

Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests

Faustinato Behivoke, Marie-Pierre Etienne, Jérôme Guitton, Roddy Michel Randriatsara, Eulalie Ranaivoson, Marc Léopold

▶ **To cite this version:**

Faustinato Behivoke, Marie-Pierre Etienne, Jérôme Guitton, Roddy Michel Randriatsara, Eulalie Ranaivoson, et al.. Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests. Ecological Indicators, 2021, 123, pp.107321. 10.1016/j.ecolind.2020.107321 . hal-03116922

HAL Id: hal-03116922
https://hal.science/hal-03116922

Submitted on 26 May 2021

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.

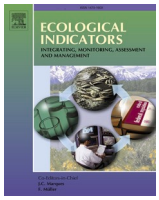
Ecological Indicators 123 (2021) 107321



Contents lists available at ScienceDirect

Ecological Indicators

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



Short Note



Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests

Faustinato Behivoke [a], Marie-Pierre Etienne [b], Jérôme Guitton [c], Roddy Michel Randriatsara [a], Eulalie Ranaivoson [a], Marc Léopold [d,*]

[a] Institut Halieutique et des Sciences Marines (IH.SM), University of Toliara, BP 141, 601 Toliara, Madagascar
[b] University of Rennes, Agrocampus Ouest, CNRS, UMR 6625 IRMAR, F-35000 Rennes, France
[c] ESE, Agrocampus Ouest, INRAE, 35042 Rennes, France
[d] ENTROPIE (IRD, University of La Reunion, CNRS, University of New Caledonia, Ifremer), 97400 Saint-Denis, La Reunion c/o IH.SM, University of Toliara, BP 141, 601 Toliara, Madagascar

ARTICLE INFO

Keywords:
Boat movement
Fishery map
GPS track
Madagascar
Spatial data
Speed threshold

ABSTRACT

During the last decade spatial patterns of industrial fisheries have been increasingly characterized using tracking technologies and machine learning analytical algorithms. In contrast, for small-scale fisheries, fishers' behaviour for estimating and mapping fishing effort has only been anecdotally explored. Following a comparative approach, we conducted a boat tracking survey in a small-scale reef fishery in Madagascar and investigated the performance of a learning random forest algorithm and a speed threshold for estimating and mapping fishing effort. We monitored the movements of a sample of 31 traditional sailing fishing boats at around 45 s time interval using small GPS trackers. A total of 306 daily tracks were recorded among five gear types (beach seine, mosquito trawl net, gillnet, handline, and speargun). To ground-truth GPS location data, fishers' behaviour was simultaneously recorded by a single on-board observer for 49 tracks. Typical, gear-specific track patterns were observed. Overall, the random forest model was found to be the most reliable, generic, and complex method for processing boat GPS tracks and detecting spatially-explicit fishing events regardless gear type. Predictions of mean fishing effort per trip showed that both methods reached from 89.4% to 97.0% accuracy across gear types. Our findings showed that boat tracking combined with on-board observation would improve the reliability of spatial fishing effort indicators in small-scale fisheries and contribute to more efficient management. Selection of the most appropriate GPS data processing method is dependent on local gear use, fishing effort indicators, and available analytical expertise.

1. Introduction

Accurate spatial and quantitative information on gear use is key for assessing the sustainability of fisheries and their impact on marine resources and the environment, as well as for supporting adaptive, area-based regulation of harvesting (Wilen, 2004; McCluskey and Lewison, 2008). To investigate the dynamics of fishers' behaviour across gears, both in time and space, research has opportunistically benefitted from on-time monitoring systems of large-scale fishing vessels, i.e. the Vessel Monitoring System (VMS) and the Automatic Information System (AIS). These systems record boat location at regular time intervals (usually set from 30 min to 2 h) in a large number of industrial fisheries worldwide.

During the last decade, characterizing the spatial patterns of industrial fisheries has been well established through such common tracking technologies. Location data is automatically processed using three types of analytical tools: i) discrimination methods based on boat speed threshold (Lee et al., 2010; de Souza et al., 2016; Le Guyader et al., 2017), ii) segmentation methods including state space models (Vermard et al., 2010; Walker and Bez, 2010; Peel and Good, 2011; Gloaguen et al., 2015), and iii) machine learning algorithms including neural networks and random forests (Russo et al., 2011; Joo et al., 2013) with behavioural data recorded simultaneously.

These processing methods allow for segmenting a large number of vessel trajectories into fishing and non-fishing activities based on behavioural criteria (e.g. boat speed and direction), thus quantifying and locating harvest during fishing trips at high spatial resolution

* Corresponding author.
E-mail address: marc.leopold@ird.fr (M. Léopold).

https://doi.org/10.1016/j.ecolind.2020.107321
Received 15 May 2020; Received in revised form 16 December 2020; Accepted 23 December 2020
Available online 13 January 2021
1470-160X/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

(<100 m) and hourly temporal scale. Accurate maps of fishing effort indicators (typically, the number of fishing hours per unit area) may then be produced in those industrial fisheries to estimate the distribution of fishing pressure across gear types as well as fishery-dependent abundance indices such as spatial catch-per-unit-effort (CPUE) with more precision.

However, in most small-scale fisheries, scarcity of numerical spatial data has been a recurrent bottleneck as VMS and AIS have little applicability in these contexts. This is due to the high operating and maintenance costs of those systems and the typically-low technological capacity of small-scale fishing boats, that often lack permanent on-board energy supply for instance. Consequently, fishers' movements have poorly been investigated for estimating and mapping fishing effort in those fisheries, while coarse spatial indicators of fishing effort have been proposed (e.g. Daw, 2008; Stewart et al., 2010).

While GPS (Global Positioning System) geolocation devices were sporadically used to map specific fishing spots or hauls to explore the spatial dynamics of resources, exploitation, and management in small-scale fisheries (e.g. Begossi, 2001; Stelzenmüller et al., 2007), the emergence of low-cost, sophisticated GPS trackers has recently opened new avenues for collecting trajectory data and accurately estimating fishing effort. In published surveys, expert-based boat speed threshold was used as the unique, pre-defined classifier to predict fishing and non-fishing activities from boat trajectories (Burgos et al., 2013; Navarrete Forero et al., 2017; ICES, 2019). However, we argue that using such a single-parameter speed method may not be suitable, a priori, for classifying fishing and non-fishing activities over a large range of small-scale fishery contexts and gear types. Because machine learning procedures would perform this classification task relying on inference and a large panel of descriptors of boat trajectory patterns, they are expected to provide a more effective and generalizable analytical framework of boat movements in small-scale fisheries, as they have done so already in large-scale fisheries (de Souza et al., 2016).

In this paper, we investigate the performance of a random forest algorithm and a speed threshold method for automatically classifying boat movements into fishing and non-fishing activities in a small-scale reef fishery in Madagascar. For this case study we monitored fishing boat movements using GPS trackers and ground-truth GPS data across five fishing gear types. Fishing boat trajectories were then processed per gear type through the random forest model and a speed threshold method following a comparative approach. Results were interpreted in terms of the reliability and generalization capacity of both analytical methods to estimate fishing effort across gear types. To our knowledge, this is the first study using a random forest model and boat trajectory data to characterize fishing effort in small-scale fisheries.

2. Materials and methods

2.1. Study site and tracking data collection

The study area was located in the bay of Toliara, southwestern Madagascar (Fig. 1). Owing to the proximity of the city of Toliara (326,000 inhabitants in 2018), this 157 km^2-wide reef and lagoon complex has been intensively exploited to feed the growing urban demand in fish (Bruggemann et al., 2012). We registered a total of 892 2.5–7 m long, outrigger sailing canoes targeting reef fish. Those traditional fishing boats were operated by one to five fishers using the typical gear types of coastal fisheries in the Western Indian Ocean region (Bruggemann et al., 2012).

We monitored the movements of a sample of 31 boats using small GPS trackers (i-GotU GT600 and Catlog) in 2017 and 2018. We provided an explicit explanation of future information use and a small financial

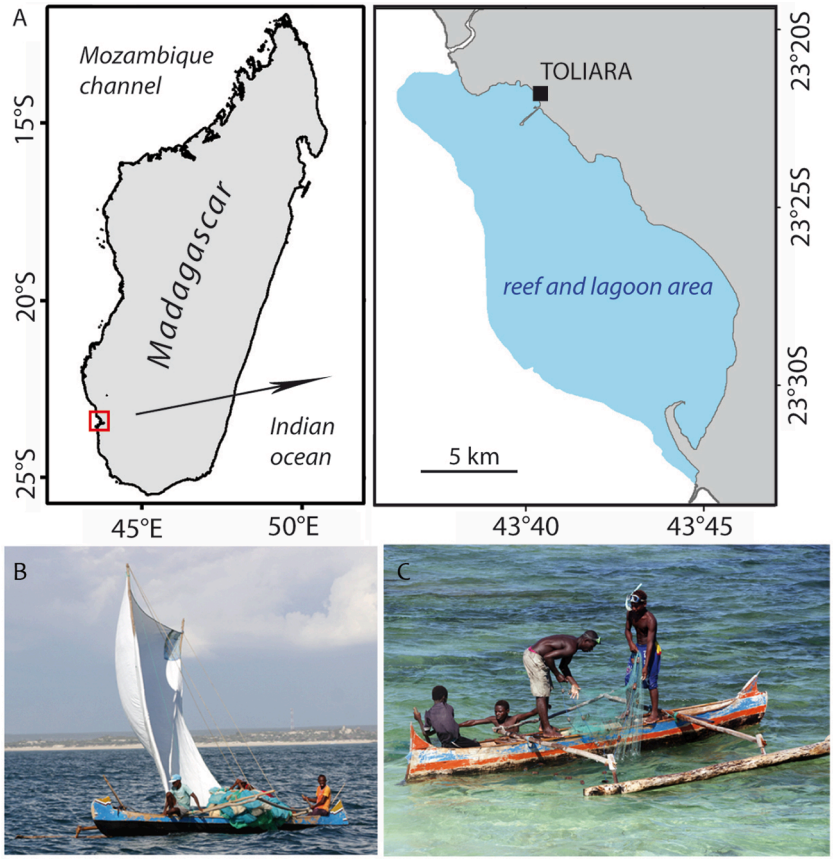


Fig. 1. Location of the survey site in southwestern Madagascar (A). Grey area: land mask (see Section 2). The small-scale fishery is operated by hundreds of traditional sailing boats (B) whose sail was usually folded while fishing (C).

incentive (USD 0.1.day^{-1}) to participant fishers. The GPS tracker recorded boat position at around 45 s time intervals to detect potential rapid change in fishers' activities during eight to fifteen consecutive days, according to battery life. A total of 306 daily tracks were recorded using the five major gear types of the fishery: beach seine, mosquito trawl net, gillnet, handline, and speargun (Table 1).

To ground-truth GPS location data (e.g. Alvard et al., 2015), fishers' behaviour was simultaneously recorded by a single on-board observer for 15 and 49 of these boats and tracks, respectively (Table 1). The observer recorded the nature, time, and duration of all fishing and non-fishing activities in a logbook using a watch synchronized with GPS time. He then combined this data with the corresponding GPS tracks to determine typical fishing trajectory patterns per gear type (see Section 3.1). Such typical, gear-specific track patterns showed that GPS data was acquired at appropriate time resolution for monitoring the fishers' behaviour. They further allowed the observer to cautiously pre-classify the 257 unseen GPS tracks into fishing and non-fishing activities at a high accuracy using a geographical information system (GIS), which validated the ground-truth approach and the analytical outputs of this survey.

Travel between landing sites and fishing spots was also categorized for each track. Average pre-classification time of the total 306 fishing tracks ranged between 30 and 90 min per track depending on trip duration and gear use.

GPS data was uploaded into a PostgreSQL/PostGIS relational database in order to facilitate data filtering and processing. Because GPS trackers were turned on prior to fishers' departure at sea, GPS boat positions that were located inland were removed using a land mask (Fig. 1).

2.2. Data analysis

Data were cleaned of obvious GPS errors and spatial inconsistencies. Because GPS trackers recorded boat position at varying time intervals due to satellite signal reception, standardization of boat trajectories was required prior processing data. A dataset of reconstructed GPS positions was generated by linearly inferring boat positions at a regular 60 s time interval using the adehabitatLT R package (Calenge, 2006).

2.2.1. Analysis via speed threshold method

Boat GPS positions of each recorded track were classified into fishing and non-fishing events using a speed threshold. Since preliminary information on boat's speed was not available in the fishery surveyed, this threshold was estimated using our empirical data. Boat speed was calculated at each GPS position by dividing the distance by the time interval between two consecutive boat positions (i.e. 60 s). Using the whole dataset, the speed threshold was then determined as that speed that corresponded to the best trade-off between sensitivity and specificity, defined as the rate of true positives (i.e. proportion of fishing events that were correctly predicted as such) and true negatives (i.e. proportion of non-fishing events that were correctly predicted as such), respectively. The following speed thresholds were found and used for processing all boat tracks per gear type: beach seine (0.54 km.h^{-1}), mosquito trawl net (0.84 km.h^{-1}), gillnet (0.63 km.h^{-1}), handline (0.48 km.h^{-1}), and speargun (1.62 km.h^{-1}). Boat GPS positions of each gear type were therefore classified as 'fishing' if boat speed was lower than the corresponding speed limit, and 'not-fishing' where boat speed exceeded that threshold.

2.2.2. Analysis via random forest algorithm

The 306 pre-classified fishing tracks were used as a ground-truthed GPS dataset to train a random forest algorithm in a cross-validation setting (Boehmke and Greenwell, 2019). A total of 24 covariates were calculated at each boat GPS position to account for the local track geometry. Those covariates described i) the area of the polygon obtained by connecting 5–10 consecutive positions (Conv), ii) the latter covariates divided by the square of the polygons' perimeter (ConvP), iii) the number of GPS positions within a circle of a 20–100 m radius (Circle), iv) the angle between three consecutive positions (Angle), and v) the average speed between 5 and 10 consecutive positions (Speed) (Table 2). The choice of these covariates was derived from the boat

Table 2
Description of the 24 covariates that were calculated at each boat GPS position and used by the random forest algorithm.

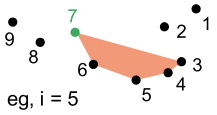
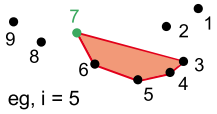
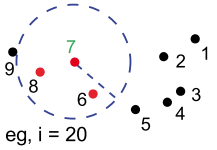
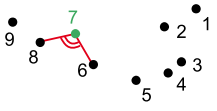
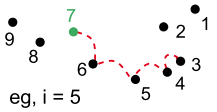
Covariates	Number of covariates		
Conv(X)	Area (in m^2) of the convex hull of the polygon obtained by connecting the positions X, X − 1, …, and X − i + 1	6 i∈[5,6,7,8,9,10]	
ConvP(X)	Conv(X) divided by the square of the polygon's perimeter	6 i∈[5,6,7,8,9,10]	
Circle(X)	Number of positions within an i-meter radius circle centred at position X	5 i∈ [20,40,60,80,100]	
Angle(X)	Angle (in degrees) between positions X -1, X, and X + 1	1	
Speed(X)	Average speed over ground (in m/s) computed using distance and time between positions X and X − i + 1	6 i∈[5,6,7,8,9,10]	

Table 1
Performance measures of the speed and random forest models per gear type based on GPS data records and pre-classified fishing tracks. AC, SE, and SP stand for accuracy, sensitivity, and specificity. The highest scores of performance criteria between the two methods are underlined.

Gear types	GPS data records (with on-board observation)		Pre-classified fishing tracks			Performance criteria					
	Boats	Tracks	Total fishing events	Mean trip duration (h)	Time spent fishing (%)	Speed threshold			Random forest		
						AC	SE	SP	AC	SE	SP
Beach seine	5 (4)	30 (10)	6036	6.07	55.3%	0.66	0.65	0.67	0.74	0.81	0.66
Gillnet	4 (3)	62 (9)	15278	6.40	64.2%	0.86	0.86	0.85	0.88	0.94	0.78
Handline	2 (2)	34 (9)	8520	5.95	70.1%	0.89	0.88	0.89	0.89	0.97	0.71
Speargun	12 (3)	72 (10)	16690	6.54	59.1%	0.85	0.85	0.85	0.89	0.92	0.85
Mosquito trawl net	8 (3)	108 (11)	17772	4.79	57.2%	0.77	0.77	0.77	0.88	0.95	0.80

trajectory geometric patterns observed among the pre-classified fishing tracks (see Section 3.1) and obviously impacted models' outputs. While the location of gear use was likely to be linked to geographical and environmental conditions (e.g. marine habitat, depth, and distance to shore), only GPS data was incorporated in the learning algorithms to enhance the general applicability of the models.

These covariates were then used as features in a random forest algorithm. The random forest hyperparameters were optimized as follows: the Gini impurity index was used as the node splitting criteria while keeping at least 10 GPS positions at each node of the decision trees; the number of trees (200, 500, 1000, and 1500) and the random subsets of covariates to possibly split at each node (2, 4, 6, 8, 10, and 12) were determined to maximize model performance using the area-under-the ROC curve (AUC). A cross-validation procedure was followed to prevent over-fitting of the model. For each gear type, that procedure consisted of i) selecting the pre-classified tracks of all except one fisher as a learning dataset, ii) training a random forest model and using that model to predict the probability of a fishing event at each boat GPS position of the remaining fisher's tracks, iii) repeating the previous step so as to perform model prediction for each fisher, and iv) producing a global confusion matrix of model predictions. The hard classification of GPS positions into fishing and non-fishing events was then obtained by optimizing the trade-off between sensitivity and specificity of the predictions.

2.2.3. Performance of analytical methods

Three types of performance measures of fishing effort estimation were used to compare both analytical methods to account for the diversity of objectives, needs, and constraints of fishery monitoring programmes according to management context. First, the outputs of each method were assessed through a confusion matrix that was calculated on the basis of three quantitative performance criteria: accuracy (i.e. proportion of correct predictions of fishing and non-fishing events), sensitivity (true positive rate, i.e. proportion of fishing events that were correctly identified as such), and specificity (true negative rate, i.e. proportion of non-fishing events that were correctly identified as such). Criterion value ranged from 0 to 1, values close to 1 indicated high score of the corresponding criteria.

Second, the methods' spatial sensitivity was explored through a mapping approach. For each method and gear type, spatial sensitivity was calculated using a $0.005^\circ \times 0.005^\circ$ cell grid as the total number of true positives in each cell divided by the observer's counts of fishing events in that cell. Grid cell size was defined to allow for depicting reef geomorphology in the survey area. Those cells whose spatial sensitivity was lower than 80% were mapped and compared between methods.

Third, fishing effort per trip (in hours) was estimated as the product of the predicted number of fishing events of the corresponding track by 60 s (i.e. standardized time interval between two consecutive GPS positions). Fishing effort estimates per trip were then compared between methods per gear type using boxplot and plot representations. The mean difference in fishing effort between observer's measures and predictions per gear type was calculated and compared between methods using kernel density plots.

Data analysis was performed using R caret (Kuhn et al., 2018) and ranger (Wright and Ziegler, 2017) packages. A sample of track data and a methodological brief of the R code is available at https://doi.org/10.23708/IBZJGD.

3. Results and discussion

3.1. Definition of fishing activity by gear type

While diverse gear types were reported overall in the fishery, fishers mostly used one single gear type for any single trip. Each trip of the 31 participant fishers was then assigned to a specific gear type following fishers' declaration.

Boat speed varied according to fishing factors (e.g. gear use, boat size and equipment, number of fishers on-board, distance to fishing spot) as well as non-fishing factors (e.g. wind strength and direction, tidal current). Fishing and non-fishing activities were then characterized as follows. The shape of boats' tracks varied according to fishing activity (i.e. for setting, operating, and hauling gear) and therefore varied across gear types (Fig. 2). Specifically, GPS patterns corresponding to fishing hauls were characterized by large, dense, and well-defined curved track segments for beach seine use (Fig. 2A), thin loops or irregular segments associated with dense groupings for gillnet use (Fig. 2B), irregular groupings corresponding to boat anchoring locations for handline use (Fig. 2C), and sinuous, dense segments for speargun and mosquito trawl net uses (Fig. 2D, E).

In contrast, non-fishing activities were broadly similar across gear types. They consisted in i) travel from and to landing sites and fishing spots over <1 km to several km distance sailing or paddling, and ii) boat operation (e.g. setting or folding sail). Short incidental non-fishing events caused by sea conditions (e.g. tidal currents) and unpredictable, both environmental and human factors were also occasionally detected. Corresponding GPS patterns consistently displayed thin track segments with varied between-point distance and irregular groupings (Fig. 2A–E).

Typical gear-specific trajectory patterns were reported in previous surveys using geolocation devices in small-scale and large-scale fisheries (e.g. Joo et al., 2013; Alvard et al., 2015; de Souza et al., 2016). Recording boat position at 60 s time interval was found appropriate to depict small-scale fishers' activities in a reef environment such as our case study.

3.2. Performance and outputs of analytical methods

The predictions of both the boat speed threshold method and the random forest model had high to very high performance criteria scores for four gear types (gillnet, speargun, handline, and mosquito trawl net), though in most cases the random forest model performed better (Table 1). However, results highlighted that low-speed and high-speed movements of beach seine fishers were far from systematically associated with fishing and non-fishing events, respectively, making boat speed a poor predictor of fishing activity for that gear type. There was no evidence of linkage between the random forest model's performance and sample size across gear types (Table 1), suggesting that the dataset was large enough across each gear type for the random forest model to be effective in our case study.

The spatial distribution of the prediction of fishing events further showed that the random forest model, more consistently detected fishing events over the fishing area for all gear types, while the speed threshold method yielded very high spatial sensitivity for gillnet and handline only (Fig. 3). The number of cells with sensitivity <80% was indeed noticeably lower for the random forest model than for the speed threshold method for speargun (−14 cells, 19.7%), mosquito trawl net (−7 cells, 17.5%), and beach seine (−8 cells, 44.4%), showing that the latter was less effective in tracking high resolution fishing practises than the former for those gear types (Fig. 3).

Following the observer's measurements, mean trip duration ranged between 4.8 h and 6.5 h across gear types and fishers spent on average 2.8 h–4.2 h per trip (55.3%–70.1% of trip duration) while fishing (Table 1, Fig. 4A). Predictions of mean fishing effort per trip showed that both methods reached from 89.4% to 97% accuracy across gear types. However, difference in fishing effort estimations were observed between methods. The speed threshold method slightly underestimated that indicator by 3.0%–7.2% (Fig. 4A, B). Specifically the mean difference in fishing effort per trip between the observer's measurements and speed threshold method's predictions ranged from +0.09 h (+3.3%) to +0.13 h (+3.4%) for handline, speargun, and mosquito trawl net and reached as high as +0.24 h (+7.2%) and +0.27 h (+6.6%) for beach seine and gillnet. On the contrary, the random forest model overestimated fishing

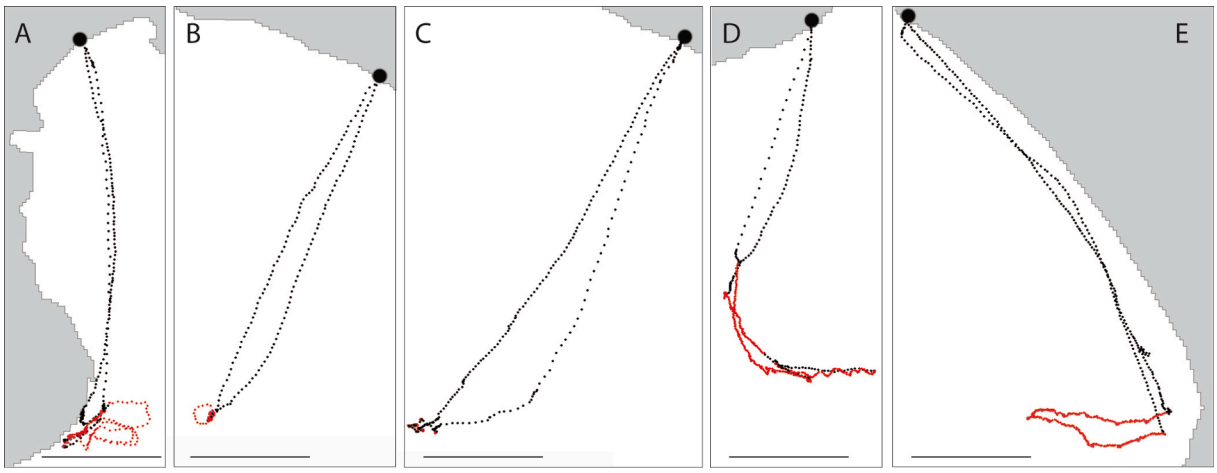


Fig. 2. Presentation of raw GPS tracks for five small-scale boats using different fishing gear types and representing potential fishers' behaviour for beach seine (A), gillnet (B), handline (Cc), speargun (D), and mosquito trawl net (E). Dots represent individual observed GPS positions. Red and black dots represent observed fishing and non-fishing events, respectively. The scale bar represents 1 km. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

effort per trip to a varied extent while yielding comparatively lower accuracy for handline, mosquito trawl net, and beach net (Fig. 4A, B). Specifically the range in mean difference in fishing effort per trip between the observer's measurements and random forest model's predictions varied from −0.42 h (−10.4%) to −0.26 h (−6.3%) for handline, beach seine, mosquito trawl net, and gillnet, and reached as low as −0.12 h (−3.1%) for speargun. The precision of fishing effort estimate was similar between the two methods and was higher for gillnet, handline, speargun, and mosquito trawl net than for beach seine (Fig. 4C).

Our results show that the perceived performance of the random forest model and the speed threshold method varied according to whether overall or individual fishing effort events were considered. Indeed, because the fishing effort estimate is the product of true positives, false positives (that misplace and overpredict fishing events), and false negatives (that misplace and overpredict non-fishing events), the trade-off between accuracy, specificity, and sensitivity may lead to over-optimistic interpretation of the reliability of speed threshold method predictions of fishing activity, as showed for four gear types in this survey. For instance, the relative error of speed threshold method's estimate of mean fishing effort per trip of mosquito trawl net users (−3.3% on average) was lower than that of the random forest model (+10.6% on average, Fig. 4A, B) despite the fact that performance scores and spatial sensitivity of the latter were much higher than those of the former for that gear type (Table 1, Fig. 3). To some extent, the perceived overall accuracy of the speed threshold method therefore occurred by chance. Setting the speed threshold is a difficult decision which strongly influences the prediction results and requires to balance the trade-offs between accuracy, specificity, and sensitivity. One should therefore not exaggerate the effectiveness of the speed threshold method for fitting the diversity of gear types and its uses in small-scale fisheries.

Difference in performance and outputs between both methods challenges available literature that considers boat speed a reliable, fine-scale spatial predictor of small-scale fishers' behaviour and, consequently, of fishing activity (e.g. Burgos et al., 2013; Alvard et al., 2015; Navarrete Forero et al., 2017). This difference likely derives from the different characteristics of the gear types and of the fisheries surveyed. Indeed, the above-mentioned case studies focused on hook fisheries (i.e.

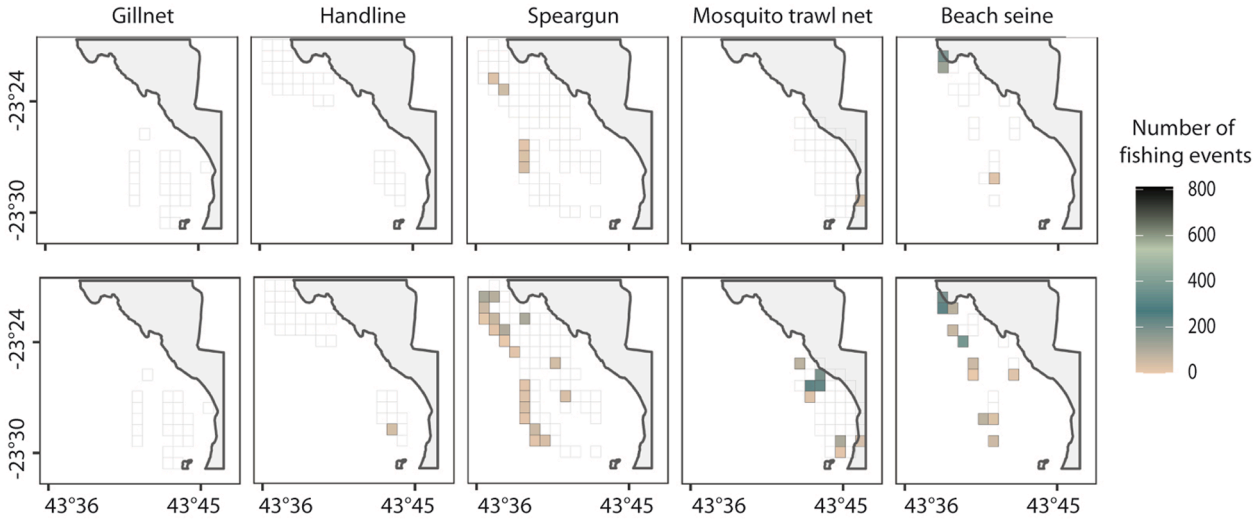


Fig. 3. Maps of spatial sensitivity of the random forest model (A) and the speed threshold method (B) for each gear type using a 0.005° × 0.005° (555 m × 555 m) cell grid. Only cells with sensitivity <80% are coloured for better clarity (see Section 2 for details). The colour gradient shows the difference in the number of fishing events between the observer's measurements and the true positives predicted by both methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

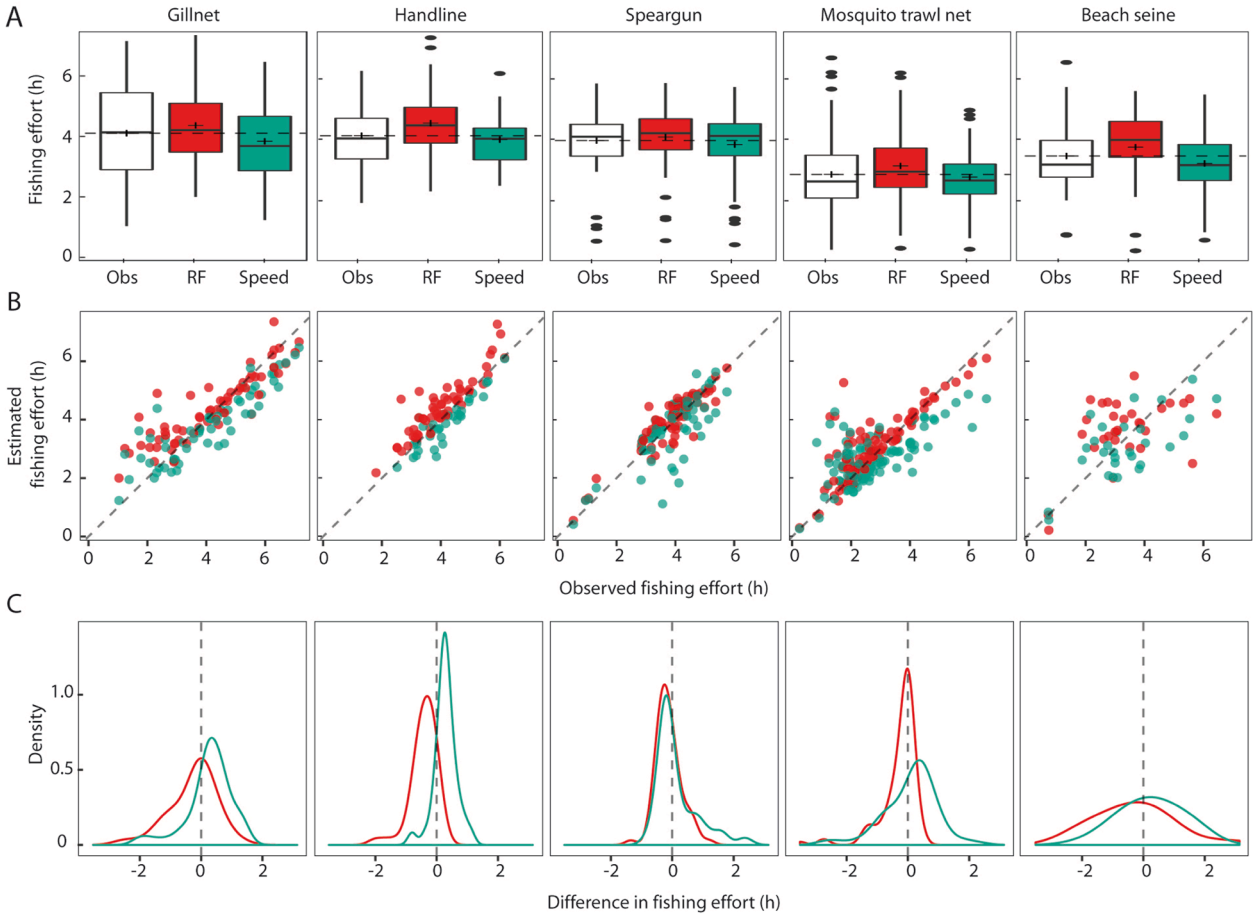


Fig. 4. Fishing effort per trip (h) per gear type. The boxplots (A) show the median (solid line), mean (black cross), first to third quartiles, and outliers of the observer's measurements (Obs) and the predictions of the random forest model (RF) and the speed threshold method (Speed). Means of the observer's measurements are showed for comparison (dotted line). The diagrams (B) display the predictions of fishing effort of the random forest model (red dots) and the speed threshold method (green dots) as compared to the observer's measurements for each individual fishing trip. The x = y function is showed (dotted line). The kernel density plots (C) show the dispersion of the difference in fishing effort (h) between the observer's measurements and the predictions of the random forest model (red line) and the speed threshold method (green line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hand line, trolling line, bottom and vertical longlines) that were operated by motorized boats that were obviously less affected by non-fishing factors (e.g. wind's direction and strength, tidal current) during fishing trips than the sailing boats monitored in the present survey. Moreover, our investigation covered a larger diversity of gear types and corresponding spatial use patterns.

Our results suggest that selecting which prediction method is appropriate for fishery management would depend on local gear use, relevant fishing effort indicators, and available analytical expertise. The random forest learning algorithm provided consistent outputs across gear types and was found to be a more generic, although more complex, analytical method of small-scale boat tracking data. Overall that model yielded higher reliability of spatially-explicit fishing effort indicator than that of the speed threshold for four out of five gear types. In addition, the cross-validation procedure allows for estimating the performance of predictions, which is a major advantage of machine-learning methods such as random forests over the speed threshold methods. Preliminary boat tracking data coupled with on-board observation would be required for calibrating both methods to local fishery context as described in the above method section.

In a small-scale fishery management setting, the GPS track-based analytical method provided precise and accurate fishing effort estimates (in hours per trip) that have not been achieved through alternative fishery monitoring methods. For instance, fisher map-based interview survey methods have become increasingly popular during the last 15 years as a cost-effective solution to data deficiency in small-scale fisheries (Close and Brent Hall, 2006; Léopold et al., 2014; Gill et al., 2019). However, one limitation of that method comes from the fact that it commonly uses the unit "fishing trip" as a measure of fishing effort. Indeed, that unit cannot capture change in trip duration and time effectively spent fishing, that are two relevant indicators of fishing pressure. Furthermore, as showed by other authors (Harley et al., 2001; McCluskey and Lewison, 2008), the unit "fishing trip" enhances hyperstability of the relationship between resource abundance and widely-used fishery-dependent abundance indices such as CPUE, which hinders detection of change in resource status and potential effects of fishing. Our results show that boat tracking and GPS track data processing through a random forest model (or a speed threshold method in certain conditions as described above) would overcome those limitations. Moreover, by locating fishing events per trip, boat GPS tracking surveys that follow appropriate sampling design and spatial temporal coverage of fishers' activity may allow for inferring distribution of fishing effort (and corresponding catches if need be) at fine spatial resolution (e.g. Burgos et al., 2013). Such mapping is required for spatial management of small-scale fisheries.

4. Conclusion

We have compared the performance of a speed threshold method and an analytical learning algorithm for estimating and mapping fishing

effort per trip in a small-scale reef fishery in Madagascar using boat GPS trackers. Overall, the random forest model was found to reliably and precisely detect fishing events and non-fishing events of boat tracks regardless gear type (i.e. beach seine, mosquito trawl net, gillnet, handline, and speargun), which makes it a powerful and generic tool for GPS track-based survey analysis in small-scale fisheries. Nevertheless, if analytical expertise precludes the use of such a sophisticated analytical procedure, using a speed threshold method may be suitable for those gear types that induce marked change in boat speed between fishing and non-fishing activities. To our knowledge, no previous study has compared machine learning and speed-based methods for processing boat trajectory data to assess the spatial distribution and intensity of fishing in small-scale fisheries.

The participation of fishers in the survey allowed for optimizing GPS data collection, learning procedures, and consecutive model's predictions of fishing effort per trip. By incorporating fishers' knowledge in fishery research (Stephenson et al., 2016) and improving the accuracy and precision of fishing effort indicator both quantitatively and spatially, this study suggests that boat tracking combined with appropriate analytical procedure may relevantly contribute to small-scale fishery management.

CRediT authorship contribution statement

Faustinato Behivoke: Methodology, Investigation, Resources, Data curation, Writing - original draft, Supervision, Project administration, Funding acquisition. **Marie-Pierre Etienne:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision. **Jérôme Guitton:** Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization. **Roddy Michel Randriatsara:** Methodology, Resources, Data curation. **Eulalie Ranaivoson:** Supervision, Project administration. **Marc Léopold:** Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgements

This study was conducted as part of the IH.SM–IRD research partnership program about Aquaculture and management of Coral ecosystems in Madagascar (JEAI ACOM). Funding was provided by the Laboratory of Excellence "Corail" (GEOFIA project, grant N°2AO2017), France and the Critical Ecosystem Partnership Fund (CEPF-IH.SM, grant N°66341). We thank the participant fishers who actively contributed to the implementation of this survey. We thank Abdoulrahman Farah, Jean-Eudes Hollebecq, and Mathis Rouzoul for their exploration of machine learning models. We are grateful to Elizabeth Hodson for proofreading the original manuscript. We also would like to thank the anonymous reviewers who suggested significant improvements to the original manuscript.

References

Alvard, M., Carlson, D., McGaffey, E., 2015. Using a partial sum method and GPS tracking data to identify area restricted search by artisanal fishers at moored fish aggregating devices in the Commonwealth of Dominica. PLoS One 10. https://doi.org/10.1371/journal.pone.0115552.

Begossi, A., 2001. Mapping spots: fishing areas or territories among islanders of the Atlantic Forest (Brazil). Reg. Environ. Change 2, 1–12. https://doi.org/10.1007/s101130100022.

Boehmke, B., Greenwell, B.M., 2019. Hands-on Machine Learning with R. CRC Press.

Bruggemann, J. Henrich, Rodier, Martine, Guillaume, Mireille M.M., Andréfouët, Serge, et al., 2012. Wicked social–ecological problems forcing unprecedented change on the latitudinal margins of coral reefs: the case of Southwest Madagascar. Ecol. Soc. 17 (4) https://doi.org/10.5751/ES-05300-170447.

Burgos, C., Gil, J., del Olmo, L.A., 2013. The Spanish blackspot seabream (Pagellus bogaraveo) fishery in the Strait of Gibraltar: spatial distribution and fishing effort derived from a small-scale GPRS/GSM based fisheries vessel monitoring system. Aquat. Living Resour. 26, 399–407. https://doi.org/10.1051/alr/2013068.

Calenge, C., 2006. The package "adehabitat" for the R software: a tool for the analysis of space and habitat use by animals. Ecol. Model. 197, 516–519. https://doi.org/10.1016/j.ecolmodel.2006.03.017.

Close, C.H., Brent Hall, G., 2006. A GIS-based protocol for the collection and use of local knowledge in fisheries management planning. J. Environ. Manage. 78, 341–352. https://doi.org/10.1016/j.jenvman.2005.04.027.

Daw, T.M., 2008. Spatial distribution of effort by artisanal fishers: exploring economic factors affecting the lobster fisheries of the Corn Islands, Nicaragua. Fish. Res. 90, 17–25. https://doi.org/10.1016/j.fishres.2007.09.027.

de Souza, E.N., Boerder, K., Matwin, S., Worm, B., 2016. Improving fishing pattern detection from satellite AIS using data mining and machine learning. PLoS One 11. https://doi.org/10.1371/journal.pone.0158248.

Gill, D.A., Oxenford, H.A., Turner, R.A., Schuhmann, P.W., 2019. Making the most of data-poor fisheries: low cost mapping of small island fisheries to inform policy. Mar. Policy 101, 198–207. https://doi.org/10.1016/j.marpol.2017.10.040.

Gloaguen, P., Mahévas, S., Rivot, E., Woillez, M., Guitton, J., Vermard, Y., Etienne, M.P., 2015. An autoregressive model to describe fishing vessel movement and activity: an autoregressive model to describe fishing vessel movement and activity. Environmetrics 26, 17–28. https://doi.org/10.1002/env.2319.

Harley, S.J., Myers, R.A., Dunn, A., 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. 58, 1760–1772. https://doi.org/10.1139/cjfas-58-9-1760.

ICES, 2019. Working Group on Fishing Technology and Fish Behaviour (WGFTFB). ICES Scientific Reports 1:61. 363pp. http://doi.org/10.17895/ices.pub.5592.

Joo, R., Bertrand, S., Tam, J., Fablet, R., 2013. Hidden Markov models: the best models for forager movements? PLoS One 8. https://doi.org/10.1371/journal.pone.0071246.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., R Core Team, 2018. Package 'caret': Classification and Regression Training. R Package Version 6.0-81.

Le Guyader, D., Ray, C., Gourmelon, F., Brosset, D., 2017. Defining high-resolution dredge fishing grounds with Automatic Identification System (AIS) data. Aquat. Living Resour. 30, 39. https://doi.org/10.1051/alr/2017038.

Lee, J., South, A.B., Jennings, S., 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. ICES J. Mar. Sci. 67, 1260–1271. https://doi.org/10.1093/icesjms/fsq010.

Léopold, M., Guillemot, N., Rocklin, D., Chen, C., 2014. A framework for mapping small-scale coastal fisheries using fishers' knowledge. ICES J. Mar. Sci. 71, 1781–1792. https://doi.org/10.1093/icesjms/fst204.

McCluskey, S.M., Lewison, R.L., 2008. Quantifying fishing effort: a synthesis of current methods and their applications. Fish Fish. 9, 188–200. https://doi.org/10.1111/j.1467-2979.2008.00283.x.

Navarrete Forero, G., Miñarro, S., Mildenberger, T.K., Breckwoldt, A., Sudirman, Reuter, H., 2017. Participatory boat tracking reveals spatial fishing patterns in an Indonesian artisanal fishery. Front. Mar. Sci. 4, 409. https://doi.org/10.3389/fmars.2017.00409.

Peel, D., Good, N.M., 2011. A hidden Markov model approach for determining vessel activity from vessel monitoring system data. Can. J. Fish. Aquat. Sci. 68, 1252–1264. https://doi.org/10.1139/f2011-055.

Russo, T., Parisi, A., Prorgi, M., Boccoli, F., Cignini, I., Tordoni, M., Cataudella, S., 2011. When behaviour reveals activity: assigning fishing effort to métiers based on VMS data using artificial neural networks. Fish. Res. 111, 53–64. https://doi.org/10.1016/j.fishres.2011.06.011.

Stelzenmüller, V., Maynou, F., Martín, P., 2007. Spatial assessment of benefits of a coastal Mediterranean marine protected area. Biol. Conserv. 136, 571–583. https://doi.org/10.1016/j.biocon.2007.01.002.

Stephenson, R.L., Paul, S., Pastoors, M.A., Kraan, M., Holm, P., Wiber, M., Mackinson, S., Dankel, D.J., Brooks, K., Benson, A., 2016. Integrating fishers' knowledge research in science and management. ICES J. Mar. Sci. 73, 1459–1465. https://doi.org/10.1093/icesjms/fsw025.

Stewart, K.R., Lewison, R.L., Dunn, D.C., Bjorkland, R.H., Kelez, S., Halpin, P.N., Crowder, L.B., 2010. Characterizing fishing effort and spatial extent of coastal fisheries. PLoS One 5. https://doi.org/10.1371/journal.pone.0014451.

Vermard, Y., Rivot, E., Mahévas, S., Marchal, P., Gascuel, D., 2010. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. Ecol. Model. 221, 1757–1769. https://doi.org/10.1016/j.ecolmodel.2010.04.005.

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. https://doi.org/10.1016/j.ecolmodel.2010.05.007.

Wilen, J.E., 2004. Spatial management of fisheries. Mar. Resour. Econ. 19, 7–19. https://doi.org/10.1086/mre.19.1.42629416.

Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77, 1–17. https://doi.org/10.18637/jss.v077.i01.