Ecological Modelling 222 (2011) 1048–1059

Contents lists available at ScienceDirect

Ecological Modelling

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

Optimization of an artificial neural network for identifying fishing set positions from VMS data: An example from the Peruvian anchovy purse seine fishery

Rocío Joo[a,b,*], Sophie Bertrand[a,b], Alexis Chaigneau[c,b], Miguel Ñiquen[b]

[a] Institut de Recherche pour le Développement (IRD), Centre de Recherche Halieutique Méditerranéenne et Tropicale, Avenue Jean Monnet, BP 171, 34203 Sète Cedex, France
[b] IMARPE, Esquina Gamarra y General Valle S/N Chucuito, Callao, Lima, Peru
[c] Institut de Recherche pour le Développement (IRD), Laboratoire d'Océanographie et de Climatologie: Expérimentation et Analyse Numérique (LOCEAN), UMR CNRS, Université P. et M. Curie, IRD, MNHN, Paris, France

ARTICLE INFO

Article history:
Received 8 January 2010
Received in revised form 24 August 2010
Accepted 26 August 2010
Available online 29 September 2010

Keywords:
Artificial neural networks
Sensitivity analysis
Fishing set locations
Vessel Monitoring System
Anchovy purse seine fishery

ABSTRACT

The spatial behavior of numerous fishing fleets is nowadays well documented thanks to satellite Vessel Monitoring Systems (VMS). Vessel positions are recorded on a frequent and regular basis which opens promising perspectives for improving fishing effort estimation and management. However, no specific information is provided on whether the vessel is fishing or not. To answer that question, existing works on VMS data usually apply simple criteria (e.g. threshold on speed). Those simple criteria generally focus in detecting true positives (a true fishing set detected as a fishing set); conversely, estimation errors are given no attention. For our case study, the Peruvian anchovy fishery, those criteria overestimate the total number of fishing sets by 182%. To overcome this problem an artificial neural network (ANN) approach is presented here. In order to set both the optimal parameterization and use "rules" for this ANN, we perform an extensive sensitivity analysis on the optimization of (1) the internal structure and training algorithm of the ANN and (2) the "rules" used for choosing both the relative size and the composition of the databases (DBs) used for training and inferring with the ANN. The "optimized" ANN greatly improves the estimates of the number and location of fishing events. For our case study, ANN reduces the total estimation error on the number of fishing sets to 1% (in average) and obtains 76% of true positives. This spatially explicit information on effort, provided with error estimation, should greatly reduce misleading interpretations of catch per unit effort and thus significantly improve the adaptive management of fisheries. While fitted on Peruvian anchovy fishery data, this type of neural network approach has wider potential and could be implemented in any fishery relying on both VMS and at-sea observer data. In order to increase the accuracy of the ANN results, we also suggest some criteria for improving sampling design by at-sea observers and VMS data.

1. Introduction

The ecosystem approach to fisheries management is increasingly calling for spatially explicit indicators (e.g. Pikitch et al., 2004; Babcock et al., 2005). Among those indicators, fishing effort is crucial for at least two reasons: (1) the need to control the compliance with spatially explicit management measures (such as inshore restrictions or Marine Protected Areas), and (2) the need to improve the interpretation of catch-per-unit of effort (CPUE) data in terms of fish stock abundance (see for instance problems related with spatial hyper aggregation, Rose and Kulka, 1999).

The effort deployed by fishing fleets can be spatially examined by VMS data. VMS provide high-resolution records of vessel positions on a regular time basis, ranging from few minutes to few hours depending on the fishery. However, while they are available for numerous fisheries and vessels (www.fao.org), this type of data has been barely used in fisheries science and management. One of the reasons is that VMS data do not provide explicit information on whether a vessel is fishing or not. Then, a first step for processing those data consists in estimating from position records where fishing events (set, haul, trawl, etc.) probably occurred. This is not a trivial task and several approaches have been undertaken, from very coarse ones (such as unique speed thresholds applied to different activities, Witt and Godley, 2007; and areas of high VMS poll density summed up with kernel home ranges, Harrington et al., 2007) to more refined ones (such as a combination of speed thresholds, directionality and other complementary rules; Deng et al., 2005; Mills et al., 2007). Although those methods usually detect

* Corresponding author. Tel.: +33 0676535663.
E-mail addresses: kaoridrummer@gmail.com (R. Joo), Sophie.Bertrand@ird.fr (S. Bertrand), alexis.chaigneau@ird.fr (A. Chaigneau), mniquen@imarpe.gob.pe (M. Ñiquen).

0304-3800/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2010.08.039

quite properly true positives (a true fishing set detected as a fishing set), false positives and over (or under) estimation errors are rarely assessed. For the present case study, the Peruvian purse seine anchovy fishery, Bertrand et al. (2008) showed that the use of a simple speed threshold on raw VMS data leads to an overestimation of the number of fishing sets of 182%. Alternatively, a general linear regression modeling approach (GLM) identified 65% of true positives and 16% of false positives, leading to a global underestimation of the total number of fishing sets of 19%. Thus, none of these two approaches are satisfying since they strongly bias the estimation of the number of fishing sets.

To solve this problem for the Peruvian anchovy fishery, Bertrand et al. (2008) proposed an ANN approach based on a Multilayer Feed-Forward Network (MFN). This methodology was chosen because (1) ANNs do not require to know nor to assume any probability distribution function, (2) ANNs are adapted for working with large datasets linked by complex non linear relations, and (3) among ANNs, MFNs are commonly used because of their simplicity and the wide availability of software tools. This particular ANN is first trained on a subset of fishing trips for which fishing set positions are documented by at-sea observers (~1% of the total fishing trips) and then used to estimate the location of fishing sets for the remaining trips monitored by VMS only. This ANN is designed to overcome the overestimation problem as it aims at: (1) accurately assessing the total number of fishing sets and (2) maximizing the true-to-false positives ratio. Bertrand et al. (2008) applied this tool on a rather limited spatial area (7°S–10°S along the Peruvian coast) and temporal window (2000–2002), correctly identifying 83% of the real fishing sets (true positives) with a total overestimation of 0.5%.

Based on those satisfying preliminary results, and before implementing such a tool in routine for the monitoring dashboard of the Peruvian anchovy fishery, there is a critical need to check the behavior and validate an optimal parameterization of the ANN when confronted to variable situations in time and space. We are particularly concerned in answering 3 main questions:

(1) At-sea observers data are usually not available on a day-to-day basis; then, when using the ANN in real-time during the fishing season, what is the effect of estimating fishing sets using an ANN trained on an earlier period?
(2) Two types of vessels, steel and wooden hulls, are participating to the industrial anchovy reduction fishery; although wooden vessels are only recently and gradually incorporated to the VMS monitoring. What may be the effect in the ANN behavior of this change in the fleet composition in VMS data?
(3) Management rules differ between two large regions in the coast of Peru (coastal restrictions, fishing bans, total allowable catch); does the ANN need different optimizations for the two regions or can we use a single ANN along the entire Peruvian coast?

Those questions are addressed performing a global optimization of the ANN. ANN optimization ranges from trial-and-error sensitivity analysis (e.g. Dedecker et al., 2004) to more complex and efficient approaches such as genetic algorithm or simulated annealing (e.g. Mühlenbein, 1990; Sexton et al., 1999; Bernardos and Vosniakos, 2007). These latter optimizations mainly concern network architecture (e.g. number of neurons and layers, and shape of activation functions) and convergence rules for the training algorithm (Bernardos and Vosniakos, 2007). Here, we refer to the optimization of (1) the internal structure and training algorithm of the ANN and (2) the "rules" used for choosing the relative size and composition of the DBs for the ANN training and inference. To address these two aspects with comparable methods, we use a trial-and-error sensitivity analysis.

In the next section, we present some characteristics of the Peruvian fishery, the data used in this study, the ANN architecture and training algorithm; and then we describe and perform a series of sensitivity tests. From the results obtained in terms of ANN performance, we draw practical recommendations for an optimal and robust use of this tool, in the specific case of the Peruvian anchovy fishery. Finally, we discuss on how this type of neural network approach can have wider potentials and could be implemented, with adjustments on the input variables, in any fishery relying on both VMS and at-sea observer data.

2. Materials and methods

2.1. Some insights into the Peruvian anchovy reduction fishery, its monitoring and management and the data used

The Peruvian anchovy fishery is characterized by the remarkable size of its production (~7 millions t.y^{-1} since 1999) and its sensitivity to the intense regional climatic variability on various spatio-temporal scales (Chavez et al., 2008). Indeed, climatic scenarios such as El Niño or la Niña events directly condition the extent of the anchovy habitat, modifying its catchability and driving its population dynamics (Bertrand et al., 2004a). To cope with this strong natural variability, fishing authorities adopted an adaptive management for the industrial fishery (Chavez et al., 2008). This management is adaptive since catch limits are re-assessed every ~6 months and opening and closure periods decided on the basis of daily monitoring of the ecosystem, the fish population and the fishery. The fishing activity is monitored by the Instituto del Mar del Peru (IMARPE) through landings statistics, VMS and at-sea observer data. At-sea observer data provide, for a small sample (~1%) of fishing trips, detailed information on time spent on steaming and searching, and on the position and catch composition of the fishing sets, among others (Bertrand et al., 2004b). Accepting observers on board is not a legal obligation for fishing companies. The observer program is therefore run on a voluntary basis by the vessels and relies on "gentlemen's agreement" between IMARPE and the fishery. While landings and VMS data are daily available to fishery authorities, at-sea observer data need several weeks to be centralized and formatted. This difference in timing of availability for the different datasets means that fishing set positions need to be estimated mainly with an ANN trained on an earlier period. The management of this fishery also depends on the region. In the north-centre region of Peru (NC, from the frontier with Ecuador at ~3°S–16°S), where most of the landings take place, industrial fishing is forbidden within the first 5 nm from the coast. Total landings are limited by a total allowable catch (TAC) and effort is limited by fairly long fishing bans (at least for the period under study; Peru implemented a new individual quota system in 2009 which should lead to longer fishing seasons). In the southern region (S, from 16°S to the frontier with Chile), coastal restrictions vary from 1.5 to 3 nm from the coast. As total landings in that area are not limited by a TAC, fishing closures are much more limited in duration and are mainly implemented if the number of juveniles in the catches exceeds a given threshold. The pelagic industrial fleet is composed of two types of vessels, both of them delivering their catches to fish meal plants for reduction: a steel fleet (steel-hulled and at least 110 m^3 of fish-hold capacity), and a wooden fleet (wooden-hulled and with a fish-hold capacity ranging from 30 to 110 m^3). Both fleets are legally obliged to use VMS tracking devices since 1999. While the steel fleet was almost entirely covered with VMS by 2000, the coverage of the wooden fleet has been much more gradual.

In this study, we use the complete VMS and at-sea observer data from late 1999 to 2007 available along the Peruvian coast. VMS data provide on a ~1-h basis, precise geographical position records for the entire steel fleet and a continuously growing part