Climate induced changes in benthic macrofauna–a non-linear model approach

Karin Junker, Dusan Sovilj, Ingrid Kröncke, Joachim W. Dippner

PII: S0924-7963(12)00051-6
DOI: doi: 10.1016/j.jmarsys.2012.02.005
Reference: MARSYS 2202

To appear in: *Journal of Marine Systems*

Received date: 19 October 2011
Revised date: 31 January 2012
Accepted date: 9 February 2012

Please cite this article as: Junker, Karin, Sovilj, Dusan, Kröncke, Ingrid, Dippner, Joachim W., Climate induced changes in benthic macrofauna–a non-linear model approach, *Journal of Marine Systems* (2012), doi: 10.1016/j.jmarsys.2012.02.005

Climate induced changes in benthic macrofauna–a non-linear model approach

Karin Junker[1], Dusan Sovilj[2], Ingrid Kröncke[3], Joachim W. Dippner[1]

[1]Leibniz Institute for Baltic Sea Research Warnemünde, Seestr.15, 18119 Rostock, Germany. Email: Karin.Junker@io-warnemuende.de

[2]Department of Information and Computer Science, Aalto University, Espoo, Finland

[3]Senckenberg am Meer, Marine Research Dept., Südstrand 40, 26382 Wilhelmshaven, Germany

Key words: benthic macrofauna communities, climate indices, neural network, climate variability, time series forecasting, regime shift

Submitted to Journal of Marine Systems

Revised Version 31.01.2012

Abstract

The non-linear methods “optimally pruned extreme learning machine” (OPELM) and “optimally pruned *k*–nearest neighbours” (OPKNN) are applied to relate various climate indices to time series of biomass, abundance and species number of benthic macrofauna communities in the southern North Sea for the period 1978-2005. The results of these methods show that the performance in forecasting macrofauna communities is as poor as linear statistical downscaling if only one climate index is used as a predictor. If a multivariate predictor is used, OPKNN shows a good forecast for biomass and species number, but not for abundance. The improvement of the forecast is of major relevance especially in the presence of biological and climate regime shifts which occurred in the considered period.

Highlights

- Two types of single layered neural networks (“optimally pruned extreme learning machine” and “optimally pruned k-nearest neighbours”) are used to forecast benthic macrofauna.
- Climate regime shift in 2001/2002 shifts the ecosystem of the southern North Sea to a new state regarding benthic macrofauna
- Benthic macrofauna biomass and species number is predictable for the period 1978-1999 using climate descriptors
- Benthic macrofauna variability cannot be forecasted after 2000
- Combination of large scale and regional scale climate descriptors results in good hindcast of benthic macrofauna biomass and species number for the complete period 1978-2005
- Benthic macrofauna abundance remains unpredictable

Introduction

In the last decades, various climate indices have been developed to understand inter-annual and inter-decadal climate variability and climate teleconnections or to identify the response of ecosystems to climate variability. E.g., in the Atlantic sector, the North Atlantic Oscillation (NAO) Index (Hurrell, 1995) has been widely used to identify the response of climate variability in terrestrial (Mysterud et al., 2003), freshwater (Straile et al., 2003) and marine ecosystems (Dippner, 2006; Drinkwater et al., 2003). Atmospheric winter circulation over the North Atlantic area has been proven to be an optimal predictor to forecast the structure of the macrofauna communities in the following spring in the southern North Sea (Kröncke et al., 1998). The mediator between climate and benthic macrofauna is sea surface temperature (SST) which is highly correlated to the NAO index (Becker and Pauly, 1996). Based on those relationships, Dippner and Kröncke (2003) developed forecast equations for the prediction of macrofauna community structure in spring from the climate during the winter before.

However, this linear relationship between climate variables and benthic macrofauna fails after 2000 which might be caused by the climate regime shift in 2001/2002 (Swanson and Tsonis, 2009; Wang et al., 2009). In a recent paper Dippner et al. (2010) demonstrated that a biological regime shift caused by a positive feedback mechanism (Drinkwater et al., 2003; Suarez and Schopf, 1988) as occurred in 1989/1990 is predictable because of the persistence of modes of the climate system dynamics. In contrast, a negative feedback mechanism as occurred in 2001/2002 results in a shift of the climate modes to another state. In such a situation, the response of the biological system to climate variability cannot be predicted any longer with methods like linear statistical downscaling (von Storch et al., 1993) due to the increase in non-linearity of the climate system (Dippner et al., 2010; Hurrell and Deser, 2010). To overcome this disadvantage two possibilities exist, the development of a "new" climate index, the application of other statistical forecast methods, or both. Here we apply two relatively new non-linear methods to relate time series of climate indices to time series of biomass, abundance and species number of benthic macrofauna in the southern North Sea.

Material and Method

Data

Macrofaunal samples were collected in the 2nd quarter of the years 1978 to 2005 in the sublitoral zone off the island of Norderney, one of the East Frisian barrier islands, at five different stations located in water depths of between 12 m and 20 m (Fig. 1). A 0.2 m² van Veen grab was used for sampling. A single grab was taken at each of the five stations. The samples were sieved over 0.63 mm mesh size and fixed in 4% buffered formaldehyde. After sorting, the organisms were preserved in 70% alcohol. Biomass was determined as ash-free dry weight (AFDW) per m². Samples were dried for 24h at 85°C and burned for 6h at 485°C. Species number, abundance, and biomass from the five stations were pooled and treated as replicates for the area, since the multivariate comparison had shown no significant difference between the macrofauna assemblages (Kröncke et al., 1998).

The following monthly climate data sets are used as winter average from December to March (DJFM) for the analyses: (1) the Arctic Oscillation (AO) index from 1899-2007 (Thompson and Wallace, 1998) which describes the leading Empirical Orthogonal Function (EOF) of monthly geopotential height anomalies at the 1000 hPa level on the Northern Hemisphere poleward from 20°N, (2) the NAO index from 1864-2009 (Hurrell, 1995) defined as the difference between the normalized monthly Sea Level Pressure (SLP) anomalies at Lisbon and Stykkisholmur, (3) the Atlantic Multidecadal Oscillation (AMO) index from 1856-2009 (Enfield et al., 2001) defined as the monthly SST anomalies in the North Atlantic area weighted from 0°–70°N, (4) area averaged monthly meridional wind anomalies (1948–2010) in the southern North Sea (53°-56°N, 2°-9°E) from NCEP/NCAR reanalysis, (5) monthly precipitation rate anomalies (1948–2010) averaged over the area 50°-57°N, 4°W-9°E from NCEP/NCAR reanalysis, (6) area averaged monthly SST anomalies (1948-2009) in the southern North Sea (53°-56°N, 2°W-9°E) from NCEP/NCAR reanalysis (Kalnay et al., 1996), (7) salinity from ICES, Marsden square 96668, (8) temperature from ICES, Marsden square 96668 and (9) weekly SST data for the German Bight from 1968-2007 south of 55.5°N and east of 6.5°E from ship-of-opportunity programs, commercial vessels, light vessels, fixed stations and buoys, coastal stations, research vessels and monitoring programs provided by the Federal Maritime and Hydrographic Agency (BSH) Hamburg (Becker et al., 1986)**.**

Methods

Neural networks have been applied to a wide variety of problems with great success (Miche et al., 2008). The main advantage of this type of models is independence from any type of functional form of the underlying process, that is, when no prior assumptions can be made and when only data is avaiable. This stems from the universal approximation property which enables them to approximate any continuous function to a desired level of accuracy. However, these networks can be considered as 'black box' models, and therefore very difficult to interpret in their trained state.

Many methods have been developed over the past decades from both statistical and machine learning perspectives. From the domain of neural networks, extreme learning machine (ELM) based models are appealing due to their very fast learning phase that does not require iterative tuning which is dominant in other neural networks. Of the range of ELM we choose to apply the “optimally pruned extreme learning machine” (OPELM) and the “optimally pruned *k*-nearest neighbours” (OP-KNN). The advantage of these neural networks is a combination of quick training time and modelling approach with no assumptions made about interactions between in- and outputs. These two methods are explained in the following.

OPELM Method

The “optimally pruned extreme learning machine” (OPELM) method belongs to the single layer neural network methods (Miche et al., 2008) and is based on the Extreme Learning Machine (ELM) algorithm from which it borrows the original single layer feed-forward network construction (Huang et al., 2006). The OPELM (Fig. 2) algorithm is introduced as a more robust method to tackle the problem of irrelevant neurons and to be more adaptive to both linear and non-linear problems. The OPELM algorithm consists of three steps:

(1) Construction of the single hidden layer

(2) Ranking the neurons in the hidden layer

(3) Selecting the appropriate number of top ranked neurons

Ranking in step 2 is performed with least angle regression algorithm (LARS) (Efron et al., 2004), an algorithm used to rank variables in regression problems in a linear setting. LARS provides exact ranking when the problem is linear, which is the case in ELM between hidden layer and output variable. Once the ranking is obtained, the selection is done with leave-one-out (LOO) estimation based on the outputs of hidden neurons. The appropriate number of neurons is selected based on the minimization of this LOO estimate (Fig. 3). These three steps are the core of the original OPELM algorithm. However, when using data sets with few samples, the LOO estimate can have high variance and can lead to complex models and over-fitting. Instead of LOO, other model selection criteria can be employed, such as information theoretic criteria. In the experiments, we resort to the corrected Akaike information criterion (Hurvich and Tsai, 1989; Sugiura, 1978). The advantage of ELM based models over other types of neural networks is their very quick training time, and at the same time they have comparable generalization capabilities.

ELM construction phase

The main concept behind the ELM lies in the random initialization of the hidden layer input weights and biases. According to Huang et al. (2006), the input weights and biases do not need to be adjusted and it is possible to calculate implicitly the hidden layer output matrix and hence the output weights. The network is obtained with very few steps and very low computational cost. Consider a set of M distinct samples (x_j , y_j) with $x_j \in \Re^d$ and $y_j \in \Re$, where d is the number of input features. For each sample $x_{j,} j \in \{1,...,M\}$, the model produces the prediction $\hat{y}_j$ based on the sum

$$\sum_{i=1}^{N} \beta_i f(w_i x_j + b_i) = \hat{y}_j \qquad (1)$$

With f being the activation function, w_i the input weights, b_i the biases and β_i the output weights. Input weights w_i and biases b_i are randomly generated. In our case, x_j is the climate time series and y_j the benthos time series, respectively. The aim is to match the prediction $\hat{y}_j$ to the actual outputs y_j which can be written in matrix form **H**β=**y**, with

$$\mathbf{H} = \begin{pmatrix} f(w_1 x_1 + b_1) & \cdots & f(w_N x_1 + b_N) \\ \vdots & \ddots & \vdots \\ f(w_1 x_M + b_1) & \cdots & f(w_N x_M + b_N) \end{pmatrix} \tag{2}$$

$\beta = (\beta_1, \cdots, \beta_N)^T$ and $y = (y_1, \cdots, y_M)^T$.

The output weights β are computed with the Moore-Penrose generalized inverse of the matrix **H** and the target values, i.e. $\beta = \mathbf{H}^{\dagger}\mathbf{Y}$. The original paper focuses on sigmoid and sine activation functions, but the kernels of neurons are not limited to these two only. In the OPELM, beside sigmoid activation function, the linear and Gaussian kernels are utilized as well. The linear kernel enables OPELM to adapt to the problems that are highly linear by adequate ranking and selection.

LARS ranking phase

As mentioned, LARS provides exact ranking when the problem is linear. This is the case in the basic ELM, where the connection is linear between hidden layer and the output. The output is given from the data y_j, while the "variables" are the outputs of the kernels $h_k = (f(w_k x_1 + b_k), ..., f(w_k x_M + b_k))^T$ (the columns of **H**). With this ranking, we can also assess the importance of different kernels in the model.

Selection phase

The main importance behind any model selection criteria is to estimate how good a model can predict future data. The error during training phase is a poor indicator of the generalization properties of the model. A penalty term has to be added to account for model complexity and number of samples available in the data. As mentioned above, we use the corrected Akaike information criterion.

The Akaike information criterion (AIC) (Akaike, 1974) is based on the information theoretic concept of Kullback–Leibler information (Kullback and Leibler, 1951) and is expressed with the formula:

$$AIC = M \log \hat{\sigma}^2 + 2K \tag{3}$$

where $\hat{\sigma}^2$ is the mean square error on the training data and K the number of parameters of the model. In OPELM case, this is the number of adjustable output weights of the hidden layer. In the situation when K is large relative to the number of samples, or when number of samples is quite low, the second-order correction term is added to the AIC formula giving the corrected AIC_c criterion (Sugiura, 1978).

$$AIC_c = M \log \hat{\sigma}^2 + 2K + \frac{2K(K+1)}{M-K-1} \tag{4}$$

The second order term accounts for the finite sample size, and when $M \to \infty$ both AIC and AIC_c are equal. The goal is to use the model, which minimizes the AIC_c value. For OPELM, the number of neurons giving lowest AIC_c value is chosen as the appropriate complexity of the network.

OPKNN Method

The "optimally pruned *k*–nearest neighbours" (OPKNN) shares a similar approach to the OPELM (Yu et al., 2008). Instead of using random initialization of input weights w and various kernels, OPKNN uses simple *k*–nearest neighbour (KNN) model as its kernel. The key idea behind KNN is that similar training samples should have similar outputs. The similarity is based on some form of distance metric, and the usual approach is to use the Euclidean metric in the input space. In OPKNN, matrix **H** is defined as:

$$\mathbf{H} = \begin{pmatrix} y_{P(1,1)} & \cdots & y_{P(1,N)} \\ \vdots & \ddots & \vdots \\ y_{P(M,1)} & \cdots & y_{P(M,N)} \end{pmatrix} \qquad \textbf{(5)}$$

Where P(j,i) is the index of the i^{th} nearest neighbour of sample x_j and $y_{P(j,i)}$ is the output of that i^{th} nearest neighbour. An important feature of OPKNN is that the model is deterministic, as it does not have any stochastic elements. A comparison of these two models in time series domain is given in (Sovilj et al., 2010).

Results

To demonstrate how the neurons are selected, Fig. 3 shows for the sake of clarity the LOO error versus the number of neurons for a run in which the AO index is projected on the macrofauna biomass. The upper limit of the number of neurons is given by the number of data pairs itself. The optimal number of neurons is estimated by searching the local minimum of the LOO error, which appears at 12 neurons in this case (Fig. 3). Finally, this number of neurons is used for the simulation.

OPELM and OPKNN are applied in a hindcast mode to the median of the biomass, abundance and species number for the whole period 1978-2005 using all climate indices and different SST time series of the North Sea as single predictors as well as in combination in a multivariate predictor. To test the performance of the hindcast, we apply OPELM and OPKNN to the same data set, but skipped the benthic data for the period 2000-2005 in order to predict this period with the models fitted for the previous years. As an example, Figure 4 shows the prediction of median of the biomass obtained from the AO winter index as predictor.

Trends and amplitudes of inter-annual variability are well reproduced for the fitting period (1978-1999). However, the prediction of the neural networks for 2000-2005 is as poor as the linear statistical downscaling method (Dippner et al., 2010).

In a next step, we combine AO, SST of the southern North Sea, precipitation and meridional wind in a multivariate predictor. This predictor is used to train new models with OPKNN with possible lags of up to 11 years. We skipped the OPELM because the OPKNN method proved to be slightly superior regarding the skills over OPELM. The OPELM inferiority is due to inherent randomness of the method, coupled with small number of samples, which leads to high variability in predictions. The OPELM method seems more susceptible to limited data and the randomness of the model. This combination makes any kind of result for OPELM very variable, and therefore inferior to OPKNN. Moreover, since OPKNN is deterministic, the model output is much easier to interpret.

We compare the results with the results of a prediction using only one predictor, the AO (Figs. 5, 6). Again, trends and amplitudes are well reproduced for both runs but the amplitudes are significantly better reproduced using the multivariate predictor. Here, the LOO error is 10.79 [number/0.2m^2] for species number time series, while for biomass it is 1.7*10^{-6} [mg AFDW /m^2]. Both errors are for OPKKN model.

Discussion

Neural networks (OPELM/OPKNN) provide an alternative framework for predicting observations, if prior knowledge of the phenomenon is lacking or completely unknown. With fast learning times, both models provide a suitable framework for testing different combinations of predictors, and ranking provides insight which predictors work best together. This can be achieved in reasonable time with many predictors.

The fact that the prediction of the benthos biomass fails after 2000 even when using non-linear methods affirms an abrupt and discontinuous biological shift after 2000 which was triggered by a climate regime shift (Dippner et al., 2010). However, if the network is allowed to learn also from the years after 2000 the prediction is much better, especially, when a multivariate predictor is considered. This is in agreement with observations that the ecosystem has shifted to a new state where other factors determining how climate acts on the ecosystem have become dominant, e.g. switch from temperature to food availability limited growth.

The biological regime shift has been associated with exceptionally high abundance of juveniles of various macrofauna species off Norderney. The simultaneously occurring climate regime shift is reflected in a lower storm frequency in winter (Loewe, 2009). The resulting calmer hydrodynamic conditions seem to favour tube building polychaetes and burrowing amphipods. These increased after 2000 whereas interface-feeding polychaetes, mobile amphipods and bivalves decreased (Kröncke et al., 2012).

The improvement of the prediction by using a multivariate predictor shows the importance of local forcing on ecosystems (Dippner et al., 2012). The benthos remains unpredictable after the shift, even for the LOO-prediction, when considering only the AO, which proved to be a sufficient predictor for the period before 2000. However, if local properties like precipitation and SST of the southern North Sea are considered, the trend and the dynamic even after 2000 are well reproduced (Fig.5). The prediction of the species number is showing this enhancement exceptionally clearly because the increase in species numbers seems to be related to the increasing SST (Fig. 6). Abundance was affected by a dramatic increase of juveniles in 2002, which was probably caused by exceptional high SST anomalies in the southern North Sea in the 1st quarter of 2002 resulting in early primary production and sufficient food availability for the larvae. But less precipitation since the shift seemed to have led to generally lower primary production (van Beusekom et al., 2009) and less food availability throughout the years, which caused the high mortality of the 2002 juveniles. Thus, primary production or carbon flux might be missing factors in the data analyses and the reason for the weak correlation for abundance data.

The low number of data points however might have the biggest influence on the low predictability of the abundance data. Abundance is dependent on climate through several processes e.g. production of juveniles under good conditions and on the other hand death due to unfavourable conditions, which are acting on different time scales. The time series thus has to have a sufficient length for the neural network to learn these relations. Our time series is not only short (28 years) but also characterised by several regime shifts that lead to different ecosystem states.

Besides showing the importance of local factors in the prediction of ecosystems, this study also shows the feasibility of non-linear methods as the ones used here for studying and predicting ecosystem. In order to mitigate the changes of an ecosystem, it is necessary to identify the dominating factors and to find out whether these are mainly climate or of anthropogenic origin. Linear methods fail to grasp the changes occurring under abrupt regime shifts even after transformation of input data, since these shifts increase the non-linearity of the system. Non-linear multivariate statistical downscaling methods have an advantage over linear methods in this case. However, the validation and interpretation of these results has to occur with great care, as over fitting is common and results sometimes misleading. Linear methods have a great advantage here. They have proven to be robust, requiring less data and computing power and to be sufficient and efficient for the purpose of identifying statistical relationships between systems. Thus, we advise not to use one method for all statistical modelling requirements, but use non-linear methods in combination with linear methods in order to avoid misinterpretation and to improve efficiency.

Acknowledgment

We thank the captains and crews of RV “Senckenberg” for their help with sampling throughout the years. We are indebted to all colleagues for assistance in sampling and sample processing, biomass determination, and taxonomic identification. Peter Loewe from BSH Hamburg has provided the weekly SST data from the German Bight, which is greatly acknowledged. We are also grateful for the very valuable comments of our two reviewers. This paper is a contribution to the BONUS+ ERANET Project AMBER (BMBF Project No. 03F0485A) and also a contribution to the Research Centre “Biodiversity and Climate (BiK-F)” funded by LOEWE.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Becker, G.A., Frey, H., Wegner, G., 1986. Atlas der Temperatur an der Oberfläche der Nordsee: Wöchentliche und monatliche Mittelwerte für den Zeitraum 1971 bis 1980, Dtsch. Hydrogr. Z., Ergebnis-Hefte B 17. Deutsches Hydrographisches Institut, pp. 1-127.

Becker, G.A., Pauly, M., 1996. Sea surface temperature changes in the North Sea and their causes. ICES J. Mar. Sci. 53, 887-898.

Dippner, J., 2006. Future aspects in marine ecosystem modelling. Journal of Marine Systems 61, 246-267.

Dippner, J.W., Junker, K., Kröncke, I., 2010. Biological regime shifts and changes in predictability. Geophys. Res. Lett. 37, L24701.

Dippner, J.W., Kornilovs, G., Junker, K., 2012. A multivariate Baltic Sea Environmental Index. AMBIO accepted.

Dippner, J.W., Kröncke, I., 2003. Forecast of climate-induced change in macrozoobenthos in the southern North Sea in spring. Climate Research 25, 179-182.

Drinkwater, K., Belgrano, A., Borja, A., Conversi, A., Edwards, M., Greene, C., Ottersen, G., Pershing, A., Walker, H., 2003. The response of marine ecosystems to climate variability associated with the North Atlantic Oscillation, in: Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M. (Eds.), The North Atlantic Oscillation, Climatic Significance and Environmental Impact. AGU Geophys. Monogr. , pp. 211-234.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. The Annals of statistics 32, 407-499.

Enfield, D.B., Mestas-Nunez, A.M., Trimble, P.J., 2001. The Atlantic Multidecadal Oscillation and its relationship to rainfall and river flows in the continental US. Geophys. Res. Lett 28, 2077-2080.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70, 489-501.

Hurrell, J.W., 1995. Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. Science 269, 676-679.

Hurrell, J.W., Deser, C., 2010. North Atlantic climate variability: The role of the North Atlantic Oscillation. Journal of Marine Systems 79, 231-244.

Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77, 437-471.

Kröncke, I., Dippner, J., Heyen, H., Zeiss, B., 1998. Long-term changes in macrofaunal communities off Norderney (East Frisia, Germany) in relation to climate variability. Marine Ecology Progress Series 167, 25-36.

Kröncke, I., Reiss, H., Dippner, J.W., 2012. Effects of cold winters and biological regime shifts on the Norderney macrofauna communities in the southern North Sea. J. Mar. Biol. Ass. UK accepted.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. The Annals of Mathematical Statistics 22, 79-86.

Loewe, P., 2009. Atmosphärenphysik. In: Loewe, P. (Ed.), System Nordsee. Berichte des Bundesamtes für Seeschiffahrt und Hydrographie, pp. 35-78.

Miche, Y., Sorjamaa, A., Lendasse, A., 2008. OP-ELM: Theory, Experiments and a Toolbox, in: Kurková, V., Neruda, R., Koutník, J. (Eds.), Artificial Neural Networks - ICANN 2008. Springer Berlin / Heidelberg, pp. 145-154.

Mysterud, A., Stenseth, N., Yoccoz, N., Ottersen, G., Langvatn, R., 2003. The response of terrestrial ecosystems to climate variability associated with the North Atlantic Oscillation, in: Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M. (Eds.), The North Atlantic Oscillation, Climatic Significance and Environmental Impact. AGU Geophys. Monogr., pp. 235-262.

Sovilj, D., Sorjamaa, A., Yu, Q., Miche, Y., Séverin, E., 2010. OPELM and OPKNN in long-term prediction of time series using projected input data. Neurocomputing 73, 1976-1986.

Straile, D., Livingstone, D., Weyhenmeyer, G., George, D., 2003. The response of freshwater ecosystems to climate variability associated with the North Atlantic Oscillation, in: Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M. (Eds.), The North Atlantic Oscillation, Climatic Significance and Environmental Impact. AGU Geophys. Monogr., pp. 263-279.

Suarez, M., Schopf, P., 1988. A delayed action oscillator for ENSO. Journal of the atmospheric Sciences 45, 3283-3287.

Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Communications in Statistics -Theory and Methods 7, 13-26.

Swanson, K., Tsonis, A., 2009. Has the climate recently shifted. Geophysical Research Letters 36.

Thompson, D.W.J., Wallace, J.M., 1998. The Arctic oscillation signature in the wintertime geopotential height and temperature fields. Geophys. Res. Lett. 25, 1297-1300.

van Beusekom, J., Bot, P., Carstensen, J., Goebel, J., Lenhart, H., Pätsch, J., Petenati, T., Raabe, T., Reise, K., Wetsteijn, B., 2009. Eutrophication, in: Marencic, H., de Vlas, J. (Eds.), Quality Status Report 2009. Wadden Sea Ecosystem No. 25, p. 21.

von Storch, H., Zorita, E., Cubasch, U., 1993. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. Journal of Climate 6, 1161-1171.

Wang, G., Swanson, K.L., Tsonis, A.A., 2009. The pacemaker of major climate shifts. Geophys. Res. Lett. 36.

Yu, Q., Sorjamaa, A., Miche, Y., Lendasse, A., Sèverin, E., Guillen, A., Mateo, F., 2008. Optimal Pruned K-Nearest Neighbors: OP-KNN Application to Financial Modeling, HIS 2008, 8th international Conference on Hybrid Intelligent Systems. IEEE, pp. 764-769.

List of Figures

Fig. 1: Study area off the island of Norderney with stations sampled in spring from 1978 to 2005.

Fig. 2: General structure of the OPELM model: X are the climate input data, Y the benthic output data, w the input weights and β the output weights.

Fig. 3: LOO error versus the numbers of neurons. The local minimum gives the optimal number of neurons.

Fig. 4: Median of biomass [mg AFDW /m^2] of benthic macrofauna (full line) and their prediction from AO index using OPELM model (dashed line) and OPKNN model (dotted line).

Fig. 5: Median of biomass [mg AFDW /m^2] of benthic macrofauna (full line) and their prediction from a multivariate predictor using OPKNN model and the AO as a single predictor (dashed line) and a multivariate predictor consisting of AO, SST of the southern North Sea, precipitation and meridional wind (dotted line). Fitting period is the complete time series; shown are the results of the LOO validation.

Fig. 6: Median of species number [number/0.2m^2] of benthic macrofauna (full line) and their prediction from a multivariate predictor using OPKNN model and the AO as a single predictor (dashed line) and a multivariate predictor consisting of AO, SST of the southern North Sea, precipitation and meridional wind (dotted line). Fitting period is the complete time series; shown are the results of the LOO validation.

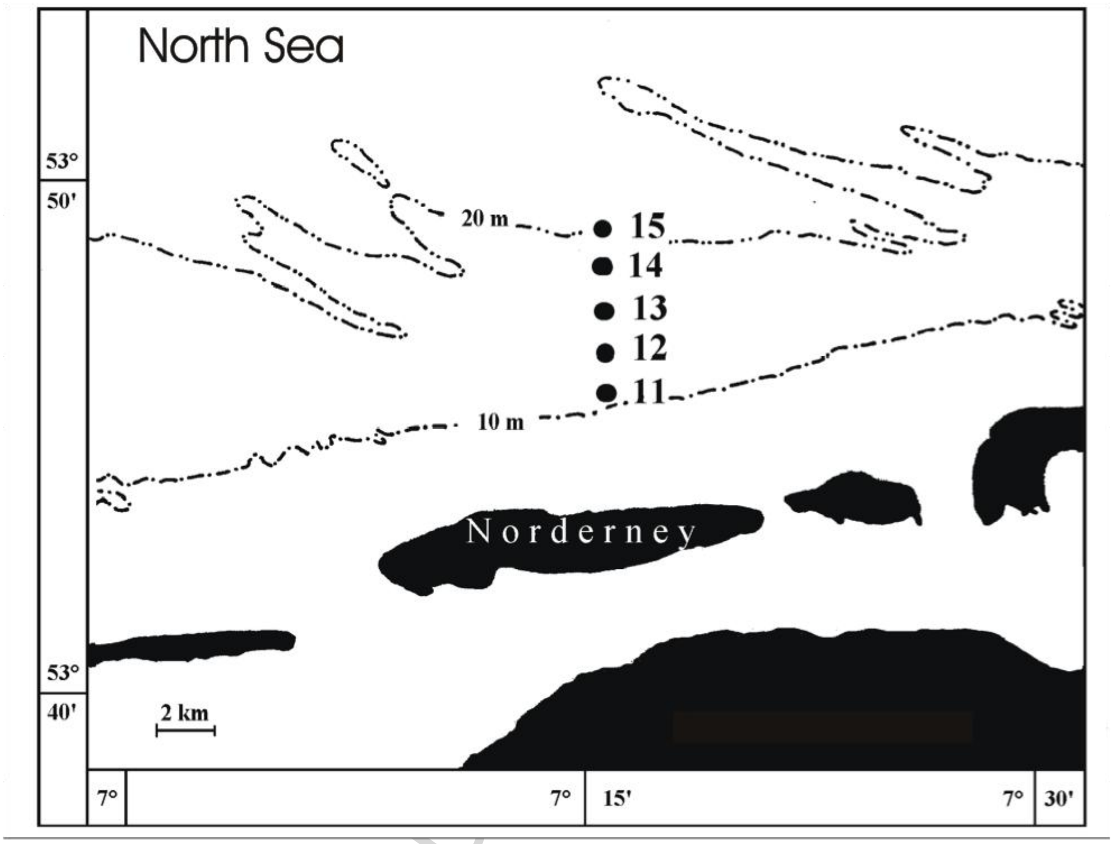
North Sea
53° 50'
20 m
15
14
13
12
11
10 m
Norderney
53° 40'
2 km
7°
7° 15'
7° 30'

Fig. 1

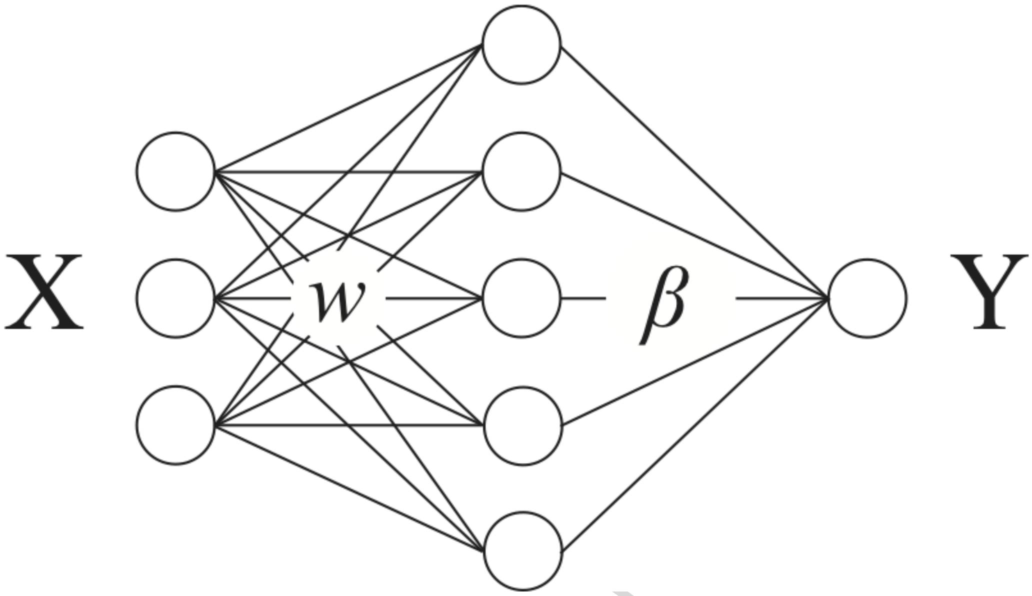

Fig. 2

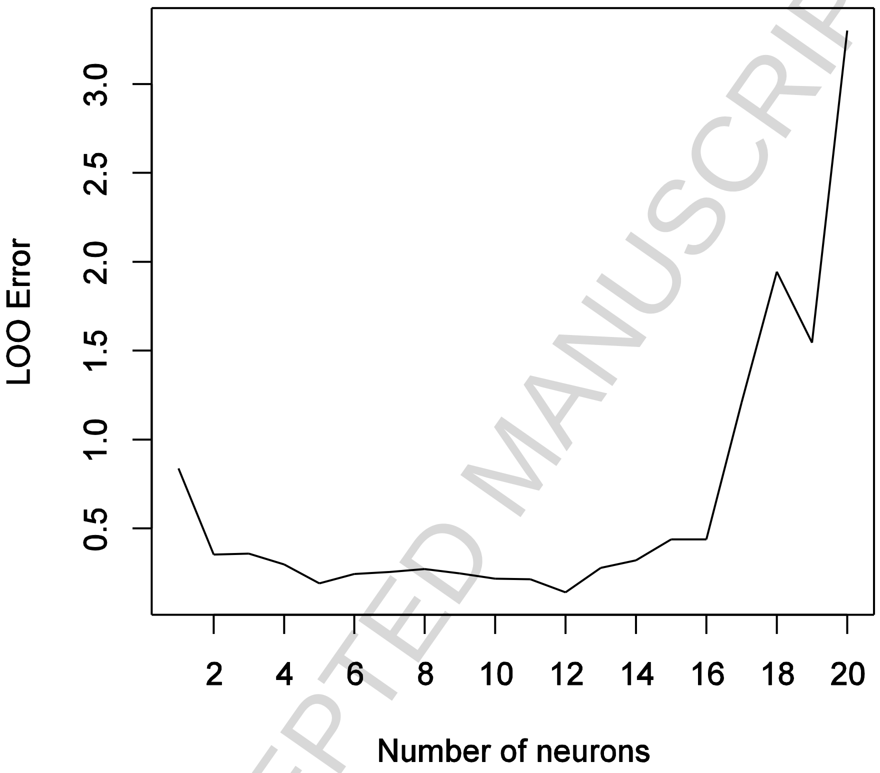
LOO Error
0.5
1.0
1.5
2.0
2.5
3.0
2
4
6
8
10
12
14
16
18
20
Number of neurons

Fig. 3

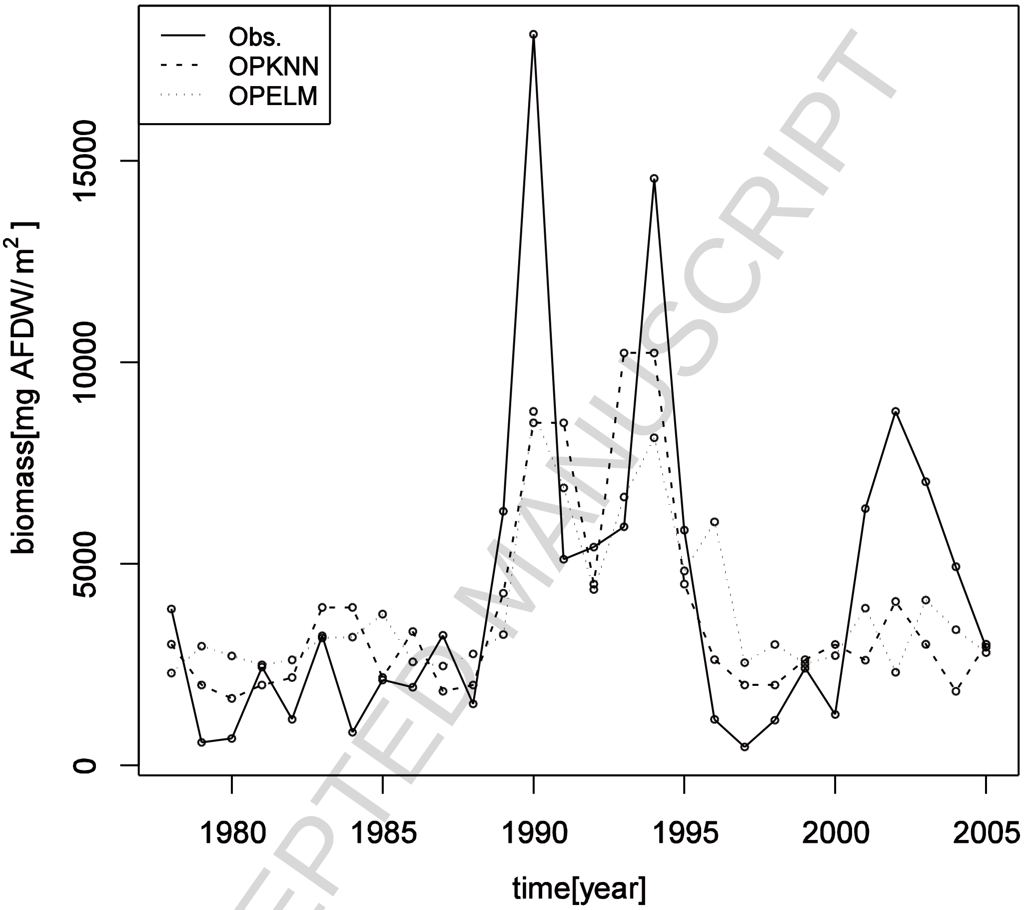
Obs.
OPKNN
OPELM
biomass[mg AFDW/ m^2]
0
5000
10000
15000
1980
1985
1990
1995
2000
2005
time[year]

Fig. 4

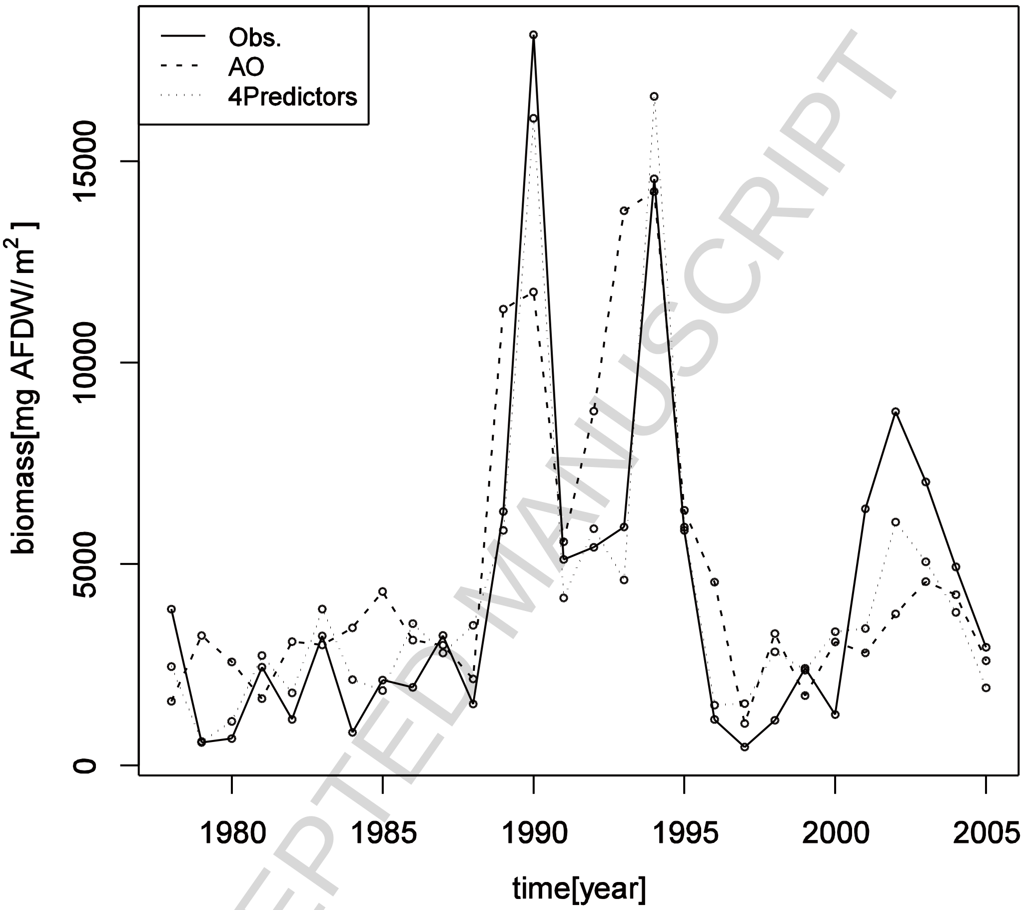
Obs.
AO
4Predictors
biomass[mg AFDW/m^2]
0
5000
10000
15000
1980
1985
1990
1995
2000
2005
time[year]

Fig. 5

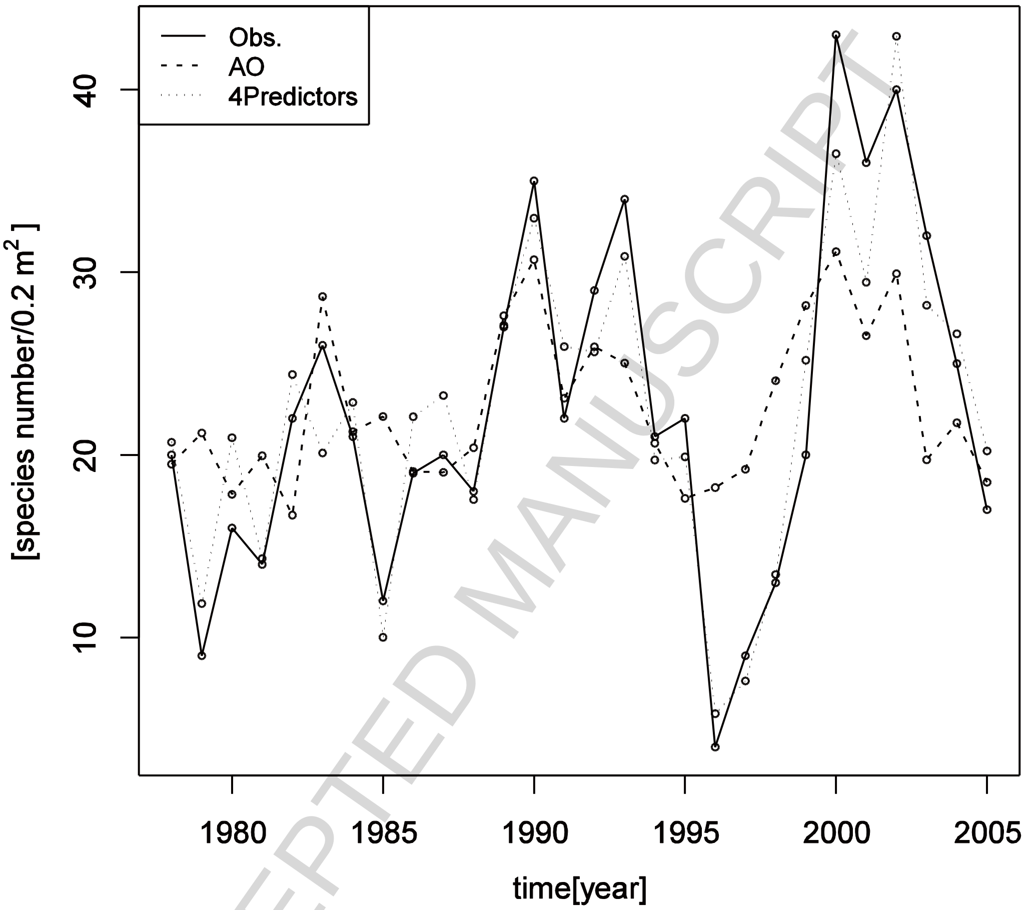
Obs.
AO
4Predictors
[species number/0.2 m^2]
10
20
30
40
1980
1985
1990
1995
2000
2005
time[year]

Fig. 6