Can Species Distribution Models Inform Us About Future Ecosystems?

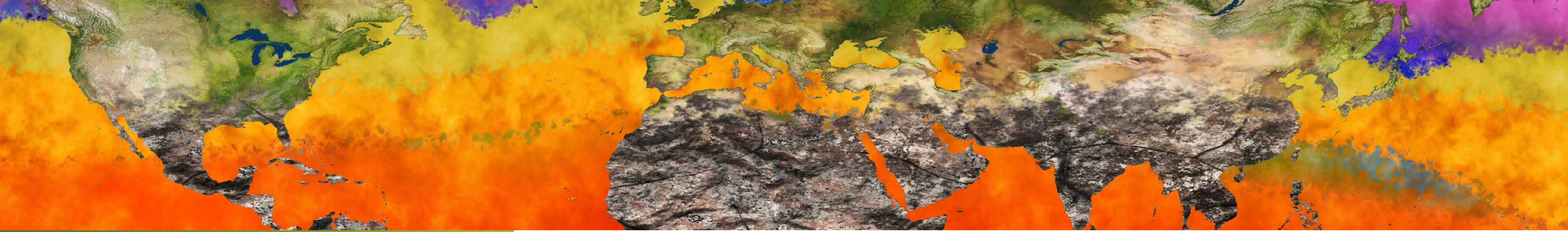
Dr Lifei Wang

JUNE 2024

doi.org/10.33548/SCIENTIA1038

EARTH & ENVIRONMENTAL SCIENCE

Scientia



Can Species Distribution Models Inform Us About Future Ecosystems?

The world is buzzing with news about how human activities and climate shifts are reshaping our ecosystems. Have you ever wondered how life will adapt to this rapidly changing world? Ecologists might be able to predict how different species will live in future using computer simulations. Dr Lifei Wang at the University of Toronto Scarborough investigates how different stimulations work under varying conditions to provide new insights into what may lie ahead.

Ecosystems and Ecological Patterns

Ecosystems are interconnected communities of living organisms interacting with each other and their physical environment. These interactions involve the exchange of energy and nutrients, which support the growth and survival of organisms within the ecosystem. Ecosystems vary in size, from small ponds to vast rainforests, and they constantly change, both in response to natural processes and human activities. Changes in one ecosystem, such as species migration or biogeochemical cycling, can have varying effects on others at different scales. When these changes accumulate, they can lead to dramatic shifts in ecological patterns.

Scientists have been closely monitoring these shifts in ecological patterns to understand emerging behaviours, such as novel traits and shifts in species diversity, and to predict the scale of impact on existing ecosystems in the future. Species distribution models (SDMs) are powerful tools used to predict how species distributions might change across space and time.

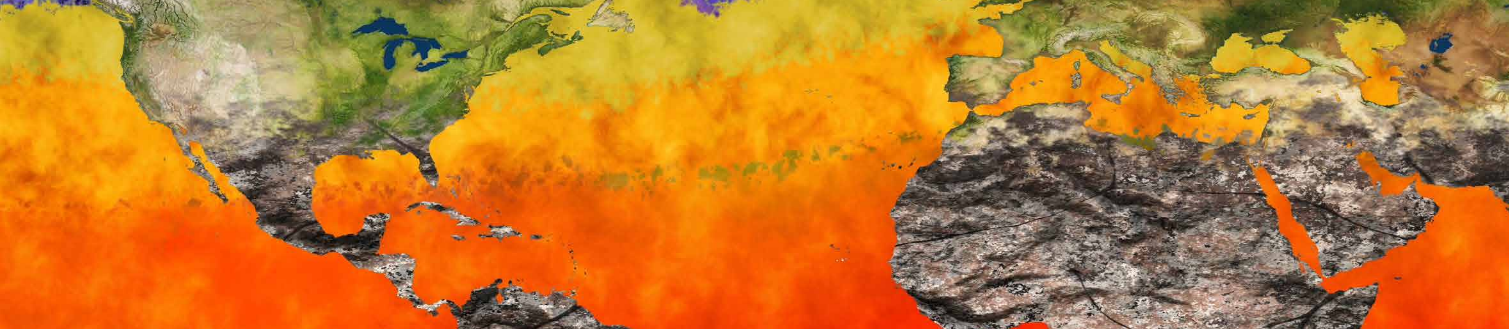
By analysing the relationships between species distributions and environmental factors, SDMs can reveal current patterns and forecast potential shifts. They help researchers understand how organisms choose their habitats, assess their vulnerability to changing environments, and plan conservation efforts. These modelling approaches are widely applied across fields such as climate change, biogeography, conservation biology, and epidemiology.

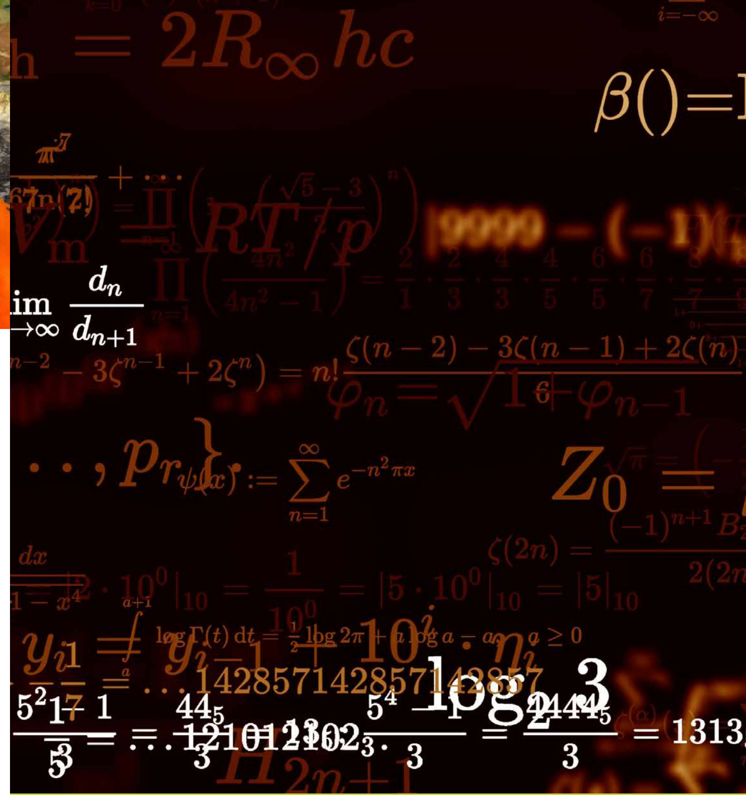
Enhancing Ecological Understanding and Predictive Accuracy

With advancements in technology and data collection, various modelling techniques, such as regression approaches (which analyse relationships between different variables) and ordination methods that help visualise data in an organised manner to look for patterns or similarities, can be utilised to enhance ecological understanding and predictive accuracy. However, there is a growing concern that different models may yield contrasting predictions even when based on the same data and application. This variability prompts the need to carefully consider data quantity (e.g., sample size) and quality (e.g., detection limit), which can influence model performance.

Dr Lifei Wang at the University of Toronto Scarborough employs a new method to explore the effects of sample size, data quality, and species response in environmental space on the performance of different SDMs. The mathematical function – generalised ß-function – describes a range of possible species responses along ecological gradients. First introduced for ecological simulations in 1976, it provides a flexible way to simulate complex species responses under different ecological assumptions.

Using this tool, Dr Wang and her colleagues simulated 16 species response relationships across ecological gradients with different shapes, controlling sample size and detection limit to investigate their influence on model performance. They compared seven typical SDMs under various sample sizes and detection limits to glean important insights into how different modelling approaches perform under different data quantity and quality conditions.





Simulation Process

The simulation involved creating a dataset using the generalised ß-function, which represents the species on two ecological gradients and its probability of occurrence at 1,000 sampling sites. The resulting dataset comprised 16 species populations, each with different occurrence responses to ecological gradients, controlled by adjusting parameters to influence skewness (asymmetry) and kurtosis (peakness) of the data distribution. This allowed the creation of species with various response shapes (such as symmetric, skewed, or bell-shaped, for readers who are mathematically minded).

Subsequently, modelling processes were simulated using sampled data from these populations at different sample sizes and detection limits. For sample size, they looked at datasets ranging from 50 to 800 sites. For detection limit, they considered probabilities ranging from 0 to 0.8. By testing models with these varied datasets, they mapped how the quantity and quality of data influence the predictive accuracy of seven SDMs.

Evaluating Features and Limitations

Each of the seven SDMs studied by Dr Wang presented distinct assumptions, advantages, and disadvantages. Linear Discrimination Analysis (LDA) works well when the variables used to predict the occurrence or abundance of a species (predictors) are related in a linear manner. Multiple Logistic Regression (MLR) can handle different types of predictors but falters when there is too much variation or noise in the data. Generalized Additive Models (GAM) struggle to understand how predictors are related but can trace complex patterns. Boosted Regression Trees (BRT) and Random Forests (RF) can handle large datasets and find intricate patterns. Artificial Neural Networks (ANN) and Maximum Entropy Models (MaxEnt) are better at handling complex patterns but might overfit the data. These features and limitations can better guide researchers in selecting the most suitable modelling approaches for specific datasets and research objectives.

The Importance of Sample Size, Data Quality, and Species Response Shape

Skewness and kurtosis are statistical measures that describe the shape of data distribution. In the context of SDMs, skewness refers to the asymmetry in the probability distribution of species occurrence along ecological gradients. Positively skewed distributions concentrate occurrences towards the lower end of the gradient, whereas negatively skewed distributions do the opposite. Kurtosis measures the peakness of the probability distribution. Higher kurtosis indicates a sharper peak with fewer outliers, while lower kurtosis suggests a flatter distribution with more dispersed occurrences.

Dr Wang's analysis found that SDMs generally performed better with larger sample sizes and lower detection limits, indicating that increasing data quantity and quality could help improve predictive accuracy. However, the improvement varied among different modelling approaches. LDA and MLR were strongly influenced by the shape of species responses, performing better for skewed species with greater kurtosis than for symmetric species. More complex models like BRT, RF, ANN, and MaxEnt performed better with larger sample sizes but were affected when sample sizes were small and kurtosis decreased, especially ANN and MaxEnt. GAM and BRT were influenced by both sample size and detection limit. These findings highlighted how sample size, data quality, and species response in environmental space affect the performance of different SDMs.

"Subsequently, modelling processes were simulated using sampled data from these populations at different sample sizes and detection limits. For sample size, they looked at datasets ranging from 50 to 800 sites.

Practical Applications and Future Research

Many ecological modelling studies have extensively explored various approaches and considered their strengths and weaknesses in predicting species distributions. However, one aspect usually overlooked is how the quantity and quality of data affect the model performance. This is often because there is inherent uncertainty in determining true occurrence probabilities of species, abundance, and their relationships with environmental factors based solely on field-sampled data.

To address this challenge, Dr Wang used simulated species with known distributions with predetermined relationships to environmental predictors. This allowed her to control the complexity of species-habitat interactions and eliminate the influence of confounding factors on model performance, providing a clearer understanding of the actual capabilities of ecological models.

Future studies could expand these simulations to novel modelling approaches (e.g., joint dynamic SDMs), which could incorporate dynamic processes, such as colonisation, extinction, and source-sink dynamics, in addition to climate conditions and spatiotemporal correlations. Such studies offer the potential to improve the reliability of species distribution maps, aiding in risk assessment, conservation planning, and resource management.

MEET THE RESEARCHER

Dr Lifei Wang, Department of Biological Sciences, University of Toronto Scarborough, Toronto, Canada

Dr Lifei Wang is a Postdoctoral Fellow in the Department of Biological Sciences at the University of Toronto Scarborough. She obtained her PhD in Ecology from the University of Toronto, where she specialised in species distribution modelling and time series analysis. Dr Wang has extensive research experience, having undertaken research work at the Gulf of Maine Research Institute and the University of Maine's School of Marine Sciences, where she focused on fisheries modelling using remote sensing and earth system data. She has received prestigious fellowships, such as the Mathematics of Information Technology and Complex Systems (Mitacs) Elevate Postdoctoral Fellowship and the US NASA Experimental Program to Stimulate Competitive Research (EPSCoR). Dr Wang has published and presented extensively on her research and is named on four patents.

CONTACT

lifei.wang@utoronto.ca

https://orcid.org/0000-0003-2530-6703

https://twitter.com/Lifei_Wang

KEY COLLABORATORS

Donald Jackson, University of Toronto

Nicholas Mandrak, University of Toronto Scarborough

Derrick de Kerckhove, Ontario Ministry of Natural Resources and Forestry

Lisa Kerr, Gulf of Maine Research Institute

Andrew Pershing, University of Maine

Find out more at **scientia.global**

FUNDING

Natural Sciences and Engineering Research Council of Canada (NSERC) Canadian Aquatic Invasive Species Network (CAISN) and Discovery Grant Programs

Mathematics of Information Technology and Complex Systems (Mitacs) Elevate Postdoctoral Fellowship

FURTHER READING

L Wang, DA Jackson, Effects of sample size, data quality, and species response in environmental space on modelling species distributions, *Landscape Ecology*, 2023, 38, 4009–4031. DOI: https://doi.org/10.1007/s10980-023-01771-2

L Wang, LA Kerr, NR Record, *et al.*, Modeling marine pelagic fish species spatiotemporal distributions utilizing a maximum entropy approach, *Fisheries Oceanography*, 2018, 27, 571–586. DOI: https://doi.org/10.1111/fog.12279

L Wang, DA Jackson, Shaping up model transferability and generality of species distribution modeling for predicting invasions: implications from a study on *Bythotrephes longimanus*, Biological Invasions, 2014, 16, 2079–2103. DOI: https://doi.org/10.1007/s10530-014-0649-6

L Wang, DA Jackson, Modeling the establishment of invasive species: habitat and biotic interactions influencing the establishment of *Bythotrephes longimanus*, Biological Invasions, 2011, 13, 2499–2512. DOI: https://doi.org/10.1007/s10530-011-0071-2

UNIVERSITY OF TORONTO