Geoscientific
Model Development
Discussions

Interactive comment on "Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model" *by* Julien Brajard et al.

Julien Brajard et al.

julien.brajard@locean-ipsl.upmc.fr

Received and published: 26 August 2019

Dear Peter Düben,

we first wish to thank you for your comments and suggestions that have helped us to improve the manuscript. We report below your original remarks (in blue) followed by our responses.

Section 3.3.2: This is quite a specific network architecture that you are using. Can you

provide more detail how you discovered it? Can you speculate whether the "×" and "+" are required due to the underlying shape of the equations of the Lorenz model?

The following explanation was added to section 3.3.2: "The specific bilinear layer aims at facilitating the training in the case where multiplications are involved in the true model. This design is therefore based on a priori knowledge of the underlying model, given that multiplicative terms are ubiquitous in ODE-based geophysical models. For instance, this is the case of the model described in Eq. (9) used to illustrate the work."

I assume that you have 1-D periodic boundary conditions for the network

You are right, this point was explicitly stated in section 3.3.2: "As the numerical experiments in the following are conducted on a periodic spatial domain, (see section 3.1), the input of the neural network is accordingly 1-D periodic at the boundaries."

Figure 2: It took me a while to understand that 2a, 2b and 2c are in parallel. This is not intuitive from the figure.

It is true the figure 2 was unclear, a new figure (hopefully more intuitive) is proposed in the new version of the paper (See figure 1)

The situation of unobserved portions of the space would be more challenging. This aspect has now been clarified in section 3.1: "Note that variables have a non-zero probability to be observed in some instance. In the more challenging case where portions of the field are never observed (e.g. fixed observing network, hidden parameter, ...), the setup may have to be adapted to account for an unobserved latent subspace, as it proposed for instance in Ayed et al. (2019)."

Figure 4: This may be my ignorance but I would have expected to see the high fre-

quencies to be correct and the low frequencies to be incorrect since you are basically training on a timestep level. Do you have any comments on this?

We have elaborated on this result in section 4.3 by adding the following comment: "The errors are thus higher at higher frequencies and there is no significant improvement throughout the optimization process (see the grey and orange lines after 5 Hz in Fig. 4). The low frequencies are better observed (there are 4 times as many observations for each 2 Hz oscillation as for an 8 Hz oscillation) and better reproduced after the DA step, whereas high-frequencies are not reproduced from observations."

Section 4.5.2: Could this configuration therefore be used to tune stochastic parametrisation schemes? This could maybe be discussed. We had some success using GANs and dropout methods to develop neural network parametrisation schemes for Lorenz 95 that showed some variability.

We agree it is actually an ongoing direction of extension of the current work we are considering. The following sentence was added at the end of section 4.5.2: "Another ongoing direction of the extension of the current work could be to estimate stochastic model errors. For example generative models (generative adversarial networks or variational auto-encoders) could be used to tune stochastic parametrisation schemes."

P18: Could this also be made more efficient by training on interpolated observations in a first instance with no need to use the entire data assimilation scheme?

Actually, this is what is done (see section 3.3.3). But thank you for stressing out that this point needs clarification. Following your comment, we also stress the gain in term of computing time by adding the following sentence in section 3.3.3: "Starting with random weights as it is usually done in ML training algorithms could lead to convergence problem of the two-step algorithm and also will be inefficient in term of computing time."

- Caption Figure 7. "0. 50
- Caption Figure 8: "with with"
- P16: "if was" -> "if it was"
- P18: "parallel computing" I would suggest to call this "concurrent computing"
since you do not refer to standard MPI/OpenMP parallelisation.
- "resolvent" could be explained a bit more.
- Abstract: "applies alternatively" could be re-phrased.
- P2 l11: "precipitations" -> "precipitation"
- P5 l1 and P7 l14: There are unnecessary line breaks.
- P7 l3: It could be stated that σ
obs is a rather arbitrary choice at this stage.
- One of the references is incomplete: "E, W.:"

Thanks for pointing out those points. They have been addressed in the new version of the paper. Except for the line break P7 l14 which was intended to make the equation compatible with a two columns format. Also, the E.W. reference is correct. Although unusually short, E is the real family name of the first author.
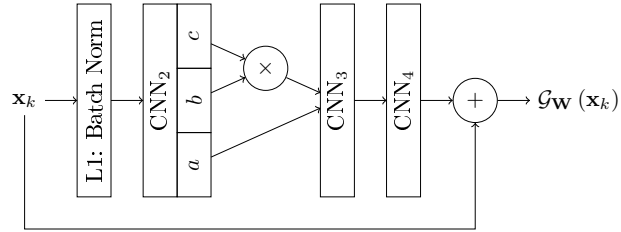

Fig. 1. Proposed neural network architecture for the surrogate model. The input layer is to the left.

C5