Analysing Temporal Variability in Spatial Distributions Using Min–Max Autocorrelation Factors: Sardine Eggs in the Bay of Biscay

Petitgas Pierre [1, *], Renard Didier [3], Desassis Nicolas [3], Huret Martin [2], Romagnan Jean-Baptiste [1], Doray Mathieu [1], Woillez Mathieu [2], Rivoirard Jacques [3]

[1] Research Unit EMH, IFREMER, Rue de l'Île d'Yeu, 44300, Nantes, France
[2] Research Unit STH, IFREMER, Pointe du Diable, 29280, Plouzané, France
[3] Centre de Géosciences, MINES ParisTech, PSL University, 35 Rue Saint Honoré, 77305, Fontainebleau, France

* Corresponding author : Pierre Petitgas, email address : pierre.petitgas@ifremer.fr

Abstract :

This paper presents a novel application of the geostatistical multivariate method known as min–max autocorrelation factors (MAFs) for analysing fisheries survey data in a space–time context. The method was used to map essential fish habitats and evaluate the variability in time of their occupancy. Research surveys at sea on marine fish stocks have been undertaken for several decades now. The data are time series of yearly maps of fish density, making it possible to analyse the space–time variability in fish spatial distributions. Space–time models are key to addressing conservation issues requiring the characterization of variability in habitat maps over time. Here, the variability in fisheries survey data series is decomposed in space and time to address these issues, using MAFs. MAFs were originally developed for noise removal in hyperspectral multivariate data and are obtained using a specific double principal components analysis. Here, MAFs were used to extract the most continuous spatial components that are consistent in time, together with the time series of their amplitudes. MAFs formed an empirical isofactorial model of the data, which served for kriging in each year using all available information across the data series. The approach was applied on the spawning distributions of sardine in the Bay of Biscay from 2000 to 2017. A multivariate approach for dealing with space–time data was adapted here, because the evolution in time was highly variable. Maps were classified using the amplitudes of the MAFs, and groups of typical distributions were identified, which showed different occurrence probabilities in different periods.

Keywords : MAF, Space–time, Habitat, Mapping, Sardine, Bay of Biscay

1 Introduction

The ecosystem approach to fisheries management intends to manage fisheries while conserving the ecosystems supporting them (Jennings 2004). In this context, conserving essential fish habitats does require to define the spatial limits of the particular areas to be protected. For that, a prior understanding of time variability in spatial distributions is needed, which calls for using space-time data analysis methods. For such analyses, an important source of information are research survey series at sea. Fisheries surveys have been undertaken for more than fifteen years now in European waters to assess marine resources each year. The surveys provide space-time data consisting of series of yearly spatial distributions of fish species.

To analyse variability of spatial distributions over time, a usual approach is to convert the density data into presence/absence and delineate the areas occupied recurrently or occasionally (Bellier et al. 2007). Presence/absence may also be defined by considering whether density values are greater than a particular threshold. Presence would refer to prevalence if the threshold is a small value or hotspots if it is a large one. In this paper, quantitative density values rather than presence/absence data are considered. A simple approach then is to map the different years on the same grid and compute mean and variance in time at each grid point. But results are often difficult to interpret because of over-dispersion and mean-variance relationships (Bellier et al. 2007). Another approach is to apply the method of empirical orthogonal functions (EOF) on the data organized for the purpose in matrix form, with grid points as columns and time dates as lines. The method is a principal components analysis (PCA) and results in estimating eigenvectors representing spatial patterns common to all time dates with principal components being their time-varying amplitudes. The EOF method has been applied on fisheries survey data series to extract spatial patterns consistent in time (Morfin et al. 2012; Petitgas et al. 2014). But as the first components in a PCA account for large parts of variance, the patterns identified are influenced by high values, which are often unstructured spatially and therefore the method is not fully adequate to estimate the structured components in a spatial distribution. To avoid this effect, min-max autocorrelation factors (Switzer and Green 1984) are considered here. MAFs are a double PCA method, which estimates common spatial components in a multivariate data set by ordering them from the most to the less autocorrelated, thus allowing to filter the latter. Further,

MAFs can be used for mapping by kriging as they are a multivariate geostatistical model. Thus, applied in a space-time context, MAFs can help map those areas used frequently in time as well as those showing more variability in their occupancy. To our knowledge, this is the first time that MAFs are used for that purpose in a space-time context. MAFs were applied here to analyse the variability of the spawning grounds of sardine, an important fish resource in the Bay of Biscay. Spawning grounds are essential fish habitats, representing the initial conditions of the drifts of eggs and larvae. And their variability thus influences the abundance of the new generation later recruiting to the stock. This paper presents the method, its application to analyse a space-time data set and provides guidelines for its application.

2 Methods

2.1 MAFs in a Space-Time Context

Min-max autocorrelation factor analysis (Switzer and Green 1984) is a multivariate method developed to filter noise in multi-channel spatial image data. The method is based on principal components analysis (PCA). MAF is designed to maximize the autocorrelation at a given lag Δ and extract the most continuous information from a multivariate data set. Consider a p-variable random field $Z(x)=(Z_1(x),...,Z_p(x))'$. Here, the index from *1* to *p* will represent time dates. The data matrix is organized with the locations as lines and the variables (time dates) as columns. The MAFs of $Z(x)$ are then obtained as follows:

1. Transform with a first PCA the original data $Z(x)$ into standardized principal components Y(x) with zero mean and unit variance;

2. Compute increments of the Y(x) for spatial lag Δ, Y(x)-Y(x+Δ), and their correlation matrix Σ_Δ;

3. Find the eigenvectors of Σ_Δ and project the Y(x) on them. Re-order these factors by increasing eigenvalue order. These are the MAFs $\chi(x)$.

By construction, the MAFs are defined with zero mean and unit variance. The eigenvalues of the MAFs correspond to the variance of the increments of the Y(x) (i.e., twice their variogram value at lag Δ). Because of the re-ordering, the first MAF has the highest autocorrelation (smallest variogram value at lag Δ), the second MAF the second highest and so one. Also by construction, increments of the MAFs at lag Δ are

4

uncorrelated, hence, MAFs are uncorrelated at lags 0 and Δ (the cross-variograms between MAFs equal zero at lags Δ and 0).

Similarly as in other methods such as for instance rotated PCA (Barnston and Livezey 1987, Dommenget and Latif 2002), MAFs are obtained by a double PCA approach, where the factors of the first PCA are optimized to comply with a particular criterion by using a second PCA. Here, the second PCA serves to order the factors by decreasing spatial continuity (increasing variogram value) at lag Δ.

In fisheries ecology, MAFs have been applied on multiple time series in one-dimension (Fujiwara and Mohr 2009; Woillez et al. 2009) to extract the most continuous common patterns in time, and in geographical space in two-dimensions (Petitgas et al. 2018) to perform kriging with a discrete approach. This study extends the application of MAFs to the analysis of space-time data. An R script is provided in Appendix A for computing MAFs on such data.

Here, the space-time data consist of the same variable repeatedly sampled at the same stations x in different years t_i, $(Z_{t1,x},\ldots,Z_{tp,x})$'. In this context the MAFs are insensitive to the ordering in time. The space-time decomposition obtained by MAFs write $Z(t,x) = m(t) + \sum_{i=1}^{p} c_i(t)\chi_i(x)$, where p is the number of years and also the total number of MAFs, $m(t)$ is the mean of Z at time t over space, $\chi_i(x)$ are the zero mean unit variance MAFs (time-invariant spatial components) and $c_i(t)$ are time-varying coefficients (amplitudes of MAFs).

When the MAFs are spatially uncorrelated at all lags (not just at lags 0 and Δ), they form an empirical isofactorial decomposition of the data (Rivoirard et al. 2014). In that case, the coefficients $c_i(t)$ are derived from the relationships $E[Z(t,x)\chi_i(x)] = c_i(t)E[\chi_i(x)^2]$ as the MAFs $\chi_i(x)$ and $\chi_j(x)$ $(i \neq j)$ are spatially uncorrelated. And the variance of $Z(t,x)$ over space at time t writes $Var_x[Z(t,x)](t) = \sum_{i=1}^{p} c_i(t)^2$, where the squared coefficient $c_i(t)^2$ with index i represents the spatial variance carried by the MAF of order i at time t. Lack of correlation at all lags was inspected by computing cross-variograms between MAFs.

2.2 Method Implementation

MAFs are computed at lag Δ. An obvious choice for Δ is the sampling mesh size (the inter-sample distance).

Switzer and Green (1984) suggest to combine lag differences from two main directions. Here, the sampling

design was anisotropic (Fig. 1) with 3 nautical miles (n.m.) for the inter-sample distance along transects and

12 n.m. for the inter-transect distance in the perpendicular direction. Thus sensitivity of the MAFs to

direction and lag was investigated. The differences $Y(x)-Y(x+\Delta)$ were computed for the two directions and

lags of the sampling design, that is, across transects (with Δ equal to 12 n.m.) and along transects (with Δ

equal to 3 n.m.). The resulting MAFs were compared by using their loadings on the original variables (the

years) (Fig. 2). The loadings were not well correlated for all MAFs and thus the choice was made to analyze

the data with MAFs computed at lag 3 n.m. along transects. The rationale was to avoid pooling differences

$Y(x)-Y(x+\Delta)$ that were too different, making MAFs less robust. Another choice to be made is the number of

MAFs to retain in the analysis. For that, two criteria are helpful. The first criterion relates to the spatial

structure of the MAFs. The high order MAFs correspond to noise in the spatial distributions as their

variograms are flat (nugget effects). One would like to consider those MAFs that are structured only to filter

the noise in the distributions. So the variogram value at first lag is a first criterion as well as the variogram

structure of each MAF. The second criterion is based on the spatial variance of Z. How many MAFs q

($q \le p$) are needed to approximate the variance of Z over space, $Var_x[Z(t,x)](t)$, in the different years? The

sum $\sum_{i=1}^{q} c_i(t)^2$ was computed for q increasing from 1 to p, in each year t. One would like to retain a

number q of MAFs to approximate with a high enough precision the spatial variance in all years. Also,

cross-validating the data by kriging using models with different MAF numbers was helpful to assess the

importance of the higher order MAFs in reproducing the data.

2.3 MAF-Based Classification of Maps

There was no or little correlation in time of the MAF amplitudes, meaning that maps close in time could

differ while maps distant in time could be alike. The coefficients $c_i(t)$ were used to compute a distance

between years in order to classify years and group maps that were alike. The distance between maps for

years u and v was $d_q^{\,2}(u,v) = \sum_{i=1}^{q} (c_i(u) - c_i(v))^2$. The clustering algorithm used Ward's criterion of

minimum variance within clusters. How clearly the clusters were separated depended on q, the number of MAFs used. The decrease in clustering height (the criterion on the dendrogram to form clusters) as a function of cluster rank was used to identify which number q of MAFs allowed for a clear distinction between groups of maps. The average map per group was then estimated. Average maps characterized typical scenarios for spatial distributions. Differences between average maps per cluster informed on habitat variability and scenario probability.

2.5 Spatio-Temporal Kriging with MAFs

As the MAFs formed an isofactorial decomposition of $Z(t,x)$, variable $Z(t, x_0)$ at known time t and unsampled location x_0, was estimated by kriging the first q MAFs separately and then linearly combining their kriged estimates

$$Z^*(t, x_0) = m(t) + \sum_{i=1}^{q} c_i(t)\, \chi^K(x_0)$$
.

Higher order MAFs ($i > q$) showing variograms with high nugget effects, were not considered for mapping the data series. Considering that they carried noise and not locally significant information, mapping without these components was then more robust (Petitgas et al. 2018). Kriging in a particular year relied on the spatially structured MAFs (the spatial patterns common to all years) and their time-varying amplitudes. The series of maps was thus coherently estimated across all years as kriging used the common structured components. Variography and kriging were performed using the RGeostats package in the R software (Renard et al. 2014). A moving rather than a unique neighborhood was used to constrain the kriging estimator to the local rather than the overall mean, resulting in the map following the local spatial variability in the data.

The performance of kriging with (the first q) MAFs was evaluated by cross-validation, that is, considering each sample as temporarily unknown and comparing the sample value to the one estimated by kriging. Following Rivoirard et al. (2013), the regression of the sampled values on the estimated ones was used to characterize whether values estimated in a given range were on average effectively observed in that range.

For comparison, cross-validation of the data was also performed using ordinary kriging and kriging with all MAFs.

3 Application

3.1 Sardine Egg Data

The data consisted of sardine egg counts along transect lines covering the French part of the Bay of Biscay shelf (Fig. 1). The data were collected in each year since 2000 during the survey PELGAS (Doray et al. 2018) in May when the fish were spawning. The survey collected data in the major ecosystem components to evaluate the pelagic fish resources together with their environment. The Fish eggs were collected at subsurface using the continuous underway fish egg sampler named CUFES (Checkley et al. 1997), which consists of a high-volume pump operated en route. Seawater was continuously pumped at 5m depth at a rate of approximately 500 l min^{-1}. Eggs were then concentrated and retained in a collector (315 μm mesh size). Samples were collected every 18 min (3 nautical miles at 10 knots, the sailing speed). Eggs were identified to species level for anchovy and sardine and counted on-board. Abundances were subsequently transformed to number of eggs per 10 m^3.

The sampling design was made of line transects perpendicular to the coast from ten meter depth to shelf-break, across the French part of the Bay of Biscay continental shelf. Transects were regularly spaced with an inter-transect distance of twelve nautical miles. The inter-sample distance along transects was three nautical miles.

Though the same transects were sailed each year their start and end were not exactly located at the same position and therefore the data locations were not positioned exactly at the same points in all years. To apply MAFs, the data in each year were migrated on one single sampling grid for all years. The grid was made of the set of transects where the start was the mean start position over the years and the number of samples per transect was the mean number of samples per transect spaced by three nautical miles. In each year, the grid node was attributed the nearest data sample value. The space-time data analyzed consisted of a grid of sample points with values in each year for a series of eighteen years, from 2000 to 2017 (Fig. 1).

The yearly means varied over one order of magnitude (from 10 to 90 eggs per 10 m^3) with no trend in time. Also, the yearly maps showed differences in their aggregation patterns (Table 1; Fig. 1). The larger the global mean, the greater was the maximum value and the smaller the slope of the curve Q(z) (smaller contribution of small values to the mean).

3.2 Computation of MAFs and Number of MAFs Retained

The sensitivity of MAFs to lag Δ and direction was analyzed by considering the MAF loadings on the original variables (Fig. 2). The loadings were correlated but with correlation coefficients varying between 0.4 and 0.8, meaning that the MAFs computed at the two different lags in the two directions were similar overall but differed slightly for particular MAFs. To avoid estimating MAFs from a too diverse pool of increments Y(x)-Y(x+Δ), MAFs were calculated at Δ = 3 n.m. (inter-sample distance along transects). The variogram values of MAFs at lag Δ (Fig. 3) showed progressive increase with increasing MAF rank, meaning that the variability could not be decomposed into a few factors only. The spatial continuity of MAFs beyond rank thirteen was lower (variogram value >0.6 while MAFs have unit variance) and the correlation range of these higher order MAFs was also shorter (see below). The number of MAFs needed to reproduce a given proportion (e.g., 0.8) of the spatial variance varied among years (Fig. 4), meaning that spatial patterns differed among years. In some years (2000, 2001, 2015) a small number of MAFs was sufficient while in other years (2006, 2011, 2012) a larger number was required. Thirteen MAFs were necessary to reproduce in each year, at least 80 percent of the spatial variance of the year.

3.2 Structural Analysis

The first five MAFs extracted clear spatial patterns at regional scale (Fig. 5) with long range structures and no nugget effect (Fig. 6; Table 2), while higher rank MAFs showed structure at a smaller scale and with high nugget effects. The spatial pattern in MAF1 corresponded to having less eggs at the shelf-break and on the northern shelf and more eggs on the southern shelf and off river Loire. The pattern in MAF2 showed less eggs at the coast and shelf-break, while that in MAF3 a patchy variation (more and less eggs) along the shelf-break and near the coast. The pattern in MAF4 corresponded to more eggs at shelf-break and less on

9

the shelf in the northern part. The pattern in MAF5 corresponded to less eggs in the southern part with a patchy variation in the northern part. Note that the sign of the variation (more or less) in the MAFs as well as in their amplitudes are interchangeable. In a given year, the sign of the variation in egg concentration relative to the mean will depend on the sign of the product $c_i(t)\chi_i(x)$.

The variograms of the MAFs were computed in the two directions of the sampling, along transects with a distance lag of 3 n.m. and across transects with a lag of 12 n.m.. No anisotropy was detected and thus isotropic variogram models were fitted. The variograms were modelled with a nugget effect and two spherical models (Table 2). The variogram models of MAF1 to MAF5 had no nugget effect (0.1), a medium range structure (20 to 30 n.m.) and a long range structure (60 to 70 n.m.). The variogram models of MAF6 to MAF13 had increasing nugget effect values (0.10 to 0.39), a shorter range (10 to 20 n.m.) and some had a long range structure (38 to 65 n.m.). The variogram models of MAF14 to MAF17 had a high nugget effect (0.40 – 0.60) and a short range structure (less than 12 n.m.) and that of MAF18 was a pure nugget effect. The first thirteen MAFs were well structured, which was consistent with the previous result indicating that they allowed to summarize the structure in the data. The variograms showed progressive destructuring, consistent with a typical MAF decomposition, where the variability modelled corresponds to a smaller and smaller spatial scale with increasing MAF rank.

Based on the inspection of the cross-variograms between MAFs (Fig. 6), lack of correlation at all lags between MAFs was adopted as a modelling choice. And the MAFs were considered to form an (empirical) isofactorial model of the data.

The amplitudes $c_i(t)$ did not show any significant linear trend in time (p-value > 0.05) for any MAF (Fig. 7) and their variograms did not show any correlation structure in time (not shown). Yet, grouping the yearly spatial distributions into clusters was investigated.

3.3 Typology of Spatial Distributions

The clearest clustering hierarchy was obtained with a distance between years involving the amplitudes of the first two MAFs (Fig. 8). The average distribution was estimated for each group (Fig. 9), where the yearly distributions were approximated with thirteen MAFs. In the group of maps G1, high egg concentrations

occurred at the shelf-break north of 45°N where as in the group of maps G2, these habitats were barely used for spawning and egg concentrations were higher in the coastal areas. Spawning distributions of type G1 had a higher probability in the period 2000-2008 while those of type G2 were more frequent in the period 2009-2017.

3.4 Mapping by Kriging with MAFs

To account for the anisotropy in the sampling, the kriging neighbourhood considered sampling points in the four quadrants and belonging to different transects. It had the following parameters: radius of 30 n.m., four samples at minimum and sixteen at maximum with a maximum of four samples by quadrant.

The mesh size of the kriging grid was 6 n.m. in latitude and longitude. The grid was defined within a polygon encompassing the survey design. The first thirteen MAFs were block-kriged separately and combined to derive one map per year (Fig. 10). The maps in the different years showed consistent patterns, probably due to the structure in the MAFs. The areas showing most variability across the series in the presence/absence of eggs were at shelf-break and coast.

3.5 Cross-Validation

Model performance was evaluated by cross-validating each data sample using the MAF models with thirteen MAFs and all MAFs, and ordinary kriging (OK). Because of the high variability in the data, rather than evaluate the model capability to reproduce the exact data values, the cross-validation considered re-estimating them in a given range. For that, the data were discretized in five classes (Table 3) chosen to contain each a sufficient number of samples. In each year, the kriged values belonging to a particular class were averaged and so were the observations for the same samples. The mean observed values were plotted versus the mean kriged values (Fig. 10). Observed and estimated values were aligned along the diagonal line, meaning that estimated and observed values were similar (belonged to similar classes). The correspondence was good with ordinary kriging (points were close to the diagonal line), except for low values, which were occasionally overestimated. In comparison, kriging with MAFs overestimated middle to low values (range 50 – 100 eggs per 10 m^3) in some years, probably because of the continuous structure of

11

the first MAFs being less adapted in particular years (some years being too different from others). Cross-validation was similar when kriging with all the eighteen MAFs or with the thirteen well-structured MAFs, meaning that on average the high order MAFs had little impact on the estimation.

4 Conclusions

The spatio-temporal data were structured in space but little in time. In this context, MAFs extracted time-consistent spatial patterns, which were ordered by decreasing smoothness in their spatial structure. Further, MAFs were assumed to be spatially uncorrelated at all lags and thus constituted an (empirical) isofactorial model of the data. In this model, the amplitudes associated with the MAFs could be estimated simply. The amplitudes allowed to compute a distance between spatial distributions and typical spatial distributions were identified from this highly variable data set. Excluding the high order MAFs filtered the less-structured components in the spatial distributions, which corresponded to a few high isolated values. The yearly maps produced by kriging with the isofactorial MAF model shared time-consistent spatial patterns. The information across years was thus used when kriging in each year.

The calculations required to choose the number of MAFs and the following guidelines were suggested for doing so. Two criteria were considered. The variogram structure of the MAFs was a first criterion allowing to filter the unstructured components in the spatial distribution. The spatial variance in each year was a second criterion as a sufficient number of MAFs was required to reproduce the spatial variance in each year. The two criteria agreed. Cross-validating the data with or without the higher order MAFs showed similar results, meaning the higher order MAFs carried mainly noise. Also, classifying the maps using the structured components only resulted in clear spatial patterns for each group of maps.

Another choice to be made was the spatial lag Δ and direction at which to compute the MAFs. The inter-sample distance was a logical choice. The important anisotropy in the sampling design imposed to analyse the sensitivity of the MAFs to lag Δ and direction. Sensitivity to direction and lag was evaluated by comparing the MAF loadings on the original variables. The fact that an isofactorial model fitted the data also showed that the MAFs were little sensitive to the spatial lag Δ because in such a model the MAFs are uncorrelated at all lags.

12

The analysis evidenced typical types of spatial distributions and revealed variability in the use of spawning habitats. Spawning depends on fish condition, fish age and length (more eggs are spawned by larger older fish and in better condition) as well as on environmental parameters (including temperature and food). The interpretation of the variability in the spatial distributions as revealed in the analysis will require further biological investigations, such as for instance, correlation of the MAFs or their amplitudes with population and environmental parameters.

References

Barnston A, Livezey R (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Month Weather Rev 115:1083-1126.

Bellier E, Planque B, Petitgas P (2007) Historical fluctuations in spawning location of anchovy (Engraulis encrasicolus) and sardine (Sardina pilchardus) in the Bay of Biscay during 1967–73 and 2000–2004. Fish Oceanogr 16:1-15.

Checkley D, Ortner P, Settle L, Cummings, S (1997) A continuous, underway fish egg sampler. Fish Oceanogr 6: 58–73.

Desassis N, Renard D (2013) Automatic variogram modeling by iterative least squares: univariate and multivariate cases. Math Geosci 45:453–470.

Doray M, Petitgas P, Romagnan J-B, Huret M, Duhamel E, Dupuy Ch, 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. Progress in Oceanography 166: 15-29.

Dommenget D, Latif M (2002) A cautionary note on the interpretation of EOFs. J Clim 15:216–225.

Fujiwara M, Mohr M (2009) Identifying environmental signals from population abundance data using multivariate time-series analysis. Oïkos 118:1712-1720.

Jennings S (2004) The ecosystem approach to fishery management: a significant step towards sustainable use of the marine environment? Mar Ecol Progr Series 274:279-282.

Morfin M, Fromentin J-M, Jadaud A, Bez N (2012) Spatio-temporal patterns of key exploited marine species in the northwestern Mediterranean sea. PLoS ONE 7(5):e37907.

Petitgas P, Doray M, Huret M, Mass, J, Woillez M (2014) Modelling the variability in fish spatial distributions over time with empirical orthogonal functions: anchovy in the Bay of Biscay. ICES J Mar Sci 71:2379–2389.

Petitgas P, Woillez M., Doray M., Rivoirard J (2018) Indicator-based geostatistical models for mapping fish survey data. Math Geosci 50:187-208.

Renard D, Bez N, Desassis N, Beucher H, Ors F, Laporte F (2014) RGeostats: the geostatistical package [version 11.1.1]. MINES ParisTech. Free download from: http://rgeostats.free.fr/.

Rivoirard J, Demange D, Freulon X, Lécureuil A, Bellot N (2013) A Top-cut model for deposits with heavy-tailed grade distribution. Math Geosci 45:967–982.

Rivoirard J, Freulon X, Demange C, Lecureuil A (2014) Kriging, indicators and non-linear geostatistics. J South Afr Inst Min Metall 114:1-6.

Switzer P, Green A (1984) Min/max autocorrelation factors for multivariate spatial imagery. Tech Rep 6, Dpt Statistics, Standford University CA.

Woillez M, Rivoirard J, Petitgas P (2009) Using min/max autocorrelation factors of survey-based indicators to follow the evolution of fish stocks in time. Aquat Liv Res 22:193-200.

Appendix A: Rscript to Compute MAFs on a Space-Time Data Set

```
maf.f <- function(data, xy, hmin, hmax) {

# Rscript to compute MAFs on a space-time data set

# Pierre Petitgas (Ifremer), Didier Renard and Nicolas Desassis (Mines-ParisTech), July 2018

# Requires library(ade4) for principle components analysis (PCA)

# data: Matrix (nx,nt) on which to apply PCA

#       nx (lines) are the spatial locations

#       nt (columns) are the time locations

# xy: Matrix of data coordinates [,1]=x, [,2]=y

# hmin, hmax: Minimum, maximum lag distance at which to compute MAFs

# Returns:

# eig= Eigenvalues

# x= MAFs (zero mean and unit variance)

# c1= Normed scores of data vectors on MAFs

### 1. Original data are transformed into normalized PCs (y)

prc1 = dudi.pca(data,center=T,scale=T,scannf=F,nf=dim(data)[2])

y = as.matrix(prc1$l1)    # Row normed scores

c1.1 = as.matrix(prc1$c1) # Normed scores

npc = dim(y)[2]

### 2. Increments of normalized PCs (datdif)

matdis=sqrt(outer(xy[,1],xy[,1],"-")^2+outer(xy[,2],xy[,2],"-")^2)

datdif=matrix(NA,nrow=sum(!(matdis<hmin | matdis>hmax)),ncol=npc)

for (i in 1:npc) {

dif.y = outer(y[,i],y[,i],"-")

dif.y[matdis<hmin | matdis>hmax]=NA

datdif[,i] = dif.y[! is.na(dif.y)]
```

```
}

datdif=datdif[apply(!is.na(datdif),1,all),]

### 3. PCA of increments, MAFs

prc2 = dudi.pca(datdif,center=F,scale=F,scannf=F,nf=npc)

c1.2 = as.matrix(prc2$c1)

eig.2 = as.matrix(prc2$eig[npc:1])

x = y %*% c1.2

x = x[,npc:1]

maf.c1 = c1.1 %*% c1.2

maf.c1 = maf.c1[,npc:1]

dimnames(x) = list(NULL,paste0("MAF",1:npc))

dimnames(maf.c1) = list(names(data),paste0("MAF",1:npc))

#

return(list(eig=eig.2, x=x, c1=maf.c1))

}
```

Table 1: Basic statistics per year of the data on the single sampling grid. Years are ranked by increasing mean value. Mean: arithmetic mean; CV: standard deviation; f0: frequency of zeroes; max: maximum value.

Year	Mean	CV	f0	max
2003	10.55	3.46	0.62	410
2015	24.06	2.42	0.24	466
2007	25.60	2.14	0.37	464
2017	37.44	2.39	0.35	757
2001	39.80	2.65	0.30	789
2016	41.26	2.45	0.43	848
2013	53.40	2.19	0.39	1152
2014	57.70	2.98	0.27	1942
2006	61.95	1.84	0.14	1280
2002	64.22	1.42	0.18	783
2009	65.28	1.91	0.41	754
2004	67.52	1.63	0.21	800
2008	70.17	3.26	0.39	2734
2012	71.22	2.89	0.49	1886
2010	71.64	2.23	0.31	1206
2000	75.79	1.46	0.12	728
2011	86.17	3.08	0.23	3513
2005	92.30	1.50	0.15	886

Table 2: Variogram models of the MAFs.

MAF rank	Nugget effect	Spherical 1 Range 1	Spherical 1 Sill 1	Spherical 2 Range 2	Spherical 2 Sill 2
1	0.00	22	0.51	69	0.49
2	0.00	31	1.00	-	-
3	0.00	22	0.72	65	0.28
4	0.00	13	0.52	70	0.48
5	0.00	15	0.71	61	0.29
6	0.10	10	0.45	38	0.45
7	0.10	18	0.90	-	-
8	0.15	21	0.73	58	0.12
9	0.18	18	0.82	-	-
10	0.18	12	0.54	58	0.28
11	0.26	15	0.61	65	0.13
12	0.38	18	0.62	-	-
13	0.39	19	0.61	-	-
14	0.40	9	0.60	-	-
15	0.40	8	0.60	-	-
16	0.50	6	0.50	-	-
17	0.60	12	0.40	-	-
18	1	-	-	-	-

Table 3: Frequency of samples in defined data classes, all data of all years pooled. Data units are sardine egg counts per 10 m^3.

Class	[0, 0.1[	[0.1, 10 [	[10, 50[	[50, 150[	>150
Frequency	0.312	0.246	0.194	0.136	0.113
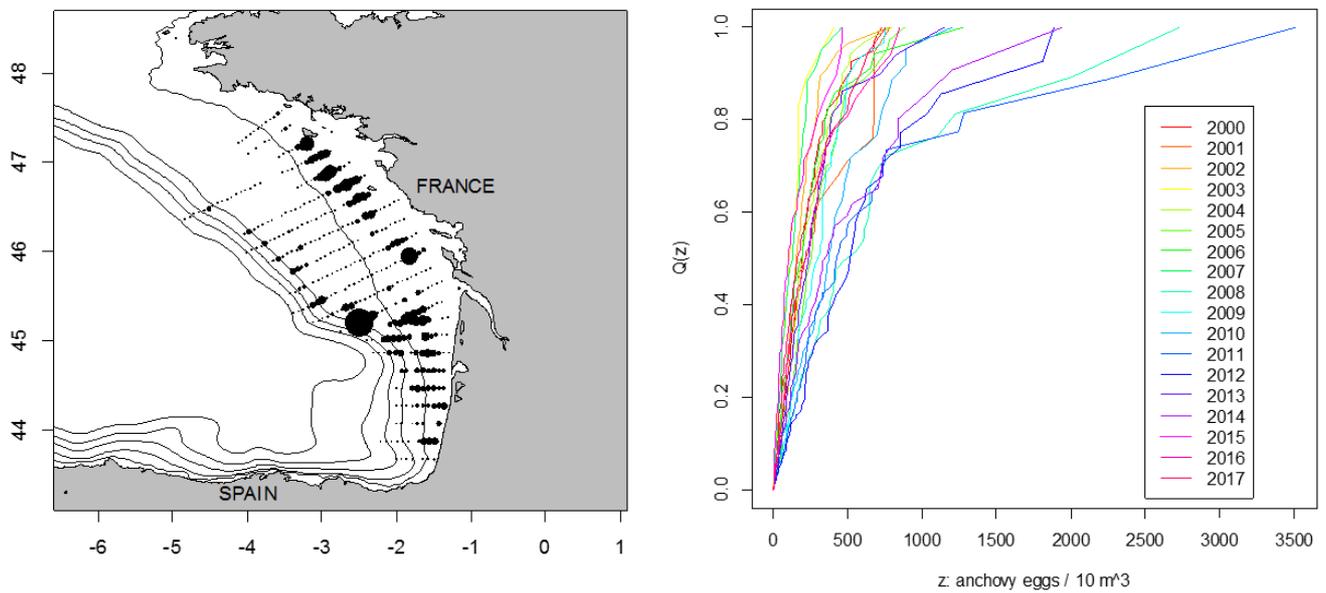
Fig. 1 Visualization of the sardine egg concentration (counts per 10 m³) data series, from 2000 to 2017. Right: design of the survey series showing data sample locations along transects and the average egg concentration. Left: selectivity curves Q(z) in each year, where Q(z) is the summed abundance proportion as a function of egg concentration value z
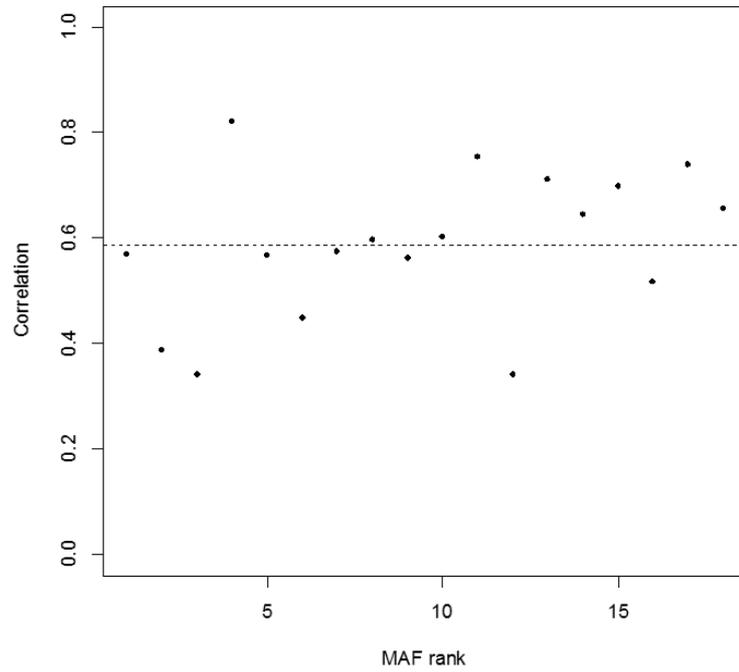
Fig. 2 Correlation between the MAF loadings when computed directionally along transects at 3 n.m. and across transects at 12 n.m.. The horizontal dashed line represents the mean correlation over all MAFs
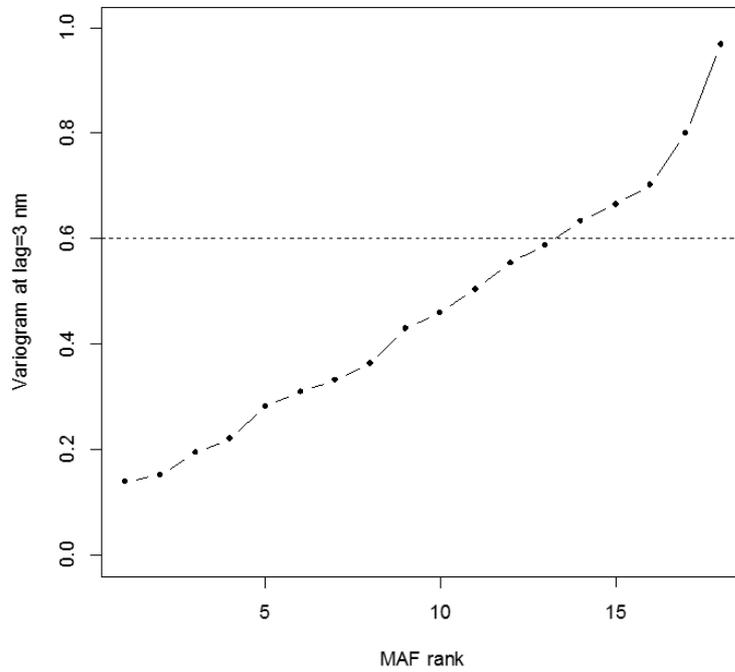
Fig. 3 Value at lag Δ (3 n.m.) of the variograms of MAFs
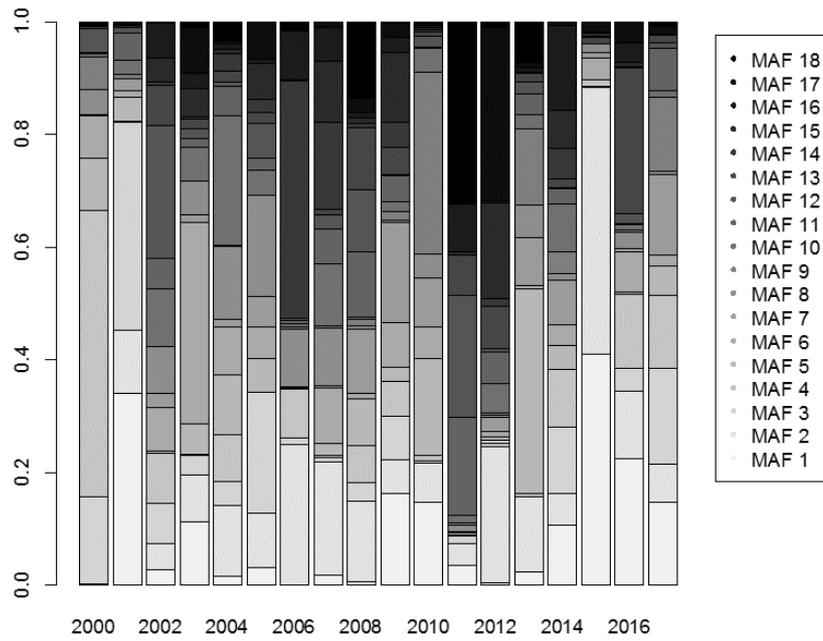
Fig. 4 Proportion of spatial variance summed for increasing MAF ranks, in each year
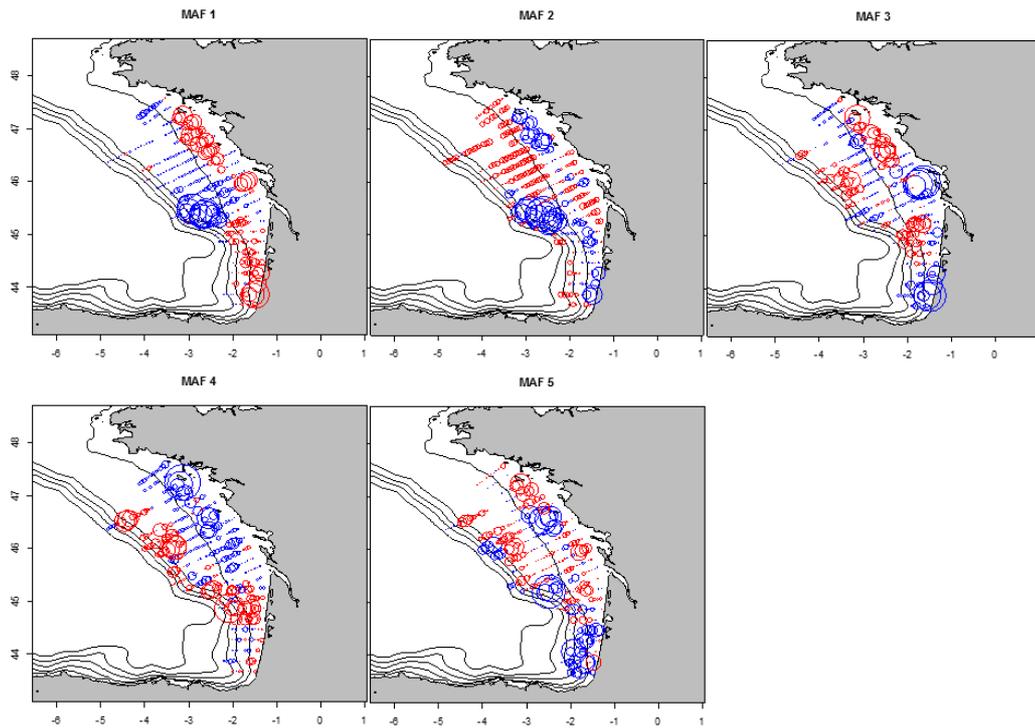
Fig. 5 Proportional representation of the first five MAFs. Each circle radius is proportional to the MAF value. Red: positive values. Blue: negative values
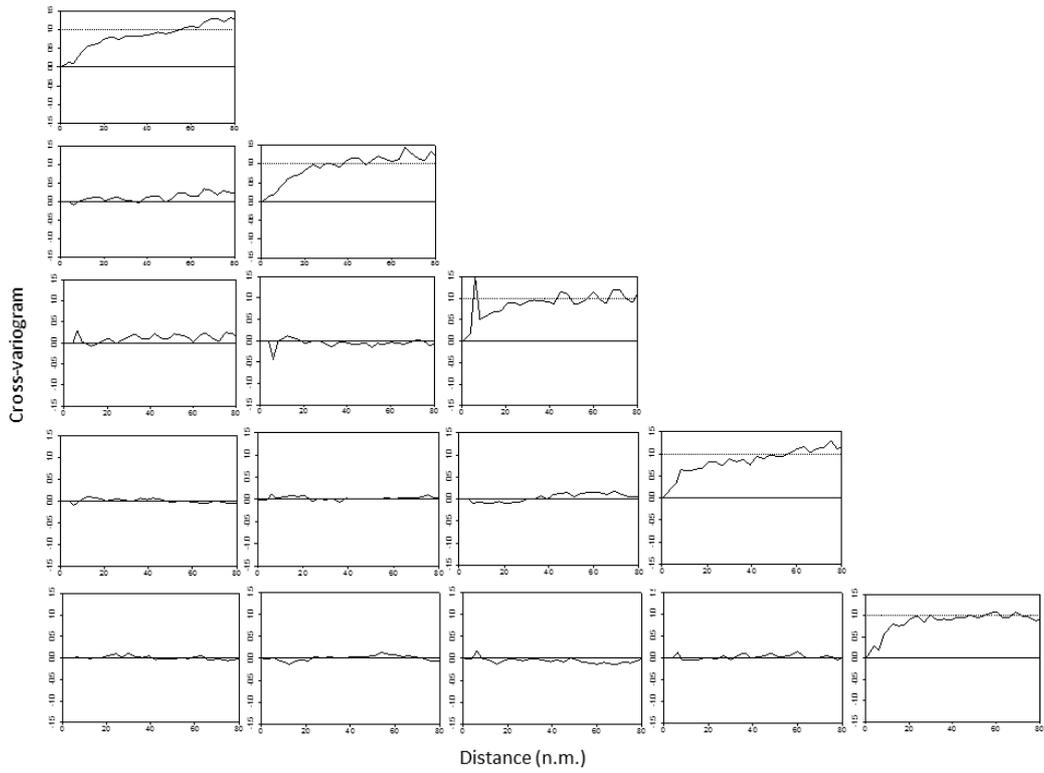
Fig. 6 Matrix of simple and cross-variograms for the first five MAFs. The variogram positioned at line i and column j is the cross-variogram between MAFs i and j. The dashed horizontal line on the simple variograms (diagonal of the matrix) represents unit variance. The horizontal line on the cross-variograms is drawn at the nil value (no correlation). The cross-variograms y-axes range from -1.5 to 1.5 and their x-axes from 0 to 80 n.m.
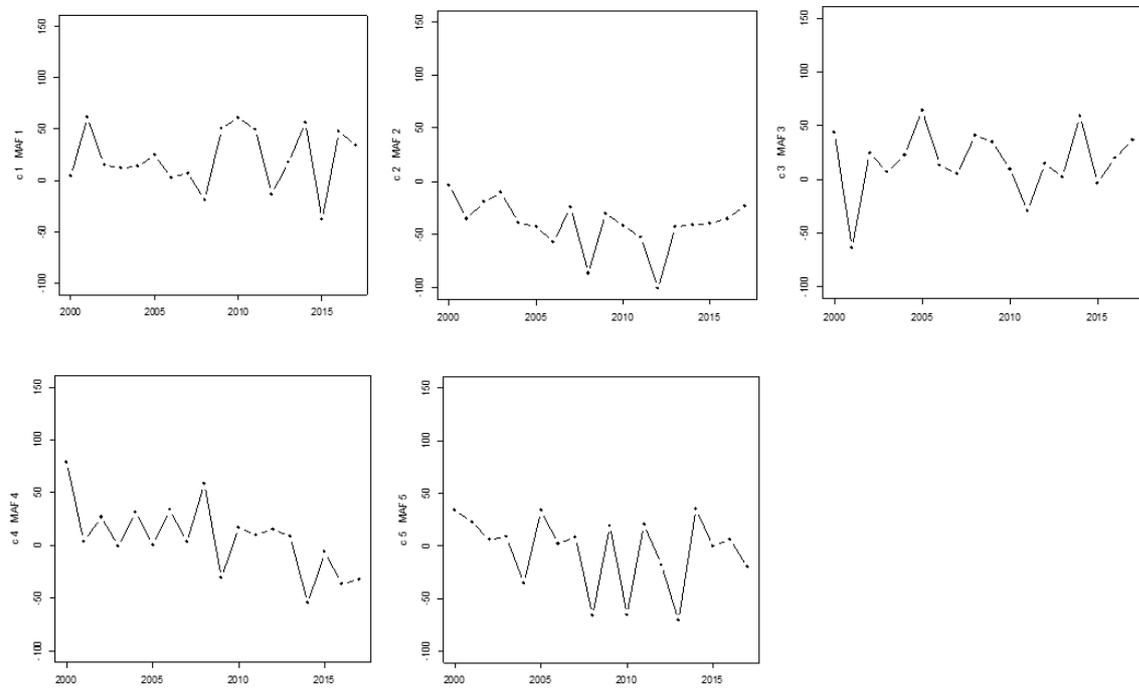
Fig. 7 Time series of amplitudes of the first five MAFs
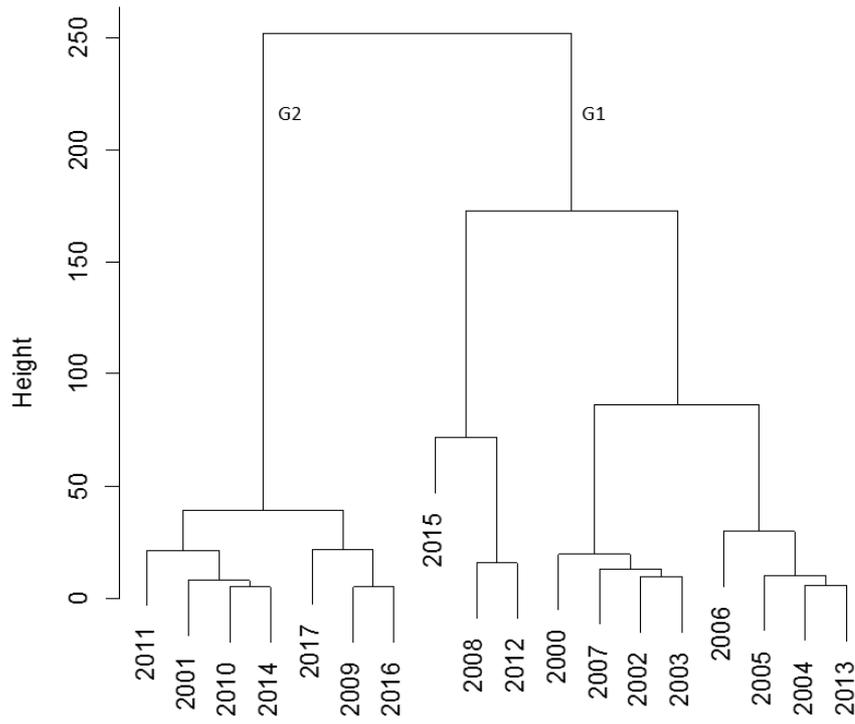
Fig. 8 Dendrogram of the clustering of spatial distributions obtained using the amplitudes of the first two
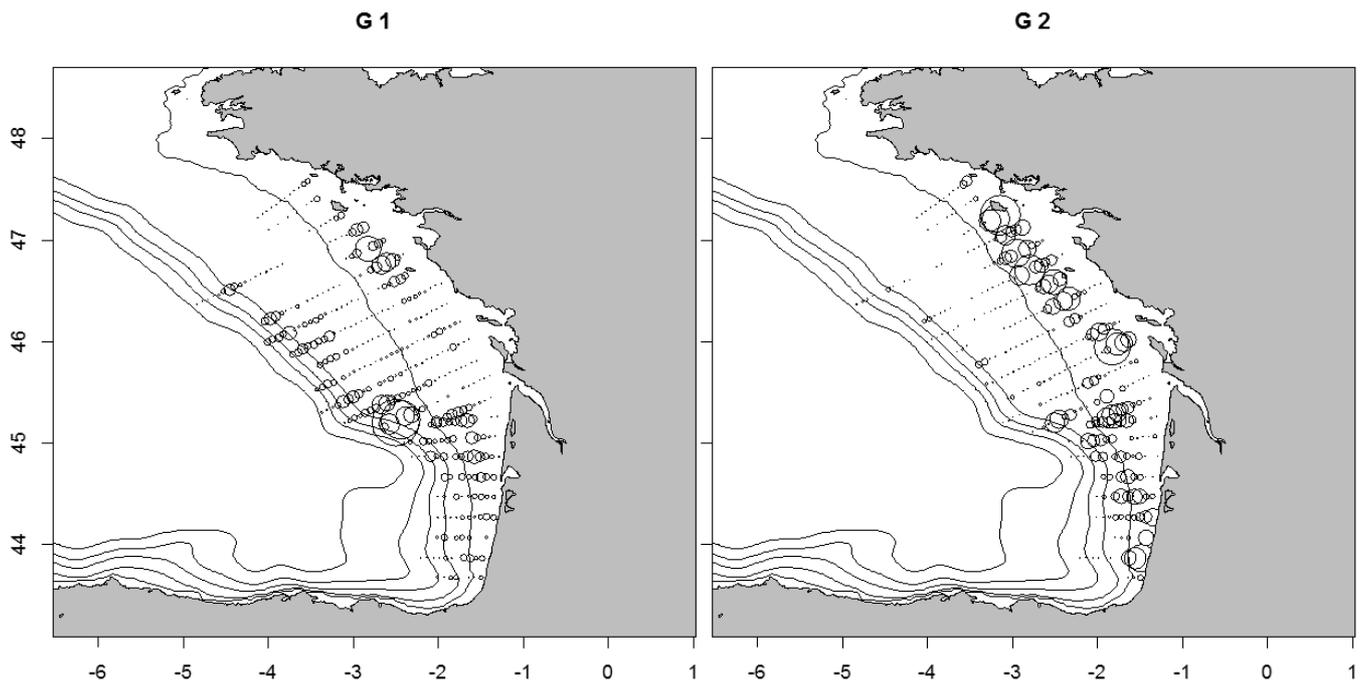
MAFs

Fig. 9 Average distributions of sardine egg concentration for the two categories of maps, as estimated with the first thirteen MAFs. The bubble radii are proportional to the egg concentration values. Years in group G1: 2000, 2002-08, and 2012, 2013, 2015. Years in group G2: 2001, 2009, and 2010-11, 2014, 2016-17
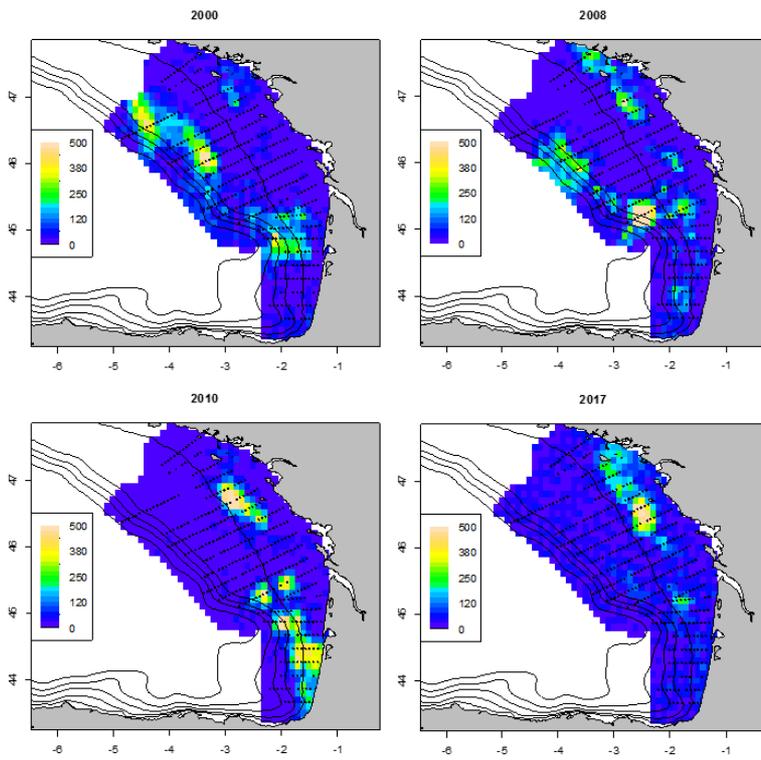
Fig. 10 Sardine egg concentration maps (counts per 10 m^3) obtained by kriging with the isofactorial MAF model in four typical years illustrating the two groups of spatial distribution. Above for group G1: years 2000, 2008. Below for group G2: years 2010, 2017
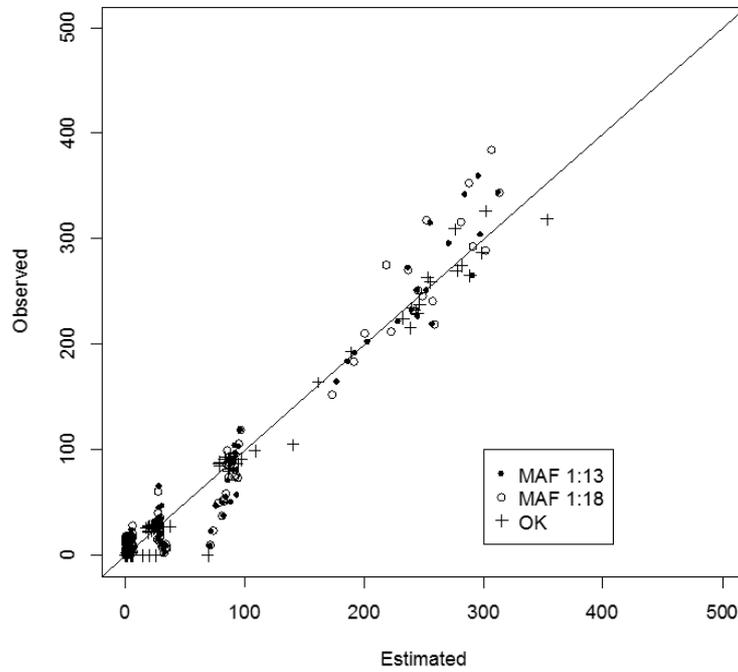
Fig. 11 Observed versus estimated mean values in each data class for the different kriging models. The line

is the first bisector. The models are: ordinary kriging (OK), MAF model with the structured components

only (MAF 1:13) and with all components (MAF 1:18)