

Answers to Damiano Pasetto

Summary: *The Authors satisfactory replied to most of my comments. I appreciate that the Authors included the analysis on the filter convergence with respect the seed and the ensemble size, together with the sensitivity to the inflating parameter. This analysis strongly improved the paper.*

Reply: We thank Damiano Pasetto for the second review and appreciate his additional comments and suggestions, which will help to improve our manuscript. In the following we provide the answers to the comments.

General comment

Comment: *I am not sure if showing the specific convergence results on the saturated conductivity of the second layer (figure 10) is the best way to demonstrate the convergence. As first, because there is a small bias on this parameter (as discussed in P16, l 25). Thus it would be important to present the convergence results also on the other parameters, however this would require too many figures. As second, because the PF provides an approximation of the posterior distribution of the parameters, thus the analysis should take into account not only the mean of the estimated parameters, but also the ensemble spread, i.e. the covariance of the ensemble for that parameter. A possible alternative would be to present in the main manuscript the convergence of the error, for example the average RMSE on the water content during the free run and its ensemble spread. I suggest including the figures on the particular convergence on each parameter (such as figure 10) in the supplementary information.*

Changes: Thank you for your suggestion. We moved the section to the appendix (Line 1, Page 24 until Line 15, Page 25) and added the figures for the convergence of all parameters as supplementary information.

Line 5, Page 16 until Line 19, Page 18: We show now the RMSE of the water content and its variance at the last assimilation time.

Detailed comments

Comment: P3, line 4: “Suppose a set?” the verb is missing.

Changes: Thank you for pointing this out.

Line 4-10, Page 3: We changed the sentence to: “For a set of observations $\mathbf{d}^{1:k} = (\mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^{k-1}, \mathbf{d}^k)$, where the superscript denotes a discrete time index, the observations are assimilated sequentially using the recursive filter equation

$$P(\mathbf{u}^{0:k} | \mathbf{d}^{1:k}) = \frac{P(\mathbf{d}^{1:k} | \mathbf{u}^k) P(\mathbf{u}^k | \mathbf{d}^{1:k-1})}{P(\mathbf{d}^k)},$$

which follows from Bayes’ theorem.”

Comment: P. 9, line 15: please delete one of the two “is” (repeated)

Changes: Line 1, Page 10: We deleted one “is”.

Comment: P10, Eq 21 : using a pedix to the number to indicate the dimension is not clear, and I think it is not necessary to repeat this pedix in all the occurrences of these values. Why not using variables γ_θ and γ_p (as in section 6.1), with the pedix just for this equation. Then, you can assign a numerical value directly to γ_θ and γ_p , without pedix. Please remove the pedix to the numerical values of γ_θ and γ_p .

Changes: Thank you for your suggestion.

Line 5-7, Page 11 & throughout the manuscript: We changed the subscript accordingly.

Comment: P. 16, line 1: in the section/subsections titles, sometimes the Authors use the upper case for all words, sometimes only for the first word. Please be consistent. Personally, I would use the upper case only for proper nouns.

Changes: Thank you very much for pointing this out. We changed the titles accordingly.

Comment: P. 16 Line 25: please replace “parameter” by “parameters”

Changes: Line 9, Page 25: We replaced “parameter” by “parameters”.

Comment: P. 17, line 3: please replace “there” by “their”. I suggest also to rephrase this sentence: model errors can have a structural or stochastic nature, different intensities, and they can manifest e.g. by bias in the results.

Changes: Thank you for your suggestion.

Line 21-23, Page 18: We follow it and rephrased the sentence accordingly: “They can have a structural or stochastic nature, different intensities, and they can manifest e.g. by biases in the results.”

Comment: P17, line 7: please add “the”: “In the course of this paper”

Changes: Line 27, Page 18: We added “the”.

Comment: P18, lines 15/20: it is not clear which is the ensemble size used. It is 600/1200 with gamma =1 or 300/600 with gamma as in Eq 21?

Reply: Thank you for pointing this out. We used 600/1200 with gamma as in Eq (21). For the 300/600 γ is $\gamma_p = 1.2$ and $\gamma_\theta = 1.1$. These results are not shown.

Changes: Line 33, Page 18 until Line 4, Page 19: We clarified which ensemble size and γ is used.

Comment: Figure 7: to better appreciate the impact of the filter, is it possible to show also the water content associated to the initial values of the parameters (as in fig. 6?).

Changes: We added the mean of an ensemble of free runs (Fig. 7, Page 14). The ensemble starts from the initial water content states (see Fig. 4), including their corresponding parameter sets and is propagated until the end of the data assimilation run. The ensemble members are not shown because it would strongly reduce the visibility of the actual results of the data assimilation (see Fig. AC1.1).

Comment: Figure 9: Please specify if the RMSE is computed for each sample of the posterior parameter distribution, or only for the simulation associated with the mean of the posterior parameter distribution.

Reply: The ensemble is propagated forward in time. The weights of the last assimilation time are used to calculate the mean of this forward run.

Changes: Line 8-9, Page 11 & Fig. 9, Page 16: We clarified this point.

Comment: Figure 10: The PF method provides a distribution of the saturated conductivity for each seed and for each ensemble size. For this analysis, are the Authors considering the mean or median of that distribution or the whole distribution? If only the mean or median are considered, it would be important to provide also the convergence of the ensemble spread.

Reply: We followed your suggestion in the general comment and show the ensemble spread and the RMSE of the water content (see also reply to the general comment). The parameter analysis is for the mean.

Changes: Fig. C1, Page 24 & Line 5, Page 24: We clarified this point and moved the section to Appendix C.

Comment: Algorithm 1: line 1: at time t_k , the ensemble of state is u_i^k . Point (d). The initialization $x=0$ is missing at the beginning. The application of the tuning parameter is missing.

Reply: The initialization is missing because the algorithm describes only

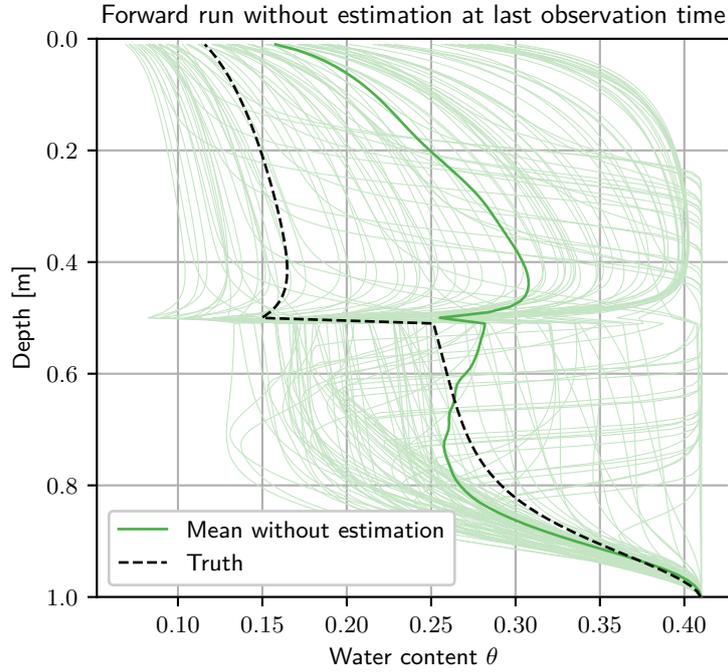


Figure AC1.1: Forward run with the initial conditions from the paper Sect. 4 without correction of the state and parameters. The ensemble (light green) is equally weighted and the mean is calculated from the ensemble at the last time step.

the resampling process. It requires the updated weights and the states and the time of the observation. The propagation of the ensemble and the actual weighting is not presented in this algorithm.

Changes: Appendix A: We corrected u_i to u_i^k , added the application of the tuning parameter and clarified that we only show one resampling step.

Comment: *References: please check the reference list: many references have the doi repeated twice*

Changes: Thank you for noticing this. We corrected the references.

Answers to Jasper Vrugt

Summary: *This is a re-review of the paper by Berg et al. on “Covariance resampling for particle filter – state and parameter estimation for soil hydrology”. I have carefully read the responses of the authors’ to my earlier comments and those of the other referee. The authors’ responses are professional and clearly indicate how and where changes were made and why. This certainly makes reviewing easier (as not all authors write a detailed response letter).*

Reply: We thank Jasper Vrugt for the second review and appreciate his additional comments and specification of his concerns, which will help to improve our manuscript. In the following we provide the answers to the comments.

General comments

*In my letter to the editor I wrote the following:
“...I can entertain/live with some of their answers to my comments, yet, I remain skeptical that the proposed method is going to work for more complex problems - and/or enlarged parameter space (current ranges are rather tight). The authors’ responses to my more theoretical questions I more or less expected. They do not deny/refute important weaknesses/difficulties of their proposed resampling method. In my view the paper focuses too much on practical application without going in depth about the underlying theory. Some changes have been made to the paper to mention problems with PDF approximation, yet, I feel that the method is not properly demonstrated with the Richards’ type flow model. Authors emphasize in their response letter their interest in the application - and that the PF is designed for soil hydrology (not mentioned this way) - and thus state/parameter dimensionality is not of their concern. In my view this really limits the potential impact of the paper. Right now, the paper is very much an Engineering solution - which may work well for simple problems with a model that tracks the data closely (transition density is more than adequate) and low state and parameter dimensionality. Beyond this simple case nothing else is demonstrated. What is more, the authors agree with my comment on the subjectivity of the inflation factor, γ (see equation 21) - but their response that this variable can be estimated on the fly during assimilation is problematic as several different values of γ will work. Some value of γ will provide a nicer*

convergence of the multivariate parameter and state distribution than others. But all those gamma values are equally likely; in other words, there is no metric that will help determine what value of gamma to use; so what value to use in practice? Personally, I find this troublesome - hence why I used the wording "Engineering Solution". I can live with this but this really diminishes the impact of the work. Of course, HESS is not a statistics journal, yet, research focused on methodological improvements should in my view be held to higher standards than more application oriented papers. Ideally, we would want people from other fields use our methodologies - that requires that improvements satisfy underlying statistical principles. This is not the case. I will not use this to say that the paper should not be published. I just want to express that I think that the paper could be much better if:

1) the authors were more concerned with "statistics", that is, satisfying/honoring important principles of convergence and approximation. The resampling method does not leave the target PDF invariant. This is something that I think we should care about as others may apply the method and present the results as if they are exactly right. This will not be the case. I see a parallel here with the poor numerics of many models used in surface hydrology. For years, this issue was denied/not investigated as people believed that the numerical error would be small compared to, say, measurement errors. Then, from 2004 papers by Kavetski, Clark, Schoups, etc. have demonstrated the effect of poor numerics on parameter and state estimates. Turned out to have a major effect; what is more, this work more or less refuted the need for global optimization methods as mass-conservative solvers turned out to produce much smoother response surfaces with a much better defined global optimum. Thus, a deeper focus on the methodology may prove very useful - and will certainly enhance tremendously the impact of the paper.

2) The application in the presented paper is rather weak - that is - the model (transition density) is (essentially) perfect and the parameter/state space is rather low. This is not realistic for real-world applications. This begs the question on how well the method is going to do in practice. In my view this is an important shortcoming.

For a hydrology paper, I can certainly accept weakness (1) above, yet then I would expect the authors to carefully examine their method; this is not the case - weakness (2) has not been addressed in the revision. I think it is important to do so - but will not object against publication if the authors do not address this. I do not want to hold up publication - as I have great respect for Prof. Roth and his co-authors. Hence my recommendation for a minor revision."

Reply: Thank you for your comments.

1) It is correct that the target PDF is not invariant under resampling using the proposed method. The PDF after resampling is a superposition of the estimated target PDF and a multivariate Gaussian of the new particles. If the model and the model error were perfectly known, we agree that it would be desirable to estimate/represent the target PDF as accurately as possible. However, if the model has an unrepresented model error (not part of the forward model) or the observation error is specified incorrectly (e.g. Gaussian distribution assumed, but lognormal distributed is correct), the target PDF from Bayes' theorem is not the true PDF since the transition/observation PDF is wrong. In those cases, which are typical for more complicated situations, considering an approximation for the target PDF can be beneficial to increase the stability of the filter and to explore the state space.

For example, the ensemble Kalman filter (EnKF) is one of the most used data assimilation methods in hydrology. The EnKF assumes for all occurring PDFs Gaussian distributions and is successfully applied in hydrology (e.g. Shi et al., 2015; Man et al., 2016; Botto et al., 2018). Zhang et al. (2017) compared the particle filter with the EnKF for the application on land surface models with the result of similar performance with slight advantages of the EnKF. Through the assumption of Gaussian distributions, the EnKF can actually correct ensemble members based on the measurement values, while the particle filter only assigns a likelihood and consequently typically requires more particles than the EnKF. This shows, in the case of the EnKF, how the advantages from assuming the Gaussian distribution in the EnKF compensate the disadvantages of the approximation of the distribution.

In the presented covariance resampling method the Gaussian approximation is only applied in the resampling to generate new particles. The approximation allows to use the covariance information in the ensemble, which facilitates the generation of meaningful new particles and improves the exploration of the state space. This reduces the required number of particles in difficult situations when many particles have to be resampled. It furthermore reduces the probability of filter degeneracy on the cost that these resampled particles alter the posterior. In less difficult situations, fewer particles are resampled and the estimated posterior comes closer to the ideal distribution.

We extended the explanation of this in the manuscript (Line 5-16, Page 6).

The factor γ is not mandatory and not required if one can afford a sufficient ensemble size. We show for our case that the covariance resampling works without applying a factor ($\gamma = 1$). However, a factor $\gamma > 1$ can strongly reduce the required number of particles. The factor helps to explore the state

and parameter space and prevents filter collapse for small effective sample sizes. We see this as an advantage of the proposed covariance resampling that it allows the easy application of the factor and offers the possibility to improve the performance. Note that the factor only applies to the resampled particles and that in less difficult situations, fewer particles are resampled and the posterior after resampling is closer to the estimated distribution. We clarify the explanation accordingly in the manuscript (see Section 6.1).

It is correct that the factor γ is heuristic. However, it cannot be chosen arbitrarily since too large values result in overinflation that cannot be compensated by the filter anymore. We do show the results for a few values of γ in the manuscript to give the reader an idea about the reasonable range for our example.

The factor has similarities to multiplicative inflation for the EnKF. It increases the uncertainty and prevents filter collapse for small effective sample sizes. In our example the initial state is the most challenging task for the filter. Perturbing the truth instead of the interpolation would result in an estimation without the factor and the same ensemble size. Therefore, it would be desirable to have an adaptive approach for γ such that it is large in the beginning and reduces to 1 once the filter has a good result for the state. In cases of a model error, γ has to increase again.

To visualize the behavior of the covariance resampling for non-Gaussian distributions, also in relation to the EnKF, we follow the comparison of deterministic and stochastic Kalman filters by Lawson and Hansen (2004).

The comparison is performed on a single analysis step. As in Lawson and Hansen (2004), the prior is a bimodal distribution. The bimodal distribution is constructed with two equally probable Gaussian distributions:

$$P(x) = \frac{1}{2\sqrt{2\pi}} \left(\exp \left[-\frac{1}{2}(x - \bar{x})^2 \right] + \exp \left[-\frac{1}{2}(x + \bar{x})^2 \right] \right), \quad (1)$$

where \bar{x} is the offset of each peak from zero. In the following example, we chose $\bar{x} = 4$. For the calculation of the Kalman gain, the bimodal prior with zero mean and a variance of $\sigma_{\text{Prior}}^2 = 17$, is equal to a Gaussian with these two statistical moments ($\mathcal{N}(0, \sigma_{\text{Prior}}^2)$).

Using Eq. (1) a rather large ensemble ($N = 5000$) is generated to represent the prior. This ensemble size is used to observe the behavior of the filters without large statistical noise. The prior distribution, including the generated ensemble, is shown in Fig. AC2.1.

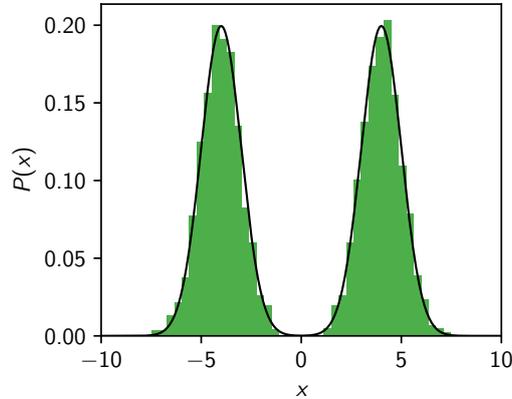


Figure AC2.1: Prior bimodal distribution $P(x)$ (black, Eq. (1)) and histogram of the 5000 ensemble members (green).

The analysis is calculated for an observation d at the position $d = 3.5$, with three different observation errors σ_{Obs}^2 . The observation errors are chosen as $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2/2$, $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2$ and $\sigma_{\text{Obs}}^2 = 2\sigma_{\text{Prior}}^2$. Figure AC2.2 shows the resulting analytical posterior distributions calculated using Bayes' theorem and the posterior ensemble of the EnKF, the particle filter with covariance resampling (PFCR) and the PFCR using $\gamma = 1.2$.

In the case of an accurate observation $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2/2$ (left), the observation determines the right mode of the bimodal prior as the posterior. The EnKF corrects the states towards the observation and the posterior becomes approximately Gaussian but with a mean shifted to lower values and a larger variance compared to the truth. The PFCR can describe the posterior in this particular case excellently. It does not need to shift the ensemble members of the left mode to the observation like the EnKF, instead almost 50% of the ensemble members are dropped and resampled. This resampling is effective because the covariance resampling samples from a Gaussian distribution and the posterior is approximately Gaussian.

In the case of a less accurate observation $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2$ (middle), the observation information is less dominant and the bimodal structure of the prior is partly visible. For the EnKF, the ensemble members of the left mode are shifted towards the observations such that the posterior becomes unimodal. The bimodal structure cannot be described by the EnKF properly. The PFCR can sample both posterior peaks with the retained ensemble. The new particles, however, are generated with a mean that lies in between both peaks, such that some of the new particles are located between the modes.

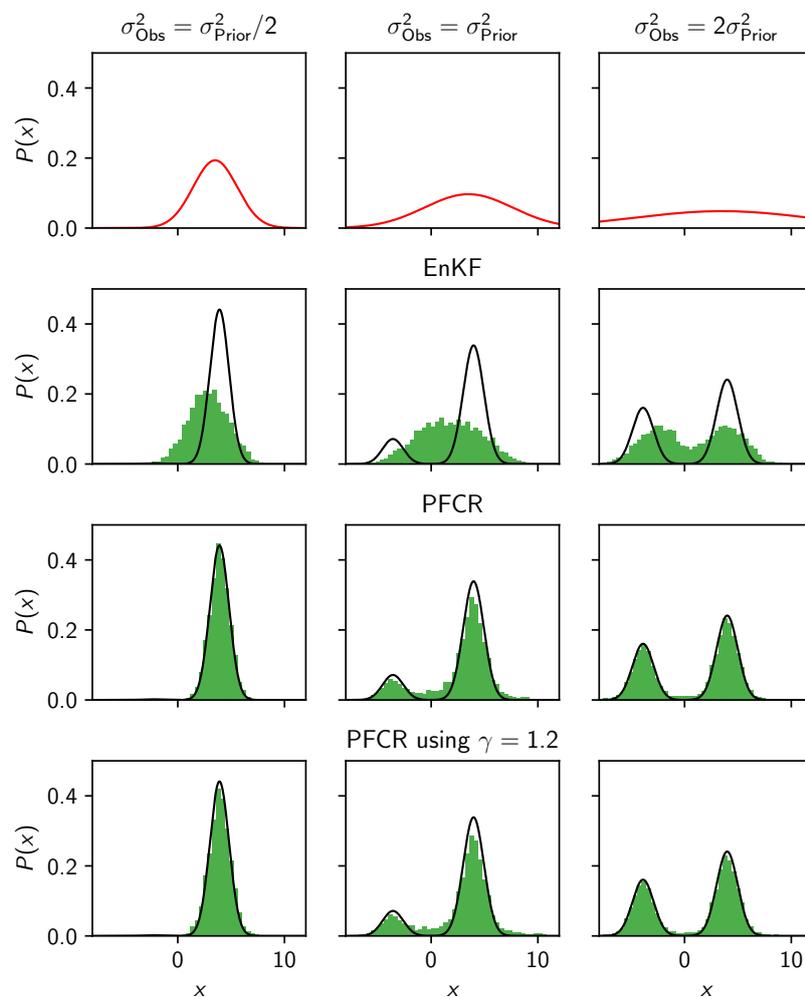


Figure AC2.2: Analysis ensemble (green) for EnKF (second row), particle filter with covariance resampling (PFCR, third row) and PFCR with $\gamma = 1.2$ (bottom row) for a bimodal prior (Fig. AC2.1), for observations (top row, red) with different observation errors: $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2/2$ (left column), $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2$ (middle column) and $\sigma_{\text{Obs}}^2 = 2\sigma_{\text{Prior}}^2$ (right column). The true posterior (black) is calculated with Bayes' theorem. In case of the PFCR, 49.7% (left), 32.6% (middle) and 10.0% (right) particles are resampled.

The deviations from the truth show how the PFCR alters the posterior distribution in this case.

For $\sigma_{\text{Obs}}^2 = 2\sigma_{\text{Prior}}^2$ (right), the posterior is similar to the prior with a less likely left mode. The large uncertainty makes it impossible to confidently decide in which peak the truth lies. The EnKF retains the bimodal distribution of the prior because the Kalman gain becomes small if the observation error is large, which leads to a small correction in the analysis step. The particles have approximately equal weights because of the large observation error. Therefore, only 10.0% of the particles are resampled which reduces the effect that particles are generated in between the peaks as in the case $\sigma_{\text{Obs}}^2 = \sigma_{\text{Prior}}^2$ (middle).

The factor γ has only a small effect in this example. Compared to the EnKF the PFCR describes the real PDF more accurately with only minor deviations from the true PDF.

2) The joint estimation of states and parameters in soil hydrology based on the Richards equation remains a challenging task. Applications with particle filters for state and parameter estimation directly to the Richards equation remain few (e.g. Montzka et al., 2011; Manoli et al., 2015), and estimate 1-5 parameters. We focus on a one-dimensional synthetic case to introduce the covariance resampling and show its behavior in detail.

We agree that the behavior in presence of model errors is relevant. We included the exemplary study of one error, the boundary condition, in the previous revision to show that the PFCR is not limited to perfectly known transition densities. However, an comprehensive investigation of model errors in general is far beyond the scope of this paper.

We improved and clarified the description of the presented example (see also reply to comment 6).

Changes: Section 6.1: We added more explanation of the factor γ , its influence on the estimated pdf and its relation to the multiplicative inflation. Line 5-16, Page 6: We extended the explanation of the influence of the covariance resampling on the estimated pdf.

Specific comments

Comment: 1. Table 1 - alpha values should have units of 1/m? Or are authors presenting values of 1/alpha instead? That explains the very large

Table AC2.1: True Mualem-van Genuchten parameters and range of the uniformly distributed initial guess.

Parameter	Truth	Lower	Upper	Lower _{prev.}	Upper _{prev.}
n_1 [-]	2.28	1.5	6.0	2.2	3.5
n_2 [-]	1.89	1.5	6.0	1.8	3.2
α_1 [m^{-1}]	-12.4	-14	-5	-14	-12
α_2 [m^{-1}]	-7.5	-10.5	-2.5	-10.5	-6.5
$\log_{10}(K_{w,1}), K_w$ in [m s^{-1}]	-4.40	-10	-3	-7	-4
$\log_{10}(K_{w,2}), K_w$ in [m s^{-1}]	-4.91	-10	-3	-7.5	-4

values of -12.4 and -7.5 for α_1 and α_2 as the air-entry value?

Reply: Thank you for noticing this. The unit is m^{-1} .

Changes: Table 1, Page 10: We corrected the unit.

Comment: 2. Parameter ranges are rather narrow. What would happen if we enlarge n to say 1.1 - 6?

Reply: We increased the parameter range for K_w , n and α for both layers (see Table AC2.1) to further investigate the behavior. Note, that for n we only increased the range to 1.5-6.0. The simplified Mualem-van Genuchten parametrization used here is only well-behaved for $n > 2$ (Ippisch et al., 2006) and undefined for $n \leq 1$. For small n the performance of the numerical solver decreases significantly such that $n > 1.1$ is necessary. Since the covariance resampling generates new ensemble members from a Gaussian distribution, it can occur that $n < 1.1$. In this case we set the value to $n = 1.1$. We used the interval 1.5 – 6.0 to reduce the occurrence of this case.

It was necessary to use more ensemble members for a converging result. In this case we used 300 ensemble members. The final state is shown in Fig. AC2.4 and the conductivity function Fig. AC2.3. Both are in good agreement with the synthetic truth.

Changes: Line 1-3, Page 11: We added the sentences: “The filter can also estimate the state and parameters for an extended range.” and “Increasing the initial uncertainty of the parameters, increases the complexity of the problem and the filter needs more ensemble member to converge.” to the manuscript.

Comment: 3. Table 1: Are the $\log(Kw)$ values natural log-values? I do not think so - otherwise we get values of $\exp(-4) * 60 * 60 * 24 = 1600 \text{ m/day}$. So they must be \log_{10} right? With $\log_{10}(Kw)$ values between -7 and -4 the range of Kw is between 0.0086 and 8.64 m/day. The lower bound of 0.86

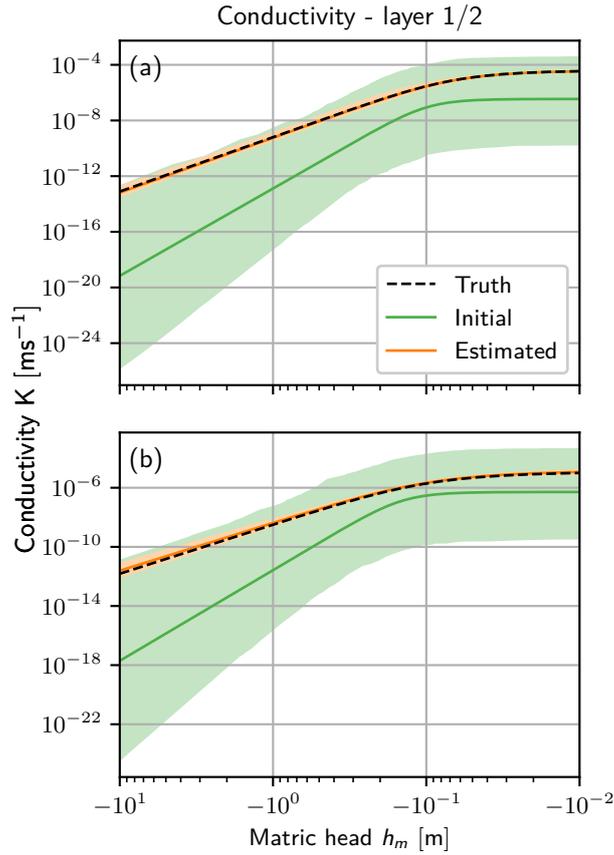


Figure AC2.3: Conductivity function $K(h_m)$ for (a): layer 1 and (b): layer 2. In this function all estimated parameters are represented. The initial 95 %-quantile of the ensemble (light green) with the mean (green) are shown. The truth (dashed black) is almost congruent with the estimated mean (orange), such that only the 95 %-quantile of the final ensemble (light orange) is visible.

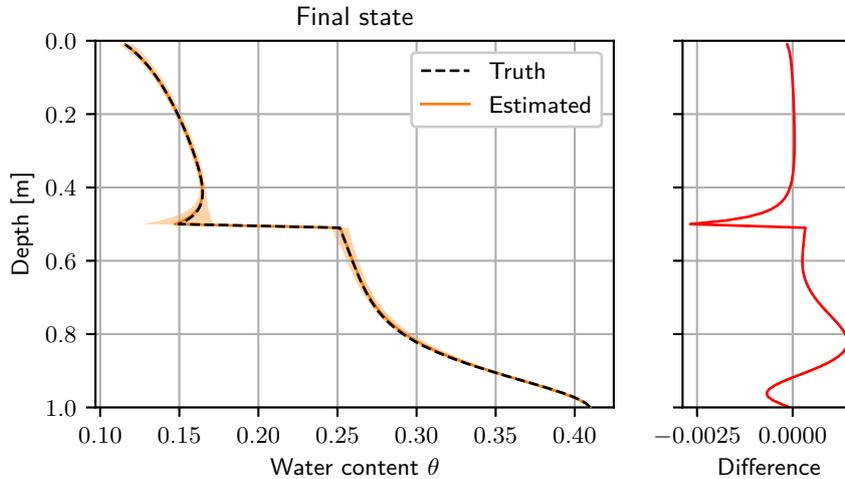


Figure AC2.4: Final water content state after the assimilation run. The truth (dashed black) is almost congruent with the estimated mean (orange), such that only the 95 %-quantile of the ensemble (light orange) is visible. Notice that even the maximal difference is well below current measurement capabilities.

cm/day may still be considered large for say fine textured soils. Therefore I would suggest using a range of $\log_{10}(Kw)$ between -10 and -4 or so. Point is - the parameter space is narrow - this simplifies tremendously inference - but may not necessarily demonstrate convergence properly; how does the filter work with enlarged parameter spaces that encompass a larger ranges of soils?

Reply: Thank you for pointing this out. \log_{10} is correct. For the application with a larger parameter spread please refer to comment 2.

Changes: Table 1, Page 10: We changed \log to \log_{10} and refer to it as decadic logarithm to clarify this (Line 8, Page 10).

Comment: 4. Line 18: *generic algorithm? Or genetic algorithm? Generic algorithm would not make sense as this has been presented before in other papers - for example Particle-DREAM. A genetic algorithm on the contrary will not maintain detailed balance unless they use the Differential Evolution variant with Metropolis step as done in DE-MC and DREAM. I have to read this new paper to understand/see what has been done and how this differs from previous DE-MC/DREAM MCMC work.*

Reply: Thank you for noticing this. It is a genetic algorithm.

Changes: Line 17, Page 2: Changed “generic” to “genetic” algorithm.

Comment: 5. Figure 11: *Those 100 ensemble members - are they randomly*

chosen? Or are these the “best” members; I would be interested to see all the members as “visibility” is not an argument here. You cannot see the 100 different lines anyway (much fewer in practice as particles have many copies). Instead of using Lines you can color the 95 % ranges of all particles.

Reply: The ensemble members in Figure 11 and Figure 12 had been randomly chosen. We follow your idea and show the 95 %-quantile of the ensemble.

Changes: Throughout the manuscript: Changed all figures with an ensemble to show the 95 %-quantile.

Comment: 6. In section 6.3, the authors attempt to address the model error by perturbing the forcing (= boundary conditions). This is not really a model error but measurement error. Indeed, the PF should rapidly address (= correct for) those errors in a few assimilation steps as this entails conservation of mass. What is much more difficult is to address actual model errors due to an incorrect/absent process representation. Those errors lead to much more complicated and predictable residual patterns, making state and certainly parameter estimation more difficult. Will produce parameter estimates that may compensate at least somewhat for the model errors. I think it is important to address, at least in writing, this difference between forcing data errors and model errors. These two errors are not equal.

Reply: We agree that there are different types of model errors. In a real case, an error in the forcing would be a measurement error. Using it in a synthetic study, the changed boundary condition is similar to an additional source/sink term. This leads to a wrong transition density (compared to the truth) since the observations are still generated using the old boundary condition. In the presented case, we have a continuous bias towards lower water content which the filter continuously needs to address. As you mentioned, this leads to a bias in the parameter estimates to compensate the error in the boundary condition. The logarithmic scale in the conductivity plot reduces the visibility of this bias.

Changes: Line 29-31, Page 18: We clarified the effect of the changed boundary condition.

Line 11-12, Page 19: We point out the bias in the parameter.

References

A. Botto, E. Belluco, and M. Camporese. Multi-source data assimilation for physically based hydrological modeling of an experimental hill-

- slope. *Hydrology and Earth System Sciences*, 22(8):4251–4266, 2018. doi:10.5194/hess-22-4251-2018.
- O. Ippisch, H.-J. Vogel, and P. Bastian. Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation. *Advances in Water Resources*, 29(12):1780–1789, 2006. doi:10.1016/j.advwatres.2005.12.011.
- W. Gregory Lawson and James A. Hansen. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Monthly Weather Review*, 132(8):1966–1981, 2004. doi:10.1175/1520-0493(2004)132<1966:IOSADF>2.0.CO;2.
- Jun Man, Weixuan Li, Lingzao Zeng, and Laosheng Wu. Data assimilation for unsaturated flow models with restart adaptive probabilistic collocation based Kalman filter. *Advances in Water Resources*, 92:258–270, 2016. doi:10.1016/j.advwatres.2016.03.016.
- Gabriele Manoli, Matteo Rossi, Damiano Pasetto, Rita Deiana, Stefano Ferraris, Giorgio Cassiani, and Mario Putti. An iterative particle filter approach for coupled hydro-geophysical inversion of a controlled infiltration experiment. *Journal of Computational Physics*, 283(Supplement C):37–51, 2015. doi:10.1016/j.jcp.2014.11.035.
- Carsten Montzka, Hamid Moradkhani, Lutz Weihermüller, Harrie-Jan Hendricks Franssen, Morton Canty, and Harry Vereecken. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of Hydrology*, 399(3):410–421, 2011. doi:10.1016/j.jhydrol.2011.01.020.
- Liangsheng Shi, Xuehang Song, Juxiu Tong, Yan Zhu, and Qiuru Zhang. Impacts of different types of measurements on estimating unsaturated flow parameters. *Journal of Hydrology*, 524:549–561, 2015. doi:10.1016/j.jhydrol.2015.01.078.
- H. Zhang, H.-J. Hendricks Franssen, X. Han, J. A. Vrugt, and H. Vereecken. State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrology and Earth System Sciences*, 21(9):4927–4958, 2017. doi:10.5194/hess-21-4927-2017.

Covariance resampling for particle filter – state and parameter estimation for soil hydrology

Daniel Berg^{1,2}, Hannes H. Bauser^{1,2}, and Kurt Roth^{1,3}

¹Institute of Environmental Physics (IUP), Heidelberg University

²HGS MathComp, Heidelberg University

³Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University

Correspondence to: Daniel Berg (daniel.berg@iup.uni-heidelberg.de)

Abstract. Particle filters are becoming increasingly popular for state and parameter estimation in hydrology. One of their crucial parts is the resampling after the assimilation step. We introduce a resampling method that uses the full weighted covariance information calculated from the ensemble to generate new particles and effectively avoids filter degeneracy. The ensemble covariance contains information between observed and unobserved dimensions and is used to fill the gaps between them. The covariance resampling approximately conserves the first two statistical moments and partly maintains the structure of the estimated distribution in the retained ensemble. The effectiveness of this method is demonstrated with a synthetic case – an unsaturated soil consisting of two homogeneous layers – by assimilating time domain reflectometry (TDR)-like measurements. Using this approach we can estimate state and parameters for a rough initial guess with 100 particles. The estimated states and parameters are tested with a forecast after the assimilation, which is found to be in good agreement with the synthetic truth.

10 1 Introduction

Mathematical models represent hydrological and other geophysical systems. They aim to describe the dynamics and the future development of system states. These models need the current state and certain system parameters (e.g. material properties, forcing) for state prediction. Both, state and system parameters, are typically ill-known and have to be estimated.

Data assimilation methods, originally used for state estimation only, are adapted to also estimate parameters and other model components like the boundary condition. The ensemble Kalman filter (EnKF) (Evensen, 1994; Burgers et al., 1998) is a popular data assimilation method in hydrology. It has the advantage of using the ensemble covariance to correlate dimensions with observations to unobserved dimensions. The EnKF with parameter estimation is applied to several hydrological systems. Moradkhani et al. (2005b) used the EnKF for a rainfall-runoff model and Chen and Zhang (2006) for saturated flow modeling. Using a hydrological model based on the Richards equation, the EnKF is mostly applied in synthetic studies (e.g. Wu and Margulis, 2011; Song et al., 2014; Erdal et al., 2015; Shi et al., 2015; Man et al., 2016). However, some applications to real data exist (e.g. Li and Ren, 2011; Bauser et al., 2016; Botto et al., 2018).

As the EnKF is based on Bayes' theorem, all uncertainties have to be represented correctly, otherwise the method has a poorer performance (Liu et al., 2012; Zhang et al., 2015). Nonlinear systems (e.g. systems described by Richards equation) violate the EnKF assumption of Gaussian probability density functions (Harlim and Majda, 2010; DeChant and Moradkhani, 2012). The

dynamics of Richards equation is generally dissipative and the Gaussian assumption is appropriate. However, jumps at layer boundaries, soliton-like fronts during strong infiltration and diverging potentials for strong evaporation deform the probability density function and lead to non-Gaussianity. In this case the probability density function requires higher statistical moments to be described correctly. Particle filter can accomplish this task.

5 The particle filter has already been used for state and parameter estimation for various hydrological systems. Since parameters do not have their own model dynamics like the state, the resampling step is crucial. Moradkhani et al. (2005a) suggested to perturb the parameters using Gaussian noise with zero mean after the resampling step. They used an unweighted variance of the ensemble modified with a damping factor such that the ensemble spread does not become too large. This method or similar has been used for instance for land surface models (Qin et al., 2009; Plaza et al., 2012), rainfall-runoff models (Weerts and
10 El Serafy, 2006) and soil hydrology (Montzka et al., 2011; Manoli et al., 2015). A common challenge is that with only a rough initial guess, perturbing only the parameters does not guarantee a sufficient ensemble spread and the filter can diverge.

Further development of the resampling for parameter estimation was done by Moradkhani et al. (2012) and Vrugt et al. (2013). They used a Markov chain Monte Carlo (MCMC) method to generate new particles. This method was further used by e.g. Yan et al. (2015) and Zhang et al. (2017). The latter compared the performance of this method with an EnKF and the
15 particle filter presented by Moradkhani et al. (2005a) and found that the performance of the filters were similar with slight advantages for the EnKF. While the MCMC is accurate, it is also expensive, as it requires additional model runs. To increase the efficiency, Abbaszadeh et al. (2018) additionally combined it with a genetic algorithm.

In this paper we introduce the covariance resampling, a resampling method that generates new particles using the ensemble covariance. This method conserves the first two statistical moments in the limit of large numbers while partly maintaining the
20 structure of the estimated distribution in the retained ensemble. With the covariance, the unobserved parameters of the new particles are correlated to the observed state dimensions. The particle filter with covariance resampling is able to estimate state and parameters in case of a difficult initial condition without additional model evaluations, which are necessary for MCMC methods.

2 Particle Filter

25 The particle filter is an ensemble-based sequential data assimilation method that consists of a forecast and an analysis step. The ensemble members are called particles. It is used to combine information from observation and model to a posterior estimate. For a detailed review consider e.g. van Leeuwen (2009).

If new information in the form of observations becomes available, the system is propagated forward to the time the observation is taken (forecast). This results in a prior probability density function. The prior is updated with the information of the
30 observation to get the posterior. This is accomplished using Bayes' theorem,

$$P(\mathbf{u}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{u})P(\mathbf{u})}{P(\mathbf{d})}, \quad (1)$$

which describes the probability of an event \mathbf{u} under the condition of another event \mathbf{d} . In data assimilation this is used to combine the information of the prior $P(\mathbf{u})$ of the state \mathbf{u} with the observation \mathbf{d} . The probability $P(\mathbf{d})$ is a normalization constant

$$P(\mathbf{d}) = \int d\mathbf{u} P(\mathbf{d}|\mathbf{u})P(\mathbf{u}) . \quad (2)$$

This describes the assimilation process for one observation. ~~Suppose For~~ a set of observations $\mathbf{d}^{1:k} = (\mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^{k-1}, \mathbf{d}^k)$, where the ~~exponent superscript~~ denotes a discrete time index. ~~To assimilate these observations sequentially, the observations are assimilated sequentially using~~ the recursive filter equation ~~;~~

$$P(\mathbf{u}^{0:k}|\mathbf{d}^{1:k}) = \frac{P(\mathbf{d}^{1:k}|\mathbf{u}^k)P(\mathbf{u}^k|\mathbf{d}^{1:k-1})}{P(\mathbf{d}^k)} , \quad (3)$$

which follows from Bayes' theorem, ~~is used~~

$$P(\mathbf{u}^{0:k}|\mathbf{d}^{1:k}) = \frac{P(\mathbf{d}^{1:k}|\mathbf{u}^k)P(\mathbf{u}^k|\mathbf{d}^{1:k-1})}{P(\mathbf{d}^k)} .$$

10 The prior distribution at time k

$$P(\mathbf{u}^k|\mathbf{d}^{1:k-1}) = \int d\mathbf{u}^{k-1} P(\mathbf{u}^k|\mathbf{u}^{k-1})P(\mathbf{u}^{k-1}|\mathbf{d}^{1:k-1}) \quad (4)$$

is calculated by propagating the posterior of the previous analysis $P(\mathbf{u}^{k-1}|\mathbf{d}^{1:k-1})$ to time k using the transition density $P(\mathbf{u}^k|\mathbf{u}^{k-1})$.

The particle filter is a Monte Carlo approach, which directly approximates the probability density functions by a weighted ensemble of realizations (particles). This direct sampling allows the particle filter to have non-Gaussian probability density functions. This is in contrast to e.g. the EnKF, which is also based on Bayes' theorem and Monte Carlo sampling but assumes Gaussian distributions.

The posterior distribution of the previous analysis $P(\mathbf{u}^{k-1}|\mathbf{d}^{1:k-1})$ is approximated by an weighted ensemble of N particles, represented by Dirac delta functions

$$20 P(\mathbf{u}^{k-1}|\mathbf{d}^{1:k-1}) = \sum_{i=1}^N w_i^k \delta_D(\mathbf{u}^{k-1} - \mathbf{u}_i^{k-1}) . \quad (5)$$

To obtain the new prior $P(\mathbf{u}^k|\mathbf{d}^{1:k-1})$ for the analysis step at time k , it is necessary to solve the integral in Eq. (4). This is achieved by propagating the ensemble forward in time to the next observation using the model equation (forecast). For this, consider the following generic model equation:

$$\mathbf{u}^k = \mathbf{f}(\mathbf{u}^{k-1}) + \boldsymbol{\beta}^k , \quad (6)$$

25 where $\mathbf{f}(\cdot)$ is the deterministic part of the model and $\boldsymbol{\beta}^k$ is a stochastic model error.

Using Eq. (4), the weights are updated according to

$$w_i^k = w_i^{k-1} \frac{P(\mathbf{d}^k|\mathbf{u}_i^k)}{P(\mathbf{d}^k)} . \quad (7)$$

After the analysis the weights are ~~normalised~~normalized using the fact that the sum has to be equal to one.

$$\sum_{i=0}^N w_i^k \stackrel{!}{=} 1 \quad \Rightarrow \quad P(\mathbf{d}^k) = \sum_{i=0}^N w_i^{k-1} P(\mathbf{d}^k | \mathbf{u}_i^k). \quad (8)$$

In general, $P(\mathbf{d}^k | \mathbf{u}_i^k)$ is an arbitrary distribution that represents the observation error. We assume Gaussian distributed observation errors which results in:

$$5 \quad P(\mathbf{d}^k | \mathbf{u}_i^k) \propto \exp \left[(\mathbf{d}^k - \mathbf{H}(\mathbf{u}_i^k))^\top \mathbf{R}^{-1} (\mathbf{d}^k - \mathbf{H}(\mathbf{u}_i^k)) \right], \quad (9)$$

where \mathbf{R}^{-1} is the inverse of the observation error covariance and \mathbf{H} is the observation operator that projects the state \mathbf{u} from state-space to observation-space.

To estimate state and parameters simultaneously we use an augmented state. In our case the augmented state \mathbf{u} consists of the state $\boldsymbol{\theta}$ (water content) and a set of parameters \mathbf{p}

$$10 \quad \mathbf{u} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{p} \end{bmatrix}. \quad (10)$$

3 Resampling

Particle filters tend to filter degeneracy, which is also referred to as filter impoverishment. After several analysis steps, one particle gets all statistical information as its weight becomes increasingly large, whereas the remaining particles only get a small weight such that the ensemble effectively collapses to this one particle. In this case, the filter does not react on new observations and follows the particle with the large weight.

Gordon et al. (1993) introduced resampling to particle filters, a technique that reduces the variance in the weights and has the potential to prevent filter degeneracy. The idea of resampling is that after the analysis, particles with large weights are replicated and particles with small weights are dropped. This helps that the regions with high weighted particles are represented better by the ensemble, which alleviates the degeneracy of the filter. Filters using resampling are referred to as sequential importance resampling (SIR). There are many different resampling algorithms (see van Leeuwen (2009) for a summary). One of these methods is the stochastic universal resampling.

3.1 Stochastic ~~Universal Resampling~~universal resampling

The stochastic universal resampling (Kitagawa, 1996) can be summarized as follows (see also Fig. 1): All weights are aligned after each other on an interval $[0, 1]$. A random number in the interval $[0, N^{-1}]$ is drawn from a uniform distribution. This number points to the first particle of the new ensemble, selected by the corresponding weight. Then N^{-1} is added $(N-1)$ -times to x . Each of the endpoints selects the corresponding particle for the new ensemble. This way some particles get duplicated and some particles are dropped. With this approach, particles with a weight smaller than N^{-1} can be chosen maximally once, whereas a weight larger than N^{-1} guarantees that the particle is at least chosen once. If all particles have equal weights, no

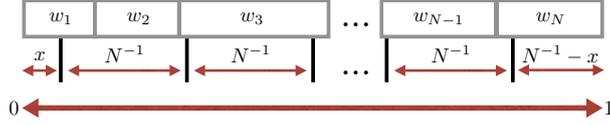


Figure 1. Illustration of the universal resampling process. A random number x is drawn from a uniform distribution in the interval $[0, N^{-1}]$. The endpoint of this number indicates the first particle. Then N^{-1} is added $(N - 1)$ -times to this random number, where every endpoint is a particle of the new ensemble. In the illustration, particle one is chosen once, particle two not once and particle three twice. This way some particles are replicated and other particles are dropped. If the model does not have a stochastic model error, it is necessary to perturb the new particles, otherwise they would be identical and the filter would degenerate.

particle is dropped. The result is a new set of N particles. After the resampling step, all weights are set to N^{-1} . The stochastic universal resampling has a low sampling noise compared to other resampling methods (van Leeuwen, 2009).

3.2 Covariance Resampling

If the model does not have a stochastic model error, like we consider in this study, it is necessary to perturb the particles, otherwise they would be identical and the filter would still degenerate. Even in the presence of a model error it can be useful to perturb the particles after the resampling step. For example if the model error is ill-known or structurally incorrect, it can help to guarantee a sufficient ensemble spread and diversity.

There are different suggestions how to do the perturbation. For example, Moradkhani et al. (2005a) used the ensemble variance to perturb the parameters with a Gaussian with zero mean. Pham (2001) proposed to sample new particles by perturbing the identical particles using a Gaussian with the (damped) ensemble covariance matrix as covariance. Xiong et al. (2006) sampled the whole ensemble from a Gaussian using the first two moments specified by the ensemble (full covariance information), which neglects the particle filter ability to use non-Gaussian distribution. All of these methods have in common, that they alter the estimated pdf-distribution to ensure a diverse ensemble.

We neither perturb the duplicated states nor draw a complete new ensemble. The covariance resampling we propose uses the universal resampling – other resampling methods can be equally used – to choose the ensemble members that are kept. Instead of duplicating the particles and setting the weights to N^{-1} , the weight of the particles is changed to

$$w_i = \frac{z}{N} \quad \text{with} \quad i \in \{1, 2, \dots, N'\}, \quad (11)$$

where the particle i is chosen z -times and N' is the number of kept particles. In the statistical limit this conserves the estimated distribution.

The total ensemble reduces to N' . To have N ensemble members again, $N - N'$ new particles have to be generated. These particles are sampled from a Gaussian $\mathcal{N}(\bar{\mathbf{u}}, \mathbf{P}^f)$ with the weighted mean

$$\bar{\mathbf{u}} = \sum_{i=1}^N w_i \mathbf{u}_i. \quad (12)$$

and the weighted covariance

$$\mathbf{P}^f = \frac{1}{1 - \sum_{i=1}^N w_i^2} \sum_{i=1}^N w_i [\mathbf{u}_i - \bar{\mathbf{u}}] [\mathbf{u}_i - \bar{\mathbf{u}}]^\top, \quad (13)$$

where the factor $\frac{1}{1 - \sum_{i=1}^N w_i^2}$ is Bessel's correction for an unbiased estimate of the weighted covariance. Mean and covariance are calculated using the weights before resampling (Eq. (7)).

- 5 ~~Sampling only the dropped particles from Gaussian conserves the mean and the covariance in the limit of large numbers with the advantage that the structure of the non-Gaussian distribution is partly conserved in the retained ensemble.~~ A weight of N^{-1} is assigned to each of the new particles, which results in a sum of all weights larger than one. Therefore, it is necessary to normalize the weights again. This results in a superposition of the estimated distribution and a Gaussian. ~~If only a few particles are resampled, the distribution remains close to the previously estimated. For-~~
- 10 Since the dropped particles are sampled from a Gaussian, the mean and the covariance are conserved in the limit of large numbers. However, the structure of the non-Gaussian distribution is only partly conserved through the retained ensemble. In more difficult situations, where an increasing fraction of ~~resampled particles~~ particles is resampled, the posterior is dominated by the approximated multivariate Gaussian. However, the approximation allows the use of the covariance information in the ensemble, which facilitates the generation of meaningful new particles and improves the exploration of the state space. In less
- 15 difficult situations, when only a few particles are resampled, the distribution remains close to the previously estimated, which includes the full structure of the estimated distribution.

Using the multivariate Gaussian utilizes the information of the covariance but sacrifices the more accurate description of the univariate distribution that could be achieved by a kernel density estimation. However, it requires a much smaller sample size compared to a multivariate kernel density estimation.

- 20 The whole resampling process is illustrated in Fig. 2. For the pseudocode of the covariance resampling please refer to Appendix A.

New particles are generated with a Cholesky decomposition of the covariance matrix. The calculation of the covariance from the ensemble can result in small numerical errors that have to be ~~regularised~~ regularized, otherwise the decomposition would fail. For details about the generation of new particles and ~~regularisation~~ regularization of the covariance matrix see Appendix

- 25 B.
- Pham (2001) introduced a tuning parameter to modify the covariance matrix and Moradkhani et al. (2005a) for the variance, respectively. They used the tuning factor to reduce the amplitude of the perturbation. For the covariance resampling we also introduce a tuning parameter. If the model dynamics does not support a sufficient spread for the ensemble, the perturbation of the covariance resampling has to be large enough to prevent the ensemble from degeneracy. One example for such a case are
- 30 parameters. The covariance matrix is modified by a multiplicative factor γ

$$\mathbf{P}^{f'} = (\gamma \gamma^\top) \circ \mathbf{P}^f, \quad (14)$$

where \circ is the entrywise product (Hadamard product). In the case of parameters the factor is chosen larger than one for the parameter space to provide a sufficient ensemble spread.

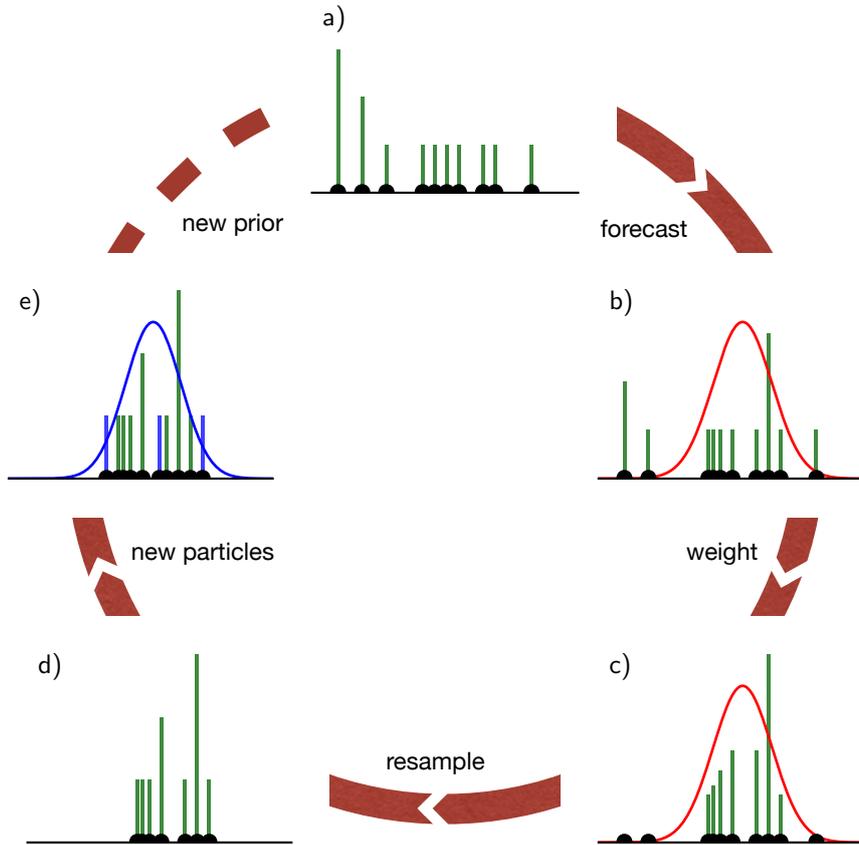


Figure 2. Illustration of the particle filter with covariance resampling. The green bars show the weight of each ensemble member (ten in this example) and the black dots the position of it. (a) The prior represented through the ensemble. (b): The ensemble is propagated to the next observation (depicted as Gaussian, red curve). (c): The particles are weighted according to the observation. At this point, some particles have already negligible weight. (d): The universal resampling drops ~~particles~~ particles with low weight (three in this example). Instead of adding new particles at the same position, only the weights of the kept particles are changed. If a particle is resampled k -times, it will get the weight $k N^{-1}$. The ensemble size is reduced and new particles have to be added to conserve the ensemble size and avoid filter degeneration. (e): The new particles are drawn from the full covariance of the ensemble (Eq. (13)) and their weight is set to N^{-1} . Since new particles with weights are added to the ensemble, it is necessary to normalize the weights to one again. This results in the posterior or the next prior respectively. The pseudocode for the algorithm can be found in Appendix A.

4 Case study

The algorithm is tested using a synthetic case study of a one-dimensional unsaturated porous medium with two homogeneous layers. The system has a vertical extent of 1 m with the layer boundary in the middle at 50 cm. The representation of the considered system is described following the structure of Bauser et al. (2016). The general representation of a system has four

components: *dynamics, forcing, subscale physics* and *state*. The dynamics propagates the state forward in time, conditioned on the subscale physics and forcing.

The dynamics in an unsaturated porous medium can be described by the Richards' equation

$$\partial_t \theta - \nabla \cdot [K(\theta) [\nabla h_m - 1]] = 0, \quad (15)$$

- 5 where $h_m(\text{L})$ is the matric head, $K(\text{LT}^{-1})$ the isotropic hydraulic conductivity and $\theta(-)$ the volumetric water content. We use the finite-element solver MuPhi (Ippisch et al., 2006) to solve Richards' equation numerically. The resolution is set to 1 cm which results in a 100-dimensional water content state.

- The macroscopic material properties represent the averaged subscale physics with the functions $K(\theta)$ and $h_m(\theta)$ and a set of parameters. In this study, the ~~Mualem-van Genuchten parametrisation~~ Mualem-van Genuchten parametrization is used
 10 (Mualem (1976), Van Genuchten (1980)):

$$K(\Theta) = K_w \Theta^\tau \left[1 - \left[1 - \Theta^{n/[n-1]} \right]^{1-1/n} \right]^2, \quad (16)$$

$$h_m(\Theta) = \frac{1}{\alpha} \left[\Theta^{-n/[n-1]} - 1 \right]^{1-1/n}, \quad (17)$$

with the saturation $\Theta(-)$

$$\Theta := \frac{\theta - \theta_r}{\theta_s - \theta_r}. \quad (18)$$

- 15 With these equations the subscale physics is described by six parameters for each layer. The parameter $\theta_s(-)$ is the saturated water content and $\theta_r(-)$ the residual water content. The matric head h_m is scaled with the parameter $\alpha(\text{L}^{-1})$ that can be related to the air entry value. The parameter $K_w(\text{LT}^{-1})$ is the saturated hydraulic conductivity, $\tau(-)$ a tortuosity factor and $n(-)$ is a shape parameter. In this study the parameters α, n and K_w will be estimated for each layer. Combining Eq. (17) and Eq. (16) results in a conductivity function

$$20 \quad K(h_m) = K_w \left[1 + (\alpha h_m)^n \right]^{-\tau(1-1/n)} \left[1 - (\alpha h_m)^{n-1} (1 + (\alpha h_m)^n)^{-1+1/n} \right]^2 \quad (19)$$

that incorporates all estimated parameters.

- For the true trajectories and the observations, parameters by Carsel and Parrish (1988) for loamy sand are used for the upper layer (layer 1) and sandy loam for the lower layer (layer 2). Table 1 gives the true values for the estimated parameters and Table 2 the values for the fixed parameters, respectively. In the following the parameters will have an index representing their
 25 corresponding layer.

Since state and parameters are estimated, we separate the model equation Eq. (6) into

$$\mathbf{u}^n = \begin{bmatrix} \boldsymbol{\theta}^k \\ \mathbf{p}^k \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\boldsymbol{\theta}^{k-1}, \mathbf{p}^{k-1}) \\ \mathbf{p}^{k-1} \end{bmatrix}, \quad (20)$$

with a constant model for the parameters \mathbf{p} and Richards' equation as $f(\cdot)$. Note that the model error of equation Eq. (6) is set to zero. In hydrology the model error is typically ill-known and can vary both in space and time.

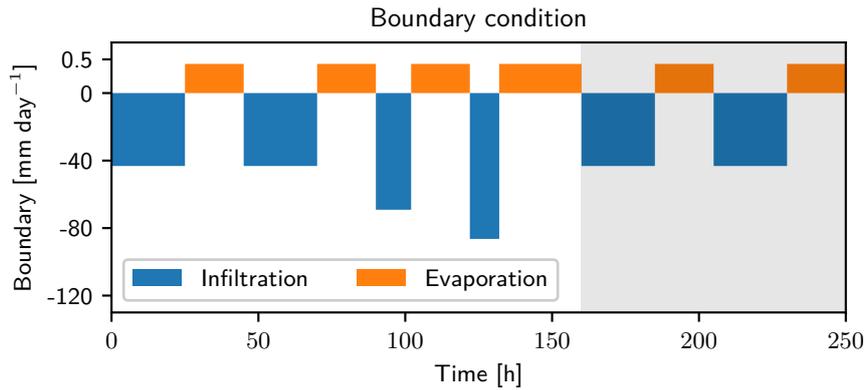


Figure 3. Upper boundary condition for the data assimilation case. Four rain events (blue) followed by a dry period (orange) are used for the test of the data assimilation run. After this run, two additional rain events and dry periods are used in a free run to test the assimilation results (grey background). The intensity and duration of these events is set equal to the first events of the data assimilation run. Note the different axes for infiltration and evaporation.

The forcing is reflected in the boundary condition of the simulation. For the lower boundary, a Dirichlet condition with zero potential (groundwater table) is used. The upper boundary condition is chosen as a flux boundary (Neumann), representing four rain events with increasing intensity and dry periods in between (see Fig. 3).

Using infiltrations with an increasing intensity has the advantage that the system has more time to adjust the parameters. This way less observations are necessary to resolve the infiltration front and the information of the observations can be incorporated in the state and parameters. The stronger infiltration front in the end is used to test the robustness of the estimated state and ~~paramters~~parameters.

The last component of the system is the state. Initially, the system is in equilibrium and will be forced by the boundary condition. The initial state is depicted in Fig. 4. Six TDR-like observations are taken equidistantly in each layer at the positions (0.1, 0.25, 0.3) m for layer 1 and (0.6, 0.75, 0.9) m for layer 2. The measurement error is chosen to be $\sigma_{\text{Obs}} = 0.007$ (e.g. Jaumann and Roth, 2017). Observations are taken hourly for the duration of 160 h.

For the initial state of the data assimilation, the observations at time zero are used. The measured water content is interpolated linearly between the measurements and kept constant towards the boundary. Additionally, the saturated water content for loamy sand, which is 0.41 is taken as the lower boundary. The approximated state is used as the ensemble mean for the particle filter. This procedure is a viable option for real data although it represents a rather crude approximation of the real initial condition.

The approximated state is perturbed by a correlated multivariate Gaussian. The main diagonal of the covariance matrix is 0.003^2 . The variance is chosen such that the ensemble represents the uncertainty of the water content in most parts (see Fig. 4). The off-diagonal entries are determined by the following two steps: (i) All covariances between the two layers are set to zero to ensure no correlations across the layer boundary. (ii) The remaining entries are the variance of the main diagonal

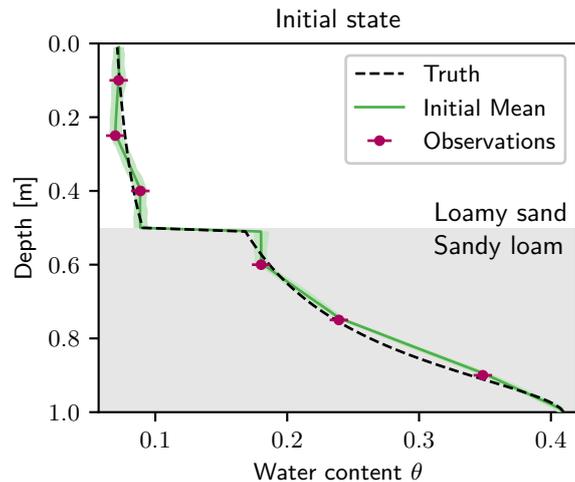


Figure 4. Initial state for the data assimilation run. Observations (purple) at time zero are connected linearly and set constant towards the layer and upper boundary. For the lower boundary, the saturated water content $\theta_r = 0.41$ of sandy loam is used for the interpolation. The ensemble (light green) with 100 ensemble members is generated by perturbing the interpolated state using a spatially correlated Gaussian. The 95 %-quantile of the initial ensemble is shown in light green. The initial truth that is used for the observations (purple) is shown as a black dashed line.

Table 1. True Mualem-van Genuchten parameters and range of the uniformly distributed initial guess.

Parameter	Truth	Lower	Upper
n_1 [-]	2.28	2.2	3.5
n_2 [-]	1.89	1.8	3.2
α_1 [m] α_1 [m^{-1}]	-12.4	-14	-12
α_2 [m] α_2 [m^{-1}]	-7.5	-10.5	-6.5
$\log(K_{w,1}), K_w$ in [m s^{-1}] $\log_{10}(K_{w,1}), K_w$ in [m s^{-1}]	-4.40	-7	-4
$\log(K_{w,2}), K_w$ in [m s^{-1}] $\log_{10}(K_{w,2}), K_w$ in [m s^{-1}]	-4.91	-7.5	-4

multiplied with the Gaspari and Cohn function (Gaspari and Cohn, 1999). The distance for the Gaspari and Cohn function is the distance of the off-diagonal-off-diagonal entry from the main diagonal and a length scale of $c = 10$ cm is used. This way, the water content is only correlated in the range of 20 cm.

The use of the covariance increases the diversity of the ensemble and also helps to avoid degeneration. Using uncorrelated Gaussian random numbers with zero mean would result in a fast degeneration of the particle filter caused by the dissipative nature of the system. The perturbation would simply dissipate and the ensemble collapses.

The initial parameters for the ensemble are uniformly distributed. The ranges of the uniform distributions are given in Tab. 1. Note that the decadic logarithm of the saturated conductivity K_w is used for the estimation, so K_w spans three order-of

Table 2. Fixed ~~Mualem-van-Mualem-van~~ Genuchten parameters.

Parameter	Layer 1	Layer 2
θ_s [-]	0.41	0.41
θ_r [-]	0.057	0.065
τ [-]	0.5	0.5

~~magnitudes. The orders of magnitude. The filter can also estimate the state and parameters for an extended range. In this study,~~
~~the ensemble size is 100. Increasing the initial uncertainty of the parameters, increases the complexity of the problem and the~~
~~filter needs more ensemble members to converge.~~ The multiplicative factor Eq. (14) is set to:

$$\gamma = \begin{bmatrix} \gamma_{\theta,100} \\ \gamma_{p,6} \end{bmatrix}, \quad (21)$$

- 5 where ~~the index γ is separated to γ_{θ} and γ_p for the water content and the parameter, respectively. The number in the subscript~~
~~denotes the dimension of the factor.~~ The covariance in the 100-dimensional state space is unchanged. For the parameter space
a factor of 1.2 is used to compensate the missing dynamics. ~~The subscript for the dimension will be omitted in the following.~~

After the assimilation, the ~~final mean of the augmented state ensemble~~ is used to run a forecast without data assimilation
~~and the mean is calculated from the propagated ensemble using the weights of the last assimilation time.~~ In this run two
10 additional infiltration events and evaporation periods (see Fig. 3) are used to test the forecasting ability of the estimated states
and parameters.

5 Results

The development of the parameters is depicted in Fig. 5. The saturated conductivity $K_{w,1}$ (Fig. 5a) can be estimated quickly
because the infiltration front induces dynamics in the first layer which is strongly dependent on K_w . Instead, $K_{w,2}$ (Fig. 5b) is
15 not sensitive to the dynamics in the first hours, as the infiltration front did not reach the second layer yet. At around 46 h, the
infiltration front reaches the first observation position in the second layer and the estimation for $K_{w,2}$ improves quickly.

If dynamic is induced in the system, the ensemble spread in the water content space increases because of different parameter
sets. This makes the particles and their corresponding parameter sets distinguishable and parameter estimation possible. The
parameters n_1 and n_2 (Fig. 5c and d) as well as α_2 (Fig. 5f) can be estimated well. One exception is α_1 (Fig. 5e). This parameter
20 is insensitive to the observations. The effect of α on the trajectory of the ensemble members is limited to a small region next to
the layer boundary. Further away, wrong values can be compensated by n . The parameter α_1 jitters randomly around a value
slightly larger than the truth. In a synthetic case, the estimation of α_1 can be improved easily by having observations directly
next to the boundary. This way a different value for α_1 would have a direct influence on the trajectory and α_1 would become
sensitive to the observations. However, in reality it is not feasible to change the measurement position or measure directly next
25 to the layer interface. We decided to retain these positions to illuminate an often encountered practical difficulty.

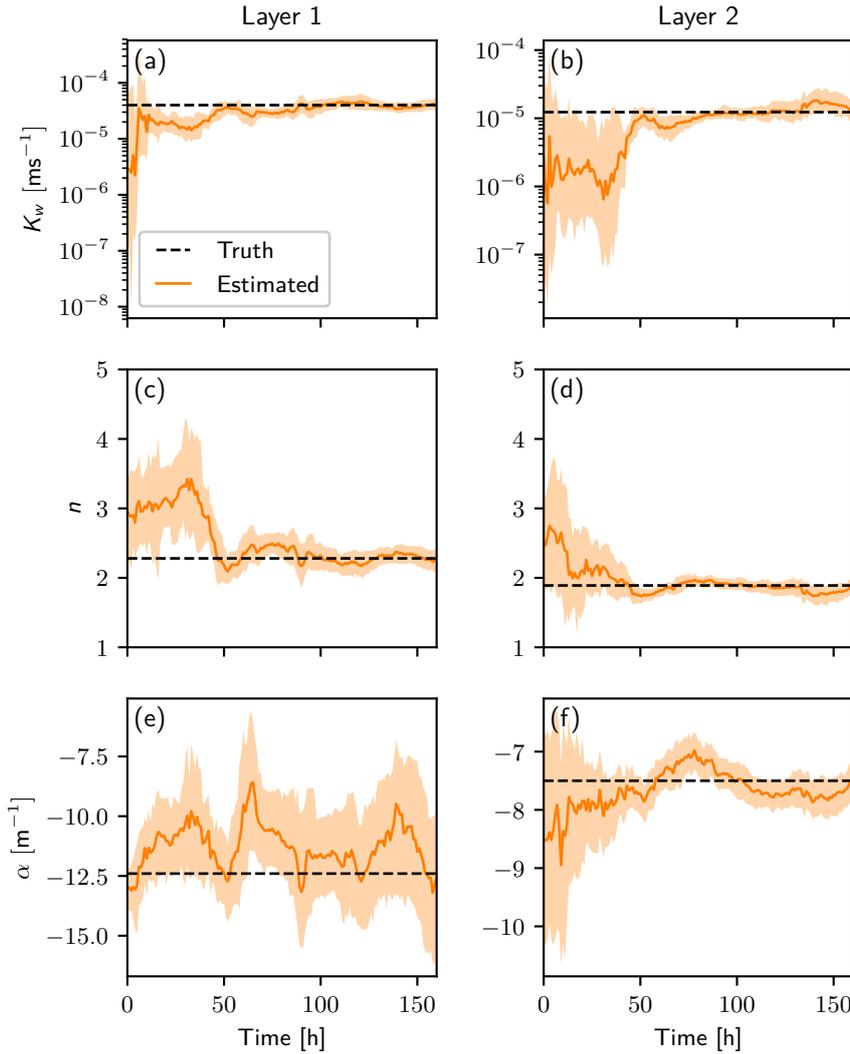


Figure 5. Estimation of all six parameters ((a): $K_{w,1}$, (b): $K_{w,2}$, (c): n_1 , (d): n_2 , (e): α_1 , (f): α_2) over time. The ensemble mean is shown in orange and the 95 %-quantile of the ensemble in light orange. The truth is a dashed black line.

To see the effect of the parameters on the forward propagation, it is necessary to have a closer look at the conductivity function Eq. (19). This function is used for the model forward propagation and many parameter sets can effectively describe the same situation in a limited regime of the hydraulic head. The function is shown in Fig. 6 for both layers ~~and reveals that the~~ for the prior and the final parameters. The difference between the truth and the estimated parameters is small ~~-even though the~~ 95 %-quantile of the prior ensemble does not include the truth for small h_m for layer 1.

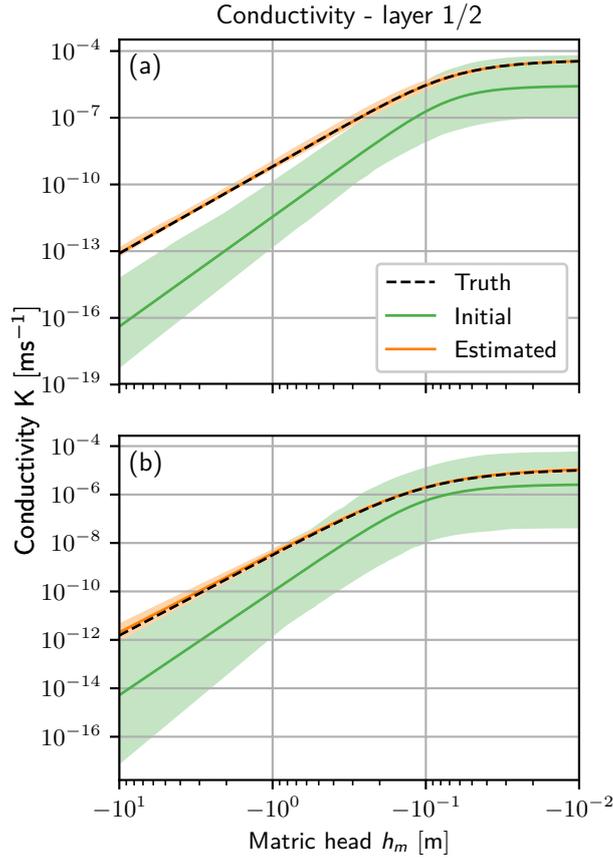


Figure 6. Conductivity function $K(h_m)$ (Eq. (19)) for (a): layer 1 and (b): layer 2. In this function all estimated parameters are represented. The initial 95 %-quantile of the ensemble (light green) with the mean (green) are shown. The truth (dashed black) is almost congruent with the estimated mean (orange), such that only the 95 %-quantile of the final ensemble (light orange) is visible.

The final water content state estimated with the particle filter agrees with the synthetic truth in a narrow band, while the mean of the ensemble propagated without data assimilation is far off (see Fig. 7). The estimated state is slightly biased to higher water contents. We checked that the direction of the bias is a random effect and is different for different seeds. The observation of a bias instead is caused by long range correlations of the model. In this case, the system has started to relax after the last infiltration and a higher water content in one part results in a higher water content in the rest of the layer. The ensemble spread next to the layer boundary is caused by the large spread of α_1 .

To analyse-analyze the ensemble, we take a closer look at the effective sample size and the number of particles that are resampled. The effective sample size is defined as (Doucet, 1998):

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^N w_i^2} \quad (22)$$

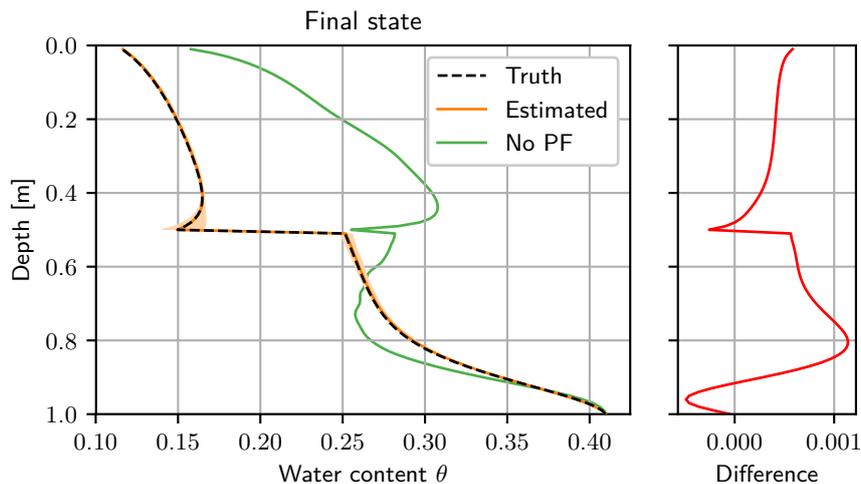


Figure 7. Final water content state after the assimilation run. The truth (dashed black) is almost congruent with the estimated mean (orange), such that only the final-95%-quantile of the ensemble (light orange) is visible. The final ensemble with the corresponding weights is used to start a free forward run afterwards. The mean without data assimilation (green) is calculated from a forecast of the initial ensemble (see Fig. 4 and Fig. 6). The difference of the estimated water content and the synthetic truth lies in a narrow band, with a small bias to larger water contents.

and gives an estimate of how many particles are significant. For example, if one particle accumulates all the weight $N_{\text{eff}} = 1$ and the ensemble is effectively described only by this particle.

Fig. 8 shows the effective sample size and the number of new particles over time. The effective sample size drops every time new information is available and the number of resampled particles increases. For times $t < 15\text{h}$, the effective sample size drops to small values. The infiltration front propagates through the first layer, which leads to a large ensemble spread caused by unknown parameters and only a few particles have a significant weight. The filter assimilates the information from the observations and resamples particles with low weight. This improves the state and parameters and leads to an increasing effective sample size until the infiltration front reaches the second layer ($t \approx 46\text{h}$). The effective sample size decreases rapidly because the parameters in the second layer are still unknown and lead to a large ensemble spread again. This variation of the effective sample size occurs every time the filter gets new information that leads to a discrepancy between states of the particles and the observations.

The effective sample size is a crucial parameter for the covariance resampling. If it drops to low values the filter can degenerate because effectively too few particles contribute to the weighted covariance (Eq. (13)) and the covariance information becomes insignificant.

For further analysis, the RMSE is calculated based on the difference of the true water content and the weighted mean at every observation time. This includes also the unobserved dimensions. The RMSE shows a similar behaviour-behavior as the

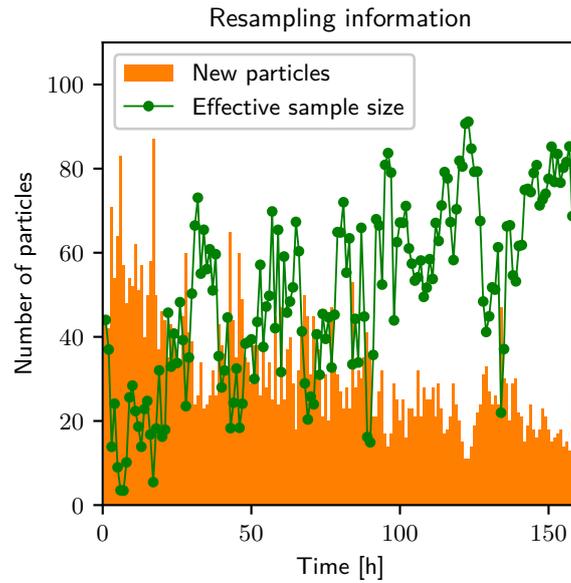


Figure 8. Amount of particles that are resampled (orange) and the effective ensemble sample size (green dots). The lines connecting the dots are for better visibility of the time dependent behaviour. The effective sample size increases while the number of resampled particles decreases. Every infiltration reduces the effective sample size and leads to more resampled particles.

parameters and the effective sample size (see Fig. 9). During the first infiltration, the state and the parameters are improved, the RMSE becomes smaller. Then the infiltration front reaches the boundary interface. The parameters of the second layer and α_1 are still wrong and diverse. This leads to a spread of the ensemble by the system dynamics and also a shift of the mean away from the truth. The parameters in the second layer are estimated and the state is corrected, which leads to a decrease in the RMSE. The increase for the next infiltration events becomes smaller since state and parameters are already improved.

After the data assimilation, the final ensemble including the weights are used for a forecast. This forward run without data assimilation shows that the RMSE oscillates in a narrow range. These oscillations are part of the unobserved space next to the boundary and are mainly caused by the wrong value of α for the first layer. In the beginning, the RMSE stays small, but when the infiltration front reaches the boundary of the two layers, the mean state and the truth begin to deviate from each other (limited to the boundary interface). After the infiltration front passed, the state starts to equilibrate and is increasingly defined by the whole parameter set, which has a certain distance to the true equilibrium.

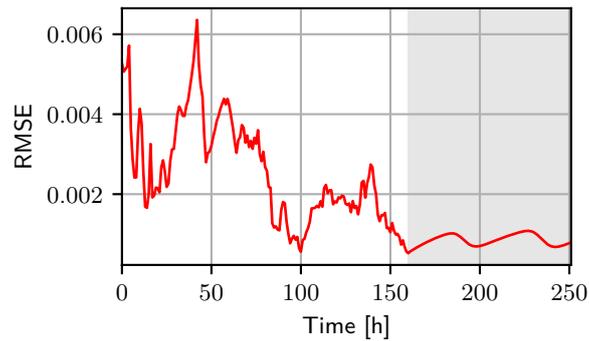


Figure 9. The RMSE (red) of the water content calculated between the truth and the estimated mean including all dimensions. After 160 hours the free run starts (grey background). The mean of the free run is calculated using the propagated ensemble members with their corresponding weights at the last assimilation time. During this forward runtime, the RMSE is about 10^{-3} . For the assimilation and the free run the RMSE increases with each infiltration.

6 Practical ~~Considerations~~considerations

For the presented case study, this section explores two issues relevant when applying the proposed covariance resampling method: (i) the choice of the factor γ in interplay with the ensemble size for different seeds and (ii) the effect of a model bias, in our case simulated using a biased upper boundary condition.

5 6.1 Tuning parameter γ

To explore the convergence of the particle filter with covariance resampling, the filter was run for 40 different seeds, varying ensemble sizes and for four different tuning parameters γ (see Eq. (14)). The tuning parameter is only changed for the parameter space γ_p , while in state space the same value is used as in the case study ($\gamma_\theta = 1.0$, $\gamma_\alpha = 1.0$). Four different tuning parameters are used: $\gamma_p = 1.06$, $\gamma_p = 1.0$ (no change in the covariance), $\gamma_p = 1.16$, $\gamma_p = 1.26$, $\gamma_p = 1.1$, $\gamma_p = 1.2$ (also used in the case study) and $\gamma_p = 1.36$, $\gamma_p = 1.3$. The remaining setup of the system (e.g. initial condition, boundary condition etc.) is identical as in Sect. 4.

Figure C1 shows the saturated conductivity in the second layer $K_{w,2}$, including the 70%-quantile (darker blue area) and the 90%-quantile (light blue area) of the 40 run with 10a shows the RMSE of the water content, calculated between the truth and the estimated mean at the last observation time. The RMSE is averaged over the 40 different seeds.

15 For all ensemble sizes, the filter either degenerates or finds the true value. Increasing the ensemble size increases the number of successful runs. The degeneration of the filter can directly be seen in the effective sample size, which drops to $N_{\text{eff}} = 1$. Therefore, we emphasize the need to control whether the filter degenerates or not, to ensure a meaningful result. Results generated with a degenerated filter must not be used.

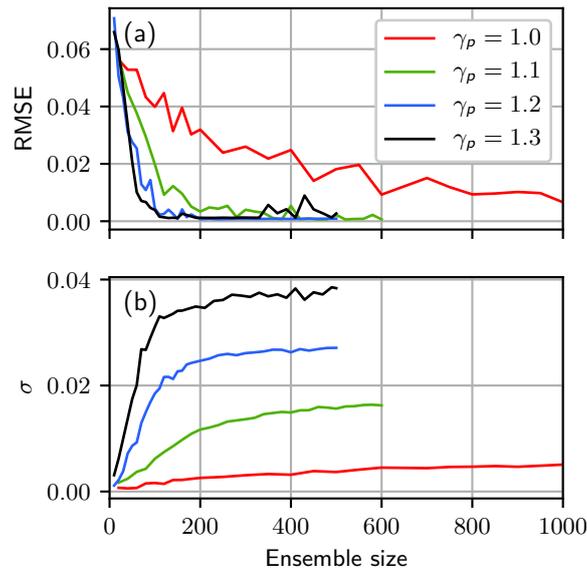


Figure 10. The saturated conductivity in RMSE and the second-layer after standard deviation of the data assimilation run for 40 water content at the last observation time (160 h), averaged over the 40 different seeds realizations. The RMSE is calculated between the truth and the estimated mean. Both quantities are shown for varying factors of γ_p (Eq. (14)): $\gamma_p = 1.0$ (red); $\gamma_p = 1.1$ (green); $\gamma_p = 1.2$ (blue); $\gamma_p = 1.3$ (black), respectively. Note the different scaling of the x-axes.

For small ensemble sizes the case $\gamma_p = 1.0$ (see Fig. C1a), which does not change the covariance matrix, the filter needs about 800 ensemble members to converge for 70% of the seeds. It still degenerates for some seeds. Increasing the tuning factor for the parameter to $\gamma_p = 1.1$ (see Fig. C1b), filter degenerates for every value of γ_p , which leads to a large RMSE. Except for the case $\gamma_p = 1.0$, the RMSE converges for less than 200 ensemble members to a common value independent of the tuning parameter. For $\gamma_p = 1.0$ the RMSE approaches this value as well, but does not reach it completely even for 1000 ensemble members. While the use of the tuning factor is not mandatory, increasing γ_p to a value slightly larger than one reduces the necessary ensemble size and the seed dependency. For 300 ensemble members, the 90%-quantile converges to the truth by an order of magnitude.

In Fig. C1c the tuning parameter is equal to the one used in the case study in Sect. 4. For less than 100 ensemble members, the behavior of the filter is seed dependent. While for some seeds the filter still converges for 20 ensemble members, it degenerates in most of them. Figure 10b shows the standard deviation σ of the ensemble in water content space at the last observation time, averaged over the 40 different realizations. For 100 ensemble members, the ensemble size used in the case study, the filter converges for every of the 40 seeds.

The apparent bias to a larger saturated conductivity for $\gamma_p = 1.26$ is compensated by the other two estimated parameter in this layer, such that the conductivity function Eq. is almost identical to the truth in the measured water content range. small ensemble sizes the filter degenerates for most of the 40 runs. In this case the standard deviation is zero. Increasing the ensemble size, increases the number of successful runs and the standard deviation converges to a final value. The convergence is similar to
5 the convergence of the RMSE in Fig. 10a. However, the ensemble converges to different σ for different values of γ_p . The tuning factor affects the covariance of the newly generated particles and thus an increasing factor results in an increased variance in the estimated distribution. The standard deviation of the ensemble is overestimated for $\gamma > 1$. The mean is not influenced for the chosen values of γ_p . However, increasing the value further will eventually increase the uncertainty too strong and influence the estimation itself (see Supplementary). For an analysis of the estimated mean for the saturated conductivity in the second
10 layer please refer to Appendix C.

Increasing the factor to $\gamma_p = 1.36$ (see Fig. C1d), does not change the result significantly compared to the case $\gamma_p = 1.26$ (C1e). The tuning factor has similarities to multiplicative inflation for the EnKF (Anderson and Anderson, 1999). It increases the uncertainty and reduces filter degeneracy. However, choosing a too large value for the tuning parameter results in an increasing uncertainty, which leads to a divergent ensemble for insensitive parameters like α_1 . The simple choice of a constant
15 multiplicative factor γ can lead to too large uncertainties. For a better uncertainty estimation it is necessary to set $\gamma = 1$. This requires a larger ensemble size. Therefore, it is important to check the results and adjust the tuning parameter accordingly. It is always possible
an adaptive factor similar to the EnKF (e.g. Wang and Bishop, 2003; Anderson, 2007; Bauser et al., 2018) is desirable to increase the ensemble size and run the assimilation without using the parameter γ efficiency of the filter further and to achieve a better uncertainty representation of the ensemble.

20 6.2 Model error

Model errors are omnipresent in real systems. These errors can manifest themselves in many different ways. They can have a structural or stochastic nature, different intensities, and they can manifest e.g. they can be stochastic or cause a bias and they can vary in there intensity by biases in the results. For data assimilation of real measurements, the consideration of model errors is an important part for the success of the methods. Several extensions and modifications to sequential data assimilation methods
25 have been discussed (e.g. Li et al. (2009), Whitaker and Hamill (2012) and Houtekamer and Zhang (2016)) to compensate and improve the filter performance in presence of model errors.

In the course of this paper, we briefly study the behavior of the particle filter with covariance resampling for the case of a biased upper boundary condition. Two cases are considered, one with a 10% and one with a 20% bias towards less water. This means the amount of rain is reduced and the evaporation is increased by these percentages. The observations are still generated
30 using the previous boundary condition (Fig. 3). This means that the ensemble is propagated with a biased model, compared to the truth, for the complete assimilation run.

Except of the ensemble size and the upper boundary condition the setup is identical as in Sect. 4. To achieve converging results with $\gamma_\theta = 1.0$ and $\gamma_r = 1.2$, the ensemble size is increased to 600 and to 1200 ensemble members for the case of the 10% and the 20% bias, respectively. The ensemble size can be reduced to 300 and 600 by By increasing the tuning factor γ for

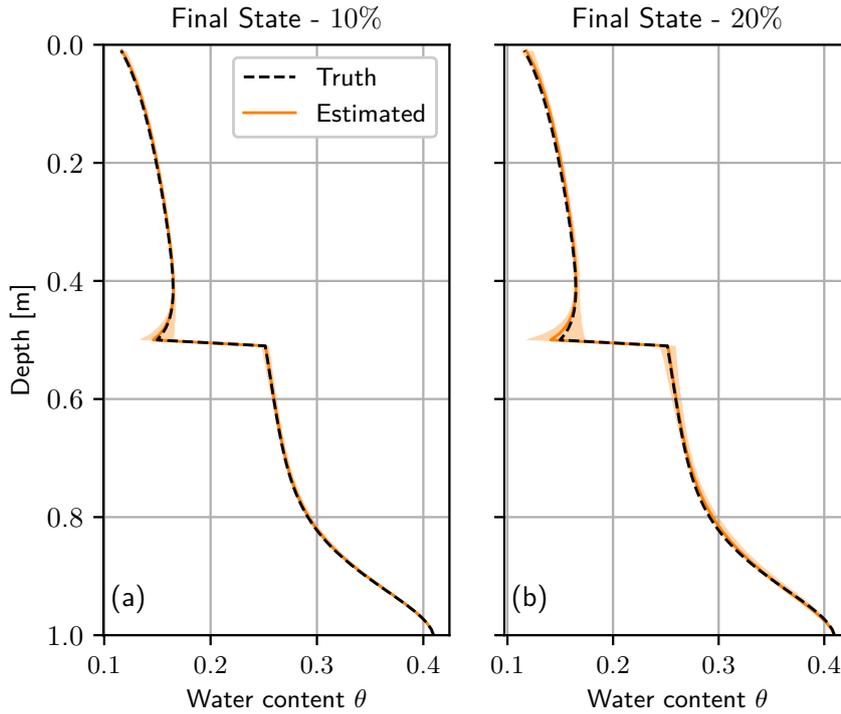


Figure 11. Final water content state after assimilation run using a bias for the upper boundary condition of (a): 10% and (b): 20%. The truth (dashed black) is almost congruent with the estimated mean (orange). For better visibility, only 100 The light orange areas represent the 95 %-quantile of the ensemble members of (a): 600 and (b): 1200 are shown ensemble members.

the state to a value slightly larger than one $\gamma_\theta = 1.1$ the necessary ensemble size can be reduced to 300 (10%) and 600 (20%). This artificially increases the uncertainty in state space, which helps the filter to compensate the bias during estimation. For better comparison with the presented case study in Sect. 4, the same values for the tuning parameter γ are used (see Eq.-) we show the results for the case with $\gamma_\theta = 1.0$ and $\gamma_\rho = 1.2$ in the following.

- 5 Figure 11 shows the final estimated state and the ensemble. The variance of the ensemble is larger compared to the case that uses the true boundary condition (see Fig. 7). The bias in the boundary condition leads to a larger uncertainty in state estimation, which increases with increasing bias (compare Fig. 11a and Fig. 11b). Although the difference to the mean slightly increaseincreases, the estimated mean still matches the truth well.

The conductivity function (see Fig. 12) shows a similar behaviour-behavior as the state. Compared to the case using the
 10 true boundary condition (see Fig. 6), the ensemble spread is larger, which increases with the bias in the boundary condition (compare Fig. 12a and Fig. 12b). The biased upper boundary condition leads to a bias in the conductivity function, which is not perfectly visible due to the logarithmic scale. The bias in the conductivity is larger for an larger error in the boundary condition. This behavior is also found for the conductivity function in the second layer.

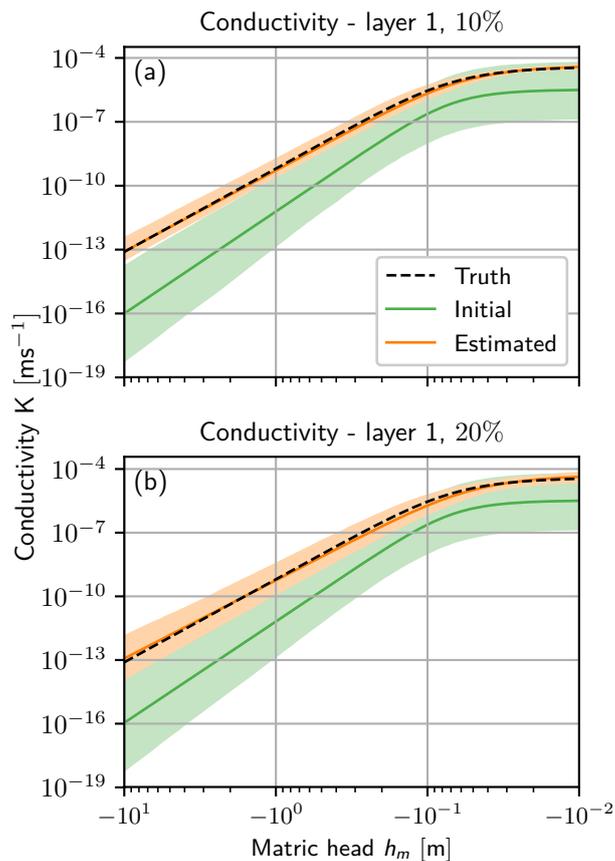


Figure 12. Conductivity function $K(h_m)$ (Eq. (19)) for a bias of (a): 10% and (b): 20%. The initial 95%-quantile of the ensemble (light green) with the mean (green) are shown. The truth (dashed black) is almost congruent with the estimated mean (orange), such that only the 95%-quantile of the final ensemble (light orange) is visible. For ~~better visibility, only 100 ensemble members of~~ (a): 600 and (b): 1200 ~~are shown for the prior and the estimated~~ ensemble members are used.

7 Summary and conclusions

We introduced a resampling method for particle filters that uses the covariance information of the ensemble to generate new particles and effectively avoids filter degeneracy. The method was tested in a synthetic one-dimensional unsaturated porous medium with two homogeneous layers. Even with just a rough initial guess, a broad parameter range and only 100 ensemble members, the estimation shows excellent results. After the assimilation, the results are verified in a free run with the final results.

The covariance connects information between observed and unobserved dimensions. This has some similarity to the ensemble Kalman filter but in our approach information of the non-Gaussian distribution is partly maintained in the retained

ensemble. Even though the RMSE of the water content includes the unobserved state dimensions, it stays in a narrow range (RMSE is about 10^{-3}) during the forecast. With every infiltration, the RMSE shows excursions caused by a wrong value of parameter α in the first layer that results in a wrong state near the layer boundary during the infiltration.

Transferring the information to the unobserved dimensions helps the filter to not degenerate when only a rough initial guess is available. The states and parameters are both altered actively. For the used initial condition, perturbing the parameters only (Moradkhani et al., 2005a), can lead to filter degeneracy because the state is only changed by the dynamics of the system. Compared to the particle filter with MCMC resampling (Moradkhani et al., 2012; Vrugt et al., 2013), the covariance resampling presented in this study has the advantage that it does not need additional model runs to generate new particles. However, the covariance resampling has to calculate the covariance matrix and perform a Cholesky decomposition every assimilation step. Similar to localization for the ensemble Kalman filter (Houtekamer and Mitchell, 2001; Hamill et al., 2001), it is possible to localize the covariance in the resampling to increase the efficiency.

The effective sample size (Eq. (22)) is a crucial parameter for this method. The covariance resampling needs a sufficient effective sample size, otherwise the calculation of the covariance matrix (Eq. (13)) becomes inaccurate and the filter may degenerate. In such a situation, the filter can be improved by resetting the weights to N^{-1} or increasing the ensemble size. In our example this was not necessary because the effective sample size was critical only for single assimilation steps.

The filter performance can be increased by a tuning parameter γ . The tuning parameter can significantly reduce the necessary ensemble size but has to be chosen carefully because otherwise the covariance can be overinflated. The mean is independent of the chosen γ , however, for $\gamma > 1$ the ensemble uncertainty is overestimated. In the presented case study, the tuning parameter reduced the necessary ensemble size by an order of magnitude. For cases with a model error, using the tuning parameter also for the state dimensions can be beneficial to stabilize the filter and reduce the necessary ensemble size further.

Different parameter sets can approximately describe the same conductivity function (Eq. (19)) in a certain matrix head regime. Model dynamics is necessary to differentiate between those sets. If the infiltration covers only a small regime, the conductivity function is only significant in the observed range and can differ from the truth otherwise. This is also reflected in the chosen boundary condition. Starting with infiltrations with low intensity but longer duration helps the filter to explore the water content range slowly and the observations can resolve the infiltration front.

The covariance resampling connects observed with unobserved dimensions to effectively estimate parameters and prevent filter degeneracy. It conserves the first two statistical moments in the limit of large numbers, while partly maintaining the structure of the non-Gaussian distribution in the retained ensemble. The method is able to estimate state and parameters in case of a difficult initial condition without additional model evaluations and using a rather small ensemble size.

30 Appendix A: Pseudocode

The following pseudocode describes the covariance resampling for a single time k , where the propagated ensemble and the calculated weights are given.

Algorithm 1 Pseudocode for the covariance resampling

Require: weights w_i^k at observation time t_k and the ensemble of N states \mathbf{u}_i^k

(a): compute weighted ensemble covariance \mathbf{Q} (Eq. (13))

(b): determine eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ of \mathbf{Q}

(c): if necessary ~~regularise~~ regularize \mathbf{Q} :

if $\min(\{\lambda_1, \lambda_2, \dots, \lambda_N\}) < 0$ **then**

$\mathbf{Q}_{\text{Reg.}} = \mathbf{Q} + \lambda_{\text{max}} \mathbf{I}$ // $\lambda_{\text{max}} \approx |\min(\{\lambda_1, \lambda_2, \dots, \lambda_N\})|$

end if

(d): optional: multiply \mathbf{Q} with the tuning parameter γ (see Eq. (14))

(e): universal resampling to determine number of child particles z (different method can be used for selection):

 draw random number x from uniform distribution $U(0, N^{-1})$

for $i = 0$ **to** $i < N$ **do**

$l = \sum_{m=0}^i w_m^k$

$z_i = 0$

while $x < l$ **do**

$z_i = z_i + 1$

$x = x + N^{-1}$

end while

end for

(f): generate new particles:

for $i = 0$ **to** $i < N$ **do**

if $z_i > 0$ **then**

 keep particle i

 assign $w = \frac{z_i}{N}$ to this particle

 generate $z_i - 1$ particles using \mathbf{u}_i and \mathbf{Q}

 assign $w = N^{-1}$ to these new particles

end if

end for

(g): ~~renormalise~~ renormalize weights $w_i = \frac{w_i}{c}$ // $c = \sum_{i=0}^N w_i$

Appendix B: Generation of correlated random numbers

B1 Cholesky ~~Decomposition~~decomposition

Correlated random numbers are generated using the Cholesky decomposition. We use the LDLT decomposition which is part of the *Eigen3* library (Guennebaud, Jacob et al., 2010). Decomposing the covariance matrix \mathbf{Q} leads to

$$5 \quad \mathbf{Q} = \mathbf{L}\mathbf{D}\mathbf{L}^\top, \tag{B1}$$

where \mathbf{D} is a diagonal matrix and \mathbf{L} is a lower unit triangular matrix. The LDLT form of the decomposition is related to the LLT-form by

$$\mathbf{Q} = \mathbf{L}'\mathbf{L}'^\top \quad \text{with} \quad \mathbf{L}' := \mathbf{L}\mathbf{D}^{\frac{1}{2}}. \tag{B2}$$

To draw a random vector \mathbf{y} from a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q})$ with mean $\boldsymbol{\mu}$, we first generate a normal distributed ($\mathcal{N}(\mathbf{0}, \mathbf{I})$) random vector \mathbf{x} . This vector is multiplied with \mathbf{L}' and the mean $\boldsymbol{\mu}$ is added:

$$\mathbf{y} = \mathbf{L}'\mathbf{x} + \boldsymbol{\mu} \tag{B3}$$

To verify that this gives the correct result the covariance matrix of \mathbf{y} is calculated:

$$\overline{(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^\top} = \overline{\mathbf{L}'\mathbf{x}(\mathbf{L}'\mathbf{x})^\top} = \mathbf{L}'\overline{\mathbf{x}\mathbf{x}^\top}\mathbf{L}'^\top = \mathbf{L}'\mathbf{I}\mathbf{L}'^\top = \mathbf{Q} \tag{B4}$$

yields \mathbf{Q} as required.

15 B2 ~~Regularisation~~Regularization of the ensemble covariance matrix

The calculation of the Cholesky decomposition (LDLT-version) is only possible if the matrix is not indefinite. Mathematically, a covariance matrix has to be positive semidefinite:

$$\mathbf{v}^\top \mathbf{Q} \mathbf{v} = \mathbf{v}^\top \sum (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top \mathbf{v} \tag{B5}$$

$$= \sum \mathbf{v}^\top (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top \mathbf{v} \tag{B6}$$

$$20 \quad = \sum (\mathbf{v}^\top (\mathbf{y}_i - \boldsymbol{\mu}))^2 \geq 0 \quad \text{with} \quad \mathbf{v} \in \mathbb{R}^d, \tag{B7}$$

but the covariance matrix calculated from our ensemble is occasionally indefinite. The reason for the covariance matrix being indefinite is a numerical error in the calculation of this matrix. In fact, the calculation of the eigenvalues λ results in negative values in the order of $\mathcal{O}(10^{-20})$.

For this purpose, the identity matrix \mathbf{I} , which is multiplied by a scalar λ_{\max} , is added to the covariance matrix. The value of λ_{\max} is in the order of magnitude of the largest negative eigenvalue of \mathbf{Q} . Thus, the ~~regularised~~regularized covariance matrix reads

$$\mathbf{Q}_{\text{Reg.}} = \mathbf{Q} + \lambda_{\max} \mathbf{I}. \tag{B8}$$

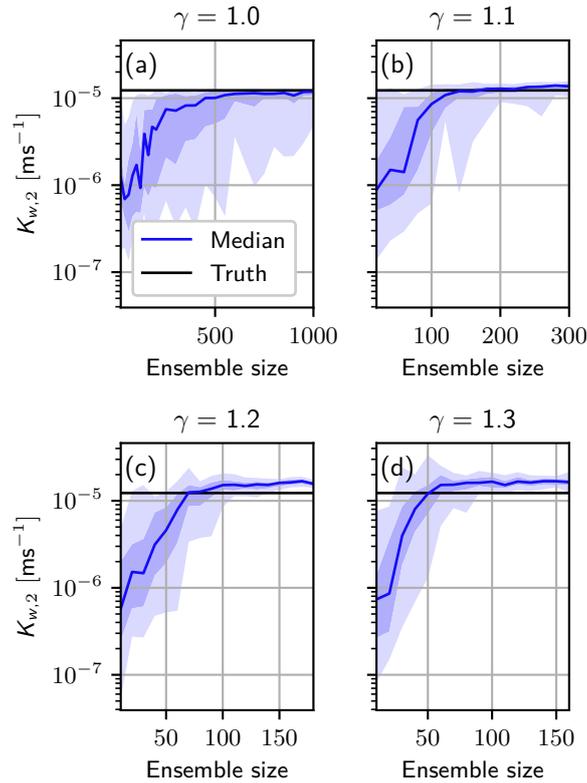


Figure C1. The mean saturated conductivity in the second layer after the data assimilation run for 40 different seeds and for varying factors of γ_p (Eq. (14)) (a): $\gamma_p = 1.0$, (b): $\gamma_p = 1.1$, (c): $\gamma_p = 1.2$ and (d): $\gamma_p = 1.3$. The blue areas represent the 70 %-quantile (darker blue) and the 90 %-quantile (light blue), respectively. Note the different scaling of the x-axes.

In our experiments, the smallest variance on the main diagonal of the covariance matrix is still 16 orders of magnitude larger than λ_{\max} such that the influence of this correction is negligible and does not change the results.

Appendix C: Dependence of $K_{w,2}$ on the tuning parameter γ

The saturated conductivity in the second layer is analyzed in the same setup as in Sect. 6.1. The assimilation is run for 40 different seeds, varying ensemble sizes and for four different tuning parameters γ (see Eq. (14)). The remaining setup of the system is identical as in Sect. 4.

Figure C1 shows the mean saturated conductivity in the second layer $K_{w,2}$ after the data assimilation run, including the 70 %-quantile (darker blue area) and the 90 %-quantile (light blue area) of the 40 runs with different seeds.

For all ensemble sizes, the filter either degenerates or finds the true value. Increasing the ensemble size increases the number of successful runs. The degeneration of the filter can directly be seen in the effective sample size, which drops to $N_{\text{eff}} = 1$.

Therefore, we emphasize the need to control whether the filter degenerates or not, to ensure a meaningful result. Results generated with a degenerated filter must not be used.

For the case $\gamma_p = 1.0$ (see Fig. C1a), which does not change the covariance matrix, the filter needs about 800 ensemble members to converge for 70 % of the seeds. It still degenerates for some seeds. Increasing the tuning factor for the parameter to $\gamma_p = 1.1$ (see Fig. C1b), reduces the necessary ensemble size and the seed dependency. For 300 ensemble members, the 90 %-quantile converges to the truth.

In Fig. C1c the tuning parameter is equal to the one used in the case study in Sect. 4. For less than 100 ensemble members, the behavior of the filter is seed dependent. While for some seeds the filter still converges for 20 ensemble members, it degenerates in most of them. For 100 ensemble members, the ensemble size used in the case study, the filter converges for every of the 40 seeds.

The apparent bias to a larger saturated conductivity for $\gamma_p = 1.2$ is compensated by the other two estimated parameters in this layer, such that the conductivity function Eq. (19) is almost identical to the truth in the measured water content range.

Increasing the factor to $\gamma_p = 1.3$ (see Fig. C1d), does not change the result significantly compared to the case $\gamma_p = 1.2$ (C1c). However, choosing a too large value for the tuning parameter results in an increasing uncertainty, which leads to a divergent ensemble for insensitive parameters like α_1 . Therefore, it is important to check the results and adjust the tuning parameter accordingly. It is always possible to increase the ensemble size and run the assimilation without using the parameter γ . The behavior of α_1 and the remaining parameters can be found in the Supplementary.

Competing interests. All authors declare that they have no competing interests.

Acknowledgements. This research is funded by Deutsche Forschungsgemeinschaft (DFG) through project RO 1080/12-1. Hannes H. Bauser and Daniel Berg were supported in part by the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences (HGS MathComp), funded by DFG grant GSC 220 in the German Universities Excellence Initiative.

References

- Abbaszadeh, P., Moradkhani, H., and Yan, H.: Enhancing hydrologic data assimilation by evolutionary particle filter and Markov chain Monte Carlo, *Advances in Water Resources*, 111, 192–204, <https://doi.org/10.1016/j.advwatres.2017.11.011>, 2018.
- Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus A*, 59, 210–224, <https://doi.org/10.1111/j.1600-0870.2006.00216.x>, 2007.
- Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Monthly Weather Review*, 127, 2741–2758, [https://doi.org/10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2), 1999.
- Bausier, H. H., Jaumann, S., Berg, D., and Roth, K.: EnKF with closed-eye period – towards a consistent aggregation of information in soil hydrology, *Hydrology and Earth System Sciences*, 20, 4999–5014, <https://doi.org/10.5194/hess-20-4999-2016>, 2016.
- 10 Bausier, H. H., Berg, D., Klein, O., and Roth, K.: Inflation method for ensemble Kalman filter in soil hydrology, *Hydrology and Earth System Sciences*, 22, 4921–4934, <https://doi.org/10.5194/hess-22-4921-2018>, 2018.
- Botto, A., Belluco, E., and Camporese, M.: Multi-source data assimilation for physically based hydrological modeling of an experimental hillslope, *Hydrology and Earth System Sciences*, 22, 4251–4266, <https://doi.org/10.5194/hess-22-4251-2018>, 2018.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, 126, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2), 1998.
- 15 Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, *Water Resources Research*, 24, 755–769, <https://doi.org/10.1029/WR024i005p00755>, 1988.
- Chen, Y. and Zhang, D.: Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Advances in Water Resources*, 29, 1107–1122, <https://doi.org/10.1016/j.advwatres.2005.09.007>, 2006.
- 20 DeChant, C. M. and Moradkhani, H.: Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, *Water Resources Research*, 48, <https://doi.org/10.1029/2011WR011011>, 2012.
- Doucet, A.: On sequential simulation-based methods for Bayesian filtering, Tech. rep., University of Cambridge, Dept. of Engineering, 1998.
- Erdal, D., Rahman, M., and Neuweiler, I.: The importance of state transformations when using the ensemble Kalman filter for unsaturated flow modeling: Dealing with strong nonlinearities, *Advances in Water Resources*, 86, 354–365, <https://doi.org/10.1016/j.advwatres.2015.09.008>, 2015.
- 25 Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, 99, 10 143–10 162, <https://doi.org/10.1029/94JC00572>, 1994.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757, <https://doi.org/10.1002/qj.49712555417>, 1999.
- 30 Gordon, N. J., Salmond, D. J., and Smith, A. F. M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F - Radar and Signal Processing*, 140, 107–113, <https://doi.org/10.1049/ip-f-2.1993.0015>, 1993.
- Guennebaud, G., Jacob, B., et al.: Eigen v3, <http://eigen.tuxfamily.org>, 2010.
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, 129, 2776–2790, [https://doi.org/10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2), 2001.
- 35 Harlim, J. and Majda, A. J.: Catastrophic filter divergence in filtering nonlinear dissipative systems, *Communications in Mathematical Sciences*, 8, 27–43, <https://projecteuclid.org/443/euclid.cms/1266935012>, 2010.

- Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, 129, 123–137, [https://doi.org/10.1175/1520-0493\(2001\)129<0123:ASEKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2), 2001.
- Houtekamer, P. L. and Zhang, F.: Review of the ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, 144, 4489–4532, <https://doi.org/10.1175/MWR-D-15-0440.1>, 2016.
- 5 Ippisch, O., Vogel, H.-J., and Bastian, P.: Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation, *Advances in Water Resources*, 29, 1780–1789, <https://doi.org/10.1016/j.advwatres.2005.12.011>, 2006.
- Jaumann, S. and Roth, K.: Effect of unrepresented model errors on estimated soil hydraulic material properties, *Hydrology and Earth System Sciences*, 21, 4301–4322, <https://doi.org/10.5194/hess-21-4301-2017>, 2017.
- Kitagawa, G.: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical*
 10 *Statistics*, 5, 1–25, <https://doi.org/10.1080/10618600.1996.10474692>, 1996.
- Li, C. and Ren, L.: Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter, *Vadose Zone Journal*, 10, 1205, <https://doi.org/10.2136/vzj2010.0159>, 2011.
- Li, H., Kalnay, E., Miyoshi, T., and Danforth, C. M.: Accounting for Model Errors in Ensemble Data Assimilation, *Monthly Weather Review*, 137, 3407–3419, <https://doi.org/10.1175/2009MWR2766.1>, 2009.
- 15 Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrology and Earth System Sciences*, 16, 3863–3887, <https://doi.org/10.5194/hess-16-3863-2012>, 2012.
- Man, J., Li, W., Zeng, L., and Wu, L.: Data assimilation for unsaturated flow models with restart adaptive probabilistic collocation based
 20 Kalman filter, *Advances in Water Resources*, 92, 258–270, <https://doi.org/10.1016/j.advwatres.2016.03.016>, 2016.
- Manoli, G., Rossi, M., Pasetto, D., Deiana, R., Ferraris, S., Cassiani, G., and Putti, M.: An iterative particle filter approach for coupled hydro-geophysical inversion of a controlled infiltration experiment, *Journal of Computational Physics*, 283, 37–51, <https://doi.org/10.1016/j.jcp.2014.11.035>, 2015.
- Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J. H., Canty, M., and Vereecken, H.: Hydraulic parameter es-
 25 timation by remotely-sensed top soil moisture observations with the particle filter, *Journal of Hydrology*, 399, 410–421, <https://doi.org/10.1016/j.jhydrol.2011.01.020>, 2011.
- Moradkhani, H., Hsu, K., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data
 assimilation using the particle filter, *Water Resources Research*, 41, <https://doi.org/10.1029/2004WR003604>, 2005a.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state–parameter estimation of hydrological models using ensemble
 30 Kalman filter, *Advances in Water Resources*, 28, 135–147, <https://doi.org/10.1016/j.advwatres.2004.09.002>, 2005b.
- Moradkhani, H., DeChant, C. M., and Sorooshian, S.: Evolution of ensemble data assimilation for uncertainty quantification using the particle
 filter-Markov chain Monte Carlo method, *Water Resources Research*, 48, <https://doi.org/10.1029/2012WR012144>, 2012.
- Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resources Research*, 12, 513–522, <https://doi.org/10.1029/WR012i003p00513>, 1976.
- 35 Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Monthly Weather Review*, 129, 1194–1207, [https://doi.org/10.1175/1520-0493\(2001\)129<1194:SMFSDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2), 2001.

- Plaza, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., and Pauwels, V. R. N.: The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter, *Hydrology and Earth System Sciences*, 16, 375–390, <https://doi.org/10.5194/hess-16-375-2012>, 2012.
- Qin, J., Liang, S., Yang, K., Kaihotsu, I., Liu, R., and Koike, T.: Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal, *Journal of Geophysical Research: Atmospheres*, 114, <https://doi.org/10.1029/2008JD011358>, d15103, 2009.
- Shi, L., Song, X., Tong, J., Zhu, Y., and Zhang, Q.: Impacts of different types of measurements on estimating unsaturated flow parameters, *Journal of Hydrology*, 524, 549–561, <https://doi.org/10.1016/j.jhydrol.2015.01.078>, 2015.
- Song, X., Shi, L., Ye, M., Yang, J., and Navon, I. M.: Numerical comparison of iterative ensemble Kalman filters for unsaturated flow inverse modeling, *Vadose Zone Journal*, 13, <https://doi.org/doi:10.2136/vzj2013.05.0083>, 2014.
- Van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil science society of America journal*, 44, 892–898, <https://doi.org/10.2136/sssaj1980.03615995004400050002x>, 1980.
- van Leeuwen, P. J.: Particle filtering in geophysical systems, *Monthly Weather Review*, 137, 4089–4114, <https://doi.org/10.1175/2009MWR2835.1>, <https://doi.org/10.1175/2009MWR2835.1>, 2009.
- 15 Vrugt, J. A., ter Braak, C. J., Diks, C. G., and Schoups, G.: Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications, *Advances in Water Resources*, 51, 457–478, <https://doi.org/10.1016/j.advwatres.2012.04.002>, 35th Year Anniversary Issue, 2013.
- Wang, X. and Bishop, C. H.: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes, *Journal of the Atmospheric Sciences*, 60, 1140–1158, [https://doi.org/10.1175/1520-0469\(2003\)060<1140:ACOBAE>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2), 2003.
- 20 Weerts, A. H. and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resources Research*, 42, <https://doi.org/10.1029/2005WR004093>, w09403, 2006.
- Whitaker, J. S. and Hamill, T. M.: Evaluating methods to account for system errors in ensemble data assimilation, *Monthly Weather Review*, 140, 3078–3089, <https://doi.org/10.1175/MWR-D-11-00276.1>, 2012.
- Wu, C.-C. and Margulis, S. A.: Feasibility of real-time soil state and flux characterization for wastewater reuse using an embedded sensor network data assimilation approach, *Journal of Hydrology*, 399, 313–325, <https://doi.org/10.1016/j.jhydrol.2011.01.011>, 2011.
- 25 Xiong, X., Navon, I. M., and Uzunoglu, B.: A note on the particle filter with posterior Gaussian resampling, *Tellus A*, 58, 456–460, <https://doi.org/10.1111/j.1600-0870.2006.00185.x>, 2006.
- Yan, H., DeChant, C. M., and Moradkhani, H.: Improving soil moisture profile prediction with the particle filter-Markov chain Monte Carlo method, *IEEE Transactions on Geoscience and Remote Sensing*, 53, 6134–6147, <https://doi.org/10.1109/TGRS.2015.2432067>, 2015.
- 30 Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C., and Jensen, K. H.: Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model, *Advances in Water Resources*, 86, 400–413, <https://doi.org/10.1016/j.advwatres.2015.07.018>, 2015.
- Zhang, H., Hendricks Franssen, H.-J., Han, X., Vrugt, J. A., and Vereecken, H.: State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter, *Hydrology and Earth System Sciences*, 21, 4927–4958, [https://doi.org/10.5194/hess-](https://doi.org/10.5194/hess-21-4927-2017)
- 35 21-4927-2017, 2017.