

# A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants

Emad Kaky<sup>a,b,c,\*</sup>, Victoria Nolan<sup>a</sup>, Abdulaziz Alatawi<sup>a,d</sup>, Francis Gilbert<sup>a</sup>

<sup>a</sup> School of Life Sciences, University of Nottingham, Nottingham NG7 2RD, UK

<sup>b</sup> Research Centre, Sulaimani Polytechnic University, Sulaymaniyah, Iraq

<sup>c</sup> Kalar Technical Institute, Sulaimani Polytechnic University, Sulaymaniyah, Iraq

<sup>d</sup> Department of Biology, University of Tabuk, Saudi Arabia

## ARTICLE INFO

### Keywords:

Species distribution modelling  
Conservation planning  
Habitat suitability  
Egypt  
Medicinal plants

## ABSTRACT

Understanding the relationship between the geographical distribution of taxa and their environmental conditions is a key concept in ecology and conservation. The use of ensemble modelling methods for species distribution modelling (SDM) have been promoted over single algorithms such as Maximum Entropy (*MaxEnt*). Nevertheless, we suggest that in cases where data, technical support or computational power are limited, for example in developing countries, single algorithm methods produce robust and accurate distribution maps. We fit SDMs for 114 Egyptian medicinal plant species (nearly all native) with a total of 14,396 occurrence points. The predictive performances of eight single-algorithm methods (*maxent*, random forest (*rf*), support-vector machine (*svm*), *maxlike*, boosted regression trees (*brt*), classification and regression trees (*cart*), flexible discriminant analysis (*fda*) and generalised linear models (*glm*)) were compared to an *ensemble* modelling approach combining all eight algorithms. Predictions were based originally on the current climate, and then projected into the future time slice of 2050 based on four alternate climate change scenarios (A2a and B2a for CMIP3 and RCP 2.6 and RCP 8.5 for CMIP5). *Ensemble* modelling, *MaxEnt* and *rf* achieved the highest predictive performances based on AUC and TSS, while *svm* and *cart* had the poorest performance. There is high similarity in habitat suitability between *MaxEnt* and *ensemble* predictive maps for both current and future emission scenarios, but lower similarity between *rf* and *ensemble*, or *rf* and *MaxEnt*. We conclude that single-algorithm modelling methods, particularly *MaxEnt*, are capable of producing distribution maps of comparable accuracy to *ensemble* methods. Furthermore, the ease of use, reduced computational time and simplicity of methods like *MaxEnt* provides support for their use in scenarios when the choice of modelling methods, knowledge or computational power is limited but the need for robust and accurate conservation predictions is urgent.

## 1. Introduction

Understanding the relationship between a species or community and its environment is a key concept in ecology and conservation. The use of species distribution models (SDMs) for this purpose has advanced the understanding of many ecological issues and is a key tool in predicting species responses to environmental change (Elith et al., 2006; Elith and Leathwick, 2009; Norberg et al., 2019; Zimmermann et al., 2010). The ability of SDM to make predictions into new spatial areas or future time periods facilitates a wide range of ecological applications including predicting habitat suitability under alternate climate change scenarios (Dormann et al., 2007; Hijmans and Graham, 2006; Pearson and Dawson, 2003), assisting with conservation and management plans (Engler et al., 2004; Zhang et al., 2012) and investigating key issues in

related disciplines such as paleobiology (Svenning et al., 2011), taxonomy (Kharouba et al., 2013) and biogeography (Guisan et al., 2006).

SDM has traditionally been focused on single-species presence data (Elith et al., 2006; Liu et al., 2011; Thuiller, 2003), but recent emphasis on the importance of community and biotic interactions in SDMs has propelled multi-species SDMs such as stacked SDMs (S-SDMs) or joint SDMs (J-SDMs) to the forefront of distribution modelling (Baselga and Araújo, 2009; Elith and Leathwick, 2009; Norberg et al., 2019). S-SDMs are more established than J-SDMs: they involve fitting models for individual species and then combining model predictions to create estimates of species richness (Benito et al., 2013; Guisan and Rahbek, 2011). In contrast, J-SDMs model a whole community of species at once, and then subsequently create predictions (Pollock et al., 2014; Warton et al., 2015). Although criticised for over-predicting species

\* Corresponding author at: Research Centre, Sulaimani Polytechnic University, Sulaymaniyah, Iraq.

E-mail address: [emadd.abbas@spu.edu.iq](mailto:emadd.abbas@spu.edu.iq) (E. Kaky).

<https://doi.org/10.1016/j.ecolinf.2020.101150>

Received 7 April 2020; Received in revised form 28 July 2020; Accepted 23 August 2020

Available online 03 September 2020

1574-9541/ Crown Copyright © 2020 Published by Elsevier B.V. All rights reserved.

richness (Guisan and Rahbek, 2011), when correctly fitted using appropriate thresholds for converting predictions into binary maps, S-SDMs can provide accurate, low-error model predictions (Benito et al., 2013; Calabrese et al., 2014) which have been shown to perform better than direct community modelling methods (Ko et al., 2016) due to their ability to capture variation in individual species-environment associations.

A wide range of SDM methods have been developed for the purpose of predicting species occurrences based on environmental characteristics (Thuiller et al., 2003; Elith and Leathwick, 2009; Norberg et al., 2019). Commonly used algorithms include generalised linear models (GLM) and generalised additive models (GAM) (Guisan et al., 2002; Hastie and Tibshirani, 2004), maximum entropy modelling (*MaxEnt*) (Elith et al., 2011; Phillips et al., 2006), random forests (*rf*) (Evans et al., 2011; Svetnik et al., 2003), boosted regression trees (*brt*) (De'ath, 2002), and other machine learning methods such as artificial neural networks (ANN) (Zurada, 1992) or Bayesian hierarchical modelling (Hefley and Hooten, 2016; Latimer et al., 2006). Many of these methods can be easily implemented simultaneously in statistical packages and programmes, and in general, the most popular tools have the more accessible and simple interfaces (Jarnevich and Young, 2015; Merow et al., 2013; Norberg et al., 2019).

*MaxEnt* in particular is heavily favoured in the scientific community. It can be implemented in a variety of packages in R including 'sdm' (Naimi and Araújo, 2016), 'dismo' (Hijmans et al., 2017) and 'ENMTools' (Warren et al., 2010), or models can be fitted using freely available, simple and user-friendly independent software programs such as 'ENMTools' and 'MAXENT'. All of these programs or packages allow many parameters to be manually determined by the user but also offer robust, well-researched default values for accurate species distribution models (Phillips and Dudík, 2008). *MaxEnt* benefits from its ability to model presence-only (PO) data (Elith et al., 2006; Phillips et al., 2006) and is thought to be robust to small sample sizes (Kaky, 2020; Kaky and Gilbert, 2016; Wisz et al., 2008), as well as being able to model complex, non-linear relationships between the response variable and predictors (Elith et al., 2006). Yet it is the ease and simplicity of its implementation that has propelled *MaxEnt* to be the most prominent, widely-used SDM technique in scientific research (Fitzpatrick et al., 2013; Fourcade et al., 2014; Hijmans and Elith, 2013). *MaxEnt* models have been criticised for being incorrectly applied and oversimplified (Morales et al., 2017; Yackulic et al., 2013), and to have produced overfitting models (Halvorsen, 2013; Merckx et al., 2011). Nevertheless, *MaxEnt* continues to be frequently used to fit models across many different taxa, geographical areas, time periods and environmental scenarios.

The choice of the most appropriate SDM method out of the large number of algorithms available for a particular organism or community has received much attention. Many studies have compared SDM performance for both single and multiple species (Beaumont et al., 2016; Bucklin et al., 2015; Elith et al., 2006; González-Irusta et al., 2015), but there is still a lack of consensus on the best choice of model (Norberg et al., 2019). Suggested sources of uncertainty in predictions include the SDM method, choice of predictor variables, spatial distribution and sample size, all of which can impact model accuracy and predictive power (Austin and Niel, 2011; Buisson et al., 2010; Edwards et al., 2006; Segurado and Araújo, 2004). Proposals to aid the process of model selection and fitting include model comparisons for specific scenarios e.g. fitting a small number of models within a cross-validation framework (Norberg et al., 2019) or ensemble modelling methods combining individual models to create one predictive output (Marmion et al., 2009; Thuiller et al., 2004).

Ensemble modelling (also known as consensus modelling or ensemble forecasting) has been gaining momentum in SDM over the past decade, and involves combining predictions from single SDM models into one predicted binary map, usually based on the average model predictions weighted by an evaluation metric e.g. AUC (Araújo and

New, 2007; Marmion et al., 2009; Thuiller et al., 2004). By combining all models, the ensemble model theoretically should produce more accurate and robust predictions than any single model (Marmion et al., 2009). Ensemble modelling has been shown to improve model predictions (Grenouillet et al., 2011; Opiel et al., 2012; Stohlgren et al., 2010), reduce overfitting when modelling rare species (Breiner et al., 2016), and has been advocated as a better alternative to single models for future climate projection modelling with large numbers of species (Araújo and New, 2007). Nevertheless there is still uncertainty in the performance of ensemble modelling; model performance is heavily influenced by the choice of the initial SDM models used for averaging (Araújo et al., 2005; Diniz-Filho et al., 2009). Furthermore, if ensemble modelling performs only a little or no better than the best-performing model, there is a strong case for choosing the single model; computation time, model complexity and unnecessary variation introduced from weaker performing models can all be reduced.

The aim of this study is to assess the comparative performance of ensemble forecasting and other common SDM techniques, particularly *MaxEnt*, for investigating the distribution of Egyptian medicinal plant species under four different future climate change scenarios. Developing countries such as Egypt often have a large number of medicinal plant species, both traditionally cultivated and wild, that are still relied upon today to treat a wide variety of diseases (Batanouny et al., 1999; Mahmoud, 2013). As with many other plant species, medicinal plants are expected to undergo life-cycle changes and distribution shifts associated with climate change (Khanum et al., 2013; Yi et al., 2016; Zhang et al., 2018; Zhao et al., 2018), but their global ecological and commercial interest and value increases the urgency with which we need to understand their population changes to mitigate potential losses (Das et al., 2016; Gairola et al., 2010). Egypt has relatively well documented medicinal plants with a broad spatial coverage collated by the Biodiversity Monitoring and Assessment Project (BioMap), which provides an excellent opportunity to explore the influence of climate change on future medicinal plant distributions, with high potential value for conservation and land management. Additionally it allows evaluation of the suitability of ensemble forecasting and other SDM methods when using a S-SDM approach to generate predictive maps of medicinal plant species richness across Egypt.

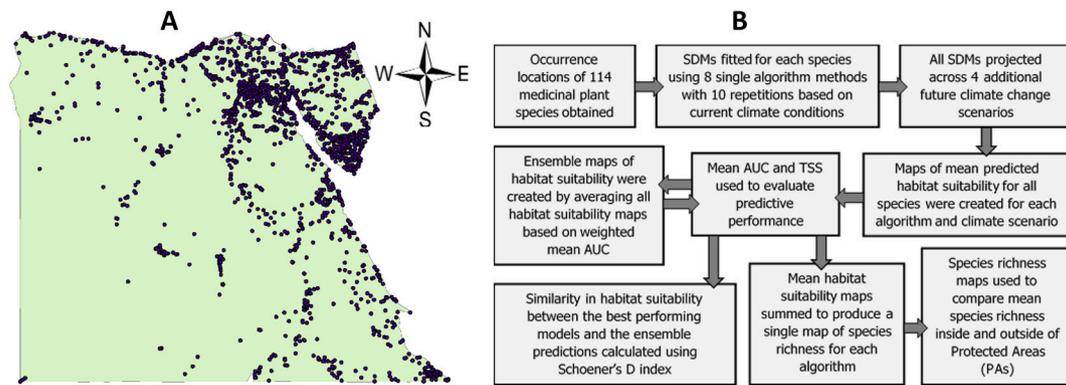
The main objectives of this study are to a) compare the performance of ensemble and single-algorithm SDM methods; b) produce accurate maps of the current distribution of medicinal plants in Egypt; and c) evaluate the impact of various projected climate-change scenarios on these distributions, with a particular emphasis on Egypt's Protected Areas (PAs) for conservation purposes.

## 2. Methods

### 2.1. Study area and occurrence localities

Egypt occupies the north-eastern region of the African continent, with a surface area in excess of one million square kilometres (in reality: 1,019,499 km<sup>2</sup>) (Hoath, 2003), or approximately 3% of the entire area of the African continent (Baha-El-Din, 2006). It lies between latitudes 32° to 22° N and longitudes 24° to 37° E, and is roughly a square with each side of length 1000 km. To a great extent it is continuous with the hyper-arid areas of the Saharan desert, with a hot and almost rainless climate: the temperature can vary between -4 and 53 °C. The Sahara has low relative humidity and is the most extensive region in the world with in excess of 10 h of daylight for every day of the year (Baha-El-Din, 2006). Normal average yearly precipitation across Egypt is less than 80 mm, ranging from 20 mm in the south to a maximum of 200 mm at the Mediterranean coast; there is basically no rain during the summer (El-Nahrawy, 2011).

The data consist of location records of the medicinal plants of Egypt, extracted from the Biodiversity Monitoring and Assessment Project (BioMap) databases. These databases were compiled over four years



**Fig. 1.** A) Locations of occurrence records of medicinal plants Obtained from the Biodiversity Monitoring and Assessment Project (BioMap) databases (2004–2008) used in this study. 14,396 occurrence records across 114 species are shown, having first excluded records from species with fewer than 10 occurrences. B) Flow chart of methodology used in this study in order to create and compare SDM predictions of habitat suitability and species richness of 114 medicinal plant species in Egypt.

(2004–2008) from systematic surveys, museum collections, private records, expedition reports and databases, and from the literature, with the aim of collating all records of Egyptian fauna and flora to make them available for analysis and research. BioMAP was funded by Italian Debt Swap and formed part of the Nature Conservation Sector of the Egyptian Environment Affairs Agency in Cairo. There are 124 medicinal plant species in this database, with a total of 15,299 records, after filtering and deleting all the records that have erroneous georeferencing, are located in ambiguous locations, and duplicated points. We selected 114 species with 14,396 occurrence points (Fig. 1A), after discarding all species with fewer than 10 records because these would result in models with low accuracy (Baldwin, 2009; van Proosdij et al., 2015). Plant nomenclature follows Boulos (1999–2005).

## 2.2. Current and future environmental data

Environmental predictors consisted of bioclimatic variables interpolated from climate data between 1950 and 2000, obtained from the Worldclim dataset (Hijmans et al., 2005; <http://www.worldclim.org>). We selected 19 initial environmental variables considered to indicate current climate circumstances (Table 1). We chose a resolution of 2.5 arc-minutes (~5 km) for the predictors for two reasons: firstly because this matched the level of accuracy of museum data; and secondly because the climate data were interpolated from a limited number of weather stations in Egypt, all concentrated around the Nile Valley and Delta - this spatial bias in the raw data results in interpolated estimates that have larger-than-usual errors. A further reason for choosing a larger resolution is that such macro-scale models are thought to maximise the impact of climate on species distributions, the underlying assumption of the technique (Pearson and Dawson, 2003). The lack of climate stations in the study region, and their highly biased distribution, are reasons to take extra care over the credibility and accuracy of the results (Martínez-Meyer, 2005). Although climate data are most commonly used in SDM, we acknowledge that our models lack alternative biotic influences such as population dynamics, species demographics and interactions that could improve the models (Urban et al., 2016). Such information is usually not available across large-scale macroecological models, and thus climate data are the best possible predictors for the distribution modelling of multiple widespread species.

Elevation data were also obtained from the SRTM Digital Elevation Database version 4.1 [available at: <http://www.cgiar-csi.org/data/elevation>]. Relevant tiles were downloaded, united together, and clipped to the borders of Egypt at resolution of 2.5 arc-minutes (for more details, see El-Gabbas et al., 2016). There is collinearity among the environmental variables, which can be a problem in any modelling method, including species distribution modelling (Guisan et al., 2002).

To reduce multicollinearity among our predictors, the Variance Inflation Factor (VIF) (Marquardt, 1970) implemented in the 'usdm' package (Naimi and Araújo, 2016; R Development Core Team, 2014) was used to exclude predictors with VIF values greater than 10. This reduced our number of environmental predictors from 20 to 8 (Table 1).

Current distribution models for each species were projected into the future time slice of 2050, using different scenarios from both Climate Model Intercomparison Project CMIP3 and CMIP5. Future predicted climate data for CMIP3 and CMIP5 were obtained from the Intergovernmental Panel on Climate Change's (IPCC) 4th and 5th assessment data (for more detail see <http://cmip-pcmdi.lnl.gov/>) taken from the International Centre for Tropical Agriculture website (see <http://www.ccafs-climate.org/>). We used data from the Global Circulation Model (GCM) generated by the UK Hadley Centre for Climate Prediction and Research (HadCM3) for two scenarios (A2a and B2a) for CMIP3. The A2a and B2a scenarios (for more details see: the IPCC special reports) used in this study have been regularly used in climate-change assessments (Hannah, 2011). Both scenarios have different assumptions about the amount of CO<sub>2</sub> emissions. The A2a 'business as usual' scenario expects that the level of CO<sub>2</sub> emissions increases without restriction because of the high growth rate in the human population, not much technological development, expanded land-use changes, and people being less environmentally aware. The B2a 'moderate mitigation' scenario expects that the level of CO<sub>2</sub> emission will not change much more than now, because human population growth will be slower, with fewer changes in land-use, people are more environmentally conscious, and there is increasing invention in technology (Saupe et al., 2011). Similarly, for CMIP5 we used two 'representative concentration pathways' (RCP 2.6 and RCP 8.5) generated by the UK Hadley Centre for Climate Prediction and Research (Hadgem2.es). RCP 2.6 characterises an optimistic prediction representing a medium level of population growth, and very low greenhouse-gas concentrations; while RCP 8.5 represents a pessimistic prediction characterized by high population growth and high levels of greenhouse-gas concentrations by the end of 2100 (Wayne, 2013).

## 2.3. Species distribution modelling

We fitted SDMs for each species using eight modelling algorithms implemented in the 'sdm' package in R (Naimi and Araújo, 2016). These were *maxlike* (Royle et al., 2012); generalised linear models (*glm*) (McCullagh and Nelder, 1989); classification and regression trees (*cart*) (Breiman et al., 1984); support vector machine (*svm*) (Vapnik, 1995); boosted regression trees (*brt*) (Friedman, 2001); maximum entropy (*MaxEnt*) (Phillips et al., 2006); mixture discriminant analysis (*mad*, or *fda*) (Hastie et al., 1994; Thuiller, 2003); and random forests (*rf*) (Breiman, 2001). The default implementations in the *sdm* package were

**Table 1**

Initial (all) and final (non-greyed) set of environmental variables used to build the models. Highlighted variables were rejected to reduce collinearity after applying the Variance Inflation Factor (VIF) with a cut-off threshold of 10.

Variable	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter
Altitude	Altitude

used (Naimi and Araújo, 2016). The data were presence-only, and pseudo-absence data were generated using default sdm package (see Naimi and Araújo, 2016), by randomly sampling 10,000 locations for each species across the study area. This approach is standard for presence-pseudoabsence modelling (Elith et al., 2011), and the large number of 10,000 points has been shown to result in high predictive accuracy (Phillips and Dudík, 2008) see (Fig. 1B).

Models were evaluated using K-fold cross-validation with 10 folds and 10 replications for each algorithm; for each replicate the data are divided randomly into 10 folds, one of which is used to evaluate the model calibrated using the other 9 folds (Peterson et al., 2011), so as to give more precise projections (Elith et al., 2011). Model performance was evaluated using two methods, a threshold-independent statistic - the area under the curve (AUC) (Fielding and Bell, 1997), and a threshold-dependent statistic - the true skills statistic (TSS) (Allouche et al., 2006). AUC varies between 0 and + 1: an AUC score between 1.0 and 0.9 = excellent, between 0.9 and 0.8 = good, between 0.8 and 0.7 = fair, between 0.7 and 0.6 = poor, and between 0.6 and 0.5 = fail (Swets, 1988). TSS scores vary between + 1 and - 1, with a score close to 1 indicating an almost perfect model, while close to zero or less than zero indicates a model no better than random (Allouche et al., 2006).

Mean AUC and TSS across the 10 replicates of each algorithm across all species were used to assess model performance. Additionally, the distribution maps for each individual species were summed to produce a single map of species richness per method (Distler et al., 2015) using ArcGIS 10.2.2 (see SI Fig. 1B).

We compared the current and the 2050 future predicted habitat suitability under both CMIP3 (A2 and B2) and CMIP5 (RCP 2.6 and RCP 8.5) to show the areas important for conservation planning under all scenarios. Each SDM fitted using the current climate data was projected into the future climate scenarios. The mean habitat suitability was calculated for each current and 2050 climate-model scenario across all pixels, taken to indicate how suitable the average cell is rather than any change in geographic range (e.g. habitat loss) (for more detail, see Wright et al., 2016). To calculate the impact of the climate scenarios on predicted habitat suitability, we measured the percent change in mean habitat suitability  $[(\text{future} - \text{current}) / \text{current}] * 100$  (Wright et al., 2016). We compared the eight modelling techniques with each other, and the current with future scenarios, using Anova for each species. Thus each species has 320 models (8 modelling techniques  $\times$  10 replications  $\times$  4 emission scenarios). For a visual summary of the methodology used, see Fig. 1B.

Finally, the *sdm* package was used to combine the distribution maps using the “ensemble” function to produce consensus ‘ensemble’ maps based on weighted AUC values (Naimi and Araújo, 2016). The mean AUC and TSS for the ensemble maps were then calculated to assess their predictive performance. The selected threshold for TSS was the one that maximised both the True Positive Rate (TPR) and True Negative Rate (TNR). ENMTools (Warren et al., 2010) was used to calculate the similarity of the predicted habitat suitability between the best performing models (*MaxEnt* and *random forest*) and the *ensemble* predictions for current and all future climate scenarios. We used the similarity measures index introduced by Warren et al. (2008), namely Schoener's D index (Schoener, 1968: for more details, see Warren et al., 2008; Warren et al., 2010). This index ranges from 0 (no similarity) to 1 (complete similarity).

The mean species richness inside and outside each Protected Area (PA) were calculated just for the *ensemble* technique and for *MaxEnt*, to compare spatial conservation priority based on these two methods. Egypt has 30 PAs established since 1983, covering about 15% of the land area (El-Gabbas et al., 2016). We created a 50-km buffer around each PA, and calculated the mean predicted species richness across all pixels inside each PA (‘inside’) and in the buffer (‘outside’). Where PAs adjoin one another, we created one buffer around both to avoid any conflict in calculating species richness. The paired difference inside-outside was calculated for each PA, and the difference became the response variable of a GLM with normal errors. Ideally, species richness could be also compared inside and outside of Egypt in order to provide a more robust picture of the distribution of medicinal plants in similar regions. However, the lack of data for these species in surrounding countries means only models for Egypt can be fitted.

### 3. Results

#### 3.1. Model performance and variable importance

Predictive accuracies of all SDM algorithms were generally good across all species in terms of both AUC and TSS, except for one species *Herniaria hirsuta* which obtained a poor AUC and TSS score with *svm* (Fig. 2, Table S1). The mean AUC varied between 0.798 (*cart*) to 0.927 (*rf*), and the mean TSS lay between 0.627 (*cart*) and 0.825 (*rf*) (Table S1). *MaxEnt* and *rf* achieved the highest performance out of all eight algorithms based on mean AUC and TSS values, while *svm* and *cart* had the poorest performances. The total number of species with model performances classified as excellent from their AUC values were 63% and 79% for *MaxEnt* and *rf* respectively (Table S2). There is of course a highly significant correlation between AUC and TSS across methods

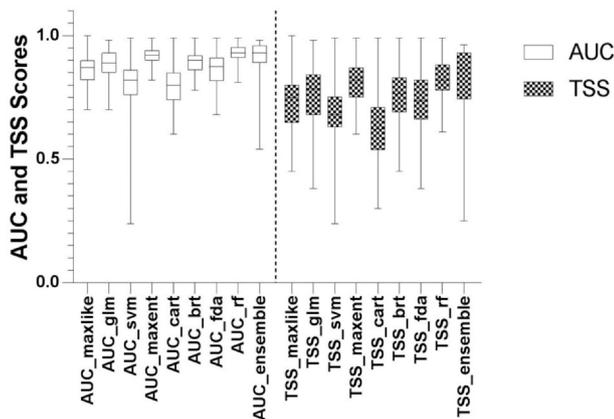


Fig. 2. Evaluation of algorithm performance based on Mean Area Under the Curve (AUC) and True Skill Statistics (TSS) scores for eight SDMs approaches (maxlike, glm, svm, MaxEnt, cart, brt, fda and rf) and for the ensemble maps from all eight algorithms across all species.

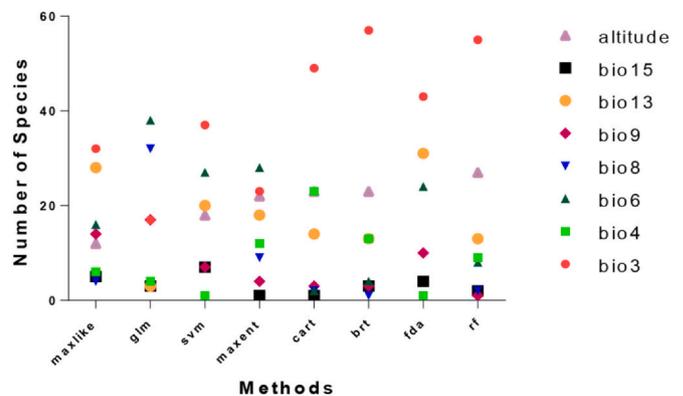


Fig. 3. Number of times each predictor variable (symbols and colours) were the most important variable for each of the separate SDM techniques.

(Fig. S1). The *ensemble* predictions produced the highest mean TSS (0.830) compared to all the single algorithms, and the second highest mean AUC (0.90) after *rf*. Although the *ensemble* mean TSS and AUC are similar to those for *MaxEnt* and *rf*, the interspecific variation for the *ensemble* predictions was much larger than either individual algorithm (Fig. 2).

Variable importance in each model differed across modelling algorithms (Fig. 3). *Bio6* (minimum temperature of the coldest month), *bio3* (isothermality), *bio13* (precipitation of the wettest month), and *altitude* were the most important predictors across all methods and species, but *bio6* was the most important in six of the SDM methods (*maxlike*, *svm*, *cart*, *brt*, *fda*, and *rf*), while *bio3* was the most important in two methods (*MaxEnt* and *glm*).

Fig. 4 shows the change in mean habitat suitability between the current and each of the future scenarios. It is obvious that there is hardly any change for the RCP scenarios, whilst for the CMIP3 scenarios the A2 scenario changes more than B2 (see Fig. 4). There were obvious

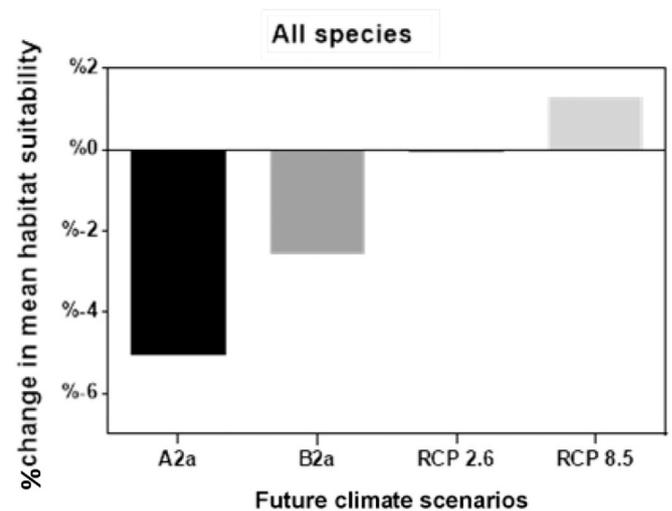
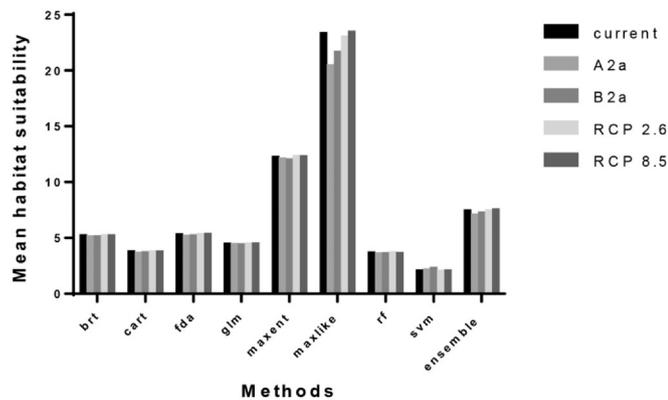


Fig. 4. Percentage change in mean habitat suitability  $[(2050 \text{ future-current}) / \text{current}] * 100$  for the four climate change scenarios (A2a, B2a, RCP 2.6 and RCP 8.5) compared to mean habitat suitability under the current climate across all species using the ensemble modelling approach. The A2a ‘business as usual’ scenario expects that the level of CO2 emissions increases without restriction, whereas the B2a ‘moderate mitigation’ scenario expects that the level of CO2 emission will not change much more than now (Hannah, 2011). RCP 2.6 characterises an optimistic prediction representing a medium level of population growth, and very low greenhouse-gas concentrations; while RCP 8.5 represents a pessimistic prediction characterized by high population growth and high levels of greenhouse-gas concentrations by the end of 2100 (Wayne, 2013).



**Fig. 5.** Mean habitat suitability across all species based on the current climate and four predicted climate change future scenarios (A2a, B2a, RCP 2.6 and RCP 8.5). The A2a ‘business as usual’ scenario expects that the level of CO<sub>2</sub> emissions increases without restriction, whereas the B2a ‘moderate mitigation’ scenario expects that the level of CO<sub>2</sub> emission will not change much more than now (Hannah, 2011). RCP 2.6 characterises an optimistic prediction representing a medium level of population growth, and very low greenhouse-gas concentrations; while RCP 8.5 represents a pessimistic prediction characterized by high population growth and high levels of greenhouse-gas concentrations by the end of 2100 (Wayne, 2013). Results are shown for each of the eight modelling methods (brt, cart, fda, glm, maxent, maxlike, rf and svm) and for the ensemble modelling method based on all of these methods. There were significant differences between SDM methods ( $F_{8, 32} = 1142, P < 0.0001$ ), but no significant differences between current and future scenarios ( $F_{4, 32} = 1.960, P = 0.1245$ ).

and significant differences between modelling methods (Fig. 5).

The *ensemble* maps (Fig. 6) show slight differences in species richness between the current scenario and both CMIP3 and CMIP5 future scenarios. *Ensemble* models predict in the current and future that all of the Mediterranean Coast, Sinai Peninsula, and various areas of the Red sea coast and eastern desert regions are of interest for conservation planning under future climate change. In contrast, the spatial patterns show differences between SDM methods, and between most SDM methods and the *ensemble* technique (see Figs. S2 to S6). However, these differences are only slight between *MaxEnt* (Fig. 7) and *ensemble* (Fig. 6) under all assumptions (see Figs. S2 to S6). This is confirmed using the quantitative comparison using Schoener’s D index: there is high similarity between *MaxEnt* and *ensemble* predictive maps (Table 2), but this is much lower between *rf* and *ensemble*, and *MaxEnt* and *rf*.

Finally, we compared species richness inside and outside the Egyptian PAs for the *MaxEnt* and *ensemble* methods, using the difference between inside and outside as a response variable in GLM. The results show that there is no difference between the two methods ( $F_{1, 249} = 0.26, P = 0.61$ ). The mean species richness inside is consistently higher than outside, ( $t = 6.86, P < 0.001$ ).

#### 4. Discussion

In response to the need for SDMs (e.g. species conservation, conservation planning for climate change) many software programs and approaches have been developed (Elith et al., 2006; Norberg et al., 2019; Thuiller, 2003). Combining and averaging models using the *ensemble* approach is thought to reduce model uncertainty and increase its robustness in modelling species distributions accurately (Araújo and New, 2007; Marmion et al., 2009; Thuiller, 2003). Nevertheless, our results show that *MaxEnt* was independently able to perform and predict comparatively well against an *ensemble* approach that combined many well-used, highly regarded algorithms to identify important areas for the conservation of Egyptian medicinal plants. Such findings do not necessarily imply that *MaxEnt* is a better technique than other approaches, and there are still cases where it is less appropriate (see

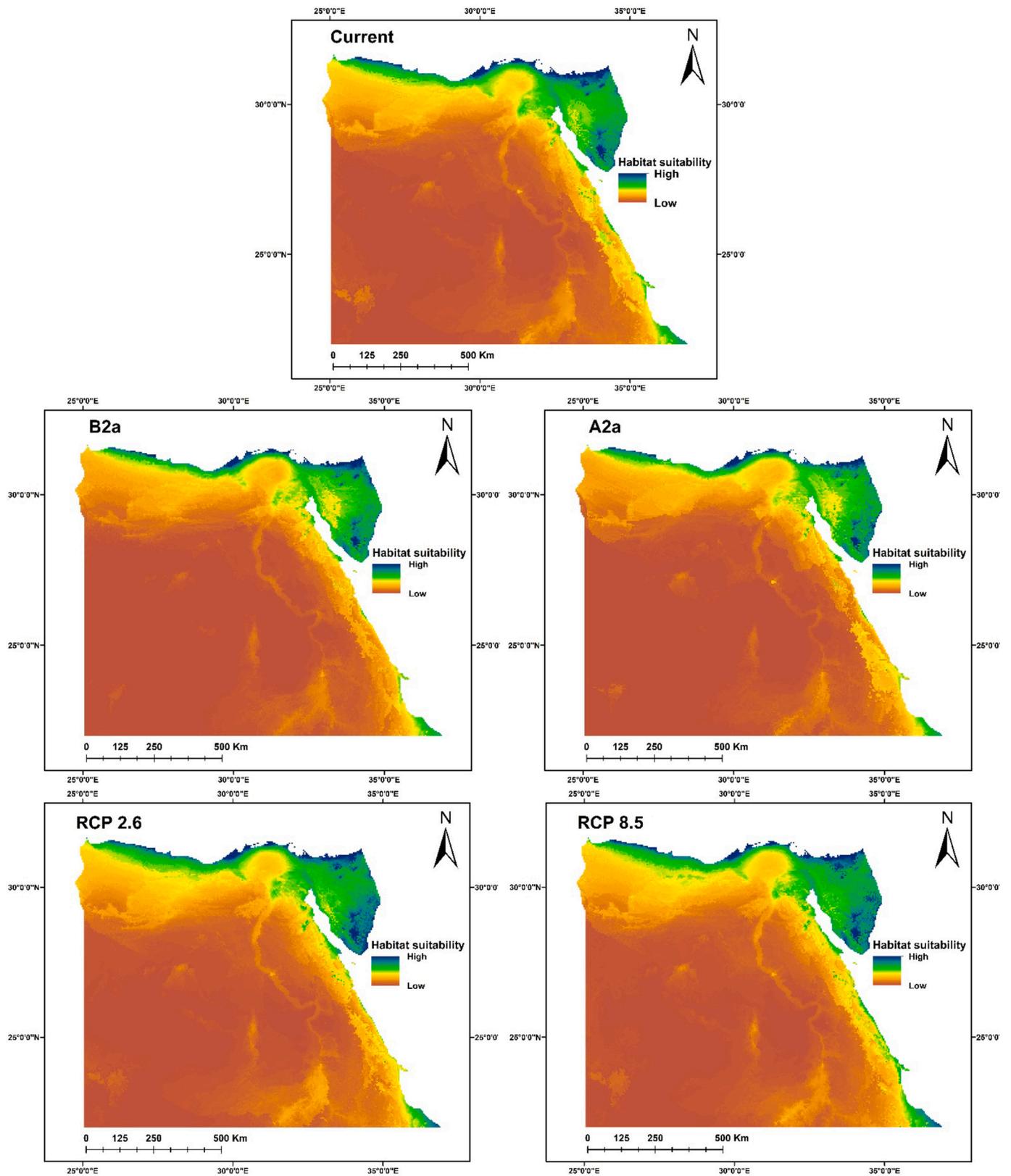
Guillera-Arroita et al., 2014). However, we propose that *MaxEnt* should still be considered as one of the most reliable and accessible techniques when modelling species distributions from incomplete data (Abdelaala et al., 2019; Fois et al., 2018; Kaky and Gilbert, 2019b). We believe that it can be applied easily to help with the issue of identifying important conservation areas, especially in developing countries where conservation efforts are less extensive.

When modelling presence-only (PO) data, *MaxEnt* is a sensible first option to consider. In this case study, *MaxEnt* generated similar habitat suitability predictions ( $> 0.90$ ) under current conditions and different climate scenarios for the Egyptian landscape in comparison with the *ensemble* approach (Table 2). This relative accuracy in comparison with most of the other techniques within the *ensemble* approach is well documented in other studies across different taxa, including terrestrial and marine species (see Elith et al., 2006; Graham et al., 2008; Monk et al., 2010; Reiss et al., 2011; Wisz et al., 2008). Thibaud et al. (2014) suggested that *MaxEnt* outperformed traditional presence-absence approaches, but Guillera-Arroita et al. (2014) demonstrated that presence-absence approaches such as GLM can predict well using small sample sizes if properly analysed, emphasizing that presence-absence is a more reliable approach because it uses evidence of species absence rather than random background points. However, this does not change the reality that *MaxEnt* is capable of generating spatial and temporal predictions of habitat suitability that are very informative, and similar to those generated by *ensemble* forecasting under a variety of climate scenarios.

It is strongly recommended to use presence-absence data when available: they are known to perform better than presence-only data, which normally contain some limitations which can limit model performance (Elith et al., 2011; Zaniewski et al., 2002). However, based on the results presented here, *MaxEnt* does not appear to be greatly influenced by these limitations (also see Baldwin, 2009). Here, we briefly discuss some reasons in relation to our dataset that influenced *MaxEnt* in generating robust spatial conservation predictions for Egyptian medicinal plants (for more detailed explanations about *MaxEnt*, see Phillips et al., 2006; Baldwin, 2009; Elith et al., 2011; Merow et al., 2013).

The first key to *MaxEnt*’s success is its regularization process to avoid overfitting, especially when using small sample sizes (Baldwin, 2009; Merow et al., 2013; Phillips et al., 2006). *MaxEnt* is able to extract useful information successfully even from incomplete data, and hence captures non-linear, complex interactions and relationships (Baldwin, 2009; Merow et al., 2013; Phillips et al., 2006). Here, plant species with variable numbers of records were accurately modelled, showing that *MaxEnt* is not sensitive to variation in sample size (detailed performance is discussed in Kaky and Gilbert, 2016). Secondly, *MaxEnt* has been shown to be relatively insensitive to moderate sampling bias (Baldwin, 2009; Phillips et al., 2006). In our study, despite using all the available resources to collate species records, there are signs of spatial bias (Fig. 1). Graham et al. (2008) found that *MaxEnt* was one of the techniques not strongly influenced by spatial errors in sampling. In this study, more than half of the plant species were classified as being modelled with ‘excellent’ accuracy (see also Kaky and Gilbert, 2016, 2017, 2019a). Correction for sampling bias is not straightforward (El-Gabbas and Dormann, 2017).

Correlative SDMs are sensitive to the chosen modelling techniques (Araújo and New, 2007), hence variation among models is expected due to their different assumptions and algorithms (El-Gabbas et al., 2016; Marmion et al., 2009). Some SDM techniques can be described as “data-hungry” in order to capture complex interactions and responses (Wisz et al., 2008), yet they can perform very well if properly handled and analysed (Guillera-Arroita et al., 2014), an advantage over *MaxEnt* since they use presence-absence data. In this study, the final S-SDM species richness maps of *Maxlike* and *svm* across current and different climate changes scenarios show signs of over-prediction and under-prediction, respectively. Random forests (*rf*) had the highest mean



**Fig. 6.** Habitat suitability maps created using the ensemble modelling method for all Egyptian medicinal plant species (using stacked SDMs) under the current climate and four predicted climate change future scenarios (A2a, B2a, RCP 2.6 and RCP 8.5). The A2a ‘business as usual’ scenario expects that the level of CO<sub>2</sub> emissions increases without restriction, whereas the B2a ‘moderate mitigation’ scenario expects that the level of CO<sub>2</sub> emission will not change much more than now (Hannah, 2011). RCP 2.6 characterises an optimistic prediction representing a medium level of population growth, and very low greenhouse-gas concentrations; while RCP 8.5 represents a pessimistic prediction characterized by high population growth and high levels of greenhouse-gas concentrations by the end of 2100 (Wayne, 2013).

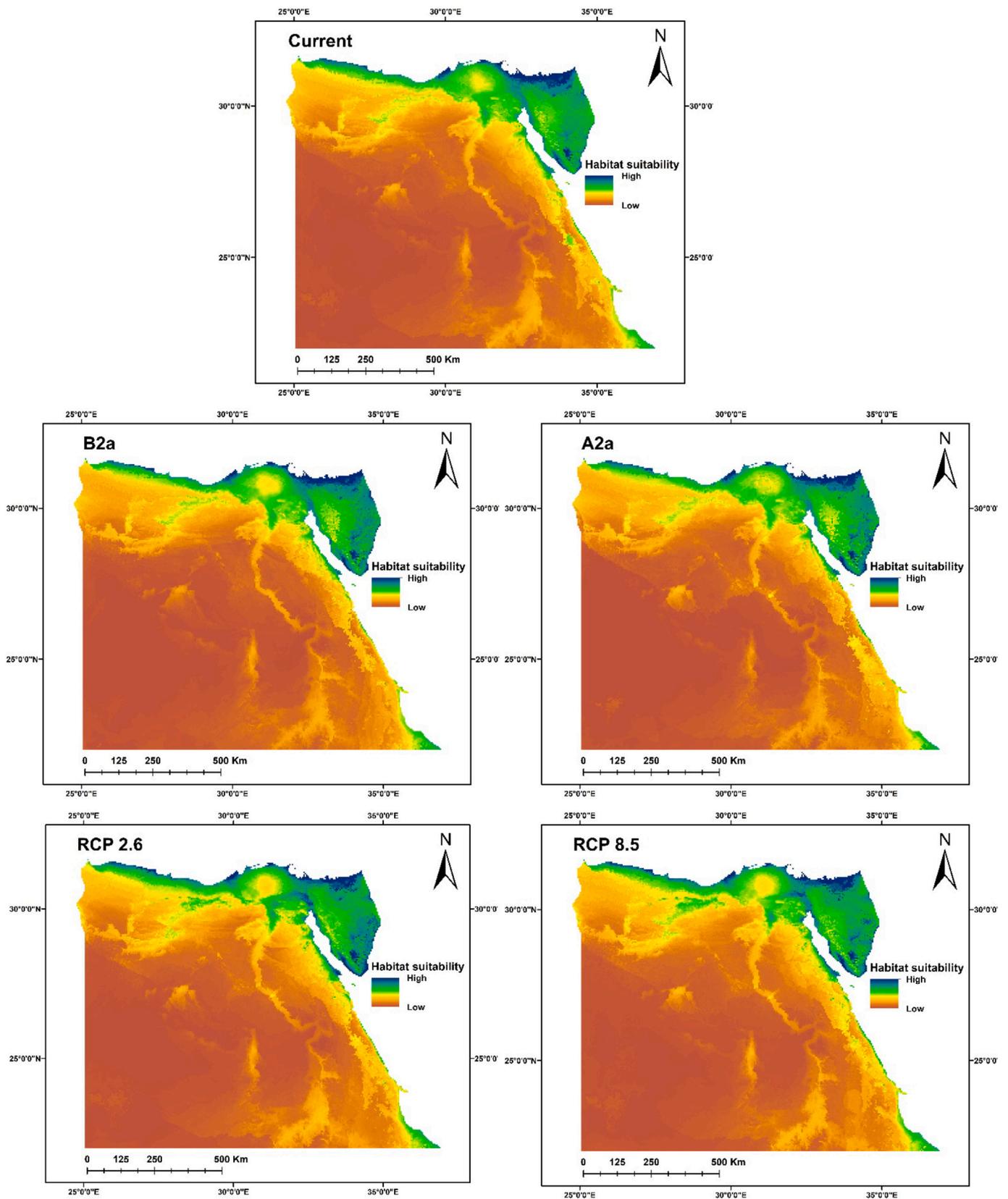


Fig. 7. Habitat suitability maps created using MaxEnt for all Egyptian medicinal plant species (using stacked SDMs) under the current climate and four predicted climate change future scenarios (A2a, B2a, RCP 2.6 and RCP 8.5). The A2a ‘business as usual’ scenario expects that the level of CO2 emissions increases without restriction, whereas the B2a ‘moderate mitigation’ scenario expects that the level of CO2 emission will not change much more than now (Hannah, 2011). RCP 2.6 characterises an optimistic prediction representing a medium level of population growth, and very low greenhouse-gas concentrations; while RCP 8.5 represents a pessimistic prediction characterized by high population growth and high levels of greenhouse-gas concentrations by the end of 2100 (Wayne, 2013).

**Table 2**

Habitat suitability similarity (Schoener's D index) between MaxEnt, Random Forest (rf) and ensemble modelling approaches, calculated using ENMTools (Warren et al., 2010).

		MaxEnt				
		Current	A2_2050	b2_2050	rcp2.6_2050	rcp8.5_2050
Ensemble	Current	0.921542	0.909275	0.911555	0.90275	0.904857
	A2_2050	0.905567	0.919812	0.909464	0.895025	0.89557
	b2_2050	0.906476	0.907022	0.925331	0.900876	0.901757
	rcp2.6_2050	0.917222	0.908953	0.91489	0.92312	0.916606
	rcp8.5_2050	0.907904	0.902441	0.907234	0.904298	0.914236
		rf				
		Current	A2_2050	b2_2050	rcp2.6_2050	rcp8.5_2050
Ensemble	Current	0.741582	0.735335	0.727987	0.745182	0.744010
	A2_2050	0.739791	0.748183	0.735602	0.748802	0.745199
	b2_2050	0.740507	0.741395	0.739552	0.749837	0.746721
	rcp2.6_2050	0.725258	0.725871	0.718657	0.743619	0.735898
	rcp8.5_2050	0.727335	0.726626	0.719102	0.739683	0.741501
		rf				
		Current	A2_2050	b2_2050	rcp2.6_2050	rcp8.5_2050
MaxEnt	Current	0.738863	0.731071	0.727251	0.743195	0.741149
	A2_2050	0.737395	0.740029	0.731214	0.744589	0.741611
	b2_2050	0.738956	0.736786	0.735257	0.746950	0.744149
	rcp2.6_2050	0.724908	0.722089	0.718025	0.740719	0.734573
	rcp8.5_2050	0.725485	0.721590	0.718119	0.738081	0.737919

accuracy across all techniques based on AUC and TSS (Fig. 2), with 79% of species having excellent accuracy (Fig. S2). However, *rf* showed lower similarity to the *ensemble* result than *MaxEnt* (Table 2).

Modelling species richness by stacking the habitat suitability maps for individual species is an important conservation tool (Benito et al., 2013; Distler et al., 2015), despite its tendency to overestimate species richness (Algar et al., 2009). Distler et al. (2015) found that an S-SDM approach produces similar richness patterns to macroecological modelling. Using *MaxEnt* to generate individual SDMs and then stacking them is an approach already used in the region and its surroundings (see Alatawi et al., 2020; El-Gabbas et al., 2016; Kaky and Gilbert, 2016, 2017, 2019a, 2019b). Newbold et al. (2010) and Alatawi et al. (unpubl. data) used *MaxEnt* to produce final S-SDM species richness maps for reptiles to test model accuracy in Egypt and Saudi Arabia, respectively: both studies showed good accuracy. In developing countries where less advanced statistical modelling techniques are being used, *MaxEnt* presents itself as an easily applied and interpreted piece of software, that uses a user-friendly interface, yet retaining confidence in the accuracy of its predictions.

There is always uncertainty about choosing the best methods to model species distributions (Elith and Graham, 2009). As a result, the *ensemble* method was introduced as a better approach over a single-technique model, because it increased the reliability of predictions (Araújo and New, 2007; Marmion et al., 2009; Thuiller, 2003). In reality, the *ensemble* approach has not caught on. Hao et al. (2019) reviewed peer-reviewed papers between 2003 and 2016 that had applied the *ensemble* method implemented in *BIOMOD*, and found only 224 eligible papers, the majority of which were concentrated in Europe and North America. In contrast, it is easy to find many thousands of *MaxEnt* applications. We conclude that the utility and ease of use of *MaxEnt* is the main reason that SDMs became an active research tool for a variety of ecological and biogeographical conservation applications.

Besides using presence-only data for this study, we used only bioclimatic variables as predictors, and it is certainly possible that the distributions of the species are also influenced by other factors (biotic factors, evolution, dispersal ability, etc.). Not incorporating some of

these factors may have limited the accuracy of our models and hence possibly the conclusions (Urban et al., 2016). However, the relevant ecological information about Egyptian medicinal plants does not exist, as with most of the world's species. Despite such unavoidable uncertainties with model predictions, SDMs represent a useful macroecological tool for exploring the dynamics of the relationship between distributions and climate conditions (Pearson and Dawson, 2003; Vasconcelos et al., 2012).

## 5. Conclusion

One last important note to point out is that we are not implying that the *ensemble* approach is unreliable, and *MaxEnt* is a better alternative. In our study the *ensemble* approach achieved a high mean accuracy (AUC = 0.90, TSS = 0.83: Table S1). Our main aim was to demonstrate that *MaxEnt* produces similar results, even for the various climate scenarios (all scenarios had a similarity index > 0.90). Due to the fact that i) not everyone can use advanced statistical modelling techniques, especially in developing countries where it is not a common practice, and ii) the majority of data from such localities are in presence-only format, we believe that in these circumstances *MaxEnt* is a better choice over complicated, computationally intensive 'black-box' *ensemble* models. *MaxEnt* can promote the cause of practical conservation much more effectively.

## Author contributions

Professor Francis Gilbert ran the Biomap project in Cairo together with his Egyptian colleague Professor Samy Zalut. On his return to Nottingham, his research group has used the collated data in a variety of ways to help Egypt's conservation efforts, and to explore how such countries with a paucity of data can maximise the potential of the data in conservation planning.

Author contributions: Gilbert designed the project; Kaky and Nolan analysed the data, Kaky, Nolan and Alatawi wrote the first draft; Gilbert and Kaky completed the manuscript.

## Declaration of Competing Interest

None.

## Acknowledgements

We would like to thank the BioMAP project for sharing the data. Emad Kaky was supported by the Research Centre at Sulaimani Polytechnic University, and Victoria Nolan was supported by the Woodland Trust (UK) and the University of Nottingham.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2020.101150>.

## References

- Abdelaala, M., Foisa, M., Fenua, G., Bacchetta, G., 2019. Using MaxEnt modeling to predict the potential distribution of the endemic plant *Rosa arabica* Crép. in Egypt. *Ecol. Inform.* 50, 68–75. <https://doi.org/10.1016/j.ecoinf.2019.01.003>.
- Alatawi, A.S., Gilbert, F., Reader, T., 2020. Modelling terrestrial reptile species richness, distributions and habitat suitability in Saudi Arabia. *J. Arid Environ.* 178, 104153. <https://doi.org/10.1016/j.jaridenv.2020.104153>.
- Algar, A.C., Kharouba, H.M., Young, E.R., Kerr, J.T., 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography* 32, 22–33. <https://doi.org/10.1111/j.1600-0587.2009.05832.x>.
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>.
- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>.
- Araújo, M.B., Whittaker, R.J., Ladle, R.J., Erhard, M., 2005. Reducing uncertainty in projections of extinction risk from climate change. *Glob. Ecol. Biogeogr.* 14, 529–538.
- Austin, M.P., Niel, K.P.V., 2011. Improving species distribution models for climate change studies: variable selection and scale. *J. Biogeogr.* 38, 1–8. <https://doi.org/10.1111/j.1365-2699.2010.02416.x>.
- Baha-El-Din, S., 2006. A guide to the reptiles and amphibians of Egypt. In: *The American University in Cairo Press, Cairo – New York*.
- Baldwin, R.A., 2009. Use of maximum entropy modeling in wildlife research. *Entropy* 11, 854–866. <https://doi.org/10.3390/e11040854>.
- Baselga, A., Araújo, M.B., 2009. Individualistic vs community modelling of species distributions under climate change. *Ecography* 32, 55–65. <https://doi.org/10.1111/j.1600-0587.2009.05856.x>.
- Batanouny, K.H., Aboutabl, E., Shabana, M., Soliman, F., 1999. Wild medicinal plants in Egypt: an inventory to support conservation and sustainable use. In: *Cairo: Academy of Scientific Research and Technology. IUCN, Swiss Development Cooperation, CH*.
- Beaumont, L.J., Graham, E., Duursma, D.E., Wilson, P.D., Cabrelli, A., Baumgartner, J.B., Hallgren, W., Esperón-Rodríguez, M., Nipperess, D.A., Warren, D.L., Laffan, S.W., VanDerWal, J., 2016. Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? *Ecol. Model.* 342, 135–146. <https://doi.org/10.1016/j.ecolmodel.2016.10.004>.
- Benito, B.M., Cayuela, L., Albuquerque, F.S., 2013. The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: guidelines to build better diversity models. *Methods Ecol. Evol.* 4, 327–335. <https://doi.org/10.1111/2041-210x.12022>.
- Boulos, L., 1999–2005. *Flora of Egypt*. 4 Al Hadara Publishing, Cairo, Egypt.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, USA.
- Breiner, F.T., Guisan, A., Bergamini, A., Nobis, M.P., 2016. Overcoming limitations of modelling rare species by using ensembles of small models. *J. Anim. Ecol.* 85, 1210–1218. <https://doi.org/10.1111/2041-210x.12403>.
- Bucklin, D.N., Bastille, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Románach, S.S., Speroterra, C., Watling, J.I., 2015. Comparing species distribution models constructed with different subsets of environmental predictors. *Divers. Distrib.* 21, 23–35. <https://doi.org/10.1111/ddi.12247>.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. *Glob. Chang. Biol.* 16, 1145–1157. <https://doi.org/10.1111/j.1365-2486.2009.02000.x>.
- Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* 23, 99–112. <https://doi.org/10.1111/geb.12102>.
- Das, M., Jain, V., Malhotra, S., 2016. Impact of climate change on medicinal and aromatic plants: review. *Indian J. Agric. Sci.* 86, 1375–1382.
- De'ath, G., 2002. Multivariate regression trees: a New technique for modeling species–environment relationships. *Ecology* 83, 1105–1117. [https://doi.org/10.1890/0012-9658\(2002\)083\[1105:mrtant\]2](https://doi.org/10.1890/0012-9658(2002)083[1105:mrtant]2).
- Diniz-Filho, J.A.F., Bini, L.M., Rangel, T.F., Loyola, R.D., Hof, C., Nogués-Bravo, D., Araújo, M.B., 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* 32, 897–906. <https://doi.org/10.1111/j.1600-0587.2009.06196.x>.
- Distler, T., Schuetz, J.G., Velásquez-Tibatá, J., Langham, G.M., 2015. Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *J. Biogeogr.* 42, 976–988. <https://doi.org/10.1111/jbi.12479>.
- Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30, 609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171>.
- Edwards, T.C., Cutler, D.R., Zimmermann, N.E., Geiser, L., Moisen, G.G., 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecol. Model.* 199, 132–141. <https://doi.org/10.1016/j.ecolmodel.2006.05.016>.
- El-Gabbas, A., Dormann, C.F., 2017. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. *Ecography* 41, 1161–1172. <https://doi.org/10.1111/ecog.03149>.
- El-Gabbas, A., Baha El Din, S., Zalat, S., Gilbert, F., 2016. Conserving Egypt's reptiles under climate change. *J. Arid Environ.* 127, 211–221. <https://doi.org/10.1016/j.jaridenv.2015.12.007>.
- Elith, J., Graham, C., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32, 66–77. <https://doi.org/10.1111/j.1600-0587.2008.05505.x>.
- Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.
- El-Nahrawy, M.A. Country Pasture/Forage Resource Profiles: Egypt. <http://www.fao.org/ag/AGP/AGPC/doc/Counprof/PDF%20files/Egypt> (Accessed 10-04-2016).
- Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J. Appl. Ecol.* 41, 263–274. <https://doi.org/10.1111/j.0021-8901.2004.00881.x>.
- Evans, J., Murphy, M., Holden, Z., Cushman, S., 2011. Modeling species distribution and change using random forest, in: *predictive species and habitat modeling in landscape*. Ecology 139–159.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49. <https://doi.org/10.1017/s0376892997000088>.
- Fitzpatrick, M.C., Gotelli, N.J., Ellison, A.M., 2013. MaxEnt versus MaxLike: empirical comparisons with ant species distributions. *Ecosphere* 4, art55.
- Fois, M., Cuenca-Lombrana, A., Fenu, G., Bacchetta, G., 2018. Using species distribution models at local scale to guide the search of poorly known species: review, methodological issues and future directions. *Ecol. Model.* 385, 124–132. <https://doi.org/10.1016/j.ecolmodel.2018.07.018>.
- Fourcade, Y., Engler, J.O., Rödder, 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, e97122. <https://doi.org/10.1371/journal.pone.0097122>.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- Gairola, S., Sharif, N., Bhatt, A., Kala, C.P., 2010. Influence of climate change on production of secondary chemicals in high altitude medicinal plants: issues needs immediate attention. *J. Med. Plant Res.* 4, 1825–1829.
- González-Irusta, J.M., González-Porto, M., Sarraalde, R., Arrese, B., Almón, B., Martín-Sosa, P., 2015. Comparing species distribution models: a case study of four deep sea urchin species. *Hydrobiologia* 745, 43–57. <https://doi.org/10.1007/s10750-014-2090-3>.
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loisele, B.A., 2008. The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* 45, 239–247. <https://doi.org/10.1111/j.1365-2664.2007.01408.x>.
- Grenouillet, G., Buisson, L., Casajus, N., Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34, 9–17. <https://doi.org/10.1111/j.1600-0587.2010.06152.x>.
- Guillera-Arroita, G., Lahoz-Monfort, J., Elith, J., 2014. MaxEnt is not a presence absence method: a comment on Thibaud et al. *Methods Ecol. Evol.* 5, 1192–1197. <https://doi.org/10.1111/2041-210x.12252>.
- Guisan, A., Rahbek, C., 2011. SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *J. Biogeogr.* 38, 1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>.
- Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* 157, 89–100. [https://doi.org/10.1016/s0304-3800\(02\)00204-1](https://doi.org/10.1016/s0304-3800(02)00204-1).
- Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R., Hastie, T., 2006. Making better biogeographical predictions of species' distributions. *J. Appl. Ecol.* 43, 386–392. <https://doi.org/10.1111/j.1365-2664.2006.01164.x>.
- Halvorsen, R., 2013. A strict maximum likelihood explanation of MaxEnt, and some

- implications for distribution modelling. *Sommerfeltia* 36, 1–132. <https://doi.org/10.2478/v10208-011-0016-2>.
- Hannah, L., 2011. *Climate Change Biology*. Elsevier Ltd.
- Hao, J., Elith, J., Guillera-Arroita, G., Lahoz-Monfort, J., 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Divers. Distrib.* 25, 839–852. <https://doi.org/10.1111/ddi.12892>.
- Hastie, T., Tibshirani, R., 2004. Generalized additive models. In: *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc.. <https://doi.org/10.1002/0471667196.ess0297.pub2>.
- Hastie, T., Tibshirani, R., Buja, A., 1994. Flexible discriminant analysis by optimal scoring. *J. Am. Stat. Assoc.* 89, 1255–1270. <https://doi.org/10.1080/01621459.1994.10476866>.
- Hefley, T.J., Hooten, M.B., 2016. Hierarchical species distribution models. *Curr. Landscape. Ecol. Rep.* 1 (8797), 87–97. <https://doi.org/10.1007/s40823-016-0008-7>.
- Hijmans, R., Elith, J., 2013. Species distribution modeling with R. *Encycl. Biodivers.* 6.
- Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob. Chang. Biol.* 12, 2272–2281. <https://doi.org/10.1111/j.1365-2486.2006.01256.x>.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., et al., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. <https://doi.org/10.1002/joc.1276>.
- Hijmans, R.J., Steven, P., John, L., Jane, E., 2017. *dismo: Species Distribution Modeling*. R Package Version 1.1-4. <https://CRAN.R-project.org/package=dismo>.
- Hoath, R., 2003. *A Field Guide to the Mammals of Egypt*, The American University in Cairo Press, Cairo – New York.
- Jarnevich, C.S., Young, N.E., 2015. Using the MaxEnt program for species distribution modelling to assess invasion risk. pp. 65–81.
- Kaky, E., 2020. Potential habitat suitability of Iraqi amphibians under climate change. *Biodiversitas* 21, 731–742. <https://doi.org/10.13057/biodiv/d210240>.
- Kaky, E., Gilbert, F., 2016. Using species distribution models to assess the importance of Egypt's protected areas for the conservation of medicinal plants. *J. Arid Environ.* 135, 140–146. <https://doi.org/10.1016/j.jaridenv.2016.09.001>.
- Kaky, E., Gilbert, F., 2017. Predicting the distributions of Egypt's medicinal plants and their potential shifts under future climate change. *PLoS One* 12, e0187714. <https://doi.org/10.1371/journal.pone.0187714>.
- Kaky, E., Gilbert, F., 2019a. Allowing for human socioeconomic impacts in the conservation of plants under climate change. *Plant Biosyst.* 154 (3), 295–305. <https://doi.org/10.1080/11263504.2019.1610109>.
- Kaky, E., Gilbert, F., 2019b. Assessment of the extinction risks of medicinal plants in Egypt under climate change by integrating species distribution models and IUCN Red List criteria. *J. Arid Environ.* 170, 103988. <https://doi.org/10.1016/j.jaridenv.2019.05.016>.
- Khanum, R., Mumtaz, A., Kumar, S., 2013. Predicting impacts of climate change on medicinal asclepiads of Pakistan using MaxEnt modeling. *Acta Oecol.* 49, 23–31. <https://doi.org/10.1016/j.actao.2013.02.007>.
- Kharoub, H.M., McCune, J.L., Thuiller, W., Huntley, B., 2013. Do ecological differences between taxonomic groups influence the relationship between species' distributions and climate? A global meta-analysis using species distribution models. *Ecography* 36, 657–664. <https://doi.org/10.1111/j.1600-0587.2012.07683.x>.
- Ko, C.-Y., Schmitz, O.J., Jetz, W., 2016. The limits of direct community modeling approaches for broad-scale predictions of ecological assemblage structure. *Biol. Conserv.* 201, 396–404. <https://doi.org/10.1016/j.biocon.2016.07.026>.
- Latimer, A.M., Wu, S., Gelfand, A.E., Silander, J.A., 2006. Building statistical models to analyze species distributions. *Ecol. Appl.* 16, 33–50.
- Liu, C., White, M., Newell, G., 2011. Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography* 34, 232–243.
- Mahmoud, T., 2013. Traditional knowledge and use of medicinal plants in the Eastern Desert of Egypt: a case study from Wadi El-Gemal National Park. *J. Med. Plants Stud.* 1, 10–17.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15, 59–69. <https://doi.org/10.1111/j.1472-4642.2008.00491.x>.
- Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12, 591–612. <https://doi.org/10.2307/1267205>.
- Martínez-Meyer, E., 2005. Climate change and biodiversity: some considerations in forecasting shifts in species' potential distributions. *Biodivers. Inform.* 2, 42–55. <https://doi.org/10.17161/bi.v2i0.8>.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Chapman and Hall.
- Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M., Vanaverbeke, J., 2011. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecol. Model.* 222, 588–597. <https://doi.org/10.1016/j.ecolmodel.2010.11.016>.
- Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36, 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.
- Monk, J., Ierodiaconou, D., Versace, V., Bellgrove, A., Harvey, E., Rattray, A., Laurenson, L., Quinn, G., 2010. Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. *Mar. Ecol. Prog. Ser.* 420, 157–174. <https://doi.org/10.3354/meps08858>.
- Morales, N.S., Fernández, I.C., Baca-González, V., 2017. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ* 5, e3093. <https://doi.org/10.7717/peerj.3093>.
- Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39, 368–375. <https://doi.org/10.1111/ecog.01881>.
- Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W.M., Zalat, S., Baha El Din, S., Gilbert, F., 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos* 119, 1326–1334.
- Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., et al., 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecol. Monogr.* 89, e01370. <https://doi.org/10.1002/ecm.1370>.
- Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A., Miller, P., Louzao, M., 2012. Comparison of five modeling techniques to predict the spatial distribution and abundance of seabirds. *Biol. Conserv.* 156, 94–104. <https://doi.org/10.1016/j.biocon.2011.11.013>.
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 12, 361–371. <https://doi.org/10.1046/j.1466-822x.2003.00042.x>.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., et al., 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175. <https://doi.org/10.1111/j.2007.0906-7590.05203.x>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Veski, P.A., McCarthy, M.A., 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* 5, 397–406. <https://doi.org/10.1111/2041-210x.12180>.
- Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.B., et al., 2016. Improving the forecast for biodiversity under climate change. *Science* 353. <https://doi.org/10.1126/science.aad8466>.
- van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J., Raes, N., 2015. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 38, 001–011. <https://doi.org/10.1111/ecog.01509>.
- R Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria URL. <http://www.R-project.org/>.
- Reiss, H., Cunze, S., König, K., Neumann, H., Kröncke, I., 2011. Species distribution modelling of marine benthos: a North Sea case study. *Mar. Ecol. Prog. Ser.* 442, 71–86. <https://doi.org/10.3354/meps09391>.
- Royle, J.A., Chandler, R.B., Yackulic, C., Nichols, J.D., 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods Ecol. Evol.* 3, 545–554. <https://doi.org/10.1111/j.2041-210x.2011.00182.x>.
- Saupe, E.E., Papes, M., Selden, P.A., Vetter, R.S., 2011. Tracking a medically important spider: climate change, ecological niche modeling, and the Brown recluse (*Loxosceles reclusa*). *PLoS One* 6, e17731. <https://doi.org/10.1371/journal.pone.0017731>.
- Schoener, T.W., 1968. Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology* 49, 704–726.
- Segurado, P., Araújo, M., 2004. An evaluation of methods for modelling species distributions. *J. Biogeogr.* 31, 1555–1568. <https://doi.org/10.1111/j.1365-2699.2004.01076.x>.
- Stohlgren, T.J., Ma, P., Kumar, S., Rocca, M., Morissette, J.T., Jarnevich, C.S., Benson, N., 2010. Ensemble habitat mapping of invasive plant species. *Risk Anal.* 30, 224–235. <https://doi.org/10.1111/j.1539-6924.2009.01343.x>.
- Svenning, J.-C., Fløjgaard, C., Marske, K.A., Nogueira-Bravo, D., Normand, S., 2011. Applications of species distribution modeling to paleobiology. *Quat. Sci. Rev.* 30, 2930–2947. <https://doi.org/10.1016/j.quascirev.2011.06.012>.
- Svetnik, V., Liaw, A., Tong, C., Culbertson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958. <https://doi.org/10.1021/ci034160g>.
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293. <https://doi.org/10.1126/science.3287615>.
- Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A., Guisan, A., 2014. Measuring the relative effect of factors affecting species distribution model predictions. *Methods Ecol. Evol.* 5, 947–955. <https://doi.org/10.1111/2041-210x.12203>.
- Thuiller, W., 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Glob. Chang. Biol.* 9, 1353–1362. <https://doi.org/10.1046/j.1365-2486.2003.00666.x>.
- Thuiller, W., Brotons, L., Araújo, M.B., Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27, 165–172. <https://doi.org/10.1111/j.0906-7590.2004.03673.x>.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. Springer, New York.
- Vasconcelos, T.S., Rodríguez, M.A., Hawkins, B.A., 2012. Species distribution modelling as a macroecological tool: a case study using New World amphibians. *Ecography* 35, 539–548. <https://doi.org/10.1111/j.1600-0587.2011.07050.x>.
- Warren, D.L., Richard, E.G., Michael, T., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>.
- Warren, D.L., Glor, R.E., Turelli, M., 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 30, 607–611. <https://doi.org/10.1111/j.1600-0587.2009.06142.x>.
- Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., Hui, F.K.C., 2015. So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766–779. <https://doi.org/10.1016/j.tree.2015.09.007>.
- Wayne, G., 2013. *The beginners's Guide to Representative Concentration Pathways*.

- Skeptical Science.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>.
- Wright, A.N., Schwartz, M.W., Hijmans, R.J., Shaffer, H.B., 2016. Advances in climate models from CMIP3 to CMIP5 do not change predictions of future habitat suitability for California reptiles and amphibians. *Clim. Chang.* 134, 579–591. <https://doi.org/10.1007/s10584-015-1552-6>.
- Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4, 236–243. <https://doi.org/10.1111/2041-210x.12004>.
- Yi, Y., Cheng, X., Yang, Z.-F., Zhang, S.-H., 2016. MaxEnt modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China. *Ecol. Eng.* 92, 260–269. <https://doi.org/10.1016/j.ecoleng.2016.04.010>.
- Zaniewski, A.E., Lehmann, A., Overton, J., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecol. Model.* 157, 261–280. [https://doi.org/10.1016/s0304-3800\(02\)00199-0](https://doi.org/10.1016/s0304-3800(02)00199-0).
- Zhang, M.-G., Zhou, Z.-K., Chen, W.-Y., Slik, J.W.F., Cannon, C.H., Raes, N., 2012. Using species distribution modeling to improve conservation and land use planning of Yunnan, China. *Biol. Conserv.* 153, 257–264. <https://doi.org/10.1016/j.biocon.2012.04.023>.
- Zhang, K., Yao, L., Meng, J., Tao, J., 2018. MaxEnt modeling for predicting the potential geographical distribution of two peony species under climate change. *Sci. Total Environ.* 634, 1326–1334. <https://doi.org/10.1016/j.scitotenv.2018.04.112>.
- Zhao, Q., Li, R., Gao, Y., Yao, Q., Guo, X., Wang, W., 2018. Modeling impacts of climate change on the geographic distribution of medicinal plant *Fritillaria cirrhosa* D. Don. *Plant Biosyst.* 152, 349–355. <https://doi.org/10.1080/11263504.2017.1289273>.
- Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B., Svenning, J.-C., 2010. New trends in species distribution modelling. *Ecography* 33, 985–989. <https://doi.org/10.1111/j.1600-0587.2010.06953.x>.
- Zurada, J., 1992. *Introduction to Artificial Neural Systems*. West Publishing Co., St. Paul, MN, USA.