Mediterranean Marine Science

Vol 25, No 2 (2024)

Mediterranean Marine Science

Surprising widespread Cymodocea nodosa occurrence along Israel's Mediterranean coast and Implications for Seagrass Conservation in a hotspot of climate change

ORI HEPNER UCKO, EDUARDO ARLÉ, SHAHAR MALAMUD, GIDON WINTERS, JONATHAN BELMAKER

doi: 10.12681/mms.36597

Supplementary Data

Mediterranean Marine Science
Indexed in WoS (Web of Science, ISI Thomson) and SCOPUS
www.hcmr.gr
DOI: https://doi.org/10.12681/mms.36597

Surprising widespread *Cymodocea nodosa* occurrence along Israel's Mediterranean coast and Implications for Seagrass Conservation in a hotspot of climate change

Ori HEPNER UCKO, Eduardo ARLÉ, Shahar MALAMUD, Gidon WINTERS and Jonathan BELMAKER

Mediterranean Marine Science, 25 (2) 2024

SI 1 - Detailed SDM methods.

Delimitation of the study area for the SDMs

The definition of the area adopted in the SDMs was based on depth, provided the focus species is dependent on light availability. We masked the environmental variables by considering only areas where the maximum depth was within the 95th quantile of occurrences records (Fig. S2).

Occurrence data preparation

As a pre-modeling step, we moved records with coordinates on land to the closest cell corresponding to the sea using the R-package 'ellipsenm' (Cobos *et al*., 2022), provided that the distance would not surpass 10 km. Out of the 180 literature-provided records, 162 were maintained in the same location indicated by the coordinates, 14 were moved to the closest marine cell, and 4 were discarded due to the distance from the closest cell being greater than 10 km. Out of the 292 presence records collected in our surveys, 275 were kept in the same location, and 17 were moved to the closest marine cell.

As the available data is presence-only, we used additional background data to run the models. We ran a sensitivity analysis to determine the number of background data that would yield better models by generating different amounts of background data. We identified the same number of background points as presences across the study area as producing the most robust results. Background points represent the conditions available in the environment compared to those found where the species is known to occur (Barbet-Massin *et al.*, 2012; Iturbide *et al*., 2018). An alternative method would be the generation of pseudo-absence in areas where the species is not expected to be found (Barbet-Massin *et al*., 2012), which involves either geographical or ecological assumptions. The decision of not generating pseudo-absence data was due to the very nature of this study, which describes the discovery of a population thriving out of the species' previously known geographical and environmental spaces.

Model training and testing

The prediction of suitable areas for *C. nodosa* was carried out with the 'sdm' R-package (Naimi & Araújo, 2016). As SDM results may vary considerably depending on the selected algorithm (Diniz-Filho *et al*., 2009; Thuiller, 2004), we adopted an ensemble forecasting approach (Araújo & New, 2007), which combines results from different algorithms in one consensus model. We used six algorithms, representing distinct modeling approaches, according to IUCN guidelines (IUCN, 2022). We used three statistical (GAM, GLM, and MARS), one classification (DOMAIN), and two machine learning (RF and BRT) algorithms (Franklin, 2010; Rangel & Loyola, 2012; Grenouillet *et al*., 2011). We ran five repetitions for each algorithm, adopting the cross-validation approach, with five partitions of the data (80% for training and 20% for testing), amassing 300 SDMs, 150 considering only the occurrences from the literature, and 150 using those combined with the presences found in our surveys. We evaluated the models according to the True Skill Statistics (TSS) and the area under the receiver-operator curve (AUC). TSS is a threshold-dependent metric that ranges from -1 to 1, with positive values indicating model performance better than random results (Allouche *et al*., 2006). AUC is threshold-independent, and ranges from 0 to 1, AUC equals 0.5 indicate performances as good as a randomly generated model, while models with AUC $\geq$ 0.7 are considered robust predictions (Swets, 1988; Lobo *et al*., 2008). Out of the 150 SDMs based only on the occurrences from the literature, 74 were considered satisfactory (mean AUC $\approx$ 0.84; mean TSS $\approx$ 0.63). The models including the occurrences collected in our surveys yielded 124 satisfactory predictions out of 150 (mean AUC $\approx$ 0.91; mean TSS $\approx$ 0.67. In both analyses, RF produced the most satisfactory models (TSS ≥ 0.5), while DOMAIN and GLM yielded the least trustworthy predictions (Table S2).

Ensemble forecast

Models with TSS ≥ 0.5 were selected as having produced useful predictions (Zhang *et al*., 2015) and combined in the ensemble models. Subsequently, we transformed all selected models into binary predictions, according to the threshold determined by the maximum sensitivity plus specificity (Liu *et al*., 2005, 2013). We then combined the binary models into an ensemble forecast by summing the values in the binary predictions and considering only the pixels for which the majority of the models classified as suitable for the species occurrence. Finally, we verified whether the occurrence records from each dataset could be predicted correctly by the corresponding final model. We carried out this last step by simply extracting the values at the presence locations from the final binary models, with 1s referring to correct predictions and 0s to failed predictions.

Table S1. Occurrences of *C. nodosa* meadows across the Eastern Mediterranean sea compiled from the literature review, available in https://zenodo.org/records/13120938.

Table S2: GAM results. The binomial response variable was the presence of seagrass meadows in a transect (number of successes) relative to the number of absences of the meadows (failures). Depth, latitude and season were used as predictors.

Predictors	***Odds Ratios***	***CI***	***p***
(Intercept)	***0.00***	***0.00 – 0.00***	***<0.001***
season [Spring]	***0.18***	***0.09 – 0.37***	***<0.001***
season [Summer]	***1.89***	***1.22 – 2.93***	***0.005***
season [Autumn]	***5.79***	***2.94 – 11.40***	***<0.001***
Smooth term (depth)			***<0.001***
Smooth term (cordinate lat)			***<0.001***
R^2	***0.466***		

Table S3: SDM results for the models built based solely on data extracted from the literature (B), and for those including the data collected in our surveys (B). Each table shows the number of selected models according to the criterion of yielding TSS ≥ 0.5, and the means

Literature data

Algorithm	Selected models	mean AUC	mean TSS
brt	17	0.8240	0.5909
domain	2	0.8050	0.5440
gam	9	0.8013	0.5790
glm	4	0.8008	0.5623
mars	17	0.8024	0.6021
rf	25	0.8963	0.7216

All data

Algorithm	Selected models	mean AUC	mean TSS
brt	25	0.9146	0.6713
domain	13	0.8422	0.5688
gam	25	0.9046	0.6531
glm	11	0.8395	0.5650
mars	25	0.9117	0.6531
rf	25	0.9560	0.7839

Fig. S1: Mediterranean Sea. The darker shade of blue represents the study area adopted in this work - the Eastern Mediterranean Basin. The lighter shade represents the areas that were excluded from our study, namely the portions of the Mediterranean Sea located East to the Strait of Sicily, and the Adriatic Sea.

Fig. S2: Eastern Mediterranean Basin. The darkest shade of blue represents the study area masked by depth, adopted for the SDMs - the cells within the Eastern Mediterranean Basin where the maximum depth was within the 95th quantile of occurrences records. The intermediate shade of blue represents the cells within the study area that were excluded by the depth criterion. The lightest shade represents the aforementioned areas that were excluded from our study.

References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223-1232.

Araujo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22, 42-47.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecology and Evolution*, 3, 327-338.

Cobos, M.E., Osorio-Olvera, L., Soberón, J., Peterson, A.T., 2022. Ellipsenm: Ecological Niche's Characterizations Using Ellipsoids. R-Package available on https://github.com/marlonecobos/ellipsenm

Diniz-Filho, J.A.F., Bini, M.L., Rangel T.F., Loyola, R.D., Hof, C. *et al*. 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* 32, 897-906.

Franklin, J., 2010. Moving beyond static species distribution models in support of conservation biogeography: Moving beyond static species distribution models. *Diversity and Distributions*, 16, 321–-330.

Grenouillet, G., Buisson, L., Casajus, N., Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, 34, 9-17.

Iturbide, M., Bedia, J., Gutiérrez, J.M., 2018. Background sampling and transferability of species distribution model ensembles under climate change. *Global and Planetary Change*, 166, 19–29.

IUCN Standards and Petitions Committee, 2022 Guidelines for Using the IUCN Red List Categories and Criteria. Version 15.1. Prepared by the Standards and Petitions Committee.

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.

Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, 40, 778-789.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145-151.

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39, 368-375.

Rangel, T.F., Loyola, R.D., 2012. Labeling Ecological Niche Models. *Natureza e Conservação*, 10, 119-126.

Swets, J.A., 1988. Measuring the Accuracy of Diagnostic Systems. *Science*, 240, 1285-1293.

Thuiller, W., 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2020-2027.

Zhang, L., Liu, S., Sun, P., Wang, T., Wang, G. *et al*. 2015 Consensus Forecasting of Species Distributions: The Effects of Niche Model Performance and Niche Properties. *PLoS ONE*,10, e0120056.