Review of https://doi.org/10.5194/egusphere-2026-245: v1

# Elucidating the performance of data assimilation neural networks for chaotic dynamics

May 7, 2026

The manuscript explores properties of machine-learning (ML) data assimilation networks (DAN) in a simple chaotic model (L96). The DAN is used to perform the (incremental) analysis in the forecast-analysis tracking cycle, while the forecast is carried out with the true dynamics. The surprising success (from `Boc24`) that a single forecast simulation is sufficient to achieve cutting-edge accuracy performance is the main point reasoned about. Three main approaches are used in the investigation (1) Computation of the "marginal gain", defined as the sensitivity (wrt. the forecast) of the DAN (linearised wrt. the "projected" innovations). (3) Benchmarks of the tracking accuracy performance of several variations of traditional ensemble methods and ML DANs.

Furthermore, locality of the marginal gain is exposed, which enables the DAN to reduce the dimensionality of large-scale systems. In addition a linear-in-$\zeta$ DAN is constructed and used to explain that the DAN implicitly operates similar to a Kalman filter (perhaps the authors could correct me on the point here?).

The manuscript is of high quality and the results are impressive. Figure 3 is a treasure that I hope does not get lost to readers who don't feel the need of the 'elucidation' of this manuscript. I can recommend its publication upon some minor changes, as described in the following.

## Chronological comments

L2-5 : I'm really struggling to understand the difference of these two approaches.

Eqn. 5 : This and the surrounding reasoning seems somewhat circular. Does it define the expansion, $P$, or its computation or approximation? It becomes clearer eventually, but not at the outset.

L112 : The definition of $\zeta$ should be more prominent.

L116 : ~~hence~~ $\longrightarrow$ implying

L166 : $\Delta_t$ has not been specified yet, diminishing the usefulness of stated RMSE values.

L175 : this 'assumption' is usually tied to the quasi-linear character of some dynamics, not the 'premises' stated above ? Also what does "close to optimal" mean in this setting?

Eqn. 8 : I would suggest displaying the equation $x^a = x^f + P^a\zeta$ somewhere nearby here, or relating it to eqn. 2.

L185 : I'm not really clear on what was achieved here. The reasoning seems somewhat circular to me. Please clarify.

L198 : linear-in-$\zeta$ deserves a clear/prominent definition.

§3.1.3 The results and arguments put forth herein seem relevant, but it is not clear to me how they fit together (so I object to the last sentence). Despite touching on some strong theoretical results, they do not paint a clearer picture than what can be gained simply by using scatter plots of Lorenz-63 to illustrate how the trajectories drift apart, and how this depends on the initial scatter's position.

L223 and L225 : Is the following really a 'sensitivity analysis'? It may well be so, but perhaps the term could be elaborated on.

L228 : I struggle to understand the choice of the word *marginal* and would suggest that it be clarified, or renamed ('differential'?).

Eqn. 10 : Should one not use $\partial$ rather than $\nabla$?

L232 : I believe one also needs to assume optimality of the DAN and linearity of the dynamics in order for the marginal variation to coincide with the Kalman gain?

Eqn. 12 : I'm surprised that Agarwal (2021) gets cited for this result and not Raanes et al. (2019), given that it seems earlier and closer to the first author's background? Also note that Stordal et al. (2016) derived the same expression for the gradient of a pre-smoothed function.

L250 : ~~second-order moment truncation~~ $\longrightarrow$ Gaussian (as already alluded to on `L241`).

Eqns. 14, 15 : This seems like a lot of symbols (for the reader to parse) to make the innocuous approximation of $\mathbb{E}_\zeta \, . \approx .|_{\zeta=0}$ . It also does not appear to rely on Stein's lemma.

L270 : Isn't the group $\mathcal{G}_y$ of induced isometries *necessarily* isomorphic to $\mathcal{G}$ ? In other words, isn't the assumption superfluous?

§3.2.2 : I don't know enough group theory to follow this in my cursory reading. As far as I can tell, it does seem to serve to set up Eqn. 21 for $\Sigma$. My intuition also suggests that the authors are seeking to perform a kind of "circulant average" (in linear algebra terms), I'm assuming across the tensor dimension representing $\partial_x$ ? Could the authors try to provide a dumbed down summary (in the manuscript)?

Eqn. 21 : Unfortunately I got even more lost by this simple looking definition (which is a shame, since it is an object central to the manuscript). Unfortunately the typesetting makes it easy to confuse brackets and the "floor" operator used right below. But I don't even really see why brackets are in use at all; earlier cases of indexing did not appear to require them, except (I see now!) eqn. 10. But why is $r$ kept inside? And if $r$ is an index, then $\Sigma$ has 3, so how is this now a 2-tensor?

L343 : The omission of IEnKS is disappointing. Are there not (at least) some already published benchmarks out there (from the same experimental settings) that could be referenced?

L366 : ~~passed~~ $\longrightarrow$ beyond

Fig. 4 : Caption: ~~$N$~~ $\longrightarrow N_{iter}$

L459 : I'm a little sceptical of the claim that the DAN must (implicitly) be learning such a map, in light of the fact that the linear-in-$\zeta$ DAN performed so poorly (relative to nonlinear ones).

## Unsorted comments

- Is linear regression really an "expansion"? The term can be confusing if it's always restricted to the 1st order. What about "linearisation"?
- The code should be available to the reviewers. Its absence is disappointing and mystifying.
- It would be useful to have a dedicated comment elucidating the performance of the DAN linear-in-$\zeta$ and the DAN with linear activation.
- Add comment somewhere relating to the well-known closure of the moment equations in the case of linear dynamics. In addition, the state update depends on the covariance, but not the other way around. From

this one could perhaps intuit that there is a lot to be gained from introducing such dependence when generalising to the nonlinear case. Indeed, one could argue that it should be more fruitful than targeting the non-Gaussian priors gained through ensemble forecasting.

- I have not fully perused `Boc24`, but it bears repeating herein anyway: Have you investigated if there are any benefits to operating DANs based on *ensemble* forecasts?

Signed, Patrick N. Raanes

## References


Patrick Nima Raanes, Andreas Størksen Stordal, and Geir Evensen. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, 26(3):325–338, 2019. `DOI: 10.5194/npg-26-325-2019`.

Andreas S. Stordal, Slawomir P. Szklarz, and Olwijn Leeuwenburgh. A theoretical look at ensemble-based optimization in reservoir management. *Mathematical Geosciences*, 48(4):399–417, 2016. `DOI: 10.1007/s11004-015-9598-6`.