A fishery observing system for the collection of fishery and oceanographic data

P. Falco, A. Belardinelli, A. Santojanni, N. Cingolani, A. Russo, E. Arneri

To cite this version:

P. Falco, A. Belardinelli, A. Santojanni, N. Cingolani, A. Russo, et al.. A fishery observing system for the collection of fishery and oceanographic data. Ocean Science Discussions, European Geosciences Union, 2006, 3 (4), pp.827-865. hal-00298590

HAL Id: hal-00298590
https://hal.archives-ouvertes.fr/hal-00298590
Submitted on 13 Jul 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A new kind of fishery observing system for the collection of fishery and oceanographic data

P. Falco¹, A. Belardinelli¹, A. Santojanni¹, N. Cingolani¹, A. Russo², and E. Arneri¹

¹Istituto di Scienze Marine – Consiglio Nazionale delle Ricerche, Sezione di Ancona, Largo Fiera della Pesca 60125, Ancona, Italy
²Dipartimento di Scienze del Mare, Università Politecnica della Marche, Via Brecce Bianche, 60131, Ancona, Italy

Received: 18 May 2006 – Accepted: 15 June 2006 – Published: 13 July 2006
Correspondence to: P. Falco (p.falco@ismar.cnr.it)
Abstract

Fishery Observing System (FOS) was developed as a first and basic step towards fish stock abundance nowcasting/forecasting within the framework of the EU research program Mediterranean Forecasting System: Toward an Environmental Prediction (MF-STEP). The study of the relationship between abundance and environmental parameters also represents a crucial point towards forecasting. Eight fishing vessels were progressively equipped with FOS instrumentation to collect fishery and oceanographic data. The vessels belonged to different harbours of the Central and Northern Adriatic Sea. For this pilot application, anchovy (*Engraulis encrasicolus*, L.) was chosen as the target species. Geo-referenced catch data, associated with in-situ temperature and depth, were the FOS products but other parameters were associated with catch data as well. MFSTEP products numerical circulation models provide many of these data. In particular, salinity was extracted from re-analysis data of numerical circulation models. Satellite-derived sea surface temperature (SST) and chlorophyll were also used as independent variables. Catch and effort data were used to estimate an abundance index (CPUE – Catch per Unit of Effort). Considering that catch records were gathered by different fishing vessels with different technical characteristics and operating on different fish densities, a standardized value of CPUE was calculated. A spatial and temporal average CPUE map was obtained together with a monthly mean time series in order to characterise the variability of anchovy abundance during the period of observation (October 2003–August 2005). In order to study the relationship between abundance and oceanographic parameters, Generalized Additive Models (GAM) were used. Preliminary results revealed a complex scenario: the southern sector of the domain is characterised by a stronger relationship than the central and northern sector where the interactions between the environment and the anchovy distribution are hidden by a higher percentage of variability within the system which is still unexplained.

GAM analysis showed that increasing the number of explanatory variables also increased the portion of variance explained by the model. Data exchange and interdisci-
plinary efforts will therefore be crucial for the success of this research activity.

1 Introduction

The analysis of the relationships between environmental conditions and the distribution of fish stocks is a complex matter. Many studies have pointed out how the environment and its variability can influence recruitment and fish distribution (Cushing, 1996; O'Brien et al., 2000; Attrill and Power, 2002; Chavez et al., 2003). Although fishing effort has grown in the recent past and is currently increasing, there is evidence that climate and environmental variability could determine large fluctuations in fish stocks. To add more complexity, the state of relationships between climate/environment and stocks could not be conclusive (Mann, 1993), it looks like they persist for one or two decades to then evolve to another state as consequence of the changing environmental conditions (Klyashtorin, 2001).

Realistic estimates regarding how fish stock abundance changes in time and space are not easy to obtain. Acoustic surveys (Swartzman et al., 1992; Daskalov, 1999; Hedger et al., 2004) are the main experimental method used to collect information about stock abundance of small pelagic fish; also, information regarding landings (Denis et al., 2002) can be of help in this evaluation. On the other hand, temporal coverage of survey data is not appropriate because it is not always possible to carry out surveys regularly. In addition, surveys do not cover the spatial extent of the stock adequately. The main issue to overcome when using landings data is the delayed availability of this information, when, for management purposes, it would be needed almost in real time. Moreover relating position to catches it is not always possible in the case of landings data. Catches can also be obtained by means of specific forms distributed among the fishing fleet of one or more harbours. In this case skippers of the fishing vessels are required fill in the forms with all the required information at the end of each fishing day. This approach may fail because fishermen can not guarantee the required attention for time-spans long enough to create appropriate time series.
Furthermore, in order to yield abundance data with time and spatial resolutions comparable with environmental data, fishery observing systems need to be improved using more advanced tools.

Generally, a good and appropriate observing system is fundamental in order to implement models to nowcast/forecast the behaviour of the marine environment and the evolution of parameters connected with it. Nowcast of fish abundance is crucial for management. The use of traditional models (e.g., Beverton and Holt, 1957) in the past did not avoid the decline of very exploited fish stocks (Pauly et al., 2002), thus the need to accomplish models which consider each single aspect of the problem, has become more and more important.

Fisheries nowcasting/forecasting is at a very preliminary stage and the number of variables involved to predict abundances of life stages (or age classes) of one species is high. One way to address the problem would consist in the realisation of a fish module in the existing and developing models of hydrodynamics and ecosystem. Population dynamics of fish are complicated in their interactions with the natural system; mortalities, by predation and/or by fishing for each life stage, are not easy to measure. A preliminary, fundamental, step toward fishery forecasting for management purposes would thus be the set up of an automated Fishery Observing System (FOS). In this paper a new kind of FOS is presented and discussed.

Typically fishery sector was and it is still considered as user of information and products derived from research activity and its role as data source have been largely ignored (Simpson, 1994). The idea here is that they became part of the observing system, not only because they can provide us with catch information but also because of the use of their daily presence at sea for collecting environmental data could open the way to a new kind of opportunity.

The FOS was tested in the Adriatic Sea. Data collection started in August 2003 and it is still ongoing. This paper presents the data collected until the end of August 2005.

The long-term objective is that of developing a short term forecasting system for fish stock abundance in the Adriatic Sea, extendable to other marine regions and species.
In this pilot application the species selected is mainly anchovy (*Engraulis encrasicolus*, L.), one of the most important commercial species, being the target of an important fishery in the northern and central Adriatic Sea with an annual catch fluctuating, at present, between 20 000 and 30 000 tonnes (Santojanni et al., 2003). The Adriatic Sea was chosen among the Mediterranean fishing areas for anchovy for three important reasons: it is the principal fishing area for this species, it is a continental basin (so relatively easy to monitor and with limited lateral advection), it is covered by regional and shelf MFSTEP models.

The paper is organized as follows: the second section provides a brief description of the status of the anchovy fishery in the Adriatic sea; in Sect. 3, the Fishery Observing System (FOS) is described and discussed along with some information regarding the data set used, the methodology used to derive the abundance index and the statistical models used to study the relationship between abundance index and environmental parameters; preliminary results are discussed in Sect. 4.

### 2 Anchovy fishery in the Adriatic Sea

Anchovy (*Engraulis encrasicolus*, Linnaeus) caught by the Italian Adriatic fishing fleet represents 90% of the total catch in the Adriatic Sea and 24% of the total Mediterranean catch (Santojanni et al., 2003; Cingolani et al., 2004). The value of Adriatic anchovy landed catches was estimated at about 35 MECU in 1998. The importance of this species is thus obvious.

The choice of anchovy as target species of this study is not only related to its economical importance but also to two other basic reasons: (1) its short life span and the dependence of year class strength on recruitment will allow the use of these data to model possible relationships between environmental parameters and recruitment; (2) the fact that anchovy is zooplanktophagous means it is directly linked to a lower trophic level which could be the final result of current investigation on bio-physical modelling (Vichi et al., 1998).
Anchovy mainly spawns from spring to autumn throughout the northern and central Adriatic and juveniles concentrate in shallow waters (less than 30 m depth) along the Italian coasts (Regner, 1996). Few months after recruitment, anchovy already reaches commercial size. Knowledge of how anchovies interact with environmental variables, especially temperature, salinity, density, stratification, fronts and other biological variables such as zooplankton, is necessary. This has not yet been achieved because, while increasingly accurate 3-D fields of environmental variables are available daily as model forecast and hindcast outputs, there is very limited (spatial and/or temporal) information about the geographical distribution of anchovies and their daily variations. A new kind of FOS is thus also intended to fill in this gap.

The Italian fishing fleet for small pelagic fishes is distributed all along the Adriatic coast and two kinds of fishing gear are currently used: mid water pelagic trawl nets towed by two vessels (volante in Italian) and light attraction purse seines (lampara in Italian). The same fishing gear catches anchovy (*Engraulis encrasicolus* L.) but also sardine (*Sardina pilchardus* Walb.) and to a lesser extent other pelagic fish such as sprat (*Sprattus sprattus* L.), horse mackerel (*Trachurus* spp.) and mackerel (*Scomber* spp.). The volante is mainly used in the northern and central Adriatic. At present approximately 70 couples of fishing vessels use this gear; their average engine power is 342 HP, average size of the vessels is 54 GRT but there are wide variations in size and engine power. Bigger lampara vessels (25 boats) operate in the Central Adriatic, south of Ancona. Here it is almost common for a fishing vessel to switch from lampara during the summer season (when there are favourable weather conditions for this fishing technique) to pelagic trawl for the remaining part of the year. During the lampara fishing season (April/May–November) some fishing vessels registered in southern Adriatic move into the Central Adriatic increasing the lampara fishing fleet up to a total of about 50/55 boats. Smaller lampara (17 boats) operate in the Gulf of Trieste.

Anchovy fishery experienced a sudden collapse in 1987, when only 700 tons were landed. Evidence from assessments suggests that the collapse was caused by very low recruitment. This was probably due to environmental factors determining the level
of recruitment (Cingolani et al., 1996; Santojanni et al., 2006). Since then, total annual catches of anchovy has increased but complete recovery did not occur. It is thus important to devise an integrated system able to collect information regarding both fish stock abundance and environmental parameters. All environmental data collected will be also useful towards the characterisation of the overall marine environment and because a near real time availability of the data will be obtained, these data could then be used in the procedure of data assimilation in the models.

3 Material and methods

3.1 Fishery Observing System

The development of the FOS was based on the need to obtain all possible data without impacting too much on the fishing activity (condition necessary in order to obtain fishermen’s collaboration). The FOS, in its last version, consists mainly of three components: an electronic logbook (EL), a GPS and a temperature and pressure recorder (Fig. 2). The latter two components are available on the market whilst the EL was developed ad hoc for this application although standard electronic components were used to build it. Each single component of the FOS generates a file: catch and position data are stored on the EL, temperature and depth data are stored on the memory of the sensor. Temperature data are collected every time the fishing gear is hauled; they have different patterns depending on which kind of fishing technique is used as it will be pointed out subsequently. Eventually, all data are stored in a database and collated.

3.1.1 Electronic logbook and positioning

The core component of FOS is the EL, in particular this is a computer with a touch screen as user interface. The EL has been strongly modified from the initial version. In the last version, as showed in Fig. 2, it is made up by two separate components, the
screen and the central unit. In this way it does not take up too much room and it looks like common equipment installed on the deck of the fishing vessel. The central unit is very compact and allows different installation solutions. A very compact mainboard (VIA EPIA PD-Series Mini-ITX) was used in order to maintain a reduced size of the central unit. It features four serial communication ports and up to six 2.0 USB connections. In our application only one USB port for direct data downloading or to connect external devices (keyboard, mouse, etc...) and three serial com ports to get permanent connection with the screen, the GPS and an external modem are needed. The latter allows data transmission to a remote computer via a GPRS/GSM connection. The production of the EL was carried out by ASYSTEL of Trevi – Italy.

Catch data are input by means of a dedicated software, programmed to be as user friendly as possible, where only the essential information are required for input. Information regarding the species are required, too. They are indicated by the software and for each species the skipper enters only the total catch for haul, an estimate of the mean size of individuals in the catch (this information is required only for anchovy and sardine) and the discards (in terms of catches and size). In order to simplify this process, the measure unit of catches was established as boxes of fish and the mean size is expressed in number of fishes per kilogram for anchovy and in two categories for sardine: large or small. These are the common ways used by fishermen in the Adriatic to indicate catch and mean size for these two species. In the post-processing phase, catches are converted in kg using a known conversion factor between boxes and kg. This factor generally differs from harbour to harbour.

A CMC Electronics Smart GPS antenna is connected with and powered by the EL. Thus every time the EL is switched on, the GPS is as well and GPS records are stored every time catch records are. Position accuracy is about 20 m. Position, date, time and speed are recorded every minute. Catch data are stored in a different file with respect to the file containing position records. At this stage catches are not geo-referenced yet as catch data could be inputted after the haul is concluded far from the real position. Haul selection and then catch geo-referencing is accomplished in the post processing
3.1.2 Temperature and pressure recorder

The third component of the FOS is a temperature and pressure sensor (TPs hereafter Fig. 2). The probe used is produced by Star Oddi company (Iceland) and its main characteristic is its reduced dimension. This characteristic is important for the current application in that the sensor is attached to the fishing gear and small equipment limits the impact on the gear and can be positioned in a safe way. The TPs used are of the DST Milli type: only 4 cm in length with a diameter of about 1 cm. Accuracy is ±0.1°C for temperature and ±0.4% of depth range. Two different sensor ranges were used: 100 m and 300 m. Memory space and also battery consumption are related to the sampling interval. The lowest value is 1 s and we used 30 s for lampara vessels and 1 min for volante as a best compromise between data time resolution and memory and battery usage. Milli sensors are self contained with the possibility of being programmed to work only during periods when data collection is required. This characteristic is particularly useful also because fishing vessels work approximately during the same hours of the day. So sampling can be continuous during working hours (even when the fishing gear is not at sea) and on stand by for the remaining time.

In order to test the performance of TPs, a comparison with a Sea Bird 911 plus probe (two order of magnitude better than TPs in accuracy) was carried out. Being $\Delta T = T_{ms} - T_{CTD}$, results show very good agreement between the temperature collected by Milli sensors ($T_{ms}$) and CTD ($T_{CTD}$) especially in areas with low temperature gradients (Fig. 3a, b). $\Delta T \sim 3^\circ C$ when data are collected in areas of strong gradient such as across a summer thermocline (Fig. 3a, b). This is because the time response of the TPs is not as fast as that of the CTD, so they are unable to follow rapid thermal variations. Together with the rate of response, the descent velocity of the sensor is actually a fundamental parameter in determining the accuracy of the data. Tests to evaluate this contribution showed that when the descent velocity of the sensor was about 10 m/min, $\Delta T \sim 2^\circ C$ in correspondence of the thermocline whereas $\Delta T \rightarrow 0$ above and underneath...
the thermocline and in particular the TPs up-cast profile was in very good agreement with the CTD one (Fig. 3b).

TPs provide a vertical temperature profile of the water column when applied to a lampara fishing gear. In this case the vessel holds the same position for the entire haul and TPs were mounted on the lowest part of the net so its vertical displacement described the vertical temperature profile of the water column. Considering that the average descent velocity of the net during a lampara haul is about 8–9 m/min and that the thermocline is present only during few months in the Adriatic Sea, the use of these sensors was assumed to be appropriate. Volante temperature data refer essentially to the trawling path of the vessels, when the net is approximately close to the sea bottom. Data collected during lowering and rising of the fishing gear can not be considered because these operations are too fast (around 20 m/min in both cases) for the characteristics of this class of TPs.

Although the performances of these sensors was satisfactory, the need to increase the accuracy of temperature measurements up to a standard oceanographic temperature recorder, required improvements of the sensors. First of all, an increase in the response rate with respect to temperature variations; secondly the use of a pressure dependent trigger to switch on the sensor; thirdly an increase in memory space. (allowing for both a faster sampling rate and a longer maintenance interval). Such improvements were carried out by Star Oddi company following our request and a new version of the sensor, called DST Logic, is now available. With these new characteristics, the accuracy was increased by about the 20% in correspondence to strong temperature gradients (Fig. 4a, b). Furthermore, a higher sampling interval can be used during the haul because the Logic sensor works during the haul only and memory space is thus saved. Enlargement of memory space, allowed to further increase sampling interval. Logic are currently set up with a sampling interval of 1 s for lampara application and 5 s for volante.
3.1.3 Data set

FOS was installed on eight boats belonging to fleets of Chioggia, Rimini, Ancona, S. Benedetto del Tronto and Giulianova degli Abruzzi, from north to south respectively (Fig. 1). Initially the decision was to monitor the fleet of one harbour (Ancona) in order to increase the statistical relevance of the data and assuming that more vessels could guarantee a good spatial coverage. Successively, because fishing vessels of the same fleet typically operate approximately within the same fishing areas, it was assumed that one boat could be representative of the movement of the entire fleet, and the decision was taken to sample vessels belonging to fleets of different harbours (Rimini, Ancona, S. Benedetto del Tronto and Giulianova degli Abruzzi). This choice ensured a good spatial coverage at least of the middle Adriatic. Four vessels were equipped with the instrumentation within the first 3 months from the beginning of data collection (August 2003). The other four were provided with a FOS a year and half later, thanks to Italian national funding of the Directorate General for Fishery of the Ministry of Agriculture. At present two vessels per harbour are monitored, with the exception of Chioggia and Giulianova (where only one vessel is monitored). Giulianova degli Abruzzi and San Benedetto del Tronto vessels change fishing gear in spring and until the late autumn they are set up as lampara. For the remaining part of the year they fish as volante along with the other monitored fishing vessels.

FOS provides three different files which are uploaded on a database developed in MS Access. The uploading procedure ends with data collation: catch data are georeferenced and associated with temperature at depth data. This operation is done for each volante and lampara haul. The selection of haul positions is based on depth data obtained by TPs. The start time of the haul is set to coincide with the lowering of the gear and the end time coincides with the lifting of the gear. All positions included in this time interval are then selected. Catch data are input sequentially and for every fish day, there is a correspondence between the temporal sequence of catches and TPs measurement. Thus the order of the hauls (as they was stored in the catch file)
and their date allow merging of daily TPs data with catch and position data. While catch data covered all the period of measurement, temperature and pressure data were sometimes missing. The reasons for these failures are different: first of all in some cases the sensors were lost or damaged. Moreover, when the GPS does not receive the signal, TPs data cannot be associated with geographic position.

3.2 Data analysis

In the present paper, data from pelagic trawlers only were considered in the analysis as they are more numerous than purse seiners, they cover a greater area and they have a longer temporal coverage. Table 1 summarises information regarding volante hauls for the period October 2003–August 2005. In particular, the total number of hauls for each vessel is reported together with the number and percentage of hauls with catch, temperature and pressure and position data.

Data obtained from each FOS component were merged: catch data were geo-referenced and associated with the along trawler track average temperature and depth. The next step consisted in estimating an index of relative fish abundance. To do this we followed the classical method based on the use of catch and effort data (Richards and Schnute, 1992; Goñi et al., 1999; Campbell, 2004; Santojanni et al., 2005). MF-STEP satellite and model products were selected and used as explanatory variables. Generalized additive models (Hastie and Tibshirani, 1991) were used in order to study relationships between the dependent (CPUE) and independent (environmental parameters) variables.

3.2.1 Evaluation of anchovy abundance

Firstly, an estimate of fish abundance was calculated. To do this we followed the classical method based on the use of catch and effort data. Catch/abundance relationship is expressed by the equation:

\[ C = qED \]  

(1)
where $D$ indicates the fish density, $q$ is the catchability coefficient and $E$ is the fishing effort. The fishing effort can be defined as the sum of means deployed for catching fish in a defined area over a defined period of time (Annex V of SEC (90)2244, Commission Communication to the Council and Parliament of the European Communities on the Common Fisheries Policy). Basically fishing effort comprises two capacity elements (vessel and gear) and an activity measure (time); thus it can be affected by either of these components. Catchability, as effort, depends on technical parameters but also on biological factors such as fish availability and behaviour during fishing activity. From a statistical point of view it can be considered as the probability of any single fish to be caught. The Catch per unit of Effort CPUE is given by the ratio:

$$CPUE = \frac{C}{E} \quad (2)$$

If $D$ is expressed in terms of number of fish $N$ over the fishing ground area $A$, then Eq. (2) can be rewritten as:

$$CPUE = \frac{C}{E} = \frac{qN}{A} \quad (3)$$

$CPUE$ is thus a function of the coefficient $q$ and of the number of fish. If the variations of $q$ can be accounted for, CPUE depends on the number of fish $N$ only. A correct evaluation of $q$ is thus crucial in order to use catch rate as an index of abundance. $q$ depends on many variables which are not constant in time and space, and they also vary for different fishing vessels. The typical way to account for these factors is to remove their effects on catch rates. This process is known as catch effort standardization (Richards and Schnute, 1992; Goñi et al., 1999; Campbell, 2004; Maunder and Punt, 2004; Santojanni et al., 2005). Formally, all the effects are represented by a multiplicative model. Considering a fishing vessel $A$, the model relates the catch rate of $A$ to the catch rate of a reference vessel and also it has to account for the difference between the current and standard vessel in terms of fishing power, area effects, time spent at sea and other possible factors. The reference vessel can be one of those monitored
and a well accepted method to determine the proportional coefficients consists in using Generalized Linear Modelling (GLM) (Hilborn and Walters, 1992; Venables and Dichmont, 2004). This procedure provides an indication of the relative importance of the factors influencing catch rates and the computation of the standardized abundance index as well. A map of the a-temporal distribution of anchovy abundance and a monthly mean values are represented respectively in Figs. 5 and 7.

3.2.2 Environmental data (from observations and numerical models)

Temperature measurements at net depth were collected during the hauls. These measurements provide a parameter generally considered important in determining fish aggregation, directly or indirectly. Many studies have pointed out the relationship between fish abundance and SST (Simpson, 1994; Cole, 1999; Waluda et al., 2001; Yanez et al., 2001), the latter being obtained by satellite measurements. Within the framework of MFSTEP, SST data and also surface chlorophyll are made available for the entire Adriatic Sea by the ISAC-CNR, Rome (http://www.bo.ingv.it/adricosm, Böhm et al., 2003) and temperature at depth data are also available as model outputs (ADRICOSM project, funded by the Italian Ministry for Environment). Models also provide other important physical parameters such as salinity, which has been used as a further explanatory variable in the analysis.

In situ temperature data were used in the analysis. Model temperature data could be very useful for integration with in situ data or SST when they are not available. Figure 6 illustrates a comparison between in situ temperature data and model temperature reanalysis for the period October 2003–November 2004. Reanalysis data is not yet available following November 2004. For this comparison temperature data collected by pelagic trawlers were used. For each haul an along track average value was obtained considering only the data collected when the net was at fishing depth (that can be close to the sea bottom or at an intermediate level depending on where the fish school was detected). The haul average position was reported on the model grid and a temperature value was linearly interpolated on each haul position using the four nearest grid
values. In particular data gathered by one pelagic trawler from Rimini are represented in Fig. 6 where the two curves are in good agreement except in correspondence of quick warming which occurred during the first decade of June 2004 (in Fig. 6 from haul number 180 to 200) when model data seem to overestimate the actual temperature. After this transitioning phase, the two curves were, once again, in good agreement. This was also confirmed by regression analysis where $R^2$ is 0.94. Mean and standard deviation of the difference between in situ and model temperatures were calculated for the first four fishing vessels used in this experiment. The mean did not exceed 0.5°C and the standard deviation was at worst about 1.5°C.

3.2.3 Generalized Additive Models (GAM)

Generalized additive models are suitable tools for exploring a data set and pointing out relationships between the dependent variable and the independent variables (Hastie and Tibshirani, 1991). GAM have already been used in spatio-temporal stock assessment modelling of different marine species (Swartzman et al., 1992; Daskalov, 1999, Hedger et al., 2004), as well as to carry out temporal analysis of commercial trawler data (Denis et al., 2002) or to improve catch-at-age indices of target species (Piet, 2002). GAM were also proposed as one possible comparison standards for catch data obtained from commercial logbooks (Walsh and Kleiber, 2004).

GAM are more flexible tools than traditional parametric models such as linear or nonlinear regression in that some parametric assumptions are relaxed, allowing to uncover patterns in the relationships between the dependent variables and independent variables, that could otherwise be missed.

Let $Y = (y_1, \ldots, y_p)$ be the response variable and $X = (x_1, \ldots, x_p)$ be a set of predictor variables. The linear relationship between expected value of $Y$ and $X$, assumes the form:

$$E(Y) = f(x_1, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$ (4)

Additive models generalize linear models by substituting the $\beta$ parameters in Eq. (4)
with smoothing functions. In this case, if \( s_1, \ldots, s_p \) are the smoothing functions, the expectation \( Y \) can be written as:

\[
E(Y) = f(x_1, \ldots, x_p) = s_0 + s_1 x_1 + \ldots + s_p x_p
\]  

(5)

The \( s \) functions are local smoothers and are estimated in a nonparametric way. Spline functions are commonly used to represent the smoothing terms.

Generalized additive models extend linear models in another way, namely by allowing for a link between \( f(x_1, \ldots, x_p) \) and the expected value of the dependent variable. In this way an alternative distribution for the underlying variation can be considered and not just normal distribution. There are many statistical applications where Gaussian models can not be used allowing GAM to be applied to a wider range of case studies.

Generalized additive models consist of a random component, an additive component, and a link function relating first two components. The response \( Y \) is the random component and it is assumed to have a density in the exponential family:

\[
f_Y(y; \theta; \phi) = \exp \left[ y\theta - b(\theta) + c(y, \phi) \right]
\]  

(6)

where \( \theta \) is called the natural parameter and \( \phi \) is the scale parameter. Normal, binomial, and Poisson distributions are all included in this family, together with others. The additive component can be written as:

\[
\eta = s_0 + \sum_{i=1}^{p} s_i(X_i)
\]  

(7)

where \( s_i \) are the smoothing functions. Eventually, the link function \( g \) defines the relationship between the mean \( \mu \) of the response variable and \( \eta \). In general this link can be expressed by the relation \( g(\mu) = \eta \) whereas the most popular link function is the canonical link by which \( \mu = \eta \).
Generalized linear models are particularly useful when exploration of the data set and visualisation of the relationship between the dependent variable and the independent variables are of concern. Where conventional linear techniques failed to describe a possible relationship, GAM can be used because of their capability to model non-linearities by means of smoothing functions. The price to pay to be more flexible than linear models is twofold: a reduced possibility of performing statistical inference and the increase in the number of degrees of freedom used by the smoothing terms.

The goodness of the fit is estimated by means of the deviance. Technically it is defined as “likelihood ratio statistic for testing any specific model within the saturated model (a model with as many parameters as there are observations), assuming the scale parameter is known and has the value 1” (Venables and Dichmont, 2004). For the normal case the relation to evaluate the deviance is the same as the residual sum of squares. Then for identity link its distribution is proportional to the chi-squared.

4 Results

4.1 Catch per unit of effort estimation

General linear models (GLM) were applied in order to derive a standardized value of CPUE.

One of the two pelagic trawlers from Ancona was chosen as reference vessel unit. The duration of the hauls was highly variable, so in order to render each single haul catch value comparable with the others, an hourly catch value was calculated and transformed to kg.

A first explanatory variable of the model accounts for the difference in fishing power between vessels. Fish abundance is not constant in time, thus an independent variable representing time variation was introduced. The spatial variability of abundance was represented through a variable Area. As shown in Fig. 1, the domain was divided into 10 areas: 5 further inshore and 5 off-shore. The isobaths in white in Fig. 1 represent
the limit between the inshore and off-shore areas, while the areas are separated in latitudinal sense every 0.6 deg from 42.2 N to 44.8 N and the northernmost sector of the Adriatic (the Gulf of Trieste) was entirely considered. Following initial attempts at using all data available, data collected by the Chioggia trawler were discarded in order to prevent problems related to collinearity because any possibility of overlapping with other vessel existed. Catches obtained in the southernmost sector of the domain (basically the area of S. Benedetto del Tronto and Giulianova) are characterised by a lower number of boxes but anchovies of greater size than catches obtained in the northern sector (Rimini and Chioggia areas). Ancona fishing vessels can overlap on both these fishing grounds and also operate on an exclusive area. In this case, catch and size have an intermediate value between northern and southern values. Thus a further independent variable was introduced to account, in particular, for differences in mean size of individuals in the catch (the number of boxes being to some extent a consequence of the size of anchovy caught). Three size classes were chosen: the first including all catches with sizes lower than 55 anchovies/kg; the second class ranging between 55 and 70; the third including catches with more than 70 anchovies/kg. The first class was thus characterised by the biggest anchovies.

For each haul a standardized value of CPUE was calculated and used in the analysis. The a-temporal representation of the distribution of CPUE is shown in Fig. 5. The domain was divided in square cells of 0.2° × 0.2°. This choice ensured a good compromise between spatial resolution and statistical reliability. An average CPUE value was then computed for each cell along with the corresponding rms value (not shown here). Highest values of CPUE were found in the northernmost sector of the domain and moving southward, along a coastal band that could remind the path of the Western Adriatic Coastal Current and the underlying Dense Water Outflow Current (WACC – Cushman-Roisin et al., 2001).

Besides a temporally and spatially averaged distribution of CPUE, a monthly mean value was also evaluated. This information is important in order to detect a possible trend in abundance. In particular, the exponential values obtained from the GLM anal-
ysis for the variable time, are represented in Fig. 7. The first month of observation (October, 2003) was considered as reference point, so its value is set to 1 and all the other values were referred to it. Furthermore, the difference between the current and the reference value provides the abundance increase or decrease in percentage. The CPUE monthly mean curve shows a strong increase starting from June 2004 to the end of July 2004. The reason for this behaviour is not so clear especially if compared with the same month in 2005 when abundance was less than half. This may be an artefact introduced by the computation of CPUE considering that the number of observations used to evaluate CPUE for June and July 2004 were the lowest. A decrease was revealed in September 2004 (generally, in the Adriatic Sea, trawler activity is interrupted during the month of August) followed by a recovery in the subsequent two months. A weak descending trend can be observed between November 2004 and July 2005. At present it is wise to skip any speculation can be carried about the reasons of such behaviour and wait for more observations which will allow a better comprehension.

4.2 GAM explanatory analysis

Standardized CPUE values were analysed to find possible relationships with environmental parameters. Some environmental parameters were available as model outputs and others as in situ measurements. Position, depth, temperature and salinity were chosen for GAM application. The first four parameters are FOS products whilst salinity is a model output. In particular salinity re-analysis data, which were available from March 2001 to November 2004, were used; therefore for this application salinity data are available for the period October 2003–November 2004. Remote sensing data cannot be used at this stage of the analysis because of there are too many hauls without measurement and GAM require great spatial and temporal coverage. Separate considerations regarding relationships of satellite products and abundance will be made in the next paragraph. GAM were applied to three vessels: in particular one vessel from Ancona, one from Rimini and the vessel from Giulianova because they had the longest and most continuous time series.
Along with information about possible relationships between dependent and independent variables, GAM also provide information as to which could be the optimal set of explanatory variables. To do this, the Akaike's information criterion was used (AIC).

Different models were set up in order to find the optimal set of explanatory variables. Among these models, three main cases were chosen. In model 1 only FOS data were used, thus position, depth and temperature. This was also the application with the complete series of catch data (October 2003–August 2005). In model 2, salinity was introduced and the time series limited to November 2004 as in model 3, where position data were discarded and only depth, temperature and salinity were used. The size of data set used has effects on the AIC value, so this aspect has to be borne in mind during the discussion following.

No conclusive results were found and a complex scenario appeared. The best result (in terms of both deviance explained and AIC value, see Table 2) was obtained for the Giulianova data set using all the independent variables in the GAM application (Model 2). The effects of the explanatory variables on abundance are represented for this case in Fig. 8. Abundance appears to increase eastwards and southwards. This is in contrast with the indications derived from depth and salinity plots where the increase in abundance was found in correspondence with decreasing depth and salinity. Of course the latter two variables are correlated thus collinearity problems may arise. On the other hand, the deviance explained by model 1 (without salinity) was lower and AIC is greater than in model 2. Thus the decision was taken to keep salinity (or depth) in the analysis. Figure 8f reveals a stronger relationship between depth and abundance in model 1 compared to model 2. Only depth, temperature and salinity were used in model 3. AIC was higher (but the number of observations lower) than in model 1 and the explained deviance was only a little bit lower than in the model 2 applications.

GAM application on the data set of the vessel of Ancona yielded less information compared to the previous case. Model 2, again, explained a higher percentage of deviance but it had the highest AIC.

In this case, acceptable results was given by a fourth combination of independent
variables. Using longitude, temperature and salinity, AIC was 1233 and the deviance explained about 20%. The effects of the three variables on abundance are illustrated in Fig. 9a, b, c. A weak increase in abundance occurred in correspondence with high temperature and salinity values and moving westwards (near the coast). As before, the two latter results are in contrast.

The situation is more confused in Rimini. Model 1 explained only 8% of the deviance. Introducing salinity, the value jumped to the 22% (approximately the same value as Ancona). When a model including only temperature and salinity as explanatory variables was used, the salinity vs abundance plot (Fig. 9d) showed a maximum in abundance at approximately 35.5 psu to then quickly decrease. The maximum was less evident in the other GAM application. The AIC value in this case was 1622, lower than models 2 and 3.

4.3 Satellite data analysis

Although their contribution could be important, satellite products (SST and chlorophyll) were not used in GAM application, because of the paucity of data. SeaWIFS chlorophyll concentration data were available up to December 2004 thus all the 2005 hauls could not be associated with this parameter. SST data were, unfortunately, missing for 52% of the hauls from Giulianova, 55% of the hauls from Ancona and 58% of the hauls from Rimini.

Using the data available a regression analysis was carried out. SST at this stage did not appear to play any significant role in driving fish aggregation. Chlorophyll behaved approximately in the same way with the exception of Giulianova abundance data. In this case the $R^2$ was 0.49 and the correlation was statistically significant (t-test, $p>0.01$).
5 Discussion

Factors influencing small pelagic fish aggregation might rely on spawning, feeding, migration and physical-biological effects (Agostini and Bakun, 2002). Study regarding how environmental conditions influence small pelagic distribution and abundance pointed out that temperature could be an important factor (Yáñez et al., 2001; Beare et al., 2004) but also wind (LLoret et al., 2004) and salinity (Regner, 1996) could play some roles in driving aggregation. Estimates of fish biomass and distribution are generally obtained from experimental scientific survey, which are limited in time. Standard population dynamic assessment methods based on commercial data (e.g., Virtual Population Analysis – VPA, Hilborn and Walters, 1992) do not give any information on fish distribution but only on biomass and level of exploitation. The FOS is trying to bridge these gaps because a time continuous picture of fish abundance distribution is obtained by means of geo-referenced abundance index (CPUE) derived from data collected by commercial fishing vessels. Moreover these CPUE index could provide a near real time monitoring of the state of the stock.

A new kind of fishery observing system was set up in order to collect both information regarding catches of small pelagic species of the Adriatic Sea and environmental parameters. An electronic logbook (EL) was installed on eight fishing vessels of the Adriatic small pelagic fishing fleet and catch data were inputted directly by the skippers of the vessels and stored on the EL. Periodically data were downloaded and they were available in digital format. This aspect represents a considerable improvement (both in terms of quality and quantity of data) with respect to the standard methods used to recover these information.

Together with catch information, temperature data at fishing depth were also gathered for each haul by means of a temperature and pressure sensor (TPs) attached to the fishing gear. This approach and in particular the use of the actual fishing depth, can results in an increase of significance of the relationship between abundance data and explanatory variables as pointed out for large pelagic fish (Podestà et al., 1993;
Moreover temperature data, as by product, can be used as data to assimilate in numerical circulation models.

Preliminary data analysis concerned mainly catches of anchovies (*Engraulis encrasicolus*, L.) and data obtained by pelagic trawlers. FOS data (geo-referenced catches associated with temperature and depth) were merged with MFSTEP model products (mainly salinity) and SST and chlorophyll obtained by remote sensing.

Standardized catch per unit of effort (CPUE) was obtained, for each haul, using a general linear model where, together with a variable taking into account the different technical characteristics of the vessels, three other variable were introduced: a time variable (abundance is not constant in time), an area variable (fish density varies in space) and eventually a variable which accounted for the different mean sizes of anchovies caught.

Generalized additive models (GAM) (Hastie and Tibshirani, 1990) were used in order to explore the effects of each environmental variable on the standardized abundance and also to find a possible model which could represent all these effects. Preliminary results point out that in general the “optimal” models (lowest AIC) were the ones determined using all the explanatory variables (position, depth, temperature and salinity, model 2 in Table 2). The maximum deviance explained by GAM was 45% for the Giulianova data set. This value was twice the deviance explained by the application of the same model to the Ancona and Rimini data sets. Results would suggest that the oceanographic features determine abundance distribution in the area of Giulianova, this fact due probably to a dynamic which can be more favourable to fish aggregation. In fact the coastal dynamic south of Ancona is characterised by a particularly intense WACC in turn characterised also by high level of kinetic energy (Cushman-Roisin et al., 2001; Poulain, 2001). Areas of convergence could be found both where the WACC encounters the descending branch of the Middle Adriatic gyre and in correspondence of the frontal zone which separates the coastal fresh water from off-shore saltier waters. Bottom topography can be also an important factor in determining areas of strong gradient such as shelf break fronts (Garcia and Palomera, 1996). Such conditions could
contribute to fish aggregation in the Giulianova area (central Adriatic) but it is not clear yet what the actual role of the skipper in using his own knowledge to choose the best fishing area as function of environmental conditions could be. GAM-model 2 analysis of the data collected by the fishing vessel of S. Benedetto (even though they are still too scarce and thus not discussed in the previous section) which works approximately on the same fishing ground as the Giulianova vessel, shows the same results (deviance explained 46%, lowest AIC). Regression analysis of chlorophyll data obtained by satellite measurements, showed a statistically significant relationship (t-test, p>0.01) with abundance with an $R^2$ equal to 0.49. Such relationship is verified only for data from Giulianova. These results would allow to detect probable effects of the environment on abundance in the southern sector of the domain. Results obtained from data of Ancona and Rimini are more confused. Model 2 still provided the highest value of deviance. The strong increase in deviance obtained by adding salinity as explanatory variable in the analysis of the Rimini data set is worth of notice (Table 2). The Adriatic Sea is a semi-enclosed basin where the contribution of the Po River outflow, strongly influences both the physics and the biology of the environment. The importance of correlation between anchovy abundance and recruitment with river freshwater output was already proved in previous studies (e.g. LLoret et al., 2004; Santojanni et al., 2006) and river effects can be also noticeable in semi-enclosed basin (Daskalov, 1999). The characterisation of river discharge, the spreading of fresh water and haline fronts location should be analysed in relation to CPUE index at spatial and time resolution which would be probably different with respect to the present study.

Remote sensing data could provide more information in this directions. Unfortunately these were not used in GAM analysis because the percentage of hauls for which satellite data was available was low. But satellite products are actually important and so some work should be done to integrate satellite data in the set of explanatory variables as well as bio-chemical models data, that could add very important elements towards the understanding of anchovy distribution and act to reduce the uncertainty around the mechanisms which link physical and biological processes.
Forecasting fish abundance is actually a difficult and challenging activity. The inclusion of a component which also provides resource estimates into numerical models for environmental prediction, has not yet been achieved. Model data can be used to provide and/or integrate experimental data in order to study the relationships between resource patterns and the environment. In this context a resource observing system has to provide data with a spatial and temporal resolution comparable with model outputs. FOS discussed so far is a first and successful attempt towards this. It could be applied to other areas of the Mediterranean Sea where fishing activity is important in order to create a wider monitoring system. Interactions and a stronger link between the research community and the fishing industry (Simpson, 1994) as well as regulatory authorities, would be a first and consistent step towards the institution of an operational fishery oceanography framework aiming to a better management of the resources.

Acknowledgements. This work was carried out within the framework of the Mediterranean Forecasting System: Toward an Environmental Prediction (MFSTEP) activities, project funded by the European Commission V Framework Program: Energy, Environment and Sustainable Development and also supported by a national program funded by the Italian Ministry of Agriculture. The authors would like to acknowledge G. Giannetti, B. Morello, M. La Mesa and D. Flocco for valuable discussions The authors would like to acknowledge also M. Loprejato for the valuable help given to check the correct working of the database and E. Paschini for the CTD data comparison. GAM analysis were carried out using R software (R Development Core Team, 2004). Generic Mapping Tools (Wessell and Smith, 1991) were used to produce maps.

References

Beare, D., Burns, F., Jones, E., Peach, K., Portilla, E., Greig, T., McKenzie, E., and Reid, D.: An


Cushing, D. H.: Toward a science of recruitment in fish populations, Oldendorf/Luhe, 175, 1996.


853
Table 1. Pelagic trawlers catch data information.

<table>
<thead>
<tr>
<th>Vessel</th>
<th>Total number of hauls</th>
<th>Number and percentage with T/P data</th>
<th>Number and percentage with catch data</th>
<th>Number and percentage with position data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancona 1</td>
<td>1376</td>
<td>980 (71%)</td>
<td>1137 (83%)</td>
<td>820 (60%)</td>
</tr>
<tr>
<td>Ancona 2</td>
<td>530</td>
<td>508 (96%)</td>
<td>447 (84%)</td>
<td>475 (89%)</td>
</tr>
<tr>
<td>Chioggia</td>
<td>346</td>
<td>322 (93%)</td>
<td>291 (84%)</td>
<td>314 (91%)</td>
</tr>
<tr>
<td>Giulianova</td>
<td>635</td>
<td>534 (84%)</td>
<td>528 (83%)</td>
<td>444 (70%)</td>
</tr>
<tr>
<td>Rimini 1</td>
<td>1343</td>
<td>1131 (84%)</td>
<td>988 (73%)</td>
<td>824 (62%)</td>
</tr>
<tr>
<td>Rimini 2</td>
<td>554</td>
<td>473 (85%)</td>
<td>527 (95%)</td>
<td>364 (65%)</td>
</tr>
<tr>
<td>S.Benedetto 1</td>
<td>166</td>
<td>148 (89%)</td>
<td>131 (79%)</td>
<td>120 (72%)</td>
</tr>
<tr>
<td>S.Benedetto 2</td>
<td>215</td>
<td>209 (97%)</td>
<td>143 (66%)</td>
<td>144 (67%)</td>
</tr>
</tbody>
</table>
Table 2. Comparison of three GAM applications. Explanatory variables used are for the application of Ancona (AN) and they are the same for the other two cases of Giulianova (GIU) and Rimini (RIM).

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Deviance explained</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN-Model 1 (lon,lat,temp,depth)</td>
<td>1068</td>
<td>20%</td>
<td>17</td>
</tr>
<tr>
<td>AN-Model 2 (lon,lat,temp,depth, sal)</td>
<td>1279</td>
<td>23%</td>
<td>21</td>
</tr>
<tr>
<td>AN-Model 3 (depth,temp,sal)</td>
<td>1273</td>
<td>17%</td>
<td>13</td>
</tr>
<tr>
<td>GIU-Model 1</td>
<td>1117</td>
<td>33%</td>
<td>17</td>
</tr>
<tr>
<td>GIU-Model 2</td>
<td>1003</td>
<td>45%</td>
<td>21</td>
</tr>
<tr>
<td>GIU-Model 3</td>
<td>1055</td>
<td>37%</td>
<td>13</td>
</tr>
<tr>
<td>RIM-Model 1</td>
<td>1549</td>
<td>8%</td>
<td>17</td>
</tr>
<tr>
<td>RIM-Model 2</td>
<td>1657</td>
<td>22%</td>
<td>21</td>
</tr>
<tr>
<td>RIM-Model 3</td>
<td>1627</td>
<td>17%</td>
<td>13</td>
</tr>
</tbody>
</table>
Fig. 1. Pelagic trawler tracks collected during the period October 2003–August 2005. White tracks indicate the hauls carried out by the fishing vessel from Giulianova, in red and cyan the hauls of the two trawlers from S. Benedetto del Tronto, in black and orange the tracks of two vessels from Ancona, in purple and green the tracks of Rimini vessels and in light blue the tracks of Chioggia vessels. The areas (labelled from 1 to 10) used in the CPUE standardization processes are also represented.
Fig. 2. FOS components.
Fig. 3. Temperature profiles collected by a Sea Bird CTD (red curve) and Star Oddi Milli sensor (blue curve). In panel (a) Milli sensor sampling interval is 1 s and the descent velocity is approximately 1 m/s. In panel (b), Milli sensor sampling interval is 30 s and the descent velocity is approximately 10 m/min.
Fig. 4. Two examples of temperature profiles collected using Sea Bird CTD (red curve), Star Oddi Milli sensor (black curve) and Star Oddi Logic Sensore (blue curve).
Fig. 5. Spatial distribution of standardized CPUE.
Fig. 6. Comparison between in situ temperature measured during Rimini trawler hauls (red curve) and re-analysis temperature obtained from MFSTEP numerical models (blue curve).
Fig. 7. Monthly standardized CPUE time series.
Fig. 8. Relationship between longitude (a), latitude (b), depth (c), temperature (d) and salinity (e) and abundance obtained applying GAM (Model 2) to Giulianova data set. In the last panel (f) the effects of depth on abundance data obtained applying (Model 1) are represented.
Fig. 9. Relationships between longitude (a), salinity (b) and temperature (c) on abundance applying GAM to the Ancona data set. In panel (d) effects of salinity on abundance for the case study of Rimini are shown. This result was obtained applying GAM with temperature and salinity as explanatory variables only.