From forager tracks to prey distributions: an application to tuna vessel monitoring systems (VMS)
Emily Walker, Jacques Rivoirard, Philippe Gaspar, Nicolas Bez

To cite this version:
Emily Walker, Jacques Rivoirard, Philippe Gaspar, Nicolas Bez. From forager tracks to prey distributions: an application to tuna vessel monitoring systems (VMS). Ecological Applications, Ecological Society of America, 2015, 25 (3), pp.826-833. 10.1890/14-0193.1.sm. hal-01143796

HAL Id: hal-01143796
https://hal-mines-paristech.archives-ouvertes.fr/hal-01143796
Submitted on 27 May 2020

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.
From forager tracks to prey distributions: an application to tuna vessel monitoring systems (VMS)

EMILY WALKER,1,2 JACQUES RIVOIRARD,3 PHILIPPE GASPAR,4 AND NICOLAS BEZ1,5

1UMR MARBEC, Institut de Recherche pour le Développement (IRD), Centre de Recherche Halieutique Méditerranéenne et Tropicale, Avenue Jean Monnet, CS 30171, 34203 Sète, France
2INRA, BioSP, 228 Route de l’Aérodrome, CS 40509 Domaine St Paul Site, Agroparc 84914 Avignon cedex 9, France
3MINES ParisTech, Centre de Géosciences, 35 rue Saint Honoré 77305 Fontainebleau, France
4CLS, Parc technologique du Canal, rue Hermès, 31520 Ramonville St-Agne, France

Abstract. In the open ocean, movements of migratory fish populations are typically surveyed using tagging methods that are subject to low sample sizes for archive tags, except for a few notable examples, and poor temporal resolution for conventional tags. Alternatively, one can infer patterns of movement of migratory fish by tracking movements of their predators, i.e., fishing vessels, whose navigational systems (e.g., GPS) provide accurate and frequent VMS (vessel monitoring system) records of movement in pursuit of prey. In this paper, we develop a state-space model that infers the foraging activities of fishing vessels from their tracks. Second, we link foraging activities to probabilities of tuna presence. Finally, using multivariate geostatistical interpolation (cokriging) we map the probability of tuna presence together with their estimation variances and produce a time series of indices of abundance. While the segmentation of the trajectories is validated by observers’ data, the present VMS-index is compared to catch rate and proved to be useful for management perspectives. The approach reported in this manuscript extends beyond the case study considered. It can be applied to any foragers that engage in an attempt of capture when they see prey and for whom this attempt is linked to a tractable change in behavior.

Key words: GPS; multivariate geostatistics; presence index; spatiotemporal distribution; trajectometry; tropical tuna; vessel monitoring system (VMS).

INTRODUCTION

Effective conservation and management of resource species require information on changes in population abundances and spatial distributions. It is optimal for this information to be based on direct observations. However, in many cases, direct observations are not possible, and one is left with indirect approaches. In marine fisheries, these indirect approaches generally involve the use of catch rates of commercial boats to indicate abundance both at the scale of a population’s entire range, and at the scale of some subarea within that range. Catch rates are computed on the basis of the compulsory declarations of catch and effort made by fishermen. The use of catch rates is not without serious and long-acknowledged problems (Paloheimo and Dickie 1964, Gulland 1974, Laurec 1977, Allen and Punsly 1984, Winters and Wheeler 1985, Rose and Kulka 1999, Gaertner and Dreyfus-Leon 2004, Polacheck 2006, Maund et al. 2006, Pianet et al. 2008, Soto et al. 2009). Nonetheless, in the numerous situations where scientific surveys are impossible to perform, we are compelled to use indirect indices of abundance. This is the case, for instance, for highly migratory species like tuna. Given the large geographical extent of tropical tuna, scientific cruises dedicated to sample and assess tuna distribution are currently impossible. Fisheries data are thus the only available information to estimate biomasses of tuna and of other exploited marine top predators (Langley et al. 2009).

The worldwide development of systems for tracking individual vessels, i.e., routine acquisition of GPS locations (in fisheries, the devices are called vessel monitoring systems [VMS]), makes it possible to estimate spatial distributions of resource fish populations using a novel approach. The basic principle is to link the foraging behaviors of fishermen, routinely monitored by GPS, to the abundance of their prey, at least qualitatively. Fishermen have been shown to behave like marine top predators (Begossi 1992, Mangel and Clark 1988, Bertrand et al. 2007, Wise et al. 2012). Using tools developed in movement ecology (Nathan et al. 2008), it may be possible to develop analytical frameworks that produce distribution maps and biomass of prey (tuna) from trajectories of their predator (fishing vessels). Such predator trajectories have been used to describe key behavioral features of the predator and to infer the environmental variables affecting behavior (e.g., Pinaud and Weimerskirch 2007). Here we use known links between foragers’ behaviors and prey
presence to infer the prey field from foragers’ movements.

Several types of analytical methods have been developed to estimate behavior states from tracking data. One approach concerns the detection of “area restricted search” (ARS) (Knoppien and Reddingius 1985, Fauchald and Tveraa 2003, Tremblay et al. 2007). However, these approaches restrict the analysis to the part of the trajectories that are considered most interesting in terms of foraging intensity, leaving the rest of the trajectory unused (e.g., Knell and Codling 2012, Louzaol et al. 2013). Classification techniques (e.g., artificial neural networks, random forests) allow for the estimate of several behavioral states (Dalziel et al. 2008, Bertrand et al. 2008, Joo et al. 2011). However, these techniques exhibit statistical or associational links that cannot easily be turned into causal links (Pearl 2009) or that are not based on biological processes. State-space models using hidden Markov models are now widely implemented in marine and terrestrial ecology to take into account prior knowledge of behavior and physiological states (Morales et al. 2004, Jonsen et al. 2005, Patterson et al. 2008, Vermaard et al. 2010, Walker and Bez 2010). Recently Jonsen et al. (2013) reviewed state-space modeling approaches for biologging data. The possible behavioral states are fixed and based on biological knowledge. Most commonly the three possible states are feeding or fishing on prey, intensive searching or foraging for prey and, extensive searching of prey or migration. Even if continuous-time versions exist (Johnson et al. 2008, Harris and Blackwell 2013), state-space models mostly rely on discrete Markovian switches between states over time and require regular time step acquisition of data.

The framework put forward in the present paper comprises four steps. First, a state-space model was parameterized to estimate foraging activity of tuna purse-seiners based on their hourly GPS positions over the period 2006–2010 (Walker and Bez 2010, Bez et al. 2011). This model was validated by on-board observer data. As in many studies, we documented the existence of two contrasting search strategies, hereafter referred to as extensive and intensive searches. These strategies correspond to intra- or inter-patch behaviors (Knoppien and Reddingius 1985). Second, proportions of time spent in each type of foraging activity were computed for the entire fleet. Third, in each pixel, the probability of tuna presence (P) was estimated by cokriging (Chilès and Delfiner 1999) the three foraging activities. Finally, monthly indices of presence were finally obtained by integration of the cokriging maps.

**Materials and Methods**

The data used for this study were collected from 2006 to 2010, from the French-flagged purse seiners (15–20 vessels). They target three tropical tuna species in the Western Indian Ocean. The global positioning systems (GPS) positions of the vessels were registered every hour and provided accurate positions (errors smaller than a few tens of meters).

Any hourly segment (step) of the vessels’ trajectories was assigned to one of the three movement states defined for this study (fishing, intensive searching, extensive searching). The model consisted of a state-space model (Buckland et al. 2004, Royer et al. 2005, Patterson et al. 2008) where the movement states were assumed to follow a first-order Markovian process. Although this framework has already been applied in other ecological studies (Morales et al. 2004, Jonsen et al. 2005), this is one of the first applications to VMS data (see also Vermaard et al. 2010). States were inferred using a Bayesian framework (Gelman et al. 2004) based on both vessel speeds and turning angles distributions. Speeds were modeled by beta distributions and turning angles by wrapped Cauchy distribution.

Fishing was associated with low speeds (right-skewed distribution) and a uniform distribution of turning angles. In contrast, extensive search was characterized by high speeds and a relatively straight path. Intensive search was intermediate. A thorough description of the model structure, calibration, and validation can be found in Walker and Bez (2010).

Model outputs (states) were aggregated over pixels of 0.2° × 0.2° × month, providing proportions of time spent per activity and per spatiotemporal pixel; these were denoted F, IS, and ES for fishing, intensive searching, and extensive searching respectively. Proportions of time spent in each type of foraging activity were computed for the entire fleet and for pixels of 0.2° × 0.2° on a monthly basis. This pixel size corresponds to 12 nautical miles (i.e., 22 km) and is the distance up to which fishing boats can detect tuna aggregations. The three proportions spent in each foraging activity sum to one, so knowing two proportions amounts to knowing the third one. In each visited pixel, the probability of tuna presence (P) was estimated by assigning a weight to the proportion for each type of foraging activity.

The final objective was to estimate the probability of tuna presence. The rules to link foraging activities to tuna presence (P) were based on the following well known and recognized fishing practices: a fishing operation is performed if and only if a tuna school is observed (F); intensive tracking tuna schools in areas of aggregations (IS) is driven by visual clues of probable tuna presence (such as presence of birds); search for prey is maintained all day long from sunrise to sun set, including during directed movement (ES), until some detection of tuna presence are revealed by one of the observation means (sonar, radar, and/or binocular surveillance). Based on these considerations, the probabilities that tuna are present during F and ES foraging activities (fishing and extensive search) were logically set to 1 and 0, respectively. The probability of tuna presence associated with intensive search (IS) was set to 0.5. So, the probability of tuna presence is defined by P = F + 0.5IS. As intensive
search only occurs in conjunction with strong indicators of tuna presence, it is likely associated with a probability of tuna presence that is larger than 0.5. To test for the sensitivity of the final results to this last parameter, we also performed a full analysis with $P = F + 0.75\text{IS}$.

Because the proportions of time spent in the three movement states sum to one, the dimension of the model could be reduced to two independent cross correlated variables. By construction, cokriging insures by definition, consistency of the estimation of any linear transformation of the variables (Chiles and Delfiner 1999). Thus, cokriging $F$ and $\text{IS}$ is also rigorously equivalent to cokriging $P$ and $F$, thereby providing a direct estimate of the target variable, namely the probability of tuna presence ($P$). A linear model of co-regionalization (Wackernagel 1995) common to each month was thus constructed for $P$ and $F$ (Appendix: Fig. A1), using a nugget effect and three spherical models with spatial ranges approximately equal to 0.5° (30 nautical miles), 1° (60 nautical miles), and 5.8° (350 nautical miles). Cokriging was performed with a moving neighborhood of 5° of diameter, on a $0.1° \times 0.1°$ grid (local estimation). VMS-based indices of presence (global estimation) were obtained by integrating the cokriging maps over the area where the standard deviation of the estimate was smaller than 0.25. While the standard deviation for the global estimate cannot be obtained by integration of the local ones (e.g., Chiles and Delfiner 1999), a global cokriging is necessary to quantify the estimation variance of the indices of presence.

The extensive searching activity did not contribute directly to the estimation $P$ as it received a weight of 0 in the final calculation. However, it had a strong impact on the estimation of the other proportions thanks to the coherent multivariate model and contributed strongly to the estimations performed.

All the computations were made under R (R Core Development Team 2010) with the R library RGeostats (Renard et al. 2014).

**RESULTS**

Information provided by observers on board 5–10% of the vessels showed that 90% of the fishing operations were properly classified by the state-space model (Walker and Bez 2010). At the fleet level, coherent patches of fishing activities emerged (Fig. 1) that provided a qualitative validation of the model. These patches included areas of intensive and extensive search. These spatial features were stable over time as indicated by the empirical variograms (compact boxplots of the 5 years $\times$ 12 months = 60 empirical variograms: Appendix: Fig. A1). The short scale (0.8°, ~50 nautical miles) dominant in the variograms of the fishing activities was clearly characteristic of the scale of prey aggregations. The medium scale associated with the intensive search

---

**Fig. 1.** Vessels’ trajectories with estimated foraging activities represented for every hourly step of the trajectory (June 2007). Vessels are in the Indian Ocean. Search strategies are described in Materials and Methods.
areas (6°, ~350 nautical miles) appeared to characterize the scale of their apparent niche.

Distribution maps of tuna probability of presence in the western Indian Ocean (June 2007 provided as an example; Fig. 2A) reveal clear mesoscale structures. These spatial features were consistent with empirical prior knowledge on cluster size and on coherent oceanic features (e.g., cyclonic and anticyclonic eddies).

Due to the migratory behavior of tuna, fishing grounds change from month to month. Selecting a fixed area over which to assess tuna presence was thus impossible. We decided to compute an index of presence over the restricted area where the estimation of tuna presence was above confidence limit defined ex-post on the basis of the standard errors. The spatial model (variogram) being common to all months considered in this study, the standard errors made during the cokriging interpolations were only dependent on location and the number of available samples. These standard errors were comparable among months. The threshold was thus defined on the basis of the histogram of all the standard errors, which showed a clear mode of large values corresponding to sectors of the map where data were missing, to estimate the tuna presence with sufficient confidence (Appendix: Fig. A2). We decided to remove the part of the map where standard errors were above 0.25 (Fig. 2A and 2B). The resulting time series of the indices of presence showed strong month to month variability (Fig. 3). This variability was stronger than the year to year variability, which is the time scale at which indices of presence are usually built and used for conservation measures. After a three-year decrease, the index of presence increased again in 2009 and 2010.

Index of tuna presence has been also calculated with a probability of tuna presence associated with intensive search (IS) equal to 0.75. The two indices were strongly correlated (Appendix: Fig. A3), the main difference between the two being that the index of tuna presence calculated with a coefficient of 0.75 was shifted by a constant value (the average and the standard deviation of the difference between the two indices were 0.1 and 0.019, respectively).

**DISCUSSION**

The lack of data for ground-truthing inferred that animal behavior is problematic when analyzing tracking data. Among the problems that this generates, the lack of validation of the model output is predominant, making any model (e.g., state-space models, artificial neural networks, ARS, regression trees) as good as any other, provided that it runs. The primary validation of model output is thus the confrontation to independent field data when possible. In the present case, a sound validation of the segmentation of trajectories was made possible by the presence of observers on board some of the vessels. This led us to evaluate that 90% of the fishing sets were properly detected by the algorithm (Walker and Bez 2010). It is worth noting that observers’ data were also used to determine empirically the order of the Markovian structure of the state-space model.

Despite the fact that one of the major variables controlling populations’ dynamics is their abundance, we often cannot measure it in practice. As a consequence, models for populations’ dynamics rely on indices of abundance rather than absolute abundance, i.e., statistics or measures that vary in time proportionally to the unknown true abundance. In essence, this makes it impossible to choose one index of abundance as opposed to other ones as we lack objective means to validate them. However we compared the VMS-index of presence developed in this paper with the traditional index of abundance used by the assessment working groups, that is, the catch rate (Fig. 3). The catch rates were computed...
as the monthly sums of the catch of the three major tuna species recorded for the French fleet (namely yellowfin tuna \([Thunnus albacares]\), bigeye tuna \([Thunnus obesus]\), and listao \([Katsuwonus pelamis]\)) divided by the number of fishing operations performed during that month. Indices were well correlated (Fig. 3), with a coefficient of correlation significantly different from 0 \((P = 0.01 \text{ after correcting for the autocorrelation of order one that was present in both time series [Pyper and Peterman 1998]})\). Both indices had similarly strong month to month variability. The spatiotemporal window chosen to analyze VMS data was the month (best compromise between too small, not enough data, and too large, not enough concomitance). We were thus left with monthly outputs.

However, the abundance of the tuna populations considered here, and the indices used to measure it, are not supposed to change strongly from one month to the next. This is why we have computed annual means, which conforms to the usual practice, and show similar trends. However, at that scale, the observations were too few (five points) to draw solid conclusions.

The model outputs also reproduce expected spatial patterns in the distribution of the tuna populations. Two are particularly worth mentioning. First, spatial structures (variograms) were strongly stable over time. This is consistent with the assumption that the tuna aggregative behavior is stable over the time scale of few years, i.e., far below evolutionary time scales. Distinction must be
made here between the stable spatial structure (social behavior) and the spatial distributions (habitat), which is more likely to vary in time. This latter variability is not inconsistent with the stability of the variograms. Second, the final probability maps exhibited meso-scale patterns that mimic similar meso-scale physical cyclonic and anti-cyclonic structures (Bakun 2006). Eddies have been shown to aggregate seabirds (Weimerskirch et al. 2004, Tew-Kai et al. 2009), potentially attracted by epipelagic communities, which is also a cue for the distribution of the tuna. These emerging properties did not correspond to any constraint of the models and were thus considered as indications of the internal consistency of the overall approach.

Vast tagging programs using archive tags (either internal tags or pop-up satellite archival tags [PSAT]), such as the ones undertaken for bluefin tuna or the one made in the Pacific (Block et al. 2005, Leroy et al. 2013), could certainly provide complementary and very useful information. Unfortunately, the tagging program undertaken in the Indian Ocean (e.g., Fonteneau and Hallier 2014) was based on traditional conventional tags and was thus not useful in the present context. Trajectometry, i.e., the transcription of tracking data into ecological concepts, was not the end product in itself of our work. It was rather an intermediate step for assessing tuna presence based on expert predefined rules connecting foraging activities of vessels to tuna presence. Fishing operations occur if and only if tuna is observed. Thus, the assumed association between fishing activities and tuna presence was well founded. Phases of extensive search correspond to active but unfruitful search of tuna aggregations. Thus associating extensive search with the absence of tuna was also well founded. Phases of extensive search correspond to active but unfruitful search of tuna aggregations. Thus associating extensive search with the absence of tuna was also well founded. However, the probability of tuna presence that was assigned to periods of intensive search (0.5) was speculative. Theoretically, the cokriging weights only depend on the form of the variogram but not its level in terms of variance–covariance, and the model of the pairs of variables \(P,F\) is nothing but a linear combination of the variogram models for \(F\) and IS. The models for \(P = F + 0.5\)IS and for \(P = F + 0.75\)IS are thus only different in the proportion given to the individual variograms but their ranges are strictly equal. Changes in cokriging estimates are thus not expected to be strong unless the two models are strongly different, i.e., unless the probability of presence associated with IS varies strongly.

The estimation variance computed here corresponds to the uncertainty coming from the spatial interpolation between tracks. It was driven by the strength of the spatial auto and cross-correlations of the study variables, the geometry of the polygon to estimate and, the location and the number of samples in the polygon. It did not include other uncertainties like, for instance, the one associated with the estimation of activities coming from the state-space model (posterior distributions) or uncertainties in the link between foraging activities and tuna presence.

On average, only one-quarter of the daytime is devoted to fishing (Bez et al. 2011). Three quarters of the vessels activity are thus not recorded in logbooks where only fishing sets are mandatory. Even though uncertain, the use of the information contained in this dominant part of the vessels’ activity represents, without a doubt, a significant improvement for inferring tuna presence, at least by indicating their absence. Observed absences of tuna aggregation are explicitly used in the present analysis, contrary to traditional catch rates. However, VMS data, i.e., GPS locations of vessels, share the same drawbacks of the fisheries statistics used to produce catch rate (Table 1). They only relate to the exploited fraction of the population, and no quantitative assessment of the size of this fraction as compared to the entire population can be made. The advantage of non declarative data like VMS is that they are free of all the bias associated with the declarations (e.g., rigor, intentional over or under declaration) and that they are near-real-time data. VMS data is useful because its collection of data does not depend on the cooperation of fishermen, and depending on the accuracy of recorded positions (thanks to GPS) contrary to the rough positions noted by the skippers. They are also the unique source of timely, exhaustive, spatially explicit and error free information on fleet dynamics.

The present system suffers one important limit that relates to the three major tuna species jointly (i.e., yellowfin tuna, bigeye tuna, listao). These highly marketable species are targeted by fishing operations either in free schools or in schools aggregated beneath floating objects (Moreno et al. 2007a, b, Dagorn et al. 2012). These three species can be mixed or not in schools and have similar two-dimensional habitats. Vessel trajectories do not hold enough information by themselves to differentiate the three targeted species and the two fishing modes (sets made on free schools or on

### Table 1. Comparison between traditional (catch rate) and new (vessel monitoring system [VMS]-based) indices of abundance.

<table>
<thead>
<tr>
<th>Index characteristic</th>
<th>Catch rate</th>
<th>VMS based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative data</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Timeliness</td>
<td>1–2 years delay</td>
<td>near-real time</td>
</tr>
<tr>
<td>Population representativeness</td>
<td>exploited fraction</td>
<td>exploited fraction</td>
</tr>
<tr>
<td>Estimation variance</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Specific</td>
<td>yes</td>
<td>generally not</td>
</tr>
<tr>
<td>Connection to real unknown</td>
<td>if fishing is purely at random</td>
<td>if foraging behavior is well connected to prey abundance</td>
</tr>
<tr>
<td>abundance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
floating objects called fish aggregated devices (FADs). Therefore, without an extra step based on catch data to estimate species composition, the probability values calculated here aggregate the three tuna species together without distinguishing between free-schools sets and FAD-associated sets.

Integrating the distribution maps led to monthly indices of presence. As the matter of fact, the maps produced represent probability of tuna presence and not abundance. Assuming that school characteristics are stable over time (no trend in school mean abundance) and that tuna presence/absence fluctuate in time like abundance (no change in tuna aggregative behavior), the probability of presence could be considered as proportional to the abundance, which is the definition of an index of abundance.

**Conclusion**

Stock assessment models tend to incorporate more and more data sources (Taylor et al. 2011), because available indices are all imperfect indices of abundance. The present paper provides a new VMS-based index that could be incorporated in this type of model. Catch rates and VMS-based index can not be directly compared but are rather complementary.

The approach presented in this paper is not specific to this particular case of study. In particular, such a VMS-based index may be readily applied worldwide to all (temperate and tropical) tuna purse-seine fisheries. Furthermore the model is already applied in Peru in small pelagic fishery (Joo et al. 2013). This approach could also be applied to other foragers, for which fine-scale tracking data are available and for which the behavior can be connected to biological states (e.g., prey searching, feeding, breeding).

**Acknowledgments**

The work was co-funded by the European project “Catch, effort and discard estimates in real time” (CEDER, FP7-222615) and by CLS (Collecte Localisation Satellite in Toulouse, France). Thanks to cooperative skippers from ORTHONGEL, data from scientific observers were kindly provided by the French “Observatoire thonier tropical” from IRD. VMS data were provided by the French Fisheries Department (Direction des Pêches Maritimes).

**Literature Cited**


