Improving Transferability of Introduced Species' Distribution Models: New Tools to Forecast the Spread of a Highly Invasive Seaweed

Heroen Verbruggen[1*], Lennert Tyberghein[2,3], Gareth S. Belton[4], Frederic Mineur[5], Alexander Jueterbock[6], Galice Hoarau[6], C. Frederico D. Gurgel[4,7,8], Olivier De Clerck[2]

1 School of Botany, University of Melbourne, Parkville, Victoria, Australia, 2 Phycology Research Group, Ghent University, Ghent, Belgium, 3 Flanders Marine Institute (VLIZ), Ostend, Belgium, 4 School of Earth and Environmental Sciences, University of Adelaide, Adelaide, South Australia, Australia, 5 Mediterranean Institute of Oceanography (MIO), Aix-Marseille University, Marseille, France, 6 Marine Ecology Group, University of Nordland, Bodø, Norway, 7 South Australia State Herbarium, Department of Environment Water and Natural Resources, Adelaide, South Australia, Australia, 8 South Australia Research and Development Institute (Aquatic Sciences), Henley Beach, South Australia, Australia

Abstract

The utility of species distribution models for applications in invasion and global change biology is critically dependent on their transferability between regions or points in time, respectively. We introduce two methods that aim to improve the transferability of presence-only models: density-based occurrence thinning and performance-based predictor selection. We evaluate the effect of these methods along with the impact of the choice of model complexity and geographic background on the transferability of a species distribution model between geographic regions. Our multifactorial experiment focuses on the notorious invasive seaweed *Caulerpa cylindracea* (previously *Caulerpa racemosa* var. *cylindracea*) and uses Maxent, a commonly used presence-only modeling technique. We show that model transferability is markedly improved by appropriate predictor selection, with occurrence thinning, model complexity and background choice having relatively minor effects. The data shows that, if available, occurrence records from the native and invaded regions should be combined as this leads to models with high predictive power while reducing the sensitivity to choices made in the modeling process. The inferred distribution model of *Caulerpa cylindracea* shows the potential for this species to further spread along the coasts of Western Europe, western Africa and the south coast of Australia.

Citation: Verbruggen H, Tyberghein L, Belton GS, Mineur F, Jueterbock A, et al. (2013) Improving Transferability of Introduced Species' Distribution Models: New Tools to Forecast the Spread of a Highly Invasive Seaweed. PLoS ONE 8(6): e68337. doi:10.1371/journal.pone.0068337

Editor: Mark R. Muldoon, Manchester University, United Kingdom

Received January 14, 2013; **Accepted** May 28, 2013; **Published** June 28, 2013

Funding: Funding was provided by the Australian Research Council (FT110100585 to HV, LP0882109 and LP0991083 to CFDG), Research Foundation–Flanders, the AXA Research Fund (to FM), the Adelaide & Mt Lofty Ranges Natural Resource Management Board, the Australian Census of Coral Reef Life and the Australia Biological Research Study (ABRS 209-62 to CFDG). LT is a doctoral fellow of the Agency for Innovation by Science and Technology (IWT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: Author Heroen Verbruggen is a PLoS One Editorial Board member. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: heroen.verbruggen@gmail.com (HV)

Introduction

Species distribution models (SDMs) help us understand and map species' distributions, play a key role in forecasting range expansion of introduced species and can help us predict the effects of climate change on species distributions [1-4]. An SDM characterizes the species' response to relevant environmental variables, using either physiological information from experimental work (mechanistic models) or by relating the presence and/or absence of the species to environmental information (correlative models) [5]. 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 approaches dominate species distribution modeling. Furthermore, it is quite troublesome to assess the absence of species from an area while species occurrence data are abundant in museum databases and the literature. As a consequence, most SDMs rely on presence-only techniques [1].

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 [6]. 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 distribution predictions as uncertain [7,8]. In addition, the ability of presence-only methods to capture a species' ecological response is affected by the choice of background points [7], predictor variables [9], model complexity [10,11] and the geographic spread of occurrence records in relation to environmental gradients [12,13]. Besides these problems, it is also possible that biotic interactions limit the utility of models based on abiotic predictors [14] and, of course, there is always the possibility that the fundamental niche of the introduced population has changed due to natural selection [15,16].

This study focuses on the choices made during the modeling process that affect the transferability and overall predictive performance of the resulting model. We introduce two new methods that have the potential to increase the transferability of correlative SDMs: density-based occurrence thinning and performance-based predictor selection. As a case study, we apply these to the highly invasive seaweed species, *Caulerpa cylindracea*, in order to assist in assessing the risk of further spreading as well as predicting areas with suitable environmental conditions worldwide.

Methods

Experimental Design

The overarching goal of the present study is to examine and improve the overall performance and the transferability between regions 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, leading to model misspecification [12,17] and issues related to its evaluation [18]. 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 with and without occurrence thinning are compared.

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 [9,19,20]. This has also led to the recognition of predictor variables as more conserved or relaxed, depending on whether they match between native and invaded species occurrences or not [9,21]. We introduce a method that surveys the performance of all possible predictor sets (explained below) and evaluate the transferability between regions 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 [10,22,23]. By default, Maxent determines the types of features it allows automatically, based on the number of samples available for model training [24], but this standard behavior has been reported to result in overfitted models [11]. 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 outcome of presence-only SDMs [7,25,26]. To examine this, we compare SDMs built with global background points to models built with a regional background.

Using *C. cylindracea* as a case study, model transferability was 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 occurrences from either range to that of models combining occurrences from both ranges.

Study Species and Environmental Data

This study focuses on the introduced and highly invasive seaweed species *Caulerpa cylindracea* Sonder [27]. 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 confusing nomenclature. Most of this confusion has existed around the *C. racemosa/peltata* complex that has more than 30 described varieties and forms [28]. 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 [29] and it is now due to be reinstated as an independent species [28].

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 [30,31]. The species has been reported from all kinds of substrata and depths, as part of a variety of benthic assemblages, and thrives in disturbed habitats of the heavily urbanized Mediterranean coastlines [30,32]. Invasive populations of *C. cylindracea* establish dense and compact monospecific stands, which easily overgrow and outcompete and/or negatively impact other seaweed [33,34], seagrass [35] and invertebrate species [36,37] leading to biotic homogenization [38] and an overall decrease of species diversity in affected areas [30]. To date only partial recovery of the assemblages could be observed after eradication of *C. cylindracea* in Italy and France [33,39].

Unlike *C. taxifolia*, which was accidentally introduced from a public aquarium [40], the vector of introduction of *C. cylindracea* to the Mediterranean Sea is unknown. It was initially hypothesized to be a Lessepssian immigrant [41,42], or a hybrid between *C. racemosa* var. *turbinata* and an unknown

tropical variety [43], until molecular investigations identified a potential source population in southwestern Australia [29]. 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 [44].

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 [29,45], and has more recently been reported from two locations on the southern coast of Australia (Adelaide, SA and Portland, VIC, e.g. references 46,47 and unpublished data GSB). As *C. cylindracea* is only found near shipping ports and had not been reported from this area prior to 2003 [48], it is most likely that this species is a recent introduction. The rapid spread of this species through the European invaded range makes it a suitable case study for the question at hand.

A total of 191 distribution records were assembled from the native range in and around Australia (65 records), the invaded range in Europe (111) and the recently invaded areas in southern Australia (15). The data sources for these records are: Australia Virtual Herbarium (http://chah.gov.au/avh/), new collections from Victoria by GSB deposited in the AD herbarium, the data gathered by FM for the ERC FP5 ALIENS project, and the literature [31,41,44,48-60]. The absence of the species in various DNA bar coding surveys of *Caulerpa* from some other parts of the Indo-Pacific (Philippines, Japan, Tanzania, Red Sea) suggests that the native range may be limited to Australia and some closeby locations (unpublished data: Stefano Draisma, Thomas Sauvage, Heroen Verbruggen).

We used the Bio-ORACLE dataset [61] as a source of marine environmental grids (90° N–90ºS, 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 and as a result, the dataset reduced to 95 distribution records.

Occurrence Thinning

Geographical biases in the occurrence records were dampened by thinning the distribution points with OccurrenceThinner 1.03 [62]. 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 [63,64], 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 three from the southern Australian invasive populations.

To evaluate whether occurrence thinning influences model transferability and performance, we compared Maxent models based on a thinned subset of samples with models using all occurrence records (but limited to one per cell as mentioned above).

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 [65], 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 (DAmean) as a measure of water transparency, dissolved oxygen (dissox) and the phosphate concentration. Nitrate concentration was not included because it is correlated with the phosphate concentration [61].

In the second step, the predictive ability of those eight variables was explored using Maxent Model Surveyor (MMS) version 1.03 [66]. We developed this software to evaluate the performance of all possible subsets of variables (2^8 - 1 = 255 for our eight predictors), using the test AUC (Area Under the receiver operating characteristic Curve) to measure model performance [67]. 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 worked from the thinned set of occurrences and restricted the model complexity to linear and quadratic features. Each of the five runs listed above was repeated ten times (i.e., on each of the ten 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 ranges. 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. We retained only those variables that were present in more than 60% of the top-scoring models for at least two out of three regions (native, Europe, combined, i.e. conditions 1, 2 and 3 described above). The 60% threshold criterion is essentially arbitrary – we chose it because it halved the number of predictor variabes from eight to four (specified in results). Retaining variables important in at least two regions was done because it would prefer variables of global, rather than regional, relevance.

In order to evaluate whether this predictor selection approach can improve the transferability of models across regions, Maxent models were run with all eight variables listed

above as well as the subset of four variables generated with the predictor selection procedure.

Model Complexity

Model complexity was varied to verify its impact on the predictive power and transferability. The first condition used the default behavior of Maxent (auto-features), which determines which features are used based on the number of samples [24]. The second condition 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 [61]. 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 western Europe extending to Africa and the Mediterranean Sea, between latitude 20° N and 60° N and longitude 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 combined samples from the native and invaded ranges always used the global background.

Niche Model Inference

Niche models were inferred with Maxent 3.3.3f [24,68,69]. 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, test and background points, were provided as SWD files and the resulting models were projected onto the Bio-ORACLE grid [61]. 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 other interpretations were based on the individual models.

Downstream Analyses

Models were compared to identify which choices lead to better-performing models. In order to evaluate the transferability of models, we compared models built on the native and invaded ranges in a pairwise fashion, using the Schoener's D niche similarity measure [70] and reciprocal test AUC (i.e. native training samples with test samples in invaded range and vice versa). The overall predictive power of models was compared with the test AUC, taking care to only compare models built with identical geographic background datasets.

Results

Exploration of new methods

We implemented two methods that tackle issues related to the overall quality and transferability of niche models. The first of these, 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).

The results of the second method, which surveyed all combinations 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 in Figure 2, 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.

The effect of these two methods on model performance was evaluated by including them as factors in our experimental design. So all Maxent analyses were run with all samples and thinned samples. Similarly, models were run with all eight 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 reducing the number of predictors, based on our 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 both 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, suggesting 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, $N = 8$). The higher AUC with global backgrounds may thus be a consequence of the sensitivity of AUC to background choice rather than an actual increase in predictive power with global backgrounds.

Model complexity and occurrence thinning did not have a large effect on transferability between regions. However, the

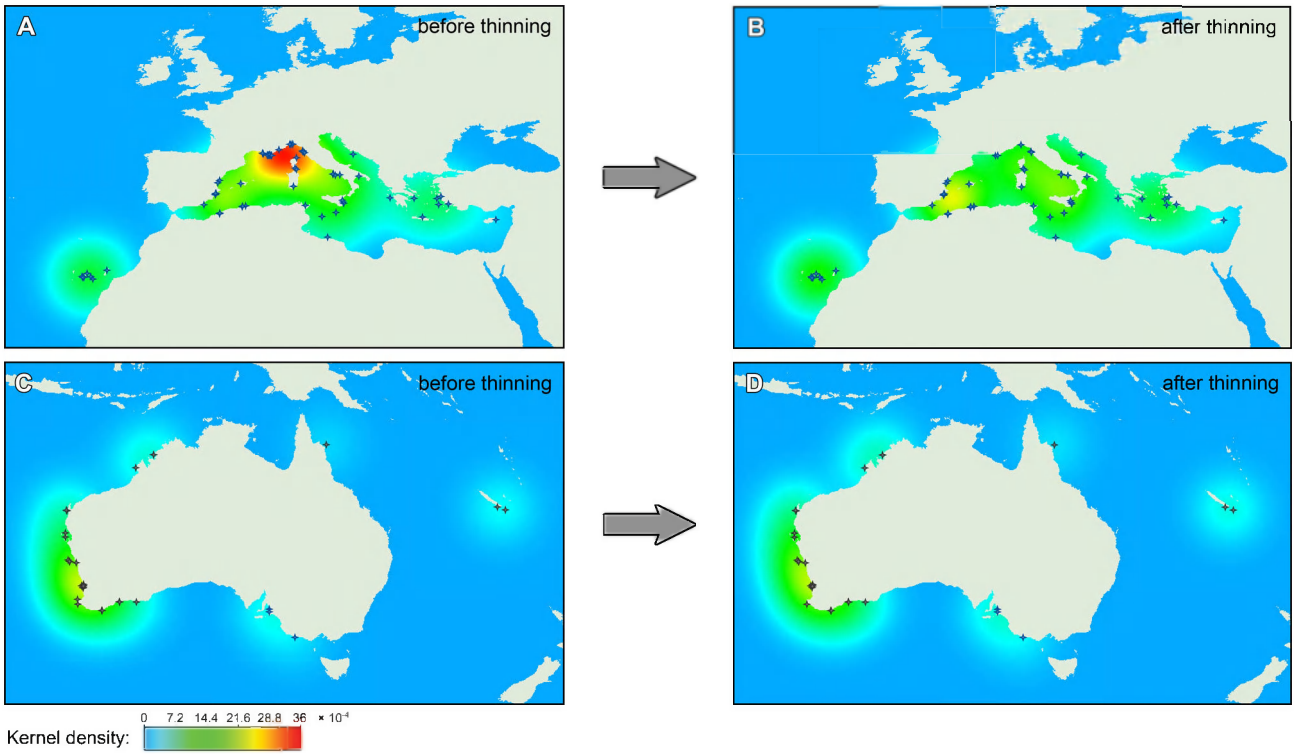

Figure 1. Effect of occurrence thinning on geographical sample bias. The colors on the map represent the regional sampling density, warmer colors indicating higher sample densities. Occurrence thinning substantially reduces the geographic sampling bias, as illustrated by the disappearance of the red 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). Note that the slightly elevated density close to the Spanish-French border in the Bay of Biscay (panels A and B) is caused by samples in the Mediterranean of which the kernel extends across land; there are no occurrences of *C. cylindracea* known from that area.

doi: 10.1371/journal.pone.0068337.g001

second row in Figure 3B 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). This difference was not present for the regional background case (3rd vs. 4th row).

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_{global} vs. $AUC_{native \rightarrow invaded}$ and AUC_{global} vs. $AUC_{invaded \rightarrow native}$ 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 (Table 1), indicating strong overall predictive power. 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).

An SDM for *Caulerpa Cylindracea*. The various SDMs with high predictive power were visually similar, and we present environmental suitability maps of one of the top-scoring models in Figure 4. 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 environmental conditions 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 environmental suitability is predicted in the various embayments of southern Australia where the species has recently established and become a

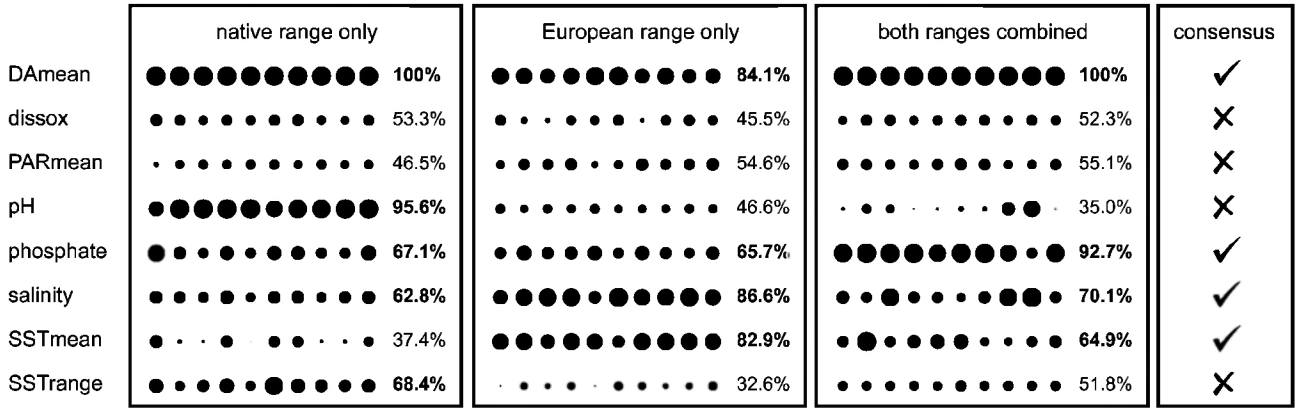

Figure 2. Results of the surveying procedure to identify the predictors present in top-scoring models. Each box contains the results of the survey for occurrence records from the native range, the invaded European range, or both ranges combined. Each column within a box represents a single survey carried out on one set of thinned coordinates. The circle diameter represents how often the variable in question occurred in the top 10 highest-scoring models (test AUC) for that set of thinned occurrences. The representation of each predictor in the top 10 is also summarized across columns (percentage indicates how many of the top 10 models had the predictor), and the consensus predictor set across ranges is indicated in the box on the right.

doi: 10.1371/journal.pone.0068337.g002

A. test AUC

predictor selection →		yes	yes	no	no
occurrence thinning →		yes	no	yes	no
global	simple	0.9245	0.9255	0.8432	0.8455
global	auto	0.9017	0.8925	0.8809	0.8085
regional	simple	0.8287	0.7985	0.5917	0.5905
regional	auto	0.8263	0.8195	0.7044	0.7090

↑ background ↑ model complexity

B. Schoener's D

yes	yes	no	no
yes	no	yes	no
0.3635	0.3748	0.1422	0.0878
0.1369	0.0902	0.0952	0.0529
0.3142	0.2480	0.1213	0.0516
0.3541	0.3503	0.0924	0.09377

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 $AUC_{native\text{-}invaded}$ and $AUC_{invaded\text{-}native}$ for the corresponding condition.

doi: 10.1371/journal.pone.0068337.g003

conspicuous member of the benthic community. The multivariate environmental similarity surface (MESS) map is positive in almost the entire range (Figure 4D, blue colors), which indicates that the conditions present in the region were observed in the training data and gives extra credibility to the model prediction. Given that the MESS map is mostly positive the "most dissimilar" (MoD) variable map is nearly blank (Figure 4F).

In the invaded region (Figure 4C), the model also predicted beyond the known occurrences of the species (Figure 1A),

Table 1. Predictive performance of models built with occurrences from native and invaded ranges as a function of choices made in the modeling process.

occurrence thinning	predictor selection	model complexity	performance (test 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

The overall predictive performance, as measured by the test AUC, is very high and the factors have only a minor influence on the outcome. All models compared in this table use the same set of 10,000 background points (global, equal area).

including 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 (Figure 4E), with the MoD map highlighting the (high) salinity occurring in the northern Red Sea as the most dissimilar variable.

The entire Maxent run including input data and all outputs is available for examination on FigShare (http://dx.doi.org/10.6084/m9.figshare.681723). Besides showing the main results presented here in more detail, this resource also allows examining limiting factors and exploring the components of the prediction for particular sites with Maxent's explain tool.

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. We will first highlight the effects of the different distribution modeling practices on model transferability and performance, as well as some limitations of the procedures described here. Then we will discuss the meaning of our SDMs for the spread of *C. cylindracea* in Europe and Australia.

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 has been observed in many studies led to the conclusions that ecological niches can shift in association with introductions outside of the native range (e.g. [19,21,71-73], but see 6). 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 [9,74]. However, changes in the fundamental niche of introduced species are certainly possible [4,15].

Regardless of whether niche shifts observed in correlative SDMs are a consequence of changes in the realized or fundamental niche, it would certainly be useful to have a set of procedures that improve the predictive power of SDMs outside the training range in order to inform conservation planning and decision making. The methods used here were applied hoping they would improve the transferability of the SDMs of introduced species built using the popular presence-only method Maxent. We found that reducing the number of predictor variables drastically improved the transferability of our SDMs. Limiting the model complexity, reducing geographic sampling bias by occurrence thinning and choosing a global background had comparably small effects.

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., [4,9,19,20,21]). The method used here, 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. Although the use of procedures to select predictors and model complexity in an automated manner is common practice in many types of modeling including niche modeling [75-77], to our knowledge such approaches have not been used commonly in combination with Maxent. However, we do acknowledge that such predictor selection methods are no substitute for physiological knowledge of the organism [78], 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 as such, the models achieve better overall performance and have higher transferability [7,10,11,79]. For these reasons, the use of simple environmental response surfaces to avoid overfitting has been recommended for SDMs of invasive species [4,7,11]. Generally, the complexity of maximum entropy models is adjusted by using L_1 regularization [68], which varies along a continuous scale and has been used in other studies aimed at improving the performance of Maxent SDMs [10]. We chose to use a simple dichotomy between Maxent's auto-features versus 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 and thus we did not observe the improvement of predictions with simpler models that other

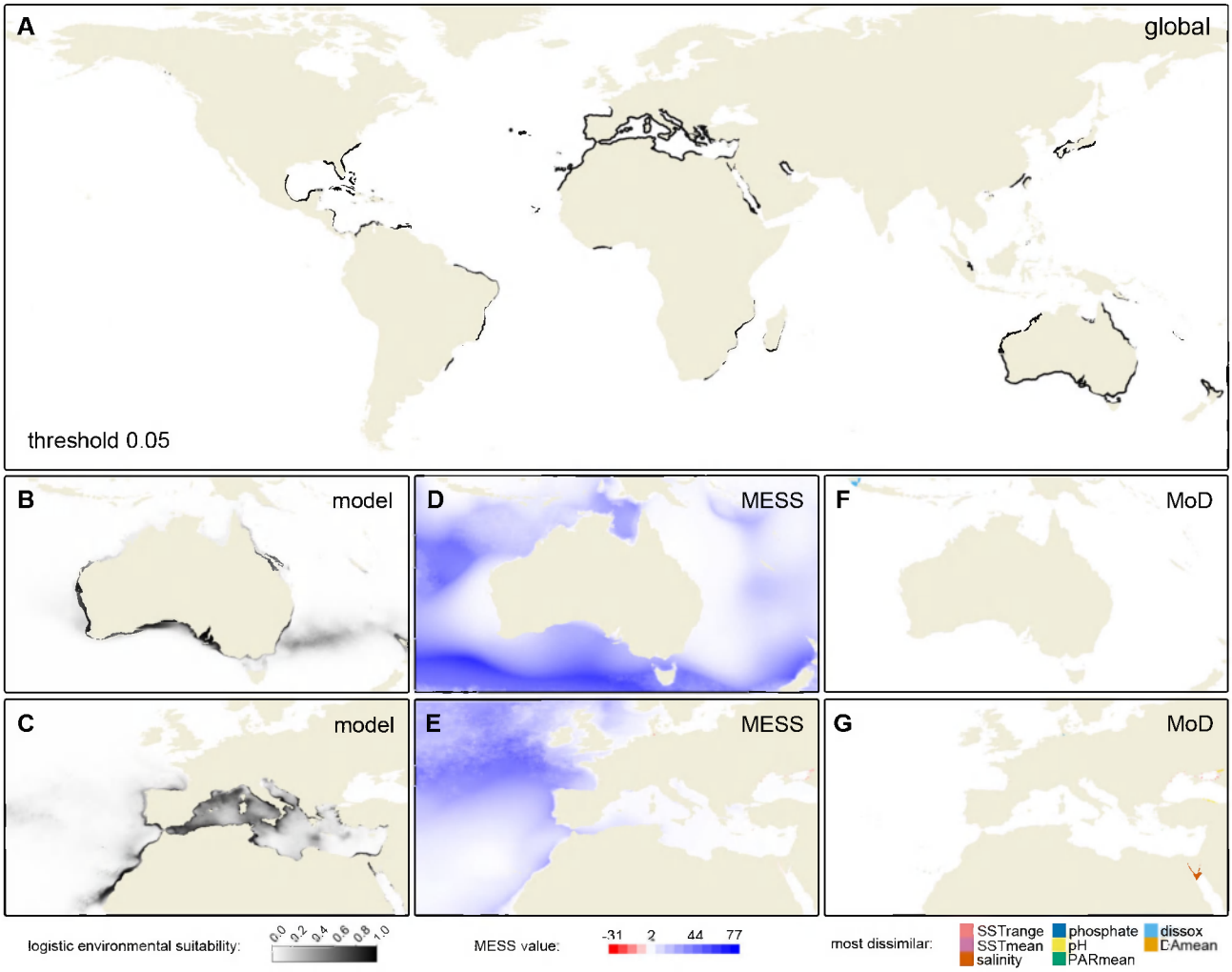

Figure 4. Species distribution model for *Caulerpa cylindracea*. Panel A shows global areas predicted to have suitable macroecological conditions for the species. This map uses a threshold for Maxent's logistic suitability corresponding to the 10% training presences (threshold = 0.053) and predictions 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.

doi: 10.1371/journal.pone.0068337.g004

studies have [7,10]. This can probably be attributed to the fact that model complexity does not differ much between the two conditions in our experimental setup: the auto-features condition only differed in having hinge features in addition to the linear and quadratic features used in the "simple" condition. Nonetheless, we follow previous authors in their conclusion that correlative models with smooth responses will generally outperform those with complex responses. This is especially true if the number of occurrence points used to build models is large, because this increases the potential for overfitting. 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 commonly done in classical model selection procedures [75].

The use of thinned occurrences generally resulted in SDMs with better transferability, but the effect was not significant in a Wilcoxon signed-rank test and small compared to that obtained from predictor selection. Nonetheless, we anticipate that this approach may be useful in situations where the geographical bias is stronger than in our dataset and/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, using bias grids in Maxent, or through application of trend surface analysis [7,25,26]. Various statistical approaches to address spatial autocorrelation have also been used [13]. In our case study, the background

selection had a rather limited effect on the transferability of SDMs and 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 (compare second row in Figure 3A with second row in Figure 3B). It is well known that AUC is sensitive to background choice, with larger backgrounds inflating AUC values while not yielding more informative models [26,80,81]. Our observation of higher AUC values for global backgrounds compared to regional backgrounds, which was not paralleled in Schoener's D, is completely in line with this. As such, for comparisons of the transferability of models built with different backgrounds, we suggest the use of Schoener's D rather than test AUC. 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" [7]. Our experiments did not confirm this conclusion but suggested that transferability (as measured by Schoener's D) is indifferent to the choice of background.

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 < 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). In this case, it is appropriate to use test AUC to compare performance, as all these models are built and evaluated using identical background points. Similar conclusions regarding the better predictive power of models using combined native-invaded datasets were reached in studies of other species (e.g., [4,82]). Our results also suggest that the combined data have the advantage of being more insensitive to the modeling choices that need to be made, but this generalization should be verified with other case studies.

Potential limitations

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

Firstly, 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 with the aim of avoiding variables for which the invaded samples are outside of the range of values of native samples. It may also be beneficial to upweight the scarce samples from the invaded range in the model-building step. It is worth noting that we used an essentially arbitrary threshold to retain predictor variables, i.e. they had to be present in 60% or more of the top-scoring models for at least two out of three regions (Figure 2). This approach was chosen because variables important in multiple regions are more likely to be of global importance, and secondly because the 60% threshold resulted in a halving of the number of predictors. However, this raises the question of how these criteria influence the results and whether more objective criteria could be used. The evaluation of all 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. However, 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 and thus lead 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 [74].

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. Also, if sampling reflects population densities, geographic autocorrelation of records can add a potentially desirable quantitative aspect to the model. This will, of course depend on the specific goal and the 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 and have had the chance to disperse widely in the invaded range are more likely to have spread through their entire potential niche and thus make good case studies.

An additional approach towards testing the degree to which these methods can be generalized, 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 [17,74,79,83-87]. 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. Such insights would obviously be beneficial to 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 of Europe and Africa as well as a substantial potential expansion along the southern coast of Australia (Figure 4A). Admittedly, the logistic values in Maxent lack a clear-cut interpretation [88] and determining thresholds for presence-only SDMs is not an exact science [89,90]. Based on several thresholds tested (e.g. 10-percentile training presence, equal training sensitivity and specificity), the inferred range boundaries are quite far beyond the known occurrences of the species (Figure 1 vs. Figure 4A). 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 only about 20 years and, despite the species' relatively rapid colonization rate [91], it is likely that it has not reached its distributional limits yet. In Australia, the native area of the species, it was known best from the Western Australian coast [48]. However, the recent observations of invasive populations of this species along the southern coast, where it did not previously occur (reference [46] and pers. obs.), prompted us to generate SDMs for this species in order to investigate whether the species could potentially colonize more of the coast. Our models do indeed suggest that the macroecological conditions are highly favorable and that *C. cylindracea* could colonize the entire southern coastline of Australia. Besides these potential expansions in regions where the species is present already, several other coastlines are predicted to be suitable environment where the species could establish if it were to be introduced (Figure 4A).

Needless to say our models only incorporate macroecological predictor variables. Besides this, the microhabitat, as well as possible biotic interactions, also need to be favorable for the species to establish itself in the areas that are predicted to be suitable. In its native range, *C. cylindracea* is usually found on rocky substrata close to the low-tide mark but in more tropical locations (NW Australia and the Great Barrier Reef) it is typically found growing on sand in lagoons and around reefs. 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 [29]. 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 beds [92-94], and that it tolerates near-bottom orbital velocities below 15 cm s^{-1} [93]. 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 if there are no biotic interactions inhibiting its settlement and expansion.

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 in the main paper, these are available as supplementary materials on FigShare (http://dx.doi.org/10.6084/m9.figshare.681723). Most correspond to our expectations based on physiological knowledge of other algae, including other *Caulerpa* species [95], but some do not. For example, the correlative model indicates that the species is mainly found in phosphate-poor waters with the response curve rapidly dropping at concentrations over 0.4 µmol 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 [96-99]. This suggests that our correlative model may be misled in this case. It is also interesting to note that models built from occurrences in the native range predicted a much broader range of suitable temperatures than models from occurrences in the invaded range. More specifically, the model from invasive occurrences has a response curve that peaks at ca. 20°C, dropping off quickly at higher temperatures. The curve from a model with native occurrences also peaks at ca. 20°C, but drops much more gently at higher temperatures. Whether this should simply be interpreted as an indication that warmer areas are yet to be colonized in the invaded range (i.e. that the model is biased towards colder temperature due to the current distribution), or that the introduced strain has a reduced range of temperature tolerance compared to the native population, remains to be investigated. 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, a number of problems related to their accuracy and transferability have to be overcome. The methods introduced, explored and evaluated here aim to improve the situation. 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 apparently reduces their sensitivity to choices made in the modeling process. We also presented an SDM of *Caulerpa cylindracea* that achieves very

high predictive power, illustrating the applicability of these methods in the marine realm for which comparably little niche modeling has been done [100]. The procedures introduced here are available for further evaluation with other case and simulation studies, which should provide further insights into the degree to which our results can be generalized. We hope and anticipate that they will form a useful strategy to improve predictive SDMs and in turn, help to better inform environmental decision makers.

Supporting Information

Figure S1. Model surveying results indicating qualitatively similar results when analyses are carried out with global or regional backgrounds.
(PDF)

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*. We thank Stefano Draisma and Willem Prud'Homme van Reine for sharing information about their DNA bar codes survey of *Caulerpa* collections and John Huisman for providing many of the occurrence records of *C. cylindracea* from NW Australia. Niels Raes and Dirk Maes provided valuable feedback on a previous version of this manuscript.

Author Contributions

Conceived and designed the experiments: HV LT. Performed the experiments: LT GB FM HV. Analyzed the data: HV LT. Contributed reagents/materials/analysis tools: AJ CFDG GH ODC. Wrote the manuscript: HV GB.

References

1. Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40: 677-697. doi:10.1146/annurev.ecolsys.110308.120159.
2. Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. Q Rev Biol 78: 419-433. doi: 10.1086/378926. PubMed: 14737826.
3. Wiens JA, Stralberg D, Jongsomjit D, Howell CA, Snyder MA (2009) Niches, models, and climate change: Assessing the assumptions and uncertainties. Proc Natl Acad Sci U S A 106 Suppl 2: 19729-19736. doi: 10.1073/pnas.0901639106. PubMed: 19822750.
4. Jimenez-Valverde A, Peterson AT, Soberon J, Overton JM, Aragon P et al. (2011) Use of niche models in invasive species risk assessments. Biol Invasions 13: 2785-2797. doi:10.1007/s10530-011-9963-4.
5. Franklin J, Miller JA (2009) Mapping species distributions. Cambridge: Cambridge University Press. 320pp.
6. Petitpierre B, Kueffer C, Broennimann O, Randin C, Daehler C et al. (2012) Climatic niche shifts are rare among terrestrial plant invaders. Science 335: 1344-1348. doi:10.1126/science.1215933. PubMed: 22422981.
7. Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. Methods Ecol Evolution 1: 330-342. doi:10.1111/j.2041-210X.2010.00036.x.
8. Zurell D, Elith J, Schroder B (2012) Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. Divers Distrib 18: 628-634. doi:10.1111/j.1472-4642.2012.00887.x.
9. Rödder 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. doi:10.1007/s00114-010-0694-7. PubMed: 20617298.
10. Warren DL, Seifert SN (2010) Environmental niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol Applications 21: 335-342.
11. Rodda GH, Jarnevich CS, Reed RN (2011) Challenges in identifying sites climatically matched to the native ranges of animal invaders. PLOS ONE 6: e14670. doi:10.1371/journal.pone.0014670. PubMed: 21347411.
12. Wolmarans R, Robertson MP, Van Rensburg BJ (2010) Predicting invasive alien plant distributions: how geographical bias in occurrence records influences model performance. J Biogeogr 37: 1797-1810. doi: 10.1111/j.1365-2699.2010.02325.x.
13. Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J et al. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30: 609-628. doi: 10.1111/j.2007.0906-7590.05171.x.
14. Leathwick JR, Austin MP (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. Ecology 82: 2560-2573. doi:10.1890/0012-9658(2001)082[2560:CIBTSI]2.0.CO;2.
15. Whitney KD, Gabler CA (2008) Rapid evolution in introduced species, 'invasive traits' and recipient communities: challenges for predicting invasive potential. Divers Distrib 14: 569-580. doi:10.1111/j.1472-4642.2008.00473.x.
16. Alexander JM, Edwards PJ (2010) Limits to the niche and range margins of alien species. Oikos 119: 1377-1386. doi:10.1111/j.1600-0706.2009.17977.x.
17. Santika T, Hutchinson MF (2009) The effect of species response form on species distribution model prediction and inference. Ecol Modell 220: 2365-2379. doi:10.1016/j.ecolmodel.2009.06.004.
18. Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93: 679-688. doi:10.1890/11-0826.1. PubMed: 22624221.
19. Rödder 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.
20. Peterson AT, Nakazawa Y (2008) Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. Glob Ecol Biogeogr 17: 135-144.
21. Rödder D, Lötters S (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (Hemidactylus turcicus). Glob Ecol Biogeogr 18: 674-687. doi:10.1111/j.1466-8238.2009.00477.x.
22. Randin CF, Dirnbock T, Dullinger S, Zimmermann NE, Zappa M et al. (2006) Are niche-based species distribution models transferable in space? J Biogeogr 33: 1689-1703. doi:10.1111/j.1365-2699.2006.01466.x.
23. Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol Evolution 3: 260-267. doi:10.1111/j.2041-210X.2011.00170.x.
24. Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161-175. doi:10.1111/j.0906-7590.2008.5203.x.
25. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A et al. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19: 181-197. doi:10.1890/07-2153.1. PubMed: 19323182.
26. Acevedo P, Jiménez-Valverde A, Lobo JM, Real R (2012) Delimiting the geographical background in species distribution modelling. J Biogeogr 39: 1383-1390. doi:10.1111/j.1365-2699.2012.02713.x.
27. Sonder G (1845) Nova algarum genera et species, quas in itinere ad oras occidentales Novae Hollandiae, collegit L Priess, Ph. Dr Bot Zeitung 3: 49-57.

28. Belton GS, Huisman JM, Prud'homme van Reine WF, Draisma SGA, Sauvage T et al. (2013) Species delimitation in the *Caulerpa racemosa* complex (Chlorophyta, Caulerpaceae). J Phycol (in review).
29. Verlaque M, Durand C, Huisman JM, Boudouresque CF, Le Parco Y (2003) On the identity and origin of the Mediterranean invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta). Eur J Phycol 38: 325-339. doi:10.1080/0967026031000161259 2.
30. Klein J, Verlaque M (2008) The *Caulerpa racemosa* invasion: A critical review. Mar Pollut Bull 56: 205-225. doi:10.1016/j.marpolbul. 2007.09.043. PubMed: 18061625.
31. Verlaque M, Afonso-Carillo J, Gil-Rodriguez MC, Durand C, Boudouresque CF et al. (2004) Blitzkrieg in a marine invasion: *Caulerpa racemosa* var. *cylindracea* (Bryopsidales, Chlorophyta) reaches the Canary islands (North-East Atlantic). Biol Invasions 6: 269-281. doi:10.1023/B:BINV.0000034589.18347.d3.
32. Piazzi L, Meinesz A, Verlaque M, Akcali B, Antolic B et al. (2005) Invasion of *Caulerpa racemosa* var. *cylindracea* (Caulerpales, Chlorophyta) in the Mediterranean Sea: an assessment of the spread. Cryptogam Algol 26: 189-202.
33. Piazzi L, Ceccarelli D (2006) Persistence of biological invasion effects: recovery of macroalgal assemblages after removal of *Caulerpa racemosa* var. *cylindracea*. Estuarine Coast Shelf Sci 68: 455-461. doi: 10.1016/j.ecss.2006.02.011.
34. Piazzi L, Ceccarelli D, Cinelli F (2001) Threat to macroalgal diversity: effects of the introduced green alga *Caulerpa racemosa* in the Mediterranean. Mar Ecol Prog S 210: 149-159. doi:10.3354/ meps210149.
35. Ceccherelli G, Campo D (2002) Different effects of *Caulerpa racemosa* on two co-occurring seagrasses in the Mediterranean. Bot Marina 45: 71-76.
36. 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. doi: 10.1007/s00338-008-0358-7.
37. 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). Mar Ecol 30: 337-345. doi:10.1111/j.1439-0485.2009.00282.x.
38. Piazzi L, Balata D (2009) Invasion of alien macroalgae in different Mediterranean habitats. Biol Invasions 11: 193-204. doi:10.1007/ s10530-008-9224-3.
39. Klein J (2007) **Impact of *Caulerpa racemosa* var. *cylindracea* (Caulerpales, Chlorophyta) on macrophyte assemblages of the north-western Mediterranean Sea**. PhD thesis Marseille: University of Aix-Marseille II. 315 p
40. Jousson O, Pawlowski J, Zaninetti L, Meinesz A, Boudouresque CF (1998) Molecular evidence for the aquarium origin of the green alga *Caulerpa taxifolia* introduced to the Mediterranean Sea. Mar Ecol Prog S 172: 275-280. doi:10.3354/meps172275.
41. 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
42. 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
43. Durand C, Manuel M, Boudouresque CF, Meinesz A, Verlaque M et al. (2002) Molecular data suggest a hybrid origin for the invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta) in the Mediterranean Sea. J Evol Biol 15: 122-133. doi:10.1046/j.1420-9101.2002.00370.x.
44. Sauvage T, Payri C, Draisma SGA, Prud'homme van Reine WF, Verbruggen H et al. (2013) Molecular diversity of the *Caulerpa racemosa–Caulerpa peltata* complex (Caulerpaceae, Bryopsidales) in New Caledonia, with new Australasian records for *C. racemosa* var. *cylindracea*. Phycologia 52: 6-13. doi:10.2216/11-116.1.
45. Nuber N, Gornik O, Lauc G, Bauer N, Zuljević A, et al (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
46. Collings G, Westphalen G, Rowling K, Eglinton Y (2004) ***Caulerpa racemosa* var. *cylindracea* occurrence in western South Australia. Report to PIRSA Marine Habitat Program. South Australian Research and Development Institute (Aquatic Sciences)** Report Number RD04/0169
47. Wiltshire KH, Rowling KP, Deveney MR (2010) Introduced marine species in South Australia: a review of records and distribution mapping. Adelaide: South Australian Research and Development Institute (Aquatic Sciences). 236 p.
48. Womersley HBS (1984) The marine benthic flora of southern Australia. Part I. Government Printer, South Australia: 329.
49. Huisman JM (2000) Marine plants of Australia. Nedlands. Western Australia: University of Western Australia Press. 300pp.
50. Piazzi L, Balestri E, Cinelli F (1994) Presence of *Caulerpa racemosa* in northwestern Mediterranean. Cryptogam Algol 15: 183-189.
51. 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
52. Verlaque M, Boudouresque CF, Meinesz A, Gravez V (2000) The *Caulerpa racemosa* complex (Caulerpales, Ulvophyceae) in the Mediterranean Sea. Bot Marina 43: 49-68.
53. Zuljevic A, Antolic B, Onofri V (2003) First record of *Caulerpa racemosa* (Caulerpales: Chlorophyta) in the Adriatic Sea. J Mar Biol Assoc UK 83: 711-712. doi:10.1017/S0025315403007689h.
54. Gambi MC, 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.
55. Buia MC, Petrocelli A, Saracino OD (1998) *Caulerpa racemosa* spread in the Mediterranean Sea: first record in the Gulf of Taranto. Biol Mar Méditerranée 5: 527-529.
56. Nizamuddin M (1991) **The green marine algae of Lybia**: Alga Publications. 227 p
57. Serio D, Pizzuto F (1999) Studio di un prato a *Caulerpa* racemosa (Forsskål) J. Agardh (Caulerpales, Chlorophyta) di Brocoli (SR) con osservazioni in coltura della specie. Boll Accad Gioenia Sci Nat Catania 31: 201-209.
58. Djellouli AS (2000) *Caulerpa racemosa* (Forskaal) J. Agardh en Tunisie. In: RACSPA PP, editor. Actes du 1er Symposium Méditerranéen sur la Végétation Marine. Ajaccio. pp. 124-127.
59. Cossu A, Gazale V (1997) Sulla presenza di *Caulerpa racemosa* (Forsskål) J. Agardh in Sardegna. In: Cossu A, Meloni MM, editors. SOS Caulerpa? Introduzione di nuove specie net Mediterraneo e compatibilità con quelle presenti. Cagliari: Poseidon. pp. 87-97.
60. Pena-Martín C, Cristóbal-Fernanz JC, Crespo B, [!(surname)!] (2003) Caulerpa racemosa (Forssk.) J. Agardh (Caulerpaceae, Chlorophyceae), nueva para la flora de Alicante. Anales Jardin Botanico de Madrid 60: 448-449.
61. Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F et al. (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. Glob Ecol Biogeogr 21: 272-281. doi:10.1111/j. 1466-8238.2011.00656.x.
62. Verbruggen H (2011). OccurrenceThinner 1: 03. http:// www.phycoweb.net/software/OccurrenceThinner/. Accessed: 2012 October 7.
63. Wand MP, Jones MC (1995) Kernel smoothing. London: Chapman and Hall.
64. Ripley B (2011) KernSmooth. http://cran.r-project.org/web/packages/ KernSmooth/. Accessed: 2012 October 7.
65. Lüning K (1990) Seaweeds: their environment, biogeography, and ecophysiology. C YarishH Kirkman. New York: Wiley & Sons. p. 527.
66. Verbruggen H (2011) Maxent Model Surveyor 1.03. http:// phycoweb.net/software/MMS/. Accessed: 2012 October 7.
67. Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24: 38-49. doi:10.1017/S0376892997000088.
68. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190: 231-259. doi:10.1016/j.ecolmodel.2005.03.026.
69. Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE et al. (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17: 43-57. doi:10.1111/j.1472-4642.2010.00725.x.
70. Schoener TW (1968) The *Anolis* lizards of Bimini: resource partitioning in a complex fauna. Ecology 49: 704-726. doi:10.2307/1935534.
71. Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson AT et al. (2007) Evidence of climatic niche shift during biological invasion. Ecol Lett 10: 701-709. doi:10.1111/j.1461-0248.2007.01060.x. PubMed: 17594425.
72. Medley KA (2010) Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. Glob Ecol Biogeogr 19: 122-133. doi: 10.1111/j.1466-8238.2009.00497.x.
73. Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Glob Ecol Biogeogr 16: 24-33. doi:10.1111/j.1466-8238.2006.00258.x.

74. Soberon J, Peterson AT (2011) Ecological niche shifts, and environmental space anisotropy: a cautionary note. Rev Mexicana Biodiversidad 82: 1348-1355.
75. Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer Verlag. 488pp.
76. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. New York: Springer Verlag. 745pp.
77. Bedia J, Busqué J, Gutiérrez JM (2011) Predicting plant species distribution across an alpine rangeland in northern Spain. A comparison of probabilistic methods. Appl Veg Sci 14: 415-432. doi:10.1111/j.1654-109X.2011.01128.x.
78. Araujo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33: 1677-1688. doi:10.1111/j.1365-2699.2006.01584.x.
79. Reineking B, Schroder B (2006) Constrain to perform: Regularization of habitat models. Ecol Modell 193: 675-690. doi:10.1016/j.ecolmodel.2005.10.003.
80. Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17: 145-151. doi:10.1111/j.1466-8238.2007.00358.x.
81. Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Glob Ecol Biogeogr 21: 498-507. doi:10.1111/j.1466-8238.2011.00683.x.
82. Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. Biol Lett 4: 585-589. doi:10.1098/rsbl.2008.0254. PubMed: 18664415.
83. Rödder D, Engler JO (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. Glob Ecol Biogeogr 20: 915-927. doi:10.1111/j.1466-8238.2011.00659.x.
84. Vaclavik T, Meentemeyer RK (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. Divers Distrib 18: 73-83. doi:10.1111/j.1472-4642.2011.00854.x.
85. Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher SP et al. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Modell 222: 1810-1819. doi:10.1016/j.ecolmodel.2011.02.011.
86. Meynard CN, Kaplan DM (2012) The effect of a gradual response to the environment on species distribution modeling performance. Ecography 35: 499-509. doi:10.1111/j.1600-0587.2011.07157.x.
87. Meynard CN, Kaplan DM (2013) Using virtual species to study species distributions and model performance. J Biogeogr 40: 1-8. doi:10.1111/jbi.12006.
88. Royle JA, Chandler RB, Yackulic C, Nichols JD (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods in Ecology and Evolution: in press.
89. Bean WT, Stafford R, Brashares JS (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography 35: 250-258. doi:10.1111/j.1600-0587.2011.06545.x.
90. Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385-393. doi:10.1111/j.0906-7590.2005.03957.x.
91. Mineur F, Davies AJ, Maggs CA, Verlaque M, Johnson MP (2010) Fronts, jumps and secondary introductions suggested as different invasion patterns in marine species, with an increase in spread rates over time. Proc Biol Sci 277: 2693-2701. PubMed: 20410039.
92. Katsanevakis S, Issaris Y, Poursanidis D, Thessalou-Legaki M (2010) Vulnerability of marine habitats to the invasive green alga *Caulerpa racemosa* var. *cylindracea* within a marine protected area. Mar Environ Res 70: 210-218. doi:10.1016/j.marenvres.2010.05.003. PubMed: 20621771.
93. Infantes E, Terrados J, Orfila A (2011) Assessment of substratum effect on the distribution of two invasive *Caulerpa* (Chlorophyta) species. Estuarine Coast Shelf Sci 91: 434-441. doi:10.1016/j.ecss.2010.11.005.
94. Katsanevakis S, Salomidi M, Panou A (2010) Modelling distribution patterns and habitat preference of the invasive green alga Caulerpa racemosa in the Saronikos Gulf (Eastern Mediterranean). Aquatic Biol 10: 57-67. doi:10.3354/ab00267.
95. Theil M, Westphalen G, Collings G, Cheshire A (2007) *Caulerpa taxifolia* responses to hyposalinity stress. Aquat Bot 87: 221-228. doi:10.1016/j.aquabot.2007.06.001.
96. Hanisak MD (1979) Nitrogen limitation of *Codium fragile* ssp. *tomentosoides* as determined by tissue analysis. Mar Biol 50: 333-337. doi:10.1007/BF00387010.
97. Lobban CS, Harrison PJ (1994) Seaweed ecology and physiology. Cambridge: Cambridge University Press. 366pp.
98. Lapointe BE, Bedford BJ (2010) Ecology and nutrition of invasive *Caulerpa brachypus* f. *parvifolia* blooms on coral reefs off southeast Florida, USA. Harmful Algae 9: 1-12. doi:10.1016/j.hal.2009.06.001.
99. Maloney B, Iliffe TM, Gelwick F, Quigg A (2011) Effect of nutrient enrichment on naturally occurring macroalgal species in six cave pools in Bermuda. Phycologia 50: 132-143. doi:10.2216/09-83.1.
100. Robinson LM, Elith J, Hobday AJ, Pearson RG, Kendall BE et al. (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Glob Ecol Biogeogr 20: 789-802. doi:10.1111/j.1466-8238.2010.00636.x.