# Interactive comment on “Application of ensemble transform data assimilation methods for parameter estimation in nonlinear problems" by Sangeetika Ruchi et al.

Sangeetika Ruchi[1], Svetlana Dubinkina[1], and Jana de Wiljes[2]

[1]Centrum Wiskunde & Informatica, P.O. Box 94079, 1098 XG Amsterdam, the Netherlands
[2]University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476, Potsdam, Germany

**Correspondence:** Jana de Wiljes (wiljes@uni-potsdam.de)

We would like to thank Femke Vossepoel and Marc Bocquet for carefully reading the manuscript and for their insightful comments and suggestions that definitely improved our article.

**Point-by-point answer to the comments by Femke Vossepoel**

Specific comments:

1. Terminology and description of example: As this paper could be of particular use to practitioners in the reservoir-engineering domain, I would encourage the authors to make the text more accessible to those. This could be done by changing or clarifying the use of certain terms and adding key references to explain the methods. For example, in reservoir engineering, the term Ensemble Kalman Filter is more commonly used than the term Ensemble Kalman Inversion; adding a number of key publications on this method and derived methods would help to set the scene and provide the reader with further background information. Also, those using data assimilation in practical applications will be interested in the actual values of the properties, and less likely to work on dimensionless problems. Relating the symbols to physical quantities would make this manuscript more accessible and relevant to them.

   **Response:** Thank you for raising this important point. It is crucial for us to reach practitioners and make the paper accessible to a wider audience and we have taken your suggestions into account. For instance we now use the term *Ensemble Kalman filter* instead of *Ensemble Kalman Inversion* as it is much more prevalent in the applied communities. Further we added a short discussion on the different terms and methods including randomised maximum likelihood, which is very popular with practitioners, and how they relate to each other.

2. Presentation of the methodologies: The mathematical rigour and expertise of the authors would allow them to not only compare the performance of the methods in an empirical sense, but also place them in the overall framework of data-assimilation methods for parameter estimation. The manner

in which the hybrid EKI-TETPF method is presented, is presented as a particle filter with several "fixes" (namely a) tempering, b) a Sinkhorn approximation, and c) an EKI proposal).

Can the authors think of a way to present the methods from a holistic viewpoint, making clear that these "fixes" are essential ingredients of the methods in order to perform a consistent and also effective parameter estimation? The abstract reads "Gaussian approximations [....] often produce astonishingly accurate estimations despite the inherently wrong underlying assumptions." Can you discuss more explicitly how the assumption of Gaussianity affects the outcome, perhaps by illustrating how non-Gaussian the distributions really are, or how the different methods deal with non- Gaussianity and/or non-linearity?

**Response:** The chosen presentation via particle filters (or Sequential Monte Carlo) allowed us to introduce all considered methods within one overarching family of filters. In order to make clear which techniques are standalone methods and which fixes they required to make the feasible in a challenging setting, we added the following text:

> In the following we will introduce a range of methods that can be employed to estimate solutions to the presented inverse problem under the overarching mantel of tempered Sequential Monte Carlo filters. Alongside these methods we will also proposed several important add-on tools required to achieve feasibility and higher accuracy in high-dimensional non-linear settings.

The sentence on Gaussian approximations has been adjusted (see our response to comment 3).

Technical corrections (language, minor items)

1. Please pay attention to the use of hyphens in compound modifiers. For example, the title could read "Application of ensemble-transform data-assimilation methods for parameter estimation in nonlinear problems". Other places where this would help: "high-dimensional problems", "entropy-inspired", "highly-correlated samples", "an easy-to-sample forml', etc.

   **Response:** Thank you this suggestion. We now use the hyphens high-dimensional, entropy-inspired, highly-correlated samples and easy-to-sample within the manuscript in order to increase readability.

   The original title was changed to "Fast hybrid tempered ensemble transform filter formulation for Bayesian elliptical problems via Sinkhorn approximation" and we preferred not to have the hyphen ensemble-transform.

2. The term "ensemble Kalman inversion" is used to a method that is known by many as "ensemble Kalman filtering". I suggest to clarify that EKI is used as equivalent to the ensemble Kalman filter. Page 2, line 38 and/or in the paragraph starting on p.8, line 201: suggest to add one of the key references for ensemble Kalman inversion or ensemble Kalman filtering, so readers can find out more about the method.

   **Response:** We have changed Ensemble Kalman Inversion to Ensemble Kalman filter everywhere in the manuscript. Furthermore, we now mention that the method is known under different names

in different communities: randomized maximum likelihood, multiple data assimilation, ensemble of data assimilation, ensemble Kalman inversion. The following text has been added to the manuscript:

> "As a side remark, EnKF was originally proposed for estimating a dynamical state of a chaotic system (e.g., Burgers et al., 1998). It was latter shown by Anderson (2001) that EnKF can be used for parameter estimation by introducing a trivial dynamics to the unknown static parameter. We note that EnKF is well known under different names in different scientific communities. In the reservoir community it is Ensemble Randomized Maximum Likelihood (Chen and Oliver, 2012), multiple data assimilation (Emerick and Reynolds, 2013), and Randomize-Then-Optimize (Bardsley et al., 2014). In the numerical weather prediction community, it falls under a large umbrella of Ensemble of Data Assimilation, see Carrassi et al. (2018) for a recent review. In the inverse problem community, it is ensemble Kalman inversion (Chada et al., 2018)."

3. Page 1, line 3: abstract: "inherently wrong": the Gaussian assumptions are not always wrong, so suggest to reformulate: "despite the simplifying assumptions" or something along these lines. Alternatively, demonstrate in the manuscript that these assumptions are actually wrong.

   **Response:** We have changed "inherently wrong" to "despite the simplifying assumptions".

4. Page 2, line 55: the number of required intermediate steps and the efficiency of ETPF still depends on it. What does *it* refer to?

   **Response:** Here *it* refers to the dependence on the initialisation. The corresponding text in the revised manuscript is "Although tempering restrains any sharp fail in the importance sampling step due to a poor initial ensemble selection, the number of required intermediate steps and the efficiency of ETPF still depends on the initialisation."

5. Page 5, line 118: Crank-Nicholson pcn-MCMC: explain what pcn means here.

   **Response:** pcn-MCMC means the preconditioned Crank-Nicolson MCMC.

6. Page 5 line 130: the scalar theta -> the scalar theta in Equation 5.

   **Response:** Thank you, we fixed it.

7. Page 6, line 152: where the minimum is compute -> where the minimum is. computed

   **Response:** Thank you, we fixed it.

8. Page 7 line 181 One the other hand -> on the other hand.

   **Response:** Thank you, we fixed it.

9. Page 8, line 204: estimation of posterior -> estimation of the posterior.

   **Response:** Thank you, we fixed it.

10. Page 8, line 205-215: make sure to list and clarify all symbols used.

**Response:** Please see the clarification (the same text is added to the revised manuscript):

The intermediate measures $\{\mu_t\}_{t=0}^T$ are approximated by Gaussian distributed variables with empirical mean $\boldsymbol{m}_t$ and empirical variance $\mathbf{C}_t$. Empirical mean $\boldsymbol{m}_{t-1}$ and empirical covariance $\mathbf{C}_{t-1}$ are defined in terms of $\{\boldsymbol{u}_{t-1,i}\}_{i=1}^M$ as following

$$\boldsymbol{m}_{t-1} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{u}_{t-1,i}, \qquad \mathbf{C}_{t-1} = \frac{1}{M-1}\sum_{i=1}^{M} (\boldsymbol{u}_{t-1,i} - \boldsymbol{m}_{t-1}) \otimes (\boldsymbol{u}_{t-1,i} - \boldsymbol{m}_{t-1}),$$

where $\otimes$ denotes Kroneker product. Then the mean and the covariance are updated as

$$\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \mathbf{C}_{t-1}^{\mathrm{u\,F}} (\mathbf{C}_{t-1}^{\mathrm{FF}} + \Delta_t \mathbf{R})^{-1} (\boldsymbol{y}_{\mathrm{obs}} - \overline{\boldsymbol{F}}_{t-1}) \qquad \text{and} \qquad \mathbf{C}_t = \mathbf{C}_{t-1} - \mathbf{C}_{t-1}^{\mathrm{u\,F}} (\mathbf{C}_{t-1}^{\mathrm{FF}} + \Delta_t \mathbf{R})^{-1} (\mathbf{C}_{t-1}^{\mathrm{uF}})',$$

respectively. Here $'$ denotes the transpose,

$$\mathbf{C}_{t-1}^{\mathrm{uF}} = \frac{1}{M-1}\sum_{i=1}^{M} (\boldsymbol{u}_{t-1,i} - \boldsymbol{m}_{t-1}) \otimes (F(\boldsymbol{u}_{t-1,i}) - \overline{\boldsymbol{F}}_{t-1}), \quad \mathbf{C}_{t-1}^{\mathrm{FF}} = \frac{1}{M-1}\sum_{i=1}^{M} [F(\boldsymbol{u}_{t-1,i}) - \overline{\boldsymbol{F}}_{t-1}] \otimes [F(\boldsymbol{u}_{t-1,i}) - \overline{\boldsymbol{F}}_{t-1}],$$

$$\overline{\boldsymbol{F}}_{t-1} = \frac{1}{M}\sum_{i=1}^{M} F(\boldsymbol{u}_{t-1,i}), \qquad \text{and} \qquad \Delta_t = \frac{1}{\phi_t - \phi_{t-1}}.$$

We recall that the nonlinear forward problem is $\boldsymbol{y} = F(\boldsymbol{u})$, the observation $\boldsymbol{y}_{\mathrm{obs}}$ has a Gaussian observation noise with zero mean and the covariance matrix $\mathbf{R}$, and $\phi_t$ is a temperature associated with the measure $\mu_t$.

11. Page 9, line 232: make clear how to choose beta.

**Response:** We agree that the choice of $\beta$ needs to be discussed. For our concrete numerical setting we added the following information in the numerical section:

"Note that $\beta \in [0, 1]$ and should be tuned according to underlying forward operator. "

We later also address the choice of $\beta$ in more detail in the conclusion (please see our response to comment 24).

12. Page 9, line 239: EKI as an more elaborate -> as a more elaborate.

**Response:** Thank you, we fixed it.

13. Page 9, line 240: Computational complexity: the estimates of computational complexity of the various methods is very useful. I suggest to include a table that illustrates the computational complexity of all methods/variations and include a few sentences on this in the 'Conclusions' part.

**Response:** Thank you for this suggestion. We added the following table (Table 1) to the appendix and now elaborate on the computational complexity in the conclusion.

14. Page 9, line 244: the example is dimensionless. Suggest to relate this to an example in which you list a number of typical values. You can then also mention that channels such as shown in Fig 1 can be found in fluvial rock formations that form aquifers or reservoirs.

| **Algorithm** | *Complexity* |
|---|---|
| TETPF | $\mathcal{O}[T(M\mathcal{C} + M^3 \log M + \tau_{\max} M\mathcal{C})]$ |
| TESPF | $\mathcal{O}[T(M\mathcal{C} + M^2 C(\alpha) + \tau_{\max} M\mathcal{C})]$ |
| EnKF | $\mathcal{O}[T(M\mathcal{C} + \kappa^2 n + \tau_{\max} M\mathcal{C})]$ |
| Hybrid EnKF-TETPF | $\mathcal{O}[T(M\mathcal{C} + \kappa^2 n + M\mathcal{C} + M^3 \log M + \tau_{\max} M\mathcal{C})]$ |
| Hybrid EnKF-TESPF | $\mathcal{O}[T(M\mathcal{C} + \kappa^2 n + M\mathcal{C} + M^2 C(\alpha) + \tau_{\max} M\mathcal{C})]$ |
| Forward model | $\mathcal{O}(M\mathcal{C})$ |
| pcn-MCMC mutation | $\mathcal{O}(\tau_{\max} M\mathcal{C})$ |
| FastEMD | $\mathcal{O}(M^3 \log M)$ |
| Sinkhorn approximation | $\mathcal{O}(M^2 C(\alpha))$ |

**Table 1.** The table provides an overview of the computational complexity of all the algorithms considered in the manuscript.

**Response:** We relate now to the paper by Zovi et al. (2017): "We note that a single-phase Darcy flow model, though not a steady-state, is widely used to model the flow in a subsurface aquifer and to infer uncertain permeability using data assimilation. For example, Zovi et al. (2017) used an EnKF to infer permeability of an existing aquifer located in North-East Italy. The area of this aquifer is 2.7 $km^2$ and exhibits several channels, such as the one depicted in Fig. 1. There a size of a computational cell was ranging from 2 m (near wells) to 20 m away from the wells."

15. Page 9, line 247: please make clear what physical variable (rate, pressure) the source term represents.

    **Response:** The source term $f$ accounts for groundwater recharge. This text is added to the revised manuscript.

16. Page 9 line 267: on an $N \times N$ grid: a potential user would like to know what is the scale, and spatial dimension. Suggest to give the value of $N$ earlier than you do now (on page 10, line 273).

    **Response:** The details of the numerical approximation are now given right after the continuous formulation.

17. Page 10, line 255: the choice of P for parameterisation is not very practical, as you are also using this letter for pressure.

    **Response:** It is changed to F for parameterisation.

18. Page 11, line 285: please make clear what property is being observed.

    **Response:** We observe the pressure at a few grid points. We have changed the text accordingly.

19. Page 12, line 310: we plot box plot -> we plot a box plot; using Sinkhorn approximation -> using a Sinkhorn approximation.

    **Response:** Thank you, we fixed it.

20. Page 12, line 313: TESPF outperforms: has a lower RMSE? Is smoother? How do you define a good performance?

**Response:** TESPF outperforms TETPF as the RMSE error is lower. The corresponding text is added to the revised manuscript.

21. Page 12 line 320: estimate well mot -> estimate well not.

**Response:** Thank you, we fixed it.

22. Page 13, figure 2: it is good to see the box plots for permeability, I would have liked to also see this for rate (observed state variable).

**Response:** We compute the pressure of the mean log permeability and plot a corresponding box plot for the RMSE in Figure 1.

First, we see that the smaller is the $\beta$, the smaller is the error. Next, we see that at the large ensemble size $M = 500$ the optimal transport resampling (shown in Figure 1(b)) outperforms the Sinkhorn approximation (shown in Figure 1(a)) in terms of smaller error. These two conclusions hold for the mean log permeability shown in Figure 2. The difference is that at the smaller ensemble size $M = 100$ the optimal transport resampling (shown in Figure 1(b)) outperforms the Sinkhorn approximation (shown in Figure 1(a)) in terms of smaller error for *the pressure*. However, for the *log permeability* it is the Sinkhorn approximation (shown in Figure 2(a)) that outperforms the optimal transport resampling (shown in Figure 2(b)) for $\beta \geq 0.6$ in terms of smaller error. We attribute this inconsistency to the cancellation of errors when computing the pressure of the mean log permeability.

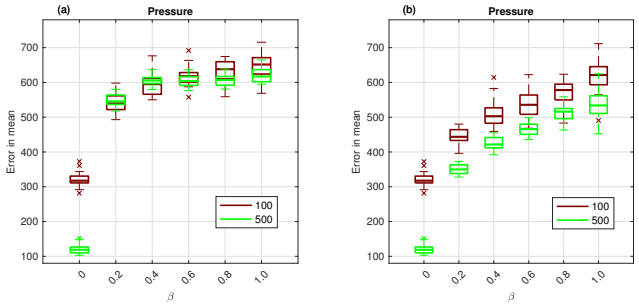


**Figure 1.** Application to F1 parameterization: using Sinkhorn approximation (a) and optimal transport resampling (b). Box plot over 20 independent simulations of RMSE of the pressure of the mean log permeability. X-axis is for the hybrid parameter, where $\beta = 0$ corresponds to EnKF and $\beta = 1$ to TET(S)PF. Ensemble size $M = 100$ is shown in red, and $M = 500$ in green. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers.

23. Page 14 figure 4: please label the x axis (it is described in the caption but would be good to see in the figure, too).

**Response:** Thank you, we fixed it.

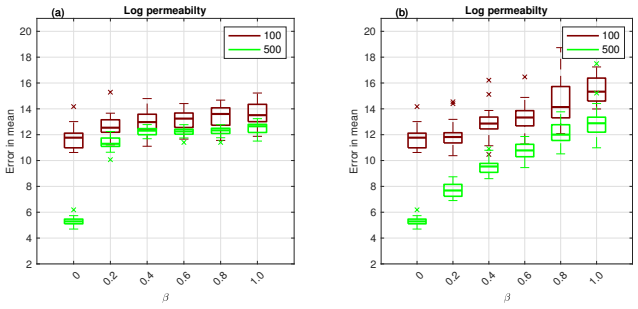


**Figure 2.** Application to F1 parameterization: using Sinkhorn approximation (a) and optimal transport resampling (b). Box plot over 20 independent simulations of RMSE of mean log permeability. X-axis is for the hybrid parameter, where $\beta = 0$ corresponds to EnKF and $\beta = 1$ to TET(S)PF. Ensemble size $M = 100$ is shown in red, and $M = 500$ in green. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers.

24. Page 14, line 332: these are very interesting results. I would value a discussion on how to find the best $\beta$ value in a realistic application of the hybrid method. In a synthetic case, this value can be determined, but how would you deal with this when assimilating real data? This discussion could be added in "conclusions: (page 17).

    **Response:** Thank you for raising this point, we added the following discussion to the conclusion:

    > Note that we have considered a synthetic case, where the truth is available, and thus chose $\beta$ in terms of accuracy of an estimate. However, in a realistic application the truth is not provided. In the context of state estimation with an underlying dynamical system it has been suggested to adaptively change the hybrid parameter with respect to the effective sample size. As the tempering scheme is already changed according to the effective sample size this ansatz would require to define the interplay between the two tuning variables. An ad-hoc choice for $\beta$ could be $0.2$ or $0.3$. This is motivated by the fact that the particle filter is too unstable in high dimensions and it is therefore sensible to use a tuning parameter prioritising the EnKF. The ad-hoc choice is supported by the numerical results in Section 3 and in Acevedo et al. (2017); de Wiljes et al. (2020) in the context of state-estimation.

25. Page 14 line 344: we plot box plot -> we plot a box plot (or "the box plot shows...").

    **Response:** Thank you, we fixed it.

26. Page 15, line 363: the application that you show, would be referred to as a "reservoir engineering" application, or a "hydrological" application, not as a "geophysical application". (In oil- and gas industry, reservoir engineering is about flow in porous media, while geophysics is about the use of seismic and other geophysical data and propagation of sound waves. In hydrology, permeability is usually replaced by hydraulic conductivity, so by using permeability your example would be more familiar to those working in reservoir engineering.)

**Response:** We are very grateful for you comment as we belief that these specifics are crucial to make our manuscript comprehensible for readers from the various community. Thank you also for taking the time to clarify the specifications of the fields. We changed geophysical application to reservoir engineering.

**Point-by-point answer to the comments by Marc Bocquet**

1. Page 1: I believe that the title of the paper is too generic, not specific enough. It could suit dozens of papers already published. I strongly suggest that you revise it. I understand that this is not easy since you use a large collection of methods. Although quick to amend, I believe this point is problematic for the visibility/identification of the paper.

   **Response:** We agree and have changed the title to "Fast hybrid tempered ensemble transform filter formulation for Bayesian elliptical problems via Sinkhorn approximation"

2. Page 1, line 2: "Kalman inversion" is not a widespread terminology, "randomized maximum likelihood" is better known, even beyond the reservoir community. See Oliver et al. (1996) and many references since then.

   **Response:** We have changed ensemble Kalman inversion to a better known ensemble Kalman filter. Furthermore, we added text about different name in different scientific communities. Namely:

   > "As a side remark, EnKF was originally proposed for estimating a dynamical state of a chaotic system (e.g., Burgers et al., 1998). It was latter shown by Anderson (2001) that EnKF can be used for parameter estimation by introducing a trivial dynamics to the unknown static parameter. We note that EnKF is well known under different names in different scientific communities. In the reservoir community it is Ensemble Randomized Maximum Likelihood (Chen and Oliver, 2012), multiple data assimilation (Emerick and Reynolds, 2013), and Randomize-Then-Optimize (Bardsley et al., 2014). In the numerical weather prediction community, it falls under a large umbrella of Ensemble of Data Assimilation, see Carrassi et al. (2018) for a recent review. In the inverse problem community, it is ensemble Kalman inversion (Chada et al., 2018)."

3. Page 1, line 4: "of the associated statistics.": I am not sure to get what you mean.

   **Response:** We mean that we can go beyond Gaussian approximations even with ensemble Kalman filter. We have adjusted the text accordingly:

   > Yet there is a lot of room for improvement specifically regarding a correct approximation of a non-Gaussian posterior distribution.

4. page 1, line 18, "a just approximation": do you mean a "correct approximation"?

   **Response:** Yes, we have adjusted the text correspondingly.

5. page 2, line 28, "The main drawback of MCMC is that this approach is not parallelizable.": You know that there are parallel (multiple tries) versions of MCMCs. It actually seems that you are yourself using multiple parallel MCMCs. So I believe you should mitigate that statement.

   **Response:** Indeed, we have omitted this statement. Instead we emphasise that MCMC samples are highly correlated. The following text has been added:

   > Typically, MCMC methods provide highly correlated samples. Therefore, in order to sample the posterior correctly MCMC requires a long chain, especially in the case of a multimodal or a peaked distribution. A peaked posterior is associated with very accurate observations.

6. Page 2, line 41-43: "However for nonlinear problems, Ernst et al. (2015); Evensen (2018) showed that in the large ensemble size limit an EKI approximation is not consistent with the Bayesian approximation.": To the best of my knowledge this is has been pointed out first by Oliver et al. (1996). The mathematical problem has also been clearly defined by Bardsley et al. (2014), and nicely named 'Randomize-Then-Optimize". There is also a recent discussion on the issue in Liu et al. (2017), p. 2894.

   **Response:** We have now included these references. Namely:

   > However for nonlinear problems, it has been shown by Oliver et al. (1996); Bardsley et al. (2014); Ernst et al. (2015); Liu et al. (2017) that an EnKF approximation is not consistent with the Bayesian approximation.

7. Page 2, line 57-58: Yes, but you should at this point mention here that the idea originates from the optimal transport community, and that it is by now widespread.

   **Response:** Thank you for pointing this out. The sentence is indeed misleading. Additionally we add a more extensive literature survey of hybrid filters as we felt we did not do it justice before. The following text address both issues and is now added to the revised version of the manuscript:

   > The lack of robustness in high dimensions can be addressed via a hybrid approach that combines a Gaussian approximation with a particle filter approximation (e.g., Santitissadeekorn and Jones, 2015). Different algorithms are created by Frei and Künsch (2013); Stordal et al. (2011), for example. In this paper, we adapt a hybrid approach of Chustagulprom et al. (2016) that uses EnKF as a proposal step for ETPF with a tuning parameter. Furthermore, it is well established that the computational complexity of solving an optimal transport problem can be significantly reduced via a Sinkhorn approximation by Cuturi (2013). This ansatz has been been implemented for the ETPF in Acevedo et al. (2017).

8. Page 3, line 73: even though obvious, it would be better to mention explicitly that $\mathcal{N}$ is the Gaussian distribution.

**Response:** Thank you, we fixed it.

9. Page 4, line 114: "Mutation" is applied mathematics Pierre Del Moral's terminology. You could briefly explain what it corresponds to in the geophysics particle filter community (rejuvenation?)

   **Response:** We have added the following text:

   > In the framework of particle filtering for dynamical systems, ensemble perturbation is achieved by rejuvenation, when ensemble members of the posterior measure are perturbed with a random noise sampled from a Gaussian distribution with zero mean and a covariance matrix of the prior measure. The covariance matrix of the ensemble is inflated and no acceptance step is performed due to the associated high computational costs for a dynamical system.
   >
   > Since we consider a static inverse problem, for ensemble perturbation we employ a Metropolis–Hastings method (thus we mutate samples) but with a proposal that speeds up an MCMC method for estimating a high-dimensional parameter.

10. Page 4, line 122: "we use random walk" $\rightarrow$ "we use the following random walk".

    **Response:** Thank you, we fixed it.

11. Page 5, line 135: "where C is computational cost of a forward model F" $\rightarrow$ "where C is the computational cost of the forward model F".

    **Response:** Thank you, we fixed it.

12. Page 6, line 136: 'is not effected"$\rightarrow$"is not affected".

    **Response:** Thank you, we fixed it.

13. Page 6, line 150: "we seak" $\rightarrow$ "we seek".

    **Response:** Thank you, we fixed it.

14. Page 6, line 160, Eq.(10): What is the definition of the norm of the random variables used in this equation?

    **Response:** Thank you for pointing this out. We have changed the equation to the following:

$$\omega^* = \arg\inf \left\{ \int_{\tilde{\mathcal{U}}\times\tilde{\mathcal{V}}} \|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|^2 d\omega(\boldsymbol{u},\tilde{\boldsymbol{u}}) : \quad \omega \in \prod(\mu,\nu) \right\}. \tag{1}$$

15. Page 7, line 181: "One the other hand"$\rightarrow$"On the other hand".

    **Response:** Thank you, we fixed it.

16. Page 7, line 193: "where Z is matrix with entires"$\rightarrow$ "where Z is the matrix with entries".

    **Response:** Thank you, we fixed it.

17. Page 7: It would be worth referring to the monograph by Peyré and Cuturi (Peyré and Cuturi, 2019) on optimal transport (in particular section 4), since it is very well done and freely available.

**Response:** Thank you for the suggestion. We now refer the reader to the monograph.

18. Page 8, line 8, "ensemble Kalman inversion (EKI) is one of the widely used algorithm." it has other (better known) names such as Randomized Maximum Likelihood (RML) and Randomize-Then-Optimize. Its sequential variant is known as the very well known EDA (Ensemble of Data Assimilation) in the numerical weather prediction/data assimilation community.

**Response:** We have addressed this in the revised manuscript. Please see our response to comment 2. above.

19. Page 8, line 218: "By implementing a sequential observation update of Whitaker et al. (2008),": what do you mean by this statement?

**Response:** We have omitted this statement, since it is irrelevant for a small number of observations. Instead, we state that

> The computational complexity of solving Eq.(13) is $\mathcal{O}(\kappa^2 n)$, where $n$ is the parameter space dimension, and $\kappa$ is the observation space dimension.

20. Page 8, line 224: "is remarkable robust" $\rightarrow$ "is remarkably robust".

**Response:** Thank you, we fixed it.

21. Page 9, Eqs.(14,15): I don't understand the intermediate member of both equations. The $\beta$ or $1-\beta$ should be powers of g, and not multiply g. Or is this a notation? What did I miss?

**Response:** Thank you for pointing out this typo. Indeed, $\beta$ or $1-\beta$ should be powers of $g$. It is now fixed in the revised manuscript.

22. Page 9, line 238-239: "This ansatz can also be understood as using the EKI as an more elaborate proposal density for the importance sampling step within SMC.": Using RML as a proposal density was already proposed and tested by Oliver et al. (1996).

**Response:** Thank you for pointing out this reference, which we now add in the revised manuscript: "This ansatz can also be understood as using the EnKF as a more elaborate proposal density for the importance sampling step within SMC (e.g., Oliver et al., 1996)."

23. Page 9, line 244-245: Are $(x, y)$ horizontal dimensions or is $y$ the depth? I believe it is worth explaining.

**Response:** $(x, y)$ are horizontal dimensions. We now add this in the revised manuscript.

24. Page 10, 258: "$\delta$ Dirac function" $\rightarrow$ "$\delta$ the Dirac function".

**Response:** Thank you, we fixed it.

25. Page 10, line 262: "We assume log permeability for" $\rightarrow$ "We assume that the log permeability for".

**Response:** Thank you, we fixed it.

26. Page 10, line 267: Ok, but which type of solver did you use? (multigrid, linear algebra solver, etc.)

**Response:** We use a linear algebra solver (backslash operator in MATLAB). The corresponding text in now added to the revised version.

27. Page 11, line 272: "The grid dimension is 70" $\rightarrow$ "The grid dimension is $N = 70$".

**Response:** Thank you, we fixed it.

28. Page 11, line 277: "The grid dimension is 50"$\rightarrow$ "The grid dimension is $N = 50$".

**Response:** Thank you, we fixed it.

29. Page 11, lines 293-295: "Such a small noise makes the data assimilation problem hard to solve, since the likelihood is very peaked and a non-iterative data assimilation approach fails.": the explanation is very unclear to me. Please clarify.

**Response:** With such a small noise the likelihood is a peaked distribution. Therefore a non-iterative data assimilation approach requires a computationally unfeasible number of ensemble members to sample the posterior. This text is now added to the revised manuscript.

30. Page 12, line 301: "An MCMC solution was obtained by combining 50 independent chains each of length 106": this contradicts to some extent the statement made about its serial nature in the introduction.

**Response:** We have omitted the statement about MCMC serial nature made earlier in the introduction.

31. Page 12, line 223: 9 observation seem too few, are they? Your experiments might rely too much on the prior. I guess for reservoir or hydrological applications, there are indeed just a few points, but they are many measurements over time at the same well.

**Response:** This is a fair criticism. Fever observations allow for a multi-modal posterior. More observations (either due to more wells or due to measurements at a few wells but over some time interval) decrease the uncertainty resulting in a uni-modal posterior.

32. Page 12, line 323: "distributed observations. which are displayed" $\rightarrow$ "distributed observations, which are displayed"

**Response:** Thank you, we fixed it.

33. Page 13, Figure 2: please add a label ($\beta$) to the x-axis.

**Response:** Thank you, we fixed it.

34. Page 13, Figure 2: At $\beta = 0$ there is quite a discrepancy between the $M = 100$ and the $M = 500$ experiments. This could show that EKI (alone) is not working very well here. Moreover, quite often, the whiskers for $M = 100$ and $M = 500$ have no overlap. We would expect some overlap, would we? Do you have an interpretation?

**Response:**

> A discrepancy between the $M = 100$ and the $M = 500$ experiments at $\beta = 0$ (thus EnKF alone) is related to the curse of dimensionality. The ensemble size $M = 100$ is too small to estimate an uncertain parameter of the dimension $10^3$ using 36 accurate observations. However, at the ensemble size $M = 500$ EnKF alone ($\beta = 0$) gives an excellent performance compared to any combination ($\beta > 0$).

We now add the above text to the revised manuscript.

Indeed, the overlap in the whiskers is to be expected. Therefore we have performed experiments with EnKF ($\beta = 0$) at the ensemble sizes $M = 200$ and $M = 400$. The box plot of the error is shown in Figure 3. We see that as ensemble size increases the whiskers get close to each other and we see an overlap between $M = 400$ and $M = 500$.

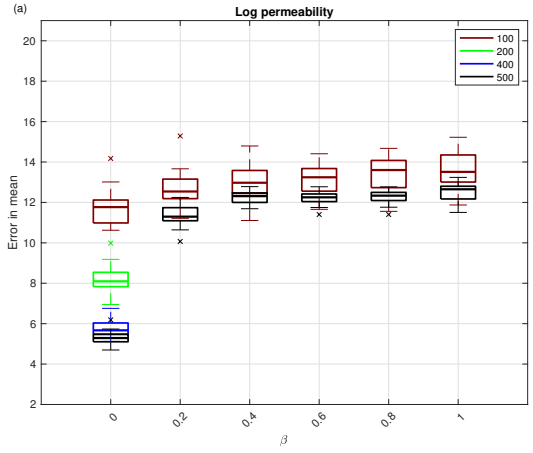


**Figure 3.** Application to F1 parameterization: using Sinkhorn approximation. Box plot over 20 independent simulations of RMSE of mean log permeability. X-axis is for the hybrid parameter, where $\beta = 0$ corresponds to EnKF and $\beta = 1$ to TET(S)PF. Ensemble size $M = 100$ is shown in red, $M = 200$ in green, $M = 400$ in blue, and $M = 500$ in black. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers.

35. page 13, Figure 3: By “Optimal” in the labelling of the panels, do you mean optimal transport, or something else?

**Response:** We see the confusion. Indeed, by "Optimal" in the labelling of the panels we mean optimal transport. We now change it to "OT".

36. page 14, Figure 4: Now, there is some consistent overlap between the $M = 100$ and $M = 500$ experiments, because, I guess, of the limited number of parameters (the curse of dimensionality is avoided in this case).

**Response:** Indeed, in this numerical experiment we have less observations (only 9). The P2 parameterization has 5 geometrical parameters and of order $10^3$ permeability parameters.

37. Page 14, line 333: "is lowest though" $\rightarrow$ "is lowest although".

**Response:** Thank you, we fixed it.

38. Page 15, lines 362-363: "This makes the proposed method a promising option for the high dimensional nonlinear problems one is typically faced with in geophysical applications.". Your problem do not have time dependence (does it?) which often makes many geophysical applications (like meteorology and ocean forecasting) very difficult. So you could mitigate that statement.

**Response:** We agree that we need to be more specific in this sentence and replaced "geophysical applications" with "reservoir engineering".

39. Page 16, Figure 6: What about the prior? How does it compare to the posterior?

**Response:** The prior is uniform, namely $d^1 \sim U[0.3,\ 2.1]$, $d^2 \sim U[\pi/2,\ 6\pi]$, $d^3 \sim U[-\pi/2,\ \pi/2]$, $d^4 \sim U[0,\ 6]$, $d^5 \sim U[0.12,\ 4.2]$. We see that the posterior is not a uniform distribution.

40. Page 17: "approach provides all the desirable properties required to obtain robust and highly accurate approximate solutions of nonlinear high dimensional Bayesian inference problems.": You cannot really make such a bold statement from one (however nice) example. Please mitigate your statement.

**Response:** We have changed this statement to the following:

> "This suggests a hybrid approach has a great potential to obtain robust and highly-accurate approximate solutions of nonlinear high-dimensional Bayesian inference problems."

41. General question which is worth discussing a bit: In practice, how fast is the Sinkhorn numerical solution compared to the exact optimal transport?

**Response:** This is a good question. Yet it is difficult to provide a general answer as the computational complexity of the Sinkhorn approximation is highly dependent on the choice of the regularisation parameter $\alpha$, i.e, specifically $\mathcal{O}(M^2 C(\alpha))$. It is know that $C(\alpha)$ grows with $\alpha$ as the original transport problem is approached yet the associated computational complexity can vary considerably. Therefore it is important to find an acceptable trade-off between a good approximation and the improvement in computational complexity. This fact was not discussed in the manuscript and we now added the following discussion to the conclusions:

> "Note however that $C(\alpha)$ depends on the chosen regularisation and grows with $\alpha$. Therefore, one needs to balance between reducing computational time and finding a reasonable approximate solution of the original transport problem when choosing a value for $\alpha$."

# ~~Application of~~ Fast hybrid tempered ensemble transform ~~data assimilation methods~~ filter formulation for ~~parameter estimation in nonlinear~~ Bayesian elliptical problems via Sinkhorn approximation

Sangeetika Ruchi[1], Svetlana Dubinkina[1], and Jana de Wiljes[2]

[1]Centrum Wiskunde & Informatica, P.O. Box 94079, 1098 XG Amsterdam, The Netherlands
[2]University of Potsdam, Karl-Liebknecht-Str. 24/25, D-14476, Potsdam, Germany

**Correspondence:** Jana de Wiljes (wiljes@uni-potsdam.de)

**Abstract.** Identification of unknown parameters on the basis of partial and noisy data is a challenging task in particular in ~~high dimensional~~ high-dimensional and nonlinear settings. Gaussian approximations to the problem, such as ensemble Kalman ~~inversion~~filtering, tend to be robust, computationally cheap and often produce astonishingly accurate estimations despite the ~~inherently wrong~~ simplifying underlying assumptions. Yet there is a lot of room for improvement specifically regarding ~~the description of the associated statistics~~a correct approximation of a non-Gaussian posterior distribution. The tempered ensemble transform particle filter is an adaptive sequential Monte Carlo method, where resampling is based on optimal transport mapping. Unlike ensemble Kalman ~~inversion~~ filtering it does not require any assumptions regarding the posterior distribution and hence has shown to provide promising results for ~~non-linear~~ nonlinear non-Gaussian inverse problems. However, the improved accuracy comes with the price of much higher computational complexity and the method is not as robust as the ensemble Kalman ~~inversion in high dimensional~~ filtering in high-dimensional problems. In this work, we add an ~~entropy inspired~~ entropy-inspired regularisation factor to the underlying optimal transport problem that allows to considerably reduce the high computational cost via Sinkhorn iterations. Further, the robustness of the method is increased via an ensemble Kalman ~~inversion~~ filtering proposal step before each update of the samples, which is also referred to as hybrid approach. The promising performance of the introduced method is numerically verified by testing it on a steady-state single-phase Darcy flow model with two different permeability configurations. The results are compared to the output of ensemble Kalman ~~inversion~~filtering, and Markov Chain Monte Carlo methods results are computed as a benchmark.

## 1 Introduction

If a solution of a considered partial differential equations (PDE) is ~~highly sensitive~~ highly-sensitive to its parameters, accurate estimation of the parameters and their uncertainties is essential to obtain a ~~just~~ correct approximation of the solution. Partial observations of the solution are then used to infer uncertain parameters by solving a PDE-constrained inverse problem. For instance one can approach such problems via methods induced by Bayes's formula (Stuart, 2010). More specifically the posterior probability density of the parameters given the data, is then computed on the basis of a prior probability density and a likelihood which is the conditional probability density associated with the given noisy observations. Well-posedness of an

inverse problem and convergence to the true posterior in the limit of observational noise going to zero was proven for different priors and under assumptions on the parameter-to-observation map by Dashti and Stuart (2017), for example.

When aiming at practical applications as in oil reservoir management (Lorentzen et al., 2020) and meteorology (Houtekamer and Zhang, 2016) for example, the posterior is approximated by means of a finite set of samples. Markov chain Monte Carlo (MCMC) methods approximate the posterior with a chain of samples–a sequential update of samples according to the posterior. ~~The main drawback of MCMC is that this approach is not parallelizable. Therefore~~Typically, MCMC methods provide highly-correlated samples. Therefore, in order to sample the posterior correctly MCMC requires a long chain, especially in the case of a multi-modal or a peaked distribution. A peaked posterior is associated with very accurate observations. Therefore, unless a speed up is introduced in a MCMC chain (e.g., Cotter et al., 2013), MCMC is impractical for computationally expensive PDE models.

Adaptive Sequential Monte Carlo (SMC) methods ~~, on the contrary, are parallelizable since they~~ are different approaches to approximate the posterior with an *ensemble* of samples by computing their probability (e.g, Vergé et al., 2015). Adaptive intermediate probability measures are introduced between the prior measure and the posterior measure to improve upon method divergence due to the curse of dimensionality following Del Moral et al. (2006); Neal (2001). Moreover, sampling from an invariant Markov kernel with the target intermediate measure and the reference prior measure improves upon ensemble diversity due to parameters stationarity as shown by Beskos et al. (2015). However, when parameter space is ~~high dimensional~~high-dimensional, adaptive SMC requires computationally prohibitive ensemble sizes unless we approximate only the first two moments (e.g., Iglesias et al., 2018) or we sample ~~highly correlated~~ highly-correlated samples (Ruchi et al., 2019).

Ensemble Kalman ~~inversion (EKI~~filtering (EnKF) approximates only the first two moments of the posterior, which makes it computationally attractive for estimating ~~high dimensional~~ high-dimensional parameters. For linear problems, Blömker et al. (2019) showed well-posedness and convergence of ~~EKI~~ EnKF for a fixed ensemble size and without any assumptions of Gaussianity. However for nonlinear problems, ~~Ernst et al. (2015); Evensen (2018) showed that in the large ensemble size limit an EKI~~ it has been shown by Oliver et al. (1996); Bardsley et al. (2014); Ernst et al. (2015); Liu et al. (2017) that an EnKF approximation is not consistent with the Bayesian approximation.

As a side remark, EnKF was originally proposed for estimating a dynamical state of a chaotic system (e.g., Burgers et al., 1998). It was latter shown by Anderson (2001) that EnKF can be used for parameter estimation by introducing a trivial dynamics to the unknown static parameter. We note that EnKF is well known under different names in different scientific communities. In the reservoir community it is Ensemble Randomized Maximum Likelihood (Chen and Oliver, 2012), multiple data assimilation (Emerick and R and Randomize-Then-Optimize (Bardsley et al., 2014). In the numerical weather prediction community, it falls under a large umbrella of Ensemble of Data Assimilation, see Carrassi et al. (2018) for a recent review. In the inverse problem community, it is ensemble Kalman inversion (Chada et al., 2018).

In order to sample ~~highly correlated~~ highly-correlated samples, one can employ optimal transport resampling that lies at the heart of the ensemble transform particle filter (ETPF) proposed by Reich (2013). An optimal transport map between two consecutive probability measures provides a direct sample-to-sample map with maximized sample correlation. Along the lines of an adaptive SMC approach a probability measure is described via the importance weights and the deterministic mapping re-

places the traditional resampling step. A so-called tempered ensemble transform particle filter (TETPF) was proposed by Ruchi et al. (2019). Note that this ansatz does not require any distributional assumption for the posterior and it was shown by Ruchi et al. (2019) that TETPF provides encouraging results for nonlinear ~~high dimensional~~ high-dimensional PDE-constrained inverse problems. However, the computational cost of solving an optimal transport problem in each iteration is considerably high.

In this work we address two issues arisen in the context of TETPF: (i) the immense computational costs of solving the associated optimal transport problem and (ii) the lack of robustness of the TETPF with respect to ~~high dimensional~~ high-dimensional problems. More specifically, the performance of ETPF has been found to be ~~highly dependent~~ highly-dependent on the initial guess. Although tempering restrains any sharp fail in the importance sampling step due to a poor initial ensemble selection, the number of required intermediate steps and the efficiency of ETPF still depends on ~~it. Chustagulprom et al. (2016) suggested that the~~ the initialisation. The lack of robustness in high dimensions can be addressed via a hybrid approach that combines a Gaussian approximation with ~~the ETPF~~ a particle filter approximation (e.g., Santitissadeekorn and Jones, 2015). Different algorithms are created by Frei and Künsch (2013); Stordal et al. (2011), for example. In this paper, we adapt a hybrid approach of Chustagulprom et al. (2016) that uses EnKF as a proposal step for ETPF with a tuning parameter. Furthermore, ~~Acevedo et al. (2017) sug~~ it is well established that the computational complexity of ~~the ETPF~~ solving an optimal transport problem can be significantly reduced via a Sinkhorn approximation ~~to the underlying transport problem.~~ by Cuturi (2013). This ansatz has been been implemented for the ETPF Acevedo et al. (2017).

Along the lines of Chustagulprom et al. (2016); de Wiljes et al. (2020), we propose a tempered ensemble transform particle filter with Sinkhorn approximation (TESPF) and a tempered hybrid approach.

The remainder of the manuscript is organised as follows: in Sect. 2, the inverse problem setting is presented. There we describe the tempered ensemble transform particle filter (TETPF) proposed by Ruchi et al. (2019). Furthermore, we introduce the tempered ensemble transform particle filter with Sinkhorn approximation (TESPF), a tempered hybrid approach that combines ~~EKI~~ EnKF and TETPF (hybrid ~~EKI-TETPF~~EnKF-TETPF), and a tempered hybrid approach that combines ~~EKI~~ EnKF and TESPF (hybrid ~~EKI-TESPF~~EnKF-TESPF). We discuss computational complexities of all the presented techniques and provide corresponding pseudocodes in Appendix A. In Sect. 3, we apply the adaptive SMC methods to an inverse problem of inferring ~~high dimensional~~ high-dimensional permeability parameters for a steady-state single-phase Darcy flow model. Permeability is parameterized following Ruchi et al. (2019), where one configuration of parametrization leads to Gaussian posteriors, while another one to non-Gaussian posteriors. Finally, we draw conclusions in Sect. 4.

## 2 Bayesian inverse problem

We assume $\boldsymbol{u} \in \tilde{\mathcal{U}} \subset \mathbb{R}^n$ is a random variable that is related to partially observable quantities $\boldsymbol{y} \in \mathcal{Y} \subset \mathbb{R}^\kappa$ by a nonlinear forward operator ~~$F : \tilde{\mathcal{U}} \to \mathcal{Y}$, namely~~ $G : \tilde{\mathcal{U}} \to \mathcal{Y}$, namely

$$\boldsymbol{y} = FG(\boldsymbol{u}).$$

Further $\boldsymbol{y}_{\text{obs}} \in \mathcal{Y}$ denotes a noisy observation of $\boldsymbol{y}$, i.e.,

$$\boldsymbol{y}_{\text{obs}} = \boldsymbol{y} + \boldsymbol{\eta}$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and $\mathcal{N}(\mathbf{0}, \mathbf{R})$ is a Gaussian distribution with zero mean and $\mathbf{R}$ covariance matrix. The aim is to determine or approximate the posterior measure $\mu(\boldsymbol{u})$ conditioned on observations $\boldsymbol{y}_{\text{obs}}$ and given a prior measure $\mu_0(\boldsymbol{u})$, which is referred to as Bayesian inverse problem. The posterior measure is absolutely continuous with respect to the prior, i.e.,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\boldsymbol{u}) \propto g(\boldsymbol{u}; \boldsymbol{y}_{\text{obs}}), \tag{1}$$

where $\propto$ is up to a constant of normalisation and $g$ is referred to as the likelihood and depends on the forward operator ~~$F$~~$G$. The Gaussian observation noise of the observation $\boldsymbol{y}_{\text{obs}}$ implies

$$g(\boldsymbol{u}; \boldsymbol{y}_{\text{obs}}) = \exp\left[-\frac{1}{2}(G(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})' \mathbf{R}^{-1} (G(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})\right], \tag{2}$$

where $'$ denotes the transpose. In the following we will introduce a range of methods that can be employed to estimate solutions to the presented inverse problem under the overarching mantel of tempered Sequential Monte Carlo filters. Alongside these methods we will also proposed several important add-on tools required to achieve feasibility and higher accuracy in high-dimensional nonlinear settings.

### 2.1 Tempered Sequential Monte Carlo

We consider sequential Monte Carlo (SMC) methods that approximate the posterior measure $\mu(\boldsymbol{u})$ via an empirical measure

$$\mu^M(\boldsymbol{u}) = \sum_{i=1}^{M} w_i \delta_{\boldsymbol{u}_i}(\boldsymbol{u}).$$

Here $\delta$ is the Dirac function, and the importance weights for the approximation of $\mu$ are

$$w_i = \frac{g(\boldsymbol{u}_i; \boldsymbol{y}_{\text{obs}})}{\sum_{j=1}^{M} g(\boldsymbol{u}_j; \boldsymbol{y}_{\text{obs}})}.$$

An ensemble $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M\} \subset \tilde{\mathcal{U}}$ consists of $M$ realizations $\boldsymbol{u}_i \in \mathbb{R}^n$ of a random variable $\boldsymbol{u}$ that are independent and identically distributed according to $\boldsymbol{u}_i \sim \mu_0$.

When an ~~easy to sample~~ easy-to-sample from prior $\mu_0$ does not approximate the complex posterior $\mu$ well, only a few weights $w_i$ have significant value resulting in a degenerative approximation of the posterior measure. Potential reasons for this effect are high dimensionality of the uncertain parameter, large number of observations, or accuracy of the observations. An existing solution to a degenerative approximation is an iterative approach based on tempering by Del Moral et al. (2006) or annealing by Neal (2001). The underlying idea is to introduce $T$ intermediate artificial measures $\{\mu_t\}_{t=0}^{T}$ between $\mu_0$ and $\mu_T = \mu$. These measures are bridged by introducing $T$ tempering parameters $\{\phi_t\}_{t=1}^{T}$ that satisfy $0 = \phi_0 < \phi_1 < \ldots < \phi_T = 1$. An intermediate measure $\mu_t$ is defined as a probability measure that has density proportional to $g(\boldsymbol{u})$ with respect to the previous measure $\mu_{t-1}$

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}\mu_{t-1}}(\boldsymbol{u}) \propto g(\boldsymbol{u}; \boldsymbol{y}_{\text{obs}})^{(\phi_t - \phi_{t-1})}.$$

Along the lines of Iglesias (2016) the tempering parameter $\phi_t$ is chosen such that effective ensemble size (ESS)

$$\mathrm{ESS}_t(\phi) = \frac{\left(\sum_{i=1}^M w_{t,i}\right)^2}{\sum_{i=1}^M w_{t,i}^2}$$

with

$$w_{t,i} = \frac{g(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}})^{(\phi_t - \phi_{t-1})}}{\sum_{j=1}^M g(\boldsymbol{u}_{t-1,j}; \boldsymbol{y}_{\text{obs}})^{(\phi_t - \phi_{t-1})}}, \tag{3}$$

does not drop below a certain threshold $1 < M_{\text{thresh}} < M$. Then an approximation of the posterior measure $\mu_t$ is

$$\mu_t^M(\boldsymbol{u}) = \sum_{i=1}^M w_{t,i} \delta_{\boldsymbol{u}_{t-1,i}}(\boldsymbol{u}). \tag{4}$$

A bisection algorithm on the interval $(\phi_{t-1}, 1]$ is employed to find $\phi$. If $\mathrm{ESS}_t > M_{\text{thresh}}$ we set $\phi_t = 1$ which implies that no further tempering is required.

The choice of ESS to define a tempering parameter is supported by results of Beskos et al. (2014) on stability of a tempered SMC method in terms of ESS. Moreover, for a Gaussian probability density approximated by importance sampling, Agapiou et al. (2017) showed that ESS is related to the second moment of the Radon-Nikodym derivative Eq. (1).

~~An~~ SMC method with importance sampling Eq. (4) does not change the sample $\{\boldsymbol{u}_{t-1,i}\}_{i=1}^M$, which leads to the method collapse due to a finite ensemble size. Therefore each tempering iteration $t$ needs to be supplied with resampling. Resampling provides a new ensemble $\{\tilde{\boldsymbol{u}}_{t,i}\}_{i=1}^M$ that approximates the measure $\mu_t$. We will discuss different resampling techniques in Sect. 2.3.

## 2.2 Mutation

Due to stationarity of the parameters ~~an SMC method requires~~ SMC methods require ensemble perturbation. ~~We~~ In the framework of particle filtering for dynamical systems, ensemble perturbation is achieved by rejuvenation, when ensemble members of the posterior measure are perturbed with a random noise sampled from a Gaussian distribution with zero mean and a covariance matrix of the prior measure. The covariance matrix of the ensemble is inflated and no acceptance step is performed due to the associated high computational costs for a dynamical system.

Since we consider a static inverse problem, for ensemble perturbation we employ a Metropolis–Hastings method (thus we mutate samples) but with a proposal that speeds up MCMC method for estimating a high-dimensional parameter. Namely, we use ensemble mutation of Cotter et al. (2013) with the target measure $\mu_t$ and the reference measure $\mu_0$. The mutation phase is initialized at $\boldsymbol{v}_{0,i} = \tilde{\boldsymbol{u}}_{t,i}$, and at the final inner iteration $\tau_{\max}$ we assign $\boldsymbol{u}_{t,i} = \boldsymbol{v}_{\tau_{\max},i}$ for $i = 1, \ldots, M$.

For a Gaussian prior we use the preconditioned Crank-Nicolson MCMC (pcn-MCMC) method

$$\boldsymbol{v}_i^{\text{prop}} = \sqrt{1-\theta^2}\boldsymbol{v}_{\tau,i} + (1 - \sqrt{1-\theta^2})\boldsymbol{m} + \theta\boldsymbol{\xi}_{\tau,i} \quad \text{for} \quad i = 1, \ldots, M. \tag{5}$$

Here $\boldsymbol{m}$ is the mean of the Gaussian prior measure $\mu_0$ and $\{\boldsymbol{\xi}_{\tau,i}\}_{i=1}^M$ are from a Gaussian distribution with zero mean and a covariance matrix of the Gaussian prior measure $\mu_0$.

For a uniform prior $U[a,b]$ we use the following random walk

$$\boldsymbol{v}_i^{\text{prop}} = \boldsymbol{v}_{\tau,i} + \boldsymbol{\xi}_{\tau,i} \quad i = 1,\ldots,M. \tag{6}$$

Here $\{\boldsymbol{\xi}_{\tau,i}\}_{i=1}^M \sim U[a-b, b-a]$ and $\{\boldsymbol{v}_i^{\text{prop}}\}_{i=1}^M$ are projected onto the $[a,b]$ interval if necessary. Then the ensemble at the inner iteration $\tau+1$ is

$$\boldsymbol{v}_{\tau+1,i} = \boldsymbol{v}_i^{\text{prop}} \quad \text{with the probability} \quad \rho(\boldsymbol{v}_i^{\text{prop}}, \boldsymbol{u}_{t-1,i}) \quad \text{for} \quad i = 1,\ldots,M; \tag{7}$$

$$\boldsymbol{v}_{\tau+1,i} = \boldsymbol{v}_{\tau,i} \quad \text{with the probability} \quad 1 - \rho(\boldsymbol{v}_i^{\text{prop}}, \boldsymbol{u}_{t-1,i}) \quad \text{for} \quad i = 1,\ldots,M. \tag{8}$$

Here $\boldsymbol{v}_i^{\text{prop}}$ is from Eq. (5) for the Gaussian measure and from Eq. (6) for the uniform measure, and

$$\rho(\boldsymbol{v}_i^{\text{prop}}, \boldsymbol{u}_{t-1,i}) = \min\left\{1, \frac{g(\boldsymbol{v}_i^{\text{prop}}; \boldsymbol{y}_{\text{obs}})^{\phi_t}}{g(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}})^{\phi_t}}\right\}.$$

The scalar $\theta \in (0,\ 1]$ in Eq. (5) controls the performance of the Markov chain. Small values of $\theta$ lead to high acceptance rates but poor mixing. Roberts and Rosenthal (2001) showed that for high-dimensional problems it is optimal to choose $\theta$ such that the acceptance rate is in between 20 % and 30 % by the last tempering iteration $T$. Cotter et al. (2013) proved that under some assumptions this mutation produces a Markov kernel with an invariant measure $\mu_t$.

*Computational complexity.* In each tempering iteration $t$ the computational complexity of the pcn-MCMC mutation is $\mathcal{O}(\tau_{\max} M \mathcal{C})$, where $\mathcal{C}$ is the computational cost of ~~a forward model $F$~~ the forward model $G$. For the pseudocode of the pcn-MCMC mutation please refer to the Algorithm 1 in Appendix A. Note that the computational complexity is not ~~effected~~ affected by the length of $\boldsymbol{u}$ which is a very desirable property in high dimensions as shown by Cotter et al. (2013) and Hairer et al. (2014).

### 2.3 Resampling phase

As we have already mentioned in Sect. 2.1, an adaptive SMC method with importance sampling needs to be supplied with resampling at each tempering iteration $t$. We consider a resampling method based on optimal transport mapping proposed by Reich (2013).

#### 2.3.1 Optimal transformation

The origin of the optimal transport theory lies in finding an optimal way of redistributing mass which was first formulated by Monge (1781). Given a distribution of matter, e.g., a pile of sand, the underlying question is how to reshape the matter into another form such that the work done is minimal. A century later the original problem was rewritten by Kantorovich (1942) in a statistical framework that allowed to tackle it. Due to these contributions it was later named the Monge-Kantorovich minimization problem. The reader is also referred to Peyré and Cuturi (2019) for a comprehensible overview.

Let us consider a scenario where the initial distribution of matter is represented by a probability measure $\mu$ on the measurable space $\tilde{\mathcal{U}}$, that has to be moved and rearranged according to a given new distribution $\nu$, defined on the measurable space $\tilde{\mathcal{V}}$. Then

we ~~seak~~ seek a probability measure that is a solution to

$$\text{mininf}\left\{\int_{\tilde{\mathcal{U}}\times\tilde{\mathcal{V}}} c(\boldsymbol{u},\tilde{\boldsymbol{u}})d\omega : \quad \omega\in\prod(\mu,\nu)\right\}, \tag{9}$$

where the minimum is ~~compute~~ computed over all joint probability measures $\omega$ on $\tilde{\mathcal{U}}\times\tilde{\mathcal{V}}$ with marginals $\mu$ and $\nu$, and $c(\boldsymbol{u},\tilde{\boldsymbol{u}})$ is a transport cost function on $\tilde{\mathcal{U}}\times\tilde{\mathcal{V}}$. The joint measures achieving the infinum are called optimal transport plans.

Let $\mu$ and $\nu$ be two measures on a measurable space $(\Omega,\mathcal{F})$ such that $\mu$ is the law of random variable $U:\Omega\to\tilde{\mathcal{U}}$ and $\nu$ is the law of random variable $V:\Omega\to\tilde{\mathcal{V}}$. Then a coupling of $(\mu,\nu)$ consists of a pair $(U,V)$. Note that couplings always exist, an example is the trivial coupling in which the random variables $U$ and $V$ are independent. A coupling is called deterministic if there exists a measurable function $\Psi_M:\tilde{\mathcal{U}}\to\tilde{\mathcal{V}}$ such that $V=\Psi_M(U)$ and $\Psi_M$ is called transport map. Unlike general couplings, deterministic couplings do not always exist. On the other hand there may be infinitely many deterministic couplings. One famous variant of Eq. (9), where the optimal coupling is known to be a deterministic coupling, is given by

$$\omega^* = \arg\inf_{\omega\in\Pi(\mu,\nu)}\sqrt{\mathbb{E}\|U-V\|^2}\left\{\int_{\tilde{\mathcal{U}}\times\tilde{\mathcal{V}}}\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|^2 d\omega(\boldsymbol{u},\tilde{\boldsymbol{u}}) : \quad \omega\in\prod(\mu,\nu)\right\}. \tag{10}$$

~~where $\Pi(\mu,\nu)$ denotes the set of measures joining $\mu$ and $\nu$ (see Villani (2008) for details).~~ The aim of the resampling step is to obtain equally probable samples. Therefore, in resampling based on optimal transport of Reich (2013), the Monge-Kantorovich minimization problem Eq. (10) is considered for the current posterior measure $\mu_t^M(\boldsymbol{u})$ given by its samples approximation Eq. (4) and a uniform measure (here the weights in the sample approximation are set to $1/M$). The discretized objective functional of the associate optimal transport problem is given by

$$J(\mathbf{S}) := \sum_{i,j=1}^{M} s_{ij}\|\boldsymbol{u}_{t-1,i}-\boldsymbol{u}_{t-1,j}\|^2$$

subject to $s_{ij}>0$ and constraints

$$\sum_{i=1}^{M} s_{ij} = \frac{1}{M}, \quad j=1,\dots M; \qquad \sum_{j=1}^{M} s_{ij} = w_{t,i}, \quad i=1,\dots M,$$

where matrix $\mathbf{S}$ describes a joint probability measure under the assumption that the state space is finite. Then samples $\{\tilde{\boldsymbol{u}}_{t,i}\}_{i=1}^M$ are obtained by a deterministic linear transform, i.e.,

$$\tilde{\boldsymbol{u}}_{t,j} := M\sum_{i=1}^{M}\boldsymbol{u}_{t-1,i}s_{ij} \quad \text{for} \quad j=1,\dots,M. \tag{11}$$

Reich (2013) showed weak convergence of the deterministic optimal transformation Eq. (11) to a solution of the Monge-Kantorovich problem Eq. (9) as $M\to\infty$.

*Computational complexity.* The computational complexity of solving the optimal transport problem with an efficient earth mover distance algorithm such as FastEMD of Pele and Werman (2009) is of order $\mathcal{O}(M^3 \log M)$. Consequently the computational complexity of the adaptive tempering SMC with optimal transport resampling (TETPF) is $\mathcal{O}[T(M\mathcal{C} + M^3 \log M + \tau_{\max} M\mathcal{C})]$, where $T$ is the number of tempering iterations, $\tau_{\max}$ is the number of pcn-MCMC inner iterations, and $\mathcal{C}$ is computational cost of a forward model ~~F~~G. For the pseudocode of the TETPF please refer to the Algorithm 4 in Appendix A.

### 2.3.2 Sinkhorn approximation

As discussed above solving the optimal transport problem has a computational complexity of $\mathcal{O} = M^3 \log(M)$ in every iteration of the tempering procedure. Thus the TETPF becomes very expensive for large $M$. ~~One~~ On the other hand an increase in the number of samples directly correlates with an improved accuracy of the estimation. In order to allow for as many samples as possible one needs to reduce the associate computational cost of the optimal transport problem. This can be achieved by replacing the optimal transport distance with a Sinkhorn distance and subsequently exploiting the new structure to elude the immense computational time of the EMD solver as shown by Cuturi (2013). More precisely the ansatz is built on the fact that the original transport problem has a natural entropic bound that is obtained for $\mathbf{S} = [\frac{1}{M} \boldsymbol{I}_M \boldsymbol{w}^\top]$ where $\boldsymbol{w} = [w_1, \ldots, w_M]$ and $\boldsymbol{I}_M = [1, \ldots, 1] \in \mathbb{R}^M$ which constitutes an independent joint probability. Therefore, one can consider the problem of finding a matrix $\mathbf{S} \in \mathbb{R}^{M \times M}$ ~~,~~ that is constraint by an additional lower entropic bound (Sinkhorn distance). This additional constraint can be incorporated via a Lagrange multiplier, which leads to the above regularised form, i.e.,

$$J_{\mathrm{SH}}(\mathbf{S}) = \sum_{i,j=1}^{M} \left\{ s_{ij} \|\boldsymbol{u}_{t-1,i} - \boldsymbol{u}_{t-1,j}\|^2 + \frac{1}{\alpha} s_{ij} \log s_{ij} \right\} \tag{12}$$

where $\alpha > 0$. Due to additional smoothness the minimum of Eq. (12) can be unique and has the form

$$\mathbf{S}^\alpha = \mathbf{diag}(\boldsymbol{b}) \exp\Big( -\alpha \mathbf{Z} \Big) \mathbf{diag}(\boldsymbol{a})$$

where $\mathbf{Z}$ is matrix with ~~entires~~ entries $z_{ij} = \|\boldsymbol{u}_{t-1,i} - \boldsymbol{u}_{t-1,j}\|^2$ and $\boldsymbol{b}$ and $\boldsymbol{a}$ non-negative vectors determined by employing Sinkhorn's fixpoint iteration described by Sinkhorn (1967). We will refer to this approach as tempered ensemble Sinkhorn particle filter (TESPF).

*Computational complexity.* Solving this regularise optimal transport problems rather than original transport problem given in Eq. (9) reduces the complexity to $\mathcal{O}(M^2 C(\alpha))$. Note however that $C(\alpha)$ depends on the chosen regularisation and grows with $\alpha$. Therefore, one needs to balance between reducing computational time and finding a reasonable approximate solution of the original transport problem when choosing a value for $\alpha$. For the pseudocode of the Sinkhorn adaptation of solving the optimal transport problem please refer to the Algorithm 3 in Appendix A. For the pseudocode of the TESPF please refer to the Algorithm 4 in Appendix A.

## 2.4 Ensemble Kalman ~~inversion~~Filter

For Bayesian inverse problems with Gaussian measures, ensemble Kalman ~~inversion (EKI~~filter (EnKF) is one of the widely used ~~algorithm. EKI~~ algorithms. EnKF is an adaptive SMC method that approximates the first two statistical moments of a

posterior distribution. For a linear forward model, ~~EKI~~ EnKF is optimal in a sense it minimizes the error in the mean (Blömker et al., 2019). For a nonlinear forward model, ~~EKI~~ EnKF still provides a good estimation of the posterior (e.g., Iglesias et al., 2018). Here we consider ~~EKI~~ EnKF method of Iglesias et al. (2018), since it is based on the tempering approach.

The intermediate measures $\{\mu_t\}_{t=0}^{T}$ are approximated by Gaussian distributed variables with empirical mean $\boldsymbol{m}_t$ and empirical variance $\mathbf{C}_t$. ~~Given empirical~~ Empirical mean $\boldsymbol{m}_{t-1}$ and empirical covariance $\mathbf{C}_{t-1}$ are defined in terms of $\{\boldsymbol{u}_{t-1,i}\}_{i=1}^{M}$ as following

$$\boldsymbol{m}_{t-1} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{u}_{t-1,i}, \qquad \mathbf{C}_{t-1} = \frac{1}{M-1}\sum_{i=1}^{M} (\boldsymbol{u}_{t-1,i} - \boldsymbol{m}_{t-1}) \otimes (\boldsymbol{u}_{t-1,i} - \boldsymbol{m}_{t-1}),$$

where $\otimes$ denotes Kroneker product. Then the mean and the covariance are updated ~~by~~ as

$$\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \mathbf{C}_{t-1}^{\text{u FuG}}(\mathbf{C}_{t-1}^{\text{FFGG}} + \Delta_t \mathbf{R})^{-1}(\boldsymbol{y}_{\text{obs}} - \overline{\boldsymbol{G}}_{t-1}), \quad \text{and} \quad \mathbf{C}_t = \mathbf{C}_{t-1} - \mathbf{C}_{t-1}^{\text{u FuG}}(\mathbf{C}_{t-1}^{\text{FFGG}} + \Delta_t \mathbf{R})^{-1}(\mathbf{C}_{t-1}^{\text{uF}\prime\text{uG}})',$$

~~where~~ respectively. Here $\prime$ denotes the transpose,

$$\mathbf{C}_{t-1}^{\text{uFuG}} = \frac{1}{M-1}\sum_{i=1}^{M}(\boldsymbol{u}_{t-1,i} - \boldsymbol{m}_{t-1}) \otimes (F G(\boldsymbol{u}_{t-1,i}) - \overline{\boldsymbol{G}}_{t-1}), \quad \mathbf{C}_{t-1}^{\text{FFGG}} = \frac{1}{M-1}\sum_{i=1}^{M} [F(\boldsymbol{u}_{t-1,i}) - \overline{\boldsymbol{F}}_{t-1}]^{M} [G(\boldsymbol{u}_{t-1,i}) - \overline{\boldsymbol{G}}_{t-1}] \otimes$$

$$\Delta_t = \frac{1}{\phi_t - \phi_{t-1}}, \overline{\boldsymbol{G}}_{t-1} = \frac{1}{M}\sum_{i=1}^{M} F G(\boldsymbol{u}_{t-1,i}), \quad \text{and} \quad \Delta_t = \frac{1}{\phi_t - \phi_{t-1}}.$$

We recall that the nonlinear forward problem is $\boldsymbol{y} = G(\boldsymbol{u})$, the observation $\boldsymbol{y}_{\text{obs}}$ has a Gaussian observation noise with zero mean and the covariance matrix $\mathbf{R}$, and $\phi_t$ is a temperature associated with the measure $\mu_t$.

Since we are interested in an ensemble approximation of the posterior distribution, we update the ensemble members by

$$\tilde{\boldsymbol{u}}_{t,i} = \boldsymbol{u}_{t-1,i} + \mathbf{C}_{t-1}^{\text{uFuG}}(\mathbf{C}_{t-1}^{\text{FFGG}} + \Delta_t \mathbf{R})^{-1}[\boldsymbol{y}_{t,i} - F(\boldsymbol{u}_{t-1,i})]^{-1}[\boldsymbol{y}_{t,i} - G(\boldsymbol{u}_{t-1,i})] \quad \text{for} \quad i = 1, \ldots, M. \tag{13}$$

Here $\boldsymbol{y}_{t,i} = \boldsymbol{y}_{\text{obs}} + \boldsymbol{\eta}_{t,i}$ and $\boldsymbol{\eta}_{t,i} \sim \mathcal{N}(\mathbf{0}, \Delta_t \mathbf{R})$ for $i = 1, \ldots, M$.

*Computational complexity.* ~~By implementing a sequential observation update of Whitaker et al. (2008), the~~ The computational complexity of solving Eq. (13) ~~can be reduced to $\mathcal{O}(2M\kappa n)$~~is $\mathcal{O}(\kappa^2 n)$, where $n$ is the parameter space dimension, and $\kappa$ is the observation space dimension~~, and $M$ is the ensemble size~~. Then the computational complexity of ~~EKI is $\mathcal{O}[T(M\mathcal{C} + 2M\kappa n + \tau_{\max} M\mathcal{C})]$~~EnKF is $\mathcal{O}[T(M\mathcal{C} + \kappa^2 n + \tau_{\max} M\mathcal{C})]$, where $T$ is the number of tempering iterations, $\tau_{\max}$ is the number of pcn-MCMC inner iterations, and $\mathcal{C}$ is computational cost of a forward model ~~$FG$~~. For the pseudocode of the ~~EKI~~ EnKF method please refer to the Algorithm 5 in Appendix A.

### 2.5 Hybrid

Despite the underlying Gaussian assumption the ~~EKI is remarkable robust in non-linear~~ EnKF is remarkably robust in nonlinear high-dimensional settings opposed to consistent SMC methods such as the TET(S)PF. For many ~~non-linear~~ nonlinear problems

it is desirable to have better uncertainty estimates while maintaining a level of robustness. This can be achieved by factorising the likelihood given by Eq. (2), e.g,

$$g(\boldsymbol{u};\boldsymbol{y}_{\text{obs}}) = g_1(\boldsymbol{u};\boldsymbol{y}_{\text{obs}}) \cdot g_2(\boldsymbol{u};\boldsymbol{y}_{\text{obs}}),$$

where

$$g_1(\boldsymbol{u};\boldsymbol{y}_{\text{obs}}) = \beta g(\boldsymbol{u};\boldsymbol{y}_{\text{obs}})^{\beta} = \exp\left[-\frac{1}{2}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})'(\beta\mathbf{R})^{-1}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})\right] \tag{14}$$

and

$$g_2(\boldsymbol{u};\boldsymbol{y}_{\text{obs}}) = (1-\beta) g(\boldsymbol{u};\boldsymbol{y}_{\text{obs}})^{(1-\beta)} = \exp\left[-\frac{1}{2}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})'[(1-\beta)\mathbf{R}]^{-1}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})\right]. \tag{15}$$

Then it is possible to alternate between methods with complementing properties such as the ~~EKI~~ EnKF and the TET(S)PF updates e.g., likelihood

$$\exp\left[-\frac{\beta}{2}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})'\mathbf{R}^{-1}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})\right]^{(\phi_t - \phi_{t-1})}$$

is used for an ~~EKI~~ EnKF update followed by an update with a TET(S)PF on the basis of

$$\exp\left[-\frac{(1-\beta)}{2}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})'\mathbf{R}^{-1}(FG(\boldsymbol{u}) - \boldsymbol{y}_{\text{obs}})\right]^{(\phi_t - \phi_{t-1})}.$$

Note that $\beta \in [0,1]$ and should be tuned according to underlying forward operator. This combination of an approximative Gaussian method and a consistent SMC method has been referred to as hybrid filters in the data assimilation literature[1] ~~(Chustagulprom et al., 201~~ This ansatz can also be understood as using the ~~EKI as an~~ EnKF as a more elaborate proposal density for the importance sampling step within SMC (e.g., Oliver et al., 1996).

*Computational complexity.* The computational complexity of combining the two algorithms is ~~$\mathcal{O}[T(M\mathcal{C} + 2M\kappa n + M\mathcal{C} + M^3 \log M +$~~ $\mathcal{O}[T(M\mathcal{C} + \kappa^2 n + M\mathcal{C} + M^3 \log M + \tau_{\max} M\mathcal{C})]$ for the hybrid ~~EKI-TETPF and $\mathcal{O}[T(M\mathcal{C} + 2M\kappa n + M\mathcal{C} + M^2 C(\alpha) + \tau_{\max} M\mathcal{C})]$~~ EnKF-TETPF and $\mathcal{O}[T(M\mathcal{C} + \kappa^2 n + M\mathcal{C} + M^2 C(\alpha) + \tau_{\max} M\mathcal{C})]$ for the hybrid ~~EKI-TESPF~~EnKF-TESPF. For the pseudocode of the hybrid methods please refer to the Algorithm 6 in Appendix A.

## 3 Numerical experiments

We consider a steady-state single-phase Darcy flow model defined over an aquifer of two-dimensional physical domain $D = [0,6] \times [0,6]$, which is given by

$$-\nabla \cdot [k(x,y)\nabla P(x,y)] = f(x,y), \quad (x,y) \in D, \tag{16}$$

[1] Note that the terminology is also used in the context of data assimilation filters combining variational and sequential approaches.

where $\nabla = (\partial/\partial x\ \partial/\partial y)'$, $\cdot$ the dot product, $P(x,y)$ the pressure, $k(x,y)$ the permeability, ~~and~~ $f(x,y)$ the source term ~~. The source term is~~

$$f(x,y) = \begin{cases} 0 & \text{if} \quad 0 < y \leq 4, \\ 137 & \text{if} \quad 4 < y < 5, \\ 274 & \text{if} \quad 5 < y \leq 6. \end{cases}$$

~~The~~ which accounts for groundwater recharge, and $(x,y)$ are horizontal dimensions. The boundary conditions are

$$P(x,0) = 100, \quad \frac{\partial P}{\partial x}(6,y) = 0, \quad -k(0,y)\frac{\partial P}{\partial x}(0,y) = 500, \quad \frac{\partial P}{\partial y}(x,6) = 0, \tag{17}$$

where $\partial D$ is the boundary of domain $D$. The source term is

$$f(x,y) = \begin{cases} 0 & \text{if} \quad 0 < y \leq 4, \\ 137 & \text{if} \quad 4 < y < 5, \\ 274 & \text{if} \quad 5 < y \leq 6. \end{cases}$$

We implement a cell-centered finite-difference method and a linear algebra solver (backslash operator in MATLAB) to solve the forward model Eqs. (16)–(17) on an $N \times N$ grid.

We note that a single-phase Darcy flow model, though not a steady-state, is widely used to model the flow in a subsurface aquifer and to infer uncertain permeability using data assimilation. For example, Zovi et al. (2017) used an EnKF to infer permeability of an existing aquifer located in North-East Italy. The area of this aquifer is 2.7 km$^2$ and exhibits several channels, such as the one depicted in Fig. 1. There a size of a computational cell was ranging from 2 m (near wells) to 20 m away from the wells.

### 3.1 Parameterisation of permeability

We consider the following two parameterisations of the permeability function $k(x,y)$

~~P1~~F1: log permeability over the entire domain $D$, $u(x,y) = \log k(x,y)$;

~~P2~~F2: permeability over domain $D$ that has a channel, $k(x,y) = k^1(x,y)\delta_{D_c}(x,y) + k^2(x,y)\delta_{D\setminus D_c}(x,y)$ as by Iglesias et al. (2014).

Here $D_c$ denotes a channel, $\delta$ the Dirac function, $k^1 = \exp(u^1(x,y))$ and $k^2 = \exp(u^2(x,y))$ denote permeabilities inside and outside the channel. The geometry of the channel is parameterized by five parameters $\{d^i\}_{i=1}^5$: amplitude, frequency, angle, initial point, and width, correspondingly. The lower boundary of the channel is given by $y = d^1\sin(d^2x/6) + \tan(d^3)x + d^4$. The upper boundary of the channel is given by $y + d^5$. These parameters are depicted in Fig. 1.

We assume that the log permeability for both ~~P1 and P2~~ F1 and F2 is drawn from a Gaussian distribution $\mu_0 = \mathcal{N}(m,C)$ with mean $m$ and covariance $C$. We define $C$ via a correlation function given by the Wittle-Matern correlation function defined

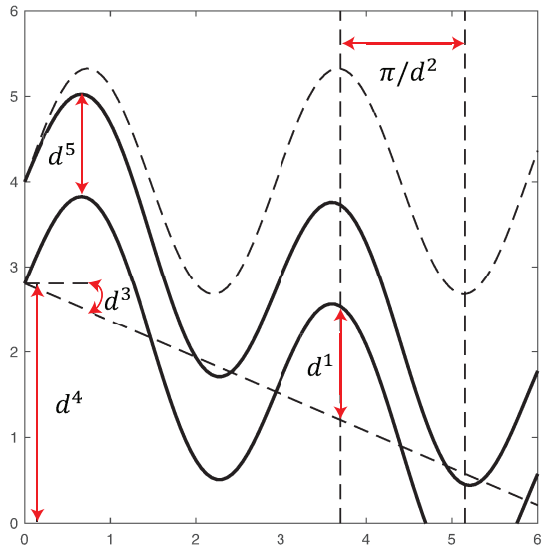


**Figure 1.** Geometrical configuration of channel flow: amplitude $d^1$, frequency $d^2$, angle $d^3$, initial point $d^4$, and width $d^5$.

by Matérn (1986)

$$c(x,y) = \frac{1}{\gamma(1)} \frac{\|x-y\|}{\upsilon} \Upsilon_1 \left( \frac{\|x-y\|}{\upsilon} \right),$$

where $\gamma$ is the gamma function, $\upsilon = 0.5$ is the characteristic length scale, and $\Upsilon_1$ is the modified Bessel function of the second kind of order 1.

We ~~implement a cell-centered finite difference method to solve the forward model Eqs. –on an $N \times N$ grid. We~~ denote by $\boldsymbol{\lambda}$ and $\mathbf{V}$ eigenvalues and eigenfunctions of the corresponding covariance matrix $\mathbf{C}$, respectively. Then, following a Karhunen-Loeve expansion, log permeability is

$$\log(k^l) = \log(m) + \sum_{\ell=1}^{N^2} \sqrt{\lambda^\ell} V^{\ell l} u^\ell \quad \text{for} \quad l = 1, \ldots, N^2,$$

where $u^\ell$ is i.i.d. from $\mathcal{N}(0,1)$ for $\ell = 1, \ldots, N^2$.

For ~~P1~~F1, the prior for log permeability is a Gaussian distribution with mean 5. The grid dimension is ~~70~~$N = 70$, and thus the uncertain parameter $\boldsymbol{u} = \{u^\ell\}_{\ell=1}^{N^2}$ has dimension 4900.

For ~~P2~~F2, we assume geometrical parameters $\boldsymbol{d} = \{d^i\}_{i=1}^5$ are drawn from uniform priors, namely $d^1 \sim U[0.3,\ 2.1]$, $d^2 \sim U[\pi/2,\ 6\pi]$, $d^3 \sim U[-\pi/2,\ \pi/2]$, $d^4 \sim U[0,\ 6]$, $d^5 \sim U[0.12,\ 4.2]$. Furthermore, we assume independence between geometric parameters and log permeability. The prior for log permeability is a Gaussian distribution with mean 15 outside the channel and with mean 100 inside the channel. The grid dimension is ~~50.~~ $N = 50$. Log permeability inside channel $\boldsymbol{u}^1 = \{u^{1,\ell}\}_{\ell=1}^{N^2}$

and log permeability outside channel $\boldsymbol{u}^2 = \{u^{2,\ell}\}_{\ell=1}^{N^2}$ are defined over the entire domain $50 \times 50$. Therefore, for ~~P2~~ F2 inference the uncertain parameter $\boldsymbol{u} = \{\boldsymbol{d}, \boldsymbol{u}^1, \boldsymbol{u}^2\}$ has dimension 5005. Moreover, for ~~P2~~ F2 we use the Metropolis-within-Gibbs methodology following Iglesias et al. (2014) to separate geometrical parameters and log permeability parameters within the mutation step, since it allows to better exploit the structure of the prior.

### 3.2 Observations

Both the true permeability and an initial ensemble are drawn from the same prior distribution as the prior includes knowledge about geological properties. However, an initial guess is computed on a coarse grid and the true solution is computed on a fine grid that has twice the resolution of the coarse grid. The synthetic observations of pressure are obtained by

$$\boldsymbol{y}_{\text{obs}} = \boldsymbol{L}(\boldsymbol{P}^{\text{true}}) + \boldsymbol{\eta}.$$

An element of $\boldsymbol{L}(\boldsymbol{P}^{\text{true}})$ is a linear functional of pressure, namely

$$L^j(\boldsymbol{P}^{\text{true}}) = \frac{1}{2\pi\sigma^2} \sum_{i=1}^{N_{\text{f}}} \exp\left(-\frac{\|\boldsymbol{X}^i - \boldsymbol{h}^j\|^2}{2\sigma^2}\right) (P^{\text{true}})^j \Delta x^2 \quad \text{for} \quad j = 1, \ldots, \kappa.$$

Here $\sigma = 0.01$, $\Delta x^2$ is the size of a grid cell $\boldsymbol{X}^i = (X^i, Y^i)$, $N_{\text{f}}$ is resolution of a fine grid, $\boldsymbol{h}^j$ is the location of the observation and $\kappa$ is the number of observations. This form of the observation functional and the parameterization ~~P1 and P2~~ F1 and F2 guaranty the continuity of the forward map from the uncertain parameters to the observations and thus the existence of the posterior distribution as shown by Iglesias et al. (2014). The observation noise $\boldsymbol{\eta}$ is drawn from a normal distribution with zero mean and known covariance matrix $\mathbf{R}$. We choose the observation noise to be 2 % of L2-norm of the true pressure. ~~Such~~ With such a small noise ~~makes the data assimilation problem hard to solve, since the likelihood is very peaked and~~ the likelihood is a peaked distribution. Therefore, a non-iterative data assimilation approach ~~fails~~requires a computationally unfeasible number of ensemble members to sample the posterior.

To save computational costs, we choose ESS threshold $M_{\text{thresh}} = M/3$ for tempering, and the length of Markov chain $\tau_{\text{max}} = 20$ for mutation.

### 3.3 Metrics

We conduct numerical experiments with ensemble sizes ~~100 and 500~~ $M = 100$ and $M = 500$, and 20 simulations with different initial ensemble realizations to check the robustness of results. We analyze the method's performance with respect to a pcn-MCMC solution from here on referred to as reference. An MCMC solution was obtained by combining 50 independent chains each of length $10^6$, $10^5$ burn-in period and $10^3$ thinning. For log permeability, we compute RMSE of the mean

$$\text{RMSE} = \sqrt{(\overline{\boldsymbol{u}} - \overline{\boldsymbol{u}}^{\text{ref}})'(\overline{\boldsymbol{u}} - \overline{\boldsymbol{u}}^{\text{ref}})}, \quad \text{where} \quad \overline{\boldsymbol{u}} = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{u}_i, \tag{18}$$

and $\boldsymbol{u}^{\text{ref}}$ is the reference solution.

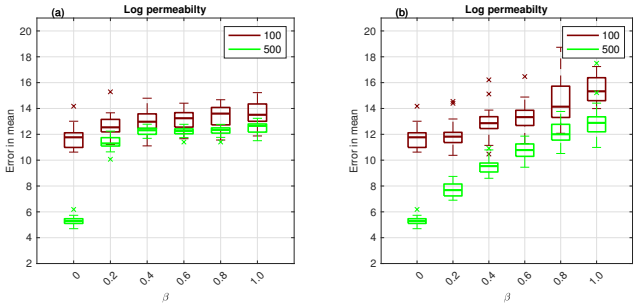


**Figure 2.** Application to ~~P1~~ F1 parameterization: using Sinkhorn approximation (a) and optimal transport resampling (b). Box plot over 20 independent simulations of RMSE of mean log permeability. X-axis is for the hybrid parameter, where $\beta = 0$ corresponds to ~~EKI~~ EnKF and $\beta = 1$ to TET(S)PF. Ensemble size ~~100~~ $M = 100$ is shown in red, and ~~500~~ $M = 500$ in green. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers.

For geometrical parameters $\boldsymbol{d}$, we compute the Kullback-Leibler divergence

$$\mathrm{D}_{\mathrm{KL}}^{i}(p^{\mathrm{ref}} \parallel p) = \sum_{j=1}^{M_{\mathrm{b}}} p^{\mathrm{ref}}(d_j^i) \log \frac{p^{\mathrm{ref}}(d_j^i)}{p(d_j^i)}, \tag{19}$$

where $p^{\mathrm{ref}}(d^i)$ is the reference posterior, $p(d^i)$ is approximated by the weights, and $M_{\mathrm{b}} = M/10$ is a chosen number of bins.

### 3.4 Application to ~~P1~~ F1 inference

For ~~P1~~F1, we perform numerical experiments using 36 uniformly distributed observations, which are displayed in circles in Fig. 3(a). We plot a box plot of RMSE given by Eq. (18) over 20 independent simulations in Fig. 2(a) using Sinkhorn approximation and in Fig. 2(b) using optimal transport. The x-axis is for the hybrid parameter $\beta$, whose value 0 corresponds to ~~EKI~~ EnKF and 1 to an adaptive SMC method with either a Sinkhorn approximation (TESPF) or optimal transport (TETPF). Ensemble size $M = 100$ is shown in red and $M = 500$ in green. First, we observe that at a small ensemble size ~~100~~ $M = 100$ and a large $\beta$ (namely $\beta \geq 0.6$) TESPF outperforms TETPF as the RMSE error is lower. Since Sinkhorn approximation is a regularization of an optimal transport solution, TESPF provides a smoother solution than TETPF that can be seen in Fig. 3(c) and Fig. 3(f), respectively, where we plot mean log permeability. Next, we see in Fig. 2 that the hybrid approach decreases RMSE compared to TET(S)PF: the smaller $\beta$ the smaller median of RMSE. ~~EKI~~ EnKF gives the smallest error due to the Gaussian parametrization of permeability. The advantage of the hybrid approach is most pronounced at a large ensemble size ~~500~~ $M = 500$ and optimal transport resampling. Furthermore, we note a discrepancy between the $M = 100$ and the $M = 500$ experiments at $\beta = 0$, thus EnKF alone. This is related to the curse of dimensionality. It appears that the ensemble size $M = 100$ is too small to estimate an uncertain parameter of the dimension $10^3$ using 36 accurate observations. However, at the ensemble size $M = 500$ EnKF alone ($\beta = 0$) gives an excellent performance compared to any combination ($\beta > 0$).

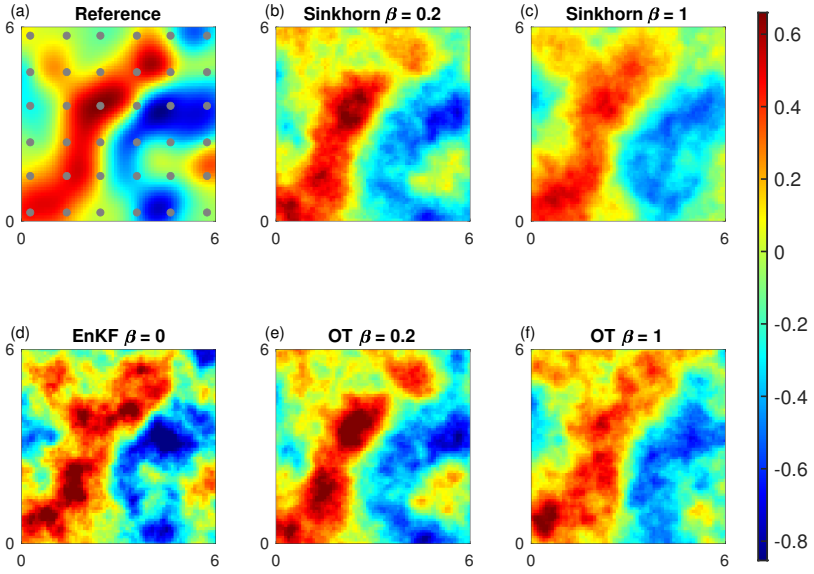


**Figure 3.** Mean log permeability for ~~P1~~ F1 inference for the lowest error at ensemble size ~~100.~~ $M = 100$. Observation locations are shown in circles. Reference (a), TESPF(0.2) (b), TESPF (c), ~~EKI~~EnKF (d), TETPF(0.2) (e), and TETPF (f).

We plot mean log permeability at ensemble size ~~100~~ $M = 100$ and a smallest RMSE over 20 simulations in Fig. 3(b)–(f) and of reference in Fig. 3(a). We see that ~~EKI~~ EnKF and TETPF(0.2) estimate well ~~mot~~ not only large-scale feature but also small-scale feature (e.g., negative mean at the top right corner) unlike TET(S)PF and TESPF(0.2).

### 3.5 Application to ~~P2~~ F2 inference

For ~~P2~~F2, we perform numerical experiments using 9 uniformly distributed observations. which are displayed in circles in Fig. 9(a). First, we display results obtained by Sinkhorn approximation. In Fig. 4, we plot box plot over 20 independent runs of KL divergence given by Eq. (19) for amplitude (a), frequency (b), angle (c), initial point (d), and width (e) that define channel. We see that ~~EKI~~ EnKF outperforms any TESPF($\cdot$) including TESPF for amplitude (a) and width (e). This is due to Gaussian-like posteriors of these two geometrical parameters displayed in Fig. 6(c) and Fig. 6(o), respectively. Due to Gaussian-like posteriors the hybrid approach decreases RMSE compared to TESPF: the smaller $\beta$ the smaller median of RMSE.

For frequency, angle, and initial point, whose KL divergence is displayed in Fig. 4(b), (c), and (d), respectively, the behaviour of adaptive SMC is nonlinear in terms of $\beta$. This is due to non Gaussian-like posteriors of these three geometrical parameters shown in Fig. 6(f), (i), and (l), respectively. Due to non Gaussian-like posteriors the hybrid approach gives an advantage

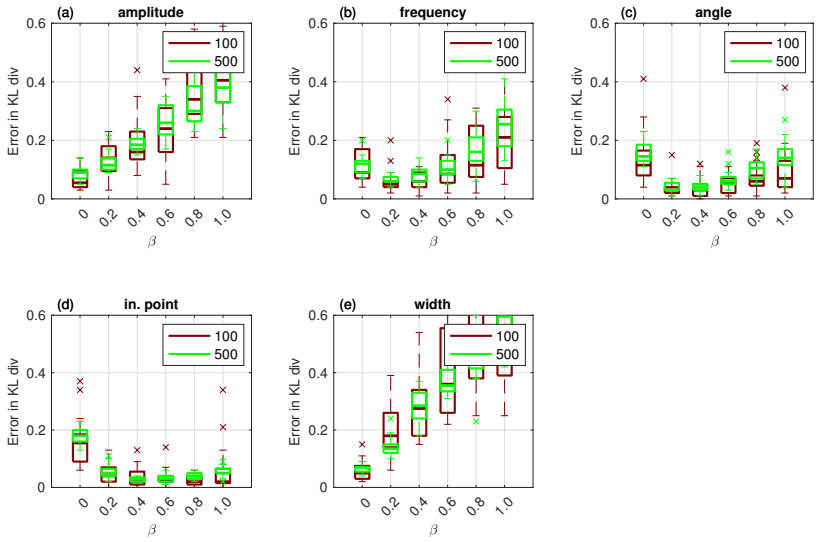


**Figure 4.** Application to ~~P2~~ F2 parameterization using Sinkhorn approximation. Box plot over 20 independent simulations of KL divergence for geometrical parameters: amplitude (a), frequency (b), angle (c), initial point (d), and width (e). X-axis is for the hybrid parameter, where $\beta = 0$ corresponds to ~~EKI~~ EnKF and $\beta = 1$ to TET(S)PF. Ensemble size ~~100~~ $M = 100$ is shown in red, and ~~500~~ $M = 500$ in green. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers.

over both TESPF and ~~EKI — there~~ EnKF — there exists a $\beta \neq 0$ for which the KL divergence is lowest ~~though~~ although it is inconsistent between geometrical parameters.

When comparing TESPF$(\cdot)$ to TETPF$(\cdot)$, we observe the same type of behaviour in terms of $\beta$: linear for amplitude and width, whose KL divergence is displayed in Fig. 5(a) and (e), respectively, and nonlinear for frequency, angle, and initial point, whose KL divergence is displayed in Fig. 5(b), (c), and (d), respectively. However, the KL divergence is smaller when optimal transport resampling is used instead of Sinkhorn approximation.

In Fig. 6, we plot posterior of geometrical parameters: amplitude (a)–(c), frequency (d)–(f), angle (g)–(i), initial point (j)–(l), and width (m)–(o), where on the left TESPF(0.2), in the middle TETPF(0.2), and on the right ~~EKI~~ EnKF are shown. In black is the reference, in red 20 simulations of ensemble size ~~100~~ $M = 100$, in green 20 simulations of ensemble size ~~500.~~ $M = 500$. The true parameters are shown as black cross. We see that as ensemble size increases posteriors approximated by TET(S)PF converge to the reference posterior unlike ~~EKI~~EnKF.

Now we investigate adaptive SMC performance for permeability estimation. First, we display results obtained by Sinkhorn approximation. ~~We plot box plot~~ The box plot shows over 20 independent simulations of RMSE given by Eq. (18) for log permeability outside channel in Fig. 7(a) and inside channel in Fig. 7(b). Even though log permeability is Gaussian distributed, for a small ensemble size ~~100~~ $M = 100$ there exists a $\beta \neq 0$ that gives lowest RMSE both outside and inside channel. As ensemble size increases, methods performance becomes equivalent.

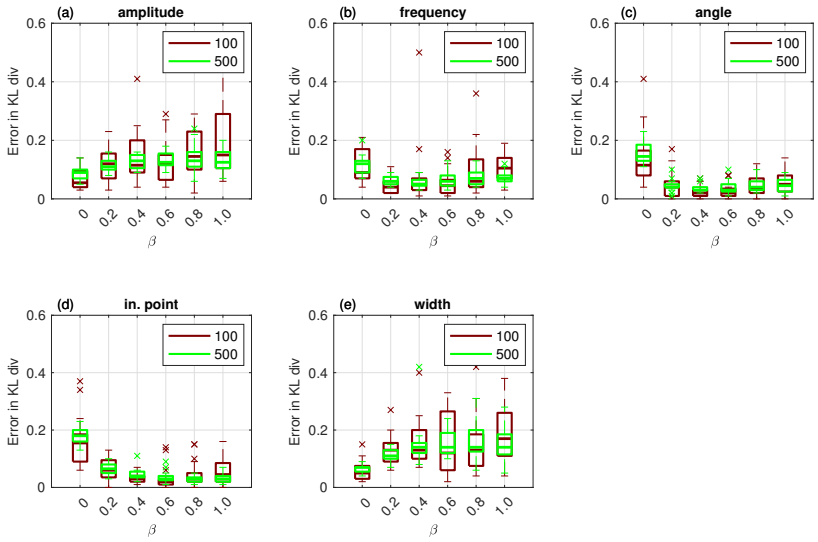


**Figure 5.** The same as Fig. 4 but using optimal transport resampling.

Next, we compare TESPF(·) to TETPF(·) for log permeability estimation outside and inside channel whose RMSE is displayed in Fig. 8(a) and (b), respectively. We observe the same type of behaviour in terms of $\beta$: nonlinear for a small ensemble size ~~100~~ $M = 100$, and equivalent for a larger ensemble size ~~500~~ $M = 500$. Furthermore, at a small ensemble size ~~100~~ $M = 100$ TESPF outperforms TETPF, which was also the case for ~~P1~~ F1 parameterization Sec. 3.4.

In Fig. 9, we show mean field of permeability over the channelized domain for reference for the lowest error at ensemble size ~~100~~ $M = 100$ for TESPF(0.2) (b), TESPF (c), ~~EKI~~EnKF (d), TETPF(0.2) (e), and TETPF (f). We plot mean log permeability over the channelized domain at ensemble size ~~100~~ $M = 100$ and a smallest RMSE over 20 simulations in Fig. 9(b)–(f) and of reference in Fig. 9(a). We see that TESPF(0.2) does an excellent job at such a small ensemble size by estimating well log permeability outside and inside channel, and parameters of the channel itself.

## 4 Conclusions

A Sinkhorn adaptation, namely the TESPF, of the previously proposed TETPF has been introduced and numerically investigated on a parameter estimation problem. The TESPF has ~~considerable smaller computational complexity than the TETPF, namely $\mathcal{O}[T(MC + M^2C(\alpha) + \tau_{\max}MC)]$ vs $\mathcal{O}[T(MC + M^3 \log M + \tau_{\max}MC)]$, yet has~~ similar accuracy results than the TETPF (see Fig. 7, 8 and 6) ~~.~~ while it can have considerable smaller computational complexity. Specifically, the TESPF has compelxity $\mathcal{O}[T(MC + M^2C(\alpha) + \tau_{\max}MC)]$ and the TETPF $\mathcal{O}[T(MC + M^3 \log M + \tau_{\max}MC)]$, (for a complete overview see table B1). In particular, the TESPF outperforms the ~~EKI~~ EnKF for non-Gaussian distributed parameters (e.g., initial point and angle in ~~P2~~F2). This makes the proposed method a promising option for the ~~high dimensional~~ high-dimensional nonlinear problems one is typically faced with in ~~geophysical applications~~reservoir engineering. Further, to counter balance potential

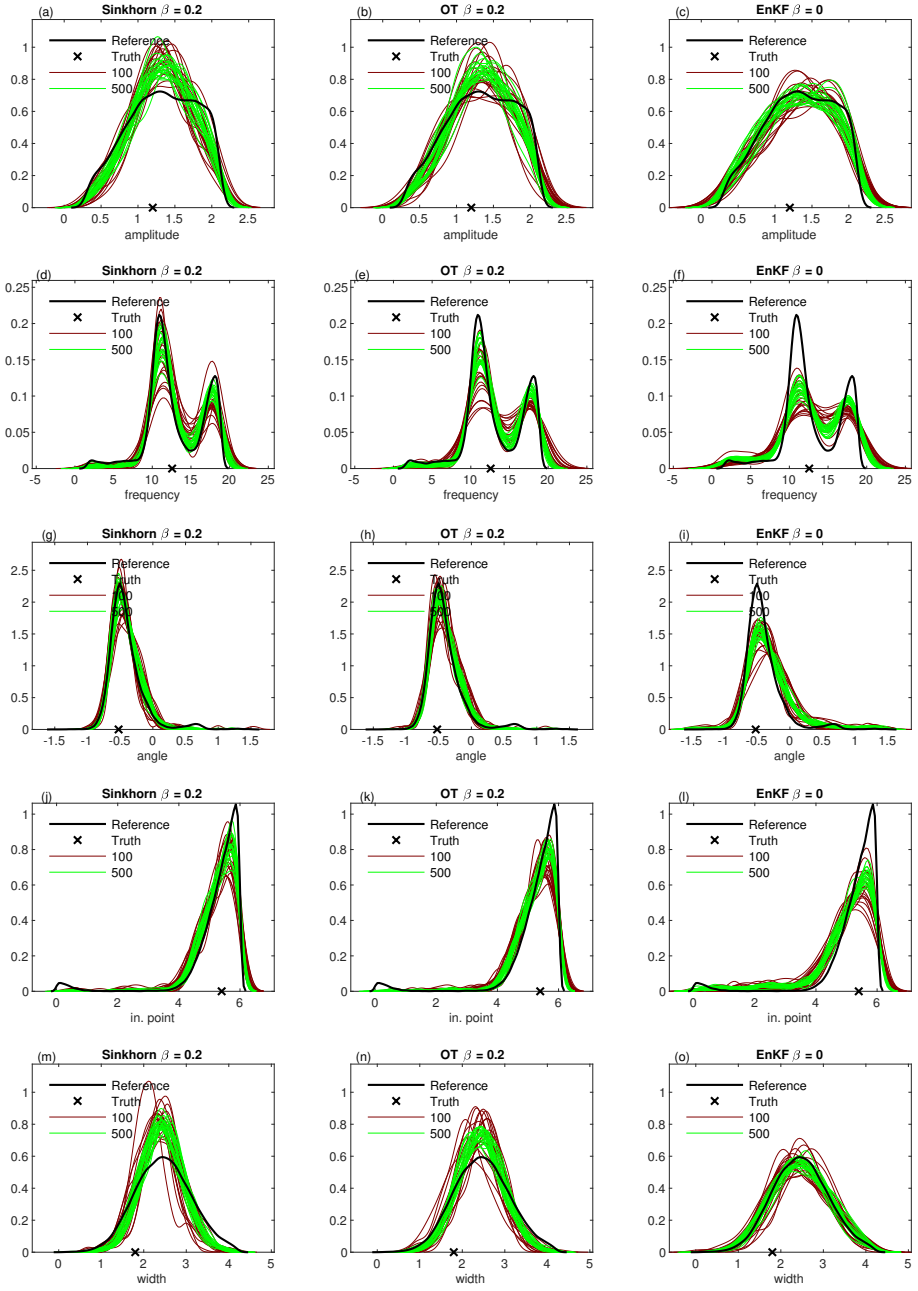


**Figure 6.** Posterior of geometrical parameters for ~~P2~~ F2 inference: amplitude (a)–(c), frequency (d)–(f), angle (g)–(i), initial point (j)–(l), and width (m)–(o). On the left is TESPF(0.2); in the middle is TETPF(0.2), and on the right is ~~EKI~~EnKF. In black is reference, in red 20 simulations of ensemble size ~~100~~$M = 100$, in green 20 simulations of ensemble size ~~500.~~ $M = 500$. The true parameters are shown as black cross.

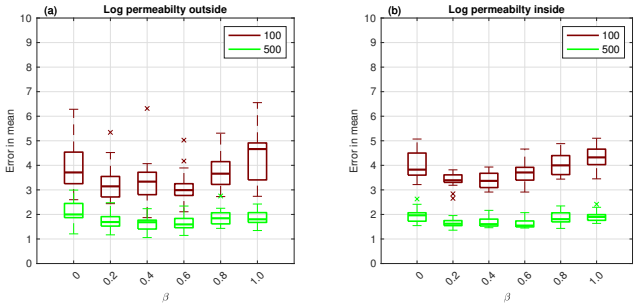


**Figure 7.** Application to ~~P2~~ F2 parameterization with Sinkhorn approximation. Box plot over 20 independent simulations of RMSE of mean log permeability outside channel (a) and inside channel (b). X-axis is for the hybrid parameter, where $\beta = 0$ corresponds to ~~EKI~~ EnKF and $\beta = 1$ to TET(S)PF. Ensemble size ~~100~~ $M = 100$ is shown in red, and ~~500~~ $M = 500$ in green. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme data points, and crosses are outliers.

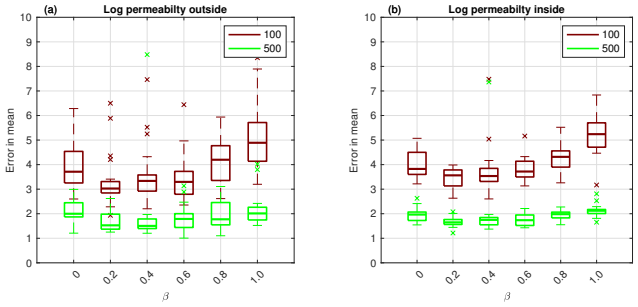


**Figure 8.** The same as Fig. 7 but using optimal transport resampling.

robustness problems of the TETPF and its Sinkhorn adaptation a hybrid between ~~EKI~~ EnKF and TET(S)PF is proposed and studied by means of the two configurations of the steady-state single-phase Darcy flow model. The combination of the two adaptive SMC methods with complementing properties, i.e., $\beta \in (0, 1)$, is superior to the individual adaptive SMC method, i.e., $\beta = 0$ or 1, for all non-Gaussian distributed parameters and performs better than the pure TETPF and the TETSPF for Gaussian distributed parameters in ~~P1~~F1. This suggests a hybrid approach ~~provides all the desirable properties required~~ has a great potential to obtain robust and ~~highly accurate~~ highly-accurate approximate solutions of nonlinear ~~high dimensional~~ high-dimensional Bayesian inference problems. Note that we have considered a synthetic case, where the truth is available, and thus chose $\beta$ in terms of accuracy of an estimate. However, in a realistic application the truth is not provided. In the context of state estimation with an underlying dynamical system it has been suggested to adaptively change the hybrid parameter with respect to the effective sample size. As the tempering scheme is already changed according to the effective sample size this ansatz would require to define the interplay between the two tuning variables. An ad-hoc choice for $\beta$ could be 0.2 or 0.3. This is motivated by the fact that the particle filter is too unstable in high dimensions and it is therefore sensible to

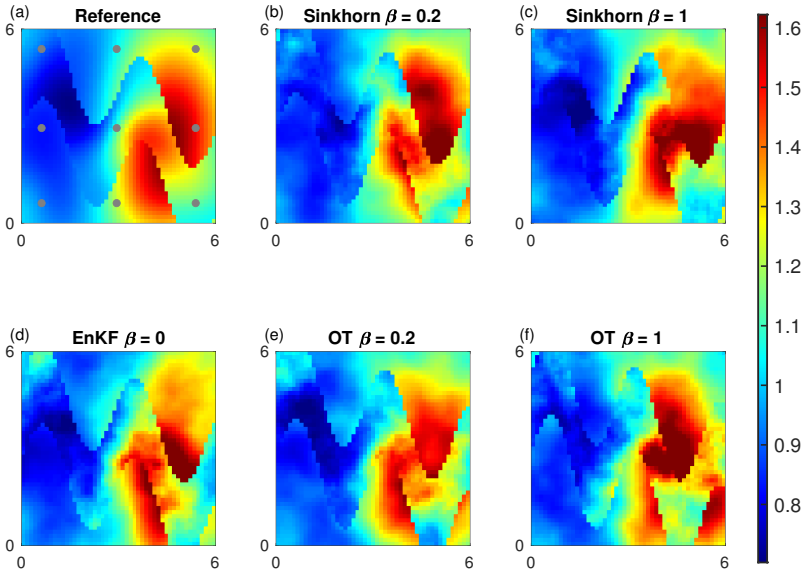


**Figure 9.** Mean log permeability for ~~P2~~ F2 inference for the lowest error at ensemble size ~~100.~~ $M = 100$. Observation locations are shown in circles. Reference (a), TESPF(0.2) (b), TESPF (c), ~~EKI~~EnKF (d), TETPF(0.2) (e), and TETPF (f).

use a tuning parameter prioritising the EnKF. The ad-hoc choice is supported by the numerical results in Section 3 and in Acevedo et al. (2017); de Wiljes et al. (2020) in the context of state-estimation.

## Appendix A: Pseudocode

**Algorithm 1** Sample mutation

**Require:** $\theta \in (0,1)$ and an integer $\tau_{\text{max}}$

**for** $i = 1, \ldots, M$ **do**

Initialize $\boldsymbol{v}_i(0) = \tilde{\boldsymbol{u}}_{t,i}$

**while** $\tau \leq \tau_{\text{max}}$ **do**

Propose $\boldsymbol{v}_i^{\text{prop}}$ using Eq. (5) for Gaussian probability or Eq. (6) for uniform probability

Set $\boldsymbol{v}_i(\tau+1) = \boldsymbol{v}_i^{\text{prop}}$ with probability Eq. (7) and set $\boldsymbol{v}_i(\tau+1) = \tilde{\boldsymbol{u}}_{t,i}$ with probability Eq. (8)

$\tau \leftarrow \tau + 1$

**end while**

Set $\boldsymbol{u}_{t,i} = \boldsymbol{v}_i(\tau_{\text{max}})$

**end for**

**Algorithm 2** Resampling based on optimal transport

**Require:** $\{\boldsymbol{u}_{t-1,i}\}_{i=1}^{M}$ and $\boldsymbol{w}_{t-1} = \{w_{t-1,1}, \ldots, w_{t-1,M}\}$

Compute $\mathbf{Z}$ with $z_{ij} = \|\boldsymbol{u}_{t-1,i} - \boldsymbol{u}_{t-1,j}\|^2$

Supply $\mathbf{Z}$ and $\boldsymbol{w}_{t-1}$ to the FastEMD algorithm of Pele & Werman with the output being the coupling $\mathbf{S}$

Compute new samples $\{\tilde{\boldsymbol{u}}_{t,i}\}_{i=1}^{M}$ from Eq. (11)

**Algorithm 3** Sinkhorn iteration for optimal transport problem with entropic regularisation

**Require:** regularisation parameter $\alpha$, $\{\boldsymbol{u}_{t-1,i}\}_{i=1}^{M}$ and $\boldsymbol{w}_{t-1} = \{w_{t-1,1}, \ldots, w_{t-1,M}\}$

Compute $\mathbf{Z}$ with $z_{ij} = ||\boldsymbol{u}_{t-1,i} - \boldsymbol{u}_{t-1,j}||^2$

Normalise $\mathbf{Z}$ with respect to its maximum entry

**while** $\varepsilon \geq 1.0e-8$ **do**

$\boldsymbol{b} = \boldsymbol{w}_{t-1} ./ [\exp(-\alpha \mathbf{Z})\boldsymbol{a}]$

$\boldsymbol{a} = \left(\frac{1}{M}\boldsymbol{I}_M / M\right) ./ [\exp(-\alpha \mathbf{Z})\boldsymbol{b}]$

$\mathbf{S} = \mathbf{diag}(\boldsymbol{b}) \exp(-\alpha \mathbf{Z}) \mathbf{diag}(\boldsymbol{a})$

$\hat{\boldsymbol{w}} = \mathbf{S}\boldsymbol{I}_M$

$\varepsilon = \|\hat{\boldsymbol{w}} - \boldsymbol{w}_{t-1}\|$

**end while**

**return** $\mathbf{S}^* = \mathbf{S}$

## Appendix B: Computational Complexity

**Algorithm 4** Adaptive SMC: TET(S)PF

**Require:** an initial ensemble $\{\boldsymbol{u}_{0,i}\}_{i=1}^{M} \sim \mu_0$, $\theta \in (0,1)$ and integers $\tau_{\text{max}}$ and $1 < M_{\text{thresh}} < M$

Set $\phi_0 = 0$

**while** $\phi_t \leq 1$ **do**

$t \to t+1$

Compute the likelihood $g(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}})$ from Eq. (2) (for $i = 1, \ldots, M$)

Compute the tempering parameter $\phi_t$:

**if** $\min_{\phi \in (\phi_{t-1},1)} \text{ESS}_t(\phi) > M_{\text{thresh}}$ **then**

set $\phi_t = 1$

**else**

compute $\phi_t$ such that $\text{ESS}_t(\phi) \approx M_{\text{thresh}}$ using a bisection algorithm on $(\phi_{t-1}, 1]$

**end if**

Compute weights $\boldsymbol{w}_{t-1} = \{w_{t-1,1}, \ldots, w_{t-1,M}\}$ from Eq. (3)

Create new samples $\{\tilde{\boldsymbol{u}}_{t,i}\}_{i=1}^{M}$ using optimal (Sinkhorn) resampling via Algorithm 2(3)

Compute $\{\boldsymbol{u}_{t,i}\}_{i=1}^{M}$ using sample mutation via Algorithm 1

**end while**

**Algorithm 5** EnKF

**Require:** an initial ensemble $\{\boldsymbol{u}_{0,i}\}_{i=1}^{M} \sim \mu_0$, $\theta \in (0,1)$ and integers $\tau_{\text{max}}$ and $1 < M_{\text{thresh}} < M$

Set $\phi_0 = 0$

**while** $\phi_t \leq 1$ **do**

$t \to t+1$

Compute the likelihood $g(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}})$ from Eq. (2) (for $i = 1, \ldots, M$)

Compute the tempering parameter $\phi_t$:

**if** $\min_{\phi \in (\phi_{t-1},1)} \text{ESS}_t(\phi) > M_{\text{thresh}}$ **then**

set $\phi_t = 1$

**else**

compute $\phi_t$ such that $\text{ESS}_t(\phi) \approx M_{\text{thresh}}$ using a bisection algorithm on $(\phi_{t-1}, 1]$

**end if**

Create new samples $\{\tilde{\boldsymbol{u}}_{t,i}\}_{i=1}^{M}$ using Eq. (13)

Compute $\{\boldsymbol{u}_{t,i}\}_{i=1}^{M}$ using sample mutation via Algorithm 1

**end while**

**Algorithm 6** Hybrid EnKF-TET(S)PF

**Require:** initial initial ensemble $\{\boldsymbol{u}_{0,i}\}_{i=1}^{M} \sim \mu_0$, $\theta \in (0,1)$, hybrid parameter $\beta$ and integers $\tau_{\max}$ and $1 < M_{\text{thresh}} < M$

Set $\phi_0 = 0$

**while** $\phi_t \leq 1$ **do**

$t \to t+1$

Compute the likelihood $g_1(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}})$ from Eq. (14) (for $i = 1, \ldots, M$)

Set $g(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}}) = g_1(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}})$ (for $i = 1, \ldots, M$)

Compute the tempering parameter $\phi_t$:

**if** $\min_{\phi \in (\phi_{t-1}, 1)} \text{ESS}_t(\phi) > M_{\text{thresh}}$ **then**

set $\phi_t = 1$

**else**

compute $\phi_t$ such that $\text{ESS}_t(\phi) \approx M_{\text{thresh}}$ using a bisection algorithm on $(\phi_{t-1}, 1]$

**end if**

Create new samples $\{\tilde{\boldsymbol{u}}_{t,i}^{\beta}\}_{i=1}^{M}$ using Eq. (13)

Compute the likelihood $g_2(\tilde{\boldsymbol{u}}_{t,i}^{\beta}; \boldsymbol{y}_{\text{obs}})$ from Eq. (15) (for $i = 1, \ldots, M$)

Set $g(\boldsymbol{u}_{t-1,i}; \boldsymbol{y}_{\text{obs}}) = g_2(\tilde{\boldsymbol{u}}_{t,i}^{\beta}; \boldsymbol{y}_{\text{obs}})$ (for $i = 1, \ldots, M$)

Compute weights $\boldsymbol{w}_{t-1} = \{w_{t-1,1}, \ldots, w_{t-1,M}\}$ from Eq. (3)

Create new samples $\{\tilde{\boldsymbol{u}}_{t,i}\}_{i=1}^{M}$ using optimal (Sinkhorn) resampling via Algorithm 2(3)

Compute $\{\boldsymbol{u}_{t,i}\}_{i=1}^{M}$ using sample mutation via Algorithm 1

**end while**

| **Algorithm** | *Complexity* |
|---|---|
| TETPF | $\mathcal{O}[T(MC + M^3 \log M + \tau_{\max} MC)]$ |
| TESPF | $\mathcal{O}[T(MC + M^2 C(\alpha) + \tau_{\max} MC)]$ |
| EnKF | $\mathcal{O}[T(MC + \kappa^2 n + \tau_{\max} MC)]$ |
| Hybrid EnKF-TETPF | $\mathcal{O}[T(MC + \kappa^2 n + MC + M^3 \log M + \tau_{\max} MC)]$ |
| Hybrid EnKF-TESPF | $\mathcal{O}[T(MC + \kappa^2 n + MC + M^2 C(\alpha) + \tau_{\max} MC)]$ |
| Forward model $G$ | $\mathcal{O}(MC)$ |
| pcn-MCMC mutation | $\mathcal{O}(\tau_{\max} MC)$ |
| FastEMD | $\mathcal{O}(M^3 \log M)$ |
| Sinkhorn approximation | $\mathcal{O}(M^2 C(\alpha))$ |

**Table B1.** The table provides an overview of the computational complexity of all the algorithms considered in the manuscript.

*Data availability.* Data and MATLAB codes for generating the plots are available in Ruchi et al. (2020).

*Author contributions.* S.R., S.D. and J.dW. designed the research, S.D. ran the numerical experiments, S.R., S.D. and J.dW. analyzed the results and wrote the manuscript.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The research of J.dW. and S.R. have been partially funded by Deutsche Forschungsgemeinschaft (DFG) - SFB1294/1 - 318763901. Further J.dW. has been supported by Simons CRM Scholar-in-Residence Program and ERC Advanced Grant ACRCC (grant 339390). S.R. has been supported by the research programme Shell-NWO/FOM Computational Sciences for Energy Research (CSER) with project number 14CSER007 which is partly financed by the Netherlands Organization for Scientific Research (NWO).

## References


Acevedo, W., de Wiljes, J., and Reich, S.: Second-order Accurate Ensemble Transform Particle Filters, SIAM J. Sci. Comput., 39, A1834–A1850, 2017.

Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., and Stuart, A. M.: Importance sampling: computational complexity and intrinsic dimension, Statistical Science, 32, 405–431, https://doi.org/10.1214/17-STS611, 2017.

Anderson, J.: An ensemble adjustment Kalman filter for data assimilation, Monthly Weather Review, 129, 2884–2903, 2001.

Bardsley, J., Solonen, A., Haario, H., and Laine, M.: Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems, SIAM J. Sci. Comput., 36, A1895–A1910, 2014.

Beskos, A., Crisan, D., and Jasra, A.: On the stability of sequential Monte Carlo methods in high dimensions, Ann. Appl. Probab., 24, 1396–1445, https://doi.org/10.1214/13-AAP951, 2014.

Beskos, A., Jasra, A., Muzaffer, E. A., and Stuart, A. M.: Sequential Monte Carlo methods for Bayesian elliptic inverse problems, Statistics and Computing, 25, 727–737, https://doi.org/10.1007/s11222-015-9556-7, 2015.

Blömker, D., Schillings, C., Wacker, P., and Weissmann, S.: Well posedness and convergence analysis of the ensemble Kalman inversion, Inverse Problems, 35, 085 007, https://doi.org/10.1088/1361-6420/ab149c, https://doi.org/10.1088%2F1361-6420%2Fab149c, 2019.

Burgers, G., Leeuwen, P., 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, 1998.

Carrassi, A., Bocquet, M., Bertino, ., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, WIREs Climate Change, 9, e535, https://doi.org/10.1002/wcc.535, 2018.

Chada, N., Iglesias, M., Roininen, L., and Stuart, A.: Parameterizations for Ensemble Kalman Inversion, Inverse Problems, 34, 055 009, 2018.

Chen, Y. and Oliver, D.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother, Math. Geosci., 44, 1–26, 2012.

Chustagulprom, N., Reich, S., and Reinhardt, M.: A Hybrid Ensemble Transform Particle Filter for Nonlinear and Spatially Extended Dynamical Systems, SIAM/ASA Journal on Uncertainty Quantification, 4, 592–608, https://doi.org/10.1137/15M1040967, 2016.

Cotter, S., Roberts, G., Stuart, A., and White, D.: MCMC methods for functions: modifying old algorithms to make them faster, Statistical Science, 28, 424–446, 2013.

Cuturi, M.: Sinkhorn Distances: Lightspeed Computation of Optimal Transport, in: Advances in Neural Information Processing Systems 26, edited by Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., pp. 2292–2300, Curran Associates, Inc., http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.pdf, 2013.

Dashti, M. and Stuart, A. M.: The Bayesian Approach to Inverse Problems, pp. 311–428, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-12385-1_7, 2017.

de Wiljes, J., Pathiraja, S., and Reich, S.: Ensemble Transform Algorithms for Nonlinear Smoothing Problems, SIAM J. Sci. Comput.,, 42, A87–A114, 2020.

Del Moral, P., Doucet, A., and Jasra, A.: Sequential Monte Carlo samplers, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68, 411–436, https://doi.org/10.1111/j.1467-9868.2006.00553.x, 2006.

Emerick, A. and Reynolds, A.: Ensemble smoother with multiple data assimilation, Computers & Geosciences, 55, 3–15, 2013.

Ernst, O. G., Sprungk, B., and Starkloff, H.-J.: Analysis of the Ensemble and Polynomial Chaos Kalman Filters in Bayesian Inverse Problems, SIAM/ASA Journal on Uncertainty Quantification, 3, 823–851, https://doi.org/10.1137/140981319, 2015.

Evensen, G.: Analysis of iterative ensemble smoothers for solving inverse problems, Computational Geosciences, 22, 885–908, https://doi.org/10.1007/s10596-018-9731-y, 2018.

Frei, M. and Künsch, H.: Bridging the ensemble Kalman and particle filters, Biometrika, 100, 781–800, 2013.

Frei, M. and Künsch, H. R.: Bridging the ensemble Kalman and particle filters, Biometrika, 100, 781–800, https://doi.org/10.1093/biomet/ast020, 2013.

Hairer, M., Stuart, A., and Vollmer, S.: Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions, Ann. Appl. Probab., 24, 2455–2490, 2014.

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.

Iglesias, M., Park, M., and Tretyakov, M.: Bayesian inversion in resin transfer molding, Inverse Problems, 34, 105 002, 2018.

Iglesias, M. A.: A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems, Inverse Problems, 32, 025 002, 2016.

Iglesias, M. A., Lin, K., and Stuart, A. M.: Well-posed Bayesian geometric inverse problems arising in subsurface flow, inverse problems, 30, 114 001, 2014.

Kantorovich, L. V.: On the translocation of masses, in: Dokl. Akad. Nauk. USSR (NS), vol. 37, pp. 199–201, 1942.

Liu, Y., Haussaire, J., Bocquet, M., Roustan, Y., Saunier, O., and Mathieu, A.: Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima-Daiichi accidental releases of radionuclides, Q. J. R. Meteorol. Soc., 143, 2886–2901, 2017.

Lorentzen, R. J., Bhakta, T., Grana, D., Luo, X., Valestrand, R., and Nævdal, G.: Simultaneous assimilation of production and seismic data: application to the Norne field, Computational Geosciences, 24, 907–920, https://doi.org/10.1007/s10596-019-09900-0, 2020.

Matérn, B.: Spatial Variation, Lecture Notes in Statistics, No. 36, Springer, 1986.

Monge, G.: Mémoire sur la théorie des déblais et des remblais, Histoire de l'Académie Royale des Sciences de Paris, 1781.

Neal, R. M.: Annealed importance sampling, Statistics and computing, 11, 125–139, 2001.

Oliver, D., He, N., and Reynolds, A.: Conditioning permeability fields to pressure data, ECMOR V-5th European Conference on the Mathematics of Oil Recovery, pp. 259–269, 1996.

Pele, O. and Werman, M.: Fast and robust earth mover's distances, in: Computer vision, 2009 IEEE 12th international conference on, pp. 460–467, IEEE, 2009.

Peyré, G. and Cuturi, M.: Computational Optimal Transport: With Applications to Data Science, 2019.

Reich, S.: A nonparametric ensemble transform method for Bayesian inference, SIAM Journal on Scientific Computing, 35, A2013–A2024, 2013.

Roberts, G. O. and Rosenthal, J. S.: Optimal scaling for various Metropolis-Hastings algorithms, Statist. Sci., 16, 351–367, https://doi.org/10.1214/ss/1015346320, 2001.

Ruchi, S., Dubinkina, S., and Iglesias, M. A.: Tempered ensemble transform particle filter for non-Gaussian elliptic problems, Inverse Problems, 35, 115 005, 2019.

Ruchi, S., Dubinkina, S., and de Wiljes, J.: Data underlying the paper: Fast hybrid tempered ensemble transform filter formulation for Bayesian elliptical problems via Sinkhorn approximation, 4TU, Centre for Research Data, Dataset, https://doi.org/10.4121/12987719, 2020.

Santitissadeekorn, N. and Jones, C.: Two-Stage Filtering for Joint State-Parameter Estimation, Monthly Weather Review, 143, 2028–2042, https://doi.org/10.1175/MWR-D-14-00176.1, 2015.

Sinkhorn, R.: Diagonal equivalence to matrices with prescribed row and column sums, The American Mathematical Monthly, 74, 402–405, 1967.

Stordal, A., Karlsen, H. A., Nævdal, G., Skaug, H., and Vallès, B.: Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter, Computational Geosciences, 15, 293–305, https://doi.org/10.1007/s10596-010-9207-1, 2011.

Stuart, A. M.: Inverse problems: a Bayesian perspective, Acta Numerica, 19, 451–559, 2010.

Vergé, C., Dubarry, C., Del Moral, P., and Moulines, E.: On parallel implementation of sequential Monte Carlo methods: the island particle model, Statistics and Computing, 25, 243–260, https://doi.org/10.1007/s11222-013-9429-x, 2015.

Villani, C.: Optimal Transport: Old and New, Grundlehren der mathematischen Wissenschaften, Springer, 2009 edn., 2008.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z.: Ensemble Data Assimilation with the NCEP Global Forecast System, Monthly Weather Review, 136, 463–482, https://doi.org/10.1175/2007MWR2018.1, 2008.

Zovi, F., Camporese, M., Hendricks Franssen, H.-J., Huisman, J. A., and Salandin, P.: Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter, Journal of Hydrology, 548, 208–224, https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.02.056, 2017.