Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-7-AC1, 2019

Nonlinear Processes in Geophysics Discussions
Open Access

Interactive comment on "Data assimilation as a deep learning tool to infer ODE representations of dynamical models" *by* Marc Bocquet et al.

Marc Bocquet et al.

marc.bocquet@enpc.fr

Received and published: 28 May 2019

May 28, 2019

Subject: Manuscript npg-2019-7: "Data assimilation as a deep learning tool to infer ODE representations of dynamical models", Response to Reviewer 1.

Dear Reviewer,

We wish to thank you for your comments and suggestions, that we have taken into

Printer-friendly version

Discussion paper

account to improve the manuscript. We respond below to these comments and tell how we modified the manuscript accordingly. A pdf file showing the differences between the original and the revised manuscript is provided.

Problem of the title: The authors should not use "deep learning" in their title. This is a confusing point because some readers might think that this paper is dealing with ML algorithms but this is not the case. However, the discussion in Sect. 3 about the link between the presented method (based on DA) and deep learning is interesting, especially when comparing equations (17) and (18). Moreover, the authors nicely show that using a weak constraint 4D-Var is a way of controlling the backpropagation (important part of deep learning methods).

We have removed “deep” from the title. As your subsequent comments suggest, the frontier between ML and data assimilation is not clear-cut. The method presented in this paper can certainly be seen as a learning method, while the optimisation of deep leaning networks is similar to that used in variational data assimilation.

I think the authors should write more context in Sect. 2. First, in Sect. 2.1, the monomial basis should be discussed, giving more explanation and illustrating for instance with the Lorenz-63 (as Brunton did).

As an example, we now list the monomials in the case of the L63 model.

I have the same remark in Sect. 2.2.1 where the drastic reduction of columns of A due to locality could be illustrated using for instance the L96 system.

As an example, we now list the monomials in the case of the L96 model for the case $L = 2$.

As for $\mathbf{A}$, we provide an example a few lines later in section 2.2.2.

Printer-friendly version

Discussion paper

Note also that we have improved Appendix A, which is closely related.

Then, Sect. 2.2.2 about the homogeneity (same behavior of the model at different locations) is hard to follow: again, authors should try to illustrate.

We have clarified the text and illustrated the enumeration of the coefficients of $\mathbf{A}$ in the L96 case, with $L = 2$, and assuming homogeneity in addition to locality.

Finally, entire Sect. 5 is very technical and I suggest to keep the important results/equations in the main text and move the rest to the appendix.

Following your suggestion, we have moved sections 2.5.1 and 2.5.2 to appendix B.

p3, l4-5 $\rightarrow$ can you give a quick explanation of the difference with the neural network approach?

We have slightly improved the sentence: "This marks a key distinction with respect to our approach where the dynamics to be determined are explicitly represented, as will be clarified later.", though we cannot go into much detail in the introduction.

p6, l5 $\rightarrow$ what do you mean by one-dimensional state space models? The models you introduce in Sect. 4 are all multi-dimensional, please clarify.

One-dimensional refers to the dimension of the physical space over which the model is defined (extend models), not the number of variables (resulting from the discretisation of the model). The sentence was indeed confusing. We have clarified this point in the revised manuscript.

p16, l8-9 $\rightarrow$ this is not always the case, especially when you use recurrent neural *networks, that are not deep in practice.*

Thank you for spotting this issue. We have rephrased the sentence in order to avoid confusion: "Finally, we note that, as opposed to most practical deep learning strategies with a huge amount of weights to estimate, we have reduced the number of control variables (i.e. $\mathbf{A}$) as much as possible."

p21, l16-17 → in addition to Fig. 8, for different values of observation noise, I would like to have a look at the estimated coefficients of A along the DA cycles.

In the revised manuscript, we now plot the value of the coefficients of $\mathbf{A}$ as a function of the minimisation iteration index for the L96 model with and without noise in the observations.

p23, l20 and Fig. 9 → what do you mean by "long space-time stripes"? This point needs clarification with maybe a zoom on Fig. 9 to make this point clear.

We have clarified the sentence which becomes: "The emergence of error, i.e. the divergence from the reference, appears as long darker stripes on the density plot of the difference (close to zero difference values appear as white or light colour)."

The figure has been modified and a zoom over the difference density plot is now shown, where the stripes are patent.

Typo: - p23, l2 → "the the"

Corrected. Thank you.

Marc Bocquet, Julien Brajard, Alberto Carrassi, Laurent Bertino.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-7, 2019.