Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2020-38-RC1, 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

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

Marc Bocquet (Referee)

marc.bocquet@enpc.fr

Received and published: 24 October 2020

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.
- A few relevant and more recent references could be added (*recent* is very short in this subject).

Printer-friendly version

Discussion paper

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

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.
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).
3. l.17: “Janjić (2016),Zeng et”: a space is missing.
4. “Artificial neural networks (NN), are powerful tools” $\longrightarrow$ “Artificial neural networks (NN) are powerful tools”

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

6. l.28: “based on example” $\longrightarrow$ “based on examples”?

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

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.

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?

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).

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?

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.

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).

14. l.123-124: “We set the batch size to 96 and do 100 epochs.” ⟶ “We set the batch size to 96 and run 100 epochs.”?

15. You should have use the latex package linenofix.sty. Your line numbering has issues!

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.

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 TensorFlow/Keras?

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).

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.

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.

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

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.

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

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

References

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2020a. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model. J. Comput. Sci. 44, 101171. doi:10.1016/j.jocs.2020.101171.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2020b. Combining data assimilation and machine learning to infer unresolved scale parametrisation. Philosophical Transactions A 0, 0. URL: https://arxiv.org/pdf/2009.04318.pdf. accepted.

Chollet, F., 2017. Deep Learning with Python. Manning Publications Company.

Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B.R., Girvan, M., Ott, E., 2018. Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model. Chaos 28, 041101. doi:10.1063/1.5028373.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2020-38, 2020.