Ecological Informatics 48 (2018) 37–47

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Mapping ecological indicators of human impact with statistical and machine learning methods: Tests on the California coast

A. Stock[a,*,1], A.J. Haupt[b], M.E. Mach[c], F. Micheli[d]

[a] *Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, CA 94305, USA*
[b] *School of Natural Sciences, California State University Monterey Bay, Seaside, CA 93955, USA*
[c] *DataONE, University of New Mexico, Albuquerque, NM 87106, USA*
[d] *Hopkins Marine Station and Stanford Center for Ocean Solutions, Stanford University, Pacific Grove, CA 93950, USA*

ARTICLE INFO

Keywords:
Multiple stressors
Cumulative impacts
Regression
Spatial prediction
Model selection
Kelp forests

ABSTRACT

Coastal ecosystems are exposed to multiple anthropogenic stressors such as fishing, pollution, and climate change. Ecosystem-based coastal management requires understanding where the combination of multiple stressors has large cumulative effects and where actions to address impacts are most urgently needed. However, the effects of multiple stressors on coastal and marine ecosystems are often non-linear and interactive. This complexity is not captured by commonly used spatial models for mapping human impacts. Flexible statistical and machine learning models like random forests have thus been used as an alternative modeling approach to identify important stressors and to make spatial predictions of their combined effects. However, tests of such models' prediction skill have been limited. Therefore, we tested how well ten statistical and machine learning methods predicted three ecological indicators of coastal marine ecosystem condition (kelp biodiversity, fish biomass, and rocky intertidal biodiversity) off California, USA. Spatial data representing anthropogenic stressors and ocean uses as well as natural gradients were used as predictors. The models' prediction errors were estimated by double spatial block cross-validation. The best models achieved mean squared errors about 25% lower than a null model for kelp biodiversity and fish biomass; none of the tested models worked well for rocky intertidal biodiversity. The models captured general trends, but not local variability of the indicators. For kelp biodiversity, the best performing method was principal components regression. For fish biomass, the best performing method was boosted regression trees. However, after tuning, this model did not include any interactions between stressors, and ridge regression (a constrained linear model) performed almost as well. While in theory flexible machine learning methods are required to represent the complex stressor-ecosystem state relationships revealed by experimental ecologists, with our data, this flexibility could not be harnessed because more flexible models overfitted due to small sample sizes and low signal-to-noise ratio. The main challenge for harnessing the flexibility of statistical and machine learning methods to link ecological indicators and anthropogenic stressors is obtaining more suitable data. In particular, better data describing the spatial and temporal distribution of human uses and stressors are needed. We conclude by discussing methodological implications for future research.

1. Introduction

Most of the world's coastal ecosystems are affected by multiple human stressors (Halpern et al., 2008, 2015). These stressors include local disturbances by direct use (e.g., recreational fishing, coastal construction activities) as well as long-term, broad-scale drivers like overfishing (Jackson et al., 2001) and climate change (Hoegh-Guldberg and Bruno, 2010; Poloczanska et al., 2016). Understanding the cumulative effects of multiple stressors is thus a priority for marine researchers, managers and resource users alike (Borja et al., 2014; Mason et al., 2017). One approach to mitigate the cumulative impacts of multiple stressors and balance human uses with environmental protection is ecosystem-based management (EBM; Levin et al., 2009). EBM requires mapping the overall condition of marine ecosystems, understanding how they are affected by anthropogenic stressors, and predicting their response to different management decisions (Borja et al., 2016; Foley et al., 2017; Thrush et al., 2016). However, laboratory and field experiments suggest that the effects of multiple stressors

---

* Corresponding author.
E-mail address: astock@ldeo.columbia.edu (A. Stock).

[1] Present affiliation: Lamont-Doherty Earth Observatory, The Earth Institute, Columbia University, Palisades, NY 10964, USA.

https://doi.org/10.1016/j.ecoinf.2018.07.007
Received 15 March 2018; Received in revised form 27 July 2018; Accepted 30 July 2018

on marine ecosystems are complex, non-linear, and hard to predict (Crain et al., 2008; Hunsicker et al., 2016; Côté et al., 2016).

Several studies have proposed spatial models for mapping human impacts on marine ecosystems. Most are simple, additive, linear models that calculate a human impact index and are unlikely to capture important ecological non-linearities and interactions (Korpinen and Andersen, 2016). The most commonly used of these human impact index models was proposed by Halpern et al. (2008), and subsequently applied in many regional studies: e.g. for the waters of British Columbia (Ban et al., 2010; Murray et al., 2015), the Baltic Sea (Korpinen et al., 2012), parts of the North Sea (Andersen et al., 2013), the Mediterranean and Black Seas (Micheli et al., 2013), the Great Lakes (Allan et al., 2013), the Bering Strait region (Afflerbach et al., 2017) and our study area, the coast of California (Halpern et al., 2009; Mach et al., 2017). When cautiously interpreted, maps generated with this model can broadly highlight potential priority areas for research and management (Andersen et al., 2017; Stock et al., 2018). However, like other existing models for mapping human impacts on marine ecosystems, Halpern et al.'s (2008) model makes various simplifying assumptions (e.g. that stressors do not interact) that are at odds with experimental results (Halpern and Fujita, 2013). In combination, these assumptions and data quality can have large effects on model outputs (Gissi et al., 2017; Stock and Micheli, 2016) and lead to ineffective environmental management decisions.

For example, a model that ignores stressor interactions may suggest reducing a stressor that in reality counteracts the effects of another stressor; in this case, reducing the former will have limited benefits (Brown et al., 2014). Halpern et al.'s and similar human impact index models also rely on expert judgment to link human stressors to ecosystem-level effects, but it would be preferable to extract relationships between stressors and ecosystem condition from data (Doubleday et al., 2017). Furthermore, first attempts to test Halpern et al.'s (2008) model against ecological field data reported mixed results (Sala et al., 2012; Andersen et al., 2015; Clark et al., 2016). While more theoretically grounded approaches for spatial modeling of multiple stressor effects are under development (Hodgson and Halpern, 2018), they have so far considered few stressors (e.g. Coll et al., 2016) or are not yet fully operational (e.g. Giakoumi et al., 2015). Simple, linear and additive models like Halpern et al.'s thus remain the most widely used (Korpinen and Andersen, 2016).

A notable, empirical exception to the established additive models and the mechanistic models under development was proposed by Parravicini et al. (2012). These authors created maps of marine habitat status and various anthropogenic stressors for a small (few km^2) coastal study area and used random forests, a statistical learning (or machine learning, which we here consider synonymous) model, to link habitat status and stressors. They then predicted how habitat status would change under different management scenarios. Statistical learning methods have also been used to identify important gradients, pressures, thresholds, and interactions that affect marine ecosystems (Jones et al., 2017; Large et al., 2013, 2015; Samhouri et al., 2017; Teichert et al., 2016). In the future, such methods in combination with "big environmental data" are expected to shed more light on and allow for better spatial prediction of multiple stressor effects. Unfortunately, currently available marine ecological data are rarely "big", and monitoring programs are not designed for this purpose (Baird et al., 2016; Dafforn et al., 2016). In particular, ecological data are often spatially autocorrelated and can be clustered in few locations (e.g., close to research centers). This violates the assumption of independent observations that many statistical learning methods make (Stojanova et al., 2013). Thus, the ability of such methods to spatially predict indicators of ecosystem state based on currently available regional ecological and stressor data sets remains unexplored.

While the use of statistical and machine learning methods for human impact mapping is a nascent field of research, these methods have at least three advantages in comparison to the more common human impact index models. First, in contrast to predicting an abstract index, statistical learning methods allow direct prediction of quantitative ecological indicators or composite indicators chosen to reflect specific management goals. Second, the accuracy of human impact indices is difficult to assess because it is not clear which measurable quantities the index values should be compared to, especially at broad spatial scales (Stock et al., 2018). In contrast, for statistical and machine learning methods, appropriate cross-validation methods provide sound estimates of spatial prediction error (Roberts et al., 2017). Third, human impact index models at present cannot incorporate natural disturbances, or possible beneficial effects of some human activities on certain biota. In contrast, statistical learning methods can incorporate any predictor for which data are available. This is important if the resulting maps are used to inform coastal management, because human uses of the coast are managed within a background of natural disturbance (Micheli et al., 2016).

The potential of statistical learning to link indicators of marine ecosystem condition to multiple stressors, and to make spatial predictions of ecosystem responses, therefore deserves further development and testing. Our paper expands the previous work by Parravicini et al. (2012) in three ways. First, we test how well statistical learning methods work for spatial prediction of ecological indicators related to coastal human impacts at a regional scale, and with accordingly coarser data. Second, we compare the performance of different model types. This is necessary because which type of model performs best depends on the characteristics of the data. On the one hand, flexible models may be required to adequately represent the interactive and non-linear relationships present in marine ecosystems under stress. However, if sample sizes and the signal-to-noise ratio in the data are small (as in ecosystems with high natural variability), more flexible models are more likely to fit to the noise (overfitting). Thus, we tested model types ranging from relatively inflexible models like ridge regression to very flexible ones like boosted regression trees. Third, while Parravicini et al. (2012) reported good model fit, models used for prediction should be tested against data that were not used to train them. Furthermore, standard methods to test prediction accuracy are over-optimistic if observations are not independent. We therefore estimate the different models' prediction accuracy based on spatial block cross-validation (Roberts et al., 2017).

We used this expanded approach to map three ecological indicators along the coast of California: Biodiversity in kelp forests and in rocky intertidal habitats, and fish biomass in kelp forests. We calculated these indicators based on species-level field observations. As input data (i.e., predictors), we used spatial data representing human activities like fishing effort with different gear types, an overall human impact index for the California Current (Halpern et al., 2009), as well as data representing environmental gradients in the study region. We conclude by discussing the potential of statistical learning for mapping and understanding the cumulative effects of human activities on coastal ecosystems, point out potential statistical pitfalls, and suggest improvements in data collection that are required to make use of this potential.

2. Methods

2.1. Overview

We used a spatially explicit double cross-validation approach to estimate the prediction accuracy of different model types (Fig. 1). We trained ten types of regression models to predict three field-measured ecological indicators based on spatial data representing anthropogenic stressors as well as natural gradients. Table 1 summarizes the responses (dependent variables) and predictors (features or independent variables). Table 2 summarizes the model types tested in this study. To account for non-independent observations, we estimated each model type's prediction accuracy (mean squared error, MSE), by means of spatial block cross-validation (Roberts et al., 2017) using 12 exhaustive

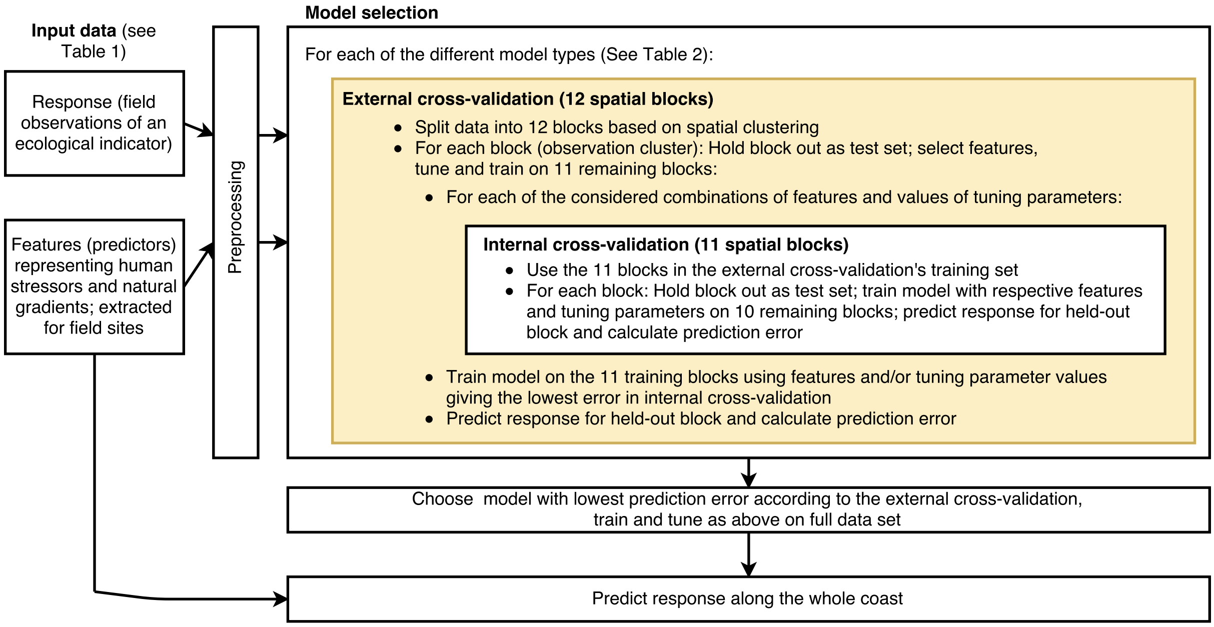

Fig. 1. Overview of the spatially explicit double cross-validation approach to estimate the prediction accuracy of different model types.

and non-overlapping spatial clusters of observations as blocks. All model types required explicit selection of predictors or setting of tuning parameters (e.g. the learning rate for boosted regression trees). However, tuning models to minimize their cross-validation error biases this error estimate. We thus used a double-cross-validation approach (e.g. Filzmoser et al., 2009), selecting variables and values for the tuning parameters based on an internal cross-validation using the 11 spatial locks in the external cross-validation's training sets.

2.2. *Responses: ecological indicators*

We calculated 3 community-level indicators of coastal ecosystem condition (benthic biodiversity and fish biomass of kelp forests, and biodiversity of rocky intertidal habitats) based on field observations collected along the West Coast of the United States by the Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO; Fig. 2). Data were requested directly from PISCO representatives, however since initial analyses were completed, the original count data collected by PISCO have been assigned permanent DOIs (included in Table 1). Few field observations were available for Oregon and Washington, with most concentrated in southern and central California. We therefore limit our analyses to the California coast.

Kelp forest benthic biodiversity was calculated based on species-level invertebrate and algal density data collected along SCUBA swath transects. Species diversity was summarized for each field site by calculating the Shannon index (H; R package *vegan*, Oksanen et al., 2013), which incorporates both species richness and relative abundance. In total, the data yielded 683 field observations of biodiversity for 160 distinct sites between 1999 and 2009. Fish abundance and size was measured from PISCO fish transects, which are conducted at different depth levels. We combined data from benthic and midwater transects and estimated biomass for each transect by means of species-specific

Table 1
Sources and descriptions of the response and predictor data used in this study.

Data	Source	Description
Responses		
Kelp biodiversity	PISCO, doi:https://doi.org/10.6085/AA/pisco_subtidal.161.2	Shannon diversity index calculated based on species-level macroalgae and invertebrate swath SCUBA transects data. Note that the DOI is for raw data, but we obtained a cleaned and summarized version from PISCO representatives (also for fish biomass and rocky intertidal diversity).
Fish biomass	PISCO, doi:https://doi.org/10.6085/AA/pisco_subtidal.150.2	Total biomass estimated based on transect surveys in kelp forests
Rocky intertidal biodiversity	PISCO, doi:https://doi.org/10.6085/AA/pisco_intertidal.52.7	Shannon diversity index calculated based on species-level macroalgae and invertebrate quadrat data
Predictors (features)		
Human impact index	Halpern et al. (2009), https://www.nceas.ucsb.edu/ globalmarine/ca_current_data	Human impact index for the California Current region, downloaded February 10th, 2017.
Human use and stressors	As above	Included: Ocean acidification, ocean deposition, inorganic pollution, organic pollution, invasive species, nutrient runoff, ocean-based pollution, sediment runoff increase, sediment runoff decrease, shipping, sea surface temperature change, coastal trash, UV radiation change, commercial fishing (split into 5 gear categories, e.g. pelagic high-bycatch), recreational fishing
Population density	CIESIN (2016)	Mean 2005 population density on land within 10 km of field observations based on gridded population of the world (GPW), v4
Water depth	Amante and Eakins (2009)	ETOPO1
Sea surface temperature	NODC (2017)	Mean of AVHRR 2000–2009 annual composites
Chlorophyll concentration	NASA (2017)	Mean of Aqua-MODIS 2003–2009 annual composites

Table 2
Overview of model types tested in this study, with information on tuning parameters and feature selection algorithms where applicable. For each model type, we considered the standard model and a variation including spatial weights for the observations.

Code	Model	R package	Feature selection and tuning parameters
NULL	Null model	n/a	n/a
LIN	Ordinary (linear) least squares	Base	Features selected by stepwise selection
LINPoly	Polynomial least squares regression (cubic)	Base	Features selected by stepwise selection
LINInt	Linear least squares regression with pair-wise interactions	Base	Features selected by stepwise selection
PCR	Principal components regression	Base	Tuning parameter: Number of components to include as features
RR	Ridge regression	glmnet	Shrinkage penalty multiplier
Lasso	Least absolute shrinkage and selection operator	glmnet	Shrinkage penalty multiplier
GAM	Generalized additive models	gam	Features and degrees of freedom for each feature selected by stepwise selection
CART	(Classification and) Regression trees	tree	Pruning parameter
RF	Random forests and bagged trees	randomForest	Proportion of predictors to try at each split
BRT	Boosted regression trees	gbm	Number of trees (1000 to 50000), learning rate (0.00005 to 0.1), interaction depth (1 to 5)

length-weight ratios (available from FishBase). In total, these field data comprised 558 field observations for 123 distinct sites between 2001 and 2009. Rocky intertidal field observations consisted of percent cover of invertebrates and macrophytes (macroalgae and the surfgrass, *Phyllospadix* sp.), observed in quadrats along transects at increasing tidal heights. We summarized these species-level observations by calculating Shannon's H. In total, these data yielded 183 field observations for 103 distinct sites between 2001 and 2010. We scaled each indicator to have mean 0 and standard deviation 1.

Many field sites had observations from several years, and were very close to other sites. To mitigate overrepresentation of places with many close-by sampling sites or repeated observations, we aggregated multi-year and close-by observations as follows. We first calculated the means of each indicator for all years with data for field sites containing observations from several years, resulting in a single value for each site and indicator. Then, we grouped close-by field observations of each indicator using agglomerative hierarchical clustering with Euclidean distance between field sites as dissimilarity function and average linkage (Hastie et al., 2009). We chose 1 km as threshold average distance between clusters, thus ensuring that the clusters' centroids fall into different grid cells of the stressor data (each 1 km^2). We summarized the field observations for each cluster by calculating the mean of all sites in a given cluster. This resulted in 136 clustered field observations for kelp biodiversity, 104 clustered field observations for fish biomass, and 95 clustered field observations for rocky intertidal biodiversity, our final sample sizes.

While our predictors covered varying time periods until 2008 (Halpern et al., 2009), the field observations were made between 1999 and 2011. We retained observations made after 2008 for two reasons. First, some sites had no data from the earlier years. Second, because the stressor data covered different years, temporal consistency between all data in the model could not be achieved anyway. We therefore considered the benefits of including data for additional sites to outweigh the consequences of including field observations made after the period covered by the stressor data.

2.3. Predictors: Anthropogenic stressors and natural gradients

In the marine human impact mapping literature, terms like “stressors”, “human activities”, “pressures” and “threats” are sometimes used inconsistently, and the data we use represent a mix of direct human use (like fishing) and proxies for broad-scale anthropogenic change (e.g. satellite-mapped sea surface temperature anomalies). Thus, we use the terms “human activities”, “human use” and “human stressors” broadly and interchangeably to refer to these data sets.

We downloaded the original, normalized data sets used by Halpern et al. (2009) to map human impacts on the California Current large marine ecosystem (Table 1). Some of these data sets contain presence-absence of infrastructure like oil rigs that exist only in the proximity of few or no field observations. We excluded such “rare” stressors because we expected that they would lead to overfitting at the few sites where they occurred. As proxies for overall human impact, we included the final human impact index by Halpern et al. (2009) and the mean population density (CIESIN, 2016) on land within 10 km of each field observation. Furthermore, we included three potential predictors representing broad-scale natural gradients in our study area: 1) Water depth (ETOPO1, Amante and Eakins, 2009); 2) Mean sea surface temperature based on AVHRR annual composites for 2000–2009 (NODC and Rosenstiel School of Marine and Atmospheric Science, 2017); and 3) average chlorophyll concentrations based on MODIS-Aqua annual composites for 2003–2009.

The stressor data were noisy at fine scales, especially along the coast. Thus, whether a field site fell into a given grid cell or into one of its neighbors could lead to very different values of some predictors (Fig. 2D), but these discontinuities are unlikely to represent real differences between close-by locations. To alleviate the effects of such fine-scale noise in the predictor data, we summarized the spatial data for each field observation cluster as the means of all grid cells within a 3 km buffer around the cluster's centroid (Fig. 1D). We then scaled each predictor to mean 0 and standard deviation 1. Effectively, averaging the predictors in buffers around each field observation acted as a low-pass filter, reducing extremes and differences between predictors for close-by sites. Thus, our final units of analysis were long-term average conditions in small coastal areas, calculated based on multiple field observations of the ecological indicators and, respectively, several grid cells of the predictor data.

2.4. Regression models

We tested various statistical and machine learning methods for regression (Table 2). These models are well described in the literature (e.g. Hastie et al., 2009; James et al., 2013). We thus provide only a basic overview of how the different models work. We compared the performance of the models against a null model for each indicator, which always predicts the mean response of the training observations, i.e. does not use any predictors.

The first class of models that we tested were different types of *linear least-squares models* (LIN, LINPoly, LINInt, PCR). For p predictors, these models have the functional form $F(X) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p$. For a given sample, the coefficients b_i are chosen to minimize the residual sum of squares (RSS). While such models are linear functions of the predictors $X_1 \ldots X_p$, the predictors can be calculated by applying non-linear functions to the original input data. For example, polynomial regression models have the same form, but include higher powers of the predictors. We included squared and cubic terms (LINPoly). The coefficients were estimated as for the standard linear model. Similarly, we added pair-wise interactions between predictors (LINInt). Variable selection for these models is described in Section 2.6. In principal

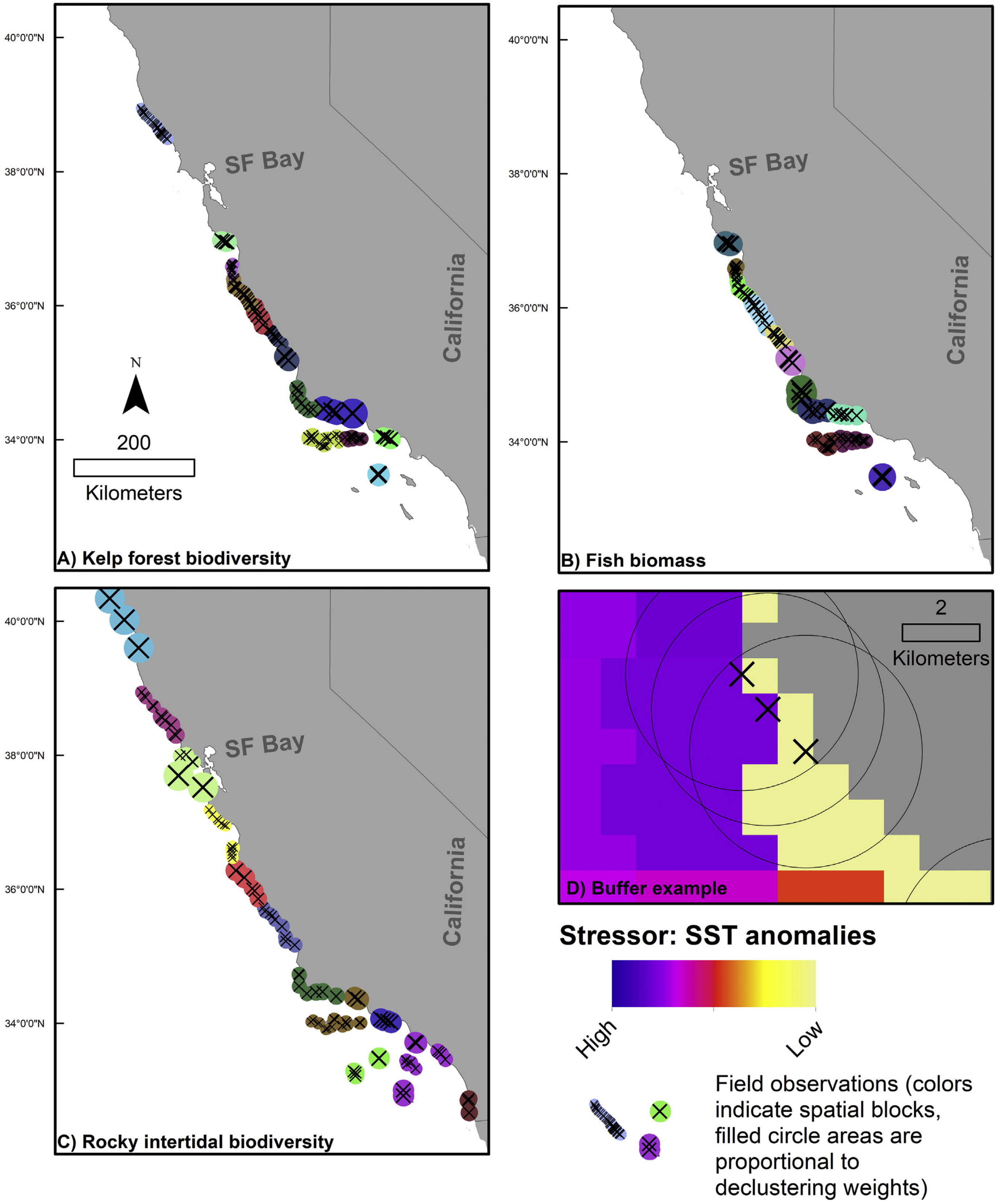

Fig. 2. Field observations and spatial clusters used as blocks in the double spatial block cross-validation (A-C); example of a stressor data set that changes values abruptly between neighboring grid cells, and 3 km buffers (hollow circles) used to smooth the predictors in each field observations' surroundings (D).

components regression (PCR), predictor X_i is the i^{th} principal component of the original predictors. The number of principal components to include is a tuning parameter.

The second class of models that we tested were *shrinkage models* (ridge regression, RR; and least absolute shrinkage and selection operator, Lasso). These models have the same functional form as standard linear models, but the coefficients $b_1 \ldots b_p$ are estimated not to minimize the RSS, but a quantity of the form $RSS + \lambda P$, where P is a penalty for large coefficients and λ is a tuning parameter determining how the penalty is weighted in comparison to the RSS. Ridge regression uses the sum of squared coefficients as penalty, whereas Lasso uses the sum of the coefficients' absolute values.

Generalized additive models (GAMs) are an extension of linear models (Hastie and Tibshirani, 1990) with the functional form $F(X) = b_0 + b_1 s_1(X_1) + b_2 s_2(X_2) + \ldots + b_p s_p(X_p)$, where $s_1 \ldots s_p$ are (possibly non-linear) functions of the predictors. In contrast to e.g. polynomial regression, these functions have flexible shapes that are learned from the data. We used smoothing splines for this purpose. Feature selection thus involved not only choosing which predictors to include, but also choosing the degrees of freedom for each included predictor.

The remaining models tested in this study were *tree-based*. These are flexible models that can per default represent non-linear relationships and interactions. Classification and Regression Trees (CART) are binary trees where at each node, the data are split based on a threshold value of one predictor. Variance is reduced by “pruning”, i.e. choosing the tree size (number of nodes) that minimizes $\Sigma RSS + \lambda\,|T|$, where ΣRSS is the sum of all leaves' RSS and $|T|$ is the size of the tree. As for the shrinkage methods, λ is a tuning parameter that controls the bias-variance tradeoff.

Bagging (bootstrap aggregation) reduces variance by averaging the predictions of many models (in this case, trees) that are fit to bootstrap samples of the original data. Random forests (RF; Breiman, 2001) can further improve prediction accuracy by reducing the correlation between individual trees. This is achieved by allowing each split in a tree to consider only a random subset of the predictors. The number of predictors considered at each split is a tuning parameter. The number of trees must be large enough to reduce the model's variance, but using more trees does not lead to overfitting. We used 2000 trees in each random forest. Because bagged regression trees are a special case of random forests (where all predictors are considered at each split), these models are not distinguished further. The choice between bagged regression trees and random forests was made by setting the tuning parameter.

Boosted regression trees (BRT) are another approach that reduces variance by combining many individual regression trees. BRT first fits a single tree with only few splits. The tree's predictions for the training observations are then multiplied by a small learning rate. A new tree is fit to the *residuals*, its predictions multiplied by the learning rate, and added to the first. Repeating this process many times, BRT approximates the function to be learned in many small steps. Tuning parameters are the learning rate, the size of each tree (interaction depth), and the number of trees (Elith et al., 2008).

2.5. Spatial autocorrelation and spatial weights

Non-independent observations can affect the performance of statistical learning methods and lead to overoptimistic estimates of their prediction error (Burman et al., 1994). While many statistical concerns arise from auto-correlated residuals (e.g. Le Rest et al., 2014), we had to fit thousands of models without knowing which predictors would be included in the final ones. Furthermore, flexible models can overfit to dependence structures in the data. In this case, no residual spatial autocorrelation may be detected. To inform spatial cross-validation, Roberts et al. (2017) thus suggest investigating dependence structures in the raw data. We estimated the spatial distances over which each response or predictor became uncorrelated with itself based on inspection of sample variograms.

We also tested if using spatial declustering weights for training the models improved their predictions. The common polygon- and Kriging-based methods for calculating such weights are sensitive to the study area's boundaries and assign too large weights to observations near the edges of clusters (Bel et al., 2009; Kovitz and Christakos, 2004). Because large parts of our data were almost linear “strings” of observations along the coast, these problems would be especially pronounced. We thus used cell declustering (Isaaks and Srivastava, 1989). However, to account for the close to linear nature of our observation locations, we did not use a regular grid. Instead, we placed square grid cells manually along the coastline, so that each cell contained as many observations as possible but did not cross the boundaries of the spatial blocks used for cross-validation. The cell size was chosen based on the sample variograms of the ecological indicators. The resulting weights are shown in Fig. 2.

For all tested model types except random forests, we used the standard implementations for fitting models with observation weights in the respective R packages (minimizing a weighted RSS). For regression with random forests, no default support for regression with weighted observations was available. We thus used the normalized observation weights as sampling probabilities for bootstrapping, similar to proposals to adjust random forests for classification with imbalanced classes (Chen et al., 2004). We compared the prediction accuracy of all model types trained with and without spatial observation weights.

2.6. Model selection and spatial prediction

We used spatial block cross-validation to estimate the different model types' prediction error. This approach uses large spatial blocks (clusters) of observations as cross-validation folds (Roberts et al., 2017). We identified spatial blocks by means of agglomerative hierarchical clustering with the Euclidean distance between field sites as dissimilarity metric and average linkage (Hastie et al., 2009). We used 12 spatial blocks because this number resulted in a good spatial partitioning of observations upon manual inspection. The spatial blocks are shown in Fig. 2. The smallest blocks contained 3 field observations, whereas the largest blocks contained about 30.

For each of the three responses, we trained the model type with the smallest cross-validated MSE on the full data set in order to make final predictions (Kuhn and Johnson, 2013). For this purpose, we created points 1 km apart along the coastline of California. At each point, we extracted the mean values for all predictors in a 3 km buffer around each point, as for the training data.

2.7. Feature selection and tuning

Feature selection and model tuning were performed using only the external cross-validation's training sets, because selecting predictors or tuning models based on cross-validation results makes these error estimates overoptimistic (Arlot and Celisse, 2010).

We used a stepwise algorithm for the models requiring explicit feature selection (LIN, LINPoly, LINInt, GAM). For linear models, this algorithm begins with the null model, and considers two actions in each step: 1) adding a predictor that was not yet included in the model; or 2) removing a predictor that was already included in the model during an earlier step. In each step, it takes the action leading to the largest reduction of cross-validated MSE. Note that this algorithm is greedy and may not find the global optimum.

For GAMs, we also had to choose the degrees of freedom (df) for each predictor. Thus, the algorithm considered four actions in each step: 1) adding a predictor that was not yet included in the model as a linear function; or 2) removing a predictor that was already included in the model with 1df, i.e. a linear function; or 3) increasing the df of one

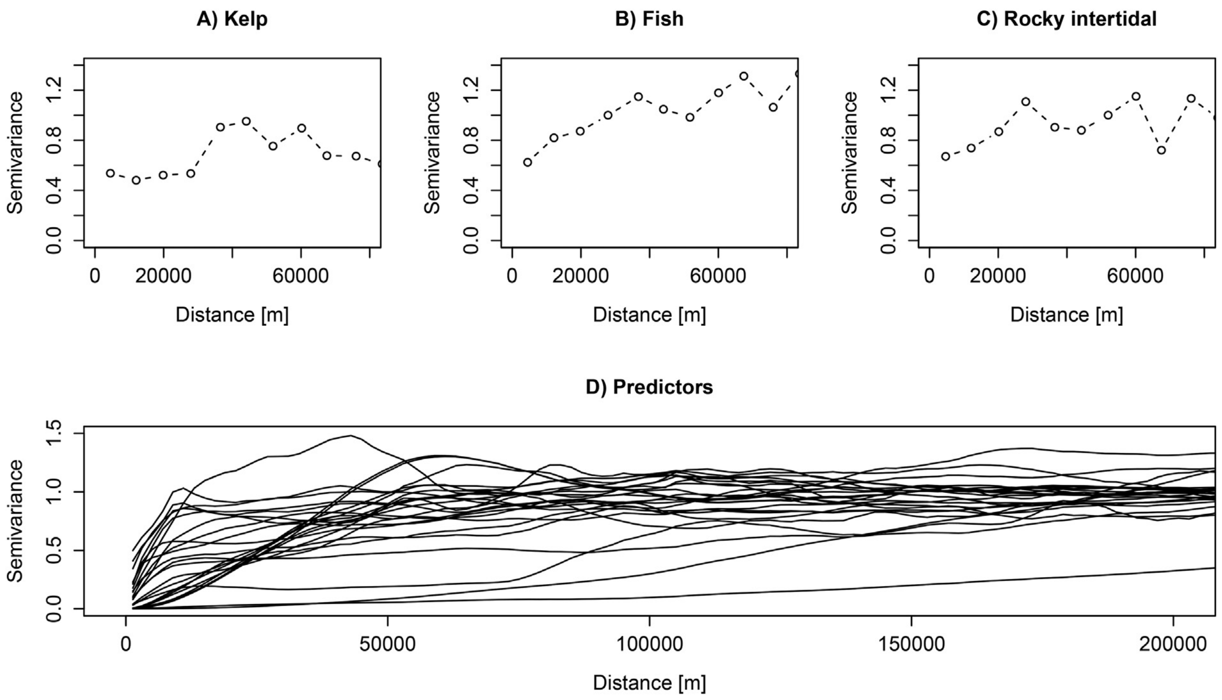

Fig. 3. Omnidirectional sample variograms of the scaled responses (A-C) and predictors (D; calculated for grid cells within 10 km of the coastline).

of the predictors by 1; or 4) decreasing the df of one of the predictors by 1. To limit the potential for overfitting the smoothing splines, we limited the maximum df for each predictor to 5.

All other models required the selection of at least one tuning parameter. As for feature selection, we compared the models based on internal spatial block cross-validation, and set the tuning parameters to the values that resulted in the lowest internal MSE.

3. Results

3.1. Spatial autocorrelation of raw data

Fig. 3 shows omnidirectional sample variograms for the scaled ecological indicators and predictors. Broadly speaking, for all three indicators, the variance first increased with separation distance, reaching a first peak at about 35-40 km, then decreasing again. At larger separation distances, the variograms behaved erratically. We thus used 40 km as cell sizes for calculating grid-based declustering weights. The predictors also showed spatial autocorrelation, with ranges varying from less than 10 to 100 s of kilometers.

Table 3
Mean squared errors (MSEs) from the spatial block cross-validation. The lowest MSEs in each column are shown in bold font. Abbreviations for models as in Table 2, and RI-rocky intertidal.

Model	No spatial weights			Declustering weights		
	KELP	FISH	RI	KELP	FISH	RI
NULL	1.39	1.03	1.16	–	–	–
LIN	2.36	0.81	1.24	2.33	1.18	1.46
LINPoly	2.36	0.81	1.24	2.33	1.18	1.46
LINInt	3.44	2.12	3.61	4.72	2.32	4.78
PCR	**1.06**	1.06	1.34	1.26	0.92	1.50
RR	1.32	**0.78**	1.16	1.39	0.81	1.19
Lasso	1.79	0.84	**1.09**	1.75	0.83	1.10
GAM	2.35	1.02	1.57	2.47	1.12	1.35
CART	1.69	1.05	1.37	1.91	0.98	**1.02**
RF	1.27	0.88	1.11	**1.24**	0.81	1.08
BRT	1.37	0.77	1.10	1.42	**0.75**	1.13

3.2. Model performance

Table 3 shows the MSEs of the tested models, estimated by means of spatial block cross-validation. The best method for predicting kelp biodiversity was principal components regression (PCR) without declustering weights, with an MSE of 1.06, or 76% of the null model MSE. Also ridge regression, random forests and boosted regression trees had a lower MSE than the null model, but considerable larger MSEs than PCR. The best method for predicting fish biomass in kelp forests was boosted regression trees (BRT) with declustering weights (MSE 0.75, or 73% of the null model MSE). However, BRT without declustering weights performed almost as well (MSE 0.77), as did ridge regression without weights (MSE 0.78). For this indicator, most models made better predictions than the null model. The best method for predicting rocky intertidal biodiversity in kelp forests was CART with spatial weights, achieving an MSE of 1.02, or 88% of the null model MSE. Because of the smaller improvement compared to the null model, we conclude that our predictors contained little relevant information for predicting this indicator, and thus do not discuss the model for rocky intertidal biodiversity further.

The following examples only serve to illustrate a potential pitfall when evaluating the performance of statistical and machine learning methods on spatial data. When using 10-fold instead of spatial block cross-validation, more flexible models consistently outperformed the less flexible ones, and estimated prediction errors were much lower. For example, for kelp biodiversity, random forest had the lowest cross-validated MSE (0.49), less than half of the best model's MSE estimated by the spatial block cross-validation and corresponding to 49% of the null model's MSE (when also estimated by 10-fold cross-validation). For fish biomass, the best models according to 10-fold cross-validation were GAM (0.65, 65% of null model MSE), boosted regression trees and random forests (0.67). While the difference between MSEs estimated by the two cross-validation approaches was smaller for fish biomass than for kelp biodiversity, the estimates were still lower than those obtained by spatial block cross-validation. Similar to the 10-fold cross-validated error estimates, out-of-bag MSEs for random forests were lower than the error estimates obtained by spatial block cross-validation (kelp: 0.46; fish: 0.63).

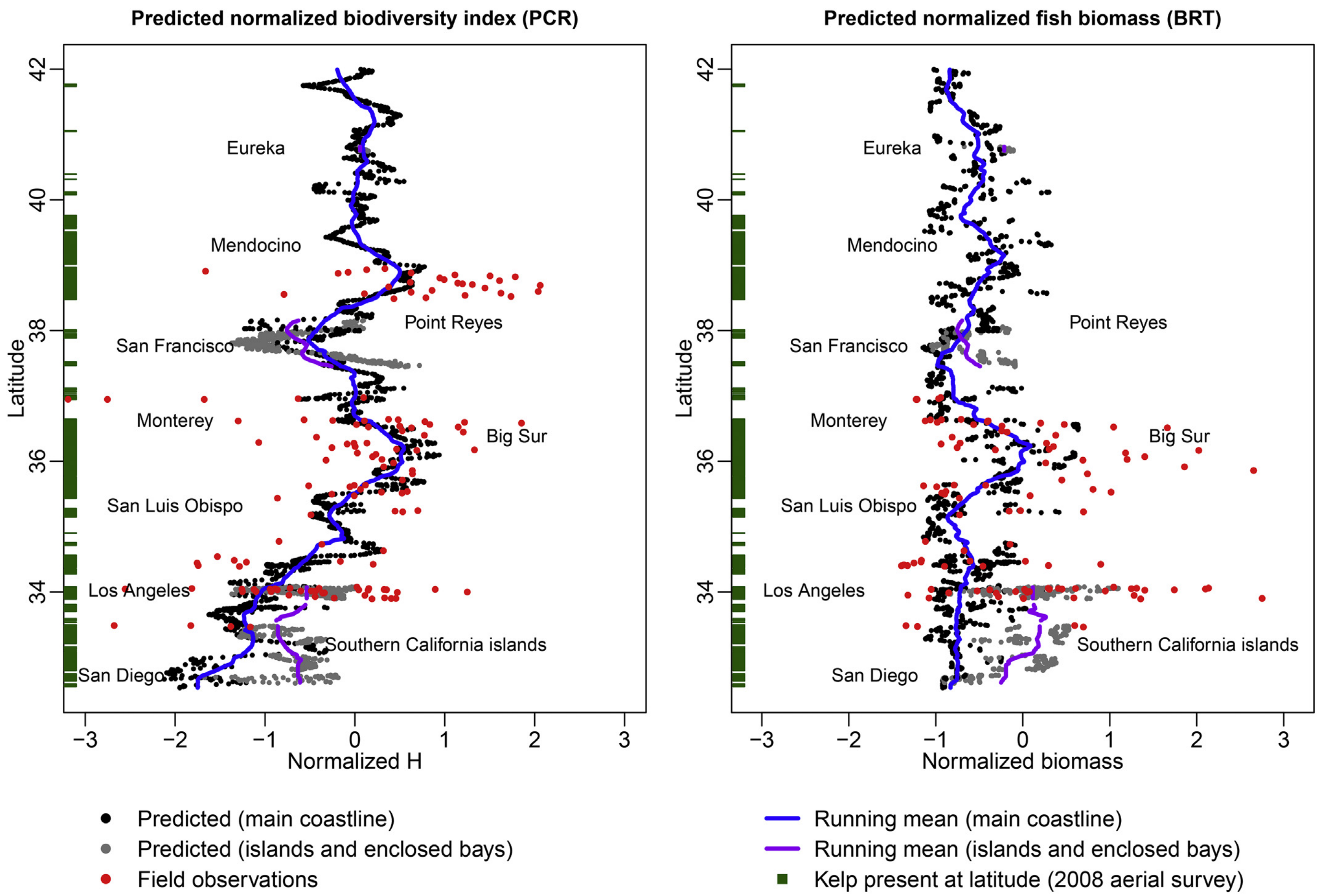

Fig. 4. Model predictions (separate for the mainland's coastline as well as islands and large enclosed bays like San Francisco Bay) for the California coast, summarized by latitude. Green blocks running along the left axes show latitudes where kelp canopy was observed during an aerial survey conducted by the California Department of Fish and Wildlife in 2008 (CDFW, 2008). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Best models

The best model for kelp forest biodiversity (trained on the full data set) was a linear function of the original predictors' first two principal components (PCs). Both PCs had high loadings for many predictors and we could thus could not disentangle the effects of individual predictors or identify important predictor groups based on this model.

The final BRT model for fish biomass had 4000 trees, a learning rate of 0.001, and interaction depth 1 (i.e. did not consider interactions between predictors). The by far most important predictors (measured as reduction in MSE attributable to them) were water depth and recreational fishing, followed (in this order) by demersal destructive fishing, chlorophyll concentration, organic pollution, the full human impact index calculated by Halpern et al. (2009), and demersal non-destructive high-bycatch fishing. The other predictors were much less important. Fish biomass decreased with increasing intensity of all important individual stressors; however, it increased with higher human impact index. Fish biomass also increased with deeper water and with decreasing chlorophyll, which may however serve as proxies for larger distances to the coastline.

3.4. Spatial predictions

Fig. 4 shows the models' predictions after training them on the full data set. It also illustrates the high variability of the ecological indicators, which none of the models captured.

From south to north, the PCR model for kelp forest biodiversity predicted mostly low biodiversity for the mainland, but higher biodiversity for Santa Catalina island, the Channel Islands and other offshore islands in southern California. North of Los Angeles, the model predicted increasing biodiversity, with some of the highest values occurring in the Big Sur area. Moving further north, it predicted decreasing biodiversity reaching low values in the San Francisco Bay Area. The bay itself had still lower predicted biodiversity overall, however with some unrealistic high values. For the coast north of San Francisco, variable but overall high biodiversity was predicted.

The predicted spatial distribution of fish biomass broadly matched that of biodiversity. The BRT model predicted low biomass for most of the mainland coast south of San Luis Obispo, but predicted high biomass for the offshore islands. North of Los Angeles, the model predicted increasing biomass, with some of the highest values again occurring in the Big Sur area. Further north towards Monterey and San Francisco, predicted biomass decreased, but again with some unrealistically high values in the San Francisco Bay. North of San Francisco, predicted fish biomass increased again, reaching a peak just south of Mendocino, then decreasing but staying mostly above the levels predicted for southern California. It should be noted, though, that the fish biomass field data did not include any observations made north of San Francisco.

4. Discussion

The objectives of this paper were 1) to investigate if statistical and machine learning methods trained on regional biological and human stressor data could make accurate spatial predictions of coastal ecological indicators; 2) to provide quantitative estimates of prediction error in comparison to a null model; and 3) to suggest appropriate methods for model selection and error estimation in this setting.

Because of many correlated predictors and high natural variability, we did not attempt to draw causal inferences about drivers of coastal ecosystem state. For example, the best model for fish biomass associated higher values of the full human impact index (Halpern et al., 2009) with higher biomass; we suspect that this is an artifact of using correlated predictors. Where we reported such relationships (e.g. that recreational fishing was one of the most important predictors of coastal fish biomass), we did so to provide a qualitative check of our models, rather than to suggest that we correctly identified what drives the respective aspects of coastal ecosystem condition in our study area.

4.1. Predictions for the California coast

The predictions for kelp forest biodiversity and fish biomass made intuitive sense, because much of California's population is concentrated in the south (Los Angeles and San Diego) as well as the San Francisco Bay area. In contrast, the northern coast, offshore islands and some stretches of the central coast like Big Sur are little developed. Because coastal and upstream population are good general indicators of human influence on coastal ecosystems (Feist and Levin, 2016), the predicted broad-scale spatial distributions appear realistic (higher biodiversity and fish biomass in less populated areas). Also, while we could not disentangle the importance of different predictors for the kelp biodiversity model, the most important predictors for the fish biomass model were related to fishing, natural gradients and pollution, which intuitively makes sense.

However, our predictions should not be seen as representing current coastal ecosystem condition in California, but as in broad strokes representing the overall potential for human impacts over a multi-year period in the first decade of the 2000s for the following reasons. First, our models were trained on field data that are several years old, and California's coastal ecosystems have since experienced major changes, for example mass mortality of invertebrates in 2011 and 2013 from toxic algal blooms and disease (Stokstad, 2014; Jurgens et al., 2015). There has also been unusually warm water throughout the northeastern Pacific from late 2013 to late 2015, which affected the composition and structure of marine and coastal biological communities along much of the North American west coast (Cavole et al., 2016). Second, we trained, tuned and tested our models using ecological indicators calculated from field observations in kelp forests and rocky intertidal habitats. While we then used the best identified models to predict the two kelp-related indicators for the whole coastline of California, kelp forests do not exist everywhere (Fig. 4). Third, while fish biomass is a common indicator for marine resource availability, and biodiversity is of direct conservation concern, other indicators may be more relevant for specific management and conservation objectives. This paper focused on model testing, and we thus chose our three ecological indictors based on data availability. Studies aiming primarily to inform management should more carefully choose and justify which indicators to use as responses (Boldt et al., 2014; Rombouts et al., 2013).

4.2. Evaluating statistical learning methods for spatial prediction

The prediction error of a model has a reducible component that can be improved by choosing an appropriate model type, and an irreducible component. The reducible error again has two components: bias and variance. Bias arises if a model is too inflexible to represent the true relationships between predictors and response. For example, a linear model will have high bias, and therefore a high prediction error, if the relationship being modeled is in fact highly non-linear. In marine ecosystems, non-linear relationships are common (Hunsicker et al., 2016), and the effects of multiple stressors are often interactive (Côté et al., 2016). One of the reasons why flexible models like random forests are promising for predicting multiple stressor effects is their ability to capture non-linear relationships and interactions, and thus achieve low bias. However, more flexible models can have errors with high variance. This means that the model represents peculiarities of the sample rather than general relationships in the population (overfitting). The variance tends to be higher if the sample size and the signal-to-noise ratio are small, as is the case for our ecological field data. It thus was not a priori clear which types of models would work best for our research problem, or how many predictors should be included. We tested various types of models in combination with dimensionality reduction methods such as stepwise variable selection and regularization in order to achieve good compromises between bias and variance, and avoid the inclusion of unnecessary predictors.

Importantly, high variance can only be detected if the models are tested on data not used for training them. This is the rationale for using cross-validation. However, in spite of high local variability of the ecological indicators, some close-by sites had rather similar values for both responses and predictors. Because the field sites were concentrated close to each other in some locations, the observations were not all independent. But in this case, standard cross-validation error estimates can be overoptimistic (Hijmans, 2012), and favor too flexible models (Le Rest et al., 2014). To obtain sound cross-validation error estimates based on spatially auto-correlated data and select the best models and predictors, observations must therefore be assigned to folds such that test observations are sufficiently far away from the training observations (Burman et al., 1994; Le Rest et al., 2013, 2014). Indeed, we found that relying on 10-fold cross-validation and out-of-bag errors would have led us to underestimate the models' prediction errors, and in the case of kelp choose a too flexible model, as expected. Future research using statistical and machine learning methods to link human pressures to ecosystem condition should carefully rule out this problem. Spatial block cross-validation is a good solution when observations are spatially dependent (Roberts et al., 2017).

4.3. Model performance

We compared the predictive skill of different models using spatial block cross-validation. Compared to null models, the best of these methods improved prediction errors by approximately 25% for biodiversity and fish biomass in kelp forests. This suggests that the tested predictors captured some relevant spatial trends. However, the models failed to predict the high local variability of the indicators. This may be caused by factors that operate at fine spatial scales (like local disease outbreaks), but which were excluded from the models because no data were available; by temporal variability of both stressors and ecological conditions that could not be captured in the models because of low temporal resolution and temporal mismatch of the available data; and by the coarse spatial resolution of some data sets used as predictors. Furthermore, modeling rocky intertidal biodiversity led to smaller improvements compared to the null model.

It should be noted, though, that spatial block cross-validation tends to produce comparatively high estimates of prediction error, because observations from a whole sub-region of the study area are withheld when training models. Because our models were intended to predict the indicators for the whole coast of California, and long stretches of the coast had no observations, we consider estimating prediction performance for unobserved sub-regions to be an appropriate test. Furthermore, maps of ecological indicators and cumulative human impacts have the potential to inform coastal management. In this case, making wrong predictions may have serious real-world consequences. Subjecting the models to a “harder” test is therefore preferable to using methods that may provide over-optimistic error estimates.

4.4. Lessons for future research

Statistical and machine learning methods are increasingly used in multiple stressor research. While works like Parravicini et al. (2012), Large et al. (2015), Teichert et al. (2016), Samhouri et al. (2017) and Jones et al. (2017) show that cumulative effects studies based on statistical learning have much potential, we found that in our study area, low signal-to-noise ratio in combination with small sample sizes lead to overfitting of the more flexible models and thereby prevented us from harnessing one of their major advantages. We conclude that data quality was a main barrier to using the flexible methods to their full potential, and to thereby gaining insights into the complex effects of multiple stressors. We thus join a growing body of research calling for data collection that is tailored to the demands of multiple stressor effect studies (Dafforn et al., 2016; Nõges et al., 2016).

Our results suggest four lessons for future research using statistical and machine learning to link marine ecosystem condition to human or environmental stressors. First, when testing models, it is important to account for non-independent observations, as explained in Section 4.2. Second, given the typically small sample sizes of ecological data, it is important to limit noise, to the greatest extent possible, in the response variables, especially if stressor data are only available as long-term averages. One way to achieve this is using composite indicators that change only slowly over time, like the marine territory status or the ecological quality ratios that Parravicini et al. (2012) and Teichert et al. (2016) predicted. Third, human use and stressor data available for our study area were noisy at fine spatial scales. Furthermore, they were only available for inconsistent time periods, a common issue in multiple stressor modeling studies (Murray et al., 2015). This is especially problematic because ecological impacts depend on the timing of stressors (Cheng et al., 2015; Wu et al., 2017), and fine temporal resolution of both predictors and response variables would be needed to detect effects of short-term stress with moderate recovery times. If this is not possible, focusing on persistent stressors or stressors with long associated recovery times is more promising; such stressors with long-term impacts are also of major conservation concern. For these reasons, while more ecological data are becoming available through legally mandated, harmonized reporting (Schinegger et al., 2016) and through promising new methods for constructing indicators (e.g. based on environmental DNA; Kelly et al., 2014), equal improvements of human use and stressor data sets are necessary if statistical and machine learning is to shed new light on and produce accurate maps of multiple stressor effects. Fourth, while we focused on spatial prediction, statistical learning methods can also identify important stressors, interactions, and thresholds (Jones et al., 2017; Large et al., 2013, 2015; Samhouri et al., 2017; Teichert et al., 2016). However, soundly disentangling the effects of several predictors is difficult if they are correlated (Freckleton, 2011; Grueber et al., 2011). For example, while random forests are able to make good predictions also with correlated predictors, estimates of predictor importance can be biased in this situation (Nicodemus et al., 2010).

Studies aiming to use statistical and machine learning to draw inferences about multiple stressors should thus approach data collection as designing a quasi-experiment. For example, if no manipulation is possible in the field, ecologists often use gradients of the variables of interest to mimic experimental designs (e.g. Wernberg et al., 2010). Unfortunately, stressors often vary along similar gradients. For example, many stressors originate from population centers (Mach et al., 2017), or follow a land-sea gradient (e.g. pollution runoff) along which also natural conditions change. Nevertheless, at least some stressors can be geographically separated if study areas and field sites are chosen appropriately. For example, along-coast pollution gradients can exist surrounding the mouths of major rivers that drain densely populated and industrialized areas, and protected area networks exclude stressors like fishing from specific locations (Sala et al., 2012). Such places where stressors are geographically separated are thus preferable study areas for investigating multiple stressor effects, and data collection sites should be chosen to capture the geographic variation of stressors. In any case, statistical and machine learning studies drawing conclusions about ecosystem responses to multiple stressors must carefully rule out the possibility that inferences are disturbed by correlated predictors.

Acknowledgments

AS and FM acknowledge the support of the US NSF-CNH Program (award no. DEB-1212124) and a grant from Stanford's Catalyst for Collaborative Solutions program. AS was also supported by a Goldman Fellowship from the Woods Institute for the Environment and MM through Stanford's Center for Ocean Solutions. Meg Caldwell, Larry Crowder and David Lobell provided helpful comments on an early draft of this paper. This study utilized data collected by the Partnership for Interdisciplinary Studies of Coastal Oceans: a long-term ecological consortium funded by the David and Lucile Packard Foundation and the Gordon and Betty Moore Foundation. We thank PISCO for providing us with these data, and NCEAS for making the stressor data used in this study available.

References

Afflerbach, J.C., Yocum, D., Halpern, B.S., 2017. Cumulative human impacts in the Bering Strait region. Ecosyst. Health Sustain. 3 (8), 1379888.

Allan, J.D., McIntyre, P.B., Smith, S.D., Halpern, B.S., Boyer, G.L., Buchsbaum, A., ... Steinman, A.D., 2013. Joint analysis of stressors and ecosystem services to enhance restoration effectiveness. Proc. Natl. Acad. Sci. 110 (1), 372–377.

Amante, C. and B.W. Eakins, 2009. ETOPO1 1 Arc-minute global relief model: procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:https://doi.org/10.7289/V5C8276M. Accessed on 4/23/2017.

Andersen, J.H., Stock, A., Heinänen, S., Mannerla, M., Vinther, M., 2013. Human uses, pressures and impacts in the eastern North Sea. 136 *Aarhus University, DCE-Danish Centre for Environment and Energy*.

Andersen, J.H., Halpern, B., Korpinen, S., Murray, C., Reker, J., 2015. Baltic Sea biodiversity status vs. cumulative human pressures. *Estuarine*. Coastal Shelf Sci. 161, 88–92.

Andersen, J.H., Berzaghi, F., Christensen, T., Geertz-Hansen, O., Mosbech, A., Stock, A., ... Wisz, M.S., 2017. Potential for cumulative effects of human stressors on fish, sea birds and marine mammals in Arctic waters. Estuar. Coast. Shelf Sci. 184, 202–206.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4, 40–79.

Baird, D.J., Van den Brink, P.J., Chariton, A.A., Dafforn, K.A., Johnston, E.L., 2016. New diagnostics for multiply stressed marine and freshwater ecosystems: integrating models, ecoinformatics and big data. Mar. Freshw. Res. 67 (4), 391–392.

Ban, N.C., Alidina, H.M., Ardron, J.A., 2010. Cumulative impact mapping: advances, relevance and limitations to marine management and conservation, using Canada's Pacific waters as a case study. Mar. Policy 34 (5), 876–886.

Bel, L., Allard, D., Laurent, J.M., Cheddadi, R., Bar-Hen, A., 2009. CART algorithm for spatial data: application to environmental and ecological data. Comput. Statistics Data Anal. 53, 3082–3093.

Boldt, J.L., Martone, R., Samhouri, J., Perry, R.I., Itoh, S., Chung, I.K., ... Yoshie, N., 2014. Developing ecosystem indicators for responses to multiple stressors. Oceanography 27 (4), 116–133.

Borja, A., 2014. Grand challenges in marine ecosystems ecology. Front. Mar. Sci. 1, 1.

Borja, A., Elliott, M., Andersen, J.H., Berg, T., Carstensen, J., Halpern, B.S., ... Rodriguez-Ezpeleta, N., 2016. Overview of integrative assessment of marine systems: the ecosystem approach in practice. Front. Mar. Sci. 3, 20.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brown, C.J., Saunders, M.I., Possingham, H.P., Richardson, A.J., 2014. Interactions between global and local stressors of ecosystems determine management effectiveness in cumulative impact mapping. Divers. Distrib. 20 (5), 538–546.

Burman, P., Chow, E., Nolan, D., 1994. A cross-validatory method for dependent data. Biometrika 351–358.

Cavole, L.M., Demko, A.M., Diner, R.E., Giddings, A., Koester, I., Pagniello, C.M., Zill, M.E., 2016. Biological impacts of the 2013–2015 warm-water anomaly in the Northeast Pacific: winners, losers, and the future. Oceanography 29 (2), 273–285.

CDFW, 2008. Kelp canopy shapefiles from 2008 aerial surveys. Available online: https://www.wildlife.ca.gov/Conservation/Marine/Kelp/Aerial-Kelp-Surveys downloaded on June 20th, 2018.

Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn imbalanced data. In: Technical Report 666. Statistics Department, University of California at Berkeley Available at. http://www.stat.berkeley.edu/users/chenchao/666.pdf.

Cheng, B.S., Bible, J.M., Chang, A.L., Ferner, M.C., Wasson, K., Zabin, C.J., ... Grosholz, E.D., 2015. Testing local and global stressor impacts on a coastal foundation species using an ecologically realistic framework. Glob. Chang. Biol. 21 (7), 2488–2499.

CIESIN, 2016. Center for International Earth Science Information Network, Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Density. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4NP22DQ. (Accessed 2017-05-17).

Clark, D., Goodwin, E., Sinner, J., Ellis, J., Singh, G., 2016. Validation and limitations of a cumulative impact model for an estuary. Ocean Coast. Manag. 120, 88–98.

Coll, M., Steenbeek, J., Sole, J., Palomera, I., Christensen, V., 2016. Modelling the cumulative spatial–temporal effects of environmental drivers and fishing in a NW Mediterranean marine ecosystem. Ecol. Model. 331, 100–114.

Côté, I.M., Darling, E.S., Brown, C.J., 2016. Interactions among ecosystem stressors and their importance in conservation. Proc. R. Soc. B 283 (1824), 20152592.

Crain, C.M., Kroeker, K., Halpern, B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecol. Lett. 11 (12), 1304–1315.

Dafforn, K.A., Johnston, E.L., Ferguson, A., Humphrey, C.L., Monk, W., Nichols, S.J., ... Baird, D.J., 2016. Big data opportunities and challenges for assessing multiple stressors across scales in aquatic ecosystems. Mar. Freshw. Res. 67 (4), 393–413.

Doubleday, Z.A., Jones, A.R., Deveney, M.R., Ward, T.M., Gillanders, B.M., 2017. Eight habitats, 38 threats and 55 experts: assessing ecological risk in a multi-use marine region. PLoS One 12 (5), e0177393.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.

Feist, B.E., Levin, P.S., 2016. Novel indicators of anthropogenic influence on marine and coastal ecosystems. Front. Mar. Sci. 3, 113.

Filzmoser, P., Liebmann, B., Varmuza, K., 2009. Repeated double cross validation. J. Chemom. 23 (4), 160–171.

Foley, M.M., Mease, L.A., Martone, R.G., Prahler, E.E., Morrison, T.H., Murray, C.C., Wojcik, D., 2017. The challenges and opportunities in cumulative effects assessment. Environ. Impact Assess. Rev. 62, 122–134.

Freckleton, R.P., 2011. Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. Behav. Ecol. Sociobiol. 65 (1), 91–101.

Giakoumi, S., Halpern, B.S., Michel, L.N., Gobert, S., Sini, M., Boudouresque, C.F., ... Pergent, G., 2015. Towards a framework for assessment and management of cumulative human impacts on marine food webs. Conserv. Biol. 29 (4), 1228–1234.

Gissi, E., Menegon, S., Sarretta, A., Appiotti, F., Maragno, D., Vianello, A., ... Barbanti, A., 2017. Addressing uncertainty in modelling cumulative impacts within maritime spatial planning in the Adriatic and Ionian region. PLoS One 12 (7), e0180501.

Grueber, C.E., Nakagawa, S., Laws, R.J., Jamieson, I.G., 2011. Multimodel inference in ecology and evolution: challenges and solutions. J. Evol. Biol. 24 (4), 699–711.

Halpern, B.S., Fujita, R., 2013. Assumptions, challenges, and future directions in cumulative impact analysis. Ecosphere 4 (10), art131.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., ... Fujita, R., 2008. A global map of human impact on marine ecosystems. Science 319 (5865), 948–952.

Halpern, B.S., Kappel, C.V., Selkoe, K.A., Micheli, F., Ebert, C.M., Kontgis, C., ... Teck, S.J., 2009. Mapping cumulative human impacts to California current marine ecosystems. Conserv. Lett. 2 (3), 138–148.

Halpern, B.S., Frazier, M., Potapenko, J., Casey, K.S., Koenig, K., Longo, C., ... Walbridge, S., 2015. Spatial and temporal changes in cumulative human impacts on the world/'s ocean. Nat. Commun. 6.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. vol. 43 CRC press.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning. Springer, New York, pp. 485–585.

Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93 (3), 679–688.

Hodgson, E.E., Halpern, B.S., 2018. Investigating cumulative effects across ecological scales. Conserv. Biol. https://doi.org/10.1111/cobi.13125.

Hoegh-Guldberg, O., Bruno, J.F., 2010. The impact of climate change on the world's marine ecosystems. Science 328 (5985), 1523–1528.

Hunsicker, M.E., Kappel, C.V., Selkoe, K.A., Halpern, B.S., Scarborough, C., Mease, L., Amrhein, A., 2016. Characterizing driver-response relationships in marine pelagic ecosystems for improved ocean management. Ecol. Appl. 26 (3), 651–663.

Isaaks, E.H., Srivastava, R.M., 1989. Applied Geostatistics. Oxford University Press, New York.

Jackson, J.B., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J., ... Hughes, T.P., 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science 293 (5530), 629–637.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. vol. 6 Springer, New York.

Jones, F.C., Plewes, R., Murison, L., MacDougall, M.J., Sinclair, S., Davies, C., ... Gunn, J., 2017. Random forests as cumulative effects models: a case study of lakes and rivers in Muskoka, Canada. J. Environ. Manag. 201, 407–424.

Jurgens, L.J., Rogers-Bennett, L., Raimondi, P.T., Schiebelhut, L.M., Dawson, M.N., Grosberg, R.K., Gaylord, B., 2015. Patterns of mass mortality among rocky shore invertebrates across 100 km of northeastern Pacific coastline. PLoS One 10 (6), e0126280.

Kelly, R.P., Port, J.A., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F., ... Crowder, L.B., 2014. Harnessing DNA to improve environmental management. Science 344 (6191), 1455–1456.

Korpinen, S., Andersen, J.H., 2016. A global review of cumulative pressure and impact assessments in marine environment. Front. Mar. Sci. 3, 153.

Korpinen, S., Meski, L., Andersen, J.H., Laamanen, M., 2012. Human pressures and their potential impact on the Baltic Sea ecosystem. Ecol. Indic. 15 (1), 105–114.

Kovitz, J.L., Christakos, G., 2004. Spatial statistics of clustered data. Stochastic Environ. Res. 18, 147–166.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.

Large, S.I., Fay, G., Friedland, K.D., Link, J.S., 2013. Defining trends and thresholds in responses of ecological indicators to fishing and environmental pressures. ICES J. Mar. Sci. 70 (4), 755–767.

Large, S.I., Fay, G., Friedland, K.D., Link, J.S., 2015. Quantifying patterns of change in marine ecosystem response to multiple pressures. PLoS One 10 (3).

Le Rest, K., Pinaud, D., Bretagnolle, V., 2013. Accounting for spatial autocorrelation from model selection to statistical inference: application to a national survey of a diurnal raptor. Ecol. Informatics 14, 17–24.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Glob. Ecol. Biogeogr. 23 (7), 811–820.

Levin, P.S., Fogarty, M.J., Murawski, S.A., Fluharty, D., 2009. Integrated ecosystem assessments: developing the scientific basis for ecosystem-based management of the ocean. PLoS Biol. 7 (1), 23.

Mach, M.E., Wedding, L.M., Reiter, S.M., Micheli, F., Fujita, R.M., Martone, R.G., 2017. Assessment and management of cumulative impacts in California's network of marine protected areas. Ocean Coast. Manag. 137, 1–11.

Mason, J.G., Rudd, M.A., Crowder, L.B., 2017. Ocean research priorities: similarities and differences among scientists, policymakers, and fishermen in the United States. Bioscience 67 (5), 418–428.

Micheli, F., Halpern, B.S., Walbridge, S., Ciriaco, S., Ferretti, F., Fraschetti, S., ... Rosenberg, A.A., 2013. Cumulative human impacts on Mediterranean and Black Sea marine ecosystems: assessing current pressures and opportunities. PLoS One 8 (12), e79889.

Micheli, F., Heiman, K.W., Kappel, C.V., Martone, R.G., Sethi, S.A., Osio, G.C., ... Tanner, J.M., 2016. Combined impacts of natural and human disturbances on rocky shore communities. Ocean Coast. Manag. 126, 42–50.

Murray, C.C., Agbayani, S., Alidina, H.M., Ban, N.C., 2015. Advancing marine cumulative effects mapping: an update in Canada's Pacific waters. Mar. Policy 58, 71–77.

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, 2017. Aqua MODIS Chlorophyll concentration, OCI a.

Nicodemus, K.K., Malley, J.D., Strobl, C., Ziegler, A., 2010. The behaviour of random forest permutation-based variable importance measures under predictor correlation. BMC Bioinforma. 11 (1), 110.

NODC, Rosenstiel School of Marine and Atmospheric Science, 2017. AVHRR Pathfinder Level 3 Annual Daytime SST Version 5. Ver. 5. PO.DAAC, CA, USA. Dataset accessed [2017-04-23] at. https://doi.org/10.5067/PATHF-AND50.

Nõges, P., Argillier, C., Borja, Á., Garmendia, J.M., Hanganu, J., Kodeš, V., ... Birk, S., 2016. Quantified biotic and abiotic responses to multiple stress in freshwater, marine and ground waters. Sci. Total Environ. 540, 43–52.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., ... Wagner, H., 2013. Package 'Vegan'. R Community Ecology Package, Version. vol. 2(9).

Parravicini, V., Rovere, A., Vassallo, P., Micheli, F., Montefalcone, M., Morri, C., ... Bianchi, C.N., 2012. Understanding relationships between conflicting human uses and coastal ecosystems status: a geospatial modeling approach. Ecol. Indic. 19, 253–263.

Poloczanska, E.S., Burrows, M.T., Brown, C.J., García Molinos, J., Halpern, B.S., Hoegh-Guldberg, O., ... Sydeman, W.J., 2016. Responses of marine organisms to climate change across oceans. Front. Mar. Sci. 3, 62.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., ... Warton, D.I., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40 (8), 913–929.

Rombouts, I., Beaugrand, G., Artigas, L.F., Dauvin, J.C., Gevaert, F., Goberville, E., ... Travers-Trolet, M., 2013. Evaluating marine ecosystem health: case studies of indicators using direct observations and modelling methods. Ecol. Indic. 24, 353–365.

Sala, E., Ballesteros, E., Dendrinos, P., Di Franco, A., Ferretti, F., Foley, D., ... Guidetti, P., 2012. The structure of Mediterranean rocky reef ecosystems across environmental and human gradients, and conservation implications. PLoS One 7 (2), e32742.

Samhouri, J.F., Andrews, K.S., Fay, G., Harvey, C.J., Hazen, E.L., Hennessey, S.M., ... Stier, A.C., 2017. Defining ecosystem thresholds for human activities and environmental pressures in the California Current. Ecosphere 8 (6).

Schinegger, R., Palt, M., Segurado, P., Schmutz, S., 2016. Untangling the effects of multiple human stressors and their impacts on fish assemblages in European running waters. Sci. Total Environ. 573, 1079–1088.

Stock, A., Micheli, F., 2016. Effects of model assumptions and data quality on spatial cumulative human impact assessments. Glob. Ecol. Biogeogr. 25 (11), 1321–1332.

Stock, A., Crowder, L.B., Halpern, B.S., Micheli, F., 2018. Uncertainty analysis and robust areas of high and low modeled human impact on the global oceans. Conserv. Biol. https://doi.org/10.1111/cobi.13141.

Stojanova, D., Ceci, M., Appice, A., Malerba, D., Dzeroski, S., 2013. Dealing with spatial autocorrelation when learning predictive clustering trees. Ecol. Informa. 13, 22–39.

Stokstad, E., 2014. Death of the stars. Science 344 (6183), 464–467.

Teichert, N., Borja, A., Chust, G., Uriarte, A., Lepage, M., 2016. Restoring fish ecological quality in estuaries: implication of interactive and cumulative effects among anthropogenic stressors. Sci. Total Environ. 542, 383–393.

Thrush, S., Lewis, N., Le Heron, R., Fisher, K., Lundquist, C., Hewitt, J., 2016. Addressing surprise and uncertain futures in marine science, marine governance, and society. Ecol. Soc. 21 (2).

Wernberg, T., Thomsen, M.S., Tuya, F., Kendrick, G.A., Staehr, P.A., Toohey, B.D., 2010. Decreasing resilience of kelp beds along a latitudinal temperature gradient: potential implications for a warmer future. Ecol. Lett. 13 (6), 685–694.

Wu, P.P., Mengersen, K., McMahon, K., Kendrick, G.A., Chartrand, K., York, P.H., Rasheed, M.A., Caley, M.J., 2017. Timing anthropogenic stressors to mitigate their impact on marine ecosystem resilience. Nat. Commun. 8, 1263.