PeskAAS: A near-real-time, open-source monitoring and analytics system for small-scale fisheries

Alexander Tilley[a]*, Shaun P. Wilkinson[bc]

[a]WorldFish, Jalan Batu Maung, Batu Maung, 11960 Bayan Lepas, Penang, Malaysia

[b]School of Biological Sciences, Victoria University of Wellington. P.O. Box 600, Wellington, New Zealand

[c]Wilderlab, Wellington, New Zealand

*Corresponding author
E-mail: alex.tilley@gmail.com

Abstract

Small-scale fisheries are responsible for half of the global fish catch, and fisheries-dependent data underpin scientific guidance of management and conservation, yet we have almost no data on their activities and production over time and space. We use open source software components including the *Shiny* R package to build `peskAAS`, an adaptable and scalable digital workflow to collate, classify, integrate, analyse and visualise small-scale fisheries data. The main features of `peskAAS` are: (i) it is fully open-source and free to deploy and use (ii) it is component-based to be flexible for integration of vessel tracking data and existing catch documentation elements; (iii) it provides spatial and temporal filtering of relative fishing productivity by fishing method and habitat (iv) it integrates *FishBase* open-source and regularly updated species-based growth parameters (v) the end user click-button dashboard is specifically co-designed for easy interpretation of fisheries production data. With limited training and code adaptation, the `peskAAS` workflow can be deployed for use by fisheries stakeholders, managers and scientists as a platform for high resolution, near-real-time production data from small-scale fisheries for the first time.

INTRODUCTION

Approximately half of the global catch of fish is landed in small-scale fisheries (SSF), yet the ability for science to guide the sustainable exploitation of these resources is inhibited by a massive data gap. There are 40 million people actively fishing in inland and coastal systems worldwide, yet due to the challenges involved in collecting data in such dispersed, informal and diverse contexts of SSF, the contributions of this fish food system to livelihoods and food security are hidden and undersold [1]. As a result, fish has been largely absent in the development of food-based approaches for greater food security and nutrition [2]. However, recent research shows that nutrient quality of fish is determined by species composition not quantity [3], so for many countries whose populations suffer from inadequate nutrient intake, fish-based food strategies could be an effective strategy.

To improve fisheries management, and realise the nutritional and food security potential of fish we require data on fish production over time and space, but most small-scale fishing occurs in low-income countries that lack the resources and capacity to collect, store or analyse these data [4]. Even with adequate data, fisheries management relies on statistical models and analytics that were developed for single species fisheries, rather than biodiverse tropical reef fisheries [5]. Many valuable tools have been developed for single species stocks (see [6], but very few exist to assist in the analysis and management of mixed species fisheries. This leads to a situation where management and policies are built on poor government estimates or reconstructed trade and consumption statistics which are often not relevant at the local scale, and hence are highly susceptible to governance and management failures [7].

Over the past two decades the proliferation of information and communication technologies (ICTs) has revolutionised the collection, communication and storage of data in the industrial fishing sector, as well as more broadly in agriculture [8] through design and development of mobile smartphone applications and digital survey forms [9]. There are some examples of applications and tools to collect fisheries data on a small scale, the most successful of which are generally 'high touch', meaning they involve significant contextual development that cannot easily be scaled to other systems or geographies (see [10–13]. The danger in developing scalable technologies is that they are often imposed as prescribed 'solutions' on low-income countries, and can merely reinforce the capacity gap, alienate managers and stakeholders, and be ill-suited to the contextual reality [14,15]. There is an urgent need for a light touch, scalable and integrated approach to data collection and analytics, but thus far this has been elusive in terms of getting simple, usable data in the hands of fisheries managers.

Here, we describe a new digital pipeline called `peskAAS` - a pseudo-acronym for fisheries (*peskas*) in the national language of Timor-Leste, Tetum, combined with Automated Analytics System. `PeskAAS` connects open source programs to collect, communicate, analyse and visualise SSF movement and catch data. This application was designed to test how scripted analytics can assist fisheries management by generating digestible, summarised information rapidly for decision-making in data-deficient and low capacity systems. We provide examples to illustrate the types of analytics and machine learning that the system can be used to carry out. Our test site of Timor-Leste was chosen due to the very limited level of fisheries development, the paucity of information, and the limited fisheries governance capacity.

`peskAAS` OVERVIEW

PeskAAS is an interactive web-hosted R *Shiny* application developed to facilitate data exploration and decision-making processes in small-scale fisheries. This application accesses the 'peskaDAT' database in real-time using the `DBI` and `RMySQL` R packages [16,17], and pulls landing records into a web-based interactive R session (hosted by https://www.shinyapps.io), from which users can apply dynamic location, gear, habitat and date-range filters to create informative plots for catch per unit effort (CPUE), total catch, species composition and total national catch estimates (Figure 1).

Figure 1. Screenshot from the interactive PeskAAS Shiny platform developed to visualise near-real time fish landings data in Timor-Leste. https://worldfish.shinyapps.io/peskAAS/

The user interface incorporates tick-box widgets for dynamic selection of sites, fishing habitats and gear types, and a slider widget is included for users to fit splines to the CPUE data and dynamically adjust the smoothing parameter to aid in visualization. Users also have an option to download a month-aggregated summary table in .csv format for further analysis, using a download-button widget.

The application is hosted remotely by shinyapps.io (https://www.shinyapps.io/) on the free pricing tier, which currently allows 25 active hours of use per month. Users can also run the application locally, by cloning the repository from GitHub and launching the application from a local R session, or by installing RStudio Shiny Server (https://www.rstudio.com/products/shiny/shiny-server/) and hosting the application on a local machine.

The digital analysis pipeline consists of a network of freely available components for data collection, storage, manipulation and analysis, catering for humanitarian organizations with low usage requirements (< 25 active hours per month) (Figure 2). The pipeline is scalable for higher usage levels with modest subscription levies for KoBo toolbox, Shiny (https://www.shinyapps.io), Heliohost (https://www.heliohost.org) and Google Cloud Platform (GCP; https://cloud.google.com).

Figure 2. Automated digital analytics pipeline for Timor-Leste fisheries catch data

Catch documentation at landing sites

Fisheries catch data were collected on 3G-enabled tablets using a digital survey form developed in *KoBo toolbox* (http://www.**kobotoolbox**.org/), by locally-employed data collectors at 25 key landing sites across 11 coastal districts in Timor-Leste (figure 3). One collector was hired in each coastal municipality. Atauro Island to the north of Dili falls within the municipality of Dili, and is a disproportionately important area for national fisheries with a high population of fishers so 3 collectors record landings across 12 small landing sites. Collectors interview individual fishers on a daily basis to collect data on transport type (boat type/shore), gear type, habitat type where fishing took place (reef, fish aggregating device, deep, shore or mangrove), number of fishers (men, women and children), trip duration, species or species group[1], size and number caught by species or species group, and sale price if applicable. The landings form evolved through time to integrate a growing list of fishers and identified species, and in response to various feedback mechanisms.

Figure 3. Map of Timor-Leste illustrating landing sites where small-scale fisheries catch was recorded using tablets.

peskaDAT database

[1] All except the most important species landed in the fishery were grouped by ISSCAAP groups within families, (FAO International Standard Statistical Classification of Aquatic Animals and Plants), or by family (see Supplementary information 2).

We developed a cloud-based MySQL database of filtered landings records called 'peskaDAT', to host cleaned and checked fisheries landings data, geo-located boat tracks and ancillary tables such as species, boat, gear and habitat information in an easily accessible and mutable format (Supplementary information 1). This database is hosted by Heliohost, a free hosting platform with several MySQL relational database servers. A monthly database backup is created and stored locally in `.rds` format (an R binary data file) in a Dropbox folder to ensure data perpetuity and backwards compatibility.

peskaPARSE.R script

To automatically access landing records from KoBo toolbox, and filter, manipulate and submit the cleaned records to the peskaDAT database on a regular (daily) basis, we developed an R script called 'peskaPARSE.R' (supplementary file 2) that is scheduled on a daily cron job (a command to a server to be executed at a specified time) run on a GCP virtual machine. The script first accesses the KoBo toolbox API (application programming interface - that allows access to the data) and pulls the new catch records into the R environment. The catch weight for each new record is estimated using von Bertalanffy growth function (VBGF) parameters obtained from FishBase (www.fishbase.org) using the rfishbase R package [18]. Several filters are applied, including species-specific length, number of individuals, weight and price boundary checks, formatting error checks, and invalid gear/habitat combination checks. Suspect entries are flagged for curation by a moderator, and their values are automatically suppressed from contributing to downstream applications (including the `peskAAS` data exploration tool described below) until records are amended in the KoBo toolbox dashboard. This is achieved *via* prompt follow-up clarification with field-based observers, to ascertain whether data entry errors were made during form submission or whether the record in question is a genuine outlier.

To automatically run the peskaPARSE.R script on a daily basis, we deployed a free Ubuntu 18.04 minimal virtual machine (VM) instance on the Google Cloud Platform (GCP), transferred the script to a convenient user directory (specified by the [address] argument below), and scheduled a 24-hourly cron job by appending the following line to the /etc/crontab file:

```
0 0 * * * root Rscript -e  'source("[address]")'
```

where `[address]` is replaced by the web location of the R script. In our case we stored the script in a publicly available dropbox directory and modified the public link to terminate with `download=1` to enable direct sourcing. See the GitHub repository at https://github.com/shaunpwilkinson/peskAAS for further details. For the script to successfully import the raw trip- and species-landing records from the KoBo and Pelagic Data Systems APIs, and deposit the filtered records in the peskaDAT MySQL database, three authorization files must be stored in the same directory as the peskaPARSE.R file: a single-line text file named KoBo-auth.txt containing the KoBo toolbox username and password, delimited by a colon operator, and similarly-formatted files containing log-on credentials for the Pelagic Data Systems dashboard (PDS-auth.txt) and the peskaDAT MySQL database (peskaDB-auth.txt; user with SELECT privileges). The regular gleaning and filtering of new landings records ensures that the data available for analysis and visualization are timely and trustworthy.

Geospatial vessel tracking

Shore-based data collectors are, of course, unable to sample catch from all fishing trips exhaustively, so to extrapolate fish production from individual trips to community, municipal and national levels, and to visualise geospatial fishing effort, we installed tamperproof solar-powered GPS units developed by Pelagic Data Systems Inc. (San Francisco, CA) on to 5-15 boats per landing site (N = 437). Trackers recorded point location data every 5 seconds and communicated those data when in range of a cellular network (Figure 4). Vessel tracks were also linked to the trip's catch data, where available, using the unique GPS unit ID, allowing us to train a model to predict unknown variables for trips with GPS data only, such as gear and habitat type).

Figure 4. The relative effort heat map of small-scale fishing vessel tracks between August 2018 and February 2020 in Timor-Leste from the `peskAAS` dashboard

Fisheries analytics

To quantify relative fishing success over time and space, the catch-per-unit-effort (CPUE) is used as a metric for relative fish abundance, and to track the response of fish stocks to fishing pressure. For CPUE, a standard unit of *effort* is required. For `peskAAS`, we use raw CPUE, with effort standardized into the unit of *fisher-hours* on each fishing trip, calculated as the trip duration in hours multiplied by the number of fishers on board[2].

To generate accurate regional and national catch and CPUE estimates, it is necessary to estimate the frequency and duration of fishing trips for each boat type, known as a vessel activity coefficient (VAC). To generate an accurate VAC accounting for differences in boat type, vessel tracks were used to generate monthly VAC values according to motorized and non-motorized vessels, and were used to calculate the first accurate national scale estimates of fish production in Timor-Leste (equation 1).

The national monthly catch (in tons; C) is estimated as:

$$C = \sum_b CPUE_b \times EPT_b \times VAC_b \times N_b \times 0.001 \qquad \text{(Equation 1)}$$

where b refers to the boat type (canoe or motor-powered), $CPUE$ is the monthly catch per unit effort (averaged across all sites, in Kg per hour), EPT is the median effort per trip (person-hours), VAC is the vessel activity coefficient (average number of trips per month) and N_b is the total number of boats of type b in the national fleet (national fisher and boat census data obtained from the Timor-Leste Ministry of Agriculture and Fisheries).

<u>Supervised prediction of missing trip attributes</u>

The integration of shore-based observer data with high-resolution geo-located vessel fishing tracks from on-board GPS units from *Pelagic Data Systems* Inc. allowed us to validate a supervised

[2] The complexity of more robust CPUE standardization (see [19]) of multiple gear types in tropical, multi-habitat, mixed species fisheries makes it impractical in a livelihoods context, so fisher hours is used as a unit applicable across all fisheries and scales.

classification approach for predicting gear and habitat types in the case where only GPS data are available (i.e. trips that are tracked but not directly observed). For each tracked but unobserved trip, a donor trip was selected from the subset of trips for which both GPS and observer data are available, based on spatial nearest-neighbour analysis. To achieve this, ten temporally-spaced, two-dimensional GPS points from each trip were used to generate training- and query-data matrices for input into the `nn2` function in the `RANN` R package [20]. Donor trips were assigned based on minimum Euclidean distance and were only assigned if the distance was < 0.20. To validate this approach we randomly allocated the linked trip set to 80% training data and 20% query data, and tested the predictive success rate in terms of number of gear and habitat types correctly predicted in the query set. The nearest-neighbour classification approach correctly predicted 81% of gear types and 90% of habitat types based on a 491 observation training set and a 120 observation query set.

RESULTS & DISCUSSION

`peskAAS` was developed concurrently with the co-generation of a national database of SSF landings with the Timor-Leste National Fisheries Directorate. As of February 2020, 45,737 fishing trips have been geotracked, representing 560,000 km travelled over 350,000 hours. Catch data was recorded for 22,375 of these trips, resulting in 34,440 species level landings records representing a total catch weight of 1076 tonnes. This system has already been utilised to test the effectiveness of fish aggregating devices at boosting catch rates by nearshore artisanal fishers [21], and to calculate the first national fisheries production estimates for the Timor-Leste SSF fleet based on landings [22]. In May 2019, the Timorese government adopted and launched this technology as their official national fisheries monitoring system, so on a national level, has passed an important test of legitimacy, and shows potential for sustainability.

`PeskAAS` was designed to be a free, online data exploration tool to enable rapid development of simple, near-real time fisheries monitoring systems to generate accurate and timely fish production data to use in adaptive, evidence-based management of fisheries resources. The tracking hardware and data service from Pelagic Data Systems is an optional extra and represents additional costs to users that must be considered in terms of sustainability for users. However, the ability to add spatial context to fisheries data is paramount, and facilitates coastal zone planning that is increasingly politicised in the development of the blue economy.

Our initial case study from Timor-Leste indicates that this type of solution is very much needed in situations of low management capacity and resources to track dispersed fisheries. The catch form and scripted analytics require no specialised equipment and the dashboard is displayed in a web browser. However, we acknowledge that `peskAAS` has several limitations. Most notably, the scripted nature of the tool limits its use to those with programming expertise, or implies a need to hire help. Therefore, ongoing work to scale up `peskAAS` will develop its modularity and customisation through templates for catch forms, analytics and dashboards displays, to allow for more flexibility to other fish production systems and countries.

The co-design of `peskAAS` with fisheries managers was crucial in building a financially and institutionally sustainable system. Hence, in scaling this tool to other countries and contexts, the need to build local legitimacy and a feeling of ownership among stakeholders will necessitate a process of co-design. The technical capacity in Timor-Leste for digital systems is improving, but as in most low-income country governments the long term maintenance of the system will still require external support. Also, further capacity building is necessary to ensure `peskAAS` outputs are being correctly interpreted to inform policy and guide extension services for fishing communities in Timor-Leste. It remains to be seen if other low-income countries will independently adapt the script to deploy the system for fisheries monitoring elsewhere.

In conclusion, `peskAAS` has already influenced national fisheries policy in Timor-Leste as a result of its inclusion in the newly drafted National Fisheries Strategy and revised Fisheries Decree law. We believe it will be an important resource for government and non-government managers, researchers and students looking for a low cost solution to collect and automate the analysis and visualisation of SSF data over space and time.

Acknowledgements

We are very grateful to the Timorese fishers and community members who collaborated with us and Timor-Leste's Ministry of Agriculture and Fisheries in this project, and particularly to WorldFish team members Joctan Dos Reis Lopes, Mario Gomes, Mario Pereira, Pedro Rodrigues and Agustinha Duarte who have provided important local insights throughout the co-design process. The development of `peskAAS` was funded by the Royal Norwegian Embassy in Jakarta and the CGIAR Big Data Platform Inspire Challenge 2018.

Funding statement

This work was undertaken as part of the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The program is supported by contributors to the CGIAR Trust Fund. Funding support was provided by an Inspire Challenge grant from the CGIAR Research Program on Big Data led by CIAT and IFPRI, and the Fisheries Sector Support Program funded by the Royal Norwegian Embassy in Jakarta.

Author contributions

A.T. conceived and designed the study. S.P.W. coded the various elements of the `peskAAS` pipeline. A.T. and S.P.W. contributed equally to this manuscript.

Data Accessibility

The `peskAAS` catch questionnaire, R script and data used to test the application in this study are archived in Dataverse https://doi.org/10.7910/DVN/TVGIJJ [23]. The `peskAAS` code and documentation is also hosted at https://github.com/shaunpwilkinson/peskAAS.

References

1. Mills DJ, Westlund L, Graaf G de, Kura Y, Willman R, Kelleher K. Under-reported and

undervalued: small-scale fisheries in the developing world. In: Pomeroy RS, Andrew NL, editors. Small-scale fisheries management: frameworks and approaches for the developing world. Wallingford: CABI; 2011. pp. 1–15. doi:10.1079/9781845936075.0001

2. FAO. The state of world fisheries and aquaculture: Meeting the sustainable development goals. Rome: The Food and Agriculture Organisation of the UN (FAO); 2018.

3. Hicks CC, Cohen PJ, Graham NAJ, Nash KL, Allison EH, D'Lima C, et al. Harnessing global fisheries to tackle micronutrient deficiencies. Nature. 2019;574: 95–98. doi:10.1038/s41586-019-1592-6

4. Kolding J, Béné C, Bavinck M. Small-scale fisheries: Importance, vulnerability and deficient knowledge. Governance of Marine Fisheries and Biodiversity Conservation: Interaction and Co-evolution. 2014; 317–331. doi:10.1002/9781118392607.ch22

5. Kolding J, Van Zwieten PAMM. The Tragedy of Our Legacy: How do Global Management Discourses Affect Small Scale Fisheries in the South? Forum Dev Stud. 2011;38: 267–297. doi:10.1080/08039410.2011.577798

6. Mildenberger TK, Taylor MH, Wolff M. TropFishR : an R package for fisheries analysis with length-frequency data. Price S, editor. Methods Ecol Evol. 2017;8: 1520–1527. doi:10.1111/2041-210X.12791

7. Gutiérrez NL, Hilborn R, Defeo O. Leadership, social capital and incentives promote successful fisheries. Nature. 2011;470: 386–389. doi:10.1038/nature09689

8. E-Agriculture Strategy Guide. FAO; 2017.

9. Information and Communications Technologies Benefit Fishing Communities: New Directions in Fisheries - A series of Policy Briefs on Development Issues. Rome: FAO; 2007. Report No.: 7. Available: http://www.fao.org/3/a-a0991e.pdf

10. Jeffers VF, Humber F, Nohasiarivelo T, Botosoamananto R, Anderson LG. Trialling the use of smartphones as a tool to address gaps in small-scale fisheries catch data in southwest Madagascar. Mar Policy. 2019;99: 267–274. doi:10.1016/j.marpol.2018.10.040

11. Petrik M, Raemaekers DRS. The case for supporting small-scale fisheries governance through ICT. 2018.

12. Tilley A, Roscher MB, editors. Information and communications technologies to secure sustainable small-scale fisheries in Asia: Successes and Failures. FAO/WorldFish; 2019.

13. James M, Mendo T, Jones EL, Orr K, McKnight A, Thompson J. AIS data to inform small scale fisheries management and marine spatial planning. Mar Policy. 2018;91: 113–121. doi:10.1016/j.marpol.2018.02.012

14. Mallalieu KI, Rocke S. Selecting Sustainable ICT Solutions for Pro-poor Intervention: Latin American and Caribbean Perspectives. Digital Poverty: Latin American and Caribbean Perspectives. 2007. pp. 115–136.

15. Donnelly M. Implementing ICTs in Indonesia ' s Small-Scale Fisheries Development Paradigms that Shape Them. malmo. 2018.

16. R Special Interest Group on Databases (R-SIG-DB), Wickham H, Müller K. DBI: R Database Interface. 2018. Available: https://CRAN.R-project.org/package=DBI

17. Ooms J, James D, DebRoy S, Wickham H, Horner J. RMySQL: Database Interface and "MySQL" Driver for R. 2019. Available: https://CRAN.R-project.org/package=RMySQL

18. Boettiger C, Lang DT, Wainwright PC. rfishbase: exploring, manipulating and visualizing FishBase data from R. J Fish Biol. 2012;81: 2030–2039. doi:10.1111/j.1095-8649.2012.03464.x

19. Maunder MN, Punt AE. Standardizing catch and effort data: a review of recent approaches. Fish Res. 2004;70: 141–159. doi:10.1016/j.fishres.2004.08.002

20. Arya S, Mount D, Kemp SE, Jefferis G. RANN: Fast Nearest Neighbour Search (Wraps ANN Library) Using L2 Metric. 2019. Available: https://CRAN.R-project.org/package=RANN

21. Tilley A, Wilkinson SP, Kolding J, López-Angarita J, Pereira M, Mills DJ. Nearshore Fish Aggregating Devices Show Positive Outcomes for Sustainable Fisheries Development in Timor-Leste. Frontiers in Marine Science. 2019;6: 487. doi:10.3389/fmars.2019.00487

22. López-Angarita J, Hunnam KJ, Pereira M, Mills DJ, Pant J, Teoh SJ, et al. Fisheries and aquaculture of Timor-Leste in 2019: Current knowledge and opportunities. Penang, Malaysia: WorldFish; 2019.

23. Tilley A. PeskAAS - National Fisheries Monitoring System of Timor-Leste. Harvard Dataverse; 2016. doi:10.7910/DVN/TVGIJJ


PeskAAS
Automated Analytics System for Small Scale Fisheries in Timor-Leste
Start date
01-Feb-2019
End date
31-Jan-2020
Tracked activity
CPUE by Month
CPUE by site
CPUE by habitat
CPUE by gear
Catch by species
Summary
Location
Viqueque
Lautem
Manatuto
Liquica
Bobonaro
Covalima
Manufahi
Ainaro
Baucau
Dili
Oe-Cusse
Atauro
select/deselect all
Boat type
Canoe
Motor
select/deselect all
Habitat
Reef/Ahu ruin
FAD/Rumpon
Deep/Tasi kle'an
Beach/Tasi ninin
Mangrove/Aiparapa
Gleaning/Meti
select/deselect all
Gear type
Gillnet/Redi
Handline/Hakail
Longline/Hakail naruk
Spear/Kilat
Cast net/Dai
Manual/Meti
Beach seine/Redi tasi ninin
Seine net/Lampara
Trap/Bubur
select/deselect all
Smoothness
0.41
Download (csv)
Daily Catch Per Unit Effort (KG/hour)
Gear type
Reported hours
WorldFish
Gabinete do
Director Geral das Pescas

Figure 1

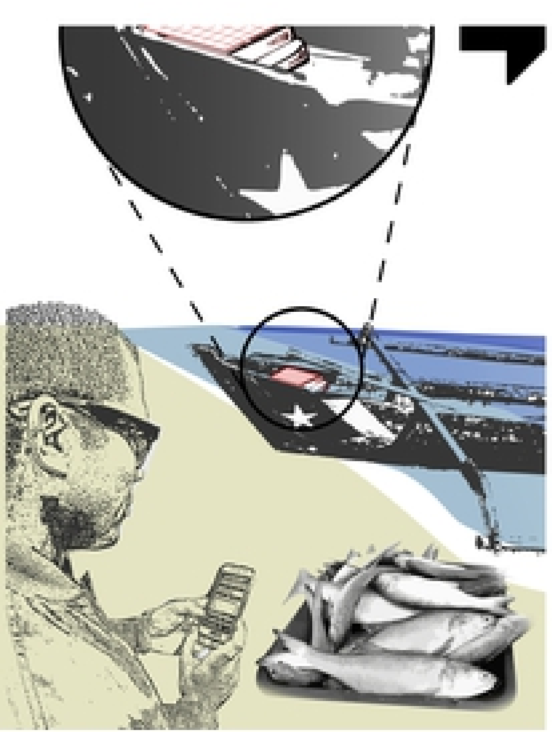
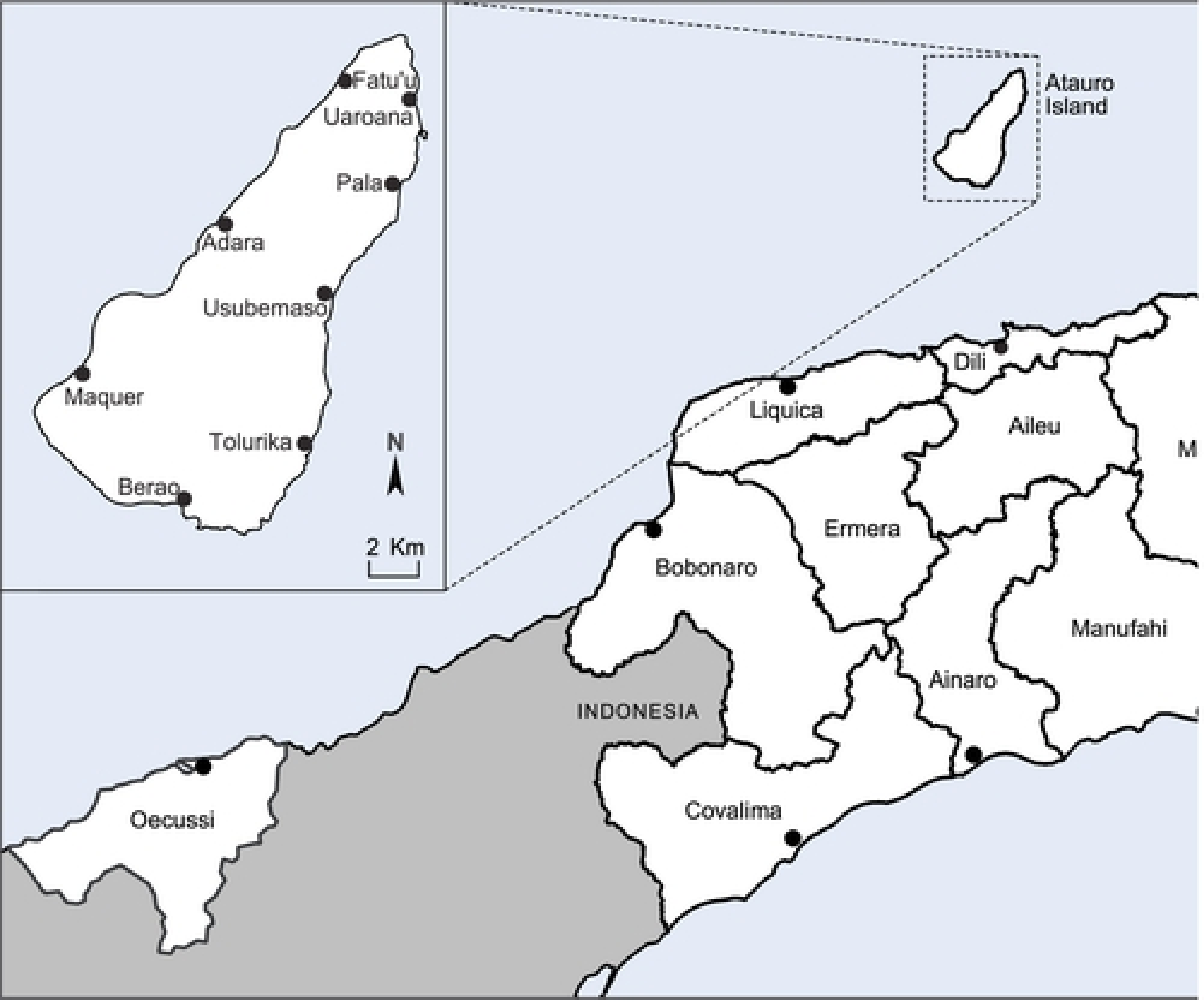


Figure 2



Fatu'u
Uaroana
Pala
Adara
Usubemaso
Maquer
Tolurika
Berao
N
2 Km
Atauro
Island
Dili
Liquica
Aileu
Ermera
Bobonaro
Manufahi
Ainaro
INDONESIA
Covalima
Oecussi

Figure 3

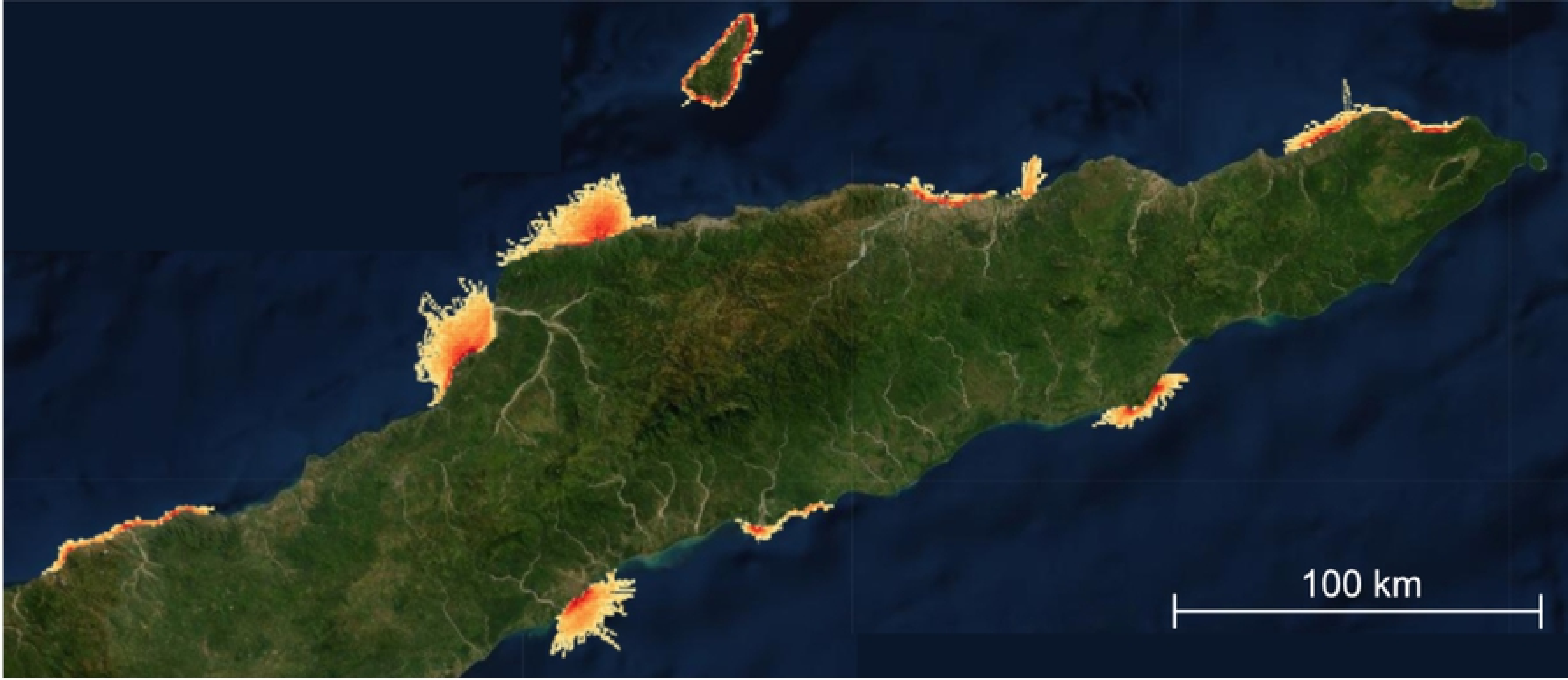


Figure 4