

# CHAPTER 2

## IMPROVING THE TRANSFERABILITY OF INTRODUCED SPECIES' DISTRIBUTION MODELS: NEW TOOLS TO FORECAST SPREADING OF A HIGHLY INVASIVE SEAWEED

Heroen Verbruggen<sup>1</sup>, Lennert Tyberghein<sup>2</sup>, Gareth S. Belton<sup>3</sup>, Frederic Mineur<sup>4</sup>, Alexander Jueterbock<sup>5</sup>, Galice Hoarau<sup>5</sup>, C. Frederico D. Gurgel<sup>3</sup>, Olivier De Clerck<sup>2</sup>

**1** School of Botany, University of Melbourne, Victoria 3010, Australia

**2** Phycology Research Group, Ghent University, Krijgslaan 281 (S8), 9000 Ghent, Belgium

**3** School of Earth and Environmental Sciences, Faculty of Science, University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia

**4** Mediterranean Institute of Oceanography (MIO), Aix-Marseille University, 13288 Marseille cedex 09, France

**5** Marine Ecology Group, Faculty of Biosciences and Aquaculture, Høgskølen I Bodø, 8049 Bodø, Norway

Manuscript in prep.

# ABSTRACT

Globalization has led to an increased spread of alien invasive species. Species distribution modeling (SDM) can help us understand and map species' distributions, and play a key role in forecasting the range expansion of such species. In order for MaxEnt presence-only SDMs to be useful in predicting and managing introduced and invasive species, some problems have to be overcome. This study focuses on the choices made during the modeling process that potentially affect the transferability and overall predictive performance of the resulting model. Two methods were introduced to increase the transferability of correlative SDMs, more specifically a tool that reduces geographical bias in occurrence records and one that identifies suitable sets of predictors. Our results showed that reducing the set of predictors to those anticipated to be of global significance yielded a strong improvement of SDM transferability, with occurrence thinning, model complexity and background choice having relatively minor effects. If both occurrences from the native and invaded regions are available, they should be combined, as this results in the best-performing models and reduces the sensitivity to choices made in the modeling process. We also presented an SDM of the highly invasive species *Caulerpa cylindracea* that achieves very high predictive power.

# INTRODUCTION

A large number of species are experiencing range modifications, either because humans have spread them to new areas or as a response to changing environmental conditions. Species distribution models (SDMs) help us understand and map species' distributions, and play a key role in forecasting the range expansion of introduced species and the effects of climate change on distributions (Elith & Leathwick, 2009; Peterson, 2003; Wiens *et al.*, 2009; Jimenez-Valverde *et al.*, 2011). An SDM characterizes the species' response to relevant environmental variables, either using physiological information from experimental work (mechanistic models) or by relating the presence and/or absence of the species to environmental information (correlative models) (Franklin & Miller, 2009). This response is subsequently projected into geographic space using gridded environmental layers, resulting in a map showing the potential distribution of the species. Because experimental physiological work has not been carried out for a great majority of species, correlative modeling approaches are by far the most common. Whereas species occurrence data are abundant in museum databases and the literature, it is far more troublesome to assess the absence of species from an area, hence most SDMs rely on presence-only techniques (Elith & Leathwick, 2009).

A crucial assumption in using SDMs to forecast the spread of introduced species or distribution changes in response to environmental change is that the model is transferable to the new conditions. In the case of introduced species, models trained primarily on distribution data from the species' native range need to be transferred to the region where it has been introduced. This often implies projecting the species response to climatic conditions that are not present in the native (training) range, which is an innately difficult task. For such situations, it is valuable to visualize those areas where extrapolation beyond observed conditions was required and consider those predictions as uncertain (Elith *et al.*, 2010; Zurell *et al.*, 2012). In addition, the ability of presence-only methods to capture a species' ecological response is affected by the choice of background points (Elith *et al.*, 2010), predictor variables (Rödder & Lötters, 2010) and model complexity (Warren & Seifert, 2010; Rodda *et al.*, 2011) as well as the geographic spread of occurrence records in relation to environmental gradients (Wolmarans *et al.*, 2010). Besides these problems, it is also possible that biotic interactions limit the utility of models based on abiotic predictors (Leathwick & Austin, 2001) and, of course, there is always the possibility that the fundamental niche of the introduced population has changed due to natural selection (Whitney & Gabler, 2008; Alexander & Edwards, 2010).

This study focuses on the second set of problems, i.e. the choices made during the modeling process that affect the transferability and overall predictive performance of the resulting model. We introduce two methods that have the potential to increase the transferability of correlative SDMs, more specifically a tool that reduces geographical bias in occurrence records and one that identifies suitable sets of predictors. We apply these methods to the introduced and highly invasive seaweed species *Caulerpa cylindracea*, aiming to infer a solid niche model for this species that can assist in predicting areas with suitable habitat worldwide and assessing the risk of further spreading. The methods are evaluated based on their predictive performance for the case study.



# METHODS

## EXPERIMENTAL DESIGN

Before providing detailed methods we will sketch and explain the overall design of our study. The overarching goal is to examine and improve the transferability and overall performance of maximum entropy (MaxEnt) presence-only models of introduced species. The experimental design centers on the impact of four important choices that have to be made during the modeling process: (1) the amount of geographic autocorrelation in occurrence records, (2) the choice of predictor variables, (3) the complexity of the model, and (4) the selection of background points.

Because most environmental variables show spatial autocorrelation, geographically biased sampling of occurrence records (e.g. heterogeneous accessibility and local expertise) naturally results in environmental biases in the data used to train the SDM (Dormann *et al.*, 2007), leading to model misspecification (Santika & Hutchinson, 2009; Wolmarans *et al.*, 2010). We introduce a method that thins occurrence records in densely sampled regions to obtain a more even geographic distribution (details given below). To examine the effect of this method, models are built before and after occurrence thinning.

The choice of predictor variables is arguably one of the most studied elements affecting the transferability of SDMs, with several papers showing differences in transferability depending on which predictor set is used (Rödder *et al.*, 2009; Rödder & Lötters, 2010; Peterson & Nakazawa, 2008) and leading to the identification of more conserved and relaxed predictors (Rödder & Lötters, 2009; Rödder & Lötters, 2010). We introduce a method that surveys the performance of all possible predictor sets (explained below) and evaluate the transferability of models built with two different sets of predictors.

The complexity of an SDM is also known to impact on its predictive performance, with overfitting often leading to poor transferability (Warren & Seifert, 2010; Randin *et al.*, 2006; Wenger & Olden, 2012). By default, MaxEnt determines the types of features it allows automatically, based on the number of samples available for model training (Phillips & Dudik, 2008), but this standard behavior has been reported to result in overfitted models (Rodda *et al.*, 2011). We compare models with automatically determined model complexity to models forced to be simple.

Finally, the selection of background points is known to affect the performance of presence-only SDMs (Phillips *et al.*, 2009; Elith *et al.*, 2010). To examine this, we compare SDMs built with global background points to models built with a regional background.

The study is carried out on a case study, an introduced seaweed species. Model transferability is assessed by training models on samples from either the native or the invaded range and measuring the overlap of the two models, as well as by calculating how well they predict presences in the other range. We also compare the overall predictive performance of SDMs trained with occurrence from either range to that of a model combining occurrences from both ranges.

## STUDY SPECIES AND ENVIRONMENTAL DATA

This study focuses on the introduced and highly invasive seaweed species *Caulerpa cylindracea* Sonder (Sonder, 1845). Specimens of the *Caulerpa* genus are well known for their rampant morphological plasticity that, due to the inconsistent use of varieties and forms amongst

taxonomists, has resulted in a highly confusing and cumbersome nomenclature. Most of this confusion has existed around the *C. racemosa/peltata* complex that has more than 30 described varieties and forms (Belton *et al.*, 2012). Until recently this included *C. cylindracea*, which, although originally described as an independent species, had long been considered a form of *C. racemosa* var. *laetevirens* until it was raised to varietal status (Verlaque *et al.*, 2003) and finally reinstated as an independent species based on molecular data (Belton *et al.*, 2012).

Since the early 1990s *C. cylindracea* has rapidly and aggressively spread in the Mediterranean Sea and Canary Islands, representing one of the most dramatic marine invasions in terms of establishment and ecological dominance (Klein & Verlaque, 2008; Verlaque *et al.*, 2004). The species has been reported from all kinds of substrata and depths, with a variety of benthic assemblages, and thrives in disturbed habitats of the heavily urbanized coastline of the Mediterranean Sea (Piazzi *et al.*, 2005; Klein & Verlaque, 2008). Invasive populations of *C. cylindracea* establish dense and compact monospecific stands, which easily overgrow and outcompete and/or negatively impact other seaweed (Piazzi & Ceccarelli, 2006; Piazzi *et al.*, 2001), seagrass (Ceccherelli & Campo, 2002) and invertebrate species (Kruzic *et al.*, 2008; Baldaconi & Corriero, 2009) leading to the biotic homogenization of affected areas (Piazzi & Balata, 2009) and an overall decrease in the total number of species in affected areas (Klein & Verlaque, 2008). Only partial recovery of the assemblages could be observed after eradication of *C. cylindracea* in Italy and France (Piazzi & Ceccarelli, 2006; Klein, 2007).

Unlike *C. taxifolia*, which was accidentally introduced from a public aquarium (Jousson *et al.*, 1998), the vector of introduction of *C. cylindracea* to the Mediterranean Sea is unknown. It was initially hypothesized to be a Lessepsian immigrant (Alongi *et al.*, 1993; Giaccone & Di Martino, 1995), or a hybrid between *C. racemosa* var. *turbinata* and an unknown tropical variety (Durand *et al.*, 2002), until molecular investigations identified a potential source population in southwestern Australia (Verlaque *et al.*, 2003). However, recent findings indicate that the native range of *C. cylindracea* is much larger than previously thought (extending from Western Australia around northern Australia into the Great Barrier Reef and New Caledonia), and that the source of the invasive *C. cylindracea* populations in the Mediterranean Sea is not known with certainty (Sauvage *et al.*, 2012).

Whatever the vector and source population, *C. cylindracea* is spreading rapidly with reports of its presence in 12 Mediterranean countries including all the large islands (Verlaque *et al.*, 2003; Nuber *et al.*, 2007), and has more recently been reported from two locations on the southern coast of Australia (reference (Collings *et al.*, 2004) and unpublished data GSB). Its rapid spread through the invaded range makes this species a suitable case study for the question at hand.

A total of 191 distribution records were assembled from the Australia Virtual Herbarium (<http://chah.gov.au/avh/>), new collections by GSB that will be deposited in the AD herbarium, the data collection gathered by FM for the ERC FP5 ALIENS project, and the literature (Sauvage *et al.*, 2012; Womersley, 1984; Huisman, 2000; Piazzi *et al.*, 1994; Piazzi *et al.*, 1997; Verlaque *et al.*, 2000; Verlaque *et al.*, 2004; Zuljevic *et al.*, 2003; Gambi & Terlizzi, 1998; Alongi *et al.*, 1993; Buia *et al.*, 1998; Nizamuddin, 1991; Serio & Pizzuto, 1999; Djellouli, 2000; Cossu & Gazale, 1997; Pena-Martin *et al.*, 2003). This included 65 records from the native range in Australia and New Caledonia, 111 from the invaded range in Europe, and 15 from recently invaded areas in South Australia.

We used the Bio-ORACLE dataset (Tyberghein *et al.*, 2012) as a source of marine environmental grids (90°N–90°S, real values). To make the distribution records compatible with the grids, occurrence coordinates situated on land according to the Bio-ORACLE grids were moved to the closest cell in the ocean. When multiple records were situated in the same Bio-ORACLE grid cell, a single record was retained, resulting in 96 records being discarded.



## OCCURRENCE THINNING

Geographical biases in the occurrence records was dampened by thinning the distribution points with OccurrenceThinner 1.03 (Verbruggen, 2011b). We developed this program to filter occurrence records using a probability-based procedure. The probability that any specific occurrence record is removed is proportional to the density of occurrence records in the area as defined by a kernel density grid. The two-dimensional binned kernel density grid used in this procedure was computed from the occurrence records with the `bkde2D` function in the R package `KernSmooth` v.2.23-7 (Wand & Jones, 1995; Ripley, 2011), with a bandwidth of 3.0. The thinning procedure with thresholds  $t_1=0.5$  and  $t_2=1.0$  was repeated 10 times, resulting in 10 occurrence-thinned datasets. These datasets had on average 25 records from the native range, 46 from the European invaded range, and 3 from the South Australian invasive populations.

To evaluate whether occurrence thinning influences model transferability and performance, we compared MaxEnt models based on all occurrence records to SDMs using a thinned subset of samples.

## PREDICTOR SETS

The predictor variables were chosen in two steps. The first step consisted of *a priori* selection of a set of 8 predictors. This selection was based on knowledge of the physiological determinants of seaweed distributions (Lüning, 1990), and takes the structure of the Bio-ORACLE dataset into account by not using multiple closely correlated predictors. The eight resulting predictors were mean sea surface temperature (SSTmean), the range in sea surface temperature (SSTrange) as a measure of seasonality, mean photosynthetically active radiation (PARmean), salinity, pH, mean diffuse attenuation (DAmeyn) as a measure of water transparency, dissolved oxygen (dissox) and the phosphate concentration. The nitrate concentration was not included because it is correlated with the phosphate concentration (Tyberghein *et al.*, 2012).

In the second step, the predictive ability of those eight variables was explored using MaxEnt Model Surveyor (MMS) version 1.03 (Verbruggen, 2011a). We developed this software to evaluate the performance of all possible subsets of variables, using the test AUC to measure model performance. The program was run multiple times: (1) on samples from native range with global background, (2) on samples from invaded range with global background, (3) on samples from both ranges with global background, (4) on samples from native range with background restricted to native range, and (5) on samples from invaded range with background restricted to invaded range. The program used 50% of the samples for training and 50% for testing, it restricted the model complexity to linear and quadratic features and used thinned occurrences. Each of the five runs listed above was repeated 10 times (i.e., on each of the 10 replicate sets of thinned occurrences). The training and test data were randomly drawn from the occurrence records and do not represent a subdivision into the native vs. invaded range. As a consequence, the model performance used to evaluate predictor combinations does not represent transferability between regions. From the MMS results, a consensus was derived as to which variables are most important across the different runs. This resulted in a set of four predictor variables (specified in results).

In order to evaluate whether this predictor selection approach can improve the transferability of models across regions, MaxEnt models were ran with all eight variables listed above as well as the subset of four generated with the predictor selection procedure.

## MODEL COMPLEXITY

Model complexity was varied to verify its impact on the predictive power and transferability. Two different settings were used. The first used the default behavior of MaxEnt, which determines which features are used based on the number of samples. The second setting forced the use of smooth response curves by allowing only linear and quadratic features to be fitted.

## BACKGROUND SELECTION

Data for background points was extracted from the Bio-ORACLE grids (Tyberghein *et al.*, 2012). Three sets of 10,000 random background points were created: (1) from the entire globe, (2) from the native range defined as a box around Australia with latitude between 5°S and 45°S and longitude between 100°E and 175°E, and (3) from the invaded range defined as West Europe extending to Africa and the Mediterranean Sea, as a box with latitude between 20°N and 60°N and longitude between 35°W and 40°E. These boxes roughly correspond to the maps of the native and invaded ranges presented in the results. In each of these three cases, the background selection corrected for unequal areas at different latitudes, i.e. they correspond to random pixel draws from equal area grids.

To compare the effect of background selection on transferability, regional models with corresponding regional backgrounds were compared to regional models with global background. Models trained with samples from the native and invaded range combined always used the global background.

## NICHE MODEL INFERENCE

Niche models were inferred with MaxEnt 3.3.3f (Phillips *et al.*, 2006; Phillips & Dudik, 2008; Elith *et al.*, 2011). The analyses were automated via a Perl script and carried out on a multicore linux server. All analyses were run with 10,000 random background points as specified above. The training and test points, as well as the background points, were provided as SWD files and the resulting models were projected onto the Bio-ORACLE grid (Tyberghein *et al.*, 2012). MaxEnt's jackknife function was activated and samples were not added to the background to avoid complicating model comparisons. The models resulting from the ten replicate occurrence-thinned training sets were averaged for visualization but all other interpretations were based on the individual models.

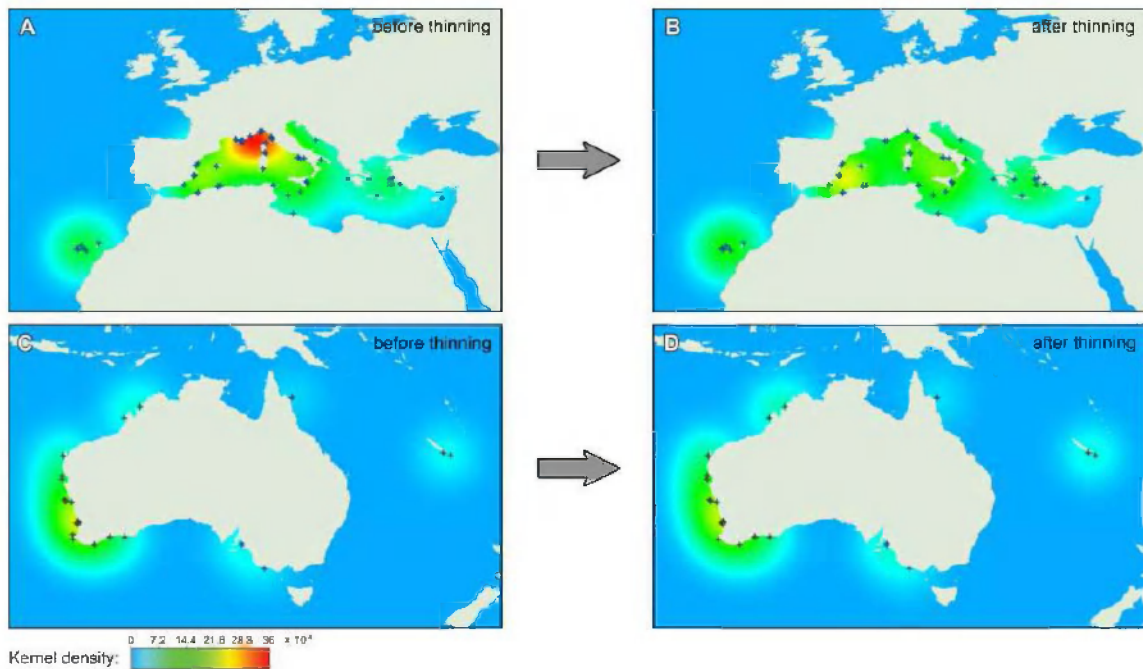
## DOWNSTREAM ANALYSES

Models are compared to identify which choices lead to better-performing models. In order to evaluate the transferability of models, models built on the native and invaded ranges were compared in a pairwise fashion, using the Schoener's D niche similarity measure and reciprocal test AUC (i.e. native training samples with test samples in invaded range and vice versa). The overall predictive power of all models was compared with the test AUC.

# RESULTS

## EXPLORATION OF NEW METHODS

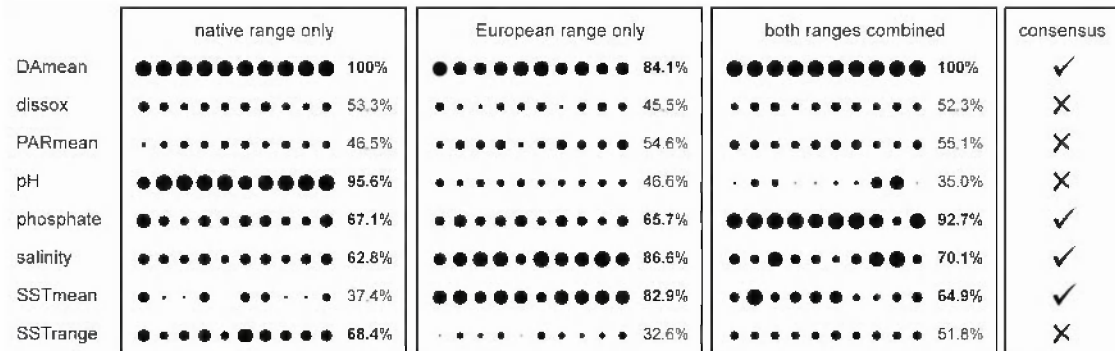
We implemented two methods that tackle issues related to the overall quality and transferability of niche models. Occurrence thinning clearly reduced the geographic sampling bias present in the occurrence points, as indicated by the kernel density plots before and after occurrence thinning (Figure 1). In this figure, the red blob with dense sampling along the French Riviera and nearby localities disappears entirely after the thinning procedure (Figure 1A-B). Geographic sampling bias was less of a problem in the native range (Figure 1C-D).



**Figure 1: Effect of occurrence thinning on geographical sample bias.** The colors on the map represent the regional sampling density, warmer colors indicating higher sample densities. Occurrence thinning substantially reduces the geographic sampling bias, as illustrated by the disappearance of the red blob along the French Riviera and closeby localities (panel A → B). There is less geographic sampling bias in the native range, so occurrence thinning does not have a big influence on the kernel density maps of that region (panel C → D).



The survey of predictor combinations is summarized in Figure 2. As could be anticipated from previous studies, the representation frequency of variables among the top-scoring models is sensitive to whether the analysis was done on the native range, the invaded range, or both combined. Using local or global background points resulted in qualitatively similar results (Figure S1). The consensus made across the three boxes, including only variables that are likely to be of global significance (present in at least 60% of the top-scoring models for at least 2 out of 3 regions), consisted of 4 predictors: DAmean, phosphate, salinity and SSTmean.



**Figure 2:** Results of the surveying procedure to identify the predictors present in top-scoring models. Each box represents the results of the predictor set survey for occurrence records from the native range, the invaded European range, or both ranges combined using global background data. The ten columns within each box represent independent runs on different sets of thinned coordinates, and the circle diameter represents how often the variable in question occurs in the top 10 highest-scoring models (test AUC) for that set of occurrences. The representation of each predictor in the top 10 is also summarized across columns (percentage), and the consensus predictor set across ranges is indicated in the box on the right.

The effect of these two methods on model performance was evaluated by including them as factors in our experiment. So all MaxEnt analyses were run with all samples and thinned samples. Similarly, models were run with all 8 variables included and with only the four consensus variables selected from the survey.

## TRANSFERABILITY AS A FUNCTION OF MODELING CHOICES

Our multifactorial experiment showed that reduction of the number of predictors based on a surveying method yielded much better models with higher test AUCs (Figure 3A) and Schoener's D (Figure 3B) than models with the full set of eight predictors. This is clearly visible in the figures: the leftmost two columns of both panels of the figure have warmer colors than the rightmost two columns. A Wilcoxon signed-rank test (WSRT) indicated that the difference in test AUC and Schoener's D between matching models is significant ( $p = 0.0078$  in both cases,  $N = 8$ ).

With test AUC as the measure of transferability (Figure 3A), the two upper rows had warmer colors than the lower two rows, indicating better performance of models that use global background samples compared to models in which background samples are restricted to the region in which the model is trained. This pattern was not present in the Schoener's D values (Figure 3B), where models with global background and auto-features had remarkably low values of D, and the WSRT outcomes conflicted strongly ( $p = 0.0078$  for AUC,  $p = 0.9453$  for Schoener's D,  $N = 8$ ).

Model complexity and occurrence thinning did not seem to have a large effect on transferability, except for the second row in Figure 3B, which shows substantially lower Schoener's D for a set of models with auto-features compared to the same set of models with enforced simple models (the row above), which led to a p-value of 0.055 for the WSRT (N = 8) even though the values with auto-features are not lower for the regional background case.

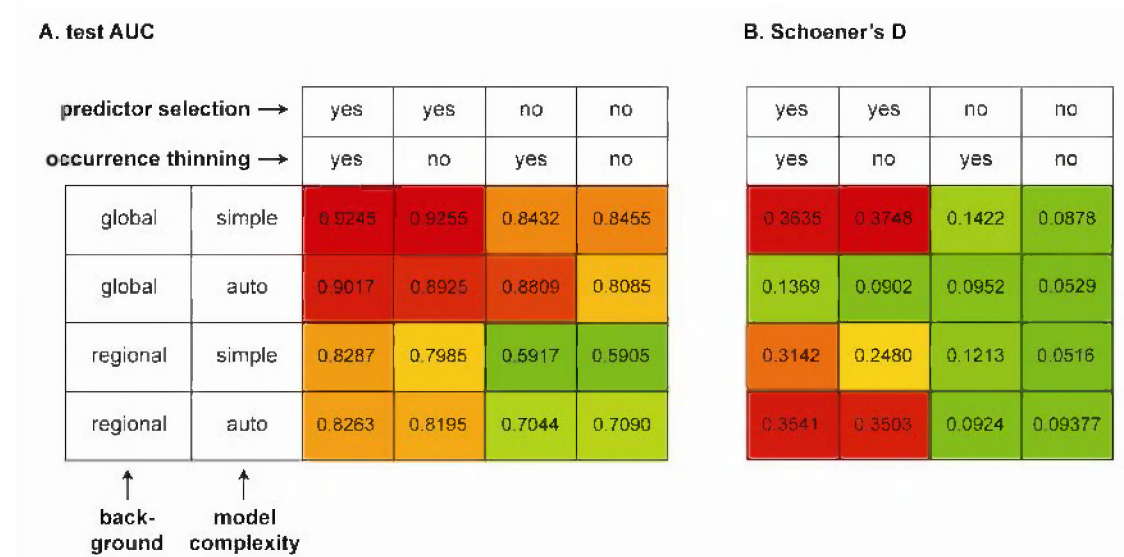


Figure 3: Impact of modeling choices on the transferability of SDMs. The transferability of models is approximated by test AUC (panel A) and the global niche overlap (Schoener's D, panel B). Columns and rows represent the combinations of the four factors that were varied in our experimental design and are identical in both panels. The values are also plotted as colors along a color gradient to permit rapid visual assessment of the important factors, with warmer colors indicating higher values. Each AUC value in panel A represent the average of the  $AUC_{\text{native-invaded}}$  and  $AUC_{\text{invaded-native}}$  for the corresponding condition.

## OVERALL PREDICTIVE PERFORMANCE OF SDMS

Models built with occurrences from throughout the native and invaded ranges have considerably higher predictive power than models trained on one range and projected onto the other (WSRT,  $p = 0.0156$  and  $0.0078$  for  $AUC_{\text{global}}$  vs.  $AUC_{\text{native} \rightarrow \text{Invaded}}$  and  $AUC_{\text{global}}$  vs.  $AUC_{\text{invaded} \rightarrow \text{native}}$  respectively,  $N = 8$ , for pairs with global background only). These models' test AUC values, calculated on 50% random test occurrences from throughout the range, are all close to 1, so these models nearly perfectly predict the distribution of the species in question (Table 1). The predictive performance of models based on pooled occurrences from native and invaded regions barely differ between conditions, indicating that models built with occurrences from both ranges are less sensitive to choices made during the modeling process (Table 1).

**Table 1: Predictive performance of models built with occurrences from native and invaded ranges as a function of choices made in the modeling process. The overall predictive performance, as measured by the test AUC, is very high and the factors have only a minor influence on the outcome.**

Occurrence thinning	Predictor selection	Model complexity	Performance (AUC)
yes	no	simple	0.975
yes	no	auto	0.990
yes	yes	simple	0.982
yes	yes	auto	0.988
no	no	simple	0.972
no	no	auto	0.991
no	yes	simple	0.974
no	yes	auto	0.992

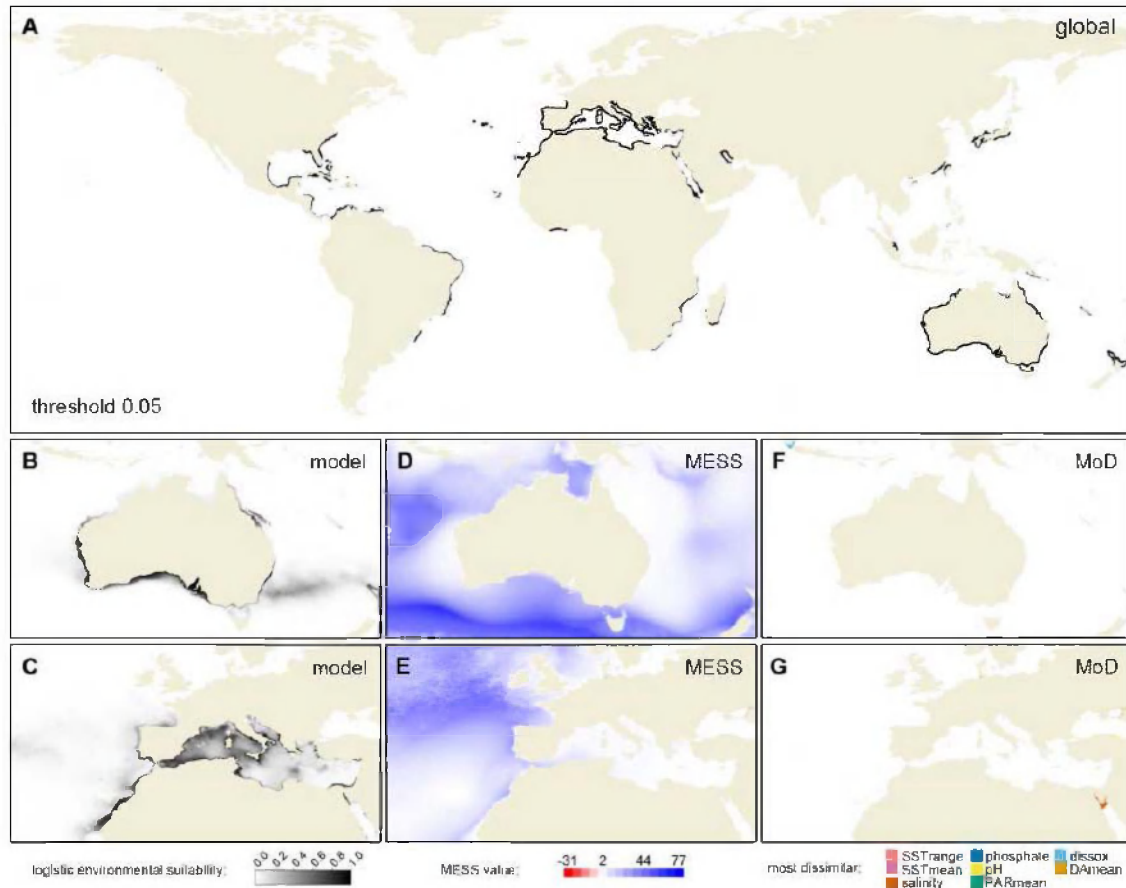
## AN SDM FOR CAULERPA CYLINDRACEA

The various SDMs with high predictive power were visually similar, and we present habitat suitability maps of one of the top-scoring models in Figure 4 and other outputs of the model as online supplementary material. The global map, which uses a threshold to indicate predicted suitable areas, clearly highlights large parts of the coasts of Australia (native region) and the Mediterranean Sea (invaded region) as having suitable macroecological conditions. In addition, the model predicts suitable habitat along the East coast of the USA, parts of the Caribbean region, the tropical to warm-temperate coast of Brazil, parts of the coasts of Madagascar and Southeast Africa, as well as Taiwan and the main Japanese islands.

Within the native region (Australia, Figure 4B), the model predicts suitable macroecological conditions along almost the entire coast of Southern Australia, including northern Tasmania, the West and East Australian coasts except for a region in SE Queensland, and parts of the North coast, where some regions had intermediate predicted suitability. These predictions are a considerable extension of the presently known range of the species (Figure 1C), and high habitat suitability is predicted in the various embayments of South Australia where the species has recently established and shows invasive behavior. The multivariate environmental similarity surface (MESS) map is positive in almost the entire range (Figure 4D, blue colors), indicating that the conditions present in



the region were observed in the training data, and giving extra credibility to the model predictions. Given that the MESS map is mostly positive the "most dissimilar" (MoD) variable map is nearly blank (Figure 4F).



**Figure 4: Species distribution model for *Caulerpa cylindracea*.** Panel A shows global areas predicted to have suitable macroecological conditions for the species. This map uses a threshold for MaxEnt's logistic suitability corresponding to the 10% training presences (threshold = 0.050) and predictions are plotted only for coastal areas (less than 7 pixels from shore), with predictions in the open ocean masked. Panels B and C show the continuous logistic model output for the native and invaded ranges, respectively. The corresponding multivariate environmental similarity surface (MESS) maps are shown in panels D and E, and the most dissimilar (MoD) variables in those areas that require extrapolation are shown in panels F and G.

In the invaded region (Figure 4C), the model also predicted beyond the known occurrences of the species (Figure 1A), including the West coast of Portugal, the NW of Spain and the NW of Africa. In the East, suitable macroecological conditions were inferred for the Northern Red Sea, although the MESS map indicates that there is extrapolation beyond observed environmental conditions there (Figure 4E), with the MoD map highlighting the (high) salinity occurring in the Northern Red Sea as the most dissimilar variable.

# DISCUSSION

Our results have implications for the invasion biology of *Caulerpa cylindracea* as well as the more general question of how best to model the distribution of species introduced outside their native range. Before discussing the meaning of our SDMs for the spread of this highly invasive seaweed species in Australia and Europe, we will highlight the effects of the different distribution modeling practices on model transferability and performance, as well as some limitations of the procedures described here.

## BUILDING MORE RELIABLE SDMs OF INTRODUCED SPECIES

Niche conservatism is a central assumption when extrapolating correlative SDMs of introduced species to an area outside the bounds of training occurrences. The poor predictive power of SDMs trained in the native range and projected onto the invaded range that was observed in many studies led to the conclusions that ecological niches can shift in association with introductions outside of the native range (e.g. Broennimann *et al.*, 2007; Medley, 2010; Rödder & Lötters, 2009; Rödder *et al.*, 2009; Fitzpatrick *et al.*, 2007). In interpreting such niche shifts, it is important to realize that correlative models estimate a species' realized niche and that, as a consequence, observed niche shifts do not necessarily reflect physiological changes (i.e., modifications of the fundamental niche). In other words, the perceived niche shift can result from two different realizations of the same fundamental niche in different areas, and it has been argued that this scenario is more parsimonious than that in which the fundamental niche changes (Soberon & Peterson, 2011; Rödder & Lötters, 2010). Even though this may be the rule, changes in the fundamental niche of introduced species are certainly possible (Jimenez-Valverde *et al.*, 2011; Whitney & Gabler, 2008).

Regardless of whether niche shifts observed in correlative SDMs are a consequence of changes in the realized or fundamental niche, it would be useful to have a set of procedures that can improve the predictive power of SDMs outside the training range to inform conservation planning and decision making. The methods used here were applied aiming at improving the transferability of the SDMs of introduced species built using the popular presence-only method MaxEnt. We found that meticulous selection of predictor variables drastically improved the transferability of our SDMs, with much smaller effects of limiting the model complexity, reducing geographic sampling bias by occurrence thinning and choosing a global background.

The effect of the choice of predictors has long been known to have a drastic effect on the transferability of SDMs of introduced species (e.g. Rödder & Lötters, 2009; Rödder *et al.*, 2009; Rödder & Lötters, 2010; Peterson & Nakazawa, 2008; Jimenez-Valverde *et al.*, 2011). The method we have used, which surveys all combinations of variables for the native as well as the invaded region, attempts to identify variables that are likely to be of global rather than regional significance. Models based on the set of variables identified by this approach were more transferable than models with a more comprehensive set of variables, irrespective of whether reciprocal test AUC or Schoener's D were used to measure transferability. The use of procedures to select predictors and model complexity in a more or less automated manner is common practice in many other types of modeling (Burnham & Anderson, 2002; Hastie *et al.*, 2009) but to our knowledge such approaches have not been used in combination with MaxEnt. It is clear that such predictor selection methods are no substitute for physiological knowledge of the organism (Araujo & Guisan, 2006), and here



they were used to further refine a set of predictors that was already reduced from the full Bio-ORACLE dataset based on what we know are important factors determining algal growth.

Previous studies have also shown that reducing the complexity of models to fit smoother responses yields the best correspondence to physiological knowledge and improves the overall performance and transferability of SDMs (Elith *et al.*, 2010; Warren & Seifert, 2010; Reineking & Schroder, 2006; Rodda *et al.*, 2011), and the use of simple environmental response surfaces to avoid overfitting has been recommended for SDMs of invasive species (Jimenez-Valverde *et al.*, 2011; Elith *et al.*, 2010; Rodda *et al.*, 2011). We chose to use a simple dichotomy (between MaxEnt's auto-features vs. the use of only linear and quadratic features) to keep the experimental setup simple. Our results did not show a meaningful difference between the transferability of models built under both conditions. The reason that we have not observed improvements of our models with simpler models while other studies have (Elith *et al.*, 2010; Warren & Seifert, 2010), can probably be attributed to the fact that for the number of samples in our training sets (25 in native area, 48 in invaded area), the model complexity does not differ much between both conditions (auto-features only adds hinge features). Nonetheless, we follow these previous authors in their conclusion that correlative models with smooth responses will generally outperform those with complex responses, especially if the number of occurrence points used to build models, and hence the potential for overfitting, is large. Furthermore, since the identification of suitable predictors and an appropriate level of model complexity are related to one another, it may be advisable to integrate these two into a single procedure as is done in classical model selection procedures (Burnham & Anderson, 2002), although this procedure remains to be explored.

The use of thinned occurrences, which should reduce potential effects of geographical sampling bias, generally resulted in SDMs with better transferability, but the effect was small compared to that obtained from predictor selection and not significant in a Wilcoxon signed-rank test. Nonetheless, we anticipate that this approach may be useful in situations where the geographical bias is stronger as in our dataset or in situations with stronger spatial autocorrelation in the environmental grids. Other approaches that have been proposed to deal with geographic bias in occurrence records are to introduce the same sort of bias in the background points by specifying a target-group background or by using bias grids in MaxEnt (Phillips *et al.*, 2009; Elith *et al.*, 2010). In our case study, the background selection also had a rather limited effect on the transferability of SDMs. In this context, it is worth noting that there were differences between the transferability results depending on whether they were measured as test AUC or as Schoener's D. The difference was most pronounced for models with global backgrounds and auto-features where test AUC values were among the largest observed (second row in Figure 3A) while Schoener's D was among the lowest observed (second row in Figure 3B). This indicates that, for those conditions, the models have good predictive power within the native and invaded ranges, but differ strongly from one another outside these regions. Whether this is a problem clearly depends on the purpose of the model that is being built. If it is intended to be used within the native and/or invaded regions, a model with a high test AUC should yield accurate predictions. If, on the other hand, it is intended to identify other areas where the species may be able to establish a successful population if introduced there, a model with better Schoener's D can be expected to give more reliable forecasts. Regarding the transferability of models as a function of the background selection, a previous study concluded that using background in reachable areas provides a "less risky prediction space" (Elith *et al.*, 2010). Based on the results for test AUC, it would appear that the opposite is true in our case study, but the difference is minor and contradicted by the Schoener's D results.

From the results discussed above it is clear that the usefulness (i.e., the predictive power) of reciprocal niche models is quite variable and strongly depends on the choices made. While they barely outperform random models under some conditions (some test AUC < 0.6 in Figure 3A), making the right choices outlined above improves the predictive power of models trained in one



range and projected onto the other ( $0.90 < \text{test AUC} < 0.93$  for the best models, Figure 3A). Nevertheless, if distribution data are available from both the native and invaded ranges, it is advisable to build models from a combined set of occurrences. For our data, models based on combined occurrences outperformed reciprocal models (test AUC  $> 0.99$  for best models). Similar conclusions regarding the use of combined native-invaded datasets were reached in studies of other species (e.g. Broennimann & Guisan, 2008; Jimenez-Valverde *et al.*, 2011). In addition to yielding better predictions, the combined data have the advantage of being relatively insensitive to the modeling choices that need to be made.

## POTENTIAL LIMITATIONS

Besides discussing the performance of the various methods applied, it is also useful to point out their assumptions and potential caveats.

First, our case study had the advantage of having relatively large sets of occurrence records for the native as well as the invaded range. In many cases, however, one will want to build reliable predictive models for species that were recently introduced and for which only a few occurrences have been recorded in the invaded range. How could a suitable set of predictors be identified in this case? Our approach relied on having sufficient data to identify those variables with predictive power in both geographic regions separately and combined. As an alternative, one could first identify the predictors achieving predictive power in the native range and subsequently compare the frequency distribution of those variables between samples from the native and invaded ranges and avoiding variables for which the invaded samples are outside of the range of values of native samples. In the model-building step, it may be beneficial to upweight the scarce samples from the invaded range. The evaluation of these ideas as well as other possible approaches is an attractive avenue for further research.

Our general approach towards increasing the transferability of SDM does not make explicit assumptions about whether or not a niche shift between ranges is present, or if it is, whether it is situated at the level of the fundamental or the realized niche. The ideal scenario is that there are no niche shifts between the populations and transferability is not an issue. If a niche shift is present, our predictor reduction approach will eliminate those predictors that have poor predictive power in one or both ranges, regardless of whether any changes in predictive power between regions are due to differences in the realized or fundamental niche. While we expect that eliminating predictors that have regional rather than general relevance will be sound in a majority of cases, there are scenarios imaginable where this will not work. For example, if the correlation structure of predictor variables differs between regions, an indirect variable (i.e. one that does not affect the distribution but is correlated with another one that does affect it) may be identified as important in both regions but have very different response curves in both areas, leading to poor transferability. Similarly, variables that are directly relevant to the distribution may differ systematically between regions, decreasing the transferability of the SDMs built from them (Soberon & Peterson, 2011).

Even though it can be expected that the distance-based thinning will improve most models, this may not always be the case. In fact, this procedure may discard useful data when regions of dense sampling coincide with steep ecological gradients over short geographic distances. This will of course depend on the specific dataset being studied.

Finally, our evaluation of methods is based on a single case study, and there are no guarantees that our results will extrapolate to other introduced species. A logical next step is to apply these methods to a range of suitable case studies. The time since the introduction and dispersal potential of the species should be prime criteria in selecting species to further test these methods. Species

that were introduced a long time ago have had the chance to disperse widely in the new region they invaded. This implies that they are more likely to inhabit a substantial portion of their potential niche in the new region. As a consequence, such species are more frequently in equilibrium with the environment in the new region and their observed range limits in the new region will better correspond to their environmental limits.

An additional approach towards testing the generalizability of the methods, as well as to explore the various other questions raised in the discussion, is to carry out simulation experiments. Simulation is a powerful tool for testing the logical consistency of ideas as well as the efficiency and reliability of methods. They have not been widely used to evaluate presence-only SDM methods, although there appears to be a trend towards their increased use in recent years (Santika & Hutchinson, 2009; Reineking & Schroder, 2006; Soberon & Peterson, 2011; Rödder & Engler, 2011; Vaclavik & Meentemeyer, 2012; Barve *et al.*, 2011; Meynard & Kaplan, 2012). Besides identifying the circumstances in which niche modeling algorithms perform well and those in which they are more likely to fail, simulation is a powerful tool to assess the effectiveness of procedures such as those described here and elsewhere. Such insights would obviously benefit the whole SDM field.

## INVASION AND SPREAD OF *CAULERPA CYLINDRACEA*

The distribution model presented for *C. cylindracea* predicted potential expansions in the invaded range along East Atlantic coastlines as well as a substantial potential expansion along the South coast of Australia (Figure 4A). Admittedly, the logistic values in MaxEnt lack a clear-cut interpretation (Royle *et al.*, 2012) and determining thresholds for presence-only SDMs is not an exact science (Bean *et al.*, 2012; Liu *et al.*, 2005). Based on the relatively conservative threshold (the 10-percentile training presence), the inferred range boundaries are quite far beyond the known occurrences of the species (e.g., East Atlantic, South Australia). This suggests that our current knowledge may underestimate the potential range of this species in these areas. In the Mediterranean and East Atlantic region, the species has only been present for about 20 years and, despite the species' relatively rapid colonization rate (Mineur *et al.*, 2010), it is likely that it has not reached its distributional limits yet. In Australia, the native area of the species, the species was known best from the West coast (Womersley, 1984). However, the recent observations of invasive populations of this species along the South coast where it did not previously occur (reference (Collings *et al.*, 2004) and pers. obs.), prompted us to generate SDMs for this species and investigate whether the species would be likely to colonize more of the coast. Our models do indeed suggest that the macroecological conditions are highly favorable and that this species could colonize the entire South coast of the country. Besides these potential expansions in regions where the species is present already, several other coastlines worldwide are predicted to be suitable habitat where the species could establish if it were to be introduced.

Needless to say our models only incorporate macroecological predictor variables, and suitable microhabitat as well as possible biotic interactions also need to be favorable for the species to establish itself in the areas that are predicted as suitable. In the native range it is usually found on rocky substrata close to the low-tide mark, although in more tropical locations (NW Australia and Great Barrier Reef) it is typically found growing on sand in lagoons and bays. In the Mediterranean Sea, it has been found between 1 and 60 meters depth, on all types of hard and soft substrata and in different communities, with the only exception being unstable sandy substrata (Verlaque *et al.*, 2003). A number of studies have studied the microhabitat preferences of the species in some detail in the Mediterranean, showing that it thrives on rocky substrata among other macroalgae as well as in dead seagrass mattes (Katsanevakis *et al.*, 2010a; Infantes *et al.*, 2011; Katsanevakis *et al.*, 2010b), and that they tolerate near-bottom orbital velocities below 15 cm s<sup>-1</sup> (Infantes *et al.*, 2011).



In summary, the species occurs in a wide range of common microhabitats, so it is likely that it could establish in the great majority of areas predicted by our SDM.

The correlative model from this study can also be used to inform experimental studies on the physiological tolerances of *C. cylindracea*. Even though we have not shown or discussed detailed response curves, most correspond to our expectations based on physiological knowledge of other algae, including other *Caulerpa* species (Theil *et al.*, 2007). For example, the correlative model indicates that the species is mainly found in phosphate-poor waters, the response curve rapidly dropping at concentrations over  $0.4 \mu\text{mol L}^{-1}$ . Studies on other species indicate that macroalgae have an increasing response curve for macronutrients and that low rather than high concentrations may be limiting seaweed species in nature (Hanisak, 1979; Lobban & Harrison, 1994; Lapointe & Bedford, 2010; Maloney *et al.*, 2011). To further characterize the most relevant features determining the species' range, it would be informative to evaluate the gradients of predictors occurring across the inferred range boundaries and put those to the test in physiological experiments.

## CONCLUSIONS

In order for MaxEnt presence-only SDMs to be useful in predicting and managing introduced and invasive species, some problems related to their accuracy and transferability have to be overcome. The methods introduced, explored and evaluated here aim to improve the situation. Especially reducing the set of predictors to those anticipated to be of global significance resulted in a strong improvement of SDM transferability, with occurrence thinning, model complexity and background choice having relatively minor effects. If available, occurrences from the native and invaded regions should be combined, as this yields the best-performing models and reduces their sensitivity to choices made in the modeling process. We also presented an SDM of the highly invasive species *Caulerpa cylindracea* that achieves very high predictive power. The procedures introduced here are available for further evaluation with other case studies and in simulation studies, which will provide further insights into how generalizable our results are. We hope and anticipate that they will form a useful strategy to improve predictive SDMs and inform environmental decision makers.



## ACKNOWLEDGEMENTS

We thank Thomas Sauvage and Claude Payri for sharing their work on New Caledonian *Caulerpa* and providing New Caledonian occurrence records of *C. cylindracea*. Funding was provided by the Australian Research Council (FT110100585 to HV), Research Foundation–Flanders and the AXA Research Fund (to FM). LT is a doctoral fellow of the Agency for Innovation by Science and Technology (IWT).

## AUTHOR CONTRIBUTIONS

HV and LT designed the study and analyzed the data. LT, GSB, FM and HV gathered occurrence records. HV and AJ wrote code. HV and GSB wrote the paper. All authors have read and edited the paper.

# REFERENCES

- Alexander, J. M. & Edwards, P. J. (2010) Limits to the niche and range margins of alien species. *Oikos*, 119, 1377-1386.
- Alongi, G., Cormaci, M., Furnari, G. & Giaccone, G. (1993) Prima segnalazione di *Caulerpa racemosa* (Chlorophyceae, Caulerpales) per le coste Italiane. *Bollettino dell'Accademia Gioenia di Scienze Naturali Catania*, 26, 9-53.
- Araujo, M. B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677-1688.
- Baldacconi, R. & Corriero, G. (2009) Effects of the spread of the alga *Caulerpa racemosa* var. *cylindracea* on the sponge assemblage from coralligenous concretions of the Apulian coast (Ionian Sea, Italy). *Marine Ecology*, 30, 337-345.
- Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberon, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222, 1810-1819.
- Bean, W. T., Stafford, R. & Brashares, J. S. (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, 35, 250-258.
- Belton, G. S., Huisman, J. M., Prud'homme Van Reine, W. F., Draisma, S. G. A., Sauvage, T. & D., G. C. F. (2012) Species delimitation in the *Caulerpa racemosa* complex (Chlorophyta, Caulerpales). *Journal of Phycology*, in review.
- Broennimann, O. & Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters*, 4, 585-589.
- Broennimann, O., Treier, U. A., Muller-Scharer, H., Thuiller, W., Peterson, A. T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, 10, 701-709.
- Buia, M. C., Petrocelli, A. & Saracino, O. D. (1998) *Caulerpa racemosa* spread in the Mediterranean Sea: first record in the Gulf of Taranto. *Biologie Marine Méditerranée*, 5, 527-529.
- Burnham, K. P. & Anderson, D. R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, edn. Springer, New York.
- Ceccherelli, G. & Campo, D. (2002) Different effects of *Caulerpa racemosa* on two co-occurring seagrasses in the Mediterranean. *Botanica Marina*, 45, 71-76.
- Collings, G., Westphalen, G., Rowling, K. & Eglinton, Y. (2004) *Caulerpa racemosa* var. *cylindracea* occurrence in western South Australia. pp Report to PIRSA Marine Habitat Program. South Australian Research and Development Institute (Aquatic Sciences) Report Number RD04/0169.
- Cossu, A. & Gazale, V. (1997) Sulla presenza di *Caulerpa racemosa* (Forsskål) J. Agardh in Sardegna. *SOS Caulerpa? Introduzione di nuove specie nel Mediterraneo e compatibilità con quelle presenti* (ed. by A. Cossu & M.M. Meloni), pp 87-97. Poseidon, Cagliari.
- Djellouli, A. S. (2000) *Caulerpa racemosa* (Forskaal) J. Agardh en Tunisie. *Proceedings of the First Mediterranean Symposium on Marine Vegetation* (ed. by P.P. Racspa), pp 124-127, Ajaccio.
- Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kuhn, I., Ohlemuller, R., Peres-Neto, P. R., Reineking, B., Schroder, B., Schurr, F. M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609-628.
- Durand, C., Manuel, M., Boudouresque, C. F., Meinesz, A., Verlaque, M. & Le Parco, Y. (2002) Molecular data suggest a hybrid origin for the invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta) in the Mediterranean Sea. *Journal of Evolutionary Biology*, 15, 122-133.
- Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330-342.
- 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, 677-697.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43-57.
- Fitzpatrick, M. C., Weltzin, J. F., Sanders, N. J. & Dunn, R. R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, 16, 24-33.
- Franklin, J. & Miller, J. A. (2009) *Mapping species distributions*, edn. Cambridge University Press, Cambridge.



- Gambi, M. C. & Terlizzi, A. (1998) Record of large population of *Caulerpa racemosa* (Forsskål) J. Agardh (Chlorophyceae) in the gulf of Salerno (Southern Tyrrhenian Sea, Italy). *Biologie Marine Méditerranée*, 5, 553-556.
- Giaccone, G. & Di Martino, V. (1995) Le Caulerpe in Mediterraneo: un ritorno del vecchio bacino Tetide verso il dominio Indo-Pacifico. *Biologia Marina Mediterranea*, 2, 607-612.
- Hanisak, M. D. (1979) Nitrogen limitation of *Codium fragile* ssp. *tomentosoides* as determined by tissue analysis. *Marine Biology*, 50, 333-337.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The elements of statistical learning*, edn. Springer, New York.
- Huisman, J. M. (2000) *Marine plants of Australia*, edn. University of Western Australia Press, Nedlands, Western Australia.
- Infantes, E., Terrados, J. & Orfila, A. (2011) Assessment of substratum effect on the distribution of two invasive *Caulerpa* (Chlorophyta) species. *Estuarine Coastal and Shelf Science*, 91, 434-441.
- Jimenez-Valverde, A., Peterson, A. T., Soberon, J., Overton, J. M., Aragon, P. & Lobo, J. M. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, 13, 2785-2797.
- Jousson, O., Pawlowski, J., Zaninetti, L., Meinesz, A. & Boudouresque, C. F. (1998) Molecular evidence for the aquarium origin of the green alga *Caulerpa taxifolia* introduced to the Mediterranean Sea. *Marine Ecology Progress Series*, 172, 275-280.
- Katsanevakis, S., Issaris, Y., Poursanidis, D. & Thessalou-Legaki, M. (2010a) Vulnerability of marine habitats to the invasive green alga *Caulerpa racemosa* var. *cylindracea* within a marine protected area. *Marine Environmental Research*, 70, 210-218.
- Katsanevakis, S., Salomidi, M. & Panou, A. (2010b) Modelling distribution patterns and habitat preference of the invasive green alga *Caulerpa racemosa* in the Saronikos Gulf (Eastern Mediterranean). *Aquatic Biology*, 10, 57-67.
- Klein, J. (2007) Impact of *Caulerpa racemosa* var. *cylindracea* (Caulerpales, Chlorophyta) on macrophyte assemblages of the north-western Mediterranean Sea. pp 315. University of Aix-Marseille II, Marseille.
- Klein, J. & Verlaque, M. (2008) The *Caulerpa racemosa* invasion: A critical review. *Marine Pollution Bulletin*, 56, 205-225.
- Kruzic, P., Zuljevic, A. & Nikolic, V. (2008) The highly invasive alga *Caulerpa racemosa* var. *cylindracea* poses a new threat to the banks of the coral *Cladocora caespitosa* in the Adriatic Sea. *Coral Reefs*, 27, 441.
- Lapointe, B. E. & Bedford, B. J. (2010) Ecology and nutrition of invasive *Caulerpa brachypus* f. *parvifolia* blooms on coral reefs off southeast Florida, USA. *Harmful Algae*, 9, 1-12.
- Leathwick, J. R. & Austin, M. P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, 82, 2560-2573.
- Liu, C. R., Berry, P. M., Dawson, T. P. & Pearson, R. G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.
- Lobban, C. S. & Harrison, P. J. (1994) *Seaweed ecology and physiology*, edn. Cambridge University Press, Cambridge.
- Lüning, K. (1990) *Seaweeds: their environment, biogeography, and ecophysiology*, edn. Wiley & Sons, New York.
- Maloney, B., Iliffe, T. M., Gelwick, F. & Quigg, A. (2011) Effect of nutrient enrichment on naturally occurring macroalgal species in six cave pools in Bermuda. *Phycologia*, 50, 132-143.
- Medley, K. A. (2010) Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. *Global Ecology and Biogeography*, 19, 122-133.
- Meynard, C. N. & Kaplan, D. M. (2012) The effect of a gradual response to the environment on species distribution modeling performance. *Ecography*, 35, 499-509.
- Mineur, F., Davies, A. J., Maggs, C. A., Verlaque, M. & Johnson, M. P. (2010) Fronts, jumps and secondary introductions suggested as different invasion patterns in marine species, with an increase in spread rates over time. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2693-2701.
- Nizamuddin, M. (1991) *The green marine algae of Lybia*, edn. Alga Publications.
- Nuber, N., Gornik, O., Lauc, G., Bauer, N., Zuljević, A., Papes, D. & V., Z. (2007) Genetic evidence for the identity of *Caulerpa racemosa* (Forsskål) J. Agardh (Caulerpales, Chlorophyta) in the Adriatic Sea. *European Journal of Phycology*, 42, 113-120.
- Pena-Martín, C., Cristóbal-Fernández, J. C., Crespo, B. & M., S.-P. (2003) *Caulerpa racemosa* (Forssk.) J. Agardh (Caulerpales, Chlorophyceae), nueva para la flora de Alicante. *Anales Jardín Botánico de Madrid*, 60, 448-449.
- Peterson, A. T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, 78, 419-433.

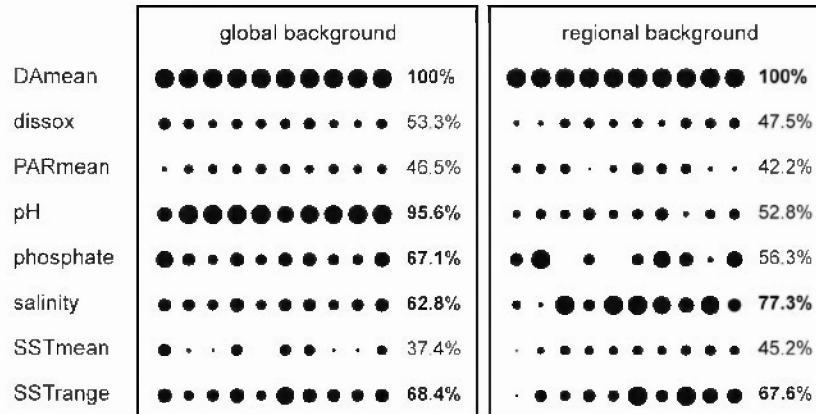


- Peterson, A. T. & Nakazawa, Y. (2008) Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography*, 17, 135-144.
- Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.
- Phillips, S. J. & Dudik, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.
- Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19, 181-197.
- Piazzi, L. & Balata, D. (2009) Invasion of alien macroalgae in different Mediterranean habitats. *Biological Invasions*, 11, 193-204.
- Piazzi, L., Balestri, E. & Cinelli, F. (1994) Presence of *Caulerpa racemosa* in northwestern Mediterranean. *Cryptogamie Algologie*, 15, 183-189.
- Piazzi, L., Balestri, E., Magri, M. & Cinelli, F. (1997) Expansion de l'algue tropicale *Caulerpa racemosa* (Forsskal) J. Agardh (Bryopsidophyceae, Chlorophyta) le long de la côte toscane (Italie). *Cryptogamie Algologie*, 18, 343-350.
- Piazzi, L. & Ceccarelli, D. (2006) Persistence of biological invasion effects: recovery of macroalgal assemblages after removal of *Caulerpa racemosa* var. *cylindracea*. *Estuarine, Coastal and Shelf Science*, 68, 455-461.
- Piazzi, L., Ceccarelli, D. & Cinelli, F. (2001) Threat to macroalgal diversity: effects of the introduced green alga *Caulerpa racemosa* in the Mediterranean. *Marine Ecology Progress Series*, 210, 149-159.
- Piazzi, L., Meinesz, A., Verlaque, M., Akcali, B., Antolic, B., Argyrou, M., Balata, D., Ballesteros, E., Calvo, S., Cinelli, F., Cirik, S., Cossu, A., D'archino, F., Djellouli, A. S., Javel, F., Lanfranco, E., Mifsud, C., Pala, D., Panayotidis, P., Peirano, A., Pergent, G., Petrocelli, A., Ruitton, S., Zuljevic, A. & Ceccherelli, G. (2005) Invasion of *Caulerpa racemosa* var. *cylindracea* (Caulerpales, Chlorophyta) in the Mediterranean Sea: an assessment of the spread. *Cryptogamie Algologie*, 26, 189-202.
- Randin, C. F., Dirnbock, T., Dullinger, S., Zimmermann, N. E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689-1703.
- Reineking, B. & Schroder, B. (2006) Constrain to perform: Regularization of habitat models. *Ecological Modelling*, 193, 675-690.
- Ripley, B. (2011) KernSmooth.
- Rodda, G. H., Jarnevich, C. S. & Reed, R. N. (2011) Challenges in identifying sites climatically matched to the native ranges of animal invaders. *PLoS One*, 6, e14670.
- Rödger, D. & Engler, J. O. (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography*, 20, 915-927.
- Rödger, D. & Lötters, S. (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography*, 18, 674-687.
- Rödger, D. & Lötters, S. (2010) Explanative power of variables used in species distribution modelling: An issue of general model transferability or niche shift in the invasive greenhouse frog (*Eleutherodactylus planirostris*)? *Naturwissenschaften*, 97, 781-796.
- Rödger, D., Schmidtlein, S., Veith, M. & Lötters, S. (2009) Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? *PLoS One*, 4, 9.
- 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 in Ecology and Evolution*, in press.
- Santika, T. & Hutchinson, M. F. (2009) The effect of species response form on species distribution model prediction and inference. *Ecological Modelling*, 220, 2365-2379.
- Sauvage, T., Payri, C., Draisma, S. G. A., Prud'homme Van Reine, W. F., Verbruggen, H., Belton, G. S., Gurgel, C. F. D., Gabriel, D., Sherwood, A. R. & Fredericq, S. (2012) Molecular diversity of the *Caulerpa racemosa-peltata* complex (Caulerpales, Chlorophyta) in New Caledonia, with new Australasian records for the variety *cylindracea*. *Phycologia*, provisionally accepted.
- Serio, D. & Pizzuto, F. (1999) Studio di un prato a *Caulerpa racemosa* (Forsskal) J. Agardh (Caulerpales, Chlorophyta) di Brocoli (SR) con osservazioni in coltura della specie. *Boll. Accad. Gioenia Sci. Nat. Catania*, 31, 201-209.
- Soberon, J. & Peterson, A. T. (2011) Ecological niche shifts, and environmental space anisotropy: a cautionary note. *Revista Mexicana de Biodiversidad*, 82, 1348-1355.

- Sonder, G. (1845) Nova algarum genera et species, quas in itinere ad oras occidentales Novae Hollandiae, collegit L. Priess, Ph. Dr. *Botanische Zeitung*, 3, 49-57.
- Theil, M., Westphalen, G., Collings, G. & Cheshire, A. (2007) *Caulerpa taxifolia* responses to hyposalinity stress. *Aquatic Botany*, 87, 221-228.
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, 21, 272-281.
- Vaclavik, T. & Meentemeyer, R. K. (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions*, 18, 73-83.
- Verbruggen, H. (2011a) Maxent Model Surveyor 1.03. <http://phycoweb.net/software/MMS/index.html>.
- Verbruggen, H. (2011b) OccurrenceThinner 1.03. <http://phycoweb.net/software/OccurrenceThinner/index.html>.
- Verlaque, M., Afonso-Carillo, J., Gil-Rodriguez, M. C., Durand, C., Boudouresque, C. F. & Le Parco, Y. (2004) Blitzkrieg in a marine invasion: *Caulerpa racemosa* var. *cylindracea* (Bryopsidales, Chlorophyta) reaches the Canary islands (North-East Atlantic). *Biological Invasions*, 6, 269-281.
- Verlaque, M., Boudouresque, C. F., Meinesz, A. & Gravez, V. (2000) The *Caulerpa racemosa* complex (Caulerpales, Ulvophyceae) in the Mediterranean Sea. *Botanica Marina*, 43, 49-68.
- Verlaque, M., Durand, C., Huisman, J. M., Boudouresque, C. F. & Le Parco, Y. (2003) On the identity and origin of the Mediterranean invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta). *European Journal of Phycology*, 38, 325-339.
- Wand, M. P. & Jones, M. C. (1995) *Kernel smoothing*, edn. Chapman and Hall, London.
- Warren, D. L. & Seifert, S. N. (2010) Environmental niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335-342.
- Wenger, S. J. & Olden, J. D. (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3, 260-267.
- Whitney, K. D. & Gabler, C. A. (2008) Rapid evolution in introduced species, 'invasive traits' and recipient communities: challenges for predicting invasive potential. *Diversity and Distributions*, 14, 569-580.
- Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A. & Snyder, M. A. (2009) Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19729-19736.
- Wolmarans, R., Robertson, M. P. & Van Rensburg, B. J. (2010) Predicting invasive alien plant distributions: how geographical bias in occurrence records influences model performance. *Journal of Biogeography*, 37, 1797-1810.
- Womersley, H. B. S. (1984) *The marine benthic flora of southern Australia. Part I*, edn. Government Printer, South Australia.
- Zuljevic, A., Antolic, B. & Onofri, V. (2003) First record of *Caulerpa racemosa* (Caulerpales: Chlorophyta) in the Adriatic Sea. *Journal of the Marine Biological Association of the United Kingdom*, 83, 711-712.
- Zurell, D., Elith, J. & Schroder, B. (2012) Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions*, 18, 628-634.

# APPENDIX

## A. native range (Australia)



## B. invaded range (Europe)

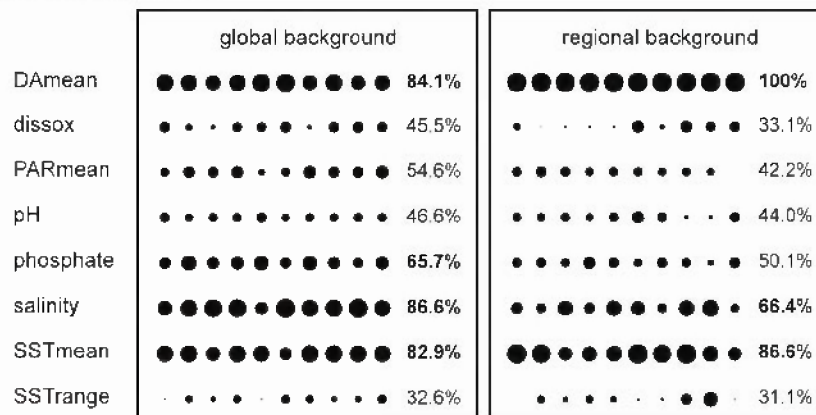


Figure S1: Additional results of the surveying procedure to identify the predictors present in top-scoring models. Boxes represent results of the predictor set survey for occurrence records from the native range (A) and the invaded European range (B) using global or regional background points. The ten columns within each box represent independent runs on different sets of thinned coordinates, and the circle diameter represents how often the variable in question occurs in the top 10 highest-scoring models (test AUC) for that set of occurrences. The representation of each predictor in the top 10 is also summarized across columns (percentage).