Estuarine, Coastal and Shelf Science
September 2021, Volume 258, Pages 107410 (12p.)
https://doi.org/10.1016/j.ecss.2021.107410
https://archimer.ifremer.fr/doc/00695/80754/

Using self organizing maps to analyze larval fish assemblage vertical dynamics through environmental-ontogenetic gradients

Álvarez I. [1, *], Font Munoz Joan Salvador [1, 3], Hernández-Carrasco I. [1], Díaz-Gil C. [1, 2], Salgado-Hernanz P.M. [1], Catalán I.A. [1]

[1] IMEDEA (UIB-CSIC). Miquel Marqués 21, 07190, Esporles, Spain
[2] Laboratori d' Investigacions Marines i Aqüicultura, LIMIA (Balearic Government), Port d'Andratx, Illes Balears, Spain
[3] IFREMER, French Institute for Sea Research, DYNECO PELAGOS, 29280, Plouzané, France

* Corresponding author : I. Alvarez, email address : itziar@imedea.uib-csic.es

Abstract :

We analyzed the influence of the stratification process in the vertical distribution of larval fish in a microtidal coastal Mediterranean zone. By applying a Self Organizing Maps (SOM) technique, we could analyze a complex dataset accounting for non-linear processes. The analysis integrated multivariate data on larval fish and environmental parameters in two depth strata through two-time components (nictimeral and fortnightly through the main spawning seasons), and considered size-based information. Although causal relationships cannot be constructed, the use of SOM analyses enabled the description of the whole system evolution overcoming the constraints of linear approaches in complex multivariate datasets with multiple dependencies in the data. We contend that this approach can help to unveil the intricate patterns of settlement/recruitment of young fish, which is often hampered by the rigidity of some formal statistical approaches.

Highlights

► Larval fish assemblages analysis with Self Organizing Maps allows overcoming linear constraints. ► Self Organizing Maps analysis allowed the description of the whole coastal multispecific system. ► Self Organizing Maps also allowed identifying ontogenetic differences in vertical positions. ► Differences in larval fish vertical strategies before & after the establishment of stratification.

Keywords : Keywords: larval fish assemblages, self-organizing maps, neural networks, seasonal thermocline, vertical distribution

1. Introduction

The persistence and resilience of fish populations depend on their adaptation success at multiple scales through space and development. Bottlenecks in fish populations are often related to variability in the survival of larvae and juveniles (Houde, 2008; Nash and Geffen, 2012) and the study of their ecological adaptations has been paramount in marine ecology and fisheries science since Hjort (1926, 1914). Juvenile and adult stages of fish mainly inhabit a world dominated by inertial forces. Under those conditions, they can control their position at the horizontal scale and have more options to respond to environmental stressors. On the contrary, eggs and early larval stages inhabit a world of frictional forces, and their low (and variable) accumulated survival depends on a combination of parental adaptations to spawning (Cianelli et al., 2015) and larval adaptations to an environment over which they have more limited control (Houde and Schekter, 1980). Due to the importance of larval fish distribution and vertical positioning for recruitment (e.g., Sinclair and Iles, 1985), much research has been devoted to the analysis of larval fish movement at the vertical dimension and, at more advanced stages, at the horizontal scale (reviewed in Kingsford et al., 2002; Leis, 2006; Montgomery et al., 2001). Larval fish have a species-specific and ontogenetic-dependent capacity to conduct vertical migrations, which probably responds to trade-offs that maximize survivorship (Vikebø et al., 2005). Vertical migrations have been documented both at coastal/estuarine systems and over deep water columns. These migrations are interpreted as responses to food distribution and light, tides, stratification, predator abundance, or turbulence (see cited reviews). Studies on vertical movements usually concentrate on one or a few species and typically analyze their findings into one or a few stages or size classes. The temporal analysis of vertical patterns of multiple species and development at a high resolution is seldom conducted (but see Rodriguez et al., 2011). In addition, the usual approach with multivariate statistical methods that tend to extract only linear dependencies (or non-linear, limited to single species), limits our ability to understand how assemblages respond to complex environments.

One of the key controllers of larval fish dynamics at the vertical scale is stratification (Ropke, 1993; Smith et al., 1999). Thermocline formation during the stratification period acts as a physical barrier to vertical injection of subsurface nutrients to the upper layer, reducing energy inputs and generating vertical segregation of resources (Mann and Lazier, 1991). Therefore, it affects prey fields and causes physiological adaptative effects on growth, respiration, mortality and nutrient recycling (Arvola et al., 2017;

Nøttestad et al., 2016). In shallow coastal environments, surface water reaches high temperatures in summer, intensifying the segregation of resources and affecting thermal tolerances of some organisms, including fish larvae (Bensoussan et al., 2010).

The vertical distribution of fish larvae in these shallow systems may influence dispersal (Paris and Cowen, 2004; Sponaugle et al., 2002). This is the case in microtidal areas where, for example, larval retention near reefs is favoured close to the bottom layers due to the weaker flow conditions (Leis, 1986).

In the last decades, increased stratification of coastal waters in temperate areas, such as the Mediterranean, has been documented (e.g., Coma et al., 2009). The knowledge of the mechanisms through which fish larvae manage to inhabit stratified waters and how juveniles find the way to recruit to the adult's habitat gains interest. In the Mediterranean, the location of fish spawning is determined by the structure of the bathymetry, types of sea bottoms, diversity of adult fish habitats, and mechanisms conditioning the primary production of the region. In contrast, the physical processes determine the final distribution patterns of fish eggs and larvae (e.g., shelf-slope density front and associated current, continental water inflows, winter mixing, stratification of the water column) (Sabatés et al., 2007). Whereas some understanding exists on vertical movements of key species in open waters (e.g., Olivar et al., 2018, 2014; Olivar and Sabatés, 1997; Sabatés, 2004), little is known about the temporal-spatial ecology at the vertical scale within larval assemblages (Vargas-Yáñez and Sabatés, 2007). We contend that monitoring many species of larval fish at the vertical scale and their development during the advent of critical physical processes in coastal systems, can fill a gap in the knowledge on the relevant mechanisms influencing coastal fish populations.

We examined the relationship between the environmental variables during the thermocline formation and the distribution of fish larvae in the water column regarding stratification and larval ontogenetic changes in size. The complexity of the involved processes in larval fish dynamics requires analytical methods capable of including the non-linearity of the data such as Self Organising Maps (SOM) (Kohonen, 1982). This technique has been used in several disciplines, including climate and meteorology (Kolczynski and Hacker, 2014), physical and biological oceanography (Basterretxea et al., 2018; Hernández-Carrasco et al., 2018; Hernández-Carrasco and Orfila, 2018; Liu et al., 2006), genetics (Sharon et al., 2013) and even aspects of larval fish ecology (Russo et al., 2014). In our case, we used SOM to elucidate patterns in complex larvae assemblages. SOM is essentially used as a dimension reduction technique, and provides

advantages as compared to other widely used statistical techniques such as PCA, K-means, or other multivariate analysis (e.g., NMDS, CCA, or RDA) for several reasons: i) it allows an appropriate analysis of high dimensional data, such as the high-throughput results of the analysis of the samples; ii) it can identify and extract the underlying non-linear data structures, i.e., using step-like or Gaussian functions for the neighborhood relationship; iii) it is an efficient method for feature extraction and afterward classification, providing patterns that explain the overall behavior of the system and iv) since SOM preserves topology, patterns are ordered in the neural network according to their similitude, which finally can be easily interpreted by visual inspection (Liu et al., 2006; Mörchen and Ultsch, 2005).

In this work, physical and biological variables such as temperature, current speed, wave height, phytoplankton, micro, and mesozooplankton biomass, were vertically sampled through the water column and through time, so the distribution of early stages of fish in relation to stratification characteristics could be determined. We hypothesized that using SOM on this complex system we could reveal if i) the vertical distribution of (some) larval fish species changes along with the multivariate environmental shift towards a stratified system including day/night variation, and ii) if the ontogenetic stage plays an essential role in the changes in larval fish vertical distribution.

2. MATERIALS AND METHODS

2.1 Field Sampling and Data Acquisition

The field study was conducted at a single station located over a depth of 40 m (2.67ºE; 39.48ºN) in the inner part of Palma Bay (Mallorca Island, NW Mediterranean, Fig. 1). Palma Bay was chosen due to the available abundant information on plankton dynamics in relation to physics (Alós et al., 2014; Álvarez et al., 2015b; Basterretxea et al., 2012; Font-Muñoz et al., 2018).

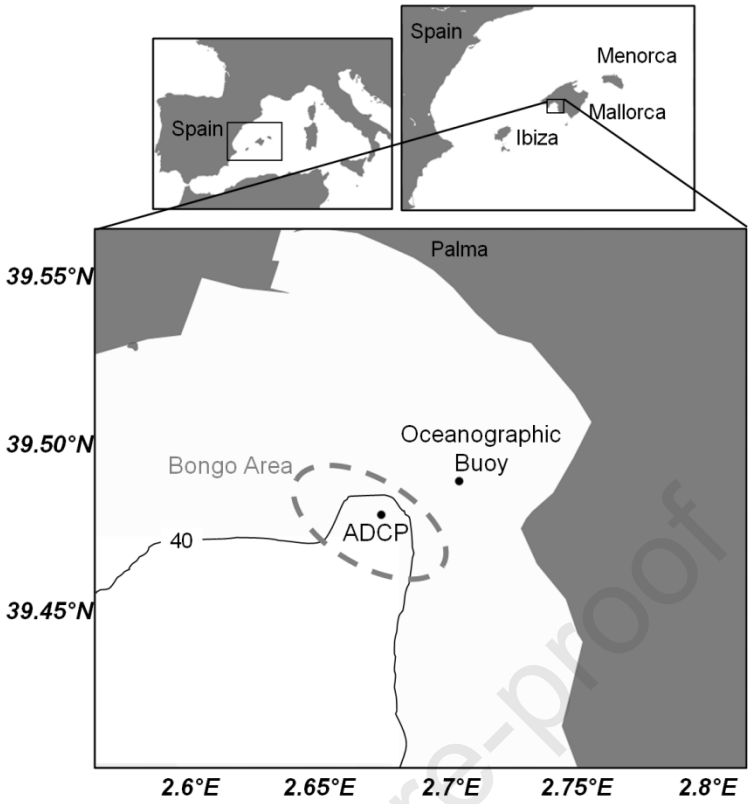

Figure 1: Map of the study area including position of the Sampling Station where the ADCP was moored and the Oceanographic Buoy. The grey ellipse indicates the area covered by the Bongo 40 tows. The 40 m isobath is also displayed

Twelve surveys were fortnightly performed between April and September 2014 (from spring to autumn). Each survey consisted of two sampling periods, conducted both at day (around 10 GMT) and night time (just after dusk). A bottom-mounted Acoustic Doppler Current Profiler (ADCP, 1 MHz Nortek Aquadopp) was moored at ~ 40 m depth to characterise flow variations during all the time series in the sampling area.

Wave height information was obtained from an oceanographic buoy located in Palma Bay (2.70°E; 39.49°N, Fig. 1). Data were extracted from http://thredds.socib.es/thredds/catalog/mooring/waves_recorder/buoy_bahiadepalma-scb_wave002/L1/2014/catalog.html. The wave height (significant height) value considered for each sampling was the mean value of the 3 hours prior to each sampling.

During each sampling, continuous temperature and salinity profiles were obtained with a SBE-25 CTD and an attached handheld deployable YSI-CastAway CTD profiler. In addition, water samples for chlorophyll analysis were collected at five depths (5, 10, 20, 30 and 40 m) using a 2.5 L Niskin bottle. Once on-board, water was filtered with GFF filters that were immediately frozen at -20°C.

The instantaneous profiles of temperature from the CastAway were used to determine the thermocline depth and adapt the zooplankton sampling.

For micro and mesozooplankton, two consecutive hauls were conducted: one covering the whole water column (from 2 m above the sea floor to the surface) and the second above the thermocline depth. Microzooplankton was sampled with vertical hauls of a 30cm diameter WP2 net fitted with a 53 µm nylon mesh. Mesozooplankton samples were obtained with a Bongo-40 net. One side was equipped with a 335 µm mesh net for ichthyoplankton identification and abundance estimates, and the other side was equipped with a 200 µm mesh for the mesozooplankton biomass estimates. The Bongo-40 tows over the whole water column consisted of a double-oblique tow down to 2 m above the seafloor. The above-thermocline hauls consisted of triple-oblique tows from thermocline depth to surface. Both types of tows were conducted at a boat speed of ~2 knots. The volume of water filtered was estimated by two GO 230 flowmeters mounted in the mouth of each net. Samples were preserved immediately after collection in 2% seawater borax buffered formalin.

2.2 Sample treatment

Chlorophyll (Chl-a, mg m^{-3}) was estimated from 24-hour acetone (90%) extracts using a Turner Designs bench fluorometer.

Microzooplankton samples were fractionated through 53 and 200 µm mesh filters, and only the fraction between 53-200 µm was retained. Dry weight from this fraction was obtained following Lovegrove (1966), and data were standardized to mg m^{-3}.

The 200 µm mesh of the Bongo samples was used to determine mesozooplankton dry weight. Samples were filtered through 200 and 500 µm meshes, and the 200-500µm fraction retained. Dry weight was estimated and standardized as for microzooplankton.

Fish larvae were sorted from the 335 µm net of the Bongo samples and identified to the lowest possible taxonomic level using a stereoscopic microscope. Larval abundance values were standardized to surface area (ind. 10 m^{-2}). A picture of each larva was taken with a digital camera mounted on the stereomicroscope, and total length (TL) measures were calculated with ImageJ software (Schneider et al., 2012) to the nearest 0.01mm. No correction for shrinkage was made. In order to facilitate the description of the taxa collected, class lengths were described as follows: L1 groups all larvae with sizes comprised between 1 and 2 mm; L2 groups the 2.01-3 mm larvae, and so forth until L14

for larvae with lengths ranging between 14.01 and 15 mm. In addition, for each taxon, larvae were grouped into pre and postflexion groups following the available bibliography (Table 1).

Table 1. Notochord flexion lengths from published literature sources. **FL**: Flexion length in mm. *Somarakis and Nikolioudakis, 2010

Family	Lower Taxa Identified	Reference Taxa	FL (mm)	Area	Reference
Ammodytidae	*Gymnammodytes cicerelus*	*Gymnammodytes cicerelus*	11	Balearic Sea	Alemany, 1997
Argentinidae	Sphyraenidae	*Argentina sphyraena*	13-17	Portugal	Ré and Meneses, 2009
Blenniidae	Blenniidae	*Parablennius tentacularis*	6.4	Balearic Sea	Alemany 1997
Bothidae	*Arnoglossus* spp.	*Arnoglossus laterna*	6	Balearic Sea	Alemany 1997
Bothidae	*Bothus* spp.	*Bothus podas*	7	Balearic Sea	Alemany 1997
Callionymidae	Callyonimidae	Callyonimidae A	4.1	Balearic Sea	Alemany 1997
Carangidae	*Trachurus* spp.	*Trachurus mediterraneus*	4.8	Balearic Sea	Alemany 1997
Centracanthidae	*Spicara flexuosa*	*Spicara smaris*	5.5	Balearic Sea	Alemany 1997
Centracanthidae	*Spicara smaris*	*Spicara smaris*	5.5	Balearic Sea	Alemany 1997
Clupeidae	*Sardinella aurita*	*Sardinella aurita*	7.5-10	Atlantic	Richards, 2006
Coryphaenidae	*Coryphaena hippurus*	*Coryphaena hippurus*	7.5-9	Gulf of Mexico	Ditty et al., 1994
Engraulidae	*Engraulis encrasicolus*	*Engraulis encrasicolus*	10	Aegean	Som. and Niko. 2010*
Gobiesocidae	Gobiesocidae	*Lepadogaster lepadogaster*	7.1	Portugal	Tojeira et al., 2012
Gobiidae	Gobiidae	*Lebetus guilleti*	3.1	Balearic Sea	Alemany 1997
Gobiidae	Gobiidae	*Pomatoschistus microps*	3.5	Balearic Sea	Alemany 1997
Gonostomatidae	*Cyclothone braueri*	*Cyclothone braueri*	6	Balearic Sea	Alemany 1997
Labridae	*Coris julis*	*Coris julis*	5	Balearic Sea	Alemany 1997
Labridae	Labridae	*Symphodus* spp.	3.1	Balearic Sea	Alemany 1997
Labridae	*Thalassoma pavo*	*Thalassoma bifasciatum*	3	Atlantic	Richards 2006
Mullidae	*Mullus barbatus*	*Mullus barbatus*	3.5	Balearic Sea	Alemany 1997
Mullidae	*Mullus surmuletus*	*Mullus surmuletus*	4	Balearic Sea	Alemany 1997
Myctophidae	*Ceratoscopelus maderensis*	*Ceratoscopelus maderensis*	5.5	Balearic Sea	Alemany 1997
Ophidiidae	Ophidiidae	*Ophidion barbatum*	9	Balearic Sea	Alemany 1997
Paralepididae	Paralepididae	*Paralepis coregonoides*	10-15	Atlantic	Richards 2006
Pomacentridae	*Chromis chromis*	*Chromis chromis*	3.5	Balearic Sea	Alemany 1997
Scombridae	*Auxis rochei*	*Auxis rochei*	5.4	Balearic Sea	Alemany 1997
Scombridae	*Thunnus thynnus*	*Thunnus thynnus*	6	Atlantic	Richards 2006
Scophthalmidae	*Lepidorhombus boscii*	*Lepidorhombus boscii*	5.9	Catalan Sea	Sabatés, 1989
Scorpaenidae	*Scorpaena porcus*	*Scorpaena porcus*	3.4	Balearic Sea	Alemany 1997
Scorpaenidae	*Scorpaena* spp.	*Scorpaena notata*	4	Balearic Sea	Alemany 1997
Serranidae	*Anthias anthias*	*Anthias anthias*	3.3	Balearic Sea	Alemany 1997
Serranidae	*Serranus hepatus*	*Serranus hepatus*	5.5	Balearic Sea	Alemany 1997
Serranidae	*Serranus* spp.	*Serranus cabrilla*	4.8	Balearic Sea	Alemany 1997
Soleidae	Soleidae	*Solea senegalensis*	4.5	NW Portugal	Ramos et al., 2010
Sparidae	*Boops boops*	*Boops boops*	5.5	Balearic Sea	Alemany 1997
Sparidae	*Diplodus* spp.	*Diplodus annularis*	5.2-	Catalan Sea	Sabates 1989

			5.6		
Sparidae	*Oblada melanura*	*Oblada melanura*	4.8	Catalan Sea	Sabates 1989
Sparidae	*Pagrus pagrus*	*Pagrus pagrus*	4.1	Atlantic	Richards 2006
Sparidae	Sparidae spp.	*Pagellus acarne*	5	Catalan Sea	Sabates 1989
Synodontidae	*Synodus saurus*	*Synodus*	9	French polinesia	Leis et al., 2003
Trachinidae	*Trachinus draco*	*Trachinus draco*	5.1	Balearic Sea	Alemany 1997
Triglidae	Triglidae	*Prionotus* spp.	6	New Jersey	McBride et al., 2002

2.3 Data analysis

Self-Organizing Maps (SOM) were used to explore the interaction between environmental variables and the vertical distribution of fish larvae. SOM is a non-linear mapping implementation used to unveil underlying patterns or structures in complex datasets, reducing the high dimensional feature space of the input data to a lower dimensional (usually two dimensional) networks of maps called neurons. SOM is based on unsupervised machine learning algorithms for the training of an artificial neural network (Kohonen, 1997, 1982) that simulates the topological distribution of neuron responses in the brain during the learning process, in particular by the relationship between neighboring neurons.

The learning process requires two main prerequisites. First, an input vector, representing the data to be analyzed. In our case, the input vector was composed of abundance taxa concatenated with each sampling's associated environmental variables (see below a detailed description of the analyzed data matrix). Second, the parameters that control the initialization and training processes.

Following Liu et al., 2006, we used an hexagonal map lattice to avoid anisotropy artifact, and a linear combination of the first two leading Empirical Orthogonal Functions (EOFs) modes for the initialization, although other initial configurations (e.g., random values) yield similar results. The training was performed using the imputation batch training algorithm (Vatanen et al., 2015) and a Gaussian neighborhood function.

The visualization of the SOM output often forms the key interface between the algorithm output and the user (Schreck, 2010). For clarity, a diagram of the SOM output used in this work, and its interpretation, is presented in Fig. 2.

The SOM analysis consists basically of an iterative presentation of the input data to a preselected neuronal network, which is iteratively modified. Each neuron (or unit) in the preselected network, is represented by a weight vector with the number of components equal to the dimension of the input sample data (abundance data and related

environmental data in our case). In each iteration, the neuron whose weight vector is closest (more similar) to the presented sample input data vector, called Best-Matching Unit (BMU), is transformed together with its topological neighbours towards the input sample following a specific neighbouring function. At the end of the training process, the probability density function of the input data is approximated by the SOM, and each neuron (unit) is associated with a reference pattern, with a number of components equal to the number of variables in the dataset so that it can be interpreted as a local summary or generalization of similar observations (Fig. 2a).

The explored matrix of environmental data includes the following variables above and below thermocline: the current speed (C, m s^{-1}), temperature (T, ºC), wave height (W, m), micro- (Mi, mg m^{-3}), and mesozooplankton biomass (Me, mg m^{-3}) and Chl-a (Chl, mg m^{-3}). A second matrix was built, consisting of larval abundances, disaggregated by taxa, length (pre and post-flexion), depth stratum, day/night, and sampling day. Each variable was separately normalized in the range between [0 1]. The response of larvae to environmental variables was analyzed, coupling physical and biological patterns through SOM. A preliminary test was performed to determine the optimal number of units in the neural network. The selection of the number of neurons is not evident and depends on the complexity of the data, on the features which one wants to examine in the dataset, and also in order to minimize the quantization and topographic errors. These two errors are the most relevant and broadly quality metrics used to evaluate the performance of SOM algorithm (Vatanen et al., 2015). When using a reduced number of neurons to 4 or 6, the associated errors increased with respect to a neural network of 9 neurons. On the other hand, when using 16 neurons, the quantization and topographic errors were very low, but a significant number of the patterns had 0% probability of occurrence to be represented among our data. Finally, we chose nine neurons as the best balance between associated errors and pattern representation without losing variability of the multivariate dataset (see Hernández-Carrasco and Orfila, 2018 for a more detailed analysis on the size of SOM). Finally, the potential of the technique for analyzing particularly frequent single-species with detailed larval length information (several size bins) was assessed in a similar way.

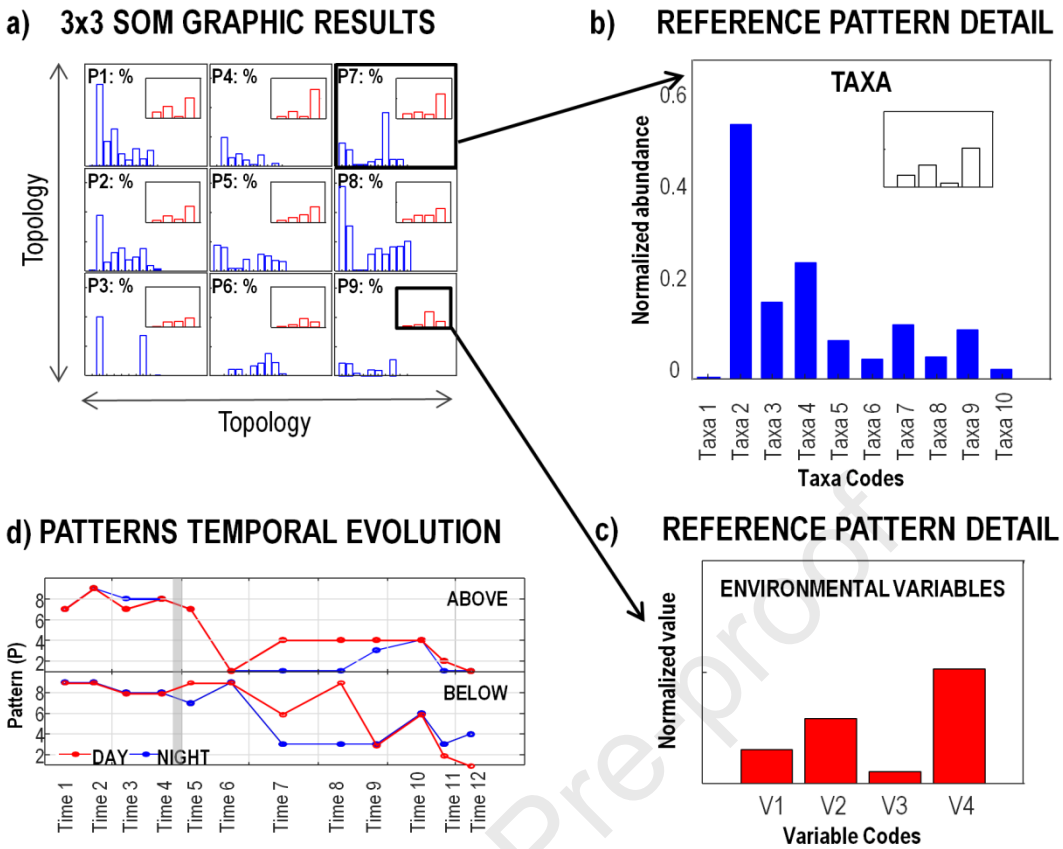

Figure 2: Diagram of the SOM output. a) Nine reference patterns extracted from a 3x3 coupled SOM analysis of taxa abundances (blue) and associated environmental parameters (red). Pn: probability of finding that particular pattern n in the data. b) Detail of the reference pattern output for the abundance data (normalized). In this pattern, Taxa 2 and 4 would be the most abundant and Taxa 1 abundance is almost incidental. c) Detail of the reference pattern output for the environmental data associated with a determined output pattern for abundance data. The number of environmental variables (normalized) corresponds to the number in the input vectors (V). d) Representation of the temporal succession of the extracted patterns. This output allows to identify the reference patterns that better describe the data at a given sampling time (Time n), namely position in relation to the thermocline (above/below) and moment of the day (day/night). The shaded area has been included to help the readers to identify the onset of the seasonal thermocline.

3. RESULTS

3.1. Hydrographic Scenario

Hydrographic data shows the typical spring to summer transition in Palma Bay where water column stratification is governed by the seasonal evolution of temperature (Fig. 3a). Since salinity was mostly uniform during the study period (Figs. 3 b, e), we selected the 20ºC isotherm as indicative of the thermocline depth. From late May to September, progressive seasonal warming of the mixing surface layer generated a sharp

thermal gradient at depths of 20 to 30 m. At the end of September, the mixing surface layer reached the bottom. Differences between average temperature above or below the thermocline showed maxima in July-August, and almost no differences between the initial and final times of the sampling period, confirming a correct sampling of the stratification process (Fig. 3d). Day/night differences in temperature were not detected and were not included as a variable in the SOM analysis. Salinity values were very uniform throughout the season, showing typical rainfall events at the end of summer (Fig. 3b, e).

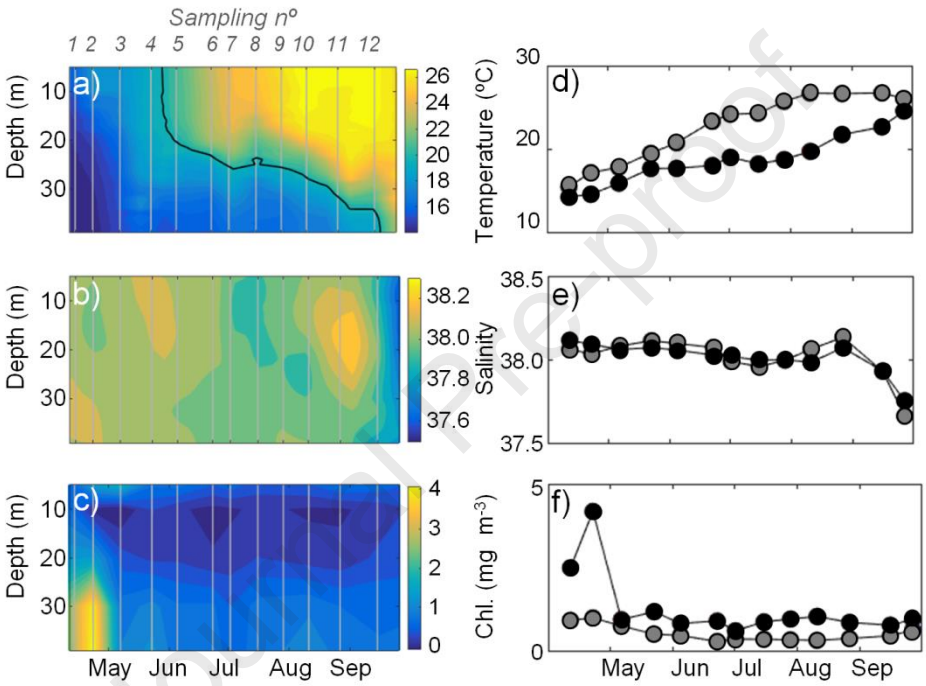

Figure 3: Environmental variables. Left Column: Vertical-Temporal evolution; Numbers 1-12 indicate the sampling number, in chronological order. Right Column: Above- (grey) and Below-Thermocline (black) values of: Temperature (ºC; a and d), Salinity (b and e) and Chlorophyll (mg m^{-3}; c and f).

Surface current velocities (not shown) were consistently higher than the bottom ones except at the end of the sampling period when velocities were almost the same. The maximum surface current velocity (0.09 m s^{-1}) was recorded in August, and the minimum (0.02 m s^{-1}) was found in April. Minimum and maximum values for the bottom current speed (0.006 and 0.03 m s^{-1}) were recorded in July and September, respectively. Wave height values were consistently higher during daytime than during night-time, with maximum differences in summer, coinciding with the breeze regime of the area. The maximum wave height (0.75 m) occurred at the end of May.

3.2. Biological Scenario

Figure 3c shows the evolution of phytoplankton biomass, approximated from Chl-a measurements. This evolution from high to very low values was consistent with the seasonal variation described for the coastal waters of the Balearic Islands (e.g., Álvarez et al., 2012). The maximum Chl-a concentration was observed during April close to the bottom (depth >30m, Chl-a > 2.2 mg m^{-3}). The progressive seasonal water warming induces the development of the thermocline, and its presence and position in the water column prevented the vertical supply of nutrients to the surface layer, limiting surface phytoplankton biomass production during summer (< 0.1 mg m^{-3}).

Microzooplankton biomass values above and below the thermocline showed the highest values at the beginning (April) and at the end (September) of the study period (Fig. 4a). In the absence of thermocline (at the beginning and end of the sampling period), high microzooplankton values were observed (12.4 to ~29 mg m^{-3}). At the beginning of April, they were confined to the first 20 meters (both night and day). For the heavily stratified period, June to August, values above and below the thermocline ranged between 0.4 and 4.4 mg m^{-3}, although below the thermocline, values were still higher during the daytime during May (7-10.5 mg m^{-3}).

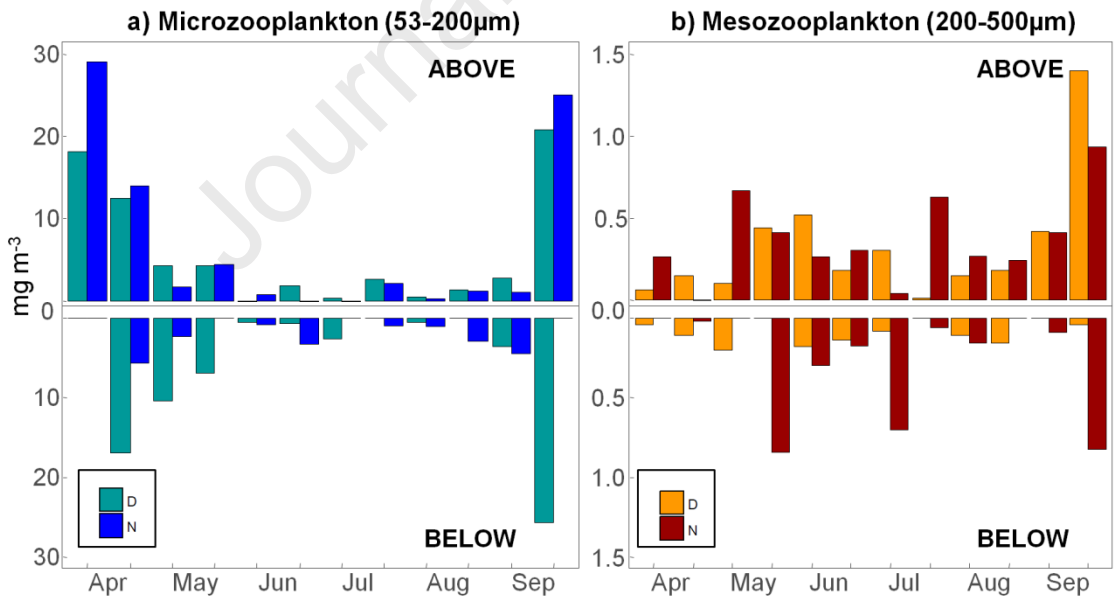

Figure 4: Biotic scenario. Above- and Below-thermocline Micro and Mesozooplankton day and night values of dry weight (biomass, mg m^{-3}) at each sampling day. D: Day; N: Night

Mesozooplankton biomass values (Fig. 4b) oscillated between 0.012 and 1.703 mg m^{-3}. Higher values developed approximately one month later with respect to microzooplankton. In general, daytime values were higher above the thermocline,

whereas the highest night-time values could be high both above and below the thermocline during the stratified period. At the end of the sampling (the entire water column was heated, see Fig. 3a), a clear difference pattern emerged, with daytime mesozooplankton being captured in the first 20 m and spreading throughout the column during the night (Fig. 4b).

3.3. Larval Fish Assemblages (LFA)

A total of 12284 fish larvae were found. Maximum larval fish abundances were observed at the end of May (Fig. 5a) followed by intermediate values during the summer months (June, July, August). In general, night abundances were higher than during the day, and no apparent differences between above and below thermocline larval abundances were found (Fig. 5b). The number of species (R) followed a dome-shaped curve, with the lowest number of taxa at the beginning (overall minimum in April in all sampling combinations) and at the end of the sampling period (Fig. 5c).

The dome-shaped relationship of R with time, similar to the total abundances, involved clear dependencies of nictimeral patterns with date, i.e., high/low abundances in one date tended to be reflected both above and below the thermocline and during day and night.

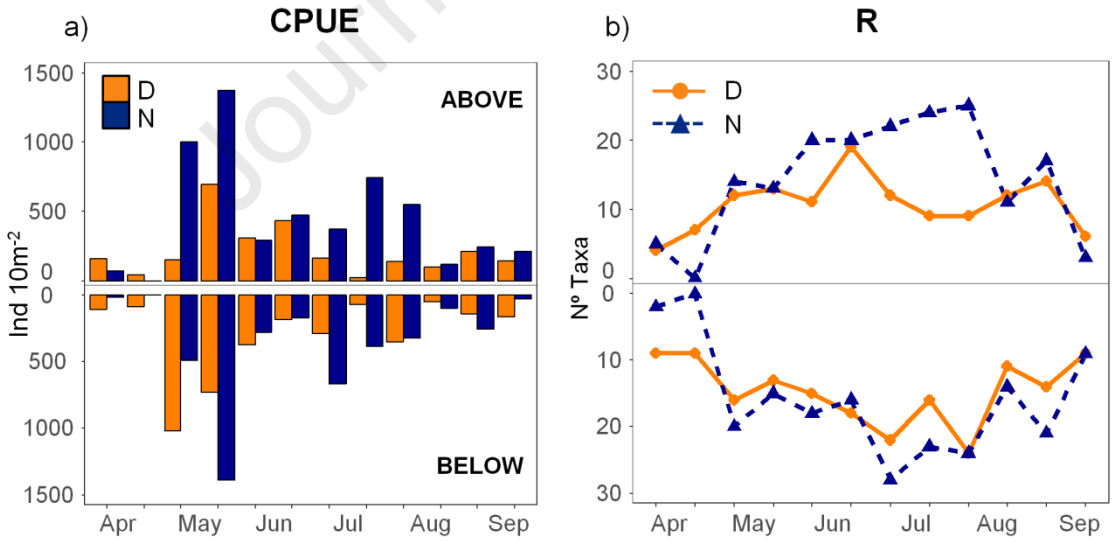

Figure 5: Larval fish assemblages. Above- and Below-thermocline day and night values at each sampling day of a) Catch-per-unit-effort (CPUE) and b) Richness (number of different taxa, R). D: Day; N: Night

The species *Spicara smaris*, *Engraulis encrasicolus* and larvae from the Gobiidae family represented 62.07 % of the total abundance (Table 2). Some taxa were found

equally at day and night tows (e.g., *S. smaris*, Labridae, *Arnoglossus* spp.), whereas some groups presented higher abundances during the night than during the day (Gobiidae, *Sardinella aurita*, *Chromis chromis*, Table 2). *Boops boops* and *Thunnus thynnus* followed the opposite pattern with higher daytime abundances. Few groups were only found at night hauls (e.g., *Thalassoma pavo*) or day hauls (e.g., *Gymnammodytes cicerelus*), representing just < 0.06 % of the total abundance.

Table 2: Different taxa occurring in the study area. **SOM code**: number assigned to each taxon in the SOM figure (Figure 6). **TA**: Total abundance (ind 10 m^{-2}) of each taxa**; RA**: Relative abundance %; **NL**: Number of Class Lengths found for each taxa; **L** Range: Min and max measured length in mm; **%D**: % of TA during day hauls; **%N**: % of TA during night hauls; **L Range D:** Length classes present in day hauls; **L Range N:** Length classes present in night hauls; In bold: RA>1%. *****: Taxa with preflexion stages included in the SOM (RA>0.2%); ******: Taxa with pre- and postflexion stages included in the SOM

Family	Lower taxa Identified	SOM code	TA	RA (%)	No L	L Range (mm)	%D	%N	L Range D	L Range N
Ammodytidae	*Gymnammodytes cicerelus*		12.9	0.06	3	2.45-4.47	100.0	0.0	L2-L4	
Belonidae	*Belone belone*		1.4	0.01	0		100.0	0.0		
Blenniidae	Blenniidae		384.9	**1.71**	8	1.75-8.79	22.1	77.9	L1-L7	L1-L8
Bothidae	*Arnoglossus* spp.*	11	113.7	0.51	8	1.24-9.56	49.7	50.3	L1-L3. L5. L9	L1-L4. L6.L7
	Bothus spp.		29.6	0.13	3	1.54-3.35	4.6	95.4	L1	L1-L3
Callionymidae	Callionymidae*	17	60.9	0.27	3	1.25-3.73	67.5	32.5	L1-L3	L1-L3
Carangidae	*Seriola dumerili*		0.1	0.00	0		100.0	0.0		
	Trachurus spp.*	6	403.2	**1.80**	4	1.27-4.13	22.9	77.1	L1-L3	L1-L4
Centracanthidae	*Spicara flexuosa**	7	225.9	**1.01**	5	1.59-5.85	33.2	66.8	L1-L5	L1-L5
	*Spicara smaris***	1,23	6971.5	**31.06**	8	1.16-8.98	41.8	58.2	L1-L8	L1-L8
Cepolidae	*Cepola* spp.		1.4	0.01	0		0.0	100.0		
Clupeidae	*Sardinella aurita***	3,24	1517.9	**6.76**	14	1.68-15	21.0	79.0	L1-L7. L9. L11	L2-L14
Coryphaenidae	*Coryphaena hippurus*		7.5	0.03	2	2.03-3.03	35.5	64.5	L2	L2. L3
Engraulidae	*Engraulis encrasicolus***	2,25	3953.3	**17.61**	14	1.93-14.3	40.3	59.7	L1-L9	L2-L14
Gobiesocidae	Gobiesocidae		38.1	0.17	5	2.86-6.23	64.3	35.7	L2-L6	L3. L4
Gobiidae	Gobiidae		3008.4	**13.40**	8	1.16-7.63	27.7	72.3	L1-L6	L1-L7
Gonostomatidae	*Cyclothone braueri*		48.6	0.22	8	1.91-12.6	36.9	63.1	L3. L4. L6	L1-L5. L7. L12
Labridae	*Coris julis**	10	200.8	0.89	4	1.38-4.15	14.6	85.4	L1-L3	L1-L4
	Labridae		777.5	**3.46**	5	1.29-5.17	55.6	44.4	L1-L5	L1-L4

	Thalassoma pavo		13.1	0.06	2	2.63-3.93	0.0	100.0		L2. L3
Mullidae	*Mullus barbatus**	13	126.0	0.56		1.83-11.6	50.4	49.6	L1-L4	L2-L5. L11
	*Mullus surmuletus**	22	55.1	0.25	4	2.43-5.74	57.2	42.8	L2-L4	L2-L5
Myctophidae	*Ceratoscopelus maderensis**	16	68.5	0.30	5	2.96-6.52	22.0	78.0	L3-L5	L2-L6
	Lampanyctus crocodilus		2.7	0.01	0		0.0	100.0		
	Myctophum punctatum		1.3	0.01	0		100.0	0.0		
Ophidiidae	Ophidiidae*	12	121.5	0.54	3	2.17-4.66	60.6	39.4	L2-L4	L2-L4
Paralepididae	Paralepididae		3.1	0.01	1	4.25-4.62	0.0	100.0		
Pomacentridae	*Chromis chromis***	4,26	599.6	**2.67**	5	1.37-5.44	26.4	73.6	L1-L4	L1-L5
Scombridae	*Auxis rochei**	14	92.4	0.41	6	1.53-6.7	28.8	71.2	L1-L6	L1-L6
	Thunnus alalunga		1.9	0.01	0		100.0	0.0		
	Thunnus thynnus		8.7	0.04	2	4.23-5.86	75.3	24.7	L4. L5	L5
Scophthalmidae	*Lepidorhombus boscii*		4.4	0.02	3	2.32-5.27	5.4	94.6	L5	L4. L5
Scorpaenidae	*Scorpaena porcus**	18	51.1	0.23	3	1.59-3.86	44.6	55.4	L1-L3	L1-L3
	Scorpaena spp.		26.7	0.12	2	1.58-2.82	38.4	61.6	L1. L2	L1. L2
Serranidae	*Anthias anthias*		4.4	0.02	1	1.67-1.78	50.8	49.2	L1	L1
	*Serranus hepatus**	8	211.6	0.94	7	1.04-7.61	41.2	58.8	L1-L4. L6	L1-L7
	Serranus spp.*	5	374.0	**1.67**	7	1.35-7.48	39.5	60.5	L1-L6	L1-L7
Soleidae	Soleidae*	21	42.9	0.19	5	1.58-5.55	32.9	67.1	L1-L3. L5	L2-L4
Sparidae	*Boops boops**	20	48.8	0.22	6	1.97-6.05	88.8	11.2	L1-L6	L4.L5
	Diplodus spp.*	9	199.2	0.89	8	1.35-12.6	47.2	52.8	L1-L7	L1-L5.L12
	*Oblada melanura**	15	123.8	0.55	5	2.21-6.87	35.8	64.2	L2-L5	L2-L6
	*Pagrus pagrus**	19	43.5	0.19	4	1.914.61	20.1	79.9	L2. L3	L1-L4
	Sparidae spp.		194.6	0.87	6	1.28-6.18	76.5	23.5	L1-L6	L1-L3. L5
Sphyraenidae	Sphyraenidae		22.7	0.10	5	1.78-5.88	12.8	87.2	L1-L3	L2. L3. L5
Sygnathidae	Sygnathidae		1.6	0.01	0		100.0	0.0		
Synodontidae	*Synodus saurus*		41.2	0.18	3	2.79-4.68	3.2	96.8	L3	L1-L4
Trachinidae	*Trachinus draco*		52.1	0.23	4	1.96-4.61	36.1	63.9	L2-L4	L1-L4
Triglidae	Triglidae		7.8	0.03	3	1.38-5.86	100.0	0.0	L1. L3. L5	
Unidentified	Unidentified		2146.3	**9.56**	0		40.6	59.4		

The wider length ranges were represented by clupeiform fish, which were also the most abundant; *E. encrasicolus* and *S. aurita* larvae ranged between 2.01 and 15 mm SL (L2-L14). The sparid *S. smaris*, also highly abundant, oscillated between 1.16-8.98 mm (L1-L9). Taxa collected in lower abundances presented fewer length classes, including the sparid *Pagrus pagrus* (L2, L3) or the myctophid *Cyclothone braueri* (L3, L4, L6, Table 2). Univariate analyses were not attempted due to the high complexity of the data, including species-specific triggers of appearance in the plankton, non-independence of several factors, etc. (see Discussion).

3.4. SOM analysis: patterns of the environment-LFA relationship

We used a coupled SOM analysis to extract general patterns of LFA response to environmental forcing below and above the thermocline, at different stages and taking into account temporal variability. Only taxa with abundances representing more than 0.19% of the total abundance (22 taxa) were included in the SOM computations (Table 2). The unidentified Gobiidae, Labridae, and Blenniidae families and the Sparidae spp. taxa were not used in this analysis as they comprise several taxa with possible different day/night patterns, reproductive cycle and different behaviour against stratification. Only *S. smaris*, *S. aurita*, *E. encrasicolus* and *C. chromis* had enough individuals at the post-flexion as to be included in the SOM analysis.

Figure 6 (top panel) shows the 9 patterns extracted from the 3x3 coupled SOM analysis of taxa abundances and environmental parameters, where the probability of finding that particular pattern in the data is indicated. The temporal succession of the different patterns is also displayed in the bottom panel. The topological ordering of the patterns in the neural network performed by SOM is clearly depicted: patterns with higher larval abundances, higher presence of post-flexion stages (taxa 23-26 in Fig. 6, see also Table 2) and stronger influence of water temperature are located around the bottom-left corner of the neural network, and those with lower taxa abundances, absence of post-flexion stages and environmental scenarios characterized by the high velocity of the current and lower temperatures are located around the top-right corner. As a consequence of the preservation of topology, the closer to these extremes, the more similar the patterns are (e.g., pattern 2 (P2) is more similar to P1 than to P8). Patterns with the highest probability of occurrence are located in the corners of the neural networks being P9 the most dominant (ca. 21%) followed by P1, P8, P3, and P7. Note that P5 with zero probability of occurrence is included by SOM to preserve topological continuity.

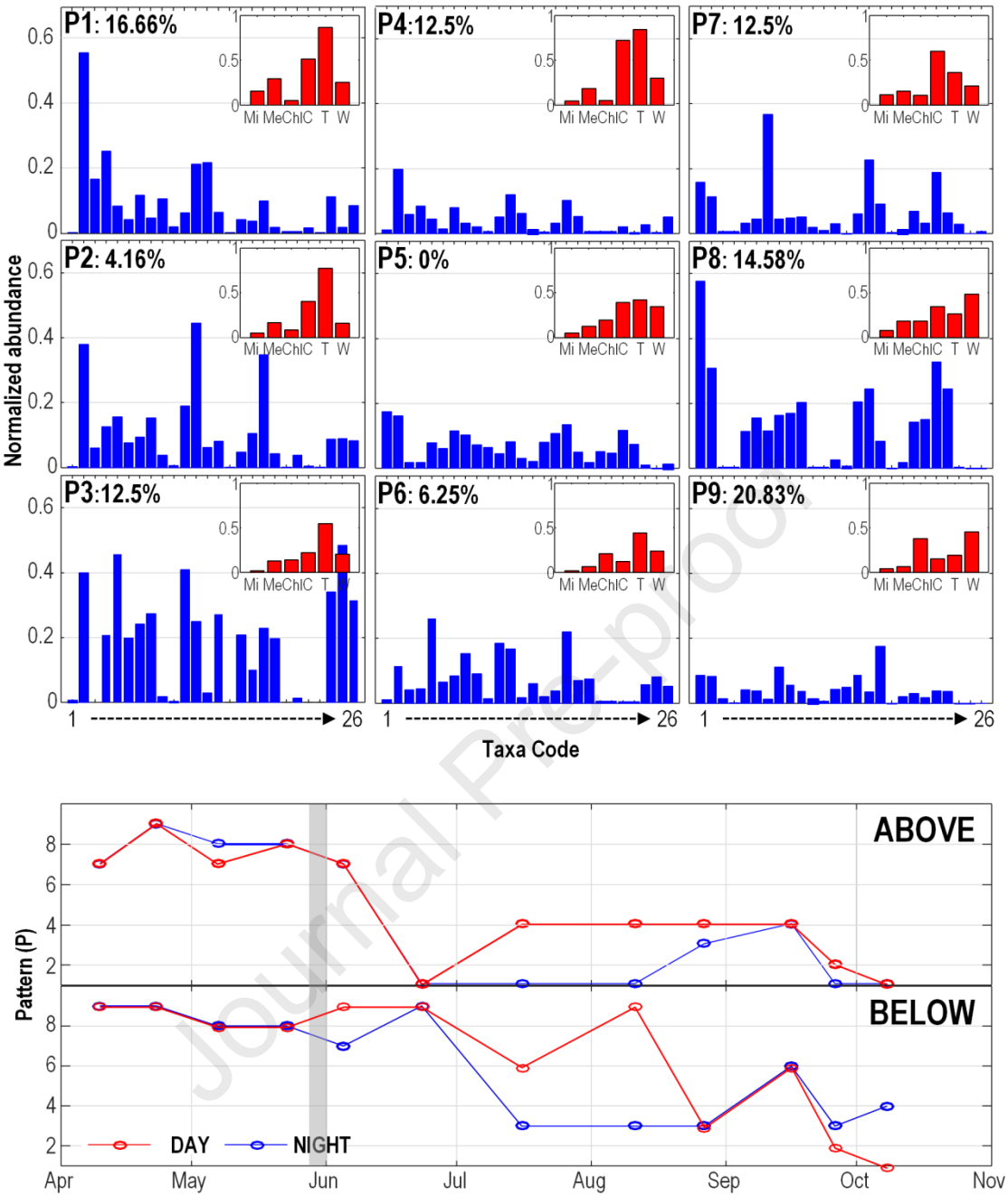

Figure 6: Results of the SOM analysis. Top: Abundance distribution patterns of each taxa (ordered 1 to 26, Table 2) given by a 3x3 SOM analysis. Number in each pattern label (P1 to P9) is the corresponding total occurrence probability. Inside each taxa pattern (blue), the corresponding environmental scenario pattern (red). Bottom: Evolution of each taxa and environmental pattern shown in the top plot during the study period. The shaded area represents the onset of stratification. Taxa numbers (Table 2): Pre-flexion stages (1-22); Post-flexion stages (23-26); Environmental variables: Mi: Microzooplankton dry weight (biomass, mg m^{-3}); Me: Mesozooplankton dry weight (biomass, mg m^{-3}); Chl: Chlorophyll (mg m^{-3}); C: module of the current (m s^{-1}); T: water temperature (ºC) and W: wave height (m)

The temporal succession of the extracted patterns also follows the bottom-right to top-left distribution (Fig. 6, bottom). That figure shows that minimum day-night differences were observed above or below the thermocline before the onset of stratification (~

June). The patterns are grouped representing different situations occurring at different times, i.e., the group of nearby patterns (P7-P9) are more dominant from April to May, and patterns (P1-P6) dominate from June onwards, in coincidence with the development of the thermocline in the area (Figs 3 a, d). This transition implies that before the thermocline establishment, prevailing patterns are formed by few taxa and high abundances of *S. smaris*. Those taxa spawn at low temperatures and exploit a situation of high chlorophyll. On the contrary, after the thermocline is formed, high temperature, relatively low Chl-a and plankton food constitute the prevailing environmental conditions for P1-4. Noticeably, the SOM reveals, in the temporal evolution of the Best-Matching Units (BMUs), that samples taken during the day and above the thermocline in June-September (P4, and to a lesser extent P2 and P1) differ largely from the patterns emerging from below the thermocline; in general, the abundance of species above the thermocline during the more stratified period was lower than below the thermocline. Finally, post-flexion larvae (species 24-26, Fig. 6 top panel) appear fundamentally in P3 and are detected mainly during the night and below the thermocline in July-August.

3.5. SOM application to single species evolution.

We here demonstrate the utility of SOM analyses for species for which enough detailed information (through time and length abundances) exists: the spring/summer-spawning European anchovy *E. encrasicolus* and the abundant winter-spawning sparid *S. smaris*. In these cases the input data used in the SOM computations is composed of time series of the abundance of a particular species for different size classes. The temporal patterns of all the *E. encrasicolus* and *S. smaris* length classes were analyzed separately through two 2x2 SOM. In the *E. encrasicolus* analysis, pattern 3 showed zero probability of being present among our data (Fig. 7). The pattern representing most data (P4, 46.15%) showed the trend of capturing more individuals during the night during all the time series, shared to a lesser extent by P1 and P2. Daytime larval abundances were higher in Pattern 1, with a 23% probability of being present in our data and corresponding mainly to the smallest individuals (L1 to L3, Fig. 7 bottom). A progressive reduction in the daytime presence of larvae in the tows is represented by Pattern 2, corresponding to length classes 4-7, whereas the largest size classes (L8-L13) are clearly captured basically in the night, according to P4 (Fig. 7, bottom).

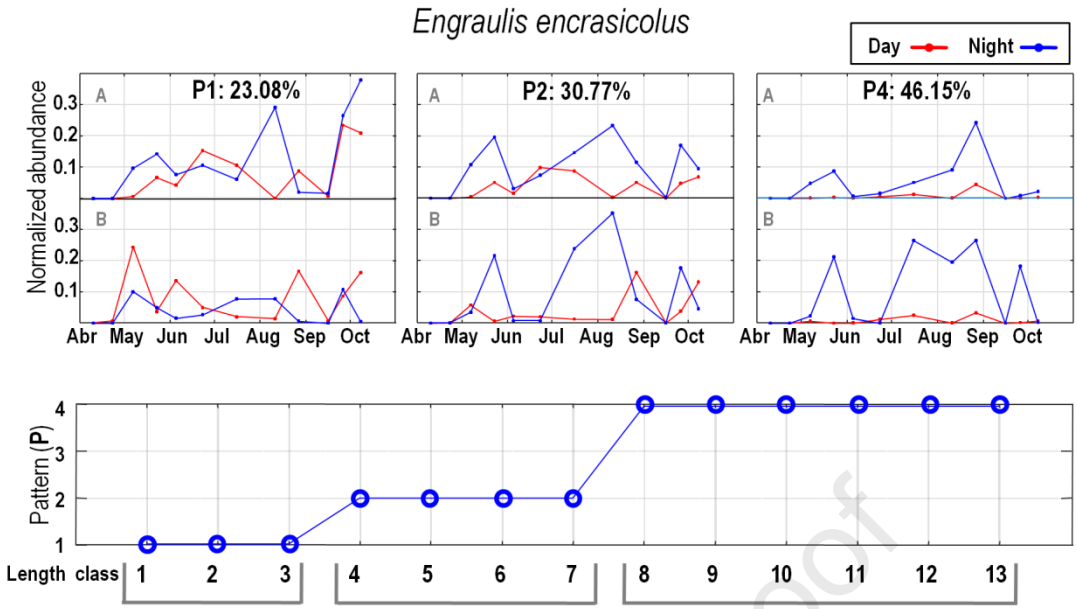

Figure 7: Results of the SOM analysis for *E. encrasicolus* length classes. Top: Temporal patterns of *E. encrasicolus* abundance above (A) and below (B) the thermocline given by a 2x2 SOM analysis during the day (red) and night (blue). Number in each pattern label (P1 to P4) is the corresponding total occurrence probability. Bottom: Correspondence of the patterns shown in the top plot to each of the length classes.

The results of the SOM applied tor *S. smaris* are shown in Fig. 8. In this case, pattern 3 showed no probabilities of being present among our data. The most probable pattern (57.14%) was P4, followed by P2 (28.57%) and P1 (14.29%). The largest individuals (length classes L4 to L7) followed P4, whose main differences with P1 and P2 (L1 to L3) were lower abundances above the thermocline, but fundamentally a high concentration below the thermocline during daylight, and not so much during the night. On the contrary, they were almost absent above the thermocline during daylight. The other patterns showed that for smaller larvae, this absence over the thermocline during daylight was less pronounced.

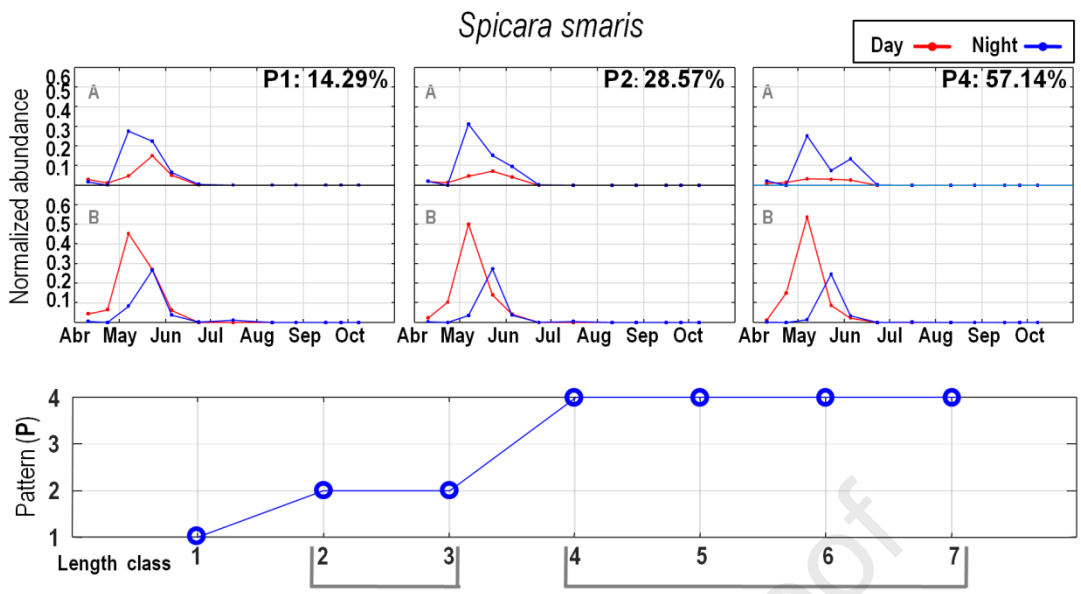

Figure 8: Results of the SOM analysis for *S. smaris* length classes. Top: Temporal patterns of *S. smaris* abundance above (A) and below (B) the thermocline given by a 2x2 SOM analysis during the day (red) and night (blue). Number in each pattern label (P1 to P4) is the corresponding total occurrence probability. Bottom: Correspondence of the patterns shown in the top plot to each of the length classes.

4. DISCUSSION

The results of the present study revealed differences in the patterns of abundance and vertical distribution of fish larvae before and after the establishment of the water column stratification. The use of Self Organised Maps analysis (SOM) made possible the reduction of the multiple non-linear relationships among taxa and environmental variables into more understandable patterns of probability, showing that valuable information can be extracted when applied to data sets of different complexity.

There are at least three drivers of complexity in the way to investigate our main hypothesis of the vertical distribution of fish larvae being associated with the properties of the environment. They are basically the natural turnover of taxa that follows seasonality, the differences in behavior between ontogenetic stages of the same taxa, and finally, the choice of the different environmental parameters that might influence the processes. Trying to formally analyze how each species behaves with regard to these factors would have probably implied the construction of several complex models (e.g., state-space models) for each species, given that typical GLMM or GAMS are inappropriate due to the underlying properties of the data.

The succession of taxa in the area, characterized by a progressive change in assemblage composition throughout spring, and a sharp change (mainly due to the increase of

benthopelagic species) in early summer were already described in the NW Mediterranean (Álvarez et al., 2012; Fernández de Puelles et al., 2007; Masó et al., 1998).

The seasonal differences detected before and after the onset of stratification, with higher abundances but lower number of taxa in spring, and an opposite pattern in summer are also in concordance with previous descriptions in the area (Álvarez et al., 2012; Fernández de Puelles et al., 2007). These differences in reproductive strategies between spring and summer assemblages have also been described for other areas in the Mediterranean such as the Catalan sea (Sabatés, 1990), northern Aegean Sea (Koutrakis et al., 2004) and also in distant areas such as the South China Sea (Huang et al., 2017). In general, the succession of taxa and differences in spawning strategies between taxa have been attributed to the reduction of inter- and intra-specific competition for food resources in the NW Mediterranean (Sabatés et al., 2007), and more specifically to shifts in the prevailing trophic pathway exploited by reproductive adults; from a more productive pelagic environment during spring, vs. a more productive benthic environment (rocky bottom and seagrass meadows) in summer(Álvarez et al., 2012).

Most larvae in our study were small and undeveloped due to i) the sampling method (Bongo-40), which selects for mainly preflexion stages (Catalán et al., 2014)and ii) the fact that Palma Bay is a productive spawning ground for many species (Álvarez et al., 2015a, 2012). Due to the limitation in size classes, the analysis of ontogenetic-dependent migration (e.g., Borges et al., 2007; Huebert et al., 2011; Irisson et al., 2009) could only be tested for few species. However, the potential of SOM for this type of analysis is clear, particularly for the combined use of non-linear methods and the use of probabilistic proxies in the interpretation.

The division into pre and post notochord flexion, used to differentiate among swimming abilities and widely applied in the literature (Olivar et al., 2018; Olivar and Sabatés, 1997, among others), has been proven not to be the best criterion to depict vertical positioning, as revealed by the two SOM analysis conducted on individual taxa (Figs 7 and 8). For *Spicara smaris*, the literature proposed 5.5 mm SL as flexion length (Table 1, Alemany, 1997) whilst we found differences in the vertical behavior after the 3 mm (Fig. 8). In the *Engraulis encrasicolus* case, 10 mm SL was used to group individuals into pre and post flexion (Table 1, Somarakis and Nikolioudakis, 2010). However, in the SOM analysis, three different size patterns with respect to the thermocline were recognized: 1-3 mm, 4-7 mm, and 8-13 mm (Fig. 7). The onset of the notochord flexion

in European anchovy was determined around 7 mm SL of preserved length by Somarakis and Nikolioudakis (2010) and was related to the initiation of swim-bladder inflation (considered to trigger the onset of vertical migrations and subsequent schooling behavior). Our results suggest that SOM can be a useful method to summarize vertical distribution data through time and ontogeny. However, the full power of SOM would be further revealed in even more complex data sets, for example, using fine-scale image-based vertical sampling techniques. In this respect, artificial intelligence that uses complex neural network approaches (e.g., Deep Learning) is irrupting in the analyses of larval fish ecology (e.g., Axler et al., 2020; Catalán et al., 2020).

The wide variety of environmental variables that can influence the vertical position of fish larvae, added to the potential non-linearity of their possible influence, makes the statistical or numerical approximations to the problem an uphill climb. Most of the research in the vertical position of fish larvae is thus centered on the description of the environment and the larval fish associations and the discussion of possible relationships among them (e.g Muhling and Beckley, 2007; Sánchez-Velasco et al., 2007). Some authors have addressed their research in a numerical way, with correlation coefficients (Gray and Kingsford, 2003; Rodríguez et al., 2006) and multiple linear regression (Vargas-Yáñez and Sabatés, 2007). We show that the SOM analysis can be a useful tool to handle both the high number of variables and the nonlinearity of the processes under analysis. The application of the current analysis, however, does not preclude the need to conduct a priori elimination of confounding variables.

Both dependence and independence of vertical distributions from the water column's vertical thermal structure have been observed for coastal and oceanic waters in different regions of the world. As an example, the relationship between the vertical distribution of fish larvae and the seasonal evolution of the structure of the water column was studied using static stability and the pycnocline position as a stratification indicator in Bahia de La Paz, SW Gulf of California (Sánchez-Velasco et al., 2007). The authors concluded that significant differences in taxa and larval abundances were found among the water column's most stable stratum and the below-pycnocline stratum. Our results support those findings, with differences found before and after the stratification period in the area, with most of the species found in the stratified (stable) period.

However, Rodríguez et al. (2006) reported an almost negligible influence of the thermocline on the vertical distribution of fish larvae in the Canaries-African coastal

transition zone and related that result to the relatively weak thermocline in the area. Gray and Kingsford (2003) also provided evidence that thermoclines had no detectable effect on vertical distributions of fish larvae and mesozooplankton in dynamic coastal waters, although in that case, perturbations in the position and intensity of thermoclines were frequent. Laboratory studies on the effect of vertical gradients on larval fish distribution show that the thermocline can exert a variable effect depending on the species, stage, and other variables such as fish condition or visibility (Catalán et al., 2011; Reglero et al., 2018; Vollset et al., 2013).

When analyzed together, all studies seem to converge at one point: stability of the water column in terms of mixing activity conditions can be related to the vertical position of fish larvae (here, larval behavior plays a more prominent role), whereas the lack of stability (e.g. weak or perturbed thermocline) weakens the detection of patterns at the vertical scale. The behavioral control will, in turn, be dependent on larval swimming capacities (e.g., Olivar et al., 2001; Ospina-Alvarez et al., 2012). Siegel et al., (2008) concluded that larval settlement is inherently a stochastic process driven by the interaction between coastal circulations and organism life histories. Siegel's "settlement equation" (coastal circulations; organism life histories) can be seen as a balance between its terms, depending on the moment in which the system is observed: coastal circulation might gain prominence for settlement in unstable scenarios (spring mixed column water in our study) whereas the influence of the organism life histories on settlement gains weight in stable scenarios (summer stratification period in our study). Although the sampling of a unique station throughout time was used to assume stationary conditions in our study, the influence of dynamical processes induced by lateral and vertical stirring, vertical shear of horizontal velocities, and advection of larvae cannot be discarded, adding complexity to the system. Other hypotheses confer a relevant role to larval behavior and justify the need for vertically-resolved surveys: the Sense Acuity and Behavioral hypothesis (Teodósio et al., 2016) proposes that individuals develop different behaviors and skills depending on the coastal cues they face. The importance of behavior for Mediterranean fish was recently proved by.

Summarizing, we could determine differences in the vertical strategies of fish larval assemblages before and after the establishment of the stratified period in a very complex data set through the use of an automatic data-analysis method (Self Organizing Map, SOM). The method neither generates cause-effect relationships nor extracts species-specific patterns (unless specified, as shown in some examples) in the way it has been

applied. However, SOM analysis allowed the description of the whole multispecific system overcoming the difficulties inherent in describing non linear processes. The possibility of analyzing these intricate multivariate patterns without the restriction of linear approaches is a useful tool in understanding the complexity of recruitment processes and settlement success of fish.

5. Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness (Ref: CTM2012-39476-C02-01). The same funding scheme partially supported I.C. through the project REC2 (Ref: CTM2011-23835). J.S. F-M. received a French state aid managed by the National Research Agency under the ISblue program (ANR-17-EURE-0015). I.H-C. acknowledges the Vicenç Mut grant funded by the Government of the Balearic Islands, Spain, and the European Social Fund (ESF) Operational Programme. C.D.G. was funded by a fellowship (FPI-INIA-2012) from the National Institute of Agricultural and Food Research and Technology. P.M. S-H. was supported by a Ph.D. research fellowship FPI (Formación Personal Investigación) fellowship BES-2013-067305 from MINECO. We thank Dr. G. Basterretxea for his help during the elaboration of this work.

6. REFERENCES

Alemany, F., 1997. Ictioplancton del Mar Balear. PhD Thesis. Universitat de les Illes Balears.

Alós, J., Palmer, M., Catalan, I.A., Alonso-Fernández, A., Basterretxea, G., Jordi, A., Buttay, L., Morales-Nin, B., Arlinghaus, R., 2014. Selective exploitation of spatially structured coastal fish populations by recreational anglers may lead to evolutionary downsizing of adults. Mar. Ecol. Prog. Ser. 503, 219–233. https://doi.org/10.3354/meps10745

Álvarez, I., Catalán, I.A., Jordi, A., Alemany, F., Basterretxea, G., 2015a. Interaction between spawning habitat and coastally steered circulation regulate larval fish retention in a large shallow temperate bay. Estuar. Coast. Shelf Sci. 167, 377–389. https://doi.org/10.1016/j.ecss.2015.10.015

Álvarez, I., Catalán, I.A., Jordi, A., Palmer, M., Sabatés, A., Basterretxea, G., 2012. Drivers of larval fish assemblage shift during the spring-summer transition in the coastal Mediterranean. Estuar. Coast. Shelf Sci. 97, 127–135. https://doi.org/10.1016/j.ecss.2011.11.029

Álvarez, I., Rodríguez, J.M., Catalán, I.A., Hidalgo, M., Álvarez-Berastegui, D., Balbín, R., Aparicio-González, A., Alemany, F., 2015b. Larval fish assemblage structure in the surface layer of the northwestern Mediterranean under contrasting oceanographic scenarios. J. Plankton Res. 37. https://doi.org/10.1093/plankt/fbv055

Arvola, L., Rask, M., Forsius, M., Ala-Opas, P., Keskitalo, J., Kulo, K., Kurkilahti, M., Lehtovaara, A., Sairanen, S., Salo, S., Saloranta, T., Verta, M., Vesala, S., 2017. Food web responses to artificial mixing in a small boreal lake. Water (Switzerland) 9, 1–21. https://doi.org/10.3390/w9070515

Axler, K.E., Sponaugle, S., Briseño-Avena, C., Jr, H.F., Warner, S.J., Dzwonkowski, B., Dykstra, S.L., Cowen, R.K., 2020. Fine-scale larval fish distributions and predator-prey dynamics in a coastal river-dominated ecosystem. Mar. Ecol. Prog. Ser. 650, 37–61.

Basterretxea, G., Font-Muñoz, J.S., Salgado-Hernanz, P.M., Arrieta, J., Hernández-Carrasco, I., 2018. Patterns of chlorophyll interannual variability in Mediterranean biogeographical regions. Remote Sens. Environ. 215, 7–17. https://doi.org/10.1016/j.rse.2018.05.027

Basterretxea, G., Jordi, A., Catalán, 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. https://doi.org/10.1111/j.1365-2419.2012.00625.x

Bensoussan, N., Romano, J.C., Harmelin, J.G., Garrabou, J., 2010. High resolution characterization of northwest Mediterranean coastal waters thermal regimes: To better understand responses of benthic communities to climate change. Estuar. Coast. Shelf Sci. 87, 431–441. https://doi.org/10.1016/j.ecss.2010.01.008

Borges, R., Beldade, R., Gonçalves, E.J., 2007. Vertical structure of very nearshore larval fish assemblages in a temperate rocky coast. Mar. Biol. 151, 1349–1363. https://doi.org/10.1007/s00227-006-0574-z

Catalán, I., Vollset, K.W., Morales-Nin, B., Folkvord, A., 2011. The effect of temperature gradients and stomach fullness on the vertical distribution of larval herring in experimental columns. J. Exp. Mar. Bio. Ecol. 404, 26–32. https://doi.org/10.1016/j.jembe.2011.05.005

Catalán, I.A., Dunand, A., Álvarez, I., Alós, J., Colinas, N., Nash, R.D.M., 2014. An evaluation of sampling methodology for assessing settlement of temperate fish in seagrass meadows. Mediterr. Mar. Sci. 15. https://doi.org/10.12681/mms.539

Catalán, I.A., Reglero, P., Álvarez, I., 2020. Research on early life stages of fish: A lively field. Mar. Ecol. Prog. Ser. 650. https://doi.org/10.3354/meps13491

Cianelli, L., Bailey, K., Olsen, E.M., 2015. Evolutionary and ecological constraints of fish spawning habitats. ICES J. Mar. Sci. 72, 285–296. https://doi.org/10.4135/9781412953924.n678

Coma, R., Ribes, M., Serrano, E., Jimenez, E., Salat, J., Pascual, J., 2009. Global warming-enhanced stratification and mass mortality events in the Mediterranean. Proc. Natl. Acad. Sci. 106, 6176–6181. https://doi.org/10.1073/pnas.0805801106

Ditty, J.G., Shaw, R.F., Grimes, C.B., Cope, J.S., 1994. Larval development, distribution, and abundance of common dolphin, *Coryphaena hippurus*, and pompano dolphin, *C. equiselis* (family: Coryphaenidae), in the northern Gulf of Mexico. Fish. Bull. 92, 275–291.

Fernández de Puelles, M.L., Alemany, F., Jansá, J., 2007. Zooplankton time-series in the Balearic Sea (Western Mediterranean): Variability during the decade 1994-2003. Prog. Oceanogr. 74, 329–354. https://doi.org/10.1016/j.pocean.2007.04.009

Font-Muñoz, J.S., Jordi, A., Anglès, S., Ferriol, P., Garcés, E., Basterretxea, G., 2018. Assessing phytoplankton community composition using combined pigment and particle size distribution analysis. Mar. Ecol. Prog. Ser. 594, 51–63. https://doi.org/10.3354/meps12559

Gray, C.A., Kingsford, M.J., 2003. Variability in thermocline depth and strength, and relationships with vertical distributions of fish larvae and mesozooplankton in dynamic coastal waters. Mar. Ecol. Prog. Ser. https://doi.org/10.3354/meps247211

Hernández-Carrasco, I., Orfila, A., 2018. The Role of an Intense Front on the Connectivity of the Western Mediterranean Sea: The Cartagena-Tenes Front. J. Geophys. Res. Ocean. 123, 4398–4422. https://doi.org/10.1029/2017JC013613

Hernández-Carrasco, I., Solabarrieta, L., Rubio, A., Esnaola, G., Reyes, E., Orfila, A., 2018. Impact of HF radar current gap-filling methodologies on the Lagrangian assessment of coastal dynamics. Ocean Sci. 14, 827–847. https://doi.org/10.5194/os-14-827-2018

Hjort, J., 1926. Fluctuations in the year classes of important foodfishes. J. du Con-seil Int. pour l'Exploration la Mer 1, 5–38.

Hjort, J., 1914. Fluctuations in the great fisheries of northern Europe viewed in the lightof biological research. Rapp. Procès-Verbaux des Réunions du Cons. International pour l'Exploration la Mer 20, 1–228.

Houde, E.D., 2008. Emerging from Hjort's shadow. J. Northwest Atl. Fish. Sci. 41, 53–70. https://doi.org/10.2960/J.v41.m634

Houde, E.D., Schekter, R.C., 1980. Feeding by marine fish larvae: developmental and functional responses. Environ. Biol. Fishes 5, 315–334. https://doi.org/10.1007/BF00005186

Huang, D., Zhang, X., Jiang, Z., Zhang, J., Arbi, I., Jiang, X., Huang, X., Zhang, W., 2017. Seasonal fluctuations of ichthyoplankton assemblage in the northeastern South China Seainfluenced by the Kuroshio intrusion. J. Geophys. Res. Ocean. Res. 122, 7253– 7266. https://doi.org/10.1002/2017JC013313

Huebert, K.B., Cowen, R.K., Sponaugle, S., 2011. Vertical migrations of reef fish larvae in the straits of florida and effects on larval transport. Limnol. Oceanogr. 56, 1653–1666. https://doi.org/10.4319/lo.2011.56.5.1653

Irisson, J.-O., Paris, C.B., Guigand, C.C., Planes, S., Org, I., 2009. Vertical distribution and ontogenetic ' migration ' in coral reef fish larvae. Mar. Biol. 55, 1–29. https://doi.org/10.43119/lo2009.55.2.0909

Kingsford, M., Shanks, A.L., Morgan, S.G., Pineda, J., 2002. Sensory environments, larval abilties and local self recruitment. Bull. Mar. Sci. 70, 309–340.

Kohonen, T., 1997. Self-Organizing Maps. Springer Series in Information Sciences.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69. https://doi.org/10.1007/BF00337288

Kolczynski, W.C., Hacker, J.P., 2014. The potential for self-organizing maps to identify model error structures. Mon. Weather Rev. 142, 1688–1696. https://doi.org/10.1175/MWR-D-13-00189.1

Koutrakis, E.T., Kallianiotis, A.A., Tsikliras, A.C., 2004. Temporal patterns of larval fish distribution and abundance in a coastal area of northern Greece *. Sci. Mar. 68, 585–595. https://doi.org/10.3989/scimar.2004.68n4585

Leis, J.M., 2006. Are Larvae of Demersal Fishes Plankton or Nekton?. Adv Mar Biol 51, 58–141.

Leis, J.M., 1986. Vertical and horizontal distribution of fish larvae near coral reefs at Lizard Island, Great Barrier Reef. Mar. Biol. 90, 505–516.

Leis, J.M., Trnski, T., Dufour, V., Harmelin-Vivien, M., Renon, J.P., Galzin, R., 2003. Local completion of the pelagic larval stage of coastal fishes in coral-reef lagoons of the Society and Tuamotu Islands. Coral Reefs 22, 271–290. https://doi.org/10.1007/s00338-003-0316-3

Liu, Y., Weisberg, R.H., Mooers, C.N.K., 2006. Performance evaluation of the self-organizing map for feature extraction. J. Geophys. Res. Ocean. 111, 1–14. https://doi.org/10.1029/2005JC003117

Mann, K.H., Lazier, J.R.N., 1991. Dynamics of marine ecosystems: biological physical interactions in the oceans, Dynamics of Marine Ecosystems, Wiley Online Books. Blackwell Scientific Publications, Boston, MA. https://doi.org/https://doi.org/10.1002/9781118687901.fmatter

Masó, M., Sabatés, A., Pilar Olivar, M., 1998. Short-term physical and biological variability in the shelf-slope region of the NW Mediterranean during the spring transition period. Cont. Shelf Res. 18, 661–675. https://doi.org/10.1016/S0278-4343(98)00011-9

McBride, R.S., Fahay, M.P., Able, K.W., 2002. Larval and settlement periods of the northern searobin (*Prionotus carolinus*) and the striped searobin (*P. evolans*). Fish. Bull. 100, 63–73. https://doi.org/10.111/j.1439-0485.2006.00146.x

Montgomery, J.C., Tolimieri, N., Haine, O.S., 2001. Active habitat selection by pre-settlement reef fishes. Fish Fish. 2, 261–277. https://doi.org/10.1046/j.1467-2960.2001.00053.x

Mörchen, F., Ultsch, A., 2005. Discovering temporal knowledge in multivariate time series, in: Studies in Classification, Data Analysis, and Knowledge Organization. pp. 272–279. https://doi.org/10.1007/3-540-28084-7-30

Muhling, B.A., Beckley, L.E., 2007. Seasonal variation in horizontal and vertical structure of larval fish assemblages off south-western Australia, with implications for larval transport. J. Plankton Res. 29, 967–983. https://doi.org/10.1093/plankt/fbm072

Nash, R.D.M., Geffen, A.J., 2012. Mortality through the early life-history of fish: What can we learn from European plaice (*Pleuronectes platessa* L.)? J. Mar. Syst. 93, 58–68. https://doi.org/10.1016/j.jmarsys.2011.09.009

Nøttestad, L., Diaz, J., Pena, H., Søiland, H., Huse, G., Ferno, A., 2016. Feeding strategy of mackerel in the Norwegian Sea relative to currents, temperature, and prey. Leif 73, 1127–1137.

Olivar, M.P., Contreras, T., Hulley, P.A., Emelianov, M., López-Pérez, C., Tuset, V., Castellón, A., 2018. Variation in the diel vertical distributions of larvae and transforming stages of oceanic fishes across the tropical and equatorial Atlantic. Prog. Oceanogr. https://doi.org/10.1016/j.pocean.2017.12.005

Olivar, M.P., Sabatés, A., 1997. Vertical distribution of fish larvae in the north-west Mediterranean Sea in spring. Mar. Biol. 129, 289–300. https://doi.org/10.1007/s002270050169

Olivar, M.P., Sabatés, A., Alemany, F., Balbín, R., Fernández de Puelles, M.L., Torres, A.P., 2014. Diel-depth distributions of fish larvae off the Balearic Islands (western Mediterranean) under two environmental scenarios. J. Mar. Syst. 138, 127–138. https://doi.org/10.1016/j.jmarsys.2013.10.009

Olivar, M.P., Salat, J., Palomera, I., 2001. Comparative study of spatial distribution patterns of the early stages of anchovy and pilchard in the NW Mediterranean Sea. Mar. Ecol. Prog. Ser. 217, 111–120. https://doi.org/10.3354/meps217111

Ospina-Alvarez, A., Parada, C., Palomera, I., 2012. Vertical migration effects on the dispersion and recruitment of European anchovy larvae: From spawning to nursery areas. Ecol. Modell. 231, 65–79. https://doi.org/10.1016/j.ecolmodel.2012.02.001

Paris, C.B., Cowen, R.K., 2004. Direct evidence of a biophysical retention mechanism for coral reef fish larvae. Limnol. Oceanogr. 49, 1964–1979. https://doi.org/10.4319/lo.2004.49.6.1964

Ramos, S., Ré, P., Bordalo, A.A., 2010. Recruitment of flatfish species to an estuarine nursery habitat (Lima estuary, NW Iberian Peninsula). J. Sea Res. 64, 473–486. https://doi.org/10.1016/j.seares.2010.01.010

Ré, P., Meneses, I., 2009. Early stages of marine fishes occurring in the Iberian Peninsula. Lisbon.

Reglero, P., Ortega, A., Balbín, R., Abascal, F.J., Medina, A., Blanco, E., de la Gándara, F., Alvarez-Berastegui, D., Hidalgo, M., Rasmuson, L., Alemany, F., Fiksen, Ø., 2018. Atlantic bluefin tuna spawn at suboptimal temperatures for their offspring. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rspb.2017.1405

Richards, W.J., 2006. Early stages of Atlantic fishes: an identification guide for the western central north Atlantic. CRC Press.

Rodriguez, J.M., Gonzalez-Pola, C., Lopez-Urrutia, A., Nogueira, E., 2011. Composition and daytime vertical distribution of the ichthyoplankton assemblage in the central cantabrian Sea shelf, during summer: An eulerian study. Cont. Shelf Res. 31, 1462–1473. https://doi.org/10.1016/j.csr.2011.06.007

Rodríguez, J.M., Hernández-León, S., Barton, E.D., 2006. Vertical distribution of fish larvae in the Canaries-African coastal transition zone in summer. Mar. Biol. 149, 885–897. https://doi.org/10.1007/s00227-006-0270-z

Ropke, A., 1993. Do larvae of mesopelagic fishes in the Arabian Sea adjust their vertical distribution to physical and biological gradients. Mar. Ecol. Prog. Ser. 101, 223–236. https://doi.org/10.3354/meps101223

Russo, T., Scardi, M., Cataudella, S., 2014. Applications of self-organizing maps for ecomorphological investigations through early ontogeny of fish. PLoS One 9. https://doi.org/10.1371/journal.pone.0086646

Sabatés, A., 2004. Diel vertical distribution of fish larvae during the winter-mixing period in the Northwestern Mediterranean. ICES J. Mar. Sci. 61, 1243–1252. https://doi.org/10.1016/j.icesjms.2004.07.022

Sabatés, A., 1990. Changes in the heterogeneity of mesoscale distribution patterns of larval fish associated with a shallow coastal haline front. Estuar. Coast. Shelf Sci. 30, 131–140. https://doi.org/10.1016/0272-7714(90)90059-Z

Sabatés, A., 1989. Sistemática y distribución espacio-temporal del ictioplancton en la costa catalana. PhD Thesis. Universitat de Barcelonsa.

Sabatés, A., Olivar, M.P., Salat, J., Palomera, I., Alemany, F., 2007. Physical and biological processes controlling the distribution of fish larvae in the NW Mediterranean. Prog. Oceanogr. 74, 355–376. https://doi.org/10.1016/j.pocean.2007.04.017

Sánchez-Velasco, L., Jiménez-Rosenberg, S.P.A., Lavín, M.F., 2007. Vertical distribution of fish larvae and its relation to water column structure in the southwestern Gulf of California. Pacific Sci. 61, 533–548. https://doi.org/10.2984/1534-6188(2007)61[533:VDOFLA]2.0.CO;2

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089

Schreck, T., 2010. Visual-Interactive Analysis With Self-Organizing Maps - Advances and Research Challenges. Self-Organizing Maps. https://doi.org/10.5772/9171

Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., Banfield, J.F., 2013. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Res. 23, 111–120. https://doi.org/10.1101/gr.142315.112

Siegel, D.A., Mitarai, S., Costello, C.J., Gaines, S.D., Kendall, B.E., Warner, R.R., Winters, K.B., 2008. The stochastic nature of larval connectivity among nearshore marine populations. Proc. Natl. Acad. Sci. 105, 8974–8979. https://doi.org/10.1073/pnas.0802544105

Sinclair, M., Iles, T.D., 1985. Atlantic Herring (*Clupea harengus*) Distributions in the Gulf of Maine – Scotian Shelf Area in Relation to Oceanographic Features . Can. J. Fish. Aquat. Sci. 42, 880–887. https://doi.org/10.1139/f85-112

Smith, K.A., Gibbs, M.T., Middleton, J.H., Suthers, I.M., 1999. Short term variability in larval fish assemblages of the Sydney shelf: Tracers of hydrographic variability. Mar. Ecol. Prog. Ser. 178, 1–15. https://doi.org/10.3354/meps178001

Somarakis, S., Nikolioudakis, N., 2010. What makes a late anchovy larva? the development of the caudal fin seen as a milestone in fish ontogeny. J. Plankton Res. 32, 317–326. https://doi.org/10.1093/plankt/fbp132

Sponaugle, S., Cowen, R.K., Shanks, A., Morgan, S.G., Leis, J.M., Pineda, J., Boehlert, G.W., Kingsford, M.J., Lindeman, K.C., Grimes, C., Munro, J.L., 2002. Predicting self-recruitment in marine populations: Biophysical correlates and mechanisms. Bull. Mar. Sci. 70, 341–375. https://doi.org/10.1016/j.csr.2006.01.020

Teodósio, M.A., Paris, C.B., Wolanski, E., Morais, P., 2016. Biophysical processes leading to the ingress of temperate fish larvae into estuarine nursery areas: A review. Estuar. Coast. Shelf Sci. https://doi.org/10.1016/j.ecss.2016.10.022

Tojeira, I., Faria, A.M., Henriques, S., Faria, C., Gonçalves, E.J., 2012. Early development and larval behaviour of two clingfishes, Lepadogaster purpurea and *Lepadogaster lepadogaster* (Pisces: Gobiesocidae). Environ. Biol. Fishes 93, 449–459. https://doi.org/10.1007/s10641-011-9935-7

Vargas-Yáñez, M., Sabatés, A., 2007. Mesoscale high-frequency variability in the Alboran Sea and its influence on fish larvae distributions. J. Mar. Syst. 68, 421–438. https://doi.org/10.1016/j.jmarsys.2007.01.004

Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Honkela, T., Lähdesmäki, H., 2015. Self-organization and missing values in SOM and GTM. Neurocomputing 147, 60–70. https://doi.org/10.1016/j.neucom.2014.02.061

Vikebø, F., Sundby, S., Ådlandsvik, B., Fiksen, Ø., 2005. The combined effect of transport and temperature on distribution and growth of larvae and pelagic juveniles of Arcto-Norwegian cod. ICES J. Mar. Sci. 62, 1375–1386. https://doi.org/10.1016/j.icesjms.2005.05.017

Vollset, K.W., Catalán, I.A., Fiksen, Folkvord, A., 2013. Effect of food deprivation on distribution of larval and early juvenile cod in experimental vertical temperature and light gradients. Mar. Ecol. Prog. Ser. 475, 191–201. https://doi.org/10.3354/meps10129

HIGHLIGHTS

USING SELF ORGANIZING MAPS TO ANALYZE LARVAL FISH ASSEMBLAGE VERTICAL DYNAMICS THROUGH ENVIRONMENTAL-ONTOGENETIC GRADIENTS.

Álvarez, I[1*]; Font-Muñoz, J S[1,3]; Hernández-Carrasco, I[1]; Díaz-Gil, C[1,2]; Salgado-Hernanz, P. M.[1]; Catalán, I. A.[1]

- Larval fish assemblages analysis with Self Organizing Maps allows overcoming linear constraints
- Self Organizing Maps analysis allowed the description of the whole coastal multispecific system
- Self Organizing Maps also allowed identifying ontogenetic differences in vertical positions
- Differences in larval fish vertical strategies before & after the establishment of stratification

YECSS-D-20-00384 Author statement

All the authors in the list have agreed to be listed and have approved the submitted version of the manuscript. I. Alvarez and I.A. Catalán proposed the research hypothesis, analyzed the larval fish assemblages, wrote and edited the manuscript. JS Font-Muñoz and I Hernández-Carrasco did the statistical analyses and actively participated in the writing of the MS. Finally, PM Salgado-Hernanz and C Diaz-Gil participated in all the surveys and in the manuscript writing.

Declaration of interests

X The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: