Hydrol. Earth Syst. Sci. Discuss., 10, C2951–C2957, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C2951/2013/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter" *by* V. R. N. Pauwels et al.

V. R. N. Pauwels et al.

valentijn.pauwels@monash.edu

Received and published: 3 July 2013

We thank the three Referees for the generally positive comments on our manuscript. We have replied to the questions that were raised in one single document.

Answers to the queries raised by Anonymous Referee 1

1. The authors provide their argument for using the gamma and kappa tuning factors in the bias correction framework. However, it is desirable to see the effects of using or not using the tuning factors in the bias estimates. Could the authors mention in which cases the tuning factors are compulsory needed?

This is a very basic question about the theoretical framework of the method. In Section 7.2 we describe the way we have obtained the parameter values for this application. Here, we also describe how an erroneous value leads to a drift in the bias estimates. This erroneous bias estimate will clearly lead to a degradation of the other results. To answer the specific question of the Reviewer, we will add in this section that an analysis similar to the one we performed (investigating the autocorrelation of the innovations) should always be performed when applying the method.

2. The authors test the bias correction filters using both synthetic and real observations. It is known that when testing a framework validated with synthetic experiments, it suffers degradation in its performance. Could the authors provide insights about the factors or parameters that were not included in the synthetic study but affect the bias correction filters in real scenarios?

This is another good point. Any synthetic experiment will always be limited by its assumptions. More specifically, real observation and process noises are unlikely to be Gaussian and stationary. On the other hand, synthetic experiments do provide the opportunity to investigate every aspect of the method. More specifically, they allow a modeller to assess the impact of the methodology on every variable of the model, even those who cannot be observed.

In order to assess the applicability of the method, which was up till then validated synthetically, we have described a real-world experiment in Section 7.8. The hardest part for a real-world study is the determination of good values for gamma and kappa. From this experiment, we can conclude that, at least for the test site studied, the method provides realistic results. In this section, we will add the statement that further investigation of the method is needed, using more complicated models in more challenging environments (for example in the presence of snow, frozen soils, etc.). We will also add that this investigation is outside the scope of the paper, which has as objective the development of the theoretical framework and the initial assessment using a very simple model.

Answers to the queries raised by Dr. Kollat

Dr. Kollat raises a number of editorial improvements, which we have all implemented. He also suggests to add a figure with the observations, the baseline results, and the assimilation results. We will add this figure, which is now Figure 8 and is discussed in Section 7.3.

Answers to the queries raised by Anonymous Referee 3

1. One important concern is about some key operating hypotheses necessary to the formulation and the application of the two-stage DKF- EnKF approach devised by the Authors. These hypotheses require the covariance matrix of the model bias to be a fraction of the biased state error covariance matrix, and the covariance matrix of the observation bias to be a fraction of the biased state error covariance matrix projected onto measurement locations. These assumptions may seem quite far-fetched unless a reasonable physical basis is provided for them. From this perspective, the Authors need to critically address the validity of these hypotheses much more thoroughly, and explain when, in their view, they are valid and when they are not.

This question is related to the first question of Referee 1. Theoretically, it is trivial to show that the unbiased error covariance is always the sum of the biased error covariance and the error covariance of the forecast bias. The question that then remains is how to partition the biased error covariance. Essentially, if the biased error covariance would consist only of the error covariance of the bias (thus gamma is equal to zero), that would imply that there is no random error in the model results, and that all the error is caused by bias. On the other hand, if gamma would be equal to one, this would imply that all the error in the model results is random and that there is no systematic error. The latter is the case in the Kalman filter without taking into account bias.

As shown in the appendix and in Section 2.5, if an initial value of the unbiased error covariance and the covariance of the biases would be known, these could be propagated separately, at least for a linear system. But then the results would be dependent on

these initial values, and a similar reasoning could be made regarding the partitioning of the biased error covariance (which fraction is attributed to bias and random error). For a nonlinear system, this becomes even more complicated, since the model matrix A_k is now varying over time, and the results would become even more sensitive to this initial fraction. For this reason, it is better to develop a more robust methodology. As stated in Section 3.3, the autocorrelation length of the innovations should be zero, if the biases are correctly estimated. Thus, if good values for the filter parameters (gamma and kappa) are chosen, the filter works in a theoretically consistent manner. We will add this explanation in Section 3.3.

2. Another important concern is about the quality of the presentation of the "Results". Several figures are low quality and unclear, which makes it very difficult to draw conclusions based on the application results. In several instances, these figures are not adequately presented and commented so that the reader is left to look into them on her/his own without adequate guidance.

Please see our replies to the specific questions in the annotated manuscript.

3. In broader terms, I suggest the Authors to expand Section 7 and make Sections 2 and 3 more concise.

Again, please see our replies to the specific questions in the annotated manuscript.

4. The Referee also suggested some improvements in an annotated copy of the manuscript. The small editorial changes we have all implemented, and answers to the more substantial questions are:

a) P2 nr 1. To avoid confusion, we will add the reference to Drecourt to the reference to Kollat.

b) P2 nr 2. We do not agree with the statement that in state augmentation, it is possible to implement independence between the biases and the state variables. In state augmentation the equations for the Kalman filter are applied without modification, which implies that the relationship between variables is determined by the model.

c) P3 nr. 1. We have tried to develop our methodology as closely as possible to the way models are applied. More specifically, if nothing is known about the bias, the state vector is propagated through time, using the model equations. The challenge is to estimate the bias using this framework. This should also clarify remark 9 on the same page.

d) P3 nr. 2. We will delete the part of the sentence as requested.

e) P3 nr. 5. We would really prefer to keep these equations, as they are a foundation of the theoretical development. What these equations imply is that, if no data are used to update the model states and biases, the bias at the next time step is simply assumed to be the same as in the previous time step. This is also reflected in Equation 8. However, if there are data assimilated into the model, then the biases are updated. In order to avoid confusion, we will add an explanation regarding the bias updates in Section 2.1.

f) P4 nr. 2. We agree that it would be a good idea to replace the listing of equations by a figure. We will remove the listing of equations by Figure 1, and the references later in the paper to these equations will be replaced by references to the figure.

g) P4 nr. 3. We will add the explanation that this could be obtained by repeatedly running the model with the same forcings, until convergence is obtained.

h) P5 nr. 2. We will add the explanation that this is the difference between the unbiased observations and simulations.

i) P5 nr. 5. We will, as suggested, remove this paragraph.

j) P6 nr. 1. If the model is unbiased (and the observations as well), then this should be reflected in the choise of the parameters gamma and kappa, which should both be estimated to be zero, looking at the autocorrelation of the innovations. We will add this explanation to the last paragraph of Section 3.3.

k) P7 nr. 1. We will, as suggested, rephrase this paragraph.

l) P7 nr. 5. We will add this advantage to the Section with the conclusions, and also in the abstract.

m) P7 nr. 6. We are not perfectly clear about the question. At this point we are already discussing a separated state and bias estimation, with an ensemble and discrete Kalman filter, respectively. If state augmentation is applied, then indeed the biases need to be estimated with an ensemble as well.

n) P7 nr. 7. Please see our reply to the first general remark by the Referee.

o) P8 nr. 3. We will add a clarification to this paragraph. Basically, the filter parameters are modified until the innovations are proven to consist of white noise (thus with autocorrelation zero).

The ensemble of unbiased states is not stationary, since the statistics can and will change over time. However, they do need to be white noise.

p) P8 nr. 6. Here, we would prefer to keep the equations. They are a part of the description of the method, and the reader can easily use this summary.

q) P8 nr. 7. Please see our earlier reply to question 1 on page 3. q) P9 nr. 2. We will rearrangedthis explanation as suggested.

r) Thank you for noting this. We will use the symbol 'psi' instead of 'gamma'.

s) P11 nr. 1. We will add, in Section 7.1 (Ensemble generation) the explanation that we have calculated the ensemble statistics with larger ensembles, and that similar results were obtained.

t) P11 nr. 2. We will clarify that this is the case.

u) P11 nr. 4. We will, as required, provide a more detailed clarification.

v) P12 nr. 2. We will clarify that the baseline run is the model application without data assimilation.

w) P12 nr. 3. Since we are dealing with discharge observations, which follow a seasonal cycle, we have attempted to mimic this by using a periodical bias. While this may be a strong simplification of the true bias, it is still more realistic than simply using a constant bias. We will add this explanation to this paragraph.

x) P13 nr. 1 and 2. We will clarify what the observability matrix is. We would really like to keep this section separate from the other sections. This section presents an extra analysis of the method, and a justification to develop it further. Adding this section to other sections would make the other sections more confusing.

y) P14 nr. 1. We will, as suggested, redraw the figure. This should eliminate any misunderstandings.

5. Regarding the figures that should be improved: We will increased the font size of all figures. Where extra explanation is demanded, we will provide this. This should eliminate the confusion about the results.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 5169, 2013.