



EGUsphere, referee comment RC1
<https://doi.org/10.5194/egusphere-2022-105-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-105

Matthew Knepley (Referee)

Referee comment on "Causal deep learning models for studying the Earth system" by Tobias Tesch et al., EGU sphere, <https://doi.org/10.5194/egusphere-2022-105-RC1>, 2022

This paper was intended to "propose a novel methodology combining deep learning (DL) and principles of causality research". However, I do not believe it does so. It reiterates a standard theorem from causal models describing a causally sufficient set for some node X of a probabilistic graphical model. Then the authors claim to choose carefully such a set. If it were possible to do so apriori, there would be no confounding and no need for the causality formalism. After choosing this set, the interpolation of the joint probability distribution with a neural network follows standard practice. Since there is no real use of the mathematical formalism of causality, this cannot justify publication. Moreover, since "An extensive discussion of our results on soil moisture-precipitation coupling in terms of physical processes (e.g. Seneviratne et al., 2010; Santanello et al., 2018) and a comparison with results from other studies (e.g. Seneviratne et al., 2010; Taylor et al., 2012; Guillod et al., 2015; Tuttle and Salvucci, 2016; Imamovic et al., 2017) are postponed to a second paper", no new physical results are presented. Thus I recommend that the paper be rejected, and the authors submit a paper with the new physical insights included.

In the paper itself, some claims could be better supported by evidence. The authors claim that simulations are always more expensive than their deep learning scheme, but no data is provided. Simulations at what resolution? Is the cost of DNN training included? More nuance here would be helpful. Derivatives calculated from the DNN solution are used to quantify sensitivities and errors, but how accurate are these estimates? On page 17, the authors state that "In our example, the null hypothesis was rejected at a confidence level of 99 %", however it is later stated that only two samples were taken. This seems misleading at best. Clarification of what is meant by the 99% confidence level in this case would be very helpful.