Report on species distribution modeling for *Phoenix dactylifera L.*

Dwin Grashof – s2110768 – 12 December 2018

Introduction

This report will first focus on the current distribution of the date palm *Phoenix dactylifera L.* obtained from GBIF, and secondly use the current distribution together with climate data from WorldClim [1] to predict the effects of climate change on the distribution of this species. Furthermore, this report will briefly touch on the importance of data selection and parameter selection used in distribution modelling.

Phoenix d. is one of the oldest crops grown in northern Africa and the Middle East. Dates are a major food source in those areas and with that also of vital importance for the economy, environment, and society [2].

Literature discusses that the occurrence limits of the date palm lie between 10°N (Somalia) and 39°N (Elche/Spain or Turkmenistan). Favorable areas are between 24° and 34°N. However, outside these ranges the date palm can be expected because these reports suggest that the palm can survive winters below the freezing point and withstand one week at -15°C. The palm will bring itself in a resting stage and regrow damaged leaves when the temperature rises above 7°C again. This in combination with the ability to withstand temperatures of above 56°C and prolonged droughts makes this tree extremely versatile [3]. This could explain the occurrences in the higher longitudinal areas observed in GBIF (Figure 1). However, these observations can also be due to the botanical gardens growing palm trees, since the locations of these observations correlate with larger cities.

GBIF occurrence data of *Phoenix dactylifera L.*

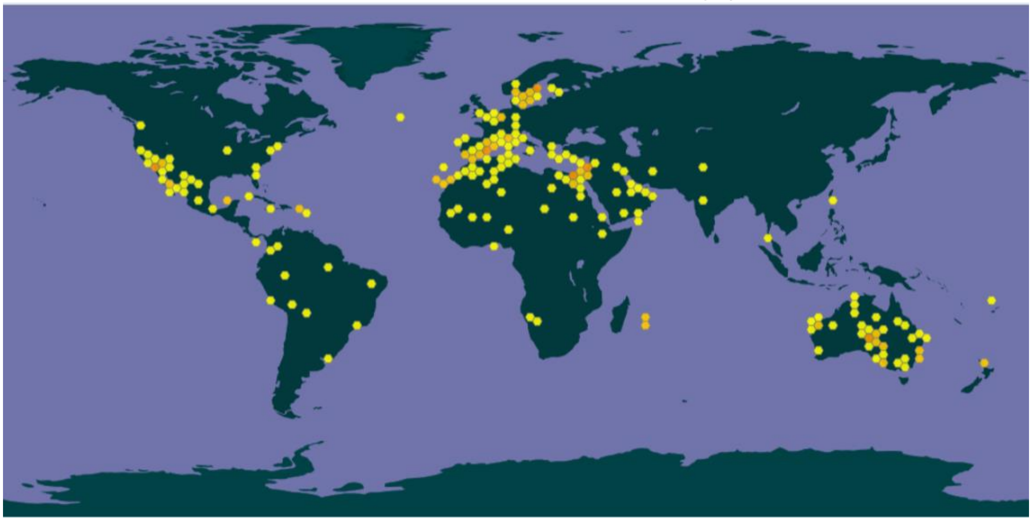

Figure 1: Current GBIF occurrence data of Phoenix dactylifera L. as per December 2018. The polygons represent occurrence and the intensity represents the density in that area.

Methodology

For the species distribution modeling two datasets from GBIF were downloaded. The first dataset holds all observations of *Phoenix d.* of which coordinates are known. The second dataset is a subset of the first dataset but only observations between 10°N and 39°N to possibly minimalize the bias of botanical gardens.

For the present and future scenarios data from WorldClim was downloaded. For the future scenario this included the rcp45 scenario projection HadGEM2-AO. Bioclim variables selected for the models is based on the preferences and limits reviewed by Zaid A. and de Wet P.F [3]. These variables include: bioclim 8 (mean temperature of wettest quarter), bioclim 10 (mean temperature of warmest quarter), bioclim 11 (mean temperature of coldest quarter), bioclim 12 (annual precipitation) and bioclim 19 (precipitation of coldest quarter). After correlation correction bioclim 10 was removed. This choice can also be supported by the fact that date palms are not significantly affected by high temperatures for prolonged times as described in literature and displayed by the model.

The model was constructed using Maximum Entropy Species Distribution Modeling (MAXENT) [4]. For MAXENT the parameters are shown in Table 1.

Table 1: MAXENT parameters that were used for both distribution models. The parameters are divided by the 'basic' and 'advanced' tabs.

MAXENT parameters

Basic		**Advanced**	
Random test percentage	0	Add samples to background	Yes
Regularization multiplier	1	Add all samples to background	No
Max number of background points	10,000	Extrapolate	Yes
Replicates	10	Do clamping	Yes
Replicated run type	Cross validation	Maximum number of iterations	500
		Convergence threshold	0.00001
		Adjust sample radius	0
		Default prevalence	0.5

Complete dataset model output

The first dataset, with all GBIF occurrences with known localities [5], in MAXENT resulted in the following models: Figure 2 for the present distribution probability prediction and Figure 3 for the future distribution probability prediction. For both maps the point-wise mean of the models from the cross validation were choses as representatives of the model.

The map of the present data shows that the highest probabilities the date palm distribution is in northern Africa, north side of the Middle East and south Europe. However, southern Australia has a high occurrence as well, which is expected. The most notable thing about this model is the high occurrence probabilities in northern Europe and even more surprisingly the south of Scandinavia. However, this may be due to the data points of botanical gardens possibly creating a bias towards colder climates.

Lastly, the map of the predicted future distribution seems comparable to the present distribution. Notable is that is seems like the distribution of the date palm shifts more to northern Europe further into Scandinavia. This might be due to one or both the presence of high longitudinal points and the predicted rise in temperature for these areas.

Present distribution probability map of *Phoenix dactylifera L*.

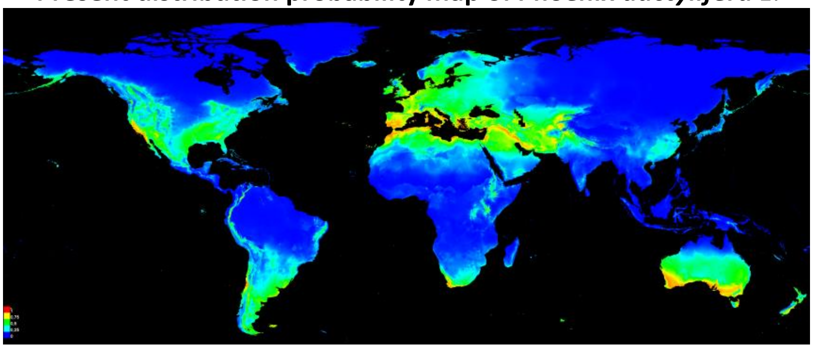

Figure 2: Showing a distribution probability model of the present based on Phoenix dactylifera L data obtained from GBIF. Warmer colors represent areas with better conditions and increased occurrence prediction.

Future distribution probability map of *Phoenix dactylifera L.*

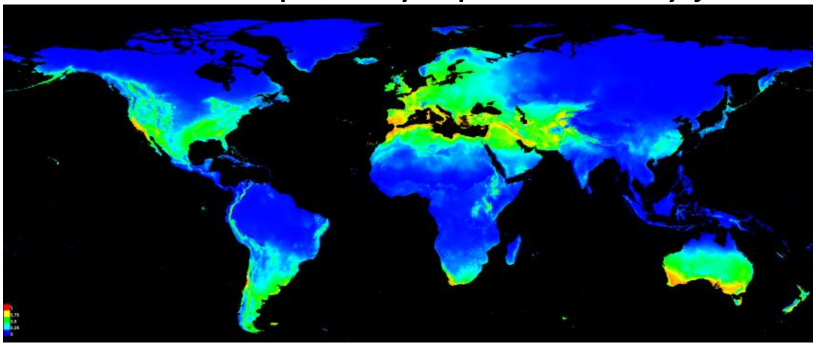

Figure 3: Showing a distribution probability model of the future based on Phoenix dactylifera L data obtained from GBIF. The future scenario was obtained as described in the methods section. Warmer colors represent areas with better conditions and increased occurrence prediction.

The model seems to perform well according to MAXENT with an average AUC of 0.884 and a standard deviation of 0.010.

Bio11 (Temperature of Coldest Quarter) shows to be the driving variable for this model as it has the most contribution (Table 2). This is also known from literature, where cold winters are a barrier for the natural spread and cultivation of the palm, while they can still be grown in colder areas as discussed earlier.

Table 2: Variance importance table as given by MAXENT for the complete dataset model.

Variable	Percent contribution	Permutation importance
Bio11	62.5	53.9
Bio19	16.6	22.9
Bio12	11.3	14.7
Bio8	9.7	8.5

Longitude filtered dataset model output

Since the model of the full dataset shows a distribution with more northern probabilities of occurrence than expected a new model was built with a dataset limited to the natural occurrence and cultivation of the date palm (between 10°N and 39°N) [6]. This excludes the northern data points and therefore possibly changing the model reaction to occurrences at low temperatures. This resulted in the following two figures (Figure 4 & 5).

Present distribution probability map of *Phoenix dactylifera L. based on longitude filtered data*

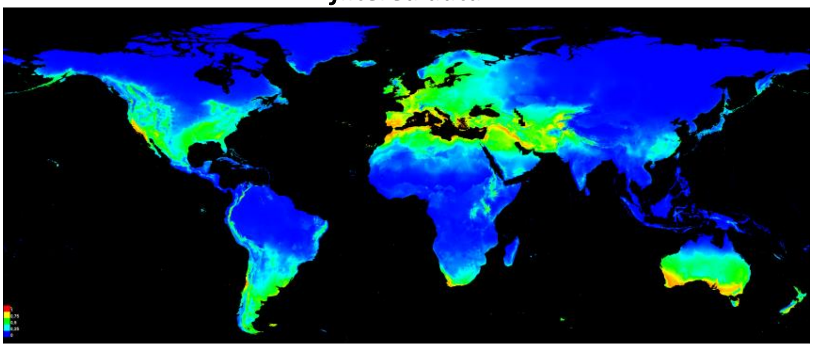

Figure 4: Showing a distribution probability model of the present based on longitude filtered Phoenix dactylifera L data obtained from GBIF. Warmer colors represent areas with better conditions and increased occurrence prediction.

Future distribution probability map of *Phoenix dactylifera L. based on longitude filtered data*

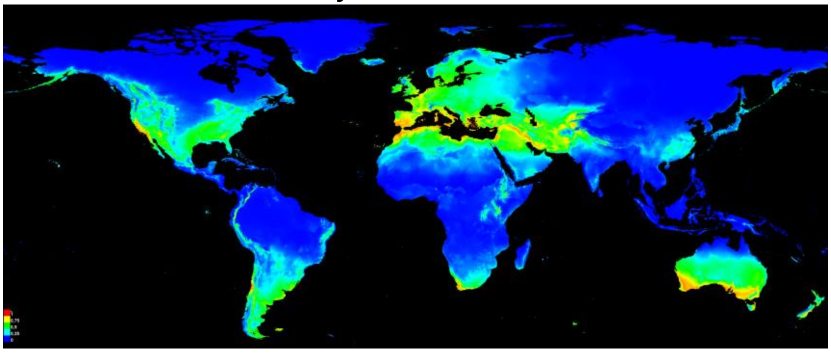

Figure 5: Showing a distribution probability model of the future based on longitude filtered Phoenix dactylifera L data obtained from GBIF. The future scenario was obtained as described in the methods section. Warmer colors represent areas with better conditions and increased occurrence prediction.

These maps surprisingly show the same if not more distribution probability in northern Europe and Scandinavia. This is unexpected since there are no northern data points linking occurrence with cold temperatures. However, what could be happening is that, looking at the full dataset (Figure 1), there is are decreasing numbers of occurrences the further north one goes. The model can interpret this as the optimum of the distribution in warmer climates, with decreasing but still somewhat probable suitability in the north. However, the filtered dataset only shows observations in northern Spain, where the number of observations actually increase towards Spain compared to northern Africa. The model might interpret this as an increasing suitability in Spain, where it is colder than in northern Africa. Therefore linking the suitability with somewhat colder areas as there are no observations to counter this linkage. This might result in the model not being able to find a suitable optimum of Bio11 and plotting the distribution of the date palm higher than expected. This is also visualized by the response curves of both models where the response curve of Bio11 for the full dataset model has a clear optimum and then declines to almost zero, while the response curve for the filtered dataset has a flattened and extended optimum and does not decrease back to zero (Figure 6A & B).

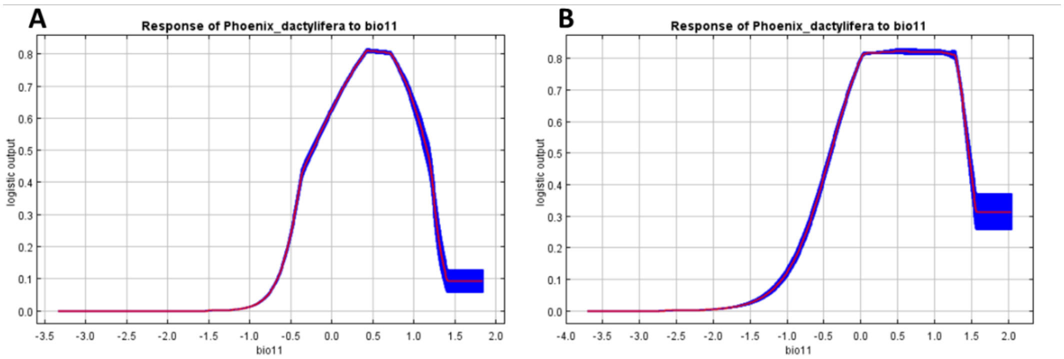

Figure 6: Response curves of Bio11 (Temperature of Coldest Quarter) A) Response curve for the complete dataset model, showing a clear optimum. B) Response curve for the longitude filtered dataset model showing a stretched optimum and no complete decline after the optimum.

The average AUC of this model is 0.892, with a standard deviation of 0.027. This is comparable to the distribution model of the full dataset.

Much like the first model, the variable with the most contribution is Bio11 (Table 3). However, this time the contribution is more evenly spread throughout the variables. This is expected since the model lacks the decreasing occurrence data for which a strong predictor can be set.

Table 3: Variance importance table as given by MAXENT for the longitude filtered dataset model.

Variable	Percent contribution	Permutation importance
Bio11	34.7	37.5

Bio8	28.4	16
Bio19	23.5	29.3
Bio12	13.4	17.2

Responses to future scenario

Since the model of the full dataset performs little different form the filtered dataset model and because this dataset has no alterations, the first model was used to look at the response to the future scenarios. When the present and future distribution models are compared to calculate the rate of change the outcome shows that there is not that much variation between the two models (Figure 7). The decrease in suitability on the upper side of the map might be due to a combination of increase in precipitation and cold temperature. According to literature the combination of high precipitation and freezing temperatures might be critical for the date palm [7]. The decrease in the Africa could be an effect of less precipitation and desertification. This might be a real threat towards the date palm. There is also an increased probability in eastern Europe, which is hard to explain.

Range change plot between present and future climate scenarios for *Phoenix dactylifera L.*

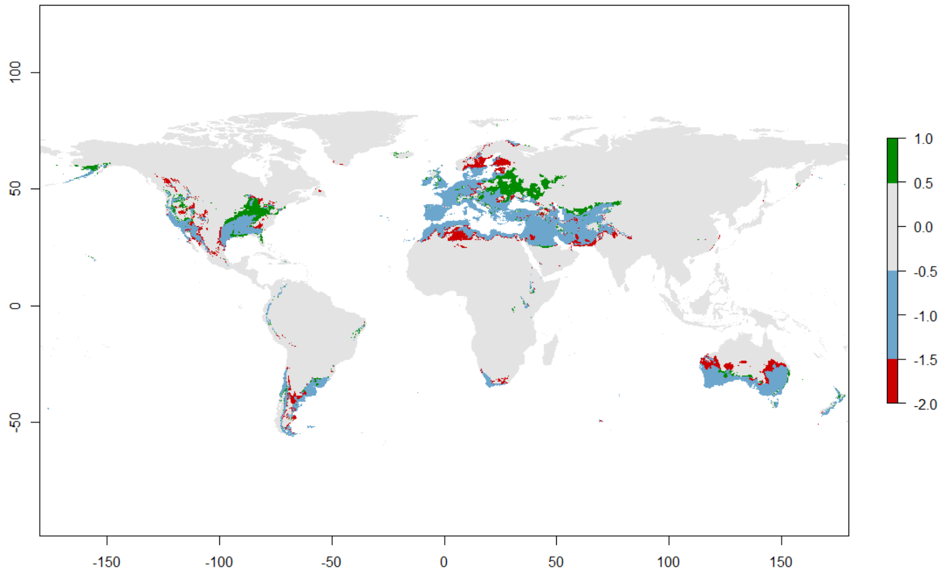

Figure 7: Range change plot between present and future climate scenarios of the complete dataset models for Phoenix dactylifera L. Red shows decrease in probability, Blue shows no increase and no decrease in probability, Grey shows areas where probabilities were low and are still low, and Green shows areas where the probability increased.

Biological interpretation

Looking at the model and combining that with the literature and biology of the date palm, this species is unlikely to be very affected by climate change, therefore not changing the distribution to a great extent. The date palm is extremely durable and has a low dispersion, contributing to the expectation that the species will not change in its distribution. However, the optimal cultivation range is slowly moving upwards as climate change continues. This can either lead to date farms following the preferences of the date palm and move upwards as well. Or secondly, it can lead to selection by farmers for date palm that have a high yield in higher temperatures. The selection of the date palm might however be extremely slow since most palm trees grow easily to over a hundred years old, and take around 20 years to fully mature and produce a lot of fruit. Resulting in a slow generation time and likely results in a slow evolving species. Because of this longer generation time and its durability the date palm is improbable to acquire biological changes.

In conclusion, the model does not show new insights and neither confirms previous hypotheses. The dataset used to model the date palm could be further optimized by extending the previously used longitudinal ranges or by manually removing biased data points. This will increase the search for the Bio11 optimum and might results in a more accurate model. Lastly, literature suggests that other important factors should be taken into account for the distribution of the date palm [3]. Other parameters should therefor include the humidity of different quarters, soil information and probably even the absence and presence of certain pollinators. With all the extra information an accurate model can be built to predict future distributions of the *Phoenix dactylifera*.

References

1. Fick S.E. and Hijmans R.J. (2017) Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology.*
2. Chao C.T., Krueger R.R. (2007) The Date Palm (Phoenix dactylifera L.): Overview of Biology, Uses, and Cultivation. *HortScience*, 42(5):1077-1082
3. Zaid A. and de Wet P.F. (2002) Date Production Support Programme. *Date Palm Cultivation*.
4. Phillips S.J., Dudík M., Schapire R.E. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2018-12-07.
5. Complete Phoenix dactylifera L. dataset. GBIF.org (07 December 2018) GBIF Occurrence Download: https://doi.org/10.15468/dl.wt2jys
6. Longitude filtered Phoenix dactylifera L. dataset. GBIF.org (07 December 2018) GBIF Occurrence Download: https://doi.org/10.15468/dl.jccnas
7. Shabani F., Kumar L. (2014) Sensitivity Analysis of CLIMEX Parameters in Modeling Potential Distribution of Phoenix dactylifera L. *PLoS ONE* 9(4): e94867. https://doi.org/10.1371/journal.pone.0094867