Review

Title: Potential for Equation Discovery with AI in the Climate Sciences
Author(s): Chris Huntingford, Andrew J. Nicoll, Cornelia Klein, and Jawairia A. Ahmad
MS No.: esd-2024-30
MS type: Perspective
Iteration: Minor revision

Verdict: Minor Revisions

SUMMARY:
This interesting paper introduces the promising research field of AI-led equation discovery to climate science. The authors provide an extensive overview over previous and current statical methods including machine learning based approaches in the field of climate science. As a remedy to the current issues such as transparency of most fully data-driven approaches and computational limitations of physics-based numerical solutions, the authors suggest the application of "equation discovery" algorithms. These AI-based algorithm enable equation generation for unknown dynamical system as well as for systems with limited dynamical information. Focussing on symbolic regression methods and specifically the SINDy algorithm, the authors provide a comprehensive, understandable and detailed description of the procedure of equation discovery. Lastly, the examples of potential applications, such as in atmospheric convection, carbon cycle parameters, and ocean feature modelling for assessing tipping point risks, further strengthen the author's conclusion and outline promising research avenues.

RESPONSE:
Overall, I find this submission to be a relevant contribution to the climate science community as it provides a comprehensive overview of ongoing research, outlines current road-blocks and specifically suggests promising approaches which were previously over-looked or unknown in the field. Therefore, I highly suggest to improve upon the clarity and structure of the abstract, introduction and conclusion. This is necessary, as I find the importance of this contribution sometimes gets lost in overly long and disjunct paragraphs and sentences. While, I like the designated re-iteration of the potential application examples in the conclusions, I suggest to improve these by focussing on the discussed ways of using equation discovery in each application. I address my concerns in details below and discuss further individual remarks.

1. *Relevance* (Does the paper address relevant scientific questions within the scope of ESD?)

This manuscript puts forward a new avenue of potential machine learning based earth system model (EMS) research, by suggesting AI-based equation discovery. Based on three highly relevant potential application examples, the authors demonstrate the relevance and motivate the integration of the proposed research direction.

2. *Novelty* (Does the paper present novel concepts, ideas, tools, or data?)

While the idea of equation discovery is an established sub-domain of machine learning and the discussed methods long-standing, their application to earth system modelling is a novel idea, to the best of my knowledge.

3. *Substantial conclusions* (Are substantial conclusions reached?)

The authors clearly establish their conclusions regarding the capabilities and potential of the AI-led equation discovery, throughout the manuscript

4. *Clarity and Validity* (Are the scientific methods and assumptions valid and clearly outlined?)

While the manuscript does exhibit structural and literary weaknesses, the algorithms in section 2.2 as well as potential applications in Section 3.3. are relayed in details and in an understandable manner.

5. *Support of the interpretations and conclusions* (Are the results sufficient to support the interpretations and conclusions?)

The presented potential applications sufficiently support and highlight the relevance of the presented methods for climate science and earth system modelling.

6. *Traceability of results* (Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists ?)

As this paper aims to be more of a perspectives/review paper the authors do not present new experiments and therefore do not require exact reproducibility.

7. *Consistency of related work* (Do the authors give proper credit to related work and clearly indicate their own new/original contribution?)

The authors consistently provide necessary citations and clearly establish the reviewed material as well as their own contributions.

8. *Title* (Does the title clearly reflect the contents of the paper?)

The manuscript title fully aligns with the manuscript content.

9. *Abstract* (Does the abstract provide a concise and complete summary?)

The abstract provides a full summary of the manuscript content. However, I highly recommend to improve structure and overall writing, as it is hard to read and difficult to follow, which does not reflect the relevance and value of this submission.

10. *Structure and Clarity* (Is the overall presentation well structured and clear?)

While I enjoyed the explanatory figures and graphs, especially introduction and discussion should be improved to further strengthen the value of this manuscript. Overly long sentences alongside sometimes disjunct paragraphs and sentences make these sections hard to follow. Exceptions are Section 2.2, 3.1, and 3.2 which, while long were easy to read.

11. *Language* (Is the language fluent and precise?)

I do have some concerns regarding the clarity of the paper (see previous point).

12. *Math* (Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?)

In Section 2.2 and Section 3.3, most mathematical expression are well-defined and explained. I only have minor concerns, which I detailed below.

13. *Possible Reduction* (Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?)

The manuscript would profit from a more precise and reduced section 2.1. (Background), discussion and introduction. In addition these parts should also be rewritten to improve clarity and readability, which currently hampers the value of this interesting contribution.

14. *Number and quality of references* (Are the number and quality of references appropriate?)

The authors clearly cite all relevant works and choose relevant works out of the respective fields. However, while SINDy and symbolic regression is a well-renowned, the field of equation discovery also extends to more novel and promising algorithms, e.g., neural operators (Lu et al. 2021, Cao et al. 2023)

15. *Supplement* (Is the amount and quality of supplementary material appropriate?)

I find this paper to be fully self-contained and therefore see no need for supplementary material.

MINOR CONCERNS:

1. *Disjunct sentences*: I find some sentences to be very hard to read, e.g. (p. 21 l. 8-10) “First, to calibrate…”

TECHNICAL CORRECTIONS:

1. Sec. 2.2: Please clarify the dimensionality of ξ (l. 6 p.8) is it the same as ξ_2. In addition I think a further specification of y might be helpful (l.7 p.8), since apparently $y = \Theta\xi$?
2. Sec. 3.3: Please add definition/descriptions of β and Tr0, since they appear to not be defined.
3. L. 13-15, p. 12: “A major concern…” -> This sentence is not understandable please check the sentence structure.

Cao, Qianying, Somdatta Goswami, and George Em Karniadakis. "LNO: Laplace neural operator for solving differential equations." *arXiv preprint arXiv:2303.10528* (2023).

Lu, Lu, et al. "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators." *Nature machine intelligence* 3.3 (2021): 218-229.