Nonlinear Processes
in Geophysics

Open Access

Discussions

EGU

NPGD

Interactive
comment

Interactive comment on "Training a convolutional neural network to conserve mass in data assimilation" *by* Yvonne Ruckstuhl et al.

Yvonne Ruckstuhl et al.

yvonne.ruckstuhl@lmu.de

Received and published: 15 December 2020

article xcolor

Review response

December 15, 2020

Thank you for taking the time to thoroughly read our manuscript and for the positive feedback. We agree with your comments and have adjusted the manuscript accordingly, see below.

1 Possible improvements

This is a nicely written paper with a clear-cut organisation. The paper is convincing and well illustrated. Among possible improvements, I would list:

- The manuscript may be a bit short and could benefit from more in-depth or additional experiments if relevant.
  We performed additional experiments to investigate the trade off between mass conservation and RMSE, which are now summarized in Figure 4. Note that we have changed the definition of the penalty term by comparing the mean fields of h, not the sum, so the penalty term is divided by n=250 now.

- A few relevant and more recent references could be added (recent is very short

in this subject).

We added the following references:

- Bocquet, M., Brajard, J., Carrassi, A., and Bertino, L.: Bayesian inference of chaotic dynamics by merging data assimilation, machinelearning and expectation-maximization, Foundations of Data Science, 2, 55–80, https://doi.org/10.3934/fods.2020004, 2020.

- Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L.: Combining data assimilation and machine learning to infer unresolved scale parametri-sation, URL:https://arxiv.org/pdf/2009.04318.pdf, 2020b

- Farchi, A., Laloyaux, P., Bonavita, M., and Bocquet, M.: Using machine learning to correct model error in data assimilation and forecastapplications, 2020

- Watson, P. A. G.: Applying Machine Learning to Improve Simulations of a Chaotic Dynamical System Using Empirical Error Correction,Journal of Advances in Modeling Earth Systems, 11, 1402–1417, https://doi.org/https://doi.org/10.1029/2018MS001597, 2019

- Yuval, Janni and O'Gorman, Paul A: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions,Nature communications, 11, 1, 1–10, 2020, Nature Publishing Group

- Stephan Rasp and Nils Thuerey:Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for WeatherBench, 2020,arXiv preprint arXiv:2008.08626

- It would be much better to make the codes available for the sake of repeatability, as is customary in the machine learning community; maybe not all of them, since that may become tedious, but for instance the model and the machine learning code pieces.

We will provide the code.

- The line and equations numbering could/should be corrected/improved.
  fixed

Please see below for the details about these suggestions. Overall, I believe the manuscript only requires minor revisions but that they should be very carefully addressed.

2 Suggestions and typos:

1. l.4-6: "In order to produce from a less computationally expensive, unconstrained analysis, a solution that is closer to the constrained analysis, we propose to use a convolutional neural network (CNN) trained on analyses produced by the QPEns.": The sentence is difficult to understand because: (i) there should not be a comma in between "expensive, unconstrained" (ii) "closer": what do you compare to? This is confusing because of the beginning of the sentence; "close" may work better here.
   We rephrased to "We therefore propose to use a convolutional neural network (CNN) trained on the difference between the analysis produced by a standard ensemble Kalman Filter (EnKF) and the QPEns to correct any violations of imposed constraints."

2. l.8-9: "To obtain these positive results, it was in one case necessary to add a penalty term to the loss function of the CNN training process.": This is too vague a statement for an abstract. In my opinion, you should make it more precise or remove it (since the abstract is not long, the former is better).
   We removed it.

3. l.17: "Janjic (2016),Zeng et": a space is missing. ′
   fixed

Printer-friendly version

Discussion paper

4. "Artificial neural networks (NN), are powerful tools" $-\to$ "Artificial neural networks (NN) are powerful tools"
   fixed

5. l.27: "non-linear": nonlinear is much more common (check the title of the journal).
   fixed

6. l.28: "based on example" $-\to$ "based on examples"?
   fixed

7. l.45: Brajard et al. (2019). has actually been accepted as Brajard et al. (2020a). Can you please update the reference?
   fixed

8. l.36: "combining NN with a knowledge based model as a hybrid forecasting approach (Pathak et al., 2018b)": I believe Brajard et al. (2020b), which recently appeared, is also a very relevant citation to your manuscript because as opposed to Pathak et al. (2018) who rely on only one degree of freedom in model error and reservoir computing, Brajard et al. (2020b) have many degrees of model error freedom and rely on CNNs, like you do.
   We added the reference

9. l.75: "Gaussian stochastic forcing β_u has a half width of 4 grid points": Is this remark about correlation length of the covariance matrix?
   No, a Gaussian shaped term β_u is added to the wind field at each model time step at a random location (see line 72 of the new manuscript).

10. l.82: "with parameters $\mu = -8$ and $\sigma = 1.5$.": You have to be more precise. What are μ and σ? You know that it can be ambiguous for log-normal distributions (whether you consider the variable of the log-variable).
    Good point. We rephrased: "a lognormal error is added to the rain field with parameters of the underlying normal distribution $\mu=-8$ and $\sigma= 1.5$"

11. l.87: "using 5-th order polynomial function (Gaspari and Cohn, 1999)": I believe that what you use is actually a 5-th piecewise rational function, is it? Thank you. We fixed it.

12. l.94-95: "the analysis error is larger than that of an arbitrary model state.": Do you mean larger than the climatological standard deviation of the model state? It's unclear to me.
Yes, we have rephrased.

13. l.117-119: I believe that you should give a reference for the selu activation function because giving those values would seem strange to typical readers of Nonlinear Processes in Geophysics (in particular they cannot really guess that they are meant to be optimal in some sense).
We have added a reference: "These values are chosen such that the mean and variance of the inputs are preserved between two consecutive layers (Klambauer et al., 2017)"

14. l.123-124: "We set the batch size to 96 and do 100 epochs." $-\rightarrow$ "We set the batch size to 96 and run 100 epochs."?
fixed

15. You should have use the latex package linenofix.sty. Your line numbering has issues!
We use the Copernicus Publications Manuscript Preparation Template for LaTeX Submissions. Now that all equations are numbered, the line numbering is also fixed.

16. Please number all of your equations. This is customary – this facilities the study of your paper by colleagues and students. Systematic numbering may be avoided in reports and book to avoid cluttering.
Agreed, we have fixed this.

17. p.5: Equation defining the loss function (no number and line numbers skipped): Why do you take the square root and not the MSE which is available in Tensor-Flow/Keras?

Because we also look at RMSE when verifying the data assimilation results, so this was just an easier direct comparison. We checked that using the MSE of TensorFlow/Keras yields similar results. However, the tables look very different when expressed in terms of MSEs instead of RMSEs. For example, the improvement in terms of RMSE is 32%, whereas in terms of MSE it is 59%.

18. l.119: "The python library Keras (Chollet et al., 2015).": (i) You are actually using TensorFlow/Keras or TensorFlow 2.x. – your statement is a bit weird. (ii) Please give the reference to Chollet's book instead, which is the Keras bible as well as an excellent introduction to TensorFlow/Keras and more generally deep learning (Chollet, 2017).

fixed

19. It would be better to provide your codes. Maybe not all pieces, but for instance the original ones like the convection model and the TensorFlow code.

We will provide the code.

20. l.135 and Figure 2: Did you average your RMSEs over several learning and/or test experiments? It is possible that the curves are significantly dependent on the initial random seed. If not, I do not expect any unpleasant surprises but more reliable (and less noisy) curves, potentially with error bars. Please clarify.

We averaged over 500 experiments. This information is now included in the Figure captions.

21. p.9; Table 2 caption: "As table 1, but for" $\longrightarrow$ "Same as table 1, but for". Same remark for Figures 5 and 6, and maybe others(?).

fixed

22. l.156-165: It may be that the CNN is actually correcting for other sources of model errors such as the impact of localisation. That would explain why EnKF+CNN can outperform QPEns.

Yes, that is a good point since the CNN has an influence radius of 5 grid points and the data assimilation a radius of 8 grid points. We therefore trained an additional CNN with the kernel of all layers of size 5, so that the influence radius is 10. This gave us however similar results as in Figure 5 and 6. We added this discussion in text: "Since the CNN only has an influence radius of 5 grid points and the localisation cut-off radius of the data assimilation is 8 grid points, it is possible that the better results of the CNN stem from this shorter influence radius. However, a CNN trained on the same data but with kernel sizes of 5 instead of 3 (leading to an influence radius of 10 grid points) yields similar results as in Figures 5 and 6 (not shown)."

23. l.175: the sentences are a bit awkward, I suggest (2 corrections): "the CNN was able to reduce the mass violation significantly. Moreover,"

fixed

24. Acknowledgements: There seems to be a useless " at the beginning.

fixed