FUTURE AND PRESENT OCCURRENCE MODDELING ON PALM DATES

Introduction

Date palms (*Phoenix dactylifera* L.) belongs to one of the most important plant families for human use (Naeem et al. 2014; Chao and Krueger 2007) as they produce dates (Figure 1). This important fruit, with over 1000 cultivars, is serving as direct fruit source in countries over the whole world. This fruit can tolerate periods of high temperature and drought, and will also require very hot summers with almost no rain and low humidity. An old saying goes that the date palm need his feet wet, but his head in the fire, as it will need a sufficient amount of groundwater (Chao and Krueger 2007).

Figure 1 Photo of 5 dates
Source: http://www.smulweb.nl/wiki/4/Dadels

The GBIF website contains 1,534 records on *P. dactylifera* around the whole globe (Figure 2). After viewing the dataset 3 records were deleted as they did not represent *P. dactylifera*, so a total of 1531 records were used in this report. As can be seen in Figure 2 most of the distribution is around the Mediterranean Sea, the Middle East and Australia.

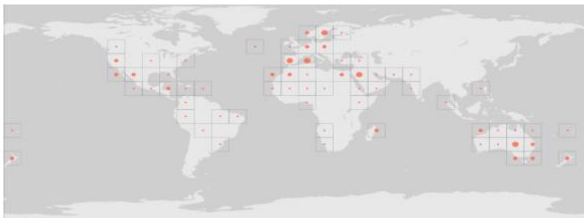


Figure 2 Occurrences of *P. dactylifera from GBIF*. Accessed on 2017-12-08.

Methodology

The '*Current*' and '*Present*' BioClim (http://www.worldclim.org/ Version: 1.4) climatic variables were used as climatic variables for the *P. dactylifera* occurrence distribution. Present data was downloaded in the '*5-min*' format and future data of 50 years later was the "*2050 RCP 4.5 HadGEM2-AO*" projection in the "*5-min"* format. These BioClim rasters were clipped to the species occurrence data, as climatic data outside *P. dactylifera* location data is not needed. Climatic variables were selected using Spearman's correlations. Correlations of 0.7 and higher were left out and in the end 5 variables were selected that had no correlation with each other. Table 1 shows the selected variables and the reason why they were selected. Future, present, and training BioClims were then exported as '*asc*' format for species modelling.

Table 1 Five climatic variables that were selected for modelling of *P. dactylifera* occurrence data

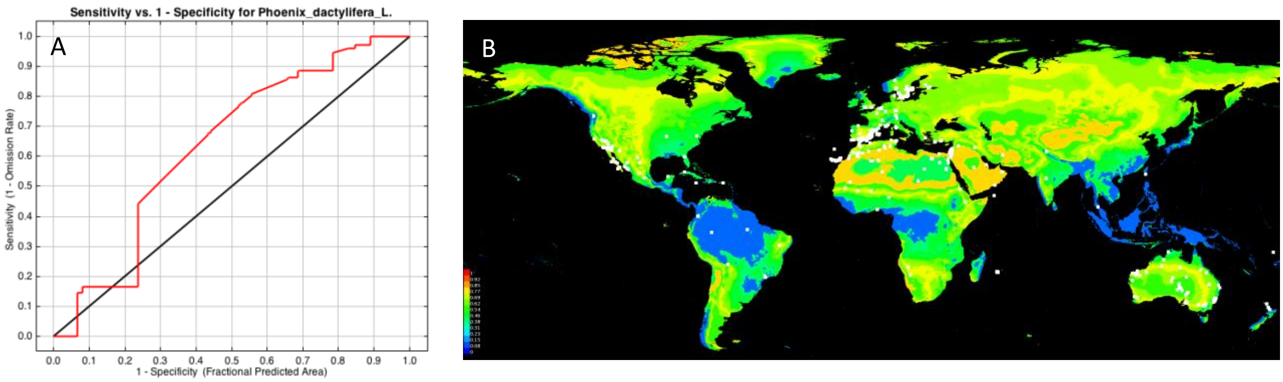
Variable	**Climatic Variable**	**Reason why used**
Bio2	Mean Diurnal range (mean of monthly (max temp – min temp))	Variable 2 showed no correlations with other variables. *P. dactylifera* is also expected to have high diurnal temperatures, because it prefers warmth.
Bio5	Max temperature of the warmest month	As *P. dactylifera* prefers high temperatures and warm summers, this variable will be important.
Bio6	Min temperature of the coldest month	Warm temperatures are preferred, low temperatures could indicate absence.
Bio12	Annual precipitation	Low humidity and almost no rain is important for *P. dactylifera.*
Bio14	Precipitation of the driest month	*P. dactylifera* prefers drought and dry climates, so this variable could be important for presence data.

For species modelling MAXENT was used (Phillips, Dudík, and Schapire 2017) with the settings as follows: '*Linear, Quadratic, Product, Threshold and Hinge features*', '*Create response curves*' and '*Make pictures of predictions*' were all checked. The output format and file type were '*Logistic*' and '*asc*'. In "*Settings"* the parameters '*Random test percentage'* was set to '*0*', '*Max number of background points*' was set to '*10,000'* and the option 'Remove duplicate presence records' was checked. While running, 19 samples were skipped because their locations were outside the climate data raster. These samples are probably located on coastal areas. A total of 542 records were used for training.

Model output

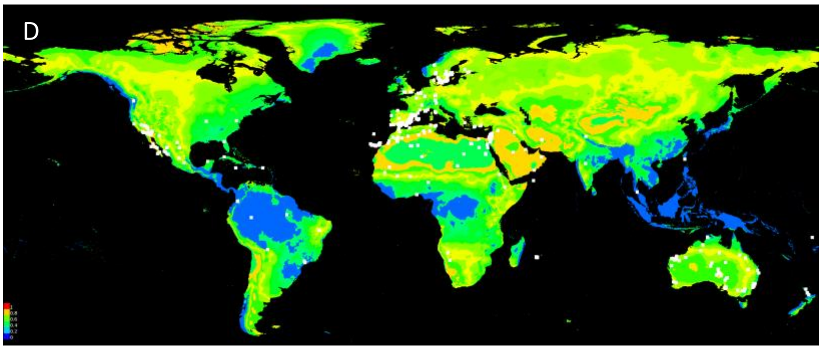
The model showed an 'area under the curve' (AUC) value of 0.665, which is better than random (0.5), but not outstanding and fairly low (Figure 3A). However, the MAXENT AUC value needs to be viewed as a relative value, to compare with other models, as it is actually using pseudo-absences (Yackulic et al. 2013).

As can be seen in Figure 3B climatic conditions are highest around the equator and around desserts (North Africa, Australia, Middle East) and lowest around tropical forests (Central Africa, Amazon, and South-East Asia). This supports the statements made in the introduction on preferred climate for *P.*

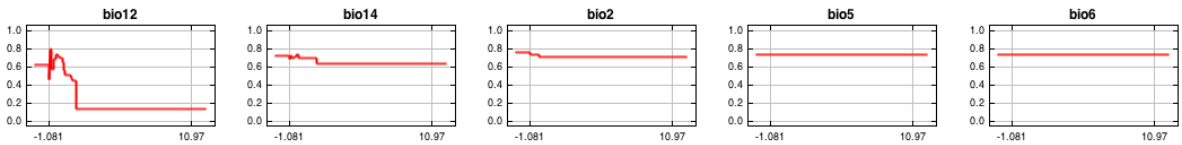


C

Var.	Percent contri-bution	Permuta-tion im-portance
bio12	97.5	97.6
bio14	2.1	2.4
bio2	0.4	0
bio5	0	0
bio6	0	0



E



F

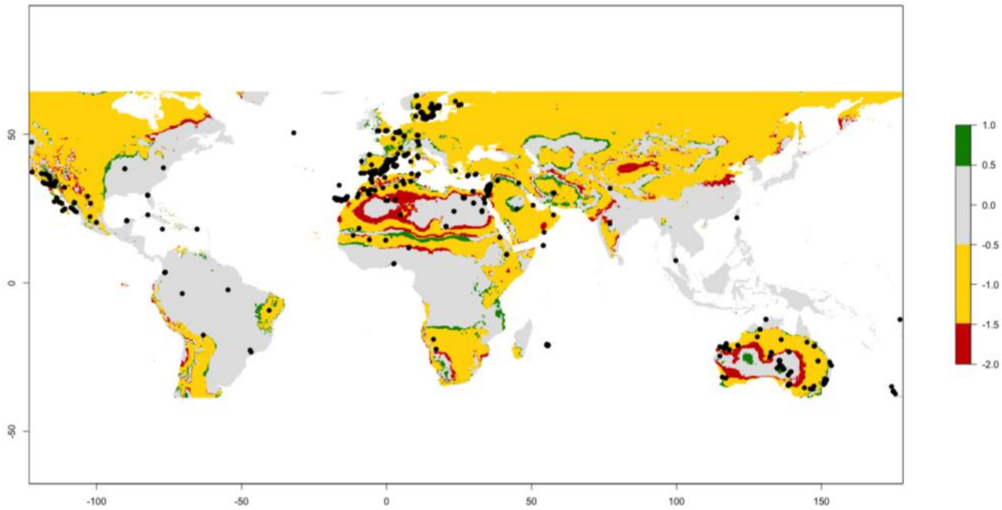


Figure 3 MAXENT output for *P. dactylifera*. A: ROC curve for training data (red line) and random (black line). The red line shows the fit of the training data to the model. The more the red line shifts to the top left, the better predictions can be made for presence data. B: Projection map of present climate data. Warmer colours show better predicted conditions. White dots show the presence locations used for training. C: The variable importance table for occurrence data. Data is ranked from high to low contributing importance. D; Projecting map of future climate data. Warmer colours show better predicted conditions. White dots show the presence locations used for training. E: Climatic variable curves showing the optimum and how their values could change the model. F. Climatic envelope of *P. dactylifera*. Grey is unsuitable habitat, yellow remains suitable and green is extra gained habitat. Black dots represent the presence locations from the dataset.

Dactylifera. This remains mostly the same in the future distribution (Figure 3D), but the differences are covered in the next chapter (Response to Future Scenario).

The threshold selection '*maximum training sensitivity plus specificity'* had a value of 0.558. This selection method is viewed as a good tool for presence data only (Norris 2014).

The Bio12 variable is by far the most important factor in this model. It is responsible for 97.6% of the model. Next is Bio14 with 2.4%. Bio2, Bio5 and Bio6 show no contributions to the model (Figure 3C). We can conclude that Precipitation (Bio12) is the most important factor in this model in predicting suitable conditions and precipitation factors cover the entire model.

Also interesting is that Bio12 clearly has an optimum for its impact on the suitability (Figure 3E), so there should be some precipitation, just enough, but not too much. This is also found as a requirement for *P. dactylifera.* There is a small optimum for Bio14, but as this is variable is not really contributing that much, it is not that important.

Response to Future Scenario

With the chosen future climate projection, *P. dactylifera* will lose suitable habitat around lots of dessert regions, especially in North Africa and the Australian desserts. But, it will also gain suitable habitat in different places, like in Central Africa, as can be seen in Figure 3F. It also gains lots of suitable habitat in the Middle East and in some in Australia, especially in North-West Australia. It can be seen that most of the changes occur near deserts and mountains, probably on the border of steppe like lands / grass lands.

Biological interpretation

The model is not outstanding, but it does a fairly good job. It could probably be improved by choosing different climatic variables. Other climatic variables, that may not be that logic for *P. dactylifera* habitat may contribute to a better model. As there mostly one variable responsible (Bio12 for 97.6%) for the model, it would be wise to try to include others, and leave out Bio5 and Bio6, which actually are not important at all. And even now, with this super important variable (in this model) the AUC value is not that good. The next model should include one more on temperature information, as the model is now only explained by precipitation (99.6%) (Figure 3C).

It can be seen that most of the changes occur near deserts and mountains, probably on the border of steppe like lands / grass lands. The world-wide distribution for the crop will not change that much according to this model. The gained suitable habitat is in almost all cases adjacent to remained suitable habitat, the same for lost suitable habitat. The only big addition in gained suitable habitat can be viewed in North West Australia.

For now, this model is only interesting, when trying to find the right amount of precipitation for this crop. The downside is that there are many cultivars and it can be expected that there is variation in the amount of precipitation. At the moment, it is not possible to incorporate this, as there could only be 11 cultivars found on GBIF.org. Outside this view, the model is not very useful at all.

Interesting is that there are some occurrences of *P. dactylifera* in non-suitable habitat (Figure 3F). A quick search, at for example the dots in grey area in North America and the Caribbean on GBIF, showed that these records are from preserved specimens and stored in a museum or in botanical gardens. These specimens should be left out in the next model.

References

Chao, Chih Cheng T, and Robert R. Krueger. 2007. "The Date Palm (Phoenix Dactylifera L.): Overview of Biology, Uses, and Cultivation." *HortScience* 42 (5): 1077–82.

Naeem, A., A.A. Khan, H.M.N. Cheema, I.A. Khan, and A. Buerkert. 2014. "DNA Barcoding for Species Identification in the Palmae Family." *Genetics and Molecular Research* 13 (4): 10341–48. doi:10.4238/2014.December.4.29.

Norris, Darren. 2014. "Model Thresholds Are More Important than Presence Location Type: Understanding the Distribution of Lowland Tapir (Tapirus Terrestris) in a Continuous Atlantic Forest of Southeast Brazil." *Tropical Conservation Science* 7 (3): 529–47. doi:10.1177/194008291400700311.

Phillips, Steven J., Miroslav Dudík, and Robert E. Schapire. 2017. "Maxent Software for Modeling Species Niches and Distributions (Version 3.4.1)." Accessed December 8. http://biodiversityinformatics.amnh.org/open_source/maxent/.

Yackulic, Charles B., Richard Chandler, Elise F. Zipkin, J. Andrew Royle, James D. Nichols, Evan H. Campbell Grant, and Sophie Veran. 2013. "Presence-Only Modelling Using MAXENT: When Can We Trust the Inferences?" *Methods in Ecology and Evolution* 4 (3): 236–43. doi:10.1111/2041-210x.12004.