Nonlinear Processes in Geophysics Discussions

Nonlin. Processes Geophys. Discuss., referee comment RC1
https://doi.org/10.5194/npg-2021-20-RC1, 2021

Comment on npg-2021-20

Julien Brajard (Referee)

Referee comment on "Using neural networks to improve simulations in the gray zone" by Raphael Kriegmair et al., Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2021-20-RC1, 2021

This article presents an application of a neural net to represent a parametrization of features that are not resolved by a low-resolution numerical model but that can occur at a similar scale than the low-resolution. The method is applied to the forecast of the state of a 1D shallow-water model (height, wind speed, rain mass fraction). The effect of adding a physical constraint is addressed.

The article is a valuable contribution to the field of machine learning-based parametrization. It is well written, easy to follow, and the conclusions are convincing and physically interpreted.
In my opinion, this work deserves publication. Nevertheless, I have 2 main comments and other secondary comments

1. about to construction of the datasets and the ANN:
- L90-91: This choice of train/validation split is very surprising. If you select one of every two points in the training, how can you be sure the training/set and the validation set are independent? On the contrary, I would expect that one state on the training is very close to the corresponding step of the validation (one time step further for example). I would be concerned about data leakage that would make you score over the validation dataset overconfident and would be unable to detect overfitting. Can you expand a bit on this choice of train/Val split?
- section 2.2 is there a test set? (see the following point about hyperparameters tuning)
- section 2.3 Did you need to tune the hyperparameter of the ANN (e.g. size of layers, learning rate, ...)? If so did you use the validation set or did you use a part of the training set? I think it would be nice to have a bit more details about this point...

2. about physical constraint
- Eq.(1) It seems relatively "easy" to enforce strictly the water mass constraints (just

remove uniformly the mean delta h at the last layer of the ANN.). Why not test this hard constraint here? First, it seems more "natural" as the weak constraint, because it is expected that the mass is strictly conserved. Second, results suggest that despite a "strong" mass constraint there is still a mass drift that makes the model diverge (Figure 9).

Other comments:

- L27-28 "but the resolution of 2-4 km does not give accurate results for typical convective cloud structures are often less than 10 km in size". I am not a native English speaker, but this sentence is a bit unclear to me.
- L71-L72 "In this study, a small but significant model intrinsic drift in the domain mean of u is accounted for by adding a relaxation term." Do you mean there is a systematic drift of u in the model? Is this at all resolutions? When it says "accounted for", does it mean that the drift is corrected?
- L72, you could say here that the overbar designs the domain average.
- L89 the index of the time t is 'i' but it was 'n' few lines above. This is still correct of course, but I feel that consistency in the notation can make the article even clearer.
- Table 1: Is the time step the same for the LR and HR run? Due to CFL, I would expect the time step to be smaller for HR.
- L85. it is mentioned that 2 orographies are used, but I don't understand what are the 2 orographies setups here, there seems to be an ensemble of orography. It is a bit clearer in the conclusion, but I think it should also be detailed here.
- L118: "...with the standard loss function, the MSE": maybe you could add "(w_mass=0)" here (instead of mentioning it L149)
- L128: How do you define the LR_ANN simulation? Is it the average of the 25 LR simulations? (Maybe this is what is meant L131, but I am not sure to understand)
- L130: Initial conditions being selected every 2 hours, do you expect them to be independent? If they are not, that could bias the average and standard deviation.
- Figure 4: It seems that, after around 20 hours, the dispersion of the RMSE around the mean is greater for LR_ANN than for LR. Could you comment on that?
- L155-160: Maybe it is worth mentioning that the overall effect of w_mass on the RMSE is very low as the improvements are similar (e.g. between 97.55% and 97.70% for h, regular case)
- Figure 10, that's a really nice point!
- L178 "Based on the subjective interpretation of the human brain of a a hand full of animations of the forecast evolution, it appears that convective events produced in the LR run are wider and shallower". Would it be some theoretical reason or literature to support this assertion?
- Figure 12: what are the dotted red lines?
- Figure 12: Is the example taken from the training set/validation set/test set?