Nonlin. Processes Geophys. Discuss., 1, C78–C81, 2014
www.nonlin-processes-geophys-discuss.net/1/C78/2014/

Nonlinear Processes in Geophysics
Discussions
Open Access

Interactive comment on "Data assimilation of two-dimensional geophysical flows with a Variational Ensemble Kalman Filter" *by* Z. Mussa et al.

Anonymous Referee #2

Received and published: 25 April 2014

The article follows on from work by Solonen et. al. (2012), which introduced the Variational Ensemble Kalman Filter (VEnKF) , and Amour et al. (2013) which looked the improvement gained over a free running shallow water model by applying the scheme. This article looks at the performance of the scheme in two examples – a QG model and the same shallow water model as was used by Amour et al. (2013).

My main issue with the article is that it is unclear how much it adds to the literature over the previous work. Although a lot of detail has been given introducing the scheme and the models used the actual results do not substantially add to the understanding of the performance or ability of the scheme beyond the results given in previous papers and are rather confusing in the message that they convey.

(i) The first example of the QG model is given as a proof of concept example. However the work by Solonen et al. (2012) has already introduced the scheme and shown how it performs in comparison with the EnKF and EKF in a low dimensional nonlinear example and a high dimensional linear model. It is unclear from reading the article what more using the QG model with the scheme reveals about the performance of the scheme. There doesn't seem to be any real discussions about what the results given demonstrate about the scheme and why they are significant, either in this section or the conclusion. Why is this model chosen? What can we learn about the scheme by using it? What results back any conclusions made? Why is the convergence with the number of ensemble members important?

(ii) The second example looks at whether cross flows introduced into the model are picked up in the analysis solution of the data assimilation. Again, I'm unclear as to why the results shown currently are important (this is probably partly due to my lack of knowledge about hydrological models). A discussion section here could be used to explain this and the results be put into context. From the results shown it seems to me that the point here is rather more to do with the ability of data assimilation to also enable the capture of cross flows rather than the particular ability or advantage of the VEnKF. If the point is to demonstrate that the VEnKF is a scheme that should be adopted then again evidence should be shown as to why I would want to use this particular scheme (currently there is no comparison to any other method).

(iii) Throughout the paper it is also not clear from the description given how the observations are incorporated into the data assimilation scheme: (a) In the first example it appears that every variable at every time step is observed. This needs to be clarified and also means that the system being studied is a very highly observed one. Do the results shown hold true when less variables are observed at more infrequent time intervals? Are more ensemble members required with less observations before convergence is achieved? (b) In the second example it seems that time interpolation was used to ensure observations at every time point. Does this mean that when the observation interval was increased there are bigger time steps used in the model. This appears to be what is shown by Fig. 11-13. If not, what happens to the solution between the observations in these figures? If the time interval of the model is increased then this needs to be explained in more detail. This leads to the question of how does the scheme work if there are multiple time steps between the observations? (c) I am also seriously concerned by the interpolation of data done in the second experiment. In data assimilation the point is to spread information that is available to unobserved variables through the Cp matrix not by interpolating observations. Doing this not only strongly violates the independent observation error assumption made but also confuses the issue of the effect of replacing the full matrix in the EKF with its ensemble representation in the VEnKF. In particular in this dam break experiment does it not impact the observations on the release of the water to then interpolate this information in time and space? The aim in data assimilation is to improve the information coming from sparse observations and a numerical model by combining both. In the results shown it would appear that the data is used as the best estimate. Would it not be possible to potential run some sort of synthetic experiment so that a truth was available. From these results there is no evidence that just interpolating the data does not give the best solution without the need for a model at all.

(iv) The final section of the paper (3.4) links the observation interpolation to the Kalman filter stability. I'm not entirely sure what is being done in this section. My understanding from the previous set-up was that the biggest gap between observations was 0.14s for a 0.1s time step. This implies that the majority of time steps have observations. The results of section 3.4 seem to suggest that actually observations were available a lot less frequently. Exactly what is being done needs to be explained in a lot more detail to fully understand the relevance of what is being presented and before it is possible to comment on the results seen. However, an initial assessment would be that the observations play an important role when minimising the cost function (equation 4). Hence it is to be expected that there is a link between the observation interpolation distance and the spread in the ensemble. A highly observed system is much more

constrained by the observations and so the spread in an ensemble would be expected to be much smaller.

To me it feels as though this is almost two papers that have been combined into one. Although there is some potential for good work in the article it is unclear exactly what overall question the paper is attempting to address. It may be better to consider splitting it into two separate articles. The first could look at the VEnKF as a data assimilation scheme using the QG example and explore the scheme in a lot more detail. The second article could then look at the Shallow Water model from the point of view of improving results in comparison to observations from hydrological models. If the paper is going to be kept as one article then consideration should be given as to the overriding question that is being answered by the paper and then results tailored to demonstrate the conclusions drawn. If the idea is to demonstrate why the VEnKF should be chosen above other schemes then this may change considerably the emphasis of the second example and require new results to be generated.

Other comments are that, although there is a lot of good and useful information in the introduction, the relevance of the information to the article is not always clear. A more pointed introduction that explains exactly why previous work is lacking and how this scheme is of benefit would make the introduction much stronger. Also the English does not always flow very well and although the article is understandable it could be immediately improved by having a good or native English speaker read through and correct it.

There may be additional comments on top of these but I feel that the high level content is in need of such substantial re-working that I haven't looked into the small scale issues.

Interactive comment on Nonlin. Processes Geophys. Discuss., 1, 403, 2014.