Hydrobiologia (2011) 670:165–188
DOI 10.1007/s10750-011-0676-6

ECOSYSTEMS AND SUSTAINABILITY

Fish distribution predictions from different points of view: comparing associative neural networks, geostatistics and regression models

A. Palialexis · S. Georgakarakos · I. Karakassis · K. Lika · V. D. Valavanis

Published online: 25 March 2011

Abstract Accurate prediction of species distributions based on sampling and environmental data is essential for further scientific analysis, such as stock assessment, detection of abundance fluctuation due to climate change or overexploitation, and to underpin management and legislation processes. The evolution of computer science and statistics has allowed the development of sophisticated and well-established modelling techniques as well as a variety of promising innovative approaches for modelling species distribution. The appropriate selection of modelling approach is crucial to the quality of predictions about species distribution. In this study, modelling techniques based on different approaches are compared and evaluated in relation to their predictive performance, utilizing fish density acoustic data. Generalized additive models and mixed models amongst the regression models, associative neural networks (ANNs) and artificial neural networks ensemble amongst the artificial neural networks and ordinary kriging amongst the geostatistical techniques are applied and evaluated. A verification dataset is used for estimating the predictive performance of these models. A combination of outputs from the different models is applied for prediction optimization to exploit the ability of each model to explain certain aspects of variation in species acoustic density. Neural networks and especially ANNs appear to provide more accurate results in fitting the training dataset while generalized additive models appear more flexible in predicting the verification dataset. The efficiency of each technique in relation to certain sampling and output strategies is also discussed.

Keywords Species distribution predictions · Habitat modelling · Models comparison · Geostatistics

Guest editors: Graham J. Pierce, Vasilis D. Valavanis, M. Begoña Santos & Julio M. Portela / Marine Ecosystems and Sustainability

A. Palialexis (✉) · I. Karakassis · K. Lika
Department of Biology, University of Crete, Vassilika Vouton, P.O. Box 2208, 71 409 Heraklion, Crete, Greece
e-mail: andreaspal@her.hcmr.gr

A. Palialexis · V. D. Valavanis
Marine GIS Lab, Hellenic Centre for Marine Research, Thalassocosmos, 71 003 Heraklion, Crete, Greece

S. Georgakarakos
Department of Marine Sciences, University of the Aegean, University Hill, 81 100 Mytilini, Lesvos, Greece

Introduction

The need for scientifically documented studies, regarding characterization of marine ecosystems, fluctuation in species biodiversity and abundance and climate change effects on marine ecosystems, has led to an increase in studies focusing on prediction of species distribution, utilizing different approaches to

habitat modelling and identification of habitat heterogeneity (Guisan et al., 2002; Elith et al., 2006; Valavanis et al., 2008). Such tools should be easily accessible to managers and stakeholders involved in decision-making and management processes while the development of powerful analytical tools and the evolution of geographic information system (GIS) and remote sensing, combined with more accurate and extensive time series data sets, constitute a promising background for the evolution of statistical modelling techniques and the extraction of applicable results. Modelling approaches, such as resource selection functions (RSF) (Valavanis et al., 2008), quantitative structure–activity relationships (QSAR) (Tetko et al., 1995), or geostatistics (Petitgas, 2001) are commonly used for habitat modelling or species distribution predictions. Extensive reviews on species distribution models are available (Guisan & Zimmermann, 2000; Redfern et al., 2006; Richards et al., 2007; Schröder, 2008; Valavanis et al., 2008; Elith & Leathwick, 2009). These approaches could additionally be used for identifying species–environment relationships or inter- and intra-species interactions by utilizing species life history data.

Knowledge on species life-history data becomes crucial during model development and model evaluation. There is always a trade-off among model complexity–overfitting and simplicity–generality of prediction efficiency. According to Guisan & Zimmermann (2000), nature is too complex and heterogeneous to be predicted accurately in every aspect of time and space from a single, although complex, model. Levins (1966) formulated the principle that only two out of three preferable model properties (generality, reality, precision) can be improved simultaneously while the third property has to be sacrificed. The selection of the appropriate modelling technique should be based on the aim of each modelling application.

Accurate estimation of species distribution, based on sampling data, is essential for further scientific analysis, such as stock assessment, understanding of abundance fluctuation due to climate change or overexploitation, and application to management and legislation processes. This study aims to compare several fish distribution prediction techniques based on sampling data with respect to their statistical performance. Advantages and disadvantages of each modelling approach are demonstrated. Furthermore, an attempt is made to optimize the prediction of species distribution by combining the output of the most informative techniques. Species distribution predictions are applied on a training dataset, on an independent verification dataset and on a dataset derived from a wider area (compared to the sampling area). The accuracy and applicability of each modelling technique, the comparisons amongst them and the biological interpretation of small pelagic species are discussed.

Materials and methods

Study area

The study area (Fig. 1) is Thermaikos Gulf, located in the North Aegean Sea (Northeastern Mediterranean). Thermaikos Gulf is a semi-enclosed basin, and relatively productive due to run-off from four major rivers. Bottom relief is smooth due to the continuous sediment input. Thermaikos Gulf forms a wide continental shelf, which extends to the south into the 1,400 m deep Sporades Basin. Water mass circulation is predominantly cyclonic (Poulos et al., 2000). Aegean water masses enter the gulf from deeper layers (Kourafalou & Tsiaras, 2007) along the eastern coast and move counter-clockwise towards the Gulf of Thessaloniki. Riverine waters usually move to the south along the western coast.

Acoustic data

Acoustic data have been collected during April–May 1998 in Thermaikos Gulf using a calibrated 38 kHz SIMRAD EK 500/BI 500 system (Bodholt et al., 1989). The echograms were scrutinized, allocating the nautical area scattering coefficient (s_A, m^2 n mi^{-2}, NASC, MacLennan et al., 2002) to the target pelagic species, such as sardine and anchovy. The integration values, with a horizontal resolution of 1 nautical mile (nm), have been transformed to the presence–absence data (Fig. 1). Acoustic data have not been converted to biomass to avoid the propagation of uncertainty in species composition and length distribution from the trawl sampling in the response variable (Walline, 2007). Species identification based on biological sampling, as well as concurrent catch data, indicated that the majority of the target species were *Sardina*

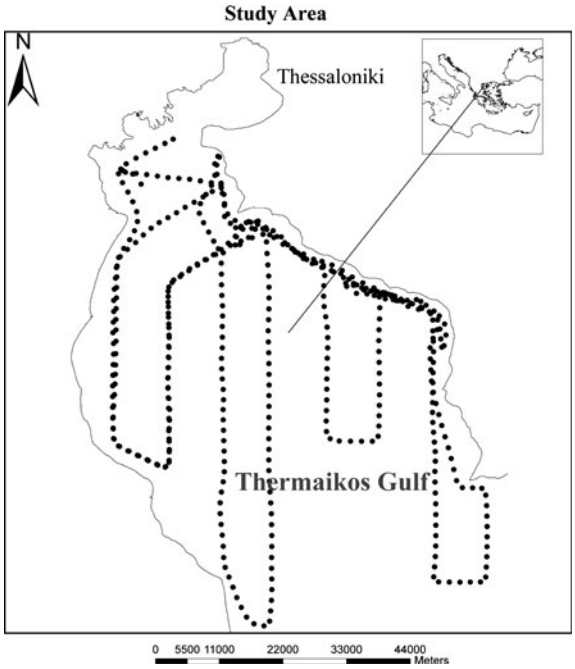

Fig. 1 Study area and sampling transects

pilchardus (~55%), *Engraulis encrasicolus* (~25%) and *Trachurus* spp. (<10%). Measurements were carried out along predetermined sampling transects, until the isobath of 100 m, near Cape Kassandra, in an area of about 1,600 nm^2 (Tsimenides et al., 1992). The acoustic data records allowed the extraction of a well-structured verification data set that was used for model evaluation. This set was constructed following an optimal dataset partitioning process (Tetko et al., 1995) where two adjacent points are separated and one included in the control and the other in the learning set. In our study, the validation dataset was formulated using one value per four sequential sampling values used for the training set. The distance between the points of the verification set (5 nm) was selected based on the fact that at this distance no spatial autocorrelation was detected. The specific selection of the validation set overcomes the underestimation of predicted errors that could caused by cross-validation approaches, especially when acoustic data are spatially autocorrelated (Hastie et al., 2009). In addition, there is no risk of comparing different sampling strategies, since the verification set is a proportion of the raw data. Semi-variograms (Matheron, 1971), calculated using ESRI's ArcGIS geostatistical analyst software (GAS) and auto-correlation function estimation plots in R statistical software (R Development Core Team, 2005), were used for identifying the spatial patterns of the raw data, the training and the verification sets and model residuals. Moran's *I* spatial autocorrelation statistic (Moran, 1950) was also used for estimate the spatial pattern of the two data sets. Furthermore, the homoscedasticity of the residuals was tested by plotting response and explanatory variables against the residuals. By these processes, the suitability of the verification set was tested. In addition, the use of this particular verification set in the validation process overcomes several assumptions that accrued from the use of cross-validation approaches. These differences between the training and the selected verification set, in addition to the fact that the verification set was not introduced to the models, makes the verification set spatially independent.

Environmental and geographic data

Models presented here were developed taking into account the ability of the available environmental parameters to explain species distribution, based on species life history characteristics. Three data sets were extracted for each explanatory variable. The first (DS1) represents the sampling points across transects, the second (DS2) represents each point of the grid that covers the sampling area at a resolution of 0.01 decimal degrees (~1 km) and the third (DS3) corresponds to the verification set. These variables as well as their sources are shown in Table 1. The spatial resolution of the explanatory variables varied from 0.01 to 0.04 decimal degrees according to spatial resolution of the raw data. For modelling purposes, all data sets were re-sampled to the highest available resolution of 0.01 decimal degrees.

A detailed exploration process was performed on the assembled data sets to identify potential spatial patterns, outliers, correlations and interactions. This process is essential for parameter selection during model development to avoid violation of model assumptions, apply any required variable transformations and gain a better understanding of the explanatory data sets. The exploration process included: (a) pair-plots, to identify collinearity, which could lead to biased parameter selection during model development (Zuur et al., 2007), (b) dot-plots and box-plots, for identification of extreme values and outliers, (c) Quantile–Quantile plots (Q–Q plots)

Table 1 Remote sensed data, metadata and their sources

Data variable	Abbreviation	Data type/sensor	Archive source
Acoustic data	s_A	Total acoustic integration (area backscattering coefficient s_A per ESDU = 1 nm)	SIMRAD EK500/BI500 system on April/May 1998 in Thermaikos Gulf
Sea surface temperature	SST	Grid/Aqua MODIS	German Aerospace Agency (DLR)
Chlorophyll-*a* concentration	CHL	Grid/Aqua MODIS	Distributed Active Archive Center (NASA)
Photosynthetically available radiation	PAR	Grid/SeaWiFS	Distributed Active Archive Center (NASA)
Sea level anomaly	SLA	Grid/Merged Jason-1, Envisat, ERS-2, GFO, T/P	AVISO
Precipitation	PRE	Grid	Mediterranean Oceanic Database (MODB)
Sea surface salinity	SSS	Grid/CARTON-GIESE SODA and CMA BCC GODAS models	Mercator operational oceanography
Bathymetry	DEP	Grid/Processed ERS-1, Geostat and historical depth soundings	Laboratory for Satellite Altimetry (NOAA)
Wind stress and direction	WS & WD	Grid & cover	Mercator operational oceanography
Coastline	Coast	Cover/Digitisation of nautical charts and aerial photography	Hellenic Ministry of Environment
Distance to coast	DCoast	Grid and cover	Extracted from coastline
Depth slope	DEPsl	Grid	Extracted from bathymetry grid
Temperature slope (thermal fronts)	SSTsl	Grid	Extracted from SST grid
Marine productivity hotspots	MPH	Grid	According to Valavanis et al. (2004)
Mesoscale thermal fronts	MTF	Cover	According to Valavanis et al. (2005)
Longitude and latitude of stations	LON, LAT	Cover in decimal degrees and metres	SIMRAD EK500/BI500 system on April/May 1998 in Thermaikos Gulf
Current speed and direction	CURSP & CURDR	Grid & cover/NEMO (OPA9 + LIM)	Mercator operational oceanography
Day-dark-night-dawn categorical factor	DDND	Cover and grid	Based on sampling date and hour
Date	DT	Cover	Based on sampling date

and histograms, for an overview of data distribution and (d) conditional box-plots, for displaying potential relationships and interactions (Zuur et al., 2010). Finally, nine variables were selected for model development, according to data exploration results, as descriptors of species distribution. These are sea surface chlorophyll-*a* concentration (CHL), sea surface temperature (SST), sea level anomaly (SLA), bathymetry (DEP), photosynthetically available radiation (PAR), distance-to-coast (DCoast), sea surface temperature slope (SSTsl), depth slope (DEPsl), and day-dark-night-dawn categorical factor (DDND) which is used for explaining the variance of the data that corresponds to behavioural variation of small pelagic species between day and night (Giannoulaki et al., 1999). Dark and dawn represent the times of onset of migration. During the night, small pelagic fish are more scattered and close to the surface while during the day, they form schools in deeper layers. SSTsl and DEPsl were generated by the SST and

DEP grids, respectively, and corresponded to the maximum rate of change in value from each grid-cell to its neighbours. Variables that were excluded after the exploration process (i.e. not used in model development) included date of sampling, precipitation, sea surface salinity, currents speed and direction, wind stress and direction, marine productivity hotspots (Valavanis et al., 2004), mesoscale thermal fronts and hour when each point was sampled. Most of these variables were excluded because they were highly correlated with variables that were actually used in the modelling process. The remotely sensed data were used in the finest temporal resolution available to describe the real-time environmental conditions during sampling. GIS routines (ESRI ArcInfo, version 9.1) were utilized for the conversion of satellite images into grids (Valavanis, 2002) and for the extraction of the environmental values at each sampling point.

Regression models

Regression-based modelling techniques are widely used for predicting species distribution. The most commonly used are generalized additive models (GAMs), generalized linear models (GLM), classification and regression trees (CART) and multivariate adaptive regression splines (MARS). An extensive description of these particular methods can be found in Elith et al. (2006) and Valavanis et al. (2008). In this study, GAMs and generalized additive mixed models (GAMMs) were selected amongst the regression models for species distribution predictions and method comparison purposes, based on the fact that GAMs are widely applied in fisheries science (e.g. Zuur et al., 2007; Palialexis et al., 2009), and they are the most common and well-developed tools for habitat modelling (Valavanis et al., 2008). GAMs are generalized models, involving a sum of smooth functions of covariates (Hastie & Tibshirani, 1990; Wood, 2006). The main advantage of GAMs over traditional regression methods is their ability to model non-linearity using non-parametric smoothers (Hastie & Tibshirani, 1990; Wood, 2006). In addition, according to Moisen & Frescino (2002) and Elith et al. (2006), GAMs perform marginally better than other regression techniques (CART and GLM). They are able to identify species–environment relationships and interactions and to provide biologically interpretable relationships between the response and explanatory variables. The advantage of GAMMs compared to GAMs is that they can deal with spatial autocorrelation, which could lead to biased models and predictions.

The total acoustic integration, natural log-transformed, was used as a response variable. The appropriate transformation method was selected by using Q–Q plots (Cleveland, 1994). Transformation permitted a Gaussian distribution to be assumed for the response variable. The identity link function was used. Depending on the corresponding Q–Q plots, some of the explanatory variables were transformed. The selection of GAM smoothing predictors was based on the method proposed by Wood & Augustin (2002), using the 'mgcv' library in R statistical software (R Development Core Team, 2005). The degree of smoothing was selected based on the observed data and the generalized cross validation method (Wood, 2006). The best-fitting model was selected by using Akaike's Information Criterion (AIC) (Akaike, 1974) and a stepwise forward selection was applied to restrict collinearity amongst the explanatory variables. The GAMMs were developed based on the final GAM model, assuming that a specific correlation structure exists amongst all points of the study area. This structure was modelled using the Gaussian distribution and Gaussian spatial correlation. The GAM and GAMM that better fitted the training dataset were finally selected to predict values in the DS1, DS2 and DS3 data sets.

Associative neural networks (ANNs)

ANNs combine the complexity of some of the statistical techniques with the machine learning objective at a more 'unconscious' level, non-transparent to the user (Michie & Spiegelhalter, 1994). According to Haykin (1994), a neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: Knowledge is acquired by the network through a learning process and, then, interneuron connection strengths known as synaptic weights are used for storing the stored knowledge. ANNs comprise algorithms for mapping the input vector (predictors) to an output vector (responses) through processing elements called 'neurons' via a

training process. The most common type of ANN is a three-layer back-propagation network, which consists of three basic groups of neurons (Ripley, 1996): (a) the input neurons collect the external information and send it to the hidden neurons in the middle layer; (b) the hidden neurons accumulate the previous signals, adjusting each input by certain weights and applying certain threshold functions; and (c) the output neurons accumulate in a second step the input from the hidden neurons, applying again a set of weighting parameters and threshold functions.

The ASNN is a method with improved predictive abilities including combination of feed-forward neural networks and a *k*-nearest neighbour technique, which is a classification method based on distances amongst input samples in space. ASNNs were initially applied in chemistry (Tetko et al., 1995) providing more accurate predictions than artificial neural network ensembles (ANNEs). More detailed information on ASNN development can be found in Tetko (2002a, b). Since this is the first application of ASNN in marine science, both ANNE and ASNN were developed to compare ASNN performance to traditional ANNE and to other modelling approaches.

In this study, an ANNE, which is an averaging neural network prediction process over several independent networks, has been developed using one hidden layer with three neurons. The ASNN model was developed on the ANNE by including the number of the nearest neighbour, k, and parameter σ for the Parzen-window regression, which is a non-parametric way of estimating the probability density function of a random variable. These factors represent smoothing parameters of ASNN to minimize the ASNN error for the training set (Tetko et al., 1995).

The ANNE and ASNN models were selected based on selection processes that include the algorithm, the number of neurons and hidden layers, and the iterations and number of ensembles. The early stopping over ensemble (ESE) method was used for training the neural networks (Bishop, 1995; Tetko & Tanchuk, 2002). In ESE, initial training sets were randomly constructed with equal size learning and validation sets for each neural network in the ensemble. Thus, each neural network had its own learning and validation sets. The learning set was used for adjusting neural network weights. The training was stopped when a minimum error for the validation set was calculated ('early stopping' point). Following ensemble learning, a simple average of all networks was used for predicting the test patterns. Networks ensembles presenting the minor root mean squared error (RMSE) and mean absolute error (MAE) were finally selected (Tetko et al., 2008). These were then processed by using the second-order Levenberg–Marquardt optimization algorithm, i.e. both first- and second-order derivatives of the error function are required for the weight optimization. The Levenberg–Marquardt algorithm usually does not fall into local minima and provides the smallest errors for a fixed number of hidden neurons (Shepherd, 1997; Tetko et al., 2008). Both ASNN and ANNE were developed using three neurons in one hidden layer, 600 networks in each ensemble and 80 iterations in neural network training. These setting were optimized in a modelling selection process according to the lower RMSE and MAE. ANNE and ASNN were developed using the ASNN software VCCLAB (Virtual Computational Chemistry Laboratory, http://www.vcclab.org, 2005), while the final selected model was used for predicting the DS1, DS2 and DS3 data sets.

Kriging

The inverse distance weighted (IDW) and spline methods (i.e. GAMs) are referred to as deterministic interpolation methods because they are directly based on the surrounding measured values or on specified mathematical formulae that determine the smoothness of the resulting surface. A second family of interpolation methods consists of geostatistical methods, such as kriging, which are based on statistical models that include autocorrelation (the statistical relationships amongst the measured points). Thus geostatistical techniques have the capability of producing a prediction surface as well as provide a measure of the certainty or accuracy of the predictions (Georgakarakos & Kitsiou, 2008). The purpose of geostatistics (Matheron, 1971) is to model the spatial variability of a given variable and then utilize the model to estimate the value of the variable at given locations, with extensive applicability to estimating abundance of fish populations from scientific surveys (Petitgas, 2001).

Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used for explaining variation in the

surface. Kriging fits a mathematical function to a specified number of points or all points within a specified radius to determine the output value for each location. It is a multistep process including exploratory statistical analysis of the data, variogram modelling, creation of the surface, and (optionally) exploring a variance surface. Kriging is most appropriate in the presence of a spatially correlated distance or directional bias in the data. Ordinary kriging is the most general and widely used amongst the kriging methods. It assumes that the constant mean of the data is unknown. This is a reasonable assumption unless there is a scientific reason to reject it.

In this study, kriging results are based on geostatistical analysis using ESRI's ArcGIS GAS, which provides an extensive set of tools for performing different kriging methods. Kriging selects weights so that estimations are unbiased and the estimation variance is minimized. The best-fitted variogram is selected after running and evaluating a large number of models, a procedure supported by GAS (Georgakarakos & Kitsiou, 2008).

Finally, the developed kriging models were evaluated by analysing the leave-one-out cross-validation residuals and their statistics were compared and tested in relation to modelling assumptions and whether standard errors estimated by the model are accurate (Isaaks & Srivastava, 1989). The comparison amongst the different developed models was carried out based on the calculation of the RMSE, the average standard error (ASE) and the coefficient of determination (R^2). An ASE close to the root-mean-squared prediction error indicates a correctly assessed variability in prediction. If the ASE is above/below the root-mean-squared prediction error, the variability of the predictions is over/under estimated.

For the purposes of this study, two ordinary kriging models were finally selected. The first one (DefKrig) was developed to minimize errors by using the default settings provided by GAS. The second one (Krig) was parameterized according to data information, regarding trends, data transformations, near neighbour selection and variogram optimization. The above features were incorporated in Krig model to optimize the prediction based on data characteristics that derived through the exploration process. Both models were used in the comparison process because of their differences in explaining the training data variation. While the first model (DefKrig) is error minimization-oriented, the second one (Krig) integrates the results of the exploration process on the training data-set and thus could be characterized as data-oriented.

Prediction optimization

Each modelling technique explains the variance of the observed data from a different point of view. Additive models use a smoothing factor; ANNs are trained based on the observed data, estimating weights and subsequently retrained, and kriging estimates similarities based on distances. A combination of these predictions was applied by creating a GAM model, where the response value is the observed data and explanatory variables correspond to the predictions of each method. This GAM was developed using forward selection and AIC, as described above. The aim of this process is to utilize the predictive capacity of each technique by creating a totally new model, which explains a larger proportion of the variance of the sampling data. This could be characterized as a model averaging approach. The use of a GAM exploits predictions of different techniques in an additive way. This model is referred to as AverMod.

Method comparison

Different aspects of models' predictive performance could be compared by applying a range of evaluation statistics (Potts & Elith, 2006). Pearson's correlation coefficient, r, was used amongst the observed and the predicted values as an index of relative similarity, though a perfect correlation does not necessarily imply exact prediction. In case of perfect correlation, all predictions might be biased in a consistent direction. Spearman's rank correlation coefficient, r_s, was used for indicating the similarity between the ranks of observed and predicted values. The use of ranks means that, as long as the order of predictions is correct, the value of the statistic will be high. Kendall's rank correlation coefficient, τ, is similar to Spearman's r_s and was used for measuring the degree of correspondence between observed and predicted rankings.

The model calibration process used in this study is analogous to Pearce & Ferrier (2000) calibration analysis for binomial data (see also Potts & Elith, 2006) and is applied to complement the previously mentioned correlation indices. A simple linear

regression between observed and predicted values (e.g. observed $= m$ (predicted) $+ b$) was applied, providing information for the bias and consistency of the prediction. In this equation, b provides information regarding the bias, while m provides information about the spread of predictions compared to the spread of the observations. A perfectly calibrated model would be a 1:1 regression line. A parallel regression line ($m = 1$, $b \neq 0$) to the 1:1 line indicates consistent bias. When $b \neq 0$ and $m \neq 1$, then the predicted values are both biased and distributed in a broader or narrower range than the observed values (Fig. 2).

Several summary statistics were estimated by comparing the predicted to the observed values. The objective of these statistics is to make an informative decision on which model provides more accurate prediction. Mean error (ME) should be close to 0, if the predictions are unbiased, and the root-mean-square standardized prediction error (RMSqSt) should be close to 1, if the standard errors are accurate. The root-mean-square prediction error (RMSE) should be relatively low, if the predictions are close to the measured values. Mean square error (MSqE) and mean standardized error (MStE) should be close to zero for accurate predictions while lower values of MSqE are generally interpreted as best explaining the variability in the observations. MStE is used as an assessment of uncertainty. MAE and ASE are quantities used for measuring how close predictions are to the eventual outcomes (the lower the ASE, the better the predictions). All the above statistics were estimated from the prediction errors of the final models selected.

In addition, some more summary statistics that describe the observed and the predicted values have been estimated, like the average value (Ave) and the standard deviation (StDev).

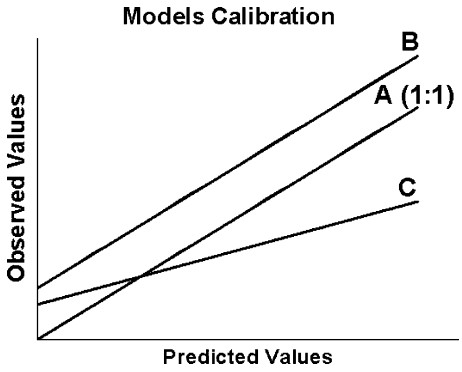

Fig. 2 Model calibration: *A* a perfectly calibrated model, *B* a model with consistent bias, *C* predicted values derived from this model are both biased and distributed in a broader or narrower range than the observed values

Results

Species distribution predictions

Regression models

After data exploration, parameter selection and model optimization, a final GAM model (GAM1) was selected. Another model (GAM2), nested within GAM1, was generated by including a temporal categorical variable as a factor. The quality characteristics of GAM2 were slightly improved compared to GAM1 (AIC and deviance explained, Table 2), however, GAM2 used more degrees of freedom than GAM1, which indicates a more complex model. Both GAM1 and GAM2 were used in the comparison process to examine if the increase in complexity of GAM2 improves the predictive efficiency of the model without decreasing model generality. Together with the final selected mixed model (GAMM1) these two additive models, were used for predicting species distribution and were compared based on characteristics related to their predictive efficiency, generality and biological interpretability (Table 2).

All models were statistically significant (Table 2). According to the AIC criterion (the lower the better), GAM2 is better than GAM1. In addition, the deviance explained (not available in GAMM1) and the adjusted R^2 suggest that GAM2 explains a higher proportion of the response variance than other models. GAM2 performs better in describing the relationship between acoustic data and environmental factors.

Predictions on DS1 (Fig. 3) indicate that in certain areas models failed to accurate predict the observed data (in the north and northwest). This is also apparent in predictions for the DS3 set. Other areas were predicted relatively well in both data sets. Bar charts indicate that GAM2 and GAMM1 perform almost equally efficiently while there are visible differences between these two and GAM1.

Table 2 Final generalized additive models and their characteristics

Model's code	Explanatory variables	Dev. exp.	Res. d.f.	R^2_a	AIC	*P*-value
GAM1	s(SLA) + s(DEP:PAR)	38.9%	23.687	0.345	570.35	$\ll$0.05
	s(SLA)		5.105			$\ll$0.05
	s(DEP:PAR)		18.682			$\ll$0.05
GAM2	s(SLA) + s(DEP:PAR) + as.factor(DDND)	47.9%	29.678	0.458	510.76	$\ll$0.05
	s(SLA)		7.051			$\ll$0.05
	s(DEP:PAR)		22.627			$\ll$0.05
	as.factor(DDND)					$\ll$0.05
GAMM1	s(SLA) + s(DEP:PAR) + correlation factor	NA	24.597	0.444	NA	$\ll$0.05
	s(SLA)		5.561			$\ll$0.05
	s(DEP:PAR)		19.036			$\ll$0.05
	Correlation factor					$\ll$0.05

Level of significance was set at 0.05. The ':' sign denotes interaction

Dev. Exp. deviance explained, *Res. d.f.* residual d.f., R^2_a adjusted R^2, *AIC* Akaike Information Criterion value, *P-value (chi-square)* significance values, *s* denotes smooth function of predictors

ANNs

Both ANNE and ASNN were developed using the following explanatory variables: DDND, SLA, CHL, PAR, SST, DEP, DCoast, DEPsl and SSTsl. The performances of ANNE and ASNN models for the training set are presented in Table 3.

For DS1 (Fig. 4, right), both models performed almost equally well. Visually, they seem to provide very accurate predictions in the whole area, while the local variation of the sampling data is also accurately predicted. Predictions for DS3 (Fig. 4, left) indicate that both models describe the global trend of the data accurately, although they under- or over-estimate the relatively high or low observed values of the independent set, respectively.

Kriging

Characteristics of the ordinary kriging models are presented in Table 4. The two final selected models differ significantly. The Krig model could be characterized as a locally oriented model compared to DefKrig, which is more globally oriented. This observation is also depicted in Fig. 5 where predictions differ significantly. The right map (Fig. 5) confirms that the Krig model where predictions depicted high and low observed values more accurately at a local scale, compared to DefKrig, which under-estimates the observed values. This is also obvious in predictions on the DS3 (Fig. 5, left).

Prediction optimization

The final selected model and its characteristics are presented in Table 5. The AverMod is a relatively simple GAM with two explanatory variables and relatively low degrees of freedom. However, the deviance explained is relatively high (45.2%). GAM2 was also tested, as an explanatory variable, since it performs statistically better than the other additive models but it was not significant ($P > 0.05$) and it was dropped during the selection process. Figure 6 indicates that predictions derived by AverMod for the DS1 set are relatively accurate over the whole extent of the study area (Fig. 6, left). Predictions for the independent set DS3 showed a tendency to under-predict high observed values and over-predict low values (Fig. 6, right).

Comparison

Model fit comparison

Model fit comparison was achieved by applying the previously mentioned estimates to DS1 and comparing predicted values from each model to the sampling observed values. In order to identify the most accurate fit to the training data, several model

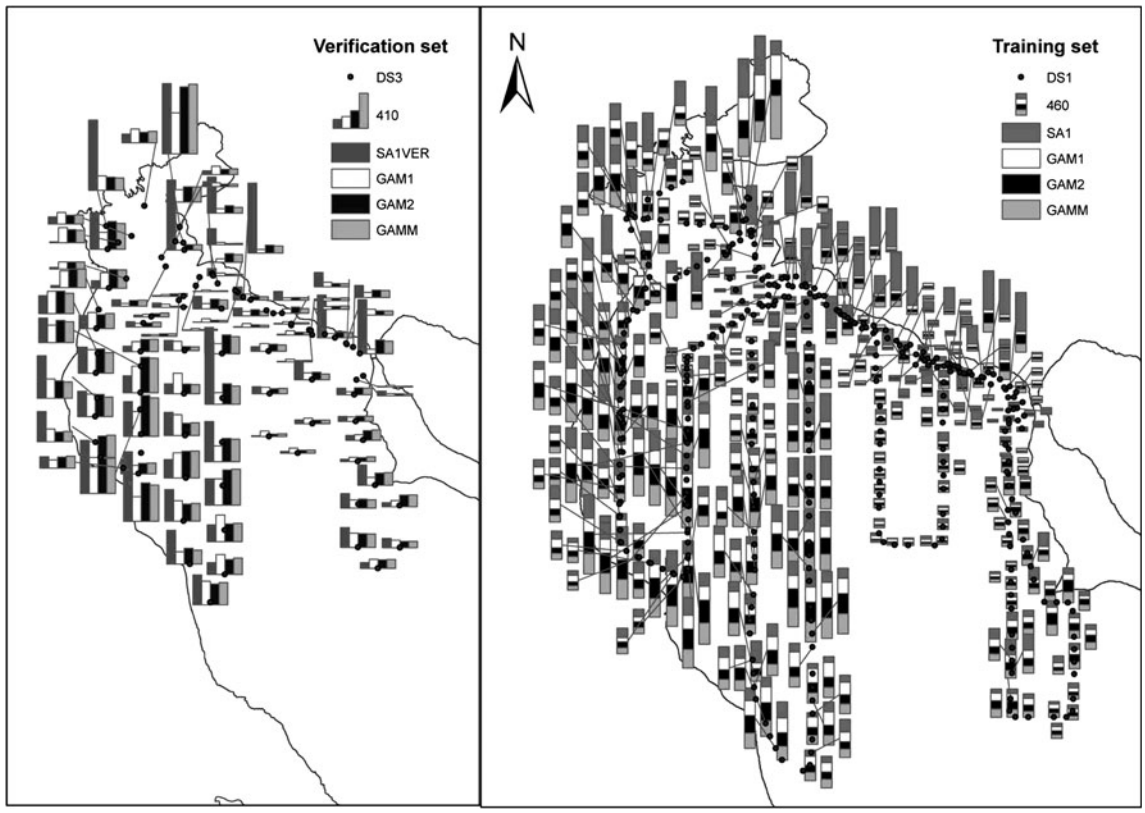

Fig. 3 A comparative representation of the observed (SA1VER for the verification set and s_A1 for the training set) and predicted values (GAM1, GAM2, GAMM) derived from generalized additive models and mixed model. Hued bar's height represents the observed and predicted fish density in each sampling point. *Left map* corresponds to the verification set (DS3-black points) and *right map* corresponds to the training set (DS1-black points). The *scale of the bar* which is shown in legend is 410 and 460, respectively

Table 3 Performance characteristics of ANNE and ASNN

Network, k, (σ)	LOO results		
	RSME	MAE	r^2
ANNE	191.10	136.64	0.219
ASNN, 54, (0.01)	176.77	116.25	0.325

RMSE root mean square error, *MAE* mean absolute error, r^2 square of correlation coefficient, *LOO* leave one out, k k near neighbour, σ sigma

comparison indices were used and presented in Table 6.

According to Pearson's correlation coefficient (Table 6), the similarities amongst observed and predicted values are better described by Krig while ASNN and ANNE are also efficient in their predictions. GAM2 performs better than the other additive models. On the other hand, Spearman and Kendall correlations indicate that the rank correlation between observed and predicted values was higher for Krig and GAM2 and lower for neural network techniques.

The model calibration process (Fig. 7; Table 6) indicates that neural networks are almost perfectly calibrated by minimizing the bias compared to the other approaches and by predicting values in the same range as observed values, as shown by the overlap with the 1:1 regression line in Fig. 7. There are great similarities amongst the calibrations of additive models where they present an almost constant bias and the ranges of their predicted values are quite close. The regression lines of kriging models vary greatly and seem to produce under-calibrated models.

Error analysis of predictions on DS1

Summary statistics (Ave, StDev) of the observed (s_A1-DS1) and the predicted sets as well as some

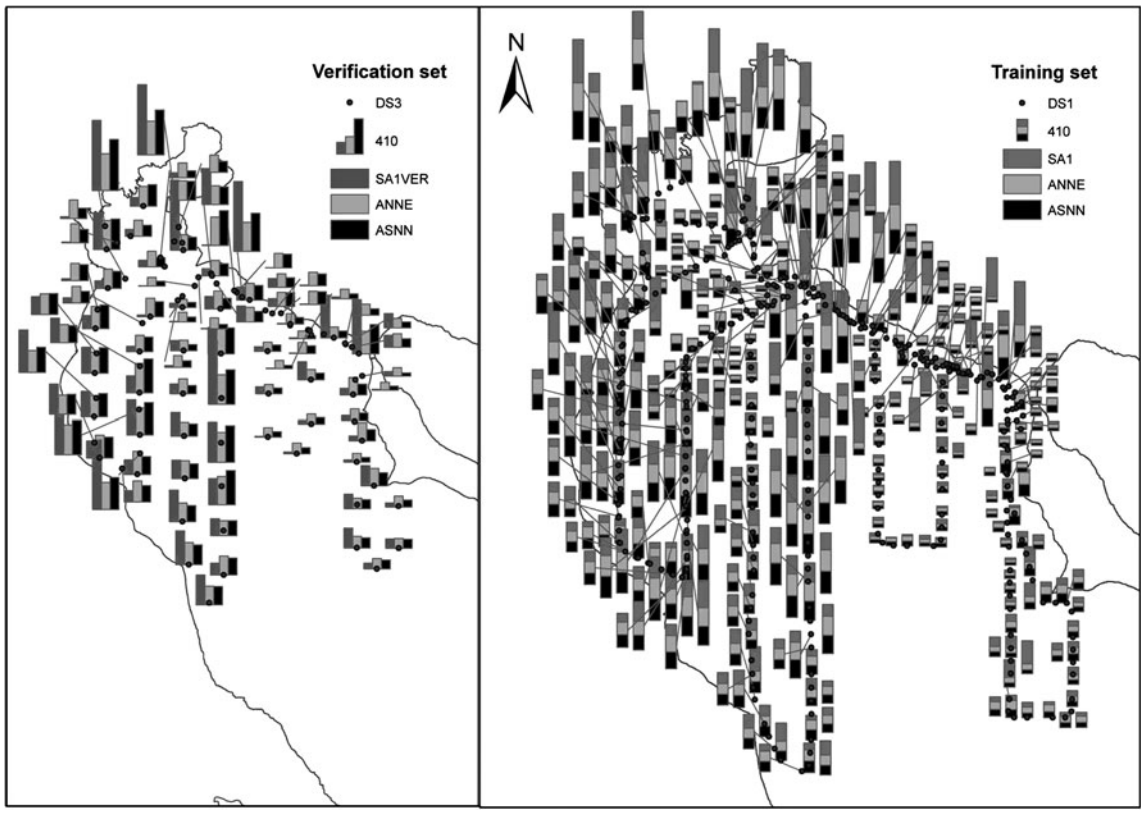

Fig. 4 A comparative representation of the observed (SA1VER for the verification set and s_A1 for the training set) and predicted values (ASNN and ANNE) derived from associative neural networks and artificial neural networks ensemble. Hued bar's height represents the observed and predicted fish density in each sampling point. *Left map* corresponds to the verification set (DS3-black points) and *right map* corresponds to the training set (DS1-black points). The *scale of the bar* which is shown in legend is 410

Table 4 Ordinary kriging models' characteristics

Krig	DefKrig
First order trend removal	No trend removal
Variogram	Variogram
15 number of lags	12 number of lags
Lag size 1852 m	Lag size 6965.1
Nugget 1.184	Nugget 46257.18
Model spherical	Model spherical
Range 26571.13 m	Range 79065.67
Anisotropy	No anisotropy
Minor range 5100.6	
Direction 3.35°	

error indices of the predictions are presented in Table 7. AverMod and DefKrig average predictions are close to the average of observed values. The ASNN and ANNE averages are also close to s_A1-DS1. The average predictions of additive models are lower than the observed while Krig over-predicts the observed average. The StDev of the observed data is relatively high, which generally characterizes the nature of acoustic data. Only the Krig model generates a similar StDev value while other models all predict lower StDev.

Error indices show that generally AverMod, DefKrig, ASNN and ANNE (in decreasing order) provide more accurate, unbiased predictions and better explain the variability of observed values, according to ME, MSqE, MStE, RMSqSt and RMSE. MAE and AverMod show that GAM2 also predicts values close to the observed. According to ME and the sum of errors, only Krig over-estimates the observed data while all other models either predict accurately or under-estimate acoustic data. In conclusion, error analysis indicates that AverMod, ASNN, ANNE and DefKrig fit the data on which

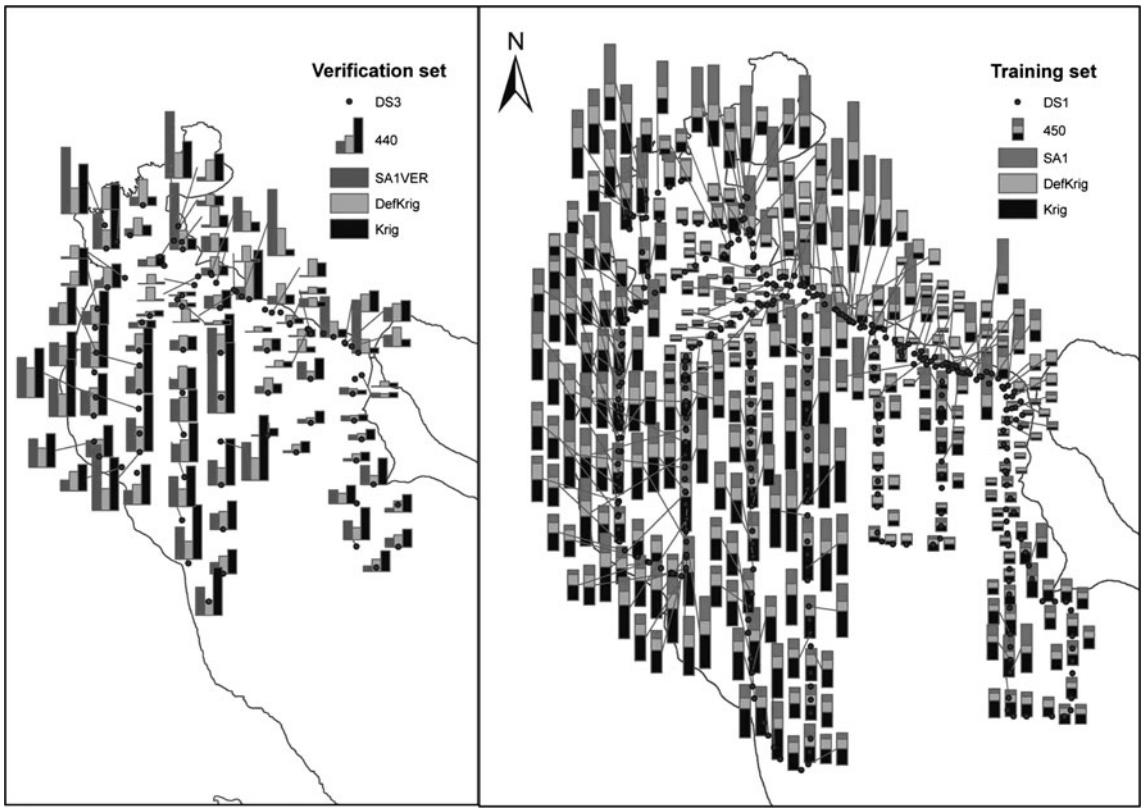

Fig. 5 A comparative representation of the observed (SA1VER for the verification set and s_A1 for the training set) and predicted values (DefKrig and Krig) derived from ordinary kriging approaches. Hued bar's height represents the observed and predicted fish density in each sampling point. *Left map* corresponds to the verification set (DS3-black points) and *right map* corresponds to the training set (DS1-black points). The *scale of the bar* which is shown in legend is 440 and 450, respectively

Table 5 Characteristics of the GAM model used for the optimization of predictions

Model's code	Explanatory variables	Dev. exp.	Res. d.f.	R^2_a	*P*-value
AverMod	s(Krig predictions) + s(ASNN predictions)	45.2%	11.093	0.436	$\ll$0.05

Level of significance was set at 0.05

Dev. Exp. deviance explained, *Res. d.f.* residual degrees of freedom, R^2_a adjusted R^2, *P-value (chi-square)* significance values, *s* denotes smooth function of predictors

they are trained more accurately than did the other models.

Comparing predictions on the verification dataset

Evaluation of model predictions on the verification dataset was achieved by applying the above-mentioned estimates to DS3. The best performing models would be characterized by generality and could predict species distribution along extended spatial and temporal axes more efficiently. Table 8 presents several model comparison indices to identify the most accurate fit to the independent verification set.

According to Pearson's correlation coefficient (Table 8), the similarities amongst observed and predicted values are better described by neural networks and especially by ASNN. GAM2 performs better than the other additive models while Krig performs better than DefKrig. On the other hand, Spearman and Kendall correlation analyses indicate that the correlation between the ranks of observed and predicted values is higher for Krig while there is

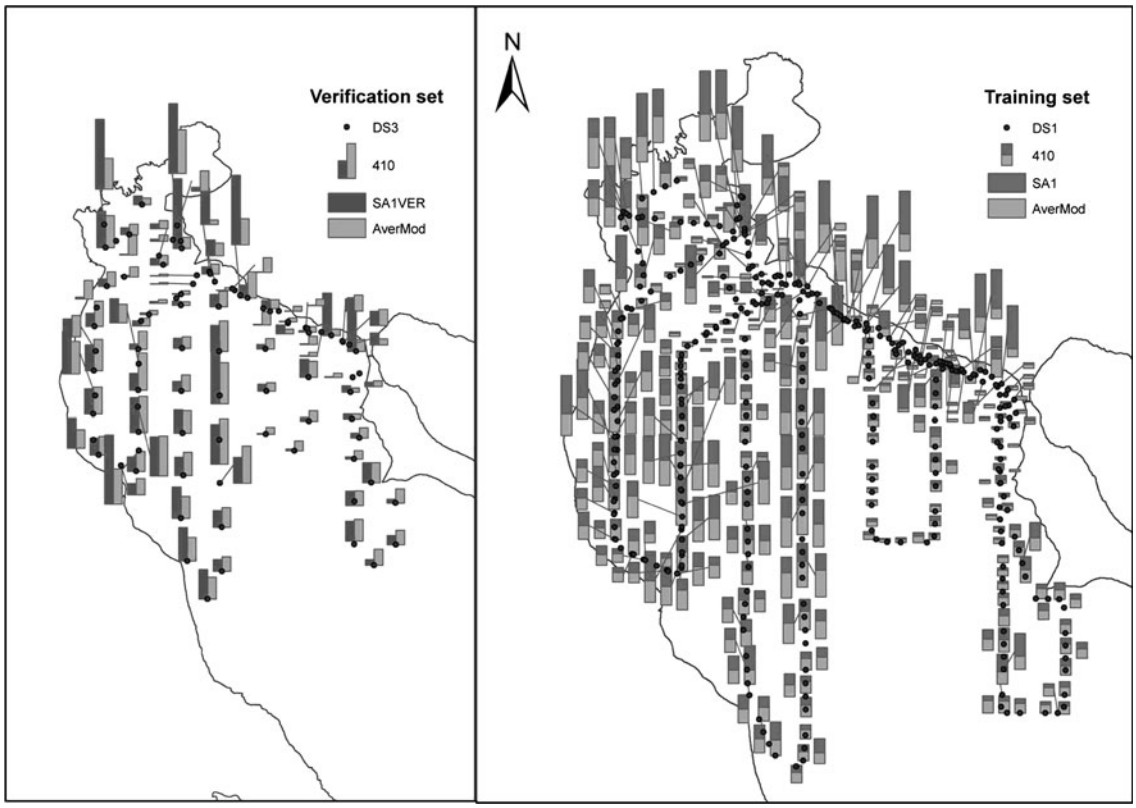

Fig. 6 A comparative representation of the observed (SA1VER for the verification set and s_A1 for the training set) and predicted values (AverMod) derived from prediction optimized model. Hued bar's height represents the observed and predicted fish density in each sampling point. *Left map* corresponds to the verification set (DS3-black points) and *right map* corresponds to the training set (DS1-black points). The *scale of the bar* which is shown in legend is 410

Table 6 Model comparison indices used for identifying the most accurate fit on training data

	GAM1	GAM2	GAMM1	DefKrig	Krig	ANNE	ASNN	AverMod
Correlation								
Pearson's r	0.38	0.49	0.45	0.50	0.59	0.57	0.56	0.67
Spearman's r_s	0.62	0.71	0.69	0.54	0.79	0.55	0.56	0.81
Kendall's τ	0.45	0.53	0.51	0.39	0.62	0.39	0.40	0.64
Calibration								
Intercept, m	0.821	0.819	0.823	1.160	0.592	0.980	0.967	1.019
Slope, b	92.89	82.65	85.51	−30.22	37.07	12.31	15.42	−3.67
R^2	0.145	0.224	0.206	0.250	0.351	0.327	0.316	0.453

little difference amongst the other techniques. The results of the model calibration process (Fig. 8; Table 8) indicate that additive model predictions are characterized by a constant bias, although the spread of the predicted values is similar to the spread of the observed values. The predictions obtained by means of neural networks for the verification set are generally less accurate than in training set, presenting bias and wider range of values, compared to the observed. The predictions derived from Krig showed a significant divergence from the 1:1 regression line while DefKrig predicts more accurately than Krig according to the evaluation process in the DS3.

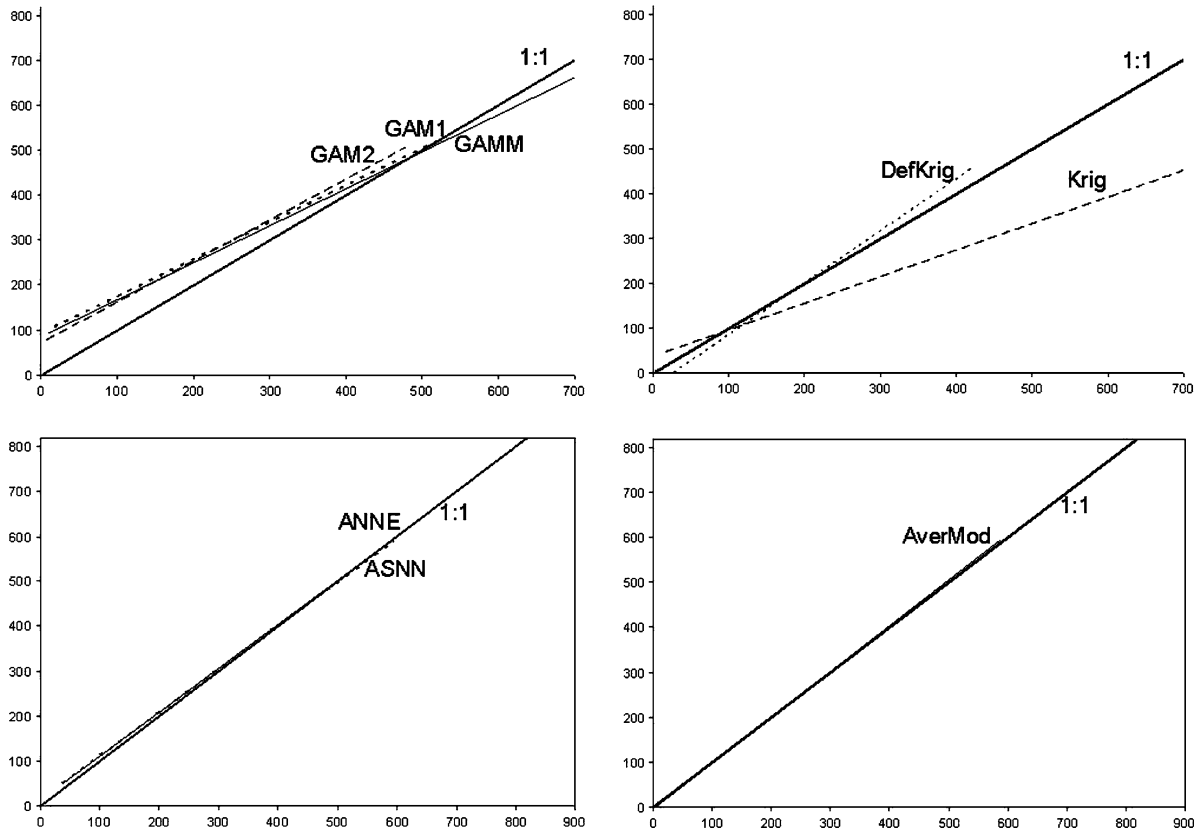

Fig. 7 Models' calibration indicated by regression lines amongst observed (*y*-axis) and predicted (*x*-axis) values on training dataset (DS1). The 1:1 regression line corresponds to a perfect calibrated model

Table 7 Models' comparison error indices used for identifying the best fit on training data

DS1	s_A1-DS1	GAM1	GAM2	GAMM1	DefKrig	Krig	ANNE	ASNN	AverMod
Ave	193.414	122.348	133.445	131.089	192.869	263.935	184.856	184.104	193.414
StDev	215.192	99.740	117.158	118.775	92.748	215.353	125.587	125.078	142.058
RMSE		1335.194	1126.701	1170.972	10.232	1324.978	160.789	174.904	~0.000
MSqE		1782742.121	1269454.155	1371174.907	104.696	1755565.923	25853.085	30591.264	~0.000
MAE		117.781	107.436	110.484	130.249	152.782	116.248	116.187	97.469
ASE		0.752	0.728	0.739	0.728	0.742	0.707	0.710	0.672
MStE		−0.356	−0.320	−0.323	−0.003	0.363	−0.048	−0.052	~0.000
RMSqSt		6.684	6.016	6.072	0.055	6.820	0.911	0.982	~0.000
ME		−71.065	−59.968	−62.325	−0.545	70.521	−8.558	−9.309	~0.000

Error analysis of predictions on DS3

Summary statistics (Ave, StDev) for the observed (s_A-DS3) and the predicted sets as well as error indices for the predictions are presented in Table 9. The error analysis of the predictions on the verification set DS3 generated different results to those obtained from the error analysis of the predictions on the training set (DS1). In this case, DefKrig and ANNE predicted the average observed values more accurately while ASNN and AverMod are also similar. Again, Krig is the only model that provides

Table 8 Models' comparison indices used for identifying the most accurate fit on the verification dataset

	GAM1	GAM2	GAMM1	DefKrig	Krig	ANNE	ASNN	AverMod
Correlation								
Pearson's r	0.48	0.58	0.56	0.53	0.58	0.67	0.73	0.67
Spearman's r_s	0.62	0.68	0.67	0.62	0.71	0.62	0.65	0.73
Kendall's τ	0.44	0.50	0.48	0.43	0.56	0.45	0.50	0.57
Calibration								
Intercept, m	1.090	0.981	0.954	1.317	0.586	1.871	1.317	1.154
Slope, b	62.92	61.83	67.97	60.09	37.33	175.28	40.79	22.14
R^2	0.233	0.338	0.315	0.281	0.332	0.454	0.536	0.445

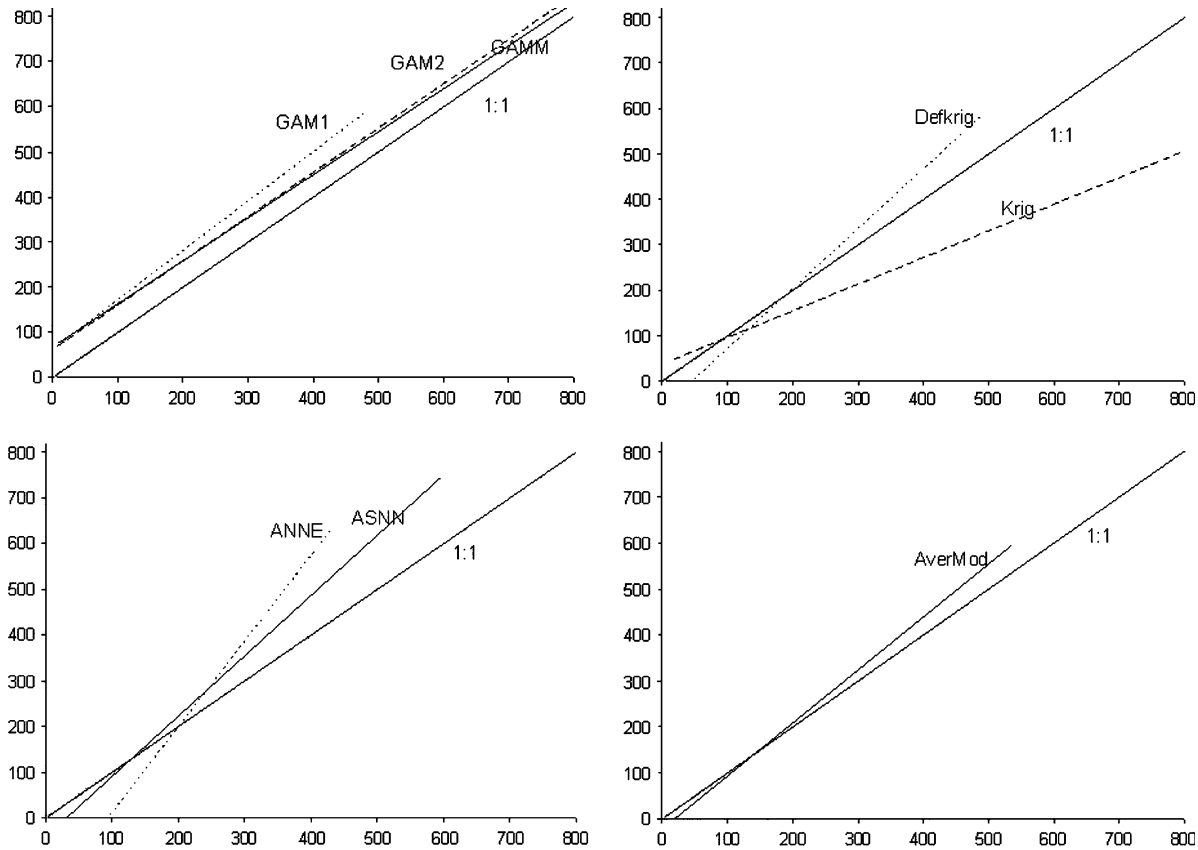

Fig. 8 Models' calibration indicated by regression lines amongst observed (y-axis) and predicted (x-axis) values on the independent verification dataset (DS3). The 1:1 regression line corresponds to a perfect calibrated model

predictions with StDev near to that of the observed values. Error indices show that DefKrig and ANNE provide the most accurate and unbiased predictions while AverMod and ASNN are also efficient. According to MAE, ASNN and GAM2 predict values similar to the observed values (AverMod, as well). ME and the sum of errors show that Krig over-estimates the observed data. Predicted values from DefKrig and ANNE are very similar to the observed values while all other models are under-estimating the observed acoustic data. The analysis of the errors derived from the predictions on DS3 indicates that ANNE outperforms ASNN. ANNE and DefKrig provide the most unbiased and accurate predictions on an independent

dataset, even if AverMod and ASNN fit the training data more effectively. In both cases, GAM2 performs best amongst the additive models.

Comparing predicted grids

A global trend in acoustic density, which is described by low values to the South–Southeast and high values to the North–Northwest and in the centre of the study area, is observed in Figs. 4, 5 and 6. Locally, there is significant variation of observed values, especially in the northeast. The predicted grids (Figs. 9, 10, 11 and 12) were generated to evaluate model predictions on an extended spatial scale and depict species potential distribution in a biologically and oceanographically distinct region, such as the Thermaikos Gulf. The crucial question is which model prediction describes species habitats and distribution more accurately, both globally and locally. Values in grids (Figs. 9, 10, 11 and 12) represent acoustic fish density (black: high values, white: low values). In addition, circles represent the sampled acoustic fish density using an opposite colour-scale to that used for the grids (black circles: low values, white circles: high values).

Predicted grids generated by GAMs are shown in Fig. 9. According to model calibration, correlations, summary statistics and error analysis, GAM2 performs better than the others techniques while GAMM1 performs better than GAM1. Figure 9 provides the possibility of a visual comparison amongst predicted grids and observed values (dots). The three models could be characterized by a common pattern in predicting potential species distribution. However in GAM2 and GAMM1, there is significant evidence of the DDND factor, which causes sudden changes in the predicted acoustic density. This observation reflects the effect of day-dark-dawn-night pattern in the species distribution. Since the environmental data were selected to approximate the real sampling time, the predicted grids correspond to species distribution at the temporal scale of the sampling effort. Model generality was tested mostly in spatial mode, rather than in temporal mode, by using DS2 and DS3. GAM1 depicts a more smooth acoustic density distribution, which is in agreement with the observed global trend but fails to accurately depict the local variation and heterogeneity of the potential predicted species habitats. There are no significant differences between GAMM1 and GAM2 ($r = 0.98$).

DefKrig and Krig generated grids (Fig. 10) differ significantly ($r = 0.36$), showing the different way that these models were developed. Error analysis provides evidence of this particular difference, which is also noticeable in the scale of the predicted density in each map (Fig. 10). DefKrig does not clearly present the global trend of the acoustic density distribution, which is more obvious in Krig. On the other hand, Krig seems to over-fit the training data especially in the centre of the sampling area where the predicted high acoustic density areas mostly overlap the observed high density sampling points.

The generated predicted grids (Fig. 11) derived from ASNN and ANNE are very similar ($r = 0.98$). The DDND pattern is less obvious than in GAMM or GAM2. In general, both grids provide a smooth and un-clustered potential density distribution and they preserve model generality, which is shown by the efficient predictions in DS3. The expected global

Table 9 Models' comparison error indices and summary statistics used for identifying the best fit on the verification dataset

DS3	s_A1-DS3	GAM1	GAM2	GAMM1	DefKrig	Krig	ANNE	ASNN	AverMod
Ave	195.011	121.167	135.830	133.212	193.649	269.270	197.925	179.003	188.242
StDev	218.712	96.831	129.724	128.699	87.956	215.041	78.762	121.495	126.426
RMSE		696.642	558.315	583.008	12.852	700.551	27.485	151.023	63.861
MSqE		485310.678	311715.763	339898.565	165.183	490772.162	755.408	22808.006	4078.247
MAE		119.579	106.025	109.288	134.417	154.808	130.066	104.537	101.902
ASE		1.568	1.468	1.414	1.647	1.452	1.498	1.325	1.315
MStE		−0.338	−0.309	−0.347	−0.006	0.396	0.015	−0.102	−0.044
RMSqSt		3.185	2.912	3.276	0.053	3.735	0.138	0.967	0.415
ME		−73.844	−59.181	−61.799	−1.362	74.258	2.913	−16.008	−6.769

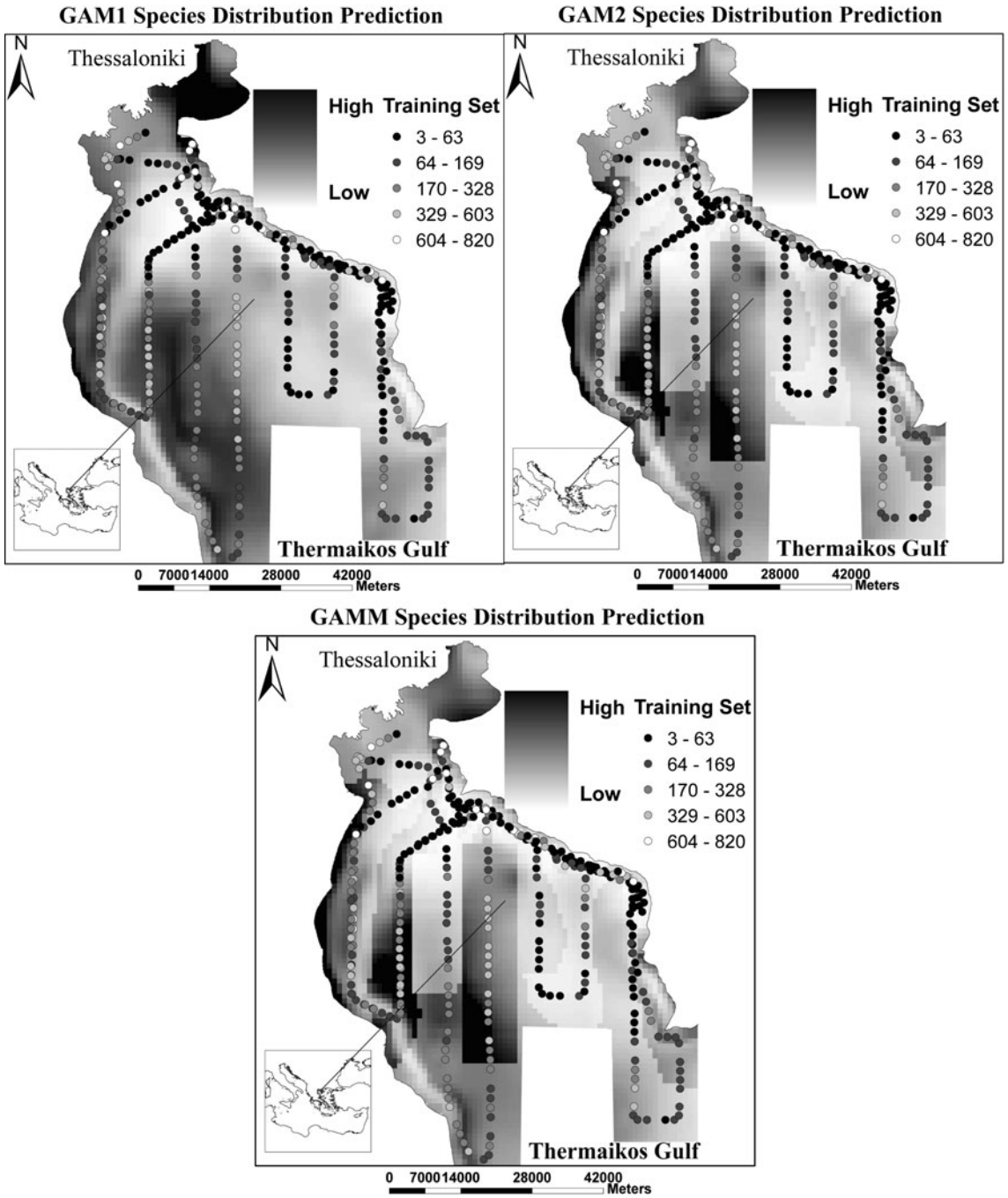

Fig. 9 Predicted grids from GAMs and GAMM that represent acoustic fish density (*black* high values, *white* low values). *Circles* represent the sampled acoustic fish density s_A1 in an opposite colour-scale than grids (*black circles* low values, *white circles* high values). *White dots* on black grid indicates accurate prediction and vice versa

trend is depicted in both grids, especially in coastal areas, but it is not very clear in the centre of the sampling area probably because of the smoother gradient of predictions.

The grid derived by AverMod prediction is shown in Fig. 12. The model calibration process, the error analysis and the correlations provide evidence that this model is characterized by a better predictive

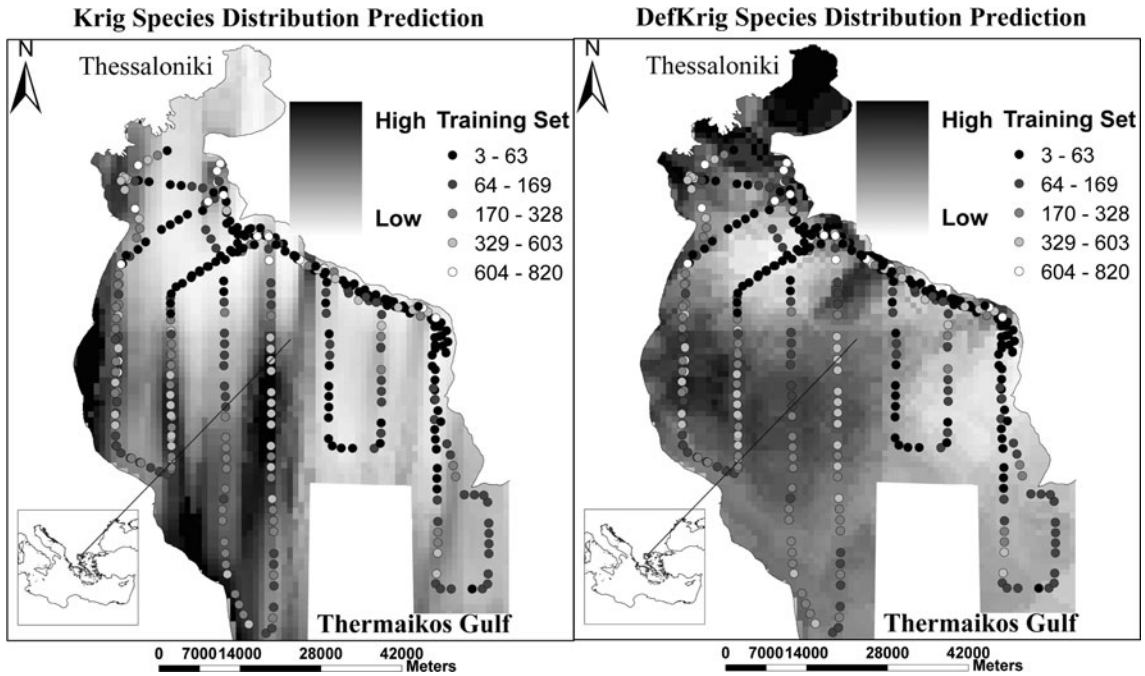

Fig. 10 Predicted grids from Krig and DefKrig that represent acoustic fish density (*black* high values, *white* low values). *Circles* represent the sampled acoustic fish density s_A1 in an opposite colour-scale than grids (*black circles* low values, *white circles* high values). *White dots* on black grid indicates accurate prediction and vice versa

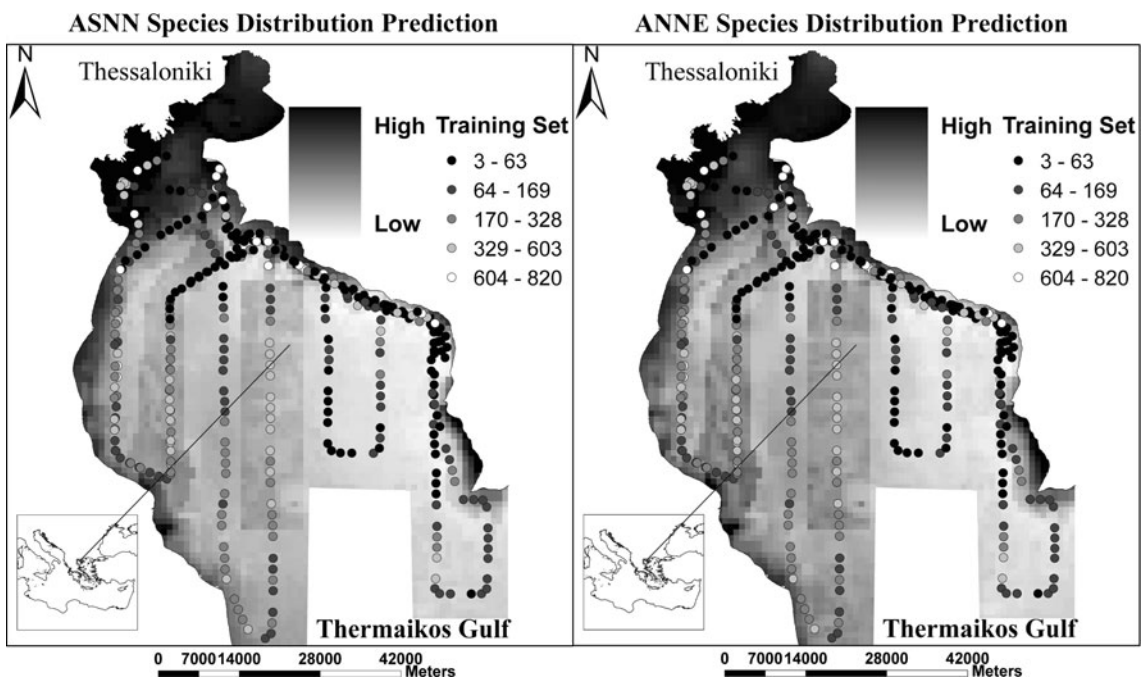

Fig. 11 Predicted grids from ASNN and ANNE that represent acoustic fish density (*black* high values, *white* low values). *Circles* represent the sampled acoustic fish density s_A1 in an opposite colour-scale than grids (*black circles* low values, *white circles* high values). *White dots* on black grid indicates accurate prediction and vice versa

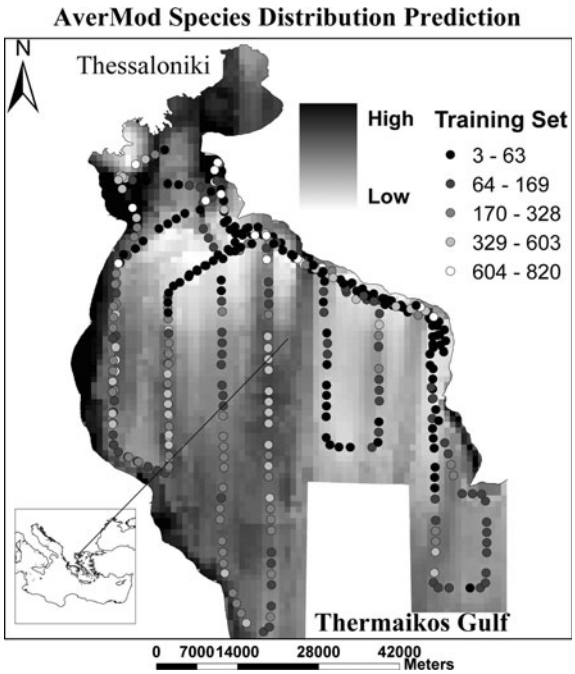

Fig. 12 Predicted grid from AverMod that represents acoustic fish density (*black* high values, *white* low values). *Circles* represent the sampled acoustic fish density s_A1 in an opposite colour-scale than grids (*black circles* low values, *white circles* high values). *White dots* on black grid indicates accurate prediction and vice versa

capacity than the other techniques. This grid combines the generality of the ASNN with the local nature of Krig in an additive way, which results in a more efficient potential density distribution. The global trend of the data is preserved while several patches across the study area indicate heterogeneity in species distribution.

Discussion

Regression models

The comparison of GAMs showed that GAM2 performs better than GAM1 and GAMM1 in all aspects. During the selection process, GAM2 showed the best quality characteristics (i.e. lowest AIC, highest deviance explained) and this is reflected in the predictive capacity of GAM2 for all three data sets used for the predictions. Thus GAM2 was the most appropriate model amongst GAMs in fitting the data, predicting an independent dataset and predicting a dataset that covered a larger area than the one used for model training. GAMM1 performs almost equally to GAM2 but the inclusion of autocorrelation did not succeed in improving the predictive ability of the mixed model. GAM1 was formulated to be the simpler model to preserve generality but predictions on the independent set did not meet the original expectations.

Generality, reality and precision are the features that group modelling techniques and only two out of the three can be achieved by a model each time (Levins, 1966). Although GAM1 is a more general model and could be used in a wide range of spatial and temporal predictions, GAM2 is the model that describes the variance of the acoustic data more accurately and precisely. On the other hand, GAMM1 is the only model that deals with spatial autocorrelation issues (Keitt et al., 2002), which can cause bias when modelling acoustic data, even if it does not perform equally to GAM2 in respect of the fitting efficiency and predictive capacity.

In general, GAMs and GAMMs are able to identify specific relationships between the response and the explanatory variables. In this case, high acoustic backscattering is related to areas with high or low values of SLA and either a combination of low DEP in a wide range of PAR or a combination of high DEP and high PAR. According to SLA, which is affected by winds and surface currents in the study area, the above conditions are met along the west and east coasts of the study area. These areas are generally nutrient-rich due to upwelling and river outflows, maintaining high concentrations of small pelagic species. The same areas are pinpointed by the interaction between effects of PAR and DEP. Additional areas in the open sea of the study area, where gyres and other oceanographic features are generated, are indicated by the latter interaction. These features are related to the life history of small pelagic species (Valavanis et al., 2005).

Kriging

The kriging models used in the present study were significantly different. DefKrig is error minimizing oriented, performing much better in the calibration process and error analysis. The predicted grid, though, provides evidence that it is a relatively conservative model. On the other hand, Krig, which

is the only model that preserves the variance and the StDev of the training set in its predictions, generates a more realistic potential distribution map, including important distribution heterogeneity but it tends to over-fit the training dataset.

Neural networks

ASNN and ANNE perform almost equally. However, ASNN is slightly better in fitting the observed data while ANNE provides better predictions in the independent dataset. In this study, contrary to the original expectation, ASNN's increased predictive ability did not overcome that of ANNE.

Model characteristics indicate that ASNN outperforms ANNE. A crucial feature of ASNN is the ability to interpret ANNE results by analysing the correlations between data cases in model space (Tetko, 2002b). This innovative approach provides a more complicated simulation of species–environment relationships since it is able to model local and global trends in the data (Tetko, 2002a). If new data become available, ASNN further improves its predictive ability and provides a reasonable approximation of the unknown function without the need to retrain the neural network (Tetko, 2002a). In this study, since there is no evidence of over-estimation or under-estimation of the training set, ASNN could be characterized by generality while conserving the ability to identify local variation in data. This is a result of the combination of an artificial feed-forward neural network, which is a memory-less approach, and *k*-nearest neighbours and the Parzen-window regression that represent the memory-based approaches (Tetko, 2002a). The neural networks can be considered as global models while the other two approaches are usually regarded as local models (Lawrence et al., 1996).

ANNs are generally characterized as a 'black box' approach. The output of ASNN, and the use of approaches like evaluation strip proposed by Elith et al. (2006), which enable the inspection of responses and the effects on the explanatory variables make ASNN more informative and less of a 'black box' approach.

Prediction optimization

Model optimization was performed by combining the predictions of the above techniques in an additive way, by training a new GAM. AverMod is the best calibrated model with increased similarities to the observed data without over-fitting the data and with the greatest predictive capacity. Since, each modelling approach aims to explain a portion of observed data's variation, the combination of two models could provide a model with increased ability to explain data's deviance. In addition, AverMod indicates that the optimal model provides an analogous increase in the efficacy of predictions. The potential species distribution on the predicted grids was justified by other published approaches that include oceanographic processes that are related to species occurrence and species habitat preferences derived from their life-history characteristics.

Elith & Leathwick (2009) stated that the realized species distribution is placed in both environmental and geographic space. In this study the Krig model explains the variability of the sampling data by utilizing exclusively geographic information. On the other hand, ASNN was developed by using mainly environmental data. As a result, the increased accuracy of the AverMod could partially be explained by the fact that it combines a 'geographic' model (Krig) with an 'environmental' model (ASNN) to generate a hybrid model that refers to a parameter space with both environmental and geographic dimensions.

Small pelagic species distribution, biology and comparison to other studies in the area

Small pelagic fish species are highly exploited species, characterized by large fluctuations in abundance and spatial distribution that mainly depend on fluctuations in environmental factors. Oceanographic features such as temperature fronts, eddies, rings and upwelling areas have been related to fish biomass concentration, at least at certain spatial scales (Laurs et al., 1984; Fiedler & Bernard, 1987; Chen et al., 2005). In general, warm, nutrient-depleted water has low chlorophyll-*a* content and cold, nutrient-rich water has high chlorophyll-*a* levels (Georgakarakos & Kitsiou, 2008). Some species, such as anchovy and sardine, which prefer phytoplankton-rich waters during some periods of their life cycle may exhibit a stronger linkage to chlorophyll-*a* concentrations (Ware & Thomson, 2005) than other species at higher trophic levels. European anchovy is also related to the influence of river outflow as shown in

the Bay of Biscay, the Adriatic Sea and the Bay of Tunis (Motos et al., 1996; Agostini & Bakun, 2002), in the Catalan Sea and the Gulf of Lions (Palomera et al., 2007; Sabates et al., 2007).

The grids predicted for Thermaikos Gulf identify two distinct areas where fish biomass is concentrated: the west coastline from north to south, which is characterized by the presence of riverine waters, and the central study area, which is related to gyre formation (Somarakis et al., 2002). Both areas could be characterized as nutrient-rich.

Several modelling approaches have been used in Thermaikos Gulf for small pelagic species. Georgakarakos & Kitsiou (2008) applied kriging and co-kriging methods on acoustic fish density data and the conclusions are in agreement with the findings of the present study, especially in the identification of the central part of the study area with high acoustic density. Schismenou et al. (2008) have characterized Thermaikos Gulf as a known fishing ground for anchovy, sardine and sardinela. However, the output of their study was in lower resolution and could not directly be compared to data presented here, since there is no distinguishable distribution fluctuation. This is the case in Tsagarakis et al. (2008) where discriminant function analysis (DFA) is applied on sardine. Despite the low resolution mapping, which is not suitable for habitat heterogeneity identification, the output of DFA is generally in agreement with the outputs of the present study. Giannoulaki et al. (2008) developed GAMs for habitat identification of anchovy for a different time frame to the one used in this study, and they resulted to similar fish distribution trend as the one that identified in this study by the most efficient models. This agreement confirms the long-term persistence of oceanographic features, such as gyres and upwelling that favour species concentration in specific areas.

Comparison and evaluation of the modelling techniques

One step towards improving evaluation of model performance in predicting species distributions is to use independent, well-structured data sets for validation (Elith et al., 2006). For presence data, the best model evaluation is achieved by withholding data (*k*-fold partitioning) for testing model predictions or by comparing RSF predictions using models developed for different periods and study areas (prospective sampling) (Boyce et al., 2002). According to Lehmann et al. (2002), cross-validation or bootstrapping is generally more practical because it creates relatively independent random data subsets and allows the use of all available data in the modelling process. By using entirely independent data sets, there is a risk of comparing different sampling strategies instead of evaluating a model (Lehmann et al., 2002). In this study, the verification set was derived from the initial sampling strategy. On the other hand, the use of such data as a verification set presents similarities to cross-validation techniques. The extensive exploration of the raw data resulted in the construction of a verification set that is unknown to models. The selected validation dataset contains measurements omitted at each step h units (h equals at least 5 nm), whilst h is chosen according to the empirical variogram of both validation set and prediction residuals. The empirical variograms revealed a low autocorrelation, even at distances below the h limit. A similar spatial structure has been encountered in previous surveys (October 1996, May 1997) even in different seasonal conditions (Georgakarakos & Kitsiou, 2008). Results from a comparative study using series of acoustic survey data from five different locations in Europe suggested that the spatial organization of the stock would be more dependent on environmental parameters than on fish abundance (Petitgas, 2001). Thus, DS3 represents an unknown set for the training process of modelling techniques. However, from a biological point of view it maintains the sampling information. The predictive performance of species potential distribution is presented for each separate dataset.

In general, neural networks and especially ASNN are more accurate than the other techniques in fitting the training dataset, while GAMs and especially GAM2 are more flexible in predicting the independent dataset. Kriging on the other hand, is a useful tool for species distribution predictions. However, the approach of predicting distribution by utilizing the spatial correlation of sampling might not be as strong as utilizing a large number of environmental variables that are biologically related to species distribution. In any case, better quality and resolution of sampling data would provide better results since all the above modelling approaches are data sensitive. Amongst biological data sets, acoustic density

abundance represents a challenging dataset for modelling studies since it is characterized by important variation. Most modelling techniques use smoothers for fitting the data (spline for GAMs, Parzen-window for ASNN) and, thus, they tend to under-predict high observed values and over-predict low values. As a result, the global error is minimized but the accuracy of predictions regarding local variation of acoustic data declines. The use of high resolution informative explanatory variables partially overcomes the above effect and reveals a more accurate variance for species distribution and thus the potential habitat heterogeneity.

ASNN and ANNE are suggested as the appropriate methods to model sampling data efficiently. Both methods represent the original data more realistically than the other techniques and they are able to identify local and global variation of the data without over-fitting the sampling/training dataset. These characteristics make ASNN and ANNE optimal approaches for presenting the realistic species distribution derived from sampling data and thus, for management purposes, marine protected areas designation, and sampling strategy selection. This study also indicates the predictive ability of GAMs, especially when predictions are required on an expanded spatial scale beyond the sampling area. Thus, GAMs could be used in studies that are oriented towards identifying potential species habitats at a larger spatial scale than in high resolution habitat heterogeneity identification. In addition, GAMs output are able to exhibit species–environment relationships that are easier to interpret than other techniques' outputs.

The overall results indicate that AverMod is the most accurate approach for predicting species distribution and thus could be characterized as the most suitable for the purposes mentioned above. The drawback of this approach is the requirement of developing several different models before combining their predictions. In addition, the combination of different model predictions should be always tested against the initial predictions since it is not a straightforward process. In this study, a GAM was used for combining different models' predictions in an additive way. Instead of GAMs, other techniques could also be used, e.g. training an ASNN by using several predictions as input. In any case, the optimization of species distribution prediction by the approach proposed in this study seems to be very promising.

Conclusions

GAMs are a very flexible approach to model fish density acoustic data and are able to identify specific relationships between the response and the explanatory variables. Kriging on the other hand, is a useful tool for species distribution predictions. However, the approach of predicting species distribution by utilizing the spatial correlation of sampling might not be as strong as utilizing a large number of environmental variables that are biologically related to species distribution. ASNN and ANNE are suggested as the appropriate methods to model sampling data efficiently. Both methods represent the original data more realistically than the other techniques and they are able to identify local and global variation of the data without over-fitting the sampling/training dataset. The prediction optimization approach over-performed the other methods, since it combines a 'geographic' model (kriging) with an 'environmental' model (ASNN) to generate a hybrid model that refers to a parameter space with both environmental and geographic dimensions.

References

Agostini, V. N. & A. Bakun, 2002. 'Ocean triads' in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (with example application to European anchovy, *Engraulis encrasicolus*). Fisheries Oceanography 11: 129–142.

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

Bishop, M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford.

Bodholt, H., H. Nes & H. Solli, 1989. A new echo sounder system. Proceedings of the Institute of Acoustics (UK) 11(3): 123–130.

Boyce, M. S., P. R. Vernier, S. E. Nielsen & F. K. A. Schmiegelow, 2002. Evaluating resource selection functions. Ecological Modelling 157: 281–300.

Chen, I. C., P. F. Lee & W. N. Tzeng, 2005. Distribution of albacore (*Thunnus alalunga*) in the Indian Ocean and its relation to environmental factors. Fisheries Oceanography 14: 71–80.

Cleveland, W. S., 1994. The Elements of Graphing Data. Hobart Press, Summit. ISBN:0-9634884-1-4.

Elith, J. & J. R. Leathwick, 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40: 677–697.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. C. Overton, A. T. Peterson, S. J. Phillips, K. S. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz & N. E. Zimmermann, 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151.

Fiedler, P. C. & H. J. Bernard, 1987. Tuna aggregation and feeding near fronts observed in satellite imagery. Continental Shelf Research 7: 871–881.

Georgakarakos, S. & D. Kitsiou, 2008. Mapping abundance distribution of small pelagic species applying hydroacoustics and co-kriging techniques. Hydrobiologia 612(1): 155–169.

Giannoulaki, M., A. Machias & N. Tsimenides, 1999. Ambient luminance and vertical migration of the sardine *Sardina pilchardus*. Marine Ecology Progress Series 178: 29–38.

Giannoulaki, M., V. D. Valavanis, A. Palialexis, K. Tsagarakis, A. Machias, S. Somarakis & C. Papaconstantinou, 2008. Modelling the presence of anchovy *Engraulis encrasicolus* in the Aegean Sea during early summer, based on satellite environmental data. Hydrobiologia 612(1): 225–240.

Guisan, A. & N. E. Zimmermann, 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147–186.

Guisan, A., J. Edwards, C. Thomas & T. Hastie, 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157: 89–100.

Hastie, T. & R. Tibshirani, 1990. Generalized Additive Models. Chapman & Hall, London.

Hastie, T., R. Tibshirani & J. Friedman, 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. Springer, Berlin.

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. Macmillan, New York.

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

Keitt, T. H., O. N. Bjornstad, P. M. Dixon & S. Citron-Pousty, 2002. Accounting for spatial pattern when modelling organism–environment interactions. Ecography 25: 616–625.

Kourafalou, V. & K. Tsiaras, 2007. A nested circulation model for the North Aegean Sea. Ocean Science 3: 1–16.

Laurs, R. M., P. C. Fiedler & D. R. Montgomery, 1984. Albacore tuna catch distributions relative to environmental features observed from satellites. Deep-Sea Research 31: 1085–1099.

Lawrence, S., A. C. Tsoi & A. D. Back, 1996. Function approximation with neural networks and local methods: bias, variance and smoothness. Australian Conference on Neural Networks. Australian National University: 16–21.

Lehmann, A., C. Overton & J. R. Leathwick, 2002. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling 157: 189–207.

Levins, R., 1966. The strategy of model building in population ecology. American Scientist 54: 421–431.

MacLennan, D. N., P. G. Fernandes & J. Dalen, 2002. A consistent approach to definitions and symbols in fisheries acoustics. ICES Journal of Marine Science 59: 365–369.

Matheron, G., 1971. The Theory of Regionalized Variables and its Applications. Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.

Michie, D., D. J. Spiegelhalter & C. Taylor, 1994. Machine Learning, Neural and Statistical Classification. Prentice Hall, Englewood Cliffs.

Moisen, G. G. & T. S. Frescino, 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling 157: 209–225.

Moran, P. A. P., 1950. Notes on continuous stochastic phenomena. Biometrika 37: 17–23.

Motos, L., A. Uriarte & V. Valéncia, 1996. The spawning environment of the Bay Biscay anchovy (*Engraulis encrasicolus* L.). Scientia Marina 60: 117–140.

Palialexis, A., S. Georgakarakos, K. Lika & V. D. Valavanis, 2009. Use of GIS, remote sensing and regression models for the identification and forecast of small pelagic fish distribution. Proceedings of the Second International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 09), June 21–26, Mykonos, Greece.

Palomera, I., M. P. Olivar, J. Salat, A. Sabates, M. Coll, A. Garcia & B. Morales-Nin, 2007. Small pelagic in the NW Mediterranean Sea: an ecological review. Progress in Oceanography 74: 377–396.

Pearce, J. & S. Ferrier, 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133: 225–245.

Petitgas, P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries 2: 231–249.

Potts, J. M. & J. Elith, 2006. Comparing species abundance models. Ecological Modelling 199: 153–163.

Poulos, S. E., G. T. Chronis, M. B. Collins & V. Lykousis, 2000. Thermaikos Gulf Coastal System, NW Aegean Sea: an overview of water/sediment fluxes in relation to air-land-ocean interactions and human activities. Journal of Marine Systems 25: 47–76.

R Development Core Team, 2005. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria [available on internet at http://www.Rproject.org].

Redfern, J. V., M. C. Ferguson, E. A. Becker, K. D. Hyrenbach, C. Good, J. Barlow, K. Kaschner, M. F. Baumgartner, K. A. Forney, L. T. Ballance, P. Fauchald, P. Halpin, T. Hamazaki, A. J. Pershing, S. S. Qian, A. Read, S. B. Reilly, L. Torres & F. Werner, 2006. Techniques for cetacean–habitat modeling: a review. Marine Ecology Progress Series 310: 271–295.

Richards, C. L., B. C. Carstens & L. Knowles, 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography 34: 1833–1845.

Ripley, B. D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.

Sabates, A., M. P. Olivar, J. Salat, I. Palomera & F. Alemany, 2007. Physical and biological processes controlling the distribution of fish larvae in the NW Mediterranean. Progress in Oceanography 74: 355–376.

Schismenou, E., M. Giannoulaki, V. D. Valavanis & S. Somarakis, 2008. Modeling and predicting potential

spawning habitat of anchovy (*Engraulis encrasicolus*) and round sardinella (*Sardinella aurita*) based on satellite environmental information. Hydrobiologia 612(1): 201–214.

Schröder, B., 2008. Challenges of species distribution modeling belowground. Journal of Plant Nutrition and Soil Science 171: 325–337.

Shepherd, A. J., 1997. Second-Order Methods for Neural Networks. Springer-Verlag, London: 145.

Somarakis, S., P. Drakopoulos & V. Filippou, 2002. Distribution and abundance of larval fishes in the northern Aegean Sea-Eastern Mediterranean—in relation to early summer oceanographic conditions. Journal of Plankton Research 24: 339–357.

Tetko, I. V., 2002a. Associative neural network. Neural Processing Letters 16: 187–199.

Tetko, I. V., 2002b. Neural network studies. 4. Introduction to associative neural networks. Journal of Chemical Information in Computer Science 42: 717–728.

Tetko, I. V. & V. Y. Tanchuk, 2002. Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. Journal of Chemical Information in Computer Science 42: 1136–1145.

Tetko, I. V., D. J. Livingstone & A. I. Luik, 1995. Neural network studies. 1. Comparison of overfitting and overtraining. Journal of Chemical Information in Computer Science 35: 826–833.

Tetko, I. V., I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches & A. Varnek, 2008. Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. Journal of Chemical Information and Modeling 48(9): 1733–1746.

Tsagarakis, K., A. Machias, S. Somarakis, M. Giannoulaki, A. Palialexis & V. D. Valavanis, 2008. Habitat discrimination of juvenile sardines in the Aegean Sea using remotely sensed environmental data. Hydrobiologia 612(1): 215–223.

Tsimenides, N., G. Bazigos, S. Georgakarakos & A. Kapantagakis, 1992. Distribution of acoustic pelagic fish populations in the northern Aegean Sea. Proceedings of the 1st World Fisheries Congress 5: 33–42.

Valavanis, V. D., 2002. Geographic Information Systems in Oceanography and Fisheries. Taylor & Francis, London: 240.

Valavanis, V. D., Kapantagakis, A., Katara, I., Palialexis, A. 2004. Critical regions: A GIS-based model of marine productivity hotspots. Aquatic Sciences 66(1): 139–148.

Valavanis, V. D., Katara, I., Palialexis, A. 2005. Marine GIS: Identification of mesoscale oceanic thermal fronts. International Journal of Geographical Information Science 19(10): 1131–1147.

Valavanis, V. D., G. J. Pierce, A. F. Zuur, A. Palialexis, A. Saveliev, I. Katara & J. Wang, 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. Hydrobiologia 612(1): 5–20.

Walline, P. D., 2007. Geostatistical simulations of eastern Bering Sea walleye pollock spatial distributions, to estimate sampling precision. ICES Journal of Marine Science 64: 559–569.

Ware, D. M. & R. E. Thomson, 2005. Bottom-up ecosystem trophic dynamics determine fish production in the Northeast Pacific. Science 308: 1280–1285.

Wood, S. N., 2006. Generalized Additive Models: An Introduction with R. CRC Press, London.

Wood, S. N. & N. H. Augustin, 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological Modelling 157: 157–177.

Zuur, A. F., E. N. Ieno & G. M. Smith, 2007. Analysing Ecological Data. Springer Series: Statistics for Biology and Health. Springer, New York.

Zuur, A. F., E. N. Ieno & C. S. Elphick, 2010. A protocol for data exploration to avoiding common statistical problems. Methods in Ecology and Evolution 1: 3–14.