



Fine-tuning niche models matters in invasion ecology. A lesson from the land planarian *Obama nungara*.

Yoan Fourcade

► To cite this version:

Yoan Fourcade. Fine-tuning niche models matters in invasion ecology. A lesson from the land planarian *Obama nungara*.. Ecological Modelling, 2021, 457, pp.109686. 10.1016/j.ecolmodel.2021.109686 . hal-03471136

HAL Id: hal-03471136

<https://hal.sorbonne-universite.fr/hal-03471136>

Submitted on 8 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fine-tuning niche models matters in invasion ecology. A lesson from the land planarian *Obama nungara*.

Yoan Fourcade

Univ. Paris Est Creteil, Sorbonne Université, CNRS, INRAE, IRD, Université de Paris, Institute of Ecology and Environmental Sciences Paris, 94010 Créteil, France

Correspondence: yoan.fourcade@u-pec.fr

ORCID ID: 0000-0003-3820-946X

Despite the apparent simplicity of species distribution modelling approaches, the reliability of their predictions depends on the application of a number of good practices regarding the input data and the parametrisation of algorithms. In the context of invasion biology, inadequate modelling procedures may lead to erroneous conclusions regarding the potential spread of introduced species. However, clear guidelines for implementing these recommendations are often lacking, confusing or simply unknown by non-modeller end-users. Here, taking as an example the introduced land planarian *Obama nungara*, I fitted MaxEnt models applying six recommended processing steps with respect to sampling bias, predictor choice, training area, evaluation and hyperparameter tuning, separately or implemented together. I compared the resulting outputs to a model fitted with all default settings. All models differed from one another and from the default model, highlighting the importance of considering all these parameters when fitting species distribution models. However, the model that incorporated all fine-tuning methods was by far the most dissimilar, predicting much larger suitable areas globally, including in Africa where *O. nungara* has not been found so far. A closer examination suggested that it is likely a result of lower overfitting. This is a demonstration that modelling settings matter a lot, to the point that fined-tuned or default models may lead to considerably different conclusions when applied to invasive species.

Keywords: Invasion biology, species distribution modelling, MaxEnt, hyperparameter tuning, Platyhelminthes, climate

1. Introduction

The discipline of invasion biology frequently makes use of biogeography methods and theories to predict the potential spread of non-native species and map invasion susceptibility for species of concern (Elith & Leathwick, 2009; Venette et al., 2010). The approaches known as species distribution modelling (SDM) or ecological niche modelling (ENM) correlate the environmental conditions at occurrence records against those where the focal species is known to be absent (in the case of presence-absence data) or where its status is unknown (in the case, more frequent, of presence-pseudoabsence/background modelling). The statistical relationships obtained can be projected in space (and time) to produce spatial predictions of the potential distribution of a species outside its native range (Elith & Leathwick, 2009). Numerous research papers and applied studies have followed this framework in a context of invasion biology, both for invasion risk assessment (Bellard et al., 2016; Thuiller et al., 2005) and to test hypotheses of niche conservatism or shift (Atwater et al., 2018; Liu et al., 2020).

Despite such a widespread use for at least two decades now (Guisan & Zimmermann, 2000), it is remarkable that guidelines for producing accurate models of species niches and distributions are still evolving, and that rigorous evaluations of modelling practices often conclude that different modelling settings offer different advantages in different contexts (Guillera-Arroita et al., 2015; Hao et al., 2020; Morán-Ordóñez et al., 2017). This may be a source of confusion for inexperienced modellers, especially since standard modelling software packages and algorithms offer default settings that produce seemingly satisfactory results (Fourcade et al., 2018). However, notwithstanding these uncertainties, a number of known issues with SDMs can be, at least partly, solved with the application of a set of good practices (Zurell et al., 2020). For example, there are several possible ways to deal with sampling bias in the input occurrences that can avoid drawing erroneous conclusions because of uneven sampling effort (e.g. Boria et al., 2014; Fourcade et al., 2014; Vollerling et al., 2019). Similarly, the extent of background data – which influences spatial predictions and evaluation metrics – can be

adequately chosen by following the appropriate guidelines (Barve et al., 2011). The selection of biologically relevant and uncorrelated environmental variables, as well as the choice of a meaningful evaluation strategy, are also important modelling decisions that must be carefully considered in order to make useful predictions (Dormann et al., 2013; Fourcade et al., 2018).

The above recommendations can be applied to any type of statistical method used to fit SDMs. However, most of modern SDM studies, including in invasion biology, employ machine learning algorithms that depart from the standard, regression-based, statistical approaches most ecologists are trained in (Ryo et al., 2021). As a consequence, there are additional parameters, often overlooked, that can be chosen and that may strongly influence the resulting modelling outputs. Among them are, for tree-based methods, the number of trees to fit, or for maximum entropy (MaxEnt) the features classes and the regularisation multiplier (Merow et al., 2014; Vignali et al., 2020). Even though several authors have advocated for a strict, model-based, tuning of these parameters (Radosavljevic & Anderson, 2014; Warren & Seifert, 2011), there is evidence that this is not always applied in practice (Morales et al., 2017). The question that arises from these observations is thus: how much does it matter to follow these guidelines or not? In the context of invasion biology, this translates into: would predictions of the potential spread of non-native species vary dramatically between default and fine-tuned models?

The present study was inspired by a recent paper published by Negrete et al. (2020), who used a maximum entropy approach to predict the potential distribution of *Obama nungara*, a land planarian native to South America, which has been recently found in Europe where it raised concerns about its possible impact on the native soil fauna (Justine et al., 2020). Despite the excellent occurrence dataset gathered by the authors and the relevance of conducting such a modelling study for this species, the whole approach seemed to be largely based on the raw input data and the default parameters of the MaxEnt software (Phillips et al., 2006). Therefore, one may wonder how much the article's conclusions would be altered if the models were fine-tuned using the latest up-to-date recommendations. In this context, I reanalysed Negrete et al.'s (2020) dataset with a series of optimizations of the modelling method in order to produce predictions of *O. nungara*'s potential distribution that are based on a more state-of-the-art procedure, which I compared to the output of a model produced with all default settings. I also implemented each of the tuning strategies individually to identify which one contributed the most to the divergence between fine-tuned and default

models. Herein, I did not aim to criticize Negrete et al.'s (2020) study, nor did I intend to provide a definitive and perfect distribution model of *O. nungara*. However, my goal was to demonstrate with a recent example that modelling settings matter in fact a lot, to the point that fine-tuned or default models may lead to largely dissimilar conclusions regarding the potential spread of an introduced species.

2. Methods

Most of the methods follow those of Negrete et al. (2020), but at each modelling step two alternative approaches were tested: one that uses default settings such as in Negrete et al. (2020), and one that applies state-of-the-art recommendations to fine-tune the model. This includes manipulating the input data or selecting the best fitting parameter among a series of models. As a result, eight different models were fitted: a 'default settings' model, a 'fully tuned' model that used all the proposed refinements, and six models that differ from the 'default settings' model by one tuned parameter (Table 1).

2.1 Occurrence dataset

Occurrence records for *O. nungara* were collected by Negrete et al. (2020) from various sources, including citizen science programs, literature search and their own observations. The existence of a bias in the data is not known, but opportunistic records frequently suffer from unequal sampling effort where areas close to towns or roads benefit from a more complete survey than remote areas (Geldmann et al., 2016; Ruete, 2015). Several methods have been proposed to improve modelling performance even in the presence of bias, most of which involve manipulating either the background (Vollering et al., 2019) or the species' occurrences (Boria et al., 2014). Here, I used the simplest approach which consists in subsampling the occurrence dataset in a spatial thinning approach to remove records that are too close from each other (Boria et al., 2014). The minimum distance between two points was set to 50 km, because it appeared visually to be a good trade-off between information loss and the removal of dense clusters of records. The subsampling was carried out using the 'spThin' R package (Aiello-Lammens et al., 2015). The full set of occurrences as used in the 'default settings' models includes a total of 144 records, which has been reduced to 93 after spatial thinning.

2.2 Environmental variables

The environmental predictors used to model the potential distribution of *O. nungara* were bioclimatic variables obtained from the Worldclim project at a resolution of 5 arc-min (Hijmans et al., 2005). In their

study, Negrete et al. (2020) selected a subset of the 19 original bioclimatic variables by removing highly correlated variables (Pearson's $r > 0.8$) and variables that contributed little to the first exploratory models. For comparison, I reused for the 'default settings' model the same set of 10 variables: BIO 1, BIO 3, BIO 4, BIO 6, BIO 7, BIO 9, BIO 10, BIO 11, BIO 17 and BIO 19. However, despite the initial effort to limit multicollinearity in the predictors' dataset, the remaining 10 bioclimatic variables still exhibited some high level of correlation (5 of these variables had a variance inflation factor (VIF) > 1000). Therefore, I attempted to reduce even further the risk of overfitting caused by multicollinearity (Dormann et al., 2013), by selecting through a stepwise procedure the set of variables (among the 10 previously described) such that their VIF remained < 4 . In this case, the final set of bioclimatic variables was reduced to four predictors only: BIO 7, BIO 10, BIO 17, and BIO 19. The calculation of variance inflation factors and the selection of variables according to this criterion were performed with the 'usdm' R package (Naimi et al., 2014). The same bioclimatic variables were also obtained for the years 2050 (period 2041-2060) and 2070 (period 2061-2080) under two scenarios of climate change: the representative concentration pathways RCP 2.6 (the most optimistic) and RCP 8.5 (the most pessimistic). These variables were downloaded for three different global circulation models (GCM): CCSM4, GFDL-CM3, and MPI-ESM-LR; the future suitability maps were averaged across these three GCM.

2.3 Background data

When using presence-only data, the environmental conditions at the location of occurrences are compared to those at a set of background points that represent the available environment (Phillips et al., 2006). By default, MaxEnt samples 10,000 background locations from the whole study area, i.e. if global rasters such as those from the Worldclim dataset are supplied, the background will cover the entire extent of the Earth's land surfaces. This was the approach retained for the 'default settings' model; however, this is generally not recommended (Barve et al., 2011). Indeed, in this case, presence points are compared to distant locations that may provide little information with regard to the ecological niche of the species. There is, for example, little value in the observation that a South American species is not present in Greenland, where the climatic conditions are strikingly different and where the species was unlikely to disperse. Ideally, background points must represent locations that the species could have occupied given its dispersal ability, i.e. the accessible area (Barve et al., 2011). This accessible area is not straightforward to estimate for a species that is currently expanding because of human transport.

However, a possible strategy for invasive species that are not at equilibrium with climate is to restrict the background area within a certain distance from occurrence points (Elith et al., 2010). Following this approach, I sampled 10,000 background locations at random in a radius of 500 km around occurrences. The exact distance was arbitrary but ensures that the training area comprises the region around native and introduced ranges (northwest Argentina, Uruguay, southern Brazil on the one hand, and western Europe on the other hand) while not extending further.

When SDMs are used for projecting potential ranges in space and time, it is important to know how much predictions outside the training area can be trusted. Especially, projecting in regions characterised by environmental values outside those experienced in the region of calibration may lead to erroneous extrapolations. Here, in order to quantify and visualise extrapolations, I adopted the extrapolation risk analysis recommended by Owens et al. (2013), in the form of mobility-oriented parity (MOP) maps. This analysis maps the environmental similarity between areas of projection and the calibration area (i.e. background extent). MOP was computed with the ntbox R package (Osorio-Olvera et al., 2020).

2.4 Model settings

Species distribution models for *O. nungara* were fitted using the MaxEnt algorithm, version 3.4.4, one of the most widespread SDM methods (Phillips et al., 2006). Two important parameters control the modelling process: the features produced by MaxEnt, which are transformations of the predictors that can take different functional forms, and the regularization multiplier which is a penalty term implemented to limit model complexity and thus overfitting (Merow et al., 2013). The default settings of MaxEnt, in versions $< 3.4.0$, allow all feature classes to be combined: linear (L), quadratic (Q), product (P), threshold (T) and hinge (H), which is what I used for the 'default settings' model (Negrete et al. (2020) use MaxEnt 3.3.3k). However, better transferability can often be achieved by using simpler response curves (Merow et al., 2014; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011); for this reason, it is recommended to conduct a prior investigation of various combinations of feature classes. Similarly, although MaxEnt uses a regularization multiplier of 1 by default, selecting a different, model-specific, regularization multiplier can provide better performance. Therefore, it is advised to select the combination of feature classes and regularization that gives the best fit while limiting overfitting. Here, I used the ENMeval R package (Muscarella et al., 2014) to test all combinations of the following feature classes: L, LQ, H, LQH, LQHP, LQHPT and regularization multipliers: 0.5, 1.0, 1.5, 2.0, 2.5, 3.0,

3.5, 4.0, 4.5, 5.0. The final model was selected as the one that provided the lowest Akaike Information Criterion (AIC).

For all models I extracted response curves (change in suitability along the range of a predictor when the others are kept at their median value) and variable contribution was assessed using both permutation importance and percent contribution. Models were projected into current and future conditions (RCP 2.6 – 2050, RCP 2.6 – 2070, RCP 8.5 – 2050, RCP 8.5 – 2070), and suitability maps were produced with a cloglog transformation, as recently recommended (Phillips et al., 2017), which results in suitability values ranging from 0 (totally unsuitable) to 1 (fully suitable). In order to compare modelling outputs, I calculated the overlap between the 'default settings' model and each of the tuned models, in current and future climate, using the Schoener's *D* index of niche overlap (Rödder & Engler, 2011; Schoener, 1968).

2.5 Evaluation of model performance

The evaluation of the predictive performance of SDMs has been a topic a debate for quite a while. Usually this is achieved by splitting the input data into training and evaluation sets, and the ability of the model to correctly predict the evaluation dataset is then summarised by the area under the receiver operating curve (AUC), an index routinely used in the medical field to evaluate the discrimination ability of diagnostic tests (Fielding & Bell, 1997; Jiménez-Valverde, 2012). In the 'default settings' model, I used a random k-fold approach in which 75% of the dataset was chosen to train the model and the remaining 25% were used to evaluate its performance. In order to make use of the entire dataset, this approach was repeated four times, each time using a different quarter of the data for evaluation, and I reported the mean AUC across the four replicates.

Even if this approach is relatively standard, it has been criticized for at least two reasons: the AUC is not a good measure of performance in presence/background modelling (Jiménez-Valverde, 2012; Lobo et al., 2008), and a random split generates training and evaluation sets that are not independent, artificially increasing the AUC (Bahn & McGill, 2013; Fourcade et al., 2018). Besides the inherent problems with AUC, a better approach is thus to partition the input data into spatially independent training and evaluation datasets. For this purpose, I used a spatial block approach to split data into four bins based on the latitude and longitude of occurrences, in such a way that the whole dataset was partitioned into four geographical rectangles containing equal amount of data (Radosavljevic & Anderson, 2014). The evaluation was then performed using the AUC in the same way as for the random k-

fold approach, except that now the evaluation dataset was always spatially independent from the training dataset. In addition to the apparent model performance, the final model predictions may also differ between these two alternative approaches because suitability maps were averaged across the four replicates, which were trained using different types of data (random subsets vs. geographically structured subsets).

To complement the AUC, I calculated two alternative measures of model performance that were not reported in Negrete et al. (2020). First, I used the threshold-dependent true skill statistics (TSS), computed after applying the threshold that maximizes the sum of specificity and sensitivity (Allouche et al., 2006). Second, I also reported the continuous Boyce index (Hirzel et al., 2006). Both indices were averaged across the four evaluation bins.

2.6 Suitable range area

In order to quantify the area of suitable habitat that is predicted under various scenarios, the continuous suitability predictions must be binarized, i.e. converted into suitable / not suitable areas. This involves defining a suitability threshold below which the environment will be considered unsuitable for the species and, reciprocally, above which the species may be potentially present. A natural approach would be to define a threshold *a priori*, which can be for example 0.5 – if suitability ranges from 0 to 1 – or a smaller value. However, the actual suitability values that would correspond in nature to a switch from negative to positive growth rate is difficult to assess beforehand, and would likely depend on many factors, including the ecology of the species and its prevalence, and the modelling method (Jiménez-Valverde & Lobo, 2007). Therefore, there are approaches that rely instead on the output suitability maps and the input occurrence data to find the most appropriate threshold for a given model. Here, I tested both a fixed threshold (set at 0.2 as in Negrete et al. (2020)) and a model-specific threshold that was defined as the threshold that maximizes the sum of specificity and sensitivity, as recommended by Liu et al. (2013). Since this threshold is model-dependent, and was thus computed four times per model type (for each cross-validation fold), I used the averaged threshold value to binarize the final continuous suitability map obtained by averaging predictions from each cross-validation fold. For both strategies, I calculated the area of suitable habitat predicted in each continent and each climate scenario.

3. Results

The performance of species distribution models for *O. nungara* was in all cases evaluated with high

values of AUC (> 0.8) (Table 1). Among them, the 'default settings', the 'regularisation multiplier selection' and the 'feature classes selection' models reached AUC values > 0.99 . The 'fully tuned' model, however, had the lowest AUC (0.81), followed by the 'background tuning' model (AUC = 0.87). All other models had AUC > 0.98 (Table 1).

Prediction maps of *O. nungara*'s potential distribution highlighted in all modelling settings highly climatically suitable areas in Europe and South America (Fig. 1). Specifically, France and northern Spain on the one hand, and the region that covers eastern Argentina, Uruguay and southern Brazil on the other hand, appeared as being among the most suitable areas at the global scale whatever the model considered. However, compared with the 'default settings' model, the 'fully tuned' model revealed the existence of suitable habitats in eastern and southern Africa, all along the southern coast of Australia, in central America, in the south of China and, more generally, showed larger areas of climatic suitability everywhere, including in Europe and south America (Fig. 1 and 4). Most of the other model settings resulted in prediction maps (in current climate) that largely resembled the 'default settings' model. There were, however, some regional differences, such as the 'uncorrelated variables' model that predicted larger suitable regions in south America (Fig. 1 and 4). Moreover, despite the apparent agreement between models overall, some suitability maps had only limited overlap with the 'default settings' model (Table 2). Notably, the most dissimilar model was the 'fully tuned' one, which exhibited an overlap < 0.5 in all climate scenarios (Table 2).

Examining response curves and variable contributions revealed substantial differences between model types in their modelled relationships between suitability and climatic predictors, even for models that otherwise produced relatively similar output maps (Fig. 2 and 3). Looking at response curves, the 'default settings' model was mostly influenced by BIO 6 (the minimum temperature of the coldest month) and BIO 17 (the precipitation of the driest quarter). However, measures of variable importance (Fig. 2) showed instead that this model was determined by BIO 1 (annual mean temperature, according to percent contribution) or BIO 3 (isothermality, according to permutation importance). The 'fully tuned' model, on the other hand, revealed smooth quadratic relationship with all the four selected predictors (Fig. 3). The 'uncorrelated variables only' model, which is fitted with the same reduced set of variables, showed relationships but with more complex functional forms, BIO 7 (temperature annual range) being the most influential variable (Fig. 3). Models where the regularisation multiplier or the feature classes were fine-tuned

('regularisation multiplier selection' and 'feature classes selection' models), or where the dataset was partitioned in space ('spatial block evaluation' model), exhibited the same response with BIO 6 and BIO 17 as the 'default settings' model (Fig. 3). Notably, the 'background tuning' models displayed very complex relationships between climatic suitability and all the 10 predictors (Fig. 3).

Selecting threshold from the model instead of a fixed 0.2 value resulted in larger suitable areas predicted with the 'default model' (Fig. 4). This was, however, the opposite for the 'fully tuned' model where the fixed threshold strategy led to much larger suitable areas in current and future climate. Looking at the whole range of model settings, there did not seem to be a consistent tendency for a larger or smaller area predicted with either threshold types (Fig. 4). When models were projected into future scenarios of climate change, the 'fully tuned' model predicted a strong decrease in the area of suitable habitats in virtually all continents and scenarios, both for the fixed and model-specific thresholds. This was, however, associated with new suitable habitats in previously unsuitable regions such as northern Europe (Fig. 1 and 4). This 'default settings' model predicted relatively stable areas in the future (Fig. 4), although with a shift towards higher latitudes in the stronger scenarios of climate change (Fig. 1). Generally, most other models behaved similarly to the 'default settings' model, with only little changes in the amount of suitable habitat in future climate (Fig. 4).

4. Discussion

The present study used as an example a recent study conducted by Negrete et al. (2020) that took advantage of a compilation of *O. nungara* occurrences in its native and introduced range to predict the species' potential distribution. I attempted to apply the same methods as described in Negrete et al. (2020) to fit the 'default model', using raw data and default settings. Interestingly, although the output prediction maps closely match those of Negrete et al. (2020), there are some notable differences between their and my models, especially in terms of the predicted changes in suitable range in future scenarios. While Negrete et al. (2020) models predicted an increase in the area of suitable range in all future scenarios compared to current climate, I failed to replicate this pattern (see the 'default model' with fixed threshold in Fig. 4). This suggests that even slight changes in methodology, so small that they are in fact indiscernible, may eventually lead to different ecological conclusions. Here, a possible reason for this discrepancy could be the version of MaxEnt (3.3.3k vs. 3.4.4, which explicitly adopts an inhomogeneous Poisson process approach (Phillips

et al., 2017)), the implementation of the MaxEnt algorithm (GUI vs. run from R via the ENMeval package), the type of transformation of the raw output (logistic vs. cloglog), or any other parameter not documented in Negrete et al. (2020).

This is remarkable that among the whole set of models tested, the 'fully tuned' model clearly stood out from the others. It was the only one that exhibited smooth changes in suitability across all continents with large areas of intermediate suitability, and that predicted vast climatically suitable regions in Africa. This is likely a sign that this model was the one that suffered the least from overfitting, as also evidenced by its simple and easy to interpret response curves (Merow et al., 2014). As a matter of fact, several of the tuning methods used here are explicitly designed to limit the effect of overfitting. Overfitting frequently arises when models are too complex (hence the selection of a simpler set of predictors and the fine-tuning of feature classes and regularisation multiplier (Merow et al., 2014; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011)), are biased towards oversampled regions (which is solved here by spatially thinning occurrences (Aiello-Lammens et al., 2015; Boria et al., 2014)) or include too large and irrelevant calibration areas (instead of restricting the background to the species' accessible area (Barve et al., 2011)). Despite that reducing the background area may lead to higher risk of extrapolating into non-analogous climates, we observed that the new regions found suitable in the 'fully tuned' model such as Africa and southern Australia are not characterised by strong extrapolation (see Fig. S2 in supplementary materials). In a similar study, Morales et al. (2017) examined the consequences of fine-tuning regularisation multiplier and feature classes compared to using default settings, for a set of previously published modelling studies. Very often, their best model used simpler response curves (e.g. selecting only linear or quadratic feature classes) and larger regularisation multiplier (up to five, compared to the default value of one) than the original studies, resulting in large differences in suitability maps. This suggests that many instances of overfitted distribution models exist in the literature, which may hold dangerous consequences when they are used in decision-making processes, such as underestimating invasion risk for exotic species.

I focused here on six different types of parametrization that are known to affect modelling performance. Despite that each of them was also tested separately, it was difficult to conclude about which setting was the most influential. Looking at Schoener's *D* values, it appears that models based on the restriction of background extent and uncorrelated variables had the lowest overlap with the 'default model'. This makes sense as these

parameters are related to the input data that are used to fit models, since they directly control the process of species' niche modelling by determining both which and from where environmental values are sampled. In this regard, the choice of a background extent (Iturbide et al., 2018) and the selection of environmental variables (Petitpierre et al., 2017) have been shown to be important predictors of SDMs' transferability. The second group of parameters that affected resemblance with the 'default model' included the selection of feature classes and regularisation multiplier. These settings control the relationship between suitability and environmental predictors, effectively changing the way suitability is projected in space and time (Brun et al., 2020). In contrast, the spatial block evaluation had less impact on the resulting suitability maps, which again can be explained by the fact that it only changed the way data was partitioned (Muscarella et al., 2014; Radosavljevic & Anderson, 2014). The only strong deviation from the 'default model' occurred when all settings were fine-tuned, using altogether all the approaches tested in this study. Therefore, it seems that, although each parameter can by itself change MaxEnt models and their projections, this is the combination of all of them that generates a drastically different species distribution model. A partial implementation of these recommendations may thus be insufficient to produce accurate predictions of species' potential distributions.

Although I highlighted the need for fine-tuning SDMs, this recommendation is not necessarily straightforward to apply. Some tools have been developed recently to test a range of parameters, in order to select the 'best' combination of settings in SDM studies. The first of this kind is certainly ENMeval (Muscarella et al., 2014), which I used in this study, that provides an automatic way of running multiple MaxEnt models differing by their feature classes and regularisation parameters. Since then, other tools have applied the same approach to other SDM methods, such as the SDM tune (Vignali et al., 2020) and biomod2 (Thuiller et al., 2009) R packages that allow tuning e.g. the number of trees in random forest or boosted regression trees algorithms. However, these procedures imply that there is an objective way to select the best parameters among a set of models. Originally, this type of model selection has been performed using AIC (Warren & Seifert, 2011), which is also the approach I employed here. Unfortunately, information metrics may be inappropriate to assess the predictive performance of SDMs (Velasco & González-Salazar, 2019). The alternative is to compare models by their AUC; despite being widely used, AUC is also a poor predictor of model performance and transferability (Fourcade et al., 2018; Lobo et al., 2008). In this regard, what I considered here as the best model (the

'fully tuned model') was not evaluated particularly well with regard to its AUC, but also using alternative measures of performance, even if they were computed based on spatially-independent partitioning of data. There is thus a need to develop better guidelines for model selection in a SDM context.

Some of the settings used here to fine-tune models were in part subjective. For example, the choice of a maximum VIF for selecting uncorrelated variables is entirely up to the modeller's decision since there is no universally accepted threshold. Ideally, SDMs should be fitted using environmental predictors that have a recognised effect of the species' ecology (Fourcade et al., 2018; Petitpierre et al., 2017). As shown here, such a choice changes the way species' presence correlates with the environment, which may strongly influence model transferability. The same subjectivity often exists in the choice of the training area and in the thinning distance for filtering occurrences. There are also alternative methods to deal with sampling bias (e.g. thinning in environmental space (Castellanos et al., 2019), manipulating the background (Vollering et al., 2019), etc.) that may eventually lead to different predictions of suitability. Finally, it must be noted that many SDM algorithms exist besides MaxEnt which I used here as a comparison with Negrete et al. (2020). Each of these alternative methods may provide different outcomes, such that ensemble approaches are sometimes favoured (Thuiller et al., 2009). However, it is not necessarily proven that ensemble modelling always gives better results than fine-tuned individual SDMs (Hao et al., 2020); in this case, it can be argued that expert-knowledge of the modelled species and of modelling tools are of crucial importance.

Despite the fact that this study was intended to be a modelling exercise applied to a methodological question, it provides some interesting insights into *O. nungara* invasion ecology. In contrast to Negrete et al. (2020), the fine-tuned model I developed revealed a large invasion potential in Africa in current and future climate. Since the species has not, to date, been detected on the continent, it may be essential to monitor carefully the introduction of exotic plants that may be a pathway for invasion in Africa. Another important outcome of the model is that a large part of central and south America, including the Brazilian Atlantic forest, appeared suitable, and not only the region that covers Uruguay and eastern Argentina, where the species has been observed (Negrete et al., 2020). It may reflect the limited dispersal ability of the species that constrains its natural range in a smaller region or, alternatively, *O. nungara* may occupy a larger range than currently thought. Luckily, these large areas of invasion risk in current climate are

predicted to decline relatively strongly in the future; even the currently suitable regions will shift northwards, which may contribute to extirpate the species from already invaded areas. Of course, the distribution that the species can actually reach depends on economic activity and trade networks that largely determine its medium- and large-scale dispersal (Justine et al., 2020). In this regard, incorporating dispersal constraints directly in the modelling process can help in defining realistic suitable ranges that avoid overprediction (Mendes et al., 2020).

5. Conclusion

The use of species distribution modelling approaches has grown enormously in the last twenty years (Elith & Leathwick, 2009; Guisan & Zimmermann, 2000); these methods have become standard in the analysis of distribution patterns and for predicting future range shifts for invasive species or under a changing climate. Paralleling this growth, methodological developments have resulted in a profusion of methods and recommendations for fitting models (e.g. Guillera-Aroita et al., 2015; Merow et al., 2013; Zurell et al., 2020), in such a way that it may be difficult to choose a modelling approach that unambiguously follows the latest up-to-date standards (especially since such standards do not necessarily exist). Here, I showed that different ways of processing data and choosing modelling settings may lead to contrasted predictions of the potential spread of an introduced flatworm species, *O. nungara*. Especially, failure to implement a suite of established recommendations – i.e. using raw data and default settings of the MaxEnt algorithm – conducted models to underestimate the area that the species can occupy.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. I thank Lisandro Negrete and his collaborators for having provided an interesting example to discuss issues related to SDM/ENM methodology, and for having made their data available.

References

- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541–545. <https://doi.org/10.1111/ecog.01132>
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill

- statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Atwater, D. Z., Ervine, C., & Barney, J. N. (2018). Climatic niche shifts are common in introduced plants. *Nature Ecology & Evolution*, 2, 34–43. <https://doi.org/10.1038/s41559-017-0396-z>
- Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, 122(3), 321–331. <https://doi.org/10.1111/j.1600-0706.2012.00299.x>
- Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222(11), 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Bellard, C., Leroy, B., Thuiller, W., Rysman, J.-F., & Courchamp, F. (2016). Major drivers of invasion risks throughout the world. *Ecosphere*, 7(3), e01241. <https://doi.org/10.1002/ecs2.1241>
- Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>
- Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z., & Zimmermann, N. E. (2020). Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, 47(1), 130–142. <https://doi.org/10.1111/jbi.13734>
- Castellanos, A. A., Huntley, J. W., Voelker, G., & Lawing, A. M. (2019). Environmental filtering improves ecological niche models across multiple scales. *Methods in Ecology and Evolution*, 10(4), 481–492. <https://doi.org/10.1111/2041-210X.13142>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Elith, J., Kearney, M., & Phillips, S. J. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38–49. <https://doi.org/10.1017/S0376892997000088>
- Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology & Biogeography*, 27, 245–256. <https://doi.org/10.1111/geb.12684>
- 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(5), e97122. <https://doi.org/10.1371/journal.pone.0097122>
- Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, 22(11), 1139–1149. <https://doi.org/10.1111/ddi.12477>
- Guillera-Arroita, G., Lahoz-Monfort, J., Elith, J., Gordon, A., Kujala, H., Lentini, P., McCarthy, M., Tingley, R., & Wintle, B. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3), 276–292. <https://doi.org/10.1111/geb.12268>
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, 43(4), 549–558. <https://doi.org/10.1111/ecog.04890>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>

- Iturbide, M., Bedia, J., & Gutiérrez, J. M. (2018). Background sampling and transferability of species distribution model ensembles under climate change. *Global and Planetary Change*, 166, 19–29. <https://doi.org/10.1016/j.gloplacha.2018.03.008>
- Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21(4), 498–507. <https://doi.org/10.1111/j.1466-8238.2011.00683.x>
- Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, 31, 361–369. <https://doi.org/10.1016/j.actao.2007.02.001>
- Justine, J.-L., Winsor, L., Gey, D., Gros, P., & Thévenot, J. (2020). Obama chez moi! The invasion of metropolitan France by the land planarian *Obama nungara* (Platyhelminthes, Geoplanidae). *PeerJ*, 8, e8385. <https://doi.org/10.7717/peerj.8385>
- Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, 40(4), 778–789. <https://doi.org/10.1111/jbi.12058>
- Liu, C., Wolter, C., Xian, W., & Jeschke, J. M. (2020). Most invasive species largely conserve their climatic niche. *Proceedings of the National Academy of Sciences*, 117(38), 23643–23651. <https://doi.org/10.1073/pnas.2004289117>
- Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2), 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Mendes, P., Velazco, S. J. E., Andrade, A. F. A. de, & De Marco, P. (2020). Dealing with overprediction in species distribution models: How adding distance constraints can improve model accuracy. *Ecological Modelling*, 431, 109180. <https://doi.org/10.1016/j.ecolmodel.2020.109180>
- Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37, 1267–1281. <https://doi.org/10.1111/ecog.00845>
- 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>
- Morales, N. S., Fernandez, I. C., & Baca-Gonzalez, 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>
- Morán-Ordóñez, A., Lahoz-Monfort, J. J., Elith, J., & Wintle, B. A. (2017). Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: What factors influence the reliability of predictions? *Global Ecology and Biogeography*, 26, 371–384. <https://doi.org/10.1111/geb.12545>
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. a, Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198–1205. <https://doi.org/10.1111/2041-210X.12261>
- Naimi, B., Hamm, N. a S., Groen, T. a, Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191–203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- Negrete, L., Francavilla, M. L., Damborenea, C., & Brusa, F. (2020). Trying to take over the world: Potential distribution of *Obama nungara* (Platyhelminthes: Geoplanidae), the Neotropical land planarian that has reached Europe. *Global Change Biology*, 26(9), 4907–4918. <https://doi.org/10.1111/gcb.15208>
- Osorio-Olvera, L., Lira-Noriega, A., Soberón, J., Peterson, A. T., Falconi, M., Contreras-Díaz, R. G., Martínez-Meyer, E., Barve, V., & Barve, N. (2020). ntbox: An r package with graphical user interface for modelling and evaluating multidimensional ecological niches. *Methods in Ecology and Evolution*, 11(10), 1199–1206. <https://doi.org/10.1111/2041-210X.13452>
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E., & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263(July), 10–18. <https://doi.org/10.1016/j.ecolmodel.2013.04.011>
- Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., & Guisan, A. (2017). Selecting predictors to maximize the transferability of species distribution models: Lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, 26(3), 275–287. <https://doi.org/10.1111/geb.12530>

- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40, 887–893. <https://doi.org/10.1111/ecog.03049>
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629–643. <https://doi.org/10.1111/jbi.12227>
- Rödder, D., & Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: Advances and possible drawbacks. *Global Ecology and Biogeography*, 20(6), 915–927. <https://doi.org/10.1111/j.1466-8238.2011.00659.x>
- Ruete, A. (2015). Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. *Biodiversity Data Journal*, 3, e5361. <https://doi.org/10.3897/BDJ.3.e5361>
- Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., & Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, 44, 199–205. <https://doi.org/10.1111/ecog.05360>
- Schoener, T. W. (1968). The Anolis lizards of Bimini: Resource partitioning in a complex fauna. *Ecology*, 49(4), 704–726. <https://doi.org/10.2307/1935534>
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, 32(3), 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Thuiller, W., Richardson, D. M., Pyšek, P., Midgley, G. F., Hughes, G. O., & Rouget, M. (2005). Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, 11(12), 2234–2250. <https://doi.org/10.1111/j.1365-2486.2005.001018.x>
- Velasco, J. A., & González-Salazar, C. (2019). Akaike information criterion should not be a “test” of geographical prediction accuracy in ecological niche modelling. *Ecological Informatics*, 51, 25–32. <https://doi.org/10.1016/j.ecoinf.2019.02.005>
- Venette, R. C., Kriticos, D. J., Magarey, R. D., Koch, F. H., Baker, R. H. A., Worner, S. P., Gómez Raboteaux, N. N., McKenney, D. W., Dobesberger, E. J., Yemshanov, D., De Barro, P. J., Hutchison, W. D., Fowler, G., Kalaris, T. M., & Pedlar, J. (2010). Pest Risk Maps for Invasive Alien Species: A Roadmap for Improvement. *BioScience*, 60(5), 349–362. <https://doi.org/10.1525/bio.2010.60.5.5>
- Vignali, S., Barras, A. G., Arlettaz, R., & Braunisch, V. (2020). SDMtune: An R package to tune and evaluate species distribution models. *Ecology and Evolution*, 10(20), 11488–11506. <https://doi.org/10.1002/ece3.6786>
- Vollering, J., Halvorsen, R., Auestad, I., & Rydgren, K. (2019). Bunching up the background betters bias in species distribution models. *Ecography*, 42(10), 1717–1727. <https://doi.org/10.1111/ecog.04503>
- Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21(2), 335–342. <https://doi.org/10.1890/10-1171.1>
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277. <https://doi.org/10.1111/ecog.04960>

Table 1: Summary of model settings and evaluation of the models. Model performance was evaluated with the area under the receiver operating curve (AUC), the true skill statistics (TSS) and the continuous Boyce index (CBI) using random k-fold or spatial block cross-validation approaches. Values of evaluation metrics are averaged across the four evaluation folds.

Model name	Model description	AUC	TSS	CBI
Default settings	Fitted with raw input data and all MaxEnt default settings (feature classes: LQHPT, regularisation multiplier = 1, all 10 variables, all 144 occurrences, background points sampled globally, model evaluated through random k-folds)	0.992	0.939	0.900
Fully tuned	Fitted with all up-to-date recommendations (see below for details): spatial thinning of occurrences, incorporation of uncorrelated variables only, background sampled close to observed occurrences, evaluation conducted in spatial blocks, model selection to find the best feature classes (LQ) and regularisation multiplier (0.5)	0.807	0.555	0.855
Spatial thinning	Same as 'default settings', except that occurrence data were filtered so that they are all separated by at least 50 km (final, spatially thinned, dataset: 93 occurrences)	0.988	0.938	0.864
Uncorrelated variables only	Same as 'default settings', except that a subset of bioclimatic variables was used in modelling in such a way that their variance inflation factor was < 4 (final set of predictors: BIO 7, BIO 10, BIO 17 and BIO 19)	0.984	0.902	0.844
Background tuning	Same as 'default settings', except that background points were selected at random within 500 km of occurrence locations	0.869	0.635	0.818
Spatial block evaluation	Same as 'default settings', except that occurrences were split into training/evaluation sets using a spatial block approach	0.987	0.931	0.639
Feature classes selection	Same as 'default settings', except that feature classes were selected from the best fitted model across multiple models (tested: L, LQ, H, LQH, LQHP, LQHPT; selected: LQHP)	0.993	0.952	0.881
Regularisation multiplier selection	Same as 'default settings', except that the regularisation multiplier was selected from the best fitted model across multiple models (tested: 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, selected: 2)	0.992	0.932	0.925

Table 2: Overlap between default and each fined-tuned model, calculated using Schoener's *D* niche overlap index.

	Fully tuned	Spatial thinning	Uncorrelated variables only	Background tuning	Spatial block evaluation	Feature classes selection	Regularisation multiplier selection
Current	0.372	0.882	0.604	0.709	0.907	0.794	0.858
RCP 2.6 - 2050	0.435	0.888	0.595	0.746	0.920	0.820	0.854
RCP 2.6 - 2070	0.440	0.891	0.598	0.749	0.922	0.821	0.856
RCP 8.5 - 2050	0.455	0.892	0.590	0.760	0.920	0.823	0.846
RCP 8.5 - 2070	0.464	0.895	0.567	0.765	0.922	0.816	0.835

Figure 1: Suitability maps for *O. nungara* obtained from each of the eight MaxEnt models (see Table 1 for settings), in current climate (left column) and in future climate (two right columns) under the RCP 8.5 scenario. Projections under the RCP 2.6 scenario are available in Supplementary materials, Figure S1.

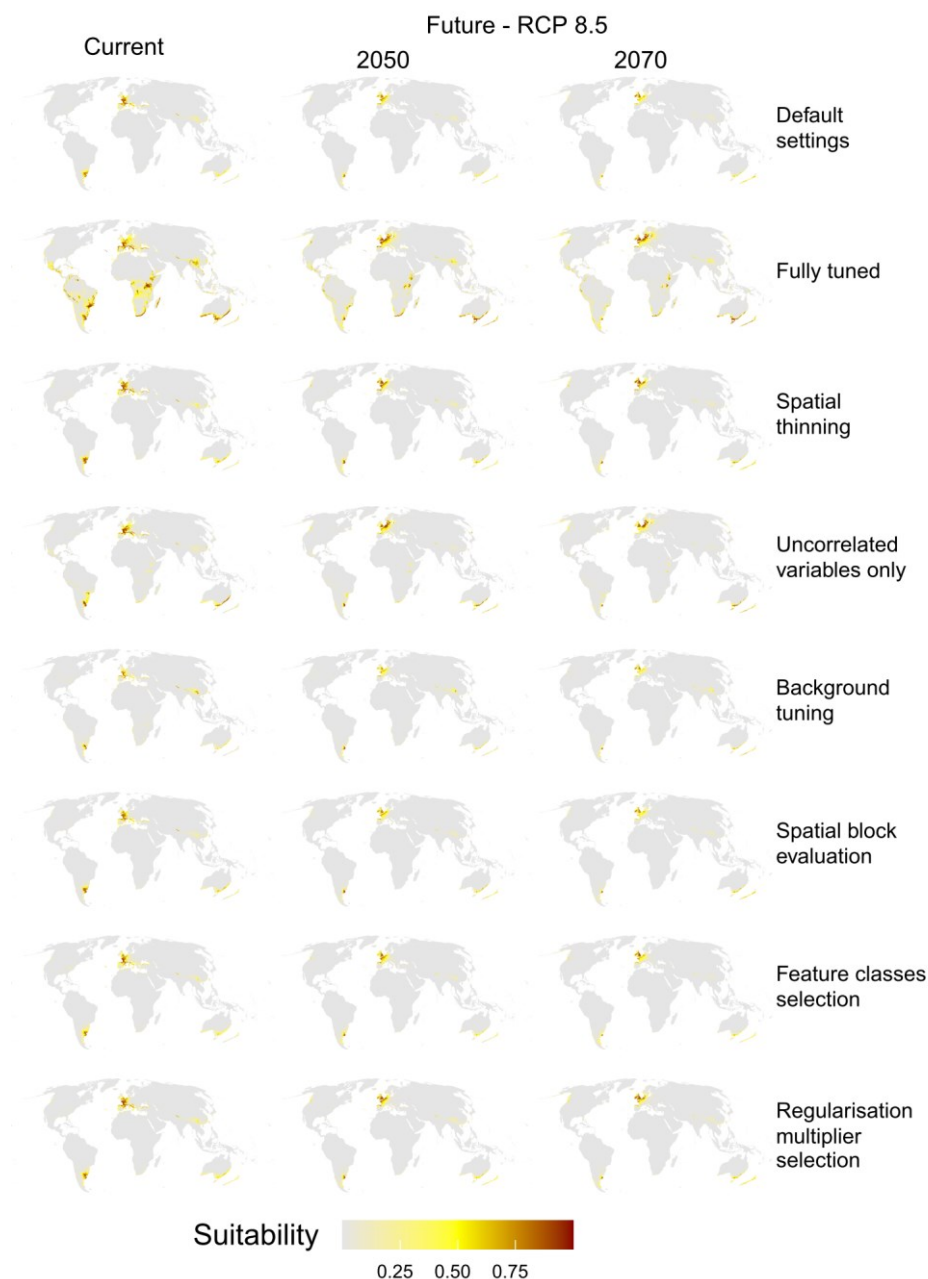


Figure 2: Variable importance obtained from each of the eight models (see Table 1 for settings), calculated using either percent contribution or permutation importance.

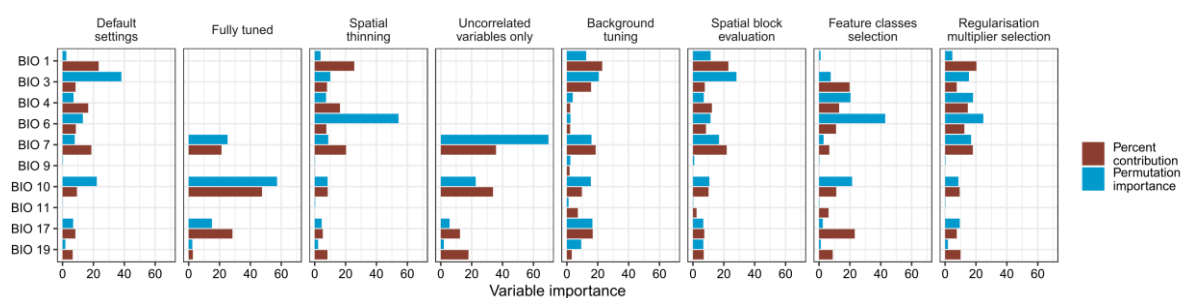


Figure 3: Modelled response curves obtained from each of the eight models (see Table 1 for settings), showing the relationship between environmental suitability and each predictor when the others are set at their mean values.

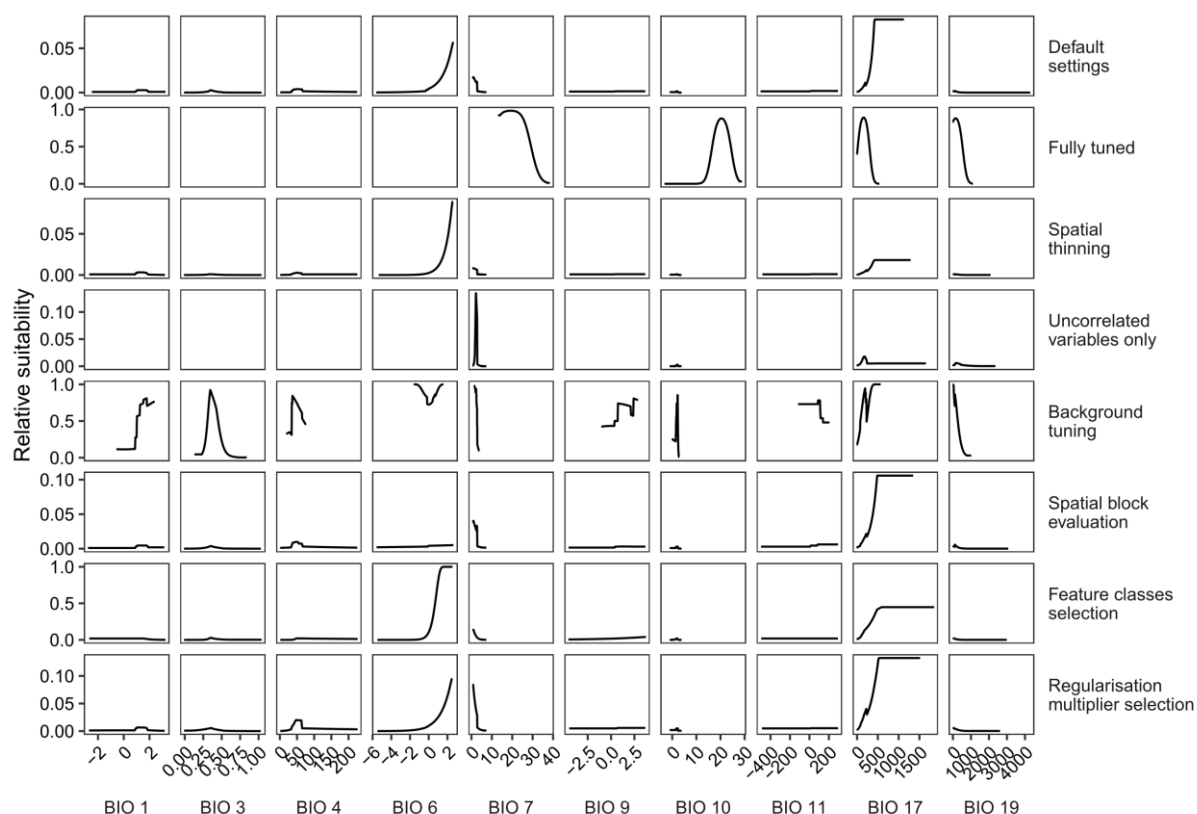
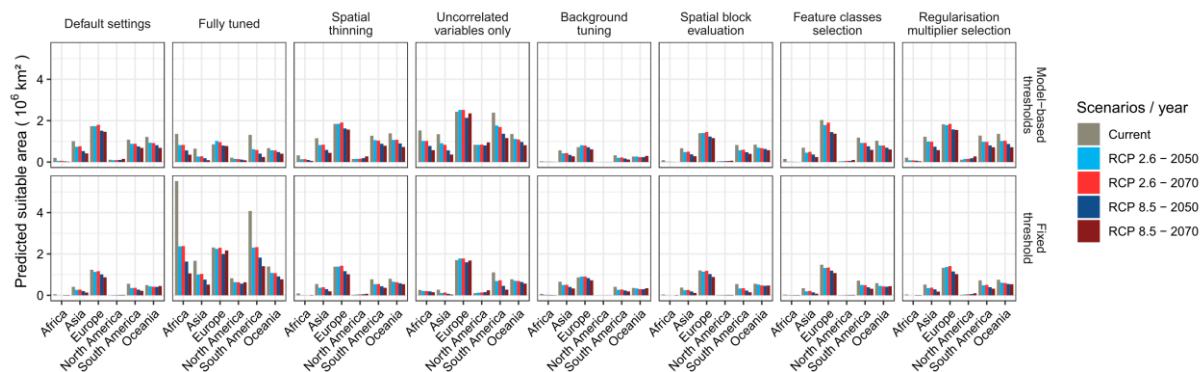


Figure 4: Predicted suitable area obtained from each of the eight models (see Table 1 for settings), in current (grey bar) and future (coloured bars) climate, when continuous suitability values were converted to binary predictions using the threshold that maximises specificity and sensitivity (top) or a fixed threshold of 0.2 (bottom).



Supplementary materials

Figure S1: Suitability maps for *O. nungara* obtained from each of the eight MaxEnt models (see Table 1 for settings), in current climate (left column) and in future climate (two right columns) under the RCP 2.6 scenario. Projections under the RCP 8.5 scenario are available in Supplementary materials, Figure S1.

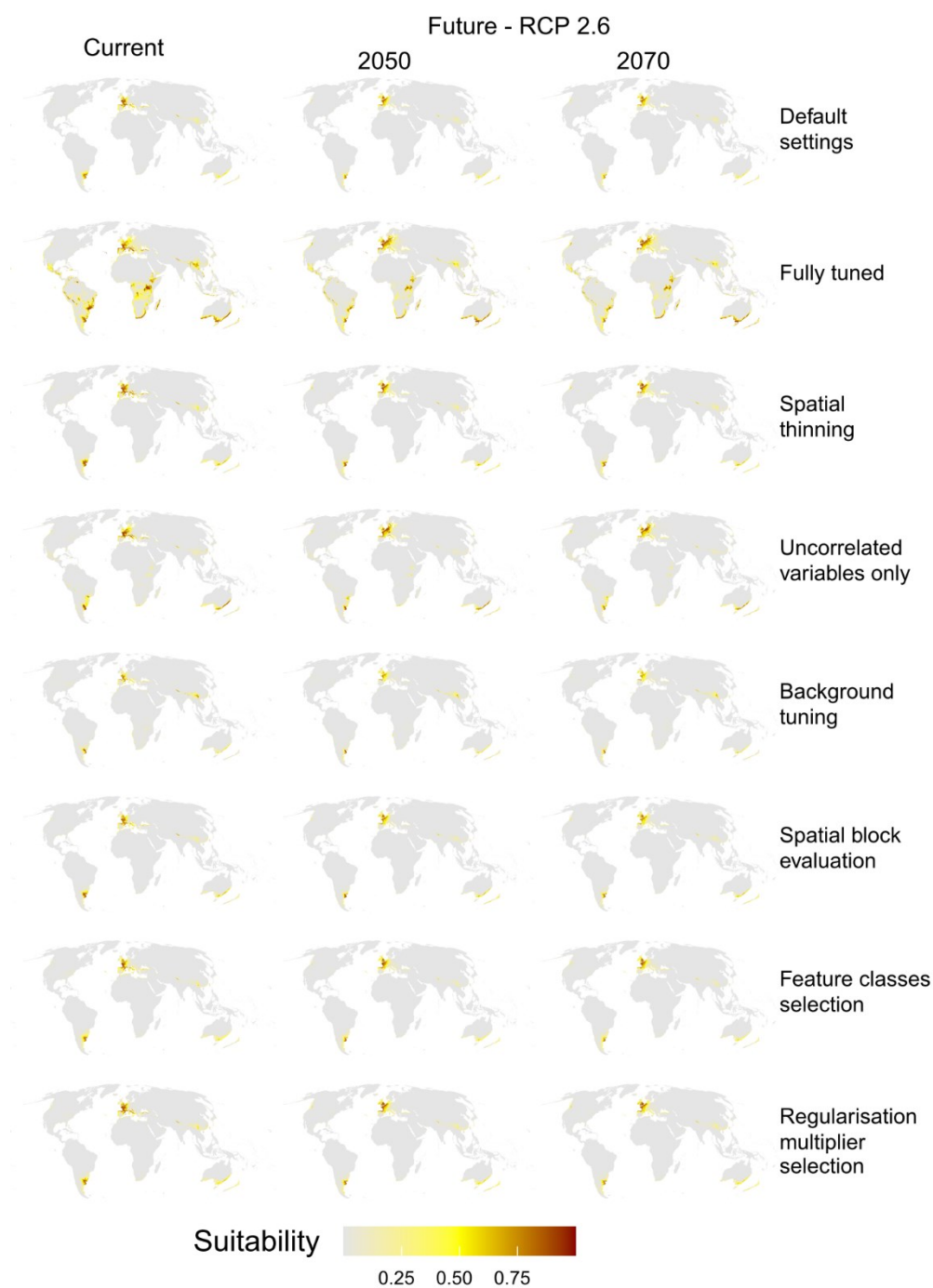


Figure S2: Extrapolation risk analysis represented as Mobility-Oriented Parity (MOP) maps. Colour scale shows the similarity between the climatic conditions at each grid cell (in current climate) and the climatic conditions in the training data. Areas of strict extrapolation are displayed in red.

