Data Assimilation and Online Parameter Optimization in Groundwater Modeling Using Nested Particle Filters

M. Ramgraber[1,2] (ID), **C. Albert[3]** (ID), **and M. Schirmer[1,2]**

[1]Department Water Resources and Drinking Water, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Zürich, Switzerland, [2]Centre d'hydrogéologie et de géothermie (CHYN), University of Neuchâtel, Neuchâtel, Switzerland, [3]Department Systems Analysis, Integrated Assessment and Modelling, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Zürich, Switzerland

Abstract Over the past decades, advances in data collection and machine learning have paved the way for the development of autonomous simulation frameworks. Among these, many are capable not only of assimilating real-time data to correct their predictive shortcomings but also of improving their future performance through self-optimization. In hydrogeology, such techniques harbor great potential for informing sustainable management practices. Simulating the intricacies of groundwater flow requires an adequate representation of unknown, often highly heterogeneous geology. Unfortunately, it is difficult to reconcile the structural complexity demanded by realistic geology with the simplifying assumptions introduced in many calibration methods. The particle filter framework would provide the necessary versatility to retain such complex information but suffers from the curse of dimensionality, a fundamental limitation discouraging its use in systems with many unknowns. Due to the prevalence of such systems in hydrogeology, the particle filter has received little attention in groundwater modeling so far. In this study, we explore the combined use of dimension-reducing techniques and artificial parameter dynamics to enable a particle filter framework for a groundwater model. Exploiting freedom in the design of the dimension-reduction approach, we ensure consistency with a predefined geological pattern. The performance of the resulting optimizer is demonstrated in a synthetic test case for three such geological configurations and compared to two Ensemble Kalman Filter setups. Favorable results even for deliberately misspecified settings make us hopeful that nested particle filters may constitute a useful tool for geologically consistent real-time parameter optimization.

1. Introduction

Parameter estimation is an essential part of any simulation in which the demand for parameter information outweighs its availability. In few environmental disciplines is this discrepancy as pronounced as in hydrogeology: Any attempt to faithfully capture the intricacies of groundwater flow demands a realistic representation of often highly heterogeneous geology (Rubin & Hubbard, 2006). Unfortunately, the extent, nature, or even existence of these features can scarcely be derived from the surface (e.g., De Marsily et al., 2005). Direct measurements to ascertain their properties offer little relief, constituting a time- and resource-intensive endeavor with no guarantee to adequately delineate the flow-relevant structure (Schöniger et al., 2012).

The task of parameter estimation is the inference of the properties from information of dependent states. In hydrogeology, transient hydraulic heads often take this role. This topic has been explored in a large body of literature over the years (e.g., Hill et al., 2000; McLaughlin & Townley, 1996; Yeh, 1986). Among these techniques, *batch-calibration* (or *history-matching*) approaches—which rely on bulk processing a preexisting set of observations—have been state of the art for decades and still remain highly popular today. Many of these techniques, like the parameter estimation and uncertainty analysis tool PEST (Doherty, 1994, 2015), have become widespread industry standard.

More recently, however, there has been a growing interest in the development of *real-time* (or *online*) parameter estimation techniques (e.g., Hendricks Franssen & Kinzelbach, 2008). Sparked by increasing availability of real-time data from wireless sensor networks (e.g., Cardell-Oliver et al., 2005) and satellite-based remote sensing (e.g., Houser et al., 1998), such algorithms assimilate a data stream of state measurements to gradually improve parameters during active model operation. The advantage over batch-calibration approaches is evident: There is no need to wait until a sufficiently large body of data is collected. Instead,

such algorithms may autonomously process data as they become available, theoretically providing a best guess estimate at all times.

1.1. Limitations of the State of the Art

In pursuit of data assimilation and real-time calibration, the Ensemble Kalman Filter (EnKF; Evensen, 1994, 2003) has established itself as one of the most popular approaches in environmental science. The reasons for this algorithm's success are manifold: Ease of implementation (Iglesias et al., 2013), high computational efficiency (Hu et al., 2013; Zovi et al., 2017), and relative robustness to violations of its fundamental assumptions of *Gaussianity* and *linearity* (Iglesias et al., 2013; Katzfuss et al., 2016) equally contributed to its prevalence in the data assimilation community.

Nonetheless, the EnKF is not without shortcomings: Its core assumptions are essentially never met in hydrogeological practice, and in such cases its high efficiency comes at the price of excluding many possible solutions. When mean and variance do not provide sufficient statistics for the probability density function (pdf) under investigation, or the system propagation is nonlinear, the assumption of Gaussianity may be violated and the EnKF will yield only approximate solutions (e.g., Amezcua & Van Leeuwen, 2014; Schillings & Stuart, 2017), the usefulness of which has to be evaluated on a case-by-case basis.

Unfortunately, the pursuit of realistic geology during parameter estimation is a case in which these approximations often prove insufficient. Parameter uncertainty due to ignorance of geological and sedimentary features can take on several shapes: Often, not only the hydraulic properties of these features are unknown but also their spatial extent and arrangement. From a mathematical perspective, the uncertainty originating from the unknown arrangement of the geological features is characterized by distinct multimodality of the parameters' pdf and thus poorly reflected by the EnKF's assumption of unimodal Gaussianity. It has been shown that if a latent geological structure of the prior is not sufficiently informed by the observed states, the EnKF updates tend to not (or only vestigially) preserve its characteristics (e.g., Zovi et al., 2017).

In a bid to address this issue, it has been proposed to employ Gaussian anamorphosis (GA, sometimes also called *normal score transformation*), an approach converting non-Gaussian marginal distributions to unimodal Gaussian ones for the duration of the assimilation step (e.g., Schöniger et al., 2012; Zhou et al., 2011). While GA has been reported to alleviate this structural degeneration to a certain degree (Zovi et al., 2017), it may come at the price of increased nonlinearity of the observation operator (Amezcua & Van Leeuwen, 2014) or the relationship between transformed variables (Zhou et al., 2011). When applied only to state space, Schöniger et al. (2012) observed a more linear relationship between transformed states and untransformed parameters. However, they remark that univariate transformations may only transform the marginals but cannot alter the multivariate dependence structures between the variables. For some types of data (e.g., concentration data; see Schöniger et al., 2012) this dependence structure can be far from Gaussian and in such cases univariate transformations will not yield multivariate Gaussian values.

An interesting alternative approach was proposed by Hu et al. (2013). Rather than calibrating the model parameters directly, their EnKF implementation filters white noise fields instead. These noise fields are subsequently used as random seeds in the generation of geostatistical parameter fields from multipoint statistics (MPS; e.g., Caers & Zhang, 2004). Since MPS generates fields consistent with a predefined geology—the training image—their EnKF implementation retains consistency with the desired structure throughout model calibration. Unfortunately, such indirect approaches introduce (further) nonlinearity. This might affect the filter's performance, since the convergence of EnKF-based parameter estimation depends on the strength of the linear correlation between state observations and filtered parameters (Jafarpour & Tarrahi, 2011).

1.2. Beyond Gaussianity

In an effort to overcome the assumption of Gaussianity and its limitations, the particle filter is a natural alternative to the EnKF. Based on a direct Monte Carlo representation of the underlying pdf, this filter makes no assumptions about the shape of the pdf or the nature of the system dynamics (e.g., Doucet & Johansen, 2009; Doucet & Tadić, 2003) and updates its ensemble through adjustments of the particles' retrieval weights. Unfortunately, this flexibility comes at the price of the so-called *curse of dimensionality*: The number of Monte Carlo samples (particles) required to adequately represent the pdf increases exponentially with the number of unknown variables (e.g., Bengtsson et al., 2008; Farchi & Bocquet, 2018; van Leeuwen, 2010).

Hydrological numerical models may have millions of cells, and each cell may have different hydraulic parameters, so that naive applications of the particle filter have been deemed computationally infeasible (e.g., Aanonsen et al., 2009; Ruiz et al., 2013; Schöniger et al., 2012). Over the past decades, particle filters have been assigned a niche role in hydrology, largely limited in their application to conceptual or lumped models with only few unknown parameters (e.g., Moradkhani et al., 2005). Recently, their scope of application has widened to include distributed models such as drought forecasting frameworks (Yan et al., 2017; Yan et al., 2018).

However, much progress has been made in the development of highly efficient filter techniques (Morzfeld et al., 2017). Particularly for state estimation, the number of required particles can be drastically reduced: Through manipulation of the proposal pdf, van Leeuwen (2010) reports a successful application of a filter with only 20 particles to a 1000-dimensional problem—an otherwise thoroughly hopeless task—and surmises that "the curse of dimensionality may have a cure." Recently, van Leeuwen et al. (2019) give an overview of particle filter variants for high-dimensional geophysical applications, mostly focused on state estimation, and Farchi and Bocquet (2018) provide an overview over local particle filters, suggested by Snyder et al. (2008) to overcome obstacles of high-dimensional particle filters. For parameter optimization, some authors (e.g., Abbaszadeh et al., 2018; Zhu et al., 2018) transferred elements from genetic algorithms to particle filters in order to alleviate the curse of dimensionality.

In general, the raw number of unknown variables may be a poor measure of the system's true complexity. Especially in hydrogeology, the required number of model parameters may be significantly smaller than the number of model cells. Such considerations already form the basis for dimension reduction in hydrogeological parameter estimation (e.g., Doherty et al., 2010) and would naturally extend to particle filtering. For more complex priors, the size and shape of the geological features can be uncertain, too, so that geometrical parameters may need to be included.

In the following sections, we will provide a brief conceptual overview of the particle filter and its limitations and present a formal derivation and algorithmic implementation of the nested particle filter. With the framework established, we will discuss hyperparameterization as a tool to exploit latent structural simplicity in the numerical grid. To conclude, we will illustrate the performance of such an algorithm for different geological conceptualizations using synthetic examples and compare it to results obtained from an EnKF.

2. Theory and Methods

In this section we will provide a general introduction to particle filters and their limitations. Subsequently, we will present a formal derivation of the nested particle filter algorithm and a blueprint for its implementation. We conclude this section with a brief discussion of hyperparameterization and artificial parameter dynamics. First, however, we will introduce the nomenclature used in this study.

2.1. Nomenclature

In many filtering applications a distinction is made between *parameters* and *states*. Parameters, in this work denoted θ, are usually static model variables, such as hydraulic conductivities or porosities, properties which generally do not depend on other variables. States, represented by x, are often time-varying system variables depending on parameters and model forcings: Hydraulic heads, temperatures, or concentrations all are common examples. Observations are treated as a third variable type, y, and are generally measurements of states. For implementation-related purposes, all model variables of a type are combined into a vector and interpreted as coordinates of a point in high-dimensional space (*parameter space* or *state space*, respectively). Particles occupy one such point and thus represent a full set of their respective variables required for a model. Individual particles, or variables related to them, are assigned a superscript index in brackets, for example, $x^{(\text{index})}$. We use a point instead of an explicit index, for example, $x^{(\cdot)}$, if we refer to all indices of a given type. Specific time points are denoted by subscripts x_{time}, and ranges between a start and end point are represented by $x_{\text{start : end}}$. A ~ should be read as *sampled from*.

2.2. The Particle Filter

Like many other parameter optimization approaches, the particle filter is based on a probabilistic framework. It serves as a tool to represent uncertainty about (and inability to perfectly replicate) what we

assume is an unknown, but fundamentally deterministic natural system. In a probabilistic framework, variables are not assigned a single value (the porosity is 0.24) but can theoretically take on all mathematically possible values (the porosity is somewhere between $-\infty$ and $+\infty$). Since this formulation carries no specific information, knowledge or belief about the plausibility of different variables is specified exclusively by a pdf defined over all numeric values (plausible porosity values lie between 0 and 1 with these probability densities).

In such a framework, any change in knowledge of the system must be reflected in a change of the corresponding pdf. This process is formally described by Bayes' theorem:

$$p(A|B) = \frac{p(B|A)\,p(A)}{p(B)}.$$

(1)

Bayes' theorem updates the prior pdf $p(A)$, representing the belief about a given parameter predating an observation B, to its more informed posterior pdf $p(A|B)$. This operation is achieved by introducing the likelihood $p(B|A)$ of the observation B for a given value A. In order to ensure that the posterior probability density function $p(A|B)$ integrates to unity, the numerator has to be normalized by the marginal likelihood $p(B)$.

In the general case it is practically impossible to pursue this Bayesian inference analytically. As a consequence, one may either confine the analysis to a special case like the Kalman filter (Kalman, 1960), which is restricted to Gaussian priors and likelihoods so that an analytic solution is available, or to forfeit an analytic solution altogether and approximate the pdfs involved through an ensemble of weighted, deterministic Monte Carlo samples (e.g., Arulampalam et al., 2002).

The key idea behind the latter approach is that an unattainable analytic target distribution $p(A)$ may be approximated by a set of N deterministic independent and identically distributed (i.i.d.) samples thereof

$$\widehat{p}(A) = \frac{1}{N}\sum_{i=1}^{N}\delta_{A^{(i)}}(A) \overset{N\to\infty}{\to} p(A)$$

(2)

where $\delta_{A^{(i)}}(A)$ is the Dirac delta distribution centered about $A^{(i)}$, with $(i) = 1,\ldots,N$ being the particle's index. This superposition of Dirac delta distributions illustrates the surrogate properties of the Monte Carlo set: Instead of sampling $p(A)$, one could (at least in the limit of $N \to \infty$) equivalently draw with probability $\frac{1}{N}$ from a sufficiently large, preexisting set of realizations thereof.

Particle filtering, then, is the technique of recursively updating this set of realizations to retain its surrogate properties along otherwise intractable Bayesian inference operations. Provided that one could ensure that the ensemble of Monte Carlo samples remains representative of $p(A|B)$, it is theoretically possible to proceed indefinitely without the need to ever recover an analytic expression. These recursive updates are achieved with a procedure based on importance sampling; by adjusting the particles' retrieval weights w_i (initially uniform $\frac{1}{N}$), a set of particles drawn from one pdf could approximate a different distribution altogether. It should be intuitive that this reweighting is useful in the context of Bayesian inference: The posterior $\widehat{p}(A|B)$ may be approximated by the particles of the prior, $\widehat{p}(A)$. Employing the nomenclature of importance sampling, one can interpret $\widehat{p}(A)$ as the importance distribution, $\widehat{p}(A|B)$ as the nominal distribution and the normalized likelihood as the importance weight. This yields:

$$\widehat{p}(A) = \sum_{i=1}^{N} w_i \delta_{A^{(i)}}(A)$$

(3)

with weights w_i and $\sum_{i=1}^{N} w_i = 1$.

In theory, this operation could be repeated indefinitely, continuously adjusting the particle weights to reflect the latest posterior. In practice, however, this approximation will become increasingly inefficient and without an infinite ensemble size eventually only one particle will retain any significant weight. To counteract this *particle degeneracy*, it is common practice to resample the particles (Figure 1). This duplicates highly weighted particles but leads to a loss of variation in the ensemble (*sample impoverishment*).
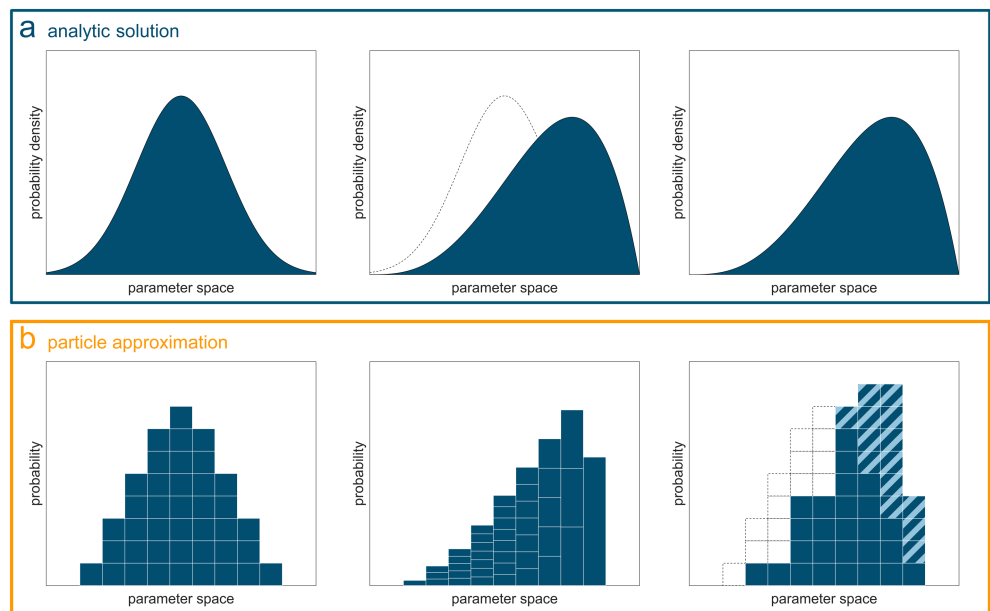
Figure 1. Conceptual scheme of Bayesian inference with a particle filter. The lower row (b) depicts a histogram representing the particle approximation of the analytic solution depicted in the upper row (a). First, the prior is approximated by equally weighted i.i.d. samples, visualized as blocks (left); during Bayesian updating, the particles are reweighted (illustrated by adjusting the block heights) so that the ensemble approximates the posterior pdf (middle); finally, new samples may be drawn from the weighted particles to yield a new, equally weighted ensemble, resetting the cycle (right).

In transient stochastic systems, this loss of diversity is counteracted by the forecast's random component, mapping duplicated state particles to slightly different positions. For the physically often static parameters, however, no such dynamic exists. Common approaches to reintroduce diversity to the parameters (*rejuvenation*) include MCMC steps (e.g., Chopin et al., 2013), jittering, or a combination of the two. In our study we will follow the second path by explicitly defining artificial parameter dynamics, which we will outline in section 2.4. First, however, we introduce the specific particle filter used in this study.

2.3. The Nested Particle Filter

The key idea behind a nested particle filter is to separate the filtering of states and parameters (Chopin et al., 2013; Crisan & Miguez, 2018). This is achieved in a *hierarchical* manner, whereby a single outer particle filter performs parameter inference, while several inner particle filters—one assigned to each parameter particle—conduct the state inference. This arrangement constitutes the nested structure and provides valuable information to both variable types: The state particle filters inherit their parent's parameter values and may treat them as fixed for the forecast, and the parameter particle filter can evaluate its particles' otherwise intractable likelihoods.

It may be worthwhile to note how the nested particle filter relates to the classic particle filter approach for joint state-parameter estimation in hydrology. By reducing the ensemble size of the inner filter to a single particle, one would retrieve the more commonly used particle filter with an augmented state vector (e.g., Montzka et al., 2011; Moradkhani et al., 2005). Larger inner ensemble sizes trade computational efficiency for better parameter likelihood estimates and can furthermore enable the use of more sophisticated particle filter algorithms (see van Leeuwen et al., 2019, for examples), which may not trivially extend across parameter space.

As explained in section 2.2, the introduction of new diversity is crucial to the reversal of ensemble collapse. Since numerical models are never perfect replications of reality, knowledge about states should degrade during the forecast between successive inference steps. In particle filters, this loss of information (increase of entropy, pushing the pdf toward uniformity) is generally represented by diffusion through an additive random error component, which—from a practical perspective—rejuvenates the particles. For the physically static parameters, however, no such natural source of diversity exists. To avoid nonreversible

ensemble collapse, we employ artificial parameter dynamics in this study. We note that this source of noise is introduced out of computational convenience rather than reflecting a physical process and account for it explicitly in the filter's derivation in the following section. The consequences of and opportunities created by artificial parameter dynamics are explored in more detail in section 2.4. In this section, we will first present the formal background of the nested particle filter with time-varying parameters, followed by its algorithmic implementation.

2.3.1. Formal Justification

The core objective of the algorithm presented in this study is the estimation of model parameters θ given a predefined geological characterization. Due to the introduction of noise through artificial parameter dynamics, these parameters become a time-varying quantity described by their trajectory through time $\theta_{0\,:\,m} := (\theta_0, ..., \theta_m)$. Further introducing a sequence of state observations $y_{1\,:\,t} := (y_1, ..., y_t)$ and a likelihood function $p(y_{1\,:\,t}|\theta_{1\,:\,m})$ allows for basic Bayesian inference:

$$p(\theta_{0:m}|y_{1:t}) = \frac{p(\theta_{0:m})p(y_{1:t}|\theta_{1:m})}{p(y_{1:t})} \tag{4}$$

where the discrete time step indices $c = 0,...,m$ and $s = 1,...,t$ illustrate the possibility for parameter and state dynamics, respectively, to operate on different time scales. In this instance we define parameter dynamics slower than its state counterpart, with each cycle c being comprised of L state time increments. The likelihood of the observation trajectory conditional on the parameter trajectory $p(y_{1\,:\,t}|\theta_{1\,:\,m})$ is a property that can be difficult to infer directly. Instead, one may expand the numerator by introducing the evolution of the predicted model state vector $x_{0\,:\,t} := (x_0, ..., x_t)$ to the likelihood and integrating it out immediately. We further recognize that the denominator is the integral of the numerator over the parameter trajectory, whose purpose is the normalization of the numerator, allowing us to formulate the expression to proportionality ($\propto$). This yields:

$$p(\theta_{0:m}|y_{1:t}) \propto \int p(\theta_{0:m})p(y_{1:t},x_{0:t}|\theta_{1:m})dx_{0:t} \tag{5}$$

Reformulating the likelihood in terms of the observation trajectory conditional on the trajectory of the predicted model states yields

$$p(\theta_{0:m}|y_{1:t}) \propto \int p(\theta_{0:m})p(y_{1:t}|x_{1:t})p(x_{0:t}|\theta_{1:m})dx_{0:t}. \tag{6}$$

Now, we may introduce a sequential solution by defining cycles $c = 1,...,m$, each composed of substeps $u = 1, ...,L$. The state time increments s are replaced by cycle-dependent subscripts $z = (c-1)L+u$. We further introduce the state forecast operator $f(x_z|x_{z-1},\theta_c)$, the likelihood function $g(y_z|x_z)$, and the artificial parameter dynamics $k(\theta_c|\theta_{c-1})$:

$$p(\theta_{0:m}|y_{1:t}) \propto \int p(\theta_0)p(x_0) \prod_{c=1}^{m} \left\{ k(\theta_c|\theta_{c-1}) \prod_{u=1}^{L} \left[f(x_z|x_{z-1},\theta_c)g(y_z|x_z) \right] \right\} dx_{0:z} \tag{7}$$

2.3.2. Algorithmic Implementation

We implement equation (7) as a nested particle filter. As previously elaborated, this approximates the analytically intractable parameter pdf by a set of N_θ particles drawn from the prior, which are then recursively reweighted, resampled, and mutated. Equivalently, the state pdf is approximated by N_θ ensembles of N_x state particles each. Superscripts $(n_\theta) = 1,...,N_\theta$ and $(n_x) = 1,...,N_x$ refer to specific particles of the parameter and state ensembles. Particles with multiple superscripts, for example, $x^{(n_\theta,n_x)}$, should be interpreted like this: State particle (n_x) of the inner filter belonging to parameter particle (n_θ). To aid intuition, we will discuss the implementation subdivided into its two main constituents: The outer particle filter for the parameters and the inner particle filters for the states.

2.3.2.1. Inner Particle Filters

Since each inner particle filter is assigned to a *parent* parameter particle, there are a total of N_θ such filters in parallel. In the following, we will consider only one of these filters as a blueprint for all others and hence assume its parent particle $\theta_c^{(n_\theta)}$ as given. We start by drawing N_x i.i.d., equally weighted initial state

particles from a suitable prior. This prior does not necessarily have to have an analytic form but could be obtained from, say, an interpolation of randomly perturbed observations or steady state simulations with a probabilistic error.

Then, the state particles are propagated individually via the forecast operator $f\left(x_z^{(n_\theta,n_x)}|x_{z-1}^{(n_\theta,n_x)},\theta_c^{(n_\theta)}\right)$ to the next time step at which observations are available. In practice, the forecast operator is constructed of two parts: First, the deterministic numerical groundwater model $M\left(x_{z-1}^{(n_\theta,n_x)},\theta_c^{(n_\theta)}\right)$, which uses $x_{z-1}^{(n_\theta,n_x)}$ as the initial conditions and $\theta_c^{(n_\theta)}$ as the parameters. Uncertain forcings or boundary conditions could also be considered at this point but are assumed known in this synthetic study for the sake of simplicity. The second part is an error term $\varepsilon_{\text{model}} \sim \mathcal{N}\left(0,\sigma_{\text{model}}^2\right)$, which, in our case, consists of a homogeneously applied Gaussian noise with mean 0 and variance σ_{model}^2:

$$x_z^{(n_\theta,n_x)} = M\left(x_{z-1}^{(n_\theta,n_x)},\theta_c^{(n_\theta)}\right) + \mathcal{J}\varepsilon_{\text{model}} \tag{8}$$

where $\mathcal{J} = (1,...,1)^T$ is a vector of ones with the same dimensions as the vector x. The choice of this error term is not trivial to determine and thus often left to the modeler's discretion. The choice of error we made above bears two advantages over other commonly used error types: (i) A homogeneous error may only alter large-scale absolute deviation in the model's state budget but has little influence on relative state distributions, thereby reducing interference of the model error on the characteristic flow responses of different parameter particles to a minimum: A sufficiently large spatially correlated error (for example, a Gaussian random field) could create artificial gradients in the head field by allowing the filter to repeatedly resample state particles with perturbations, which add or remove water in certain regions of the model domain, whereas a sufficiently large spatially uncorrelated error could cause numerical instability. (ii) Applying the error homogeneously, on the other hand, restricts the state perturbation onto a diagonal line of slope 1 in state space irrespective of its dimensionality, effectively rendering the error one-dimensional at the cost of limiting the algorithm's ability to *correct* state predictions.

Once all state particles are propagated and a new observation vector y_z is obtained, the particles can be weighted. First, determine the likelihood $g\left(y_z|x_z^{(n_\theta,n_x)}\right)$ of the observations conditional on the predictions for each updated state particles. Assuming independent normal observation errors $\varepsilon_{\text{obs}} \sim \mathcal{N}\left(\mu=0,\sigma_{\text{obs}}^2\right)$, the full likelihood function can be treated as a composite likelihood (e.g., Varin et al., 2011) calculated from the product of N_{obs} likelihoods—one for each individual observation. Each such component is evaluated according to

$$l_z^{(n_\theta,n_x,o)} = \frac{1}{\sqrt{2\pi\sigma_{\text{obs}}^2}} exp\left(-\frac{\left(x_z^{(n_\theta,n_x,o)}-y_z^{(o)}\right)^2}{2\sigma_{\text{obs}}^2}\right) \tag{9}$$

where superscript $(o) = 1,...,N_{\text{obs}}$ denotes an index of a specific observation in the observation vector y_z or the index of the corresponding prediction in the state vector $x_z^{(n_\theta,n_x)}$. The composite likelihood $\ell_z^{(n_\theta,n_x)}$ for each state particle is calculated as the product over all independent likelihood components:

$$\ell_z^{(n_\theta,n_x)} = \prod_{o=1}^{N_{obs}} l_z^{(n_\theta,n_x,o)}. \tag{10}$$

This composite likelihood can be used to determine the unnormalized weight of the corresponding state particle $x_z^{(n_\theta,n_x)}$. In this study, we use the prior as the proposal distribution (Chopin et al., 2013) but note that other choices are possible (e.g., van Leeuwen, 2010). Assuming that the prior weights are equal, we can normalize the likelihoods to retrieve normalized weights:

$$w_z^{(n_\theta,n_x)} = \frac{\ell_z^{(n_\theta,n_x)}}{\sum_{n_x=1}^{N_x} \ell_z^{(n_\theta,n_x)}}. \tag{11}$$

At the end of each data assimilation step, the state particles are resampled. We select N_x new state particles from the multinomial distribution weighted according to $w_z^{(n_\theta,\cdot)}$ (e.g., Gordon et al., 1993; Figure 1). After

each state particle $x_z^{(n_\theta, n_x)}$ is independently assigned an ancestral index $a = 1,...,N_x$, it copies the states of its respective ancestor $x_z^{(n_\theta, a)}$, and the state particle weights are reset to uniformity. The filter then proceeds to the next time step increment $z+1$ and repeats from equation (9).

2.3.2.2. Outer Particle Filter

Similarly to the state particle filters, we initiate the parameter particle filter by drawing samples from a suitable prior. Like the inner particle filters, the process starts with mutating the particles. As previously established, this is achieved through the artificial parameter dynamics $k(\theta_c | \theta_{c-1})$. The precise nature of these dynamics will be explored in detail in section 2.4, but assume for now that they yield a slightly mutated particle $\theta_c^{(n_\theta)}$ for each progenitor $\theta_{c-1}^{(n_\theta)}$:

$$\theta_c^{(n_\theta)} \sim k\left(\theta_c^{(n_\theta)} | \theta_{c-1}^{(n_\theta)}\right). \tag{12}$$

The next step is the evaluation of the likelihoods, which are extracted from composite likelihoods of the inner particle filters $\ell_z^{(n_\theta, \cdot)}$ before the state particles are resampled. Since each cycle is composed of L substeps $u = 1,...,L$, we form the product of the composite likelihoods over the current cycle, then form the Monte Carlo integral over the inner particle filter to retrieve the cycle's marginal likelihood $\mathscr{L}_c^{(n_\theta)}$ for each parameter particle:

$$\mathscr{L}_c^{(n_\theta)} = \frac{1}{N_x} \sum_{n_x=1}^{N_x} \prod_{u=1}^{L} \ell_z^{(n_\theta, n_x)}. \tag{13}$$

Assuming prior equal weights, this marginal likelihood may serve as an unnormalized weight for the parameter particles and yields normalized weights according to

$$w_c^{(n_\theta)} = \frac{\mathscr{L}_c^{(n_\theta)}}{\sum_{n_\theta=1}^{N_\theta} \mathscr{L}_c^{(n_\theta)}}. \tag{14}$$

Finally, the weighted parameter particles are resampled equivalently to the procedure of the inner particle filters. Each resampled particle inherits not only its ancestor's parameter values but also its ancestor's state particle filter. After resampling, the filter proceeds to the next cycle $c+1$ and repeats from equation (13). The pseudocode for the algorithm is provided in Figure S1 in the supporting information.

2.4. Artificial Parameter Dynamics

Introducing artificial parameter dynamics carries the explicit assumption that the model parameters evolve (randomly) in time. This approach is not a new idea (Doucet & Tadić, 2003; Li et al., 2014) and has been used in a variety of applications (e.g., Crisan & Miguez, 2018; Kitagawa, 1998; Reichert & Mieleitner, 2009). Some EnKF variations, such as the one suggested by Pathiraja et al. (2018), also employ explicit random parameter dynamics for active covariance inflation. Other EnKF inflation methods such as damping of the analysis (e.g., Hendricks Franssen & Kinzelbach, 2008; Keller et al., 2018) or linear scaling (Anderson & Anderson, 1999) achieve a similar effect deterministically. Even if no covariance inflation is used, practically all EnKF applications with joint parameter inference nonetheless render the parameters time-varying through the analysis step, although this is rarely stated. A notable exception to this is the Restart EnKF (Gu & Oliver, 2007), designed specifically to address this issue by restarting the simulation with a time-constant parameter set following the analysis step.

Rendering otherwise static parameters time varying is a potentially dangerous assumption: In general, these artificial dynamics will have no physical equivalent and may thus risk leading to inconsistencies between states and parameters. Time-varying parameters can, however, be justified in sufficiently dissipative settings, which tend to forget their history and thus prevent the accumulation of error (groundwater flow, following a diffusion equation, is generally dissipative).

We note that alternative approaches to address ensemble collapse in particle filters exist, such as the Metropolis-Hastings Markov chain Monte Carlo jumps employed by the SMC[2] algorithm (Chopin et al., 2013). This technique does not require the assumption of time-varying parameters and is invariant with
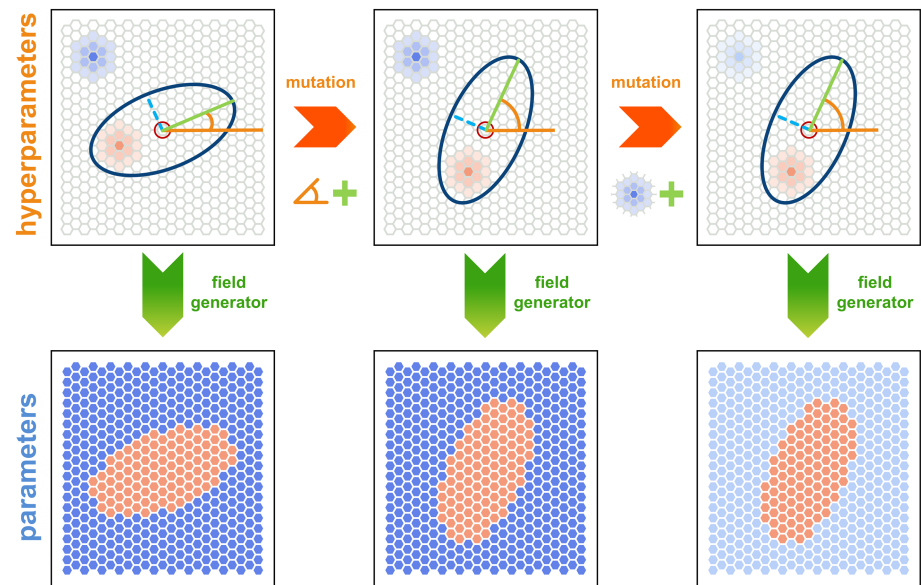
Figure 2. Conceptual scheme of hyperparameterization: A limited number of hyperparameters (primary axis length, secondary axis length, position, rotation, and two conductivities) can generate a full parameter field via a field generator (green arrows). Randomly mutating the hyperparameters (orange arrows) can efficiently alter aspects of the corresponding parameter field while maintaining the prescribed geological pattern—in this case, a lens embedded in a background sediment.

respect to the target distribution but comes at the cost of steadily increasing computational effort, rendering it ill-suited for online application.

Beyond these considerations, however, the introduction of artificial dynamics is attractive in several ways. Primarily, it permits the parameter particle ensemble to rejuvenate itself by introducing a form of covariance inflation. Cycles of resampling followed by slight mutations allow the ensemble to gradually explore regions of parameter space not sampled by the initial particles. Unfortunately, an efficient exploration of parameter space based on random mutation is still precluded by the curse of dimensionality. Hyperparameterization offers a way to alleviate this issue and provides additional interesting opportunities we will explore in the following.

2.4.1. Hyperparameterization

Recall that it is often possible to reduce the effective dimensionality of systems, for example, by principle component analysis (e.g., Wold et al., 1987). In the case of discretized subsurface models, dimension reduction is often achieved through the use of pilot points (e.g., Doherty et al., 2010; RamaRao et al., 1995), more conventional zonation into geological units (or facies) of shared properties (e.g., Carrera & Neuman, 1986; Franssen et al., 2009; Yeh, 1986), or interpolation techniques such as kriging, inverse distance weighting, or splines (Robinson & Metternicht, 2006; Yeh, 1986). These reduced parameter sets are able to generate the fully dimensional parameter field required by the model via a predefined set of rules. They can thus be interpreted as *hyperparameters*—parameters describing other parameters. The set of rules by which the hyperparameters relate to the full parameter field equivalent will subsequently be called a *field generator*. An illustration of this process is provided in Figure 2.

Adopting a hyperparameterization carries several distinct advantages. First, it allows the user to reduce the dimensionality of the model's parameterization to the degree of complexity demanded by the geological features rather than the numerical grid. The schema in Figure 2 illustrates that the number of hyperparameters is substantially lower than the number of parameters (six opposed to 460) and the parameterization becomes independent of grid resolution. More importantly, the field generator guarantees conformance with prescribed geological structures through construction by restricting the exploration of parameter values to a different, possibly simpler space. Conversely, all parameter fields, which cannot be created by the field generator, are no longer possible outcomes, which may be welcome in pursuit of geological realism. We

note that many more sophisticated object-based field generators such as HyVR (Bennett et al., 2019) or FLUVSIM (Deutsch & Tran, 2002) have been developed and could be adapted to interface with optimizers such as the one presented in this study.

2.4.2. Random Dynamics in Hyperspace

The reduced dimensionality of the auxiliary hyper (parameter)space permits a simpler exploration of the parameter space, for example, via Gibbs sampling (e.g., Kruschke, 2014). To implement the artificial parameter dynamics, we employ custom *mutation operators* designed to efficiently and randomly explore the hyperspace. These operators are largely based on Gibbs updates but may include more intricate actions designed to capitalize on predefined expectations of system response. Beyond the Gibbs updates, some of the more complex operators like *swapping node conductivities* (a reflection in hyperspace) find equivalents in mutation operators of genetic algorithms (*interchanging*, *swap*, or *Twors* mutation; see, e.g., Abdoun et al., 2012; Sivanandam & Deepa, 2007). Others like *removing lenses* or *adding nodes* (a removal or addition of hyperspace dimensions) are only rarely encountered in evolutionary algorithms, since most problem statements assume a fixed dimensionality of the parameter space (Lee & Antonsson, 2000). Formally incorporating hyperparameterization into the framework derived in section 2.3 could be achieved in two ways: (i) To entirely replace the parameters with hyperparameters, and interpret the field generator as a deterministic part of the numerical model (a hyperspace-based perspective), or (ii) to incorporate them as a *latent logic* underlying the artificial random dynamics in parameter space (a parameter space-based perspective). Both views are different interpretations of the same process.

2.5. EnKF Setups

To provide a comparison to more prevalent algorithms in hydrogeological literature, we also consider two different EnKF setups: A classical augmented state-vector EnKF and an augmented state-vector EnKF with GA. Both EnKF setups are initialized with parameter fields created by the field generators but subsequently operate in joint state-parameter space (for the standard EnKF setup) or the transformed joint state-parameter space (for the GA-EnKF setup). This approach is a widely used in subsurface hydrology (e.g., Tang et al., 2015; Zhou et al., 2011; Zovi et al., 2017), often with initial parameter fields generated by MPS. In our study, both setups are implemented with covariance localization following the approach of Hamill et al. (2001) with a length scale d_{loc} of 120 m:

$$
Z = \begin{cases}
1 - \frac{1}{4}\left(\frac{d_{ij}}{\lambda}\right)^5 + \frac{1}{2}\left(\frac{d_{ij}}{\lambda}\right)^4 + \frac{5}{8}\left(\frac{d_{ij}}{\lambda}\right)^3 - \frac{5}{3}\left(\frac{d_{ij}}{\lambda}\right)^2 & d_{ij} \leq \lambda \\
\frac{1}{12}\left(\frac{d_{ij}}{\lambda}\right)^5 - \frac{1}{2}\left(\frac{d_{ij}}{\lambda}\right)^4 + \frac{5}{8}\left(\frac{d_{ij}}{\lambda}\right)^3 + \frac{5}{3}\left(\frac{d_{ij}}{\lambda}\right)^2 - 5\left(\frac{d_{ij}}{\lambda}\right) + 4 - \frac{2}{3}\left(\frac{d_{ij}}{\lambda}\right)^{-1} & \lambda < d_{ij} \leq 2\lambda \\
0 & d_{ij} > 2\lambda
\end{cases}
\tag{15}
$$

where Z is the localization matrix, $\lambda = d_{\text{loc}}\sqrt{10/3}$, and d_{ij} is the Euclidian distance between cells i and j. Furthermore, both setups employ a Kalman gain damping factor of $\alpha = 0.5$. In combination, we calculate the Kalman gain K as follows:

$$
K = \alpha(C \circ Z)H^T\left(H(C \circ Z)H^T + R\right)^{-1}
\tag{16}
$$

where C is the augmented state vector covariance matrix, $\circ$ denotes elementwise multiplication, H is the operator extracting the observation locations, and R is the (diagonal) observation error covariance matrix. For the GA-EnKF, C and R are replaced by their respective transformed equivalents. The GA was implemented following the procedure prescribed in Schöniger et al. (2012), using the anamorphosis function extrapolation rule of Keller et al. (2018) and the observation error covariance transformation approach of Geppert (2015) and transforming both state and parameter spaces, back-transforming after each analysis step.

We further consider a hybrid setup, replacing the nested particle filter's inner filter with EnKFs for the states only, with $d_{\text{loc}} = 120\ m$, $\alpha = 0.01$, and no GA. This setup is designed to capitalize on the comparatively linear dynamics and Gaussian uncertainties in state space, while retaining the particle filter's flexibility for the
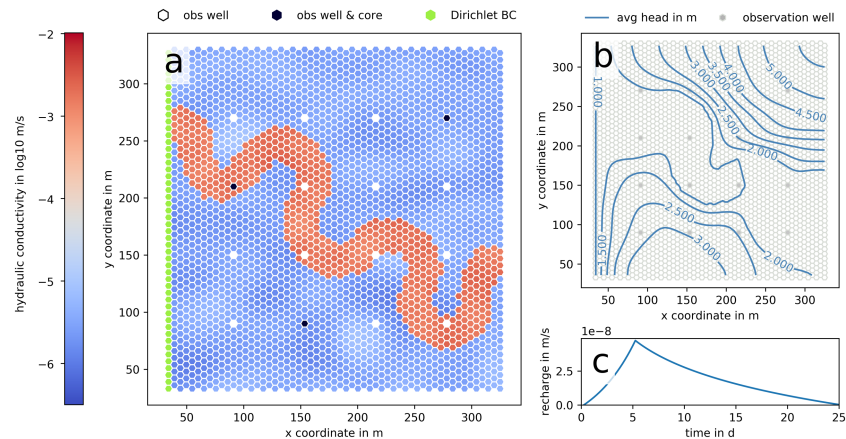
Figure 3. Schematic overview of the synthetic reference. The geology consists of a high-conductive paleochannel in low-conductive background material (a). Observations are taken at 16 locations, but the geology is assumed known at only three wells. The northern, eastern, and southern boundaries are no-flow, while the western boundary with a fixed head of 1 m serves as the system's only sink. The system's mean water table (b) is established by periodic recharge boundary condition (c).

more demanding parameter inference. The state particle likelihoods $\ell_z^{(n_\theta, n_x)}$ required for the outer filters (equation (13)) are extracted from the inner EnKFs just before the analysis step.

3. Synthetic Case Study

3.1. Model Setup

We test the algorithm in a synthetic, horizontal 2-D case study allowing for perfect knowledge of the reference *reality*. The model domain is tessellated with hexagonal grids to maximize structural isotropy while minimizing cell count. The studied aquifer is unconfined, and its synthetic geology features a high-conductive paleochannel embedded within low-conductive background material (Figure 3). Observations of hydraulic head are taken at 16 wells distributed regularly in a 4×4 grid by extracting the corresponding nodal values and adding uncorrelated Gaussian noise with zero mean and a standard deviation of $\sigma_{obs} = 0.02$ *m*. The geology was assumed known at only three of these wells. Flow in the system is driven by periodic recharge draining to the western fixed-head (Dirichlet) boundary. This causes the hydraulic heads to approach dynamic steady state after a sufficiently large number of cycles, implying that they are transient but time periodic. The initial state particles were generated by individually perturbing the initial state observations with the observation error and interpolating and extrapolating to all other cells. Further relevant model and filter properties are listed in Table S1 (supporting information). The algorithm was implemented in Python, interfacing with MODFLOW-USG (USG: unstructured grid; Panday et al., 2013) through the Python package FloPy (Bakker et al., 2016). Parameters to be estimated were hydraulic conductivities, and the observed and predicted states were hydraulic heads.

3.2. Field Generators and Mutation Operators

We test the nested particle filter algorithm for three hyperparameterizations corresponding to three different geological conceptual models against the same synthetic reference. The motivation for investigating these scenarios is that geological characterizations of field sites are essentially always imperfect. The hyperparameterizations considered are based on pilot points, lenses embedded in background sediment, and a meandering paleochannel embedded in background sediment. The latter two scenarios feature Gaussian heterogeneity independently within the background sediments and features (channels/lenses), which was deliberately mischaracterized: The correlation length of the synthetic test case is 3 times larger than its representation in the model used in data assimilation (Figure 4).

Following the concepts introduced in section 2.4, each scenario corresponds to a unique hyperparameterization with its own field generator and mutation operator. When called, the mutation operators carry out none (5% chance), one (55% chance) or two (40% chance) of a selection of actions which are listed in Table S2
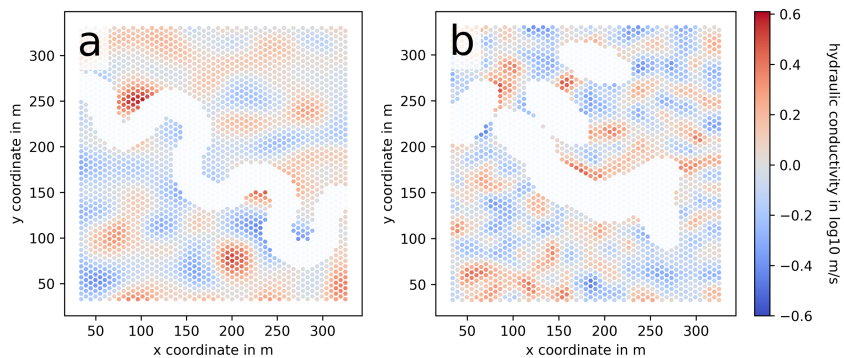
Figure 4. Mischaracterization of the synthetic reference's facies heterogeneity. The isotropic correlation length of the Gaussian heterogeneity in the synthetic reference's geology (a) is 3 times as large as assumed in the geological characterization (b).

(node-based), Table S3 (lens-based), and Table S4 (meander-based) in the supporting material. The no-mutation chance is added to ensure that the ensemble can remain within an optimum, once located.

3.2.1. Node-Based Field Generator

The first field generator is based on a classical interpolation approach: A selection of nodes is randomly placed in the model domain, each of which is assigned a hydraulic conductivity. Creation of the full parameter field follows from interpolation and extrapolation according to the nodes' and cells' spatial positions using inverse distance weighting with a high power factor (Shepard, 1968). This yields parameter fields very similar to what would be obtained through Voronoi tessellation (e.g., Aurenhammer & Klein, 2000). The hyperparameters are the number of nodes and the position and hydraulic conductivity of each node. Points of known geology are represented by fixed nodes with immutable, perfectly known hydraulic conductivity. On average, this scenario features 100 hyperparameters.

3.2.2. Lens-Based Field Generator

This field generator creates geological patterns based on distributing elliptical lenses of defined geometric properties over the model domain. Creation of the full parameter field is achieved by assigning each cell one of two sediment facies, depending on its placement with respect to the lenses. Parameters are then assigned based on two separate facies-specific conductivity maps for each particle, each of which is defined over the whole model domain with a specified mean, standard deviation, and variogram. The hyperparameters are the conductivity mean for both facies maps, the number of lenses, and the position, size, rotation, and aspect of each lens. Points of known geology are limits enforcing one of the two facies types, whose internal heterogeneity is—as described above—mischaracterized. On average, this scenario features 42 hyperparameters.

3.2.3. Meander-Based Mutation Kernel

The third field generator is conceptually similar to the lens-based field generator, but it generates meander-like features. First, the field generator generates a meander from the hyperparameters. The creation of the full parameter field is based on assigning each grid cell a sediment facies depending on whether it is located within or outside the meander. This scenario also employs two facies maps, with the same fundamental mischaracterization of internal heterogeneity. The creation of the meander from the hyperparameters is a multi-step process illustrated in Figure S2 (supporting information). Hyperparameters are the mean hydraulic conductivity for both facies maps and several parameters specifying the meander: The position and orientation of the start point and end point, the meander phase shift, the meander width, and the channel width. Points of known geology are represented as guaranteed adherence to one of the two facies types. Due to the complex nature of this field generator, an iterative procedure is required to ensure conformance with points of known geology. After mutation, the meander's phase shift is randomly adjusted until a conforming facies distribution is found. If no such move is possible, the proposed mutation is rejected. This scenario features nine hyperparameters.

3.3. Computational Setup

The algorithm was tested on two desktop computers using a 64-bit Windows 7 OS, with Intel® Core™ i7-2600 CPU @ 3.4GHz and Intel® Core™ i7-3770 CPU @ 3.4GHz processors and 8 GB of RAM. The simulation of

the full synthetic calibration period of 750 days for the nested particle filter scenarios and the hybrid scenario was achieved using 200 parameter particles with five state particles each. For the EnKF scenarios, the ensemble was initialized with 1,000 joint state-parameter realizations.

Computation times for the nested particle filter varied depending on the hyperparameterization used: The node- and lens-based scenarios required about 18 core hours for a full run, well below the available time in a field application. The meander-based hyperparameterization runs required significantly more time due to the iterative nature of its mutation operator, demanding about 49 core hours of computation time. The simulation times reported above were obtained for a sequential implementation of the algorithm. A parallelized setup was tested but dismissed as ineffective after increasing the computation time by a factor of 4 due to overhead. Each scenario was repeated with 10 different *random seeds* to test the reproducibility of the results obtained. A full run for the EnKF setups required about 18 core hours each. This time requirement was constant for all hyperparameterizations, since the field generators were only called once during the initialization of the algorithm.

4. Results and Discussion

We first report the parameter fields obtained at the end of the data assimilation period. Subsequently, we investigate the head and conductivity discrepancies between the ensemble results and the synthetic reference. Finally, we investigate the deterministic prediction performance of the obtained parameter fields and conclude with a discussion of the results.

4.1.. Optimized Parameter Fields

Figure 5 depicts a selection of parameter fields at different points in time as obtained by the nested particle filters. The first column (Figures 5a, 5d, and 5g) illustrates selected samples from the initial parameter fields, created by the different field generators with randomized hyperparameter input. The second column (Figures 5b, 5e, and 5h) shows the expectation of the final cycle's ensemble for selected random seeds. The third column (Figures 5c, 5f, and 5i) illustrates the average parameter field over all 10 random seeds obtained at the end of the calibration period. A comparison with Figure 2 reveals that even the structurally mischaracterized field generators (*node-based* and *lens-based*) seem to evolve their hyperparameters in a way allowing for the creation of a structure functionally similar to the reference meander.

Results for the hybrid nested particle filters with inner EnKFs match the findings of these scenarios (Figures S3 to S5, supporting information). This can be explained by collapse of state uncertainty in the first quarter of the simulation time (Figures S6 to S8, supporting information), after which there are no more functional differences between the two particle filter setups.

Note that the mean parameter fields (Figure 5, second column; Figures S3 to S5) are more *crisp* than one would expect from an ensemble-based optimizer. This results from a weight-based ensemble collapse of the outer filter during every calibration cycle, right before new diversity is created by the artificial parameter dynamics. Contributing factors are the parameter cycle length, the number of observation points, the observation error standard deviation, the sensitivity of the predictions to changes in the parameter values, and the magnitude of parameter changes proposed by the artificial parameter dynamics. While this study's model and setup resulted in repeated collapses, other systems and sites may be less prone to degeneration and could preserve more of the probabilistic information. For interested readers, a quantitative analysis of this collapse considering the setup of the nested particle filter is provided in Appendix A. It is worth noting, however, that the optimization of the parameter fields proceeds despite the loss of most probabilistic information. This becomes evident in the posterior parameter fields' tendency to express hydraulic conductivity distributions functionally similar to the synthetic reference field (Figure 5, third column).

Figure 6 summarizes the results for the two EnKF scenarios. The first, second, and third rows correspond to the node-based, lens-based, and meander-based scenarios. Columns one to four illustrate results of the standard EnKF scenarios, columns five to eight those of the GA-EnKF scenarios. The first and fifth columns show individual parameter realizations at the end of the assimilation period, the columns to their right the corresponding ensemble average. The third and seventh columns illustrate the final standard deviation of the hydraulic conductivities, and the fourth and eighth columns the standard deviations of the hydraulic heads at the end of the assimilation period.
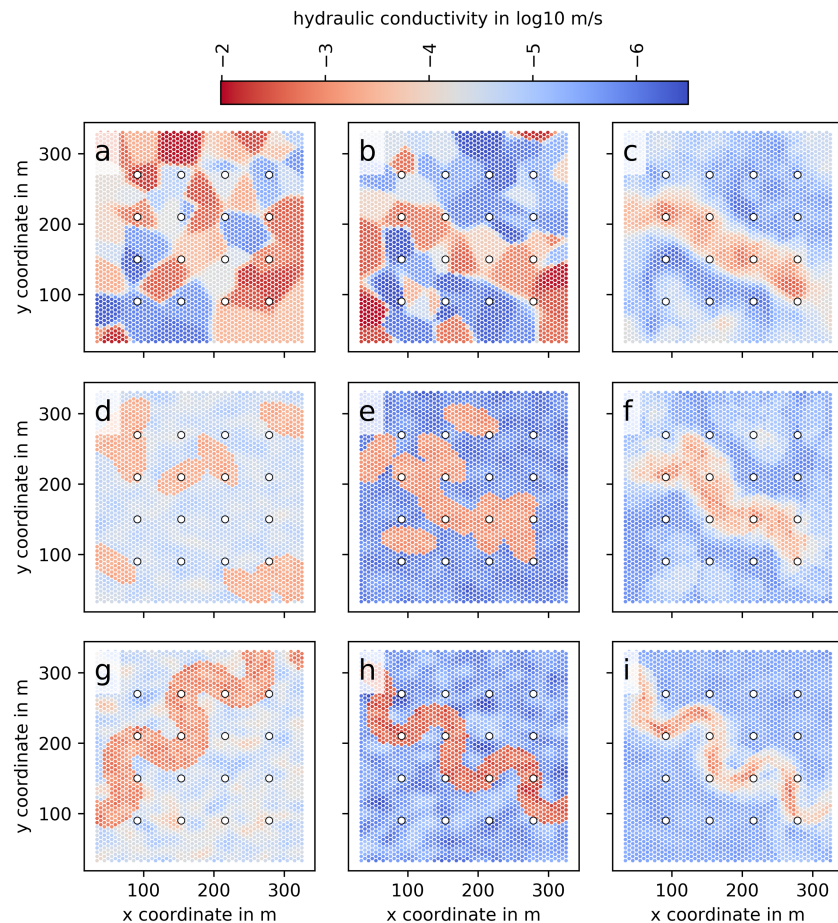
Figure 5. Selected parameter fields for the three field generators: Node-based (a–c), lens-based (d–f), and meander-based (g–i); the first column (a, d, and g) illustrates selected initial parameter particles; the second column (b, e, and h) shows expected final parameter fields for selected random seeds; the third column (c, f, and i) depicts mean final parameter fields through all 10 random seeds.

Inspecting the optimized parameter fields of the EnKF scenarios, we observe that both scenarios have identified the general structure of the reference conductivity field. Unfortunately, however, we also note that the realizations of the final ensemble (Figure 6, columns 1 and 5) have all but lost the geological features defined in the initial field generation. In the standard EnKF, remnants of the initial structural features still remain (Figure 6, column 1), whereas the GA-EnKF seems to have completely erased the structural differences. The standard deviations of parameter uncertainty (Figure 6, columns 3 and 7) for the lens-based and meander-based hyperparameterizations (Figure 6, rows 2 and 3) further indicate regions in the southeastern quadrant for which the ensemble seemed to have collapsed against the prescribed lower bound of the hydraulic conductivity, in turn also collapsing the local state uncertainties (Figure 6, columns 4 and 8). This effect is more pronounced for the standard EnKF than for the GA-EnKF scenario. Finally, we note that for both EnKF scenarios the node-based hyperparameterization appears to retain the highest state and parameter uncertainty at the end of the data assimilation period.

4.2.. Parameter Estimation Performance

We now analyze the optimized parameter fields and their resulting state predictions. The perfect knowledge of the synthetic model allows us to investigate the deviations from the reference even at locations where otherwise no information would be available. In order to avoid overinterpreting statistical artifacts, we investigate the parameter and state fields averaged over all 10 random seeds. Figure 7 illustrates the state deviations (Figure 7, first column) and the parameter deviations (Figure 7, second column) for the node-based (Figure 7, first row), lens-based (Figure 7, second row), and meander-based (Figure 7, third row)
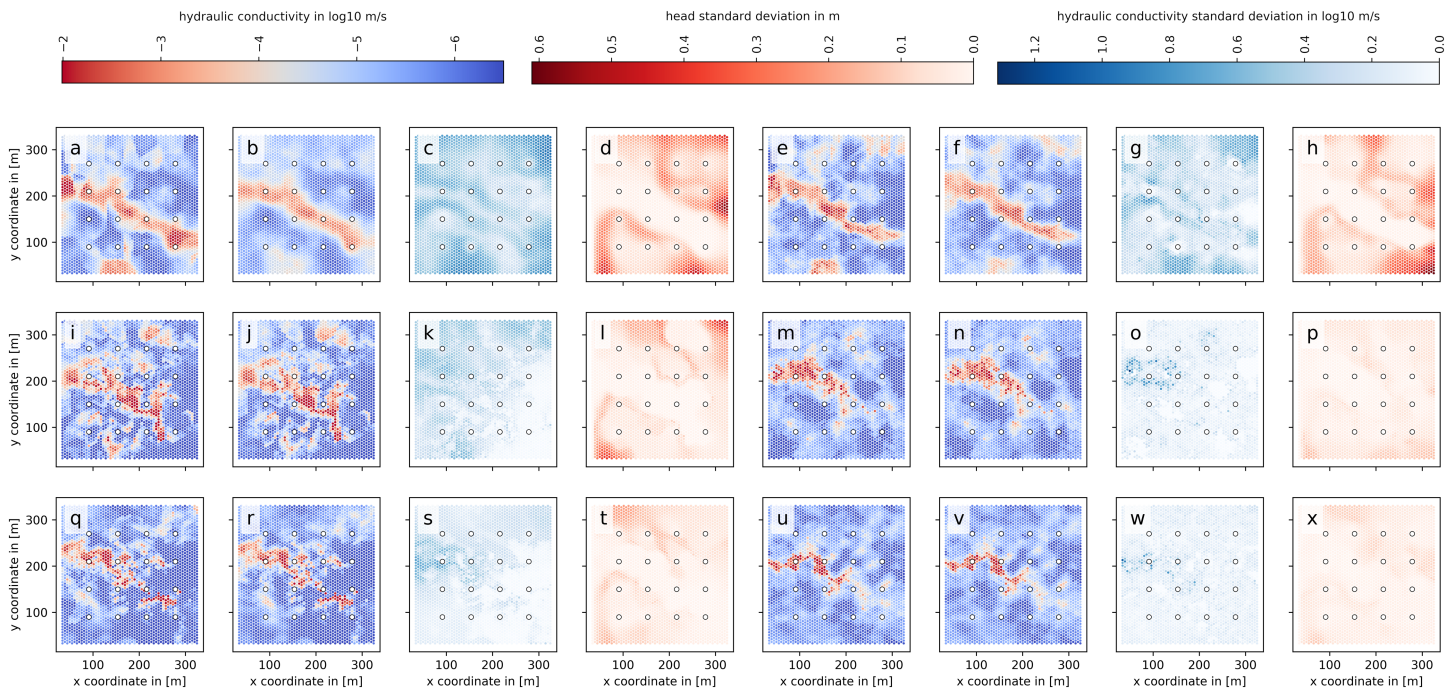
Figure 6. Selected parameter fields (columns 1, 2, 5, and 6) and the standard deviations of conductivities (columns 3 and 7) and heads (columns 4 and 8) obtained by the EnKF for the three field generators at the end of the data assimilation period: Node-based (row 1), lens-based (row 2), and meander-based (row 3); the first column (a, i, and q) illustrates selected realizations of the standard EnKF scenario; the second column (b, j, and r) shows the ensemble mean of the standard EnKF scenario; the third (c, k, and s) and fourth column (d, l, and t) illustrate the standard deviations of parameters and heads; the fifth (e, m, and u), sixth (f, n, and v), seventh (g, o, and w), and eighth (h, p, and x) are the corresponding entries of the GA-EnKF scenario.

scenarios. The corresponding standard deviations are illustrated in Figure S9, their counterparts for the two EnKF scenarios are provided in Figures S10 to S13, and deviations for the hybrid nested particle filter are illustrated in Figure S14 (supporting information).

Comparing the state deviations of the nested particle filters (default and hybrid setups), we note a few interesting aspects: All three scenarios managed to reproduce the state observations along the synthetic meander relatively well. This is a result of the high correlation between observed heads within the meander and results in a heavy likelihood penalty should this continuous, high-conductive structure be broken or not created. We further note that on average both the node-based and lens-based scenarios identify conductivity fields, which reproduce head observations at the observation points well but deviate from the reference between measurement points. It is worth noting that the node-based hyper-parameterization develops larger state deviations in between observation points than the lens-based case. A possible explanation is that the node-based approach allows an almost continuous adjustment of hydraulic conductivity, whereas the lens-based method is essentially binary. The meander-based case succeeds in reproducing observations within the meander but deviates substantially for outlying observation wells.

To explain this inability to reproduce the state observations farther into the background facies, it may help to consider the parameter field deviations in the light of the mischaracterization of heterogeneity (Figure 4). While the optimization algorithm identifies the correct mean background conductivity, the state mismatch suggests that this mischaracterization results in an underprediction of states compared to the reference. This phenomenon is likely rooted in an inability to replicate the effects of the reference's internal heterogeneity patterns. The lens-based scenario is subject to the same fundamental mischaracterization but the independent placement of high-conductive lenses at strategic locations within the background facies allows compensating this structural error. A closer look at the internal background heterogeneity reveals that the optimizer places clusters of high-conductive lenses along the northeast or southwest of the model domain in the node-based and lens-based scenarios. We believe that these placements compensate the unresolved internal variability of the background field. In the node-based scenario, we find parameter fields with features similar to
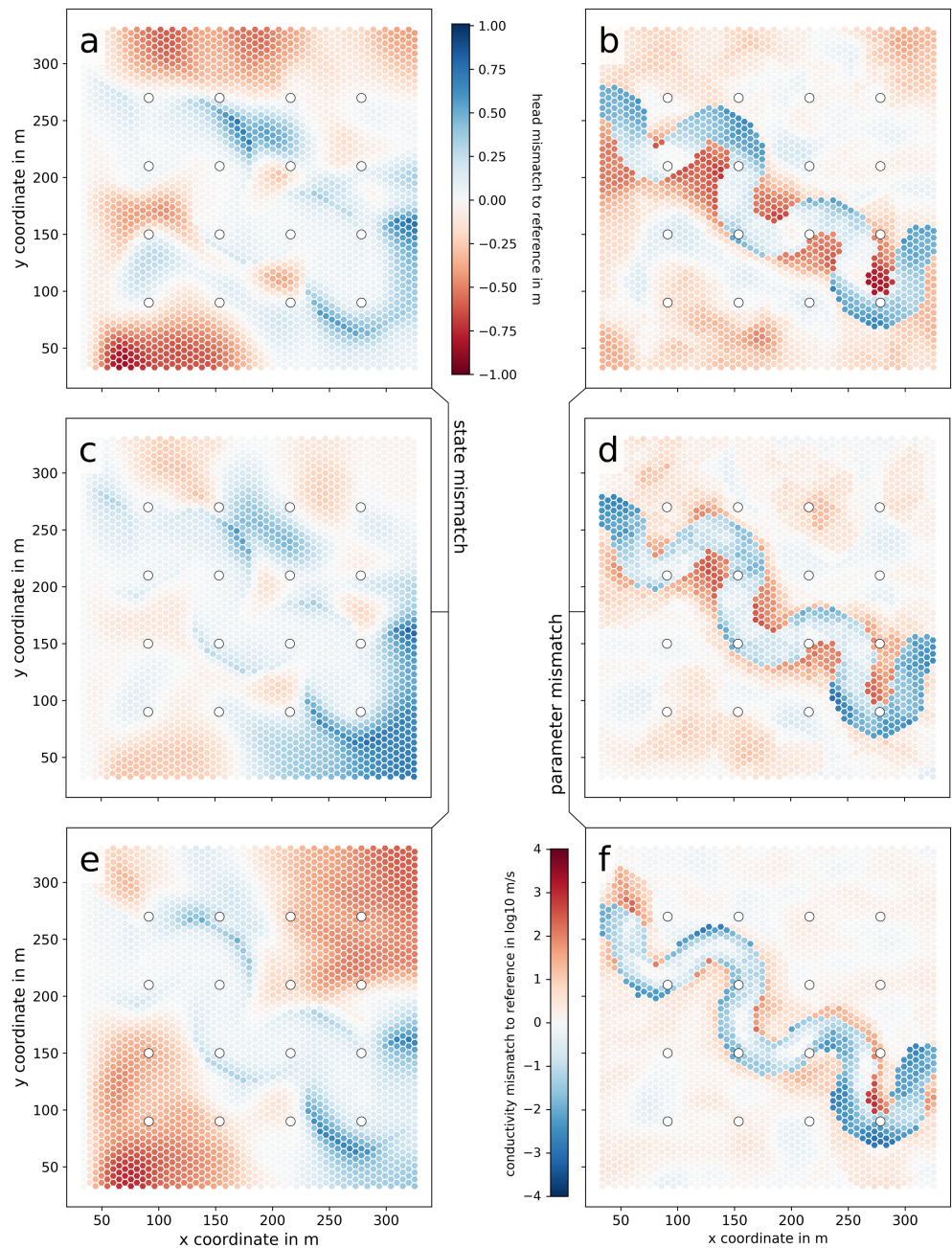
Figure 7. Mismatch between the synthetic reference and ensemble mean at the end of the data assimilation period for hydraulic heads (a, c, and e) and hydraulic conductivities (b, d, and f), averaged over 10 random seeds for the nested particle filter scenarios. Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyper-parameterizations. Mind that the relation of colors to quantities is reversed between the state mismatch (a, c, and e) and the parameter mismatch (b, d, and f) columns. This was a deliberate choice to visually underline the common relation of parameter underestimation to state overestimation, and vice versa.

the lens-based case, despite much greater freedom in the distribution of conductivities. Recurring features include high-conductive clusters in the southwest, an underestimation of conductivity between the second, third, sixth, and seventh observation wells (counting left to right, top to bottom) or in the southeast. These recurrences suggest that a number of features of the reference conductivity field may not be sufficiently informed by the head observations and that there likely are more stable, or more easily identifiable, equivalent solutions to the reference.

4.3.. Predictive Performance

As the gradual improvement of model predictions is a major objective of the nested particle filter, it may be worth investigating prediction performance without data assimilation and the explicit assumption of time-varying parameters. To do so, we extract the expected parameter values at the end of each cycle and repeat a deterministic simulation over 90 recharge cycles for both the particle filter and EnKF scenarios. After a dynamic steady state has been established, we determine the average root-mean-square error between the states $x_{\text{predicted}}$ and the corresponding synthetic reference $x_{\text{synthetic}}$ at the observation points over the final recharge period $s = 1,...,100$. We refer to this quantity as the prediction root-mean-square error (pRMSE) and calculate it according to

$$\text{pRMSE} = \frac{1}{N_{\text{obs}}} \sum_{o=1}^{N_{\text{obs}}} \sqrt{\frac{1}{100} \sum_{s=1}^{100} \left(x_{\text{predicted}}^{(o,s)} - x_{\text{synthetic}}^{(o,s)} \right)^2} \tag{17}$$

We furthermore calculate the percentage bias of the predicted states according to

$$\text{pbias} = \frac{1}{N_{\text{obs}}} \sum_{o=1}^{N_{\text{obs}}} \sum_{s=1}^{100} \frac{x_{\text{predicted}}^{(o,s)} - x_{\text{synthetic}}^{(o,s)}}{x_{\text{synthetic}}^{(o,s)}} \tag{18}$$

The results of this evaluation are visualized in Figure 8.

On average, we observe an increase in predictive performance across all hyperparameterization scenarios, albeit with varying degrees of stability and fidelity. For the nested particle filter, the final pRMSEs of the node-based, lens-based, and meander-based scenarios fall approximately between 0.05 and 0.2 m, 0.1 and 0.4 m, and 0.15 and 0.3 m, respectively. The final relative biases of the predictions for the node-based, lens-based, and meander-based scenarios fall between ±10%, −10% to +20%, and ±10%, respectively. Initially, all three scenarios display—for most random seeds—a positive bias but vary in the magnitude of their respective pRMSEs. The discrepancies in initial pRMSEs can be explained by the nature of the different hyperparameterizations, and how much they are constrained by prior geological information. The node-based scenario is barely constrained, with the contribution of the three nodes of known conductivity lost among the effect of 50 randomized nodes. The lens-based scenario featured the highest prior pRMSEs, as a replication of the reference's state distribution requires a specific arrangement of lenses unlikely to emerge by chance. The meander-based case displayed the lowest prior pRMSEs, owing to the strong constraint placed on the meander's path—it has to pass through a specific well in the northwest—and the overall low number of hyperparameters, increasing the chances of creating a prior parameter particle near an optimum. It is worth noting that we are free to adapt the ensemble size during the optimization process. This would, for example, allow us to initialize the filter with a large ensemble of prior particles, identify the best among them through weighting, and then resample (and proceed with) a smaller ensemble.

Aside from long-term optimization, however, we observe short-term deviations from prevailing optima. The occurrence of such instabilities is not surprising given the probabilistic nature of the artificial parameter dynamics but is exacerbated by the limited cycle length. Across the scenarios, we observe a qualitative drop in optimization stability from the node-based, through the lens-based, to the meander-based hyperparameterization. The relative stability of the node-based scenario is based in the absence of hyperparameters with a large-scale impact on the conductivity field, effectively limiting the *damage* from resampling a subpar mutation. Conversely, the instability of the meander-based scenario (observed in both the default and hybrid setup) can be explained by the strong interaction between its hyperparameters, none of which can be altered without large-scale consequences. This also provides an explanation for the apparent presence of local optima, which become evident by prolonged sequences during which proposed mutations are repeatedly rejected (e.g., Figure 8 c, random seed 5). The lens-based scenario appears to be as stable as the node-based one, but occasionally expresses larger deviations, most likely because some of its hyperparameters have a larger impact on the resulting parameter field. We note that the magnitude of this instability could be alleviated by reducing the strength of the mutations over time, a common practice in methods like simulated annealing
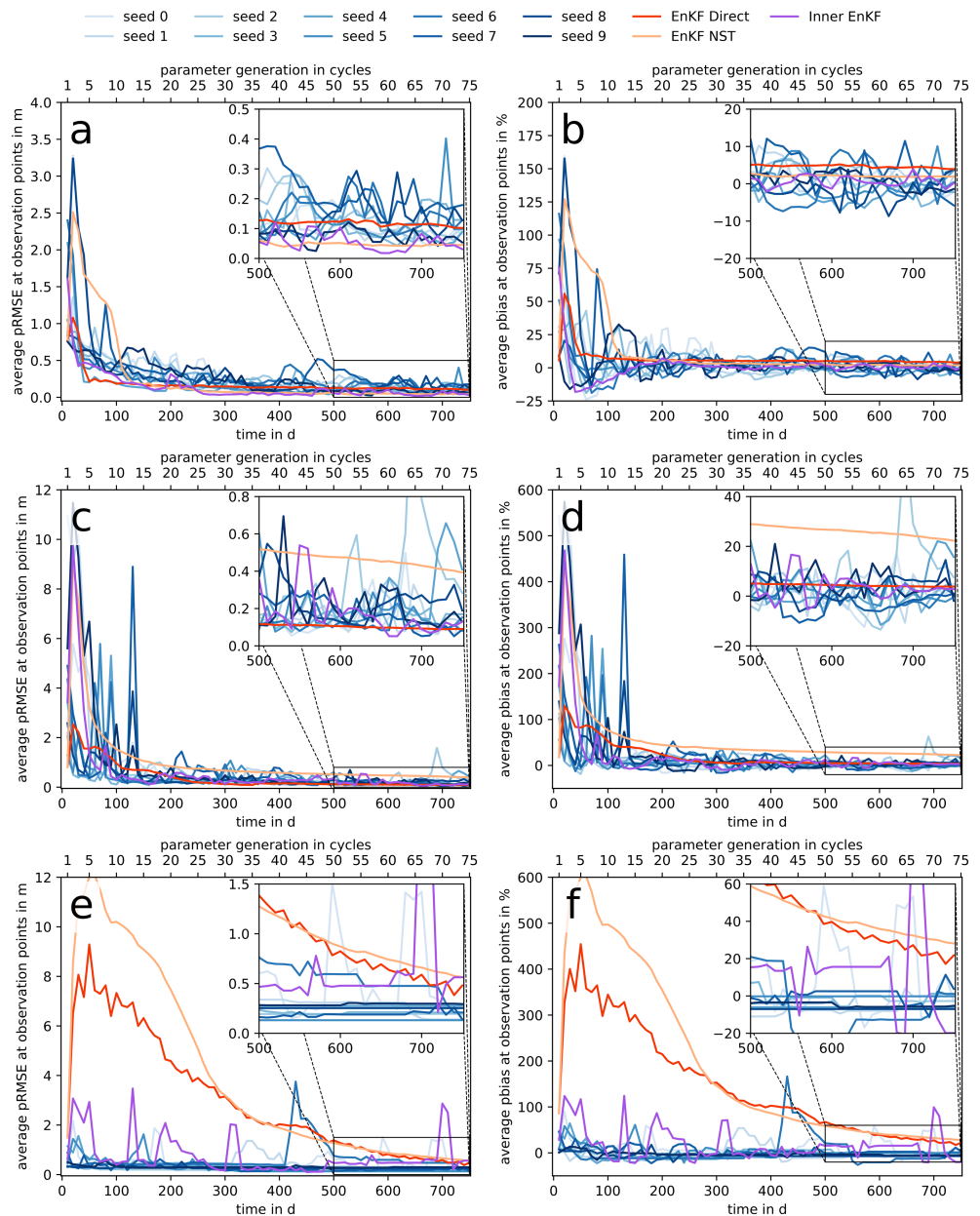
Figure 8. Development of pRMSE (a, c, and e) and percentage bias (b, d, and f) through time for the node-based (a, b), lens-based (c, d) and meander-based (e, f) scenarios and 10 different random seeds, the Inner EnKF scenario, and the two EnKF scenarios.

(e.g., Kirkpatrick et al., 1983). In general, the hybrid scenarios with inner EnKFs yield similar performance to the other nested particle filter setups.

For the full EnKF scenarios, optimization performance was generally more stable. Both standard EnKF and GA-EnKF showed similar performance, although the standard EnKF seemingly optimized quicker. More significant deviations in performance were found between the different geological scenarios. Since the EnKF scenarios only uses the field generators during the generation of the prior ensemble and then updates the ensembles as if they were multi-Gaussian, the observed differences must originate from variations in the properties of the prior covariance matrices. Extracting the eigenvalues of these matrices (Figure S15), we note that the node-based covariance matrix has the highest, the meander-

based case the lowest, and the lens-based case an intermediate effective dimensionality. For the node based scenarios—offering the largest initial parameter uncertainty with the least spatial correlation—optimization performance was quick and precise, matching or surpassing the nested particle filters. For the lens-based hyperparameterization, the standard EnKF appears to outperform the GA-EnKF in both pRMSE and relative bias and matches the best of the nested particle filters. The GA-EnKF performs worse than all nested particle filters, although Figures 8c and 8e suggest steady improvement until the end. This is likely owed to filter collapse, as Figures 6f and 6h show that the parameter uncertainty collapsed for both scenarios toward the end. A similar collapse seemingly occurred for the meander-based hyperparameterization (Figures 6j and l): Both EnKF scenarios feature worse pRMSEs and relative biases than the nested particle filters, although steady improvements are observed until the end. This ensemble collapse could be remedied with active covariance inflation.

4.4.. Discussion

The results obtained suggest the nested particle filter can optimize even highly complex hyperparameterizations, but different degrees of optimization performance divide the investigated scenarios in three classes: An overflexible setup (node-based), a balanced setup (lens-based), and an overconstrained setup (meander-based).

Despite given the correct geological structure, the meander-based scenario performed poorest. Unable to compensate for a fundamental mischaracterization of the background facies heterogeneity, we found that the hyperparameterization could not generate parameter fields adequately replicating the synthetic reference's hydraulic head distribution. Furthermore, high dependence among the hyperparameters suggests the presence of numerous local optima and unstable optimization performance.

Despite featuring an erroneous geology and being subject to the same fundamental mischaracterization of heterogeneity, the lens-based hyperparameterization performed favorably. Its optimized parameter fields revealed that the lenses can arrange themselves to form features functionally similar to the reference's high-conductive meander and even compensated for the mischaracterized heterogeneity. While the prediction performance did not fully measure up to the node-based scenario, its geological characterization resulted in smaller deviations from the reference at points where no information would otherwise be available.

Unencumbered by constraints of complex geology and with many virtually independent hyperparameters, the node-based scenario permitted the most flexible adjustment of the conductivity field. While this scenario best replicated the reference's state observations, a comparison to the synthetic reference's latent states revealed substantial deviation in-between observation wells. This suggests that the quality of the fit was at least partially afforded by over-parameterization.

Results for the hybrid nested particle filter with inner EnKFs were similar to those obtained with the normal nested particle filters, owed to an irreversible collapse of state uncertainty, which eliminated differences between the two setups early on. In a real setting with nonnegligible forecast errors and uncertain forcing, we expect that these sources of entropy would retain more uncertainty and permit the hybrid filter to better leverage the efficiency of its inner EnKF.

For the full EnKF reference scenarios, we note that the best performance was achieved for the node-based scenario, closely followed by the lens-based scenario. The lens-based and the meander-based scenarios both suffered from ensemble collapse. This suggests that the EnKF optimization performs best if the initial ensemble encloses a large volume of parameter space and may perform poorly in scenarios with larger spatial correlations (lens-based and meander-based). Where the ensembles did not collapse, the EnKF can yield parameter fields on par to or better than those from the nested particle filters. Unfortunately, the EnKF's update procedure abandons the support of the geological prior in the process: The EnKF did not sufficiently honor the initial geological features in all scenarios considered, yielding posterior parameter fields strongly deviating from the prescribed geology. This suggests that if conformance with a geological prior is of the essence, the use of data assimilation and calibration schemes, which implicitly rely smoothness or regularity assumptions in conflict with the prior's support (e.g., Gaussianity for the EnKF), may not be advisable. In the present study, we suggest geology-obeying hyperparameterizations, but MPS with geologically realistic training images may also work.

Summarizing the results, we find that the nested particle filter can successfully optimize hyperparameters whose relation to the state response is highly nonlinear or discontinuous. The scenarios investigated in this study revealed a number of important aspects: While conformance to arbitrary geology can be enforced with ease, it is important to leave the algorithm sufficient flexibility to compensate for potential (and in practice inevitable) mischaracterizations of geology. In the case of the meander-based field generator, further hyperparameters adjusting the mischaracterized heterogeneity might have improved the performance. A general way to achieve this is by providing options to locally adjust parameter fields, as shown by the lens-based and node-based scenarios. Acknowledging that this compensation for structurally wrong parameters (*parameter surrogacy*; Doherty & Christensen, 2011) is often undesirable, its occurrence may nonetheless reveal fundamental conceptual errors.

5. Conclusions

In this study, we explored the use of a nested particle filter framework, a generalization of the classic particle filter for joint state-parameter estimation, for real-time parameter optimization in distributed groundwater models. We made use of hyperparameterized field generators to reduce the dimensionality of the optimization problem and to guarantee conformance to a prescribed geology throughout the optimization process, using variance inflation through artificial parameter dynamics to rejuvenate the parameter particles. Examining the performance for three simple field generators, we identified versatile hyperparameterization as a prerequisite for the algorithm's success in mischaracterized settings.

We then compared the optimizer to two classic EnKF setups—a standard state-vector augmented EnKF and a state-vector augmented EnKF with GA, both initialized with samples from the different field generators. While their optimization performance can be equal or even superior with sufficient initial variability in the parameter ensemble, none of the scenarios considered sufficiently preserved the prescribed geological structures. This is a consequence of the EnKF updating its ensembles as if they were multi-Gaussian in combination with the highly non-Gaussian support of the parameter priors in grid parameter space. Where geological fidelity is essential, we thus suggest to combine field generators (e.g., MPS, object-based generators) with optimization routines capable of reliably navigating the prior's support. As such, our results make us believe that the nested particle filter—if adequately (hyper)parameterized—could constitute a valuable complement to other real-time parameter estimation methods, particularly in scenarios where the conservation of complex geological features is critical.

We would like to remark that it would theoretically also be possible to have an EnKF operate on a joint state-hyperparameter space. This would likely impose a number of restrictions on the design of the field generator to ensure sufficient regularity and possibly require postanalysis sanity checks to ensure consistency with geological conditioning data. It is furthermore to be expected that the linear relation between state response and hyperparameters would be weakened. The design of such EnKF-friendly hyperparameterizations was beyond the scope of the present study. It remains open to what degree the efficiency of the EnKF justifies these self-imposed restrictions, but we note that this could be an interesting avenue for future research. Hybrid solutions such as the inner EnKF scenarios considered in this study are also a promising way to leverage the EnKF's efficiency for the state updates while retaining the particle filter's generality for the parameter updates.

A limitation of this study is the lack of uncertainty retained in the particle ensemble, a consequence of the filter's ensemble collapse, although the framework is in principle capable of sustaining such information and may do so in different settings or with different likelihood functions. A further limitation is the use of artificial parameter dynamics, which may preclude an application in systems with longer memory if physical consistency between states and parameters is of concern. Continuing research could address these issues by adapting the cycle length across the optimization period, adjusting mutation magnitude and initializing the filter with a larger parameter ensemble size. Applications in less dissipative settings, such as the simulation of remediation efforts, could be approached by employing a nested particle filter with a different rejuvenation mechanism such as SMC^2. The arbitrary nature of parameter dynamics also provides an adaptive interface to other field generators, for example, MPS. The flexibility of the nested particle filter framework encourages experimentation with different numerical models, parameter dynamics, states, and

(hyper)parameters. The algorithm presented herein extends trivially to 3-D, as none of its elements places a restriction on the spatial dimensionality of the numerical model.

Appendix A: Investigation of the Ensemble Collapse

In section 4.1. we noted that the outer particle filter in this study collapses during each parameter cycle. In pursuit of uncertainty information pertaining the parameter distribution, this is of course undesirable. In this appendix, we investigate the ensemble collapse with a back-of-the-envelope calculation to identify the reasons for this persistent collapse and how the framework would have to be parameterized to allow for a better conservation of parameter uncertainty.

As we established in section 2.2, the ensemble collapses if only one particle retains any significant nonzero weight. According to equation (14), the parameter particle weights are proportional to the marginal likelihood derived over the arithmetic mean of its inner particle filter's likelihoods. We summarize:

$$w_c^{(n_\theta)} \propto \mathcal{L}_c^{(n_\theta)} = \frac{1}{N_x} \sum_{n_x=1}^{N_x} \prod_{u=1}^{L} \prod_{o=1}^{N_\mathrm{obs}} l_z^{(n_\theta, n_x, n_o)} \tag{A1}$$

Since we want to learn about the relations required to prevent degeneracy, it may be useful to consider a simplified case in which $l_z^{(n_\theta, n_x, n_o)} = \overline{l^{(n_\theta)}}, \forall n_x, n_o, u$, which leads to the following simplification:

$$w_c^{(n_\theta)} \propto \frac{1}{N_x} \sum_{n_x=1}^{N_x} \prod_{u=1}^{L} \prod_{o=1}^{N_\mathrm{obs}} l_z^{(n_\theta, n_x, n_o)} = \overline{l^{(a)}}^{LN_\mathrm{obs}} \tag{A2}$$

The prevention of the parameter filter collapse requires that at least some parameter particles retain a nonzero weight ratio. Omitting the time subscripts from the notation for the sake of simplicity, we define a weight ratio r:

$$r = \frac{w^{(b)}}{w^{(a)}}, 0 \le r \le 1, w^{(b)} \le w^{(a)} \tag{A3}$$

where $a, b \in 1, ..., N_\theta$. Using the simplification introduced in equation (20), we can adapt the expression for the likelihood (equation (9)):

$$\overline{l^{(n_\theta)}} = \frac{1}{\sqrt{2\pi\sigma_\mathrm{obs}^2}} e^{\left(-\frac{\left(\overline{(n_\theta)} \right)^2}{2\sigma_\mathrm{obs}^2} \right)} \tag{A4}$$

where $\overline{(n_\theta)}$ denotes the mismatch between state prediction and observation, which—in accordance with the simplification introduced in equation (20)—is assumed equal for all state particles, observations, and time steps for the sake of this exercise. Combining equations (A2)–(A4), we can derive an expression for the weight ratio depending on the observation mismatches between two parameter particles a and b:

$$R = e^{\left(LN_\mathrm{obs} \frac{\left({(a)} \right)^2}{2\sigma_\mathrm{obs}^2} - LN_\mathrm{obs} \frac{\left({(b)} \right)^2}{2\sigma_\mathrm{obs}^2} \right)} \tag{A5}$$

which can be reformulated to yield an expression for the required relation between observation mismatches given a weight ratio R:

$$\frac{2\sigma_\mathrm{obs}^2 \ln R}{LN_\mathrm{obs}} = \left({(a)} \right)^2 - \left({(b)} \right)^2. \tag{A6}$$

As R is bounded between 0 and 1 and all other variables are positive, the term on the left-hand side is either zero or negative. Since we further defined $w^{(b)} \le w^{(a)}$, we get ${(a)} \le {(b)}$. We can reformulate equation (A6) into an expression for ${(b)}$ given ${(a)}$:
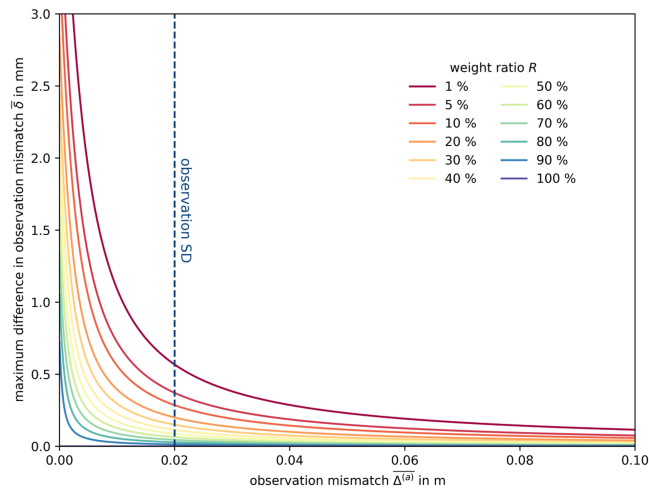
Figure 9. Relation between average observation mismatch and the maximum observation mismatch discrepancy to a second particle for several weight ratios, derived under the assumption $\overline{\Delta^{(b)}} > \overline{\Delta^{(a)}}$. Parameters used to obtain these results correspond to the ones used in this study: $L = 10$, $N_{obs} = 16$, $\sigma_{obs} = 0.02$.

$$\overline{\Delta^{(b)}} = \sqrt{\left(\overline{\Delta^{(a)}}\right)^2 - \frac{2\sigma_{obs}^2 \ln R}{LN_{obs}}}. \tag{A7}$$

Results for the values used in this study ($L = 10$, $N_{obs} = 16$, $\sigma_{obs} = 0.02$) are illustrated in Figure 9 for different weight ratios. The larger a given particle's average observation mismatch $\overline{\Delta^{(a)}}$, the less a second particle's average observation mismatch $\overline{\Delta^{(b)}}$ is allowed to deviate from it to conform to the desired weight ratio. The order of magnitude of δ reveals the reason for the observed ensemble collapse. Assuming that we deem a weight ratio of 10% sufficient, and assuming that one particle has an average observation mismatch of 5 cm, a second (inferior) parameter particle could deviate at most 1.15 μm from the average observation mismatch before causing degeneracy—a highly improbable event.

Based on this simplified representation, we can determine a number of variables that can help prevent weight degeneracy: Shortening the cycle length L is a possibility, but risks carry over effects arising from insufficient dissipation. Decreasing N_{obs} or increasing σ_{obs}^2 could also yield improvements, but both variables are generally out of the user's hand. A more viable solution strategy may be to reduce the magnitude of changes proposed in the artificial parameter dynamics to reduce the resulting deviation in state predictions.

References

Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C., & Vallès, B. (2009). The Ensemble Kalman Filter in reservoir engineering—A review. *SPE Journal*, *14*(03), 393–412. https://doi.org/10.2118/117274-PA

Abbaszadeh, P., Moradkhani, H., & Yan, H. (2018). Enhancing hydrologic data assimilation by evolutionary particle filter and Markov chain Monte Carlo. *Advances in Water Resources*, *111*, 192–204.

Abdoun, O., Abouchabaka, J., & Tajani, C. (2012). Analyzing the performance of mutation operators to solve the travelling salesman problem. arXiv *preprint arXiv:1203.3099*.

Amezcua, J., & Van Leeuwen, P. J. (2014). Gaussian anamorphosis in the analysis step of the EnKF: A joint state-variable/observation approach. *Tellus A: Dynamic Meteorology and Oceanography*, *66*(1), 23,493.

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, *127*(12), 2741–2758.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian\non-Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. https://doi.org/10.1109/78.978374

Aurenhammer, F., & Klein, R. (2000). Voronoi diagrams. *Handbook of Computational Geometry*, *5*(10), 201–290.

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW model development using Python and FloPy. *Groundwater*, *54*(5), 733–739. https://doi.org/10.1111/gwat.12413

Bengtsson, T., Bickel, P., & Li, B. (2008). *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In Probability and statistics: Essays in honor of David A. Freedman* (pp. 316–334). Institute of Mathematical Statistics.

Bennett, J. P., Haslauer, C. P., Ross, M., & Cirpka, O. A. (2019). An open, object-based framework for generating anisotropy in sedimentary subsurface models. *Groundwater*, *57*(3), 420–429.

Caers, J., & Zhang, T. (2004). Multiple-point geostatistics: A quantitative vehicle for integrating geologic analogs into multiple reservoir models.

Cardell-Oliver, R., Kranz, M., Smettem, K., & Mayer, K. (2005). A reactive soil moisture sensor network: Design and field evaluation. *International journal of distributed sensor networks*, *1*(2), 149–162.

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research*, *22*(2), 199–210.

Chopin, N., Jacob, P. E., & Papaspiliopoulos, O. (2013). SMC2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(3), 397–426.

Crisan, D., & Miguez, J. (2018). Nested particle filters for online parameter estimation in discrete-time state-space Markov models. *Bernoulli*, *24*(4A), 3039–3086.

De Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., & Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, *13*(1), 161–183.

Deutsch, C., & Tran, T. (2002). FLUVSIM: A program for object-based stochastic modeling of fluvial depositional systems. *Computers & Geosciences*, *28*(4), 525–535.

Doherty, J. (1994). PEST: A unique computer program for model-independent parameter optimisation. Water Down Under 94: Groundwater/Surface Hydrology Common Interest Papers; Preprints of Papers, 551.

Doherty, J. (2015). *Calibration and uncertainty analysis for complex environmental models*. Brisbane, Australia: Watermark Numerical Computing.

Doherty, J., & Christensen, S. (2011). Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. *Water Resources Research*, *47*, W12534. https://doi.org/10.1029/2011WR010763

Doherty, J., Fienen, M. N., & Hunt, R. J. (2010). Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions (2328-0328). Retrieved from

Doucet, A., & Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, *12*(656-704), 3.

Doucet, A., & Tadić, V. B. (2003). Parameter estimation in general state-space models using particle methods. *Annals of the institute of Statistical Mathematics*, *55*(2), 409–422.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143–10162.

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Farchi, A., & Bocquet, M. (2018). Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, *25*(4), 765–807.

Franssen, H. H., Alcolea, A., Riva, M., Bakr, M., Van der Wiel, N., Stauffer, F., & Guadagnini, A. (2009). A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. *Advances in Water Resources*, *32*(6), 851–872.

Geppert, G. (2015). Analysis and application of the ensemble Kalman filter for the estimation of bounded quantities.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Paper presented at the IEE Proceedings F (Radar and Signal Processing).

Gu, Y., & Oliver, D. S. (2007). An iterative Ensemble Kalman Filter for multiphase fluid flow data assimilation. *SPE Journal*, *12*(4), 438–446. https://doi.org/10.2118/108438-pa

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790.

Hendricks Franssen, H. J., & Kinzelbach, W. (2008). Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, *44*, W09408. https://doi.org/10.1029/2007WR006505

Hill, M. C., Banta, E. R., Harbaugh, A. W., & Anderman, E. R. (2000). MODFLOW-2000, the US Geological Survey modular ground-water model: User guide to the observation, sensitivity, and parameter-estimation processes and three post-processing programs (2331-1258). Retrieved from

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, *34*(12), 3405–3420.

Hu, L. Y., Zhao, Y., Liu, Y., Scheepens, C., & Bouchard, A. (2013). Updating multipoint simulations using the ensemble Kalman filter. *Computers & Geosciences*, *51*, 7–15. https://doi.org/10.1016/j.cageo.2012.08.020

Iglesias, M. A., Law, K. J. H., & Stuart, A. M. (2013). Ensemble Kalman methods for inverse problems. *Inverse Problems*, *29*(4). https://doi.org/10.1088/0266-5611/29/4/045001

Jafarpour, B., & Tarrahi, M. (2011). Assessing the performance of the ensemble Kalman filter for subsurface flow data integration under variogram uncertainty. *Water Resources Research*, *47*, W05537. https://doi.org/10.1029/2010WR009090

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35–45.

Katzfuss, M., Stroud, J. R., & Wikle, C. K. (2016). Understanding the Ensemble Kalman Filter. *The American Statistician*, *70*(4), 350–357. https://doi.org/10.1080/00031305.2016.1141709

Keller, J., Hendricks Franssen, H. J., & Marquart, G. (2018). Comparing seven variants of the Ensemble Kalman Filter: How many synthetic experiments are needed? *Water Resources Research*, *54*, 6299–6318. https://doi.org/10.1029/2018WR023374

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

Kitagawa, G. (1998). A self-organizing state-space model. *Journal of the American Statistical Association*, 1203–1215.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Lee, C., & Antonsson, E. (2000). Variable length genomes for evolutionary algorithms. Paper presented at the GECCO.

Li, T., Sun, S., Sattar, T. P., & Corchado, J. M. (2014). Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches. *Expert Systems with applications*, *41*(8), 3944–3954.

McLaughlin, D., & Townley, L. R. (1996). A reassessment of the groundwater inverse problem. *Water Resources Research*, *32*(5), 1131–1161.

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J. H., Canty, M., & Vereecken, H. (2011). Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of Hydrology*, *399*(3-4), 410–421.

Moradkhani, H., Hsu, K.-L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, *41*, W05012. https://doi.org/10.1029/2004WR003604

Morzfeld, M., Hodyss, D., & Snyder, C. (2017). What the collapse of the ensemble Kalman filter tells us about particle filters. *Tellus A: Dynamic Meteorology and Oceanography*, *69*(1).

Panday, S., Langevin, C. D., Niswonger, R. G., Ibaraki, M., & Hughes, J. D. (2013). MODFLOW-USG Version 1: An unstructured grid version of MODFLOW for simulating groundwater flow and tightly coupled processes using a control volume finite-difference formulation. U.S. Geological Survey (Techniques and Methods 6-A45).

Pathiraja, S., Anghileri, D., Burlando, P., Sharma, A., Marshall, L., & Moradkhani, H. (2018). Time-varying parameter models for catchments with land use change: The importance of model structure. *Hydrology and Earth System Sciences*, *22*(5), 2903–2919.

RamaRao, B. S., LaVenue, A. M., De Marsily, G., & Marietta, M. G. (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. *Water Resources Research*, *31*(3), 475–493.

Reichert, P., & Mieleitner, J. (2009). Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. *Water Resources Research*, *45*, W10402. https://doi.org/10.1029/2009WR007814

Robinson, T., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and electronics in agriculture*, *50*(2), 97–108.

Rubin, Y., & Hubbard, S. S. (2006). *Hydrogeophysics* (Vol. 50). Springer Science & Business Media.

Ruiz, J. J., Pulido, M., & Miyoshi, T. (2013). Estimating model parameters with ensemble-based data assimilation: A review. *Journal of the Meteorological Society of Japan. Ser. II*, *91*(2), 79–99.

Schillings, C., & Stuart, A. M. (2017). Analysis of the Ensemble Kalman Filter for inverse problems. *SIAM Journal on Numerical Analysis*, *55*(3), 1264–1290.

Schöniger, A., Nowak, W., & Hendricks Franssen, H. J. (2012). Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*, *48*, W04502. https://doi.org/10.1029/2011WR010462

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. Paper presented at the Proceedings of the 1968 23rd ACM national conference.

Sivanandam, S., & Deepa, S. (2007). *Introduction to genetic algorithms*. Springer Science & Business Media.

Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. (2008). Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, *136*(12), 4629–4640.

Tang, Q., Kurtz, W., Brunner, P., Vereecken, H., & Franssen, H.-J. H. (2015). Characterisation of river–aquifer exchange fluxes: The role of spatial patterns of riverbed hydraulic conductivities. *Journal of Hydrology*, *531*, 111–123.

van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, *136*(653), 1991–1999.

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., & Reich, S. (2019). Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, *145*(723), 2335–2365. https://doi.org/10.1002/qj.3551

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37–52.

Yan, H., Moradkhani, H., & Zarekarizi, M. (2017). A probabilistic drought forecasting framework: A combined dynamical and statistical approach. *Journal of Hydrology*, *548*, 291–304.

Yan, H., Zarekarizi, M., & Moradkhani, H. (2018). Toward improving drought monitoring using the remotely sensed soil moisture assimilation: A parallel particle filtering framework. *Remote sensing of environment*, *216*, 456–471.

Yeh, W. W. G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, *22*(2), 95–108.

Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H. J., & Li, L. (2011). An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, *34*(7), 844–864. https://doi.org/10.1016/j.advwatres.2011.04.014

Zhu, G., Li, X., Ma, J., Wang, Y., Liu, S., Huang, C., et al. (2018). A new moving strategy for the sequential Monte Carlo approach in optimizing the hydrological model parameters. *Advances in Water Resources*, *114*, 164–179. https://doi.org/10.1016/j.advwatres.2018.02.007

Zovi, F., Camporese, M., Hendricks Franssen, H. J., Huisman, J. A., & Salandin, P. (2017). 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/10.1016/j.jhydrol.2017.02.056