Anonymous Referee #1

Thank you for your comments. We have here tried to answer these point-by-point.

Comment: The introduction does not have enough depth. Significantly more work has been done on bias estimation through data assimilation in hydrology, and hardly any of this work has been discussed. At least a good effort is needed to improve this.

Changes: We have added more references to parameter estimation in data assimilation as well as bias estimation and have hereby placed our work in a broader perspective.

Comment: A number of issues regarding the data assimilation algorithm are very unclear. On page 8137, the authors state that a linear operator H is used when assimilating data. But the authors assimilate discharge, and the relationship between discharge and ground water levels and soil moisture is nonlinear. We need more detail on how the data assimilation system is set up. Clearly define which variables are in the state vector, and which in the observation vector.

Reply: A linear operator is used as stream discharge is observed directly and so is groundwater head. The state vector contains both of these variables plus stream water level, all of which are model states. As such, there is no need for a nonlinear operator. However, we agree that the state vector is not clearly defined.

Changes: In section 2.4.1 (State variables), the following is inserted: "In this study, the state vector contains groundwater head, stream discharge and stream water level, all of which are updated at each updating time step."

Comment: Also on page 8137, what is the "observation covariance"? Is this the observation error covariance?

Reply: Yes.

Changes: Changed.

Comment: On page 8139, it is explained how the localization weight is calculated. But it is still not clear to me how exactly this is used. Please provide some more explanation.

Reply: We have expanded on the description, and refer the reader to Rasmussen et al. (2015).

Changes: Inserted into section 2.3.2 (Localization): "As each state vector member is analyzed in the ETKF, the ensemble of model observations (i.e. the ensemble of model states at the observation locations) is generated, and the sample correlation coefficient between each of the model observations and the state member is determined. The localization weights of the observations to the state member being analyzed are then calculated from the correlation coefficients as follows."

Comment: Also on page 8139, make it more clear how H is extended. It is stated that H is extended accordingly, but that is not enough detail. Please also see my earlier comment on the linearity of H.

Reply: We agree that the description is not adequate.

Changes: The sentence is changed to "The mapping matrix H is extended according to equation 10", with equation 10 showing the relationship between the extended state vector and the vector of model observations.

Comment: I have a serious problem with using covariance inflation. A number of papers have shown that when the ensemble is adequately generated, this is not necessary. More-over, this inflation will make the algorithm inconsistent with its theoretical derivation, and therefore will make it work suboptimal. This needs to be at least mentioned and discussed.

Reply: We agree that the use of covariance inflation is unfortunate. However, the issue has been widely researched and is commonly used in a multitude of data assimilation applications. As the reviewer remarked, covariance inflation is unnecessary if the ensemble is adequately generated, but this is not likely to be the case in this study, as the uncertainty of different aspects of the physical system is unknown.

Changes: The issue with using covariance inflation as described by the reviewer has been noted in the paper: "Using covariance inflation is, like using localization, inconsistent with the derivation of the filter and only necessary due to inadequate or incorrect noise description and ensemble generation. However, due to the complex nature of the model, generating an ensemble that perfectly represents the uncertainty of the model is difficult and particularly in the test using real data outside the scope of this paper." References to applications of covariance inflation have been added to the first paragraph of the section.

Comment: I also have a question with using damping (page 8140). This is (the way I understand it) not consistent with covariance inflation. First, this inflation is applied to make sure the updates are sufficiently large, but then the damping is applied to reduce the updates. This needs a much stronger justification. Again, this is inconsistent with the theoretical derivation of the filter, and will thus make the results suboptimal. And I really think that both these tricks (inflation and damping) could be avoided through a better ensemble generation.

Reply: Damping, like inflation and localization, does indeed make the filter suboptimal. However, with the complexity of the physical system, the model, and the state vector structure, we found that "tricks" such as these are necessary, and that mathematical optimality is not feasible with the current knowledgebase. This also seems to be the general opinion in the scientific community, as witnessed by the multitude of data application research in which inflation, localization and damping is used without much discussion. Damping does reduce the updates, but it does not cancel out the inflation; not partly nor completely. After the update of states and parameters, damping actually helps to maintain the spread of the ensemble, as the update is likely to reduce the spread of the ensemble to a near collapse.

Changes: None.

Comment: Below equation 15, please again provide more details on the augmentation of H.

Changes: The equation showing the relationship between the unbiased model observation and the biased model state has been included.

Comment: Equation 21: The updated bias is calculated as the old one minus the gain multiplied by the innovation. I have a question about this minus. In all papers on observation bias estimation there is a plus here. Assume that there is a large bias between observations and results. From equation 21 this bias will reduce. Plugging this reduced bias into equation 22 the unbiased states will increase while they shouldn't. Perhaps it is a typing error but please double check this. If it is correct, I would suggest to explain why the minus is there, as opposed to the other papers on observation bias estimation.

Reply: Correct; the minus is a simple typing error.

Changes: Corrected.

Comment: Section 2.3.7.: why not update the states each time a discharge observation is available? This would or at least should lead to better results.

Reply: Updating the states every time observations are available is very time consuming and more importantly was found to induce numerical instability in the model after the update. Updating causes certain spin-up effects, in which the groundwater part of the model attempts to return to stable state, and updating the states before the equilibrium has occurred led to increasingly erroneous updates and eventual numerical instability. Furthermore, the slow dynamics of groundwater means that little or no change in groundwater head occurs in one day, and updating would be unnecessary. Any “Stand-alone” updating of stream discharge (without updating the groundwater head) would quickly be cancelled out by the groundwater-stream interaction.

Comment: Section 2.4.1.: This is a mistake that is made in a large number of papers in hydrology on discharge assimilation. Discharge is NOT a state variable, it is a diagnostic variable. State variables have to be seen as initial conditions, to which you apply the model equations, and then you get the results that you update with the Kalman filter. Discharge is simply NOT a state variable, it is a model output. If you enter the discharge in the state vector, and you check the requirements of observability and controllability, they would not be fulfilled. If one wants to assimilate discharge into a hydrologic model, the discharge has to be entered in the observation vector, and the soil moisture and water table levels in the state vector. The observation system is in this case NONLINEAR. There are a number of papers on this, that the authors really should read. Please note that this is the same principle as for example assimilating radar backscatter values or surface brightness temperatures into a hydrologic model.

Reply: It is correct that for typical lumped, conceptual rainfall-runoff models discharge is not a state variable. In this case the state typically consists of the water content in different conceptual reservoirs. However, since the model is predicting the discharge (i.e. the model includes the inverse observation operator), one can include discharge in the state vector and thereby obtain a linear observation operator (see e.g. Clark et al., 2008). With the MIKE SHE model the state description is different. The MIKE SHE model includes the MIKE 11 river model, which uses an alternating calculation scheme, where calculations of discharge and water levels are performed in every other calculation point, and both variables are required to be updated and used as initial conditions. Both stream discharge and stream water level are therefore included in the state vector and updated simultaneously with the groundwater head. The observation operator is in this case linear, as both discharge and groundwater head are observed directly.

Comment: The last sentence before section 3.1.2. is a little bizarre: "Test have shown that the length of the assimilation window is of little importance and therefore no other assimilation window was tested." Doesn’t this sound a bit contradictional?

Reply: Agree.

Changes: Changed to “Precursive tests have shown that the length of the assimilation window is of little importance and therefore no other assimilation window lengths were evaluated in this study.”

Comment: Section 3.1.2.: I do not agree with the statement that updating every observation is too often, I actually think the opposite is true. This may be the result of issues in the setup of the filter, as I explained in my earlier comments.

Reply: See earlier reply.

Comment: Last sentence before section 3.3.: If you plug an ensemble of forcings and parameters into a nonlinear model, and even if these ensembles are unbiased, it is very likely that the resulting ensemble of model results will be biased, because of the nonlinearity. Please add this explanation.

Changes: Added: “Note that due to the nonlinearity of the model, the noise, while unbiased, is likely to cause some bias in the ensemble of model results.”

Comment: Just a detail: section 3.4.: why are data this old being used?

Changes: Added to section 2.2.1 (The Karup catchment): The catchment was the subject of an extensive monitoring campaign between 1970 and 1990, and the large amount of observations available makes the catchment an excellent subject for the study of hydrological data assimilation.

Comment: Section 3.4.2.: are the RMSEs biased or unbiased?

Reply: The RMSEs are biased, but are still (with reservations) an expression of the ability of the model to recreate the dynamics of the process as observed in the observations.

Changes: None.

Comment: Page 8155: It is true that both methods were tested in Drecourt et al, but they looked at model biases, not observation biases. This should be clarified.

Changes: Clarification inserted: “[is similar to the one derived and presented in Drecourt et al (2006)] but modified to estimate observation bias rather than model bias”

Comment: End of page 8156 and top of page 8157: you could actually calibrate this gamma parameter. Why not try this?

Reply: Calibrating the gamma-parameter is a good idea, but we feel that it falls outside the scope of the current study. It is however something that we will consider in follow-up studies.

Changes: None.

Anonymous Referee #2

Thank you for your comments. We have here tried to answer these point-by-point.

Comment: Page 8131, Lines 12-20. There have been other data assimilation studies that have focused on updating system state variables and system parameters in an integrated hydrological (groundwater-surface water) framework. These probably should be mentioned. They include:

Kurtz, W., H.-J. Hendricks Franssen, P. Brunner, and H. Vereecken (2013), Is High-Resolution Inverse Characterization of Heterogeneous River Bed Hydraulic Conduc-tivities Needed and Possible? Hydrol. Earth Syst. Sci. 17 (10): 3795–3813. doi:10.5194/hess-17-3795-2013.

Bailey, R.T., and Baù, D. (2012), Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. Hydrology and Earth System Sciences, 16, 287-304.

Changes: The above references have been added to the paragraph.

Comment: Section 2.2.2 – what is the discretization of the stream network? Is it the same that is used for the aquifer? Are the groundwater and surface water processes coupled, or just linked? (linked = no iteration during the time step, just passing values between the stream model and the aquifer flow model)?

Changes: The following has been inserted into the section "The stream model network is set up using an alternating calculation scheme in which discharge and water level is calculated respectively in alternating points, and is independent from the groundwater model discretization Exchanges of water between the two processes is taken place in the groundwater model grid cells where a river branch is present.. The exchange takes place at every groundwater model time step.".

Comment: Section 2.2.3 – spatial variability of streambed parameters (i.e. "leakage coefficient controls") has been a focus of research during the past few years, particularly in integrated hydrological modeling. How does using spatially-uniform stream model parameters influence the model results? Could this have an impact on the data assimilation results, particularly since some of the observation wells are close to the stream network and hence could be influenced by spatially-variable groundwater-surface water interactions?

Reply: The use of a spatially variable leakage coefficient would without a doubt have improved model performance and would have been preferable, but we believe it not to be feasible at the current scale and especially at the current discretization of the groundwater model. Furthermore, we believe that the current discretization is a much bigger source of error in the groundwater-stream flow interaction than the uniform parameterization of the leakage coefficient. Partly due to the errors in the groundwater-stream flow interaction, the observation points closest to the stream network was omitted from the assimilation as described in section 3.1.1.

Changes: None

Comment: Page 8142, Line 18. Why choose a standard deviation of 0.6 m? Is this based on field data? Were other values tested?

Reply: The chosen value is based on the multitudes of tests and experiments that were performed before the results shown in this paper. The impact of the parameter was not deliberately tested, as the parameter did not appear to be particularly important to the estimation of bias.

Changes: The following is inserted in to the section: “The standard deviation vas chosen based on precursive testing in the synthetic test environment, that showed that this value generally led to the best estimates of bias.”

Comment: Page 8145, Lines 20-21. I am confused by this. Isn’t the point of the DA methodology to estimate the parameters? (i.e. “calibrate” the model?) So then why is the model calibrated using AutoCal? I am not sure how this fits into the general aims of the study.

Reply: We believe there has been a misunderstanding. The calibrated model is only used as a reference point to compare the parameter estimation of the data assimilation algorithm.

Changes: The sentence is extended for clarification: “[the model is also calibrated using AutoCal] in order to be able to compare the parameter estimation through data assimilation with parameter estimation through more common method, such as inverse modelling”

Comment: Section 2.4.2 – Is Hydraulic conductivity spatially-uniform throughout the catchment? Is this realistic? It seems that K should be specified as spatially-variable (ac-cording to geostatistics), and the K field should be updated using the system-response measurements.

Reply: As described in section 2.2.3, the parameterization of the groundwater model, including K, is based on a 3D geological model. The parameters are assigned to each unit of the geological model, and this parameterization is mapped to the 2D computational grid in a preprocessing step (built into the model code), that follows each parameter update.

Changes: Added to section 2.4.2: “Note that the estimated hydraulic conductivities are those of the geological units, that are gridded to the computational grid before further propagation of the ensemble (see section 2.2.3)”.

Comment: Section 2.3.7 - Please provide more information regarding the “Asynchronous assimilation”. Are the daily discharge measurements averaged over the 28 days, and then the average discharge is assimilated at the update time?

Reply: The method does not average discharge, but involves saving the individual asynchronous observations and model results, and thus improving the update at the time of updating through correlation in time.

Changes: Added to the section: “The state vector is extended with model results for asynchronous observation times and the observation vector is extended with the asynchronous observations. After that, the asynchronous observations and model results are simply treated as normal model states.”

Comment: Page 8140, Line 21. Change “hereby” to “thereby”

Changes: Changed.

Comment: Section 3.4. A 1-year warm-up period does not seem long enough to provide a significant spread in the ensemble, given the slow travel time of groundwater. Could you please quantify the spread of the ensemble at the end of 1969, to demonstrate that enough spread occurred?

Reply: The catchment is geologically dominated by sand, and thus highly permeable. An ensemble spread therefore builds up very fast.

Changes: Added to section 3.4: “At the end of the year 1969, the spread of the ensemble of groundwater head is between 0.8 and 2.1 m (depending on the location in the catchment), which is considered sufficient for assimilation to commence.”

Comment: In the Results section, please provide a 1:1 plot (simulated vs. observed) of groundwater head for some of the scenarios. Perhaps show a “before” and “after” plot (without and with data assimilation) to demonstrate the improvement of the hydrologic system when DA is used. Also, a plot to compare the results of the different scenarios, with the ensemble mean used for the simulated results.

Reply:

Changes: 1:1 plots of groundwater head in selected observations in selected observation locations has been added, including a comparison of the base model (i.e. no DA) and SepFil and ColFil_ens200 scenarios. Plots comparing the results of different scenarios (mean of the ensemble) are already present in the paper, namely figures 5 and 8 showing head as a function of time in the synthetic test and real data tests respectively.

Comment: Section 5: please provide conclusions, rather than just a summary of the study and discussion of results. What are the implications of the results? How can results be used in future studies, particularly in applications to real-world watershed systems?

Changes: Added to the end of the conclusion: “The study has shown that hydrological observational bias can be corrected in a data assimilation scheme and that it can improve state updating and parameter estimation. With both model- and observational bias being significant sources of error in hydrological modelling which will have a negative impact on the performance of data assimilation in hydrological models, the results provide an important advancement of application of hydrological data assimilation in large scale, integrated hydrological models.”

Anonymous Referee #3

Thank you for your comments. We have here tried to answer these point-by-point.

Comment: The first issue is that, despite the authors' claim that they "discuss the challenges associated with coupling two processes (groundwater and stream flow) in a single filter", the discussion is strongly biased towards the performance of the filters, little attempt is made to relate the results to the physical processes occur-ring within the catchment, and model results are never showed in terms of stream discharge. The scarcity of information given about the model itself does not help in this respect. I understand the model is not new and has been used in many studies before, but I think a more detailed description is warranted, especially concerning the coupling between the various hydrological components.

Reply: We agree that the wording cited in the comment is not representative of the paper's content and focus. However, the focus on the filter performance rather than the specific physical processes modelled in this study, which is reflected in the content and structure of the paper, is intended. The previous paper, Rasmussen et al. (2015), has a broader focus in which the two processes are evaluated separately and compared.

Changes: The sentences cited have been replaced by the following: "We discuss the challenges associated with observational bias in hydrological data assimilation for both state updating and parameter estimation. Two existing methods of estimating observation bias, the separate bias Kalman Filter and the augmented state vector approach, are tested and the results compared. The novelty of the study lies in the focus on data assimilation bias estimation in a complex, integrated hydrological model as well as the impact of bias on parameter estimation using both synthetic tests and real world observations." Furthermore, more information has been added on the model coupling in section 2.2.2.

Comment: Another main issue of this paper is that it refers extensively to another (companion?) paper from the same authors, also submitted (and very recently published) to HESS. As far as I understand, the only difference between this paper and the other is that here bias correction is considered, while all the remaining contents of the present manuscript (methods, model, study area) have been already included in the previous paper. In other words, the results of this study provide only a marginal contribution to the literature. It is not for me to decide whether the contents of this manuscript alone warrant publication in HESS, but I honestly think that the results presented here could (and should) have been included in the previously published paper as additional sections.

Reply: It is true that the model, study area and part of the method are similar to that of the previous paper, which is why it is referenced so extensively. The added information of this paper is indeed bias estimation, but also the assimilation of real observations, which is still not common at this scale – particularly in conjunction with bias estimation. As such, we believe that the new findings presented in this paper can stand alone and that the findings are novel and thus warrant a separate paper. However, as stated in the comment above, we agree that the objective of the paper, as described in the introduction, does not convey this, and it has therefore been changed.

Changes: Same as in the above comment.

Comment: Page 8136: most of Section 2.2.2 is a model description, not setup. I suggest merging it with Section 2.1.

Changes: The section has been merged into section 2.1.

Comment: Also, more details about the model are needed, especially as regards the coupling. E.g., what are the parameters "drain level" and "drain time constant" and what is their physical meaning? Are they relevant for the coupling of surface and subsurface flow?

Changes: A description of the parameters has been added to section 2.2.3 (Model parameterization): “The drain level and drain time constant parameters control the amount of groundwater drained to the nearest stream once the groundwater table exceeds the drain level, and are as such linking the groundwater module and the stream flow module of the model. The drain simulation represents the subsurface tile drain systems installed under most farmlands as well as the natural lateral flow processes that often occur in the topsoil. The leakage coefficient is another coupling parameter, which represents the hydraulic properties of the thin layer of the sediments at the bottom of the stream. This parameter is of particular importance with regard to river base flow.”. Also, the following has been added regarding the coupling between the processes: “The stream network model is set up using an alternating calculation scheme in which discharge and water level is calculated respectively in alternating points, and is independent from the groundwater model in discretization, but exchanges of water between the two processes is made available in model grids of the two processes that physically overlap.”.

Comment: Regarding the setup, what are the initial and boundary conditions used in this study?

Changes: The following is added to section 2.1 (Model): “The groundwater model initial conditions is based on an extended warm-up of the model, in which a quasi-steady state develops, while the stream flow initial conditions was calculated assuming a steady state condition.”

Comment: Page 8137, lines 5-7: as far as I know, even a standard EnKF does not require the full covariance matrix, as the product HP(PH)T can be assembled directly. Please rephrase the sentence.

Changes: The sentence has been changed to: “It is computationally more efficient than the EnKF and is furthermore deterministic, meaning that the observations in the filter are not perturbed, which reduces the impacts introduced by sampling uncertainty”.

Comment: Page 8141: Eqs (12), (13), and (14) can be merged into a single equation.

Changes: The equations have been merged into a single equation.

Comment: Page 8142, lines 9-13: in my experience, discharge observations in natural rivers can be as biased (if not more) than groundwater head observations, due to the need of a rating curve that is often accurate only for low flow rates and extrapolated for high flow rates. This statement should be relaxed, or at least appropriate references should be provided to justify it.

Reply: We agree that bias on discharge observations can be as significant as groundwater head bias and that the sentence should be relaxed.

Changes: The sentence has been changed to: “This study assumes no bias in discharge observations, meaning that the only biased observations are the groundwater head observations. In real world observations, discharge observations would usually also be biased, but this bias is not considered in this paper in order to simplify the problem.”

Comment: Page 8142, lines 14-18: this is not clear. Either the initial bias is zero in all the locations or it is generated from a distribution with 0.6 m standard deviation and 0 mean. Please clarify.

Reply: We agree that the sentence is unclear.

Changes: The sentence has been changed to “In this study, the initial estimate of bias in all observation points is generated by sampling from a normal distribution with a standard deviation of 0.6 m and a mean of 0.”

Comment: Page 8148, line 16: I can see seven scenarios in Table 1, not five.

Changes: Corrected.

Comment: Page 8152, lines 2-5: this explanation for the reduction of NS coefficient when passing from an ensemble size of 50 to 100 and increase when passing from 100 to 200 is not convincing. Definitely more details are needed here to explain the model behavior. For instance, I suggest adding to Figure 3 (in another panel) some comparison between the true discharge and the discharge in the assimilation scenarios. Also, why don’t you show in Figure 3 the results of the other scenarios (SepFil, SepFil NoQEst, and NoBias)? Finally, the open loop results (simulations without data assimilation) should be added as well, to evaluate the real benefits of data assimilation in this series of simulations.

Reply: The same issue with stream discharge (i.e. the peaks caused by spurious correlation) was observed and discussed repeatedly in Rasmussen (2015). The peaks in stream discharge from spurious correlation dominate the performance indicator, which makes comparison between scenarios difficult due to the random nature of spurious correlations. We do however agree that this point is not sufficiently well described in the paper.

Changes: The following is inserted into the section: “As shown by Rasmussen et al. (2015), these spurious correlations are likely to result in increased drainage to the stream model, resulting in large errors in stream flow. The errors from spurious correlations in the stream flow model dominate the performance indicator and are, due to the nature of spurious correlation, random. [As a result, the Nash-Sutcliffe coefficient is reduced when using an ensemble size of 100, but increased with an ensemble size of 200.]”

Comment: Page 8152, line 26: the reference to Figure 5 is given before any reference to Figure 4.

Changes: The sequence of figures has been corrected (i.e. figures 4 and 5 have been switched).

Comment: Also, I would expect more detailed comments about Fig. 5 other than “little drifting behavior is observed in the model”.

Reply: The figure is referenced for the drifting behavior for each scenario (three) and finally referenced, described and discussed in depth in the section in which the scenarios are compared.

Changes: The following is added to the discussion of the figure: “Figure 5 shows that the drifting behavior is generally most pronounced in the NoBiasEst scenario and least pronounced in the ColFilEns200 scenario, with the drifting behavior of the SepFil scenario falling in between these two scenarios.”

Comment: Page 8153, lines 4-7: I don’t see many differences in the drain level bias between SepFil and SepFil NoQEst, only in the drain time constant. Also, if these parameters are so important, they must be defined and discussed in more detail in the model description.

Reply: Correct; the bias in estimated drain level is not significant.

Changes: The sentence has been changed to: “The reduction in NS is explained by a bias in the estimated drain constant (Figure 4) and by a poorer description of the groundwater head level as indicated by the head RMSE”. Furthermore the parameters have been discussed in more detail (se previous comment)

Comment: Page 8153, line 26 to page 8154, line 5: is this comment based on results showed in some figures?

Reply: The statement is based on experience with the model and the data assimilation framework. But we agree that the sentence does not convey this, and should be changed.

Changes: The sentence is changed to: “It is observed that updating the groundwater head to a biased observation level causes the head to return to an unbiased level when model propagation is resumed (i.e. it is drifting as seen in Figure 5). The model behavior becomes unnatural in the sense that it is not controlled primarily by the input forcings, but rather by the model trying to retain equilibrium. This can result in deteriorated estimation of parameters and updates of model states not only in the observation points but in the entire model domain.”

Comment: Page 8155, Section 4.2: is the “base” model an open loop simulation? Please clarify.

Reply: The base model is described in section 3.1.1 (Synthetic test observations) as a deterministic model. However, we agree that this needs clarification.

Changes: Added to section 3.1.1: “Note that both the true model and the base model are deterministic models, that is, single, propagated models without any noise added.”

Comment: Also, as for the synthetic tests, I suggest adding and discussing a figure showing the model results in terms of stream discharge. In my opinion, as the subject is an integrated hydrological model, it is important to investigate the model behavior with respect to all its hydrological components.

Reply: As the model is integrated, we agree that it is important to evaluate both groundwater head and stream discharge, which is why we use the stream discharge Nash-Sutcliffe coefficient as one of the three indicators of performance. It is our belief that this indicator serves the purpose of evaluating the modelled stream flow, and that graphical representation of the stream flow would not add significant information to the paper.

Changes: None.

Technical corrections

Comment: Page 8156, lines 5-12: this paragraph is repeated twice, please delete.

Changes: Deleted

Comment: Figures 4 and 10: please add units to the parameters.

Changes: Units added.

Comment: Figure 9: correct the caption. This figure does not refer to the synthetic tests.

Changes: Corrected.

Manuscript prepared for J. Name
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 17 January 2016

Data assimilation in integrated hydrological modelling in the presence of observation bias

Jorn Rasmussen[1], Henrik Madsen[2], Karsten Hogh Jensen[1], and Jens Christian Refsgaard[3]

[1]Department of Geosciences and Natural Resource Management, University of Copenhagen, Denmark.
[2]DHI, Horsholm, Denmark.
[3]Geological Survey of Denmark and Greenland, Copenhagen, Denmark.

Correspondence to: Jorn Rasmussen (jr@ign.ku.dk)

Abstract. The use of bias-aware Kalman filters for estimating and correcting observation bias in groundwater head observations is evaluated using both synthetic and real observations. In the synthetic test, groundwater head observations with a constant bias and unbiased stream discharge observations are assimilated in a catchment scale integrated hydrological model with the aim of updating stream discharge and groundwater head, as well as several model parameters relating to both stream flow and groundwater modeling. The Colored Noise Kalman filter (ColKF) and the Separate bias Kalman filter (SepKF) are tested and evaluated for correcting the observation biases. The study found that both methods were able to estimate most of the biases and that using any of the two bias estimation methods resulted in significant improvements over using a bias-unaware Kalman Filter. While the convergence of the ColKF was significantly faster than the convergence of the SepKF, a much larger ensemble size was required as the estimation of biases would otherwise fail. Real observations of groundwater head and stream discharge were also assimilated, resulting in improved stream flow modeling in terms of an increased Nash-Sutcliffe coefficient while no clear improvement in groundwater head modeling was observed. Both the ColKF and the SepKF tended to underestimate the biases, which resulted in drifting model behavior and sub-optimal parameter estimation, but both methods provided better state updating and parameter estimation than using a bias-unaware filter.

1 Introduction

Sequential assimilation of observations in models is a widely used method in several fields, including meteorology and hydrology. The method has repeatedly been shown to improve forecasting performance, reduce uncertainty, and optimize parameter values, and is still a topic subject to on-going research.

Data assimilation in hydrological models has been studied in a number of settings, from single process models, modelling only a limited part of the hydrological cycle (e.g., Franssen et al., 2011; Albergel et al., 2008; Moradkhani and Sorooshian, 2005), to integrated models incorporating all the relevant processes including precipitation, evapotranspiration, recharge and streamflow (e.g., Camporese et al., 2009; Shi et al., 2014; Rasmussen et al., 2015). The latter presents a number of challenges that have yet to be comprehensively addressed; particularly relating to the differences in process time scales, e.g. between groundwater flow and surface runoff, and the coupling between these processes. An integrated approach to hydrological modeling is however important in many applications due to the exchange of water between the hydrological components; thus it remains important to explore these aspects. In Camporese et al. (2009) the Ensemble Kalman filter (EnKF) was applied to an integrated model of a synthetic tilted v-catchment and both stream discharge and groundwater hydraulic head observations were assimilated to update both groundwater and stream states. Shi et al. (2014) applied the EnKF to an integrated land surface hydrological model of a small catchment and, using seven different observation types, successfully estimated six parameters and sequentially updated the model states. Rasmussen et al. (2015) used the Ensemble Transform Kalman Filter to assimilate groundwater head and stream discharge in a catchment scale integrated hydrological model for both state updating and parameter estimation. Other studies that focus on joint state updating and parameter estimation in integrated hydrological modelling include Bailey and Baù (2012), in which a smoother was used to calibrate hydraulic conductivity using stream flow and head observations, and Kurtz et al. (2013), which used head observations to calibrate heterogenous river bed conductivities.

Biases in both models and observations pose challenges to data assimilation in hydrology, and have previously partly been studied (e.g., Dee and da Silva, 1998; Dee, 2005; Reichle et al., 2004; Lannoy et al., 2007; Bosilovich et Bias is found in all components of the hydrological cycle, and take a variety of forms. Notable examples are model bias stemming from model structure or parameter errors, and observation errors, which is due to the difference in scale between point observations and gridded model variables. The latter is a significant source of bias in many groundwater models, as the horizontal discretization of the models is often large. If one is to update the groundwater head in a hydrological model using sequential data assimilation, this observation bias must be taken into account.

While the EnKF, and any derivation thereof, implicitly accounts for both model- and observation uncertainty in the form of zero-mean white noise, model and observation biases remains an issue that requires modifications to the filter. A few methods have been developed that attempt to estimate biases online, and they have been applied successfully in many settings. With few exceptions, the bias aware filters can be grouped in two: Separate filter methods and augmented state methods. The Separate bias Kalman Filter (e.g., Dee and da Silva, 1998; Pauwels et al., 2013; Drecourt et al., 2006) uses a second Kalman filter for updating the biases. This second filter is independent from the filter that updates the states, and the method can therefore not account for correlation between states and biases. Alternatively, augmenting the state space with bias estimates (e.g., Derber and Wu, 1998; Dee, 2005; Drecourt et al., 2006; Fertig et al., 2009) allows the filter to account for the correlation between states and biases, and is therefore useful when the bias is dependent on the observed values. While most implementations of bias estimation assumes that the model is unbiased and that the observations are biased, or vice versa, Pauwels et al. (2013) presented a method for estimating both model bias and observation bias simultaneously using a double Separate bias Kalman Filter.

This study uses both a synthetic test setup and real observations to test the application of bias correction to a data assimilation framework that assimilates groundwater head and stream discharge observations in an integrated hydrological model for joint state updating and parameter estimation. We discuss the challenges associated with ~~coupling two processes (groundwater and stream flow) in a single filter and the estimationof parameters in the presence of significant observation bias~~observational bias in hydrological data assimilation for both state updating and parameter estimation. Two existing methods of estimating observation bias, the Separate bias Kalman Filter and the augmented state vector approach, are tested and the results compared. The novelty of the study lies in the ~~holistic approach to sequential hydrological data assimilation , which accounts for parameter errors and uncertainty, different process time scales in both model states and observations, and the presence of observation bias~~focus on data assimilation bias estimation in a complex, integrated hydrological model as well as the impact of bias on parameter estimation in both synthetic test and using real world observations. While each of these aspects have previously been studied individually ~~,~~ the combination of the aspects creates new challenges, which require particular attention. This paper shares several similarities with the preceeding Rasmussen et al. (2015), notably the model catchment and setup, but differs in the focus on bias and the application of data assimilation to real world observations. Rasmussen et al. (2015) presents a synthetic study of data assimilation in integrated hydrological modelling in which the filter performance as a function of ensemble size is investigated and the current paper expands on this and adds the complexity of bias estimation and real world data.

2 Methods

2.1 Model

This study uses a transient, spatially distributed hydrological model based on the MIKE SHE code (Graham and Butts, 2005). This code considers all major components of the land phase of the hydrological cycle and the code allows the hydrological components to be dynamically coupled, meaning that feedback (i.e. exchange of water) between the processes is possible at each time step. The feedback is of particular importance for the groundwater-stream interaction in areas where these processes are closely linked, and it makes the model code particularly suited for investigation of data assimilation in integrated hydrological modelling.

2.2 ~~Study area~~

2.1.1 ~~The Karup catchment~~

~~This study is based on the Karup catchment (Figure 1), which is located in the northern part of the Danish Jutland peninsula. The catchment has an area of 440 km^2 and agriculture is~~ The coupling between the unsaturated and groundwater zones in MIKE SHE is complicated, as the processes in the two zones are interdependent. When water is exchanged from the unsaturated zone to the groundwater, the ~~dominant land use, while the geology is dominated by highly permeable quaternary sand. It is a very flat catchment, with a gentle south-north slope ranging from 93 m.a.s.l. in the southern part to 22 m. a.s.l. in the northern part. The Karup river is the primary drainage feature and it springs in the southern part and exits in the northern edge of the catchment. Along its path, the Karup river is joined by seven smaller tributaries. The flat topography and sandy sediments implies that the Karup river is primarily groundwater fed, which emphasizes the importance of an integrated approach to the hydrological modelling of the catchment, as the exchange between the groundwater and the river is a predominant process of the hydrological response of the catchment~~groundwater table rises, thereby changing the flow of the unsaturated zone. This complex interdependence is in MIKE SHE simplified, as the two processes only exchange water at every time step of the groundwater model. As the time step of the groundwater model is often much longer than the time step of the unsaturated zone model, the groundwater table is kept constant during several unsaturated zone time steps. This may lead to water balance errors, and in an attempt to reduce these errors, MIKE SHE has a coupling control that adjusts the groundwater table and recalculates the unsaturated zone states if the water balance error is above a user-specified threshold.

~~The Karup catchment with locations of discharge and hydraulic head observations.~~

2.1.1 ~~Model setup~~

An integrated model, which includes groundwater flow, vadose zone flow, evapotranspiration, surface and streamflow is used in this study. Vertical groundwater flow components are neglected in the study and groundwater flow is simulated based on the 2D Boussinesq equation. Each numerical element of the groundwater flow model is coupled to a one dimensional model for vertical flow in the vadose zone. For numerical and computational convenience capillary forces are neglected and only gravity driven flow is considered, which is an option in the MIKE SHE code (Graham and Butts, 2005). Stream flow is simulated using the kinematic routing option. The stream network model is set up using an alternating calculation scheme in which discharge and water level is calculated respectively in alternating points, and is independent from the groundwater model in discretization, but exchanges of water between the two processes is made available in model grids of the two processes that physically overlap. The exchange takes place at every groundwater model time step. Evapotranspiration is modelled using the Kristensen and Jensen (1975) model. The groundwater model initial conditions is based on an extended warm-up of the model, win which a quasi-steady state develops, while the stream flow initial conditions was calculated automatically assuming a steady state condition.

A horizontal grid size of 1 km x 1 km is used, with a vertical discretization of the unsaturated zone gradually increasing from 0.05 m at the top to 1 m below a depth of 10 m. Further details of the MIKE SHE model application to the Karup catchment can be found in Rasmussen et al. (2015).

2.2 Study area

2.2.1 The Karup catchment

This study is based on the Karup catchment (Figure 1), which is located in the northern part of the Danish Jutland peninsula. The catchment has an area of 440 km^2 and agriculture is the dominant land use, while the geology is dominated by highly permeable quaternary sand. It is a very flat catchment, with a gentle south-north slope ranging from 93 m.a.s.l. in the southern part to 22 m.a.s.l. in the northern part. The Karup river is the primary drainage feature and it springs in the southern part and exits in the northern edge of the catchment. Along its path, the Karup river is joined by seven smaller tributaries. The flat topography and sandy sediments implies that the Karup river is primarily groundwater fed, which emphasizes the importance of an integrated approach to the hydrological modelling of the catchment, as the exchange between the groundwater and the river is a predominant process of the hydrological response of the catchment.

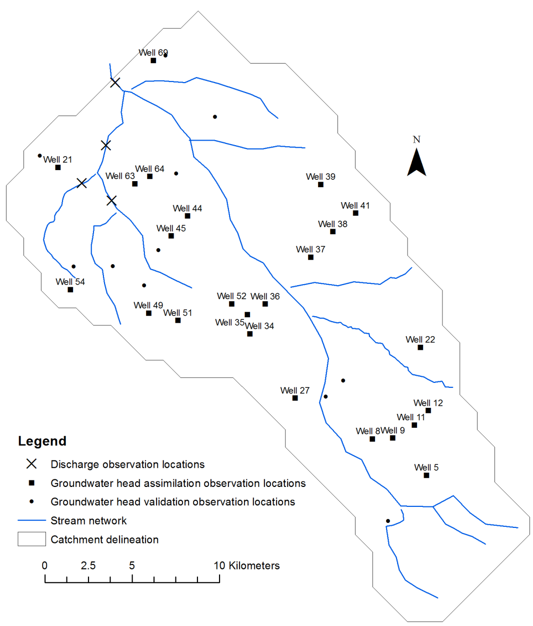

Figure 1. The Karup catchment with locations of discharge and hydraulic head observations.

2.2.2 Model parameterization

The geological model used in this study is a 3D model, which contains one dominant geological unit (meltwater sand) and five lenses (clay, quartz sand, mica clay/silt and limestone), each with assigned parameters of hydraulic conductivity, specific yield and specific storage. The geological model is in a preprocessing step converted into a 2D model by interpolating the parameter values and gridding them to the computational grid, resulting in a spatially variable field of hydraulic conductivity. The parameter values of the stream model are assumed uniform throughout the model ~~domanin~~domain. The drain level and ~~the~~ drain time constant ~~control the drainage flow to the river, while the leakage coefficient controls the river-groundwater interaction.~~ parameters control the amount of groundwater drained to the nearest stream once the groundwater table exceeds the drain level, and are as such linking the groundwater module and the stream flow module of the model. This models the artificial drain systems installed under most farmlands as well as the natural drainage processes that often occur in the topsoil, and the parameters are therefore particularly important for the drain flow of the river. The leakage coefficient is another coupling parameter, which represents the hydraulic properties of the thin layer of the sediments at the bottom of the stream. This parameter is of particular importance with regard to river base flow. For more details of the model parameterization, reference is made to Rasmussen et al. (2015).

2.3 Data Assimilation

2.3.1 Ensemble Transform Kalman Filter

The algorithm used for assimilating data in this study is the Ensemble Transform Kalman Filter (ETKF) (Bishop and Hodyss, 2009), which is a square root formulation of the EnKF. It is more computationally efficient than the EnKF ~~, as it does not require a full error covariance matrix to be determined unlike the EnKF, which also requires a computationally expensive inversion of the error covariance~~and is furthermore deterministic, meaning that the observations are in the filter not perturbed, which reduces the issues introduced by sampling. While the ETKF was first presented by Bishop and Hodyss (2009), the implementation used in this study is that of harlim and Hunt (2005). Vectors of the forecasted state variables of the ensemble members are structured in an $m\ x\ k$ matrix, X^f, where m is the number of states and k is the number of ensemble members:

$$X^f = [x_1^f, ..., x_k^f] \tag{1}$$

A $s\ x\ k$ matrix Y^f of model observations (s is the number of observations) is formed by applying a linear operator H that maps the state space into observation space to each column of X^f. This matrix is averaged over the columns to form a $s\ x\ 1$ vector of mean model observations, $\bar{y}^f$, which is then columnwise subtracted from Y^f to form the $s\ x\ k$ matrix of model observation anomalies, Y^b.

Next, X^f is averaged columnwise to form the $m\ x\ 1$ vector of mean model states $\bar{x}^f$ and this vector is subtracted from each column of X^f to create an $m\ x\ k$ matrix of model state anomalies X^b.

An $k\ x\ s$ matrix, C, is computed as follows:

$$C = (Y^b) \cdot R^{-1} \cdot p_{obs} \tag{2}$$

where R is a $s\ x\ s$ matrix of observation error covariance, and P_{obs} is a $s\ x\ s$ diagonal matrix with the localization weights of each observation on the diagonal. The $k\ x\ k$ error covariance matrix is computed by

$$\tilde{P}^a = [(k-1) \cdot I + CY^b]^{-1} \tag{3}$$

where I is a $k\ x\ k$ identity matrix. The $k\ x\ k$ matrix of analysis error covariance is computed as

$$W^a = [(k-1)\tilde{P}^a]^{1/2} \tag{4}$$

The $k\ x\ 1$ vector w^a is calculated as

$$w^a = \tilde{P}^a C \cdot (y - \bar{y}^b) \tag{5}$$

where y is a $s\ x\ 1$ vector of observations, and $\bar{y}^b$ is a $s\ x\ 1$ vector of the mean model observations. w^a is then added each column of W^a, forming the $k\ x\ k$ matrix of updated error covariance W. The $m\ x\ k$ matrix is calculated:

$$X^c = X^b W \tag{6}$$

Finally, the updated model ensemble, X^u, is calculated by adding $\bar{x}^b$ to each column of X^c.

2.3.2 Localization

Rasmussen et al. (2015) showed that the common distance-based localization methods do not suffice for localization in integrated hydrological models; instead an adaptive localization method first developed by Miyoshi (2010) will be used. Rather than removing correlation based on the physical distance from an observation, this localization method is a combination of cross-validating the sample correlation (as estimated from the ensemble) and raising the correlation coefficient to a power in an attempt to distinguish true correlation and spurious correlation. As each state vector member is analyzed in the ETKF, the ensemble of model observations (i.e. the ensemble of model states in the observation locations) is generated, and the sample correlation coefficient between the each of the model observations and the state member is determined. The localization weights of the observations to the state member being analyzed are then calculated from the correlation coefficients as follows.

For each state variable, the ensemble is split into two sub-ensembles of equal size. The sample correlation between the state variable and each observation state variable is calculated for both sub-ensembles. These correlation coefficients are then combined using the following expression:

$$p_{obs,a} = \left(1 - \frac{|c_1 - c_2|}{2}\right)^a \tag{7}$$

where $p_{obs,a}$ is the localization weight, c_1 and c_2 are the correlation coefficients from the first and second sub-ensembles, and a is a constant used for tuning the localization.

Another localization weight, $p_{obs,b}$, is determined using the sample correlation coefficient for the entire ensemble, c, and another tuning constant, b, as follows:

$$p_{obs,b} = |c|^b \tag{8}$$

The final (applied) localization weight, p_{obs} (equation 2), is calculated as the product of $p_{obs,a}$ and $p_{obs,b}$. Rasmussen et al. (2015) found that parameter values of $a = 2$ and $b = 2$, produced the lowest root mean square error in the groundwater head in a similar model, and these parameter values will also be used in this study.

2.3.3 Parameter estimation with the ETKF

Parameters are in this study estimated sequentially using the augmented state vector approach (Liu and Gupta, 2007; Rasmussen et al., 2015). The state vectors (equation 1) are extended to also contain the parameters that are to be estimated:

$$X^f = \begin{bmatrix} x_1^f & & x_n^f \\ & \cdots & \\ \Theta_1^f & & \Theta_n^f \end{bmatrix} \tag{9}$$

where Θ_i^f is the set of parameters used to propagate the i'th ensemble member. The mapping matrix H is extended ~~accordingly~~according to equation 10.

$$x = H \cdot \begin{bmatrix} x_1^f \\ \Theta_n^f \end{bmatrix} \tag{10}$$

2.3.4 Inflation

In order to compensate for the systematic underestimation of error variance that is endemic to ensemble based Kalman filtering, covariance inflation ~~(Anderson and Anderson, 1999)~~ (Anderson and Anderson, 1999; Whitaker and Ham applied to both the groundwater head states and the stream discharge states. The inflation is applied by adding a percentage to the ensemble of forecast anomalies:

$$X^f = (1+\alpha)X^f \tag{11}$$

where α is the inflation factor. The inflation factor used in this study is 0.2, which is based on tests of different inflation factors, and has been shown to help maintain a good spread of the ensemble of states.

The ensemble of parameter values is also inflated using equation 11 but instead of using a constant inflation factor, the inflation factor for the ensemble of parameter values is calculated at each update and for each parameter to match a target spread (as described by the standard deviation):

$$\alpha = \frac{\sigma_{target}}{\sigma_{forecast}} \tag{12}$$

where σ is the standard deviation. σ_{target} denotes the desired spread of the ensemble of parameter values and $\sigma_{forecast}$ denotes the spread of the ensemble before updating. This method is only applied if the forecast standard deviation of the ensemble of parameters is smaller than the target standard deviation, which in this study is set to 10% of the initial standard deviation of the ensemble. This value has shown to produce the best results, by maintaining a sufficient spread that does not create instabilities in any of the ensemble members.

Using covariance inflation is, like using localization, inconsistent with the deriviation of the filter and only necessary due to inadequate or incorrect noise description and ensemble generation. However, due to the complex nature of the model, Generating an ensemble that perfectly represents the uncertainty of the model is difficult and particularly in the test using real data outside the scope of this paper.

2.3.5 Damping

A simple damping mechanism is implemented in the modeling framework to reduce the magnitude of the state- and parameter updates and ~~hereby~~ thereby reduce the shock introduced to the system in the form of instantaneous changes of model states and parameter values at the time of updating. Furthermore, damping has the same effect as inflation, as it helps maintain an ensemble spread and thus combats the tendency for the ensemble to collapse. Damping of parameter updates is common, and has been studied in Franssen and Kinzelbach (2008) and used in Rasmussen et al. (2015).

Damping is pragmatically applied post-updating as follows. For each ensemble member, the ~~magnitude of the update, U,~~ post-damping and final state vector is calculated as

$$\cancel{U} x_i^{u,D} = x_i^f + D \cdot (x_i^u - x_i^f) \tag{13}$$

~~The dampened update is subsequently calculated as~~

$$U_i^D = D \cdot U_i$$

where D is a user specified $m\ x\ 1$ vector of damping coefficients. Note that the values in D may vary depending on the variable type (i.e. hydraulic head, stream discharge or water level) or parameter type.

~~Finally, the post-damping and final state vector is calculated:~~

$$x_i^{u,D} = x_i^f + U_i^D$$

A damping coefficient of 0.1 was used for all parameters in all scenarios studied, while different damping coefficients for the states have been analysed in the tests described below.

2.3.6 Bias estimation

This study compares two different methods for estimating observation bias: the Colored noise Kalman filter (ColKF) and the Separate bias Kalman filter (SepKF).

The ColKF methodology for estimating bias follows that of Fertig et al. (2009), in which the biases are estimated online by augmenting the state vector, in a similar way as for estimating parameters. That is, the augmented state vector, which contains both states and parameter values is further augmented by an ensemble of observation biases as follows:

$$X^f = \begin{bmatrix} x_1^f & & x_n^f \\ \Theta_1^f & \cdots & \Theta_n^f \\ \beta_1^f & & \beta_n^f \end{bmatrix} \tag{14}$$

where β_i^f is the set of observation biases of the i'th ensemble member The linear operator H is modified such that when it is applied to the columns of X^f, the bias is added to the appropriate model observations:

$$\hat{x}_i^f = x_i^f + \beta_i^f \tag{15}$$

where $\hat{x}$ is the $i'th$ unbiased model observation. Note that a constant bias forecast model is used, meaning that

$$\beta_i^{f,t+1} = \beta_i^{u,t} \tag{16}$$

where the super script u indicates an updated value, and t refers to the time step.

This study assumes no bias in discharge observations, meaning that the only biased observations are the groundwater head observations. In real world observations, discharge observations would usually also be biased, but this bias is generally small compared to the random error of the observations and compared to biases in groundwater head observations.

The method requires an initial bias estimate based on á priori information. Furthermore, as with estimation of parameters, a spread in bias estimates needs to be generated. In this study, the initial estimate of bias in all observation points is ~~zero and the spread is~~ generated by sampling from a normal distribution with a standard deviation of 0.6 m and a mean of 0. The standard deviation vas chosen based on precursive testing in the synthetic test environment, that showed that this value generally led to the best estimates of bias.

The implementation of the SepKF in this study is similar to the one derived and presented in Drecourt et al. (2006) but modified to estimate observation bias rather than model bias and to be implemented for use in a square root formulation of the filter. The bias filter is a discrete filter that is coupled to the ensemble based filter used for updating the states and the parameters as follows. The forecasted model observation error covariance, P is estimated from the ensemble of anomalies:

$$P = \frac{1}{n-1} Y^b \cdot (Y^b)^T \tag{17}$$

The bias error covariance is estimated as being proportional to the ensemble model observation forecast error covariance, P, through a parameter γ ($0 \leq \gamma \leq 1$):

$$P_b = \gamma P \tag{18}$$

where γ is a tuning parameter that controls the fraction of information from the observations that is used to bias and states respectively. Tests using different values of γ revealed that this parameter had little impact on the final estimated bias, but a value of 0.1 was chosen for this study, as it performed slightly better than other values tested. The bias error covariance is furthermore conditioned to the assumption of no bias in discharge observations.

The Kalman gain for the bias filter is then calculated as:

$$K_b = P_b (H P_b H^T + H P H^T + R)^{-1} \tag{19}$$

The bias Kalman gain is localized as follows:

$$K_{b,local} = K_b (H \cdot L) \tag{20}$$

where L is a s x s matrix containing the localization weights for each state as determined by the adaptive localization algorithm. The updated biases are calculated as:

$$\beta^u = \beta^f \cancel{-} + K_b(y - H\bar{x}^b) \tag{21}$$

Finally, the updated states are calculated using the following modification of equation 5:

$$w^a = \tilde{P}^a C \cdot [(y - \beta^u) - \bar{y}^b] \tag{22}$$

The augmented state method has the advantage that it can take any interaction between the bias and the states into account, as the full forecast covariance matrix is used. On the other hand, the SepKF filter ignores any cross-correlation between bias and states.

While ignoring the correlation between state error and bias error may be problematic where such correlation exists, the price of using the augmented state method is the increase in the state-space that needs to be spanned by the ensemble. To describe the uncertainty of the augmented state, an $(m+p+s)$ x $(m+p+s)$ (states, parameters and observations) covariance matrix is necessary, while an $(m+p)$ x $(m+p)$ plus a $(s$ x $s)$ matrix is necessary for the SepKF. This is likely to increase the required ensemble size when using the augmented state method and thus increase computational demands.

2.3.7 Asynchronous assimilation

Due to the differences in frequency between the two observation types, this study uses asynchronous assimilation (Sakov et al., 2010). This way, the more frequent stream discharge observations can be assimilated along with the less frequent groundwater head observations, without the states having to be updated each time a discharge observation is available. The ~~method has been previously successfully used in Rasmussen et al. (2015). The observationsavailable between two updates, as well as their corresponding model observations, are collected and assimilated~~ state vector is extended with model results for asynchronous observation times and the observation vector is extended with the asynchronous observations. After that, the asynchronous observations and model results are simply treated as normal model states. The information contained in the extensions are then used to improve the update at the time of ~~the update~~updating. This is done by saving the individual observations and the model results for the time steps in which observations are available, and calculating the ensemble error as if the model results for the different time steps are model states. So, given a set of j observations at times $t_1, ..., t_j$ collected, the model observations is formulated as follows:

$$HX^f = [(Hx^f)_1^T, ..., (Hx^f)_j^T] \tag{23}$$

Similarly, the observation vector is extended to correspond to the ensemble observations. While the asynchronous observations and model observations are saved and used in the filter at the time of updating, they are afterwards discarded and no retrospective updating of states is performed.

2.4 Filter setup

2.4.1 State variables

In this study, ~~groundwater hydraulic~~ the state vector contains groundwater head, stream discharge ~~,~~ and stream water level, all of which are updated at each updating time step. The states are updated every 4 weeks, when groundwater head observations are available. The daily discharge observations available in between updates are included as asynchronous observations while the discharge observations available at the time of updating are assimilated normally.

2.4.2 Estimated parameters

The horizontal hydraulic conductivities of meltwater sand (HK_mws) and quaternary sand (HK_qs) are estimated, with the vertical conductivities tied to them at a ratio of 10:1. Note that the estimated hydraulic conductivities are those of the geological units, that are gridded to the computational grid before further propagation of the ensemble (see section 2.2.2). Furthermore, the two parameters that control drainage, the drain level and the drain time constant, are estimated, and so is the leakage coefficient, which controls the groundwater-stream flow interaction. These parameters were selected based on their scaled sensitivities as determined by using the AUTOCAL software (Madsen, 2003), with HK_mws being by far the most sensitive towards both stream flow and groundwater head. For a full list of sensitivity coefficients, the reader is referred to Rasmussen et al. (2015). HK_mws, HK_qs, the drain time constant and the leakage coefficient were transformed logarithmically, as their uncertainty is expected to span several decades.

3 Inverse modelling

In order to evaluate the performance of the data assimilation algorithm for parameter estimation using real observations, the model is also calibrated using AutoCal in order to be able to compare the parameter estimation through data assimilation with parameter estimation through more common method, such as inverse modelling. A multi-objective calibration approach is used, in which both groundwater head observations and stream discharge observations are aggregated and optimized. The setup of parameters is similar to the one used in the data assimilation approach (see section 2.4.2), with the same variable- and dependent parameters and initial values, in order to make the results of the inverse modelling and the data assimilation directly comparable.

Root mean square error is used as objective function of both groundwater head observations and stream discharge observations, and the two are aggregated using transformation to a common distance scale (Madsen, 2003). Both objective functions are weighted equally in the aggregation, to ensure an equal importance on optimizing both the stream flow and the groundwater head of the model.

3.1 Data availability

The Karup catchment was between 1970 and 1990 the subject of an extensive monitoring campaign in which stream discharge and groundwater head were rigorously measured. As a result, groundwater head observations are available in 35 locations (Figure 1) with a frequency of 14 days^{-1}, and daily stream discharge observations are available in four locations in the stream network.

3.1.1 Synthetic test observations

A twin test approach is used in the first part of this study, meaning that a "true" model is defined, and that the observations to be assimilated are generated from the results of this true model. The same model, but with perturbed parameter values, denoted the base model, forms the basis of the ensemble that is used for data assimilation. Note that both the true model and the base model are deterministic models, that is, single, propagated models without any noise added. The setup is identical to that of Rasmussen et al. (2015), and the reader is referred thereto for a detailed description and a list of parameter values. Groundwater observations are made available at 24 locations that form a subset of the 35 locations in which real observations are available (Figure 1). The reason for omitting some of the observation locations is that they are located too close to the stream network, and act as exchange between the groundwater model and the stream model. It was found that the groundwater head of these grid cells are very sensitive to the stream flow simulation, and small changes in the head lead to significant changes in the stream flow. As such, they are not suitable for assimilation and were used only as validation observations. Furthermore, one observation did not reflect the dynamics of the model due to its proximity to the model boundary and was therefore omitted. In the twin test experiment the groundwater observations are generated with a frequency of 28 days^{-1} and are added a time-varying, normally distributed white noise with a standard deviation of 0.05 m and each are added a randomly generated (normally distributed) constant bias with a standard deviation of 0.5 m.

Four stream discharge observations that coincide with the locations of real observations are included. The discharge observations are made available on a daily basis, and are added a normally distributed white noise that is proportional to the observed value using a standard deviation of 5% of the observed discharge, which is a common error observed in real world observations of discharge (Herschy, 1999).

The states and parameters are updated every time groundwater head observations are available, i.e. every 28 days, and the daily discharge observations available in between updates are assimilated asynchronously. Tests have shown that the length of the assimilation window is of little importance and therefore no other assimilation window was tested.

3.1.2 Real observations

Like in the synthetic test, the same 24 groundwater head observation locations are chosen for assimilation, while the remaining locations are used for validation. The real groundwater head observations are available with a frequency of 14 days^{-1}, but to avoid updating the states and parameters too often every other observation is assimilated asynchronously, allowing an assimilation window of 28 days like in the synthetic test. All four discharge observation locations are used for assimilation and are assimilated asynchronously.

3.2 Model noise

Model noise is added to the ensemble through the forcings, i.e. precipitation and reference evapotranspiration, and the parameters. Noise on forcings is added as a Gaussian noise with a standard deviation of 20% of the observed value, while no spatial correlation of the noise is considered.

Noise is added in the form of a Gaussian zero mean distribution to a large number of model parameters relating to all model processes and not just to the estimated parameters. In total noise is added to 66 parameters, only five of which are estimated. Adding noise to parameters that are not estimated helps maintain the spread of the ensemble even as the spread of the estimated parameters is reduced. Note that the zero mean of parameter noise means that if the filter successfully estimates all of the five included parameters, the ensemble of models is unbiased except for any bias there may have been introduced through the sampling of parameter- and forcing noise.

3.3 Test scenarios

For studying the performance of the data assimilation using synthetic observations, the study includes the ~~five~~ seven scenarios listed in Table 1. All scenarios include bias estimation, joint state updating and parameter estimation and simultaneous assimilation of groundwater head and stream discharge observations.

When assimilating real observations, three scenarios are studied: ColFil and SepFil and NoBiasEst (Table 2). The ColFil uses the ColKF, a damping factor of 0.1 and an ensemble size of 200, making it a combination of the Ens200 and Hdampen scenarios studied in the synthetic test. The SepFil uses the SepKF and an ensemble size of 100. The increase in ensemble size used when using real observations is due to the more complex nature of the model and observation error caused by differing dynamics of the observations and the model. For comparison, the NoBiasEst scenario uses no bias estimation.

3.4 Performance indicators

The model simulation period is from January 1st 1968 to December 31st 1973, and is divided into the following periods:

Table 1. Overview of setups studied in the synthetic tests.

Setup	ColFil Ens50	ColFil Ens100	ColFil Ens200	ColFil Hdamp	SepFil	SepFil NoQ	NoBias Est
Ensemble size	50	100	200	50	50	50	50
H damping factor	1	1	1	0.1	1	1	1
Q damping factor	1	1	1	1	1	1	1
Parameter damping factor	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Head observation stdv. (m)	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Discharge observation stdv. (-)	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Observation types assimilated *	Q, H	Q, H	Q, H	Q, H	Q,H	H	Q, H
States updated *	Q, h,H	Q, h,H	Q, h,H	Q, h,H	Q, h, H	Q,h,H	Q, h,H
Bias correction method	ColKF	ColKF	ColKF	ColKF	SepKF	SepKF	-

* Q: stream discharge, h: stream water level, H: groundwater head

Table 2. Scenarios studied in the real data tests.

Setup	ColFil	SepFil	NoBias Est
Ensemble size	200	100	100
H damping factor	0.1	1	1
Q damping factor	1	1	1
Parameter damping factor	0.1	0.1	0.1
Head observation stdv. (m)	0.05	0.05	0.05
Discharge observation stdv. (-)	0.05	0.05	0.05
Observation types assimilated	Q, H	Q, H	Q, H
States updated	Q, h,H	Q, h,H	Q, h,H
Bias correction method	ColKF	SepKF	-

- 1969: Warm-up, in which the ensemble is propagated without being updated in order to allow a spread in the ensemble of states to develop. At the end of the year 1969, the spread of the ensemble of groundwater head is between 2.1 m and 0.7 m (depending on the location in the catchment), which is considered sufficeint for assimilation to commence.

- 1970: Preliminary assimilation of observations, which allows the filter to constrain the states and parameters. The results of this period are not included in the performance evaluation.

- 1971-1972: Assimilation of observations for evaluation. The results of this period are included in the performance evaluation as an indicator for how well the filter performs. In the remainder of the report described as the "Assimilation period".
- 1973-1974: Validation period, in which the ensemble is propagated but not updated. It is used to assess the improvement in long term forecasting due to the filter update.

3.4.1 Synthetic test performance indicators

The performance of the filter when using synthetic observations is measured using three indicators:

- The mean estimated bias error ("Mean Bias Error"), calculated as the average difference (in all observation points) between the actual bias used to generate the biased observation and the mean of the ensemble of estimated biases at the end of the assimilation period.
- The average root mean square error of the groundwater head ("Head RMSE") in all calculation points of the groundwater model domain for the assimilation period.
- The Nash-Sutcliffe coefficient of the stream discharge at the outlet of the catchment ("NS") for the assimilation period.

3.4.2 Real data performance indicators

The performance of the filter when using real observations is measured using two indicators:

- The mean RMSE of all 35 groundwater head observation points for
    1. The assimilation period
    2. The validation period
- The Nash-Sutcliffe coefficient for stream discharge in the outlet of the catchment for
    1. The assimilation period
    2. The validation period

Furthermore, a deterministic model with the optimal parameter set (as determined by the data assimilation algorithm) is used to evaluate the estimated parameters. This model is designated "optimal model" and is evaluated using the above indicators. For comparison, the results of the optimized model using AUTOCAL is included (hereafter designated "AutoCal model").

4 Results and discussion

4.1 Synthetic tests

4.1.1 Bias correction using the Colored noise Filter

The filter setup that is considered the baseline setup is ColFilEns50 in which the ensemble size is 50 and the parameter updates are dampened by a factor of 0.1, while no damping of the state updating is performed. The baseline setup is adopted from Rasmussen et al. (2015) as this setup performed satisfactorily for the same catchment and similar number of observations. However, Rasmussen et al. (2015) did not consider bias correction.

The ColFilEns50 performed poorly in all three performance indicators as seen in Figure 2. The average error in estimated bias is 0.47 m; worse than the average absolute bias of the observations (0.38 m), and the filter often estimates a bias that is in the wrong direction. This suggests that better, or at least similar poor results, could be obtained by not correcting the bias. Furthermore, the updating of groundwater head is often erroneous, as evident from the spikes in groundwater head RMSE (Figure 3) that occur at the time of updating. This wrong updating may be explained by two issues: The wrongly estimated bias, which compels the filter to update the states wrongly as it does not know the unbiased observations, or the appearance of spurious correlation. Rasmussen et al. (2015) observed the same spikes in head RMSE when using unbiased observations and concluded that they are caused by spurious correlation.

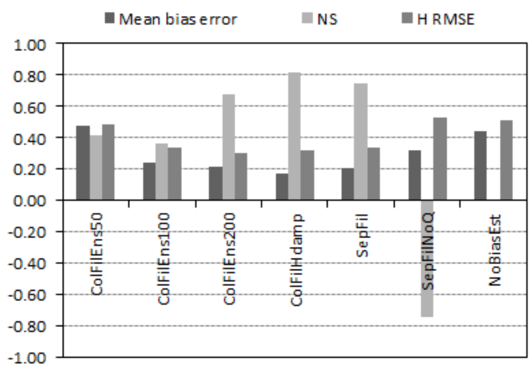

Figure 2. Mean bias error, NS and H RMSE for the years 1971-1972 in the synthetic test.

The poor performance of the ColFilEns50 is unexpected, as an almost identical setup was successfully used in Rasmussen et al. (2015), albeit using unbiased observations. However, adding bias correction to the filter increases the state space that must be spanned by the ensemble, thus potentially requiring a larger ensemble size.

Doubling or quadrupling the ensemble size to 100 and 200 respectively (ColFilEns100 and ColFilEns200 scenarios) resulted in major improvements in almost all indicators (Figure 2). In terms of estimating bias, the error is reduced by approximately 50% to 0.24 m and 0.22 m respectively, and the head RMSE is reduced by 26% and 31%. However, as visible in Figure 3, incorrect state updates still occur even with an ensemble size of ~~200, and these result in the same peaks in stream discharge as observed when using the baseline setup~~200. As shown by Rasmussen et al. (2015), these spurious correlations are likely to result in increased drainage to the stream model, resulting large errors in stream flow. The errors from spurious correlations in the stream flow model dominate the performance indicator and are, due to the nature of spurious correlation, random. As a result, the Nash-Sutcliffe coefficient is reduced when using an ensemble size of 100, but increased with an ensemble size of 200.

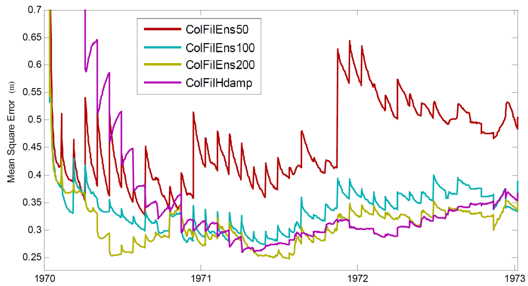

Figure 3. The temporal variation of Head RMSE in the synthetic test.

The increased performance, and the reduction in the spikes in head RMSE, supports the hypothesis that the poor performance of the ColFilEns50 setup is caused primarily by spurious correlation.

Dampening the update of groundwater head (ColFilHdamp scenario) had a profound effect on all the performance indicators (Figure 2). The mean bias error is reduced by 63% compared to the baseline setup, and the NS is nearly doubled. Finally, the head RMSE is reduced by 28% to 0.32 m, which is higher than what is obtained by increasing the ensemble size or retuning the localization algorithm, but still a significant improvement.

Dampening reduces the instant change in groundwater head, and as such reduces the problems that arise due to the non-linear relationship between states as well as reducing spurious correlation. Furthermore, it reduces the numerical effects that come from changing model states and parameters, in which the model attempts to regain equilibrium. However, dampening the state updates causes a slower reduction in head RMSE (Figure 3), the value approximately converges to the RMSE of Ens100 and Ens200 within one year of assimilation.

4.1.2 Bias correction using the SepKF

Using the SepKF (scenario SepFil) resulted in significant improvements over the ColFilEns50 setup in all performance indicators compared to the ColKF setup with the same number of ensemble members (ColFilEns50) (Figure 2). The mean bias error is reduced to 0.20 m, which is comparable to

ColFilEns200 and ColFilHdamp setups and little drifting behavior is observed in the model (Figure 4). NS is increased to 0.75, and head RMSE is reduced to 0.34.

~~Spread of estimated parameters at the final update (Synthetic test). Thin blue lines show the total spread of the ensemble and thick blue lines show the 25th and 75th percentile. Dots show the mean of the ensemble. The horizontal lines show the true parameter value (black line) and the base parameter value (magenta line).~~

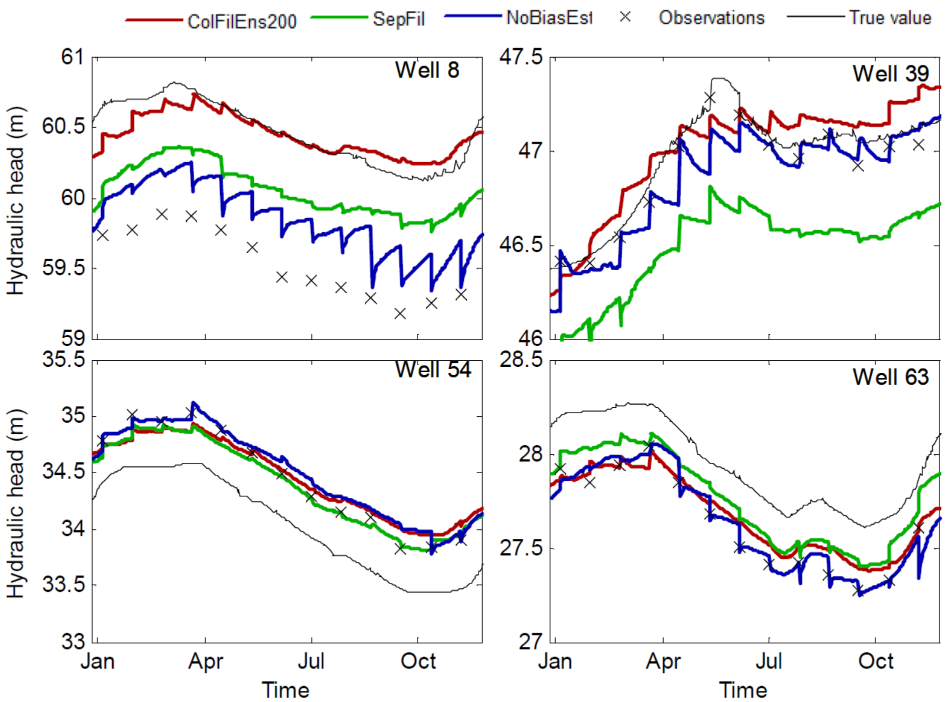

Figure 4. Groundwater head as a function of time in four selected observation locations for the year 1972 (Synthetic test).

4.1.3 Excluding discharge observations

When excluding the discharge observations (scenario SepFilNoQ), the filter performs worse in all three indicators. Compared to the SepFil scenario, both the mean bias error and the head RMSE is increased by 58%, and the NS is reduced to -0.74. The reduction in NS is explained by a bias in the estimated drain constant ~~and drain level~~ (Figure 5) and by a poorer description of the groundwater head level as indicated by the head RMSE.

Rasmussen et al. (2015) showed that discharge observations are particularly valuable for estimating parameters and updating stream discharge, but less valuable for groundwater head updating. They found that excluding discharge observation resulted in an improvement in groundwater head description when the spatial coverage of groundwater head observations is good, as there is a trade-

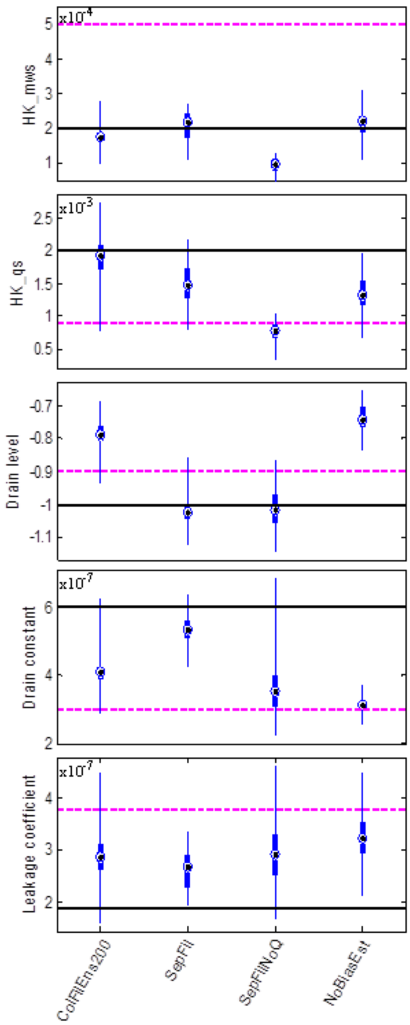

Figure 5. Spread of estimated parameters at the final update (Synthetic test). Thin blue lines show the total spread of the ensemble and thick blue lines show the 25th and 75th percentile. Dots show the mean of the ensemble. The horizontal lines show the true parameter value (black line) and the base parameter value (magenta line).

off between optimizing the stream flow and the groundwater head. However, the current results suggest that discharge observations also helps improve the estimation of groundwater head observation bias and consequently of the groundwater heads.

4.1.4 Bias-unaware filter

Excluding bias estimation from the filter (NoBiasEst scenario) results, as expected, in significant reductions in filter performance (Figure 2). This scenario may be considered as having an estimated bias of zero, and as such has a mean bias error of 0.44 m (i.e. the average absolute bias used to generate the observations), which lead to an increase in head RMSE of 6% over the already poorly performing ColFilEns50 scenario and 50% over the SepFil scenario. Furthermore, the NS was reduced to zero due to erroneous updates of the groundwater head and poorly estimated parameters; in particular the drain level and the drain constant (see Figure 5). The omission of bias estimation also resulted in significant (and expected) gradual deviation from the updated level (i.e. drifting) as seen in Figure 4. When considering the predictive power of the model, the bias-unaware filter is also more likely to esimate biased groundwater heads (Figure 6). This figure indicates that particularly in the observation point "Well 8", the bias-unaware will consistently forecast a too low groundwater head, and, as seen in Figure 4, this is to a great extent casued by the biased filter updates.

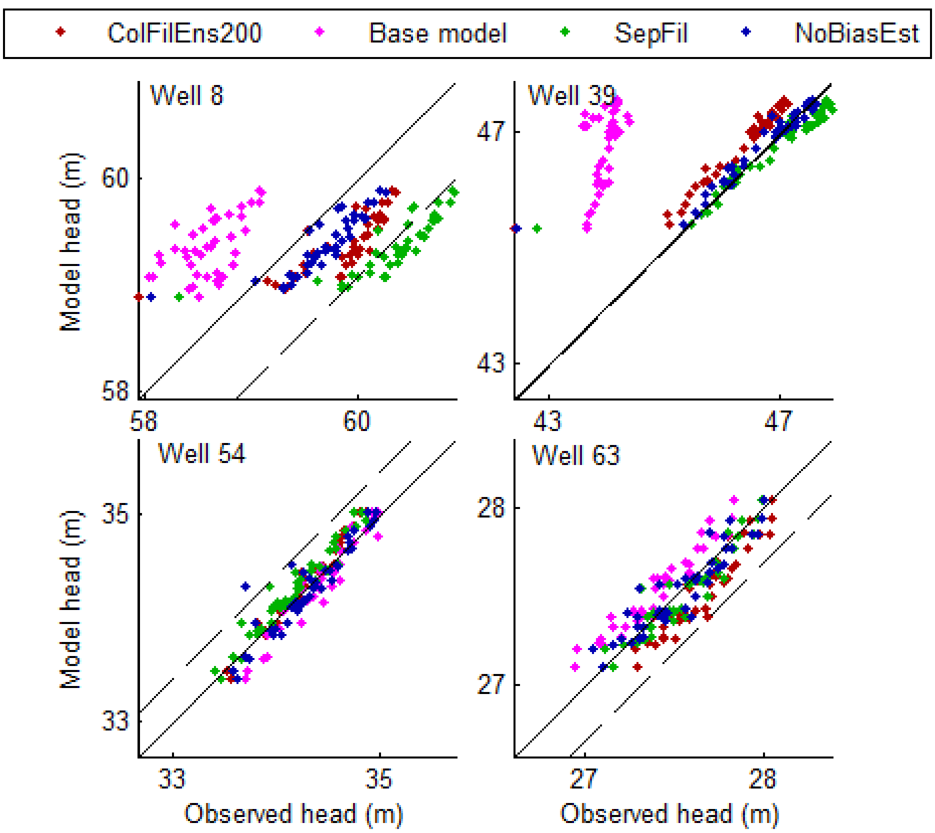

Figure 6. Model observations versus synthetic observations in selected observation locations. The dashed line indicates the 1:1 line when corrected for the applied bias. Note that the plotted model observations are the forecasted model observations, i.e. before the states are updated in the filter.

It is clear that omitting bias estimation when biases are present has a negative impact on both state updating and parameter estimation. ~~Updating~~ It is observed that updating the groundwater head to a biased observation level causes the head to return to an unbiased level when model propagation is resumed ~~and the~~ (i.e. it is drifting as seen in Figure 4). The model behavior becomes unnatural in the sense that it is not controlled primarily by the input forcings, but rather by the model trying to retain equilibrium. This ~~results~~ can result in deteriorated estimation of parameters and updates of model states not only in the observation points but in the entire model domain.

4.1.5 Comparison of the ColKF, the SepKF and the bias-unaware filter

The time varying estimated biases using the ColKF and the SepKF for each observation location are shown in Figure 7. The figure compares the ColFilEns200 and the SepFil scenarios, as they are the most easily comparable in terms of setup and performance. Both scenarios have comparable mean bias error (0.22 m and 0.20 m for ColFilEns200 and SepFil respectively), but as Figure 7 shows, there are significant differences in the estimation of bias in most observation locations. The ColKF converges significantly faster than the SepKF to the true value in most locations where the bias estimation is successful, due to the inclusion of bias-state correlation in the ColKF. The SepKF also underestimates the bias in some locations, most likely due to the simplifications and assumptions, notably the assumption that the bias error covariance is proportional to the state error covariance.

~~Both methods~~ Figure 4 shows that the drifting behavior i generally most pronounced in the NoBiasEst scenario and least pronounced in the ColFIlEns200 scenario, with the drifting behavior of the SepFil scenario falling in between the two former scenarios. Both the ColKF and the SepKF reduce the bias error in most locations except in wells 39, 54 and 63. The erroneous bias estimation may be because of the estimated parameter values. Visual inspections of the groundwater head as a function of time (Figure 4) reveals that there is no significant systematic deviation from the updated level (i.e. drifting) in the ColFilEns200 and SepFil and therefore no update of the observation bias in the filter. The lack of model drifting despite erroneous bias estimation is caused by the wrongly estimated parameters, and as such this is an equifinality issue: The filter has been able to produce non-drifting behavior of the model despite biased states, by using a biased parameter set. On the other hand, the NoBiasEst displays significant drifting in wells 8, 39 and 63, even when the updated states are unbiased (well 39) but as the filter is unaware of bias, this is not corrected.

The improvements gained from using the SepKF filter rather than the ColKF stem from the reduction in uncertainty needed to be described by the ensemble, and thus a smaller ensemble size is required. Ignoring the correlation between the bias and the state reduces the complexity of the system, and if that correlation is negligible, as in this case, there is little advantage in using the ColKF over the SepKF.

The two bias correction methods were also compared in Drecourt et al. (2006) using a simple one-dimensional groundwater model. While they did not consider the issue of ensemble size, they too found that both the ColKF and the SepKF can successfully estimate biases and improve model forecasting abilities. They also noted that the convergence of the SepKF is slower than the convergence of the ColKF, but the performances of the two methods were otherwise comparable.

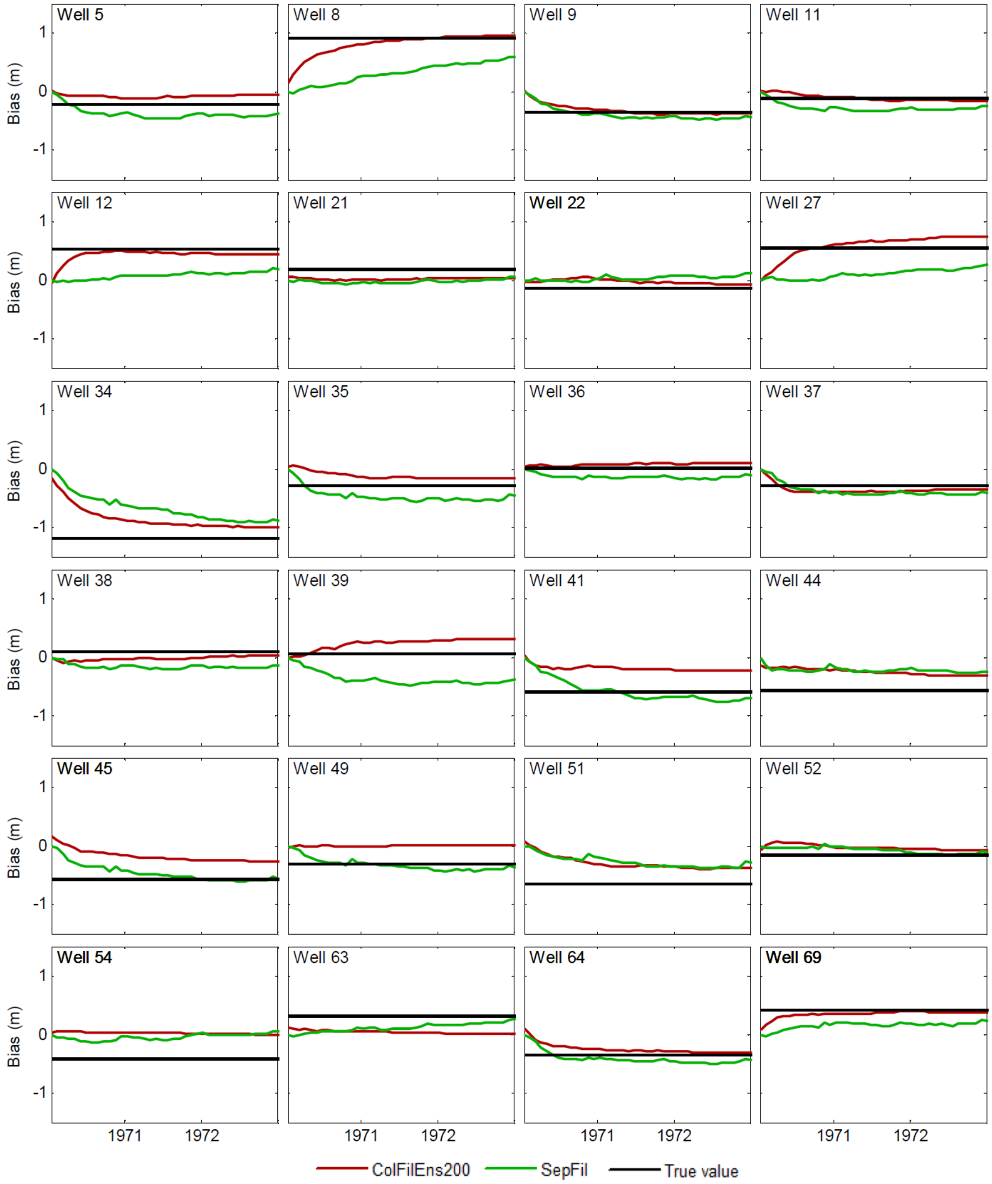

Figure 7. Estimated bias in the ColFilEns200 and SepFil scenarios as a function of time in the synthetic tests, compared to the true bias value used to generate the biased observations.

4.2 Real data tests

The Nash-Sutcliffe coefficient for stream discharge and the mean RMSE of groundwater head can be seen in Figure 8. When comparing to the base values data assimilation with the separate bias filter (scenario SepFil), the colored noise filter (scenario ColFil) and the bias-unaware filter (scenario NoBiasEst) all result in increased Nash-Sutcliffe coefficients and reduced mean head RMSE in the assimilation period.

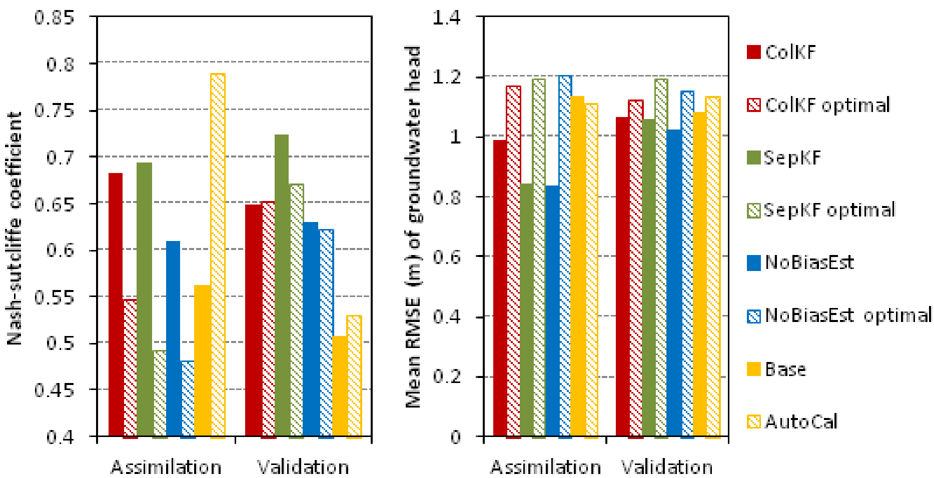

Figure 8. Nash-Sutcliffe coefficient for stream discharge (left) and mean RMSE of groundwater head observations (right) in the assimilation and validation periods respectively (real data).

In the NoBiasEst scenario, the model states are forced to match the observations as any bias is ignored, which results in a lower mean head RMSE in both the assimilation and the validation period (Figure 8). However, the assumption of unbiased head observations results in the NoBiasEst scenario having the lowest Nash-Sutcliffe coefficients of the three scenarios due to a trade-off between stream discharge observations and groundwater head observations, and it results in the drifting model behavior apparent in Figure 9, in which the model deviates strongly from the observed level in between updates. In SepFil, bias estimation is included using the SepKF, which results in a higher Nash-Sutcliffe coefficient and a comparable head RMSE to that of the NoBiasEst scenario. The effect of the bias estimation can be seen in Figure 9 as the filter does not update the groundwater head to the level of the observation but acknowledges a bias, which results in less drifting between updates compared to the NoBiasEst scenario. However, the deviation is still significant, which indicates that the bias is underestimated for this observation point. This is in line with the synthetic tests, where it was observed that the SepFil tends to underestimate large biases.

~~In SepFil, bias estimation is included using the SepKF, which results in a higher Nash-Sutcliffe coefficient and a comparable head RMSE to that of the NoBiasEst scenario. The effect of the bias~~

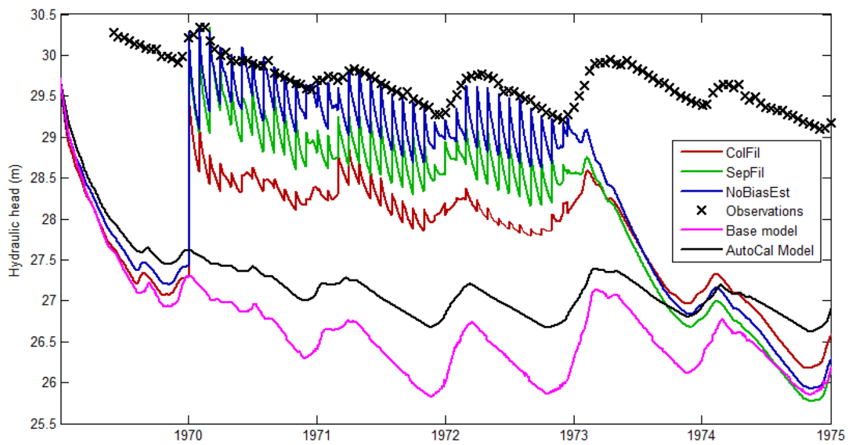

Figure 9. Groundwater head as a function of time in head observation location well 64 (real data).

~~estimation can be seen in Figure 9 as the filter does not update the groundwater head to the level of the observation but acknowledges a bias, which results in less drifting between updates compared to the NoBiasEst scenario. However, the deviation is still significant, which indicates that the bias is underestimated for this observation point. This is in line with the synthetic tests, where it was observed that the SepFil tends to underestimate large biases.~~

The ColFil scenario results in higher mean head RMSE and slightly lower Nash-Sutcliffe coefficient than the SepFil, but the ColFil optimal model (i.e. the deterministic model using the parameter set estimated by the filter) performs better than the SepFil optimal model with respect to most indicators.

The ColFil scenario estimates significantly larger biases in most observation points (Figure 10), with an average absolute estimated bias of 0.63 m, compared to 0.19 m in the SepFil scenario. With few exceptions, SepFil estimates a smaller bias than the ColFil, though in most cases in the same direction. Different bias directions are estimated by the two filters in two of the 24 observation locations, as illustrated in Figure 9, which may be caused by significant differences in the estimated parameter values (see Figure 11), as is also seen in the synthetic test.

A bias of approximately zero is estimated in seven observation locations, while biases of up to 1.8 m are estimated in others. In most locations however, the bias appears underestimated, as exemplified by Figure 9. This underestimation is observed as drifting and is likely caused by two factors. For the SepKF, the update of bias is constrained by the γ parameter, meaning that a too low value of γ may limit the update too much and thereby make the filter unable to estimate the correct bias, while a too high γ value is likely to yield unstable bias estimates. A test was made using a γ value of 0.3 (the value used in SepFil is 0.1), which resulted in increases in the estimated biases, but also resulted in unstable bias estimates that changed significantly with each time step as the filter did not properly distinguish biases, random error and model dynamics. Furthermore, as more and more

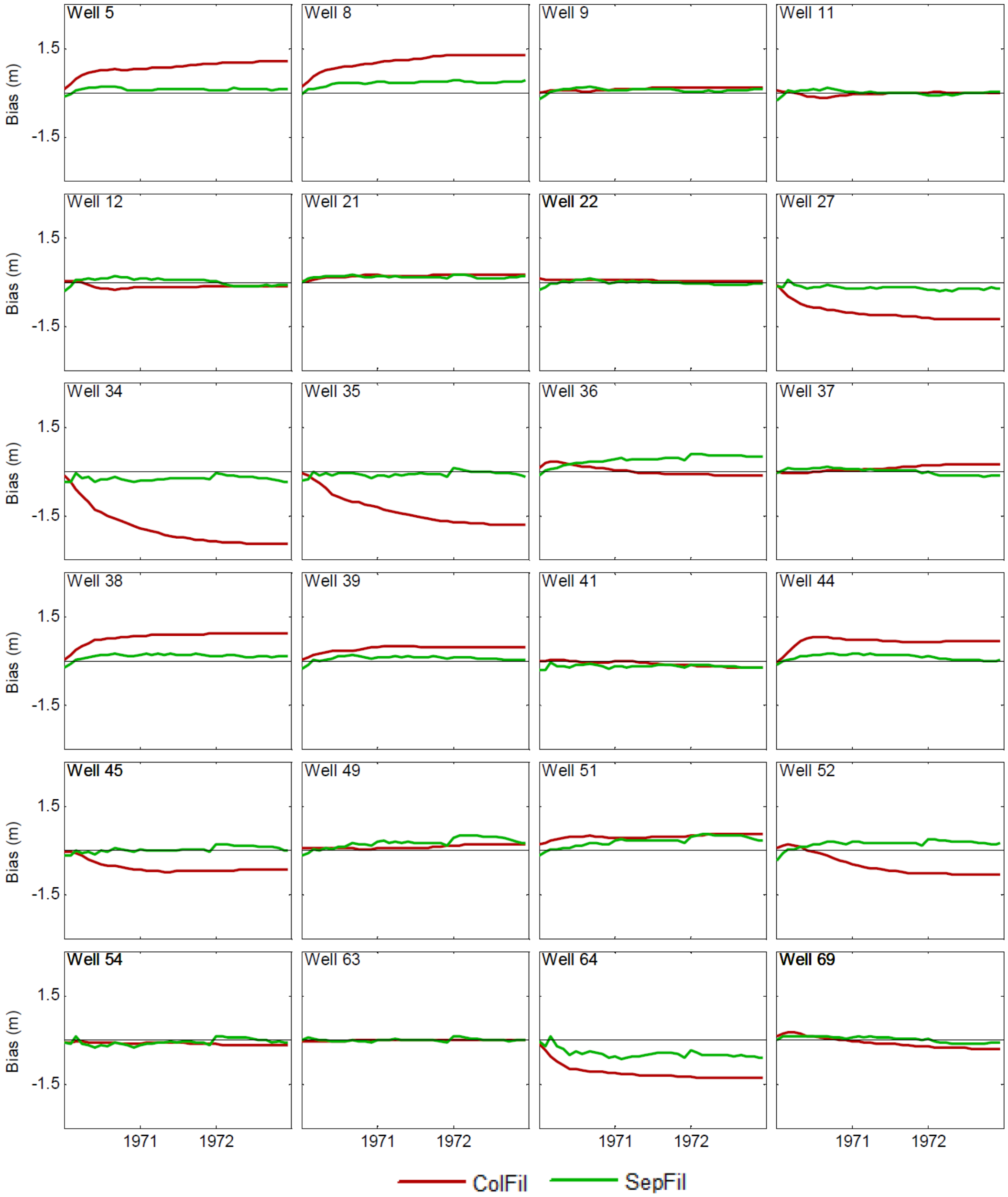

Figure 10. Estimated bias in the ColFilEns200 and SepFil scenarios as a function of time ~~in the synthetic tests, compared to the true~~ (real data). The black line indicates zero bias~~value used to generate the biased observations~~.

observations are assimilated and the spread of the ensemble of states is reduced, the update of the biases is smaller as the bias error covariance is assumed proportional to the state error covariance. If the ensemble spread of states is reduced too much, or even collapses, before correct biases are estimated, the bias estimation effectively stops. A similar consideration is applicable to the ColKF, as the ColKF operates with an ensemble of biases, and the spread of the ensemble of biases (and thereby the bias error covariance) is independent of the ensemble of states. If the spread of the ensemble of biases is too small, bias estimation effectively stops.

Comparing the optimal models of the ColFil, the SepFil and the NoBiasEst with the base model and the AutoCal model reveals a clear difference between the assimilation period and the validation period. While the optimal models produce lower NS for the assimilation time than both the base model and the AutoCal model, there is a clear improvement in the NS in the validation period over both the AutoCalModel and the base model. This suggests that AutoCal has produced a biased parameter set, which is not the case using any of the three Kalman Filters. However, the value of bias correction for parameter estimation is unclear, as there is no significant difference in the validation NS of the bias-aware Kalman filters and the bias-unaware Kalman Filter.

This tendency is not present in head RMSE, where the optimal models perform more poorly in terms of head RMSE than the base model and the AutoCal model. While it is to be expected that the AutoCal model would produce lower head RMSE than both the ColKF and the SepKF since the AutoCal model has been optimized specifically based on the head RMSE, it was expected that the optimal models of the ColKF and SepKF would produce improvements over the base model. However, it should be noted that the evaluation of model performance is based on the possibly biased observed values, and that the estimated biases have not been taken into account in the head RMSE calculations. The lack of clear improvement in the optimal models may be explained by the fact that there is little room for improvement with the current model structure as underlined by the relatively small improvements between the AutoCal model and the base model. It may also in part be explained by the underestimation of the biases in both the ColFil and SepFil scenarios. Improving the model structure and the filter setups may improve the potential of estimating parameters, but with the current results the value of data assimilation for parameter estimation is not clear.

5 Conclusions

Observation bias is a notable challenge in integrated hydrological modelling and needs to be addressed when applying data assimilation to the models. Updating the states of a model to match strongly biased observations will decrease filter performance and may even cause numerical instability. The two methods for correcting observation bias presented in this study can help reduce the bias issue in data assimilation and improve filter performance. Both methods improved the groundwater head and stream discharge of the model, and with varying degrees of success estimated the

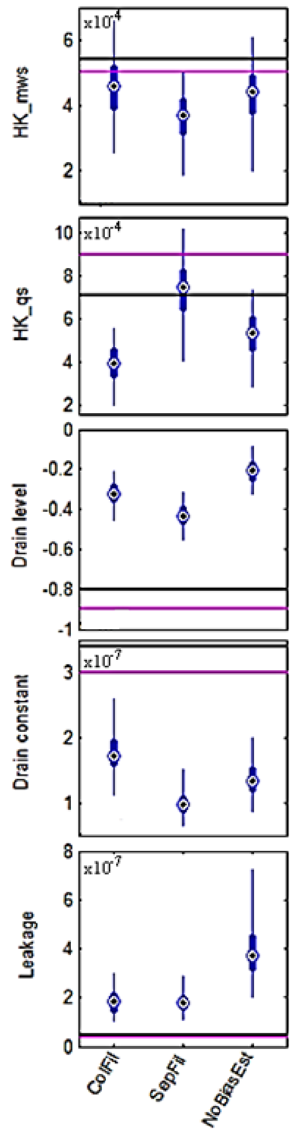

Figure 11. Spread of estimated parameters at the final update (real data). Thin blue lines show the total spread of the ensemble and thick blue lines show the 25th and 75th percentile. Dots show the mean of the ensemble. The horizontal lines show the AutoCal parameter value (black line) and the base parameter value (magenta line).

observation bias when using synthetic observations. When using real observations, both bias estimation methods resulted in improved stream flow modelling, but little improvement was seen in groundwater heads.

The main difference in the bias correction methods analysed is the interaction between the bias and the states. While the ColKF takes advantage of the full covariance matrix, the SepKF only takes into account the interaction that is present from the state to the bias and not the other way around. While this is a limitation of the SepKF, it results in a lower requirement for ensemble members, meaning that for smaller ensembles, the SepKF outperforms the ColKF. To obtain similar results to those of the SepKF when using the ColKF, the ensemble size needed to be doubled or even quadrupled, or the updates of the states needed to be dampened in an attempt to reduce the spurious correlations.

Most of the model parameters were successfully estimated in the synthetic tests, but biased observations introduces issues with equifinality. A biased parameter set may produce unbiased model behavior (i.e. without drifting) in one or more observations even if the estimated bias is incorrect. As a result, the filter does not update the bias of the observation, and the erroneous parameter set is not corrected. This resulted in significantly different parameter sets estimated by the different filters for both the synthetic tests and the tests using real data.

The study has shown that hydrological observational bias can be corrected in a data assimilation scheme and that it can improve state updating and parameter estimation. With both model- and observational bias being significant sources of error in hydrological modelling that may function as a road block for the application of data assimilation to hydrological models, these results may act as a stepping stone for the advancement of hydrological data assimilation in large scale, integrated hydrological models.

References

Albergel, C., Rudiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., , and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrol. Earth Syst. Sci., 12, 1323–1337, doi:10.5194/hess-12-1323-2008, 2008.

Anderson, J. and Anderson, S.: A Monte Carlo implementationof the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Wea. Rev., 127, 2741–2758, 1999.

Bailey, R. and Baù, D.: Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother, Hydrol. Earth Syst. Sci., 16, 287–304, 2012.

Bishop, C. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models, Tellus, 61A, 84–89, 2009.

Bosilovich, M. G., Radakovich, J. D., da Silva, A., Todling, R., and Verter, F.: Skin Temperature Analysis and Bias Correction in a Coupled Land-Atmosphere Data Assimilation System, Journal of the Meteorological Society of Japan, 85A, 205–228, doi:10.2151/jmsj.85A.205, 2007.

Camporese, M., Paniconi, C., Putti, M., and Salandin, P.: Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow, Water Resour. Res., 45, W10 421, doi:10.1029/2008WR007031, 2009.

Dee, D.: Bias and Data assimilation, Quart. J.R. Meteorol. Soc., 131, 3323–3343, 2005.

Dee, D. and da Silva, A.: Data assimilation in the presence of forecast bias, Quarterly Journal of the Royal Meteorological Society, 124, 269–295, 1998.

Derber, J. and Wu, W.: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system, Mon. Wea. Rev., 126, 2287–2299, 1998.

Drecourt, J., Madsen, H., and Rosbjerg, D.: Bias aware Kalman filters: Comparison and improvements, Adv. Water Resour., 29, 707–718, doi:10.1016/j.advwatres.2005.07.006, 2006.

Fertig, E., Baek, S., Hunt, B., Ott, E., Szunyogh, I., Aravequia, J., Kalnay, E., Li, H., and Liu, J.: Observation bias correction with an ensemble Kalman filter, Tellus, 61, 210–226, doi:10.1111/j.1600-0870.2008.00378.x, 2009.

Franssen, H. H. and Kinzelbach, W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resour. Res., 44, doi:10.1029/2007WR006505, 2008.

Franssen, H. H., Kaiser, H., Kuhlmann, U., Bauser, G., Stauffer, F., Muller, R., and Kinzelbach, W.: Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating, Water Resour. Res., 47, W02 532, doi:10.1029/2010WR009480, 2011.

Graham, D. and Butts, M.: Flexible, integrated watershed modelling with MIKE SHE, Eds. V.P. Singh and D.K. Frevert, pp. 245–272, 2005.

harlim, J. and Hunt, B.: Local Ensemble Transform Kalman Filter: An Efficient Scheme for Assimilating Atmospheric Data, Preprint, 2005.

Herschy, R.: Hydrometry - Principles and Practices, Wiley & Sons Ltd, 1999.

Kristensen, K. and Jensen, S.: A model of estimating actual evapotranspiration from potential evapotranspiration, Nordic Hydrol., 7, 1975.

Kurtz, W., Franssen, H. H., Brunner, P., and Vereecken, H.: Is High-Resolution Inverse Characterization of Heterogeneous River Bed Hydraulic Conduc-tivities Needed and Possible?, Hydrol. Earth Syst. Sci., 17, 3795–3813, doi:10.5194/hess-17-3795-2013, 2013.

Lannoy, G. J. M., Reichle, R. H., Houser, P., Pauwels, V. R. N., and Verhoest, N. E. C.: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, Water Resources Research, 43, n/a–n/a, doi:10.1029/2006WR005449, http://dx.doi.org/10.1029/2006WR005449, w09410, 2007.

Liu, Y. and Gupta, H.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43, doi:10.1029/2006WR005756, 2007.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26, doi:10.1016/S0309-1708(02)00092-1, 2003.

Miyoshi, T.: An adaptive covariance localization method with the LETKF (presentation), 14th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS) (recorded presentation), 2010.

Moradkhani, H. and Sorooshian, S.: General review of rainfallrunoff modeling: model calibration, data assimilation, and uncertainty analysis, in Hydrological Modeling and Water Cycle, Coupling of the Atmospheric and Hydrological Models, Water Science and Technology Library, 63, 1–24, doi:10.1007/978-3-540-77843-1, 2005.

Pauwels, V., Lannoy, G. D., Franssen, H. H., and Vereecken, H.: Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter, Hydrol. Earth Syst. Sci., 17, 3499–3521, doi:10.5194/hess-17-3499-2013, 2013.

Rasmussen, J., Madsen, H., Jensen, H. H., and Refsgaard, J.: Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance, Hydrol. Earth Syst. Sci. Discuss., 12, 2267–2304, doi:10.5194/hessd-12-2267-2015, 2015.

Reichle, R. H., Koster, R. D., Dong, J., and Berg, A. A.: Global Soil Moisture from Satellite Observations, Land Surface Models, and Ground Data: Implications for Data Assimilation, J. Hydrometeor, 5, 430–442, doi:http://dx.doi.org/10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2, 2004.

Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, Tellus, 62A, 24–29, doi:10.1111/j.1600-0870.2009.00417.x, 2010.

Shi, Y., Zhang, K., Duffy, C., and Yu, Z.: Parameter Estimation of a Physically-Based Land Surface Hydrologic Model Using the Ensemble Kalman Filter: A Synthetic Experiment, Water Resour. Res., doi:10.1002/2013WR014070, 2014.

Whitaker, J. and Hamill, T.: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation, Mon. Wea. Rev., doi:10.1175/MWR-D-11-00276.1, 2012.