Fisheries Research 111 (2011) 53–64

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

When behaviour reveals activity: Assigning fishing effort to métiers based on VMS data using artificial neural networks

Tommaso Russo [a,*], Antonio Parisi [b], Marina Prorgi [a], Fabrizio Boccoli [c], Innocenzo Cignini [c], Maurizio Tordoni [c], Stefano Cataudella [a]

[a] Laboratory of Experimental Ecology and Aquaculture – Department of Biology – "Tor Vergata" University of Rome, via della Ricerca Scientifica s.n.c., Rome 00133, Italy
[b] SEFeMEQ – Department of Financial Economics and Quantitative Methods – Faculty of Economics – "Tor Vergata" University of Rome, via Columbia 2, Rome 00133, Italy
[c] Fleet Register Unit – PEMACIII – General Directorate for Maritime Fisheries and Aquaculture – Ministry for Agriculture, Food and Forestry, viale dell'Arte 16, Rome 00144, Italy

ARTICLE INFO

Article history:
Received 8 March 2011
Received in revised form 17 June 2011
Accepted 21 June 2011

Keywords:
Vessel monitoring system
Fishery
Fishing gear
Fishing intention
Fishing métiers
Multilayer perceptron network

ABSTRACT

The identification of groups of vessels with the same exploitation pattern (e.g. gear used, fishing ground, target species) over time, usually referred to as a "métier", is a major topic of fishery management. However, métier detection on fishing trips is still done using the incomplete or biased information present in declaration of landings (logbooks), under the assumption that the reported landings profiles reflect intended catches. Nowadays, fishing effort can be tracked at high spatial resolution using vessel monitoring systems (VMS). VMS potentially provides information about vessels fishing activity if the frequency of signals is natively high or appropriately interpolated. An artificial neural network is used to analyse interpolated VMS tracks and Vessel Register data to identify fishing activity. A multilayer perceptron network (MPN) was trained to recognize one among 15 possible métiers from a series of 33 variables: 12 in binary form for licensed gears, 6 probability classes for vessel speed, 3 for vessel heading and 7 for sea depth, respectively. The MPN was iteratively trained on subsamples of a large dataset corresponding to the activity of the Italian fishing fleet, for which information about métier was collected and validated by on-board observations by scientific operators, and then tested on other subsets of the data. The best architecture for MPN was identified and analysed. The mean percentage of correct predictions obtained on the test datasets was very high (>94%), confirming that VMS data can provide information on vessel activity. Overall, these findings suggest that this is a promising approach to assign fishing effort, resolved at single trip scale, to specific métiers, even giving independent assessment of fishing activity with respect to those provided by logbook and capture data.

1. Introduction

In the last decade, the rapid decline of many fish resources has forced European fisheries research and management bodies to revise fundamental aspects of fisheries management, moving towards an ecosystem approach to fisheries (EAF – FAO, 2008). The European Commission, in the attempt to implement an EAF for marine management and exploitation of sea resources (COM, 2008) established a framework for surveying and management of the fisheries sector through the Data Collection Regulations and the data collection framework (DCF) of the common fishery policy (EC, 2008a,b). The previous vision of fisheries resources as independently exploited units has been replaced by an awareness of interactions among species and among fleets, while the environmental effects of various fishing gears has become evident (ICES, 2009a). For this reason, one of the concepts that received particular attention has been those of *fishing activity* (*or strategy*). The term refers to the particular spatial and temporal management of the vessel (and related fishing gears) used by each fisher (ICES, 2009b). The study of fishers' behaviour as a source of impact on ecosystems and the effects of regulations on fishers' activity has become an active area of research (Herrero, 2004; Hilborn, 1985; Salas and Gaertner, 2004). Fleets therefore need to be disaggregated into homogeneous categories of fishing activities, defined according to target species or assemblages, fishing area and fishing season (Marchal, 2008; Salas and Gaertner, 2004; Tzanatos et al., 2006). This appears to be particularly important in multi-species, multi-fleet fisheries such as those in Mediterranean Sea (Moranta et al., 2008) and specifically the Italian fleet (Colloca et al., 2003), in which a variety of species are caught together given a complex scheme of technical interactions among fleets (Castro et al., 2010; Palmer et al., 2009). Thus, a major topic of fishery management is concerned with

* Corresponding author.
E-mail address: Tommaso.Russo@Uniroma2.it (T. Russo).

0165-7836/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2011.06.011

identifying groups of vessels with the same exploitation pattern (e.g. gear used, fishing ground, target species) over time, usually referred to as a "métier" (Marchal and Horwood, 1996; Pelletier and Ferraris, 2000; Ulrich et al., 2001). The track, that is a vessel trip starting and ending in a given harbour, is the sample unit at which fishing activity is analysed and classified to métier (Marchal, 2008) within the DCF (http://datacollection.jrc.ec.europa.eu/dcf-fish/bio-metier/sampling). However, it should be stated that a track may actually correspond to more than one métier (Davie and Lordan, 2009).

The concept of métier is directly linked to fishing activity since it reflects the fishing intention (e.g. the species targeted, the area visited and the gear used) at the start of fishing trip (Marchal, 2008). In this way, the métier may be a appropriate way to classify fishing impacts, and is likely to be relevant when analysing the effects of fisheries on the ecosystems (ICES, 2009a). The DCF therefore contains a series of fishery-related targets (e.g. catch and effort), which are specified by métier. This legislation reinforces the need for suitable methods to identify the métier of fishing tracks. This issue has been historically addressed in three ways: (1) by the so-called "output-based" methods (Marchal, 2008), consisting of *a posteriori* quantitative analyses of catch or landings composition, with or without effort information (gear, season, location) to assign métier. Various multivariate procedures have been applied for this purpose: principal component analysis (PCA; Castro et al., 2010; Jabeur et al., 2000; Laurec et al., 1991), multiple correspondence analysis (Pelletier and Ferraris, 2000), and cluster analysis (Lewy and Vinther, 1994); (2) by the so-called "input-based" (Marchal, 2008) methods, based on *a priori* qualitative knowledge of the fisheries, mainly collected during face-to-face interviews, so that the allocation of each fishing trip to métier relies on a process of trial and error, by deriving discriminating thresholds based either on landings (weight or value), or mesh size (Biseau, 1998; Tétard et al., 1995; Ulrich and Andersen, 2004); or (3) by a combination of the input and output-based methods (Bastardie et al., 2010; Ulrich and Andersen, 2004). However, both types of method (as well as their combinations) have some limitations: they depend on available data, and landings data may be unreliable and may not contain enough information to infer fishing intention (Chang, 2011; Marchal, 2008). In addition, data collection about landings can require unquantifiable additional cost to be collected while VMS, which is not cost-free, are mandatory and so are automatically available.

A step forward is represented by the fact that, since 2000, European fishing vessels with a length overall (LOA) ≥10 m are obliged to keep logbooks and, since 2005, vessels with LOA ≥15 m must be equipped with a vessel monitoring system (VMS). Logbooks contain mandatory information about: (a) vessel identification; (b) gear type and mesh size used during the trip; (c) dates of operation; (d) fishing location identified with respect to ICES statistical rectangle; and (e) quantities retained on board (Pedersen et al., 2009). VMS involves the presence, on each fishing vessel, of an automatic transmitting station (the so-called *blue box*), which periodically sends information about vessel position, speed, and heading. VMS is an ideal for research purposes because it provides observations characterized by high spatial resolution, and it is independent from fishers declarations in the logbooks (Bertrand et al., 2007; Deng et al., 2005).

Although the literature about VMS is rapidly increasing, the use of VMS as a scientific tool is still limited. This is at least partially due to the fact that the VMS data are usually characterized by a relatively low frequency of signals (generally about 2 h), which means that some spatial model (e.g. Deng et al., 2005; Mills et al., 2007; Vermard et al., 2010; Walker and Bez, 2010) needs to be used to fully utilize the data. However, this limitation may be reduced since Russo et al. (2011) proposed a method, called "modified Catmull-Rom" (CRm), to interpolate fishing tracks which provides a realistic reconstruction of fishing vessel behaviour for any métier. Although the use of this method results in the addition of information to the original data (since position, speed and heading for interpolated points are not originally present in the dataset), CRm seems to be able to capture the often complex pattern of vessels course, with accuracy higher than other methods (Russo et al., 2011).

New models and procedures are now available to assess métiers associated with VMS tracks through the use of logbook data (Bastardie et al., 2010), but these are still affected by the variable quality of information recorded in logbooks (e.g. Bastardie et al., 2010; Gerritsen & Lordan, 2011; Marchal, 2008). In addition, logbooks may not reflect fine-scale tactical features (the so-called skipper effect), which can significantly contribute to the fishing strategy (Marchal, 2008). In his milestone study, Marchal (2008) extensively investigated the linkage between métier and catch profile provided by logbooks and concluded that the collection of information on fishing intention before a fishing trip is better than inferring it from logbook data. Therefore, there is a need to explore alternative strategies that do not rely on logbook data. Recently, Chang (2011) and Bastardie et al. (2010) stressed that VMS could provide reliable information about vessel operation patterns or vessel behaviour at a small spatial scales. The present study starts from the idea that if fishing intention is a good descriptor of fishing activity, the real behavioural pattern expressed by a fishing vessel, as captured by VMS, should be even a better one. Previous attempts in extracting information about fishing activity from VMS data are mainly concerned with the distinction between fishing and non-fishing points using analysis of speed profile (e.g. Mills et al., 2007; Piet et al., 2007; Vermard et al., 2010), but these studies defined *a priori* fleet segments and métier units.

This study uses artificial neural networks (ANNs; Lek et al., 1996; Scardi, 2001) to assign fishing effort, resolved at the single trip scale, to métier. ANNs represent a family of modelling techniques devised to mimic the functioning of the human brain (Stern, 1996) in order to analyse large and complex data sets characterized by non-linear relationships among variables, internal redundancy and noise (Lek et al., 1996). ANNs are a data-driven approach (no explicit relationship is assumed to exist between input and output data), which gathers information from the input data and produces output by applying learned training algorithms which are exclusively based on maximizing the performance of predictions. The type of ANN used in this study is represented by the multilayer perceptron network (MPN), since it is the simplest and most widely used ANN architecture to pursue classification issues (Haykin, 1999; Lek et al., 1996; Scardi, 2001). This tool has already been successfully used on VMS data to distinguish between fishing or not fishing points of single tracks (Bertrand et al., 2008; Joo et al., 2011), on daily landing data to infer the fishing tactics employed (Palmer et al., 2009). The advantages of ANN–MPNs comprise: (1) neither known nor assumed probability distribution functions is needed for the data, and (2) simplicity of construction and training. Our ANN is designed to identify, for a given fishing track, the associated métier among several possibilities. The variables used in the study are divided in two groups: the first describing vessel behaviour, and the second containing information about the licenses of the different vessels. The ANN is trained, validated and tested on a set of fishing tracks for which métiers and logbook information are documented by at-sea-observers.

In the next sections, we briefly describe the characteristics of the Italian fisheries, the data used in this study, and the ANN architecture and training algorithm. After training, we compare the classification provided by the neural network output with on-board observations. Then we describe and perform a series of sensitivity analyses. Finally, we discuss on advantages and limits of this approach, and on how this ANN can be implemented for present and future issues of the DCF.

2. Materials and methods

2.1. The Italian fleet structure and activity

Mediterranean fisheries are different from those in most other EU fishing regions, primarily due to the large variety of fishing gears, the multi-species targets, and the large number of species accepted by the markets (Sabatella, 2010). Although mixed fisheries are present in other EU sea regions, the Mediterranean fleets have more métiers (Sabatella, 2010). In 2007, the Italian fleet consisted of 13,804 vessels, 3500 of which were commercial vessels (the rest were recreational vessels). 59% of these 3500 vessels is equipped with VMS. Moreover, the majority (over 60%) of the 2076 Italian vessels equipped with VMS are licensed to use more than a single gear type (Fig. 1a) (Sabatella, 2010). The Italian fishing fleet is characterized by vessels that perform their activities mostly within a distance of 12 nm from the coast. Fishing trips last about 12 h on average and only one gear is used during each trip, although gear could change within a season or a year (Sabatella, 2010). The only exception is the subset of the fleet operating in the Sicily channel (Geographic Sub Area 16), where fishing trips can be up to 2 weeks long, but are characterized by the use of a single gear on any given trip. The list of licensed gears for each vessel is under the control of the central national authority (the Italian Ministry for Agriculture, Food and Forestry) through the Vessel Register (VR). The gear type currently recorded in the VR corresponds to métiers of level 4 (Table 1) of the EC classification (ICES, 2009b). In the EU DCF, métiers of level 4 are progressively disaggregated through level 5 and 6 (EC, 2008a,b) with respect to target assemblage and mesh size and other selective devices. Table 1 also shows that, in many cases, a license of level 4 corresponds to more than 3 activities at level 6 (as in the case of bottom otter trawl). Consequently, when the number of possible fishing activities is computed as the product of the number of licenses and the number of corresponding métiers at level 6, it appears that vessels with multiple licenses at level 4 can perform up to 13 fishing activities (Fig. 1b).

2.2. Rationale of the approach

The 15 métiers identifying the activity of the Italian professional fishing fleet can be technically grouped in three categories (ICES, 2009b): (1) **Passive** gears that are placed on the seabed and do not move until lifted by the fishing vessel; (2) **Towed** gears that are towed across the seabed; and (3) **Mobile** gears involve movement of the fishing vessel during deployment, but are not actively towed. The tracks belonging to different categories are generally characterized by important differences in terms of course and operation speed during fisheries operations (Flewwelling et al., 2002), as documented by the high frequency (20' between pings) tracks used in Russo et al. (2011). This is reflected, in turn, by the spatial trajectories of vessels (Fig. 2a–c). VMS data provide information about vessel position, speed, and heading, while sea depth can be inferred by vessel position. These quantities can be used to describe VMS tracks. The patterns of these variables change depending on fishing activity (Fig. 2d–f) and could be used to identify the fishing activity associated with a given track of a particular fishing vessel. It should be stressed, however, that this argument is valid only if the patterns for these variables are computed using high frequency tracks (either interpolated or originally provided by opportunely set blue boxes). Here we present a method to determine the métier of a given fishing track using (1) variables obtained by transforming VMS data into frequencies for a series of classes describing the distributions of speed, sea depth and changes in vessel direction/orientation (as measured by vessel heading) and (2) fishing license information (the list of authorized gears for the vessel) from the VR. The unique EU vessel identification number was used to join these two groups of data for each vessel. The obtained dataset is used as input to train an MPN. After training, the MPN is tested for its ability to identify the métier on an independent dataset. MPN, as member of the large ANN family, was chosen because the predictor variables (obtained using VMS and from the VR) are linked to the outcome variable (the métier) via a non-linear relationship. Thus, methods based on the assumptions of a linear model cannot fit the data well.

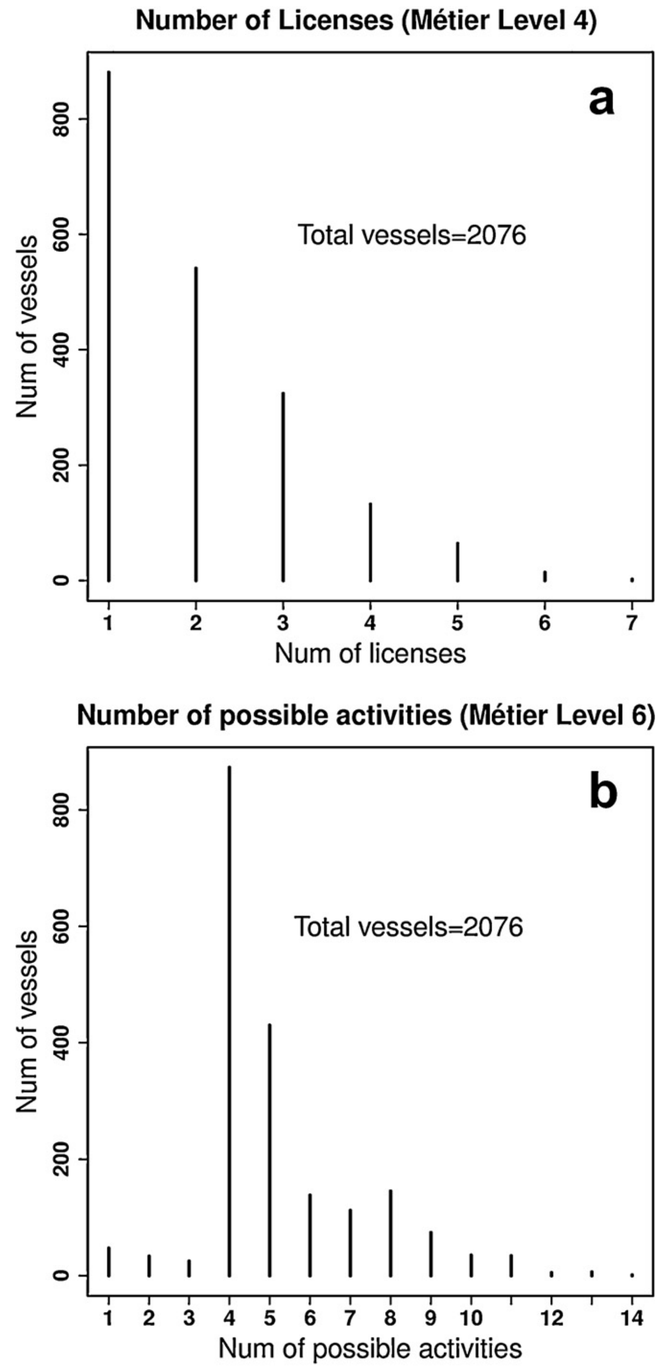

Fig. 1. Distributions of (a) number of licenses (métiers of level 4) and (b) number of possible activities (level 6 métiers) for the Italian professional fishing fleet (N = 2076 vessels).

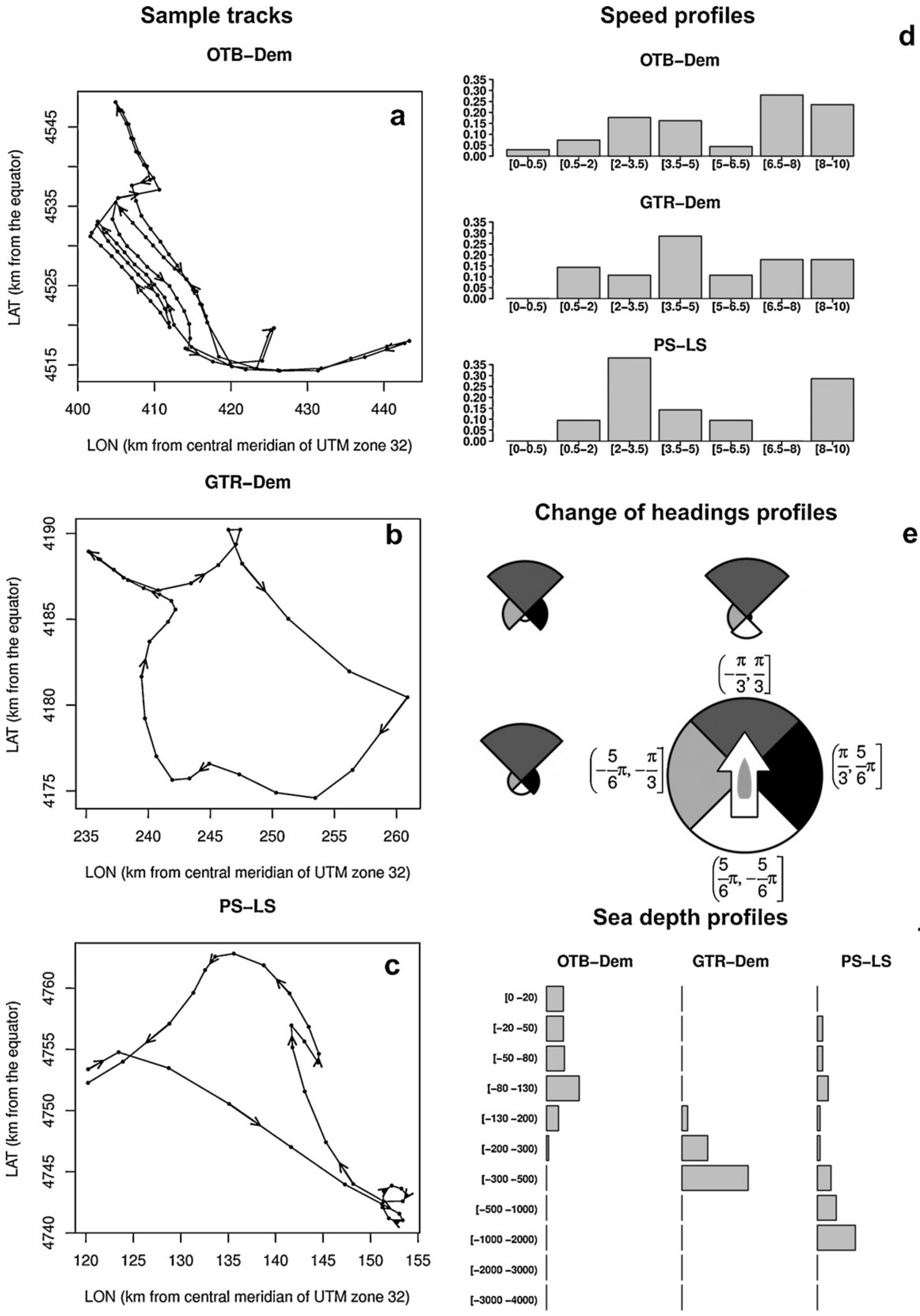

Fig. 2. Graphical representations of three sample tracks (one for each group of métiers, as assessed by on-board observers): (a) bottom otter trawl for demersal species (OTB-Dem), a type of Towed gear; (b) trammel net for demersal species (GTR-Dem), a type of Passive gear; and (c) purse seine for large pelagic fish (PS), a type of Mobile gear. The path followed by the vessel is indicated by the arrows, while points identify VMS positions. For the same sample tracks, the different groups of variables summarizing profiles of vessel speed (e), change of heading (f) and sea depth are shown. Each variable of each series is measured as a value of frequency in the range [0,1].

Table 1
Characteristics of the métiers identifying the activity of the Italian fishing fleet equipped by VMS. The table shows the different levels of classification and aggregation: that of level 4 corresponds to the licenses of Italian Vessel Register, while levels 5 and 6 (which completely match) identify the finest scale of disaggregation used within the data collection framework.

Category	Level 3 Gear group	Level 4 Gear type	Level 5 Target assemblage	Level 6 Mesh size and other selective devices	Short code
Towed	Dregdes	Boat dredge [DRB]	Molluscs	–	DRB-Moll
	Bottom trawls	Bottom otter trawl [OTB]	Demersal species	≥40	OTB-Dem
			Deep water species	≥40	OTB-Deep
			Mixed demersal species and deep water species	≥40	OTB-Mix
		Bottom pair trawl [PTB]	Demersal species	≥40	PTB-Dem
		Beam trawl [TBB]	Demersal species	≥40	TBB-Dem
	Pelagic trawls	Midwater otter trawl [OTM]	Mixed demersal and pelagic species	≥20	OTM-Mix
		Pelagic pair trawl [PTM]	Small pelagic fish	≥20	PTM-SP
Mobile	Longlines	Drifting longlines [LLD]	Large pelagic fish	[a]	LLD-LP
		Set longlines [LLS]	Demersal fish	[a]	LLS-Dem
	Surrounding nets	Lampara nets [LA]	Small and large pelagic fish	≥14	LA-SL
		Purse seine [PS]	Small pelagic fish	≥14	PS-Sm
			Large pelagic fish	≥14	PS-La
Passive	Nets	Trammel net [GTR]	Demersal species	≥16	GTR-Dem
		Set gillnet [GNS]	Demersal species	≥16	GNS-Dem

Source: Modified from ICES (2009b).
[a] Not spelled out in DCR but defined with reference to relevant EU regulation(s).

In our analysis, each fishing track was fully inspected (starting from the departure by harbour to the return), without distinguishing between fishing and non-fishing points. In this way, tracks enter into the analysis without assumption or information about métier, which is instead used to evaluate the MPN prediction on data not used for training.

The relative frequency distribution of vessel speed was described by a set of seven classes (Fig. 2d). The values of speed larger than 10 km h^{-1} were ignored since any fishing activity generally occurs below this speed and vessels travelling at or above this speed are assumed to be steaming (e.g. Gerritsen & Lordan, 2011; Mills et al., 2007; Palmer et al., 2009). The highest values of speed could simply reflect the structural characteristics of a given vessel, since maximum steaming speed is related to vessel length (Prado, 1990). Computation of frequencies with respect to four classes are applied to the changes of heading between two successive points (Fig. 2e), while 11 classes are used to describe sea depths for each track (Fig. 2f). Classes for sea depth were fixed with reference to the limits defined by the ICES (2009a). The three groups of variables obtained (classes for speed, heading change, and sea depth) sum to 22 variables but, to avoid computation problems related to the perfect colinearity within each group (Carpio and Hermosilla, 2002), the last class of each group was not included in the analysis. The final number of variables obtained by VMS was then 19 (Table 2). Two other variables were computed for each track: (1) trip duration (TD), since different métiers are likely to be characterized by different durations (Iriondo et al., 2010); and (2) maximum distance from the coast (MDC – computed as the maximum value of distances between the coast line and each of the points composing the track) since the activities related to each métier are likely to occur at different distance from the coast (Katsanevakis et al., 2010).

The final list of these 33 variables (Table 2 – 12 from VR, 19 from VMS, TD, and MDC), assumed to represent the vessel track, is submitted to the MPN which learns how to connect patterns of these variables to the list of métiers. The MPN uses these 33 variables as input, and produces 15 values of probability, one for each of the métiers. Given that data to be submitted to the MPN have to be standardized (Haykin, 1999), the last two parameters (TD and MDC, respectively) were normalized into the [0,1] interval, whereas binary data for licenses and continuous data for class frequencies did not require normalization.

2.3. Dataset used and pre-processing of data

The data used in this study, provided by the Italian Ministry for Agriculture, Food and Forestry, originated from a random subsample of the Italian fishing fleet from which information about fishing activity was collected and validated by on-board scientific observers. Much more data than provided by the logbooks were collected, including: capture composition (including discards), time of fishing, and fishing intention at the start of fishing trip. This allowed a real assessment (RA) of the level 6 métier that was then used to train and test the ANN. The final dataset contains 15,000 tracks, that is 1000 for each of the 15 métiers listed in Table 1. This large dataset was obtained by merging available data for the years 2006–2009. Only trips that used one gear were included, so that multi-gear tracks were not represented in the dataset. Also, each vessel in the dataset was characterized by at least three different licenses at level 4, as reported in the VR, for a mean number of 5.7 possible level 6 métiers for each vessel. This implies that, with respect to the information represented in Fig. 1, we removed the left part of the distribution. This constraint was imposed to avoid a situation in which the identification of métier for a given track is extremely simplified by the fact that the respective vessel is licensed and equipped for only one or two gears.

Considering that the blue boxes of the Italian fleet were generally set to send signals at a frequency of 2 h, each track needed to be temporally (and then spatially) interpolated to provide a detailed description of vessel path and behaviour (ICES, 2009a). We used the method of Russo et al. (2011) to estimate high-frequency tracks with positional data at 20 min intervals. Apart from spatial position, vessel speed and heading are also estimated for the interpolated points, using a spline algorithm in which the blue boxes signals (provided at 2-h intervals) are used to compute the tangent (see Russo et al., 2011 for an exhaustive description of the method).

2.4. Artificial neural network

An MPN (Fig. 3) consists of at least three layers of neurons (also called units or nodes). The neurons of each layer process information in parallel, while information flows in a unidirectional way from the first (input) layer to last (output) layer through a second (hidden) layer. The input layer contains as many neurons as

Table 2
List of the variables submitted to MPN: 33 of 35 were included.

Variable	Type	Unit/notes	
DRB OTB PTB TBB OTM PTM LLD LLS GTR GNS LA PS	License	Presence or absence (binary variables)	
[0–0.5) [0.5–2) [2–3.5) [3.5–5) [5–6.5) [6.5–8) [8–10)[a]	Classes for vessel speed	km h^{-1}	Cumulate sum is equal to 1
$(-\pi/3,\pi/3]$ $(\pi/3,(5/6)\pi]$ $((5/6)\pi,(-5/6)\pi]$ $((-5/6)\pi,-\pi/3]$[a]	Classes for change of vessel heading	Radian	Cumulate sum is equal to 1
[0,−20) [−20,−50) [−50,−80) [−80,−130) [−130,−200) [−200,−300) [−300,−500) [−500,−1000) [−1000,−2000) [−2000,−3000) [−3000,−4000)[a]	Classes for sea depth	Meters (m) from the sea surface	Cumulative sum is equal to 1
TD	Trip duration	Seconds from the first to the last VMS signals of each track	Normalized into the [0,1] interval after computation
MDC	Maximum distance from the coast	Meters (m) between the coast line and the farthest VMS signals of each track	

[a] Marks classes which have not been submitted to MPN in order to avoid computational problems connected to collinearity.

independent variables or descriptors used to predict the dependent variables, which in turn constitute the output layer. An MPN could contain more than a single hidden layer, so that the total number of layers could be larger than 3. However, the MPN used in this study has a single hidden neural layer, since Cybenko (1989) and Hornik et al. (1989) demonstrated that this structure is able to approximate any complex non-linear function with great accuracy. The structure of the MPN presented here is more complex than those used in other recent studies (Joo et al., 2011; Palmer et al., 2009), because both the number of variables entering the MPN (representing the neurons of the input layer) and that of métiers to be predicted (representing the output layer) are quite high (33 and 15, respectively, see Fig. 3), However, it is conceptually the same: neurons of a given layer are linked to the neurons of the next layer by activation functions. The hidden layer neurons compute a weighted sum of the input variables through a first activation function, a sigmoid function in this case, which is the most common choice because it is not linear and characterized by a very easy to compute derivative (Haykin, 1999; Lek et al., 1996); then they send a result to the output neurons through a second activation function. A multiple logistic model, also called a softmax activation function (Bridle, 1990), was used in this case. The softmax function scales the neural network outputs so that their values (one for each of the 15 output neurons, corresponding to the possible métiers) range between 0 and 1 and their sum is 1, so allowing probabilistic interpretations of the results. Thus, the output produced for each track can be regarded as a series of probabilities to belong to each one of the 15 level 6 métiers. The application of MPN consists of a *training* phase and a *test* phase. The training (also called learning process) is based on the use of two datasets (the "training" and "validation" datasets, respectively) to adjust the weights of the hidden layer neurons to minimize the error function (in our case, the mean squared error – MSE) between the observed and predicted values. In the test phase, the MPN should be able to show high performances (i.e. >80%) in classifying patterns from a "test dataset", which is different from both the training and the validation dataset. As pointed out, e.g. by Joo et al. (2011), the use of two datasets during training will limit the risk of "overfitting". This phenomenon occurs when an MPN adjusts to the training dataset so much that its ability to generalize and correctly classify new patterns (that is patterns not used during training) is partially lost. However, this risk can be efficiently avoided using an early-stopping strategy based on the use of both training and validation datasets (Caruana et al., 2000). While the training dataset is effectively used to minimize MSE, the validation dataset is used to control the performance of the MPN during the learning process: the training performs better until both MSE on training and validation dataset decrease, whereas overfitting occurs when the MSE for training dataset continues to become smaller, but that for validation dataset begins to increase. Thus, the early stop strategy is based on the rule that when a systematic increase of

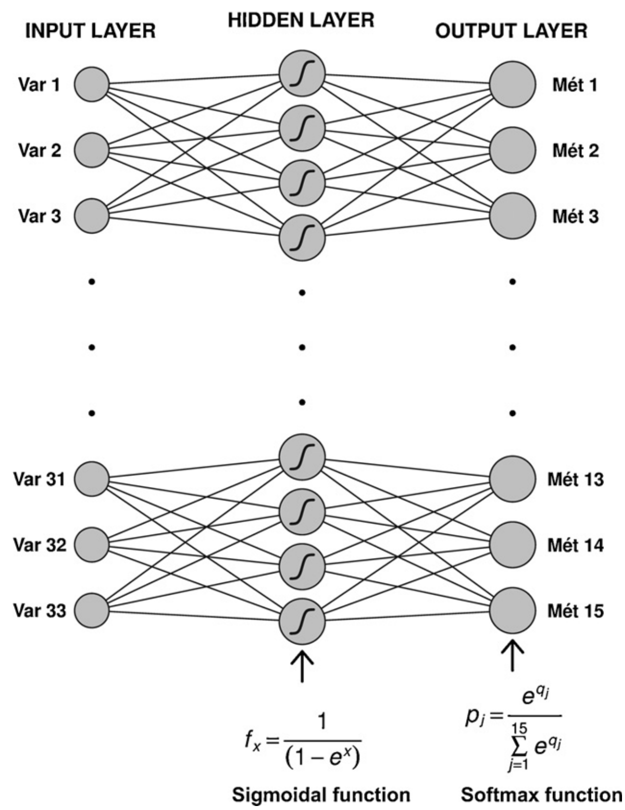

Fig. 3. Architecture of the MPN (modified from Palmer et al., 2009). In the expression of Sigmoid function, x is the value of an input variable in the first layer of MPN. In the expression of Softmax function, p is the probability, for a given track, of belong to the jth métier, while q is the value computed by the neuron of the hidden layer for the jth métier.

the validation errors occurs, the training is stopped (Caruana et al., 2000).

The dataset used in this study was partitioned in three subsets: (1) the training dataset which contained a random subset of 600 tracks for each métier (60% of the total set); (2) the validation dataset which contained 200 tracks per métier (20% of the total set); and (3) the test dataset which contained the remaining 200 tracks per métier. Thus, all the three datasets contain data on all 15 métiers, in the same proportion. Moreover, the three datasets are complementary and do not overlap. The relative sizes of the training, validation and test subsets were initially fixed to values indicated as optimal by recent studies (Joo et al., 2011) since the literature is not conclusive about an "ideal" size for each partition (Lek and Guégan, 2000; Sarle, 1995; Zhang et al., 1998). Joo et al. (2011), who effectively used a MPN for identifying fishing set positions from VMS data, demonstrated that: (1) small changes in the proportions of the partitions do not significantly affect the results; and (2) the overall size of the whole dataset is more important than the relative size of partitions. This latter definitively have marginal effects on large datasets (that is including thousands of observations, as in the case of the present study).

2.5. Sensitivity analysis

Two critical aspects arise when an MPN is applied: (1) how many hidden neurons are required (2) what is the role and relative importance of each input variable. The number of neurons in the hidden layer is an important factor, since it affects the performance of the MPN. In this study, the best architecture was selected on the basis of a heuristic test: we trained neural networks with a wide range of numbers of neurons in the hidden layer (ranging from 5 to 65; see Scardi, 2001) and then selected the one which provided the best results (highest mean percentage of correct classification on the test dataset). For each value of the number of neurons in the hidden layer, the neural network was trained and tested 10 times by randomly selected subsets for training, validation and test datasets. Point 2 was addressed for the MPN with the optimal number of neurons in the hidden layer. In a trained MPN, each input variable is associated with a set of weights (one for each neuron in the hidden layer). The relative importance of input variables is intuitively linked to the values of weights for the trained MPN (Haykin, 1999). In fact, as the weights in the hidden neurons represent a series of coefficients for an input variable, the effect of a change in the value of an input variable is closely linked to the value of its corresponding weights in the hidden layer: the larger the values, the larger is the effect of a change to the value in that input variable. It follows that, if the mean weights for two input variables are very different (i.e. one is more than 10 times the other), the relative contribution, and then importance, of the two variables is very different. Thus, large weights (i.e. ~100 in absolute value) can impair MPN generalization by lowering the reliability of the output (Geman et al., 1992). For these reasons, an MPN is generally considered "well-trained" if the weights for the hidden layer are homogeneous (Haykin, 1999).

The patterns of weights for each variable in the hidden layer were graphically explored to evaluate homogeneity and variation. We then adapted the procedure proposed by Garson (1991) to assess the influence of each input variable in the trained MPN. Considering a well trained MPN with an $m \times n \times o$ network architecture (i.e. m input nodes, n hidden nodes, and o output node), the procedure for calculating the relative importance (RI) of input variables is as follows:

1. Arrange a matrix M $(o \times n)$ for the interconnection weights between the n neurons in the hidden layer and the o neurons in the output layer;
2. Arrange a matrix W $(m \times n)$ for the interconnection weights between the m neurons in the input layer and the n neurons in the hidden layer;
3. Calculate the row vector

$$R = \sum_{j=1}^{o} MW_{i,j}^{T}$$

4. Calculate the relative importance, RI_i (in percentile), of input variable i, as given by

$$R = \frac{|r_i|}{\sum_{i=1}^{m_j} |r_i|} \times 100, \quad i = 1 \sim m$$

The relative importance, RI (ranging between 0 and 100 with a cumulative sum of 100), provides an index for evaluating the contribution or influence of each input variable to the output variables. In this way, input variables which contribute marginally could be identified and discarded. However, the effect of the removal of some input variables should be validated *a posteriori* by comparing the MSE for the MPN with and without the input variable that is a candidate to be discarded. This is because the RI index only evaluates relationships of first order among input variables. On the basis of this procedure, the relative importance of the 33 input variables was evaluated. Since 10 training processes were carried out for each value of the number of neurons in the hidden layer, we performed the sensitivity analysis for the one scoring at the median position with respect to the percent of test tracks correctly assigned.

Table 3
Confusion matrix obtained from the winner-takes-all classification of the test data subset. Percentage of correctly classified tracks was 94.4%, while the Kappa statistics was 0.949 (with $p < 1 \times 10^{-16}$). Numbers in cells are tracks.

		Classification performed by neural network															
		DRB-Moll	GNS-Dem	GTR-Dem	LA-SL	LLD-LP	LLS-Dem	OTB-Deep	OTB-Dem	OTB-Mix	OTM-Mix	PS-LP	PS-SP	PTB-Dem	PTM-SP	TBB-Dem	Total
Real assessment by on-board observators	DRB-Moll	200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	200
	GNS-Dem	0	200	0	0	0	0	0	0	0	0	0	0	0	0	0	200
	GTR-Dem	0	0	194	0	0	0	0	0	4	1	0	0	1	0	0	200
	LA-SL	0	0	0	200	0	0	0	0	0	0	0	0	0	0	0	200
	LLD-LP	1	3	3	1	165	0	4	8	1	5	5	3	1	0	0	200
	LLS-Dem	0	0	0	0	0	200	0	0	0	0	0	0	0	0	0	200
	OTB-Deep	0	0	0	0	4	0	191	0	3	2	0	0	0	0	0	200
	OTB-Dem	0	0	0	1	2	0	0	183	1	3	0	0	0	3	7	200
	OTB-Mix	0	0	0	0	1	0	4	0	186	9	0	0	0	0	0	200
	OTM-Mix	0	0	0	0	2	0	0	4	0	174	0	0	3	0	17	200
	PS-LP	0	0	0	0	1	0	0	0	0	0	193	3	0	3	0	200
	PS-SP	2	2	3	0	1	0	0	0	0	1	8	173	4	6	0	200
	PTB-Dem	0	0	0	0	0	0	0	0	0	0	0	0	198	2	0	200
	PTM-SP	0	0	0	0	0	0	0	5	0	0	5	0	8	182	0	200
	TBB-Dem	0	0	0	0	0	0	0	2	0	5	0	0	0	0	193	200
	Total	203	205	200	202	176	200	199	202	195	200	211	179	215	196	217	

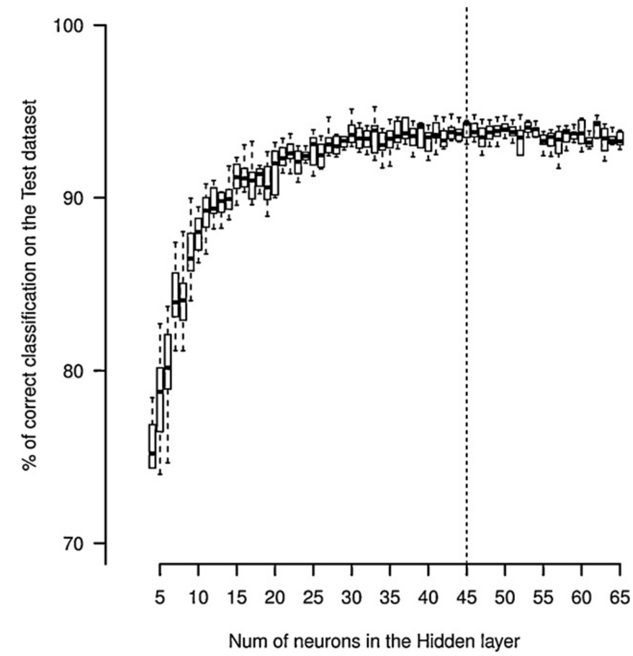

Fig. 4. Relationship between percentage of correct classification on the test dataset and number of neurons in the hidden layer.

2.6. Analysis of MPN results

Given that a softmax function was used for the neurons of the output layer, a *winner-takes-all* strategy was applied for the classification of the test tracks: each track was classified according to the highest métier membership value in the output layer, which means that the values of probability for the non-winner métiers and the hierarchy (that is the list of métiers associated with the smaller values of softmax output function) eventually arising from these data was ignored. A confusion matrix was obtained by comparing the classification provided by the neural network output and the RA by on-board observers, and this matrix was analysed by means of the Kappa statistics (K) in order to test the agreement between the two classifications and its significance (Cohen, 1960; Fleiss et al., 1969).

All the analyses and computational steps were been carried out in the R environment (R Development Core Team, 2009). The MPN training was performed by means of the adaptive versions of the gradient descent with momentum back-propagation algorithm (Haykin, 1999), implemented in the R package AMORE (Castejón et al., 2006). Constant values were set for the learning rate (0.1) and the momentum (0.5) to accelerate training while preventing the system from falling into local minima (Lek et al., 1996). Computation of the Kappa statistics was carried out by the function provided in the R package "rmac" (Fay, 2005).

3. Results

MPNs with 45 neurons in the hidden layer gave the best results in terms of percentage of correct classification ($94.0\% \pm 2.56 \times 10^{-3}$) of the tracks in the test datasets. However, MPNs with different numbers of neurons in the hidden layer provided very similar results, giving evidence of training stability (Fig. 4). In effect, the trend of the relationship between number of neurons in the hidden layer and mean percentage of correctly classified tracks (Fig. 4) were almost the same (that is a percentage of correct classification higher than 90%) for a number of neurons in the hidden layer larger than 15. It should be stressed that these results were obtained for the test dataset, which is completely different from the two (training and validation) used during the learning phase and represented new data for the trained MPN. The weights for input variables in the hidden layer of the trained MPN looked homogeneous (Fig. 5a). Although the overall range was between −20 and +20, most (80%) of the weights ranged between −5 and +5.

The sensitivity analysis showed that the variables characterized by the smallest importance were the license for PTB and MDC (Fig. 5b). However, when these variables were discarded and the MPN re-trained with the same architecture, but with 31 input variables, the resulting mean MSE was significantly larger (Student's t, $p < 0.05$). This suggested that some higher order relationship exists with the other variables, and so PTB and MDC were maintained in the list of variables submitted to the MPN in order to not reduce performance. Apart from these two special cases, the other input variables seem to be characterized by comparable importance in terms of *RI* values (Fig. 5b), all ranging between 1 and 7.

Confusion matrices based on the comparison between MPN classification and RA were obtained for both the complete series of 15 considered métiers (Table 3) and the three groups of métiers (namely, towed, mobile and passive ones, Table 4). All the tracks in the test dataset were correctly classified by the trained MPN in four cases (DRB-Moll, GNS-Dem, LA-SL and LLS-Dem). For five other cases (GTR-Dem, OTB-Deep, PS-LP, PTB-Dem and TBB-Dem), the number of misclassified tracks were less than 10 (<5%). For the remaining three cases (OTB-Dem, OTB-Mix and PTM-SP) the proportion of misclassified tracks is a bit higher (∼9%). Thus, the major part of classification errors is limited to OTM-Mix (13%), PS-SP (13.5%) and LLD-LP (17.5%). Two of these belong to Mobile gear class, while the major part of misclassified OTM-Mix tracks were assigned to OTB-Mix, which is a quite similar fishing activity (in terms of gears) focusing on the same demersal resources.

The Kappa statistics for the confusion matrix built on the complete series of 15 métiers was highly significant ($K = 0.949$, $p < 1 \times 10^{-16}$), as well as that on the grouped métiers ($K = 0.971$, $p < 1 \times 10^{-16}$) so that the overall agreement between RA and MPN outputs could be considered as "excellent", according to Landis and Koch (1977), in both cases.

4. Discussion and conclusions

One of the main challenges in fisheries research is the development of reliable methods to characterize fishing activities and to *a posteriori* assign a given unit of effort (i.e. a single fishing trip or track) to a particular category of exploitation pattern (e.g. gear used, fishing ground, target species), that is to a métier. Many studies following different approaches are available in the literature. Among these, the analysis of catch profile data provided by landings or logbooks is the commonest approach (e.g. Bastardie et al., 2010; Lewy and Vinther, 1994; Pelletier and Ferraris, 2000; Ulrich et al., 2001). Unfortunately, these methods are intrinsically limited by two inescapable drawbacks: (1) landings and logbooks are potentially biased (and not always available) sources of data, and (2) catch composition does not necessarily reflect fishing intention, the concept at the core of métier definition. Nowadays, the new possibilities opened by the rapid technical progress of VMS

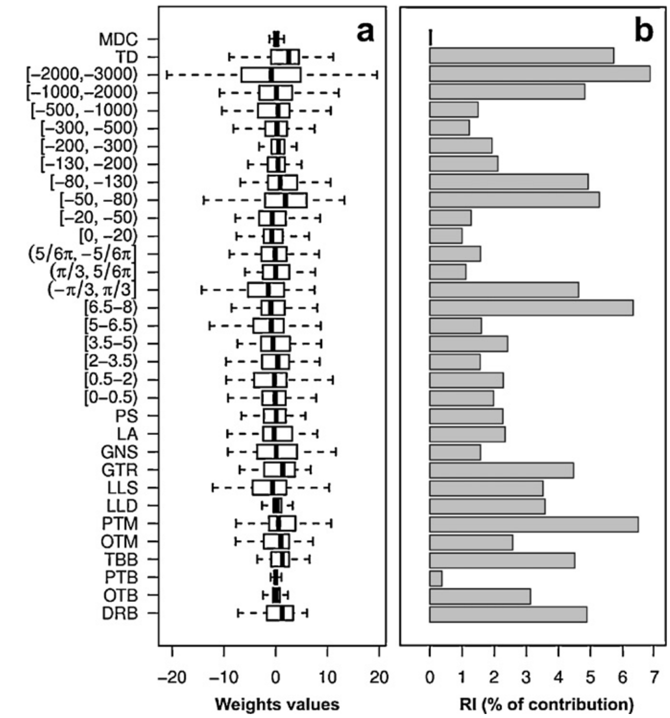

Fig. 5. (a) Boxplots of the values for the weights assigned to the each of the 33 input variables within the neurons of the hidden layer. (b) Relative importance index (*RI*) for the same variables.

in fisheries management and research is a reality (Chang, 2011). Although VMS has historically been used for vessel monitoring, in recent years it has increasingly been applied in fisheries science, as displayed by the long list of published studies. VMS data also have the advantage of being immediately available because they are collected and managed by the central authorities. In the present study, a new approach based on the analysis of VMS data by means of an artificial intelligence technique was applied and successfully tested over a large dataset representing the activity of the Italian commercial fishing fleet. The method performs well, since the best MPN architecture correctly classified around 94.0% of tracks in the test datasets. This result appears particularly good because it can be computed that, as the mean number of level 6 métier associated with each vessel in the used dataset was 5.7, the probability of correctly classifying the métier of a track by chance, when the licensed gears are known is about 17% (100/5.7). This seems to be promising for further developments and trials in different fisheries.

The present approach was carried out using interpolated VMS tracks. Although this implies a preliminary computational step in the analysis, it could be argued that: (1) the interpolation procedure does not require assumptions or information about the métier (Russo et al., 2011); and (2) if the error associated with the interpolation has some effect, which could be assessed in future studies, it probably does not impair the results since the mean percentage of correct MPN classification is very high.

Table 4
Confusion matrix obtained from the winner-takes-all classification of the test data subset, in which métiers were grouped with respect to the three typologies of towed, mobile and passive, respectively. Percentage of correctly classified tracks was 94.4%, while the Kappa statistics was 0.971, $p < 1 \times 10^{-16}$).

		MPN			
		Towed	Mobile	Passive	Total
RA	Towed	1585	36	6	1627
	Mobile	15	958	0	973
	Passive	0	11	394	405
	Total	1600	1005	400	

Apart from VMS data, this approach uses only one other source of information: that included in the VR, i.e. the list of authorized gears (corresponding to level 4 métiers), so this information drives the MPN by reducing the list of possible activities (level 6 métiers) for a given vessel. Among these possible level 6 métiers, the artificial neural network selects one on the basis of a series of variables describing the patterns of speed, change of heading and sea depth throughout the fishing vessel trip. Since the VR is managed by the same office that grants the licenses, it is regularly updated and provides a reliable source of information.

Therefore, one of the advantages of this approach over the common analysis of catch profile data provided by landings or logbooks is that it uses only two sources of data (VR and VMS) which are reliable, easy to collect, and completely independent from fishers control. In fact, the VMS signal provided by blue box can be blocked but not modified, with some exceptions documented in the literature (FAO, 1998), but never observed in the VMS dataset for the Italian fisheries, while VR is under the authority control of the State.

The idea at the basis of the approach presented in this study integrates several existing scientific methods. First it uses frequency distributions to describe the pattern of speed and change of heading have been previously used by Joo et al. (2011) and Walker and Bez (2010) (to distinguish between fishing and non fishing points along a track), by Palmer and Wigley (2009) (to identify area fished), and by Bastardie et al. (2010) (to map fishing effort). Secondly the use of ANN to analyse VMS data is not new either (e.g. Joo et al., 2011; Palmer et al., 2009). However, to our knowledge, this work represents the first attempt to provide a classification of fishing activity which is completely based on VMS and VR data.

The reason for the excellent performance shown by the MPN in identifying level 6 métiers from the transformed VMS data may be understood by looking at some real cases (Fig. 2). Vessels making different fishing activities are characterized by markedly different behaviours (Prado, 1990; Russo et al., 2011). A preliminary survey on three examples, one for each gear type (towed, mobile and passive) gave an idea of differences in speed, change of heading and sea depth (Fig. 2). The tracks described by bottom otter trawlers (Fig. 2a), that represent one of the commonest towed gears, are almost always characterized by the appearance of transects parallel to the coast since both dynamic stability and efficiency of these gears is optimized when following a bathymetric profile without abrupt changes of depth (Chun-Woo, 1995; Sainsbury, 1996). These transects correspond to fishing activity, which is the period during which the gear is towed across the seabed and the vessel moves in a straight line at an almost constant speed. As consequence, the speed spectrum of a trawler is substantially bimodal (Palmer and Wigley, 2009) and most changes in heading are small (Fig. 2f). Bottom trawlers operate along a wide bathymetric range, from the shallow shelf (deeper than 50 m) to the middle slope (down to 800–1000 m) (Palmer et al., 2009). Thus the pattern of sea depth reasonably affects the distinction between similar métiers such as OTB-Dem, OTB-Deep and OTB-Mix. The situation is different for Passive (Fig. 2b) or Mobile (Fig. 2c) gears. In both cases, the trajectories described through the space are more smooth, as reflected by the distributions of heading change (Fig. 2e): the class of largest change in heading is more dominant than that of Towed gears. This could be explained by the fact that, when a vessel is fishing with Mobile or Passive gear, it has to control its spatial position in time, even acting against drift and wind, until the gear is properly set (Russo et al., 2011; Sainsbury, 1996). In the meantime, patterns of speed and sea depth are different, supporting the existence of differences captured by the MPN. However, these discriminating differences in speed, heading change and sea depth can neither be captured by rough descriptors as central indexes (mean or median) nor by histograms computed and then compared for large groups of tracks belonging to different métiers. This because each métier can be associated with several different patterns. Consequently is difficult to distinguish between two métiers using methods based on a linear relationships among input data and métiers, as largely demonstrated in the literature for other cases of non linear relationships among independent and outcome variables (Haykin, 1999). This aspect would be graphically clear if the relevant variables were two or three. In our case, we have many variables, and this precludes an effective representation of the relationship among métiers as captured by the variables used to train the MPN.

There were negligible errors for Towed gears, whereas errors are more common for Mobile gears. This could be because Mobile gears frequently require a long search and exploring activity before gear deployment, even characterized by large movements associated with a switch among fishing ground and that, ultimately, vessel behaviour is largely conditioned by spatial distribution of fish (Bertrand et al., 2007). This could result in more complex and difficult to classify trajectories.

It should be stressed that this application was carried out under the assumption that métier does not change within a given fishing trip. Although this is the case for Italian fisheries, it might not be the case generally. This is a limitation of the method which could be theoretically overcome by partitioning tracks in homogeneous portions. Otherwise, the application should be restricted to mono gear fishing trips. Apart from this aspect, the method presented here should be tested in different contexts represented by other fleets, while its final aim is not to replace approaches based on catch analysis, but to be used in conjunction with other approaches to characterize fishing activity at the finest scale by coupling VMS and logbook data (e.g. Bastardie et al., 2010).

The trained MPN presented can be directly implemented as tool to monitor the Italian fishing fleet, and particularly to assign fishing effort, resolved at single trip scale, to specific métiers. If technical innovations or normative rules will not determine relevant changes in vessel behaviour, this method could be directly applied to future data. A re-training of the MPN could be planned if the data set increases substantially (i.e. it at least doubles) or if changes in vessel behaviour occur. In the latter case, additional data should be collected to take into account for future development of fishing strategies.

MPN, along with other member of the ANN family, has the disadvantages of requiring large amounts of data to be trained and are time consuming (that is the learning process is slow) compared to classification trees or traditional clustering procedures (Liu et al., 1996; Parlos et al., 1994). The availability of enough data to train an MPN, which could even account for future changes in fishing strategy as well as in fleet and gear structure, could be faced by routinely integrating data collected through: (1) biological sampling carried out within the DCF and (2) scientific surveys, e.g. the International Bottom Trawl Survey in the Mediterranean (MEDITS) (http://www.sibm.it/SITO%20MEDITS/principaleprogramme.htm). Both these activities are carried out yearly and are done by scientific observers boarded on chartered vessels belonging to the Italian fleet and equipped by VMS. Observers collect detailed information about fishing activity, for all the métiers, while the vessels perform normal activity. In this way, the collected data are representative of the evolution of both fishing strategy and gear equipment for the different métiers.

The black-box nature of the ANNs does not allow an explicit formalization of the process leading to the observed results. In other words, the computational process connecting inputs to outputs in a trained MPN architecture cannot be translated in a series of mathematical rules, such as those of a classification tree. Although this could seem to be a drawback, it is probably the reason for the success of this family of tools in modelling complex ecological relationships (Lek and Guégan, 2000), and particularly when compared to classical techniques (e.g. Scardi and Harding, 1999). We observed a general homogeneity in the relative importance of the input variables used, since none reached 10% and almost none was smaller than 1.5% (Fig. 5). The only two exceptions, represented by PTB and MDC, though contribute to the prediction since their removal lowered MPN performance.

In conclusion, disaggregating fishing activity represents an challenge to the scientific community, since VMS and logbook technology is still developing and no direct way exists to assign métiers with certainty. A step towards solving this problem could be the implementation of an electronic recording and reporting system of fisheries data (catch, landing, sales and transhipment) via so-called electronic logbooks (EC, 2008c). However, VMS will represent the only tool able to provide independent estimates of fishing intensity (Chang, 2011) and implementations of sound approaches in extracting and processing information from this source will likely contribute to improve quantitative evaluation of fishing effort at the finest scale, following the directive of the European Community.

Acknowledgements

Authors thank Dr. Fabio Fiorentino (Italian National Research Council – CNR) who provided some insightful ideas to the rationale of the method, Dr. Paolo Carpentieri (Department of Environmental Biology, "La Sapienza" University of Rome), for his contribution about characteristics of Italian fleet, and Prof. Michele Scardi (Department of Biology – "Tor Vergata" University of Rome), who provided much advice regarding ANNs. The revised manuscript has been edited and substantially improved by Dr. H.D. Gerritsen (Marine Institute, Galway).

References

Bastardie, F., Nielsen, J.R., Ulrich, C., Egekvist, J., Degel, H., 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fish. Res. 106, 41–53.

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., Gerlotto, F., 2007. Scale-invariant movements of fisherm: the same foraging strategy as natural predators. Ecol. Appl. 17 (2), 331–337, doi:10.1890/06-0303.

Bertrand, S., Diaz, E., Lengaigne, M., 2008. Patterns in the spatial distribution of Peruvian anchovy (Engraulis ringens) revealed by spatially explicit data. Prog. Oceanogr. 79, 379–389, doi:10.1016/j.pocean.2008.10.009.

Biseau, A., 1998. Definition of a directed fishing effort in a mixedspecies trawl fishery, and its impact on stock assessment. Aquat. Living Resour. 11, 119–136.

Bridle, J.S., 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Fogleman Soulie, F., Herault, J. (Eds.), Neurocomputing: Algorithms, Architectures and Applications. Springer-Verlag, Berlin, pp. 227–236.

Carpio, K.J.E., Hermosilla, A.Y., 2002. On multicollinearity and artificial neural networks. Compl. Int. 10, Paper ID: hermos01, URL: http://www.complexity.org.au/ci/vol10/hermos01/.

Caruana, R., Lawrence, S., Giles, C.L., 2000. Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. In: Proc. NIPS, pp. 402–408.

Castejón, M., Ordieres, J., González, A., 2006. AMORE: A MORE Flexible Neural Network Package. , http://finzi.psych.upenn.edu/R/library/AMORE/html/00Index.html.

Castro, J., Punzón, A., Pierce, G.J., Marín, M., Abad, E., 2010. Identification of métiers of the Northern Spanish coastal bottom pair trawl fleet by using the partitioning method CLARA. Fish. Res. 102, 84–190, doi:10.1016/j.fishres.2009.11.011.

Chang, S.K., 2011. Application of a vessel monitoring system to advance sustainable fisheries management—benefits received in Taiwan. Mar. Policy 35, 116–121.

Chun-Woo, L., 1995. Depth control of a midwater trawl gear using fuzzy logic. Fish. Res. 24, 311–320.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46, doi:10.1177/001316446002000104.

Colloca, F., Cardinale, M., Belluscio, A., Ardizzone, G., 2003. Pattern of distribution and diversity of demersal assemblages in the central Mediterranean sea. Estuar. Coast. Shelf Sci. 56, 469–480, doi:10.1016/S0272-7714(02)00196-8.

COM, 2008. Commission of the European Communities. Communication from the Commission to the Council and the European Parliament. The Role of the CFP in Implementing an Ecosystem Approach to Marine Management. COM, 187 final.

Cybenko, G., 1989. Approximation by superimpositions of a sigmoidal function. Math. Control Signal. 2, 303–314, doi:10.1007/BF02134016.

Davie, S., Lordan, C., 2009. Using a multivariate approach to define Irish métiers in the Irish Sea. Ir. Fish. Investig. 21, 49.

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., Die, D., 2005. Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. Can. J. Fish. Aquat. Sci. 62 (3), 611–622, doi:10.1139/F04-219.

EC, 2008a. Council Regulation (EC) No. 199/2008 concerning the establishment of a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy. Off. J. Eur. Union L60, 1–12.

EC, 2008b. Commission Decision No. 949 of 6 November 2008 adopting a multiannual Community programm pursuant to Council Regulation (EC) No. 199/2008 establishing a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy. Off. J. Eur. Union L346, 37–88.

EC, 2008c. Commission Regulation No. 1077 of 3 November 2008 laying down detailed rules for the implementation of Council Regulation (EC) No. 1966/2006 on electronic recording and reporting of fishing activities and on means of remote sensing and repealing Regulation (EC) No. 1566/2007. Off. J. Eur. Union L295, 1–21.

FAO, 1998. Fishing Operations. 1. Vessel Monitoring Systems. Fisheries Dep., Fishing Technology Series, Rome, Italy.

FAO, 2008. Fisheries Technical Guidelines for Responsible Fisheries. No. 4(Suppl. 2), Add. 1, Rome, 78p.

Fay, M.P., 2005. Random marginal agreement coefficients: rethinking the adjustment for chance in agreement coefficients. Biostatistics 6, 171–180.

Fleiss, J.L., Cohen, J., Everitt, B.S., 1969. Large sample standard errors of kappa and weighted kappa. Psychol. Bull. 72, 323–327.

Flewwelling, P., Cullinan, C., Balton, D., Sautter, R.P., Reynolds, J.E., 2002. Recent trends in monitoring, control and surveillance systems for capture fisheries. In: FAO Fisheries Technical Paper No. 415. FAO, Rome, p. 200.

Garson, G.D., 1991. Interpreting neural-network connection weights. AI Expert 6 (4), 47–51.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Comput. 4, 1–58.

Gerritsen, H., Lordan, C., 2011. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. ICES J. Mar. Sci. 68, 245–252, doi:10.1093/icesjms/fsq137.

Haykin, S., 1999. Neural Networks—A Comprehensive Foundation, second ed. Prentice Hall, New Jersey, ISBN: 0-13-273350-1.

Herrero, I., 2004. Different approaches to efficiency analysis: an application to the Spanish trawl fleet operating in Moroccan waters. Eur. J. Operat. Res. 167 (1), 257–271.

Hilborn, R., 1985. Fleet dynamics and individual variation: why some people catch more fish than others. Can. J. Fish. Aquat. Sci. 42, 2–13.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

ICES, 2009a. Report of the Working Group on the Ecosystem Effects of Fishing Activities (WGECO) , Copenhagen, Denmark, 15–21 April. ICES CM 2009/ACOM:20, p. 190.

ICES, 2009b. Report of the 3rd Meeting of the Mediterranean Planning Group for Methodological Development (PGMed) , Montpellier, 2nd–6th March, p. 53.

Iriondo, A., Prellezo, R., Santurtún, M., García, D., Quincoces, I., Mugerza, E., 2010. A multivariate approach for métier definition: a case study of Basque Country trawlers. Rev. Invest. Mar. 17 (6), 140–148.

Jabeur, C., Gobert, B., Missaoui, H., 2000. Typologie de la flottille de peche cotier dans le golfe de Gabes (Tunisie). Aquat. Living Resour. 13, 421–428.

Joo, R., Bertrand, S., Chaigneau, A., Niquen, M., 2011. Optimization of an artificial neural network for identifying fishing set positions from VMS data: an example from the Peruvian anchovy purse seine fishery. Ecol. Model. 222, 1048–1059.

Katsanevakis, S., Maravelias, C.D., Kell, L.T., 2010. Landings profiles and potential métiers in Greek set longliners. ICES J. Mar. Sci. 67, 646–656.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174, doi:10.2307/2529310.

Laurec, A., Biseau, A., Charuau, A., 1991. Modelling technical interactions. ICES Mar. Sci. 193, 225–236.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.

Lek, S., Guégan, J.F., 2000. Artificial Neuronal Networks: Application to Ecology and Evolution. Springer, New York, ISSN-10: 3540669213.

Lewy, P., Vinther, M., 1994. Identification of Danish North Sea trawl fisheries. ICES J. Mar. Sci. 51, 263–272.

Liu, K., Subbarayan, S., Shoults, R.R., Manry, M.T., Kwan, C., Lewis, F.I., Naccarino, J., 1996. Comparison of very short-term load forecasting techniques. IEEE T. Power Syst. 11 (2), 877–882.

Marchal, P., 2008. A comparative analysis of métiers and catch profiles for some French demersal and pelagic fleets. ICES J. Mar. Sci. 65, 674–686.

Marchal, P., Horwood, J., 1996. Long-term targets for the Celtic Sea mixed-species multi-métiers fisheries. Aquat. Living Resour. 9, 81–94.

Mills, C.M., Townsend, S.E., Jennings, S., Eastwood, P.D., Houghton, C.A., 2007. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. ICES J. Mar. Sci 64, 248–255, doi:10.1093/icesjms/fsl026.

Moranta, J., Quetglas, A., Massutı, E., Guijarro, B., Ordines, F., Valls, M., 2008. Research trends on demersal fisheries oceanography in the Mediterranean. In: Mertens, L.P. (Ed.), Biological Oceanography Research Trends. Nova Science Publishers, Inc., New York, pp. 9–65.

Palmer, M.C., Wigley, S.E., 2009. Using positional data from vessel monitoring systems to validate the logbook-reported area fished and the stock allocation of commercial fisheries landings. N. A. J. Fish. Manage. 29, 928–942, doi:10.1577/M08-135.1.

Palmer, M., Quetglas, A., Guijarro, B., Moranta, J., Ordines, F., Massut, E., 2009. Performance of artificial neural networks and discriminant analysis in predicting fishing tactics from multispecific fisheries. Can. J. Fish. Aquat. Sci. 66, 224–237, doi:10.1139/F08-208.

Parlos, A.G., Fernandez, B., Atiya, A.F., Muthusami, J., Tsai, W.K., 1994. An accelerated learning algorithm for multilayer perceptron networks. IEEE T. Neural Networ. 5 (3), 493–497.

Pedersen, S.A., Fock, H.O., Sell, A.F., 2009. Mapping fisheries in the German exclusive economic zone with special reference to offshore Natura 2000 sites. Mar. Policy 33, 571–590.

Pelletier, D., Ferraris, J., 2000. A multivariate approach for defining fishing tactics from commercial catch and effort data. Can. J. Fish. Aquat. Sci. 57, 51–65, doi:10.1139/cjfas-57-1-51.

Piet, G.J., Quirijns, F.J., Robinson, L., Greenstreet, S.P.R., 2007. Potential pressure indicators for fishing, and their data requirements. ICES J. Mar. Sci. 64, 110–121.

Prado, J., 1990. Fisherman's Workbook. FAO. Fishing News Books. Blackwell Scientific Publications Ltd., London.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 , http://www.R-project.org.

Russo, T., Parisi, A., Cataudella, S., 2011. New insights in interpolating fishing tracks from VMS data for different métiers. Fish. Res. 108, 184–194.

Sabatella, R.F., 2010. The 2009 Annual Economic Report on the European Fishing Fleet—Italy. Scientific, Technical and Economic Committee for Fisheries (STECF).

Sainsbury, J.C., 1996. Commercial Fishing Methods. An Introduction to Vessels and Gears, third ed. Fishing News Books, Oxford.

Salas, S., Gaertner, D., 2004. The behavioral dynamics of fishers: management implications. Fish Fish. 5, 153–167.

Sarle, W., 1995. Stopped training and other remedies for overfitting. In: Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics, Pittsburgh, PA, 21–24 June, Interface Foundation, pp. 352–360.

Scardi, M., 2001. Advances in neural network modeling of phyto-plankton primary production. Ecol. Model. 146 (1–3), 33–45.

Scardi, M., Harding Jr., L.W., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. Ecol. Model. 120, 213–223.

Stern, H.S., 1996. Neural networks in applied statistics. Technometrics 38, 205–214, doi:10.2307/1270601.

Tétard, A., Boon, M., Bennett, D., Berthou, P., Bossy, S., Casey, J., De Clerck, R., Delpech, J.P., Dintheer, C., Giret, M., Large, P., Latrouite, D., Lemoine, M., Millner, R., Morizur, Y., Ozanne, S., Palmer, D., Pawson, M., Pickett, G., Vince, M., 1995. Catalogue international des activités des flottilles de la Manche, approche des interactions techniques. Editions IFREMER, Brest, p. 337.

Tzanatos, E., Somarakis, S., Tserpes, G., Koutsikopoulos, C., 2006. Identifying and classifying small-scale fisheries métiers in the Mediterranean: a case study in the Patraikos Gulf, Greece. Fish. Res. 81 (2–3), 158–168, doi:10.1016/j.fishres.2006.07.007.

Ulrich, C., Andersen, B.S., 2004. Dynamics of fisheries, and the flexibility of vessel activity in Denmark between 1989 and 2001. ICES J. Mar. Sci. 61, 308–322.

Ulrich, C., Gascuel, D., Dunn, M.R., Le Gallic, B., Dintheer, C., 2001. Estimation of technical interactions due to the competition for resource in a mixed-species fishery, and the typology of fleets and métiers in the English Channel. Aquat. Living Resour. 14, 267–281.

Vermard, Y., Rivot, E., Mahévas, S., Marchal, P., Gascuel, D., 2010. Identifying fishing trip behavior and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. Ecol. Model. 221, 1757–1769.

Walker, E., Bez, N., 2010. A pioneer validation of a state-space model of vessel trajectories (VMS) with observers' data. Ecol. Model. 221, 2008–2017.

Zhang, G., Patuwo, E., Hu, M., 1998. Forecasting with artificial neural networks: the state of the art. Int. J. Forecasting 14, 35–62, doi:10.1016/S0169-2070(97)00044-7.