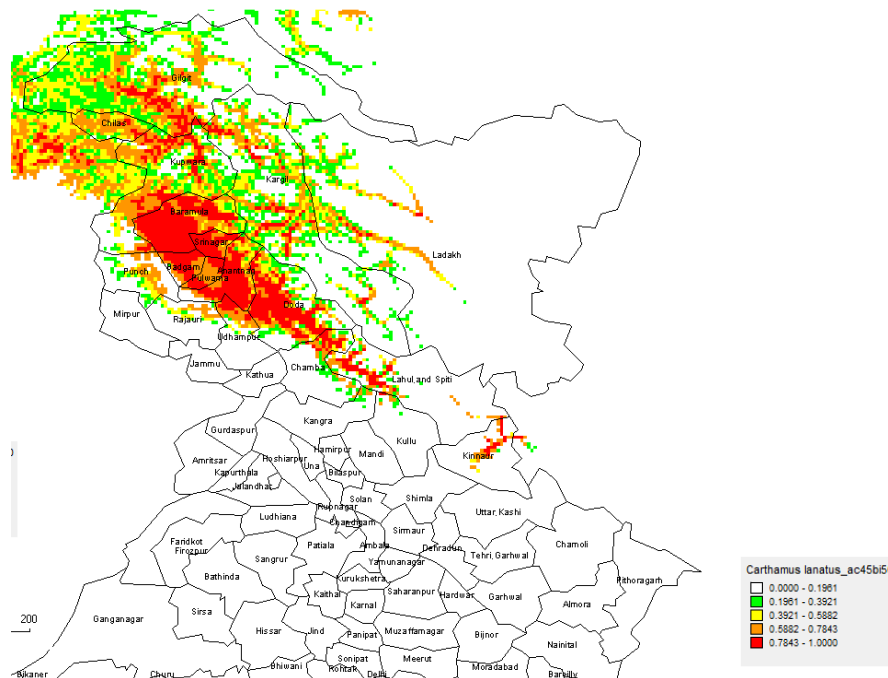
Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.003	Fixed cumulative value 1	0.172	0.000	0.000	8.771E-4
5.000	0.017	Fixed cumulative value 5	0.088	0.000	0.000	5.995E-5
10.000	0.035	Fixed cumulative value 10	0.054	0.000	0.000	8.4E-6
31.477	0.174	Minimum training presence	0.009	0.000	0.250	2.519E-6
48.373	0.644	10 percentile training presence	0.001	0.083	0.250	1.344E-8
31.477	0.174	Equal training sensitivity and specificity	0.009	0.000	0.250	2.519E-6
31.477	0.174	Maximum training sensitivity plus specificity	0.009	0.000	0.250	2.519E-6
11.653	0.043	Equal test sensitivity and specificity	0.047	0.000	0.000	4.693E-6
11.653	0.043	Maximum test sensitivity plus specificity	0.047	0.000	0.000	4.693E-6
2.040	0.007	Balance training omission, predicted area and threshold value	0.135	0.000	0.000	3.286E-4
23.469	0.104	Equate entropy of thresholded and original distributions	0.018	0.000	0.250	2.218E-5

Table 2 gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated. Annual precipitation (Bio12), Max temperature of warmest month (Bio5), Precipitation of driest quarter (Bio17) and Isothermality (Bio3) are variables which had least influence on the distribution model while other remaining environmental variables *viz.*, Precipitation of wettest month (Bio13), (Min temperature of coldest month (Bio6), Mean temperature of warmest quarter (Bio10), Precipitation of driest quarter (Bio17), Annual mean temperature (Bio1), Mean temperature of coldest quarter (Bio11), Precipitation of wettest quarter (Bio16), Precipitation of coldest quarter (Bio19), Precipitation seasonality (Bio15), Mean temperature of wettest quarter (Bio8), Mean temperature of driest quarter (Bio9), Mean diurnal range (Bio2), Temperature seasonality (Bio4), Precipitation of warmest quarter (Bio18), Precipitation of wettest month (Bio14) have profound influence on the species model generated for *C. lanatus* in this study.

**Table 2. Estimates of relative contributions of environmental variables to the Maxent model on *C. lanatus***

Variable	Percent contribution	Permutation importance
Min Temperature of coldest quarter (bio11)	20.8	0
Mean Temperature of Driest Quarter (bio9)	17.7	39.5
Precipitation seasonality (bio15)	17.2	3.6
Mean temperature of the wettest quarter (bio8)	12.1	7
Mean Temperature of Warmest Quarter (bio10)	9.3	6.2
Minimum temperature of the coldest month (bio6)	7.2	0
Mean temperature of wettest quarter (bio18)	4.9	4.3
Precipitation of coldest quarter (bio19)	4.3	3.8
Temperature annual range (bio7)	2	28.2
Temperature seasonality (bio4)	1.2	6.5
Annual mean temperature (bio1)	1.1	0.2
Precipitation of driest month (bio14)	1.1	0.1
Mean diurnal range (bio2)	0.6	0
Precipitation of wettest quarter (bio16)	0.5	0
Precipitation of wettest month (bio13)	0.1	0.7
Annual Precipitation (bio12)	0	0
Max temperature of warmest month (bio5)	0	0
Precipitation of driest quarter (bio17)	0	0
Isothermality (bio3)	0	0

Fig. 2 represents the MaxEnt model for *Carthamus lanatus* generated for future climatic conditions (2050). The model indicates that more potential climate suitable regions could be available in 2050 when compared to present climatic scenario for the distribution of *C. lanatus*. Interestingly the regions of distribution are falling in parts of Gilgit, Chilas, Baramula, Srinagar, Kupwara, Kargil, Doda, Leh (Ladakh), Udhampur of Jammu & Kashmir and Chamba and Lahul&Spiti districts of Himachal Pradesh.



**Fig.2. MaxEnt model generated for *Carthamus lanatus* (Future climate in 2050)**

MaxEnt integrates species occurrences with background data (i.e., randomly selected points) from spatial environmental variables representing different environmental gradients in the study area and generates probability of species' presence [4]. It identifies areas that have conditions most similar to species' current known occurrences and ranks them from 0 (unsuitable or most dissimilar) to 1 (most suitable or most similar). MaxEnt generates a set of feature classes (i.e., linear [L], quadratic [Q], product [P], threshold [T], and hinge [H]) using environmental variables. It uses a regularization multiplier (RM) to reduce the number of parameters and thus automatically controls the model complexity [6,7]. The default RM value is 1; a smaller value of RM (1) may potentially overfit the model and produce more restricted distribution whereas a higher value (>1) would result in simpler models with less discriminating power and broader species potential distribution [8]. Therefore, tuning of MaxEnt model settings has been suggested for selecting an optimal model to enhance predictions [9, 10].

The maximum entropy modelling technique has been successfully used to model potential plant and insect pests distributions for purposes such as monitoring invasive species and disease vectors and their likely spread due to climate change by many researchers in the recent past [11,12,13-31]. Reddy *et al.*, [32, 33] used MaxEnt for identifying suitable regions for cultivation of roselle and *Basella alba* respectively in the country to combat climate change that may affect the crop growing conditions. It executes well on small sample sizes [34], which indicates that the multiplicative methods used in MaxEnt give better predictions than the discriminative methods employed by other techniques [6,8]. However, we have attempted to identify suitable locations for distribution of this important wild safflower species in our country in the changed climatic regime.

#### BIBLIOGRAPHY

- [1] Sultan S M, Dikshit N and Sivaraj N 2015. Diversity, distribution and conservation of Saffron Thistle (*Carthamus lanatus* L.) in mid-high altitude temperate zone of Kashmir valley, India; A DIVA-GIS study. *Tropical Ecology* **56(3)**:303-310
- [2] Hijmans R J, Cameron S E, Parra J L, Jones P G, and Jarvis A 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25(15)** :1965–1978
- [3] Anderson RP, Lew D, Peterson AT. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Journal of Ecological Modelling* **162**: 211-232.
- [4] Phillips S J and Dudik M 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161-175.
- [5] Phillips S, Anderson R and Schapire R 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.
- [6] Elith J, Graham C H, Anderson R P, Dudík M, Ferrier S, Guisan A, Hijmans R J, Huettmann F, Leathwick R, Lehmann A, Li J, Lohmann L G, Loiselle B A, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J Mc C, Peterson A T, Phillips J, Richardson K, Scachetti-Pereira R, Schapire E, Soberon J, Williams S, Wisz M and Zimmermann E 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.
- [7] Phillips S J and Dudik M 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161-175.
- [8] Pearce J and Ferrier S 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**: 225-245.
- [9] Shcheglovitova M and Anderson R P 2013. Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample size. *Ecological Modelling* **269**: 9-17.
- [10] Boria RA, Olson LE, Goodman SM, Anderson RP 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol Modell* **275**: 73-77.
- [11] Villordon A, Njugana W, Simon G, Philip N, Heneriko K, Simon C J, Don L B., Bernard Y, Phinehas T and Robert OMM 2006. Using GIS based tools and distribution modelling to determine sweet potato exploration and documentation priorities in Sub-Saharan Africa. *Hort Science* **41**: 1377-1381.
- [12] Yost A C, Petersen S L, Gregg M and Miller R 2008. Predictive modeling and mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using maximum entropy and a long-term dataset from Southern Oregon. *Ecological Informatics* **3**: 375-386.
- [13] Klopfenstein N B, Kim M S, Hanna J W, Richardson B A and Lundquist J E 2009. Approaches to predicting potential impacts of climate change on forest disease: an example with Armillaria root disease. Res. Pap. RMRS-RP-76. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 10 p.

- [14] Kraft K H 2009. The domestication of the chile pepper, *Capsicum annum*: genetic, ecological, and anthropogenic patterns of genetic diversity. Ph.D. Thesis. University of California, Davis, USA.
- [15] Chamaille L, Tran A, Meunier A, Bourdoiseau G, Ready P and Dedet J P. 2010. Environmental risk mapping of *Canine leishmaniasis* in France. *Parasites and Vectors* **3**: 31. Doi: 10.1186/17563305-3-31.
- [16] Kulkarni, M.A., Desrochers, R.E. & Kerr, J.T. 2010. High resolution niche models of malaria vectors in northern Tanzania: a new capacity to predict malaria risk? *PLoS One* **5** (2): e9396. Doi: 10.1371/journal.pone.0009396.
- [17] Ali Abdullah S.2011. cogeographic, genetic and taxonomic studies of the genus *Lathyrus* L. Ph.D. Thesis. University of Birmingham, UK.
- [18] Fischer D, Moeller P, Thomas S M, Naucke J T and Beierkuhnlein C 2011. Combining climatic projections and dispersal ability: a method for estimating the responses of sandfly vector species to climate change. *PLoS Neglected Tropical Diseases* **5**: e1407. Doi: 10.1371/journal.pntd.0001407.
- [19] Gormley A M, Forsyth D M, Griffioen P, Lindeman M, Ramsey D S L, Scroggie M P and Woodford L 2011. Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology* **48**: 25-34.
- [20] Theissinger K 2011. Comparative phylogeography of two co-distributed arctic-alpine fresh water insect species in Europe. Ph.D. Thesis. Johannes Gutenberg-University Mainz, Germany.
- [21] Advani N K 2012. Thermal ecology of the Glanville Fritillary butterfly (*Melitaea cinxia*). Ph.D. Thesis. The University of Texas at Austin, USA.
- [22] David A.K, Keith L, Mc Dougall, Christopher CS and Jillian LW 2012. Spatial Analysis of Risks Posed by Root Rot Pathogen, *Phytophthora cinnamomi*: Implications for Disease Management. *Proceedings of Linnean Society of New South Wales* **134**: 147-180.
- [23] Gurgel-Gonçalves R, Galvão C, Costa J and Peterson A T 2012. Geographic distribution of Chagas disease vectors in Brazil based on ecological niche modelling. *Journal of Tropical Medicine*. Doi: 10.1155/2012/705326.
- [24] Robinson J L 2012. Patterns of Ecological Performance and Aquatic Insect Diversity in High Quality Protected Area Networks. Ph.D. Dissertation, University of Tennessee, USA.
- [25] Petersen M J 2013. Evidence of a climatic niche shift following North American introductions of two crane flies (Diptera; genus *Tipula*). *Biological Invasions* **15**: 885- 897. Doi: 10.1007/s10530-012-0337-3.
- [26] Sánchez-Guillén R A, Muñoz J, Rodríguez-Tapia G, Feria Arroyo TP and Córdoba- Aguilar A 2013. Climate-Induced Range Shifts and Possible Hybridisation Consequences in Insects. *PLoS One* **8**: e80531. doi:10.1371/journal.pone.0080531
- [27] Solhjoui-Fard S, Sarafrazi A, Moeini MM, Ahadiyat A 2013. Predicting habitat distribution of five heteropteran pest species in Iran. *Journal of Insect Science* **13**:116. Available online: www.insectscience.org/13.116.
- [28] Fourcade Y, Engler J O, Rodder D, Secondi J 2014. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLoS One* **9** (5): e97122. doi:10.1371/journal.pone.0097122.
- [29] Williams K A, Richards C S and Villet M H 2014. Predicting the geographic distribution of *Lucilia sericata* and *Lucilia cuprina* (Diptera: Calliphoridae) in South Africa. *African Invertebrates* **55**: 157-170.
- [30] Khoury C K, Heider B, Castañeda-Álvarez N P, Achicanoy H A, Sosa C C, Miller R E, Scotland R W, Wood J R I, Rossel G, Eserman L A, Jarret R L, Yencho G C, Bernau V, Juarez H, Sotelo S, de Haan S and Struik P C 2015. Distributions, *ex-situ* conservation priorities, and genetic resource potential of crop wild relatives of sweet potato [*Ipomoea batatas* (L.) Lam., I.series Batatas]. *Front. PlantSci.***6**:251. doi:10.3389/fpls.2015.00251.
- [31] Zimmermann H, Loos J, Wehrden H and Fischer J 2015. Aliens in Transylvania: risk maps of invasive alien plant species in Central Romania. *NeoBiota* **24**: 55-65.
- [32] Reddy M T, Hameedunnisa Begum, Neelam Sunil, Someswara Rao Pandravada, Natarajan Sivaraj 2015a. Assessing Climate Suitability for Sustainable Vegetable Roselle (*Hibiscus sabdariffa* var. *sabdariffa* L.) cultivation in India Using MaxEnt Model. *Agricultural and Biological Sciences Journal* **1**: 62-70.
- [33] Reddy M T, Begum H, Sunil N, Pandravada S R, Sivaraj N and Kumar S 2015b. Mapping the Climate Suitability Using MaxEnt Modeling Approach for Ceylon Spinach (*Basella alba* L.) cultivation in India. *The Journal of Agricultural Sciences* **10**: 87-97.
- [34] Pearson R G., Raxworthy C J, Nakamura M and Peterson A T 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**: 102-117.

