FEATURE ARTICLE

Using machine learning to link spatiotemporal information to biological processes in the ocean: a case study for North Sea cod recruitment

Bernhard Kühn*, Marc H. Taylor, Alexander Kempf

Thünen Institute of Sea Fisheries, Herwigstraße 31, 27570 Bremerhaven, Germany

ABSTRACT: Marine organisms are subject to environmental variability on various temporal and spatial scales, which affect processes related to growth and mortality of different life stages. Marine scientists are often faced with the challenge of identifying environmental variables that best explain these processes, which, given the complexity of the interactions, can be like searching for a needle in the proverbial haystack. Even after initial hypothesis-based variable selection, a large number of potential candidate variables can remain if different lagged and seasonal influences are considered. To tackle this problem, we propose a machine learning framework that incorporates important steps in model building, ranging from environmental signal extraction to automated variable selection and model validation. Its modular structure allows for the inclusion of both parametric and machine learning models, like random forest. Unsupervised feature extractions via empirical orthogonal functions (EOFs) or self-organising maps (SOMs) are demonstrated as a way to summarize spatiotemporal fields for inclusion in predictive models. The proposed framework offers a robust way to reduce model complexity through a multi-objective genetic algorithm (NSGA-II) combined with rigorous cross-validation. We applied the framework to recruitment of the North Sea cod stock and investigated the effects of sea surface temperature (SST), salinity and currents on the stock via a modified version of random forest. The best model (5-fold CV r^2 = 0.69) incorporated spawning stock biomass and EOF-derived time series of SST and salinity anomalies acting through different seasons, likely relating to differing environmental effects on specific life-history stages during the recruitment year.

Linking spatiotemporal information to North Sea cod recruitment via machine learning.

Image: B. Kühn, M. Taylor; Gears from https://commons.wikimedia.org/wiki/File:Gear_7.svg *(CC-BY-SA license)*

1. INTRODUCTION

Although there are still challenges ahead, machine learning (ML) is entering marine science on a broad scale (Malde et al. 2020). Automated fish identification from images (Allken et al. 2019), age-reading from otoliths (Moen et al. 2018), identification of

*Corresponding author: bernhard.kuehn@thuenen.de

plankton (Schröder et al. 2020) and advances in forecasting oceanographic phenomena such as El Niño–Southern Oscillation (Ham et al. 2019) are among the numerous applications of ML in marine science. Additionally, the rapid emergence of high-quality ocean data products from ocean model output, from both reanalysis and remote sensing, constitutes a vast resource of spatiotemporal data related to physical and low-trophic-level biological processes.

Classical ML approaches involve steps like feature extraction (e.g. important signal in high-dimensional data) and feature selection (e.g. which features to include in the final model) to be partly overseen by a human expert, before the final model is built (van Ginneken 2017). In contrast, deep learning offers the opportunity for complete automation of a model-building pipeline, with these steps being learned from the data via different layers at the expense of requiring a large amount of data (van Ginneken 2017). However, many problems in marine science do not fall in the category of big data; e.g. time series from higher trophic levels are often aggregated at coarser time steps with <100 data points. This is the case for many fish stock–recruitment (SR) time series, which are among the longest in biology, with some dating back to the 1950s, but containing only a single data point per year. In addition to their limited length, recruitment signals are largely obscured by environmental variability acting on various spatial and temporal scales throughout the life cycle of the organism. Attempts to improve recruitment prediction by taking environmental processes into account have a long history, yet are still considered to be notoriously difficult because the choice of the environmental signal to include can be subjective (Myers 1998, Houde 2008, Subbey et al. 2014). Although deep learning methods are hardly applicable in such situations, there is particular interest in more precise and robust forecasts of population parameters from living marine resources due to their importance in terms of economic value and food security. However, even in these data-limited situations, fisheries science can benefit from harnessing advances in ML as an alternative to classic statistical methods.

Various attempts have already been made to use ML methods for the prediction of stock–recruitment relationships, including neural networks (Chen & Ware 1999, Megrey et al. 2005, Krekoukiotis et al. 2016), random forest (Hansen et al. 2015, Smoliński 2019) and naïve Bayes (Fernandes et al. 2010, 2015). Although it is believed that these methods can improve predictions of recruitment (Megrey et al. 2005), they are not a standard tool among fisheries

scientists. This is attributed, in part, to unfamiliarity with ML (Olden et al. 2008) and the perception of ML as a 'black box', with limited access on how the algorithm derived its prediction (Burrell 2016). However, recent advances (e.g. Goldstein et al. 2015, Biecek 2018) in understanding variable importance and local and global model behaviour have helped to shed light on this black box and have aided in the interpretation and understanding of predictions. Additionally, step-by-step model building guidelines, like that of Fernandes et al. (2010, 2015), spanning steps like filter-based variable selection and robust cross-validation, are valuable to an understanding of, and adoption of, ML tools. Although the use of ML methods in data-limited situations is advancing, there is still a rather narrow focus on the supervised ML algorithm, i.e. the part which performs a regression or classification task on a response variable. Other aspects like feature extraction from spatiotemporal data and variable selection are often neglected.

We believe that many spatiotemporal data sources are underutilised in studies aimed at investigating their influence on fish life cycle dynamics, as it is common practice to rather use pre-processed and readily available large-scale indices (like the North Atlantic Oscillation [NAO] or Atlantic Multidecadal Oscillation [AMO]) (Dippner 1997a, Gröger & Fogarty 2011, Sguotti et al. 2020) or simple averages of environmental variables of interest (e.g. sea surface temperature [SST]). Large-scale indices offer the advantage of condensing multiple, also unobserved, variables to a simple index over a broad region, which can summarize effects on highly migratory species like Atlantic bluefin tuna (Faillettaz et al. 2019). However, the generalisation of large-scale indices might overlook more local or regional conditions of importance to a given fish stock. Therefore, we argue for a hypothesis-driven extraction of local environmental variables in the species' range, as large-scale indices like the NAO exert their effect through local changes in the physical environment (Drinkwater et al. 2003), which can reveal a more direct relationship. The use of dimension reduction methods in an ML pipeline can offer additional ways to condense the multivariate nature of spatiotemporal datasets to a few modes of variability.

A common tool used in oceanography and atmospheric sciences is empirical orthogonal function (EOF) analysis, which is able to extract dominant signals as linear combinations of the original dataset. This can be sufficient if the underlying relations are indeed linear, but there is a controversy around

whether atmospheric and ocean processes are inherently linear or non-linear in nature (Wunsch 1999, Rudnick & Davis 2003, Hsieh et al. 2005). Hsieh et al. (2005) argued that biological time series are indeed non-linear, but oceanographic processes are more or less linear stochastic processes. However, as atmospheric processes can be multimodal, the mean might not be the most frequently observed state (Hauser et al. 2015). Therefore, extending the toolbox to capture non-linear system states as well is crucial for summarizing spatiotemporal environmental data. A promising method is the use of self-organising maps (SOMs) (Kohonen 1982). These unsupervised clustering algorithms have some advantages over EOFs for recovering asymmetric and non-linear features from data (Liu & Weisberg 2005, Liu et al. 2006), albeit at the expense of only providing a discretized time series. SOMs have received some attention in the domain of oceanography in a descriptive sense, but they are not yet used for predictive modelling in the marine biological sciences.

By incorporating feature extraction via EOFs or SOMs as a first step in the ML pipeline, we want to bridge the gap between the big-data side of environmental data and the data-limited side of marine biological data in this study. As a case study, we chose North Sea cod recruitment, since our understanding of stock–recruitment relationships is still one of the unresolved fields in fisheries science (Subbey et al. 2014), which could benefit from new insights offered by ML.

We argue for a complete end-to-end ML framework that offers various tools for unsupervised feature extraction, automated feature selection and the final model-building with model interpretation to help marine biology scientists in adoption of ML approaches.

2. METHODS

2.1. North Sea cod case study

The North Sea cod *Gadus morhua* stock is one of the best-studied stocks in the North Atlantic, with data available since the early 1960s. The spawning stock biomass (SSB) peaked in the 1970s, together with other gadoids (ICES 2018). This peak is referred to as the 'gadoid outburst'. Since then, the population has generally been in decline, with an all-time low in the years 2004–2006. The population somewhat recovered until 2015, which was followed again by a decline in recent years. Recruitment of age-1 fish

generally follows the trend in SSB, but with a regime of very low recruitment for the last 20 yr (Fig. 1). In addition to a direct impact of SSB, various hypotheses have been put forward to explain the recruitment pattern of cod, with the most prominent ones being changes in temperature and zooplankton prey dynamics (Sundby 2000, Beaugrand et al. 2003, Beaugrand & Kirby 2010). Other environmental variables, like salinity and currents, have gained less attention, although there are reports of a negative association with low-salinity waters (<34‰) along the Norwegian Trench (Svendsen et al. 1991), and various studies found associations of cod larvae with frontal zones in this area (Munk et al. 1995, 1999, Munk 2007, 2014). For currents, most links have been made indirectly through the NAO, with the causal mechanism relating to the intrusion of calanoid zooplankton prey via the northern inflow (Planque & Fromentin 1996, Stephens et al. 1998), which correlates with the NAO and westerly wind patterns (Mathis et al. 2015). Recently, many studies have focussed on assessing the vulnerability of early life stages of cod to drift via individual-based drift models (Peck et al. 2009, Jonsson et al. 2016, Huserbråten et al. 2018, Kvile et al. 2018, Romagnoni et al. 2020). Since our focus here is a demonstration of the general ML framework, we limited our choice of environmental variables to spatiotemporal fields of temperature (SST), salinity and currents, for which data have been available since the beginning of the cod-recruitment time series (1963–2017) and are hypothesised to play a role in the life cycle of North Sea cod.

It is widely acknowledged that recruitment dynamics of fish can be influenced at various times during their life cycle (Houde 2008, Hare 2014), ranging
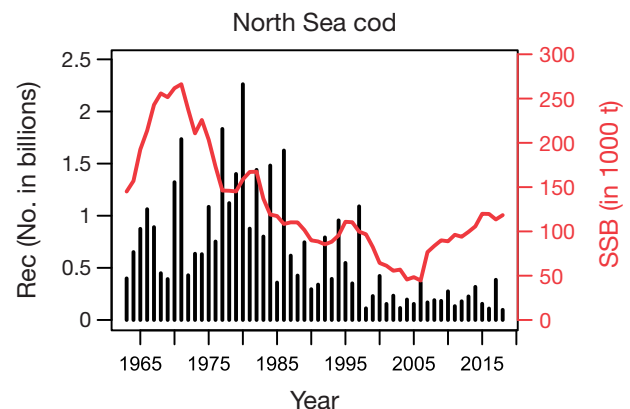

Fig. 1. Stock dynamics of North Sea cod depicting the recruitment (Rec, black bars) and spawning stock biomass (SSB, red line) based on the 2018 assessment data ICES (2018)

from the time of oocyte development in adults, to the egg, larval and juvenile phases until recruits enter the fishery. The recruitment process of cod in the North Sea starts in autumn, when oocytes develop over the winter (Kjesbu et al. 2010). Spawning takes place between January and March, with regional differences between the south and the north (González-Irusta & Wright 2016). The pelagic eggs develop to larvae and then juveniles, at which point they enter a bottom-dwelling life style involving demersal feeding at the beginning of summer (Bastrikin et al. 2014). After overwintering, the age-1 juveniles are first considered as recruits. To fully capture the potential effects of the environment on cod recruitment, we also considered lagged environmental variables at these critical phases in the life of North Sea cod (Table 1).

SST and salinity fields of the North Sea (5° W–12° E and 49–62° N) were obtained from the Adjusted Hydrography Optimal Interpolation (AHOI) (Núñez-Riboni & Akimova 2015) dataset with a spatial resolution of 0.2° × 0.2° in an extended run ranging from 1948 to 2017 (I. Núñez-Riboni pers. comm.) with a monthly temporal resolution. SST was defined as the average over the first 5 m, to be consistent with other SST datasets based on buoys and ship bulk-measurements, which are often taken between 1 and 5 m depth (Emery 2003). Salinity fields were averaged over a depth of 100 m, to reduce the effects of rainfall on the salinity signal in the coastal areas.

For spatiotemporal fields of currents in the North Sea (5° W–12° E and 49–62° N), the upper layer (5.02 m depth) of the ORAS4-reanalysis dataset was used, with a spatial resolution of 1° × 1° and monthly temporal resolution ranging from 1958 to 2017 (Balmaseda et al. 2013) and downloaded from the Integrated Climate Data Center of the University of Hamburg (https://icdc.cen.uni-hamburg.de/daten/ reanalysis-ocean/easy-init-ocean/ecmwf-ocean-reana lysis-system-4-oras4.html).

Time series of North Sea cod (stock cod.27.47 d20) recruitment at age class 1 and SSB were retrieved from the official assessment summary (http://standardgraphs.ices.dk/stockList.aspx), spanning the years 1963–2018. For further analysis, recruitment was log-transformed and SSB was lagged by 1 yr prior to match the age of recruitment.

Monthly anomalies for all environmental spatiotemporal fields were calculated by subtracting the long-term monthly mean values from each spatial grid. Afterwards, mean seasonal anomalies were calculated over 3 mo intervals (winter: December–February [DJF]; spring: March–May [MAM]; summer: June–August [JJA]; autumn: September–November [SON]).

2.2. ML framework

We developed a framework to summarise spatiotemporal information via EOF analysis or SOM feature extraction in an unsupervised way (with unlabelled data) and link them to biological variables of interest (e.g. cod recruitment time series) in a supervised learning approach (with labelled data). The framework consists of 3 main steps (Fig. 2): (1) preprocessing (e.g. anomaly calculation) of spatiotemporal data and feature extraction via unsupervised EOF or SOM analysis and joining with the response variable of interest in a common dataset; (2) feature selection with a cross-validated multi-objective wrapper in which an optimal subset of features is selected within the supervised training phase of the model; and (3) interpretation of the final model with the selected features. The different parameters used within the ML pipeline are summarized in Table 2.

Table 1. Life cycle stages of North Sea cod and time windows in which environmental effects are hypothesised to take place. SON: fall (September–November); DJF: winter (December–February); MAM: spring (March–May); JJA: summer (June–August)

Stage	Possible effect on recruitment	Time period	Season index
Pre-spawning	Environmental conditions during the pre-spawning phase in autumn affecting spawning stock conditioning	Autumn lag 2	SON_{t-2}
Spawning/egg phase	Environmental conditions during spawning and within the egg phase	Winter lag 1	DJF_{t-1}
Late egg phase/larval phase	Environmental conditions during the end of the egg phase and throughout the whole larval phase	Spring lag 1	MAM_{t-1}
Juvenile pre-recruit phase	Environmental conditions during the juvenile phase	Summer lag 1	JJA_{t-1}
		Autumn lag 1	SON_{t-1}
Juvenile recruit during first overwintering	Environmental conditions during the first winter	Winter lag 0	DJF_t

2.2.1. Feature extraction

Feature extraction is a way to describe dominant patterns of variation in a high-dimensional dataset within a lower-dimensional space. This is especially important when dealing with dynamic data in oceanography and meteorology, as there are at least 2 spatial dimensions (latitude, longitude), with depth
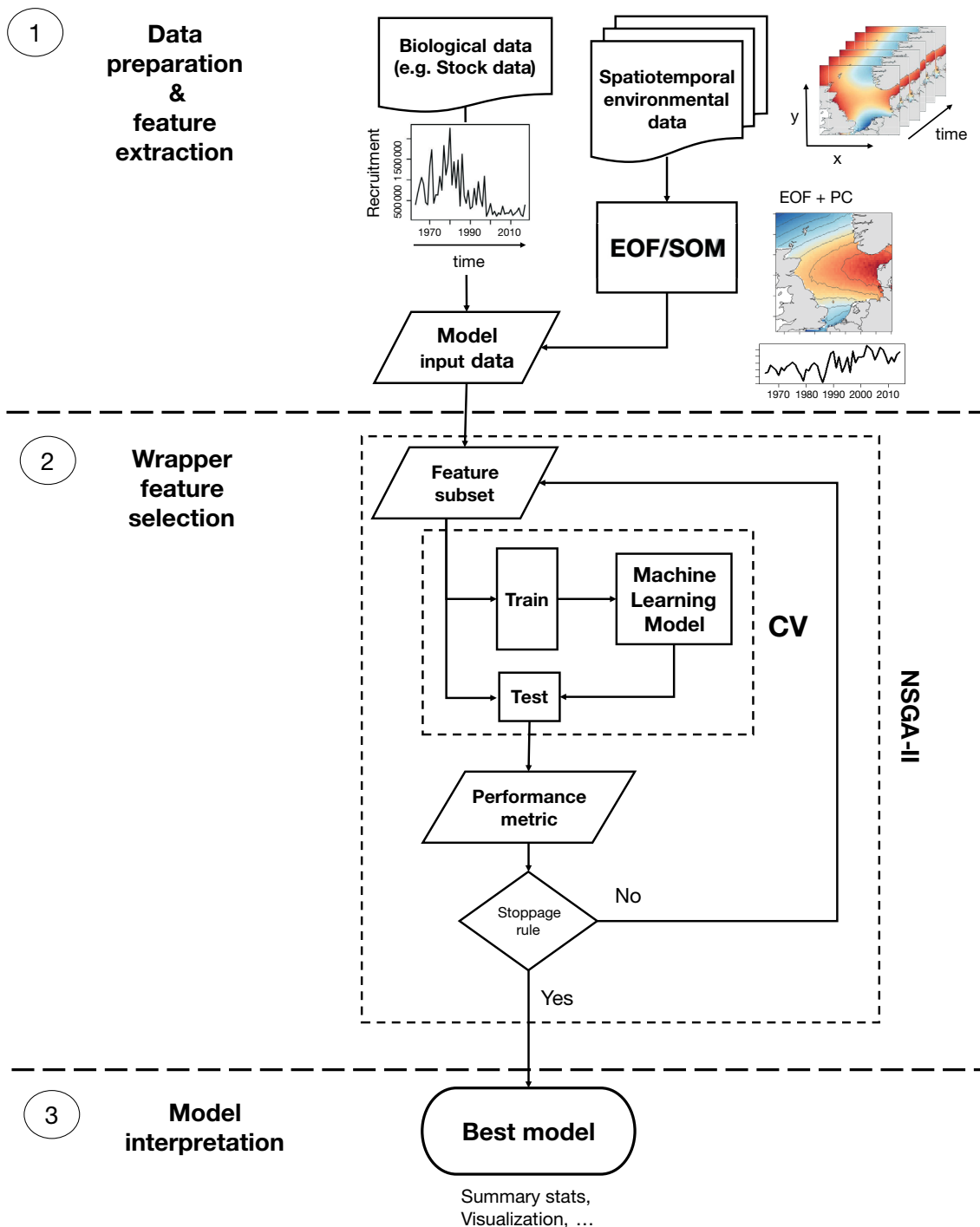

Fig. 2. Overview of the workflow in the proposed framework, combining (1) unsupervised feature extraction of spatiotemporal environmental data via empirical orthogonal functions (EOFs) or self-organising maps (SOMs) with (2) a robust cross-validated (CV) wrapper-based feature selection via a non-dominated sorting genetic algorithm (NSGA-II), to predict biological processes (e.g. recruitment) with a supervised machine learning model and (3) tools for further model interpretation

Table 2. Parameters used for different algorithms. SOM: self-organising map; NSGA-II: non-dominated sorting genetic algorithm; mtry: number of variables randomly sampled as candidates at each split; CV: cross-validation; N: total number of features; p: number of features selected in the evaluated model

Parameter	Value
SOM	
Type	Stochastic
Distance	Euclidean
Neighbourhood	Rectangular
Neighbourhood–function	Bubble
Neighbourhood–radius	Linear decreasing (2/3 of maximum unit-to-unit distance to 0)
Learning rate	Linear decreasing (0.05 to 0.01)
NSGA-II	
Crossover operator	Uniform crossover
Selection operator	Simple selection
Mutation operator	Bitflip mutation
Crossover rate	0.8
Mutation rate	1/N
Population size	100
Maximum offspring size (variable)	100
Maximum number of runs	500
Stop criterion	100
Random forest/ extreme randomized trees	
Number trees	200
mtry	$\sqrt{p}$
Number random splits	1
n-Repeated k-fold CV	5 × 30

considered in some cases, and a temporal dimension is involved.

EOFs. A common method in the earth sciences to decompose spatiotemporal environmental fields is via EOF analysis or principal component analysis (S-mode PCA) (Jolliffe 2002). It decomposes a time-dynamic spatial field into a linear combination of orthogonal spatially stationary structures (eigenvectors or EOFs) and their oscillation in time (principal components [PCs]) (Björnssin & Venegas 1997). As EOF analysis is a linear method, it has problems capturing non-linear spatiotemporal relationships. In addition, the orthogonality constraint, which ensures that the derived spatial fields are orthogonal and corresponding time series are uncorrelated, limits the physical interpretation of EOFs as true modes of climate variability and can result in mixing of modes that are not orthogonal in nature. The analysis was done as described by Wilks (2006) for scalar-valued fields (SST and salinity) and vector-valued fields (currents). To ensure the exclusion of noise in the fur-

ther analysis, the number of PCs included was truncated by a form of Horn's parallel analysis (Horn 1965), whereby the 95%-bootstrapped confidence interval of the eigenvalues was compared against the eigenvalues of an EOF null model with randomly permuted time series at each spatial location. Only PC time series with eigenvalues greater than corresponding eigenvalues of the null model were retained and used as features in the further analysis.

SOMs. SOMs (Kohonen 1982) are unsupervised neural networks that find a low dimensional embedding/clustering of large data sets (Sang et al. 2008). SOMs were introduced in climate and ocean sciences as an alternative to EOF analysis to find dominant modes and patterns (Liu & Weisberg 2011). SOMs project a multi-dimensional input to pre-defined 2-dimensional maps, so that objects which are closely related in the input space are mapped to neighbouring units on the map (topological preserving). Each node on the map is associated with a corresponding weights vector, with the same dimension as the input vector. In its stochastic version, the algorithm is trained iteratively by projecting a random vector from the input to the closest unit on the map (best-matching unit, BMU), thereby changing the weights of the selected and neighbouring units by moving them closer to the input points. This results in a self-organisation of the grid to more closely resemble the input space. To gain a low-dimensional mapping of spatiotemporal fields, the input is organised in an $m \times n$ – matrix with m (rows) denoting time and n (columns) the spatial dimension. The resulting SOM weights then describe the spatial pattern and the association of a BMU to each time step, the time progression of the spatial pattern (Liu et al. 2016b). The advantage of SOMs over EOF analysis is that they are able to detect non-linear relationships, with the cost of not exactly conserving the amplitude of the data (Reusch et al. 2005). Additionally, the produced maps can be readily interpreted as physical processes, as they are in the same units as the original dataset. Moreover, the problem of mode-mixing, i.e. the blending of distinct spatial patterns frequently observed in EOF analysis, is well handled by SOMs as they do not enforce orthogonality or require rotations (Reusch et al. 2005, Liu et al. 2006). The drawback, however, is the *a priori* decision on the number of patterns (nodes on the map) and the discreteness of the resulting time series of pattern occurrence. The latter can be an advantage in using SOMs as input for ML models, as some models show better performance with categorical features because they take non-linear relationships into account

(Myllymäki et al. 2002, Lustgarten et al. 2008, Gupta et al. 2010). However, in the case of linear relationships, one might lose statistical power by using categorical features (Myllymäki et al. 2002). For an objective *a priori* decision on the SOM grid structure and associated number of patterns, we used the minimum of the dynamic validity index (DVI) proposed by Shen et al. (2005) for *k*-means clustering. It finds a trade-off between the compactness of each cluster and its separateness from other clusters. The BMU time series were then used as covariates in the analysis. Further details on the SOM analysis and the DVI can be found in Text S1 in Supplement 1 at www.int-res.com/articles/suppl/m664p001_supp.pdf.

2.2.2. Feature selection

Search algorithms. Feature subset selection is a way of removing redundant and uninformative variables (i.e. features) in an ML model. The goal is to avoid overfitting, increase model performance, reduce running and training time and gain a better model understanding with a reduced set of features (Iguyon & Elisseeff 2003, Guyon et al. 2006, Saeys et al. 2007). There are 3 selection approaches: (1) filter methods: the number of features is reduced independently before model fitting; (2) wrapper methods: feature selection is done during model fitting; and (3) embedded methods: natural feature selection is a built-in part of the model. Filter methods select features before model training, with the aim of selecting variables with a high correlation to the response and low correlation to other variables. Filtering is computationally fast, but potentially misses out on features that are important in interactive relationships with other variables. In contrast, wrapper methods find a best feature set during model fitting and therefore use interaction between features and the model, with higher computational costs. Embedded algorithms are intrinsic to a specific learning algorithm (e.g. feature importance in random forest or boosting ridge regression), mostly with fewer computational costs than wrappers. In our framework, we chose a wrapper feature selection method to harness both the full strengths of the ML model and to allow flexibility in the choice of the model.

Wrapper algorithms consist of 3 broad categories: (1) exhaustive search, (2) greedy search and (3) metaheuristic global search algorithms (Diao & Shen 2015, El Aboudi & Benhlima 2016). Although exhaustive search guarantees the detection of a global optimal solution, it is only applicable if the search space is

small, as the computational cost grows exponentially with the number of features (2^N model evaluations). Greedy search algorithms, like stepwise selection (forward, backward) or recursive feature elimination, add/remove a feature to/from the model and look at the performance gain/drop. They are computationally fast, but the order of features added/removed plays an important role in the obtained result. The search is limited to a narrow area of the search space and is easily trapped in local optima (Xue et al. 2016). Metaheuristic global search algorithms are efficient optimisation strategies that mostly harness a way of stochastic search to escape local optima. Popular methods are population-based metaheuristics that often utilise a form of evolutionary search strategy (e.g. genetic algorithm) to track the improvement of a set of possible solutions simultaneously (Xue et al. 2013). Usually, they optimise a single goal, namely model performance. However, because feature subset selection is by definition a multi-objective problem, representing a trade-off between minimising the number of features in a model and maximising performance, multi-objective feature selection methods have gained attention. They explicitly take into account the number of features as part of the optimisation to force the algorithm to explore the lower parameter end of the search space. Various applications have used the advantages of multi-objective feature selection methods (Hamdani et al. 2007, Xue et al. 2013, Ribeiro et al. 2015), but they are still not widely known outside the field of ML and computer science, and we are not aware of any application in marine sciences.

Non-dominated sorting genetic algorithm. One commonly used multi-objective evolutionary algorithm is the non-dominated sorting-based genetic algorithm in its second improved version (NSGA-II) proposed by Deb et al. (2002). NSGA-II and its modifications have already been used in a feature selection context (e.g. Hamdani et al. 2007, Huang et al. 2010, Ribeiro et al. 2015) and despite their age, still show good performance compared to newer algorithms (Xue et al. 2013). NSGA-II is a genetic algorithm (GA), which employs the idea of Darwin's survival of the fittest to optimisation problems utilising concepts like natural selection, mating and mutation. To employ GA in a feature selection setting, the number of covariates in the model is encoded into a binary vector ('individual') denoting if a variable ('gene') is included (1) or not (0). A set of individuals with differing genes forms a 'population' of potential candidate solutions (Fig. 3). The population evolves by allowing individuals with a higher
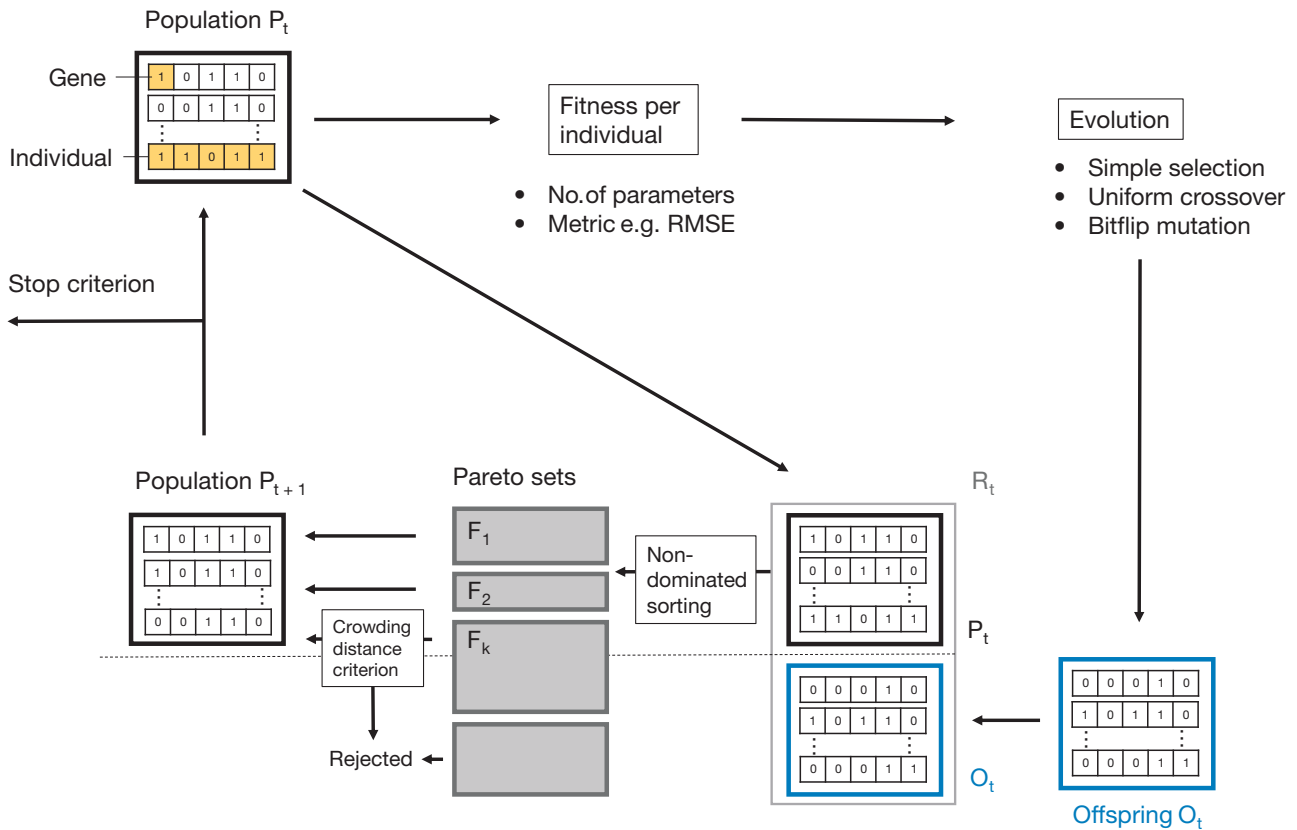
Fig. 3. Schematic of NSGA-II feature selection, beginning with a population *P* at time step *t* consisting of various candidate models (individuals), which are encoded in a binary vector showing which feature (gene) is included into the model (1) or not (0). Each individual has an assigned fitness consisting of the number of features and the metric used to evaluate the goodness of fit of a machine learning model (in this case RMSE). The population now evolves by means of selection, mating (crossover) and mutation and produces offspring. The original population and its offspring form a new population *R* and are sorted based on the concept of non-dominance into different Pareto fronts. From this ordered population, a new population for time step *t*+1 is formed, which has the same size as the original one. Pareto fronts that do not entirely fit into the new population will be either accepted or rejected based on crowding distance, which prefers solutions which lie in less occupied areas of the fitness space. The optimisation is finished if the maximum number of simulation steps is reached or a stopping condition is satisfied

'fitness', the optimisation objective, to produce 'offspring'. A crossover operator determines how the individual vectors are mixed to produce new solutions. Genes (the covariates) that produce fitter offspring are more likely to be included in the next generation. Random changes ('mutations') to the genes of an individual increase the diversity of the population and provide a chance to escape local optima. To ensure that good solutions are not lost in the evolutionary process, a portion of the most fit individuals from the parent generation pass over to the next generation ('elitism'). In the multi-objective case, the evolution of individuals with higher fitness is determined by multiple explicit goals simultaneously, which can be described in terms of Pareto-set theory. Solutions are Pareto-optimal if an improvement in one objective leads to the worsening in the other

objective and vice versa. The set of Pareto-optimal solutions is called the 'Pareto front', consisting of only non-dominated solutions, for which no equal or better solution in terms of all other objectives can be found. NSGA-II tries to guide the search towards the global Pareto front by simultaneously utilising the search strategy of GA together with a ranking of solutions based on non-dominance (Fig. 3). For that, the parent P_t and offspring population O_t are combined to form a new population R_t and sorted based on the principle of non-domination, which ensures both convergence towards the optimal Pareto front and elitism as previous and current population members are included. Individuals belonging to the same non-dominated set $F_{1...k}$ are then grouped together, and the best sets are subsequently chosen until the size of the new population equals the previous one.

For the last set F_k, which does not fully fit into the next population, the individuals are sorted based on a crowded-comparison operator, which preserves diversity by favouring solutions that are not in the same neighbourhood in the objective space. The process is repeated until the convergence of the solution (stop criterion) or until the pre-defined number of simulations steps is reached. Convergence was measured by translating the bi-objective optimisation problem into a single-objective one by means of the hypervolume indicator (Zitzler & Thiele 1998, 1999). It measures the volume (or area in a bi-objective case) enclosed by the Pareto front and a reference point, allowing a comparison of different Pareto sets with the same reference point. As the reference point, we chose $r = (1,1)$ and scaled the different objectives accordingly. Due to the discrete nature of solutions with respect to the number of parameters, the original NSGA-II algorithm suffers from an accumulation of duplicate solutions. To preserve diversity in the population, decrease selection bias and increase performance, duplicates among offspring were removed (Aguirre & Tanaka 2007). For faster computation, we included parallel computing of individuals in the population and memoisation (Michie 1968), the efficient storage of previously visited solutions in the feature space.

Fitness function. In a multi-objective case, the fitness function reflects the different objectives – in the case of feature selection, namely the number of parameters and the performance of the ML model. Here, we used the root mean squared error RMSE = $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{pred} - y_{obs})^2}$ as a metric to assess model performance and guide the evolutionary search. To track the convergence with the hyper-volume indicator, the RMSE of the model was scaled by the RMSE of a naïve prediction (in this case the mean value of the response variable, y_{obs}) and set to 1 if predictions were performing worse.

Choice of the final solution from the Pareto front. The solutions obtained from the search consist of non-dominating solutions, meaning the possibly best model under a given size of parameters. The decision on a final compromise solution was based on the Euclidean distance of each solution within the Pareto front to an ideal solution (in this case the coordinate origin [0,0]) and finding the 'knee' of this curve. We used this measure for its simplicity and interpretability; however, there are many more measures from the field of multi-criteria decision making that can be used for choosing a compromise solution within a Pareto front (see e.g. Wang & Rangaiah 2017 for an overview).

2.2.3. ML model

For the wrapper approach, the ML model is part of the fitness function. In general, the framework is flexible enough to allow for implementation of any kind of classification or regression model. For the cod case study, we decided to use a modified version of random forest (Breiman 2001), as the algorithm is flexible enough to consider non-linear relationships and interactions. Random forest is an ensemble of uncorrelated decision trees. By themselves, they are weak learners that are sensitive to small perturbations in the training data (low bias, high variance). To reduce the variance, random forest uses bootstrap aggregation (or short: 'bagging') of the training data to train each tree with a bootstrap sample and averaging across all predictions. In addition, candidate variables for each split are randomly selected among the input variables, which leads to more diversity in trees. The key idea behind random forest is to have many uncorrelated trees acting as a committee to gain a stable prediction, in which the individual errors of each tree cancel each other out. A way to further reduce correlation of individual forests is an algorithm called 'extreme randomized trees' by Geurts et al. (2006), where in addition to the random selection of features for splitting, the cut point for the split itself is also determined randomly. The original 'extreme randomized trees' algorithm builds the trees on the full dataset; however, to aid comparison with the original random forest algorithm, we decided to keep the original bagging step, characteristic for random forest. The advantage of the additional randomization is a smoother prediction. To show the effects of the proposed changes, we compared the performance and predictions for both algorithms on our case study in Fig. S1 in Supplement 2. To (1) ensure stability of results, (2) avoid overfitting and (3) assess the predictive performance on unseen data, we used repeated n-fold cross-validation (5 folds × 30 permutations) of the model within the NSGA-II algorithm.

2.2.4. Interpretation of model output

ML models are often referred to as black-box models, with a potentially large number of parameters, making it hard to understand how the model uses its inputs to derive a prediction (Biecek 2018). However, in marine science in general, and fisheries science in particular, it is crucial to understand predictions; otherwise, it would be difficult to gain the trust of

human decision makers. In recent years, advances in model-agnostic methods, which are applicable for various types of models, further fostered the interpretation of black-box ML models. In this framework, we used methods to gain understanding of feature importance and local and global model behaviour. Feature importance was assessed by permuting each feature randomly and looking at the increase in prediction error, a concept originally proposed by Breiman (2001) for random forests and extended for other models by Fisher et al. (2018). 'Pd plots' were used to assess the global effect of each individual feature on the response. Local effects were examined via 'breakdown plots', which try to decompose each model prediction into the additive contribution of each feature (Staniak & Biecek 2018). To enable a compact visualisation, we scaled the variable contribution produced by the breakdown analysis, to obtain a measure in percentages for each data point. Additionally, to understand potential mechanisms for cod recruitment, the chosen features in the best EOF-based model were correlated to various regional indices of climate variability including box averages of SST, salinity, currents, precipitation, wind, SST and salinity fronts and large-scale indices like the NAO (for the detailed methodology and results, see Text S2, Fig. S2 and Table S1, all in Supplement 3). This final step sheds light upon linking particular PCs to physical mechanisms. The choice of these additional datasets was based on the accessibility of readily available datasets and hypothesis-driven selections, and is similar in its approach to the analysis of Núñez-Riboni & Akimova (2017), who related inter-annual variability of salinity in the North Sea to box averages of wind, currents, sea level pressure, precipitation, discharge and salinity in specific regions.

2.3. Software implementation

The analysis was done in the programming language R version 3.6.1 (R Core Team 2019). Apart from the base-R functionality, several additional packages were used for this analysis, including 'kohonen' (Wehrens & Buydens 2007, Wehrens & Kruisselbrink 2018) for SOM-feature extraction; 'ecr 2.0' (Bossek 2017) for feature selection via NSGA-II; 'R.cache' (Bengtsson 2019) for memoisation; and the 'caret' (Kuhn 2008) and 'ranger' packages (Wright & Ziegler 2017) for implementing the ML model and cross-validation. Model interpretation was done with the help of the 'pdp' package

(Greenwell 2017) as well as 'DALEX' (Biecek 2018) and 'iBreakDown' (Version 1.3.1; Gosiewska & Biecek preprint, https://arxiv.org/abs/1903.11420) for breakdown analysis. The code for the framework was compiled into the R-package 'marmalaid' v0.1 (Kühn 2020) under the repository https://github.com/BernhardKuehn/marmalaid. Data and code to comprehend the main steps of the North Sea cod case study are distributed via figshare.com at https://doi.org/10.6084/m9.figshare.13473162.

3. RESULTS

3.1. Feature extraction

From the EOF analysis, the largest number of significant PCs were extracted from the salinity fields yielding 7, 6, 5 and 6 PCs for the seasons DJF, MAM, JJA and SON, respectively. SST fields yielded 5, 4, 4 and 5 PCs, whereas the current fields resulted in 2, 3, 3 and 3 extracted PCs for the respective seasons (see Figs. S3–S14 in Supplement 4 for spatial fields and time series). Combined with their lagged versions and the time series of SSB, 82 predictor variables were used as input for the feature selection algorithm.

The number of optimal patterns extracted via SOM varied between 4 (Currents JJA) and 10 (SST SON, Salinity MAM), with the most frequent number of patterns being 6 (Figs. S15–S26 in Supplement 4). Rather than having 1 abrupt minimum in the DVI, the optimal solutions were similar to neighbouring SOM complexities, indicating the validity of various grid choices within a certain range (Fig. S27 in Supplement 4). Altogether, only 19 input time series were presented to the feature selection algorithm.

The extracted spatial EOF-patterns were similar among seasons; however, the seasonal PC time series differed within a given year, indicating substantial seasonal variation (compare spatial pattern and associated time series for each environmental variable in Figs. S3–S14). The EOF1 patterns for SST and salinity show a global pattern over the whole domain, with the largest variability associated with the eastern North Sea for SST and especially coastal regions around the German Bight for salinity (Figs. S3–S10). They capture the warming trend of the North Sea in the case of SST, and for salinity, the intrusions of North Atlantic water caused by remote salinity variations in the North Atlantic (e.g. great salinity anomalies, Dickson et al. 1988, Belkin et al. 1998, Belkin

2004) and/or variation in the intensity of the inflow). Simultaneously, EOF1 of salinity shows strong variability in the German Bight, possibly attributed to the discharge of large rivers like the Elbe. For currents, the EOF1 pattern seems to reflect changes in intensity of the wind-driven cyclonic circulation together with changes in Atlantic inflow through the Dover Passage and the Norwegian Trench (Figs. S11–S14). The system shows an oscillatory structure with higher anomalies of inflow in the south and less inflow in the north and vice versa. The EOF2 pattern depicts a north–south oscillatory field in the case of SST and opposite fluctuations of the Southern/German Bight and the coast of Denmark/Skagerrak/Norwegian Trench area for salinity, likely attributable to outflow from the Baltic. For currents, this pattern mainly represents the strength of the inflow from a western direction into the Skagerrak. The SOM patterns largely reflect a clustering in the space spanned by the first 2 PCs (Fig. S28 in Supplement 4). Extreme patterns are mostly mapped at opposite corners of the grid, with an intermediate pattern of lower amplitude in between (Figs. S15–S26). Similar to EOFs, the time evolution of the SOM pattern captures the warming trend in the case of SST (Figs. S15–S18) as well as the different variations in salinity attributed to northern inflow, Baltic Sea outflow and discharge

from the main rivers (Figs. S19–S22); however, SOMs resolved the spatial differences in far more detail and asymmetrically, compared to the EOF pattern. For currents, the SOMs mainly resolved the north–south and west–east current patterns found in EOF1 and EOF2 (Figs. S23–S26).

3.2. Feature selection

EOF-derived PCs (continuous temporal variables) yielded better predictors for cod recruitment time series than SOM-derived patterns (discrete temporal variables) (Fig. 4, Table 3; Table S2 in Supplement 5). The final solution, reflecting the compromise between complexity and performance, was the 5-parameter solution in the case of EOF-derived patterns and the 3-parameter solution for SOM-derived patterns (Fig. 4, Table 3; Table S2). The SOM-based model feature selection (19 input variables) converged to the final solution after only 18 generations, whereas the EOF-based model feature selection (82 input variables) reached the final solution after 149 generations. The NSGA-II-based feature selection successfully guided the search to models with increased performance and fewer parameters in the objective space (Fig. 4b).
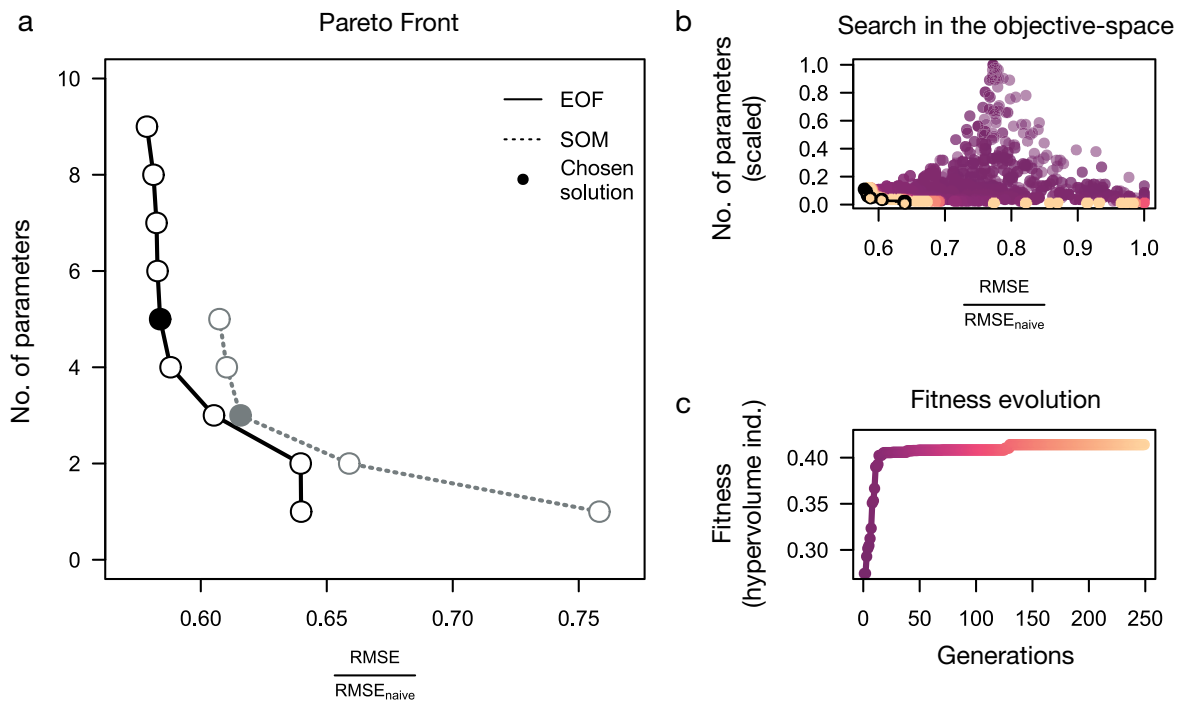

Fig. 4. (a) Pareto fronts of the NSGA-II feature selection for both empirical orthogonal function (EOF)-derived and self-organising map (SOM)-derived features. The evolution of the search for the EOF-derived features is shown in both (b) the objective space and (c) by the calculated hypervolume for tracking the convergence of the search (lighter colours depict later generations)

Table 3. Final models as part of the Pareto front for the empirical orthogonal function (EOF) analysis; the chosen solution is highlighted in **bold**. CV: cross-validation; MAE: mean absolute error

Covariates	No. of Predictors	RMSE$_{cv(5×30)}$	MAE$_{cv(5×30)}$	Pseudo-R$^2_{cv(5×30)}$	Distance to (0,0)
$(SST.PC1.JJA)_{t-1}$	1	0.528	0.42	0.591	0.640
$(SST.PC1.JJA)_{t-1}$ + $(SST.PC1.MAM)_{t-1}$	2	0.532	0.44	0.589	0.640
$(SST.PC3.DJF)_t$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC1.MAM)_{t-1}$	3	0.502	0.415	0.651	0.606
SSB$_{t-1}$ + $(Salt.PC2.JJA)_{t-1}$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC3.DJF)_t$	4	0.49	0.415	0.668	0.59
SSB$_{t-1}$ + $(Salt.PC2.JJA)_{t-1}$ + $(Salt.PC5.MAM)_{t-1}$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC3.DJF)_t$	**5**	**0.484**	**0.413**	**0.691**	**0.587**
SSB$_{t-1}$ + $(Salt.PC2.JJA)_{t-1}$ + $(Salt.PC5.MAM)_{t-1}$ + $(SST.PC1.MAM)_{t-1}$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC3.DJF)_t$	6	0.484	0.415	0.687	0.587
SSB$_{t-1}$ + $(Salt.PC2.JJA)_{t-1}$ + $(Salt.PC5.MAM)_{t-1}$ + $(Salt.PC7.DJF)_t$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC3.DJF)_t$ + $(SST.PC1.SON)_{t-1}$	7	0.484	0.412	0.705	0.589
SSB$_{t-1}$ + $(Salt.PC2.JJA)_{t-1}$ + $(Salt.PC5.MAM)_{t-1}$ + $(SST.PC1.MAM)_{t-1}$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC3.DJF)_t$ + $(SST.PC1.SON)_{t-1}$ + $(Salt.PC7.DJF)_t$	8	0.481	0.412	0.701	0.589
SSB$_{t-1}$ + $(Salt.PC2.JJA)_{t-1}$ + $(Salt.PC2.MAM)_{t-1}$ + $(Salt.PC5.MAM)_{t-1}$ + $(SST.PC1.MAM)_{t-1}$ + $(SST.PC1.JJA)_{t-1}$ + $(SST.PC3.DJF)_t$ + $(SST.PC1.SON)_{t-1}$ + $(Salt.PC7.DJF)_t$	9	0.478	0.405	0.707	0.589

3.3. Model interpretation

Since the EOF-derived model performed in total better than the SOM-based model, only the 'best' solution from the EOF-based model is described here in detail. Feature importance of the final EOF-model showed the SST-pattern in summer (denoted as SST.PC1.JJA$_{t-1}$) during the juvenile pre-recruit phase being the most important, followed by SSB$_{t-1}$ when the cohort is born, overwintering SST (SST.PC3.DJF$_t$) and 2 salinity signals in summer (Salt.PC2.JJA$_{t-1}$) and spring (Salt.PC5.MAM$_{t-1}$) linked to the juvenile pre-recruit and larval phase, respectively (Fig. 5a). All temperature effects in the final model were negatively associated with recruitment (Fig. 5b,d) and associated with high SSTs in the
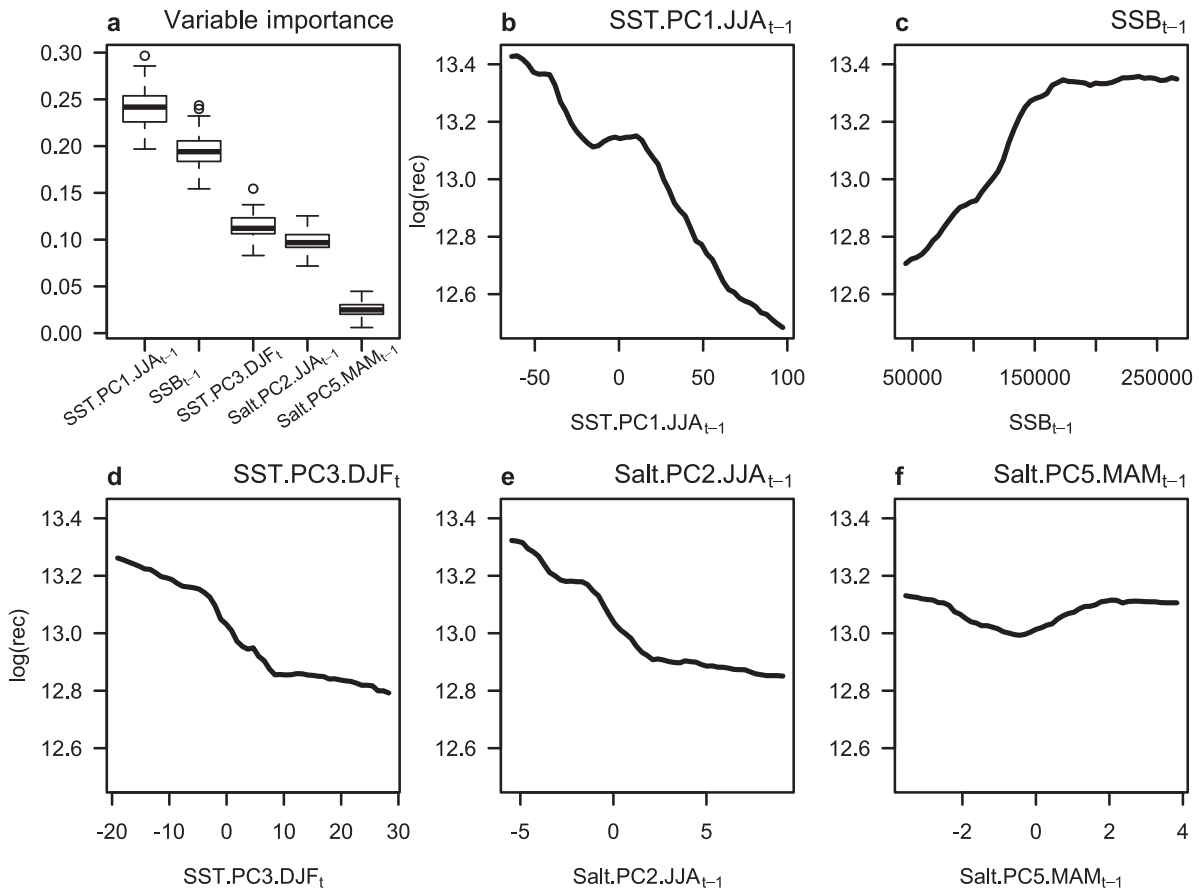
Fig. 5. (a) Variable importance for each covariate; boxplots denote median (thick line), interquartile range (IQR, box), 1.5× IQR (whiskers) and extreme values (points) for 100 successive builds of the final model. (b–f) Partial-dependence plots showing the relative change in the response (log(recruitment)) with variation in each feature; SST: sea surface temperature; PC: principal component; JJA: summer (June–August); SSB: spawning stock biomass; DJF: winter (December–February); Salt: salinity; MAM: spring (March–May)

whole North Sea in summer and high winter SSTs in the northwestern North Sea as well as low winter SSTs in the eastern North Sea and the Skagerrak (Fig. 6). The SST signal in winter ($SST.PC3.DJF_t$) was strongly associated with SST fronts in the central (r = 0.79, p < 0.05) and eastern part of the North Sea (Table S1), with particularly strong gradients in cold winters (results not shown). The salinity effect was estimated to be negative on recruitment, as the summer signal of salinity anomalies correlated positively with the original field in the Norwegian Trench/ Skagerrak area (Table S1). Correlation analysis of PC2 of salinity in the summer months during the juvenile pre-recruit phase with various local and regional climate indices revealed a strong positive correlation (r = 0.73, p < 0.05) with an index of frontal zone strength in an area along the Norwegian Trench/ Skagerrak (Table S1). This indicates that cod recruitment is linked to the gradient of salinity differences responsible for front formation rather than the

absolute salinity values. For the salinity signal in spring (PC5 MAM), no strong explanatory physical mechanism could be found via correlation analysis (Table S1). However, both salinity signals in the model were associated with the area along the Norwegian Trench/Skagerrak, indicating a possible association with the salinity gradients/changes related to the Norwegian Coastal Current and Baltic Sea outflow (Fig. 6). The model also captured the compensating influence of SSB on recruitment (Fig. 5c). Summer temperatures exert a non-linear effect on recruitment, with a strong almost linear effect in both negative and positive directions and a positive but plateau-like influence at values near 0 (Fig. 5b). Temperature during the overwintering phase in winter (Fig. 5d) and salinity (Fig. 5e) in summer during the pre-recruit juvenile phase exert almost a linear influence on cod recruitment, while salinity in spring during the larval phase revealed a non-linear, but weak, effect (Fig. 5f).
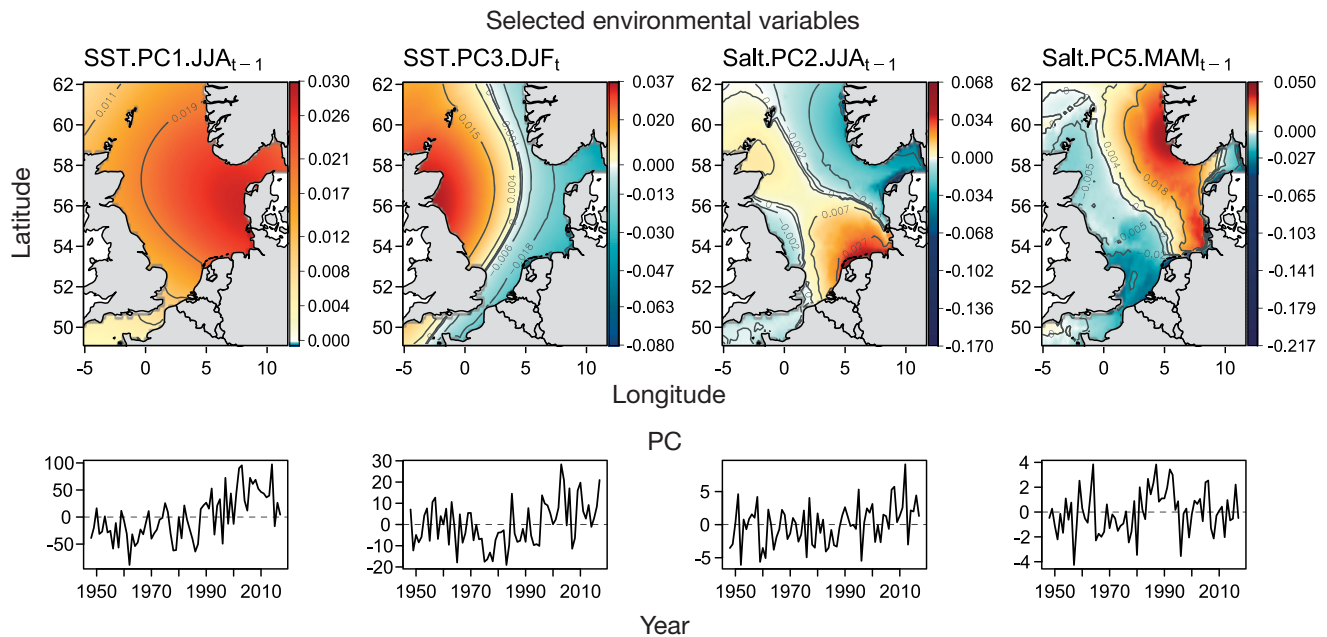
Selected environmental variables

Fig. 6. Empirical orthogonal function (EOF) pattern (top row) and principal component (PC) time series (bottom row) for the environmental variables selected in the final model (abbreviations as in Fig. 5). EOF spatial fields represent dimensionless spatial weights, whereas the amplitude of the PC time series relates to the variance explained by the extracted signal

Individual predictions of the random forest model were decomposed into additive components related to the contributing features. One key feature is the mostly positive contributions of summer SST to the recruitment prediction until the end of the 1980s and a largely negative contribution from 1998 onwards with a phase of transition in between (Fig. 7). From the end of the 1960s to the mid-1970s, SSB largely contributed positively to the recruitment prediction. The highest recruitment events in the time series from the beginning of the 1970s to 1984 were largely driven by SSB and favourable summer and overwintering temperatures. From 1985 onwards, the positive SSB effect ceased, with temperatures in summer solely driving the good recruitment of cod. In the transitioning phase, from 1990 to 1997, SSB was already estimated to exert a negative effect on recruitment, with summer temperatures being favourable for recruitment in the years 1992, 1994 and 1997, and adverse in the rest. The period is characterised by largely fluctuating influences from all environmental variables rather than a consistent positive or negative impact, with salinity in the summer months being a dominant driver in the years 1990, 1991, 1996 and 1997. From 1998 on, SSB and almost all environmental variables, particularly SSTs in summer, had a strong negative impact on recruitment.

4. DISCUSSION

4.1. General discussion of the framework

We presented a framework for deriving a robust ML model incorporating important steps, from summarising spatiotemporal information to selecting robust features and choosing an appropriate model. Although it is believed that methods like artificial neural networks or random forest can greatly help in improving predictions like SR relationships (Megrey et al. 2005, Dreyfus-León & Chen 2007, Smoliński 2019), ML modelling is still not a standard tool. Applications of 'classic' SR models (e.g. Ricker, Beverton-Holt, Cushing) are still predominantly used (Keyl & Wolff 2008), while ML methods are only briefly mentioned in a recent review on future and perspectives in SR research (Subbey et al. 2014), and no application was listed that includes ML in assessments and simulations of management strategy evaluation (Haltuch et al. 2019). Our proposed framework extends existing work (e.g. Fernandes et al. 2010, 2015) regarding research on recruitment of marine fishes and offers guidance for marine ecologists in general by starting with feature extraction from spatiotemporal data showing 2 important pathways like EOF analysis or using SOMs and a wrapper-based feature selection with NSGA-II. The framework itself can be po-
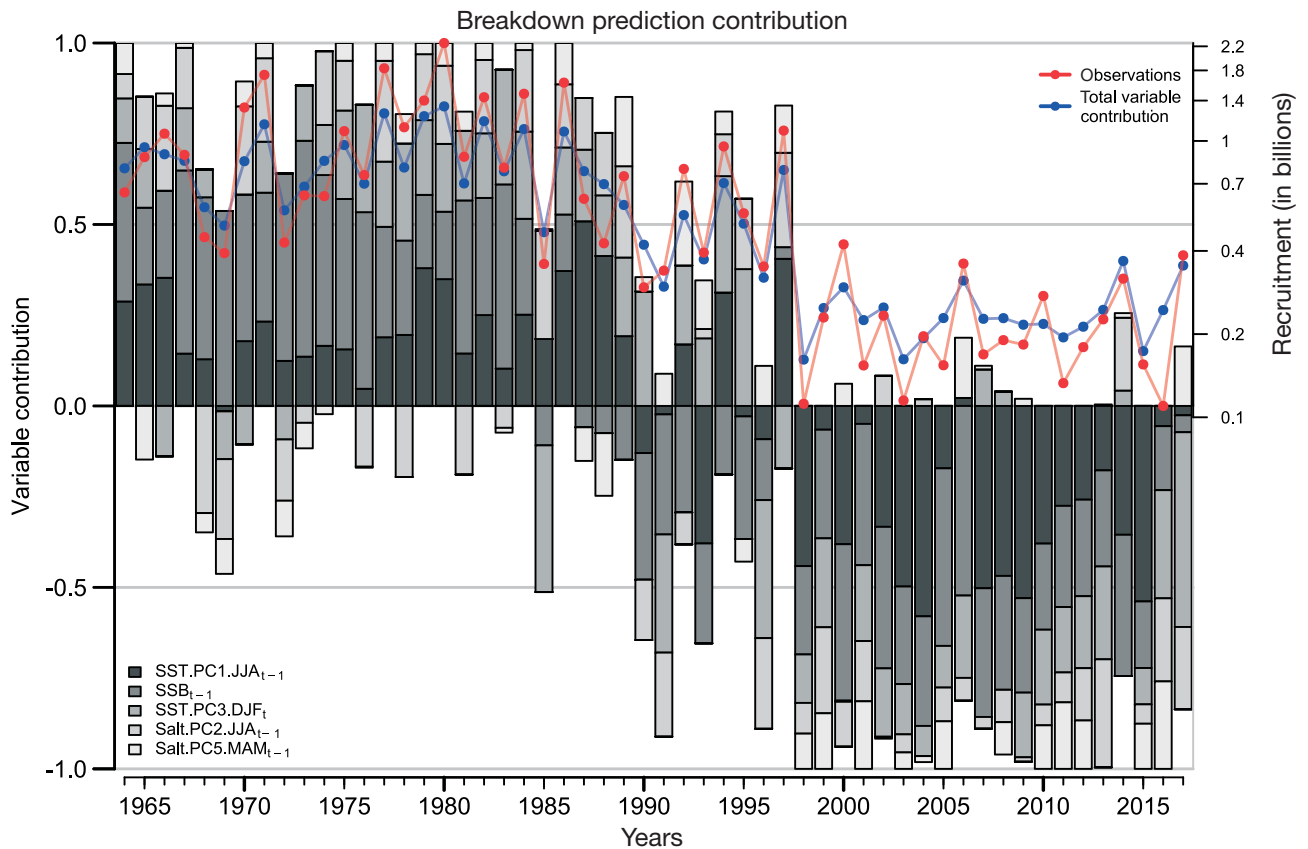
Fig. 7. Results of the 'breakdown' analysis, showing the contribution (as an additive fraction) of each variable (bars; abbreviations as in Fig. 5) to the predicted outcome (blue); the time series of cod recruitment (red) is overlayed. Note that the blue line is the prediction on the training dataset to visualise how the trained forest gets to its prediction, but is NOT meant for visualising the generalisation ability on unseen data

tentially applied to many different datasets and research questions within the marine realm.

4.2. Feature extraction

Local-scale EOF analysis has a long history of applications in fisheries science, but is mostly linked in a descriptive and correlative sense (Dippner 1997a,b, Dippner & Ottersen 2001, Petitgas et al. 2009, Huret et al. 2010) rather than predictive modelling. Similarly, large-scale oceanographic and atmospheric patterns like the NAO, AMO or the Subpolar Gyre index, which can be defined as results of EOF analysis, are frequently used, as they are easily available and often correlate with biological variables of interest. Rather than using large-scale indices, we want to emphasize the use of local EOF- or SOM-derived indices for predictive modelling that better match the spatial distribution of e.g. the fish stock. Similarly to local EOF analysis, SOMs are mostly used in a descriptive sense

in the field of oceanography (Liu & Weisberg 2005, 2011, Liu et al. 2016b), while their use is still limited in biological science applications (e.g. Liu et al. 2016a). We provide the first application on how SOM-derived patterns allowing for non-linear relationships can be incorporated into predictive models of fish recruitment. Although the performance of the SOM-based models was inferior in our case study compared to the EOF-based cod recruitment models, they present an alternative pathway to summarize spatiotemporal environmental information for use in predictive modelling and may be superior to EOFs in other applications. Comparing the resulting patterns from both analyses yielded insights into the temporal and spatial distributions of the dominant signals of SST, salinity and currents in the North Sea. While it would be desirable to identify EOF patterns with clearly separate physical mechanisms, this is not guaranteed (Emery & Thomson 1997). In our particular case, the EOF patterns of salinity seem to mix mechanisms which are not necessarily in phase (like Baltic outflow and changes in volume

transport of the Rhine; I. Núñez-Riboni pers. comm.) and to separate mechanisms which are in phase (like changes in river discharge of the Rhine and Elbe; I. Núñez-Riboni pers. comm.). Note, however, that the identification of specific physical mechanisms is not fundamental to the validity of the present analysis. What is important is simply the link to cod recruitment, which could arise from a combination of mechanisms instead of individual mechanisms.

4.3. Feature selection

In marine science in general, and fisheries science in particular, it is important to derive parsimonious models with fewer variables, as it both helps in understanding and communicating results to stakeholders and facilitating predictions. This can be done by limiting the number of variables *a priori*, selecting the possible candidate models manually or using automated methods of feature selection. Many of the earlier studies that used ML in recruitment research either selected variables *a priori* or manually (e.g. Chen & Ware 1999, Chen et al. 2000, Huse & Ottersen 2003). As manual selection of variables becomes impractical with the use of many environmental variables in different seasons, we presented an automated multi-objective evolutionary algorithm (MOEA), specifically NSGA-II, to derive a limited number of variables. It can be seen as an alternative to various other feature selection methods applied in fishery science, like correlation-based filters (e.g. Fernandes et al. 2010), the K2 algorithm specific for Bayesian networks (e.g. Trifonova et al. 2014) or Boruta for random forests (e.g. Smoliński 2019). The advantage of using a MOEA like NSGA-II is that, unlike filter-methods, it fully takes into account the model structure. The simultaneous control for model complexity and performance allows the effective exploration of solutions with low complexity. Using a heuristic search algorithm on various inter-correlated variables poses the risk of increased Type I error rates, as unimportant variables can be selected by random chance. In separate simulation experiments, conducted as a quality control measure, 4 artificial predictors and 200 additional inter-correlated non-sense variables with similar autocorrelation structure, but unrelated to the response variable, were designed to test the ability of the differing feature selection methods in correctly identifying true signals related to the response variable (Text S3 in Supplement 6). The results showed fewer cases of falsely selected variables with NSGA-II than with Boruta,

recursive feature elimination and a filter approach (Fig. S29 in Supplement 6). This is due to the effective search in the feature space combined with rigorous cross-validation. Further improvements may be achieved by decreasing the number of folds at the expense of possibly missing important variables (Type II error). Furthermore, contrary to other algorithms (e.g. Boruta for random forest), NSGA-II is not limited to a particular ML model, as any model can be included via the definition of an appropriate fitness function.

The major drawback of the NSGA-II-based feature selection is the long runtime, which was reduced via parallelisation and the option for memoisation (storing already evaluated covariate combinations). Even under these improvements, the algorithm may not scale well for ML models that involve a long training time or have a large number of observations/variables. An overview study on genetic algorithm-based feature selection estimated evolutionary approaches to be suitable for tasks with less than 1000 features (Xue et al. 2016). One option for a larger number of variables would be to use an additional step of dimension-reduction techniques like PCA or to reduce the number of permutations in the repeated *k*-fold cross-validation step. However, at least for applications where the number of observations is relatively low, we believe that the efficient exploration of the feature space outweighs the additional runtime.

4.4. North Sea cod case study

The final model (R^2 = 0.69, via 5-fold cross-validation) is among the highest performing models for North Sea cod recruitment that take into account one or more environmental variables. Early work assessing recruitment of North Sea cod used an environmental Ricker, which accounted for 42% of the variance using SST time series from February to June (Planque et al. 2003) and 47% from February to March (Cook & Heath 2005). A model combining both the Ricker and Beverton-Holt relationship together with annual-averaged SST and zooplankton abundance yielded a similar fit (45% explained variance) (Olsen et al. 2011). Sguotti et al. (2020) found that an environmental Ricker together with annual-averaged SST outperformed more sophisticated modelling approaches like empirical dynamic modelling or stochastic cusp modelling for North Sea cod (55% explained variance). Akimova et al. (2016) found that cod recruitment was strongly related to temperature in January–March at 75 m depth near the Fair Isle region, where a modified Cushing function fitted

with the data explained around 67% in recruitment variability. Pécuchet et al. (2015) used residuals from a Ricker model and fitted a generalised additive model taking into account temperature, currents and nutrients (NO_3 and PO_4 separately) in the time from February to March, explaining 67% in deviance.

Although a direct comparison of studies is difficult, as the time series of North Sea cod recruitment became longer over the years, the temperature effect on cod recruitment seems to be a dominant and prevailing feature found by our model and other models. The incorporation of various lagged temperature (JJA_{t-1} and DJF_t) and salinity signals (MAM_{t-1}, JJA_{t-1}) in different seasons in the model can represent different effects on life stages. The fact that the influence of SST is almost linear on cod recruitment explains why linear models yield a similar performance compared to the more flexible approach presented here. However, according to our results, the season in which the temperature effect is manifested in the population is variable. Many authors have found an effect in winter or spring, whereas variable importance in our model showed a higher importance of summer temperatures for cod recruitment. The effect of temperature at the beginning of the year is estimated to be an effect experienced during the larval stage, either directly through unfavourable temperatures for growth/mortality or indirectly through food availability via match–mismatch dynamics (Beaugrand et al. 2003, Beaugrand & Kirby 2010, Buckley et al. 2010, Kristiansen et al. 2011). The high importance of summer and overwintering temperatures in our model, however, might point to the pre-recruit juvenile phase as an additional bottleneck in recruitment. This can be due to direct detrimental effects of summer temperatures, as supported by Laurel et al. (2017), who found modelled summer growth stress especially in the months from July to September for 0-group coastal cod in the Skagerrak to be related to impaired recruitment from 1998 to 2012. Also, predator–prey relationships can be amplified by higher temperatures as e.g. predator–prey overlap between 0-group cod and grey gurnard increases with higher temperature in the North Sea in summer (Kempf et al. 2013).

The effect of the winter SST signal on cod overwintering juveniles might be rather complex, as it correlates with SST in the northwestern North Sea, an area influenced by Atlantic water and SST fronts in the central North Sea (Table S1), where these warm water masses meet cold water from the southern North Sea in winter. Although overwintering mortality in pre-recruits is recognized in affecting recruitment signals (Hurst 2007), no studies are yet available that systematically evaluate the conditions in the overwintering period of North Sea cod pre-recruits. Further work is needed to evaluate if this effect is a direct effect of temperature or rather an indirect effect, e.g. due to increased predation or starvation in winter. The salinity effect in summer (PC2 JJA_{t-1}) during the juvenile phase might point to conditions in the nursery habitat of cod. The signal is strongly linked to frontal zones along the Norwegian Trench and the Skagerrak area (Table S1), a region which is known for higher densities of larval and juvenile cod and their preferred food at the shelf-break front (Munk et al. 1995, 1999, Munk 2007, 2014). The causal mechanism behind this might be indirect and associated with favourable feeding conditions for late-stage larvae and juveniles or a slow-down in early life stage dispersal due to strong retention in the front. Although the effect of salinity in spring (PC5 MAM_{t-1}) is only marginal in the model, the association with approximately the same area in the northeastern North Sea provides a consistent picture of the salinity effect on cod. However, as the salinity signal in spring could only be weakly linked to any physical mechanism (Table S1), caution in making an interpretation is recommended.

An interesting aspect of the breakdown analysis is the different contribution of variables throughout the historical recruitment record, especially the time period from 1998 onwards, with an ongoing low recruitment of cod and a consistent negative contribution from the predictors in the model. There is evidence that the North Sea underwent a regime shift in the years 1997/1998, which is related to the switch from a cold-phase to a warm-phase AMO (Auber et al. 2015) and a change in the plankton community composition of the North Sea (Alvarez-Fernandez et al. 2012). Alvarez-Fernandez et al. (2012) found that the 1998 regime shift was characterised by a decrease in total copepod biomass and a larger contribution of warm-water zooplankton. Additionally, the authors noted that the increase in temperature was governed by a decrease in wind speed post-1996, lower Atlantic water inflow from the North and, in general, a situation with less mixing and more permanent stratification. This is supported by a recent study, which concluded that the reduction in the NAO in 1996, an accompanying drop of *Calanus finmarchicus* in 1997, the low spawning stock and especially the increase in summer–autumn temperatures created a 'perfect storm' of unfavourable conditions that led to the drop in juvenile Norwegian coastal cod in the Skagerrak (Perälä et al. 2020). Although our analysis did not include the effects of

the NAO or zooplankton dynamics, the breakdown analysis points in the same direction, as summer temperatures and SSB were the most influential variables explaining the low recruitment in the period after 1998. Tools such as the breakdown analysis can help to disentangle such effects of multiple drivers and to make ML models more understandable.

4.5. Lessons learned

4.5.1. There is no free lunch

In the ML community, the saying 'There is no free lunch!' (Schaffer 1994, Wolpert & Macready 1997) refers to the fact that no algorithm exists that is in all aspects superior for all problems. Indeed, random forest comes close to being a promising solution for many problems, as it has good out-of-box performance, contains comparatively few tuning parameters and is rather robust to overfitting due to ensemble averaging. However, in our case study, the 'standard' random forest algorithm showed signs of overfitting in regions where only few data points with high variance were observed. This was especially evident in strong spikes in recruitment for small changes in SSB (e.g. years 1985/1986), which is not biologically meaningful (Fig. S1). Tools like Pd plots or breakdown can help to reveal this unexpected model behaviour and should be included at the end of every ML pipeline. The additional randomization via extreme randomized trees presents a natural way to help alleviate this problem. Alternative ways would be a reduction in tree depth or a smaller bagging fraction, although finding the right amount of reduction is a rather subjective step.

The 'no free lunch' theorem is also applicable to the NSGA-II feature selection algorithm. We used the algorithm as it is widely used in many applications, easily understandable and already implemented in the programming language R. However, for differently shaped Pareto fronts, other MOEAs might outperform NSGA-II. Jiménez et al. (2017) compared NSGA-II feature selection to an evolutionary non-dominated radial slots based algorithm (ENORA) for feature selection with random forest as the ML model. Although ENORA outperformed NSGA-II in regards to the diversity of solutions obtained, the lower parameter solutions of NSGA-II almost always had lower RMSEs compared to solutions with similar complexity found by ENORA. Xue et al. (2013) compared 2 variants of their proposed multi-objective particle swarm-optimisation (PSO) for feature selection to var-

ious MOEAs, including NSGA-II, on 12 benchmark datasets. NSGA-II was among the top-performing algorithms for the small datasets (<50 features), but was outperformed for large datasets (>100 features) by the PSO, albeit with similar performance as the other MOEAs. Under the 'no-free-lunch' theorem, we see our framework as a modular tool with interchangeable parts to aid the use of ML methods in fisheries science, rather than an all-in-one solution.

4.5.2. Extrapolation

A great problem one needs to be aware of when working with flexible ML methods is the low ability for extrapolation. Predictions of tree-based methods can only lie within the range of the response values already seen in the training dataset, due to averaging over partitions of the space spanned by the response and the features. Extrapolating will result in a constant value representing the highest/lowest prediction learned from the data. In the case of stock–recruitment modelling, this becomes crucial especially in management strategy evaluations, where a meaningful extrapolation between zero and the lowest observed value may be required. Additionally, long-term forecasts, where variables are expected to increase largely outside the historical range (e.g. a strong temperature trend), pose a problem for regression trees. An alternative here would be to continuously update the training data set as new data become available. On the other hand, there are cases were this conservative extrapolation provides a better suited solution as, for example, linear polynomial regression models that grow to infinity when extrapolated. Also, parametric methods can suffer from poor predictions outside the data range, if parameter estimates are unreliable in the case of limited observations (Hillary 2012).

4.5.3. Stochasticity

Contrary to classical statistical methods, many ML methods show inherent stochasticity. In our analysis, this aspect encompasses the stochastic nature of SOMs, stochastic cross-over and mutation in genetic algorithms, the splitting in training and test set in cross-validation and the randomization of splits, the variable choice for each node, and the bagging step in the 'extreme' version of random forests. Controlling for these steps by using a specified random seed for the random number generator makes research reproducible among

different machines. Also, using enough repetitions is crucial in making sure results do not arise by pure chance. However, dependent on the feature set, one cannot entirely rule out that other similar solutions with almost equal performance than the chosen one exist.

5. CONCLUSION AND SCOPE FOR FURTHER APPLICATIONS

The proposed framework offers a modular and flexible tool for applications in which pattern extraction from a multitude of variables combined with a classification/regression-type problem is acquired. The choice to limit the number of extracted patterns, and cross-validating the feature selection, helps to avoid overfitting and identify robust models, which should be useful in many ecological marine science applications. These include linking aggregated time series of e.g. species biodiversity, community composition, stock–recruitment, catches, age/length at maturity or any other indicator of ecosystem/stock health to spatiotemporally resolved data such as fleet movements, spatial distributions of species life stages or further spatiotemporal environmental variables. Although SOMs did not perform as well as EOF analysis in our case study, we emphasize their use as exploratory tools to extract patterns from biological variables such as plankton or species counts of fish and benthos, due to their capability to identify nonlinear patterns. Furthermore, we showed that NSGA-II offers an effective way to find a parsimonious model in a large feature space, by simultaneously taking the unique structure of the ML model into account that goes beyond mere correlation. The modular nature of the framework also allows for the inclusion of parametric linear models and testing a multitude of predictor variables. However, as these models have problems dealing with correlated features, a prior filter for collinearity might be needed. Tree-based ensembles, on the other hand, offer some flexibility over conventional methods, being able to deal with correlated features and model non-linear behaviour. However, we argue that the drawbacks of these flexible methods need to be openly communicated, especially if extrapolation is the goal.

LITERATURE CITED

Aguirre HE, Tanaka K (2007) Working principles, behavior, and performance of MOEAs on MNK-landscapes. Eur J Oper Res 181:1670–1690

Akimova A, Núñez-Riboni I, Kempf A, Taylor MH (2016) Spatially-resolved influence of temperature and salinity on stock and recruitment variability of commercially important fishes in the North Sea. PLOS ONE 11:e0161917

Allken V, Handegard NO, Rosen S, Schreyeck T, Mahiout T, Malde K (2019) Fish species identification using a convolutional neural network trained on synthetic data. ICES J Mar Sci 76:342–349

Alvarez-Fernandez S, Lindeboom H, Meesters E (2012) Temporal changes in plankton of the North Sea: community shifts and environmental drivers. Mar Ecol Prog Ser 462:21–38

Auber A, Travers-Trolet M, Villanueva MC, Ernande B (2015) Regime shift in an exploited fish community related to natural climate oscillations. PLOS ONE 10: e0129883

Balmaseda MA, Mogensen K, Weaver AT (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. Q J R Meteorol Soc 139:1132–1161

Bastrikin DK, Gallego A, Millar CP, Priede IG, Jones EG (2014) Settlement length and temporal settlement patterns of juvenile cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), and whiting (*Merlangius merlangus*) in a northern North Sea coastal nursery area. ICES J Mar Sci 71:2101–2113

Beaugrand G, Kirby RR (2010) Climate, plankton and cod. Glob Change Biol 16:1268–1280

Beaugrand G, Brander KM, Lindley JA, Souissi S, Reid PC (2003) Plankton effect on cod recruitment in the North Sea. Nature 426:661–664

Belkin IM (2004) Propagation of the 'Great Salinity Anomaly' of the 1990s around the northern North Atlantic. Geophys Res Lett 31:L08306

Belkin IM, Levitus S, Antonov J, Malmberg SA (1998) 'Great Salinity Anomalies' in the North Atlantic. Prog Oceanogr 41:1–68

Bengtsson H (2019) R.cache: fast and light-weight caching (memoization) of objects and results to speed up computations. R package version 0.14.0. https://CRAN. R-project.org/package=R.cache

Biecek P (2018) DALEX: Explainers for complex predictive models in R. J Mach Learn Res 19:1–5

Björnssin H, Venegas SA (1997) A manual for EOF and SVD analysis of climatic data. CCGCR Rep 97-1. Centre for Climate and Global Change Research, McGill University, Montreal. http://www.geog.mcgill.ca/gec3/wp-content/uploads/2009/03/Report-no.-1997-1.pdf

Bossek J (2017) ecr 2.0: A modular framework for evolutionary computation in R. In: GECCO '17 Companion: Proceedings of the 2017 Genetic and Evolutionary Computation Conference, 15–19 Jul 2017, Berlin. Association for Computing Machinery, New York, NY, p 1187–1193

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

Buckley LJ, Lough RG, Mountain D (2010) Seasonal trends in mortality and growth of cod and haddock larvae result in an optimal window for survival. Mar Ecol Prog Ser 405: 57–69

Burrell J (2016) How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc 3: 1–12

Chen DG, Ware DM (1999) A neural network model for forecasting fish stock recruitment. Can J Fish Aquat Sci 56:2385–2396

Chen DG, Hargreaves NB, Ware DM, Liu Y (2000) A fuzzy logic model with genetic algorithm for analyzing fish stock–recruitment relationships. Can J Fish Aquat Sci 57:1878–1887

Cook RM, Heath MR (2005) The implications of warming climate for the management of North Sea demersal fisheries. ICES J Mar Sci 62:1322–1326

de Almeida Ribeiro L, Soares AS, Lima TW, Jorge CAC and others (2015) Multi-objective genetic algorithm for variable selection in multivariate classification problems: a case study in verification of biodiesel adulteration. Procedia Comput Sci 51:346–355

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6:182–197

Diao R, Shen Q (2015) Nature inspired feature selection meta-heuristics. Artif Intell Rev 44:311–340

Dickson RR, Meincke J, Malmberg SA, Lee AJ (1988) The 'Great Salinity Anomaly' in the northern North Atlantic 1968–1982. Prog Oceanogr 20:103–151

Dippner JW (1997a) Recruitment success of different fish stocks in the North Sea in relation to climate variability. Dtsch Hydrogr Z 49:277–293

Dippner JW (1997b) SST anomalies in the North Sea in relation to the North Atlantic Oscillation and the influence on the theoretical spawning time of fish. Dtsch Hydrogr Z 49:267–275

Dippner JW, Ottersen G (2001) Cod and climate variability in the Barents Sea. Clim Res 17:73–82

Dreyfus-León M, Chen DG (2007) Recruitment prediction with genetic algorithms with application to the Pacific herring fishery. Ecol Model 203:141–146

Drinkwater KF, Belgrano A, Borja A, Conversi A and others (2003) The response of marine ecosystems to climate variability associated with the North Atlantic Oscillation. In: Hurrell JW, Kushnir Y, Ottersen G, Visbeck M (eds) The North Atlantic Oscillation: climatic significance and environmental impact. Geophys Monogr Ser 134. American Geophysical Union, Washington, DC, p 211–234

El Aboudi N, Benhlima L (2016) Review on wrapper feature selection approaches. In: 2016 International Conference on Engineering & MIS (ICEMIS), 22–24 Sep 2016, Agadir, Morocco. IEEE, p 1–5

Emery WJ (2003) Air–sea interaction. Sea surface temperature. In: Holton JR, Curry JA, Pyle JA (eds) Encyclopedia of Atmospheric Sciences. Elsevier, Amsterdam, p 100–109

Emery WJ, Thomson RE (1997) Data analysis methods in physical oceanography. Elsevier, New York, NY

Faillettaz R, Beaugrand G, Goberville E, Kirby RR (2019) Atlantic Multidecadal Oscillations drive the basin-scale distribution of Atlantic bluefin tuna. Sci Adv 5:eaar6993

Fernandes JA, Irigoien X, Goikoetxea N, Lozano JA, Inza I, Pérez A, Bode A (2010) Fish recruitment prediction, using robust supervised classification methods. Ecol Model 221:338–352

Fernandes JA, Irigoien X, Lozano JA, Inza I, Goikoetxea N, Pérez A (2015) Evaluating machine-learning techniques for recruitment forecasting of seven North East Atlantic fish species. Ecol Inform 25:35–42

Fisher A, Rudin C, Dominici F (2018) All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J Mach Learn Res 20:1–81

Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63:3–42

Goldstein A, Kapelner A, Bleich J, Pitkin E (2015) Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J Comput Graph Stat 24:44–65

González-Irusta JM, Wright PJ (2016) Spawning grounds of Atlantic cod (Gadus morhua) in the North Sea. ICES J Mar Sci 73:304–315

Greenwell BM (2017) pdp: an R package for constructing partial dependence plots. R J 9:421–436

Gröger JP, Fogarty MJ (2011) Broad-scale climate influences on cod (Gadus morhua) recruitment on Georges Bank. ICES J Mar Sci 68:592–602

Gupta A, Mehrotra KG, Mohan C (2010) A clustering-based discretization for supervised learning. Stat Probab Lett 80:816–824

Guyon I, Gunn S, Nikravesh M, Zadeh LA (2006) Feature extraction: foundations and applications. Springer, Berlin

Haltuch MA, Brooks EN, Brodziak J, Devine JA and others (2019) Unraveling the recruitment problem: a review of environmentally-informed forecasting and management strategy evaluation. Fish Res 217:198–216

Ham YG, Kim JH, Luo JJ (2019) Deep learning for multi-year ENSO forecasts. Nature 573:568–572

Hamdani TM, Won JM, Alimi AM, Karray F (2007) Multi-objective feature selection with NSGA II. In: Beliczynski B, Dzielinski A, Iwanowski M, Ribeiro B (eds) Adaptive and natural computing algorithms (ICANNGA 2007). Lecture notes in computer science, Vol 4431. Springer, Berlin, p 240–247

Hansen GJA, Carpenter SR, Gaeta JW, Hennessy JM, Vander Zanden MJ (2015) Predicting walleye recruitment as a tool for prioritizing management actions. Can J Fish Aquat Sci 72:661–672

Hare JA (2014) The future of fisheries oceanography lies in the pursuit of multiple hypotheses. ICES J Mar Sci 71:2343–2356

Hauser T, Demirov E, Zhu J, Yashayaev I (2015) North Atlantic atmospheric and ocean inter-annual variability over the past fifty years – dominant patterns and decadal shifts. Prog Oceanogr 132:197–219

Hillary RM (2012) Practical uses of non-parametric methods in fisheries assessment modelling. Mar Freshw Res 63:606–615

Horn JL (1965) A rationale and test for the number of factors in factor analysis. Psychometrika 30:179–185

Houde ED (2008) Emerging from Hjort's shadow. J Northwest Atl Fish Sci 41:53–70

Hsieh C, Glaser SM, Lucas AJ, Sugihara G (2005) Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435:336–340

Huang B, Buckley B, Kechadi TM (2010) Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. Expert Syst Appl 37:3638–3646

Huret M, Petitgas P, Struski C, Léger F, Sourisseau M, Lazure P (2010) A 37 years hindcast of a coupled physical–biogeochemical model and its use for fisheries oceanography in the Bay of Biscay. ICES C 2010/A06

Hurst TP (2007) Causes and consequences of winter mortality in fishes. J Fish Biol 71:315–345

Huse G, Ottersen G (2003) Forecasting recruitment and stock biomass of Northeast Arctic cod using neural networks. Sci Mar 67(Suppl 1):325–335

Huserbråten MBO, Moland E, Albretsen J (2018) Cod at drift in the North Sea. Prog Oceanogr 167:116–124

ICES (2018) Cod (Gadus morhua) in Subarea 4, Division 7.d, and Subdivision 20 (North Sea, eastern English Channel, Skagerrak) ICES Advice on fishing opportunities, catch, and effort Greater North Sea Ecoregion Published 29 June 2018 cod.27.47d20 Version 3: 14 November 2018. https://doi.org/10.17895/ices.pub.4436 ICES Advice 2018

Iguyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182

Jiménez F, Sánchez G, García JM, Sciavicco G, Miralles L (2017) Multi-objective evolutionary feature selection for online sales forecasting. Neurocomputing 234:75–92

Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York, NY

Jonsson PR, Corell H, André C, Svedäng H, Moksnes PO (2016) Recent decline in cod stocks in the North Sea–Skagerrak–Kattegat shifts the sources of larval supply. Fish Oceanogr 25:210–228

Kempf A, Stelzenmüller V, Akimova A, Floeter J (2013) Spatial assessment of predator–prey relationships in the North Sea: the influence of abiotic habitat properties on the spatial overlap between 0-group cod and grey gurnard. Fish Oceanogr 22:174–192

Keyl F, Wolff M (2008) Environmental variability and fisheries: What can models do? Rev Fish Biol Fish 18: 273–299

Kjesbu OS, Righton D, Krüger-Johnsen M, Thorsen A, Michalsen K, Fonn M, Witthames PR (2010) Thermal dynamics of ovarian maturation in Atlantic cod (*Gadus morhua*). Can J Fish Aquat Sci 67:605–625

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69

Krekoukiotis D, Palacz AP, St. John MA (2016) Assessing the role of environmental factors on Baltic cod recruitment, a complex adaptive system emergent property. Front Mar Sci 3:126

Kristiansen T, Drinkwater KF, Lough RG, Sundby S (2011) Recruitment variability in North Atlantic cod and match–mismatch dynamics. PLOS ONE 6:e17456

Kuhn M (2008) Building predictive models in R using the 'caret' package. J Stat Softw 28:1–26

Kühn B (2020) marmalaid: MARine MAchine Learning AID - a collection of functions for utilising machine learning in marine biological science. R package version 0.1. https://github.com/BernhardKuehn/marmalaid doi: 10.6084/m9.figshare.13469736

Kvile KØ, Romagnoni G, Dagestad KF, Langangen Ø, Kristiansen T (2018) Sensitivity of modelled North Sea cod larvae transport to vertical behaviour, ocean model resolution and interannual variation in ocean dynamics. ICES J Mar Sci 75:2413–2424

Laurel BJ, Cote D, Gregory RS, Rogers L, Knutsen H, Olsen EM (2017) Recruitment signals in juvenile cod surveys depend on thermal growth conditions. Can J Fish Aquat Sci 74:511–523

Liu Y, Weisberg RH (2005) Patterns of ocean current variability on the West Florida Shelf using the self-organizing map. J Geophys Res 110:C06003

Liu Y, Weisberg RH (2011) A review of self-organizing map applications in meteorology and oceanography. In: Mwasiagi JI (ed) Self organizing maps: applications and novel algorithm design. InTech, Rijeka, p 253–272

Liu Y, Weisberg RH, Mooers CNK (2006) Performance evaluation of the self-organizing map for feature extraction. J Geophys Res 111:C05018

Liu Y, Weisberg RH, Lenes JM, Zheng L, Hubbard K, Walsh JJ (2016a) Offshore forcing on the 'pressure point' of the West Florida Shelf: anomalous upwelling and its influence on harmful algal blooms. J Geophys Res Oceans 121:5501–5515

Liu Y, Weisberg RH, Vignudelli S, Mitchum GT (2016b) Patterns of the loop current system and regions of sea surface height variability in the eastern Gulf of Mexico revealed by the self-organizing maps. J Geophys Res Oceans 121:2347–2366

Lustgarten JL, Gopalakrishnan V, Grover H, Visweswaran S (2008) Improving classification performance with discretization on biomedical datasets. AMIA Annu Symp Proc 2008:445–449

Malde K, Handegard NO, Eikvil L, Salberg AB (2020) Machine intelligence and the data-driven future of marine science. ICES J Mar Sci 77:1274–1285

Mathis M, Elizalde A, Mikolajewicz U, Pohlmann T (2015) Variability patterns of the general circulation and sea water temperature in the North Sea. Prog Oceanogr 135:91–112

Megrey BA, Lee YW, MacKlin SA (2005) Comparative analysis of statistical tools to identify recruitment–environment relationships and forecast recruitment strength. ICES J Mar Sci 62:1256–1269

Michie D (1968) 'Memo' functions and machine learning. Nature 218:19–22

Moen E, Handegard NO, Allken V, Albert OT, Harbitz A, Malde K (2018) Automatic interpretation of otoliths using deep learning. PLOS ONE 13:e0204713

Munk P (2007) Cross-frontal variation in growth rate and prey availability of larval North Sea cod *Gadus morhua.* Mar Ecol Prog Ser 334:225–235

Munk P (2014) Fish larvae at fronts: horizontal and vertical distributions of gadoid fish larvae across a frontal zone at the Norwegian Trench. Deep Sea Res II Top Stud Oceanogr 107:3–14

Munk P, Larsson PO, Danielsen D, Moksness E (1995) Larval and small juvenile cod *Gadus morhua* concentrated in the highly productive areas of a shelf break front. Mar Ecol Prog Ser 125:21–30

Munk P, Larsson PO, Danielssen DS, Moksness E (1999) Variability in frontal zone formation and distribution of gadoid fish larvae at the shelf break in the northeastern North Sea. Mar Ecol Prog Ser 177:221–233

Myers RA (1998) When do environment–recruitment correlations work? Rev Fish Biol Fish 8:285–305

Myllymäki P, Silander T, Tirri H, Uronen P (2002) B-Course: a web-based tool for Bayesian and causal data analysis. Int J Artif Intell Tools 11:369–387

Núñez-Riboni I, Akimova A (2015) Monthly maps of optimally interpolated *in situ* hydrography in the North Sea from 1948 to 2013. J Mar Syst 151:15–34

Núñez-Riboni I, Akimova A (2017) Quantifying the impact of the major driving mechanisms of inter-annual variability of salinity in the North Sea. Prog Oceanogr 154:25–37

Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83:171–193

Olsen EM, Ottersen G, Llope M, Chan KS, Beaugrand G, Stenseth NC (2011) Spawning stock and recruitment in North Sea cod shaped by food and climate. Proc R Soc B 278:504–510

Peck MA, Kühn W, Hinrichsen HH, Pohlmann T (2009) Inter-annual and inter-specific differences in the drift of fish eggs and yolksac larvae in the North Sea: a biophysical modeling approach. Sci Mar 73:23–36

Pécuchet L, Nielsen JR, Christensen A (2015) Impacts of the local environment on recruitment: a comparative study of North Sea and Baltic Sea fish stocks. ICES J Mar Sci 72:1323–1335

Perälä T, Olsen EM, Hutchings JA (2020) Disentangling conditional effects of multiple regime shifts on Atlantic cod productivity. PLOS ONE 15:e0237414

Petitgas P, Huret M, Léger F, Peck MA, Dickey-Collas M, Rijnsdorp AD (2009) Patterns and schedules in hindcasted environments and fish life cycles. ICES C 2009/E25

Planque B, Fromentin JM (1996) *Calanus* and environment in the eastern North Atlantic. I. Spatial and temporal patterns of *C. finmarchicus* and *C. helgolandicus.* Mar Ecol Prog Ser 134:101–109

Planque B, Fox CJ, Saunders MA, Rockett P (2003) On the prediction of short term changes in the recruitment of North Sea cod (*Gadus morhua*) using statistical temperature forecasts. Sci Mar 67:211–218

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Reusch DB, Alley RB, Hewitson BC (2005) Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. Polar Geogr 29:188–212

Romagnoni G, Kvile KØ, Dagestad K, Eikeset AM, Kristiansen T, Stenseth NC, Langangen Ø (2020) Influence of larval transport and temperature on recruitment dynamics of North Sea cod (*Gadus morhua*) across spatial scales of observation. Fish Oceanogr 29:324–339

Rudnick DL, Davis RE (2003) Red noise and regime shifts. Deep Sea Res I Oceanogr Res Pap 50:691–699

Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23:2507–2517

Sang H, Gelfand AE, Lennard C, Hegerl G, Hewitson B (2008) Interpreting self-organizing maps through space–time data models. Ann Appl Stat 2:1194–1216

Schaffer C (1994) A conservation law for generalization performance. In: Hirsh H, Cohen WW (eds) Proceedings of the 11th International Conference on Machine Learning. Morgan Kaufmann Publishers, San Francisco, CA, p 259–265

Schröder SM, Kiko R, Koch R (2020) MorphoCluster: efficient annotation of plankton images by clustering. Sensors (Basel) 20:3060

Sguotti C, Otto SA, Cormon X, Werner KM, Deyle E, Sugihara G, Möllmann C (2020) Non-linearity in stock–recruitment relationships of Atlantic cod: insights from a multi-model approach. ICES J Mar Sci 77:1492–1502

Shen J, Chang SI, Lee ES, Deng Y, Brown SJ (2005) Determination of cluster number in clustering microarray data. Appl Math Comput 169:1172–1185

Smoliński S (2019) Incorporation of optimal environmental signals in the prediction of fish recruitment using random forest algorithms. Can J Fish Aquat Sci 76:15–27

Staniak M, Biecek P (2018) Explanations of model predictions with live and breakDown Packages. R J 10:395–409

Stephens JA, Jordan MB, Taylor AH, Proctor R (1998) The effects of fluctuations in North Sea flows on zooplankton abundance. J Plankton Res 20:943–956

Subbey S, Devine JA, Schaarschmidt U, Nash RDM (2014) Modelling and forecasting stock–recruitment: current and future perspectives. ICES J Mar Sci 71:2307–2322

Sundby S (2000) Recruitment of Atlantic cod stocks in relation to temperature and advection of copepod populations. Sarsia 85:277–298

Svendsen E, Aglen A, Iversen SA, Skagen DW, Smestad O (1991) Influence of climate on recruitment and migration of fish stocks in the North Sea. ICES C 1991/H12

Trifonova N, Duplisea D, Kenny A, Maxwell D, Tucker A (2014) Incorporating regime metrics into latent variable dynamic models to detect early-warning signals of functional changes in fisheries ecology. In: Džeroski S, Panov P, Kocev D, Todorovski L (eds) Discovery Science. DS 2014. Lecture notes in computer science, Vol 8777. Springer, Cham. https://doi.org/10.1007/978-3-319-11812-3_26

van Ginneken B (2017) Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. Radiol Phys Technol 10:23–32

Wang Z, Rangaiah GP (2017) Application and analysis of methods for selecting an optimal solution from the Pareto-optimal front obtained by multiobjective optimization. Ind Eng Chem Res 56:560–574

Wehrens R, Buydens LMC (2007) Self- and super-organizing maps in R: the kohonen package. J Stat Softw 21:1–19

Wehrens R, Kruisselbrink J (2018) Flexible self-organizing maps in kohonen 3.0. J Stat Softw 87:1–18

Wilks DS (2006) Statistical methods in the atmospheric sciences, 2nd edn. Academic Press, Amsterdam

Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1:67–82

Wright MN, Ziegler A (2017) ranger: a fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 77:1–17

Wunsch C (1999) The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. Bull Am Meteorol Soc 80:245–255

Xue B, Zhang M, Browne WN (2013) Particle swarm optimization for feature selection in classification: a multi-objective approach. IEEE Trans Cybern 43:1656–1671

Xue B, Zhang M, Browne WN, Yao X (2016) A survey on evolutionary computation approaches to feature selection. IEEE Trans Evol Comput 20:606–626

Zitzler E, Thiele L (1998) Multiobjective optimization using evolutionary algorithms – a comparative case study. In: Eiben AE, Bäck T, Schoenauer M, Schwefel HP (eds) Parallel problem solving from nature – PPSN V. Proc 5th Int Conf. Amsterdam, The Netherlands, September 27–30, 1998. Lecture notes in computer science, Vol 1498. Springer, Berlin, p 292–301

Zitzler E, Thiele L (1999) Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans Evol Comput 3:257–271