Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography

A. Schöniger,[1,2] W. Nowak,[1] and H.-J. Hendricks Franssen[3]

Received 24 January 2011; revised 20 February 2012; accepted 24 February 2012; published 3 April 2012.

[1] Ensemble Kalman filters (EnKFs) are a successful tool for estimating state variables in atmospheric and oceanic sciences. Recent research has prepared the EnKF for parameter estimation in groundwater applications. EnKFs are optimal in the sense of Bayesian updating only if all involved variables are multivariate Gaussian. Subsurface flow and transport state variables, however, generally do not show Gaussian dependence on hydraulic log conductivity and among each other, even if log conductivity is multi-Gaussian. To improve EnKFs in this context, we apply nonlinear, monotonic transformations to the observed states, rendering them Gaussian (Gaussian anamorphosis, GA). Similar ideas have recently been presented by Béal et al. (2010) in the context of state estimation. Our work transfers and adapts this methodology to parameter estimation. Additionally, we address the treatment of measurement errors in the transformation and provide several multivariate analysis tools to evaluate the expected usefulness of GA beforehand. For illustration, we present a first-time application of an EnKF to parameter estimation from 3-D hydraulic tomography in multi-Gaussian log conductivity fields. Results show that (1) GA achieves an implicit pseudolinearization of drawdown data as a function of log conductivity and (2) this makes both parameter identification and prediction of flow and transport more accurate. Combining EnKFs with GA yields a computationally efficient tool for nonlinear inversion of data with improved accuracy. This is an attractive benefit, given that linearization-free methods such as particle filters are computationally extremely demanding.

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

1. Introduction

1.1. History of Ensemble Kalman Filters

[2] Robust and reliable prognoses of contaminant spreading in the subsurface require a sound knowledge of hydraulic parameters (e.g., hydraulic conductivity). Sparse measurement data often fail to satisfyingly characterize heterogeneous aquifers in a deterministic manner [e.g., *Rubin*, 2003]. Instead, stochastic inverse modeling techniques allow to assimilate measurement data, estimate parameters and quantify parameter and prediction uncertainty. Among several options for stochastic inverse modeling [e.g., *Hendricks Franssen et al.*, 2009], the ensemble Kalman filter (EnKF) was proposed by *Evensen* [1994] and was later refined by *Burgers et al.* [1998]. The EnKF updates predictions of simulation models when new observations become available. It is based on an often nonlinear prediction model, a measurement model and an updating scheme based on recursive application of Bayes rule in a linearized (first and second statistical moments) context. Only if all probability density functions involved in the updating step are multi-Gaussian, the EnKF update will be optimal. An important advantage of the EnKF over the classical or the extended Kalman Filter is that the necessary covariances between states at any given time step are estimated efficiently from a limited ensemble of stochastic realizations without requiring sensitivity analyses.

[3] Early applications of the EnKF in hydrologic sciences assimilated remote sensing data for updating (surface) soil moisture contents [*Reichle et al.*, 2002a, 2002b; *Margulis et al.*, 2002; *Crow and Wood*, 2003]. It was found that data assimilation improved hydrologic predictions [*Margulis et al.*, 2002] and that the EnKF outperformed the extended Kalman Filter [*Reichle et al.*, 2002b]. However, *Reichle et al.* [2002a] observed that a large number of stochastic realizations was needed for satisfactory estimates of ensemble variances.

[4] The first applications of EnKFs for flow in porous media were in petroleum engineering [*Naevdal et al.*, 2003; *Lorentzen et al.*, 2005; *Zafari and Reynolds*, 2005; *Liu and Oliver*, 2005]. *Skjervheim et al.* [2007] included 4-D seismic measurements in data assimilation with the EnKF. *Chen and Zhang* [2006] introduced the EnKF to subsurface hydrology, and further applications followed [e.g., *Hendricks Franssen and Kinzelbach*, 2008; *Tong et al.*, 2010]. Since only the states at the current time step are updated in the traditional formulation of the EnKF, only few of the

[1]Institute for Modelling Hydraulic and Environmental Systems, LH^2/SimTech, Universität Stuttgart, Stuttgart, Germany.
[2]Now at BoSS Consult GmbH, Stuttgart, Germany.
[3]Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany.

0043-1397/12/2011WR010462

above studies considered updating model parameters. Mostly, uncertainty was only considered with respect to the model forcings, initial conditions and measurements.

[5] Subsurface hydrologic model predictions, however, are often dominated by parameter uncertainty [e.g., *Smith and Schwartz*, 1981]. Therefore, the EnKF was reformulated as an augmented state vector approach and as a dual state parameter approach to include uncertain parameters besides the unknown states [e.g., *Moradkhani et al.*, 2005b; *Evensen*, 2009]. *Vrugt et al.* [2005] introduced the option to combine EnKF with optimization for parameter inference. *Liu et al.* [2008] included the calibration of transport parameters in an EnKF framework. The EnKF updating does not enforce consistency of the updated states and model parameters with the governing flow equation. Therefore, *Wen and Chen* [2005, 2006, 2007] proposed a confirming step to ensure the consistency, in their case for multiphase flow in porous media. *Gu and Oliver* [2007] extended this work toward an iterative EnKF which guarantees that physical constraints are met. The groundwater flow equation is less nonlinear as compared with the multiphase flow equation. This is probably the reason why *Hendricks Franssen and Kinzelbach* [2008] found very similar results for the traditional EnKF and the iterative EnKF. *Hendricks Franssen and Kinzelbach* [2009] compared the EnKF and Monte Carlo (MC) based inverse modeling with the sequential self-calibration method in two synthetic studies. These authors found that, even for a large transmissivity variance, the two methods performed similarly, but the EnKF needed a factor of 80 less CPU time.

[6] *Nowak* [2009] reformulated the EnKF such that only parameters are updated, introducing the need for an iterative procedure each time new measurement data are assimilated. This work concluded the evolution of the EnKF from a method that updates only the states, toward a geostatistical inversion algorithm that only updates parameters. If only material parameters are updated, it is assumed that some other sources of uncertainty (e.g., model forcings) are negligible as compared to material parameter uncertainty. The approach, however, does not exclude the option to parametrize uncertain model forcings and jointly update material and forcing parameters. The assimilation of observations by EnKFs in order to update only parameters, not states, will be referred to as p-space (parameter space) EnKF.

1.2. Data Transformation: Motivation and Approach

[7] An important limitation of the EnKF is that it is optimal only for multi-Gaussian relations among all considered variables. Non-Gaussianity of priors and/or nonlinear simulation models result in non-Gaussian model states and data. For subsurface applications, significant non-Gaussianity of states occurs in case of strong spatial heterogeneity of hydraulic conductivity and, in general, for the simulation of contaminant transport with sharp fronts (e.g., advection dominated processes). This occurs even if log conductivity is multi-Gaussian, which is a very frequent assumption in stochastic hydrogeology. Therefore, it is highly desirable in subsurface hydrology to find sequential data assimilation methodologies that are robust against deviations from normality in the states.

[8] One possible option for making sequential data assimilation more robust against deviations from Gaussianity in the states is Gaussian anamorphosis (GA) [e.g., *Chilès and Delfiner*, 1999]. Non-Gaussian random variables are transformed into Gaussian ones using analytical or numerical transformations. This procedure has been used in older studies on disjunctive kriging [e.g., *Matheron*, 1973; *Rivoirard*, 1994], and is referred to as normal score transform in the geostatistical software library GSLIB [*Deutsch and Journel*, 1998]. A good overview of such techniques in petroleum engineering is provided by *Aanonsen et al.* [2009], and a specific example is the work by *Gu and Oliver* [2006]. *Bertino et al.* [2003] introduced it in the context of data assimilation by the EnKF. GA is applied on both the observations and the state space, and the covariance matrices are estimated for the transformed variables. The normal EnKF scheme is applied on the transformed variables, and the states are back transformed after the updating step.

[9] *Simon and Bertino* [2009] used GA for the chlorophyll field in a biogeochemical ocean model. *Béal et al.* [2010] found that GA resulted in a substantial improvement of estimations with a 3-D coupled physical-biogeochemical ocean model as compared to the classical EnKF. In these studies, GA was not applied on parameters or in the context of parameter estimation. We are not aware of the existence of other works that applied GA in the context of sequential data assimilation or in hydrologic sciences, except from the work of *Zhou et al.* [2011]. In specific, there is no study yet that analyzed GA of observed state variables in the context of p-space EnKFs.

[10] Given the fact that the EnKF is an attractive approach because of its CPU efficiency and its relative robustness against nonlinearities [*Hendricks Franssen and Kinzelbach*, 2009], the aim of our research is to analyze and mitigate the effects of non-Gaussian data dependence on the performance of p-space EnKFs. Picking up the idea of GA, we propose nonlinear, monotonic transformations of marginal distributions to move arbitrarily distributed data closer to Gaussianity, at least in the univariate sense. Compared to previous works on this topic [e.g., *Béal et al.*, 2010; *Simon and Bertino*, 2009], we will apply GA in EnKFs for pure parameter estimation instead of state variable estimation or joint state-parameter estimation. Updating only the parameters also obliterates the need to perform a back transformation of updated transformed states. Instead, we obtain updated states from new model runs with the updated parameters. This ensures that the updated state fields always obey the governing equations and mass or energy balances, which is a substantial theoretical advantage over back transformed state fields. The proposed methodology (p-space EnKF with transformed data) will be denoted here by tEnKF. Additionally, our work covers critical issues of GA in any type of EnKF that have never been discussed before. These include transformation of measurement errors, implementation of physical distribution bounds in the GA and the influence of multivariate dependence structures among state variables on the performance gain by GA. Thus, we provide more extensive approaches to predict the success of GA in p-space EnKFs prior to application that exceed the considerations in previous studies by far. Finally, a significant difference to previous studies is that we quantify the improvement by GA with rigorous theoretical measures taken from geostatistical optimal design of

experiments, and we assess for the first time the statistical significance of the improvements in 200 repetitions of a test case.

[11] The benefit of our methodology is illustrated with a first-time application of EnKFs to hydraulic tomography (HT), featuring a multi-Gaussian log conductivity model but non-Gaussian states in a 3-D setup. Among all methods for HT reported in the literature (see section 5.1) that solve the original pressure equation, we could achieve an unprecedented numerical resolution of the inversion even on a contemporary desktop computer. It will be shown that GA leads to an apparent pseudolinearization of the dependence structure between drawdown data and multi-Gaussian log conductivity fields, which can be more accurately exploited by the filter. This yields a more accurate quantification of parameter uncertainty and hence a more accurate prognosis of flow and transport statistics. A detailed discussion of the univariate character of GA and the pseudocharacter of the linearization is provided in sections 4.6 and 7. In a study parallel to ours [*Zhou et al.*, 2011], similar transformation techniques are also applied on parameters, but aim at improving the characterization of non-multi-Gaussian hydraulic conductivity fields. In contrast, we focus on the estimation of multi-Gaussian parameter fields, but discuss in depth the conditions when GA will be beneficial, looking at both univariate and multivariate phenomena to support the decision.

[12] There are other methods available to assimilate non-multi-Gaussian data, which we, however, rate as less promising than our suggested approach. Mostly, those methods are limited in their applicability because of their much higher computational effort. For the sake of completeness, we provide a detailed overview of these alternatives and their limitations in section 2. We proceed with our suggested improvement by presenting the original analysis scheme of the p-space EnKF in section 3 and introducing the tEnKF in section 4.

2. Review of Alternative Approaches

[13] The particle filter (PF) [*Gordon et al.*, 1993] is able to handle any type of statistical distribution and is very robust for strongly nonlinear model dynamics. The PF works on the basis of a large number of particles, which are the equivalents of the stochastic realizations in the EnKF. The particles sample the state (and possibly also the parameter) space according to a given prior distribution. The particle filter allows for a maximum flexibility concerning the functional shape of this prior distribution. The particles are propagated forward in time by the simulation model. When measurements become available, the particles are weighted according to the differences between simulated and measured values.

[14] In the presence of many or strong data, the PF tends to assign very small weights to many particles and large weights to a few particles, leading to a severe degeneration of the statistics because of too small effective (weighted) sample sizes [*Kong et al.*, 1994]. This has been identified as a major obstacle to high-dimensional filtering [*Snyder et al.*, 2008]. Resampling techniques like sequential importance resampling can reduce these problems, but various papers argue that the resampling techniques will not be enough to make the particle filter feasible for large-scale geophysical applications [*van Leeuwen*, 2009; *Snyder et al.*, 2008]. A significant amount of research has been invested to limit this problem. It became common to eliminate particles with very small weights and replace them with particles that are close to the successful (high-weighted) particles, e.g., using sequential importance resampling [*Gordon et al.*, 1993; *Moradkhani et al.*, 2005a; *Leisenring and Moradkhani*, 2010], residual sampling [*Liu and Chen*, 1998], stochastic universal resampling [*Kitagawa*, 1996], Markov chain Monte Carlo (MCMC) methods [*Bocquet et al.*, 2010; submitted AGU paper] or the unscented PF, where the unscented Kalman filter (UKF) is utilized for proposal generation [*Van Der Merwe et al.*, 2001]. An alternative is to resample from a Gaussian distribution on the basis of first and second moments estimated from all the particles [*Xiong et al.*, 2006; *Nakano et al.*, 2007]. Further details on resampling in the context of the PF are given by *van Leeuwen* [2009] and *Arulampalam et al.* [2002].

[15] The PF has been applied much less in the hydrologic sciences than the EnKF. The main applications of the PF have been rainfall-runoff modeling including parameter updating [*Moradkhani et al.*, 2005a; *Weerts and El Serafy*, 2006; *Smith et al.*, 2008] and assimilation of remote sensing data [*Pan et al.*, 2008; *Han and Li*, 2008; *Qin et al.*, 2009; *Montzka et al.*, 2011; *Rings et al.*, 2010; *Dechant and Moradkhani*, 2011]. *Ng et al.* [2009] improved estimates of groundwater recharge by assimilating chloride data with the PF.

[16] In all these examples, the models used had at most some thousands of states and less than 20 unknown parameters. This points to the main current limitation of the PF: In order to sample adequately a high-dimensional state and parameter space, a very large number of particles is needed. Resampling, MCMC methods and improved proposal distributions (see above) slow down filter divergence for large-scale applications. Therefore, although the PF is one of the important alternatives for the EnKF, currently it needs a very large amount of CPU time, and hardly any examples of implementation together with large simulation models can be found in the literature. Nevertheless, recently some works [*van Leeuwen*, 2010; Vrugt et al., submitted manuscript, 2010] give hope that in the future the particle filter could also be used in combination with large simulation models.

[17] Non-Gaussian probability density functions (pdf) can also be approximated by a number of Gaussian pdfs using a Gaussian mixture model (GMM). For the purpose of data assimilation, the parameters of the GMM have to be estimated; see *Sun et al.* [2009] for a groundwater application and *van Leeuwen* [2009], *Bocquet et al.* [2010], and *Van Delft et al.* [2009] for combinations of EnKF and PF. However, none of these alternative methods has been tested for large-scale simulation models, probably because these methods still require a too large amount of CPU time for such applications.

3. Parameter-Space EnKF Analysis Scheme

[18] The set of equations needed to update an ensemble of parameter fields with the p-space EnKF shall be presented in the context of geostatistical inversion, according to *Nowak* [2009]. In hydrogeological applications, the parameter vector **s** contains, e.g., elementwise log conductivity

values of a discretized domain. An error-free model $\mathbf{y} = \mathbf{f}(\mathbf{s})$ is assumed to predict states such as drawdown, pressure head or concentration. The model's prediction quality is affected only by the uncertainty in the parameters, i.e., other sources of uncertainty are neglected in this study. Perturbed measurements are generated by adding a random measurement error $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ to the observed values in order to ensure comparability with real, noisy observations $\mathbf{y}_o$ (see also *Burgers et al.* [1998] for further details). Here $N(\cdot, \cdot)$ denotes a multivariate Gaussian distribution, and $\mathbf{R}$ is the (typically diagonal) covariance matrix of measurement errors. The deviation of simulated states from observed states for each realization $i = 1, \ldots, N$ can be written as

$$\Delta_i = \mathbf{y}_o - (\mathbf{y}_{u,i} + \epsilon_i). \tag{1}$$

[19] The state covariance matrix $\mathbf{Q}_{yy}$ is estimated from the ensemble by

$$\mathbf{Q}_{yy} = E[(\mathbf{y}_u - E[\mathbf{y}_u])(\mathbf{y}_u - E[\mathbf{y}_u])^T]. \tag{2}$$

[20] Likewise, the cross-covariance matrix between parameters and states, $\mathbf{Q}_{sy}$, is obtained, and the covariance matrix of measurements is given by $\mathbf{Q}_{yy} + \mathbf{R}$.

[21] The p-space EnKF updates each unconditioned parameter vector $\mathbf{s}_{u,i}$ to a conditioned parameter vector $\mathbf{s}_{c,i}$ according to

$$\mathbf{s}_{c,i} = \mathbf{s}_{u,i} + \mathbf{Q}_{sy}(\mathbf{Q}_{yy} + \mathbf{R})^{-1}[\mathbf{y}_o - (\mathbf{y}_{u,i} + \epsilon_i)], \tag{3}$$

where the influence function $\mathbf{K} = \mathbf{Q}_{sy}(\mathbf{Q}_{yy} + \mathbf{R})^{-1}$ is called Kalman gain and is formally equivalent to the one used in simple cokriging. Advanced elements of geostatistical models such as trends with uncertain coefficients [e.g., *Kitanidis*, 1986] can be implicitly included in this formulation [*Fritz et al.*, 2009; *Nowak et al.*, 2003].

[22] The conditional covariance matrix of parameters (called the error covariance in EnKFs) may be computed from

$$\mathbf{Q}_{ss,c} = E[(\mathbf{s}_c - E[\mathbf{s}_c])(\mathbf{s}_c - E[\mathbf{s}_c])^T]. \tag{4}$$

[23] The updated ensemble of state predictions $\mathbf{y}_c$ is determined from new model runs $\mathbf{y}_c = f(\mathbf{s}_c)$, and the conditional mean and the error covariance matrix for the predicted states are determined analogously.

[24] Like the original state space EnKF, the p-space EnKF converges to the result of the classical Kalman filter for linear models and with increasing ensemble size. It is derived to be an error covariance minimizing scheme for multi-Gaussian state variables and a linear physical model [*Burgers et al.*, 1998].

4. Parameter Space EnKF Applied to Transformed Data

[25] The assumption of multi-Gaussian dependence structures among state variables is generally not justified for subsurface flow and transport variables. Instead, the type of distribution and dependence on log conductivity is governed by physical processes and imposed boundary conditions. Only in seldom cases, theoretical distribution functions can be derived under a set of idealized boundary conditions [e.g., *Bellin and Tonina*, 2007; *Nowak et al.*, 2008; *Schwede et al.*, 2008; *Cirpka et al.*, 2011]. When desired, the parameters of these distribution functions can be fitted to data sets with the help of the maximum likelihood method or by moment matching [e.g., *Fiorotto and Caroni*, 2002].

[26] Figure 1 shows histograms of two exemplary state variables that are relevant in subsurface applications: drawdown and solute concentration. Drawdown data are taken at two different measurement locations in the 3-D test case setup that will be described in section 5.2. Concentration data are taken at two arbitrarily chosen measurement locations in a model setup not shown here. Different histogram shapes result from the different data types and locations, and only a fraction of the data follows approximately a normal distribution which is indicated by the solid lines.

4.1. Transformation Technique

[27] In order to better exploit the EnKF's potential, we apply nonlinear, monotonic univariate transformations to the state variables that render them univariate Gaussian. An arbitrarily distributed variable y and its Gaussian transform z are linked by their cumulative distribution functions (CDFs):

$$z = G^{-1}[F(y)], \tag{5}$$

with the corresponding empirical CDF $F(y)$ and the theoretical standard normal CDF G. As G is, per definition, monotonously increasing, the inverse G^{-1} exists. The operation $z = \psi(y)$ (equation (5)) is called GA function.

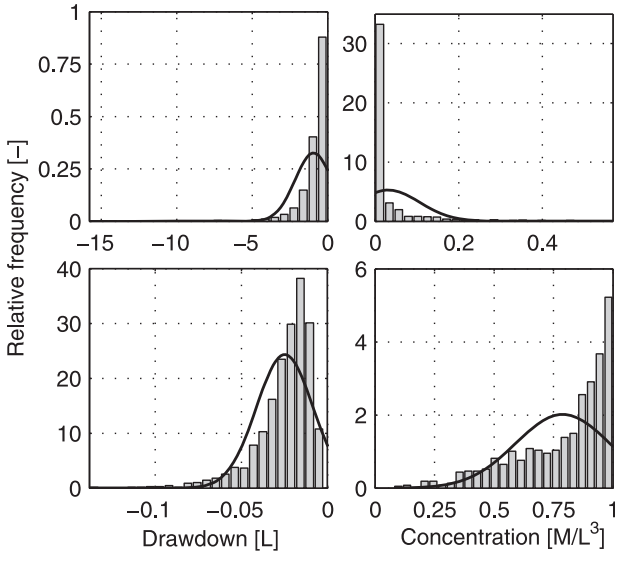

Figure 1. Histograms of two relevant variables in groundwater applications. Bars represent relative frequency, and solid lines show normal distributions that correspond to mean values and variances calculated from the respective samples. Data sets in top and bottom rows are taken from two different measurement locations.

[28] If a theoretical distribution function of y is available, the anamorphosis function $\psi(y)$ is obtained with negligible computational effort. If fitting a parametric distribution function is not appropriate, because state variables or the data do not follow any obvious theoretical distribution, applying nonparametric methods is an alternative. They produce more robust results in the sense that fewer simplifying and possibly subjective assumptions are made: Characteristics of the data are drawn from a sample of sufficient size. We extract this information from the ensemble of MC realizations $\mathbf{y}_u$ which is generated within the EnKF scheme to compute the covariances (see equation (2)), with no additional computational effort made. This allows to make the GA transformation specific to the respective local statistics at any given location in a computational domain (see below).

[29] The procedure of building the anamorphosis function comprises three main steps, in accordance with the work of *Simon and Bertino* [2009]. (1) Find Gaussian values z corresponding to each of the data values y by analyzing their ranks: construction of the empirical anamorphosis function (section 4.2). (2) Fit a continuous function to the empirical anamorphosis function: interpolation of the empirical anamorphosis function (section 4.3). (3) Define the tails of the continuous anamorphosis function by assigning an extrapolation rule (section 4.4).

[30] In the following, without loss of generality, we will assume that the distribution type of the state variable under consideration is unknown. We build an individual anamorphosis function for each measurement location from all realizations, as suggested by *Béal et al.* [2010]. This local approach of transforming the ensemble is less restrictive than constructing a global anamorphosis function for all locations [e.g., *Simon and Bertino*, 2009], because statistical stationarity for state variables is difficult to justify in subsurface applications, e.g., for heads or concentrations in bounded domains with nonsimplistic initial and boundary conditions [e.g., *Osnes*, 1995; *Rubin and Dagan*, 1988, 1989].

4.2. Empirical Anamorphosis Function

[31] To obtain an empirical CDF $F_j(y_j)$, we use the estimator, following *Johnson and Wichern* [1988]:

$$F_j = \frac{j - \frac{1}{2}}{N}, \tag{6}$$

where j are the ranks of the data and N is the sample size (in our case, the ensemble size of the EnKF). From F_j, we can determine the Gaussian transform z by rank transformation. This results in a transformed Gaussian data set that covers a data range from $z_{min} = G^{-1}[(1 - 1/2)/N]$ to $z_{max} = G^{-1}[(N - 1/2)/N]$. An exemplary empirical anamorphosis function for $N = 100$ is plotted in Figure 2. Characteristics of both the untransformed and the transformed variable are illustrated in the corresponding histograms: While, in this example, the original variable y is highly negatively skewed, the transformed variable z is symmetric around zero and displays a characteristically bell-shaped curve. By virtue of the empirical GA, the normal probability plot forms a perfectly straight line, indicating that z indeed follows a Gaussian distribution within the range covered by the sample.

4.3. Continuous Anamorphosis Function

[32] The correction of each parameter field by equation (3) is based on the deviation of the simulated measurements from the observed ones. Therefore, the observed measurements have to be transformed according to the same anamorphosis function, and an appropriate fitting technique has to be found for observed data values that are not included in the numerical Monte Carlo ensemble.

[33] Both *Béal et al.* [2010] and *Simon and Bertino* [2009] define a continuous anamorphosis function by linearly interpolating the empirical anamorphosis function. We propose interpolating the empirical CDF instead, and subsequently determine the continuous anamorphosis function. Linear interpolation of the CDF is a frequently used method in the statistics community. Furthermore, we believe that any linear assumption should be made at the earliest possible stage: Rather, the input of the nonlinear anamorphosis function shall be interpolated than its output.

[34] In an extensive numerical investigation not shown here, we found that more elaborate fitting techniques (e.g., regression by expansion in Hermite polynomials as applied in disjunctive kriging [*Rivoirard*, 1994]) do not significantly increase the accuracy when reproducing a specific distribution. Instead, they are prone to oscillations toward the tails of the distribution. For ensemble sizes larger than $N = 200$, linear interpolation of the empirical CDF proved to be a stable and reasonably accurate choice. The ensemble size required for reasonably accurate results depends strongly on the individual problem; however, ensemble

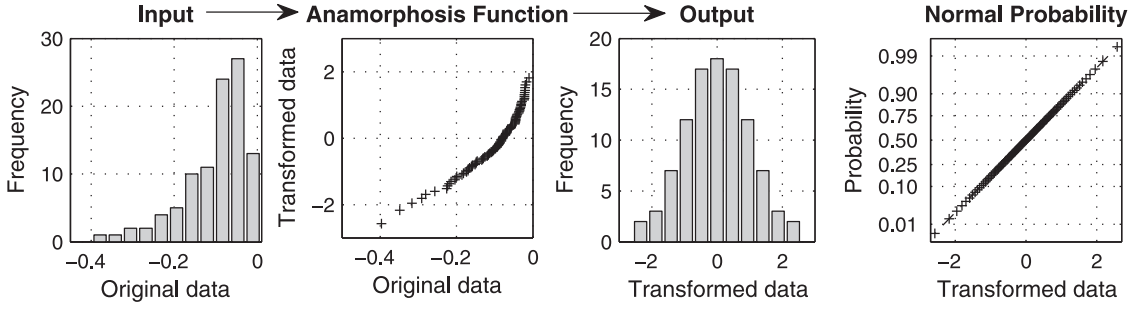

Figure 2. Histogram of untransformed variable, empirical anamorphosis function, histogram of transformed variable, and normal probability plot for transformed variable.

sizes smaller than $N = 200$ are generally not recommended for subsurface applications [*Chen and Zhang*, 2006].

4.4. Definition of the Tails

[35] It may occur that observed measurement values lie outside the range sampled by the initial MC simulation. In absence of better knowledge, we suggest linearly extrapolating the anamorphosis function with a mean slope (e.g., extrapolating with a straight line that connects the lowest and the highest original data values and their respective transformed values). This is equivalent to making a Gaussian assumption for the tails, since values are transformed linearly. This approach is more stable with regard to sampling errors in small ensembles than extrapolating the slope of the last pair of points, and is not subject to oscillations such as extrapolation with Hermite polynomials.

[36] Note that if the observations at a significant amount of measurement locations fall outside the predicted ensemble range, statistics and model assumptions should be checked as argued by *Béal et al.* [2010]. This might be an indication of model bias but can always appear legitimately because of the limited ensemble size, with small probabilities in the order of $P \approx \frac{1}{N}$ per individual data value (see the estimated CDF, equation (6)). Therefore, we consider it important to not restrict the tEnKF to error-free models a priori. Ideas for extrapolation could be refined in future research. For a well-known, but less elaborate standard procedure to determine the anamorphosis function, we refer the reader to the normal score transformation as used in the GSLIB package [*Deutsch and Journel*, 1998].

4.5. Transformation of Measurement Errors

[37] Real observations generally include measurement errors due to imperfections in the measurement process (limited accuracy of measurement instruments, human failure, scale disparity, etc.). In the untransformed space, the EnKF adds a random error $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ to the simulated states in order to simulate the error of the measurement process (see equation (1)).

[38] We suggest building an anamorphosis function based on these perturbed measurements. This ensures that simulated data and real observations are compatible also in the transformed space. At the same time, when generating many random values of ϵ for each sample member y_i, it acts as a kernel smoothing technique [e.g., *Cheng and Parzen*, 1997] for estimating the CDF, making the anamorphosis function more robust even for small ensemble sizes.

4.6. Implicit Pseudolinearization

[39] For parameter estimation, the p-space EnKF updating step interprets the dependence between the observed state variables and the unknown parameters. If the unknown parameters follow a Gaussian distribution (mostly assumed for log conductivity in traditional geostatistical inversion techniques), GA of dependent state variables implicitly triggers a pseudolinearization of the dependence structure. This effect is exploited by the filter's linear updating step and thus leads to more accurate updating results. Figure 4a (see section 5.3) exemplarily shows the dependence between drawdown data and log conductivity before and after transformation: The strongly nonlinear dependence of drawdown on log conductivity has been converted to an almost perfectly linear dependence. The data shown here are taken from the 3-D test case presented in section 5. A more detailed discussion of this topic is provided in section 7.

4.7. Updating Step in the Transformed Space

[40] The updating equation of the tEnKF can now be written in terms of the Gaussian anamorphosed variables $\hat{y} = z$:

$$\mathbf{s}_{c,i} = \mathbf{s}_{u,i} + \mathbf{Q}_{s\hat{y}}\left(\mathbf{Q}_{yy}\widehat{+\mathbf{R}}\right)^{-1}\left(\hat{\mathbf{y}}_o - \left(\mathbf{y}_{u,i}\widehat{+\epsilon_i}\right)\right), \tag{7}$$

with the hat symbolizing anamorphosed variables and their covariances. The vector of the N transformed ensemble values at each of the k measurement locations is obtained by evaluating the individual anamorphosis functions ψ_k. The perturbed measurement covariance matrix $\mathbf{Q}_{yy}\widehat{+\mathbf{R}}$ can be directly calculated from the transformed perturbed measurements $\mathbf{y}_i\widehat{+\epsilon}$. In order to determine the cross-covariance matrix $\mathbf{Q}_{s\hat{y}}$, the simulated measurements are transformed according to the same anamorphosis function, but without their respective measurement errors.

[41] After the updating step, back transformation of the state variables is not required in the context of pure parameter estimation or geostatistical inversion. Instead, the updated states are obtained from a subsequent model run on the updated ensemble of parameter fields. This constitutes a major difference to the application of tEnKFs in state estimation, where transformed states are directly updated by the EnKF at all computational nodes in the domain, and have to be back transformed after updating in the transformed space. Therefore, in state estimation, the anamorphosis function is required to be bijective and spatially continuous. In our case, the anamorphosis function does not have to be bijective (although it is when using the above methods), and we avoid the problem of having to construct a spatially continuous anamorphosis function. Also, the updated states from subsequent simulation without back transformation are guaranteed to follow the governing equations and all mass or energy balances.

[42] Up to now, in EnKF applications, observations that fall outside the physical range (e.g., negative concentrations) have been corrected manually [e.g., *Clark et al.*, 2008], which compromises the random character of the measurement error. The tEnKF as defined in section 4 does not cut unphysical observations but rather leaves them untouched in the spirit of maintaining the randomness of their measurement errors. Instead, unphysical measurement values are automatically filtered out during the course of updating, because the state field is back transformed by solving the physical differential equation. This again avoids the appearance of unphysical updated state variable values.

5. Application to Hydraulic Tomography

[43] Aquifer characterization by HT requires the assimilation of drawdown data from multiple hydraulic interference tests that depend nonlinearly on log conductivity. Our tEnKF is designed to increase the accuracy of nonlinear parameter estimation in such nonlinear problems, and the improvement will be demonstrated by application to HT, here restricted to steady state conditions. Our choice of a steady state case study is not a restriction of our method, but is caused by our intention to compare the tEnKF to

reference solutions from bootstrap filtering (see section 5.4). As a welcome side effect, this also avoids the choice among the many available methods that mitigate the computational effort of transient hydraulic tomography (i.e., using temporal moments of drawdown as done by *Li et al.* [2005]). Several methods to evaluate HT can be found in the literature. They vary in their physical background and computational effort, and shall be briefly outlined in section 5.1.

5.1. Existing Methods

[44] HT was proposed in the 1990s [*Gottlieb and Dietrich*, 1995; *Butler et al.*, 1999], but became more popular only later [e.g., *Yeh and Liu*, 2000]. Recent applications are reported by, e.g., *Vesselinov et al.* [2001], *Illman et al.* [2010], *Straface et al.* [2007], and *Li et al.* [2007].

[45] Different inverse modeling methods have been applied to HT, among which the most frequent ones are the quasi-linear geostatistical approach (QL) [*Kitanidis*, 1995] and the sequential successive linear estimator (SSLE) [e.g., *Yeh and Liu*, 2000].

[46] In order to take full profit of the multiple interference tests and the observed head responses, it is important to use the entire time series of all observed head data, which requires inverse modeling in transient state. This makes the calculations very CPU intensive. Many alternatives to speed up CPU time can be found in the literature, including the so-called steady shape solution by *Bohling et al.* [2002], a reduction of time series by *Zhu and Yeh* [2005], temporal moments by *Li et al.* [2005] and *Zhu and Yeh* [2006], a low-frequency asymptotic approach by *Vasco and Karasaki* [2006], and ray tracing methods [e.g., *Brauchler et al.*, 2007].

[47] Geostatistically based inverse modeling methods like QL and SSLE require that cross covariances between hydraulic head and aquifer parameters are known for all observation time steps or for all temporal moments. This becomes a major obstacle if the grid contains a large number of grid cells, but can be circumvented with the help of spectral methods [*Nowak et al.*, 2003; *Fritz et al.*, 2009]. *Hendricks Franssen and Gomez-Hernandez* [2002] used the sequential self-calibration method (SSC) for inversion, and *Castagna and Bellin* [2009] applied the pilot point method (PP). SSC and PP do not require the explicit estimation of (cross-) covariance matrices, but are nevertheless very CPU-intensive [*Hendricks Franssen and Kinzelbach*, 2009]. A much more detailed review on hydraulic tomography techniques and applications is provided by *Cardiff and Barrash* [2011].

[48] Using the EnKF in the context of HT will offer the following advantages in comparison with other inverse methodologies: (1) the estimation of large covariance matrices via sensitivity analyses is avoided. (2) The needed CPU time is smaller than for QL, SSLE, or SSC [*Hendricks Franssen and Kinzelbach*, 2009; *Nowak*, 2009]. (3) It is still possible to speed up the analysis of transient pumping tests using temporal moments of drawdown. (4) The use of ensemble statistics is more accurate than sensitivity-based first-order cross covariances if the ensemble is sufficiently large [*Nowak*, 2009]. Finally, (5) as opposed to impressively fast methods based on ray tracing [e.g., *Brauchler et al.*, 2007], the original physical governing equations are kept.

5.2. Description of the 3-D Test Case

[49] Drawdown $d\,[L]$ (change in hydraulic head $h\,[L]$) due to a pumping well with pumping rate $q\,[L^3/T]$ at location $\mathbf{x}_0$ in a heterogeneous aquifer with locally isotropic hydraulic conductivity $K\ [L/T]$ is governed by

$$\nabla \cdot [K(\mathbf{x})\nabla h] = -q\delta(\mathbf{x}_0) \tag{8}$$

together with adequate boundary conditions (with δ being Dirac's delta).

[50] Our considered domain represents a rectangular 3-D section of a confined aquifer. It expands over an area of 36 m × 36 m × 3 m and is discretized into $n_{grid} = 432{,}000$ elements ($\Delta x = \Delta y = 0.3$ m, $\Delta z = 0.1$ m). The eastern and western boundaries are fixed to zero drawdown (corresponding to Dirichlet boundary conditions for hydraulic head). Impermeable boundaries are assigned to the northern, southern, top and bottom boundaries. Log conductivity $Y = \ln K$ is assumed to follow a stationary multi-Gaussian distribution with an exponential covariance model and a microscale smoothing parameter [*Kitanidis*, 1997].

[51] Our hydraulic tomography scenario comprises four monitoring wells with three equidistant vertical levels each; see Figure 3. Pumping is induced in four locations given in Table 1 and marked with solid squares in Figure 3. At all monitoring locations, measurements of drawdown are taken for each pumping test at steady state, yielding a number of measurements $n_{meas} = 48$. A synthetic data set is generated by solving equation (8) for the described pumping tests with a random conductivity field. Measurement errors, assumed to be uncorrelated and Gaussian distributed with $\epsilon \sim N(\mathbf{0}, \mathbf{R})$, are added to the synthetic data set. A summary of parameter values adopted in the 3-D test case, including geostatistical parameters, is provided in Table 2.

[52] The tEnKF with $N = 2{,}000$ realizations is applied to estimate the unknown conductivity field from the observed data. The tEnKF analysis scheme is implemented in MATLAB and coupled with the same MATLAB-based FEM code and random field generator also used by *Nowak et al.* [2008] and run on a simple contemporary desktop computer. Both the original p-space EnKF and the

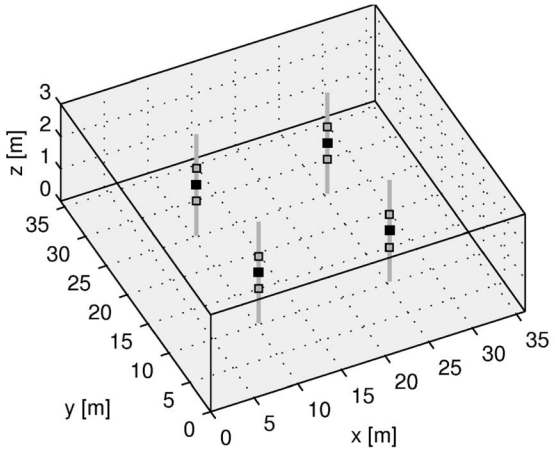

Figure 3. The 3-D domain with measurement locations (open squares) and pumping locations (solid squares).

Table 1. Pumping Locations in the 3-D Domain

Measurement Number	x (m)	y (m)	z (m)
1	10.5	10.5	1.5
14	10.5	25.5	1.5
27	25.5	10.5	1.5
40	25.5	25.5	1.5

transform variant are applied for direct comparison. Note that no iterations are implemented in order to focus on the effect of data transformation on the performance of a single EnKF updating step. Statistical analysis of the two updated ensembles includes the mean (best estimate) and the variance (prediction variance) of the updated conductivity and drawdown fields. The total (spatially averaged) prediction variance of the conditioned parameter field Y_c is defined as

$$\mathrm{var}_{tot}[Y] = \frac{1}{n_{grid}} \frac{1}{N} \sum_{i=1}^{n_{grid}} \sum_{j=1}^{N} (Y_{c,ij} - Y_{mean})^2, \tag{9}$$

with Y_{mean} being the mean value prescribed by the geostatistical model. The spatially averaged prediction variance is equivalent to the A criterion of optimal design, and serves as measure for the absorbed information in spatial estimation [*Nowak*, 2010].

[53] We apply two measures to quantify the deviation from the respective "true" fields that provide the synthetic data set: (1) parameter/prediction bias,

$$\vartheta_{bias} = \sqrt{\frac{1}{n_{grid}} \sum_{i=1}^{n_{grid}} (\vartheta_{true,i} - \overline{\vartheta_{c,i}})^2}, \tag{10}$$

with ϑ being the parameter Y or the state variable d and $\overline{\vartheta_c}$ being the estimate obtained from cellwise averaging over all realizations in each version of the p-space EnKF, and (2) measurement bias,

$$y_{bias,k} = \frac{\overline{y_{c,k}} - y_{o,k}}{\sqrt{R_k}}, \tag{11}$$

defined as deviation of the reproduced measurements $\overline{y_{c,k}}$ compared to the observed ones, normalized by the respective measurement error standard deviation. The sum over all measurement bias values squared

$$\mathrm{sum}[y_{bias}] = \sum_{k=1}^{n_{meas}} y_{bias,k}^2 \tag{12}$$

will be referred to as measurement bias sum.

Table 2. Model Parameters Used for 3-D Synthetic Test Case

Parameter	Symbol	Value	Unit
Domain discretization			
Domain size	$[L_x, L_y, L_z]$	[36, 36, 3]	m
Grid spacing	$[\Delta_x, \Delta_y, \Delta_z]$	[0.3, 0.3, 0.1]	m
Geostatistical model for Y			
Geometric mean	K_g	10^{-5}	m s^{-1}
Variance	σ_Y^2	1	–
Correlation length	$[\lambda_x, \lambda_y, \lambda_z]$	[5, 5, 0.5]	m
Microscale smoothing	d	2.5	m
Measurement error ϵ			
Standard deviation	σ_ϵ	10% + 0.01	m

5.3. Results and Discussion

[54] Before the tEnKF conditioning step, the drawdown data at the measurement locations were transformed according to the methods described in section 4. The empirical anamorphosis functions at two measurement locations (pumping at location a with $x = 10.5$ m, $y = 10.5$ m, $z = 1.5$ m, monitoring at location b with $x = 10.5$ m, $y = 25.5$ m, $z = 1.5$ m) are exemplarily plotted in Figure 4. Drawdown data from the ensemble at those two locations are plotted against Y values at the position of strongest correlation. Obviously, a pseudolinearization of dependence is achieved by GA. The degree of nonlinearity (i.e., the curvature of the anamorphosis function) and thus the gained linearization depends on the measurement location, on the prevailing dependence structure at the respective location, and the type of governing equation (here equation (8)).

[55] Figure 5 shows the "true" Y field (top) as well as the best estimate obtained by the tEnKF (center) and the prediction variance (bottom). The tEnKF is able to capture the zone of low hydraulic conductivity in the northeastern part of the domain and clearly reduces the unconditioned variance given by the geostatistical model in the surroundings of the pumping wells and in the center of the domain, where information of the four pumping tests overlaps.

[56] A comparison of the performance of the two methods is provided in Figure 6. Figure 6 shows a horizontal cross section through the northern part of the domain (y = 25.5 m, z = 1.5 m). The best estimate as well as the standard deviation of the p-space EnKF and the tEnKF are plotted for log conductivity (Figure 6, top) and for drawdown (Figure 6, bottom, logarithmic scale of ordinate) together with the synthetic truth that generated the HT data set. For the largest part of the cross section, the conventional p-space EnKF clearly underestimates hydraulic conductivity, in the eastern part it overestimates the synthetic values. The updating step was not able to cause a significant deviation from the mean conductivity prescribed by the geostatistical model. Remember that we have not used any log conductivity measurements to condition the parameter field, but only measurements of drawdown. Obviously, the tEnKF is able to capture the provided information more effectively, and thus the update of the parameter field is much closer to the synthetic truth. Also, the estimation variance is significantly reduced in the vicinity of the wells (marked in Figure 6) in comparison to the traditional method.

[57] The same holds for drawdown in Figure 6 (bottom). We show a detail of the cross section through the northeastern pumping well. Here, the p-space EnKF consequently overestimates the resulting drawdown, while the tEnKF gives a much better approximation of the true (synthetic) drawdown in the vicinity of the well.

[58] For a further assessment of both methods, the normalized measurement bias (equation (11)) together with the normalized standard deviation are plotted in Figure 7. The measurements reproduced right at the pumping wells are shown in Figure 7 (top), results at the monitoring locations can be seen in the bottom part. Obviously, our proposed method improves the estimate at the vast majority of the measurement locations, indicated by a normalized bias around zero, mostly lying within the band of plus or minus one measurement error standard deviation. The worst cases

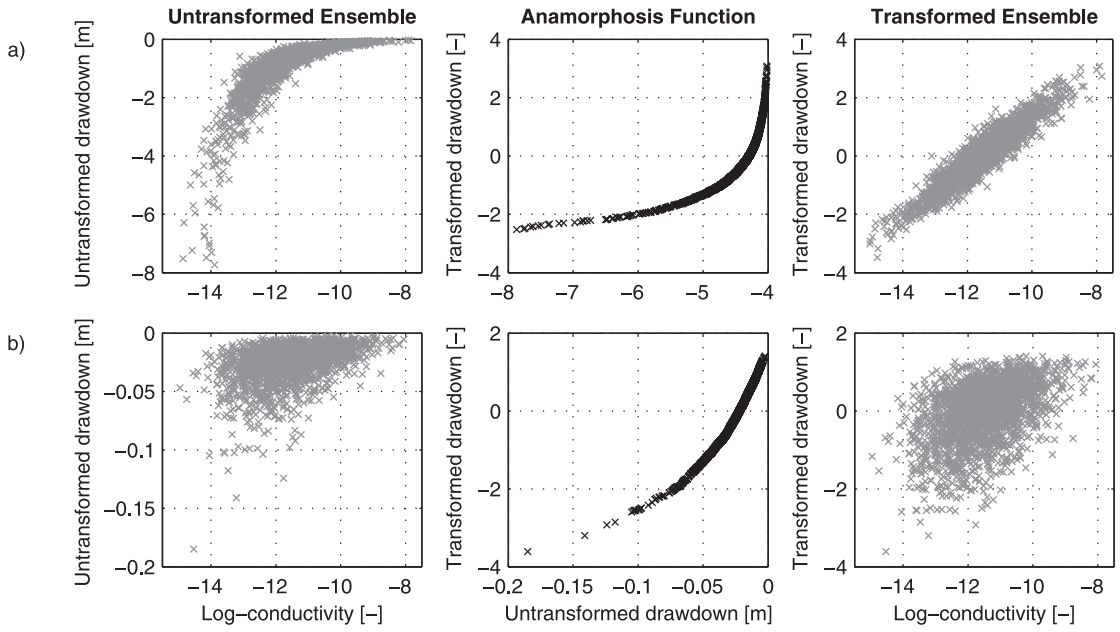

Figure 4. Empirical anamorphosis functions at two arbitrarily chosen measurement locations. (a) Location is the well where pumping is induced in the first of four pumping tests; (b) a monitoring location.

for the original p-space EnKF occur close to pumping wells or at the pumping wells themselves. This is also where the improvement by the tEnKF is largest. Hence, it becomes evident that the benefit of our suggested method is largest close to the pumping wells and decreases with distance.

[59] It is also remarkable, that the drawdown reproduced by the tEnKF is in general less pronounced than the one obtained by the p-space EnKF, which overestimates drawdown at many locations (compare also Figure 6). Our randomly chosen test case features characteristic zones of low conductivity. This property seems to be overestimated by the traditional EnKF variant, since the updated Y values are much too low, resulting in a too strong drawdown. Our suggested method is able to better handle this specific data set. Note that it is not a general property of the tEnKF's bias to lie above the p-space EnKF bias, but is an effect of this specific synthetic truth.

[60] The resulting errors made by the two EnKF variants in assimilating the synthetic data set are summarized in Table 3. The tEnKF achieved a remarkable error reduction of 94.43% when compared to the error made by the traditional p-space EnKF with regard to the estimated measurements, and a reduction of 7.02% with regard to the entire log conductivity field. The error made when estimating the state fields is averaged over all 4 pumping scenarios and was reduced by 55.17% when applying data transformation. Yet, the total prediction variance is larger in the tEnKF, possibly questioning the improvement in processing of information at first sight.

[61] It can be concluded that the proposed methodology is able to better reproduce the synthetic truth for both the parameter field and the state variable fields, while increasing prediction uncertainty. To clarify these seemingly ambiguous findings and ensure significance and independence of a specific data set, we provide a comparison with a reference solution for the EnKF in a statistical framework in section 5.4. There, we will also provide a direct visual comparison of domain-wide parameter and state fields from the different methods.

5.4. Statistical Significance of Results

[62] The above results show that GA improves the accuracy of parameter estimation in the described test application. Yet this might result from characteristics of the specific data set. Therefore, 200 repeated test cases were generated that featured different synthetic data sets each. Moreover, we compared these results with a reference solution for the best estimate rather than the synthetic data set, because this also allows assessing the quality of estimation variances. The reference solutions were obtained from a brute force bootstrap filter [*Gordon et al.*, 1993; *Smith and Gelfand*, 1992], because it does not rely on the assumptions of univariate or multivariate Gaussianity or linearity. It is a direct numerical implementation of Bayes' theorem. The bootstrap filter is the limiting case of particle filters for steady state systems, looking at parameter updating.

[63] However, bootstrap filtering is computationally very costly, which is the reason that kept us from considering a more complex or time-dependent setup. To avoid prohibitive computations, we reduced the test case to a simpler setup. By these means, we could afford $N = 100{,}000$ realizations for the bootstrap filter, which are required to obtain good and reliable results. Only for the reference analysis with the bootstrap filter, we moved from our contemporary desktop computer to a larger multinode computing cluster. N was selected on the basis of convergence experiments with the Jackknife method [*Efron*, 1982] for different ensemble sizes and is large enough to make sure that the prediction variance has a stable value. We considered a 2-D depth-averaged aquifer that spans an area of 100 m $\times$ 100 m, with a single

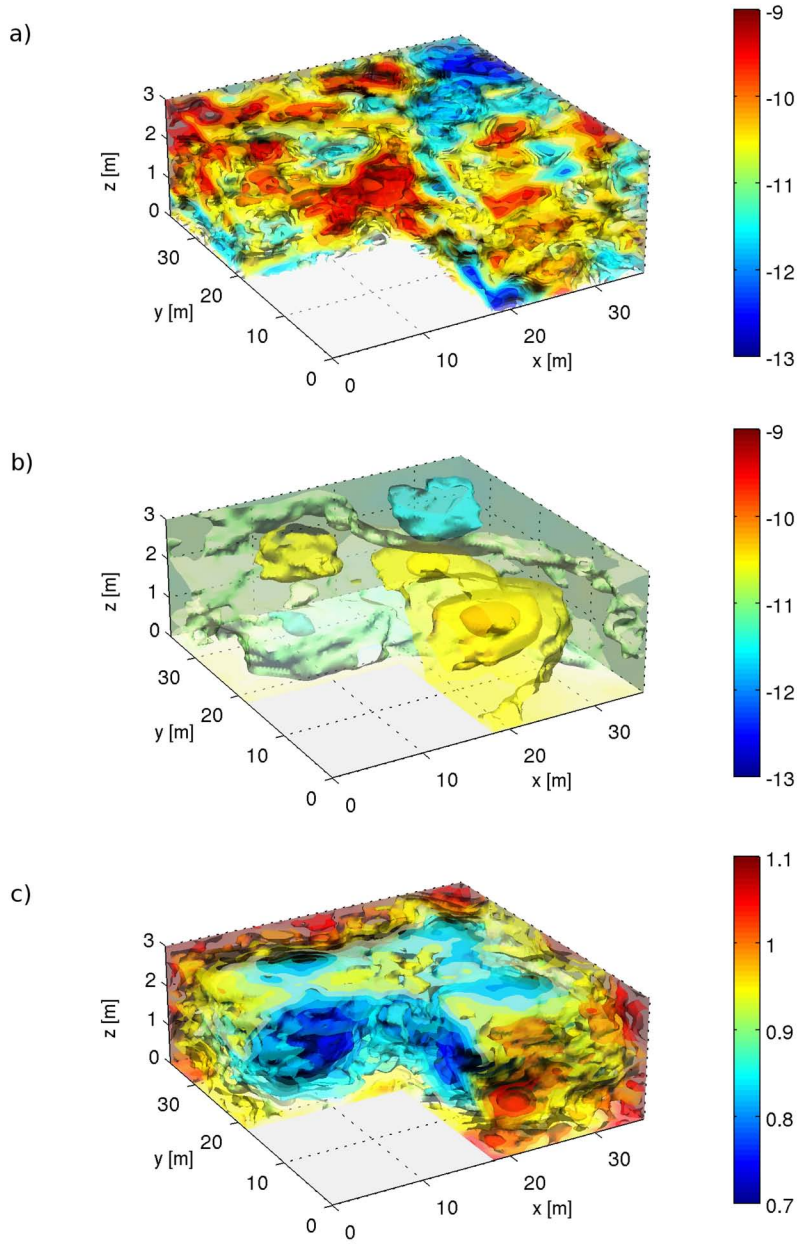

Figure 5. Illustration of 3-D test case: (a) Synthetic log conductivity field, (b) best estimate, and (c) prediction variance of the log conductivity field as obtained by the tEnKF.

pumping well installed in the center of the domain. The same geostatistical model as described in section 5.2, reduced to 2-D, was used for field generation. Eight measurement locations are spread regularly around the well, and one measurement location is positioned right at the well (see Figure 8).

[64] The synthetic log conductivity and drawdown fields as well as the results obtained by the three methods (bootstrap filter, p-space EnKF and tEnKF) are plotted in Figure 9 for one arbitrarily chosen test case among the 200 we analyzed. Before we look at performance statistics in all 200 test cases, we will discuss the performance for the specific test case plotted in Figure 9. It becomes evident that the zone of high conductivity around the pumping well is much better represented by our suggested method. The drawdown estimated by the tEnKF is less pronounced and closer to both the synthetic truth and the bootstrap reference solution when compared to the overestimation of drawdown by the p-space EnKF. The tEnKF shows hardly any improvement over the p-space EnKF for the estimation variance of log conductivity, it is significantly closer to the reference solution for the prediction variance of drawdown. The fifth row shows the normalized measurement bias as defined by equation (11) and the corresponding standard deviation that result from applying the three different methods. As in section 5.3, the largest improvement when applying the tEnKF is again visible at the pumping location (measurement 1).

[65] Now, we look at the performance statistics in all 200 cases. We found that, in the vast majority of test cases, the tEnKF update results in higher contrasts in the

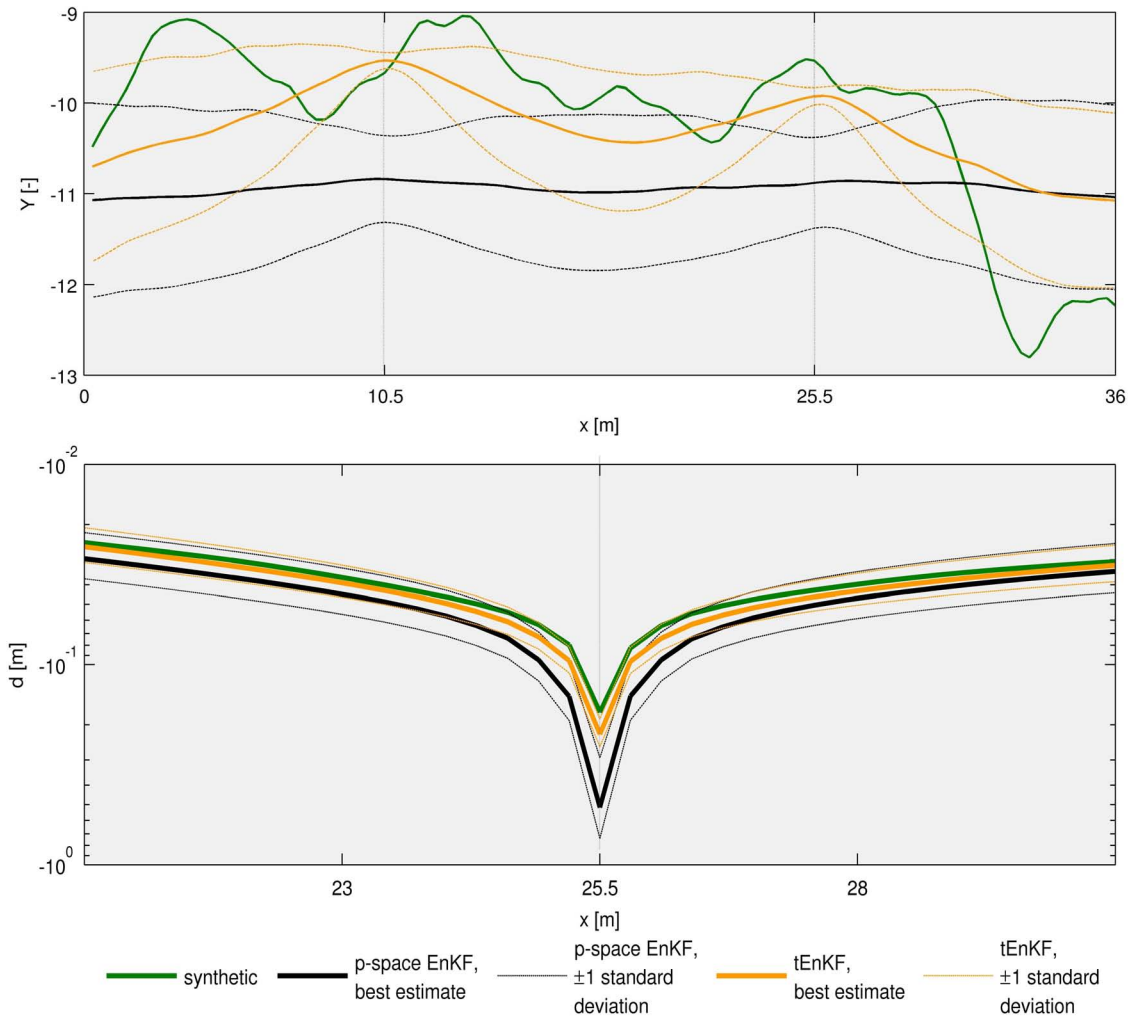

Figure 6. Cross section through 3-D domain of the test case: (top) Log conductivity and (bottom) drawdown at y = 25.5 m and z = 1.5 m. Vertical lines show the positions of drawdown measurements.

parameter field while the p-space EnKF update erroneously stays closer to the mean as prescribed by the geostatistical model. In order to quantify the deviation from the conditional mean fields obtained by the bootstrap filter instead of the "true" field, we now use equations (10) and (11) with respect to the reference solution. To assess the robustness of GA in the EnKF toward too small ensembles, we repeated the analysis for different ensemble sizes $N = 1000, 500, 250$. The deviation of the EnKF prediction variance $\mathrm{var}[Y_c]$ from the weighted bootstrap variance $\mathrm{var}[Y_{BS}]$ is defined as RMSE:

$$\Delta\mathrm{var}[Y] = \sqrt{\frac{1}{n_{grid}}\sum_{i=1}^{n_{grid}}(\mathrm{var}[Y_{BS,i}] - \mathrm{var}[Y_{c,i}])^2}. \quad (13)$$

[66] Results averaged over all 200 test cases are summarized in Table 4. The error of the tEnKF is only 29% for simulation bias, 63% for the parameter field and 43% for the state field as compared to the error of the original p-space EnKF for an ensemble size of $N = 1{,}000$. The deviation from the prediction variance obtained by the bootstrap filter can be reduced by 34%. On the basis of the 95% confidence intervals provided in Table 4, all the improvements are substantial and statistically significant. With decreasing ensemble size, the degree of improvement by GA also decreases because GA based on a small ensemble introduces additional noise into the analysis (see section 7).

[67] Note that with regard to the estimated parameter field, the prediction of the tEnKF was in some cases even worse than the prediction of the traditional p-space EnKF. In these cases, synthetic data far outside the range sampled by the EnKF occurred, so that the extrapolated part of the anamorphosis function was used. This could be controlled by, e.g., applying an iterative scheme as suggested by *Nowak* [2009] or by refining the extrapolation rule in future research.

[68] The results of the statistical analysis verify the promising findings from section 5.3. They also justify the increased prediction variance obtained with the tEnKF in the 3-D test case, since the prediction variance is in average significantly closer to the reference solution's variance. Cases may occur (as in our 3-D test case), where the traditional p-space EnKF produces a smaller prediction variance, which is, however, smaller than in the correct

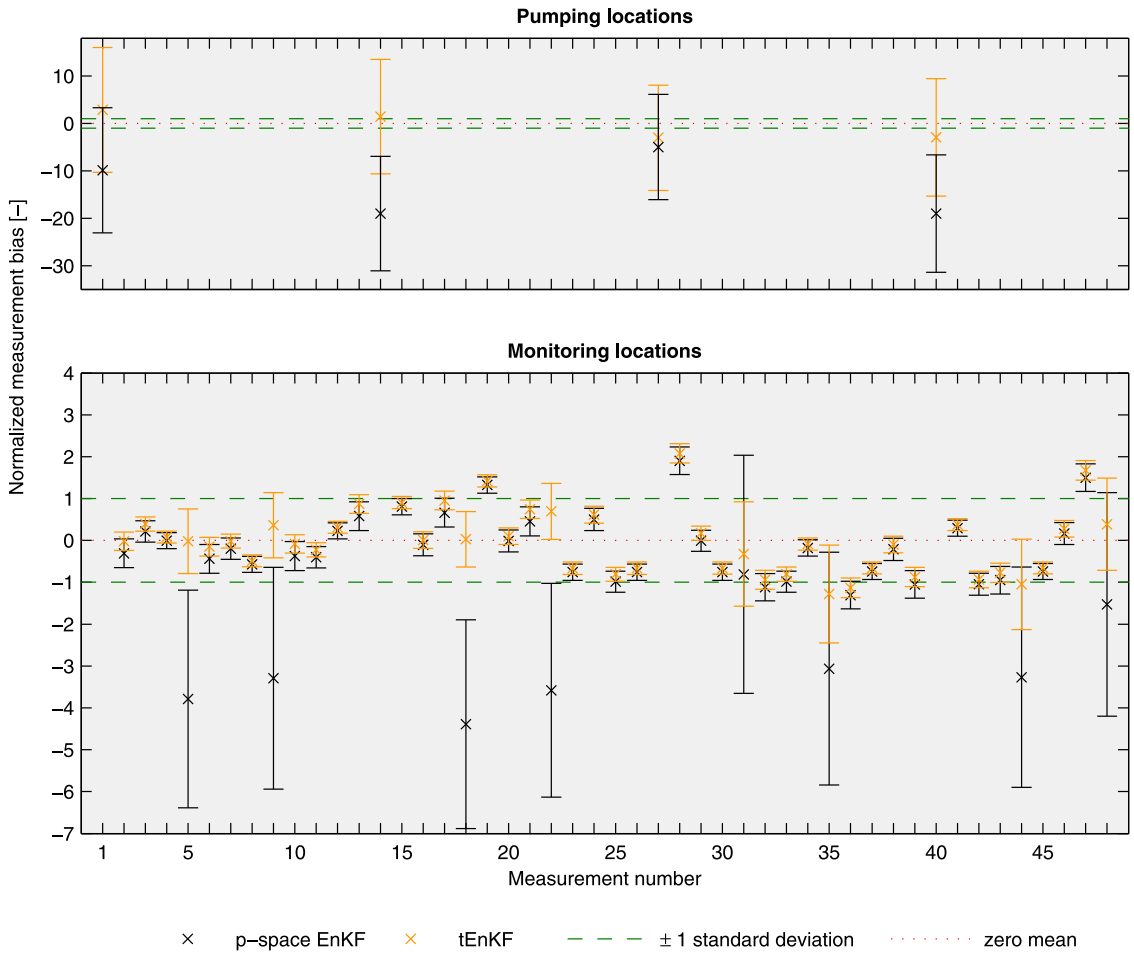

Figure 7. Normalized measurement bias of both methods (black markers, p-space EnKF; orange markers, tEnKF) distinguished for (top) measurements directly at the pumping wells and (bottom) measurements at monitoring locations. Error bars show the normalized standard deviations; the theoretical values (red dots, mean around zero; green dashed lines, standard deviation between −1 and 1) are also plotted.

solution of Bayes's theorem. This misleadingly positive property of the original p-space EnKF discussed in section 5.3 was revealed to be, actually, a drawback by comparing with the bootstrap filter reference solution. Hence, the findings of our statistical analysis firmly support the introduction of the tEnKF as an attractive alternative. It outmatches the traditional p-space EnKF with regard to prediction quality, and is preferable to bootstrap filters or PF methods with regard to computational effort.

6. Alternative Transformation Techniques

[69] Different approaches to construct the empirical anamorphosis function are possible, and the same holds for the chosen interpolation and extrapolation techniques. Instead of constructing the empirical anamorphosis function from perturbed measurements, it could be built from the noise-free simulated ensemble of states. Then, an adequate technique for the separate transformation of measurement errors has to be found. For that case, we suggest transforming the

Table 3. Prediction Error of p-Space EnKF and tEnKF: 3-D Test Case

Error Measure	p-Space EnKF	tEnKF	Error Reduction
Measurement bias sum	946.709	52.736	94.43%
Y bias	0.997	0.927	7.02%
d bias	0.0058	0.0026	55.17%
Prediction variance	1.177	1.451	−23.28%

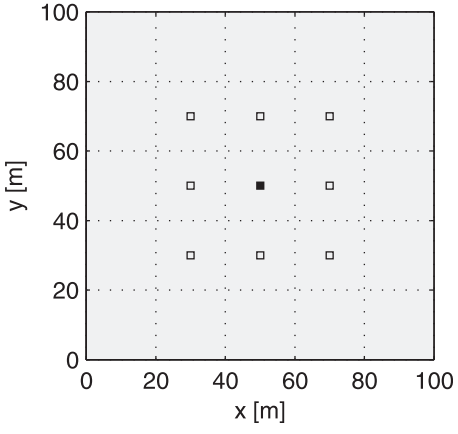

Figure 8. The 2-D domain with measurement locations (open squares) and pumping location (solid square).

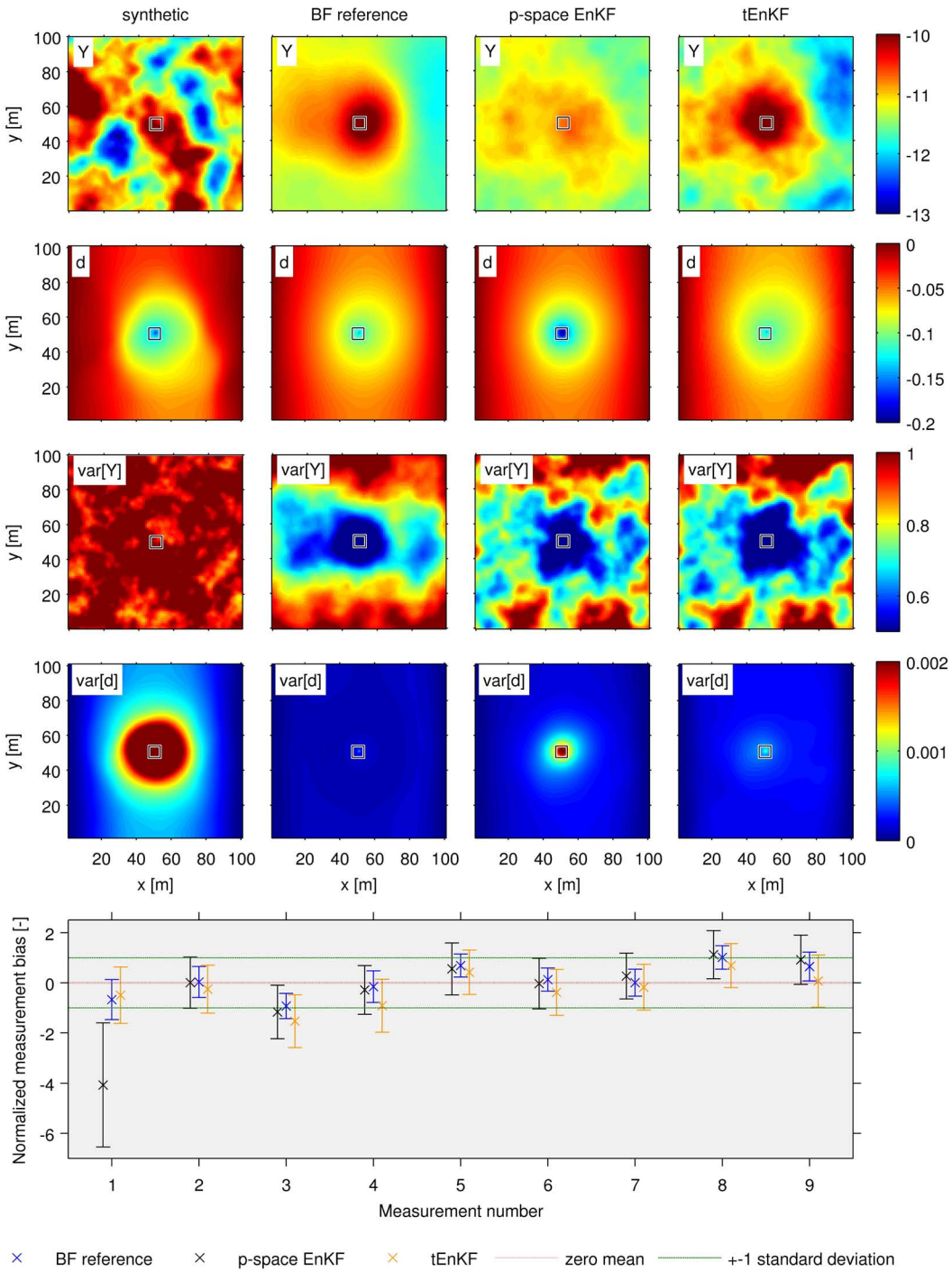

Figure 9. Illustration of one arbitrarily chosen 2-D test case: Synthetic truth and best estimates of the log conductivity and drawdown field (first and second rows), prior variances and prediction variances (third and fourth rows), and normalized measurement bias (fifth row).

Table 4. Averaged Fraction of tEnKF Error Compared to p-Space EnKF Error[a]

Ensemble Size	Measurement Bias	Y Bias	d Bias	$\Delta\text{var}[Y]$
1000	0.29 ($\pm 2 \cdot 0.04$)	0.63 ($\pm 2 \cdot 0.02$)	0.43 ($\pm 2 \cdot 0.02$)	0.66 ($\pm 2 \cdot 0.04$)
500	0.30 ($\pm 2 \cdot 0.04$)	0.65 ($\pm 2 \cdot 0.02$)	0.45 ($\pm 2 \cdot 0.02$)	0.64 ($\pm 2 \cdot 0.04$)
250	0.60 ($\pm 2 \cdot 0.09$)	0.77 ($\pm 2 \cdot 0.02$)	0.63 ($\pm 2 \cdot 0.04$)	0.74 ($\pm 2 \cdot 0.03$)

[a]The 95% confidence intervals for these values are given in parentheses (200 2-D test cases).

measurement error ϵ by maintaining the fraction of measurement error in the total measurement $\mathbf{y}_u + \epsilon$:

$$\hat{\epsilon}_k = \epsilon_k \sqrt{\frac{\hat{\mathbf{R}}_{kk}}{\mathbf{R}_{kk}}}, \tag{14}$$

with

$$\hat{\mathbf{R}}_{kk} = \mathbf{R}_{kk} \frac{\mathbf{Q}_{\hat{y}\hat{y},kk}}{\mathbf{Q}_{yy,kk}}. \tag{15}$$

[70] Note that separate transformation of simulated measurement and measurement error does not yield the same anamorphosed perturbed measurement, since the anamorphosis function is nonlinear.

[71] Test cases (not shown here for briefness) indicated that this alternative approach further improves the prediction quality of the state variable field. However, it is not as effective in reproducing the parameter field when compared to the method discussed previously.

[72] Another advantage of this alternative approach is the opportunity to include knowledge about physical bounds of the prior distributions in the construction of the anamorphosis function. Examples for state variables in groundwater applications that have physical bounds are concentration (bounded between zero and solubility) and hydraulic head (bounded between Dirichlet boundary conditions or bounded at one side because of a combination of Dirichlet and Neumann boundary conditions).

[73] As soon as physical bounds of a variable's prior distribution are reached in the prior ensemble, the values will cluster at these distribution bounds and can be transformed in a controlled manner that preserves this bound. If clustered data values are transformed according to their ranks as described in section 4.2, the problem of nonunique ranks for identical data values will arise. To avoid this problem, we suggest assigning the average rank of all clustered values to the entire cluster. This will introduce a discontinuity in the anamorphosis function at the physical bound, which ensures that all clustered values are mapped onto the same anamorphosed value. The remaining transformed data set will resemble a truncated range of the Gaussian distribution. This suggestion is merely a first approach and could be refined in future research.

7. Applicability to Other Data Types

[74] We have investigated the benefit of the tEnKF for different variable types that vary in their dependence between measured values at different locations and between measured values and parameters. In the following, we analyze the multivariate dependence structure to give recommendations for when to use the tEnKF.

7.1. Nonlinearity of Dependence Structures

[75] GA aims to convert a nonlinear dependence of state variables on parameters into a more linear one, that can be more efficiently processed by the linear EnKF analysis scheme. Thus, the degree of improvement by GA depends on its ability to pseudolinearize the dependence structure between a state variable at a specific measurement location and the entire spatial parameter field. This shall be evaluated on the basis of rank correlation. A similar analysis has already been suggested by *Béal et al.* [2010], but we extend it further to account for varying dependence structures in space.

[76] Figure 10a visualizes the dependence of drawdown measurements at the pumping well (2-D test case; see section 5.4) on the parameter field as a function of space, as determined from the rank correlation. *Béal et al.* [2010] only examined one local bivariate dependence, not the dependencies between all possible bivariate combinations as we do here. Obviously, the correlation is strongest at the well and decreases with distance. The dependence structure shows a rather local and smooth behavior. This is not the case for concentration data, shown in Figure 10b. It is not possible to distinguish a simple dependence pattern, the dependence structure is nonlocal and shows various positive and negative peaks in the domain. This is, e.g., due to the influence of the source hydraulics on the propagation of the plume [*de Barros and Nowak*, 2010].

[77] Consequently, there are several influencing processes that define the dependence of concentration data on log conductivity, but univariate transformation can only provide pseudolinearization in a spatially averaged sense.

[78] Theoretically, different anamorphosis functions for each separate contribution would be needed to address the

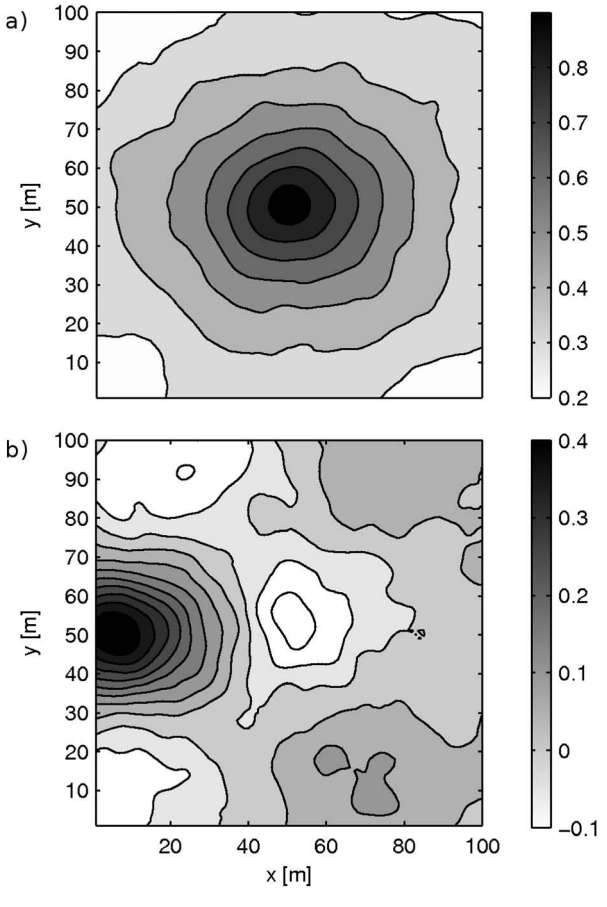

Figure 10. Spatial rank correlation between measurements in the center of the domain and the parameter field: (a) drawdown data and (b) concentration data.

different influences on dependence, each with its respective type of nonlinearity. Since this cannot be handled with univariate transformations, we do not recommend GA for state variables that show such a complex and nonlocal behavior in dependence as concentrations show here.

[79] There might also be variables that show a univariate distribution and multivariate dependence structure close to Gaussian (e.g., hydraulic heads that are hardly influenced by boundary conditions). In this case, application of GA is also not recommended because the construction of the anamorphosis function might even reduce the prevailing Gaussianity as a result of the empirical nature of the procedure.

7.2. Remaining Non-Multi-Gaussianity

[80] In order to evaluate the remaining non-multi-Gaussianity of the transformed variables, we evaluate bivariate copula densities [e.g., *Bárdossy and Li*, 2008]. Data sets at two different measurement locations are mapped onto uniform marginal distributions by sorting the data values and dividing their ranks by the total number of measurements. The resulting values are plotted against each other, and the copula density is determined on a grid of 100 × 100 elements in probability space (compare to histogram bins). Exemplary copula density plots are shown in Figure 11.

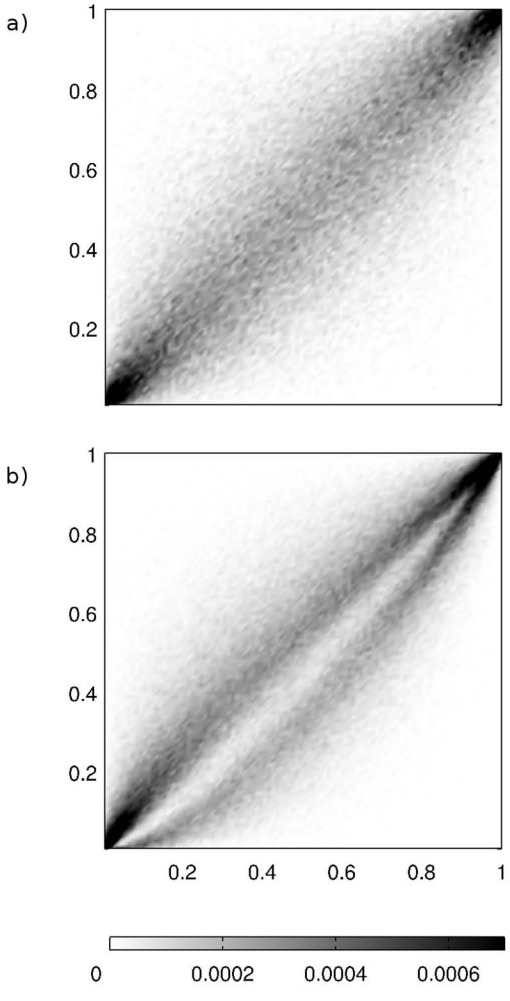

Figure 11. Exemplary copula density plot: (a) drawdown data and (b) concentration data.

[81] Drawdown data are taken from the 2-D test cases presented in section 5.4 and show a copula density similar to a Gaussian copula. Thus, the bivariate structure of drawdown data can be considered bi-Gaussian, and the non-Gaussian influence of the nonlinear dependence on log conductivity can be successfully mitigated by univariate transformation. These properties make a state variable suitable for parameter estimation with the tEnKF.

[82] Concentration data, in contrast, do not fulfill these criteria. The data used for the copula density plot in Figure 11b are taken from *Nowak et al.* [2010]. Obviously, the "coffee bean shaped" copula is far from Gaussian and prohibits a bi-Gaussian assumption, since a Gaussian copula is characterized by its symmetry to both of the diagonals, by exactly two maxima (one each at (0,0) and (1,1)) and exactly one saddle point at (0.5, 0.5). As univariate transformation does not affect the bivariate copula density, GA is not able to convert these data into multivariate Gaussian values.

[83] *Béal et al.* [2010] assess the suitability of state variables for data transformation based only on the rank correlation of the data under consideration (see section 7.1). While the rank correlation coefficient represents only a measure for the univariate mean correlation, the shape of a copula allows to draw conclusions on the type of multivariate structure as well. Therefore, we suggest analyzing the bivariate structure of a state variable first by building empirical copula densities at different combinations of measurement locations. If the resulting copulas indicate a near-Gaussian structure, data transformation is a promising approach; if the copulas are far from Gaussian, however, multivariate transformations should be considered instead. If the interpretation of the resulting Copula plots is less obvious or inappropriate because of a small ensemble size, tests for multivariate normality are given by, e.g., *Smith and Jain* [1988] and *Cox and Small* [1978]. Note that this investigative step can be performed before transforming the data values, thus it is possible to make an educated guess about the expected success of transformation beforehand.

8. Summary and Conclusions

[84] This study has investigated the application of the ensemble Kalman filter (EnKF) to transformed data in the context of parameter estimation. The unmodified EnKF is optimal only for multi-Gaussian variables, i.e., for Gaussian distributed states that are characterized by multi-Gaussian dependence. *Béal et al.* [2010] and other authors named in section 1.2 used the concept of Gaussian anamorphosis (GA) to transform state variables of arbitrary distribution type into univariate Gaussian variables and thus mitigated the effects of non-Gaussianity on the performance of the EnKF.

[85] Our work has transferred and adapted the methodology presented by these authors from EnKFs for state estimation to EnKFs for parameter estimation (p-space EnKF) or geostatistical inversion. We also extend their concept to allow the treatment of measurement errors, to provide a more extensive analysis concerning the range of applicability based on copulas and multivariate spatial statistics, and to assess the success of the method with statistical tests in a large set of 200 test cases.

[86] We call the p-space EnKF applied to transformed data tEnKF. The construction of a continuous anamorphosis function and the updating step in the transformed space are discussed in detail. We showed that the transformation of state variables implicitly leads to a pseudolinearization of dependence on the parameter field to be updated, which is itself assumed to be multi-Gaussian. This results in an improvement in prediction accuracy, since the EnKF is optimal for multi-Gaussian variables, works best for mostly linear models and thus is moved closer to optimality by GA. This supports the motivation for data transformation in the context of parameter estimation with a theoretical foundation.

[87] In contrast to state estimation, state variables are not back transformed by inverting the anamorphosis function after updating in the transformed space. Instead, we solve the physical differential equation with updated parameters, which ensures physical state variable values through all steps and fulfillment of the governing equations. Alternative transformation techniques have been touched upon that allow to include information about physical bounds of the prior distributions in the GA.

[88] We provide recommendations how to assess the degree of suitability for different types of state variables even before the data transformation step. The tEnKF is a substantial improvement in performing nonlinear updates, but it is not yet complete since multi-Gaussianity does not necessarily follow from univariate Gaussianity. We analyzed the nonlinearity and structures of multivariate dependence, as well as the remaining non-multi-Gaussianity after univariate transformation. Experiences gained throughout this study indicate that the dependence structure is a key factor in determining the degree of potential improvement. A strongly nonlinear, but rather local dependence structure allows substantial improvements. Complex and nonlocal dependence structures cannot be satisfyingly converted to multi-Gaussian dependence by univariate transformations. As the limiting factor, we identified the pure structure of multivariate dependence regardless of marginal distribution shapes. The reason is that univariate transformations can change the marginal shape as desired, but cannot change the bivariate or multivariate dependence structure of the involved variables. Multivariate transformations would be needed to address variables that do not fulfill the criteria mentioned here.

[89] To illustrate the practical success of the tEnKF, we have chosen an application to hydraulic tomography. A comprehensive 3-D steady state test case has been presented that comprises 12 measurement locations. An ensemble of 2000 multi-Gaussian log conductivity fields was updated by assimilating synthetic drawdown measurements of 4 pumping tests. Results showed that the tEnKF outperforms the untransformed p-space EnKF with regard to the quality of the best estimate. The error made when reproducing the synthetic drawdown fields was reduced by more than 50%. Our test case was a first-time application of an EnKF to parameter estimation from hydraulic tomography.

[90] To support these promising results and ensure that they are not merely an artifact of one specific data set, we additionally analyzed 200 synthetic test cases that featured a single pumping test in a 2-D domain and ensembles of 1000 realizations each. We determined the error of the tEnKF by comparing its results with a reference solution obtained by a bootstrap filter. The statistical analysis of these results strengthens the conclusions drawn from the 3-D test case: The tEnKF reduced simulation bias by 71%, the error with regard to the parameter field by 37% and the error with regard to the state field by 57% as compared to the performance of the p-space EnKF. Also the deviation from the prediction variance of the bootstrap filter was significantly reduced, indicating a more accurate processing of information. To test robustness against smaller ensemble sizes, we repeated this test also with 500 and 250 realizations.

[91] Obviously, the nonlinear function that relates drawdown data to log conductivity is sufficiently local, can be pseudolinearized by GA and efficiently exploited by the tEnKF, leading to an improved accuracy of prognosis. The degree of nonlinearity and hence the improvement increases with decreasing distance of monitoring wells from the pumping well.

[92] Altogether, the tEnKF is a valuable tool to address strong nonlinearity. It keeps the additional effort over the traditional p-space EnKF very low, while offering a substantial improvement of prediction quality, almost approaching the prediction accuracy of much more CPU-intensive bootstrap filters.

[93] **Acknowledgements.** The authors would like to thank the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/1) and within the International Research Training Group "Nonlinearities and upscaling in porous media" (NUPUS, IRTG 1398) at the University of Stuttgart.

References

Aanonsen, S., G. Naevdal, D. Oliver, A. Reynolds, and B. Valles (2009), The ensemble Kalman filter in reservoir engineering—A review, *SPE J.*, *14*(3), 393–412.

Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.*, *50*(2), 174–188.

Bárdossy, A., and J. Li (2008), Geostatistical interpolation using copulas, *Water Resour. Res.*, *44*, W07412, doi:10.1029/2007WR006115.

Béal, D., P. Brasseur, J.-M. Brankart, Y. Ourmières, and J. Verron (2010), Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: Implications for nonlinear estimation using Gaussian anamorphosis, *Ocean Sci.*, *6*, 247–262.

Bellin, A., and D. Tonina (2007), Probability density function of non-reactive solute concentration in heterogeneous porous formations, *J. Contam. Hydrol.*, *94*(1–2), 109–125.

Bertino, L., G. Evensen, and H. Wackernagel (2003), Sequential data assimilation techniques in oceanography, *Int. Stat. Rev.*, *71*(2), 223–241.

Bocquet, M., C. A. Pires, and L. Wu (2010), Beyond Gaussian statistical modeling in geophysical data assimilation, *Mon. Weather Rev.*, *138*, 2997–3023, doi:10.1175/2010MWR3164.1.

Bohling, G. C., X. Zhan, J. Butler, and L. Zheng (2002), Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities, *Water Resour. Res.*, *38*(12), 1324, doi:10.1029/2001WR001176.

Brauchler, R., J. T. Cheng, P. Dietrich, M. Everett, B. Johnson, R. Liedl, and M. Sauter (2007), An inversion strategy for hydraulic tomography: Coupling travel time and amplitude inversion, *J. Hydrol.*, *345*(3–4), 184–198.

Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*(6), 1719–1724.

Butler, J. J., C. D. McElwee, and G. C. Bohling (1999), Pumping tests in networks of multilevel sampling wells: Motivation and methodology, *Water Resour. Res.*, *35*(11), 3553–3560, doi:10.1029/1999WR900231.

Cardiff, M., and W. Barrash (2011), 3D transient hydraulic tomography in unconfined aquifers with fast drainage response, *Water Resour. Res.*, *47*, W12518, doi:10.1029/2010WR010367.

Castagna, M., and A. Bellin (2009), A Bayesian approach for inversion of hydraulic tomographic data, *Water Resour. Res.*, *45*, W04410, doi:10.1029/2008WR007078.

Chen, Y., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv. Water Res.*, *29*, 1107–1122, doi:10.1016/j.advwatres.2005.09.007.

Cheng, C., and E. Parzen (1997), Unified estimators of smooth quantile and quantile density functions, *J. Stat. Plann. Inference*, *59*(2), 291–307, doi:10.1016/S0378-3758(96)00110-3.

Chilès, J. P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley, New York.

Cirpka, O. A., F. P. J. de Barros, G. Chiogna, and W. Nowak (2011), Probability density function of steady-state concentration in two-dimensional heterogeneous porous media, *Water Resour. Res.*, *47*, W11523, doi:10.1029/2011WR010750.

Clark, M. P., D. E. Rupp, R. A. Woods, X. Zheng, R. P. Ibbitt, A. G. Slater, J. Schmidt, and M. J. Uddstrom (2008), Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Adv. Water Res.*, *31*(10), 1309–1324, doi:10.1016/j.advwatres.2008.06.005.

Cox, D. R., and N. J. H. Small (1978), Testing multivariate normality, *Biometrika*, *65*(2), 263–272, doi:10.1093/biomet/65.2.263.

Crow, W. T., and E. F. Wood (2003), The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, *Adv. Water Res.*, *26*(2), 137–149, doi:10.1016/S0309-1708(02)00088-X.

de Barros, F. P. J., and W. Nowak (2010), On the link between contaminant source release conditions and plume prediction uncertainty, *J. Contam. Hydrol.*, *116*(1–4), 24–34, doi:10.1016/j.jconhyd.2010.05.004.

Dechant, C., and H. Moradkhani (2011), Radiance data assimilation for operational snow and streamflow forecasting, *Adv. Water Res.*, *34*, 351–364, doi:10.1016/j.advwatres.2010.12.009.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed., Oxford Univ. Press, New York.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, vol. 1, 92 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, doi:10.1029/94JC00572.

Evensen, G. (2009), The ensemble Kalman filter for combined state and parameter estimation, *IEEE Control Syst. Mag.*, *29*(3), 83–104.

Fiorotto, V., and E. Caroni (2002), Solute concentration statistics in heterogeneous aquifers for finite Péclet values, *Transp. Porous Media*, *48*, 331–351, doi:10.1023/A:1015744421033.

Fritz, J., I. Neuweiler, and W. Nowak (2009), Application of FFT-based algorithms for large-scale universal kriging problems, *Math. Geosci.*, *41*(5), 509–533, doi:10.1007/s11004-009-9220-x.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993), Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proc., Part F, Radar Signal Process.*, *140*(2), 107–113.

Gottlieb, J., and P. Dietrich (1995), Identification of the permeability distribution in soil by hydraulic tomography, *Inverse Probl.*, *11*, 353–360.

Gu, Y., and D. Oliver (2006), The ensemble Kalman filter for continuous updating of reservoir simulation models, *J. Energy Res. Technol.*, *128*, 79–88.

Gu, Y., and D. Oliver (2007), An iterative ensemble Kalman filter for multiphase fluid flow data assimilation, *SPE J.*, *12*(4), 438–446, doi:10.2118/108438-PA.

Han, X., and X. Li (2008), An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation, *Remote Sens. Environ.*, *112*(4), 1434–1449, doi:10.1016/j.rse.2007.07.008.

Hendricks Franssen, H.-J., and J. J. Gomez-Hernandez (2002), 3D inverse modelling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations, *Stochastic Environ. Res. Risk Assess.*, *16*(2), 155–174.

Hendricks Franssen, H.-J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, *44*, W09408, doi:10.1029/2007WR006505.

Hendricks Franssen, H.-J., and W. Kinzelbach (2009), Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems, *J. Hydrol.*, *365*(3–4), 261–274, doi:10.1016/j.jhydrol.2008.11.033.

Hendricks Franssen, H.-J., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments, *Adv. Water Res.*, *32*(6), 851–872, doi:10.1016/j.advwatres.2009.02.011.

Illman, W. A., J. Zhu, A. J. Craig, and D. Yin (2010), Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study, *Water Resour. Res.*, *46*, W04502, doi:10.1029/2009WR007745.

Johnson, R. A., and D. W. Wichern (1988), *Applied Multivariate Statistical Analysis*, Prentice-Hall, Englewood Cliffs, N. J.

Kitagawa, G. (1996), Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comput. Graphical Stat.*, *5*(1), 1–25.

Kitanidis, P. K. (1986), Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, *22*(4), 499–507, doi:10.1029/WR022i004p00499.

Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*(10), 2411–2420, doi:10.1029/95WR01945.

Kitanidis, P. K. (1997), *Introduction to Geostatistics: Applications to Hydrogeology*, Cambridge Univ. Press, Cambridge, U. K.

Kong, A., J. Liu, and W. Wong (1994), Sequential imputations and Bayesian missing data problems, *J. Am. Stat. Assoc.*, *89*(425), 278–288.

Leisenring, M., and H. Moradkhani (2010), Snow water equivalent prediction using Bayesian data assimilation methods, *Stochastic Environ. Res. Risk Asses.*, *25*(2), 1–18, doi:10.1007/s00477-010-0445-5.

Li, W., W. Nowak, and O. A. Cirpka (2005), Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown, *Water Resour. Res.*, *41*, W08403, doi:10.1029/2004WR003874.

Li, W., A. Englert, O. A. Cirpka, J. Vanderborght, and H. Vereecken (2007), Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests, *Water Resour. Res.*, *43*, W04433, doi:10.1029/2006WR005333.

Liu, G., Y. Chen, and D. Zhang (2008), Investigation of flow and transport processes at the MADE site using ensemble Kalman filter, *Adv. Water Res.*, *31*(7), 975–986, doi:10.1016/j.advwatres.2008.03.006.

Liu, J. S., and R. Chen (1998), Sequential Monte Carlo methods for dynamic systems, *J. Am. Stat. Assoc.*, *93*(443), 1032–1044.

Liu, N., and D. S. Oliver (2005), Ensemble Kalman filter for automatic history matching of geologic facies, *J. Pet. Sci. Eng.*, *47*(3–4), 147–161, doi:10.1016/j.petrol.2005.03.006.

Lorentzen, R., G. Naevdal, B. Valles, A. Berg, and A. A. Grimstad (2005), Analysis of the ensemble Kalman filter for estimation of permeability and porosity in reservoir models, paper presented at SPE Annual Technical Conference and Exhibition, Soc. of Pet. Eng., Dallas, Tex.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, *Water Resour. Res.*, *38*(12), 1299, doi:10.1029/2001WR001114.

Matheron, G. (1973), Le krigeage disjonctif, *Intern. Note N-360*, 40 pp., Centre de Géostatistique, Ecole des Mines de Paris, Paris, France.

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

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

Moradkhani, H., S. Sorooshian, H. Gupta, and P. Houser (2005b), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Res.*, *28*(2), 135–147.

Naevdal, G., L. V. Johnsen, S. I. Aanonsen, and E. H. Vefring (2003), Reservoir monitoring and continuous model updating using ensemble Kalman filter, *Soc. Pet. Eng. J.*, *10*(1), 66–74, doi:10.2118/84372-PA.

Nakano, S., G. Ueno, and T. Higuchi (2007), Merging particle filter for sequential data assimilation, *Nonlinear Proc. Geophys.*, *14*(4), 395–408.

Ng, G. H. C., D. McLaughlin, D. Entekhabi, and B. Scanlon (2009), Using data assimilation to identify diffuse recharge mechanisms from chemical and physical data in the unsaturated zone, *Water Resour. Res.*, *45*, W09409, doi:10.1029/2009WR007831.

Nowak, W. (2009), Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.*, *45*, W04431, doi:10.1029/2008WR007328.

Nowak, W. (2010), Measures of parameter uncertainty in geostatistical estimation and geostatistical optimal design, *Math. Geosci.*, *42*(2), 199–221

Nowak, W., S. Tenkleve, and O. Cirpka (2003), Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities, *Math. Geol.*, *35*(1), 53–66.

Nowak, W., R. L. Schwede, O. A. Cirpka, and I. Neuweiler (2008), Probability density functions of hydraulic head and velocity in three-dimensional heterogeneous porous media, *Water Resour. Res.*, *44*, W12433, doi:10.1029/2007WR006668.

Nowak, W., F. P. J. de Barros, and Y. Rubin (2010), Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain, *Water Resour. Res.*, *46*, W03535, doi:10.1029/2009WR008312.

Osnes, H. (1995), Stochastic analysis of head spatial variability in bounded rectangular heterogeneous aquifers, *Water Resour. Res.*, *31*(12), 2981–2990, doi:10.1029/95WR01956.

Pan, M., E. F. Wood, R. Wójcik, and M. F. McCabe (2008), Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation, *Remote Sens. Environ.*, *112*(4), 1282–1294, doi:10.1016/j.rse.2007.02.039.

Qin, J., S. Liang, K. Yang, I. Kaihotsu, R. Liu, and T. Koike (2009), Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal, *J. Geophys. Res.*, *114*, D15103, doi:10.1029/2008JD011358.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*, 103–114.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*, 728–740.

Rings, J., J. Huisman, and H. Vereecken (2010), Coupled hydrogeophysical parameter estimation using a sequential Bayesian approach, *Hydrol. Earth Syst. Sci.*, *14*(3), 545–556, doi:10.5194/hess-14-545-2010.

Rivoirard, J. (1994), *Introduction to Disjunctive Kriging and Non-linear Geostatistics*, Oxford Univ. Press, New York.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, New York.

Rubin, Y., and G. Dagan (1988), Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 1. Constant head boundary, *Water Resour. Res.*, *24*(10), 1689–1697, doi:10.1029/WR024i010p01689.

Rubin, Y., and G. Dagan (1989), Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 2. Impervious boundary, *Water Resour. Res.*, *25*(4), 707–712, doi:10.1029/WR025i004p00707.

Schwede, R., O. Cirpka, W. Nowak, and I. Neuweiler (2008), Impact of sampling volume on the probability density function of steady state concentration, *Water Resour. Res.*, *44*, W12433, doi:10.1029/2007WR006668.

Simon, E., and L. Bertino (2009), Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: A twin experiment, *Ocean Sci.*, *5*, 495–510.

Skjervheim, J.-A., G. Evensen, S. I. Aanonsen, B. O. Ruud, and T. A. Johansen (2007), Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter, *SPE J.*, *12*(3), 95789, doi:10.2118/95789-PA.

Smith, A. F. M., and A. E. Gelfand (1992), Bayesian statistics without tears: A sampling-resampling perspective, *Am. Stat.*, *46*(2), 84–88.

Smith, L., and F. W. Schwartz (1981), Mass transport: 2. Analysis of uncertainty in prediction, *Water Resour. Res.*, *17*(2), 351–369, doi:10.1029/WR017i002p00351.

Smith, P. J., K. J. Beven, and J. A. Tawn (2008), Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach, *Water Resour. Res.*, *44*, W01410, doi:10.1029/2006WR005205.

Smith, S. P., and A. K. Jain (1988), A test to determine the multivariate normality of a data set, *IEEE Trans. Pattern Anal. Mach. Intell.*, *10*(5), 757–761, doi:10.1109/34.6789.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.*, *136*, 4629–4640, doi:10.1175/2008MWR2529.1.

Straface, S., T. C. J. Yeh, J. Zhu, S. Troisi, and C. H. Lee (2007), Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy, *Water Resour. Res.*, *43*, W07432, doi:10.1029/2006WR005287.

Sun, A. Y., A. P. Morris, and S. Mohanty (2009), Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques, *Water Resour. Res.*, *45*, W07424, doi:10.1029/2008WR007443.

Tong, J., B. X. Hu, and J. Yang (2010), Using data assimilation method to calibrate a heterogeneous conductivity field conditioning on transient flow test data, *Stochastic Environ. Res. Risk Assess.*, *24*(8), 1211–1223, doi:10.1007/s00477-010-0392-1.

Van Delft, G., G. Y. El Serafy, and A. W. Heemink (2009), The ensemble particle filter (EnPF) in rainfall-runoff models, *Stochastic Environ. Res. Risk Assess.*, *23*(8), 1203–1211, doi:10.1007/s00477-008-0301-z.

Van Der Merwe, R., A. Doucet, N. De Freitas, and E. Wan (2001), The unscented particle filter, *Adv. Neural Inf. Process. Syst.*, *96*, 584–590.

van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Mon. Weather Rev.*, *137*, 4098–4114, doi:10.1175/2009MWR2835.1.

van Leeuwen, P. J. (2010), Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Q. J. R. Meteorol. Soc.*, *136*(653), 1991–1999.

Vasco, D. W., and K. Karasaki (2006), Interpretation and inversion of low-frequency head observations, *Water Resour. Res.*, *42*, W05408, doi:10.1029/2005WR004445.

Vesselinov, V. V., S. P. Neuman, and W. A. Illman (2001), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 2. Equivalent parameters, high-resolution stochastic imaging and scale effects, *Water Resour. Res.*, *37*(12), 3019–3041.

Vrugt, J., C. Diks, H. Gupta, W. Bouten, and J. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resour. Res.*, *42*, W09403, doi:10.1029/2005WR004093.

Wen, X. H., and W. H. Chen (2005), Real-time reservoir model updating using ensemble Kalman filter, paper presented at Reservoir Simulation Symposium, Soc. of Pet. Eng., Houston, Texas.

Wen, X. H., and W. H. Chen (2006), Real-time reservoir model updating using ensemble Kalman filter with confirming option, *SPE J.*, *11*(4), 431–442, doi:10.2118/92991-PA.

Wen, X. H., and W. H. Chen (2007), Some practical issues on real-time reservoir model updating using ensemble Kalman filter, *SPE J.*, *12*(2), 156–166, doi:10.118/111571-PA.

Xiong, X., I. M. Navon, and B. Uzunoglu (2006), A note on the particle filter with posterior Gaussian resampling, *Tellus, Ser. A*, *58*(4), 456–460, doi:10.1111/j.1600-0870.2006.00185.x.

Yeh, T. C. J., and S. Liu (2000), Hydraulic tomography: Development of a new aquifer test method, *Water Resour. Res.*, *36*(8), 2095–2105, doi:10.1029/2000WR900114.

Zafari, M., and A. Reynolds (2005), Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, paper presented at Annual Technical Conference and Exhibition, Soc. of Pet. Eng., Dallas, Texas.

Zhou, H., J. Gomez-Hernandez, H.-J. Hendricks Franssen, and L. Li (2011), An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, *Adv. Water Res.*, *34*(7), 844–864, doi:10.1016/j.advwatres.2011.04.014.

Zhu, J., and T.-C. J. Yeh (2005), Characterization of aquifer heterogeneity using transient hydraulic tomography, *Water Resour. Res.*, *41*, W07028, doi:10.1029/2004WR003790.

Zhu, J., and T.-C. J. Yeh (2006), Analysis of hydraulic tomography using temporal moments of drawdown recovery data, *Water Resour. Res.*, *42*, W02403, doi:10.1029/2005WR004309.

H.-J. Hendricks Franssen, Institute of Chemistry and Dynamics of the Geosphere, ICG-4, Research Centre Jülich, Wilhelm-Johnen-Strasse, D-52425 Jülich, Germany.

W. Nowak, Institute for Modelling Hydraulic and Environmental Systems, LH^2/SimTech, Universität Stuttgart, Pfaffenwaldring 61, D-70569 Stuttgart, Germany. (wolfgang.nowak@iws.uni-stuttgart.de)

A. Schöniger, BoSS Consult GmbH, Lotterbergstrasse 16, D-70499 Stuttgart, Germany.