Matters arising

Caution over the use of ecological big data for conservation

https://doi.org/10.1038/s41586-021-03463-w

Received: 6 November 2019

Accepted: 16 March 2021

Published online: 7 July 2021

Check for updates

Alastair V. Harry[1,2] ✉ & J. Matias Braccini[1]

ARISING FROM N. Queiroz et al. *Nature* https://doi.org/10.1038/s41586-019-1444-4 (2019)

Highly collaborative and data-intensive ecology studies are at the forefront of innovative solutions to global issues in conservation and natural resource management[1,2]. In their spatial risk assessment of industrialized fishing, Queiroz et al.[3] use big data and collaborative science to outline a global conservation blueprint for pelagic sharks. In Australian waters, their analysis incorrectly identified global risk 'hotspots' in areas that are not subject to fishing and where spatial closures and other management measures are already in place to protect sharks. We highlight the potential for large-scale global analyses to misdirect conservation efforts if not aligned with regional needs and priorities.

Although ecologists have enthusiastically adopted collaborative, data-driven approaches in recent years, limited attention has been given to the challenges in this emergent field, including the potential for these often highly impactful studies to confound management and conservation actions[4]. We applaud the collaborative effort by Queiroz et al.[3] in assimilating satellite tagging data on 1,800 large pelagic and neritic sharks generated by 153 authors. However, we also caution against the use of data-intensive methods for guiding policy at the global scale without proper acknowledgement of their risks, complexities and limitations.

In their paper, Queiroz et al.[3] identify Australia's North West Shelf (NWS) as a global fishing exposure hotspot for sharks on the basis of spatial overlap with purported drifting longline and purse seine fishing vessel movements, despite no such fishing having occurred during the past two decades in this area. When we downscaled the approach of Queiroz et al.[3], we found errors in the data used to evaluate fishing exposure in these waters that were derived using a machine learning approach applied to vessel automatic identification system (AIS) location data[5].

In Western Australian state waters—an area larger than the Bering Sea—99.8% of longline and 100% of purse seine AIS data were incorrectly classified by the machine learning algorithm (Table 1 and Fig. 1). Incorrect classifications included movement data from other types of commercial fishing vessels as well as non-fishing vessels. For example, 95% of the data for purse seines in Western Australia waters were attributed to the movements of the research vessel of our agency (which, incidentally, does not undertake purse seine or drifting longline surveys).

The area of the NWS identified as highest risk falls within a spatial closure of 0.8 million km^2 in which directed shark fishing has been prohibited since 2005[6]. Although an area to the northeast remains open to shark fishing, none has occurred since 2009[6] and a network of State and Commonwealth marine reserves has since been implemented over much of that area. Fishery-independent surveys carried out over a 17-year period confirm stable or increasing relative abundance and size of large sharks in the region[6]. Historically, the waters adjacent to the NWS shelf were indeed important fishing grounds for foreign drifting longline vessels before their exclusion from Australian waters in 1997[7], and for Australian vessels in the subsequent years[8]. Contemporary longlining by a domestic tuna and billfish fishery still occurs, although these vessels were absent from the AIS data used by Queiroz et al.[3]. Since 2005, the intensity of this fishery has decreased and its footprint shifted to the southwest[9].

The approach of Queiroz et al.[3] fared better at the scale of the entire Australian Exclusive Economic Zone and offshore territories (10.2 million km^2), where the tuna and billfish longline fleet operating off eastern Australia was correctly classified (Fig. 1). However, 51% of drifting longline data were still incorrect (Table 1) and, notably, several demersal trawlers were also misclassified as being part of the longline fleet. Data from these vessels led to the incorrect identification of another pelagic longline risk hotspot within the Great Barrier Reef Marine Park (Fig. 1), where this fishing method is not permitted. In the case of both the NWS and Great Barrier Reef, the fishing exposure hotspots identified were due to fewer than five vessels being misclassified, highlighting a presumably unexpected level of sensitivity in the analysis.

As illustrated here, although patterns identified in global analyses may be broadly informative, they can also be incorrect or misinformative at regional levels where there is the scope for misallocating resources for conservation and management. Framed alternatively,

Table 1 | Summary of machine-learning classified fishing effort data

	Western Australia		Australia and offshore territories	
Total area (million km^2)	2.27		10.2	
Gear type	Longline	Purse seine	Longline	Purse seine
Total classified vessels	11	3	76	15
Incorrectly classified vessels	9	3	24	11
Fishing hours	41,074	2,650	190,355	7,511
Incorrect fishing hours (%)	99.82%	100%	51%	82%

The machine-learning classified fishing effort data used by Queiroz et al.[3] to evaluate the risks to sharks from fishing in Western Australian and Australian maritime jurisdictions. The table shows the total number of vessels classified as using longlines or purse seine, and their respective fishing hours, along with the number of vessels and percentage of fishing hours found to be incorrect. Australia and offshore territories includes all offshore and sub-Antarctic territories and the Australian Antarctic Territory.

[1]Fisheries & Agriculture Resource Management, Department of Primary Industries and Regional Development Western Australia, Hillarys, Western Australia, Australia. [2]Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, Murdoch, Western Australia, Australia. ✉e-mail: alastair.harry@gmail.com

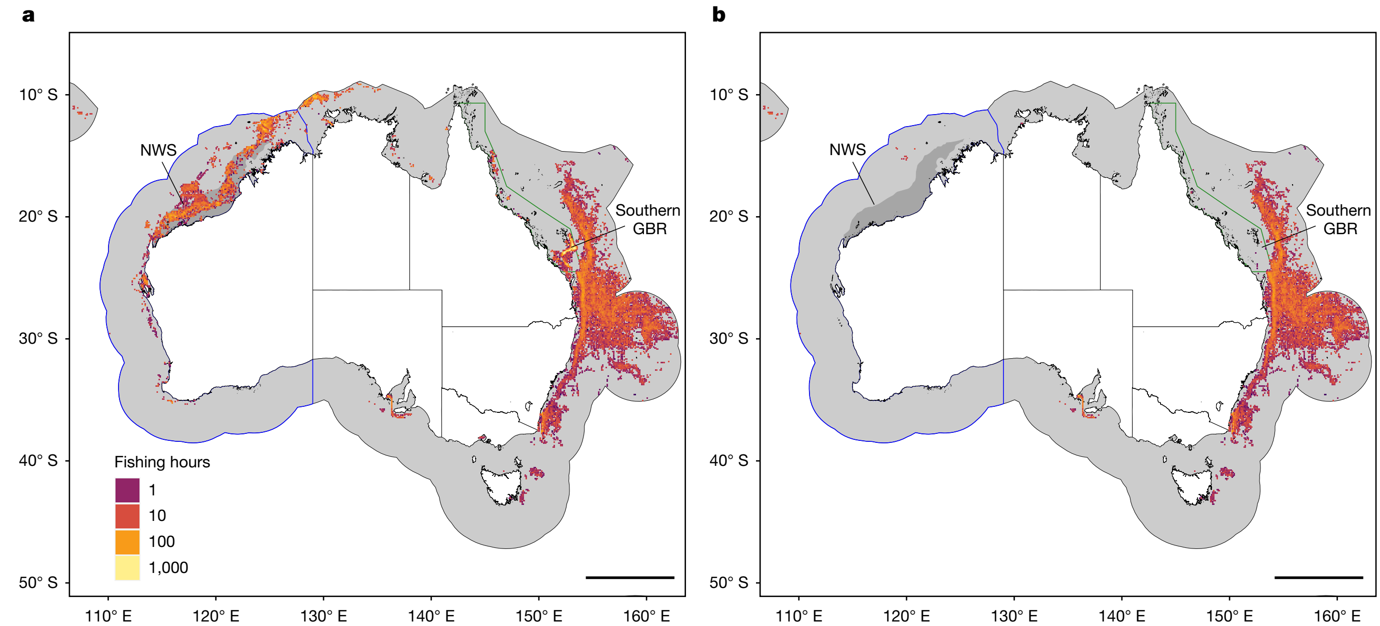

Fig. 1 | Machine-learning-classified fishing effort data (0.1° × 0.1° grid cells) used to evaluate the risk to sharks from pelagic longline and purse seine fishing in waters under Australian jurisdiction. a, Data used in the original analysis by Queiroz et al.[3]. **b**, Data that were correctly attributed to longline and purse seine fishing vessels. The NWS and the southern Great Barrier Reef (GBR) were identified as globally important 'hotspots' based on the spatial overlap of longline fishing and shark density. Grey shading shows the waters under Australian jurisdiction. The green line denotes the boundary of the Great Barrier Reef Marine Park. The blue line denotes the Western Australian maritime jurisdiction. Dark grey shading within Western Australian waters denotes the NWS. This figure was created with the statistical software R v.4.0.2[15]. Scale bars, 800 km.

what constitutes an acceptable level of accuracy at the global level may be unacceptable at the regional or local level. The sheer volume of data alone cannot overcome issues of potential bias and, in some cases, can magnify them[10,11].

These challenges point to a greater role for authors of global studies in harmonizing their research outcomes with regional needs and priorities. Strategies for aligning research that makes use of the large number of contributing authors could involve consultation with natural resource managers or the use of regional focus groups to identify errors and inconsistencies. In this case, examination of the substantial body of publicly available, annually published status reports for the relevant Australian fisheries, or engagement with Australian fisheries scientists, would have revealed the errors.

Big-data research driven by multi-author collaboration has reshaped the speed and scale at which science is conducted and delivered, with impact and reach often far exceeding traditional studies. The responsibility lies with practitioners to ensure that these methods are used appropriately given their potential to influence decision-making.

In Western Australia, the findings of Queiroz et al.[3] risk undermining confidence in the science-based management controls that are already implemented to protect the mature biomass of long-lived dusky shark (*Carcharhinus obscurus*) and sandbar shark (*C. plumbeus*) stocks in the region[12]. Off the southern Great Barrier Reef, the incorrect identification of a global longlining hotspot has the potential to undermine regional advice for the conservation of tiger sharks (*Galeocerdo cuvier*) and white sharks (*Carcharadon carcharias*), which have seen major population declines over recent decades[13].

The demand for solutions to global-scale environmental problems has necessitated changes to the prevailing culture of individual, investigator-driven ecology[14]. Queiroz et al.[3] provide a powerful demonstration of what can be achieved when ecologists work collectively by leveraging their data and expertise to approach these problems in new ways. An ongoing challenge of this and similar studies is how to provide globally relevant advice without superseding that of practitioners working at the regional level. A balanced and critical view of highly collaborative and data-intensive approaches is essential if the opportunities they provide are to be fully realized.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The results of the manual vessel review are available on GitHub (https://github.com/alharry/sharkMA).

1. Cheruvelil, K. S. & Soranno, P. A. Data-intensive ecological research is catalyzed by open science and team science. *Bioscience* **68**, 813–822 (2018).
2. Kelling, S. et al. Data-intensive science: a new paradigm for biodiversity studies. *Bioscience* **59**, 613–620 (2009).
3. Queiroz, N. et al. Global spatial risk assessment of sharks under the footprint of fisheries. *Nature* **572**, 461–466 (2019).
4. Clarke, R. Big data, big risks. *Inf. Syst. J.* **26**, 77–90 (2016).
5. Kroodsma, D. A. et al. Tracking the global footprint of fisheries. *Science* **359**, 904–908 (2018).
6. Braccini, M., Molony, B. & Blay, N. Patterns in abundance and size of sharks in northwestern Australia: cause for optimism. *ICES J. Mar. Sci.* **77**, 72–82 (2020).
7. Larcombe, J., Caton, A., Wiliams, D. M. & Speare, P. *Western Tuna and Billfish Fisheries Research* (Bureau of Resource Sciences, Canberra 1997).
8. Larcombe, J. & Begg, G. *Fishery Status Reports 2007. Status of Fish Stocks Managed by the Australian Government* (Bureau of Rural Sciences, Canberra 2007).
9. Hobsbawn, P. I., Patterson, H. M. & Blake, S. A. P. *Australian National Report to the Scientific Committee of the Indian Ocean Tuna Commission* (ABARES, Canberra 2020).
10. Kaplan, R. M., Chambers, D. A. & Glasgow, R. E. Big data and large sample size: a cautionary note on the potential for bias. *Clin. Transl. Sci.* **7**, 342–346 (2014).
11. Bayraktarov, E. et al. Do big unstructured biodiversity data mean more knowledge? *Front. Ecol. Evol.* **6**, 239 (2019).
12. Braccini, M., Blay, N., Hesp, A. & Molony, B. *Resource Assessment Report Temperate Demersal Elasmobranch Resource of Western Australia*. Fisheries Research Report No. 294 (Department of Primary Industries and Regional Development, Perth, 2018).
13. Roff, G., Brown, C. J., Priest, M. A. & Mumby, P. J. Decline of coastal apex shark populations over the past half century. *Commun. Biol.* **1**, 223 (2018).

14. Hampton, S. E. et al. Big data and the future of ecology. *Front. Ecol. Environ.* **11**, 156–162 (2013).
15. R Core Team. R: A language and environment for statistical computing. http://www.R-project.org/ (R Foundation for Statistical Computing, 2020).

Acknowledgements We thank P. Orange and the WAFMRL Library for assistance in researching historical fishing records.

Author contributions A.V.H. carried out the analysis and wrote the first draft. A.V.H. and J.M.B. conceived the idea, interpreted the results, and edited and revised the final manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03463-w.
Correspondence and requests for materials should be addressed to A.V.H.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.