Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-544-AC2, 2016


Hydrology and Earth System Sciences
Open Access
Discussions

# *Interactive comment on* "A Bayesian Consistent Dual Ensemble Kalman Filter for State-Parameter Estimation in Subsurface Hydrology" *by* B. Ait-El-Fquih et al.

**B. Ait-El-Fquih et al.**

boujemaa.aitelfquih@kaust.edu.sa

Received and published: 30 March 2016

\- **Reply to Referee #2** -

We would like to thank the referee for carefully reviewing our work and for his/her constructive comments. We have revised the manuscript taking into account all the referee comments and suggestions whose detailed replies are given below.



**Main concerns**

C1. *A one-step-ahead smoothing based dual EnKF is presented in the manuscript. The authors compared results of the new method with standard joint and dual EnKF. As mentioned by the authors, Gharamti et al (2015) proposed a new EnKF method too by combining the one-step-ahead smoothing formulation (page 3). What will the result look like if these two one-step-ahead based EnKF methods are compared? Also the dual EnKF OSA needs to be better distinguished from the one by Gharamti et al (2015).*

R1. We thank the referee for bringing this up. The referee raised two points:

- *Comparison of the filtering schemes:*

  We have examined the performance of the Joint-EnKF$_{\mathrm{OSA}}$ of *Gharamti et al. (2015)* compared to the new filtering schemes in the first section of the results. We plot the resulting state and parameter estimates on top of the previous ones in Figure 4 and Table 3. The Joint-EnKF$_{\mathrm{OSA}}$ is found more reliable during the early assimilation period but later and towards the end of assimilation, the proposed new scheme becomes clearly more accurate.

- *Differences between the algorithms:*

  The proposed Dual-EnKF$_{\mathrm{OSA}}$ in this work results from the generic algorithm of Section 3.1 by applying two random sampling properties (see Appendix A) under the Gaussian assumption. At each assimilation step of the Dual-EnKF$_{\mathrm{OSA}}$, the observed data are used three times through Kalman-like updates: twice in the smoothing step (one for smoothing the previous state as in Eq. (27) and one for updating the parameters as in Eq. (28)), and once in the "forecast" step to compute the analysis of the current state as in Eq. (35).

The work in *Gharamti et al. (2015)* follows a similar approach and applies the same two random sampling properties on the generic algorithm of Section 3.1, but under the following assumption (beside the Gaussian assumption)[1]:

$$p(x_n|x_{n-1}, \theta, y_n) = p(x_n|x_{n-1}, \theta), \tag{1}$$

which is based on the fact that given the previous state, $x_{n-1}$, and the parameters, $\theta$, the current state, $x_n$, is independent of its observation, $y_n$. Following this assumption, the "forecast" step of the generic algorithm which is composed of a Bayesian step (24) followed by a propagation step (23), reduces to the propagation step (23). As stated in page 445 of *Gharamti et al. (2015)*, the assumption (2) has been adopted to compute the analysis pdf, $p(x_n|y_{0:n})$, from the joint smoothing pdf, $p(x_{n-1}, \theta|y_{0:n})$, based on Eq. (9) (or Eq. (23) in our manuscript), by avoiding the use of the computationally demanding term $p(x_n|x_{n-1}, \theta, y_n)$ and replace it by the more easily sampled state transition pdf, $p(x_n|x_{n-1}, \theta)$. Here, we propose a more efficient approach (see pages 10-11 and Appendix B) to directly sample from the analysis pdf without explicitly computing $p(x_n|x_{n-1}, \theta, y_n)$ and without the need of any additional assumption.

Now, when applying the two random sampling properties (under the Gaussian assumption), the Joint-EnKF$_{\text{OSA}}$ in *Gharamti et al. (2015)* shares the same smoothing step as the Dual-EnKF$_{\text{OSA}}$, which, as mentioned above, involves two Kalman-like updates. However, in contrast with the Dual-EnKF$_{\text{OSA}}$, the Joint-EnKF$_{\text{OSA}}$ does not involve a Kalman-like update in the "forecast" step (because of the omission of the Bayesian step (24) from the generic algorithm). In terms of computational cost, these algorithms have almost the same cost as they require the same number of model runs; the only difference is one Kalman update for each member which is generally computationally not consequent compared to the cost of integrating the model. The differences between the proposed Dual-

[1] Refer to Eq. (16) in *Gharamti et al. (2015)*.



EnKF$_{\text{OSA}}$ and the Joint-EnKF$_{\text{OSA}}$ of *Gharamti et al. (2015)* were made clearer in lines 332-343 of the revised manuscript.

C2. *Page 8. The authors talked about one-step-ahead smoothing function but did not explain what is this function used for? What role does it play in the new algorithm dual EnKF OSA and how does it work?*

R2. We understand that the referee is referring to our statement (page 8, lines 245-246): "*...but also involves a (new) smoothing step that constraints the state with the future observation.*". This suggests that compared to the standard Dual-EnKF, the proposed Dual-EnKF$_{\text{OSA}}$ involves a new update step of the state using the future observation (hence the term "one-step-ahead (OSA) smoothing"). This smoothing "function" is given by Eq. (27), being a Kalman-like update of the state analysis members using the future observation. Line 317 of page 11 was updated to clarify the Kalman-like update character of the smoothing "function". Thank you.

C3. *Page 11. The observation data are used three time in dual EnKF OSA rather than twice as in the dual EnKF. The authors thus concluded that it is in a fully consistent Bayesian formulation. Please clarify the relationship between them. Also please relate this with the comment on standard dual EnKF that the ensemble does not represent the forecast pdf (page 8 lines 231-233).*

R3. We thank the referee for pointing this out. The proposed Dual-EnKF$_{\text{OSA}}$ is not a fully Bayesian algorithm because the observed data are used three times, but because this algorithm results from the generic (theoretically sound) Bayesian filtering algorithm presented in Section 3.1 by applying two random sampling properties (see Appendix

A), and under the (commonly used) Gaussian assumption only. As stated in our reply R1 above, at each assimilation time step of the resulting Dual-EnKF$_{\mathrm{OSA}}$, it turns out that the observed data are used three times through Kalman-like updates: one for smoothing the previous state (Eq. (27)), one for updating the parameters (Eq. (28)) and one for updating the current state (Eq. (35)). Accordingly, the use of the observed data three times is a consequence of the fully Bayesian character of the proposed scheme and not a cause. This was made clearer in the text below Eq. (35).

Regarding our statement on standard Dual-EnKF that the ensemble does not represent the forecast pdf and therefore it is a heuristic algorithm (page 8, lines 231-233), please note that this is related to the (Bayesian consistent) Joint-EnKF and not to the Dual-EnKF$_{\mathrm{OSA}}$. Indeed, as one can see from Sections 2.2.1 and 2.2.2, the Joint-EnKF and the Dual-EnKF algorithms mainly differ in the state analysis step. The analysis members of the state, $x_n^{a,(m)}$, are computed by the Joint-EnKF using a Kalman-like update of the forecast members, $x_n^{f,(m)} = \mathcal{M}_{n-1}(x_{n-1}^{a,(m)}, \theta_{|n-1}^{(m)})$ given in Eq. (5), while the Dual-EnKF computes $x_n^{a,(m)}$ following the same mechanism but using the members $\widetilde{x}_n^{f,(m)} = \mathcal{M}_{n-1}(x_{n-1}^{a,(m)}, \theta_{|n}^{(m)})$ given in Eq. (12), instead of $x_n^{f,(m)}$. However, in contrast with $x_n^{f,(m)}$, the members $\widetilde{x}_n^{f,(m)}$ are not samples from the forecast pdf (more generally, one does not know from which pdf $\widetilde{x}_n^{f,(m)}$ are sampled) since they are obtained by integrating the model with the updated parameters, $\theta_{|n}^{(m)}$ (along with $x_{n-1}^{a,(m)}$), instead of $\theta_{|n-1}^{(m)}$. Despite that, these are used in the analysis step as forecast members to compute the analysis members, $x_n^{a,(m)}$, following the same Kalman-like update as in the Joint-EnKF, hence the heuristic nature of the Dual-EnKF.

Now, to relate that with the Dual-EnKF$_{\mathrm{OSA}}$, one can easily see that this latter reduces to the Dual-EnKF in the particular case of a perfect model and in the absence of state smoothing (i.e., when Eq. (27) vanishes and $x_{n-1}^{s,(m)} = x_{n-1}^{a,(m)}$ in Eq. (33)). This was made clearer in page 11, lines 330-331 of the revised manuscript.



C4. *Page 17 line 527-528 "The proposed dual EnKF OSA efficiently recovers the reference trajectory of MW2 and MW3". I think this statement is not so proper since the trajectory or trend of the reference is not captured well by any method, including the dual EnKF OSA. The reference is barely covered by the ensemble and the peak values are late in the ensemble at MW3. But I agree that the dual EnKF OSA works better than the other two methods.*

R4. We have relaxed our argument here and rephrased the sentence accordingly. The sentence now reads as follows: "The proposed Dual-EnKF$_{\mathrm{OSA}}$ performs fairly well, providing a reasonable recovery of the reference trajectory at MW2 and MW3.". Thank you.

C5. *Page 17 line 534 "We further examine ..... against the joint- and dual- EnKFs ...." But in Figure 7 only the results by dual-EnKF and dual- EnKF OSA are shown. The results by joint EnKF are not included.*

R5. We thank the referee for pointing this out. We have now included the results from the Joint-EnKF in Figure 7.

C6. *Page 18 line 576 "the dual- EnKF OSA tends to maintain a larger variance at the edges than the dual-EnKF, which in turn increases the impact of the observations". I have two questions on this: first, it is difficult to tell (Figure 10) the "larger" variance by dual- EnKF OSA. Their results look pretty similar. Secondly, why will the variance of hydraulic conductivity at the edge increase the impact of the observations (at 9 wells) which is not so near from the edges? Furthermore, the boundary conditions are either no-flow or constant head which have limited influence on observations. One suggestions on the color bar of the figure: the white color does not appear in the bar* *but occupy a lot of area in the Figure and the contours make the figure complicated.*

R6. We have improved the quality of the figure. Essentially, we removed the contour lines and kept the colouring. We further adjusted the colour bar to better emphasize the different behaviours of the schemes. In terms of the larger variance at the north boundary, we refer to the fact that an ensemble with larger spread (variance) is expected to fit more the observations using a Kalman-based update. However, if the ensemble spread is very small and the estimate is still far from the truth then the impact of assimilating future data would be minimal.

**Minor corrections**

C1. *Page 2 line 38, "Hendricks Franssen and Kinzelbach, 2009" instead of "Franssen and Kinzelbach, 2009". At the same time, please correct the item in the reference list (page 22).*

R1. Done. Thank you.

C2. *Page 5 line 150. "Let, for an ensemble ..., $r$ denotes" should be "denote" instead of "denotes"?*

R2. Done, thank you.

C3. *Page 10. The equation 25 looks exactly the same as equation 5. Is this correct?*



R3. Yes, this is correct. However, mechanisms (5) and (25) are the same, but not their "outputs". In other words, the forecast members $x_n^{f,(m)}$ resulting from (5) are not the same as those obtained by (25) even when starting from the same "input" $(x_{n-1}^{a,(m)}, \theta_{|n-1}^{(m)})$ (which is very unlikely). This is because the noise $\eta_{n-1}^{(m)}$ in (5) are different than those of (25) even though they are sampled from the same Gaussian distribution $\mathcal{N}(0, Q_{n-1})$.

C4. *Figure 1, the black crosses represent hard measurements according to the text on Line 425. But it can also be added here to avoid any confusion.*

R4. Done. Thank you.

C5. *Title of the subsection 5.5 "further assessment of the dual- EnKF OSA scheme" does not reflect the content. From the title we expect the result by dual- EnKF OSA only. But in fact it still compares the results of three methods. It could be changed to "prediction capability assessment" or something like this.*

R5. We thank the referee for the comment. This title was changed as suggested.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/hess-2015-544/hess-2015-544-AC2-supplement.zip

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-544, 2016.