

Data-Driven Discovery of Mechanistic Ecosystem Models with LLMs

Scott Spillias^{1,2*} Jacob Rogers³ Fabio Boschetti^{2,4}
Beth Fulton^{1,2} Magda Guglielmo⁵ SukYee Yong⁵
Rowan Trebilco^{1,2}

October 17, 2025

¹CSIRO Environment, Hobart, Australia

²Centre for Marine Socio-Ecology, University of Tasmania, Hobart, Australia

³CSIRO Environment, St. Lucia, Australia

⁴CSIRO Environment, IOMRC Crawley, Australia

⁵CSIRO IM&T, Eveleigh, Australia

Acknowledgements

SS was supported by an R+ Postdoctoral Fellowship.

Code and Data Availability

The code for the LEMMA framework is available in a public GitHub repository at <https://github.com/s-spillias/EMs-with-LLMs>. The repository includes all scripts necessary to reproduce the results presented in this paper, including the genetic algorithm implementation, model evaluation tools, and analysis scripts. The time series data used in the case studies are also available in the repository. The software is released under an MIT license.

*Corresponding author: scott.spillias@csiro.au

22 Declaration on Generative AI Usage

23 During the preparation of this manuscript, we used Claude-3.5-Sonnet, a
24 large language model, to assist with code documentation, manuscript format-
25 ting, and language editing. The scientific content, analyses, interpretations,
26 and conclusions presented in this paper were developed and validated by the
27 human authors.

28 Author Contributions

29 SS: Conceptualization, Methodology, Software, Data curation, Formal analy-
30 sis, Writing - original draft, Writing - review & editing JR: Software, Method-
31 ology, Data curation, Writing - review & editing FB: Formal analysis, Super-
32 vision, Writing - review & editing BF: Formal analysis, Supervision, Writing
33 - review & editing RT: Supervision, Writing - review & editing MG: Soft-
34 ware, Methodology, Writing - review & editing SY: Software, Methodology,
35 Writing - review & editing

36 Competing Interests

37 The authors declare no competing interests.