Summary: *Particle filters (PFs) have found widespread application and use for state and/or parameter estimation of dynamic system models. The premise of such filters is that they provide an exact approximation of the state forecast distribution. Yet, particle filters are not necessarily efficient as they may require a very large number of ensemble members (so-called particles) to approximate closely the evolving state distribution. This is particularly true in high-dimensional state spaces and complicates significantly the practical and/or real-time application of particle filters. What is more, particle filters are prone to sample impoverishment, that is, after a number of so-called assimilation steps, a very large number of the particles receives a negligible weight. These particles thus contribute little to the state forecast distribution and should be discarded/eliminated to a) refocus the thrust of the filter on the high-density region of the state space, and b) maintain an adequate filter efficiency and use of CPU resources. In the past decades, different resampling methods have been proposed and/or used to periodically rejuvenate the particle ensemble and ensure an adequate tracking of the evolving state distribution. Of these, Sequential Importance Resampling (SIR) has found most application and use. This method re-samples the particle ensemble using the computed weights of the N particles. These weights are simply equivalent to the product of the prior and the likelihood of the particle?s simulated trajectory. Whereas this SIR method is computationally efficient, it typically leads to a resampled ensemble with many copies of only a few of the "best" members (those with highest weights). In theory, this should not necessarily be a problem as the model operator (transition density) would disperse identical copies of the initial states as a result of the stochastic model error. This would work well in practice if the transition density of the state vector closely approximates the underlying system behavior. Unfortunately, even a modest deviation of the model operator from the actual system dynamics would deteriorate the particle ensemble to a point that most particles receive only a negligible weight. Thus, resampling is of crucial importance to periodically rejuvenate the particle ensemble and make sure that the simulated state PDF mimics closely the observed system behavior. Note that Ensemble Kalman filters do not suffer this same problem with sample impoverishment as they use a state analysis step to update the state forecasts of the N ensemble members each time an observation is becoming available.*

In this paper, the authors present a new resampling method to improve the efficiency and practical application of particle filters. This resampling method stores only a single copy of the M "best" particles determined with a standard resampling method, say SIR, and simulates the N – M "open spots" by drawing from a m-variate normal distribution with mean and covariance from the m-dimension.

Reply: We thank Jasper Vrugt for the detailed comments and suggestions, which will help to improve our manuscript. We also welcome a dialogue on this and related topics. In the following we provide our answers to Jaspar Vrugt's comments.

Specific comments

Comment: *1. The authors should consider a more realistic or appealing case study. Indeed, the present one-dimensional Richards' type flow problem with two horizontal layers is too simple to really demonstrate the advantages of the proposed covariance resampling methodology. The authors should consider a range of different state dimensionalities, m, to demonstrate that their method does not suffer from particle impoverishment. The authors should consider the Lorenz96 model with m = 40 state variables - this would demonstrate (or not) that the proposed resampling method works well in higher dimensional state spaces and would track closely the observed system dynamics. Such case study would make the paper much stronger and more appealing to those interested in methodological developments.*

Reply: Thank you for your comment. We propose the covariance resampling for state and parameter estimation in soil hydrology. This resampling aims to tackle the challenges of absent or unknown model errors for parameter estimation. Without perturbation particles will stay identical in the absence of a model error and the filter will degenerate. In the presented case study, we showed the effectiveness of the method for a soil hydrological problem. The Lorenz-96 model is an artificial atmospheric model and the behaviour differs significantly from soil hydrology using Richards' equation. Because of our focus on soil hydrology, the Lorenz-96 model is not used.

Just like other particle filters using only resampling, the proposed filter will also suffer from filter impoverishment in high-dimensional systems. The covariance resampling does not aim to lift the 'curse of dimensionality'.

Comment: *2. The authors refer to Figure 2 for a demonstration of the proposed covariance resampling method. I do not necessarily find this illustration to be particularly informative - that is - I think the authors can do a better job in detailing the proposed resampling method. The present animation assumes as if the target distribution is already well described with the present forecast distribution. In practice, this is often not true, certainly in higher dimensional state spaces. Also, I think the authors should differentiate between the state forecast density and the "true" or "unobserved" state fore-*

cast PDF. Then detail how the resampling works in practice. Personally, I always enjoy reading well-crafted algorithmic recipes (and associated coding) as those detail a step-by-step plan of how to implement the steps detailed in the main text.
Reply: Thank you for this suggestion. We agree that pseudo code is a more detailed recipe for implementing the actual algorithm. We think that a picture is more illustrative. Therefore, we will add the pseudo code in the appendix as an additional information for the reader.

Comment: *3. On Page 5 (top part) the authors list some previous approaches that have been used to resample the particle ensemble. I think the authors should mention whether each of these listed approaches leave the target state PDF invariant - that is - they lead to an exact approximation of the evolving target PDF. This may not necessarily be of concern to most hydrologists but is a requirement for methods to find widespread application and use. The same comment applies to the Introduction Section of the paper. In other words, I think it is good to emphasize that ad-hoc methods may provide results - but that such methods may not enjoy statistical underpinning.*
Reply: All of the mentioned methods approximate the posterior PDF. Of course some approximations are closer to the 'true' PDF than others. As long as all PDFs/errors are described correctly, we agree that e.g. the MCMC-resampling presented in Vrugt et al. (2013) is one of the methods that is closer to this target PDF. However, the model errors e.g. unrepresented physics, are typically unknown. In this case, the estimated target PDF is not the 'true' PDF anymore. Even without resampling and in the limit of infinite ensemble members the estimated PDF will be erroneous. Then it can be an advantage dropping the rigorous formalism to gain stability for the method. We will clarify this.

Comment: *4. I think the paper will be better if the authors replace Equation (1) with a recursive implementation of Bayes Law. This will make clear the relationship between the prior and posterior state PDF, how Equation (8) ties into this, and defines the importance weight, incremental importance weight and normalized importance weight. As it stands the current theory section omits completely the dimension of time - and this is key to state estimation.*
Reply: Thank you for this suggestion. We described the update at one time step. We agree that this is incomplete. We will extend the description, add a time index and use the recursive Bayes filter equation to clarify the update and analysis steps.

Comment: *5. I think it may be worthwhile to tie Equation (2) to the marginal likelihood. This is what you want to maximize with parameter esti-*

mation - but is of no real concern/importance for state estimation.
Reply: Equation (2) is the marginal likelihood of the observation and is a normalisation constant in Bayes' theorem. This normalisation constant is calculated using Equation (6) (see also next reply).

Comment: *6. Do not understand the need for Equation (6) - and also do not necessarily directly understand how the normalized weights lead to the normalization constant. This denominator, or evidence, does not require the importance weights to add up to unity, right?*
Reply: Thank you for the comment. We added this equation to clarify the calculation of the factor $P(d)$. Since this factor is a normalisation constant, which is the same for all weights, it is possible to calculate this factor from the normalisation of the weights. Therefore, to calculate this factor, it is necessary that the importance weights add up to unity. We think adding the recursive formulation of Bayes' theorem, like you suggested in comment 4, will clarify this. The equations will then change to

$$w_i^k = w_i^{k-1} \frac{P(\boldsymbol{d}^k | \boldsymbol{u}_i^k)}{P(\boldsymbol{d}^k)} \tag{1}$$

and

$$\sum_{i=0}^{N} w_i^k \stackrel{!}{=} 1 \quad \Rightarrow \quad P(\boldsymbol{d}^k) = \sum_{i=0}^{N} w_i^{k-1} P(\boldsymbol{d}^k | \boldsymbol{u}_i^k) \ , \tag{2}$$

where k is the discrete time index.

Comment: *7. The authors assign a weight of 1/N to the samples drawn from the m-variate normal distribution. I am not sure whether this leaves the state PDF invariant. The authors treat as if the samples from the multivariate normal have an equal weight - this is fine if ALL N samples were drawn from the multivariate normal PDF - but the state vectors drawn from this normal PDF are combined with the existing M "best" particles of the state forecast distribution - and those latter ones do not have a weight of 1/N. This cannot be justified theoretically. So, it is of crucial importance to demonstrate that the proposed resampling method leads to the exact target PDF.*
Reply: Thank you for the comment. This is correct, the new particles do not leave the PDF invariant. The retained particles do not have the weights of $\frac{1}{N}$. The weights are changed such that the resampling is similar to the universal resampling, except that the ensemble size is reduced. At this point the PDF is invariant in the limit of large ensemble sizes. To increase the ensemble size to N again, new particles are drawn from the multivariate

normal distribution. This alters the estimated posterior distribution. The mean and variance of the distribution are invariant under this process. For a large effective sample size, the overall structure remains close to the original posterior PDF. For small effective sample sizes, a large fraction of the particles is resampled such that the posterior distribution is dominated by the approximated multivariate Gaussian. We will extend the description of the method to clarify when the distribution is altered.

Comment: *8. The present resampling method relies heavily on the simulated state forecast distribution. If this distribution does not properly approximate the actual target PDF then resampling will provide N unique samples but those state vectors are not expected to produce a proper forecast PDF at the next time when a subsequent measurement becomes available. In other words, the present resampling method assumes that the transition density (model operator) approximates closely the true system dynamics. Once the state forecast PDF is systematically biased (likely to happen in real-world application) then the present resampling method may not necessarily enhance particle filtering results.*
Reply: Thank you for your input. You are right with the conclusion that a systematically bias will provide a challenge, if this bias is not represented in the model equation. However, in this case also an accurate representation of the PDF will result in a wrong forecast PDF for the next time step. In such a case an ensemble of unique samples has probably a better chance to not degenerate in the next time step because they can explorer the state space more widely.

Comment: *9. The authors use a perturbation factor, gamma, to inflate or deflate the covariance matrix of the normal resampling PDF. There is no justification for this - that is - its value is entirely subjective - indeed, one can tune gamma to provide appealing results, yet the value of gamma should guarantee an exact approximation of the target PDF (see Vrugt et al., 2013).*
Reply: It is correct that the multiplicative factor γ is a tuning parameter to inflate or deflate the covariance matrix. This factor can be used to increase the efficiency of the method. Such tuning parameters are common in different data assimilation methods. The factor can be set to $\gamma = 1.0$, which also gives a converging result, however, the necessary particles are about one order of magnitude larger in our example. The reason for the necessity of more particles is that the parameters do not have their own dynamics. For small effective sample sizes, the variance can become very small such that the new particles do not deviate significantly. During the forward propagation the particles are not separated because of the missing dynamics and absent model error, which leads to possible filter degeneration. Using a factor larger

than 1 stabilises the filter for small effective sample sizes. We will add a paragraph about the convergence of the filter for different seeds and factors γ with respect to the ensemble size.

Comment: *10. The authors use state augmentation to estimate jointly the model's state variables and parameter values. Do I conclude correctly that the authors use the normal resampling PDF to generate new parameter vectors? So, the resampling method assumes that the state variables and parameter values are multivariate normal. Would it not make sense to implement a mixture distribution instead - and estimate this distribution from the forecasted state/parameter distribution of the N particles?*
Reply: Your conclusion is correct, we approximate the posterior as a multivariate normal distribution for the generation of the new particles. A kernel density estimation would be possible but will neglect the correlation structure of the ensemble. A multivariate kernel density estimation would need too many particles to be practical. Therefore, using the multivariate normal distribution allows us to use the correlation information of the ensemble.

Comment: *11. The synthetic case study presented in the paper satisfies the assumption of a perfect model and thus transition density; in other words, the presented resampling method should work well as the forecast PDF (state/parameter) is not expected to deviate systematically from the observed data and system behavior. A real-world case study with actual measured data would provide a much stronger test of the proposed method. For example, one can use the Lorenz96 model to create an artificial data set - and then use an alternative model formulation to test, evaluate and benchmark the proposed covariance resampling method.*
Reply: We introduced this method as a possibility for state and parameter estimation using particle filters in soil hydrology. The paper is intended to present the concept of this method, which is shown in a synthetic example. Using real data has additional challenges, which would shift the focus of the paper away from the method towards these challenges.

References

Jasper A. Vrugt, Cajo J.F. ter Braak, Cees G.H. Diks, and Gerrit Schoups. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, 51(Supplement C):457 – 478, 2013. ISSN 0309-1708. doi:10.1016/j.advwatres.2012.04.002. URL `http://www.sciencedirect.`

com/science/article/pii/S0309170812000863. 35th Year Anniversary Issue.