**SUPPLEMENTARY INFORMATION**

**Appendix 1 – Species Distribution Modelling datasets and parameters**

***Predictor datasets***

*Opuntia ficus-indica and Euphorbia tirucalli*

The choice of environmental and climatic variables selected for species distribution modelling should ideally be based on the known ecology of the species (Title & Bemmels, 2018), as this has previously demonstrated more realistic SDMs (Rodder et al., 2009; Saupe et al., 2012). Following the predictor dataset analysis for *O. ficus-indica* and *E. tirucalli* SDMs in Buckland et al. (2022), four bioclim variables (Table 1) were selected for use in the models. Whilst Title & Bemmels (2018) noted that the inclusion of more complex climatic indices, for example the aridity index, may characterise environmental conditions that are more directly physiologically relevant to species than more simple climatic parameters, previous SDM research by Buckland et al. (2022) has suggested that the inclusion of additional environmental predictor datasets (such as: Aridity index, cloud cover) to do not improve the model performance.

|  |  |
| --- | --- |
| **Bioclim Code / Metric** | **Variable** |
| Bio2 | Mean Diurnal Range (Mean of monthly (max temp – min temp)) |
| Bio6 | Min temp of coldest month |
| Bio12 | Annual precipitation |
| Bio15 | Precipitation seasonality (Coefficient of Variation) |

**Table A1.** Bioclimatic variables used as environmental predictor datasets for *O. ficus-indica* and *E. tirucalli*. See Buckland et al. (2022) for details of covariance analysis and predictor dataset selection.

*Portulacaria afra*

Based on initial SDM analyses using all 19 bioclimatic environmental predictors, a final three were selected for use in the main SDM analysis of the study for projecting the distribution of *P. afra*. Bio4 (temperature seasonality), Bio11 (mean temperature of coldest month) and Bio13 (precipitation of wettest month) were chosen based on their high levels of variable importance, relative to the other environmental predictors, and due to low levels of co-variance. Table A2 shows initial variable importance scores across all bioclimatic variables explored for explaining the relationship between *P. afra* occurrences and current climatic conditions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Bioclim Code / Metric** | **Variable** | **Variable importance score** | **Normalised percentage**  |
| Bio1 | Annual mean temp | 0.0178 | 8% |
| Bio2 | Mean diurnal range | 0.00006 | 0% |
| Bio3 | Isothermality | 0.00128 | 1% |
| Bio4 | Temp seasonality | 0.11692 | 50% |
| Bio5 | Max temp of warmest month | 0.00782 | 3% |
| Bio6 | Min temp of coldest month | 0.01304 | 6% |
| Bio7 | Temp annual range | 0.01532 | 7% |
| Bio8 | Mean temp of wettest quarter | 0.00316 | 1% |
| Bio9 | Mean temp of driest quarter | 0.00248 | 1% |
| Bio10 | Mean temp of warmest quarter | 0.00924 | 4% |
| Bio11 | Mean temp of coldest quarter | 0.01538 | 7% |
| Bio12 | Annual precipitation | 0.00648 | 3% |
| Bio13 | Precip of wettest month | 0.0182 | 8% |
| Bio14 | Precip of driest month | 0.00006 | 0% |
| Bio15 | Precipitation seasonality | 0.0011 | 0% |
| Bio16 | Precip of wettest quarter | 0.00566 | 2% |
| Bio17 | Precip of driest quarter | 0.00028 | 0% |
| Bio18 | Precip of warmest quarter | 0.00024 | 0% |
| Bio19 | Precip of coldest quarter | 0.0008 | 0% |

**Table A2.** Variable importance scores for all 19 bioclimatic variables used in initial P. afra SDM analysis. Final three bioclimatic variables used as environmental predictor datasets for *P. afra* highlighted in green.

***Occurrence data***

Occurrence data was downloaded from the Global Biodiversity Information Facility data (GBIF.org, 2020) and cleaned using the methods described in Zizka (2019). Initially 8,558, 2,589 and 1,606 occurrences were downloaded respectively for *O. ficus-indica*, *E. tirucalli* and *P. afra* on 4February 2020; and this was subsequently cleaned to only those based on ‘human observation’ and with coordinates, removing any ambiguity over specimens held in botanic gardens, herbariums, museums, etc.

Spatial bias of occurrence datasets has the potential of distorting our view on large-scale biodiversity patterns (Ballesteros-Mejia et al., 2013; Beck et al., 2014; Boakes et al., 2010; Varela et al., 2014; Yang et al., 2013). Spatially biased data in this study would have a two-fold impact on distorting SDMs: firstly, through biasing the presence data used to train and evaluate model performance; secondly in biasing the surface range envelope model used in the pseudo-absence dataset generation and therefore model performance metrics. With this in mind, we applied a geographic sampling filter, selecting up to five occurrence data points from each 1° x 1° grid cell – reducing our total occurrences datasets further. A final number of 1,266, 1,014 and 262 occurrences of the three species respectively were used in the SDMS. See Buckland et al. (2022) for further discussion on the sensitivity of SDMs to spatial bias in training datasets.

***Pseudo-Absences***

Pseudo-absence (PA) datasets are widely used in SDM analyses as a replacement for true absence records (Buckland et al., 2022; Chefaoui & Lobo, 2008; Iturbide et al., 2018; Raes & Aguirre-Gutiérrez, 2018; Václavik & Meentemeyer, 2009; Wisz & Guisan, 2009), which often do not exist. PA datasets are generated by sampling background areas from which presence (occurrence) records have not been noted, and are selected through a range of different sampling strategies: random, surface range envelope (SRE), or based on a min or max distance from the occurrence records.

Based on recommendations from the literature (e.g. Barbet-Massin et al., 2012; Iturbide et al., 2018), an equal number of PAs were sampled relative to presences for each species, with multiple PA realisations (five) created to reduce overall uncertainty. PAs were sampled from regions outside of the suitable area estimated by the SRE method (Thuiller et al., 2014). SRE models are based on presence-only data (Barbet-Massin et al., 2012); SRE quantile refers to the quantile used to remove the most extreme values of each environmental variable for determining tolerance boundaries (this study: quantile 0.025 ~ 95% confidence interval) (Hallgren et al., 2019).

***Model fitting***

Whilst SDMs can be trained according to a range of different algorithms, there is often no single ‘best’ choice of model that is selected in place of the other algorithms available. If the same algorithm was applied to the paired dataset multiple times, the model would yield different results between repeats, and the same is true for fitting the models using different algorithms. Defining the ‘best’ performance of each of these individual models and repeats, is further subjective and relative to the evaluation metric that has been selected and focuses on a specific aspect of model performance. Thus, whilst a model may perform well according to one measure, it may not be the ‘best’ model according to another metric. See Buckland et al. (2022) for a discussion on the considerations with model selection and SDM considerations.

In this study, SDMs were initially fitted across two different algorithms that both would be satisfied by the same method of PA dataset generation: Boosted Regression Trees (Elith, 2008) and Random Forests (Breiman, 2001). Default model parameters found in the biomod2 package (Georges and Thuiller, 2013) were used and 10 repeats were completed per algorithm per PA dataset, producing a total of 100 model repeats. The between and within modelling variability shown in SDM outputs has led to the widespread usage of ensemble models (Marmion et al., 2009; Qin et al., 2020; Raes & Aguirre-Gutiérrez, 2018; Senay et al., 2013).

The performance of each individually trained model was assessed based on true skill statistic (TSS) and the relative operating characteristic (ROC), and a weighted mean ensemble was generated. TSS measures the difference between sensitivity and specificity of the model (Allouche et al., 2006; Wunderlich et al., 2019); a TSS score of 0 suggests the model is no better than random, and scores closer to 1 suggest the model is better at discriminating between presence and absence points given a cut-off value generated. Binary cut-off values are calculated as the point at which specificity and sensitivity were maximised. The ROC score is a measure of how well a model is able to discriminate between two alternative outcomes, and is considered a measure of potential usefulness of the model and is graded between 0-1, with 0.5 indicating no skill, and 1 suggesting a perfect model. Both TSS and ROC are commonly used as evaluation metrics within species distribution modelling.

Individual models were combined using two ensemble-model algorithms: weighted mean of probabilities and coefficient of variation of probabilities, to provide a level of uncertainty in the former ensemble model. SDMs were trained based on ‘historical’ (near current) climate data (1970-2000 AD WorldClim 2.1) (Fick & Hijmans, 2017) and projected using ‘historical’ and ‘future’ conditions based on CMIP6 forecasts (see main text). Current occurrence and predictor datasets were split 60% for training and validation, with the remaining 40% used for testing and evaluating model performance. All models were fitted and projected using the biomod2 package version 3.3 (Thuiller et al., 2014) in R Studio version 1.2.5033.

**Appendix 2 – Multidimension response curves**

**Figure A1.** Multidimension response curves for *O. ficus-indica*, *E. tirucalli* and *P. afra*. See main text for commentary on results.

**Appendix 3 – Available land analysis**

|  |  |  |
| --- | --- | --- |
|  | ***Opuntia ficus-indica*** | ***Euphorbia tirucalli*** |
| **SSP 1 (m ha)** | **SSP 2 (m ha)** | **SSP 3 (m ha)** | **SSP 1 (m ha)** | **SSP 2 (m ha)** | **SSP 3 (m ha)** |
| Tanzania | **34.70** | **24.95** | **21.31** | **34.70** | **24.99** | **21.91** |
| Angola | **33.68** | **21.45** | **53.68** | **33.76** | **21.46** | **53.86** |
| Mozambique | **32.46** | 6.91 | 6.70 | 32.48 | 9.83 | 7.60 |
| South Sudan | **31.87** | 2.04 | 1.37 | 31.87 | 3.25 | 2.28 |
| Zambia | **24.29** | **14.18** | **34.46** | **24.32** | **16.14** | **36.36** |
| Nigeria | 23.16 | 1.54 | 0.60 | 23.01 | 3.02 | 3.41 |
| Ethiopia | 19.33 | **15.66** | 14.12 | 19.31 | 19.08 | 21.48 |
| Mali | 7.95 | 0.19 | 0.14 | 7.35 | 2.29 | 1.07 |
| Zimbabwe | 7.23 | 3.78 | 1.15 | 7.24 | 5.03 | 1.92 |
| Malawi | 5.93 | 2.03 | 0.53 | 5.93 | 2.20 | 0.61 |
| Ghana | 5.82 | 0.00 | 0.65 | 5.82 | 0.00 | 0.81 |
| Madagascar | 5.73 | 0.54 | 0.46 | 5.72 | 1.32 | 1.71 |
| South Africa | 5.27 | 0.37 | 0.94 | 4.73 | 0.35 | 0.94 |
| Senegal | 4.90 | 0.00 | 0.00 | 4.90 | 3.79 | 3.32 |
| Kenya | 4.76 | 3.94 | 1.79 | 4.76 | 3.98 | 2.04 |
| Burkina Faso | 3.77 | 0.43 | 0.00 | 3.77 | 0.43 | 0.39 |
| Benin | 3.69 | 0.33 | 0.12 | 3.69 | 0.60 | 1.04 |
| Central African Republic | 3.60 | **21.72** | **21.25** | **3.60** | **23.87** | **36.21** |
| Togo | 3.58 | 0.00 | 0.27 | 3.58 | 0.00 | 0.60 |
| Uganda | 3.52 | 0.90 | 0.46 | 3.52 | 1.00 | 0.54 |
| Guinea | 3.43 | 0.87 | 0.36 | 3.43 | 3.67 | 2.42 |
| DRC | 3.42 | 6.19 | **25.44** | **3.43** | **6.29** | **28.80** |
| Cameroon | 3.17 | 3.64 | 3.11 | 3.13 | 4.85 | 9.74 |
| Chad | 2.47 | 0.65 | 0.16 | 2.47 | 1.04 | 0.43 |
| Sudan | 1.56 | 0.86 | 0.35 | 1.56 | 1.08 | 0.67 |
| Burundi | 1.00 | 0.89 | 0.11 | 0.99 | 0.92 | 0.17 |
| Ivory Coast | 0.90 | 1.25 | 3.19 | 0.90 | 1.25 | 3.32 |
| Gambia | 0.44 | 0.00 | 0.00 | 0.44 | 0.08 | 0.00 |
| Rwanda | 0.29 | 0.29 | 0.05 | 0.29 | 0.29 | 0.12 |
| Guinea-Bissau | 0.28 | 0.00 | 0.00 | 0.28 | 0.27 | 0.49 |
| Lesotho | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Somalia | 0.03 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 |
| Comoros | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 |
| Republic of the Congo | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 | 0.18 |
| Namibia | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| Botswana | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.05 |
| Sierra Leone | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| Niger | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.15 |
| **Total (million ha)** | **282.3** | **135.6** | **193.1** | **281.6** | **162.4** | **245.7** |

**Table A3.** Individual country estimates of suitable and available land for *Opuntia ficus-indica* and *Euphorbia tirucalli* cultivation according to SDM 2081-2100 projections when combined with Daioglou et al. (2019) SSP 1, 2 and 3 available land estimates. Land areas reported in million hectares.

|  |  |
| --- | --- |
|  | ***Portulacaria afra*** |
| **SSP 1 (m ha)** | **SSP 2 (m ha)** | **SSP 3 (m ha)** |
| South Africa | 7.04 | 0.30 | 0.54 |
| Zimbabwe | 1.34 | 0.73 | 0.02 |
| Lesotho | 0.70 | 0.00 | 0.00 |
| Kenya | 0.19 | 0.15 | 0.12 |
| Tanzania | 0.08 | 0.00 | 0.00 |
| Mozambique | 0.03 | 0.00 | 0.00 |
| Angola | 0.00 | 0.00 | 0.00 |
| Ethiopia | 0.00 | 0.00 | 0.01 |
| **Total (million ha)** | **9.38** | **1.18** | **0.68** |

**Table A4.** Individual country estimates of suitable and available land for *Portulacaria afra* cultivation according to SDM 2081-2100 projections when combined with Daioglou et al. (2019) SSP 1, 2 and 3 available land estimates – reported in million hectares.