Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2020-427-RC1, 2021

Comment on gmd-2020-427

Anonymous Referee #1

Referee comment on "Copula-based synthetic data augmentation for machine-learning emulators" by David Meyer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-427-RC1, 2021

The title of the paper is: "Copula-based synthetic data generation for **machine learning emulators in weather and climate:** application to a simple radiation model". The major conclusion of the paper is that using emulated data may be of particular interest to scientists and practitioners developing **ML emulators in weather and climate**. In my opinion both the title and the main conclusion are unfocused and the main conclusion is not something new.

First of all, authors should specify how they understand **"machine learning emulators in weather and climate".** There is a long history of developing ML emulators for numerical weather and climate models, including emulations of sophisticated state-of-the-art radiation models:

Chevallier, F., Morcrette, J.-J., Chéruy, F., Scott, N. A. (2000). Use of a neural-network-based longwave radiative transfer scheme in the ECMWF atmospheric model. Quarterly Journal of the Royal Meteorological Society, 126, 761-776

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Chalikov, D. V. (2005). New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Long Wave Radiation in a Climate Model. Monthly Weather Review, 133, 1370-1383

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Belochitski, A. (2008a). Decadal climate simulations using accurate and fast neural network emulation of full, long- and short wave, radiation. Monthly Weather Review, 136, 3683-3695. doi:10.1175/2008MWR2385.1

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Hou, Y. T., Lord, S. J., & Belochitski, A. A. (2010). Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions, Monthly Weather Review, 138, 1822-1842. doi: 10.1175/2009MWR3149.1

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Belochitski, A. A. (2013). Using Ensemble of Neural Networks to Learn Stochastic Convection Parameterization for Climate and Numerical Weather Prediction Models from Data Simulated by Cloud Resolving Model, Advances in Artificial Neural Systems, 2013, Article ID 485913, 13 pages.

doi:10.1155/2013/485913

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic Validation of a Neural Network Unified Physics Parameterization, Geophysical Research Letters, 45, 6289-6298. https://doi.org/10.1029/2018GL078510

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., Yacalis G. (2018). Could Machine Learning Break the Convection Parameterization Deadlock? Geophysical Research Letters, 45, 5742–5751. https://doi.org/10.1029/ 2018GL078202

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: potential for modeling of climate, climate change and extreme events. Journal of Advances in Modeling Earth Systems. https://doi.org/10.1029/2018MS001351

Pal, A., Mahajan, S., and Norman, M. R. (2019). Using deep neural networks as cost-effective surrogate models for Super-Parameterized E3SM radiative transfer. Geophysical Research Letters, 46, 6069–6079. https://doi.org/10.1029/2018GL081646

Rasp, S., Pritchard, M. S., Gentine, P. (2018). Deep learning to represent subgrid processes in
climate models, PNAS Latest Articles. https://doi.org/10.1073/pnas.1810286115

Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0) arXiv:1907.01351v4 [physics.ao-ph]

Scher, S. (2018). Toward dataâ□□driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. Geophysical Research Letters, 45. https://doi.org/10.1029/2018GL080704

Scher, S., Messori G. (2019). Weather and climate forecasting with neural networks: using GCMs with different complexity as study-ground. Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-53

plus a quickly growing amount of publications during 2019, 2020, and even already during 2021.

In all these works, data simulated by GCM, CRM, or LES are used. Researchers, working in these fields of weather and climate, long ago recognized that they must use simulated data because there is no sufficient amount of observations, collocated in space and time to use for training their ML emulations. In some cases ECMWF or NCEP analysis or reanalysis are used to integrate in an optimal way simulated and observed data. The same sources of data are used to develop forward models.

Thus, I suggest to authors:
1. to clearly specify what subfield of "weather and climate" they target
2. to edit the title and the text correspondingly
3. to explain how their results provide researchers, working in the targeted field, with a new information