Modelling the variability in fish spatial distributions over time with empirical orthogonal functions: anchovy in the Bay of Biscay

Petitgas Pierre [1, *], Doray Mathieu [1], Huret Martin [2], Masse Jacques [1], Woillez Mathieu [2]

[1] IFREMER, Département Écologie et Modèles pour l'Halieutique, rue de l'Île d'Yeu, 44300 Nantes, France
[2] IFREMER, Laboratoire de Biologie Halieutique, CS 10070, 29280, Plouzané, France

* Corresponding author : Pierre Petitgas, tel. +33 240 37 41 63 ;
email address : pierre.petitgas@ifremer.fr

Abstract :

Characterizing the space–time variability in spatial distributions as well as understanding its drivers is basic to designing robust spatial management plans. As a prerequisite, we analyse here how this variability relates to population dynamics in conjunction with environmental conditions. For that, spatio-temporal statistical approaches are needed but seldom used in fisheries science. To fill this gap, we showcase the usefulness of the method of empirical orthogonal functions (EOFs). Guidelines are given to apply the method on a series of gridded maps as derived from fisheries survey dataseries that now span over decades. The method is applied to the series, 2000–2012, of the spatial distributions of European anchovy in the Bay of Biscay at spawning time. Across the series, the EOF decomposition allowed to identify three main types of spatial distributions. One type corresponded to an extended distribution, another to a restricted distribution in core areas, and the third to a very coastal distribution. The coastal spawning distribution corresponded to a low population growth rate as it was never followed by a large recruitment in the subsequent year. We did not attempt to explain the spatial patterns per se but the drivers of change from one type of distribution to another. Stock size and fish size as well as bottom temperature and water column stratification were the covariates that controlled the variability in the spatial distributions over time. Further, the spatial distribution at spawning time related to recruitment in the following year, meaning that variability in the spatial distribution of spawning affected population dynamics. The typology of maps based on EOF decomposition summarized this spatial variability into spatial spawning configurations, which may serve spatial planning.

Keywords : anchovy, Bay of Biscay, density-dependence, EOF, habitat, spatial distribution

1. Introduction

Habitats represent the environmental conditions that are favourable for an organism (e.g., for its presence, growth) and thus habitat maps provide the space-time envelops of suitable conditions. Statistical regression has been widely used to model the habitats of the presence of species (Guisan and Zimmermann, 2000; Austin, 2007) or fish populations (Planque et al., 2011; Le Pape et al., 2014). But although closely related, habitats and spatial distributions are different concepts. Even if habitats are potentially suitable, their occupation will rely on the ability of the fish to colonize them with varying density. Thus to link habitats (suitable conditions) to fish spatial distributions, one needs to consider the mechanisms of habitat occupancy. These mechanisms involve factors internal to the population (e.g., abundance, demography, behaviour) as well as interactions in the ecosystem (e.g., trophic interactions, connectivity across the life cycle). We shall here focus on the former factors. Density-dependent habitat selection models have been used to explain how population spatial distributions vary with overall population abundance in different ways (Shepherd and Litvak, 2004; MacCall, 1990). Also physiological and behavioural mechanisms have been invoked to explain the re55 colonisation of past habitats during the rebuilding phase of a stock after its collapse (Petitgas et al., 2010). Thus we shall here consider that variability in spatial distributions results from variability in environmental conditions and internal population behaviour (Fig.1).

Further, we hypothesize that the spatial distribution of a population at spawning time is not independent of its demographic dynamics as it affects subsequent recruitment and therefore there is a feed-back (Fig. 1) implying that a particular spatial spawning

63   configuration is associated with a given recruitment regime. Such hypothesis is perhaps

64   more relevant for short lived species, which show greater variability in their spatial and

65   demographic dynamics. Indeed, sensitivity analyses using coupled physical-biological

66   models demonstrate the importance of initial spawning conditions on larval dispersion

67   and survival for European anchovy (Ospina-Alvarez et al., 2013; Huret et al., 2010). Here

68   the hypothesis is tested by identifying particular types of spawning spatial distributions in

69   fisheries survey data series. We suggest a space-time method for doing so, that extracts

70   principle spatial modes in the distribution. Further, we showcase how the method offers

71   the possibility to relate these principal spatial modes to covariates obtained at different

72   spatial resolutions.

73

74   To characterize and understand the variability of fish spatial distributions over time, this

75   study intends to showcase the usefulness of applying the space-time decomposition

76   method of Empirical Orthogonal Functions (EOF: Preisendorfer, 1988) on fisheries

77   survey data series. In contrast to habitat models where focus is on explaining the mean

78   distribution and its potential change with external drivers only, we here focus with EOFs

79   on characterizing the observed variability around the mean and explaining it with both

80   external and internal population drivers.

81

82   In a fisheries management context, characterizing the variability in spatial distributions

83   and understanding their consequences is important for at least two reasons. Marine

84   Protected Areas (MPA) are often designed based on habitats of particular life history

85   stages (Le Pape et al., 2014). However, variation in the spatial occupancy across years

86   may generate uncertainty in the temporal effectiveness of an MPA as the fish may

87   colonize other areas than the MPA (van Keeken et al. 2007). Further, indicators of spatial

88   distributions have been shown to relate to population parameters (e.g., recruitment,

89   demography, mortality) over a large range of stocks (Woillez et al. 2006). Thus the

90   characterization of how spatial distributions vary over time will add robustness to

91   population diagnostics as well as spatial management plans.

92

93   Empricial Orthogonal Functions have long been used in meteorology and physical

94   oceanography to decompose the time and space variability of a time series of maps.

95   Fisheries survey series now span more than ten years and thus offer sufficient space-time

96   information on the spatial distributions of fish populations to consider the use of EOFs

97   for analysing the variability in their spatial distributions. Here we apply this approach to

98   the time series of European anchovy (*Engraulis encrasicolus*) spatial distributions at

99   spawning time to extract the main features of variability in spatial occupancy. On that

100   basis, we identify major spatial configurations of the spawning population. We then relate

101   these to year class strength in the subsequent year. We also explain the spawning

102   configurations with population and environmental parameters. In doing so, we showcase

103   how EOFs provide a methodological framework to understand the ecology of population

104   dynamics in its spatio-temporal dimensions.

105

106   **Material and Methods**

107   **Method of Empirical of Orthogonal Functions**

108   The method of EOF (Preisendorfer 1988) is a particular Principal Component Analysis

109   (PCA) applied to a series of gridded maps, which allows to decompose the space-time

110   (residual) variability in the time series of maps into principal spatial modes and their

111   amplitudes. The decomposition is a linear factorisation of spatial components

112   (eigenvectors) that are constant in time and amplitudes (principal components) that are

113   variable in time. The variability around the mean map is thus modelled as the sum of

114   time-invariant spatial components that are weighted by their time-varying amplitudes :

115   $$Z(t,s) = \overline{Z}(.,s) + \sum_{m=1}^{q} U_m(t) E_m^T(s) \quad [1]$$

116   where:

117   $Z(t,s)$ is the variable under study at time t and spatial coordinate s,

118   $\overline{Z}(.,s)$ the time average at each coordinate s,

119   $E_m(s)$ the eigen vectors or EOFs (principal spatial modes) scaled to unity,

120   $U_m(t)$ the EOF amplitudes (principal components) scaled to $\sqrt{\lambda_m}$, where the λ_m are the

121   q non null eigen values associated with the EOFs

4

122

123   To achieve the decomposition, the method proceeds as follows. $Z(t,s)$ is a matrix

124   containing the gridded maps as line vectors with similar spatial order, thus having $t = 1,$

125    $..., N$ lines and $s = 1, ..., K$ columns. Each grid cell must be valued. In case of missing

126   information interpolation is needed or use of a coarser grid for data averaging. In each

127   grid cell, the cell time average is subtracted, which results in a matrix of anomalies on

128   which to perform the EOF decomposition: $X(t,s)=Z(t,s)-\overline{Z}(.,s)$. Matrix $S = X^T X /N$ is

129   then the covariance in space over time and matrix $Sa = XX^T /K$ the covariance in time

130   over space. A principal components analysis of matrix S (or equivalently Sa) leads to

131   computing the eigenvalues λ_m, eigenvectors $E_m(s)$ and principal components $U_m(t)$.

132

133   To retain the most meaningful EOFs and interpret their spatial patterns we used the eigen

134   values (overall variance accounted for by the components) and in addition the 'local'

135   explained variance (Schrum et al. 2006, Woillez et al. 2010). The 'local' explained

136   variance at location s associated with EOF of order m, $\eta_m(s)$, is the proportion of variance

137   across time that $Um(t)$ and $Em(s)$ explain at that location :

138    $\eta_m(s)=Var[Y_m](s)/\sum_m Var[Y_m](s)$; where $Y_m=U_m(t)E_m^T(s)$

139   When the map of local variance shows sub-regional patterns, the EOF explains variability

140   in these areas and a biological interpretation can be looked for. Further, when the patterns

141   in the EOF (higher/ lower values) can be superimposed on that in the local explained

142   variance, the EOF is dynamically relevant in time and the EOF decomposition well suited

143   to capture the space-time variability in the series of maps. Note that the sign of $Ym(t,s)$

144   depends on the combination of the EOF and its amplitude: it is positive when $Em(s)$ and

145    $Um(t)$ are of the same sign and negative when they are of opposite signs.

146

147   **Fish survey data**

148   The survey series considered was the yearly spring acoustic survey series PELGAS,

149   2000-2012, undertaken by IFREMER on board RV "Thalassa" over the French shelf of

150   the Bay of Biscay in May. The survey design is made of parallel transects, orientated

151   perpendicular to the isobaths and regularly spaced 12 nautical miles (n.m.) apart, from

152   43.5 N to 48.8 N and from coast (10 m depth) to the shelf break. Along the transects

153   38kHz acoustic records are collected continuously by day, at 10 knots (Doray et al.,

154   2010). Opportunistic pelagic trawl hauls are performed depending on the echotraces and

155   provide information on species proportions, length distributions, weight and age. During

156   night-time, conductivity–temperature–depth (CTD) profiles are performed on a regular

157   grid of stations (Fig. 2), providing measurements of environmental condition.

158

159   The anchovy population is surveyed in May at its peak spawning time and its distribution

160   is contained inside the surveyed area (ICES 2010). Europrean anchovy is mature at age 1

161   at its first spring and spawning starts in all length groups when surface temperature is

162   above 13 °C, corresponding in the Bay of Biscay to the onset of seasonal thermal

163   stratification (Motos, 1996). Thus the surveyed population corresponds to the spawning

164   adults.

165

166   In the Bay of Biscay, because of the multi-species context and variable schooling

167   characteristics (Massé et al., 1996), the school echotraces cannot be identified to species

168   from their acoustic properties alone with the echosounder currently in use for the

169   assessment. Echotraces are identified to species at a coarser spatial resolution based on

170   pelagic trawl hauls, which are performed over several n.m. to capture aggregations of

171   echotraces. The combination of the trawl haul data with the acoustic data allow to convert

172   acoustic backscatter into fish abundance by species (MacLennan and Simmonds, 2005,

173   section 9; Doray et al., 2010). The resulting data are abundance (tonnes) by species per

174   n.m.$^{-2}$ for each n.m. along the survey track. The survey series contains rare but very high

175   values, which may mask the regional patterns of variability. There is uncertainty in these

176   very high values because of uncertainty in allocating acoustic backscatter to species

177   (measurement error) and because extreme biological aggregation is rare and thus little

178   predictable. Thus, we followed a practice used in geoscience in the study of ore deposits

179   when measured concentrations exceed the capability of the measurement device

180   (Rivoirard et al., 2012). Values over a threshold, q, are truncated to that threshold (if

181   $Z(x,t)>q$, then $Z(x,t)=q$). The threshold considered here was 200 tonnes n.m.$^{-2}$, which

182   corresponded to reported maximum concentration of anchovy school aggregations (2

183   tonnes per school, 5 schools per km: Massé et al., 1996; Petitgas et al. 2001). Values

184   greater than 200 were rare with a frequency of 0.002 in the data set. We considered the

185   spatio-temporal series of these (truncated) data, with one survey per year, 2000-2012.

186

Block averaging

188   Prior to the EOF analysis, the data were averaged by block over a grid, which was the

189   same in all years. The grid mesh size selected was 0.4° in latitude and 0.4° in longitude,

190   with origin x_0 at 43°N and 6°W. The mean in block (i, j, x_0) was the simple average of the

191   data inside the block and it was positioned at the location of the block centre. The limits

192   of block (i, j, x_0) depend on the origin x_0 and therefore the samples involved in computing

193   the block mean. To decondition the block means from the grid origin, the point origin x_0

194   was randomized in the lowest left corner block one 100 times as in Petitgas et al. (2009).

195   At each randomization k, the grid origin x_k varies and the mean in block of rank (i, j) is

196   computed. Each block of rank (i, j) has then one 100 means associated to it. The mean of

197   all 100 means was then calculated and positioned at the centre of block (i, j, x_0). This

198   blocking procedure results in an implicit kernel-like interpolation where each data takes

199   part in the block average with a frequency (weight) depending on its distance to the block

200   centre. Finally, blocks which were not valued due to lack of samples in a given year were

201   omitted in the analysis as well as the blocks which had their centre point outside the

202   polygon defining the survey area (Fig. 2). Several trials with different mesh sizes were

203   undertaken and the mesh retained was the smallest one for which the first two principal

204   components explained more than 50% of total variance. The grid mesh size retained

205   allowed a reasonable compromise between enough fine scale smoothing and sufficient

206   large and mesoscale details. Finally, instead of biomass per cell we used the percentage

207   biomass: $P(t,s)=1000\,Z(t,s)/\sum_{s} Z(t,s)$. Matrix $P(t,s)$ was then centred by column to form

208   matrix X.

209

Typology of maps

211   The EOF decomposition being a Principal Components Analysis where the years are the

212   lines (individual observations) and the map grid cells the columns (variables), it is

213   possible to perform a hierarchical clustering of the years, applying a clustering procedure

214   on the matrix of distances between years in the factorial space of the first p components

215   ($p<K$). The clustering was performed using the Ward criterium (minimize the intra group

216   variance), which aims at finding compact groups. The average map in group g, $\overline{Z}_g(.,s)$,

217   was estimated by taking the simple average of maps Z for the years of group g. The

218   clustering resulted in defining types of spatial distributions thus summarizing the

219   meaningful variability in the series of maps.

220

221   **Explicative covariates**

222   In a classical habitat statistical approach (e.g., Le Pape et al., 2014) the fish distribution

223   map (response) and the maps of (explanatory) covariates are produced on the same spatial

224   grid and each point in the maps contribute to an overall correlation model linking fish

225   presence to covariates. Here, the EOF framework proposes another approach. The EOF

226   patterns of variability are defined empirically from the data and are invariant in time. The

227   variation over time in the spatial distribution is taken in charge by the variations in the

228   weights of the EOF patterns as given by the EOF amplitude time series. We thus focussed

229   on explaining the EOF amplitudes time series with covariates. Based on our conceptual

230   approach (Fig. 1) covariates considered were population parameters affecting population

231   behaviour via density-dependent processes and hydrological conditions affecting habitat

232   suitability.

233

234   *Population parameters*. Following Cotter et al. (2009), we considered three indices to

235   characterize population status: total (spawning) biomass and two length distribution

236   percentiles. Population biomass was estimated from the survey as reported to ICES

237   (2012). The length distribution percentiles were the 25% and 75% percentiles. The length

238   distribution in the population was estimated as follows. The (local) frequency distribution

239   of length in a trawl haul was weighted by fish biomass along the acoustic transects within

240   a radius of 10 n.m. from the traul haul position. The population length distribution was

241   the weighted average of all local distributions.

242

243   *Hydrological indices*. Hydrological condition was characterized using indices calculated

244   from the CTD profiles at the stations (Fig. 2). We considered 4 indices (Table 1) to

245   characterize typical features on the shelf in spring (Huret et al., 2013): surface and bottom

246   temperature, a water column stratification index and a river plume index. These were

247   calculated following the procedures detailed in Huret et al. (2013). Surface values were

248   taken at 7 m depth. Bottom values were 5 m above bottom. The deficit of potential

249   energy, *Dep*, is the energy required to homogeneize the water column. The greater the

250   *Dep* value the greater the stratification. It was calculated from surface to 60 m depth for

251   the profiles where bottom depth exceeded 60 m. The equivalent fresh water height, H_{fw},

252   measures the height of accumulated fresh water considering a reference sea water salinity

253   (35.85 psu: mean bottom salinity). In comparison to surface salinity, H_{fw} is less affected

254   by vertical mixing and thus reflects better the past history of river discharge over a few

255   months. The indices were calculated at each station and the overall spatial mean was

256   calculated to obtain time series of mean indices.

257

258   *Correlation between amplitudes of EOFs and covariates*. The time series of each

259   amplitude $U_m(t)$ associated with EOF of order m was linearly regressed on that of the

$$U_m(t)=\sum_{j=0}^{p} a_j X_j(t)$$

260   different covariates $X_j(t)$:   ; where $X_0(t)=1$. For each m, the most

261   probable model was selected using Akaike's information criterium (AIC: Burnham and

262   Anderson, 2002). For each *m*, we considered the 127 possible models given the set of the

263   seven covariates. For each *m*, the model retained was that with the lowest AIC value. To

264   identify the most explanatory covariates (i.e., those involved in the most probable

265   models) we computed their relative importance weights. For that, each covariate *j*

266   involved in model *i* explaining amplitude Um was attributed the model probability $p_m(i)$

267   as deduced from its AIC value or zero if model *i* did not involve the covariate. The

268   relative importance weight of covariate *j* in explaining amplitude *Um* was the sum over

269   all model probabilities: $\sum_{i=1}^{127} p_m(i)$, where $p_m(i)=\exp[-0.5\Delta(i,m)]/\sum_{i=1}^{127}\exp[-0.5\Delta(i,m)]$ with

270     $\Delta(i,m)=AIC(i,m)-\min(AIC(i,m))$. The relative importance weight varies between 0 and 1.

271   Typically, strong explanatory covariates will have a relative importance weight around

272   0.9, moderatly explanatory between 0.6 - 0.9 and weakly explanatory between 0.5 - 0.6.

273   Below 0.5 covariates will often be little relevant.

274

Results

Patterns of variability

The average distribution over all years (Fig. 3) shows anchovy to be mainly located south of 46°N with low abundance in the north mainly in coastal areas. The major concentration is located off the Gironde estuary (45-46°N, 1.5-2°W), along the coast and on the shelf off Landes. The principal spatial modes of yearly variability around the mean were extracted using the EOF decomposition technique. Four principal components were retained (*m=4*). Each explained more than 10% of total variance (Table 2) and showed a strong spatial pattern in their EOF as well as in the map of local explained variance (Figs. 4 and 5). The first EOF (Fig. 4) captured the variability along the coast of Landes and off Gironde estuary: the lower the abundance along the coast of Landes, the higher the concentration off Gironde. The second EOF (Fig. 4) captured a large scale pattern where less abundance in the south was associated to more abundance in the north. More specifically, EOF 2 captured the variability at the shelf break off Landes, the central part of the shelf off Gironde and Bretagne: the lower the abundance at the coast off Gironde and along the shelf break off Landes, the greater it was in Bretagne and in the central part of the shelf off Gironde. The third EOF (Fig. 5) captured the variability along the coast of Vendée and also in the area where the shelf break is curved (45°N, 2°W): the lower the abundance on the shelf off Landes, the greater it was along the coast of Vendée. The fourth EOF (Fig. 5) captured the variability along the shelf break north of 45°N: the lower the abundance at the coast off Vendée and at 45°N, the higher at the shelf break.

Typology of maps and their relationship with recruitment

Using the first four principal components associated with the EOFs described previously, three groups of maps were identified by hierarchical clustering (Fig. 6). The differences among the average maps of each group denoted strong differences in the spatial distributions over the years (Fig. 7). Group 1 (Fig. 7: maps in years 2000, 2001, 2008, 2009, 2011, 2012) corresponded to the largest spatial extension. The distribution extended in the northern part, on the shelf break north of 45°N and along the coast of Vendée and Bretagne although the largest concentration was off Gironde and on the shelf

10

305   of Landes. In constrast, Group 2 (Fig. 7: maps in years 2002, 2005, 2010) corresponded

306   to the smallest spatial extension and showed higher concentrations in 2 areas, off Gironde

307   and at the shelf break off Landes. Group 3 (Fig. 7: maps in years 2003, 2004, 2006, 2007)

308   corrresponded to a coastal spatial distribution with high concentrations along the coast

309   south of 46°N. This typology of spatial patterns at spawning time related to biomass in

310   the current year and to recruitment (age 1 fish) in the subsequent year (Fig. 8): very

311   coastal spawning distributions of type G3 occurred when biomass was low only and at

312   low biomass level, spawning distributions G3 and G2, which had smaller spatial

313   extension than G1, were never followed by high recruitment.

314

315   **Correlates of amplitudes of EOFs**

316   Based on the AIC criteria, the selected models showed high R-square ranging 0.48 – 0.70

317   (Table 3) and made sense biologically. Temperature alone explained amplitude of EOF 4,

318   which characterized variability along the shelf break. By contrast amplitude of EOF 3

319   (which characterized variability along the coast off Vendée) was explained by population

320   parameters only. EOF amplitudes 1 and 2 were explained by a combination of population

321   and environmental indices. Population biomass intervened as covariate of the amplitudes

322   of the three first EOFs, meaning that variability in the distributions was strongly density-

323   dependent. Also, the length distribution influenced the amplitudes of EOFs 1 and 3,

324   which both involved variability in coastal waters. Bottom temperature and water column

325   stratification were covariates of the amplitudes of EOFs 1 and 2, which both involved

326   expansion to the north and variability on mid-shelf. It is noteworthy that the river plume

327   index was never selected as a covariate. The relative importance weights of the covariates

328   (Table 4) confirmed that the covariates in the models selected were the most explanatory

329   ones. Yet, for the amplitude of EOF1, the most probable model (Table 3) involved the

330   stratification index (Dep) which had a low importance weight (Table 4). The model

331   without the stratification index (i.e., with SSB, q75, and Tb only) was ranked second

332   most probable with a probability of 0.85 and thus was nearly as probable as the retained

333   model (with SSB, q75, Tb, and Dep: Table 3). The stratification index played a slightly

334   minor role relative to the other three covariates in explaining the amplitude of EOF1. For

335   the amplitude of EOF3, the most probable (Table 3) model involved SSB which also had

336　a low importance weight (Table 4). The model without SSB (with q25 and q75 only:

337　Table 3) was ranked second best model but with the low probability of 0.49, meaning that

338　SSB should nevertheless be retained as a covariate. In all, Biomass (SSB), proportion of

339　large fish (third quartile q75), bottom temperature (Tb) and water column stratification

340　(Dep) were the most explanatory covariates of the changes over time in the spatial

341　distributions.

342

343　**Discussion**

344　The EOF decomposition characterized the variability in the spatial distributions over time

345　by extracting time-invariant principal spatial modes and their time-varying amplitudes.

346　The EOF decomposition served to classify maps. The typology of the spawning spatial

347　distributions related to population subsequent recruitment. The time series of amplitudes

348　were explained by population parameters (abundance, length distribution) and

349　environmental conditions (bottom temperature, water column stratification). Thus

350　changes in the spatial distribution were modelled depending on how the time-varying

351　amplitudes (weights of the principal EOF patterns) varied with covariates.

352

353　*How meaningful are EOFs*. The EOFs characterize spatial patterns of variability around

354　the mean map that are constant in time. It is their amplitudes that vary over time not the

355　EOFs. Such mathematical decomposition is not always suited to characterize complex

356　natural space-time variability and we here discuss how to acknowledge their suitability to

357　the case study. The EOF decomposition is little adapted to situations where areas of

358　variability change location over time. The local explained variance (Schrum et al., 2006)

359　is a way to test for that. If the local explained variance displays similar spatial patterns

360　than the EOFs, the EOFs are then dynamically relevant in time. Here, EOFs and their

361　local explained variance showed similar maps. Another point of consideration is that the

362　mathematical property of spatial orthogonality in the EOF decomposition may generate

363　artifactual patterns in the spatial modes with little physical interpretation, such as, e.g.,

364　dipole structures (Dommenget and Latif, 2002). It is therefore key to be able to critically

365　interpret the spatial patterns in the EOFs by relating them to physical or biological

366　phenomena. To increase interpretation, orthogonal (Varimax) rotation of EOFs has been

367   proposed to extract more locally defined spatial patterns (e.g., Richman 1986). But this

368   approach has its limitations: the ability to extract global spatial patterns in the modes is

369   lower (Dommenget & Latif 2002) and the rotation may introduce correlation between

370   amplitudes of rotated modes (Mestas-Nunez 2000) and thus reduce the ability to

371   reconstruct the data and classify maps. Here, the EOFs identified local areas that were

372   meaningful for the stock (coast, Gironde, shelf-break), making unecessary to find more

373   localised patterns by rotation. The EOF decomposition was thus considered suitable here

374   to characterize the space-time variability in the series of anchovy spawning spatial

375   distributions.

376

377   *Data smoothing on a grid prior to the EOF decomposition.* EOFs are best interpretable

378   when their spatial patterns develop over several grid cells at regional or sub-regional

379   scale. The grid mesh size over which to block average the data should thus be adequate.

380   Here, we used the amount of variance in the EOFs to define the grid mesh size. The scale

381   in the EOF patterns was greater than 1°Latitude x 1°Longitude, which was compatible

382   with the meso-scale / subregional aggregation pattern of schools in the area (Petitgas,

383   2003). The data set contained rare extreme values, which were uncertain given current

384   knowledge of schooling aggregation in the area (Petitgas et al., 2001). To deal with them,

385   the data were truncated. The truncation threshold was not defined statistically but based

386   on known schooling behaviour, similarly as for filtering unreliable information. Morfin et

387   al. (2012) took a different approach to study the variability of spatial patterns: (i) they

388   worked on log-transformed data to diminishing the influence of high values; (ii) they

389   interpolated the transformed data by kriging on a grid of small mesh size, which amounts

390   to smoothing. In contrast, our analysis provides guidelines to work on the raw data and

391   makes full use of the local explained variance to interpret the EOFs and their dynamics in

392   time.

393

394   *Covariates.* The regression of EOF amplitudes on series of explicative covariates was a

395   flexible approach as it allowed to relate spatial patterns at regional and sub-regional

396   scales (EOFs) in the fish distribution to covariates obtained at other spatial resolutions

397   (e.g., population scale). The three major patterns of variability in the spatial distribution

398   (Table 3) amounted to range expansion to northern Biscay, coastal distribution,

399   aggregation off Gironde estuary. Their relative importance was controlled by a

400   combination of environmental conditions and population parameters. Population biomass,

401   population length distribution, bottom temperature and water-column stratification were

402   influential in making the spatial distributions vary, which agrees with the literature.

403   Range expansion with population abundance has been observed in many fish stocks

404   (Shepherd and Litvack, 2004) and was originally observed on anchovy in the California

405   current (MacCall, 1990). Also, anchovy in the bay of Biscay shows a gradient in length

406   from coast to off-shore (Petitgas et al., 2003), the smaller fish being more coastal. It is

407   therefore not surprising that the anchovy distribution varies with population abundance

408   and length distribution and our analysis models how this happens. Also bottom

409   temperature and water column stratification were influential. During day-time, anchovy

410   forms schools well below the thermocline generally 10-20 m above the bottom (Massé,

411   1996), which may explain why bottom temperature may influence habitat suitability and

412   thus anchovy spatial distribution. Water column stratification is under the influence of

413   river plumes and warming surface temperature. River plumes are related to early season

414   plankton production (Labry et al., 2001) and warming surface temperature triggers

415   spawning (Motos, 1996) that occurs at night close to surface. Water column stratification

416   is thus also naturally involved in determining spawning habitat suitability.

417

418   *EOFs and typology of maps*. Based on the EOF decompositon, we classified the varying

419   spawning distributions into 3 major types of maps: distributions extending over many

420   habitats including northern parts of Biscay (type G1), distributions contracted on two core

421   habitats in south Biscay (type G2) and distributions limited to coastal areas in south

422   Biscay (typeG3). At low biomass level the distributions of type G3 and G2 that were

423   more contracted spatially and coastal were never followed by a high recruitment in the

424   following year. The coastal distribution of type G3 occurred when biomass was lower,

425   fish length smaller and water column stratification greater. In these circumstances total

426   annual fecundity can be expected to be lower because of lower spawning biomass and a

427   shorter spawning period as predicted by bioenergetics (Pecquerie et al., 2009). Also when

428   too high, water column stratification can be detrimental for larval survival (Allain et al.,

429   2007). At low biomass level, the spawning distribution of type G2 was associated with

430   larger fish and colder bottom temperature. These conditions could also result in lower

431   total fecundity over the restricted G2 habitats. Further, the spatial initial conditions for

432   larvae drifts will be more coastal and with less spatial extension for types G3 and G2 than

433   for G1 and this may also influence laval trajectories and survival probability. Whether G2

434   and G3 spawning types at biomass level can be associated with low subsequent

435   recruitment could be mechanistically tested with a coupled bio-physical larval model

436   predicting larval transport and survival. In doing so, a given spawning map type could be

437   used as initial condition to the larval model, following the approach of Ospina-Alvarez et

438   al. (2013). Using different egg maps as initial condition for their transport larval model as

439   derived from survey data, they found significant differences in larval drift trajectories.

440

441   *Dynamic update of maps.* Variability in the spatial distributions under both population

442   and environmental conditions is a useful knowledge for increasing robustness in spatial

443   management measures. Based on the regressions of EOF amplitudes on covariates, spatial

444   distributions could be predicted to update dynamically the spawning maps depending on

445   combined population and environmental scenarios. Robustness of spatial management

446   measures could then be tested on that basis. The study provides a typology of spawning

447   distributions that can be input to dynamic population spatial models.

448

457

458   **References**

459   Allain, G., Petitgas, P. and Lazure, P. 2007. The influence of environment and spawning
460   distribution on the survival of anchovy (Engraulis encrasicolus) larvae in the Bay of
461   Biscay (NE Atlantic) investigated by biophysical simulations. Fisheries
462   Oceanography, 16: 506-514.

463   Austin, M. 2007. Species distribution models and ecological theory: a critical assessment
464   and some possible new approaches. Ecological Modeling, 200: 1-19.

465   Burnham, K. and Anderson, D. 2002. Model Selection and Multi-model Inference: a
466   Practical Information-Theoretical Approach. Springer, New York

467   Cotter, J., Mesnil, B., Witthames, P. and Parker-Humphreys, M. 2009. Notes on nine
468   biological indicators estimable from trawl surveys with an illustrative assessment for
469   North Sea cod. Aquatic Living Resources, 22: 135-153

470   Dommenget, D. and Latif, M. 2002. A cautionary Note on the Interpretion of EOFs.
471   Journal of Climatology, 15: 216-225

472   Doray, M., Massé, J. and Petitgas, P. 2010. Pelagic fish stock assessment by acoustic
473   methods at Ifremer. http://archimer.ifremer.fr/doc/00003/11446/

474   Guisan, A. and Zimmermann, N. 2000. Predictive habitat distribution models in ecology.
475   Ecological Modelling 135: 147-186

476   Huret, M., Petitgas, P. and Woillez, M. 2010. Dispersal kernels and their drivers captured
477   with a hydrodynamic model and spatial indices: A case study on anchovy (Engraulis
478   encrasicolus) early life stages in the Bay of Biscay. Progress in Oceanography, 87: 6-
479   17.

480   Huret, M., Sourisseau, M., Petitgas, P., Struski, C., Léger, F. and Lazure, P. 2013. A
481   multi-decadal hindcast of a physical–biogeochemical model and derived
482   oceanographic indices in the Bay of Biscay. Journal of Marine Systems, 109-110:
483   S77–S94.

484   ICES. 2012. Report of the Working Group on Southern Horse Mackerel, Anchovy and
485   Sardine (WGHANSA). ICES CM 2012/ACOM:16

486   Labry, C., Herbland, A., Delmas, D., Laborde, P., Lazure, P., Froidefond, J.-M., Jégou,
487   A.-M. and Sautour, B. 2001. Initiation of winter phytoplankton blooms within the
488   Gironde plume waters in the bay of Biscay. Marine Ecology Progress Series 212:
489   117-130.

490 Le Pape, O., Delavenne, J. and Vaz, S. 2014. Quantitative mapping of fish habitat: A
491   useful tool to design spatialised management measures and marine protected area
492   with fishery objectives. Ocean and Coastal Management, 87: 8-19

493 MacCall, A. 1990. Dynamic geography of marine fish populations. University of
494   Washington Press, Seattle.

495 Massé, J. 1996. Acoustic observations in the bay of Biscay: schooling, vertical
496   distribution, species assemblages and behaviour. Scientia Marina, 60 (Suppl.2): 227-
497   234

498 Massé, J., Koutsikopoulos, C. and Patty, W. 1996. The structure and spatial distribution
499   of pelagic fish schools in mutlispecies clusters: an acoustic study. ICES Journal of
500   Marine Science 53: 155-160

501 Mestas-Nuñez, A. 2000. Orthogonality properties of rotated empirical modes.
502   International Journal of Climatology, 20: 1509-1516

503 Morfin, M., Fromentin, J.-M., Jadaud, A. and Bez, N. 2012. Spatio-Temporal Patterns of
504   Key Exploited Marine Species in the Northwestern Mediterranean Sea. PlosOne,
505   7(5): e37907. doi:10.1371/journal.pone.0037907

506 Motos, L. 1996. Reproductive biology and fecundity of the Bay of Biscay anchovy
507   population (Engraulis encrasicolus, L.). Scientia Marina, 60 (Suppl. 2): 195-207

508 Ospina-Alavrez, A., Bernal, M., Catalán, I., Roos, D., Bigot, J.-L. and Palomera, I. 2013.
509   Modeling fish egg production and spatial distribution from acoustic data: a step
510   forward into the analysis of recruitment. PLoS One, 8: e73687. DOI:
511   10.1371/journal.pone.0073687.

512 Pecquerie, L., Petitgas, P. and Kooijman, S. 2009. Modeling fish growth and reproduction
513   in the context of the Dynamic Energy Budget theory to predict environmental impact
514   on anchovy spawning duration. Journal of Sea Research 62: 93–105.

515 Petitgas, P. 2003. A method for the identification and characterization of clusters of
516   schools along the transect lines of fisheries-acoustic surveys. ICES Journal of Marine
517   Science, 60: 872–884

518 Petitgas, P., Massé, J., Grellier, P. and Beillois, P. 2003. Variation in the spatial
519   distribution of fish length: a multi-annual geostatistics approach on anchovy in
520   Biscay, 1983-2002. ICES CM 2003/Q:15.

521   Petitgas, P., Goarant, A., Massé, J. and Bourriau, P. 2009. Combining acoustic and
522   CUFES data for the quality control of fish-stock survey estimates. ICES Journal of
523   Marine Science, 66: 1384-1390

524   Petitgas, P., Reid, D., Carrera, P., Iglesias, M., Georgakarakos, S., Liorzou, B. and Massé,
525   J. 2001. On the relation between schools, clusters of schools, and abundance in
526   pelagic fish. ICES Journal of Marine Science, 58: 1150-1160

527   Petitgas, P., Secor, D., McQuinn, I., Huse, G. and Lo, N. 2010. Stock collapses and their
528   recovery: mechanisms that establish and maintain life-cycle closure in space and
529   time. ICES Journal of Marine Science, 67: 1841-1848.

530   Planque, B., Loots, C., Petitgas, P., Linstrøm, U. and Vaz, S. 2011. Understanding what
531   controls the spatial distribution of fish populations using a multi-model approach.
532   Fisheries Oceanography, 20: 1-17

533   Preisendorfer, R. 1988. Principal Component Analysis in Meteorology and
534   Oceanography. Elsevier, Amsterdam

535   Richman, M. 1986. Rotation of principal components. Journal of Climatology, 6: 293-
536   335

537   Rivoirard, J., Demange, C., Freulon, X., Lécureuil, A. and Bellot, N. 2012. A Top-Cut
538   Model for Deposits with Heavy-Tailed Grade Distribution. Mathematical Geoscience,
539   44. DOI 10.1007/s11004-012-9401-x

540   Schrum, C., St. John, M. and Alekseeva, I. 2006. ECOSMO, a coupled ecosystem model
541   of the North Sea and Baltic Sea: Part II. Spatial–seasonal characteristics in the North
542   Sea as revealed by EOF analysis. Journal of Marine Systems, 61 (1–2) : 100–113

543   Shepherd, T. and Litvak, M. 2004. Density-dependent habitat selection and the ideal free
544   distribution in marine fish spatial dynamics: considerations and cautions. Fish and
545   Fisheries, 5: 141-152.

546   Simmonds, J. and MacLennan, D. 2005. Fisheries acoustics: theory and practice.
547   Blackwell Science, Oxford. 2nd Edition.

548   Van Keeken, O., van Hoppe, M., Grift, R. and Rijnsdorp, A. 2007. Changes in the spatial
549   distribution of North Sea plaice (Pleuronectes platessa) and implications for fisheries
550   management. Journal of Sea Research, 57: 187-197

551   Woillez, M., Petitgas, P., Huret, M., Struski, C. and Léger, F. 2010. Statistical monitoring
552   of spatial patterns of environmental indices for integrated ecosystem assessment:
553   Application to the Bay of Biscay pelagic zone. Progress in Oceanography, 87: 83–93

554   Woillez, M., Petitgas, P., Rivoirard, J., Fernandes, P., terHofstede, R., Korsbrekke, K.,
555   Orlowski, A., Spedicato, M.-T. and Politou, C.-Y. 2006. Relationships between
556   population spatial occupation and population dynamics. ICES CM 2006/O:05

557

558

559

559

560 Table 1: hydrological indices calculated using CTD profiles collected during the PelGas

561 cruises. ρ: water density; z: depth; H: bottom depth or 60 m when bottom depth was

562 greater; g: gravitational acceleration; s_0: reference sea water salinity

563

Index	units	Formula
Stratification index		
Deficit of potential energy (Dep)	kg m^{-1} s^{-2}	$Dep=\dfrac{1}{H}\int\limits_{-H}^{0}(\bar{\rho}-\rho(z))gzdz$; $\bar{\rho}=\dfrac{1}{H}\int\limits_{-H}^{0}\rho(z)dz$
River plume index		
Equivalent fresh water height (Hfw)	m	$Hfw=\int\limits_{-H}^{0}\dfrac{s_0-s(z)}{s_0}dz$; $s_0=35.85$
Hydrological indices		
Surface (7m) temperature (Ts)	°C	
Bottom temperature (Tb)	°C	

564

565

566 Table 2 : Cumulated percent total variance explained by the principal components of the

567 EOF decomposition.

568

PC	1	2	3	4	5	6	7	8	9	10	11	12	13
% var	0.30	0.50	0.65	0.77	0.84	0.89	0.93	0.95	0.97	0.98	0.99	1.00	1.00

569

570

571

572

573

574   Table 3 : Selected linear regression models of EOF amplitudes *Um* on covariates.

575   Covariates are: SSB: stock biomass as estimated by the survey (tonnes), q25 and q75 the

576   25 and 75 percentiles of the length distribution (cm), Ts and Tb the surface and bottom

577   temperature (°C), Dep the index of stratification, Hfw the index of river plume (see table

578   1). Values tabled are the coefficients estimated for each covariate. R-square is the

579   variance explained by the linear model (1- residual variance/total variance).

580

EOF	*EOF1*	*EOF2*	*EOF3*	*EOF4*
EOF description	*Less at coast off Landes, more off Gironde*	*Less in southern Biscay, more in North*	*Less on shelf off Landes, more at coast off Vendée*	*Less on shelf off Landes, more at shelf edge in North*
Amplitudes	U_1	U_2	U_3	U_4
Intercept	97.44	-201.3	196.3	-127.432
SSB	0.0005875	0.0002659	0.0002982	-
q25	-	-	-24.87	-
q75	12.54	-	31.93	-
Ts	-	-	-	8.789
Tb	-29.93	12.98	-	-
Dep	0.4274	0.5136	-	-
Hfw	-	-	-	-
R-square	0.68	0.70	0.43	0.52

581

582

583

583　Table 4: Relative importance weights of covariates over the 127 candidate linear models

584　for each EOF amplitude *Um*. Covariates are defined in Table 3.

585

	U_1	U_2	U_3	U_4
SSB	0.97	0.80	0.47	0.32
q25	0.47	0.40	0.63	0.47
q75	0.55	0.45	0.53	0.40
Ts	0.40	0.44	0.32	0.81
Tb	0.87	0.60	0.36	0.36
Dep	0.43	0.62	0.30	0.42
Hfw	0.38	0.38	0.34	0.48

586

587   Fig. 1: Schematics of the variability in spatial distributions. Environmental and

588   population conditions affecting habitats and behaviour respectively, which influence

589   spatial distribution. In return, the spatial organisation of the population also affects its

590   dynamics and thus feeds back into the loop of drivers affecting population behaviour.

591

592   Fig. 2 : Localization of acoustic transects, CTD stations and polygon of study

593

594   Fig. 3 : Mean distribution map, 2000-2012. The map represents the average percentage

595   (x1000) of population biomass in grid cells of 0.4°x0.4°.

596

597   Fig. 4: EOFs 1 and 2 (left) and their associated local explained variance (right). EOF1

598   (top) corresponds to the following pattern of variability: when there is less fish at the

599   coast off Landes (negative values), there is more fish off Gironde (positive values) and

600   vice versa. For EOF2 (bottom) the pattern of variability is the following: when there is

601   less fish in southern Biscay (negative values), there is more fish in northern Biscay

602   (positive values) and vice versa.

603

604   Fig. 5: EOFs 3 and 4 (left) and their associated local explained variance (right). EOF3

605   (top) corresponds to the following pattern of variability: when there is less fish on shelf

606   off Landes (negative values), there is more fish at coast off Vendée (positive values) and

607   vice versa. For EOF4 (bottom) the pattern of variability is the following: when there is

608   less fish on shelf off Landes (negative values), there is more fish at the shelf-edge in

609   North (positive values) and vice versa.

610

611   Fig. 6: Typology of maps (hierarchical clustering) based on the first four principal

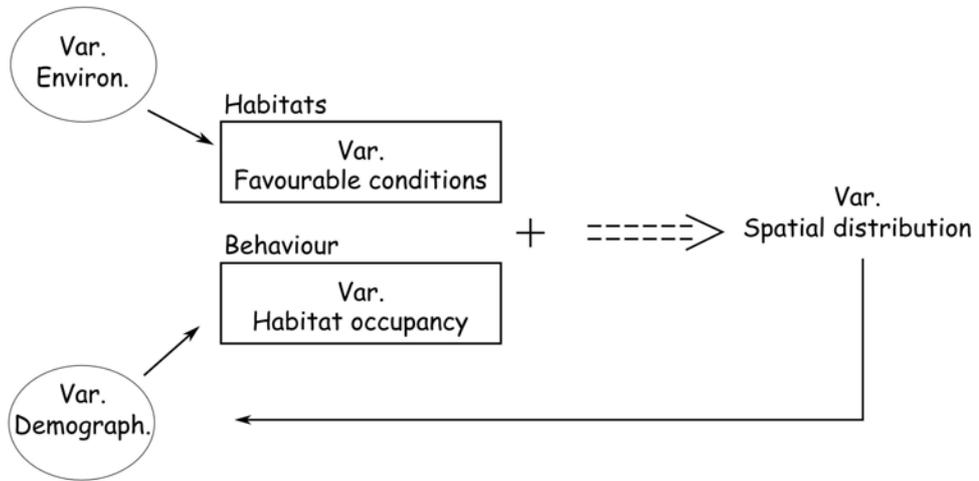612   components associated to the EOFs. Three groups were retained.

613

614   Fig. 7: Mean map in each of the three identified groups of maps: G1 (years 2000, 2001,

615   2008, 2009, 2011, 2012), G2 (years 2002, 2005, 2010), G3 (years 2003, 2004, 2006,

616   2007).

617

618   Fig. 8: Relationship between recruitment in year (t+1), spawning stock biomass and

619   spawning distribution map type (G1-G3) in year (t). At low spawning biomass level

620   maps of type G3 and G2 in year (t) are never followed with a high recruitment in year

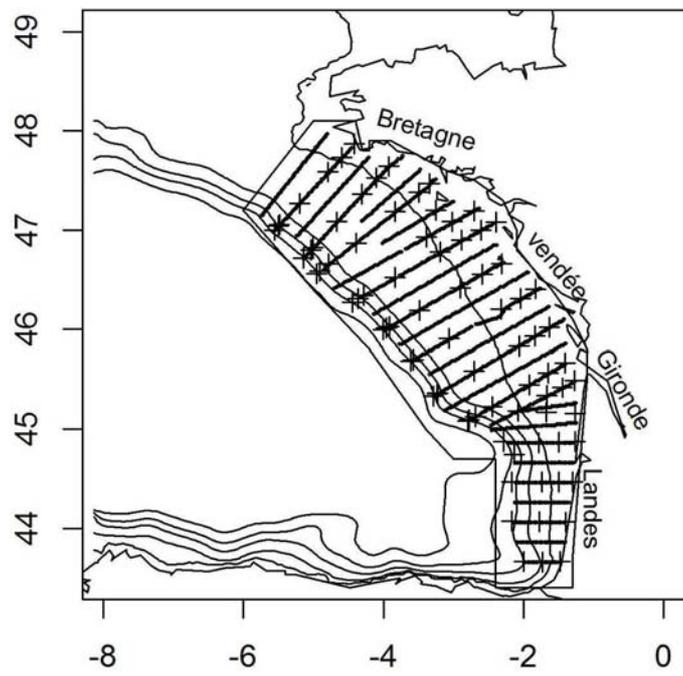621   (t+1). Map of type G3 is observed only at low biomass level.

622

622

623
624
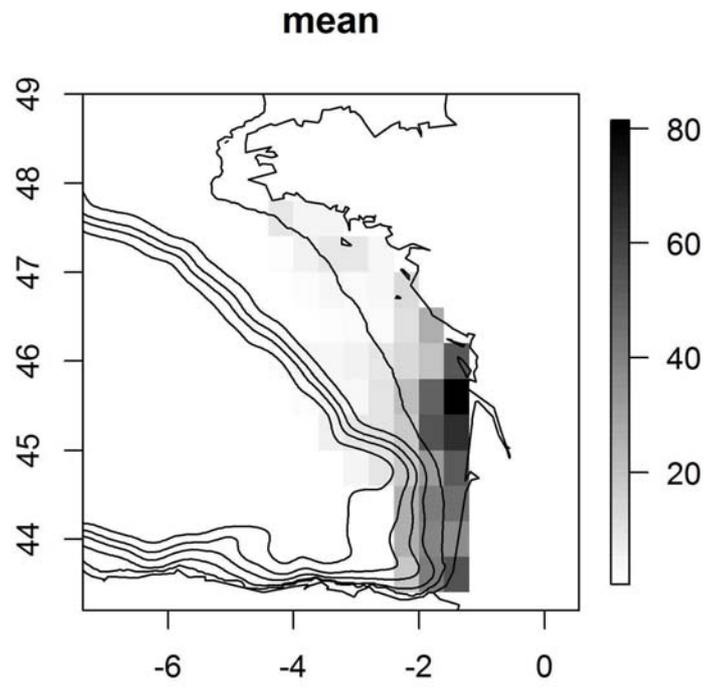625   Figure 1
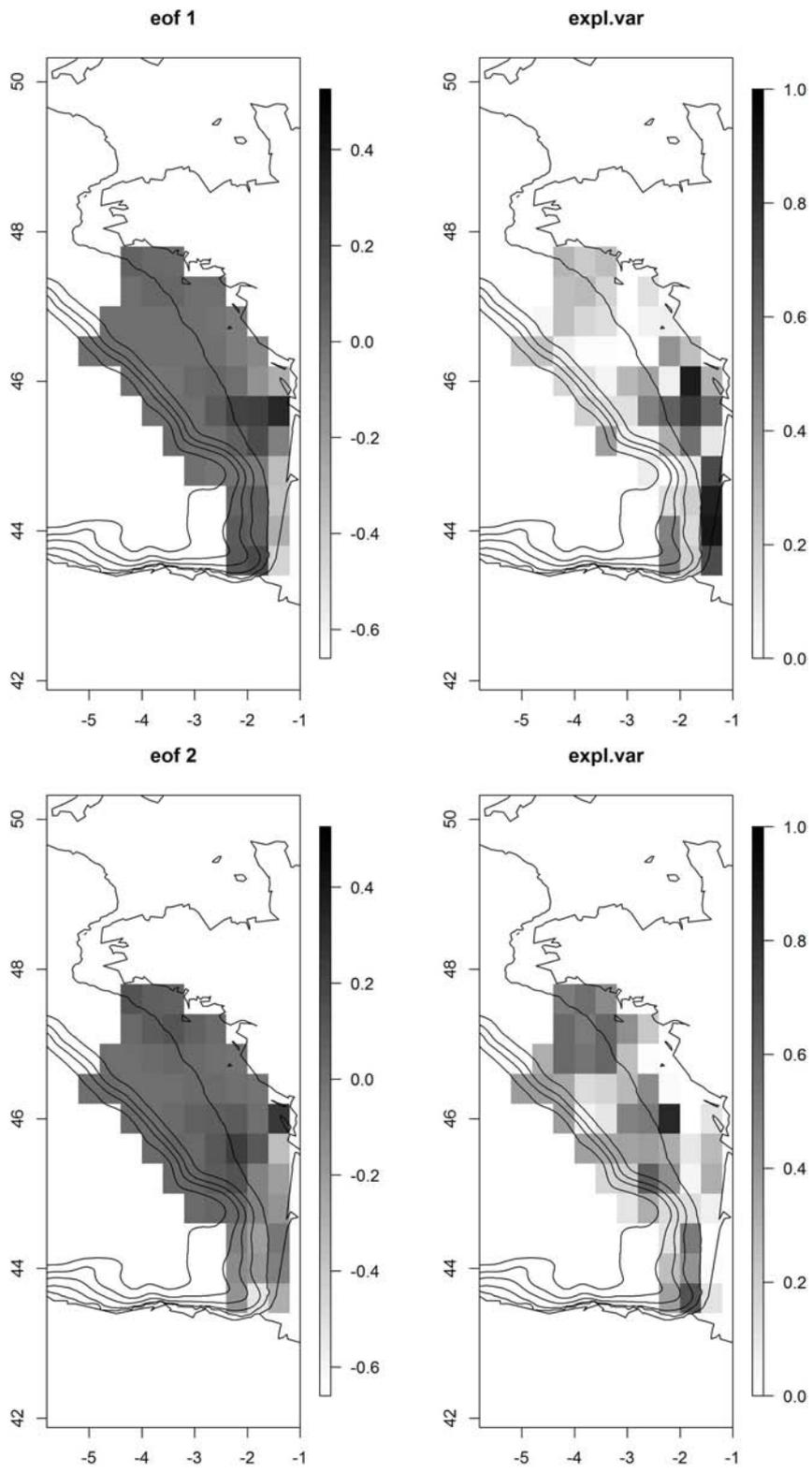626

626

627
628
629 Figure 2
630

630

mean

631

Figure 3

632

633

633

634
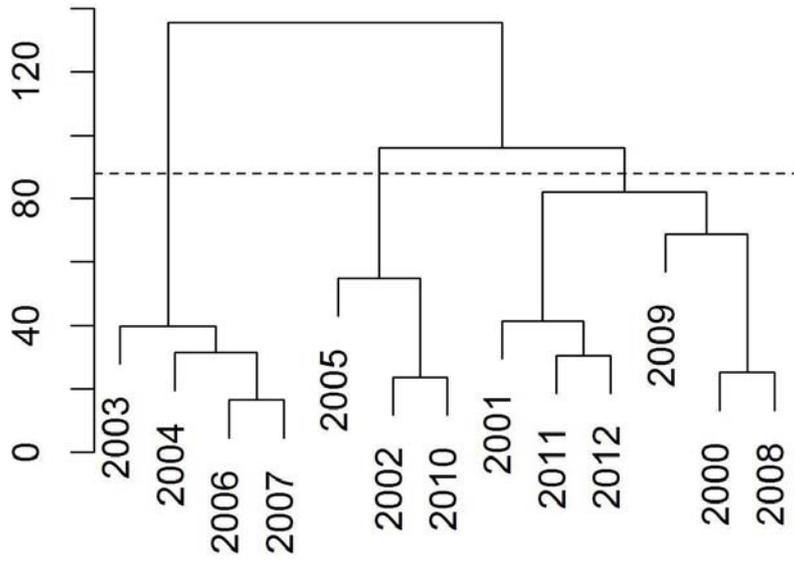635   Figure 4
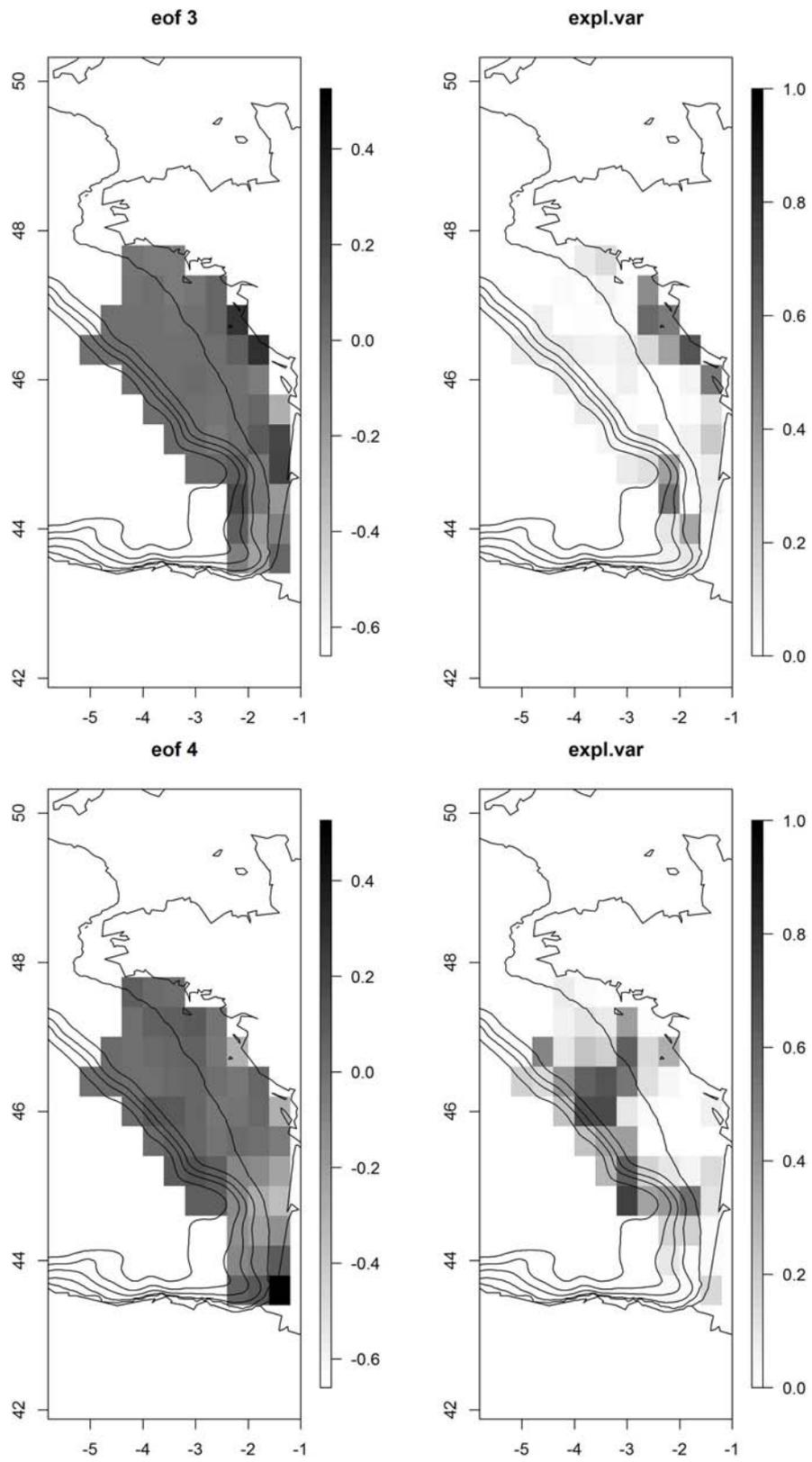
636

637   Figure 5
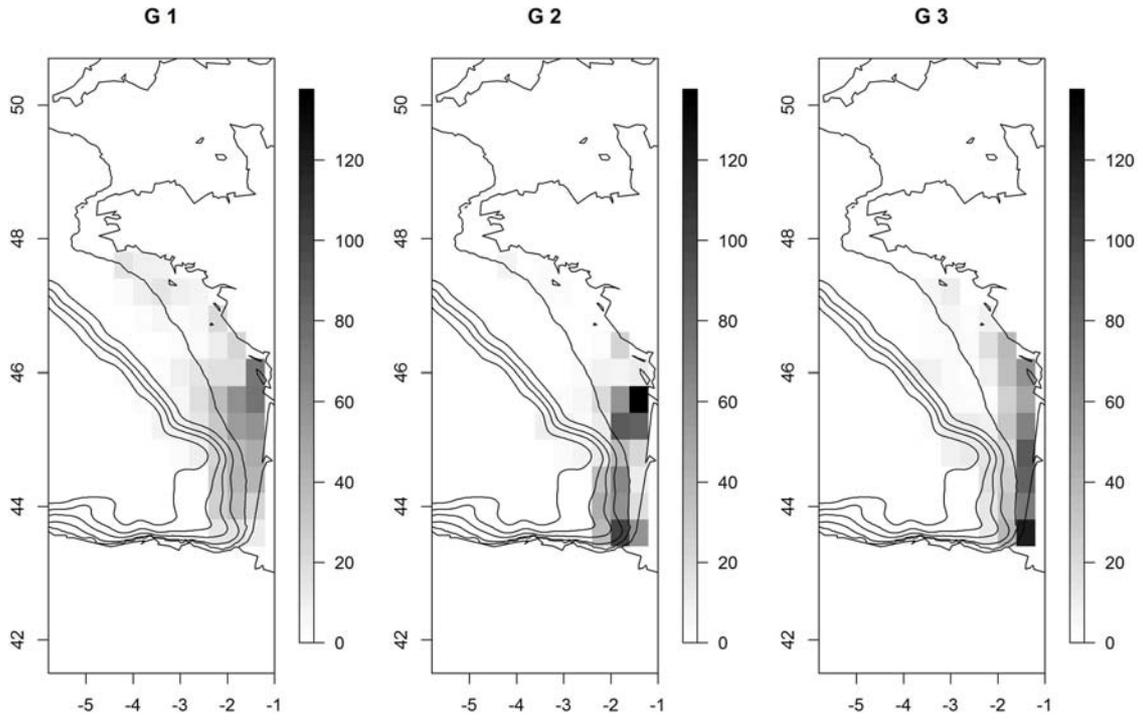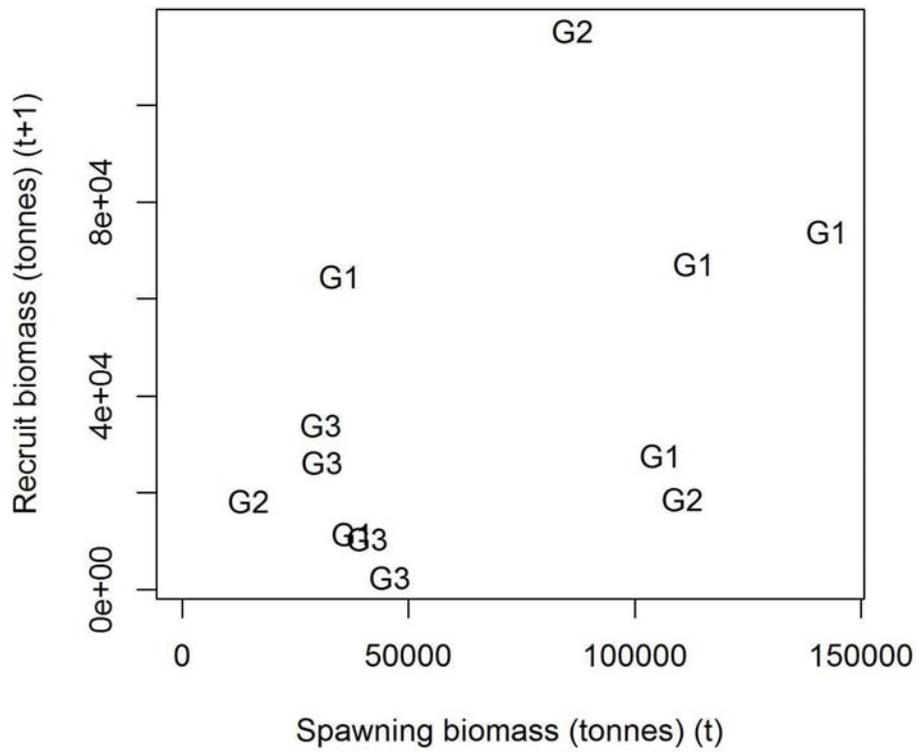
638

639

640

641

642

643   Figure 6

644

644

645
646
647   Figure 7
648
649

649

650
651
652 Figure 8
653