Ecological Modelling
June 2021, Volume 449 Pages 109501 (12p.)
https://doi.org/10.1016/j.ecolmodel.2021.109501
https://archimer.ifremer.fr/doc/00688/80044/

Which spatial interpolators I should use? A case study applying to marine species

Rufino Marta [1, 2, 3, *], Albouy Camille [1], Brind'Amour Anik [1]

[1] IFREMER - Centre Atlantique, French Research Institute for Exploitation of the Sea, Département Ecologie et Modèles pour l'Halieutique (EMH), Rue de l'Ile d'Yeu - BP 21105, 44311 Nantes cedex 3, France
[2] Portuguese Institute for the Sea and the Atmosphere (IPMA), Division of Modelling and Management of Fisheries Resources, Av. Dr. Alfredo Magalhães Ramalho, 6, 1495-165 Lisboa, Portugal
[3] Centre of Statistics and its Applications (CEAUL), Faculty of Sciences, University of Lisbon, Portugal

* Corresponding author : Marta Rufino, email address : marta.rufino@ipma.pt

Abstract :

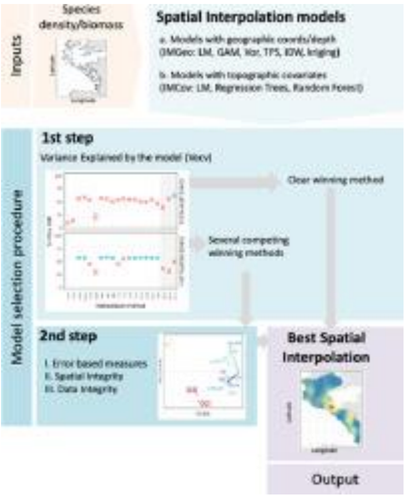
Species are spread in space, whereas sampling is sparse. Thus, to describe and map along environmental gradients, it is necessary to interpolate the species abundance. Considering the plethora of valid methods, the researcher gets easily puzzled to choose the most appropriate interpolation approach with reference to the ecological question being asked.

We propose a procedure to select among alternative spatial distribution models and we illustrate it with 175 marine species distributions (35 species * 5 years). In a first step, the distribution of the variance explained by the predictive model (VEcv) given by 10-fold cross validation is estimated for each interpolation method. When the inter-quartile range of the VEcv distribution of the different methods overlap, the selection passes to a second step, using 11 measures belonging to three criteria: 1) error based measures, 2) spatial equivalence measures (center of gravity, inertia, isotropy and index of aggregation) and 3) measures based on the data integrity after interpolation, for example the percentage of area over the maximum sampled data.

We applied our approach to marine species sampled using either stratified random survey (trawl) or systematic survey (acoustic). We found that 87% of all species distributions had overlapping VEcv and thus passed the first selection. In the second selection step, the best method varied with species and year, although general additive model (GAM), Thin Plate Spline (TPS), Universal Kriging (UKr) and Random Forest (Rfor) performed better for the trawl data and TPS, Ordinary Kriging (OKri) and UKr for the acoustic data. Further, the results differed within methods (e.g. kriging neighborhood and type of kriging) and small modifications on the specifications can have a large impact on the surfaces produced.

The proposed approach 1) is accessible and intuitive, and does not require any complex software or sophisticated methodology; 2) shows exactly in what aspects each interpolation model is prevalent over the others and permits to make a decision accordingly to the objectives of the study; 3) takes into account different criteria to evaluate each, properties of an interpolation method; 4) is universal and does not depend on the method used or the data characteristics. A detailed review on the subject is also included.

Graphical abstract


Inputs
Species density/biomass
Spatial Interpolation models
a. Models with geographic coords/depth (IMGeo: LM, GAM, Vor, TPS, IDW, kriging)
b. Models with topographic covariates (IMCov: LM, Regression Trees, Random Forest)
Model selection procedure
1st step
Variance Explained by the model (10cv)
Clear winning method
Several competing winning methods
2nd step
I. Error based measures
II. Spatial Integrity
III. Data Integrity
Best Spatial Interpolation
Output

Highlights

► A new method was developed to select among spatial distribution models using 2-steps. ► The 1st step uses the variance explained by the predictive model (10-fold cross validation). ► The 2nd step uses 3 criteria: error based measures, spatial equivalence measures and data integrity. ► The method is illustrated using 175 marine species distributions (35 species x 5. years). ► The approach is accessible, clear, multi-criteria and is universal as it does not depend on the method used or the data characteristics.

Keywords : Spatial interpolation, Geostatistics, Machine learning, Spatial distribution, SDM, Review

Introduction

Everything becomes clearer when it is presented in a figure. Maps of species distributions are required for numerous purposes, such as to visualize spatial variability, spot changes in the communities or provide estimations of variables of interest. Species distributions are mapped by interpolation, that is, predicting values at un-sampled areas using a modelling procedure applied to the sampled data. Ecological processes are inherent in species distributions (or spatial structure), such as those impacted by anthropogenic factors and climate change, and thus reflected in the respective maps. Coupled with species traits or phylogenetic information, species distribution maps can inform on the location of functional or phylogenetic hotspots.

Interpolated species distributions are widely used in a range of fields and applications, including regional biodiversity assessments, spatial conservation prioritization, evolutionary biology, epidemiology, global change biology and wildlife management (Araujo and Peterson, 2012). Given the importance of spatial interpolation, new and more powerful methods are developed on a regular basis and continuous/progressive evaluation of these statistical models is necessary (Austin, 2007).

There is a large body of literature on species distribution models (SDM, also known as bioclimatic envelope models, ecological niche models and habitat suitability models), that explore the relationship between geographical occurrences of species and corresponding environmental variables (Araújo and Guisan, 2006; Dormann et al., 2007; Elith and Leathwick, 2009; Guisan and Zimmermann, 2000; Hui et al., 2013; Olden and Jackson, 2002). These are however, more challenging to apply on marine data to model than on its terrestrial counterparts. Marine fisheries data compared to its terrestrial counterpart (Lecours et al., 2016), typically have fewer sampling stations, in face of the large costs associated with the survey operations and cover irregular survey shape areas (e.g. 3-15 m depth along the coast). Further, marine species distributions are typically characterized by a large percentage of zero observations and a huge variability. Finally, the availability of environmental covariates in the marine field is generally scarcer, both in terms of geographic span and resolution, also due to difficulties associated with sampling. Thus, in face of this particularities of marine data, the challenges arisen by producing spatial models using this type of data have been often referred (Olden and Jackson, 2002).

So, the next question that naturally arises is which method to choose? It is widely recognized that there are no magical recipes to determine the perfect interpolation model. Undoubtedly, this should be focused on the data *per se*, laying on the statistical theoretical background and respective assumptions (Bivand et al., 2013; Cressie, 1993; Li and Heap, 2008; Sluiter, 2009; Wackernagel, 1998; Webster and Oliver, 2007). Several methods pass this first selection, however we need a framework accounting for model selection and evaluation to help decision making. Such decision is a complex issue, central to ecological modelling, with huge implications (Naimi and Araújo, 2016). Model selection involves evaluation, validation, performance, accuracy, skill, efficiency or robustness. Nevertheless these concepts are hard to disentangle and have been used with different meanings on the literature (reviewed in Bellocchi et al., 2010).

Generally, interpolators are compared using error based measurements, that is predicted *vs.* observed values, preferably obtained from cross-validation or jackknife processes (Li, 2016; reviewed by Li and Heap, 2008, 2011; Richter et al., 2012; Stow et al., 2009; Willmott et al., 2015). Using primarily error-based criteria, comparisons among spatial interpolators have been done in the field of meteorology (Aalto et al., 2013), air quality (Hoffman, 2015), soil (Gasch et al., 2015; Hengl et al., 2004, 2015), marine sediment (Diesing et al., 2014; Lark et al., 2016; Li et al., 2011), bathymetry (Amante and Eakins, 2016) and environmental sciences (Li and Heap, 2011, 2008). Among these, Mean Absolute Error (MAE, bias or a measure of average error-magnitude) and Root Mean Squared Error (RMSE, for accuracy or over/under fitting) are the most commonly within the field of environmental sciences (Li and Heap, 2008; Richter et al., 2012; Willmott, 1982). However, as their magnitude depends on the scale/unit of the variable predicted, these are hardly comparable among variables or subjects. Further, most of these accuracy measures are algebraically related, being thus potentially redundant and collinear (Li, 2016; Li and Heap, 2011, 2008; Willmott et al., 2015). It has also been suggested the use of dimensionless measures, besides at least one error measure on the variable scale (modified coefficient of efficiency, Legates and McCabe Jr., 1999; coefficient of efficiency, Nash and Sutcliffe, 1970; index of aggreement, Willmott, 1981, 1982; modified index of aggreement, Willmott et al., 2012, 2015). Such approaches have been applied within the field of hydrology/climatology until recently, where Li (2016, 2017) revising Willmot's D, advocated its use as an universal tool to assess the accuracy of predictive models within environmental sciences, naming it Variance Explained by predictive models, estimated by cross-validation (VEcv), that is: how well a model is predicted, relative to the average of the observations (also called coefficient of efficiency, Nash and Sutcliffe, 1970 or G-value or goodness-of- prediction measure).

However, all these criteria often lead to overlapping results, that is several models having similar VEcv values, being difficult to select only one. Furthermore, these are based essentially on error measurements and do not take into account other fundamental aspects, such as spatial integrity (that is if the interpolation respects the spatial distribution of the data) or the spatial data limits of the interpolation relative to the original data. Additionally, whatever the criteria considered, it has long been demanded the establishment of a consistent and rationale set of procedures that should be used to compare spatial interpolation models (Fox 1981) (Willmott, 1982).

The objective of the current work, is therefore, to develop a simple and accessible protocol to compare the results given by different spatial interpolating methods, that integrates important aspects for mapping marine species distribution, namely not only error measures, but also spatial and data integrity after interpolation. The proposed protocol was applied to compare 20 interpolation methods applied to 35 species distributions from two typical fisheries surveys (trawl and acoustic), carried out during 5 years. The interpolation methods considered comprise approaches only using geographic coordinates and methods using depth and other topographic variables derived from bathymetry.

Materials and Methods

DATA

We considered two data sets for the species distributions case studies, one obtained from scientific groundfish bottom trawl surveys (EVHOE) and another from a scientific pelagic acoustic surveys (PELGAS).

The bottom trawl survey is carried out annually during Autumn in the North Atlantic ("Evaluation Halieutique de l'Ouest Européen, EVHOE cruise, RV Thalassa, IFREMER," n.d.). It ranges from the Bay of Biscay up to the Celtic sea, with a randomly stratified sampling strategy, comprising from 119 to 153 stations/year ("Evaluation Halieutique de l'Ouest Européen, EVHOE cruise, RV Thalassa, IFREMER," n.d.)(map and location of sampling stations can be found in see Supplement 1). The biomass of the 29 fish species occurring more than 10 times/year during the survey and excluding the main pelagic species was used (see Supplement 1 for the species list names and further details on the survey). For the purpose of this study we used data between 2011 and 2015 with an average of 198 sampling stations (number of hauls per year can be found in Supplement 1).

The pelagic survey (Doray, M., Duhamel E. , Huret M. , Petitgas P., 2002; Doray M., Badts V., Masse J., Duhamel E., Huret M., Doremus G., 2014) is an acoustic spring survey that aims at monitoring the Bay of Biscay pelagic ecosystem to inform fisheries and ecosystem management. Initially, PELGAS objective was to estimate biomass anchovy (*Engraulis encrasicolus*) and nowaday, the survey goals were extended to estimate the stocks of all the small pelagic fish species in the Bay of Biscay. From this survey we extracted the biomasses of six small pelagic fish species (see Supplement 1 for the species list and map), sampled over 1345 to 1997 locations obtained from 29 acoustic radials perpendicular to the coast between 2011 and 2015.

In the simplest cases, interpolation can be carried out using only the geographic coordinates as explanatory variables. However, in marine systems bathymetry influences species spatial distribution and this information is available at global scale, and thus can be added to improve interpolation models. Bathymetry data was extracted from GEBCO data base and validated with depth data obtained during the EVHOE surveys (IMGeo). Additionally, other covariates derived from bathymetry can be included to improve interpolation models. Those additional covariates were added without referring to any specific ecological hypotheses, but likely serving as proxy of other unmeasured environmental variables. We extrapolated eleven covariates (IMCov) such as derived from bathymetry/elevation such as (Lecours et al., 2016; Wilson et al., 2007): slope, aspect, northerness, easterness, rough, profile curvature (surf.curv), bathymetric position index (TPI), terrain ruggedness index (TRI), surface flow (flowdir), local Moran I (moran) and distance to the nearest coast (dist.coast)(details of each variable and respective maps can be found in Supplement 2). Thus, all models using covariates (IMCov) were produced including all variables, whereas the final model, was produced with an automatic selection of these variables for each distribution.

INTERPOLATION METHODS

Seven families of methods were applied to both case studies, aiming not be exhaustive: linear models (LM), general additive models (GAM), Inverse Distance Weighting (IDW), Thin Plate Spline (TPS), VORonoi triangulation (VOR), Kriging (Kr) and stochastic Conditional Simulation (CSim) for the methods using just geographic coordinates and eventually depth (dep)(IMGeo) and three families of methods using the 11 bottom topographical variables (IMCov): multiple regression (GLM), Regression Tree (RTre) and Random Forest (RFor) (Hengl et al., 2015, 2007, 2004, 2003; Li et al., 2011)(Supplement 3). Additionally, we used several alternatives on some methods, intended to quantify the within-model and between-model variability (Araújo and Guisan, 2006), namely considering only geographic variables as covariates or adding depth as well (GAMl/ GAMd, MKri/ UKri), changing kriging neighborhood (OK03/ OK05/ OK07/ OK10/ OK20/ OK30 or the fitting procedure as automatic vs. manual (OKri/MKri)). We provided a brief description of the methods used in Supplement 3 whereas additional details can be found in the vast literature (Bivand et al., 2013; Cressie, 1993; Fortin and Dale, 2005) and more precisely in two reviews on the subject (Li and Heap, 2008; Sluiter, 2009).

CRITERIA FOR COMPARISON OF INTERPOLATORS

Three complementary criteria were used to compare and evaluate the accuracy of interpolation models: (i) error based measures, (ii) changes in the spatial structure due to interpolation and (iii) data integrity after interpolation.

1. Error based measures

Error indices were estimated using predicted and observed values obtained by ten-fold cross-validation (10-fold CV). Ten-fold cross-validation was done by randomly splitting the data into 10 parts. We estimated the model using 9 of those 10 data set, whereas the observed values (10th split) are predicted using the model estimated. This process is repeated for each of the ten splits, obtaining predicted and observed values for the ten folds, which are then used to estimate the error measures. We performed a 10-fold CV, and instead of leave one out procedure, as this method has been considered to give too optimistic measures of error. The process of 10-fold CV was then repeated with a random split 100 times to obtain a distribution of the error indices. From predicted and observed values obtained by 10-fold, three measures were calculated: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) and the Variance explained by predicted models estimated using cross validation procedures (VEcv). For comparability with previous works, the MAE ([0, ∞], the lower the better) and Root Mean Squared Error (RMSE, [0, ∞], the lower the better) were estimated (Richter et al., 2012).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(|P_i - O_i|)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}$$

VEcv a dimensionless measure, varies between 100 for an excellent model and -∞. VEcv lower than 0, indicates that the model is worse than the average and we rounded these values to -1 for visualization purposes.

$$VEcv = 100\ \times (1 - \frac{\sum(P_i - O_i)^2}{\sum(O_i - \bar{O}\ \ )^2})$$

To perform the first step of our selection procedure (Fig. 1), we estimated the average RMSE, MAE and VEcv from the distributions obtained by the 10-fold CV, along with the upper and lower quantiles of the VEcv (probability of 0.25 and 0.75). We classified the average VEcv following Li (2016), into: 1) very poor if VEcv ≤ 10%; 2), poor if 10 < VEcv ≤ 30%; 3), average 30 < VEcv ≤ 50%; 4), good if 50 < VEcv ≤ 80%; 5) and excellent if VEcv > 80%.

To harmonize the interpolation with the other interpolators considered in this study (MAE, RMSE), we calculated an inversion of the VEcv (VEcv.inv = abs(VEcv/100-1)*100), where the lowest value, indicates the worst the model.

2. Spatial integrity

Ideally, an interpolation method should preserve the geometrical properties of the data, and therefore its spatial structure. We evaluated changes in the data spatial structure due to interpolation using deviations on four spatial indicators, relative to its results on the sampled data. Four spatial indicators were considered in the current work, following the revision and work of (Rufino et al., 2019). The center of gravity (CG) indicates the mean spatial location of the population (Bez and Rivoirard, 2001; Woillez et al., 2009b). The Euclidean distance between the CG estimated using the sampled data (without considering different areas of influence for each sample) and the interpolated data was calculated. It was used as a measure of the impact of interpolation on the geometric center of the distribution. The inertia represents the spatial dispersion of the population around its CG, i.e. the mean square distance between individual fish and the CG (Bez and Rivoirard, 2001; Woillez et al., 2009b). Isotropy/anisotropy (isotropy) represents the dispersion shape of the inertia around the CG (i.e. round or ellipsoid), and it is simply the ratio between the two inertia axes (Woillez et al., 2009a).

The Gini index quantifies the distribution's aggregation or concentration, and represents twice the area between the identity function and the Lorenz curve. It is bounded between 1 and 0, and the highest its value the most concentrated is the biomass in fewest samples.

To quantify geometric changes in the spatial distribution due to interpolation, the difference between the spatial indicators calculated for the sampled data and the indicators calculated in from the interpolated surfaces was estimated Thus, the absolute difference between inertia (log transformed for scaling), isotropy and Gini index of the interpolated surfaces and the respective ones estimated on the sampled data ($\Delta I = | I_{sample} - I_{interpolation|}$) were used as a measure of the interpolation method spatial integrity. Therefore, the higher the value of the difference on these spatial measures, the higher the shift relative to the sampled data's spatial structure cause by the interpolation method, thus the worse its performance.

3. Data limits integrity

Ideally, an interpolation method should preserve the samples data integrity and not predict values outside the data range or shifts in the mean biomass. To evaluate interpolator's data integrity, we used four measures estimated on each interpolated surface:

1) a.pix_under: relative area below/above minimum sampled biomass (pix_under = $abs[(\#pixels[max(B)_{interpolation} < max(B)_{sample}]/ \#pixels_{interpolation}]\times 100)$;

2) a.pix_over: relative area above/bellow maximum sampled prediction (pix_over) ($abspix\_over = [(\#pixels[max(B)_{interpolation} > max(B)_{sample}]/ \#pixels_{interpolation}]\times 100)$;

3) a.mean_perc: relative change on the predicted mean biomass ($mean\_perc = abs[(\mu_{interpolation} - \mu_{sample})/\mu_{sample}]\times 100)$;

4) a.over_perc: relative change on the maximum prediction ($over\_prec = abs[(\mu_{interpolation} - \mu_{sample})/\mu_{sample}]\times 100)$;

Where $B_{interpolation}$ is the interpolated biomass, B_{sample} is the samped biomass and #pixels, the number of pixels. As most of these measures were dimensionless, they can be compared among studies and methods and interpret as follow: the higher the value, the wose the performance of the interpolator (for comparability with the other indicators).

The eleven criteria were scaled such as the highest the value, the greatest the impact of the interpolation on the spatial structure and data integrity or greater the errors, bias or accuracy. Additionally, these criteria were not strongly correlated between each other, except for the RMSE/MAE (which is expected algebraically), a.mean_perc with a.over_perc (r=-63%) and CGdist with a.inertia (r = 58%). Further, these criteria were also not strongly correlated with biomass, which is another desirable property. The correlation between indicators is shown in supplement 4.

PROTOCOL TO COMPARE THE INTERPOLATORS

The 2-steps protocol proposed to compare spatial interpolators is summarized in Fig. 1.
First step:
a) Estimate the distribution of VEcv obtained by 10-fold CV. Select the interpolation method with the highest average VEcv;
b) Calculate the interquartile range (IQR) of each VEcv distribution (Q1, 25% and Q3, 75%);
c) Select all interpolation methods where the IQR overlaps with the interpolation method that showed highest VEcv;
d) If no other interpolation method IQR overlaps the highest VEcv IQR, the decision is reached, and the best method is clear. Otherwise, we continue to the second step.
Second step:

a) Calculate all measures proposed above for each interpolation method: 3 error measures (VEcv, MAE and RMSE), 4 spatial indicators (distance of the center of gravity, difference in inertia, isotropy and Gini index) and the 4 data integrity measures;
b) Do a principal component analysis (PCA), scaled and not centered on the indicators matrix for the distribution (species-year), that render the criteria ranges comparable, and that integrates all measures vs. all methods being evaluated;
c) The inverse of the distance between the methods loadings on the first two PCA axes relative to the center of the PCA, is then used to rank the interpolation methods; The PCA further shows concretely what aspects of the sampled data, the interpolation model is not respecting;

This protocol was applied to every species/year distribution, for all interpolation methods being assessed as a case study. All analyses were carried out using r-project. An R-script with a small simplified example is added in the supplement 5.

Results

FIRST SELECTION STEP: VECV CRITERIA

Twenty two percent of the models performed very poorly, producing interpolated surfaces that were worse than the mean, as shown by the negative average of the variance explained by the predictive model (VEcv < 0 in 784 models; Fig. 2). The maximum variance explained by the predictive model of the interpolations (VEcv) was 86% (EVHOE.TRISESM.2013-Kr3 -

Trisopterus esmarkii in 2013). Three species/years showed negative VEcv for all models, and thus were eliminated from further analysis (CONGCON.2011, 2012 - *Conger conger* and LOPHPIS.2014 - *Lophius piscatorius* from 2014).

The percentage of 'bad' models was higher for the bottom trawl survey data (EVHOE; 25%) than for the pelagic survey (PELGAS; 11% Fig. 2 and Fig. 3). For the bottom trawl survey, VEcv was slightly higher for GAM, TPS, geostatistical models (Kri) and RFor. As for the pelagic surveys, IDW, TPS, geostatistical models (Kri) and RFor reach better results (Fig. 2 and Fig. 3). However, overall good VEcv were lower than 50%, except for the acoustic surveys. Only few species distributions sampled with the acoustic surveys attained 'excellent' VEcv classes, whereas most distributions were classified as good for IDW, TPS, Kriging and RForest (Fig. 2). For the bottom trawl data, most models were classified as average or poor according to the VEcv criteria (Fig. 2). LM1 and LM2 were systematically worse than the mean.

Overall, the use of bottom covariates did not improve the interpolation's VEcv (Fig. 3). The variance explained by the predictive model (VEcv) varied more among species than between years (Fig. 4 and Fig. 5). However, for some benthic species the models using covariates showed higher VEcv than models without covariates (e.g. CONGCON - *Conger conger*, PHYBLE- *Phycis blennoides*, SOLESOL - *Solea solea*, HELIDAC - *Helicolenus dactylopterus*, SCYOCAN - *Scyliorhinus canicula* and LEPIWHI - *Lepidorhombus whiffiagonis*) (Fig. 4). Overall best results were obtained by species sampled in the pelagic survey, *Trisopterus esmarkii* (TRISESM) and *Merlangius merlangus* (MERNMER) were exceptions to this, attaining also higher VEcv in bottom trawl survey. Inter-annual variability of VEcv varied across species, with species (e.g. *Sardina pilchardus* - SARDPIL or *Trisopterus minutus* - TRISMIN) showing very little change in the results among the years, whereas others such as *Trisopterus luscus* (TRISLUS), showing a more variable results, although overall the patterns observed in relation to each method, across the years were relatively stable, i.e. most of the years within species showed a similar results (Fig. 5). Kriging based methods showed very similar VEcv between each other, independently of the neighborhood considered (between 3 and 30 points, i.e. OK03-OK30), the fitting of the variogram manually (MKri) or automatically (OKri), using depth as covariate (UKri) or even with conditional simulation (CSim) (Fig. 5).

In 13% of the distributions the method with the highest VEcv showed no-overlapping of the IQR with all remaining methods (23 cases representing 15 species, out of the 172 sp/year distributions)(Fig. 5). Within those, in 10 cases, RFor was the best method, 6 cases it was GAMd, 3 cases UKr, 2 cases IDW and 1 case TPS and MKr. *Sprattus sprattus* (SPRASPR) was the only species from the pelagic survey with one distribution showing a clear winning method on the first step. In none of the case studies all years for one species showed only one best method. Thus, in the remaining 149 case studies a second selection step was required.

SECOND SELECTION STEP: MULTIPLE CRITERIA

The best interpolation method according to each criteria varied widely across species-years, confirming that different aspects of the distributions are taken into account by each measure (Fig. 6; note that cases with ties were omitted). For example, RFor was the best method in terms of a.pix_under, a.over.perc, CGdist, RMSE and MAE whereas UKri was the best according to a.mean_perc and GAMd produced the maps with less deviations on the isotropy and with more similar aggregation (Gini index, Fig. 6). It is also clear the contrast between the results given by different kriging neighborhood in the best method by criteria (Fig. 6).

For a decision framework on the second step, all measures from each distribution were integrated using a principal component analysis, as illustrated for four species/years in Fig. 7.

The variance explained by PC1 was always above 90%, although PC2 also proportionated important information in discriminating the issues of the different interpolators relatively to the measures considered. In the given examples, for *Argentina* sp. From 2014 (ARGENT.2014), GAM and TPS methods showed highest deviations in terms of pixels under minimum whereas kriging and Rfor showed highest spatial distortion, although integrating all criteria the best method would be Kr30. Similar interpretation can be done for every species/year distribution, and thus conclude that *Callionymus lyra* in 2012 (CLAMLYR.2012) best method would be GAM, for *Conger conger* in 2015 (CONGCON.2015) would be Kr30 whereas for *Gadiculus argenteus* in 2013 (GADIARG.2013) would be Kr7.

The results of the protocol on the interpolator selection procedure, given the two steps together are found in Fig. 8. As already mentioned, few case studies were resolved on the first step (grey boxes). The use of the multi-criteria privileged GAM, TPS, UKr and Rfor for the trawl data set and TPS, OKri and UKr for the acoustic data set. It is also evident that the results differed within methods (e.g. kriging neighborhood and type of kriging) showing that small modifications on the methods can have a large impact on the surfaces produced. Further, it is interesting to observe such a large disparity on the best method, not only between species but also across years for the same species.

Discussion

In the current work we develop a two-step procedure to aid researchers select the best interpolation method for their data. The method uses a multi-criteria approach, that considers error-based measures, changes in the spatial structure and data integrity after interpolation and permits to determine in which particular aspect the interpolation is failing. The two-step procedure was illustrated by comparing 20 interpolation methods applied to 175 distributions, i.e. 35 species obtained during five years (2011-2015) of a typical bottom trawl survey and a pelagic acoustic survey. In the first step of the selection procedure, all interpolation methods within the highest VEcv's interquartile ranges are selected. In 13% of the case-studies no other method had overlapping VEcv and thus, the selection is complete without having to go through a second step. However, in the remaining 87% of the cases multiple methods were within the best VEcv interquartile range and thus a second step was proposed, using additional criteria. The effect of the interpolation was then integrated using PCA from which an index was extracted to rank the quality of the interpolations.

In the current work we aimed to use the simplest and most available approach to summarise the indicators in the second step, and this is why the PCA was selected. However, other multivariate statistical methods besides PCA could be used as an alternative in future works. Furthermore, it is possible to use just the first component (instead of two as in the current work) or to use the suggested indicators *per se*. Further work is needed to develop this particular aspect.

OTHER SELECTION PROCEDURES

Hengl *et al.* (2013) considered that the selection of a mapping procedure should account for accuracy (considered to be measured by RMSE), bias (considered to be measured by MAE), robustness (model sensitivity — in how many situations would the algorithm completely fail / how much artifacts does it produces?), reliability (how good is the model in estimating the prediction error, i.e. how accurate is the prediction variance considering the true mapping accuracy?) and computation burden (the time needed to complete predictions). Visual examination has been considered as equally important as accuracy measurements (Li et al., 2011), although it is largely subjective and not explicitly defined, consistent or repeatable (Stow et al., 2009). Model selection has involved evaluation, validation, performance, accuracy, skill, efficiency or robustness, although these concepts are hard to disentangle and have been used with different meanings on the literature (reviewed in Bellocchi et al., 2010). There is a clear absence of a unifying selection procedure for interpolation models, although the most recent works advocate the use of VEcv as a universal tool (Li, 2016, 2017). However, in 87% of the distributions studied in the current work such approach was not sufficient to discriminate among the interpolation methods, in view of the strong overlap between the respective distributions. Further, the measures used for model selection should be explicit, with a straightforward meaning and if possible, integrate multiple desirable properties of the interpolation procedure. Similar to the current study, other authors have suggested that method selection should be multi-criteria (Stow et al., 2009) and that the use of a single error measure may lead to incorrect interpretation (Hoffman, 2015) .

RMSE AND MAE

Each different measure comes with advantages and drawbacks. RMSE provides a measure of error size, but it is sensitive to outliers as it places a lot of weight on large errors (Hernandez-Stefanoni and Ponce-Hernandez, 2006). However, MAE and RMSE are among the best overall measures of model performance as they summarize the mean difference in the units of observed and predicted values (Willmott, 1982), although being highly correlated between each other, with biomass/occurrence and algebraically related. Variance explained by predictive model (VEcv) has been recently considered as the best error based criteria to evaluate interpolators (Li, 2017, 2016). Our results indicate that this measure is straightforward to interpret and quick evaluate thus it was included in the procedure, but for comparison with previous works, RMSE and MAE were also incorporated. .

CROSS VALIDATION

The error measures used to evaluate interpolation methods are traditionally estimated by cross validation procedures, either leave one out (LOO)(Kilibarda et al., 2014) or k-fold cross-validation (generally five or ten)(Davis, 1987; Li et al., 2011)(for a schematic overview of the re-sampling strategies for model validation see Richter et al., 2012). First of all we decided to not perform LOO cross-validation because in the case of skewness distributions and extreme values of the input data, this kind of cross-validation might produce strange outputs (Hengl, 2009). Secondly in highly clustered spatial distribution (like the species distributions considered), k (number of subsets, typically 5 or 10) should be large enough so that the data into the k-subsets contains enough information on the whole model domain and the spatial structure (Augustin et al., 2013). In our study, after a preliminary test, it was concluded that 10-fold was a good compromise for marine species distribution (not reported for brevity). Additionally, as the results of cross-validation strongly depended on the way the data is split (folds), the process should be randomly repeated several times, and as consequence it is obtained a distribution of the measure, which can then be used to compare the model's VEcv (like it was done in the first step of the procedure). It is important to note also that cross-validation is not necessarily independent, indeed, points used for cross-validation are subset of the original sampling design. Consequently, if the original design is biased and/or non-representative, then also the cross-validation might not reveal the true accuracy of a technique (Hengl, 2009). Further, error-based measures estimated by cross-validation results can be corrupted for clustered data sets on interpolator comparison (Hengl et al., 2013), highlighting the importance of having additional criteria in the selection procedure when the decision is not evident.

Cross validation can also be used to define the spatial model (Fortin and Dale, 2005; Gaetan et al., 2010; Wackernagel, 1998) and kriging neighborhood (Paramo and Roa, 2003) within the geostatistical methods. This is particularly crucial, because the effectiveness of the kriging depends on how well the selected model fits the data (Fortin and Dale, 2005). There were only small changes in the variance explained by the predictive model observed between kriging computed with different neighborhood or to the process of defining the spatial model (manual vs. automatic), when compared to other methods, but large changes were observed when the other comparison criteria were included. For example, Gini index or the percentage of over predictions changed widely with kriging neighborhood. Such comparison can also provide an idea of the variation due to the parametrization between and within each different techniques, as recommended in Araújo & Guisan (Araújo and Guisan, 2006)(within-model vs. between-model comparisons). Thus, the proposed protocol can also be applied as a tool to improve model specification by within model comparison, in future works. Similarly, the effect of a more detailed parametrization of the other interpolation methods on the quality of the predictions cannot be ignored. The method developed can also be used for such parametrization, as it was explored in the current work for kriging.

SPATIAL AND DATA INTEGRITY

Spatial indicators have been develop with the aim of quantifying distribution's spatial patterns (Bez and Rivoirard, 2001; Woillez et al., 2009b, 2007), but have also been applied as a model validation tool, to compare the model's outputs with sampled data for example (e.g. Huret et al., 2010)(Rufino et al., 2018). Additionally, these metrics are particularly sensitive to interpolation (Rufino et al., 2019) and are well suited to assess the spatial integrity of the sample data, after interpolation. The four metrics selected in our study highlight shifts in the main spatial features of the distributions, namely its location (center of gravity), dispersion (inertia), direction (isotropy) and aggregation (Gini index). It is expected (similar to what is done for modelling procedures) that a better interpolator would cause a minimum effect on those spatial aspects, when compared with the sampled data, therefore preserving the data spatial integrity. Future works can assess the use of other spatial indicators, such as the index of collocation for example, that may potentially be interesting with this aim.

Similarly, a good interpolation method should preserve the data limits of the sampled data. This is often done within works of spatial analysis, but rarely mentioned and hardly quantified. All those measures used in the selection procedure were made relative to the area and in the same direction (i.e. the larger the value, the worst the mode) for comparability purposes.

These aspects together are of outmost importance for species distributions maps, and have never been systematized previously. It can be argued that some of these measures just report the intrinsic properties of the interpolation methods and thus could be inferred solely on theoretical grounds. For example, kriging methods tend to under-estimate the maximum biomass (Bivand et al., 2013; Cressie, 1993). The proposed measures evidence those properties, and make them accessible without requiring a strong expertise on spatial analyses. On the other way, some measures, for example those with reference to the spatial integrity, are much less evident to describe theoretically.

PURELY GEOGRAPHIC METHODS *VS*. METHODS USING COVARIATES

Purely geographic methods, i.e. using only geographic coordinates to produce spatial predictions (i.e. ordinary kriging, TPS, IDW, etc.) are essential for the cases where there is a belief that geographic processes are dominant over environmental ones or in the absence of adequate environmental predictors (Elith and Leathwick, 2009). In the majority of cases the purpose of the statistical modelling is the prediction of species distribution, whereas the relationships between species and the environment tend to be a secondary consideration (Austin, 2002; Guisan and Zimmermann, 2000). This is also the focus of the current work and probably the commonest situation in marine studies or fisheries management. Thus, the applied models with covariates used only topographic variables directly derivable from depth, which is widely available. Unlike on its terrestrial counterpart, the effect of bottom topography on fish distribution is seldom tackled in fisheries ecology (Giannoulaki et al., 2006, 2003). These features are further advantageous for being relatively stable through time on these areas (Maravelias, 1999)(unlike other environmental characteristics) thus being potentially interesting also for long term studies, where other variables are not available.

It can be expected that topographic information of the sea bottom is more important for demersal species than for the pelagic ones. However, sea bottom topography features are known to be determinant for small pelagics species (Giannoulaki et al., 2006; Maravelias, 1999) and some of these species occur also near the sea bottom (e.g. *Scomber* sp. and *Trachurus trachurus*). However, the models done with topographic variables were not better than those using just geographic variables for any pelagic species, but it is imperative to see the marine environment as a continuous system where all aspects are connnected, and thereby only manageable through an ecosystem approach (Cotter et al., 2009; Doray et al., 2018). Other relevant environmental variables such as temperature and productivity, would improve the models with covariates, but can be more species specific. When a set of ecological covariates is available, whatever these are, the current method is also applicable.

SPATIAL AUTOCORRELATION

The notion of spatial autocorrelation is largely attributed to Tobler's 1st Law of Geography, “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). Spatial autocorrelation is widely present in marine species species distribution and is an essential aspect to acount for in spatial prediction (Dormann et al., 2007; Elith and Leathwick, 2009; Legendre, 1993). We verified using a large empirical data set that models accounting for spatial autocorrelation (i.e. geostatitical models) showed higher VEcv overall both for the bottom trawl data and for the acoustic pelagic survey. This difference was more pronounced on the acoustic pelagic data, where the number of samples is also much higher and the spatial models of the variogram, better defined (pers. obs. MMR). Improving the spatial model definition increases the effectiveness of kriging (Fortin and Dale, 2005). This was also observed in Rufino *et al* (2006) using simulated data, where the precision and accuracy of the kriging predictions increased with the sample size, as well as the importance of spatial autocorrelation. On the other way, in some cases, high data variability may hamper the retrieval of the spatial models and mask spatial autocorrelation (Rufino et al., 2006).

The clumped spatial patterns typical of marine species distribution can emerge simply as a result of the spatial autocorrelation of the environmental and of biotic processes (Legendre, 1993). In the current work, most species evidenced the presence of auto-correlation in the experimental variogram model. Strong residual geographic patterning generally indicates that either key environmental predictors are missing, that the model is mis-specified or that geographic factors are influential (Elith and Leathwick, 2009). In a study as broad as the current one this would be a natural consequence as it was not the aim to explore the key environmental predictors of each species, nether to parametrise in detail each method. Recent works have shown excellent results on the application of combined methods (random forest + kriging) in other areas (Appelhans et al., 2015; Diesing et al., 2014; Hengl et al., 2015; Li et al., 2011, 2013, 2016), and thus it would be interesting to explore such applications to marine SDM on future works. Further, the selection protocol is extensible interpolation methods on the spatio-temporal domain.

It is evident that the best interpolation changed widely across species and years, and thus in each case a detailed analysis is required. Furthermore, the fact that certain interpolation models performed better for some species in a certain dataset, does not imply that it will always perform better with other fisheries datasets (Davis, 1987). Nevertheless, our application of the selection protocol to the two surveys reveals general guidelines for the variability of the results given by different interpolation methods. It is clear that each model need to be parameterized in detail, individually and according to the species data for a proper spatial analysis and that neglecting ecological knowledge is a limiting factor in the use of statistical modelling to predict species distribution (Austin, 2002).

We conclude that the proposed 2-step approach for method's selection has several benefits: 1) it is accessible and does not require any complex software or sophisticated method; 2) it is explicit in the sense that it evidences the benefits of each interpolation model relative to the others, empirically, that is on the maps produced and thus, permits to make a decision accordingly to the objectives of each study; 3) it takes into account different criteria, thus integrating several desirable properties of interpolation methods; 4) it does not depend on the method used or the data characteristics, thus being universal and can be applied to virtually any method developed in the future.

Acknowledgments

The authors acknowledge all previous participants in EVHOE and PELGAS surveys, in particular Mathieu Doray and Pierre Petitgas for providing the data set used in this work and for several discussions on the subjects covered. Marta M. Rufino is funded by a *post-doctoral contract (researcher)* awarded by IFREMER and Agrocampus Ouest within the framework of the MSFD. The authors also acknowledge two anonymous referees for their useful comments on the manuscript.

Data availability

The data used in the current work can be located in

https://campagnes.flotteoceanographique.fr/series/8/

References

Aalto, J., Pirinen, P., Heikkinen, J., Venäläinen, A., 2013. Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. Theor. Appl. Climatol. 112, 99–111. doi:10.1007/s00704-012-0716-9

Amante, C.J., Eakins, B.W., 2016. Accuracy of interpolated bathymetry in digital elevation models. J. Coast. Res. 76, 123–133. doi:10.2112/SI76-011

Appelhans, T., Mwangomo, E., Hardy, D.R., Hemp, A., Nauss, T., 2015. Evaluating machine learning approaches for the interpolation of monthly air temperature at. Spat. Stat. 14, 91–113. doi:10.1016/j.spasta.2015.05.008

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688. doi:10.1111/j.1365-2699.2006.01584.x

Araujo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93, 1527–1539. doi:10.1890/11-1930.1

Augustin, N.H., Trenkel, V.M., Wood, S.N., Lorance, P., 2013. Space-time modelling of blue ling for fisheries stock management. Environmetrics 24, 109–119. doi:10.1002/env.2196

Austin, M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Modell. 200, 1–19. doi:10.1016/j.ecolmodel.2006.07.005

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Modell. 157, 101–118. doi:10.1016/S0304-3800(02)00205-3

Baddeley, A., Gregori, P., Mateu, J., Stoica, R., Stoyan, D., Bickel, E.P., Diggle, P., Fienberg, S., Gather, U., Baddeley, A., Stoyan, D., 2006. Case Studies in Spatial Point Process Modeling. Springer. doi:10.1007/0-387-31144-0

Bellocchi, G., Rivington, M., Donatelli, M., Matthews, K., 2010. Validation of biophysical models : issues and methodologies. Agron. Sustain. Dev. 30, 109–130. doi:10.1051/agro/2009001

Bez, N., Rivoirard, J., 2001. Transitive geostatistics to characterise spatial aggregations with diffuse limits: an application on mackerel ichtyoplankton. Fish. Res. 50, 41–58. doi:10.1016/S0165-7836(00)00241-1

Bivand, R., Pebesma, E., Gómez-Rubio, V., 2013. Applied spatial data analysis with R, Use R. Springer. doi:10.1007/978-0-387-78171-6

Cotter, J., Petitgas, P., Abella, A., Apostolaki, P., Mesnil, B., Politou, C.-Y., Rivoirard, J., Rochet, M.-J., Spedicato, M.T., Trenkel, V.M., Woillez, M., 2009. Towards an ecosystem approach to fisheries management (EAFM) when trawl surveys provide the main source of information. Aquat. Living Resour. 22, 243–254. doi:10.1051/alr/2009025

Cressie, N.A.C., 1993. Statistics for spatial data, Statistics for Spatial Data. John Wiley &

Sons, New York. doi:10.1002/9781119115151

Dauvin, J.-C., Thiebaut, E., Gesteira, J.L.G., Ghertsos, K., Gentil, F., Ropert, M., Sylvand, B., 2004. Spatial structure of a subtidal macrobenthic community in the Bay of Veys (western Bay of Seine, English Channel). J. Exp. Mar. Bio. Ecol. 307, 217–235. doi:10.1016/j.jembe.2004.02.005

Davis, B.M., 1987. Uses and abuses of cross-validation in geostatistics. Math. Geol. 19, 241–248. doi:10.1007/BF00897749

Diesing, M., Green, S.L., Stephens, D., Lark, R.M., Stewart, H.A., Dove, D., 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Cont. Shelf Res. 84, 107–119. doi:10.1016/j.csr.2014.05.004

Doray, M., Duhamel E. , Huret M. , Petitgas P., M.J., 2002. PELGAS. doi:10.18142/18

Doray M., Badts V., Masse J., Duhamel E., Huret M., Doremus G., P.P., 2014. Manuel des protocoles de campagne halieutique. Campagnes PELGAS (PELagiques GAScogne) / Manual of fisheries survey protocols. PELGAS surveys (PELagiques GAScogne). doi:10.13155/30259

Doray, M., Petitgas, P., Romagnan, J.B., Huret, M., Duhamel, E., Dupuy, C., Spitz, J., Authier, M., Sanchez, F., Berger, L., Dorémus, G., Bourriau, P., Grellier, P., Massé, J., 2018. The PELGAS survey: Ship-based integrated monitoring of the Bay of Biscay pelagic ecosystem. Prog. Oceanogr. 166, 15–29. doi:10.1016/j.pocean.2017.09.015

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schro¨der, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography (Cop.). doi: 10.11. doi:10.1111/j.2007.0906-7590.05171.x

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. doi:10.1146/annurev.ecolsys.110308.120159

Evaluation Halieutique de l'Ouest Européen, EVHOE cruise, RV Thalassa, IFREMER, n.d. doi:10.18142/8

Evans JS, 2017. spatialEco. R package version 0.0.1-7.

Fortin, M.M.-J., Dale, M.R.T., 2005. Spatial analysis: a guide for ecologists, 4th ed. Cambridge University Press, Cambrigde. doi:10.2277/0521804345

Gaetan, C., Guyon, X., Bleakley, K., 2010. Spatial Statistics and Modeling, Media. Springer. doi:10.1007/978-0-387-92257-7

Gasch, C.K., Hengl, T., Gräler, B., Meyer, H., Magney, T.S., Brown, D.J., 2015. Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. Spat. Stat. 14, 70–90. doi:10.1016/j.spasta.2015.04.001

Giannoulaki, M., Machias, A., Koutsikopoulos, C., Haralabous, J., Somarakis, S., Tsimenides, N., 2003. Effect of coastal topography on the spatial structure of the populations of small pelagic fish. Mar. Ecol. Prog. Ser. 265, 243–253. doi:10.3354/meps265243

Giannoulaki, M., Machias, A., Koutsikopoulos, C., Somarakis, S., 2006. The effect of coastal topography on the spatial structure of anchovy and sardine. ICES J. Mar. Sci. 63, 650–662. doi:10.1016/j.icesjms.2005.10.017

Guisan, A., Edwards Thomas C., J., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Modell. 157, 89–100. doi:10.1016/S0304-3800(02)00204-1

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–186. doi:10.1016/S0304-3800(00)00354-9

Hengl, T., 2009. A Practical guide to Geostatistical Mapping, Scientific and Technical Research series. doi:10.1016/0277-9390(86)90082-8

Hengl, T., Geuvelink, G.B.M., Stein, A., 2003. Comparison of kriging with external drift and regression kriging. Technical note, ITC, Available on-line at http://www.itc.nl/library/Academic_output/. doi:10.1016/S0016-7061(00)00042-2

Hengl, T., Heuvelink, G., Stien, A., 2004. A generic frame work for the spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93. doi:10.1016/j.geoderma.2003.08.018

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One 10, 1–26. doi:10.1371/journal.pone.0125814

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. Comput. Geosci. 33, 1301–1315. doi:10.1016/j.cageo.2007.05.001

Hengl, T., MacMillan, R.A., Nikolić, M., 2013. Mapping efficiency and information content. Int. J. Appl. Earth Obs. Geoinf. 22, 127–138. doi:10.1016/j.jag.2012.02.005

Hoffman, S., 2015. Assessment of prediction accuracy in autonomous air quality models. Desalin. Water Treat. 57, 1322–1326. doi:10.1080/19443994.2014.1002283

Hui, F.K.C., Warton, D.I., Foster, S.D., Dunstan, P.K., 2013. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time vs. separate species distribution models. Ecology 94, 1913–1919. doi:10.1890/12-1322.1

Huret, M., Petitgas, P., Woillez, M., 2010. Dispersal kernels and their drivers captured with a hydrodynamic model and spatial indices: A case study on anchovy (Engraulis encrasicolus) early life stages in the Bay of Biscay. Prog. Oceanogr. 87, 6–17. doi:10.1016/j.pocean.2010.09.023

Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perčec Tadič, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res. Atmos. 119, 2294–2313. doi:10.1002/2013JD020803

Lark, M., Dove, D., Green, S., Stewart, H., Marchant, B., Diesing, M., 2016. Uncertainty in predictions of seabed sediment classes based on grab samples and acoustic data. Geophys. Res. Abstr. 18.

Lecours, V., Dolan, M.F.J., Micallef, A., Lucieer, V.L., 2016. A review of marine geomorphometry, the quantitative study of the seafloor. Hydrol. Earth Syst. Sci. 20, 3207–3244. doi:10.5194/hess-20-3207-2016

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241. doi:10.1029/1998WR900018

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74, 1659–1673. doi:10.2307/1939924

Legendre, P., Legendre, L., 1998. Numerical ecology, 2nd ed. Elsevier, Amsterdam. doi:10.1016/S0304-3800(00)00291-X

Li, J., 2017. Assessing the accuracy of predictive models for numerical data: Not r nor r 2 , why not? Then what? PLoS One 12, 1–16. doi:10.1371/journal.pone.0183250

Li, J., 2016. Assessing spatial predictive models in the environmental sciences: Accuracy measures, data variation and variance explained. Environ. Model. Softw. 80, 1–8. doi:10.1016/j.envsoft.2016.02.004

Li, J., Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecol. Inform. 6, 228–241.

doi:10.1016/j.ecoinf.2010.12.003

Li, J., Heap, A.D., 2008. A review of spatial interpolation methods for environmental scientists, Geoscience Australia.

Li, J., Heap, A.D., Potter, A., Huang, Z., Daniell, J.J., 2011. Can we improve the spatial predictions of seabed sediments? A case study of spatial interpolation of mud content across the southwest Australian margin. Cont. Shelf Res. 31, 1365–1376. doi:10.1016/j.csr.2011.05.015

Li, J., Justy, P., Siwabessy, W., Tran, M., Huang, Z., Heap, A.D., 2013. Predicting Seabed Hardness Using Random Forest in R, in: Elsevier (Ed.), Data Mining Applications with R. pp. 299–329. doi:10.1016/B978-0-12-411511-8.00011-6

Li, J., Siwabessy, J., Tran, M., 2016. Selecting optimal random forest predictive models : a case study on predicting the spatial distribution of seabed hardness selecting optimal random forest predictive models: a case study on predicting the spatial distribution of seabed hardness. PLoS One 11, 1–29. doi:10.1371/journal.pone.0149089

Maravelias, C.D., 1999. Habitat selection and clustering of a pelagic fish: effects of topography and bathymetry on species dynamics. Can. J. Fish. Aquat. Sci. 56, 437–450. doi:10.1139/f98-176

Morfin, M., Bez, N., Fromentin, J.M., 2016. Habitats of ten demersal species in the Gulf of Lions and potential implications for spatial management. Mar. Ecol. Prog. Ser. 547, 219–232. doi:10.3354/meps11603

Naimi, B., Araújo, M.B., 2016. Sdm: A reproducible and extensible R platform for species distribution modelling. Ecography (Cop.). 39, 368–375. doi:10.1111/ecog.01881

Nash, J.E., Sutcliffe, J. V, 1970. River flow forecasting through conceptual models part I-a discussion of principles. J. Hydrol. 10, 282–290. doi:10.1016/0022-1694(70)90255-6

Nychka, D., 2016. fields: Tools for Spatial Data. doi:10.5065/D6W957CT

Olden, J.D., Jackson, D.A., 2002. A comparison of statistical approaches for modelling fish species distributions. Freshw. Biol. 47, 1976–1995. doi:10.1046/j.1365-2427.2002.00945.x

Paramo, J., Roa, R., 2003. Acoustic-geostatistical assessment and habitat-abundance relations of small pelagic fish from the Colombian Caribbean. Fish. Res. 60, 309–319. doi:10.1016/S0165-7836(02)00142-X

Richter, K., Atzberger, C., Hank, T.B., Mauser, W., 2012. Derivation of biophysical variables from Earth observation data: validation and statistical measures. J. Appl. Remote Sens. 6, 063557–1. doi:10.1117/1.JRS.6.063557

Rufino, M.M., Bez, N., Brind'Amour, A., 2019. Influence of data pre-processing on the behavior of spatial indicators. Ecol. Indic. 99, 108–117. doi:10.1016/j.ecolind.2018.11.058

Rufino, M.M., Bez, N., Brind'Amour, A., 2018. Integrating spatial indicators in the surveillance of exploited marine ecosystems. PLoS One 13, e0207538. doi:10.1371/journal.pone.0207538

Rufino, M.M., Stelzenmüller, V., Maynou, F., Zauke, G.P., 2006. Assessing the performance of linear geostatistical tools applied to artificial fisheries data. Fish. Res. 82, 263–279. doi:10.1016/j.fishres.2006.06.013

Sluiter, R., 2009. Interpolation methods for climate data literature review, KNMI intern rapport.

Stow, C.A., Jolliff, J., McGillicuddy, D.J., Jr., c S.C.D., Allen, J.I., Friedrichs, M.A.M., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76, 4–15. doi:10.1016/j.jmarsys.2008.03.011

Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. 72, 1297–1310. doi:10.1093/icesjms/fsu243

Tobler, W.R., 1970. A computer movie simulation urban growth in Detroit region. Econ. Geogr. 46, 234–240. doi:10.1126/science.11.277.620

Wackernagel, H., 1998. Multivariate geostatistics: An introduction with applications., Second. ed. Springer-Verlag, Berlin. doi:10.1007/978-3-662-03550-4

Webster, R., Oliver, M., 2007. Geostatistics for environmental scientists, 2nd ed. ed. John Wiley and Sons, New York. doi:10.1002/9780470517277

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 1309–1313. doi:10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194. doi:10.1080/02723646.1981.10642213

Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32, 2088–2094. doi:10.1002/joc.2419

Willmott, C.J., Robeson, S.M., Matsuura, K., Ficklin, D.L., 2015. Assessment of three dimensionless measures of model performance. Environ. Model. Softw. 73, 167–174. doi:10.1016/j.envsoft.2015.08.012

Wilson, M.F.J., Connell, B.O., Guinan, J.C., Grehan, A.J., 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. doi:10.1080/01490410701295962

Woillez, M., Poulard, J.-C., Rivoirard, J., Petitgas, P., Bez, N., 2007. Indices for capturing spatial patterns and their evolution in time, with application to European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES J. Mar. Sci. 64, 537–550. doi:10.1093/icesjms/fsm025

Woillez, M., Rivoirard, J., Fernandes, P.G., 2009a. Evaluating the uncertainty of abundance estimates from acoustic surveys using geostatistical simulations. ICES J. Mar. Sci. 66, 1377–1383. doi:10.1093/icesjms/fsp137

Woillez, M., Rivoirard, J., Petitgas, P., 2009b. Notes on survey-based spatial indicators for monitoring fish populations. Aquat. Living Resour. 22, 155–164. doi:10.1051/alr/2009017

Wood, N.S., 2006. Generalized Additive Models: An Introduction with R, Texts in Statistical Science Series. Chapman & Hall/CRC. doi:10.1201/9781315370279

Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Analysing Ecological Data, Profiles of drug substances, excipients, and related methodology. Springer. doi:10.1016/B978-0-12-387667-6.00013-0

FIGURE LEGENDS:

Fig. 1: Graphical abstract: Conceptual diagram of the method developed to compare the interpolators.

Fig. 2: Frequency of the variance explained by predictive model classes between spatial interpolation methods, for the bottom trawl survey (EVHOE, upper panel) and the acoustic pelagic survey (PELGAS, lower panel). Find further details on the methods codes in text.

Fig. 3 : Variance explained by the predictive model between spatial interpolation method's main families, by survey (blue triangles for pelagic survey, PELGAS and red bals for bottom trawl survey, EVHOE)(mean and respective 95% CI estimated using bootstrap). LM: linear model; GAM: generalised additive models; IDW: inverse distance weighting; Vor: voronoi triangulation; TPS: thin plate spline; Kri: kriging and conditional simulation; Covar: multiple regression, regretion tree and random forest (simple and mixed, i.e. with kriged residuals). Please find further details on the methods codes in the text.

Fig. 4 : Variance explained by the predictive model between species, for interpolation methods using topographic covariates (orange line with squares, IMCov) and for methods just using geographic coordinates and eventually depth (green line with diamonds, IMGeo). Filled symbols represent the species captured in the bottom trawl survey whereas open symbols indicate the acoustic pelagic survey. Mean and respective 95% CI estimated using bootstrap is represented. Species were ordered by IMGeo VEcv. Find further details on the species codes in the text.

Fig. 5: Variance explained by the model of each interpolation method (median), estimated by cross validation for all species-year distributions. Red points indicate that the interpolation method was within the best method interquartile range and red star indicate the best VEcv in each case whereas black dots indicate models that did not passed for thr second step. Grey shadded area correspond to methods carried out using the 11 topographic covariates (IMCov), whereas white background shows methods using only lat+long and some depth (IMGeo). Please find further details on the methods and species codes in the text.

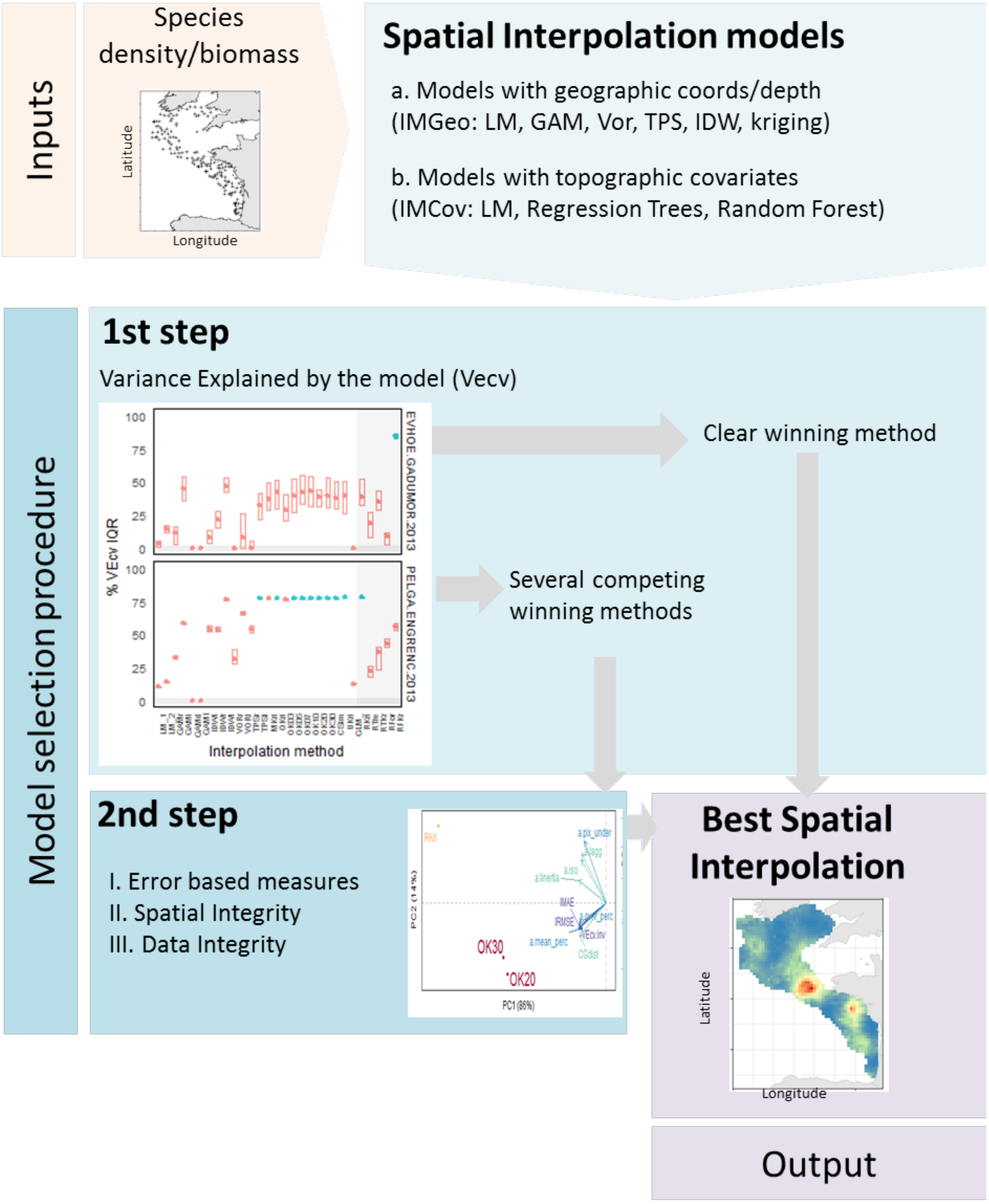
Fig. 6: Distributions and methods that required the second step. Winning interpolation method according to each measure criteria (left panel; only cases where more than one method showing its VEcv within the highest method VEcv interquartile range were selected and situations with ties were excluded, i.e. several methods showing the same classification according to the criteria). Winning method for each measure-criteria by species-year distribution (right panel). Vertical grey line separates the methods using several covariates from the others. Please find further details on the criteria and methods codes in the text. The colour legend is represented in the barplot.

Fig. 7: Second step of the spatial interpolator's selection protocol applied to four case studies (2 bottom trawl, EVHOE and 2 acoustic pelagic, PELGAS). On the right side plots, the PCA shows where each interpolation method (represented with circles, orange-red coloured, according to the distance to the centre) failed according to the measures representing the three selection criteria (error-based in violet, spatial integrity in green and data integrity in blue). On the left panels, the inverse Euclidean distance to the centre of each method, provides the quantitative decision integrative measure. Please find the details of the code's labels in the text.

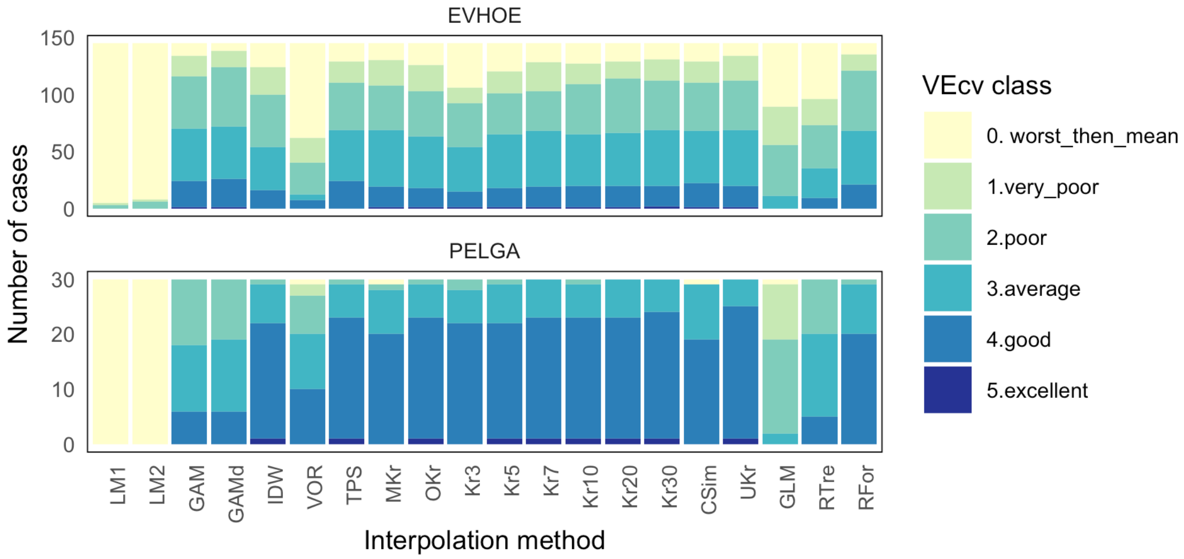
Fig. 8: Winning spatial interpolation method among the different approaches considered, for each case study (species-year distributions) according to the two step selection procedure (1st step using IQR VEcv identified with orange line and 2nd step, using the 3 criteria with 11 measures, identified with blue line). Number of cases of each selected method by survey. Vertical grey line separates the methods using several covariates from the others. See further details of the species codes and methods on text. The colour legend is represented in the barplot.

Figures

Fig. 1

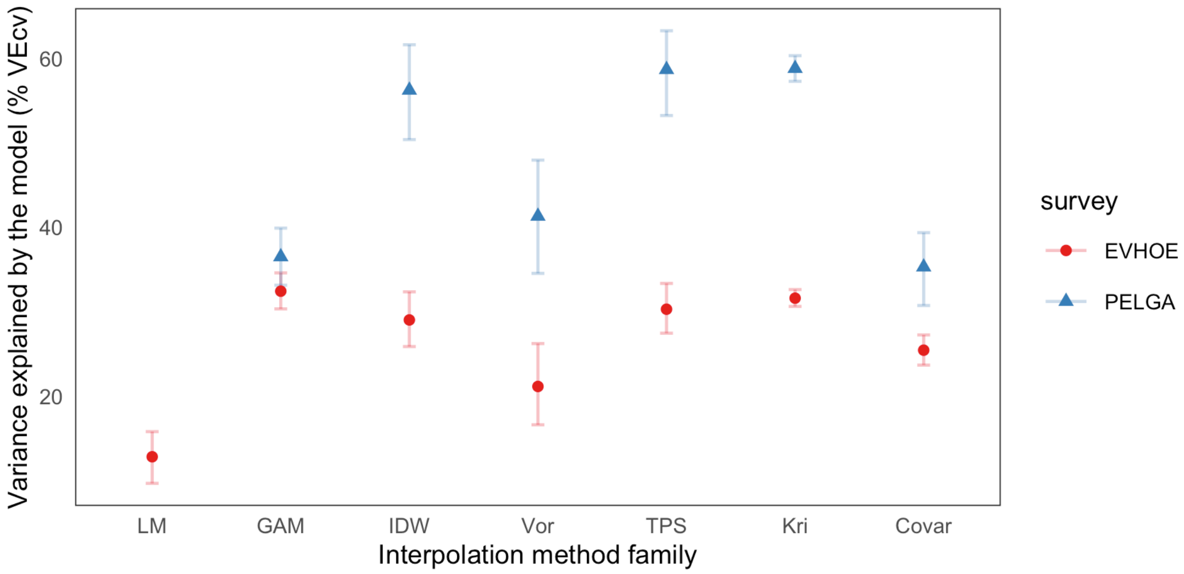

Inputs
Species density/biomass
Latitude
Longitude
Spatial Interpolation models
a. Models with geographic coords/depth
(IMGeo: LM, GAM, Vor, TPS, IDW, kriging)
b. Models with topographic covariates
(IMCov: LM, Regression Trees, Random Forest)
Model selection procedure
1st step
Variance Explained by the model (Vecv)
% VEcv IQR
EVHOE.GADUMOR.2013
PELGA.ENGRENC.2013
Interpolation method
Clear winning method
Several competing winning methods
2nd step
I. Error based measures
II. Spatial Integrity
III. Data Integrity
PC1 (86%)
PC2 (14%)
OK30
OK20
Best Spatial Interpolation
Latitude
Longitude
Output

842 **Fig. 2: VEcv classes + RT**



843
844
845

846 **Fig. 3: VEcv by family**

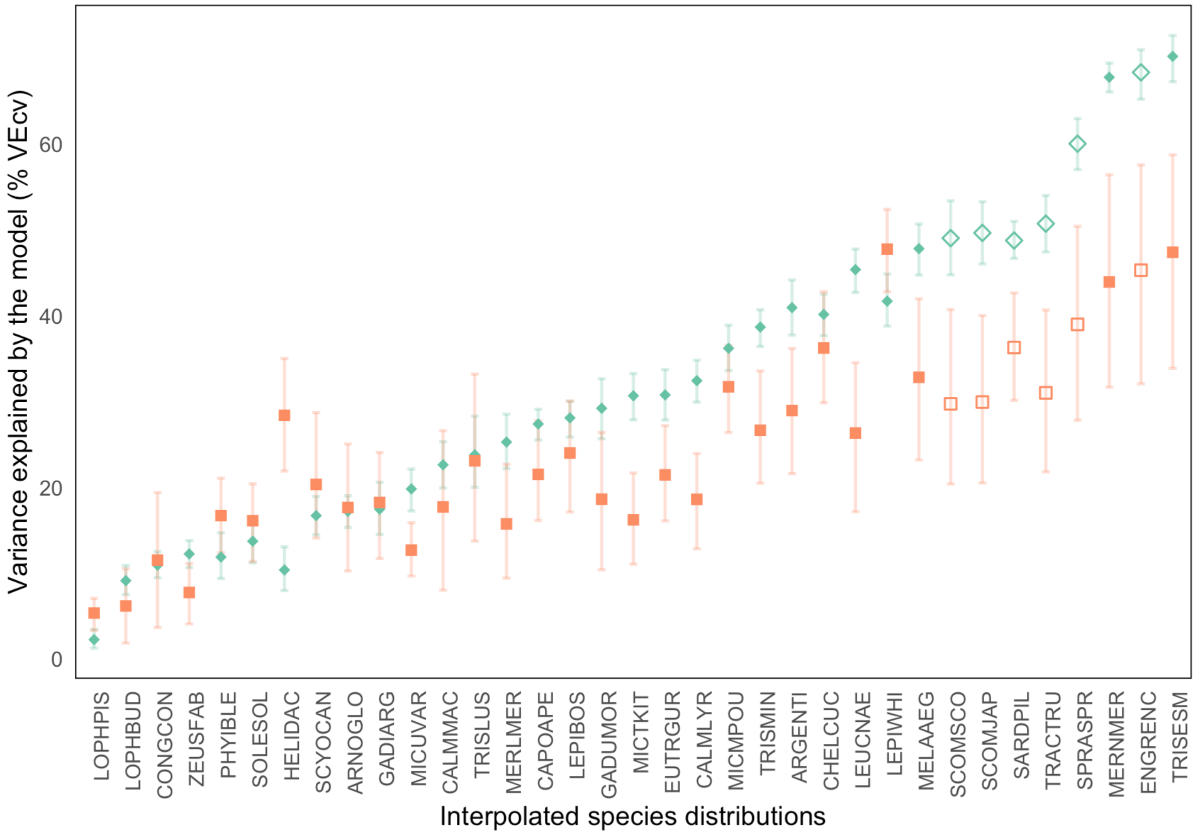


847

848 **Fig. 4: VEcv by Sp**

849
850

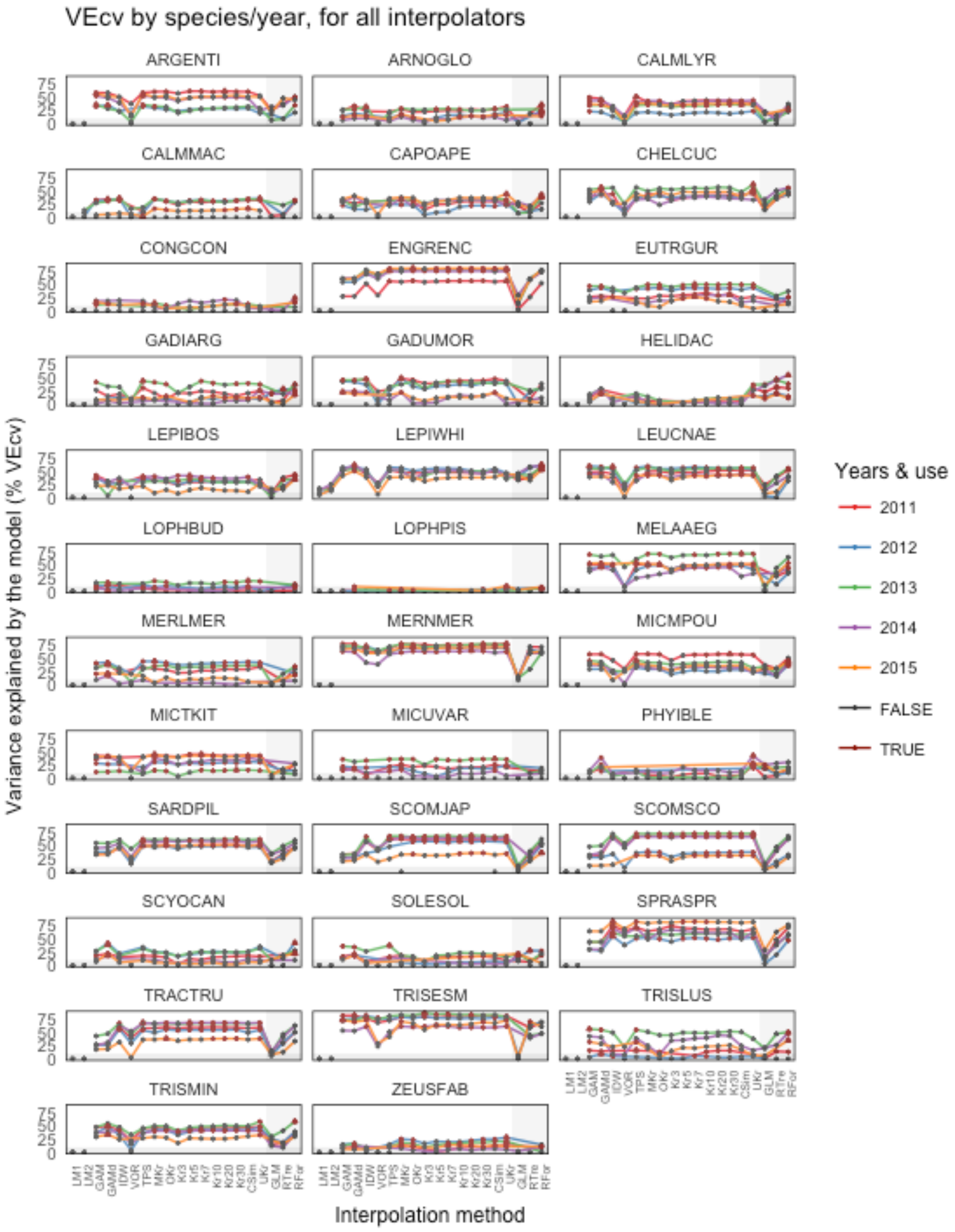
851



Variance explained by the model (% VEcv)
0
20
40
60
Interpolated species distributions
LOPHPIS
LOPHBUD
CONGCON
ZEUSFAB
PHYIBLE
SOLESOL
HELIDAC
SCYOCAN
ARNOGLO
GADIARG
MICUVAR
CALMMAC
TRISLUS
MERLMER
CAPOAPE
LEPIBOS
GADUMOR
MICTKIT
EUTRGUR
CALMLYR
MICMPOU
TRISMIN
ARGENTI
CHELCUC
LEUCNAE
LEPIWHI
MELAAEG
SCOMSCO
SCOMJAP
SARDPIL
TRACTRU
SPRASPR
MERNMER
ENGRENC
TRISESM

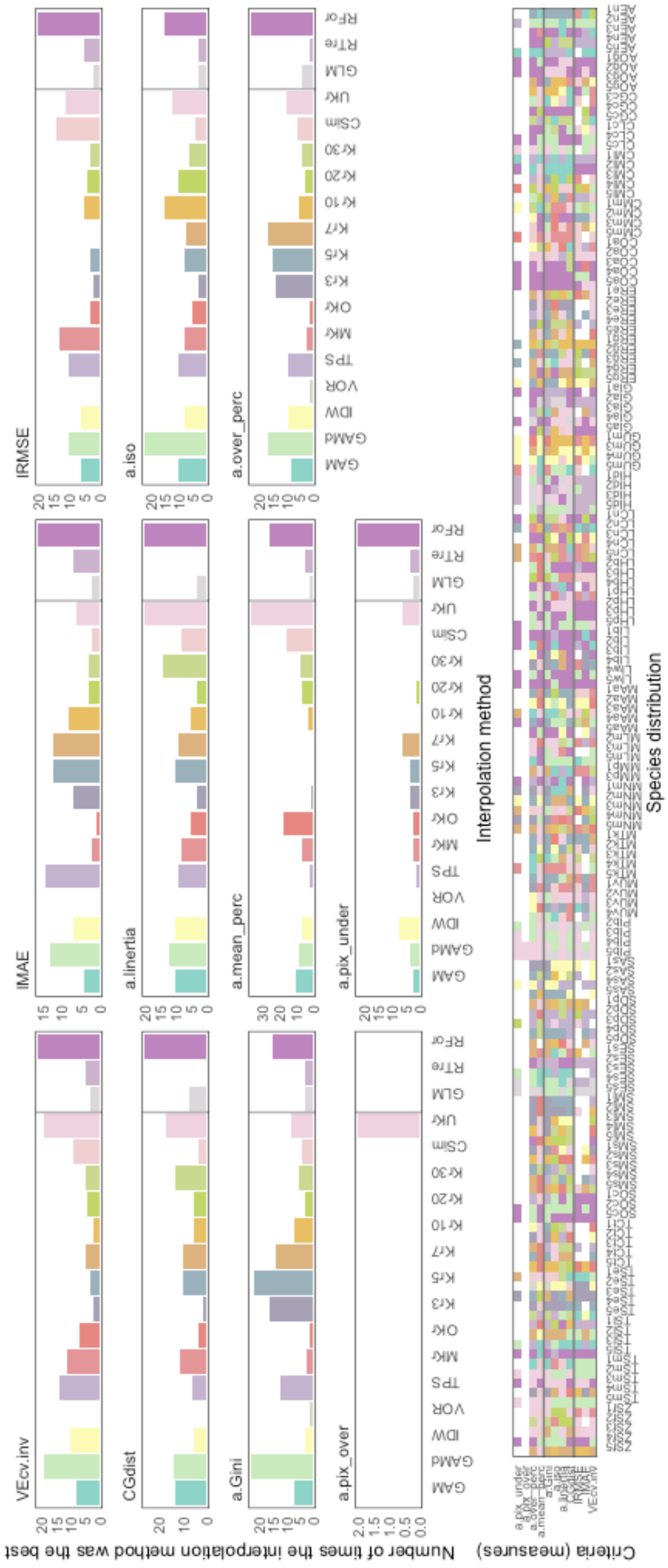
852 **Fig. 5: all VEcvs**

853

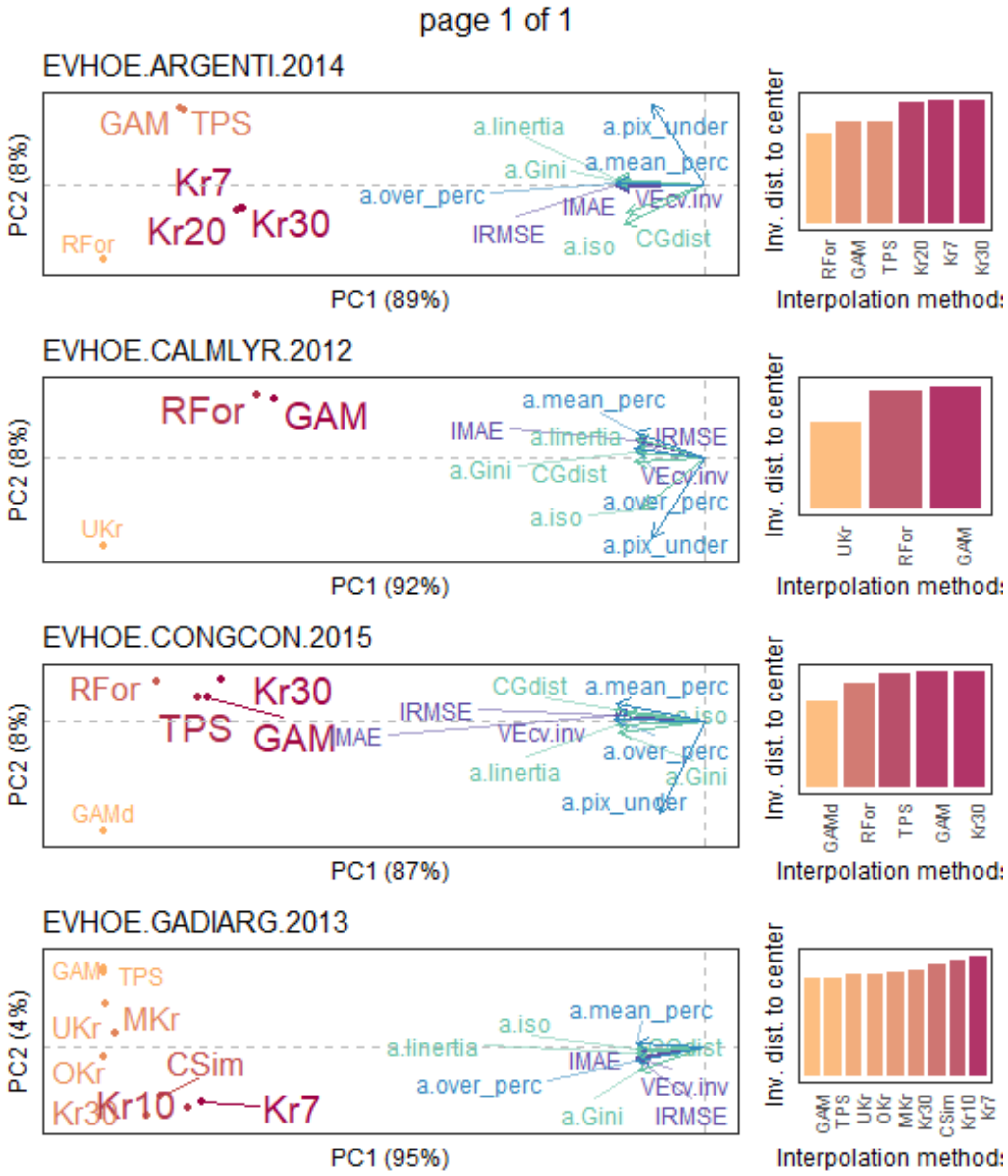


854
855

Fig. 6: Best method according to each criteria (2nd step only).



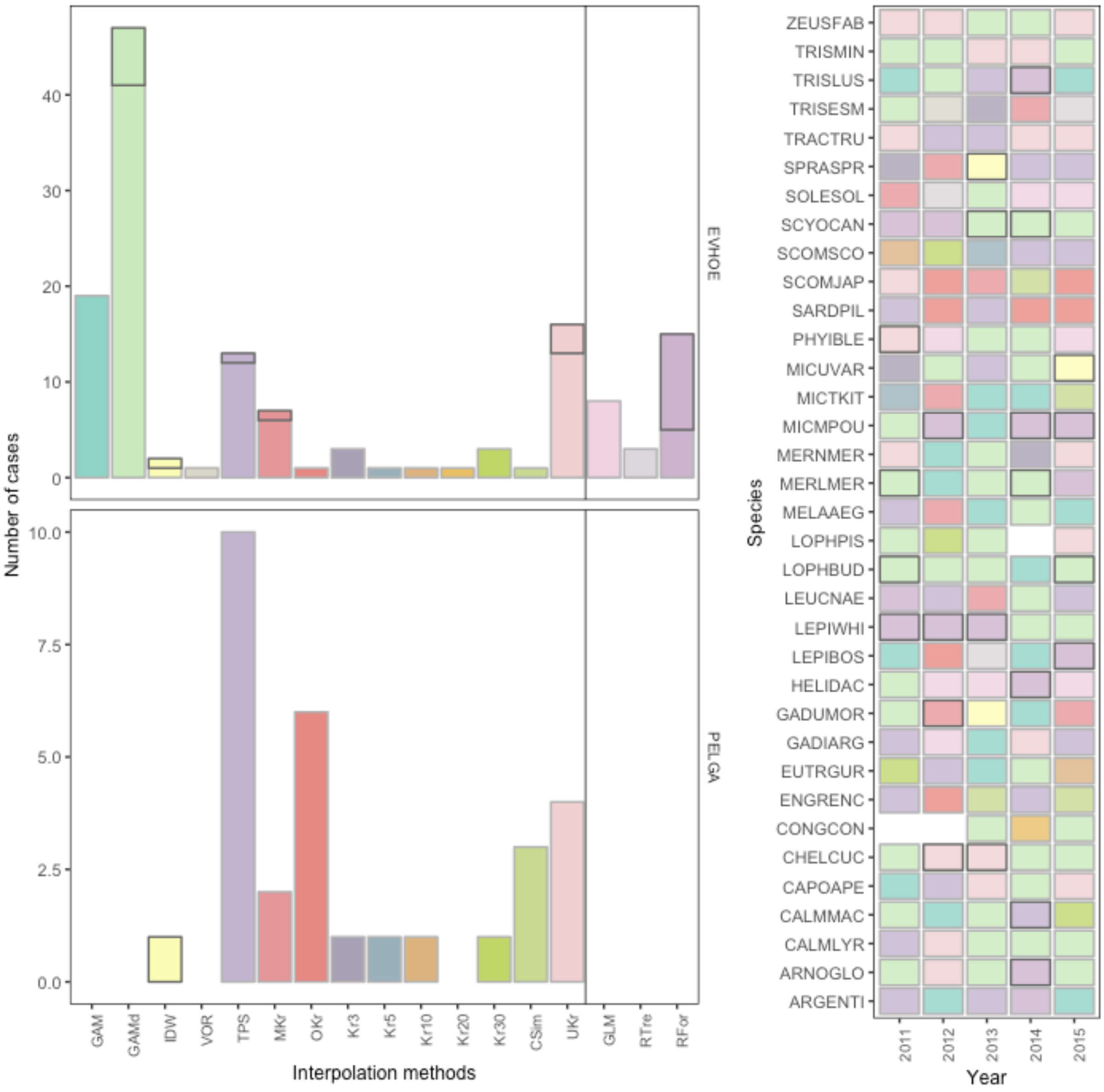
858 **Fig. 7: second step PCA examples**



859

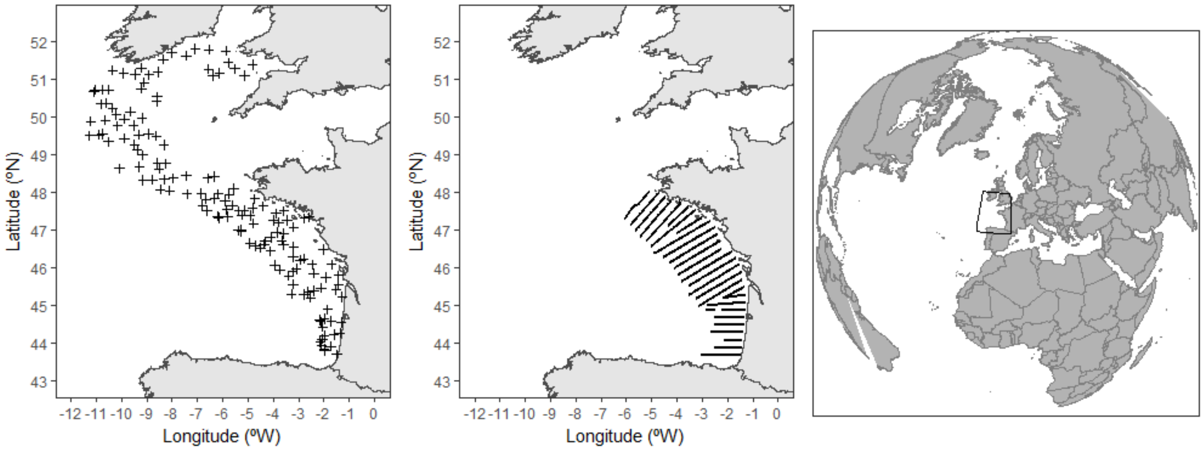
860

Fig. 8: **Winning methods overall;**



Supplement 1

Location of the sampling stations from 2015 of the bottom trawl survey (EVHOE, left panel), of the pelagic acoustic survey (PELGAS) on the Gulf of Gascogne/Bay of Biscay in the North Atlantic coast of France (right panel). Number of hauls per year on EVHOE, from 2011 to 2015 were respectively 220, 195, 208, 219 and 148.



Species list names and abreviations

EVHOE: *Argentina* sp. (ARGENTI)*, Arnoglossus* sp. (ARNOGLO)*, Callionymus lyra* (CALMLYR)*, Callionymus maculatus* (CALMMAC)*, Capros aper* (CAPOAPE)*, Chelidonichthys cuculus* (CHELCUC)*, Conger conger* (CONGCON)*, Eutrigla gurnardus* (EUTRGUR)*, Gadiculus argenteus* (GADIARG), *Gadus morhua* (GADUMOR)*, Helicolenus dactylopterus* (HELIDAC), *Lepidorhombus boscii* (LEPIBOS)*, Lepidorhombus whiffiagonis* (LEPIWHI)*, Leucoraja naevus* (LEUCNAE)*, Lophius budegassa* (LOPHBUD)*, Lophius piscatorius* (LOPHPIS)*, Melanogrammus aeglefinus* (MELAAEG)*, Merluccius merluccius* (MERLMER), *Merlangius merlangus* (MERNMER)*, Microchirus variegatus* (MICUVAR)*, Micromesistius poutassou* (MICPOU)*, Microstomus kitt* (MICKIT)*, Phycis blennoides* (PHYIBLE)*, Scyliorhinus canicula* (SCYOCAN)*, Solea solea* (SOLESOL)*, Trisopterus esmarkii* (TRISESM)*, Trisopterus luscus* (TRISLUS)*, Trisopterus minutus* (TRISMIN) and *Zeus faber* (ZEUSFAB)

PELGAS: *Engraulis encrasicolus* (ENGRENC), *Sardina pilchardus* (SARDPIL), *Scomber japonicas* (SCOMJAP), *Scomber scombrus* (SCOMSCO), *Sprattus sprattus* (SPRASSPR) and *Trachurus trachurus* (TRACTRU)

Details of the survey:

The EVHOE survey has been carried out on the R/V Thalassa, a stern trawler of 73.7 m length by 14.9 m wide (tonnage of 3022 t). The fishing gear used is a GOV 36/47 without exocet Kite which is replaced by 6 additional floats and with a horizontal opening of 20 m and a vertical opening of 4 m.

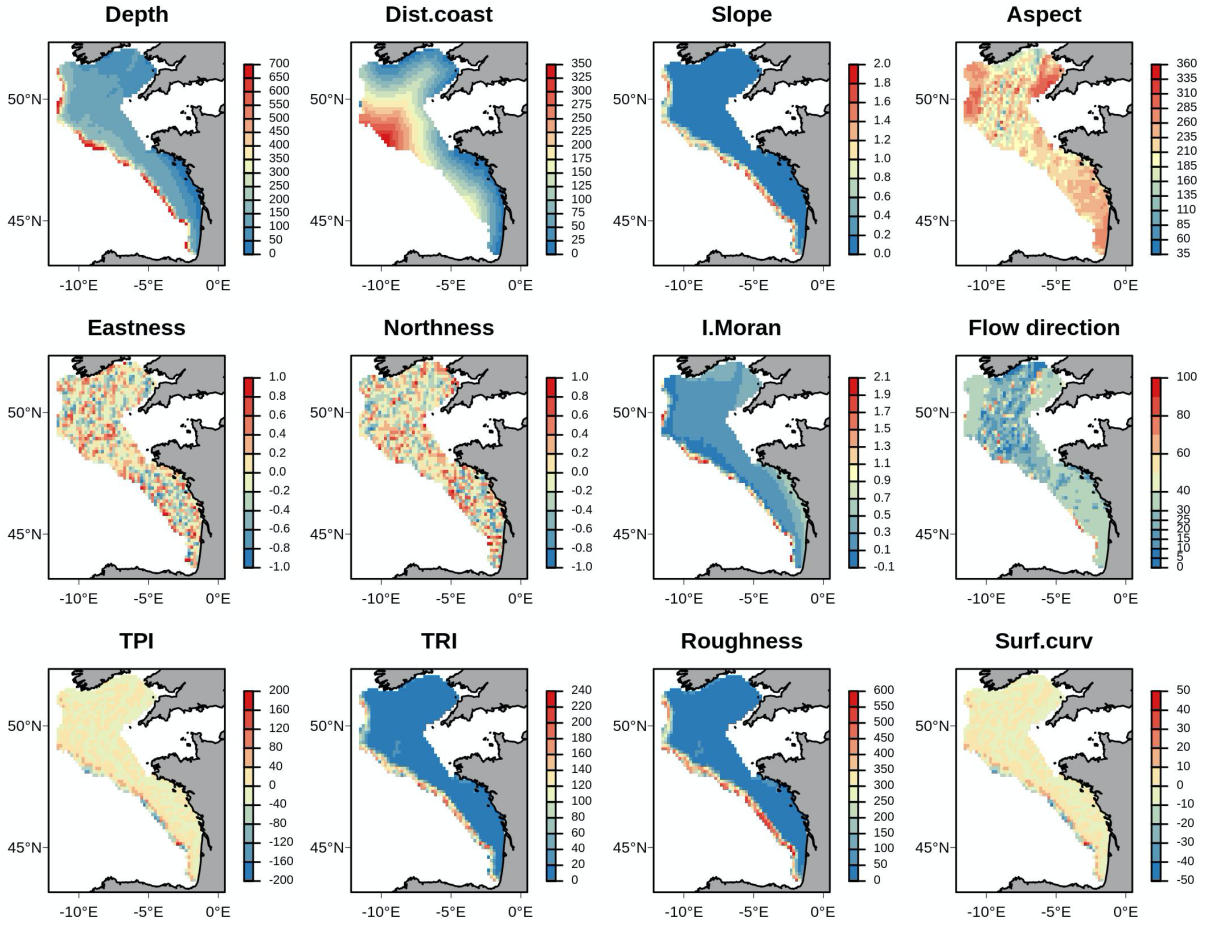
Supplement 2

Description of terrain variables:

Slope represent the terrain steepness (arrangement and magnitude of elevation differences)(slope) whereas terrain aspect (aspect) measures its orientation in degrees, relative to the north and it is particularly important to exposure to currents or water movement (Wilson et al., 2007). From aspect, northerness and easterness were derived (Wilson et al., 2007). Profile curvature defines convex/concave areas, represented by the rate of slope change along a profile, i.e. the surface of the steepest down-slope direction (surf.curv)(package 'spatialEco')(Evans JS, 2017). Bathymetric Position Index is the difference between the value of a cell and the mean value of its surrounding cells and provides an indication of whether any particular pixel forms part of a positive (e.g., crest) or negative a (e.g., trough) feature of the surrounding terrain (TPI, marine version of the topographic position index). Terrain Ruggedness Index represents terrain variability whereas roughness represents the bathymetric amplitude of a cell and its surroundings (TRI). Surface flow confluence indicates the steepest downhill path (flowdir). Local Moran I was calculated as a measure of local spatial autocorrelation in the bathymetric neighborhood (moran)(Diesing et al., 2014; Li et al., 2016). Additionally, distance to the nearest coast was also estimated (dist.coast).

Maps of the topographic variables used in the current work (bottom trawl area, EVHOE)

After gridding all variables to EVHOE prediction grid (1117 pixels, 15 km).



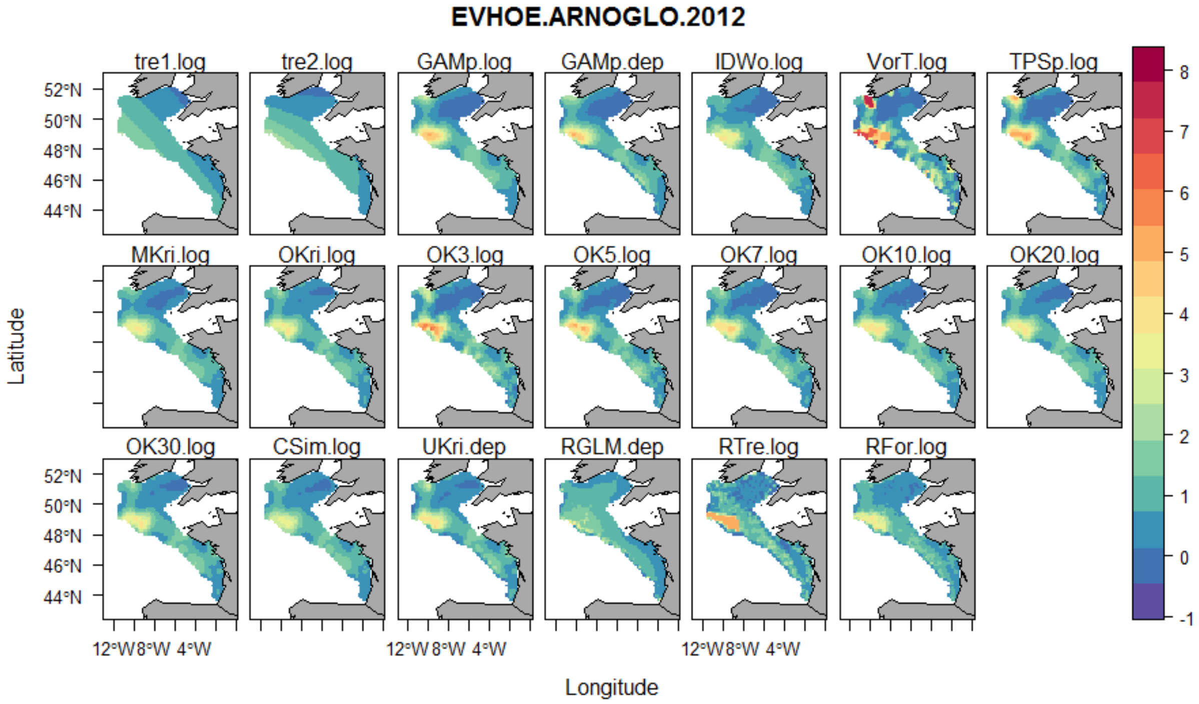
Supplement 3

Table 1: Summary of the interpolation methods used in the current work.

Interpolator	Description	R-package /Function	Covar
IMGeo			
LM1	1st-order trend surface	gstat::gstat(vari ~ 1, degree=1)	-
LM2	2nd order trend surface	gstat::gstat(vari ~ 1, degree=2)	-
GAM	Generalized Additive Model		

	(Wood, 2006)		
GAM	GAM in function of lat and long (in UTM)	mgcv ::gam (lvari ~ s(lat ,long))	
GAMd	GAM in function of lat and long (in UTM) and depth	mgcv ::gam (lvari ~ s(lat ,long)+s(Depth))	Depth
IDW	Inverse distance weight		
IDW	Optimized using cross validation	gstat::gstat(lvari~1, nmax=opt$par[1], set=list(idp=opt$par[2]))	-
VOR	Voronoi thesselation (Fortin and Dale, 2005)	dismo::voronoi(dat.s["lvari"])	
TPS	Thin Plate Spline interpolation (Nychka, 2016)		
Kriging	Methods using kriging (Bivand et al., 2013)		
MKr	Ordinary kriging interpolation (manual fitting of variogram)	gstat::krige(lvari ~ 1)	-
CSim	Stochastic conditional simulation	gstat::krige(lvari ~ 1, nsim = 1000, nmax = 20)	-
OKri		automap::autoKrige(lvari~1, model = c("Sph", "Exp")))(gstat)	-
17-22. Kr3 Kr5 Kr7 Kr10 Kr20 Kr30	OK with various neighborhood	automap::autoKrige(lvari~1, model = c("Sph", "Exp"), nmax=jj) (gstat)	-
UKr	Universal kriging, with depth as covariate	automap::autoKrige(lvari~Depth, model = c("Sph", "Exp")) (gstat)	Depth
GLM	Multiple regression	dat.s, lvari~Depth+slope+aspect+eastness+northness+surf.curv+TPI+TRI+rough+dist.coast+flowdir	Depth+slope+aspect+eastness+northness+surf.curv+TPI+TRI+rough+dist.coast+flowdir
RTre	Regression tree	rpart:: rpart (dat.s, vari ~ Depth+slope+aspect+eastness+northness+surf.curv+TPI+TRI+rough+dist.coast+flowdir)	Depth+slope+aspect+eastness+northness+surf.curv+TPI+TRI+rough+dist.coast+flowdir
RFor	Random forest	randomForest: randomForest(dat.s, vari ~ Depth+slope+aspect+eastness+northness+surf.curv+TPI+TRI+rough+dist.coast+flowdir)	Depth+slope+aspect+eastness+northness+surf.curv+TPI+TRI+rough+dist.coast+flowdir+moran

Details on the spatial interpolation methods considered



Methods using just geographic coordinates or depth (IMGeo)

1st and 2nd order trend surfaces (LM1 and LM2, respectively).

In these interpolation methods, a first or second order trend surface is fitted to the raw data, respectively. It is a simplistic approach that was included in the current work as a worst case scenario that should be slightly better than a simple overall mean.

Inverse distance weighting (IDW)

Inverse distance weighting (IDW) is an advanced nearest neighbor approach that allows including more observations than only the nearest points. The value at a certain grid cell is obtained from a linear combination of the surrounding locations and the weight of each observation is determined by the distance. IDW is an exact interpolator. The method is fast, easy to implement and easily "tailored" for specific needs, but ancillary data cannot be incorporated. The method tends to generate "bull's eye patterns" (Sluiter, 2009).

Voronoi tessellation (VorT)

The nearest neighbors method predicts the value of an attribute at an unsampled point based on the value of the nearest sample by drawing perpendicular bisectors between sampled points, forming such as Voronoi polygons (or Dirichlet/ Thiessen). This produces one polygon per sample and the sample is located in the center of the polygon, such that in each polygon all points are nearer to its enclosed sample point than to any other sample points (Legendre and Legendre, 1998; Li and Heap, 2008; Webster and Oliver, 2007). This technique is generally used with point data or categorical variables, but can also be used with densities/biomasses (Baddeley et al., 2006; Dauvin et al., 2004; Morfin et al., 2016; Thorson et al., 2015; Zuur et al., 2007).

Thin Plate Splines (TPS)

Thin plate smoothing splines (TPS), formally known as "laplacian smoothing splines". Similar to the previous method, splines are fitted to the sampled data, but in this method, the smoothing parameter is calculated by minimizing the generalized cross validation function (GCV). This method is relatively robust because the minimization of GCV directly addresses the predictive accuracy and is less dependent on the veracity of the underlying statistical model (Hutchinson, 1995) (Li and Heap, 2008). We applied this method using package::fields.

Generalized Additive Models (GAM)

Generalized additive models (GAM) are a semiparametric extension of generalized linear models (GLM), but allow nonlinear relationships between the response and explanatory variables (Wood, 2006), are very commonly used in biological studies (Guisan et al., 2002). GAMs have been often used as a method to produced spatial predictions (i.e. interpolation) by considering the geographic coordinates and its interaction as covariates (Augustin et al., 2013). In the current work we used a GAM applied with the geographic coordinates as covariates (s(x,y, bs="ts"))(GAM), a model where besides geographic coordinates, depth was also considered as covariate (GAMd). GAMs were applied using the r package::mgcv (Wood, 2006).

Kriging

From an interpolation point of view, kriging is equivalent to a thin-plate spline and is one species among the many in the genus of weighted inverse distance methods, albeit with attractive properties. However, from a statistical point of view, kriging produces the "best linear unbiased prediction" for an unknown location. It is linear since the estimated values are weighted linear combinations of the available data, unbiased because the mean of the error is 0, and it aims to minimize the variance of the errors (Cressie 1990). Several variations of kriging methods were selected following previous works, all applied in the log transformed data. A sequence of interpolation approaches was considered, starting with ordinary kriging with global mean (Okr using automatic modelling and Mkr using manual variogram fitting), ordinary kriging with local neighbourhood estimation (considering 3, 5, 7, 10, 20 and 30 neighbours. Kriging neighborhood is a defined area, in terms of shape and size. Only samples from this area are used in the computation of the local estimates using the kriging technique.

Kriging with external trend, also called universal kriging using depth as covariate (Ukr). It is an extension of OK by incorporating the local trend within the neighbourhood search widow as a smoothly varying function of the coordinates. UK estimates the trend components within each search neighbourhood window and then performs SK on the corresponding residuals.

Stochastic conditional simulations (CSim)

These techniques are used more and more, commonly to generate a series of spatial data that have a given degree of spatial dependence, in order to evaluate whether or not observed sample data show significant spatial patterns (Fortin and Dale, 2005). In this method, the parameters of the variogram model (defined previously for ordinary kriging) derived from the experimental variogram were used to generate 1000 stochastic simulations, with the same degree of spatial variance as the observed data (Fortin and Dale, 2005). These methods are known to generate maps having more spatial variability than the kriged ones and hence looking more realistic in comparison to the observed map (Fortin and Dale, 2005).

Methods using topographic covariates (IMCov)

In a mixed method approach to interpolation, the final predictions result of a combination of methods. The main trends are modelled in function of a group of selected covariates in first step (for example General Linear Model (GLM) or machine learning). In a second step, the residuals of this model are then analyzed using kriging, and then incorporated into the predictions (Hengl et al., 2007, 2004; Li et al., 2016). These methods require the availability of covariates. In the current work we used the marine topographic variables derived from GEBCO bathymetric maps, therefore widely available at a worldwide scale. Three mixed models were considered, one using general linear model and two using machine learning algorithms, regression trees and random forest.

Multiple regression (GLM)

Stepwise multiple regression with all topography-depth covariates was carried out for each distribution (lvari ~ Depth + slope + rough + moran + TRI + TPI + dist.coast + flowdir + aspect + eastness + northness). The regression model considered assumes that the residuals are generated from a normally distributed, second-order stationarity random process—i.e. a random process that has a constant mean and variance.

Regression trees (RTre)

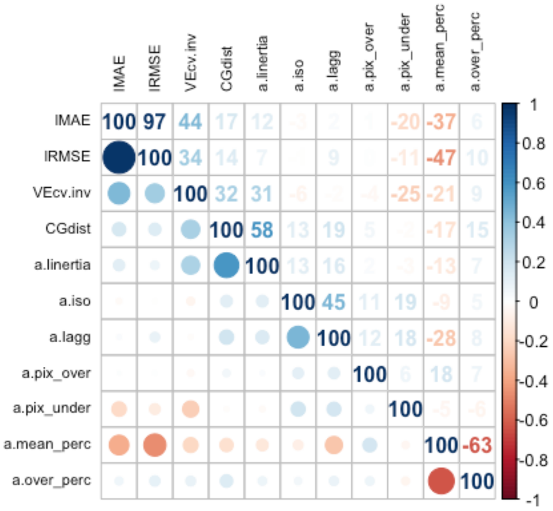
The regression tree approach (also known as binary decision trees) uses binary recursive partitioning whereby the data of the primary variable are successively split along the gradient of the explanatory variables into two descendent subsets or nodes. These splits occur so that at any node the split is selected to maximize the difference between two split groups or branches. The mean value of the primary variable in each terminal node can then be used to map the variable across the region of interest (Li and Heap, 2008). Regression tree (CART) algorithm was fitted to the data to produce a tree with optimal tree size.

Random forest (RFor)

A random forest model of each species distribution in function of all marine topographic covariates was produced (Hengl et al., 2015; Li et al., 2016, 2013, 2011)(vari ~ Depth + slope + aspect + eastness + northness + surf.curv + TPI + TRI + rough + moran + dist.coast + flowdir).

Supplement 4

Correlation plot between the indicators



Supplement 5

```
# title: "Example script for the selection of interpolation method"
# author: "Marta M Rufino^[EMH, IFREMER]"

# The aim of this script is to provide an example of the selection of an interpolation method.
This is an accompanying work of the paper.
```

```
# We will need to have theese packages intalled:
require(ggplot2); require(RColorBrewer); require(gridExtra) # ploting
require(dplyr); require(tidyr);#data manipulation
require(raster); require(rasterVis); #plot and spatial data manipulation
require(sp); # spatial data
require(gstat); #kriging and idw
require(fields) #tps
require(mgcv); #gam
require(RGeostats) #spatial indicators. Pkg can be download from here
http://rgeostats.free.fr/download.php and install manually
require(ineq) #Gini index (spatial indicators)
require(ggrepel) # PCA plot

# Get an example running with MEUSE dataset:
library(sp); library(gstat)
data(meuse)
data(meuse.grid)
gridded(meuse.grid) = ~x+y
# m <- vgm(.59, "Sph", 874, .04)

dat = meuse %>%
  dplyr::select(x,y,zinc, dist) %>%
  dplyr::rename("vari" = "zinc",
               "Depth" = "dist")
dat.grid <- meuse.grid["dist"]
names(dat.grid) = "Depth"

# 1. Interpolate the data (function interp.dat_CV)

# First we make the different spatial interpolation models. For each model, we will do cross
validation. We will only consider the models without covariates to facilitate the process.

## This chunk runs a function to interpolate the data of each species distribution
(interp.dat_CV) and estimate respective CV

## DataDir should the the directory where you have your functions file
('interp.dat_CV_script.r')

# Please run the required functions which are in the end of the script

# Run the function
kk <- interp.dat_CV(nam="zinc", dat=dat, dat.grid=dat.grid, CV=TRUE, plotit=FALSE,
                   replicate.cv = 10)
# note we only used 10 replicates of the cv instead of 100 to make a quicker test

# see the results:
head(kk)
# the result of this function is a list with three items:
# 1. res is the raster stack with all predictions from each method;
# 2. cv.results is the cross-validation summary results
# 3. model.params is some of the models parameters stored

# Extract the raster stack with all interpolator predictions and store it as a new object
pred = kk$res
names(pred)

# Plot the interpolations
coli <- function (region = rev(brewer.pal(n = 10, 'Spectral')), ...)
{theme <- rasterTheme(region = region, ...); theme}
levelplot(pred, par.settings = coli)

# Reshape the table to fit nicely in the results
cv.res <- left_join(
  data.frame(kk$cv.results) %>%
    dplyr::select(Index, median, method) %>%
    tidyr::spread(Index, median),
  data.frame(kk$cv.results) %>%
    dplyr::filter(Index=="VEcv") %>%
    dplyr::select(VEcv.Q1, VEcv.Q3, method) )

# Reorder factor levels
cv.res$method <- factor(cv.res$method,
                         levels=c("tre2","GAMp","IDWo",
                                  "TPSp", "MKri","UKri"))

# Round
cv.res[,-c(1)] <- round(cv.res[,-c(1)],2)

# Make the log of the measures
cv.res$lMAE <- log1p(cv.res$MAE)
cv.res$lRMSE <- log1p(cv.res$RMSE)

# Classify VEcv
cv.res$VEcv.class <- cv.res$VEcv
cv.res$VEcv.class <- cut(cv.res$VEcv.class, c(-2,0,10, 30, 50, 80, 100))
levels(cv.res$VEcv.class) <-
  c("0. worst_then_mean",
    "1.very_poor",
    "2.poor",
    "3.average",
    "4.good",
    "5.excellent")

# Inverted VEcv, i.e. the bigger the worst:
cv.res$VEcv.inv <- abs(cv.res$VEcv/100-1)*100

# Estimate spatial indicators
# In this chunk we will estimate the spatial indicators using the sampled data and the interpolated data (prediction rasters).

# For this we will use the pre-packed functions in the packages 'Rgeostats' and 'ineq', although the indexes are relatively simple to calculate.

# Further, the function will also estimate the 'data limits integrity indicators'.

# This chunk estimates the difference in the spatial indicators between the raw data and interpolated surfaces and the data limits integrity indicators

# Note we require RGeostats and ineq for this chunk.

# Test the function in one case
fun.inter2(ii="tre2", dat=dat, pred=pred)

# Apply to all interpolation methods:
ind.res <- lapply(as.list(levels(cv.res$method)), fun.inter2, dat=dat, pred=pred)
ind.res <- do.call("bind_rows", ind.res)

# Merge the results with cv results:
tot.res <- full_join(cv.res, ind.res, by=c("method"))

# Now, all the indicators were estimated for each interpolation method. We shall then proceed to make the first step of the selection method.

# First selection step: VEcv interquantile range

# Reorder factor levels:
tot.res$method <- factor(tot.res$method,
                         levels=c("tre2","GAMp","IDWo",
                                  "TPSp", "MKri","UKri"))
tot.res$short.method <- ordered(tot.res$method,
labels=c("LM2","GAM","IDW","TPS","MKr","UKr"))


# Which methods have the VEcv higher than the lower Q3
tot.res <- tot.res %>%
  dplyr::select(-MAE, -R2, -RMSE) %>%
  # valid methods, i.e. within range of inter-quartile:
  dplyr::mutate(
    VEcv.criteria = c(VEcv.Q3 >= max(VEcv.Q1)))

# Plot the VEcv, Q1 and Q3 and respective criteria
tot.res %>%
  ggplot(aes(x=method, y=VEcv))+
  geom_point(aes(col=VEcv.criteria))+
  geom_crossbar(aes(ymin=VEcv.Q1, ymax=VEcv.Q3,y=VEcv, col=VEcv.criteria), alpha=.5, width=.5)+
  ggtitle("First step of interpolators selection")
```

```
# Second selection step: indicators

# select the data for the PCA of indicators
row.names(tot.res) = tot.res$short.method
sec.res <- tot.res %>%
  dplyr::filter(VEcv.criteria==TRUE) %>%
  dplyr::select(VEcv.inv, lMAE, lRMSE,
                CGdist, a.linertia, a.iso, a.Gini,
                a.pix_under, a.pix_over, a.mean_perc, a.over_perc)

# Use the function to plot the results and estimate the best method of the selection
PCbiplot(datpc=sec.res, x="PC1", y="PC2")



#######################
## Functions required
#######################



interp.dat_CV <- function(nam, dat, dat.grid,
                          CV=TRUE, plotit=TRUE,
                          replicate.cv = 10){
  require(ggplot2); require(RColorBrewer);  # ploting
  require(dplyr); require(tidyr);#ploting and manover
  require(raster); require(rasterVis); #plot and manipulation
  require(sp); # spatial data
  require(gstat); #kriging and idw
  require(fields) #tps
  require(mgcv); #gam
  #require(scales) #modeling
  theme_set(theme_bw(base_size = 9));

  # Arguments:
  # nam is the label code
  # dat data frame with x, y, vari (variable of interest) and Depth
  # dat.grid # predictions grid that we want to estimate. class SpatialPixels - sp
  # plotit: produce plots for each interpolation.

  # Start the function

  # make a raster stack to fill with interpolation predictions of the different models:
  dat.pred <- raster(dat.grid)
  dat.pred[] <- NA
  dat.pred <- stack(dat.pred)
  # make a dataframe to fit in the parameters
  model.params <- data.frame(code=as.character(nam))
```

```

## make the spatial object
dat.s <- dat
coordinates(dat.s) <- ~x+y
# the warning is due to the recent change to PROJ6
proj4string(dat.s) <- CRS("+init=epsg:28992")
proj4string(dat.grid) <- CRS("+init=epsg:28992")
#dat.border <- spTransform(dat.border, utm30)
#dat.pred <- stack(raster(dat.grid)) # obj to save the data

# for plotting
coli <- function (region = rev(brewer.pal(n = 10, 'Spectral')), ...)
{theme <- rasterTheme(region = region, ...); theme}

# function to make individual plots
fun.plot <- function(ii, dat.s){
  print(levelplot(dat.pred[[ii]]+.1, main=paste(ii, nam, round(max(dat$vari))),
                  zscaleLog=FALSE,
                  par.settings = coli))}

# Function to estimate error measures
fun.eval <- function(observed, predicted){
  resi <- c(observed- predicted)
  # rmse(sim=predicted, obs=observed)
  (RMSE <- sqrt(mean(resi^2)))
  #mae(sim=predicted, obs=observed)
  (MAE <- mean(abs(resi)))
  #(RMAE = MAE/mean(observed))
  #RMAE2 = mean(abs((kk$predicted-kk$observed)/mean(kk$observed)))*100
  #(RRMSE = RMSE/mean(observed))
  #RRMSE2 = sqrt(mean((kk$predicted-kk$observed)/mean(kk$observed)^2))*100
  # R2 should be 1-sum((kk$observed-kk$predicted)^2)/sum((kk$observed-
mean(kk$observed))^2)
  (R2 <- 1-(sum((resi)^2)/sum((observed-mean(observed))^2)))
  #1-(RMSE/sqrt(mean((kk$observed-mean(kk$observed))^2))))
  #(R3 <- 1-var(resi)/var(kk$observed)) #HENGL
  (VEcv <- (1 - sum((resi)^2)/
              sum((observed-mean(observed))^2))*100)
  res.error <- data.frame(RMSE=round(RMSE,2),
                          MAE=round(MAE,2),
                          #RMAE=round(RMAE,2), RRMSE=round(RRMSE,2),
                          R2=round(R2,2),
                          VEcv=round(VEcv,2))
  return(res.error)
}

# Function to make the cross validation and estimate error measures
cvfun.replicate <- function(xx, FUN, ii=ii, nam=nam, replicate.cv=replicate.cv){
  # xx is the data frame with x,y and biom..., FUN is the fun model of each method;
  # xx=dat.s; FUN=cv1.fun.cv; replicate.cv=10
```

```
    cv2.fun.fold <-function(xx, FUN){
      set.seed(seed <- as.integer(runif(1)*2e9))
      print(seed)
      kf <- sample(rep(seq_len(10), length.out=nrow(dat)))
      # Apply fun for the 10 folds
      kk <- lapply(as.list(sort(unique(kf))),
                   FUN = FUN, xx=xx, kf=kf) %>%
        dplyr::bind_rows()
      kk$seed=seed
      ## if we want to export predicted/observed
      #write.table(kk, append=TRUE,
      #           file = paste0(paste("pred.obs_cv1000", nam, ii, sep="_"), ".xls"),
      #           sep="\t", row.names=FALSE, col.names=FALSE)
      assign("last.warning", NULL, envir = baseenv())

      kk <- kk %>%
        dplyr::group_by(fold) %>%
        do(fun.eval(observed=.$observed, predicted=.$predicted)) %>%
        dplyr::filter(is.finite(VEcv)) %>%
        ungroup() %>%
        dplyr::summarise_all(funs(mean)) %>%
        dplyr::select(RMSE:VEcv) %>%
        data.frame()
      return(kk)
      rm(kf, kk, seed)
    }

    # to test cv2.fun.fold(FUN = cv1.fun.cv, xx=dat)
    # replicate CV 100 times
    xx1 <- replicate(replicate.cv, cv2.fun.fold(FUN = FUN, xx=xx), simplify = FALSE) %>%
      bind_rows %>%
      mutate(sp=nam, method=ii)
    xx1[mapply(is.infinite, xx1)] <- NA
    xx1 <- na.exclude(xx1)
    ## plot the distribution
    #print(xx1 %>%  tidyr::gather(Index, value, RMSE:VEcv) %>%
    #       ggplot(aes(x=value, group=Index, col=Index))+geom_density()+facet_wrap(~Index,
scales="free"))
    ## if we want to export the results:
    # write.table(xx1, file = paste("indices_cv1000", nam, ii, ".xls", sep="_"), sep="\t",
row.names=FALSE)

    # Get stats
    kk1 <- xx1 %>% tidyr::gather(Index, value, RMSE:VEcv) %>%
      dplyr::group_by(Index) %>%
      dplyr::summarize(VEcv.Q1=quantile (value, probs=0.25),
                       VEcv.Q3=quantile(value, probs=0.75),
                       mean=mean(value, na.rm=TRUE), N=n(),
                       median=median(value, na.rm=TRUE), N=n(),
                       max=max(value, na.rm=TRUE),
           min=min(value, na.rm=TRUE)) %>%
  dplyr::mutate(sp=nam, method=ii)
 return(kk1); rm(xx)
}


###################################
# 2nd order trend surface
###################################
ii <- "tre2"
dat.trend2 <- gstat(formula=vari ~ 1, data=dat.s, degree=2)
dat.trend2 <- predict(dat.trend2, newdata=dat.grid)
#spplot(dat.trend2[1], contour=TRUE,main="2nd order trend surface interpolation")
dat.pred[[ii]] <- raster(dat.trend2[1])

# Cross validation replicate
if(CV==TRUE){
 # function to do CV on each fold
 cv1.fun.cv = function(xx, k, kf){
  # Function to reproduce the interpolator
  # for a part of the data and predict with the other part
  # the output MUST be a dataframe with:
  # fold/observed/predicted
  kk <- gstat(formula=vari ~ 1, data=xx[kf != k,], degree=2)
  kk1 <- predict(kk, newdata=xx[kf == k,])
  return(data.frame(fold = k, observed = xx[kf == k,]$vari,
           predicted = kk1$var1.pred))
  rm(kk, kk1, k)
 }
 kk <- cvfun.replicate(xx=dat.s, FUN=cv1.fun.cv, ii=ii, nam=nam, replicate.cv=replicate.cv)
 cv.results <- kk;
}


# ###################################
# # GAM model
# ###################################
require(mgcv)
ii="GAMp"
dat.mod <- gam(vari~s(x,y, bs="ts"), data=dat)
kk <- data.frame(coordinates(dat.grid));
names(kk)= c("x","y")
dat.mod2 <- predict(dat.mod, newdata=kk)
kk <- cbind(kk, dat.mod2)
dat.pred[[ii]] <- rasterFromXYZ(kk)
# store parameters
model.params$R2.GAMp <- summary(dat.mod)$r.sq

# Cross validation replicate
if(CV==TRUE){
    # function to do CV on each fold
    cv1.fun.cv = function(xx, k, kf){
      kk <- gam(vari ~ s(x,y, bs="ts", k=50), data=xx[kf != k,])
      kk1 <- predict(kk, newdata=xx[kf == k,])
      return(data.frame(fold=k, observed=xx[kf == k,]$vari,
                        predicted=kk1))
      # kk <- gam(vari ~ s(x, y, bs="ts", k=50), data=xx[kf != k,])
      # #print(summary(kk)); print(plot(kk))
      # kk1 <- predict(kk, newdata=xx[kf == k,])
      # return(data.frame(fold=k, observed=xx[kf == k,]$vari, predicted=kk1))
      rm(kk, kk1, k)
    }
    # test: cv1.fun.cv(xx=dat, k=1, kf=kf)
    kk <- cvfun.replicate(xx=dat, FUN=cv1.fun.cv, ii=ii, nam=nam, replicate.cv=replicate.cv)
    print(head(kk))
    cv.results <- bind_rows(cv.results, kk);
    rm(kk)
  }

  rm(dat.mod, dat.mod2, ii)


  # ####################################
  # # Inverse distance weighting interpolation OPTIMIZED
  # ####################################
  ii="IDWo"
  RMSE <- function(observed, predicted) {
    sqrt(mean((predicted - observed)^2, na.rm=TRUE))}

  f1 <- function(x, test, train) {
    nmx <- x[1]
    idp <- x[2]
    if (nmx < 1) return(Inf)
    if (idp < .001) return(Inf)
    m <- gstat(formula=vari~1, locations=train, nmax=nmx, set=list(idp=idp))
    p <- predict(m, newdata=test, debug.level=0)$var1.pred
    RMSE(test$vari, p)
  }
  # set.seed(20150518)
  i <- sample(nrow(dat.s), 0.2 * nrow(dat.s))
  tst <- dat.s[i,]
  trn <- dat.s[-i,]
  opt <- optim(c(8, .5), f1, test=tst, train=trn)

  dat.idwopt <- gstat(formula=vari~1, locations=dat.s, nmax=opt$par[1],
set=list(idp=opt$par[2]))
  dat.idwopt <- raster::interpolate(raster(dat.grid), dat.idwopt)
  ## [inverse distance weighted interpolation]
  dat.idwopt <- mask(dat.idwopt, dat.grid)
  dat.pred[[ii]] <- dat.idwopt

  # Cross validation replicate
  if(CV==TRUE){
    # function to do CV on each fold
    cv1.fun.cv = function(xx, k, kf){
      kk <- gstat(formula=vari~1, locations=xx[kf != k,], nmax=opt$par[1],
set=list(idp=opt$par[2]))
      kk1 <- predict(kk, newdata=xx[kf == k,])
      return(data.frame(fold=k, observed=xx[kf == k,]$vari,
                        predicted=kk1$var1.pred))
      rm(kk, kk1, k)
    }
    # test: cv1.fun.cv(xx=dat.s, k=1, kf=kf)
    kk <- cvfun.replicate(xx=dat.s, FUN=cv1.fun.cv, ii=ii, nam=nam, replicate.cv=replicate.cv)
    print(head(kk))
    cv.results <- bind_rows(cv.results, kk); rm(kk)
  }
  rm(i,tst, trn, opt,f1)


  ####################################
  # TPS
  ####################################
  print(ii <- "TPSp")
  kk <- Tps(coordinates(dat.s), dat.s$vari)
  dat.tps <- raster::interpolate(raster(dat.grid), kk)
  dat.tps <- mask(dat.tps, dat.grid)
  dat.pred[[ii]] <- dat.tps

  # Cross validation replicate
  if(CV==TRUE){
    # function to do CV on each fold
    cv1.fun.cv = function(xx, k, kf){
      kk <- Tps(coordinates(xx[kf != k,]), xx[kf != k,]$vari)
      kk1 <- predict(kk, coordinates(xx[kf == k,]))
      return(data.frame(fold=k, observed=xx[kf == k,]$vari,
                        predicted=c(kk1)))
      rm(kk, kk1, k)
    }
    # test: cv1.fun.cv(xx=dat.s, k=1, kf=kf)
    kk <- cvfun.replicate(xx=dat.s, FUN=cv1.fun.cv, ii=ii, nam=nam, replicate.cv=replicate.cv)
    print(head(kk))
    cv.results <- bind_rows(cv.results, kk); rm(kk)
  }


  ####################################
  # Manual kriging
  ####################################
  print(ii <- "MKri")
  dat.vgm <- variogram(vari~1, dat.s)

  kk = vgm(psill = max(dat.vgm$gamma), model="Sph", range=max(dat.vgm$dist)/2,
nugget=min(dat.vgm$gamma))
  dat.fit <- fit.variogram(dat.vgm, model = kk)
  ## plot variogram with respective model
  #plot(dat.vgm, dat.fit)

  dat.krige <- krige(vari ~ 1, dat.s, dat.grid, model = dat.fit)
  #spplot(dat.krige[1])
  dat.pred[[ii]] <- raster(dat.krige[1])
  model.params <- cbind(model.params, "MKri"=data.frame(nug=dat.fit[1,2], sill=dat.fit[2,2],
range=dat.fit[2,3]))

  # Cross validation replicate
  if(CV==TRUE){
    # function to do CV on each fold
    cv1.fun.cv = function(xx, k, kf){
      kk <- krige(vari ~ 1, xx[kf != k,], xx[kf == k,], model = dat.fit)
      return(data.frame(fold=k, observed=xx[kf == k,]$vari,
                        predicted=kk$var1.pred))
      rm(kk, kk1, k)
    }
    # test: cv1.fun.cv(xx=dat.s, k=1, kf=kf)
    kk <- cvfun.replicate(xx=dat.s, FUN=cv1.fun.cv, ii=ii, nam=nam, replicate.cv=replicate.cv)
    print(head(kk))
    cv.results <- bind_rows(cv.results, kk);
    rm(kk)
  }


  ###################################
  # Universal kriging
  ###################################
  print(ii <- "UKri")
  dat.vgm <- variogram(vari~Depth, dat.s)

  kk = vgm(psill = max(dat.vgm$gamma), model="Sph", range=max(dat.vgm$dist)/2,
nugget=min(dat.vgm$gamma))
  dat.fit <- fit.variogram(dat.vgm, model = kk)
  #plot(dat.vgm, dat.fit)

  dat.krige <- krige(vari ~ 1, dat.s, dat.grid, model = dat.fit)
  #spplot(dat.krige[1])
  dat.pred[[ii]] <- raster(dat.krige[1])
  model.params <- cbind(model.params, "UKri"=data.frame(nug=dat.fit[1,2], sill=dat.fit[2,2],
range=dat.fit[2,3]))

  # Cross validation replicate
  if(CV==TRUE){
```

```
  # function to do CV on each fold
  cv1.fun.cv = function(xx, k, kf){
    kk <- krige(vari ~ Depth, xx[kf != k,], xx[kf == k,], model = dat.fit)
    return(data.frame(fold=k, observed=xx[kf == k,]$vari,
                      predicted=kk$var1.pred))
    rm(kk, kk1, k)
  }
  # test: cv1.fun.cv(xx=dat.s, k=1, kf=kf)
  kk <- cvfun.replicate(xx=dat.s, FUN=cv1.fun.cv, ii=ii, nam=nam, replicate.cv=replicate.cv)
  print(head(kk))
  cv.results <- bind_rows(cv.results, kk);
  rm(kk)
}


#################################
## FINAL PLOTS
#################################

## exclude Depth layer and project if wanted
# res <- projectRaster(dat.pred, crs=myCRS)
res = stack(dat.pred[[-1]])

if(plotit == TRUE){
  par(ask=TRUE)
  #samples
  p1<- qplot(data=dat, x=x, y=y, size=vari, col=vari, alpha=.0)+
    ggtitle(paste(nam,
                "max:", round(max(dat$vari)),
                "mean:", round(mean(dat$vari))))

  # maps
  p2 <- levelplot(res,
                main=paste(nam, round(max(dat$vari))),
                zscaleLog=FALSE,layout=c(6, 1),
                par.settings = coli)

  # log maps
  p3 <- levelplot(res+.1, main="Log", zscaleLog=TRUE,
                layout=c(6, 1),par.settings = coli)
  # scaled maps
  p4 <- levelplot(scale(res), main=paste('Scaled', nam),
                layout=c(6, 1),par.settings = coli)

  # histogram
  p5 <- histogram(res,
                xlim=c(0,max(dat$vari)))
  # density
  p6 <- densityplot(res,
                  xlim=c(0,max(dat$vari)))

    # boxplot
    p7 <- bwplot(res)

    grid.draw(grid.arrange(p2,p3,p4,
                           layout_matrix = rbind(c(1,1,1),c(2,2,2),c(3,3,3))))
    grid.draw(grid.arrange(p5,p1,p6,p7,
                           layout_matrix = rbind(c(1,1,1),c(2,3,4))))
    par(ask=FALSE)
  }

  return(list(res = res,
              cv.results = cv.results,
              model.params = model.params))

  assign("last.warning", NULL, envir = baseenv())
  # res <- tidyr::gather(data.frame(res), method, pred, -x,-y)
}

# to test
# kk <- interp.dat_CV(nam="zinc", dat=dat, dat.grid=dat.grid, CV=TRUE, plotit=TRUE)


# Function to estimate the spatial indicators:
fun.indicators <- function(dat) {
  ## function to be used in this chunk:
  require(RGeostats);require(ineq) # For Gini index
  dat <- data.frame(dat)
  names(dat) <- c("x","y","pred")

  if(max(dat$pred, na.rm=TRUE)!=0){
    dat$pred[dat$pred<0] = 0 #to avoid issues with center of gravity
    kk <- db.create(x1=dat$x, x2=dat$y, z1=dat$pred)
    kk2 <- SI.cgi(kk) # for the centre of gravity, inertia and isotropy-
    kk5 <- ineq(dat$pred, type="Gini")# gini index
    return(data.frame(t(unlist(kk2)[c(1,3,4,5)]), Gini=kk5))
  }}

# To test the function
# fun.indicators(dat = rasterToPoints(pred[[1]]))


# Function to apply the indicators function, estimate the difference between sampled and
interpolated and estimate the data limits intergrity measures:
fun.inter2 = function(ii, dat=dat, pred=pred){

  predi = rasterToPoints(pred[[ii]])

  # Estimate the indicators of interpolated
  res2 <- fun.indicators(predi)

  # Estimate the indicators for raw data
  res1 <- fun.indicators(dat[, c("x","y","vari")]) # 0.35 sec

  # Absulute difference between sampled and interpolated
  res <- abs(res1-res2)
  names(res)[c(1,2,5)] = paste0("a.",names(res)[c(1,2,5)])
  res$method <- ii

  # Diff of center of gravity
  res$CGdist <- spDistsN1(as.matrix(res1[,c("center1","center2")]),
as.matrix(res2[,c("center1","center2")]))

  # Rescale inertia
  res$a.linertia = log1p(res$a.inertia)

  # Get number of pixels values over max biom pred
  ll <- dim(predi)[1] # number of pixels
  mx <- max(dat$vari, na.rm=TRUE)
  mm <- mean(dat$vari, na.rm=TRUE)

  res$a.pix_under <-
    abs(ifelse(is.null(dim(predi[predi[,3]<0,])[1]),0,
                round(dim(predi[predi[,3]<0,])[1]/ll*100,2)))

  res$a.pix_over <-
    abs(ifelse(is.null(dim(predi[predi[,3]> mx,])[1]),0,
                round((dim(predi[predi[,3] > mx,])[1]/ll)*100,2)))

  res$a.mean_perc <- abs(round(c(mean(predi[,3], na.rm=TRUE)- mm)/mm * 100, 2))

  res$a.over_perc <- abs(round((max(predi[,3], na.rm=TRUE)-mx)/mx*100,2))

  res <- res[,c("method", "CGdist","a.linertia","a.iso","a.Gini",
                 "a.pix_under","a.pix_over","a.mean_perc","a.over_perc")]
  return(res)
}

# Function to make the PCA and estimate the best method according to the indicators
PCbiplot <- function(datpc=sec.res,
                      x="PC1", y="PC2") {
  require(ggrepel)

  # exclude indicators with zero only
  datpc = datpc[,colSums(datpc)!=0]

  # PCA
  PC <- prcomp(datpc, scale=TRUE, center=FALSE)
  #biplot(PC)
  data <- data.frame(winner2=row.names(PC$x),PC$x)
  datapc <- data.frame(varnames=rownames(PC$rotation), PC$rotation)

  mult <- min((max(data[,"PC2"]) -  min(data[,"PC2"])/(max(datapc[,"PC2"])-
min(datapc[,"PC2"]))),(max(data[,"PC1"]) - min(data[,"PC1"])/(max(datapc[,"PC1"])-
min(datapc[,"PC1"]))))
  datapc <- transform(datapc, v1 = .7 * mult * (get("PC1")),v2 = .7 * mult * (get("PC2")))
  dev <- paste0(c(round((((PC$sdev)^2 / sum(PC$sdev^2) )*100))[1:2],"%")

  # Get distance to center of each point:
  data$dist <- apply(data[,c("PC1","PC2")], 1,
                     function(x) {
                       (sqrt((x[1] - 0)^2+(x[2]-0)^2))})

  # Reverse weights, as the closer to zero the better:
  data$dist2 <- 1/data$dist

  # Classification and col of criteria
  col.ind <- data.frame(nam=row.names(datapc), class=1)
  col.ind[col.ind$nam %in% c("lMAE","lRMSE","VEcv.inv"),2]<-"Error";
  col.ind[col.ind$nam %in% c("CGdist","a.linertia","a.iso","a.Gini"),2]<- "Spatial";
  col.ind[col.ind$nam %in% c("a.pix_under","a.mean_perc","a.over_perc","a.pix_over"),2]<-
"Integrity"
  col.ind$nam=factor(col.ind$nam)
  col.ind$col=c("#5E4FA2","#3288BD","#66C2A5")[factor(col.ind$class)]
  #rev(brewer.pal(11, "Spectral"))

  plot1 <-
    ggplot(data, aes(x=PC1, y=PC2)) +
    geom_point(aes(col=dist2), size = 1, shape=16)+
    geom_text_repel(aes(label = winner2, size=dist2, color=dist2)) +
    scale_colour_gradient(high = "#9E0142", low = "#FDAE61")+
    geom_hline(aes(yintercept=0), size=.2, color=8, linetype=2) +
    geom_vline(aes(xintercept=0), size=.2, color=8, linetype=2)+
    xlim(extendrange(c(data$PC1,datapc$PC1))[1],0.01)+
    ylim((extendrange(c(data$PC2,datapc$PC2))))+
    # plot criteria:
    geom_text_repel(data=datapc, aes(x=v1, y=v2, label=varnames),size = 3,
segment.alpha=.5,
                    color=col.ind$col)+
    geom_segment(data=datapc, aes(x=0, y=0, xend=v1, yend=v2),
arrow=arrow(length=unit(0.2,"cm")), color=col.ind$col)+
    xlab(paste0("PC1 (",dev[1],")"))+
    ylab(paste0("PC2 (",dev[2],")"))+
    ggtitle("PCA of indicators")+
    theme(line = element_blank(),
          axis.text=element_blank(),
          axis.ticks=element_blank())+
    scale_size(range = c(3, 5))+
    guides(size=FALSE, fill=FALSE, col=FALSE)
  plot2 <-
```

```
    ggplot(data, aes(x=reorder(winner2,dist2), y=dist2)) +
    geom_bar(stat="identity", aes(fill=dist2),col="White", alpha=.8)+
    scale_fill_gradient(high = "#9E0142", low = "#FDAE61")+
    theme(line = element_blank(),
          axis.text.y=element_blank(),
          axis.ticks.y=element_blank(),
          axis.text.x=
            element_text(angle=90,hjust=1))+
    xlab("Interpolation methods")+
    ylab("Inv. dist. to center")+
    ggtitle(" ")+
    guides(size=FALSE, fill=FALSE, col=FALSE)

  grid.draw(arrangeGrob(plot1, plot2, ncol=2, widths = c(3/4,1/4)))

  # Result's table
  kk <- data %>%
    dplyr::rename("method" ="winner2") %>%
    dplyr::mutate(dist2 = round(dist2,2)) %>%
    dplyr::select(method, dist2) %>%
    dplyr::arrange(dist2)
  return(kk)
}
```