Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online)
Vol.7, No.3, 2017

Conducting and Analyzing Openmodeller and Maximum Entropy Model Software for Species Distribution Modeling using Arc GIS

Haftu Abrha Mengesha
Maichew agricultural college, Maichew, Ethiopia

Abstract

This paper gives a basic procedure and analysis to use species distribution models (**SDMs**) of the Maxent 3.3.3k, open modeler version 1.1.0 with useful steps used in Arc GIS 10 for modeling of species' geographic distributions. To model species' geographic distributions Arc GIS plays great role for preparation of input data and processing of the output data. Predicting species distribution using SDMs for the map of potential suitable habitat and pest risk for endangered species is critical for monitoring and restoring in to their natural habitat, artificial introductions, conservation sites, and management of their native habitat. This paper also gives knowledge on procedures used to run the species distribution models (SDMs) in order to conserve and sustain the species. The tutorial was also conducted in Northern Ethiopia.

Keywords: Arc GIS, Openmodeller, Maxent, Representative Concentration Pathway (RCP), General circulation model (GCM)

1. Introduction

Species distribution models (SDMs) are currently the most broadly used scientific methods to study potential climate change impacts on biodiversity. The models are constructed using a variety of modeling methods and combine species geographical coordinates of the occurrence records with a set of predictor variables. SDM are made to compare current and projected future distribution of species based on climate change scenario and current environmental climate data. They consisted of locating and mapping points of occurrence using the physical characteristics of the study area (6). Methods that have been used to predict species potential distribution using Species Distribution Models (SDM) based on occurrence GPS data points and environmental variables displaying usually best results (1).

With the rapid spread of ecological niche modeling (ENM) the need of detailed dataset of environmental characterization has increased and the creation of the World-Climate data set is developed. The worldclim website has past, current and future environmental characterization at different General Circulation Methods (GCM), emission concentration scenario or Representative Concentration Pathway (RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5) and resolutions (10 minute, 5 minute, 2.5 minute and 30 arc second). It has projected world climate data in raster form using new Representative Concentration Pathway (RCP) scenario based in different global circulation models for mid and end parts of this century.

There are different scenario development stages. The recent are special report on emission scenario (5) and Representative Concentration Pathways (2). Special report on emission scenario is replaced by the RCPs (10). The replacement is due to lack of a complete set of socio-economic, emission and climate projections (8). The new and latest scenarios are called Representative Concentration Pathways (RCPs). RCPs are used as inputs for climate modeling and are affected by concentrations of a variety of greenhouse gases, land use, air pollution, changes in technology, population, energy production and a variety of additional factors and that contains four pathways of emission level which are RCP8.5 (high emission scenario), RCP6 and RCP4.5 (medium emission scenario) and RCP 2.6 (low emission scenario) (5).In addition to climate data altitude, soil group, vegetation cover uses input data for SDMs. Therefore, the modeling approach uses different environmental variables and locations of the species.

2) Materials and procedures

1. Arc GIS

The main steps required to do with Arc GIS are listed as the following:

1.1) Download environmental variables

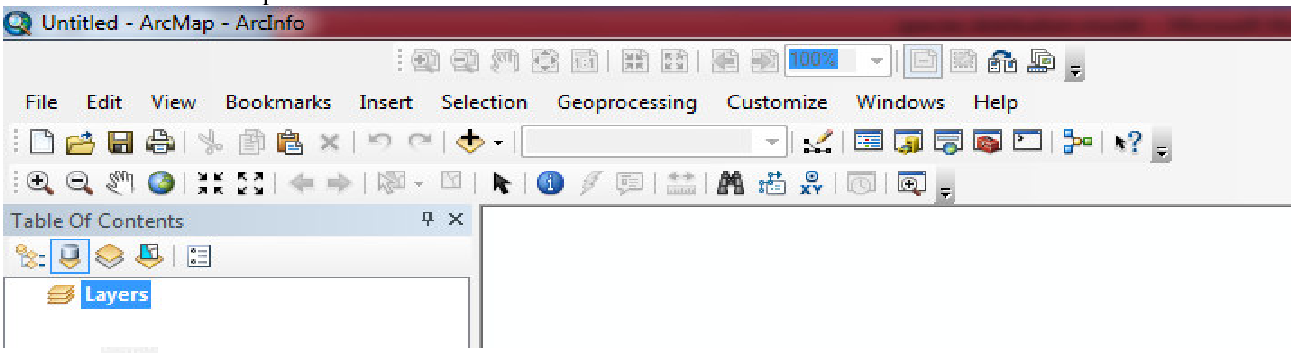
Environmental variables are used to predict current and future species distribution. They are found in their websites. Climatic variables and altitude are found in worldclim website *(www.worldclim.org)*for past, current and future data sets. Soil groups are found in FAO/UNSECO soil classification; Harmonized World Soil Data base (HWSD) and from international soil information center (ISRIC).

1.2) Extraction of environmental variables

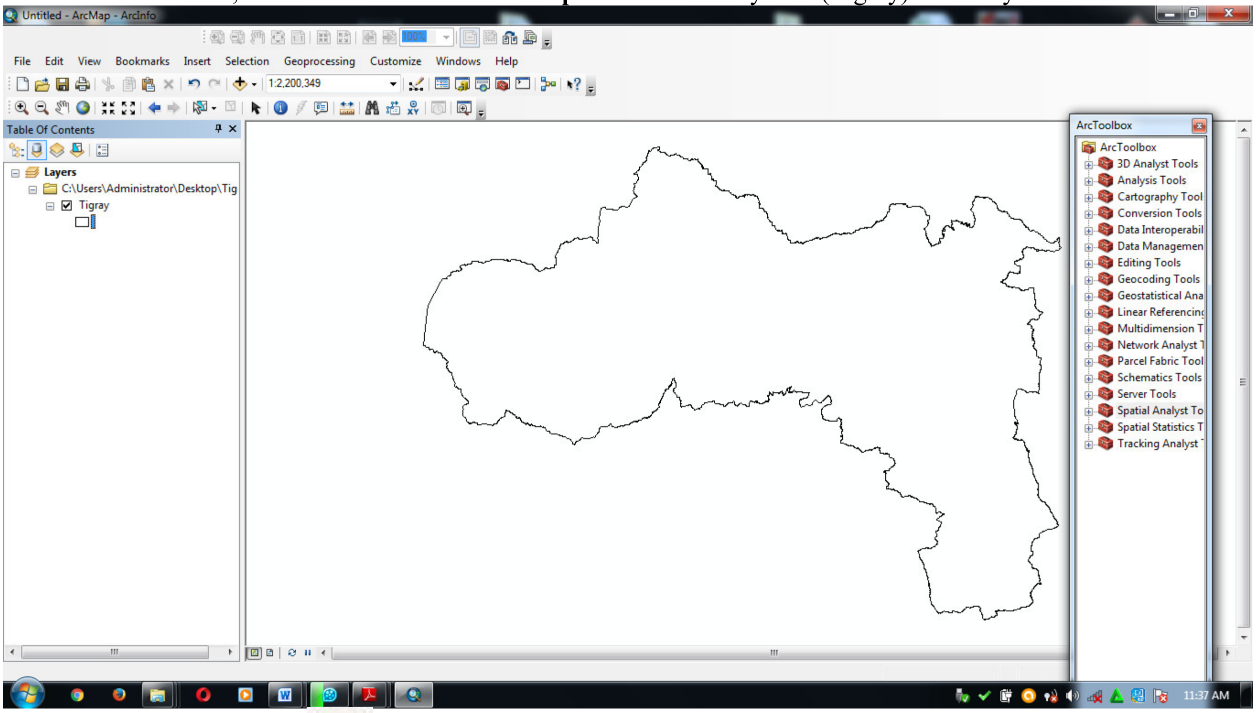
This process is used to extract the raster environmental variables to the study area as the following:

- Create folder called **Tigray** which has shape file of Tigray.
- Create folder called **worldclim** which has environmental variables downloaded from the websites.
- Create folder called **extracted** on desktop

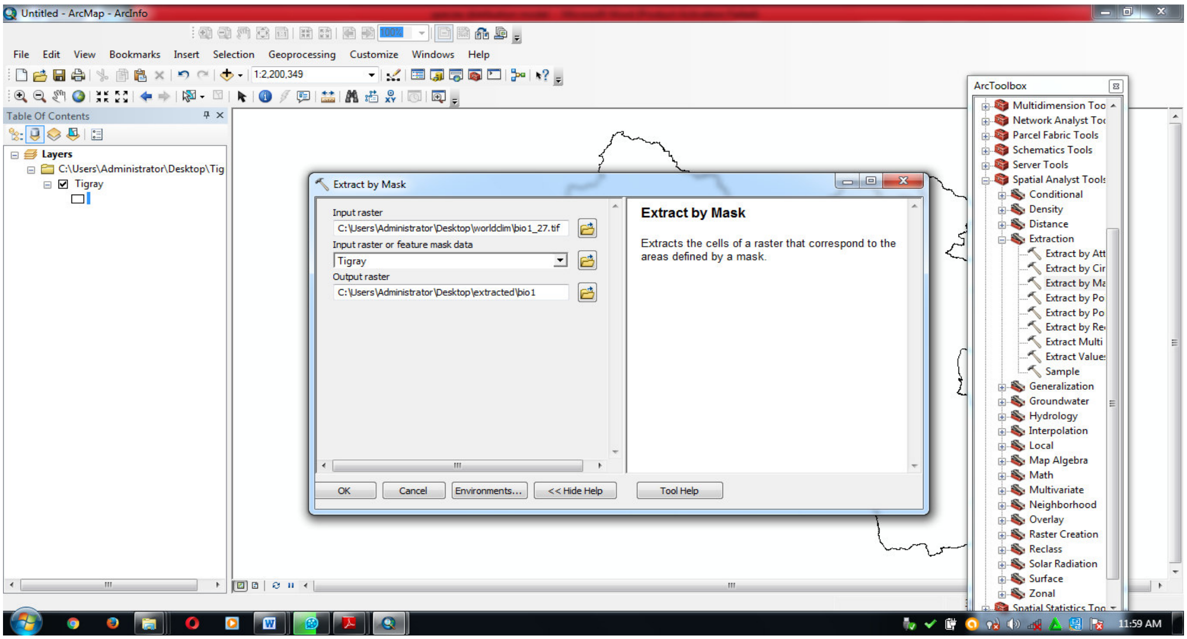
- Create folder called **ASCII** on desktop
- Open Arc GIS 10



Click on button, then search and insert the **shape file** of the study area (Tigray) to the layer.



- Click arc tool box ( ) > spatial analyst > extraction >extraction by mask
- Insert bio1 from the folder **worldclim** in input raster
- Insert Tigray shape file in input raster or feature mask data
- Insert extracted folder the extracted output names as bio1, bio2 in output raster and seems:


Extract by Mask

Input raster
C:\Users\Administrator\Desktop\worlddim\bio1_27.tif

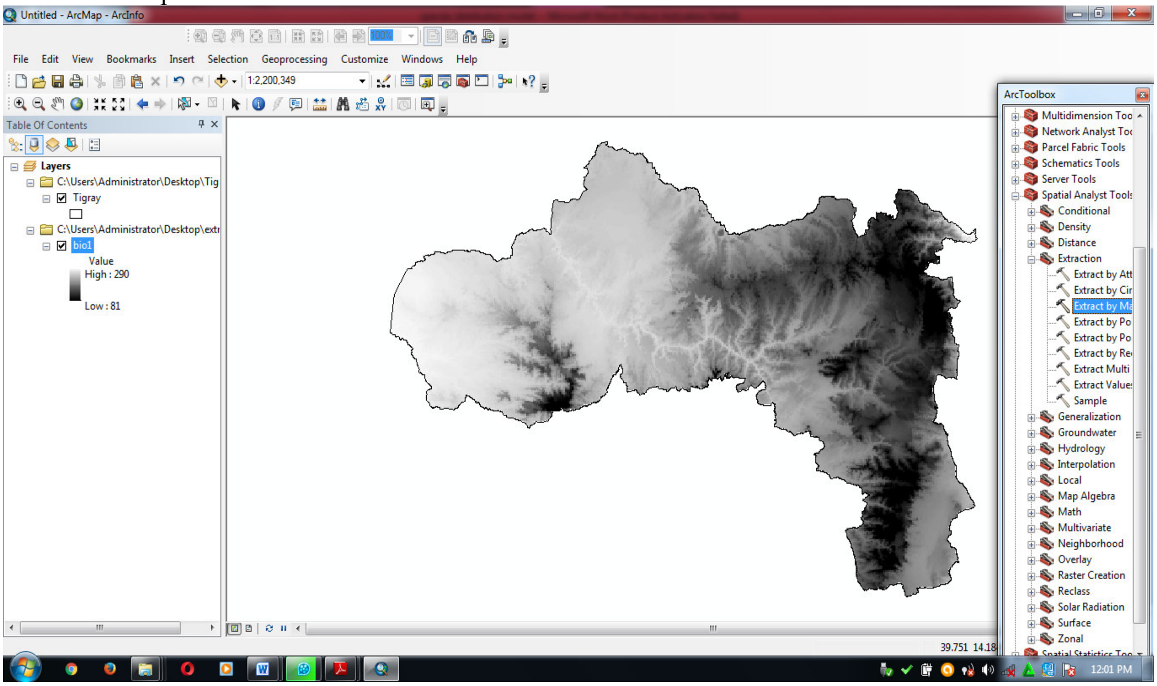
Input raster or feature mask data
Tigray

Output raster
C:\Users\Administrator\Desktop\extracted\bio1

Extract by Mask

Extracts the cells of a raster that correspond to the areas defined by a mask.

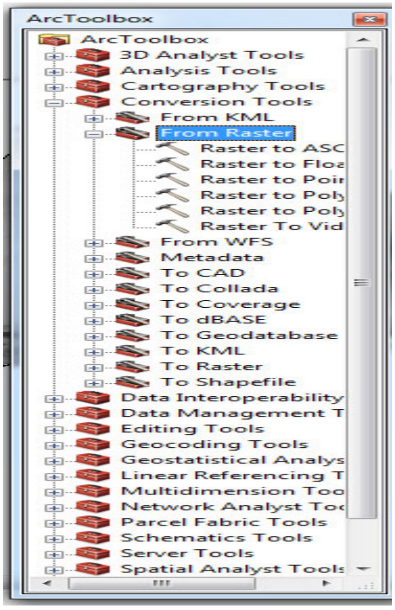
- Then press on **ok** button and seems:



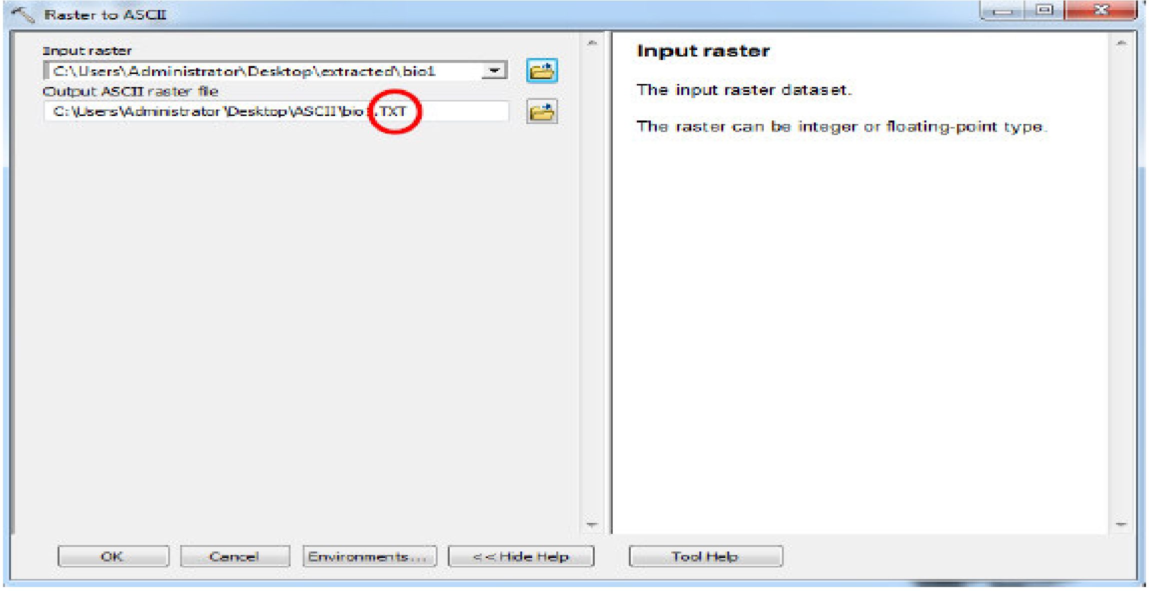
- Use the same steps for bio2 bio 3…..bio19 and other variables

1.3) Converting the environmental variables from raster to ASCII

Click on arc tool box > conversion tool >from raster > raster to ASCII



- Insert **bio1** from **extracted** folder in the input raster button
- Search the ASCII folder from desktop in the output ASCII raster file then write bio1 and seems:


Raster to ASCII

Input raster
C:\Users\Administrator\Desktop\extracted\bio1

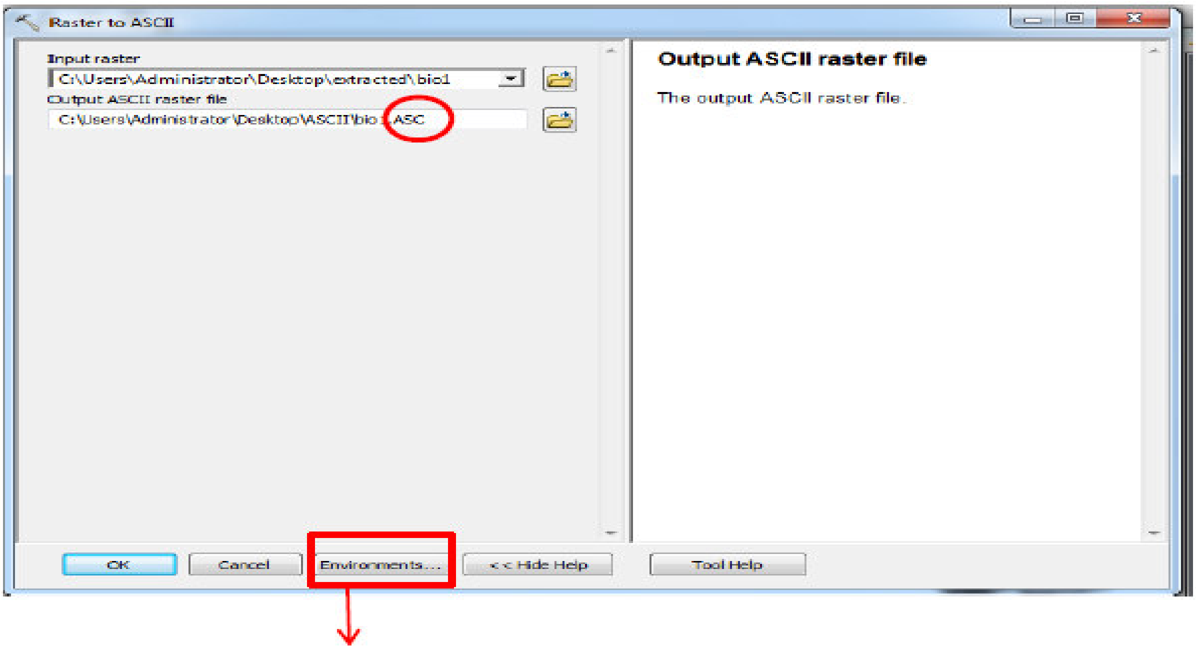
Output ASCII raster file
C:\Users\Administrator\Desktop\ASCII\bio1.TXT

Input raster

The input raster dataset.

The raster can be integer or floating-point type.

- As you see in figure of output ASCII raster file "C:\Users\Administrator\Desktop\ASCII\bio1.TXT" please change manually the TXT to ASC and seems:



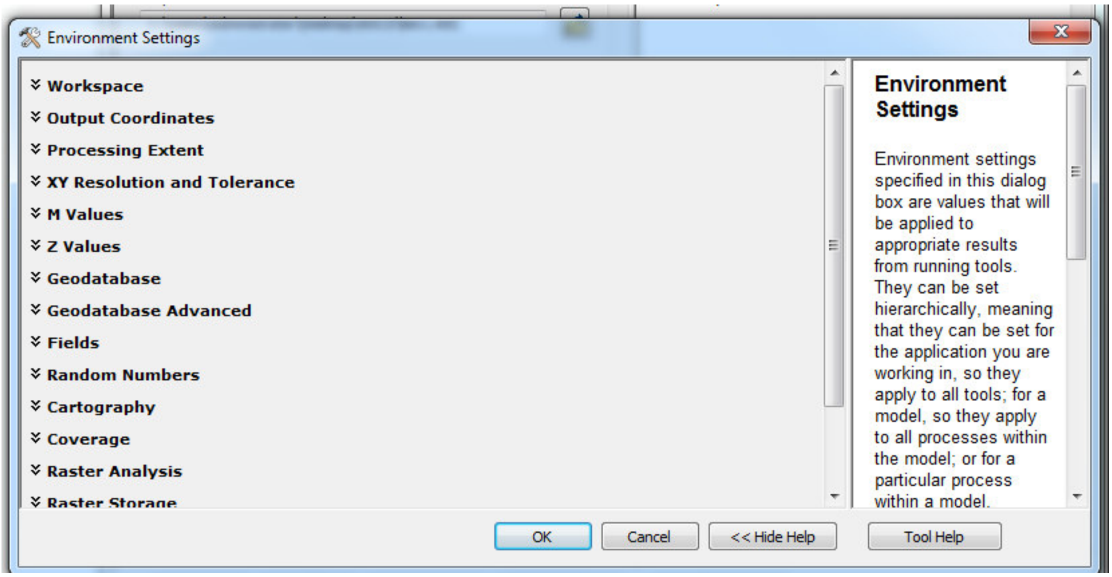
Input raster: C:\Users\Administrator\Desktop\extracted\bio1

Output ASCII raster file: C:\Users\Administrator\Desktop\ASCII\bio1.ASC

Output ASCII raster file

The output ASCII raster file.

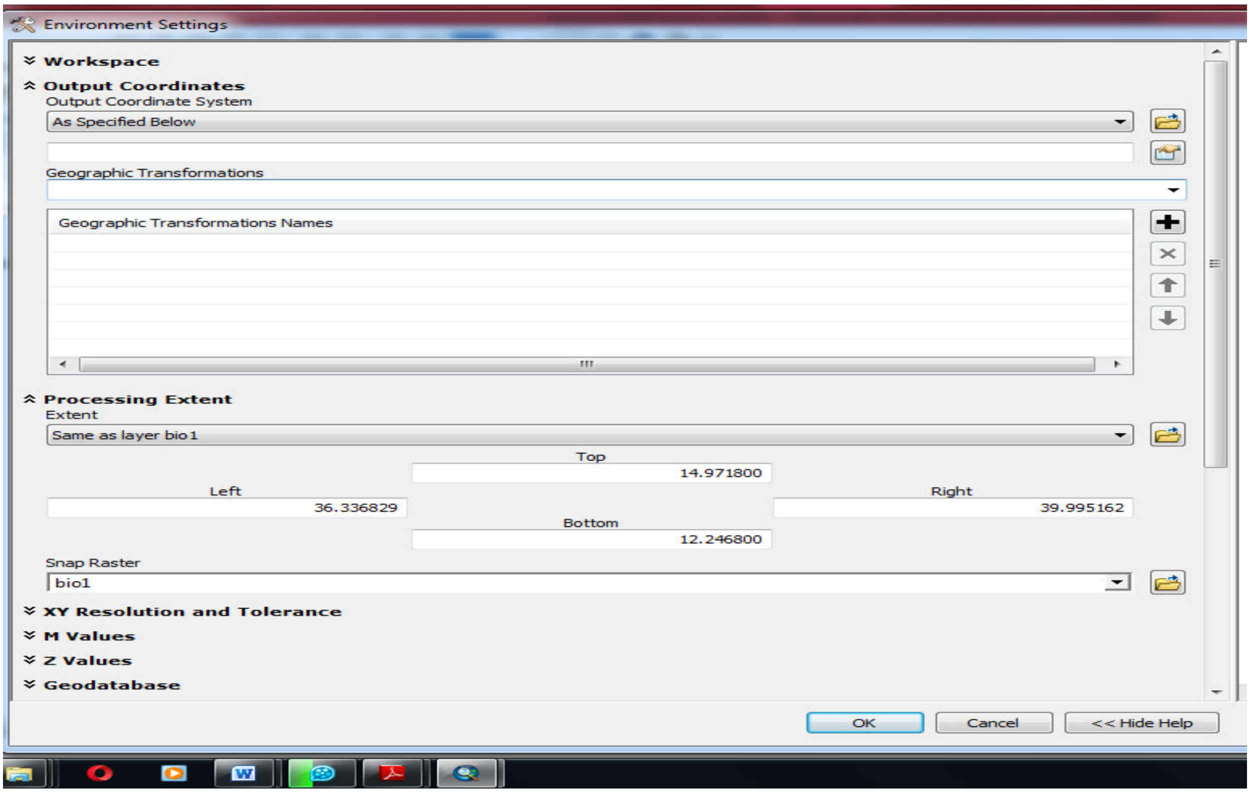
- Click on Environments and seems:



Environment Settings

Environment settings specified in this dialog box are values that will be applied to appropriate results from running tools. They can be set hierarchically, meaning that they can be set for the application you are working in, so they apply to all tools; for a model, so they apply to all processes within the model; or for a particular process within a model.

- Click on **output coordinates** and say **as specified below**
- Click on **process extent** and say it the **same as layer bio1**and in the **snap raster** say it **bio 1**
- Click on **raster analysis** and say itthe **same as layer bio1**and seems:



Environment Settings

Workspace

Output Coordinates

Output Coordinate System: As Specified Below

Geographic Transformations

Geographic Transformations Names

Processing Extent

Extent: Same as layer bio1

Top: 14.971800

Left: 36.336829

Right: 39.995162

Bottom: 12.246800

Snap Raster: bio1

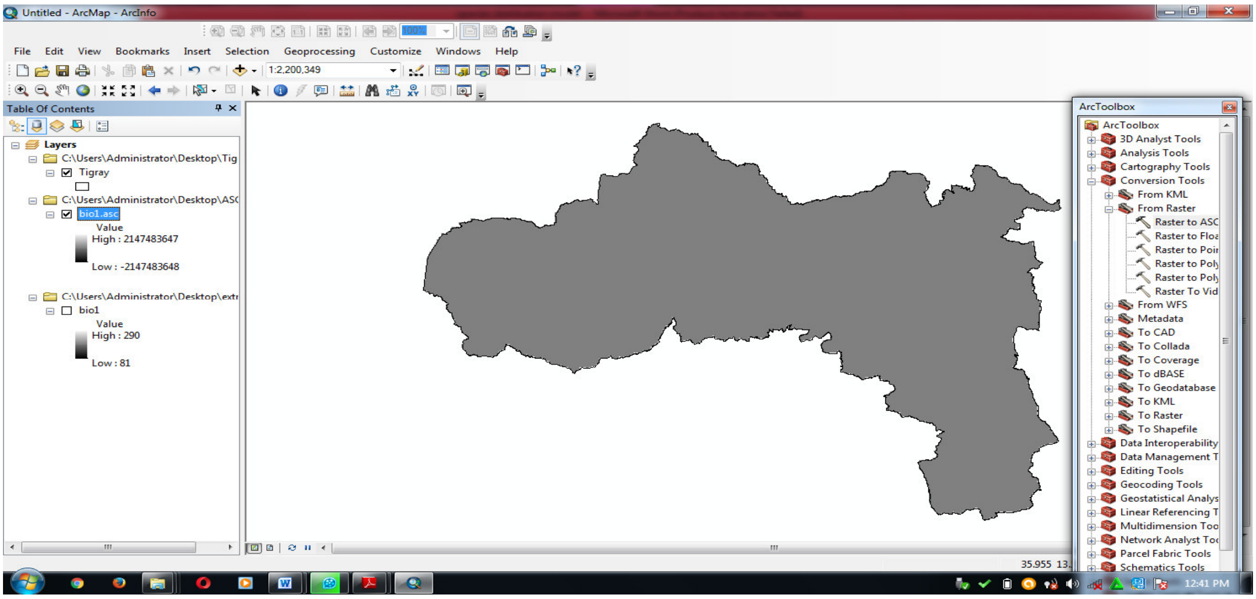
XY Resolution and Tolerance

M Values

Z Values

Geodatabase

Then click **ok** button and seems



The same as bio 1 the remain bio 2, bio3….bio 19 should adjust their environment setting as:

- Click on **output coordinates** and say it the **same as layer bio1**
- Click on **process extent** and say it the **same as layer bio1** and in the **snap raster** say it **bio 1**
- Click on **raster analysis** and say it the **same as layer bio1**

This is used to make the ASCII files to be the same coordinate location; the same cell size, if not the work will be invalid because the software cannot read it.

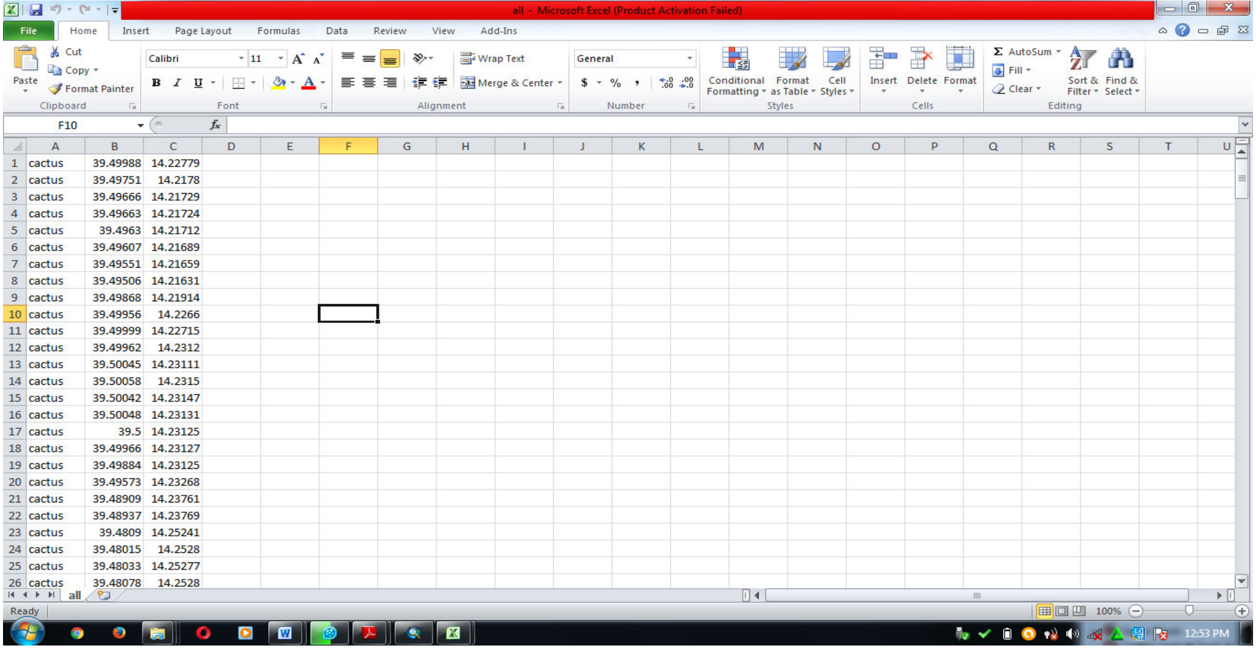
NB: These outputs are used as input data for Maxent (maximum entropy model), Diva GIS, Openmodeller, and Genetic Algorithm for Rule-set Production (GARP) and other models.

2) Maxent 3.3.3k

Maxent is a machine learning process that uses a statistical mechanics approach and requires only presence data, along with a suite of environmental variables relevant to the focal species' distribution (4). It uses the principle of maximum entropy on presence only data to estimate a set of functions that relate environmental variables and habitat suitability in order to approximate the species' niche and potential geographic distribution (7).In addition, it establishes relationships between occurrences of species and environmental conditions in the study area.

2.1) Converting the GPS location points from excels to comma separated value (CSV)

Maxent model reads the GPS location points only if they are change in to comma separated value (CSV).The excel files should be changed to decimal degree using Arc GIS. Then the excel file can change using file > save as> save as type > CSV (comma delimited) then save. They are also arranged as the following in order compatible for the software.

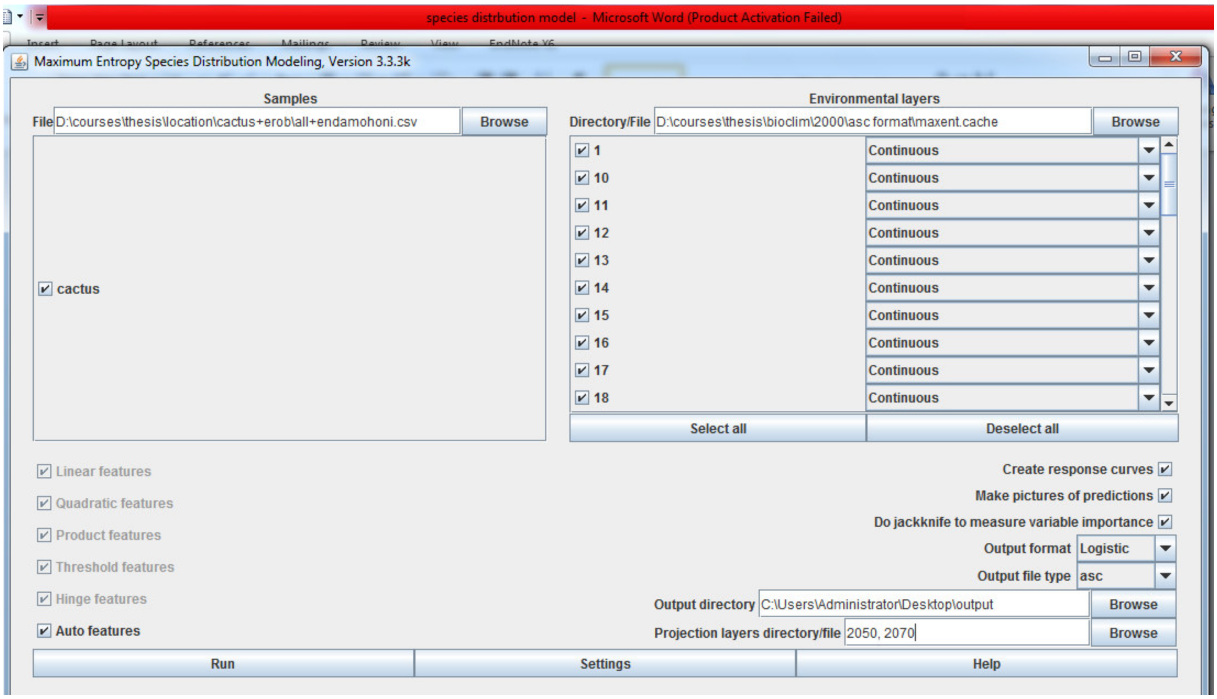


	A	B	C
1	cactus	39.49988	14.22779
2	cactus	39.49751	14.2178
3	cactus	39.49666	14.21729
4	cactus	39.49663	14.21724
5	cactus	39.4963	14.21712
6	cactus	39.49607	14.21689
7	cactus	39.49551	14.21659
8	cactus	39.49506	14.21631
9	cactus	39.49868	14.21914
10	cactus	39.49956	14.2266
11	cactus	39.49999	14.22715
12	cactus	39.49962	14.2312
13	cactus	39.50045	14.23111
14	cactus	39.50058	14.2315
15	cactus	39.50042	14.23147
16	cactus	39.50048	14.23131
17	cactus	39.5	14.23125
18	cactus	39.49966	14.23127
19	cactus	39.49884	14.23125
20	cactus	39.49573	14.23268
21	cactus	39.48909	14.23761
22	cactus	39.48937	14.23769
23	cactus	39.4809	14.25241
24	cactus	39.48015	14.2528
25	cactus	39.48033	14.25277
26	cactus	39.48078	14.2528

2.2) Adjusting the setting

Maxent has two windows which are sample window and environmental layers window. Sample window is used to insert the CSV GPS presence location of the species and it can insert different types of species but the environmental layers window is used to insert the ASCII format of the climatic variables, soil, vegetation cover, altitude and etc. This window can used to arrange categorical and continues environmental variables.

- Browse the CSV GPS location in **sample** window
- Browse ASCII folder from desktop in **environmental layers** window
- Create output folder in the desktop and browse in to output directory
- In projection layers directory/file you can insert different time prediction like 2000, 2050, 2050 by writing the times on hand comma then space and seems:
- Tick setting as the below figure and adjust setting:



species distrbution model - Microsoft Word (Product Activation Failed)

Maximum Entropy Species Distribution Modeling, Version 3.3.3k

Samples

File D:\courses\thesis\location\cactus+erob\all+endamohoni.csv [Browse]

☑ cactus

Environmental layers

Directory/File D:\courses\thesis\bioclim\2000\asc format\maxent.cache [Browse]

☑ 1	Continuous
☑ 10	Continuous
☑ 11	Continuous
☑ 12	Continuous
☑ 13	Continuous
☑ 14	Continuous
☑ 15	Continuous
☑ 16	Continuous
☑ 17	Continuous
☑ 18	Continuous

☑ Linear features
☑ Quadratic features
☑ Product features
☑ Threshold features
☑ Hinge features
☑ Auto features

Create response curves ☑
Make pictures of predictions ☑
Do jackknife to measure variable importance ☑
Output format Logistic
Output file type asc
Output directory C:\Users\Administrator\Desktop\output
Projection layers directory/file 2050, 2070

- Random seed and Random seed test

Model evaluations are essential to test the predictive performance of SDMs. The model performance is determined by means of receiver operating characteristic (ROC) plots. This also can perform by dividing the occurrences data into two parts which are training data that is used to calibrate the model and test data is used to know model accuracy. These are quantified by the area under curve (AUC). They are between 0.5 (random) and 1. AUC closest to one (1) showed that the model performance is excellent.

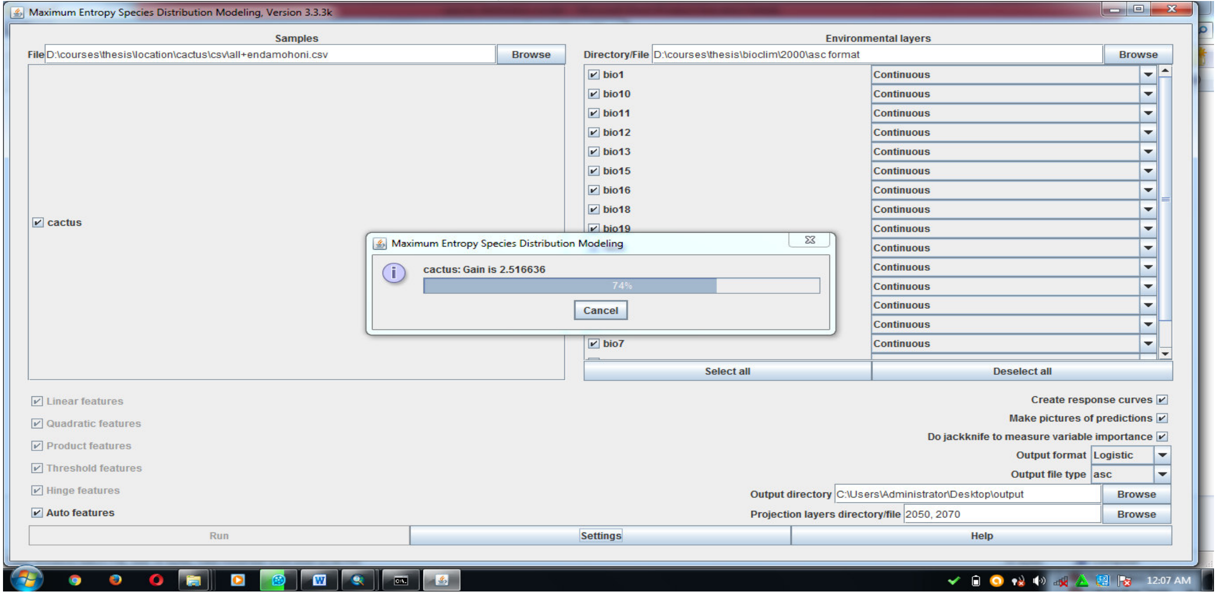
- Put regularization multiplier as default

The "regularization multiplier" is used a smoothing parameter and used to reduce model over-fitting (11). In the model, a default value of 1 is used for the regularization parameter (3).Therefore, leave it as default.

- Maximum iteration number

The other approach used to enhance model performance was adjusting maximum iteration numbers. Normally Maxent sets 500 maximum iteration numbers as default, but you can increase more than as the default (500). This allows the model to have adequate time for convergence.

- Leave the remain settings as default and click on **Run** button



Maximum Entropy Species Distribution Modeling, Version 3.3.3k

Samples

File D:\courses\thesis\location\cactus\csv\all+endamohoni.csv [Browse]

☑ cactus

Environmental layers

Directory/File D:\courses\thesis\bioclim\2000\asc format [Browse]

☑ bio1	Continuous
☑ bio10	Continuous
☑ bio11	Continuous
☑ bio12	Continuous
☑ bio13	Continuous
☑ bio15	Continuous
☑ bio16	Continuous
☑ bio18	Continuous
☑ bio19	Continuous
	Continuous
	Continuous
	Continuous
	Continuous
	Continuous
☑ bio7	Continuous

Maximum Entropy Species Distribution Modeling

cactus: Gain is 2.516636

74%

☑ Linear features
☑ Quadratic features
☑ Product features
☑ Threshold features
☑ Hinge features
☑ Auto features

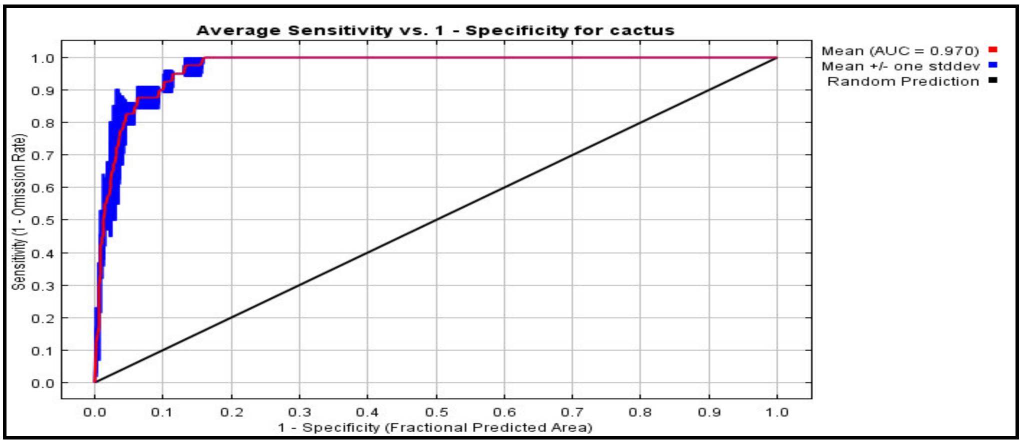
Create response curves ☑
Make pictures of predictions ☑
Do jackknife to measure variable importance ☑
Output format Logistic
Output file type asc
Output directory C:\Users\Administrator\Desktop\output
Projection layers directory/file 2050, 2070

CONGRATULATIONS

2.3) output

The main output file for the Maxent model are in the form of image (ASCII), logistic curve, and default browser of the computer, excel, file folder (called plot) and others. Besides the most useful one is default browser result (Mozilla Firefox, Google chrome, Internet Explorer) that contains information on the overall averaging of all model runs that were specified with statistical analyses, plots, model images, and links to the other files and runs. It also contains the parameters used in the model and model evaluation.

- Model evaluation

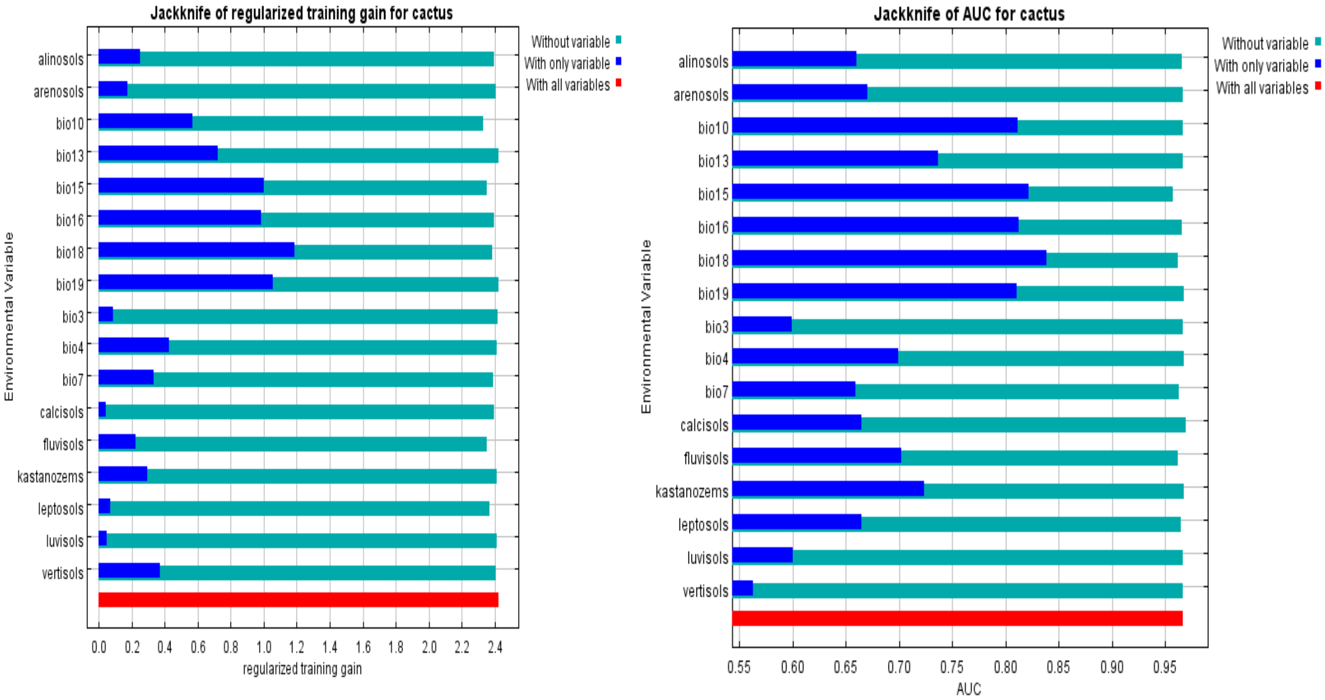


- **Variables contribution to the predictive model**

All environmental variables were inserted to Maxent to know their individual contribution. Maxent measures the environmental contribution through percent contribution table and jackknife. Jackknife measures for test data, training data and area under curve. Jackknife provides information on the performance of each variable in the model in terms of how important each variable is in explaining the species distribution and how much unique information each variable provide.

Contribution and permutation importance of different variables for the distribution of cactus

Variable	Percent contribution	Permutation importance
Bio 18	25.7	2.1
Bio 19	18.2	0
Bio 15	8.7	4.3

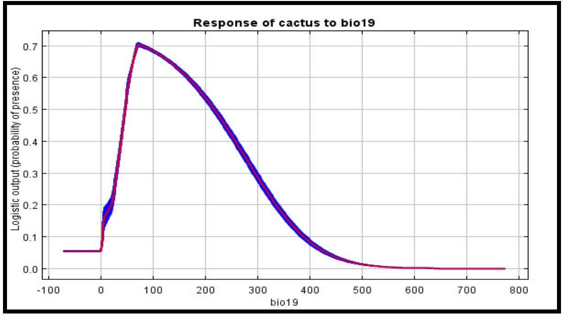


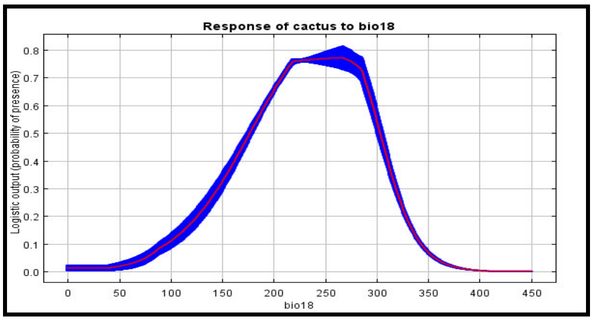
Results of jackknife evaluation of relative importance of the variables

Finally, the basic process done during the environmental predictors selections were:
1. Use all the pre-selected variables to run the model
2. Check the jackknife test (percent contribution table) results. Omit the variables which have zero or negative effects.
3. Use the remaining variables to run the model and check the jackknife (percent contribution table) results and omit other variables, and
4. Repeat the step until the variables have positive effect to the total gain.

- **Response curves**

Maxent indicates the response of species to different enviromental variables using response curves. It is also important to know in what manner each variable influences species distributions. The variable response curves are displayed by Maxent (using logistic output). The response curves consist of a chart with upward trends for variables indicate a positive association, downward movements represent a negative relationship. As the following figures:

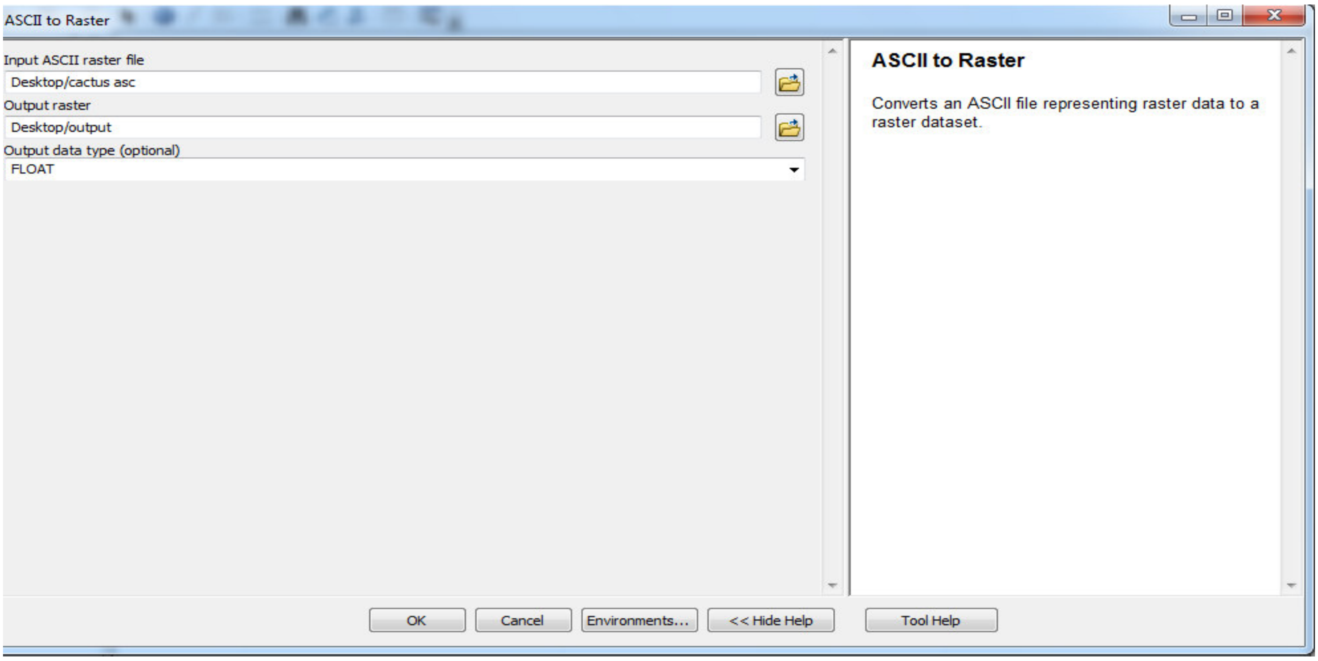




- Map output

Maxent output is also gives a continuous ASCII map. The average ASCII file output can change in Arc GIS using:

- Arc tool box > conversion > to raster > ASCII to raster


ASCII to Raster

Input ASCII raster file: Desktop/cactus asc

Output raster: Desktop/output

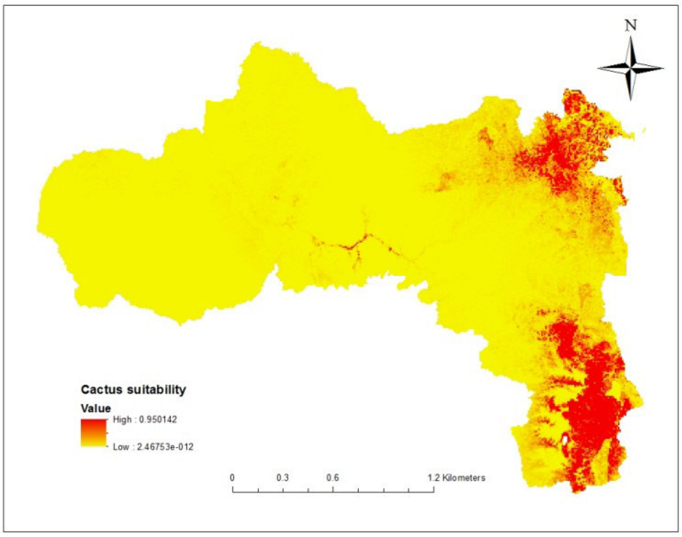
Output data type (optional): FLOAT

ASCII to Raster

Converts an ASCII file representing raster data to a raster dataset.

- Change INTEGER in to FLOAT in **Output data type**

Then Maxent default setting produces with values from 0-1 representing habitat suitability. The suitability map shows with different colors and the image uses colors to indicate predicted probability.



- The red color indicates cactus distribution area, but the yellow one indicates cactus unsuitable area. And you can classify the map according the following table:

Suitability threshold

SN	Threshold value	Threshold description
01	0.7087-1.0000	The geographical ranges of the excellent area
02	0.5315-0.7087	Optimum area
03	0.3543- 0.5315	Suitable area
04	0.1772-0.3543	Less suitable area
05	0.0000-0.1772	Unsuitable area

Sources: (9)

CONGRATULATIONS

3) Openmodeller

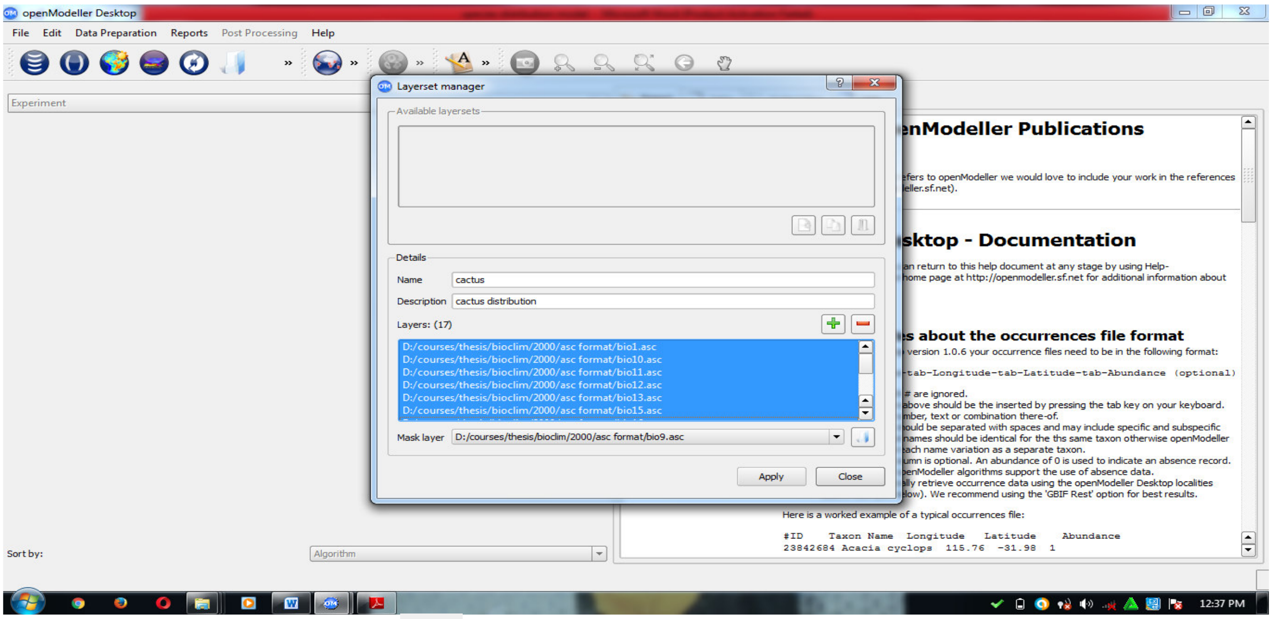
3.1) converting the GPS points from excels to word

The occurrence points of the species should be saved as compatible by open modeler software. The points are arranged in excel and excel should be saved as word (TXT). Since, open modeler reads when the occurrence points are in word format.

#id	label	long	lat	abundance

3.2) adjusting the setting

- Click on **layer sets** ()> write in **name** cactus > write in **description** cactus distribution, then click on add ()> browse environmental variables by pressing > select all climatic variables by pressing CTRL then it appears as following figure:



Details

Name: cactus

Description: cactus distribution

Layers: (17)

```
D:/courses/thesis/bioclim/2000/asc format/bio1.asc
D:/courses/thesis/bioclim/2000/asc format/bio10.asc
D:/courses/thesis/bioclim/2000/asc format/bio11.asc
D:/courses/thesis/bioclim/2000/asc format/bio12.asc
D:/courses/thesis/bioclim/2000/asc format/bio13.asc
D:/courses/thesis/bioclim/2000/asc format/bio15.asc
```

Mask layer: D:/courses/thesis/bioclim/2000/asc format/bio9.asc

enModeller Publications

efers to openModeller we would love to include your work in the references
eller.sf.net).

sktop - Documentation

an return to this help document at any stage by using Help-
home page at http://openmodeller.sf.net for additional information about

s about the occurrences file format

version 1.0.6 your occurrence files need to be in the following format:

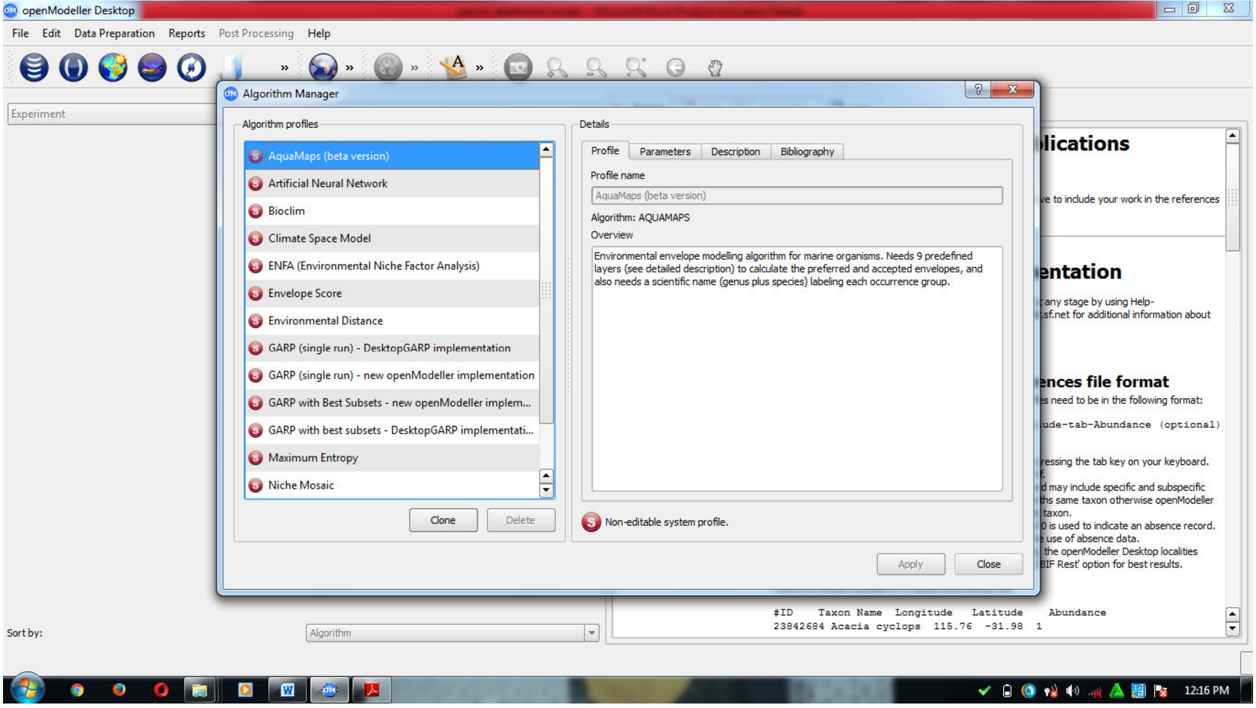
```
-tab-Longitude-tab-Latitude-tab-Abundance (optional)
```

# are ignored.
above should be the inserted by pressing the tab key on your keyboard.
mber, text or combination there-of.
ould be separated with spaces and may include specific and subspecific
names should be identical for the ths same taxon otherwise openModeller
ach name variation as a separate taxon.
umn is optional. An abundance of 0 is used to indicate an absence record.
penModeller algorithms support the use of absence data.
lly retrieve occurrence data using the openModeller Desktop localities
low). We recommend using the 'GBIF Rest' option for best results.

Here is a worked example of a typical occurrences file:

```
#ID      Taxon Name   Longitude   Latitude    Abundance
23842684 Acacia cyclops   115.76  -31.98   1
```

- Click the algorism profiles ( ). It seems:



Algorithm profiles:

- AquaMaps (beta version)
- Artificial Neural Network
- Bioclim
- Climate Space Model
- ENFA (Environmental Niche Factor Analysis)
- Envelope Score
- Environmental Distance
- GARP (single run) - DesktopGARP implementation
- GARP (single run) - new openModeller implementation
- GARP with Best Subsets - new openModeller implem...
- GARP with best subsets - DesktopGARP implementati...
- Maximum Entropy
- Niche Mosaic

Profile name: AquaMaps (beta version)

Algorithm: AQUAMAPS

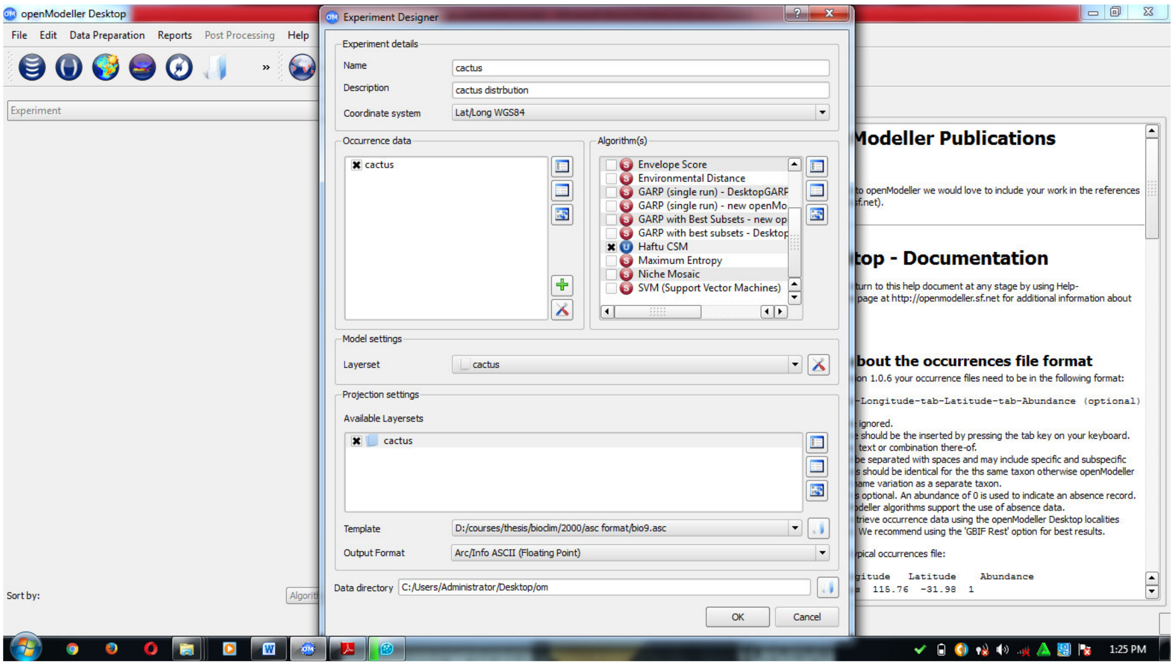
Overview

Environmental envelope modelling algorithm for marine organisms. Needs 9 predefined layers (see detailed description) to calculate the preferred and accepted envelopes, and also needs a scientific name (genus plus species) labeling each occurrence group.

S Non-editable system profile.

To adjust the setting of the algorisms or models of the open modeler it is possible by clicking on **clone.** But before selecting the clone button, select one algorism profiles. For example select Climate Space Model then select clone button. Then rename the **Copy 1 of Climate Space Model** in to **Haftu CSM** and you can edit setting of the algorism then after click **apply** and **close** button. Algorism profiles found in open modeler are Aqua Maps, Artificial Neural Network, Bioclim, Climate Space Model, Environmental Niche Factor Analysis, Envelope Score, GARP, Maxent, Niche Mosaic and Support Vector Machine.

- Select new experiment button ( ) > write in **name** cactus > write in **description** cactus distribution > in the occurrence data there is Add button ( ) > browse occurrence points (word format).
- Select **Haftu CSM** algorism from list of algorisms
- Select and make output directory folder


openModeller Desktop
File Edit Data Preparation Reports Post Processing Help
Experiment
Sort by:
Modeller Publications
top - Documentation
bout the occurrences file format

Experiment Designer

Experiment details

Name: cactus

Description: cactus distrbution

Coordinate system: Lat/Long WGS84

Occurrence data

☒ cactus

Algorithm(s)

- ☐ Envelope Score
- ☐ Environmental Distance
- ☐ GARP (single run) - DesktopGARP
- ☐ GARP (single run) - new openMo
- ☐ GARP with Best Subsets - new op
- ☐ GARP with best subsets - Desktop
- ☒ Haftu CSM
- ☐ Maximum Entropy
- ☐ Niche Mosaic
- ☐ SVM (Support Vector Machines)

Model settings

Layerset: cactus

Projection settings

Available Layersets

☒ cactus

Template: D:/courses/thesis/bioclim/2000/asc format/bio9.asc

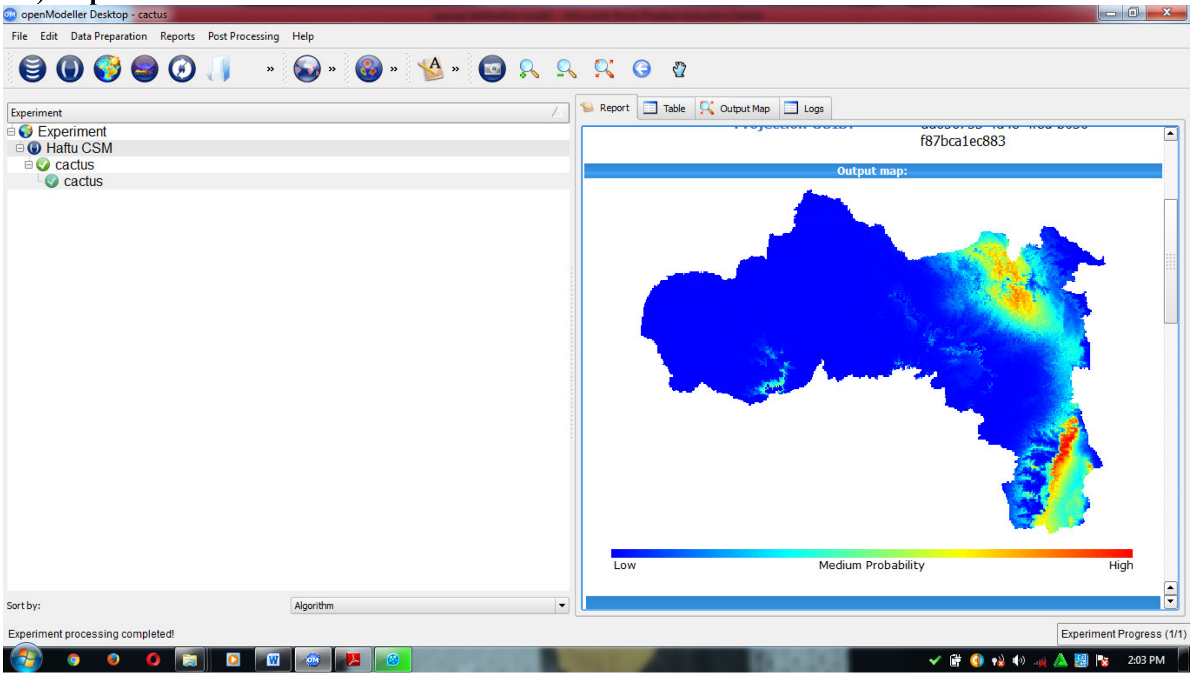
Output Format: Arc/Info ASCII (Floating Point)

Data directory: C:/Users/Administrator/Desktop/om

OK Cancel

Press **ok** button

3.3) output


openModeller Desktop - cactus
File Edit Data Preparation Reports Post Processing Help
Report
Table
Output Map
Logs
Output map:
Low
Medium Probability
High
Sort by:
Algorithm
Experiment Progress (1/1)
2:03 PM

Experiment

- Experiment
  - Haftu CSM
    - cactus
      - cactus

f87bca1ec883

Experiment processing completed!

Detailed Model Reports

Taxon	cactus
Algorithm	Haftu CSM (climate space model)
Average Area Under Curve (AUC)	0.90
Average Accuracy:	0.8979% (using 50% threshold)
Model Creation Layer set	cactus

Congratulations

Acknowledgement

- First and for most, I would like to thank the "**Almighty God**" who made it possible, to begin and finish this work successfully, and for His protection and favor in my entire life.
- I would like to thank my family Abrha Mengesha, AbebaBerhe, Senbetay Abrha, Weyni Fiseha, Mulugeta Abrha, Mihiret Abrha, Samrawit Abrha, Milana Abrha and my best friends Zinabu Mehari,

Yasin Mehomed and MehamodAwel Seid, Abadi mereseie for their fruitful contribution.

References

1.Ceccarelli, S., Balsalobre, A., Susevich, M. L., Echeverria, M. G., Gorla, D. E. & Marti, G. A. 2015. Modelling the potential geographic distribution of triatomines infected by Triatoma virus in the southern cone of South America. *Parasites & vectors,* 8**,** 153.

2. Füssel, H.-M. 2009. An updated assessment of the risks from climate change based on research published since the IPCC Fourth Assessment Report. *Climatic Change,* 97**,** 469-482.

3. Girma, A., de Bie, C., Skidmore, A. K., Venus, V. & Bongers, F. 2016. Hyper-temporal SPOT-NDVI dataset parameterization captures species distributions. *International Journal of Geographical Information Science,* 30**,** 89-107.

4. Groff, L. A. 2011. *A species distribution model for guiding Oregon spotted frog (Rana pretiosa) surveys near the southern extent of its geographic range.* Humboldt State University.

5. Intergovermental panel on climate change (IPCC).2000

6. Miola, D. T., Freitas, C. R., Barbosa, M. & Fernandes, G. W. 2011. Modeling the spatial distribution of the endemic and threatened palm shrub Syagrus glaucescens (Arecaceae). *Neotropical Biology and Conservation,* 6**,** 78-84.

7. Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling,* 190**,** 231-259.

8.Potgieter, L., Van Vuuren, J. & Conlong, D. 2013. Modelling the effects of the sterile insect technique applied to Eldana saccharina Walker in sugarcane. *ORiON: The Journal of ORSSA,* 28**,** 59-84.

9. Reddy, M. T., Begum, H., Sunil, N., Rao, P. S., Sivaraj, N. & Kumar, S. 2014. Preliminary characterization and evaluation of landraces of Indian spinach (Basella spp. L.) for agro-economic and quality traits. *Plant Breeding and Biotechnology,* 2**,** 48-63.

10. Wayne, G. 2013. The beginner's guide to representative concentration pathways. *Skeptical Sci., URL: http://www. skepticalscience. com/docs/RCP Guide. pdf.*

11. Young, N., Carter, L. & Evangelista, P. 2011. A MaxEnt model v3. 3.3 e tutorial (ArcGIS v10). *Fort Collins, Colorado*.