

Input precision, output excellence: the importance of data quality control and method selection in disease risk mapping

Shirin Taheri,^{a,*} María José Ruiz-López,^{a,b} Sergio Magallanes,^{a,b} and Jordi Figuerola^{a,b}

^aDepartamento de Biología de la Conservación y Cambio Global, Estación Biológica de Doñana (EBD), CSIC, Sevilla, Spain

^bCIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

The transmission system of vector-borne diseases is highly complex, involving the interaction between various species of vectors and reservoirs (vertebrate hosts) that affect pathogen transmission. Different abiotic factors modulate these biotic interactions. For instance, climate change can accelerate the reproductive and developmental cycles of vectors, increase their abundance, expand their geographic range, and change their biting behaviour.¹ These changes can also affect the abundance, reproduction, and migratory behaviour of natural hosts,² or increase the replication rate of the pathogens.³ This scenario may alter the transmission risk of some pathogens, concerning citizens, epidemiologists, public health officials, and authorities worldwide (Fig. 1a). To study abiotic factors' impact on vector-borne diseases, predictive modelling is frequently used. In particular, machine learning and artificial intelligence (AI) methods are becoming increasingly popular to identify high-risk pathogens and their diversity, abundance, and distribution across various scales.

In a recent study, Farooq et al. (2022)⁵ applied a component-based approach to analyse the predictive potential of eco-climatic factors for predicting West Nile Virus (WNV) outbreaks across Europe from (2010–2019). The study incorporated human WNV cases, vectors, and avian host abundance, using AI for data analysis. Their findings indicate that temperature anomalies, lower water index, and drier winter conditions are the determining factors for WNV outbreaks across Europe. While vector abundance (*Culex pipiens* and *Culex modestus*) and the 61 passerine species identified as potential reservoir hosts did not contribute to the model. This contradicts previous reports establishing a clear association between avian and vector abundance and WNV incidence.^{6,7}

Upon reviewing the dataset provided by the authors, we detected that data on *Culex* species between 2011 and 2019 were only documented in 28% of the NUTS3 regions (See Fig. 1). Among these 562 NUTS3, only 14% were assigned a value for *C. pipiens* and 24.3% for *C. modestus*, with the remaining areas imputed as 0 (Fig. 1b and c). This contradicts the widespread presence of these vectors across Europe.⁸ Although the

vectors distribution in Farooq et al. (2020) closely aligns with the ECDC database, '0' typically indicates missing data rather than vector absence. The presence of vector data in ECDC databases is not independent of pathogen incidence, as *Culex* sp. distribution data is more likely to be collected where WNV cases have been reported. In addition, the study suggests a link between the 2018 WNV outbreak and *Cx. modestus* abundance. However, this is surprising as the original dataset shows *Cx. modestus* was absent from Europe that year, what is clearly wrong. This highlights the risk of confirmatory bias when using AI methods on biased and incomplete datasets.⁹

Similar data deficiencies were observed in the 61-bird species designated as reservoir hosts, with a distribution that does not match real-world observations. For example, *Turdus merula* (common blackbird), highly susceptible to WNV infection and abundant across Europe, was only recorded in (n = 11) NUTS3 regions in southern Spain, contradicting its known widespread presence (Fig. 1d). Additionally, the authors do not clarify the selection criteria for the 61 birds listed, omitting highly susceptible species such as *Passer domesticus* and Corvids that are not included in the risk model. Twenty species such as *Aegypius monachus* (a vulture) and *Pandion haliaetus* (a raptor), were incorrectly classified as passeriformes.

The bias in the data may explain why the authors found no clear link between human WNV cases and the vectors or reservoir hosts in Europe. This implies that their conclusions might be substantially influenced by uncertainties or flawed distribution patterns of key components for assessing the risk of WNV.

Despite numerous studies attempting to document disease risk maps, challenges persist across multiple fronts. Like it happens when mapping biodiversity, when mapping pathogens spatio-temporal sampling biases hinder accurate disease risk assessment. This challenge is particularly pronounced in vector-borne diseases, which involve multiple species that host, vector, and disseminate pathogens. Therefore, prioritizing data quality assessment and expert knowledge should be



The Lancet Regional Health - Europe
2024;42: 100944

Published Online xxx
<https://doi.org/10.1016/j.lanepe.2024.100944>

DOIs of original articles: <https://doi.org/10.1016/j.lanepe.2022.100370>, <https://doi.org/10.1016/j.lanepe.2024.100947>

*Corresponding author.

E-mail address: Taheri.shi@gmail.com (S. Taheri).

© 2024 The Author(s). 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/>).

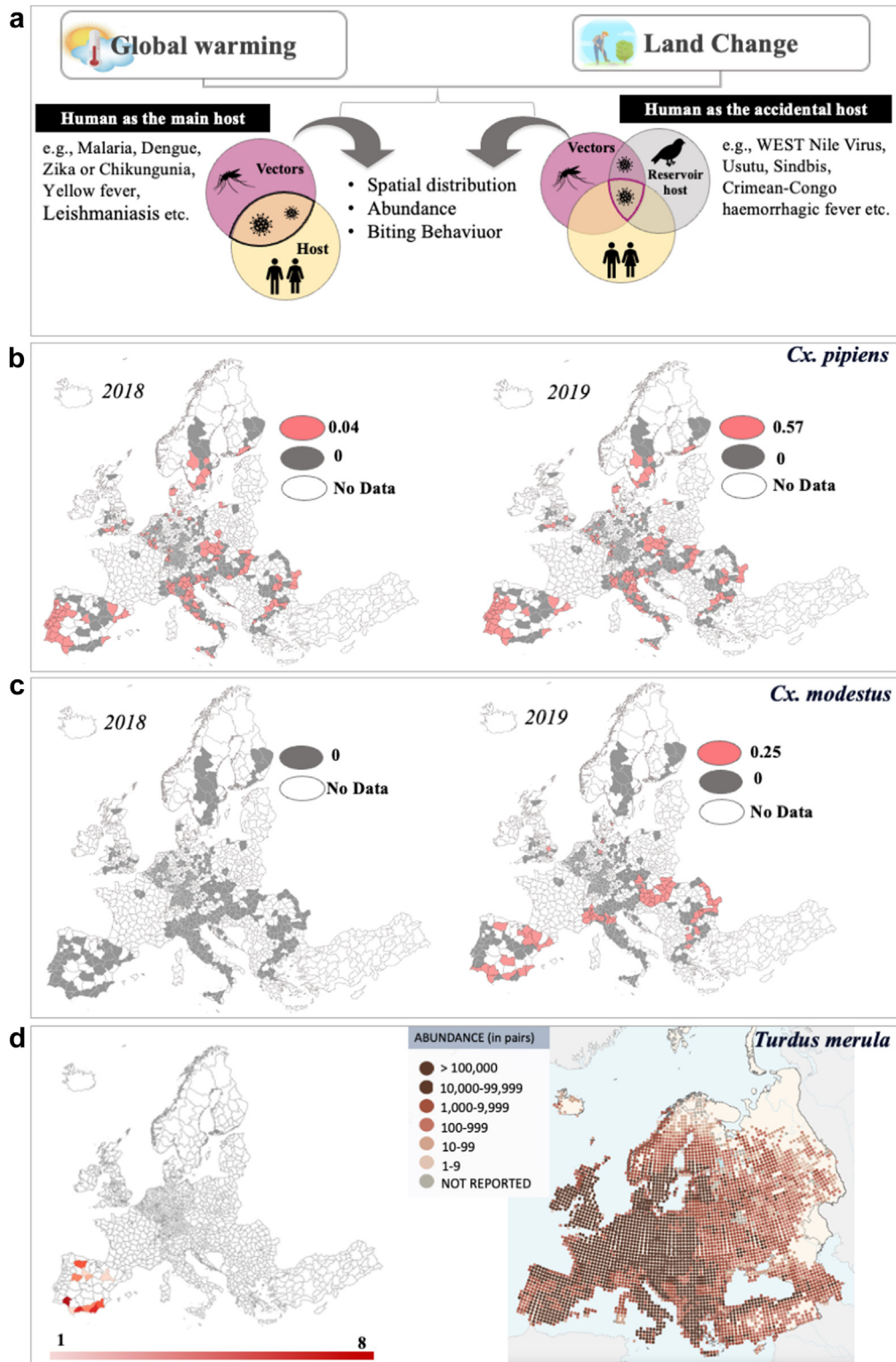


Fig. 1: a) A conceptual framework illustrating complex interactions between vectors and human as primary host, as observed in diseases like Malaria (black line on left), and the triangle analogy among, vector, reservoir host (e.g., birds) and accidental host, such as West Nile Virus (purple line on right). b) Abundance and distribution of *Culex pipiens* in Europe for 2018 & 2019 according to Farooq et al. (2022)' dataset for NUTS3 regions (Nomenclature of territorial units for statistics). c) Abundance and distribution of *Culex modestus* in Europe for 2018 & 2019 also based on Farooq et al. (2022)' dataset. d) Distribution and abundance comparisons of *Turdus merula*, with the left panel based on Farooq et al. (2022), and the right panel on the Second Atlas of European birds (EBBA2⁴).

the initial step in disease risk mapping. While advanced data analysis techniques like AI are valuable tools, they cannot perform magic in the modelling calibration process. Failure to address biases and errors in the input will directly impact the outputs. This complex process adheres to the principle of “garbage-in, garbage-out”.¹⁰

The climate and health crises are converging, yet disease outbreaks cannot occur in the absence of bridge vectors, reservoir hosts, pathogens and the interaction among them. Their surveillance and management require “One Health” approaches integrating reliable information on biological, social, and climatic factors. Misinterpreted or inaccurate data can result in misguided decisions regarding public health strategies, resource allocation, and treatment protocols.

Contributors

S.T, M.J.R.L, S.M and J.F all contributed to the writing of the manuscript.

Declaration of interests

Dr. Taheri reports support from Fundación La Caixa grant number HR22-00123. Dr. Ruiz-López reports funding from the Agencia Estatal de Investigación (project PID2020-118921RJ-100/AE1/10.13039/501100011033).

Dr. Magallanes report support from MICIU grants #PLEC-2021-007968. Dr. Figuerola reports funding from Fundación La Caixa grant #HR22-00123, Junta de Andalucía, MICIU grants #PLEC-2021-007968, #PID2021-123761OB-I00 and European Commission.

Acknowledgements

Our research is supported by Fundación La Caixa through the project ARBOPREVENT(HR22-00123).

References

- 1 Caminade C, McIntyre KM, Jones AE. Impact of recent and future climate change on vector-borne diseases. *Ann N Y Acad Sci.* 2019;1436(1):157–173.
- 2 Bartlow AW, Manore C, Xu C, et al. Forecasting zoonotic infectious disease response to climate change: mosquito vectors and a changing environment. *Vet Sci.* 2019;6:40.
- 3 Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD. Climate change and infectious diseases: from evidence to a predictive framework. *Science (1979).* 2013;341(6145):514–519.
- 4 Hagemeyer WJM, Blair MJ. *The EBCC Atlas of European breeding birds: their distribution and abundance.* London: T. & A.D. Poyser; 1997.
- 5 Farooq Z, Rocklöv J, Wallin J, et al. Artificial intelligence to predict West Nile virus outbreaks with eco-climatic drivers. *Lancet Reg Health Eur.* 2022;17:100370.
- 6 Kilpatrick AM, Daszak P, Jones MJ, Marra PP, Kramer LD. Host heterogeneity dominates West Nile virus transmission. *Proc Biol Sci.* 2006;273(1599):2327–2333.
- 7 Martínez-De La Puente J, Ferraguti M, Ruiz S, et al. Mosquito community influences West Nile virus seroprevalence in wild birds: implications for the risk of spillover into human populations. *Sci Rep.* 2018;8(1):1–7.
- 8 Kampen H, Medlock JM, Vaux AG, et al. Approaches to passive mosquito surveillance in the EU. *Parasit Vectors.* 2015;8(1):1–13.
- 9 Hernández-Triana LM, Brugman VA, Pramual P, et al. Genetic diversity and population structure of *Culex modestus* across Europe: does recent appearance in the United Kingdom reveal a tendency for geographical spread? *Med Vet Entomol.* 2020;34(1):86–96.
- 10 Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Mag.* 1996;17(3):37.