Nonlinear ecological processes driving the distribution of marine decapod larvae

M. Peña[a,*], A. Carbonell[a], A. Tor[a], D. Alvarez-Berastegui[b], R. Balbín[a], A. dos Santos[c], F. Alemany[a]

[a]*Instituto Español de Oceanografia, Palma de Mallorca, Spain*
[b]*Sistema de Observación Costero y de Predicción de las Islas Baleares, Palma de Mallorca, Spain*
[c]*Instituto Português do Mar e da Atmosfera, Lisbon, Portugal*

Abstract

The complexity of the natural processes lead to many nonlinear interacting factors that influence the distribution and survival of marine pelagic species, particularly in their larval phase. The management of these ecosystems require techniques that unveil those interactions by studying the system globally, including all relevant variables and combining both community and environmental data in a single step. Specifically, we apply an unsupervised neural network, the Self-Organizing Map (SOM), to a combined dataset of environmental and decapod larvae community data from the Balearic sea, obtained in two years with contrasting environmental scenarios, as an Exploratory Data Analysis (EDA) technique that provides a global and more detailed view of both the environmental processes and their influence on the distribution of such planktonic community.

We examine the parental influence on the initial larval distribution by aggregating data by adult habitat, which also increments the signal to noise ratio (mean data patterns over noise due to outliers or measurement errors), and consider the distribution of larvae by development stage (as a proxy of age and hence of potential dispersion). The joined study of parental effect, drifting or concentration events determined by dynamical processes in the whole water column, and lifespan, draws the possible paths followed by larvae, and highlights the more influencing variables in their distribution. Investigation of the different aspects of dynamic height (absolute values, gradients or edges and correlations) clarified the effect of the oceanographic processes on decapods' larvae.

Keywords: Nonlinear processes, unsupervised neural networks, Self-Organizing Maps, Decapod larvae, Balearic Sea., Dynamic height

*Corresponding author
Email address: `marian.pena@ba.ieo.es` (M. Peña)

Introduction

The complexity of the natural processes lead to nonlinear relationships that are increasingly considered and emphasised in the literature (Dixon et al., 1999, 2001; Hsieh et al., 2005; Largier, 2003). A nonlinear relation between alongshore dispersal and time in the plankton is described in Largier (2003). Many small scale factors interact nonlinearly with marine pelagic larvae survival and distribution (Bailey, 2002), particularly wind mixing or turbulence (Dixon et al., 1999, 2001). We need to unveil all possible linear and non-linear processes inherent in the complexity of our ecosystems in order to understand the factors influencing the distribution of individuals. The technique employed should include as many relevant variables as possible and combine both community and environmental data in a single step.

Machine learning techniques can be classified in supervised and unsupervised algorithms: the first group includes a known response variable that acts as 'teacher' for the algorithm, guiding it towards a result, and includes, among the most used techniques in larvae ecology, classification, regression, Generalized linear models (GLM) and Generalized additive models (GAM). The second group exclusively includes explanatory variables, and it is not guided towards a result, but makes use of the data to find average patterns. The most known unsupervised techniques are Principal Component Analysis (PCA), clustering such as K-means and, neural networks such as the Self-organizing Map (Kohonen, 1982; Peña et al., 2008) and other Topology preserving Mappings (Peña and Fyfe, 2006; Peña, 2007). In general, ecological modelling of fish and invertebrate larvae make use of techniques that join unsupervised projections (Principal Component Analysis, Metric Multidimensional Scaling, Correspondence analysis) to identify data relationships and supervised linear regressions (Redundancy analysis, Non-Metric Multidimensional Scaling, Canonical Correspondence Analysis) for prediction; see Petitgas et al. (2008) for a revision. Other supervised techniques employed are decision trees (Muhling et al., 2010), GLM and GAM (Dingsør et al., 2007) and neural networks (Muhling et al., 2013). Unsupervised clustering techniques are also popular, mainly through the use of dendrograms (Lindley, 1986). Other techniques such as the Single Parameter Quotient (SPQ) were used in the past to analyze the preference of larvae for certain variable ranges (Lluch-Belda et al., 1991). We are not aware of any application of an unsupervised non-linear technique such as the Self-Organizing Map (SOM) to larvae distribution, though a few examples of ecological modelling in other areas exist, such as in geology (Kosiba et al., 2010), water pollution (Shanmuganathan et al., 2003), vegetation (Foody, 1999), forest data (Giraudel and Lek, 2001) and riverine communities (Li et al., 2012; Stojkovic et al., 2013). Chon (2011) revised the applications of SOM to ecological modelling extensively. SOM application to ecological data has been done until now separately to environmental data (Basterretxea et al., 2012; Liu et al., 2006; Richardson et al., 2003) to find their main components, and biological data (Chon et al., 1996; Kwon et al., 2012; Recknagel, 2001) to study changes in spatio-temporal community patterns or for community ordination. Li et al. (2012) applied SOM to abundance data of riverine macroinvertebrates to obtain community clusters, and then calculated mean values of environmental data for each cluster. Stojkovic et al. (2013) compared an a priori clustering based on environmental data to a posteriori group of clusters based on SOM applied to riverine

fish community abundance data. To our knowledge, no direct application of SOM to a joint dataset of community abundance and environmental data has been done.

Artificial neural networks such as the SOM are able to capture the (non-linear) complexity of the ecosystem without including previous knowledge about the variables relationships, as it is done in supervised techniques. Using an unsupervised technique allows to consider all relevant variables, without pruning them in order to guarantee independence or reduce the number of predictor variables, as required in a supervised technique. Different relationships may arise in this way for intercorrelated variables such as temperature and salinity, allowing to get a more detailed picture of the water masses and their influence on larval distribution. SOM is also robust to errors/outliers, as showcased in Paini et al. (2010). The maximum number of variables used depend on the number of observations in supervised techniques (5 times less as a rule of thumb), while the SOM is often used for feature selection, even with more variables than samples. These advantages improve the ability of researchers to identify potential effects not considered a priory and helping them to establish new hypotheses about causal relations. The main drawback of the joint treatment of the data and including a large number of variables is that correlations not based on causality are more prone to appear, which can be detected with a good knowledge of the data and area. Over performance of SOM over linear ordination techniques were investigated in Giraudel and Lek (2001). The SOM is also employed in a technique called 'correlation hunting' explained in this work, useful to find correlation between variables (Vesanto and Ahola, 1999).

The Balearic Sea (east of Spain) is one of the main spawning areas for bluefin tuna and for other top predator migratory fishes. It has been hypothesised that the main driving factor determining the spatial distribution of these pelagic spawners in the Balearic sea is the interaction of saltier resident waters and new and fresher surface Atlantic waters arriving every summer after crossing the Gibraltar strait (Alemany et al., 2010). Previous studies on meroplanktonic communities in the Balearic sea were mainly focused on fish species and based on supervised techniques (Alemany et al., 2006; Reglero et al., 2012; Rodríguez et al., 2013) excepting Alemany et al. (2010) and Torres et al. (2011) that used SPQ and dendrograms respectively; a recent study dealing with the vertical distribution of the larvae of several decapod species used dendrograms and GAMs (Torres et al., 2013). Similar processes operate in fish and invertebrate larvae although sometimes at different scales -fish have greater swimming capacities while invertebrates can further delay metamorphosis (Bradbury and Snelgrove, 2001)- and thus different responses could be given to similar oceanographic processes. The aim of this study is the application of an unsupervised neural network as an Exploratory Data Analysis (EDA) to a combined dataset of community and environmental data, in order to find structure in the data not seen with linear or supervised methods. Firstly we want to see the influence of all water masses (shallower and deeper) on the distribution of larvae by considering temperatures and salinities at different depth; secondly to study possible differences in environmental impact on decapod larvae populations according to their parent habitat; and thirdly, to evaluate the difference between distribution of recently spawned individuals and older larvae, as a consequence of oceanographic processes experienced

73 during their lifespan.

Material and methods

Data and study area

76 Environmental data and plankton samples used in this study were collected during the Tunibal 2004
77 and Tunibal 2005 surveys, carried out from June 18 to July 10 and June 27 to July 23 respectively. The
78 study area was located around the Balearic islands, mainly in the Algerian Basin (see Fig. 1). Sampling
79 was done over a regular grid of stations (108 stations in 2004 and 110 stations in 2005) placed at intervals of
80 10 x 10 nautical miles, working continuously 24h a day. At each station, hydrographic data were obtained
81 using a rosette equipped with a Seabird 911+ CTD, which included sensors to measure the fluorescence
82 and dissolved oxygen, and Niskin bottles. Hydrographic profiles were realised from surface to 350 or 650 m
83 depth, or 10 metres above the bottom in the shallower stations.

84 Microzooplankton samples for biomass estimations used in this study were taken with Calvet nets
85 equipped with 55 microns meshes sizes performing vertical hauls, at a towing speed of 1m/s. mesozoo-
86 plankton samples were obtained from Bongo 60 cm mouth diameter nets equipped with 200 and 333 microns
87 meshes sizes performing oblique tows, towed at a hauling speed of 20 m min^{-1} and a vessel speed of 2 knots.
88 Samples from the 200 microns mesh size net were frozen and used for biomass measurements and that from
89 the 333 microns mesh size were preserved in 4% buffered formalin in seawater for taxonomic studies. All
90 plankton hauls were carried out from surface to 70 m depth, and General Oceanic flowmeters (model 2030)
91 were fitted to each net in order to measure the volume of water filtered. For more details of sampling pro-
92 cedures and hydrographic data analysis see Alemany et al. (2010) and Balbín et al. (2013). Table 1 shows
93 all considered variables for statistical analysis, including the environmental variables: fluorescence at Mixed
94 Layer Depth (MLD) and integrated from 0 to 100 m, maximum fluorescence and its depth, temperature and
95 salinity at 5, 10, 25, 50, 100 and 200 m depth and at the MLD and chlorophyll-a at 25 m depth (mg/m^3). In
96 order to see if direction of the currents is relevant, and not only intensity, north-south and east-west compo-
97 nents were also selected as variables. Position (latitude, longitude, distance to the coast), hour and bottom
98 depth were included in the analysis. Hour of the day was scaled from 0 to 1 representing the whole day cycle
99 (0 to 24 h). Brunt-Väisälä frequency was calculated to establish areas of higher stability (higher 'max-bfrq'
100 values) and depth of maximum instability. To complement this hydrographical information obtained in situ,
101 data from satellite imagery was also considered, as surface Chlorophyll-a concentration (mg/m^3) and sea
102 surface temperature ($^\circ C$).

103 To carry out this study decapod larvae were sorted from the samples and identified; development stages
104 were noted following Dos Santos and Lindley (2001) and Dos Santos and González-Gordillo (2004) and larvae
105 descriptions available as presented in the checklist of González-Gordillo et al. (2001). Identification was to
106 species level whenever it was possible, but genus, or family, were used otherwise. Larval taxa were further
107 classified based on their adult habitats (Crosnier and Forest, 1973; Udekem d'Acoz, 1999; Zariquiey-Alvarez,

4

Figure 1: Study area
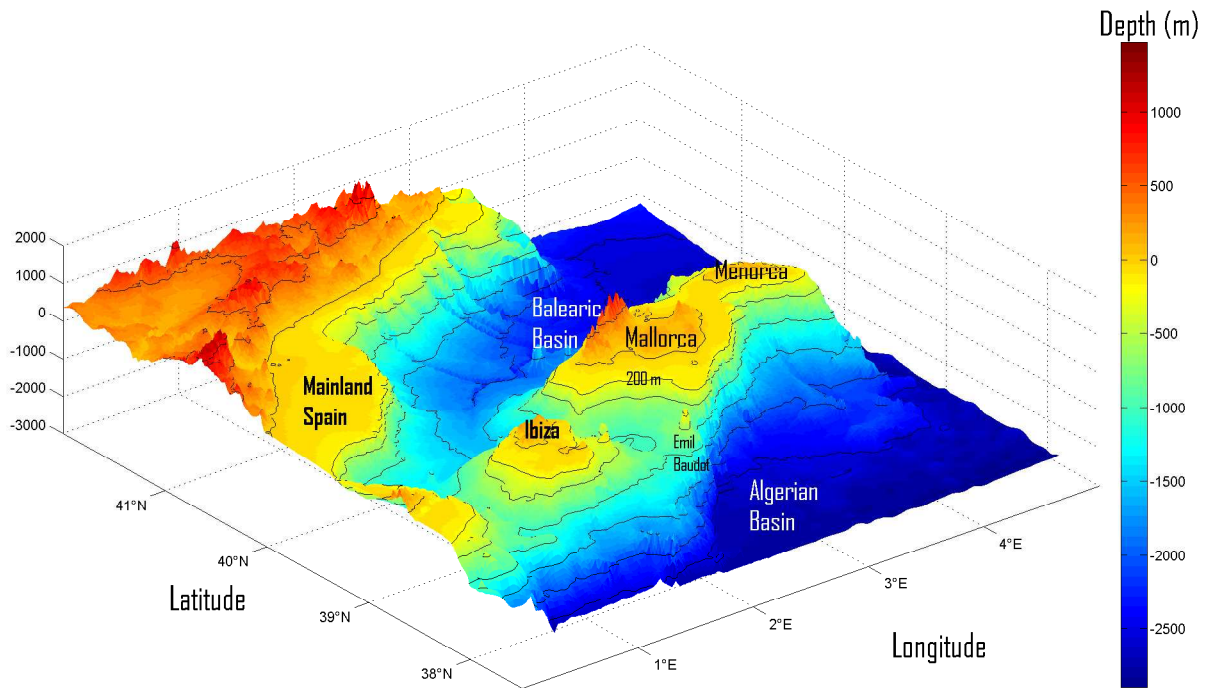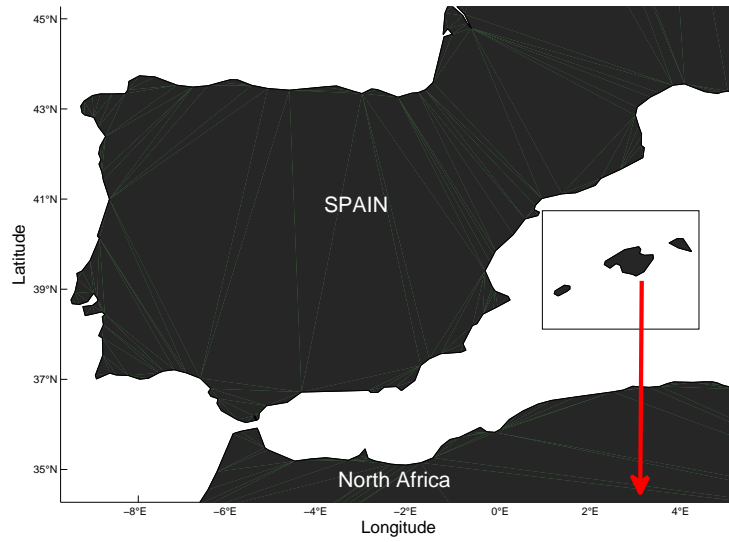
1968) as h1: epibenthic shelf (neritic), h2: epibenthic slope and bathyal (oceanic), h3: nektobenthic shelf (neritic), h4: nektobenthic slope and bathyal (oceanic), h5: mesopelagic (oceanic), and h6: endobenthic (neritic) species. Neritic habitat was considered from the coast to 200 m depth and oceanic from 200 m to 3000 m depth. An abundance index per habitat was created grouping all the larvae originating in the same adult habitat for each station and volume of water filtered. A variable showing the development stage (1 to 10) with maximum abundance per habitat and station was employed as a proxy of ontogenetic variation in the distribution. A value of 10 is given to decapodite (Dendrobranchiata) and megalopa (Pleocyemata). Groups by larvae cycle duration or guilds were also carried out in this particular study, but did not produce further results.

Self-Organizing Maps

The Self-Organizing Map (SOM)(Kohonen, 1982) computes a set of reference vectors (prototypes or neurons) that represent local means of the data. Those prototypes reduce the redundancy in the samples (stations) like in K-means, while the projection into a two dimensional space reduces redundancy of the variables, like in PCA. Averaging the dataset in reference vectors also removes noise.

The SOM is applied in a descriptive manner, that is, all available data is used as the training set. The algorithm organises the positions of the neurons in an unsupervised competitive learning mechanism that works as follows for each training step:

1. Randomly select one sample vector x, from the input data set.

2. Find the neuron, c, whose centre is closest to the input pattern; that neuron will be the winning neuron or the so called Best Matching Unit (BMU) for the pattern.

3. Adjust the centers toward the data vector for the winning neuron and all its neighbours using the following equation:

$$\Delta w_i = \eta(x - w_i)\Lambda(i, c) \tag{1}$$

where η is the learning rate, and Λ is often a monotonically decreasing function of the distance between i and c, known as the neighbourhood function. Normally this function is a Gaussian or a difference of Gaussians.

4. Repeat the steps 1 to 3 for new inputs until some convergence criterion is reached.

To establish the size of the map the usual procedure was followed: the number of units was estimated as $5 * dlen^{0.54321}$, where dlen was the number of samples. Then the two biggest eigenvalues of the training data were calculated and the ratio between sidelengths of the map grid were set to this ratio. The actual sidelengths were then set so that their product was as close to the desired number of map units as possible.

Each SOM neuron has the same dimensionality as the original data set but is located in a 2-dimensional grid; visualising separately the value of one particular variable for all neurons depicts a map of the variable

in a 2D projection (known as component plane). The same area in the map can be compared for all variables in order to see the coincidence of high or low values. To study individual relationships between variables, we applied a second SOM using the projected variables (neuron's variables) as sample points, and visualized in a SOM plot all the variables clustered by absolute correlation between them; this is the so called 'correlation hunting' introduced in (Vesanto and Ahola, 1999) that finds linear correlations, but also nonlinear and local or partial correlations. The figure includes high values (above 0.6, being the total range 0 to 1) of the distance of each map unit to its immediate neighbours in gray; these areas delimitate clusters or areas of higher similarity between the variables. Significant correlations ($p < 0.05$) with absolute correlations higher than 0.6 are depicted in a correlation matrix. In order to see which are the new relationships encountered by the SOM, we also present the correlation matrices of the original variables. The SOM toolbox[1] was employed for this study.

The data includes some stations with very high values of larvae abundance. In order to reduce their influence, we applied the sigmoidal normalisation preprocessing that maintains a linear transformation near the mean and a smooth non linearity at the extremes; this transformation produced the lower quantisation error (average distance between each data vector and its BMU) and topographic error (the proportion of all data vectors for which first and second BMUs are not adjacent units).

An analysis on the sensitivity if the SOM to the number of variables used is included in the appendix.

Results

Hydrodynamics

In summer 2004 the presence of Western Intermediate Water (WIW) in the Ibiza channel blocked the northward progress of surface recent Atlantic Waters (recent AW) and deviated them to the east of the archipelago, following the Emile Baudot escarpment and the eastern slope of Mallorca. These recent AW reached the south of Menorca island, where an important topographically induced anticyclonic eddy ($\sim$300 m deep) was formed. On the contrary, in summer 2005 the absence of Western Intermediate Water (WIW) in the channels allowed the inputs of recent AW to progress northwards through the Ibiza channel. However, most of them returned to the Algerian sub-basin through the Mallorca channel, forming and anticyclonic eddy, which instead of progressing northward following Emile Baudot escarpment remained at southern latitudes. The density front (not deeper than 200 m) separates diagonally the study area, with resident AW in the northeast and recent AW in the southwest. More detailed descriptions of the hydrodynamic scenarios found in summer 2004 and 2005 in this area have been already published in Alemany et al. (2010); Balbín et al. (2013) and Rodríguez et al. (2013).

---

[1]http://www.cis.hut.fi/projects/somtoolbox/

Data statistics and spatial distribution

172 Main currents (Gvel variable or absolute velocities) were found surrounding the anticyclonic eddy south
173 of Menorca in 2004 (currents with east-west and north-south direction approximately equally represented)
174 and delimiting the density front extending from the Mallorca channel to the south of Menorca in 2005 (clear
175 prevalence of east direction (see Fig. 2 and Table 1). In relation to 2004, mean values of chlorophyll-a,
176 microzooplankton and fluorescence went down in 2005, while mesozooplankton went up. The mixed layer
177 depth was deeper in 2005, with the maximum stability depth located 3.5 m deeper than in 2004. On the
178 other hand, surface salinities were higher in 2005.

179 Total number of larvae increased from 2004 to 2005, with 45% increment due mainly to larvae of oceanic
180 nektobenthic species (Table 2), though larvae from all habitats incremented their abundance. Larvae of
181 mesopelagic decapods were the most abundant for both years, while endobenthic were so scarce (3 stations
182 in 2004 and 11 stations in 2005) that the few correlations encountered have to be taken with caution. The
183 five more abundant species per habitat each year were approximately the same (Table 3) with the most
184 abundant species changing only for nektobenthic shelf and mesopelagic habitat. The range of possible stages
185 was maintained within each habitat, which allows to compare the stage of maximum abundance per habitat
186 between both years. Late larval stages were more abundant in 2005, with stage IV incrementing a 43% in
187 2005, and just the first two stages having their numbers reduced. Most top abundance stations were near the
188 coast for both years, except for a station located south east of the study area in 2005 (Fig. 3) corresponding
189 to mesopelagic species larvae. The further development of larvae in 2005 was noticeable for all habitats
190 (Fig. 4), with many last larval stages located along the density front were higher currents were registered,
191 particularly for larvae of oceanic epibenthic species.

192 *Seascape non-linearities*

193 Fig. 5 includes the projected variables in the 2-dimensional space ordered by absolute correlation. Shaded
194 gray areas indicates higher distances between neurons, separating clusters. Type and degree of correlation
195 for all variables are included in the corresponding correlation matrix for 2004 (Fig. 6) and for 2005 (Fig.
196 7), both for linear (before the SOM) and non-linear (after the SOM) relationships. Non linear relationships
197 found by the SOM are numerous for environmental data, but particularly prominent for abundance and
198 stage data, as no linear correlation was found.

199 Salinity linear inter-correlation stand out as the most relevant and persistent feature for both years; only
200 the salinity at 200 m in 2004 is not correlated with the salinity at other depths. Temperature linear inter-
201 correlations are not so relevant, particularly for 2005 where only surface temperatures (above MLD, i.e. at
202 5, 10, 25 m) are correlated with the immediate above and below, and no correlation for deeper temperatures
203 was found.

204 Higher dynamic heights define the anticyclonic eddy south of Menorca in 2004, and the density front
205 in 2005. Dynamic height presents correlation with 50/100 m depth temperature in 2004 and 200 m depth

Figure 2: Dynamic Height (cm) at 15 m depth and associated geostrophic currents in the area for both years
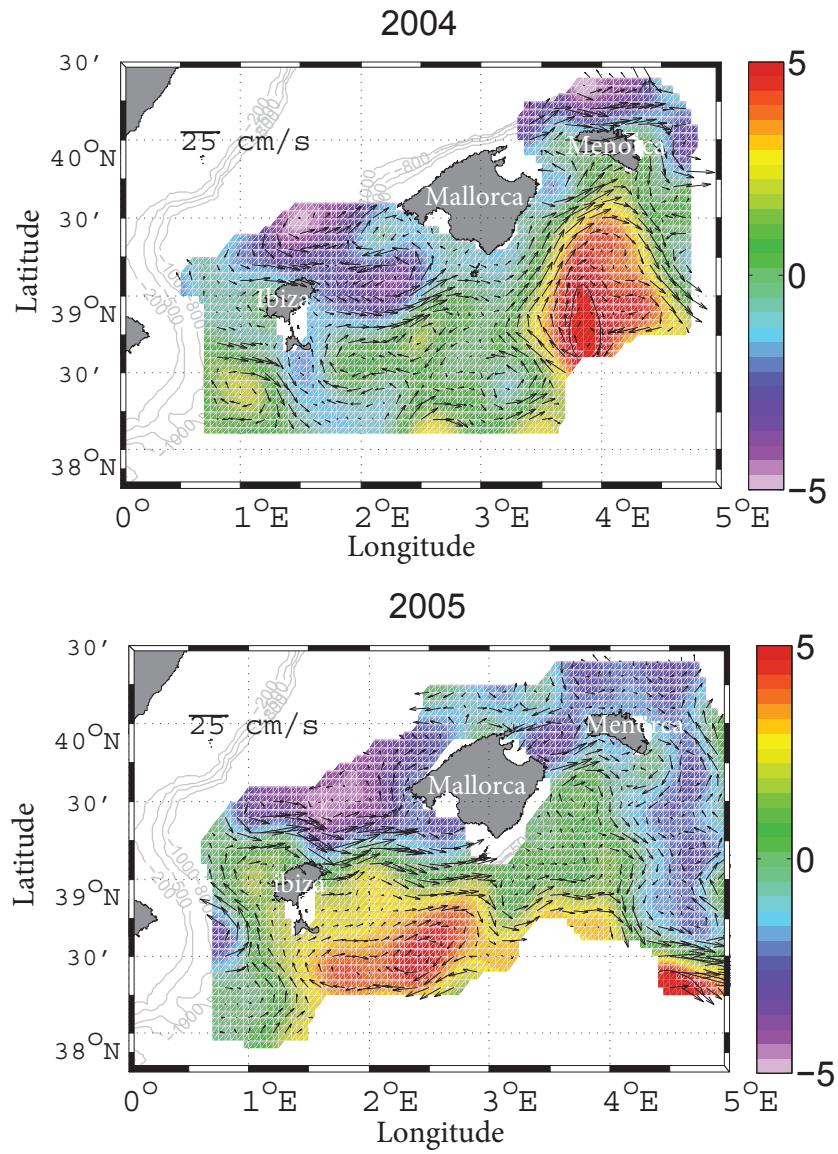
Table 1: Statistical summary of the data in 2004 (108 stations) and 2005 (110 stations). *10 corresponds to megalopal stage.
**10 corresponds to decapodite stage

Code	Variable	2004			2005		
		min	mean	max	min	mean	max
lat	Latitude (degrees)	38.16	39.09	40.33	38.16	39.13	40.34
long	Longitude (degrees)	0.78	2.96	4.70	0.78	2.99	4.92
hour	Time of day (h)	0.01	0.51	0.99	0.00	0.48	0.99
depth	Depth (m)	55.00	1425.15	2649.00	55.00	1490.36	3452.00
EWVEL	Horizontal velocity component (m/s)	-6.59	2.68	14.49	-7.68	3.93	23.37
NSVEL	Vertical velocity component (m/s)	-8.90	0.61	10.57	-9.29	0.15	7.85
GVEL	Total velocity (m/s)	0.87	6.37	14.61	0.41	6.58	23.42
dyn	Dynamic Height (J/kg)	-4.32	-3.75	-3.23	-4.30	-3.83	-3.33
chlasat	Satellite Chlorophyll-a (mg/m^3)	0.08	0.12	0.23	0.08	0.10	0.17
chlasatcv	Satellite Chlorophyll-a cv (mg/m^3)	0.87	4.36	27.08	0.62	3.58	12.23
sst	Satellite temperature $(^\circ C)$	21.42	23.52	24.86	24.17	25.31	26.02
sstcv	Satellite temperature cv $(^\circ C)$	0.10	0.53	2.81	0.10	0.34	2.21
PS-micro	microzooplankton (mg)	0.49	3.59	11.88	0.39	2.59	6.93
PS-meso	mesozooplankton (mg)	1.28	6.11	18.29	1.19	7.21	24.88
maxfluo	Maximum fluorescence (fu)	0.43	1.15	2.61	0.39	1.08	3.46
zmaxfluo	Depth of Maximum fluorescence (m)	43.00	78.55	126.00	48.00	78.52	106.00
sumfluo	0 to 100 m depth total fluorescence (fu)	6.75	32.80	62.92	10.61	32.23	67.24
sumfluoMLD	MLD total fluorescence (fu)	0.50	1.38	2.81	0.10	1.10	2.92
MLD	Mixed Layer Depth (m)	8.46	13.90	21.06	7.48	16.26	27.50
Tem5	Temperature at 5 m depth $(^\circ C)$	21.38	23.37	24.85	23.86	25.36	26.18
Tem10	Temperature at 10 m depth $(^\circ C)$	20.16	22.98	24.85	21.51	24.97	26.05
Tem25	Temperature at 25 m depth $(^\circ C)$	14.69	17.53	21.30	16.29	21.24	25.89
Tem50	Temperature at 50 m depth $(^\circ C)$	13.35	14.83	16.20	13.56	14.97	17.35
tem100	Temperature at 100 m depth $(^\circ C)$	12.85	13.39	13.88	12.87	13.30	13.88
tem200	Temperature at 200 m depth $(^\circ C)$	12.88	13.20	13.56	12.83	13.15	13.29
TemMLD	Temperature at the MLD $(^\circ C)$	21.15	23.21	24.81	22.11	25.17	26.02
Sal5	Salinity at 5 m depth (ppt)	36.87	37.53	37.97	36.77	37.67	38.14
Sal10	Salinity at 10 m depth (ppt)	36.86	37.54	37.97	36.77	37.71	38.24
Sal25	Salinity at 25 m depth (ppt)	37.10	37.67	38.12	36.87	37.80	38.30
Sal50	Salinity at 50 m depth (ppt)	37.50	37.92	38.25	37.16	37.94	38.33
sal100	Salinity at 100 m depth (ppt)	37.98	38.15	38.39	37.83	38.19	38.39
sal200	Salinity at 200 m depth (ppt)	38.16	38.34	38.48	38.24	38.39	38.48
SalMLD	Salinity at the MLD (ppt)	36.88	37.53	37.96	36.78	37.69	38.25
max-bfrq	Brunt-Väisälä coefficient (instability) (s^{-1})	0.002	0.0035	0.007	0.001	0.0035	0.008
zbfrq	Depth of maximum instability (m)	1.50	16.35	27.50	5.50	19.86	41.50
chl25m	Chlorophyll-a at 25 m depth (mg/m^3)	0.03	0.10	0.31	0.04	0.08	0.13
h1	Epibenthic shelf density (n/m^3)	0.00	0.08	2.07	0.00	0.08	1.61
h2	Epibenthic slope&bathyal density (n/m^3)	0.00	0.02	0.34	0.00	0.02	0.41
h3	Nektobenthic shelf density (n/m^3)	0.00	0.01	0.15	0.00	0.04	2.63
h4	Nektobenthic slope&bathyal density (n/m^3)	0.00	0.04	0.57	0.00	0.10	1.23
h5	Mesopelagic density (n/m^3)	0.00	0.11	0.63	0.00	0.12	4.75
h6	Endobenthic density (n/m^3)	0.00	0.00	0.02	0.00	0.00	0.12
h1-st	Epibenthic shelf modal stage*	1.00	1.63	4.00	1.00	3.00	10.00
h2-st	Epibenthic slope&bathyal modal stage*	1.00	1.50	10.00	1.00	3.64	10.00
h3-st	Nektobenthic shelf modal stage	1.00	1.56	6.00	1.00	2.42	9.00
h4-st	Nektobenthic slope&bathyal modal stage	1.00	2.34	7.00	1.00	4.05	7.00
h5-st	Mesopelagic modal stage**	1.00	2.15	10.00	1.00	5.00	10.00
h6-st	Endobenthic modal stage	1.00	1.00	1.00	1.00	1.15	4.00
dist	Distance to the coast (Kms)	11.11	54.45	123.34	11.11	60	167.79

Figure 3: Top: Total larvae distribution. Bottom: Modal stage distribution from 1 to 10 (bottom) for both years. x: stations with no larvae. A: megalopa or decapodit (stage 10)
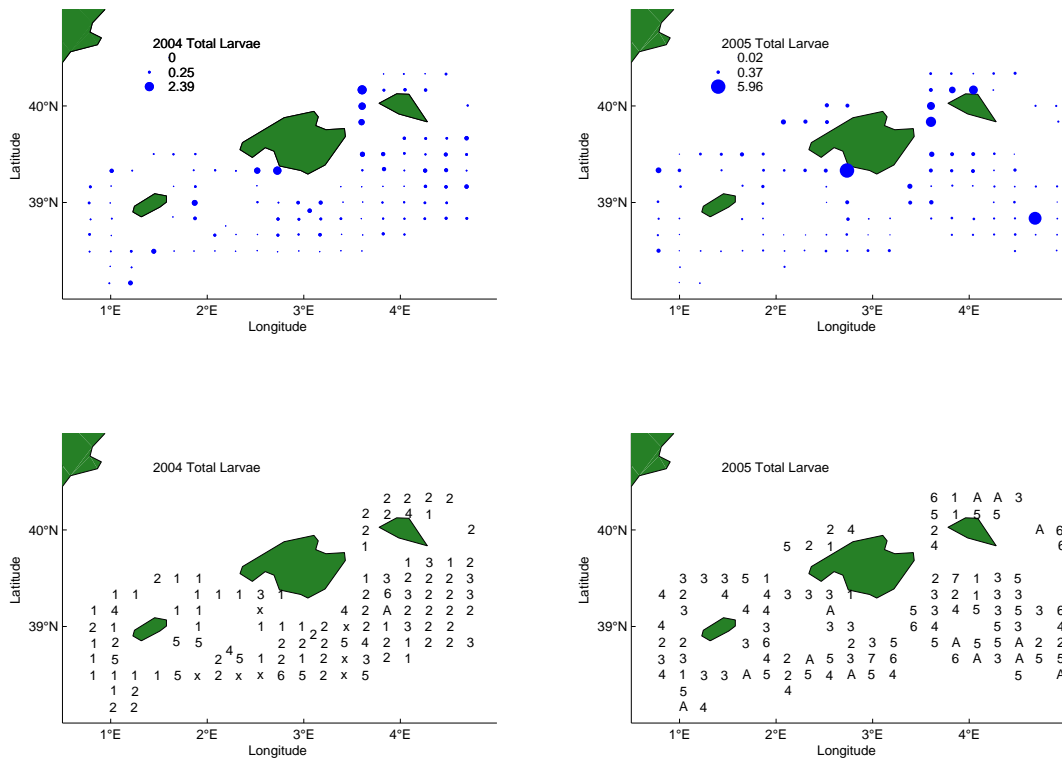
Figure 4: Modal stage (1:10) per habitat distribution. x: stations with no larvae. M: megalopa, D: decapodit (stage 10)
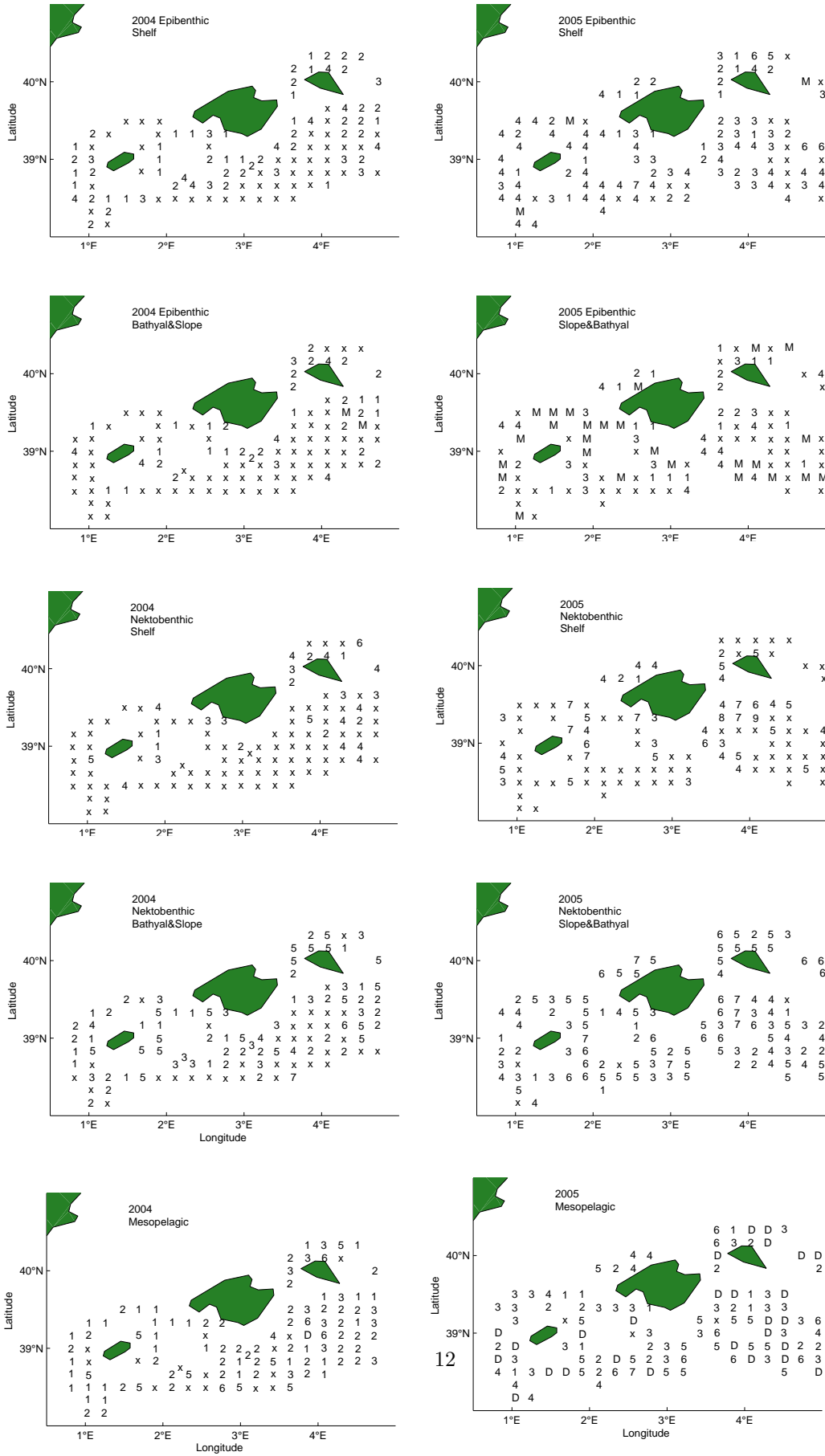
Figure 5: SOM applied to larvae data split by habitat. Projected variables are ordered by absolute correlation. Areas of higher distance between neurons (above 0.6) are depicted as shaded gray areas. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic). TempX and SalX: temperature and salinity at X metres. See table 1 for more variable explanation.
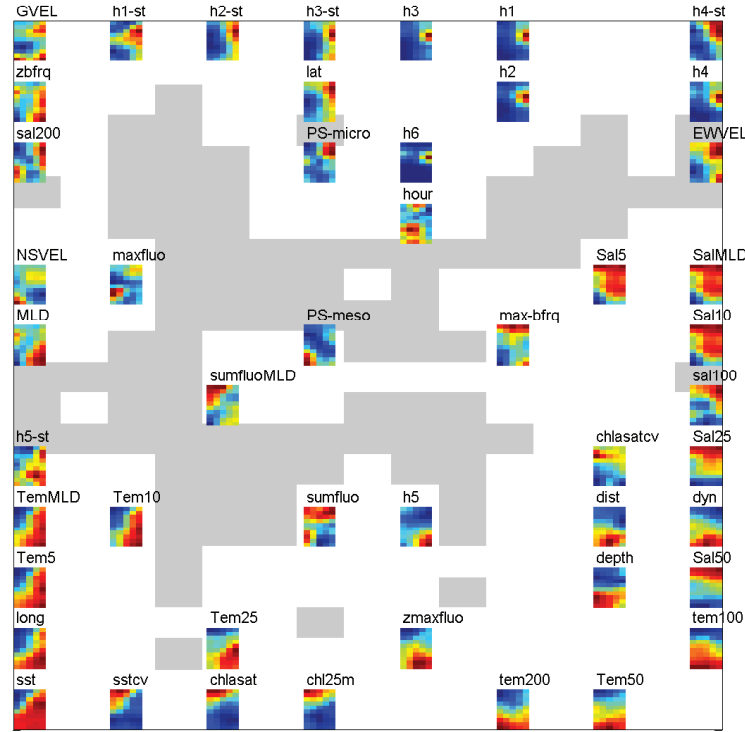
(a) 2004
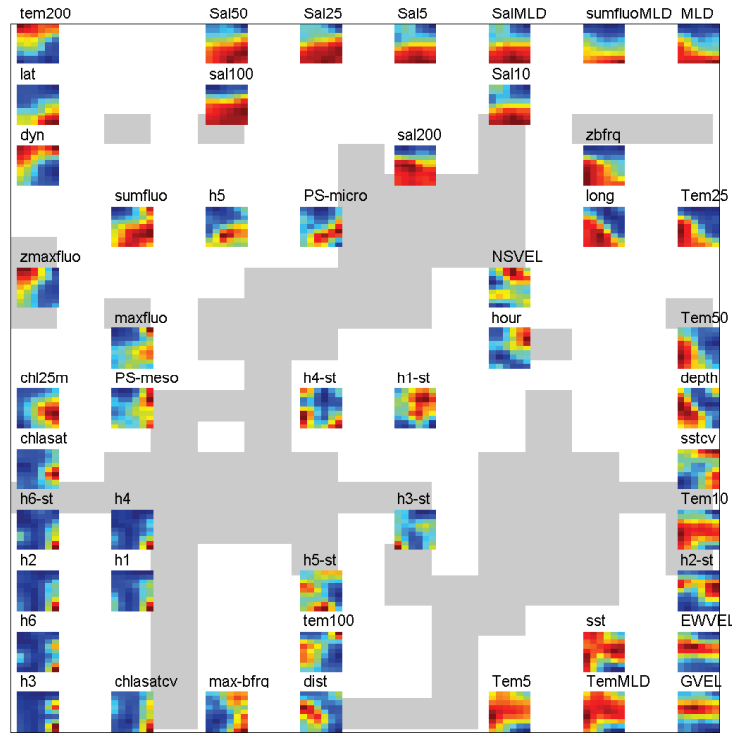
(b) 2005

Figure 6: Linear (top) and non linear (bottom) correlations for 2004. Only significant correlations ($p < 0.05$) with absolute values above 0.6 are shown for clarity. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic).
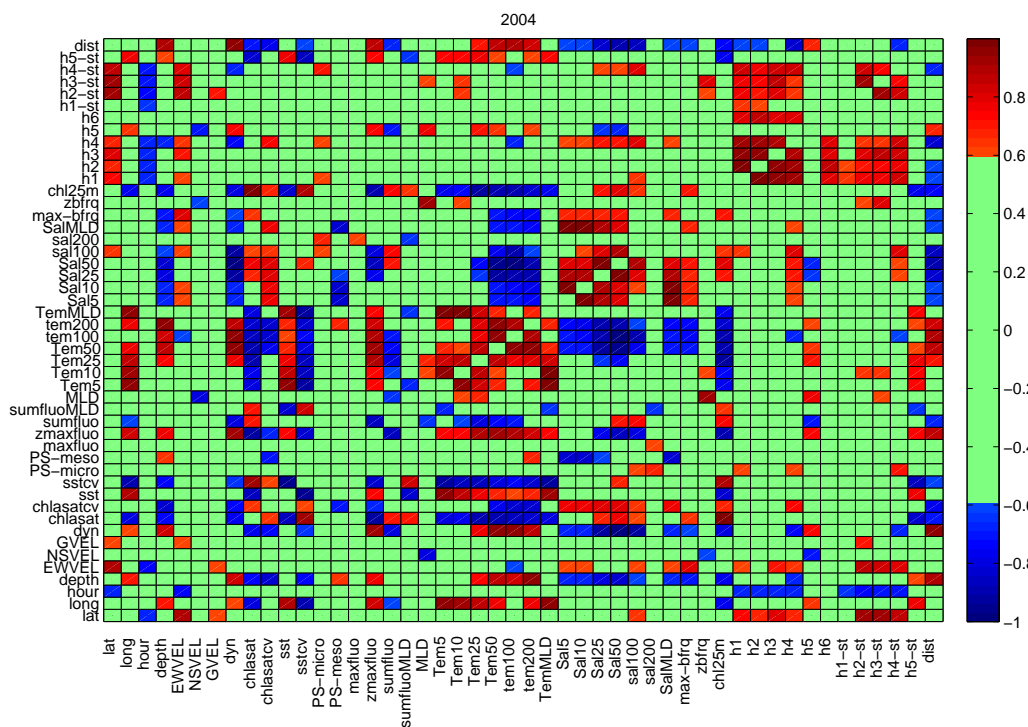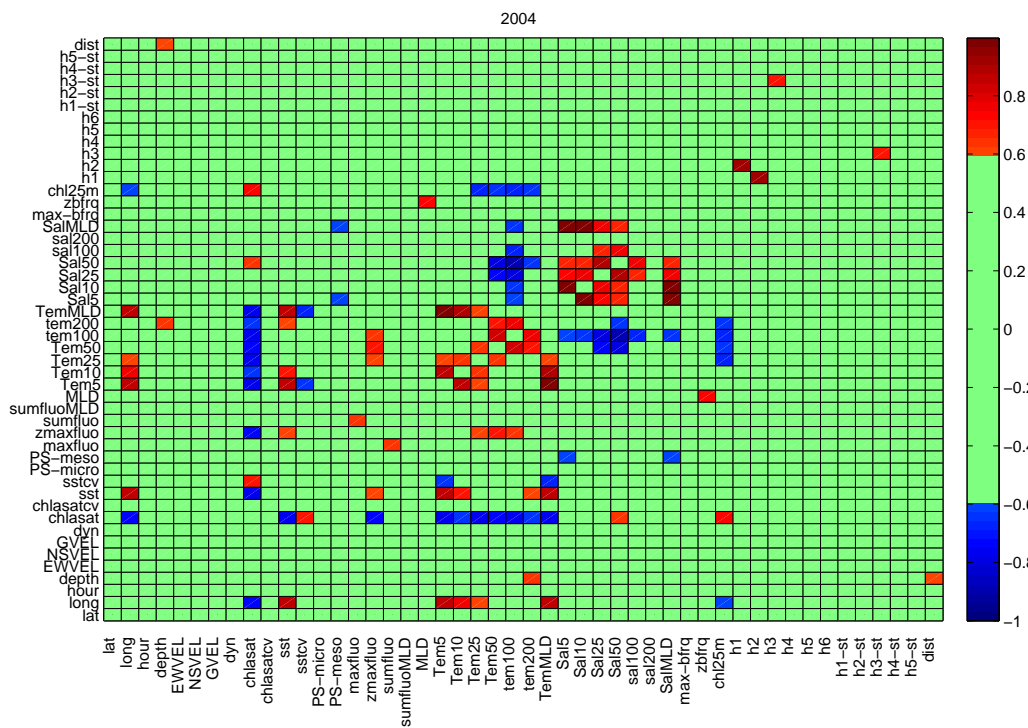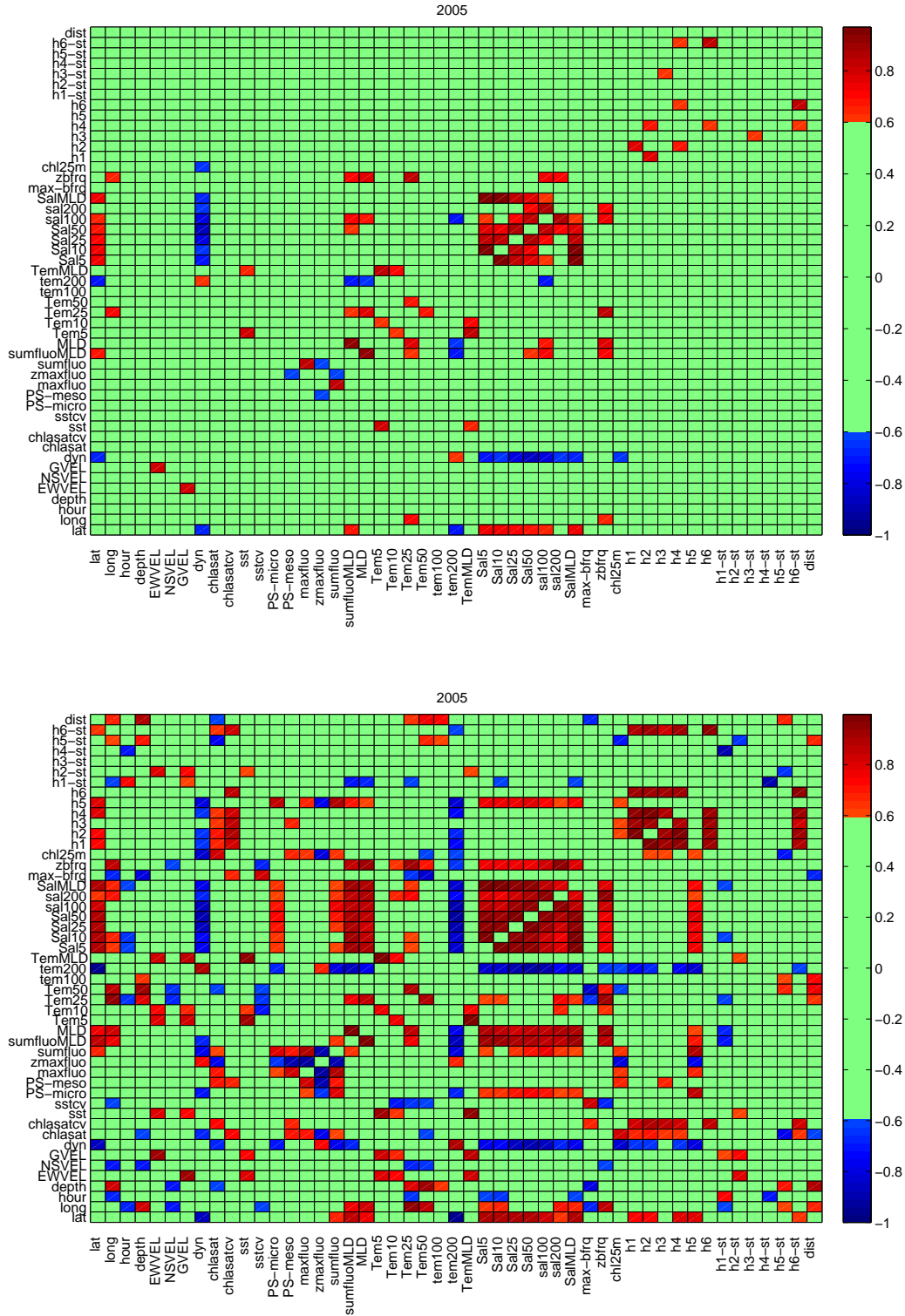
Figure 7: Linear (top) and non linear (bottom) correlations for 2005. Only significant correlations ($p < 0.05$) with absolute values above 0.6 are shown for clarity. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic).

Table 2: Abundances (number) and percentages per habitat and development stage (st1-stX) for 2004 and 2005 with h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic).

N(%)	h1	h2	h3	h4	h5	h6	Total
2004	1383 (32.3)	274 (6.4)	96 (2.2)	608 (14.2)	1919 (44.8)	6 (0.1)	4286 (100)
2005	1628 (22.6)	471 (6.5)	762 (10.6)	1917 (26.7)	2365 (32.9)	50 (0.7)	7193 (100)

N(%)	st I	st II	st III	st IV	st V
2004	1266 (30.0)	1565 (37.0)	720 (17.0)	313 (7.4)	273 (6.5)
2005	1063 (16.4)	1309 (20.3)	1389 (21.5)	1282 (19.8)	937 (14.5)

N(%)	st VI	st VII	st VIII	st IX	st X	Total
2004	88(2.1)	12 (0.3)	0 (0)	0 (0)	49 (1.2)	4286 (100)
2005	484 (7.5)	273 (4.2)	15 (0.2)	8 (0.1)	441 (6.8)	7193 (100)

in 2005. Higher salinities are always in the north, while temperatures and depth of maximum fluorescence are higher in the east. Temperatures and salinities are negatively correlated in 2004 reflecting the earlier sampling of the eastern area with resident AW (lower temperatures and saltier waters) than the western part (higher temperatures and fresher waters), but positively in 2005 at the MLD, when recent AW were sampled first on the west.

Brunt-Väisälä frequency and MLD are correlated as they are different ways of locating the thermocline. The range of Brunt-Väisälä frequency is slightly wider for 2005, depicting also a higher depth of maximum stability. While 2004 depicts spatially homogenous distributed values of maximum stability depth, in 2005 the resident waters on the northeast gathered all the maximum values, indicating deeper thermocline. Stratification was higher north of the islands and south of Menorca in 2004, coinciding with the highest EW currents. In 2005 the depth of maximum stability is clearly highest on the east (within the resident waters), while the highest NS currents are mainly to the west.

In 2004 chlorophyll-a was abundant but mostly on the west (negative correlation with longitude), and thus negatively correlated with temperature that was higher on the east. Microzooplankton is correlated with temperatures at higher depths in 2004 and with all temperatures in 2005.

Larvae non-linearities

When comparing larvae of different adult habitats (Fig. 6(a) on centre lower part) mesopelagic habitat (h5) difference is clear, with fluorescence closely related to larval abundance and stages mainly correlated with temperature at the surface for both years. On the other hand, chlorophyll-a is the main variable correlated

Table 3: Percentages of the five more abundant species per habitat for 2004 and 2005 with h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic).

2004		2005	
h1	%	h1	%
Xantho spp.	34.20	*Xantho spp.*	33.42
Ilia nucleus	17.21	*Calcinus tubularis*	14.93
Galathea spp.	15.47	*Anapagurus spp.*	14.13
Ebalia tumefacta	7.88	*Galathea spp.*	12.29
Anapagurus spp.	5.13	*Ilia nucleus*	7.19
Total	79.90	Total	81.94
h2	%	h2	%
Liocarcinus spp.	44.16	*Liocarcinus spp.*	22.51
Pagurus spp.	18.98	*Goneplax rhomboides*	14.65
Ebalia spp.	12.77	*Bathynectes maravigna*	9.13
Goneplax rhomboides	8.76	*Ebalia spp.*	8.70
Munida spp.	3.65	*Pagurus spp.*	8.49
Total	88.32	Total	63.48
h3	%	h3	%
Eualus spp.	77.08	*Eualus spp.*	88.45
Palaemon spp.	6.25	*Lysmata seticaudata*	5.91
Lysmata seticaudata	5.21	*Philocheras bispinosus*	1.31
Philocheras spp.	5.21	*Palaemon spp.*	1.18
Synalpheus spp.	2.08	*Hippolyte spp.*	1.05
Total	95.83	Total	97.90
h4	%	h4	%
Alpheus glaber	32.40	*Alpheus glaber*	38.50
Processa spp.	14.97	*Processa spp.*	15.08
Pandalina spp.	14.64	*Acanthephyra spp.*	10.75
Processa nouveli	11.35	*Processa nouveli*	9.18
Plesionika spp.	7.89	*Solenocera membranacea*	7.62
Total	81.25	Total	81.12
h5	%	h5	%
Sergestes henseni	27.51	*Parasergestes vigilax*	46.43
Eusergestes arcticus	24.75	*Sergestes spp.*	18.10
Parasergestes vigilax	22.77	*Eusergestes arcticus*	13.15
Sergestes spp.	16.83	*Gennadas elegans*	9.39
Gennadas elegans	4.74	*Sergestes henseni*	6.93
Total	96.61	Total	94.00

with the four first habitats in 2005 (Fig. 6(b) on the bottom left), while stages of neritic epibenthic and oceanic nektobenthic habitats are closely related to the hour variable (Fig. 6(b) centre area).

Mesopelagic species abundances and stages are correlated positively with dynamic height and MLD in 2004 as this group was trapped in the anticyclonic eddy. However oceanic nektobenthic are positively correlated with deeper salinities, but also with east-west currents (as neritic habitats -epi and nektobenthic-). Abundance and stages of oceanic nektobenthic are inversely correlated with dynamic height. These species are mainly located at the north edge of the eddy and north of Menorca island.

In 2005 the main oceanographic process is the density front splitting the study area diagonally from north of Ibiza to south of Menorca. Larvae from mesopelagic habitat species are positively correlated with salinity and MLD, and inversely with dynamic height (as larvae of species from oceanic epibenthic and nektobenthic habitats). Only stages of both epibenthic habitats are correlated with currents this year, of north-south component in this case.

Abundances are mainly located at higher latitudes both in 2004 (larvae of neritic epibenthic and oceanic nektobenthic species) and 2005 (larvae of epibenthic, oceanic nektobenthic and mesopelagic species), where EW currents dominate. Only abundance and stage of larvae of mesopelagic species are correlated with longitude in 2004 (location of the eddy), while larvae of coastal epibenthic and mesopelagic in 2005 are more developed in the west and east respectively. Ontogenetic stage of larvae of oceanic nektobenthic species is inversely correlated with distance to the coast in 2004.

Microzooplankton is correlated with larvae of neritic epibenthic and oceanic nektobenthic species abundances in 2004 and with larvae of mesopelagic species in 2005. Chlorophyll-a is also relevant when it was not so abundant (2005) for all larvae of epi and nektobenthic species habitat.

Prototypes in the SOM have detected patterns in a subset of time of the day within 10 am-14.30 pm for 2004 and 7.30 am-16.30 pm for 2005. In 2004 this corresponds to negative correlation with EW currents, while in 2005 it corresponds to positive correlation with NS currents (though the correlation was slightly below 0.6 and it is not included in the figures). Larvae of neritic epibenthic, oceanic nektobenthic species and total larvae in 2004 (abundances and stages) have a nonlinear negative correlation with time and positive with currents. In 2005 larvae from two habitats correlate with the currents, but while larvae of neritic epibenthic' species modal stage correlate positively, larvae of oceanic epibenthic are negatively correlated. However, the sensitivity analysis in the appendix shows how these correlations with time of the day are not significative, as they disappear when varying the number of variables.

Discussion

We have presented the application of an unsupervised neural network (the Self-Organizing Map) to a combined dataset of hydrological and community data. This provides a global view of the ecosystem, obtaining nonlinear relationships within each dataset as well as between larvae and their environment. Although 48 variables can be considered a dataset of medium dimensionality, it is a good practise to check

18

²⁶⁰ that the same correlations are given with different subsets of the data. Results on the sensitivity to number
²⁶¹ of variables in the appendix prove the SOM as a robust technique, although small changes with weaker
²⁶² correlations recommend performing such analysis. Basically strong correlations remain when varying the
²⁶³ number of variables (for the variables present), but some new weaker correlations can appear when considering
²⁶⁴ less variables, as their correlation become a more relevant pattern after removing others. Also some weak
²⁶⁵ correlations disappear when reducing the number of variables; this is due to artifacts in the data, and is
²⁶⁶ mainly related to currents and time of the day for this data.

²⁶⁷ Correlation matrices in this work present some relationships which are known beforehand and due to
²⁶⁸ local characteristics, such as higher depths and distance to the coast on the east. Also some spurious
²⁶⁹ relationships that are the result of mediation or moderation processes (Baron and Kenny, 1986), such as
²⁷⁰ correlation between temperature and geographical areas due to the sampling order and summer warming
²⁷¹ during the survey. However, many relevant relationships were encountered, particularly the ones relating
²⁷² larvae abundance and development stage with their environment, but also the correlations defining in detail
²⁷³ the physical environment, commented in the following paragraph. Although the stage range is not common
²⁷⁴ for all species, it is similar for each habitat, and coincident for the most abundant species that were similar
²⁷⁵ both years. Species with maximum stage were well spread around the area for all habitats, showing no
²⁷⁶ spatial patterns, supporting the idea that stage range was not influencing the results for intra-year or inter-
²⁷⁷ year comparisons. In other cases, grouping stages in early, medium and late development would be more
²⁷⁸ appropriate, though some subjectivity would be introduced in the data, and the information of development
²⁷⁹ diminished.

²⁸⁰ Dynamic height delimits the main hydrodynamic processes each year (eddy and density front), but they
²⁸¹ are correlated with temperature at different depths. In 2004 the eddy is characterised by higher temperatures
²⁸² at the whole water column than the rest of the stations (giving higher correlation between dynamic height
²⁸³ and temperature at middle depths). In 2005 the density front presents warm surface waters similar to the
²⁸⁴ whole area, colder waters around 50 m depth characteristic of the recent AW, but warmer waters at the
²⁸⁵ deeper values than the rest of the area possibly due to resident water sinking, which produces the positive
²⁸⁶ correlation between dynamic height and temperature at 200 m depth this year.

²⁸⁷ The main variables correlated with abundances and developmental stage were dynamic height, temper-
²⁸⁸ ature, salinity, currents and food (microzooplankton and chlorophyll-a). Different oceanographic processes
²⁸⁹ each year lead to different larvae distributions. The larvae of mesopelagic habitat species are the most dif-
²⁹⁰ ferent in terms of spatial distribution as in 2004 they were trapped in the eddy located south of Menorca
²⁹¹ showing opposite relationships with several environmental variables than the other habitats, and in 2005
²⁹² depicts correlations with temperature, chlorophyll-a and fluorescence but not with salinities or currents as
²⁹³ the others.

²⁹⁴ A possible reason for the higher development of all larvae in 2005 could be the increment in nearly 2 ° C of
²⁹⁵ the surface temperature, influenced also by the later start (9 days) and end (13 days) of the survey that year.

Temperatures are clearly more influential on the distribution of stages than on their abundance distribution, with the first five larval groups correlated positively when adding both years (results not included). An increase in temperature produces faster developments and is often related to the occurrence and abundance of decapod larvae (Anger, 2001; Pan et al., 2011). Salinity seems to have a more persistent influence both in abundance and ontogenetic distribution. This is expectable, since it is a conservative parameter of the water masses and, in our case, characterise clearly the two surface water masses found in the study area. Considering the different dynamics of both water masses we should expect differences in larval communities between them. Thus, recent AW comes from the centre of the Algerian subbasin, where only a few larvae of very offshore species can be found, though sometimes sweep local individuals located over the shelves or slopes on their way north, whereas resident AW in our study area were found closer to the coasts, where most decapod larvae are located. Moreover, these water masses are characterised by different nutrient contents, which are usually higher in resident AW because of vertical mixing processes occurring in NW Mediterranean and the inputs from the rivers Rhône and Ebro, leading ultimately to higher microzooplankton abundances, which obviously can influence decapod larval abundances. Salinity is the 'driving force affecting the water-column density and thus stability' as with high salinity in surface layers and/or cool winters strong convection and nutrient input into the photic zone may occur (García-Comas et al., 2011), which explains the higher nutrient concentration in the resident waters. The Northern Current that surrounds the coast southwards from the Gulf of Lions to the west of the Balearic islands, transport these enriched resident AW to the Balearic Islands, reaching first the western part of the archipelago, which explain why resident AW located at the west of the study area has always higher nutrient content.

Strong positive correlations were found between latitude and most of the larvae groups. However no correlation with distance to the coast appeared, indicating no influence of the coast to larvae's northward distribution, but just the sweeping effect mentioned above. The northern distribution of larvae in 2004 show a relationship with EW currents located around Menorca mainly, perhaps because the larvae were previously taken from the slope while the waters run northeastward.

In 2004 satellite chlorophyll-a was high compared to 2005, suggesting high phytoplankton abundance (which is known to be the predated by younger decapods stages), and hence this would not be a limiting factor conditioning larval distribution. On the contrary, in 2005 its lower concentration, particularly at 25 m depth, triggered this cue for larvae distribution, often found in the literature (Brandao et al., 2012). On the other hand microzooplankton was correlated with decapod larval abundance both years. Larvae correlation with microzooplankton is weaker in 2004 (neritic epibenthic and oceanic nektobenthic species) but stronger for mesopelagic species in 2005, and when adding both years together for the total larvae. microzooplankton seem to be always relevant for some decapod larvae, while chlorophyll-a and fluorescence would have a threshold below which they influence their distribution. Integrated fluorescence correlated with mesozooplankton, as stated in Rodríguez et al. (2013). Nutrients input to euphotic layer in this area depends mainly on mesoscale oceanographic processes, as there isn't any river input, though in Rodríguez et al. (2013)

no correlation between fluorescence and dynamic height or currents was found. However, relevant nonlinear correlation (above 0.6) between dynamic height and fluorescence (mainly the integrated until 100 m depth) and chlorophyll-a are depicted by the SOM for both years.

Asch and Jr. (2013) define four mechanisms to explain the possible influence of the dynamical height to larvae distribution. Mechanism 1) positive correlation with dynamic height due to their reflection of high temperatures and low salinities, Mechanism 2) negative correlation near the coast, were the upwelling proportionates nutrients to larvae, Mechanism 3) positive correlation with dynamical heights' gradients, i.e. currents and Mechanism 4) positive/negative correlation with anticyclonic/cyclonic eddies that retain the larvae in a particular area. Mechanism 1 can be related to density fronts as the one in 2005, however abundances were not higher at the front. The second mechanism was neither found, as the abundances were in general more abundant far from the coast. Instead mechanism 3 or currents associated to both the front in 2005 and the eddy in 2004 seem to be more relevant. Finally, Mechanism 4 could reflect the case of mesopelagic species larvae, trapped inside the eddy in 2004. This phenomenon occurs in other areas such as the Canary islands (Landeira et al., 2010; Landeira, 2010) and Tasman Sea (Brandt, 1981; Griffiths and Brandt, 1983). Outside the eddy, the relationship with dynamic height is negative, indicating their location outside or at the limit of the eddy in 2004, and in resident waters in 2005.

Different strategies have been found in 2004 related to the presence of the anticyclonic eddy south of Menorca, with larvae of mesopelagic species located at the core of the eddy, while the rest are associated to the currents at the edges. Decapod communities located inside, outside and at the edge of eddies were described in Griffiths and Brandt (1983). Eddies Ekman pumping can affect phytoplankton growth and community composition (Anderson et al., 2011), causing different distributions between species/groups. They highlight the importance of including the eddy/wind interaction in simulations of the physics and biology of eddies.

Currents' influence on larvae horizontal distribution is the result of a passive transport modulated by the larvae vertical swimming, associated to its ontogenetic development and vertical migration, and active swimming for older stages. Larvae were considered passive particles until recently (Herbing, 2002; Reiss and Panteleev, 2000) but active opposition to vertical flow has been proven (Genin et al., 2005; Weidberg, 2012). According to this work, larvae distribution seem to have some link to favourable currents, that could be modulated by other factors. Obviously, the spatial distribution of spawners, which determine the point of origin of the larvae, plays a key role in the spatial distribution of the larvae of different species. Abundance is not related to the higher currents intensities, which could be the result of currents gathering larvae on their way, but to particular directions and areas different for each year. The magnitude and spatial structure of an eddy's Ekman pumping (with average velocities in the interior of midlatitude anticyclones exceeding 2.5 cm/day (Gaube, 2012)) is a (nonlinear) function of both the magnitude and geographic structure of the SST anomalies and the direction and magnitude of the background wind (Gaube, 2012), and thus EW component, when located near the coast as in 2004, could increment the nutrients brought to the surface by this pumping, while NS is just employed as a land approaching/separating stream in 2005. With fronts and

21

eddies affecting this area regularly, and the reduced extension of the Balearic islands' shelf (as seen in Fig. 1) -meaning that most locations are potentially affected by offshore processes-, this could be a major issue for larvae retention.

Late larval stages are located in 2005 at the limit of the front between resident and recent AW (with higher currents), showing difference in ontogenetic distribution patterns. However, currents do not correlate significantly with stations of higher chlorophyll-a, fluorescence or microzooplankton for any year. Rodríguez et al. (2013) found that fish larvae in the same area for 2005 mainly stayed at their adult habitat location, i.e. neritic remain close to the coast and oceanic species do not approach much the coast. Decapod larvae showed in this work a distinct ontogenetic tendency in the spatial distribution, suggesting passive transport or active movement during 2005 for both epibenthic and just oceanic nektobenthic larvae; longer duration of development stages could also lead to more prolonged drifting.

Conclusions

This work highlighted how the dynamical processes in the whole water column, particularly in a mesoscale area such as the Balearic sea, influences the distribution of larvae individuals. Abundance data were grouped by adult habitat due to their influence in the location of the spawning area, that conditions the initial distribution of larvae and thus their final dispersion. Clear distribution's differences were found between these groups, that reflect the coastal or oceanic habitat of adults, but also links to different features of the mesoscale activity area in Baleares such as eddies' center/edges or front currents. Investigation of the different aspects of dynamic height (absolute values, gradients or edges and correlations) helped to better understand the influence of the oceanographic processes on decapods. Larvae of mesopelagic species showed up to be the most different group, associated to the interior of the eddy and higher depths in general, while the rest of the habitats were located at the edges of the eddy found in 2004. Most larvae were related to resident waters in the presence of a front in 2005, excepting the nektobenthonic found on the west where the front moved upwards. Deep salinities seem to drive abundance and stages while temperature mainly influences stage development. Microzooplankton was the only prey variable correlated with larvae in 2004, though Chla was also relevant when not so abundant (2005). Finally the inclusion of stage information produced an estimation of the path followed by some larvae groups during their lifespan, according to the water dynamics encountered.

Acknowledgements

22

surveys and/or carried out the lab analysis that generated some of the considered environmental variables. The use of SOCIB data and products is gratefully acknowledged. We thank the anonymous reviewers for the constructive comments on the manuscript.

**Appendix: SOM sensitivity to variation in the number of variables**

In this work we have included 48 variables with a sample size of 108-110 observations (Case I). The SOM technique has been successfully applied in previous studies to the so called "High-Dimensional and Low Sample Datasets" (HDLSD) such as in Rauber et al. (2000). Nevertheless the relation between number of variables and samples may affect the statistical significance of results (Suzuki and Shimodaira, 2006; Liu et al., 2008), as it might be an artifact of sampling fluctuations. Even thought this effect is less relevant in unsupervised learning (technique used in this study) than in supervised learning, we tested the sensitivity of SOM to the number of variables using 24 (Case II) and 12 (Case III) variables, and compared the results with Case I; they should give the same output for the variables appearing each time. In the first case, we keep the main variables giving correlations in Case I, and the second case includes just oceanographic variables. For temperatures and salinities, we keep the value at the mixing depth as a representation of the surface, and at 100 and 200 m depth, to check that the same correlations with deeper waters are found. Results appear on Figures .8, .9, .10 and .11. We can appreciate how most of the correlations are the same for all cases; slight variations in the decimals make some weaker correlations around 0.6 to appear or disappear in some cases due to the visualisation of correlations above 0.6. Note that we only consider significant correlations ($p < 0.05$), which removes some of the weaker correlations. SOM projections for different number of variables project differently the values into the map; this means that is the relative difference between variables in the same map that have to be compared.

Varying the variables considered maintains the most important patterns, but may slightly change weaker correlations. These weaker correlations may become more relevant when considered in an study case of lower number of variables. Examples are correlations between the salinity at 200 m and salinity at 100 and with temperature at the MLD, and between temperature at MLD and depth in 2004. In other cases, correlations present with more variables, disappear when considering less variables; this is due to the mentioned artifacts. Examples are the correlation between salinity at the MLD and the EW current in 2004 and between EW current and salinity at MLD in 2005, and all correlations with time of the day. Considering correlations above 0.8 for instance may be a safer choice.

References

Alemany, F., Deudero, S., Morales-Nín, B., López-Jurado, J., Jansà, J., Palmer, M., Palomera., I., 2006. Influence of physical environmental factors on the composition and horizontal distribution of summer larval fish assemblages off Mallorca island (Balearic archipelago, western Mediterranean. J. Plankton Res 28, 473–487.

Figure .8: SOM applied to larvae data split by habitat. Projected variables are ordered by absolute correlation. Areas of higher distance between neurons (above 0.6) are depicted as shaded gray areas. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic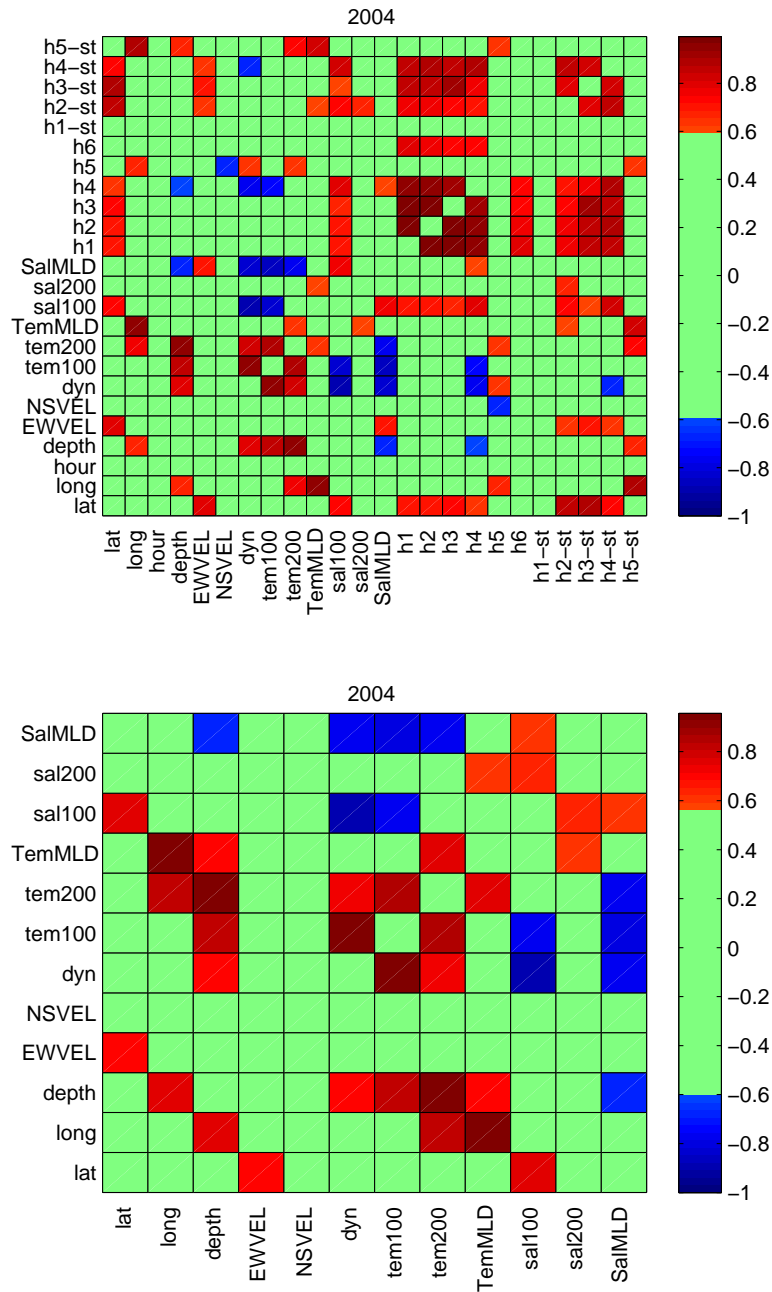), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic). TempX and SalX: temperature and salinity at X metres. See table 1 of the manuscript for more variable explanation.
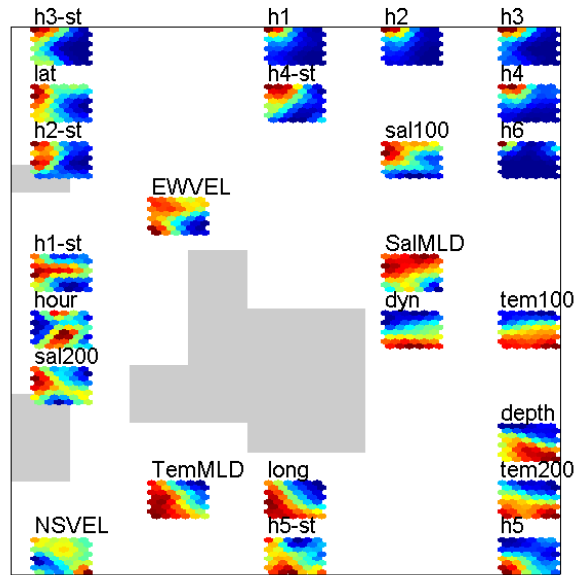
(a) 2004 24 variables
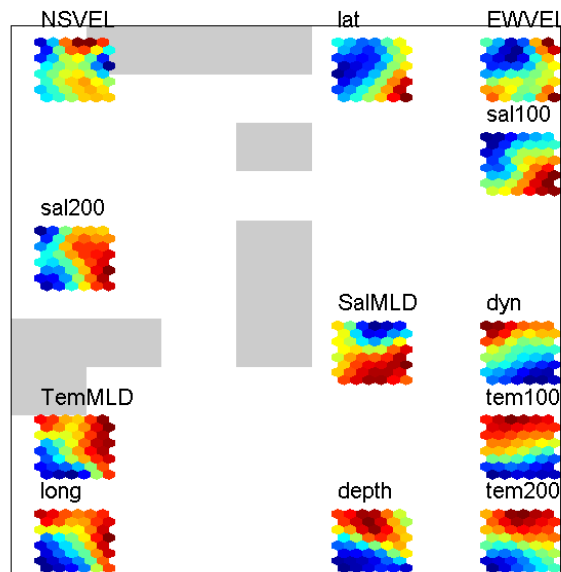
(b) 2004 12 variables

Figure .9: Correlations for 2004 with 24 (top) and 12 variables (bottom). Only significant correlations ($p < 0.05$) with absolute values above 0.6 are shown for clarity. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic).

Figure .10: SOM applied to larvae data split by habitat. Projected variables are ordered by absolute correlation. Areas of higher distance between neurons (above 0.6) are depicted as shaded gray areas. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic). TempX and SalX: temperature and salinity at X metres. See table 1 for more variable explanation.
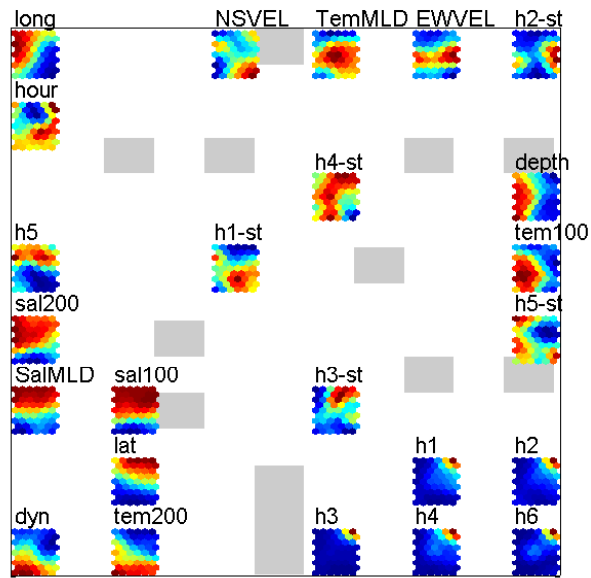

(a) 2005 24 variables
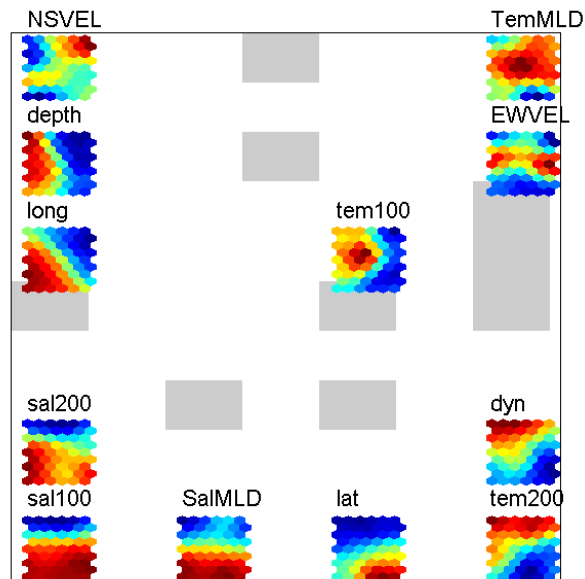
(b) 2005 12 variables

Figure .11: Correlations for 2005 with 24 (top) and 12 variables (bottom). Only significant correlations ($p < 0.05$) with absolute values above 0.6 are shown for clarity. h1: Epibenthic shelf (neritic), h2: Epibenthic slope & bathyal (oceanic), h3: Nektobenthic shelf (neritic), h4: Nektobenthic slope & bathyal (oceanic), h5: Mesopelagic (oceanic) and h6: Endobenthic (neritic).
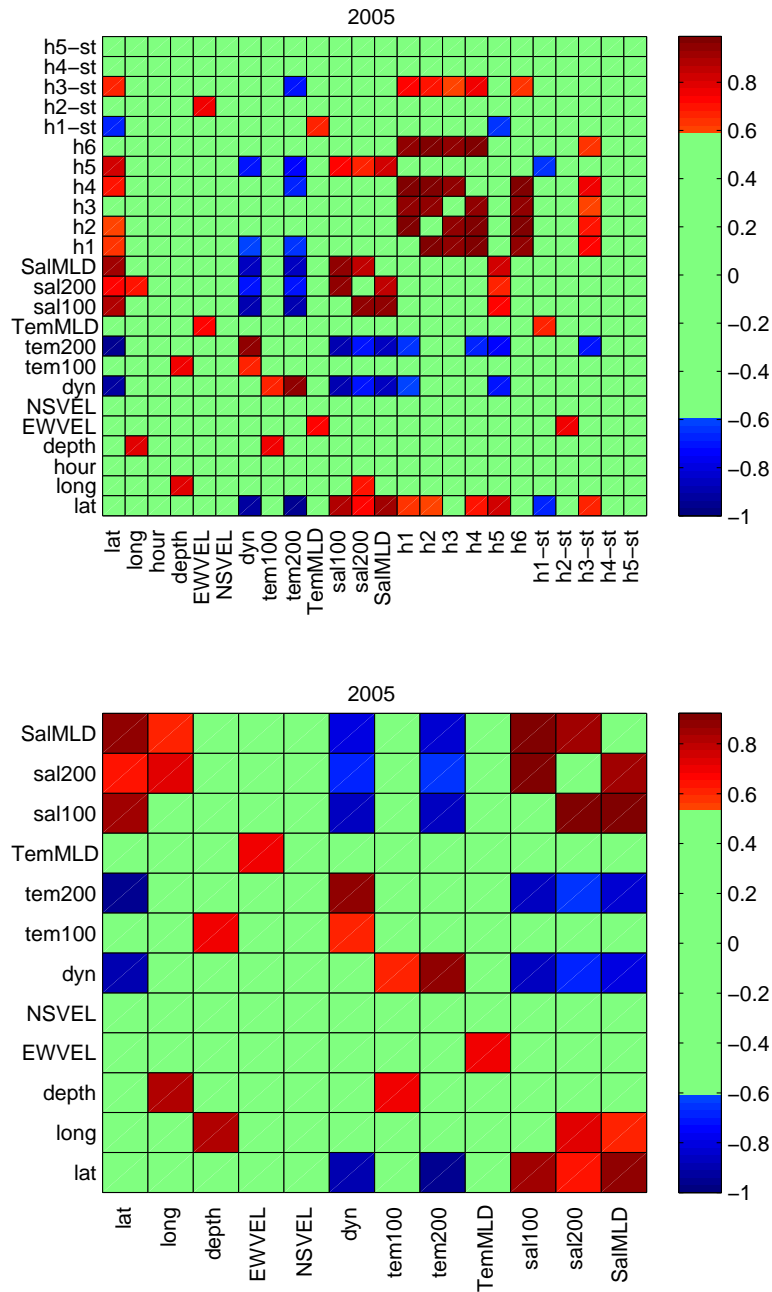
Alemany, F., Quintanilla, L., Vélez-Belchi, P., García, A., Cortés, D., Rodríguez, J., Fernández de Puelles, M., González-Pola, C., López-Jurado, J., 2010. Characterization of the spawning habitat of Atlantic bluefin tuna and related species in the Balearic Sea (western Mediterranean). Prog. Oceanogr 86, 21–38.

Anderson, L., Jr., D.J.M., Maltrud, M.E., Lima, I.D., Doney, S., 2011. Impact of eddy-wind interaction on eddy demographics and phytoplankton community structure in a model of the North Atlantic Ocean. Dynam. Atmos. Oceans 52, 80 – 94.

Anger, K., 2001. The Biology of Decapod Crustacean Larvae. volume 14. Crustacean Issues.

Asch, R.G., Jr., D.M.C., 2013. Dynamic height: A key variable for identifying the spawning habitat of small pelagic fishes. Deep-sea Res. Pt. I 71, 79 – 91.

Bailey, K., 2002. Complex processes in the survival of walleye pollock larvae and forecasting implications. Fish. Sci 68, 200–205.

Balbín, R., López-Jurado, J., Flexas, M., Reglero, P., Vélez-Velchi, P., González-Pola, C., Rodríguez, J., García, A., Alemany, F., 2013. Interannual variability of the early summer circulation around the Balearic Islands: driving factors and potential effects on the marine ecosystem. J. Mar. Syst .

Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol 51, 1173–1182,.

Basterretxea, G., Jordi, A., CatalaN, I.A., Sabates, A., 2012. Model-based assessment of local-scale fish larval connectivity in a network of marine protected areas. Fish. Oceanogr 21, 291–306.

Bradbury, I.R., Snelgrove, P.V., 2001. Contrasting larval transport in demersal fish and benthic invertebrates: the roles of behaviour and advective processes in determining spatial pattern. Can. J. Fish. Aquat. Sci 58, 811–823.

Brandao, M.C., Koettker, A.G., Freire, A.S., 2012. Abundance and composition of decapod larvae at Saint Paul's Rocks (equatorial Atlantic). Mar. Ecol , 1–15.

Brandt, S., 1981. Effects of a Warm-Core Eddy on Fish Distributions in the Tasman Sea Off East Australia. Mar. Ecol. Prog. Ser. 6, 19–33.

Chon, T.S., 2011. Self-organizing maps applied to ecological sciences. Ecol. Inform 6, 50 – 61.

Chon, T.S., Park, Y.S., Moon, K.H., Cha, E.Y., 1996. Patternizing communities by using an artificial neural network. Ecol. Model 90, 69 – 78.

Crosnier, A., Forest, J., 1973. Les crevettes profondes de l'Atlantique Oriental Tropical.

Dingsør, G., Ciannelli, L., Chan, K.S., Ottersen, G., Stenseth, N., 2007. Density dependence and density independence during the early life stages of four marine fish stocks. Usda. Ne. Exp 88(3), 625–634.

Dixon, P.A., Milicich, M., Sugihara, G., 1999. Episodic fluctuations in larval supply. Science 283, 1528–1530.

Dixon, P.A., Milicich, M.J., Sugihara, G., 2001. Noise and nonlinearity in an ecological system, in: Nonlinear dynamics and statistics. Springer, pp. 339–364.

Dos Santos, A., González-Gordillo, J., 2004. Illustrated keys for the identification of the Pleocyemata (Crustacea: Decapoda) zoeal stages, from the coastal region of south-western Europe. J. Mar. Biol. Assoc. Uk 84(1), 205–227.

Dos Santos, A., Lindley, J., 2001. Crustacea Decapoda: Larvae. II. Dendrobranchiata (Aristeidae, Benthesi-cymidae, Penaeidae, Solenoceridae, Sicyonidae, Sergestidae & Luciferidae). ICES Identif. Leafl. Plankton 186, 1–9.

Foody, G.M., 1999. Applications of the self-organising feature map neural network in community data analysis. Ecol. Model 120, 97 – 107.

García-Comas, C., L.Stemmann, Ibánez, F., Berline, L., Mazzocchi, M.G., Gasparini, S., Picheral, M., Gorsky, G., 2011. Zooplankton long-term changes in the NW Mediterranean Sea: Decadal periodicity forced by winter hydrographic conditions related to large-scale atmospheric changes? J. Mar. Syst 87, 216 – 226.

Gaube, P., 2012. Satellite Observations of the Influence of Mesoscale Ocean Eddies on Near-Surface Temper-ature, Phytoplankton and Surface Stress. Ph.D. thesis. Earth, Ocean, and Atmospheric Sciences, Oregon State University.

Genin, A., Jaffe, J.S., Reef, R., Richter, C., Franks, P.J.S., 2005. Swimming against the flow: A mechanism of zooplankton aggregation. Science 308, 860–862.

Giraudel, J., Lek, S., 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecol. Model 146, 329–339.

González-Gordillo, J., dos Santos, A., Rodríguez, A., 2001. Checklist and annotated bibliography of decapod crustacean larvae from the Southwestern European coast (Gibraltar Strait area). Sci. Mar 65, 275–305.

Griffiths, F., Brandt, S., 1983. Distribution of mesopelagic decapod Crustacea in and around a warm-core eddy in the Tasman Sea. Mar. Ecol. Prog. Ser. 12, 175–184.

Herbing, I.H.v., 2002. Effects of temperature on larval fish swimming performance: the importance of physics to physiology. J. Fish. Biol 61, 865–876.

Hsieh, C., Glaser, S.M., Lucas, A.J., Sugihara, G., 2005. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435, 336–340.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern 43(1), 59–69.

Kosiba, P., Stankiewicz, A., Mroz, L., 2010. Modelling of habitat conditions by self-organizing feature maps using relations between soil, plant chemical properties and type of basaltoides. Acta Soc. Bot. Pol 79, 315–323.

Kwon, Y.S., Li, F., Chung, N., Bae, M.J., Hwang, S.J., Byoen, M.S., Park, S.J., Park, Y.S., 2012. Response of fish communities to various environmental variables across multiple spatial scales. Int. J. Environ. Res. Public. Health 9, 3629–3653.

Landeira, J., Lozano-Soldevilla, F., Hernández-León, S., Barton, E., 2010. Spatial variability of planktonic invertebrate larvae in the Canary Islands area. J. Mar. Biol. Assoc. Uk 90, 1217–1225.

Landeira, J.M., 2010. Larvas planctónicas de crustáceos decápodos en las Islas Canarias. Ph.D. thesis. Departamento de Biologia Animal. Universidad de La laguna.

Largier, J.L., 2003. Considerations in estimating larval dispersal distances from oceanographic data. Ecol. Appl 13, 71–89.

Li, F., Cai, Q., Qu, X., Tang, T., Wu, N., Fu, X., Duan, S., Jähnig, S.C., 2012. Characterizing macroinvertebrate communities across China: Large-scale implementation of a self-organizing map. Ecol. Indic 23, 394–401.

Lindley, J., 1986. Vertical distributions of decapod crustacean larvae and pelagic post-larvae over Great Sole Bank (Celtic Sea) in June 1983. Mar. Biol 90, 545–549.

Liu, Y., Weisberg, R.H., Mooers, C.N., 2006. Performance evaluation of the self-organizing map for feature extraction. J. Geophys. Res.: Oceans (1978–2012) 111.

Liu, Y., Hayes, D.N., Nobel, A., Marron, J.S.,2006. Statistical Significance of Clustering for High-Dimension, Low Sample Size Data. J. of the American Stat. Assoc.) 103(483), 1281–1293.

Lluch-Belda, D., Lluch-Cota, D.B., Hernandez-Vazquez, S., Salinas-Zavala, C.A., Schwartzlose, R.A., 1991. Sardine and anchovy spawning as related to temperature and upwell in the California current system. Technical Report. CalCOFI Repor.

Muhling, B.A., Lamkin, J.T., Roffer, M.A., 2010. Predicting the occurrence of Atlantic bluefin tuna (Thunnus thynnus) larvae in the northern Gulf of Mexico: building a classification model from archival data. . Fish. Oceanogr 19(6), 526–539.

Muhling, B.A., Reglero, P., Ciannelli, L., Alvarez-Berastegui, D., Alemany, F., Lamkin, J.T., Roffer, M.A., 2013. Comparison between environmental characteristics of larval bluefin tuna Thunnus thynnus habitat in the Gulf of Mexico and western Mediterranean Sea. Mar. Ecol. Prog. Ser. 486, 257–276.

Paini, D.R., Worner, S.P., Cook, D.C., Barro, P.J.D., Thomas, M.B., 2010. Using a self-organizing map to predict invasive species: sensitivity to data errors and a comparison with expert opinion. J. Appl. Ecol 47, 290–298.

Pan, M., Pierce, G.J., Cunningham, C.O., Hay, S.J., 2011. Seasonal and interannual variation of decapod larval abundance from two coastal locations in Scotland, UK. J. Mar. Biol. Assoc. Uk 91, 1443–1451.

Peña, M., 2007. Latent Variable Spaces For The Construction Of Topology Preserving Mappings. Computing and Information Systems Technical Reports 41, 14299.

Peña, M., Fyfe, C., 2006. Forecasting with topology preserving maps: Harmonic Topographic Map and Topographic product of experts application., in: I International Conference on Multidisciplinay Information Sciences and Technologies (InSciT2006). Mérida.

Peña, M., Georgakarakos, S., Boyra, G., 2008. Visualisation of the clustering of acoustic echotraces using Topology Preserving Mapping. Application to juvenile anchovy in the Bay of Biscay, in: ICES CM 2008/R:15.

Petitgas, P., Vaz, S., Loots, C., Peck, M., Rockmann, C., Hal, R., Tulp, I., 2008. Cross mapping and statistical analyses quantifying the likely impacts of climate change on suitable habitat for different life stages of key species in different regions. Technical Report. Reclaim project, Habitat mapping Report.

Rauber, A., Tomsich, P., Merkl, D., 2000. parsom: A parallel implementation of the self-organizing map exploiting cache effects: making the som fit for interactive high-performance data analysis, in: Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, IEEE. pp. 177–182.

Recknagel, F., 2001. Applications of machine learning to ecological modelling. Ecol. Model 146, 303 – 310.

Reglero, P., Ciannelli, L., Alvarez-Berastegui, D., Balbín, R., López-Jurado, J., Alemany, F., 2012. Geographically and environmentally driven spawning distributions of tuna species in the western Mediterranean Sea. Mar. Ecol. Prog. Ser. 463, 273–284.

Reiss, C.S., Panteleev, G., 2000. Observations on larval fish transport and retention on the Scotian Shelf in relation to geostrophic circulation. Fish. Oceanogr 9, 195–213.

Richardson, A., Risien, C., Shillington, F., 2003. Using self-organizing maps to identify patterns in satellite imagery. Prog. Oceanogr 59, 223–239.

Rodríguez, J., Alvarez, I., López-Jurado, J., García, A., Balbín, R., Alvarez-Berastegui, D., Torres, A., Alemany, F., 2013. Environmental forcing and the larval fish community associated to the Atlantic bluefin tuna spawning habitat of the Balearic region (western Mediterranean), in early summer 2005. Deep-sea Res. Pt. I 77, 11 – 22. J. Exp. Mar. Biol. Ecol 186, 1 – 16.

Shanmuganathan, S., Sallis, P., Buckeridge, J., 2003. Ecological modelling with self-organising maps., in: Proceedings of the MODSIM2003 Biennial Conference of the Modelling and Simulation Society of Australia and New Zealand.

Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering, Bioinformatics 22, 1540 – 1542.

Stojkovic, M., Simic, V., Milosevic, D., Mancev, D., Penczak, T., 2013. Visualization of fish community distribution patterns using the self-organizing map: A case study of the Great Morava river system (Serbia). Ecol. Model 248, 20 – 29.

Torres, A.P., Reglero, P., Balbín, R., Urtizberea, A., Alemany, F., 2011. Coexistence of larvae of tuna species and other fish in the surface mixed layer in the NW mediterranean. J. Plankton Res .

Torres, A.P., Santos, A.D., Balbn, R., Alemany, F., Massut, E., Reglero, P., 2013. Decapod crustacean larval communities in the balearic sea (western mediterranean): Seasonal composition, horizontal and vertical distribution patterns. J. Mar. Syst , –.

Udekem d'Acoz, C., 1999. Inventaire et distribution des crustacés décapodes de l'atlantique nord-oriental, de la Méditerranée et des eaux continentales adjacentes au nord de 25°N. Coll. Patrim. Nat. 40, 1–383.

Vesanto, J., Ahola, J., 1999. Hunting for Correlations in Data Using the Self-Organizing Map, in: Proceeding of the International ICSC Congress on Computational Intelligence Methods and Applications (CIMA99), ICSC.

Weidberg, N., 2012. Caracterización de la distribución espacial y temporal del meroplancton en la costa asturiana. Ph.D. thesis. Universidad de Oviedo.

Zariquiey-Alvarez, R., 1968. Crustáceos decápodos ibéricos. Investigación Pesquera : publicación de ciencia marina y pesquera, Consejo Superior de Investigaciones Científicas, Patronato Juan de la Cierva.