Hydrol. Earth Syst. Sci. Discuss., 12, C2109–C2113, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C2109/2015/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Joint inference of groundwater-recharge and hydraulic-conductivity fields from head data using the Ensemble-Kalman filter" *by* D. Erdal and O. A. Cirpka

Anonymous Referee #1

Received and published: 17 June 2015

MAIN COMMENT

The paper discusses the interesting issue of the difficulty to identify simultaneously hydraulic conductivity and recharge from piezometric head observations. For this purpose the authors use the ensemble Kalman filter and build a synthetic experiment to conclude that unless the prior information used to generate the initial set of realizations is "correct" it is impossible to identify simultaneously both hydraulic conductivity and recharge.

Although the authors prove nicely their point, the arguments given to justify the final

results are flawed: the emphasis should not be put in the prior information but rather in the need of extra information to be able to single out the combination of hydraulic conductivity and recharge that is correct out of the many combinations that are coherent with the observed piezometric head.

It is wrong to say that the prior information used to generate the initial set of ensemble realizations determines what the final estimates will be (after data assimilation) and that you cannot generate realizations outside the prior random function model. The final ensemble of realizations can largely depart from the initial one, when the observations are inconsistent with the prior model. The sequential set of ensembles that are obtained after each assimilation step can be interpreted as a Markov chain that will "forget" the structure built in the initial ensemble after some assimilation steps. More so, the updating of the ensemble realizations is solely based on the covariance structure of the ensemble, and for this reason, there is a tendency for the final ensemble of estimates to converge towards realizations drawn from a multiGaussian function, even when the initial ensemble is far from being multiGaussian. Enough observation data can change completely the random function of the final ensemble with regard to the one used to generate the initial ensemble.

What happens in the example presented by the authors is that the observations are consistent with all prior models used. In fact, the reference could be the one used, or it could be a realization generated with the "wrong" model, the results would be the same. Therefore, you need some additional piece of information, to discriminate among the different prior models which one is the one consistent with your unknown reality. It is not that the EnKF does not work when the prior model is incorrect. Knowing which is the orientation of the deposition will draw to prefer a prior model over another.

My request is that the paper be rewritten removing all these comments about the importance about the prior information, and the influence that this prior information has in the final ensemble, and replacing it by talking about the importance of having additional information that would allow you to discriminate among alternative models. I contend

that assimilating data on fluxes or concentrations would also alleviate the problem of the prior model. And I insist that the prior model information will fade away as time passes and data are assimilated, and could vanish if the prior model is inconsistent with the observational data.

ADDITIONAL COMMENT

I do not think there is any need to present the extended Kalman filter equations, especially when they are not used to justify the use of the ensemble Kalman filter. In this respect, there are some conceptual misunderstandings about the ensemble Kalman filter that must be corrected. First of all, there is no need to make any multi-Gaussian assumption to get to the ensemble Kalman filter equations, like there is no need to make any multi-Gaussian assumption to get to the cokriging equations; it is true that under the multi-Gaussian assumption the ensemble mean and ensemble covariance would be the mean and covariance of the conditional distribution given the obsevations; however, from the point of view of optimal estimate in the a least-square sense, the Kalman filter equations do not need any multi-Gaussian assumption.

I do not quite understand the last paragraph in page 5576 when the authors say that the EnKF is a linearized estimate that is alleviated by the repeated application over many time steps. The authors should understand that the EnKF captures the linear relationship that there is between the parameter and state variables through the experimental covariance –that's all–, the fact that you apply the updating equations over many time steps does not "alleviate the effects of non-linearity". The reason why the EnKF works and the extended Kalman filter did not is because the covariances are computed on parameters and states which have been obtained by solving the state equation through an ensemble of realizations, and therefore are much closer to the "true" covariance than the one obtained by propagating the initial covariance in time through a linearization of the state equation.

Please revise your presentation and discussion of the EnKF.

OTHER COMMENTS

Since the state (piezometric heads) is not updated, you should explain how the state is computed after each assimilation step. Is the model rerun from time zero with the updated parameters?

Page 5577, line 2, the original prior knowledge is smeared out after the first assimilation step by the Kalman gain, it carries over at the beginning but it will eventually disappear.

Page 5577, line 8, if the prior knowledge is erroneous and there are sufficient observation data inconsistent with that prior model the estimates will converge to the "truth".

Page 5578, why is NRMSE only computed at the observation wells and not over all the aquifer, since you have the reference information?

Page 5578, it is unclear what is the denominator of the equation, is it the ensemble variance? or is it the prior measurement uncertainty (in which case it is a constant value)?

It would have been nice to see some variance maps along the ensemble mean maps.

Top of 5583, in real applications you need information to discern among alternative combinations of conductivity and recharge, this information could help you in choosing the prior model, or it could be other types of data (such as fluxes).

End of 5583. No, multiple-point statistics will not help here, that is, it will not allow you to discriminate between a good and a wrong model as long as the observation data are consistent with those models. Besides, it has been proven that the EnKF will filter out the non multi-Gaussian characteristics of the initial ensemble of realizations.

SUMMARY

I liked the paper and I think it should be published, but only after the emphasis in the conclusions is shifted.