



## Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator

W. Nowak<sup>1,2</sup>

Received 1 August 2008; revised 23 February 2009; accepted 4 March 2009; published 30 April 2009.

[1] Linearized representations of the stochastic groundwater flow and transport equations have been heavily used in hydrogeology, e.g., for geostatistical inversion or generating conditional realizations. The respective linearizations are commonly defined via Jacobians (numerical sensitivity matrices). This study will show that Jacobian-based linearizations are biased with nonminimal error variance in the ensemble sense. An alternative linearization approach will be derived from the principles of unbiasedness and minimum error variance. The resulting paradigm prefers empirical cross covariances from Monte Carlo analyses over those from linearized error propagation and points toward methods like ensemble Kalman filters (EnKFs). Unlike conditional simulation in geostatistical applications, EnKFs condition transient state variables rather than geostatistical parameter fields. Recently, modifications toward geostatistical applications have been tested and used. This study completes the transformation of EnKFs to geostatistical conditioning tools on the basis of best unbiased ensemble linearization. To distinguish it from the original EnKF, the new method is called the Kalman ensemble generator (KEG). The new context of best unbiased ensemble linearization provides an additional theoretical foundation to EnKF-like methods (such as the KEG). Like EnKFs and derivatives, the KEG is optimal for Gaussian variables. Toward increased robustness and accuracy in non-Gaussian and nonlinear cases, sequential updating, acceptance/rejection sampling, successive linearization, and a Levenberg-Marquardt formalism are added. State variables are updated through simulation with updated parameters, always guaranteeing the physicalness of all state variables. The KEG combines the computational efficiency of linearized methods with the robustness of EnKFs and accuracy of expensive realization-based methods while drawing on the advantages of conditional simulation over conditional estimation (such as adequate representation of solute dispersion). As proof of concept, a large-scale numerical test case with 200 synthetic sets of flow and tracer data is conducted and analyzed.

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

### 1. Introduction and Approach

#### 1.1. Motivation: The Biasedness of Linearizations

[2] The principle of linearized error propagation can look back on a long history of successes in stochastic hydrogeology and, in more general, in describing uncertain dynamic systems [e.g., Schweppe, 1973]. Typical hydrogeological applications include geostatistical inversion [e.g., Zimmerman et al., 1998; Keidser and Rosbjerg, 1991; Yeh et al., 1996; Kitanidis, 1995], maximum likelihood estimation of covariance parameters [e.g., Kitanidis and Vomvoris, 1983; Kitanidis and Lane, 1985; Kitanidis, 1995, 1996], and geostatistical optimal design [e.g., Cirpka et al., 2004; Herrera and Pinder, 2005; Feyen and Gorelick, 2005].

However, the crux of linearization techniques is the trade-off between their computational efficiency, conceptual ease and limited range of applicability.

[3] Jacobian-based linearizations have been shown to be biased for nonlinear problems. A first goal of this study is to find a new type of linearization that has better properties. The biasedness of a linearization is defined as the systematic deviation of a tangent from the actual nonlinear function. It can occur for all processes that depend nonlinearly on their parameters, including all flow and transport processes in heterogeneous subsurface environments.

[4] This general statement is easily supported by the fact that all scale-dependent physical process depend nonlinearly on their parameter fields. For any scale-dependent process, inserting the ensemble average of a parameter field into the original equation does not predict the ensemble mean behavior. Instead, the correctly averaged equation requires effective parameters or even has a different mathematical form [e.g., Rubin, 2003; Zhang, 2002].

[5] A powerful example is solute transport in heterogeneous aquifers, and it shall serve as illustration throughout

<sup>1</sup>Civil and Environmental Engineering, University of California, Berkeley, California, USA.

<sup>2</sup>Institute of Hydraulic Engineering, Universität Stuttgart, Stuttgart, Germany.

the entire study. For solute transport, the effective ensemble mean equation is macrodispersive, whereas the transport equation with ensemble mean parameters is only locally dispersive and underestimates dispersion [e.g., *Rubin, 2003*].

[6] The current study will revise the concept of linearization. A linearization scheme for nonlinear equations (and stochastic partial differential equations in particular) will be derived from the principles of unbiasedness and minimum approximation error in the ensemble mean sense. Because of its properties, it will be called best unbiased ensemble linearization (EL). Being based on ensemble statistics, EL will also guarantee adequate treatment of solute dispersion.

[7] The remainder of the introduction will review where biasedness occurs in geostatistical inverse modeling, and how it may be overcome by conditional simulation. Conditional simulation methods may be split into realization-based (MC-type) ones that treat individual realizations one by one, and ensemble-based ones, which work on entire ensembles at a time. The latter are computationally more efficient, yet can be stochastically rigorous and conceptually straightforward, and can outperform realization-based methods [e.g., *Hendricks Franssen and Kinzelbach, 2008b*].

[8] Recently, ensemble-based methods have been modified from pure state space data assimilation tools [e.g., *Evensen, 1994*] toward joint updating of parameters and states [e.g., *Chen and Zhang, 2006*; *Hendricks Franssen and Kinzelbach, 2008a*; *Evensen, 2007*, p. 95]. The most recent trend is further modification toward pure parameter space updating, where updated states are obtained via simulation with updated parameters [e.g., *Liu et al., 2008*]. The contribution of this work, summarized at the end of the introduction, may be seen as final step in the transformation of ensemble Kalman filters toward geostatistical inversion.

## 1.2. Biasedness in Geostatistical Estimation Techniques

[9] A large class of linearizing methods can be found in geostatistical inversion of flow and tracer data, or in the generation of conditional realizations. These methods obtain cross covariances and autocovariances and expected values by Jacobian-based linearized error propagation. Jacobians are derived in sensitivity analyses of the involved flow and transport models, e.g., by adjoint state sensitivities [e.g., *Townley and Wilson, 1985*; *Sykes et al., 1985*]. Examples are the quasi-linear geostatistical method of *Kitanidis [1995]* and the successive linear estimator of *Yeh et al. [1996]*, later revisited by *Vargas-Guzmán and Yeh [2002]*.

[10] Dependent state variables are estimated by inserting the current estimate of the parameter field into the original equation. This technique is accurate only to zeroth order, and clearly biased. Returning to the example of solute dispersion, the bias appears as a lack of dispersion: because of the scale dependence of transport, dispersion is systematically underrepresented on estimated conductivity fields that are smoother than conditional realizations. The trivial lore “not to use best estimates of conductivity fields for transport simulations” is a direct consequence. Likewise, the corresponding first-order concentration variance fails to represent the uncertainty of concentration related to macrodispersive effects. All interpretations of concentration data based on such linearized approaches are bound to produce inaccurate results.

[11] Successive linearization about an increasingly heterogeneous conditional mean conductivity field may gradually include more effects of heterogeneity [e.g., *Cirpka and Kitanidis, 2001*], but will still underrepresent dispersion and fail to interpret concentration data accurately. *Rubin et al. [1999, 2003]* derived dispersion coefficients that apply to estimated conductivity fields, but require perfect separation of scales between large blocks of estimated conductivity and small-scale dispersion phenomena. Simultaneous estimation of a space-dependent dispersivity also helps to overcome the lack of dispersion [*Nowak and Cirpka, 2006*], but entails a scale dependence on the support volume of available tracer data.

## 1.3. Realization-Based Conditional Simulation

[12] The same lore that recommends not performing transport simulations on estimated parameter fields suggests simulating solute transport on conditioned conductivity fields instead, pointing toward the Monte Carlo framework. Each conditioned random conductivity field honors both data and natural variability, and therefore represents solute dispersion accurately.

[13] Available techniques are computationally quite expensive: the pilot point method of *RamaRao et al. [1995]* and *LaVenue et al. [1995]* [see also *Alcolea et al., 2006*], sequential self-calibration by *Gómez-Hernández et al. [1997]* and *Capilla et al. [1997]* [see also *Hendricks Franssen et al., 2003*], or Monte Carlo Markov chain methods [e.g., *Zanini and Kitanidis, 2008*]. Detailed discussion of realization-based methods is provided by H. J. Hendricks Franssen et al. (A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments, submitted to *Advances in Water Research*, 2009).

[14] These realization-based methods rely less on linearized error propagation for data interpretation. For transport simulations, they avoid estimated conductivity fields and can legitimately use the hydrodynamic (local) dispersion tensor without the above dispersion and biasedness issues.

[15] The most widespread ones, i.e., the pilot point method and sequential self-calibration, substitute indirect data types (such as hydraulic heads) by a selection of pilot points or master blocks, which are then used for kriging-like interpolation just like direct data (conductivity data values). Values for the pilot points or master blocks are found by quasi-linear optimization, such that the random conductivity fields comply with the given data values. Some of their drawbacks include (1) the approximate character of this substitution, (2) a lack of options to enforce a prescribed distribution shape of measurement/simulation mismatch, and (3) the computational effort involved for optimizing individual realizations. The first two drawbacks may be minor and of little relevance to the resulting ensemble statistics, as shown in the comparison study by *Hendricks Franssen et al.* (submitted manuscript, 2009), but the issue of computational effort remains.

[16] The Monte Carlo Markov chain method of *Zanini and Kitanidis [2008]* can be seen as a postprocessor to the quasi-linear geostatistical approach (QLGA) [*Kitanidis, 1995*] to improve the quality of conditional realizations. It is stochastically rigorous, but still involves a massive computational effort for conditioning individual realizations. Without that upgrade, the QLGA either requires to

use a single linearization about the conditional mean (at the cost of inaccuracy unless the conditional covariance is very small) or to use individual linearizations for each conditional realization (at excessively high computational costs). A most rigorous and fully Bayesian high end-member of conditional simulation with a minimum of assumptions and simplifications is the method of anchored inversion by Z. Zhang and Y. Rubin (Inverse modeling of spatial random fields using anchors, submitted to *Water Resources Research*, 2009). At the current stage, its advantages come at even higher computational costs, which may be reduced by further research.

#### 1.4. Ensemble-Based Methods

[17] The advantages of conditional simulation may be exploited at substantially reduced computational costs, when conditioning entire ensembles rather than individual realizations. The current study will use the EL concept along these lines, obtaining a quasi-linear generator for conditional ensembles. Only to mild surprise, the resulting method is quite similar to an ensemble Kalman filter (EnKF), therefore called the Kalman ensemble generator (KEG).

[18] The EnKF has been proposed by *Evensen* [1994], later clarified by *Burgers et al.* [1998] and extensively reviewed by *Evensen* [2003]. EnKFs update transient model predictions whenever new data become available. Designed for real-time forecasting of dynamic systems, they have a strictly forward-in-time flow of information. In other words, they do not update the past with present data. Their key elements are a transient prediction model, a measurement model and a forward-in-time Bayesian updating scheme.

[19] Similar to other conditioning techniques, EnKFs require expected values and cross covariances and autocovariances between all model states. These are extracted from an ensemble of realizations which is constantly being updated. The most compelling motivation to use ensemble statistics is to avoid computationally infeasible sensitivity analysis and storage of excessively large autocovariance matrices of parameters. At the same time, EnKFs behave more robustly for nonlinear problems because the ensemble statistics can be accurately evolved in time with nonlinear models. *Burgers et al.* [1998] showed that EnKFs retain higher-order terms compared to the original Kalman filter or the extended Kalman filter [e.g., *Jazwinski*, 1970].

[20] The EL concept derived in this study will add a new angle to the theoretical foundation of EnKFs: It links the choice of ensemble covariances to the fundamental principles of unbiasedness and minimum approximation error. Seen from this angle, the EnKF and the KEG use linearizations that are optimal for the entire ensemble, providing them with excellent computational efficiency. Moreover, they are conceptually straightforward, stochastically rigorous, easy to implement, and require no intrusive modification of simulation software. The accuracy of the conditional statistics obtained by the KEG and its overwhelmingly low computational costs will be demonstrated later.

#### 1.5. From State Space to Parameter Space

[21] Recent developments indicate a transition of the EnKF from the state space toward the parameter space. In their mostly meteorological and oceanographic applications, EnKFs focused on the state space alone. State space

methods use measurements of state variables to update the prediction of state variables. Time-invariant physical parameter fields are insignificant, and the notion of geostatistical structures is entirely absent.

[22] This differs from hydrogeostatistical applications, which focus on the parameter space. Soil parameters are modeled as time-invariant (static) random space functions. The main motivation is to identify static parameter fields of soil properties, much less to combine real-time predictions with incoming streams of observed data. The concept of forward-in-time flow of information does not apply. Instead, great attention is paid to the geostatistical structure of variability, because it plays a major role in the effective behavior of heterogeneous porous media [e.g., *Rubin*, 2003].

[23] A somewhat intermediate concept is the ensemble-based static Kalman filter [e.g., *Herrera*, 1998; *Herrera and Pinder*, 2005; *Zhang et al.*, 2005], sKF for short. It involves a steady state rather than a forward-in-time prediction model, but is still a state space method. Its primary objective is still to improve model predictions, not to condition geostatistical parameter fields.

[24] On the basis of its past successes, EnKFs have received quickly growing attention in hydrogeological studies, as summarized by *Chen and Zhang* [2006], *Hendricks Franssen and Kinzelbach* [2008a, 2008b]. The aforementioned studies (and other works cited therein) included geostatistical parameters into the list of variables to be updated by the EnKF.

[25] *Wen and Chen* [2006] demonstrated the improvement of accuracy when restarting the EnKF once the parameter values have been conditioned. *Hendricks Franssen and Kinzelbach* [2008a] tested the restart principle and found only little or no improvement, probably because their model equations are much closer to linearity. The restart accurately reevaluates the ensemble statistics of state variables using the original equations. This moves the EnKF from a state space or mixed state/parameter space method toward a parameter space method.

[26] The KEG introduced in the current study will complete the transformation of EnKFs into classical parameter space methods. Parameters will be seen as static random space functions. Only the parameter space will be updated from measurements. Updated states will be obtained indirectly by simulation with the updated parameters, which is somewhat similar to an enforced restart in the work of *Hendricks Franssen and Kinzelbach* [2008a] and *Wen and Chen* [2006]. In the examples provided here, the KEG will generate ensembles of log conductivity fields conditional on flow and tracer data, and on measurements of log conductivity itself.

[27] The rigorous theoretical foundation via best unbiased ensemble linearization and the successful history of the EnKF strongly advocate the further use of the KEG, sKF and EnKF methods in hydrogeostatistical applications. In the tradition of geostatistical inversion methods, the KEG will allow for measurement error, but not for model error. Kalman filters require model error to conceptualize measurement/simulation mismatch. In the geostatistical tradition, this mismatch is attributed to yet uncalibrated parameters and boundary conditions.



[28] Erroneous model assumptions can lead to biased parameter estimates, and considering model error may increase the robustness in such situations. The highly flexible EnKF framework, however, gives little reason not to include additional uncertain quantities into the list of parameters for updating, thus reducing the arbitrariness and potential errors in model assumptions. A good example is the joint identification of uncertain conductivity and recharge fields of *Hendricks Franssen and Kinzelbach* [2008a], or the joint identification of unknown boundary values. But of course, extensions of the KEG toward model error will be possible.

[29] A remaining concern in the current study is the original state space character of EnKF-like methods and the KEG. Before extensively using the KEG in applications that raise high requirements to geostatistical structures, it will undergo a deep scrutiny in the current study. *Chen and Zhang* [2006] tested the ability of EnKFs to cope with inaccurate assumptions on geostatistical structures. Since their synthetic data set was almost exhaustively dense, the EnKF was still able to converge toward the reference conductivity field from which the synthetic data were obtained.

[30] Quite contrarily, the rationale behind the tests in the current study is to investigate whether the KEG maintains a prescribed geostatistical model in the absence of strong data, or whether the spatial statistics of the parameter field degenerate during the updating procedure. This is somewhat related to the filter inbreeding problem discussed, e.g., by *Hendricks Franssen and Kinzelbach* [2008a]. Filter inbreeding is the deterioration of ensemble statistics due to an insufficient ensemble size and leads to underestimated prediction variances. While *Hendricks Franssen and Kinzelbach* [2008a] and others cited therein only tested one-point statistics (the field variance), the current study will include two-point statistics (covariances) in order to test geostatistical properties. Further quality assessment includes the compliance of measurement/simulation mismatch statistics with the assumed distribution shape for measurement errors.

## 1.6. Contributions and Organization of the Current Study

[31] The new contributions of the current study can be summarized as follows.

[32] 1. The quasi-linear Kalman ensemble generator (KEG) finalizes the trend [e.g., *Chen and Zhang*, 2006; *Hendricks Franssen and Kinzelbach*, 2008a; *Liu et al.*, 2008] of ensemble Kalman filters (EnKFs) toward geostatistical conditional simulation.

[33] 2. The underlying idea of EnKFs is to use ensemble covariances in their updating equations. This concept is rederived from the principles of best unbiased linearization, providing an additional theoretical foundation to EnKF-like methods. The rederivation also clarifies the advantages of the KEG over Jacobian-based linearized conditioning techniques.

[34] 3. A two-step updating approach like the one by *Hendricks Franssen and Kinzelbach* [2008a] is used. The KEG first processes direct data (linearly related to the parameter field) to update the parameters prior to any simulation of state variables. Then, it processes indirect data (nonlinearly related) by updating the parameters to

reduce the measurement/simulation mismatch. For the indirect data, it has a quasi-linear iteration scheme, stabilized by a geostatistically driven Levenberg-Marquardt technique [*Nowak and Cirpka*, 2004].

[35] 4. The physicalness of updated model states is always guaranteed because they are updated indirectly via simulation with updated parameters. In combination with the above, this significantly improves accuracy of results.

[36] 5. The accuracy of maintaining a prescribed geostatistical structure (two-point covariances) during the conditioning step is assessed. Previous studies looked at one-point statistics only. The filter bias is shown to be zero, complying with the rederivation from best unbiased ensemble linearization.

[37] The current study is organized as follows: First, the concept of best unbiased ensemble linearization will be derived in section 2. Section 3 summarizes the geostatistical framework in brief to install the necessary notation. The quasi-linear Kalman ensemble generator is introduced in section 4, and its similarity and differences to the EnKF and sKF methods are discussed in more detail. In a computationally intensive test case, section 5 assesses the geostatistical properties of the KEG and discusses its computational efficiency.

## 2. Best Unbiased Ensemble Linearization

[38] Let  $\mathbf{s}$  be a parameter vector following a multivariate distribution  $p(\mathbf{s})$ . The value of a dependent state variable at a location  $\mathbf{x}_m$  is denoted by  $y(\mathbf{x}_m) = f(\mathbf{s})$ , where the operator  $f(\cdot)$  represents a model equation (e.g., in the form of a stochastic partial differential equation). In the hydrogeological context,  $\mathbf{s}$  is the field of log conductivity discretized on a numerical grid,  $f(\cdot)$  might represent the stochastic flow or transport equation, and  $y(\mathbf{x}_m)$  would denote a hydraulic head or tracer concentration as predicted by the model.

[39] For simplicity, the following derivation considers a single datum. The goal is to approximate  $f(\mathbf{s})$  by a linearization:

$$f(\mathbf{s}) \approx \tilde{f}(\mathbf{s}) = y_0 + \mathbf{H}(\mathbf{s} - \mathbf{s}_0). \quad (1)$$

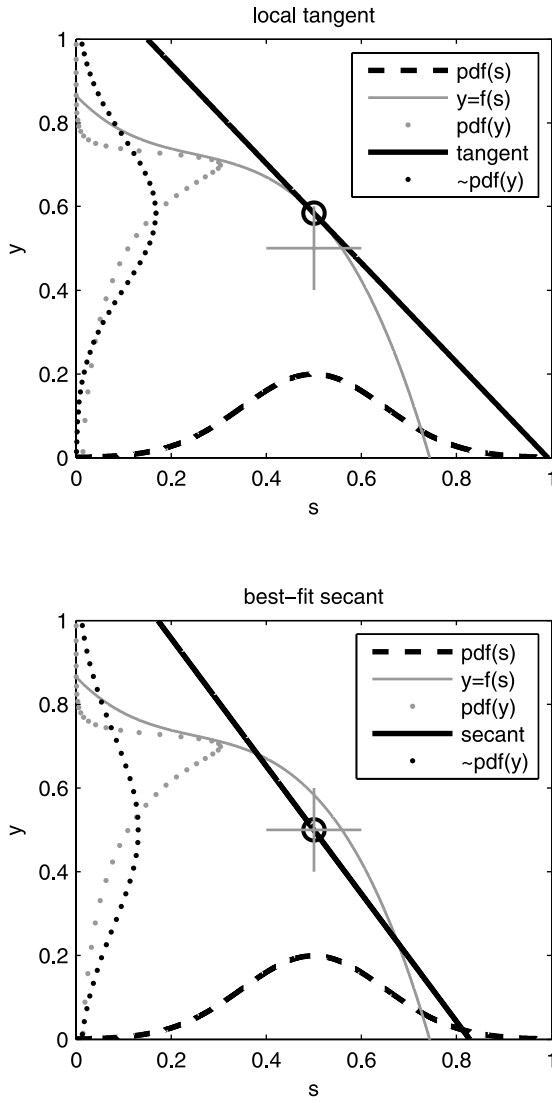
In a purely geometric interpretation,  $[\mathbf{s}_0; y_0]$  is a supporting vector to a hyperplane that approximates the surface  $y = f(\mathbf{s})$ , and  $\mathbf{H}$  is the hyperplane slope. Jacobian-based linearization would now evaluate  $\mathbf{H} = \partial f(\mathbf{s})/\partial \mathbf{s}$  via sensitivity analysis at  $\mathbf{s}_0 = \bar{\mathbf{s}}$  (the mean of  $p(\mathbf{s})$  or the current conditional mean in successive Bayesian updating schemes). Instead, this study treats  $\mathbf{s}_0$  and  $\mathbf{H}$  as free parameters, optimized for unbiasedness and minimum approximation error.

[40] The outcome is not a tangent hyperplane to  $f(\mathbf{s})$  at  $\bar{\mathbf{s}}$ , but rather a global best fit secant hyperplane. To this end, the error  $\varepsilon = f(\mathbf{s}) - \tilde{f}(\mathbf{s})$  must have zero mean and minimum variance:

$$E[\varepsilon] = 0 \quad (2)$$

$$E[\varepsilon^2] \rightarrow \min. \quad (3)$$

Here,  $E[\cdot]$  is the expected value operator over  $p(\mathbf{s})$ . Because of the conditions in equations (2) and (3), the resulting



**Figure 1.** Principles of (top) local tangent (traditional linearization) and (bottom) global best fit secant (best unbiased linearization) for the univariate case comparing the estimated distributions of a dependent variable from each approach. Circles, support points of the tangent and the secant; crosshairs, true mean values of  $s$  and  $y$ . For the local tangent, the bias is identical to the difference between circle position and crosshair position.

linearization  $\tilde{f}(\mathbf{s})$  may legitimately be called best and unbiased linearization in the ensemble sense, i.e., over the distribution  $p(\mathbf{s})$ .

[41] The advantages of the concept become apparent when estimating the distribution of the dependent variable  $y$ . By virtue of the unbiasedness condition, the estimated mean value will be exact. Because of the minimum variance of approximation error, the error of estimating the variance is automatically minimal among all possible unbiased linearizations.

[42] Figure 1 illustrates the principle for a univariate case. Please observe that for the function

$$f(s) = 0.725 - \frac{1}{4} \left( x - \frac{1}{4} \right) - 5 \left( x - \frac{1}{4} \right)^2$$

chosen in this example, the local tangent is consistently larger than  $f(s)$ . This holds everywhere except at the support point  $\bar{s}$ , where it is equal to  $f(s)$ . As a consequence, the estimated distribution  $p(y)$  based on the tangent is biased toward higher values. The true mean value and variance of  $y$  are  $\bar{y} = 1/2$  and  $\sigma_y^2 = 0.068273$ , respectively. The approximation of  $f(s)$  by the local tangent through  $[\bar{s}; f(\bar{s})]$  yields  $\bar{y} = 0.5884$  and  $\sigma_y^2 = 0.031755$ , which is a significant bias and a significant underestimation of the variance. Approximation by the global best fit secant results in  $\bar{y} = 1/2$  and  $\sigma_y^2 = 0.052327$ , which is free of bias and a substantially smaller error in estimating the variance.

[43] When inserting  $\varepsilon = f(\mathbf{s}) - \tilde{f}(\mathbf{s})$  into the unbiasedness condition, the straightforward result is that any unbiased linearization has to return  $E[f(\mathbf{s})]$  if evaluated at  $E[\mathbf{s}]$  (Appendix A). Hence,

$$\mathbf{s}_0 = E[\mathbf{s}] \quad (4)$$

$$y_0 = E[f(\mathbf{s})]. \quad (5)$$

This is independent of the slope  $\mathbf{H}$ , which is still arbitrary at this point. Instead of the true value  $\bar{y} = E[f(\mathbf{s})]$ , traditional linearization uses  $E[f(\mathbf{s})] \approx f(\bar{\mathbf{s}})$ , which is known to be biased for nonlinear functions.

[44] When adding the minimum error variance condition (Appendix B), one obtains

$$\begin{aligned} \mathbf{Q}_{ss}\mathbf{H}^T &= \mathbf{q}_{sy}, \\ \Leftrightarrow \mathbf{H} &= \left( \mathbf{Q}_{ss}^{-1} \mathbf{q}_{sy} \right)^T, \end{aligned} \quad (6)$$

in which  $\mathbf{q}_{sy}$  is the true covariance between  $\mathbf{s}$  and  $y$  in the joint distribution  $p(\mathbf{s}, y)$ , and  $\mathbf{Q}_{ss}$  is the autocovariance matrix for  $p(\mathbf{s})$ . In other words, a best linearization must choose  $\mathbf{H}$  such that the cross covariance from linear error propagation, given by  $\mathbf{Q}_{ss}\mathbf{H}^T$  [e.g., Schweppe, 1973], exactly meets the actual cross covariance  $\mathbf{q}_{sy}$ . In that sense,  $\mathbf{H}$  is an effective derivative  $\partial y / \partial \mathbf{s}$  over the distribution  $p(\mathbf{s})$ .

[45] The resulting (now minimized) expected square error is given by

$$E[\varepsilon^2] = \sigma_y^2 - \mathbf{H}\mathbf{Q}\mathbf{H}^T, \quad (7)$$

i.e., by the error in predicting the variance of  $y$  with the linearized method. Appendix B shows that, similar to the Cramer-Rao inequality [Rao, 1973, p. 324],  $\mathbf{H}\mathbf{Q}\mathbf{H}^T$  is a lower bound to  $\sigma_y^2$ . It is equal to  $\sigma_y^2$  if  $f(\mathbf{s})$  is linear in  $\mathbf{s}$ .

[46] The above analysis has assumed knowledge of three quantities: the mean value  $\bar{y} = E[y]$ , the variance  $\sigma_y^2 = E[(y - \bar{y})^2]$  and the cross covariance  $\mathbf{q}_{sy} = E[(\mathbf{s} - \bar{\mathbf{s}})(y - \bar{y})]$ . These may theoretically be taken from higher-order accurate analytical solutions in simple cases. In the remainder of the study, they will be approximated readily via Monte Carlo estimates (denoted by  $\tilde{\bar{y}}$ ,  $\tilde{\sigma}_y^2$  and  $\tilde{\mathbf{q}}_{sy}$ ), allowing to cover arbitrarily complex cases.

[47] The required MC analysis of course requires computational effort to set up a respective ensemble. The same ensemble can be used for linearization of many data dependencies. For EnKF-like methods, this ensemble is at

the same time the initial unconditional ensemble to be updated later on.

[48] From a signal processing perspective, the term  $\mathbf{Q}_{ss}^{-1} \mathbf{q}_{sy}$  in equation (6) represents a deconvolution. It can be solved at impressive speed using FFT-based PCG solvers [e.g., *Chan and Ng*, 1996], if  $\mathbf{Q}_{ss}$  is stationary and  $\mathbf{s}$  is discretized on a regular grid. *Fritz et al.* [2009] extended this class of solvers to intrinsic cases and allow for irregular grids. Deconvolution amplifies noise, so one may be concerned about noisy  $\mathbf{q}_{sy}$  from an insufficiently large ensemble, leading to an inaccurate approximation of  $\mathbf{H}$ . In that case, one can perform the deconvolution combined with geostatistically based noise-filtering [*Nowak*, 2005].

[49] The deconvolution can even be avoided entirely, because  $\mathbf{H}$  in its raw form seldom appears in common applications. The actual required quantities are the following (co)variances:

$$\mathbf{q}_{sy} \approx \mathbf{Q}_{ss} \mathbf{H}^T = \tilde{\mathbf{q}}_{sy}, \quad (8)$$

$$\sigma_y^2 \approx \mathbf{H} \mathbf{Q}_{ss} \mathbf{H}^T = \left( \mathbf{q}_{ys} \mathbf{Q}_{ss}^{-1/2} \right) \left( \mathbf{Q}_{ss}^{-1/2} \mathbf{q}_{sy} \right) \approx \tilde{\sigma}_y^2. \quad (9)$$

[50] The first expression suggests to directly use the Monte Carlo estimate  $\tilde{\mathbf{q}}_{sy}$ . The second equation requires only a semideconvolution to evaluate  $\mathbf{H} \mathbf{Q}_{ss} \mathbf{H}^T$ . One may still be concerned about a semideconvolution, or one may deem the restriction to intrinsic cases (enforced by practicalities of the deconvolution problem) as inadequate. In such cases, an inconsistent but reasonable approximation is to set

$$\mathbf{H} \mathbf{Q}_{ss} \mathbf{H}^T = \tilde{\sigma}_y^2, \quad (10)$$

where  $\tilde{\sigma}_y^2$  is the Monte Carlo approximation to the true variance of  $y$ . Note that because of the inconsistency, this does not lead to a zero error variance via equation (7).

[51] The results above suggest simply to use expected values and covariances from Monte Carlo analysis. As discussed earlier, it is common practice to do just so in EnKF methods. In their context, however, this is a rather practical choice to avoid costly evaluations of tangents, and to retain higher-order terms in comparison to the original or extended Kalman filters [e.g., *Evensen*, 2003]. There has been no link to the fundamental concepts of unbiasedness and minimum error variance of an implicit linearization. The current study supports this choice with a new context and a firm theoretical basis.

[52] Two different unbiasedness and minimum variance properties are relevant in the current study. All Kalman filters have the unbiased and minimum variance property in estimation, if the model equations are strictly linear. The later test cases will demonstrate that the unbiased minimum variance property in linearization helps to stay close to the unbiased minimum variance property in estimation even in nonlinear cases.

### 3. Geostatistical Framework

[53] Within the context of stochastic hydrogeology, the unknown parameters  $\mathbf{s}$  are typically discretized values of log

conductivity  $Y(\mathbf{x}) = \ln K(\mathbf{x})$ . These are modeled as a random space function, defined by a geostatistical model [*Diggle and Ribeiro*, 2007; *Matheron*, 1971]. For the sake of maximum generality while keeping notation short, log conductivity is here assumed to be intrinsic, generalized to uncertain rather than unknown mean and trend coefficients.

#### 3.1. Generalized Bayesian Intrinsic Model

[54] Consider the parameter vector  $\mathbf{s} | \beta \sim \mathbf{N}(\mathbf{X}\beta, \mathbf{C}_{ss})$ , i.e., multi-Gaussian with mean vector  $\mathbf{X}\beta$  and covariance matrix  $\mathbf{C}_{ss}$  which is assumed to be known.  $\mathbf{X}$  is an  $n_s \times p$  matrix containing  $p$  deterministic trend functions and  $\beta$  is the corresponding  $p \times 1$  vector of trend coefficients. In the generalized intrinsic case, these coefficients are again random variables, distributed  $\beta \sim \mathbf{N}(\beta^*, \mathbf{C}_{\beta\beta})$  with expected value  $\beta^*$  and covariance  $\mathbf{C}_{\beta\beta}$ .

[55] While  $\mathbf{s} | \beta \sim \mathbf{N}(\mathbf{X}\beta, \mathbf{C}_{ss})$  holds for known values of  $\beta$ ,  $\mathbf{s}$  for uncertain  $\beta$  follows the distribution  $\mathbf{s} \sim \mathbf{N}(\mathbf{X}\beta^*, \mathbf{G}_{ss})$ , where

$$\mathbf{G}_{ss} = \mathbf{Q}_{ss} + \mathbf{X} \mathbf{C}_{\beta\beta} \mathbf{X}^T \quad (11)$$

is a generalized covariance matrix [*Kitanidis*, 1993]. The generalization to Gaussian uncertain (rather than entirely unknown) mean and trend coefficients goes back to *Kitanidis* [1986] and has been refined for and applied in geostatistical inversion by *Nowak and Cirpka* [2004]. The notational advantage of generalized covariance matrices is the formal identity of equations to the known-mean case, i.e., the absence of symbols for estimating the trend coefficients  $\beta$ .

#### 3.2. Conditioning Conductivity Fields on Data

[56] Now, consider the  $n_y \times 1$  vector  $\mathbf{y}$  of measured state variables at locations  $\mathbf{x}_m$  according to  $\mathbf{y} = \mathbf{f}(\mathbf{s}) + \varepsilon$ . Here,  $\mathbf{f}(\mathbf{s})$  is a process model and  $\varepsilon \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$  is a vector of measurement errors with zero mean and covariance matrix  $\mathbf{R}$ . For known  $\mathbf{s}$ , the measurements have the conditional distribution  $\mathbf{y} | \mathbf{s} \sim \mathbf{N}(\mathbf{f}(\mathbf{s}), \mathbf{R})$ .

[57] Linearized error propagation yields the marginal distribution of  $\mathbf{y}$  to be  $\mathbf{N}(\bar{\mathbf{y}} = \mathbf{H} \mathbf{X} \beta^*, \mathbf{G}_{yy})$ , where

$$\mathbf{G}_{yy} = \mathbf{H} \mathbf{G}_{ss} \mathbf{H}^T + \mathbf{R} \quad (12)$$

is the generalized covariance matrix of  $\mathbf{y}$ .  $\mathbf{H}$  is a linearized representation of the process model that relates observed state variables to conductivity. Without loss of generality, the additive constant in the linearized representation is omitted from notation. Direct measurements of conductivity are included in this notation by setting the corresponding row in  $\mathbf{H}$  to all zeros with a single unit entry at the sampled position [e.g., *Fritz et al.*, 2009]. Including direct measurements of parameter values is a standard procedure in the hydrogeostatistical literature. In the original context of the EnKF, this is rarely seen because information on model parameters is not the primary focus. Any arbitrary data type can be processed if a model function  $\mathbf{f}(\mathbf{s})$  for the underlying process is available. This includes, e.g., geophysical data.



[58] Within linear(ized) approaches, the conditional distribution  $s|y$  is again multi-Gaussian with conditional mean  $\hat{s}$  and covariance  $\mathbf{G}_{ss|y}$ :

$$\begin{aligned}\hat{s} &= \mathbf{X}\beta^* + \mathbf{G}_{sy}\mathbf{G}_{yy}^{-1}(y_o - \bar{y}) \\ \mathbf{G}_{ss|y} &= \mathbf{G}_{ss} - \mathbf{G}_{sy}\mathbf{G}_{yy}^{-1}\mathbf{G}_{ys} \\ \mathbf{G}_{sy} &= \mathbf{G}_{ss}\mathbf{H}^T.\end{aligned}\quad (13)$$

[59] The accuracy of linearization may be improved by successive linearization about a current estimate [e.g., *Kitanidis*, 1995; *Yeh et al.*, 1996], but this leaves the multi-Gaussian assumption untouched. To the concern of the current study, the zeroth-order approximation of  $\bar{y}$  via  $\mathbf{H}\mathbf{X}\beta^*$  and the first-order approximations of  $\mathbf{G}_{yy}$  and  $\mathbf{G}_{sy}$  lead to a bias and nonminimal error in data interpretation. Especially the interpretation of tracer data is affected, as discussed in the introduction.

## 4. Quasi-Linear Kalman Ensemble Generator

### 4.1. Linear Kalman Ensemble Generator

[60] The previous sections presented the concept of best unbiased ensemble linearization (EL) and summarized the necessary notation for geostatistics and conditioning. The upcoming section employs the EL concept to condition random conductivity fields on flow and tracer data, leading to a new method called the quasi-linear Kalman ensemble generator (KEG). This set of physical quantities is chosen for illustration, but the method applies to arbitrary sets of random parameters and arbitrary data types.

[61] When using EL in the conditioning context, the statistics required in equation (13) are extracted from a sufficiently large ensemble as Monte Carlo estimates,  $\tilde{\mathbf{G}}_{yy}$ ,  $\tilde{\mathbf{G}}_{sy} = \tilde{\mathbf{G}}_{ys}^T$  and  $\tilde{\bar{y}}$ . Adequate ensemble sizes (about 500) for a closely related EnKF are discussed by *Chen and Zhang* [2006]. The ensemble contains random fields  $\mathbf{s}_{u,i}$  drawn from  $p(\mathbf{s})$ , their respective model outcomes  $\mathbf{y}_{u,i} = \mathbf{f}(\mathbf{s}_{u,i})$ , and an ensemble of random measurement errors  $\varepsilon_i$  drawn from  $p(\varepsilon)$ . The resulting method is called the KEG to distinguish the new (parameter space) method from the (state space) ensemble Kalman filter. The KEG conditions each unconditional realization  $\mathbf{s}_{u,i}$  according to the common equation

$$\mathbf{s}_{c,i} = \mathbf{s}_{u,i} + \tilde{\mathbf{G}}_{sy}\tilde{\mathbf{G}}_{yy}^{-1}(y_o - \mathbf{f}(\mathbf{s}_{u,i}) + \varepsilon_i).\quad (14)$$

[62] Basic similarities and differences to the EnKF have been discussed in section 1. The similarity lies in the formal identity of equation (14) to the updating equation in the EnKF. The major difference is that  $\mathbf{s}$  denotes the parameter space and follows a geostatistical model. Model states are evaluated by rerunning the simulation with updated parameters, ensuring accurate uncertainty propagation for nonlinear systems and the physicalness of model states.

[63] In the form denoted here, equation (14) is not a sequential updating scheme (yet). For trivial extension toward sequential application, the required ensemble statistics are simply updated after each step. Time-dependent problems require time-dependent models, and different data sets may resemble snapshots of the system at different times. Still, no explicit or implicit time direction is associ-

ated to equation (14) because the updated quantity  $\mathbf{s}$  is the time-invariant log conductivity field, not transient model states. The management of time-dependent data will be touched later.

### 4.2. Acceptance/Rejection Sampling

[64] The EL concept may be best and unbiased, but still remains a linearization. The actual distribution  $p(\mathbf{y})$  is not Gaussian, so the conditioning procedure will not be exact. As a test for accurate conditioning on data, the ensemble statistics of  $\mathbf{r}_i = \mathbf{y}_o - \mathbf{f}(\mathbf{s}_{c,i})$  should comply with the distribution  $p(\varepsilon)$ . The following acceptance/rejection sampling framework enforces this condition to increase the accuracy of conditional statistics.

[65] 1. For each realization, compute the critical CDF value  $Pr_i$  in the  $\chi^2$  distribution for  $\chi_i^2 = \mathbf{r}_i^T\mathbf{R}^{-1}\mathbf{r}_i$ .

[66] 2. If  $Pr_i$  is smaller than a random number drawn from the interval [0; 1], accept the  $i$ th realization as a legitimate member of the conditional ensemble. For all rejected realizations, repeatedly apply equation (14) until acceptance.

[67] Most realization-based conditional simulation methods (e.g., the pilot point method or sequential self-calibration) allow for measurement error, but do not offer a rigorous treatment of its statistics: they merely impose a convergence threshold to the measurement-simulation mismatch, which leads to conditional distributions of the mismatch with uncontrolled shape.

### 4.3. Successive Ensemble Linearization With Levenberg-Marquardt Regularization

[68] Successive linearization improves the performance of linearized approaches. For geostatistical inversion, *Carrera and Glorioso* [1991] concluded that cokriging-like techniques should be performed iteratively. Following this rationale, several iterative methods emerged, like the quasi-linear geostatistical approach by *Kitanidis* [1995] and the successive linear estimator by *Yeh et al.* [1996]. Their underlying optimization algorithms are based on the well-known Gauss-Newton method.

[69] *Dietrich and Newsam* [1989] pointed out how an artificially increased  $\mathbf{R}$  added to equation (12) stabilizes the geostatistical inverse problem while suffering a loss of information. *Nowak and Cirpka* [2004] modified this idea toward a geostatistically based Levenberg-Marquardt algorithm: the added  $\mathbf{R}$  is successively reduced to zero as the algorithm converges. This keeps the benefit of regularization while avoiding the loss of information.

[70] Within the current context, the principle of successive linearization and regularization leads to the following quasi-linear approach.

[71] 1. Generate an unconditional ensemble as described in section 4.1.

[72] 2. Choose a value  $\mathbf{R}^* > \mathbf{R}$  for use in equation (14), which leads to weaker conditioning and hence to smaller step sizes.

[73] 3. Evaluate all other ensemble statistics needed for equation (14).

[74] 4. Perform the above acceptance/rejection sampling algorithm.

[75] 5. Decrease  $\mathbf{R}^*$  toward  $\mathbf{R}$ .

[76] 6. Repeat steps 3 to 5 until  $\mathbf{R}^* > \mathbf{R}$ , or [optionally] until the overall ensemble statistics of  $\mathbf{r}_i^T\mathbf{R}^{-1}\mathbf{r}_i$  are satisfactory.

[77] This procedure improves the robustness for cases with higher variability and for data types with more non-linear relations  $\mathbf{y} = \mathbf{f}(\mathbf{s})$ .

#### 4.4. Sequential Conditioning

[78] Previous studies have shown improved efficiency and accuracy by considering separate subsets of the overall data within sequential updating schemes [e.g., *Vargas-Guzmán and Yeh, 2002*]. For example, direct measurements of parameters can be included accurately and without the inverse framework in a first conditioning stage as in the work by *Hendricks Franssen and Kinzelbach [2008a]*.

[79] The KEG follows the same approach: in the first stage, the unconditional ensemble is conditioned on all available direct measurements. Indirect data are added in a second step, using the quasi-linear setting. If desired, the second step can be split into several more steps, where indirect data could be added in their order of nonlinearity, i.e., head measurements first, then measurements of drawdown, and finally tracer data. With a smaller remaining parameter covariance after the first step, the linearized approach for indirect data has to hold only over smaller variations of the parameter field in subsequent applications of equation (14), will be more accurate, and hence will require a lower number of quasi-linear iteration steps.

[80] For application to time-dependent systems, data snapshots from different time steps may also be included sequentially, as in the original EnKF framework. While updating the snapshots sequentially, the parameters offer an increasingly good representation of the system, so that later snapshots are expected to require a lower number of quasi-linear iteration steps.

### 5. Synthetic Test Case

[81] Toward an extensive test case, the KEG is implemented in MATLAB, using standard Galerkin FEM for flow and the streamline upwind Petrov-Galerkin FEM for transport [*Hughes, 1987; Fletcher, 1996*]. The resulting equations are solved using the UMFPACK solver [*Davis, 2004*]. For random field generation, the spectral method of *Dietrich and Newsam [1993]* is implemented. The number of realizations in the ensemble is set to 2000. This relatively high number was chosen to obtain highly accurate reference statistics in the test of two-point statistics. More on the choice of ensemble size is addressed in section 7.

[82] For direct comparison, the same test case is evaluated with the quasi-linear geostatistical approach (QLGA) by *Kitanidis [1995]*, stabilized by the geostatistically driven Levenberg-Marquardt technique [*Nowak and Cirpka, 2004*], sped up with the FFT-based methods for error propagation provided by *Nowak et al. [2003]*, and equipped with adjoint state sensitivity analysis [e.g., *Townley and Wilson, 1985; Sykes et al., 1985*].

[83] The test case considers steady state groundwater flow and advective-dispersive transport of a conservative tracer in a depth-integrated confined aquifer. The domain is sized 100 m  $\times$  100 m, with Dirichlet head boundaries  $\phi = 1$  and  $\phi = 0$  on the west and east, impermeable boundaries in the north and south. A pumping well for aquifer testing is located at the domain center and pumps at 50% of the domain's total discharge. Drawdown  $\Delta\phi$  is simulated at steady state. A fixed-concentration plume from a tracer test

**Table 1.** Parameter Values Used for the Synthetic Test Case

	Parameter	Unit	Value
<i>Numerical Grid</i>			
Domain size	$[L_x, L_y]$	m	[100, 100]
Grid spacing	$[\Delta_x, \Delta_y]$	m	[1, 1]
<i>Transport Parameters</i>			
Effective porosity	$n_e$	-	0.35
Dispersivities	$[\alpha_l, \alpha_t]$	m	[2.5, 0.25]
Diffusion coefficient	$D_m$	m <sup>2</sup> /s	$10^{-9}$
<i>Geostatistical Model for <math>\ln K^a</math></i>			
Geometric mean	$K_g$	m/s	$10^{-5}$
Variance	$\sigma_Y^2$	-	1
Integral scale	$[\lambda_x, \lambda_y]$	m	[10, 10]
Microscale smoothing	$d$	m	2.5
<i>Measurement Error Standard Deviations</i>			
$\ln K$	$\sigma_{r,Y}$	-	0.25
Head $\phi$	$\sigma_{r,\phi}$	m	0.02
Drawdown $\Delta\phi$	$\sigma_{r,\Delta\phi}$	m	0.01
Concentration $c$	$\sigma_{r,c}$	-	20%

<sup>a</sup>Values are exponential.

enters the west boundary with 20 m width and  $c_0 = 1$ . Concentration  $c$  is simulated separately, not affected by the pumping test. Table 1 summarizes all relevant parameter values.

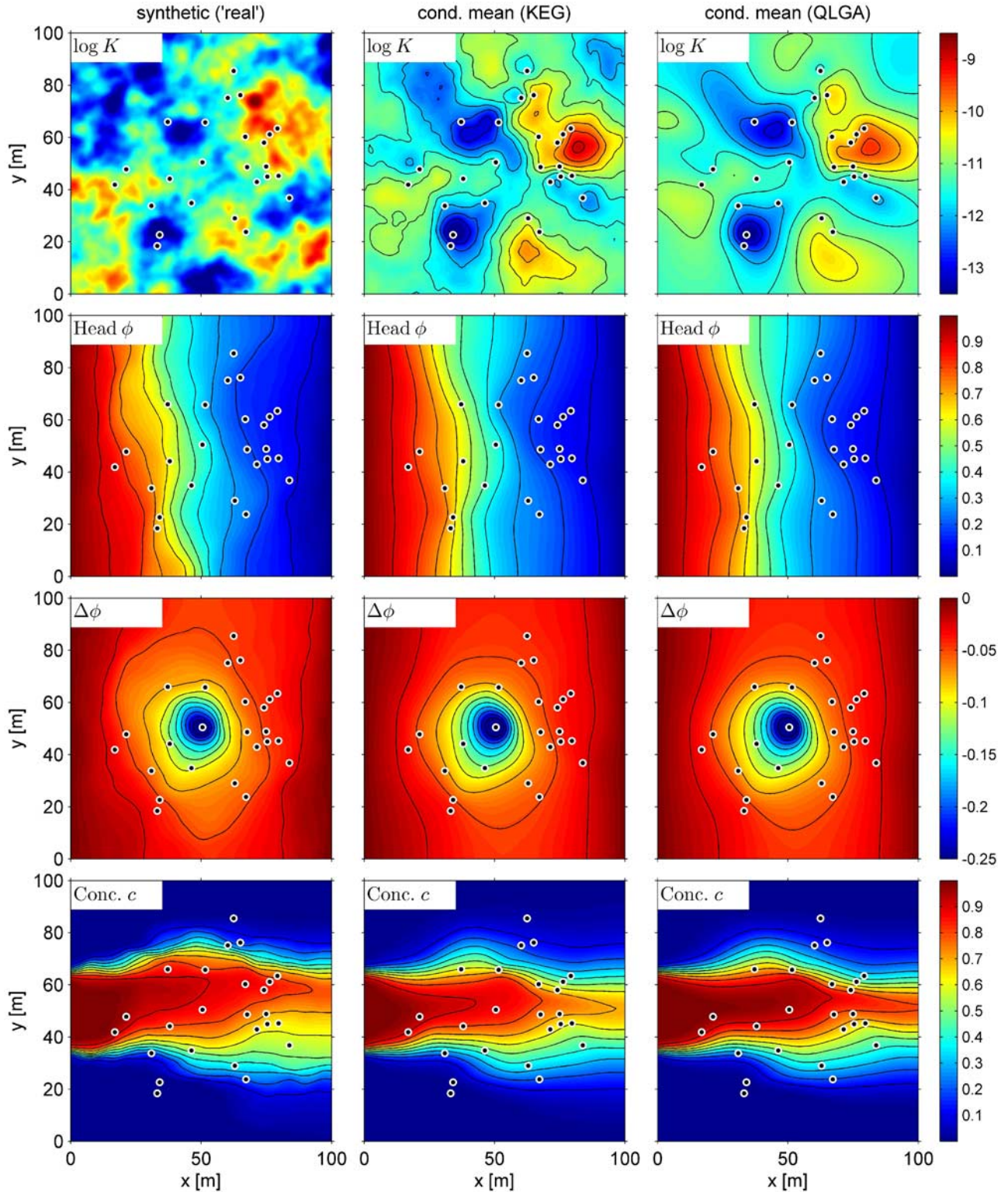
[84] Synthetic data sets are generated from random conductivity fields and their respective simulated heads, drawdowns and concentrations. The isotropic exponential model is assumed for the covariance of log conductivity, modified by a microscale smoothing parameter [e.g., *Kitanidis, 1997*]. Each data set features 25 point-like sampling locations of log conductivity, hydraulic head, drawdown and tracer concentration, summing up to 100 measurements. Measurement locations are placed randomly within the inner 80% of the domain, and are identical for all data sets. Measurement errors are assumed independent and Gaussian, defined by a standard deviation for each data type. The geostatistical parameters and measurement error levels are included in Table 1.

[85] One synthetic case is provided in the left plots of Figure 2, showing a realization of log conductivity together with its simulated head, drawdown and concentration, and the measurement locations. The results for that specific case are shown in Figures 2 and 3. The match between the synthetic field and the conditional ensemble average (Figure 2) and the standard deviation of the conditional ensemble (Figure 3) look as expected, and will not be discussed in much detail. The quality of results will be assessed in the following section on the basis of more than a trivial comparison for one single data set. The difference in the results between the KEG and the QLGA is discussed in section 8.

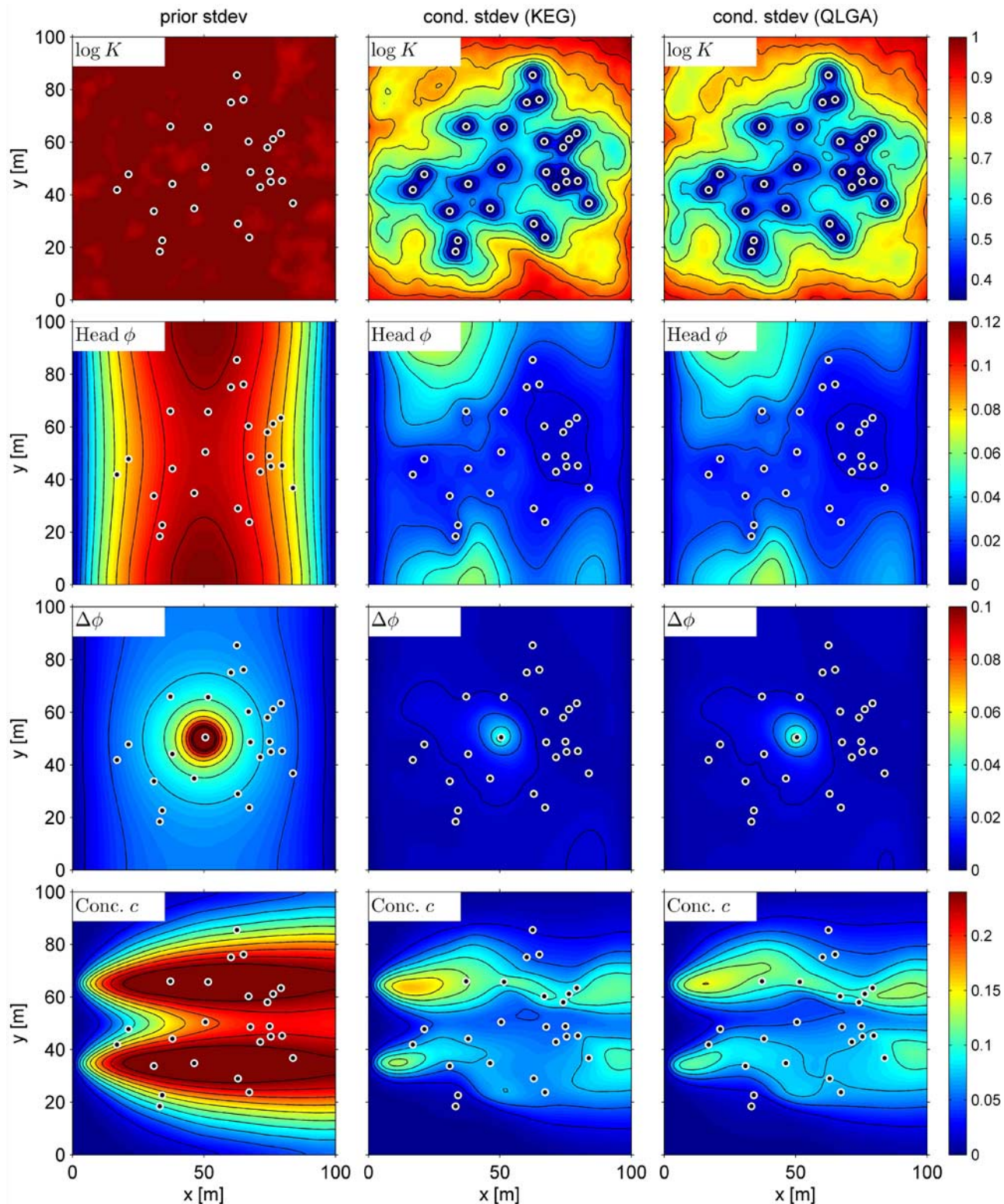
### 6. Assessment of Accuracy

[86] A single data set is insufficient for assessing the geostatistical properties and their accuracy of a conditioning method. For example, the synthetic data may display a smaller variance than average because of its limited size. The resulting log conductivity ensemble would then also display a smaller variance (because it is calibrated to values





**Figure 2.** One synthetic test case for the quasi-linear Kalman ensemble generator. (left) A synthetic log conductivity field together with simulated head, drawdown, and tracer concentration. (middle) Ensemble mean values from the quasi-linear Kalman ensemble generator. (right) Same as the middle plots but for the quasi-linear geostatistical approach. Dots are the locations of measurements; contour lines are drawn at identical levels.



**Figure 3.** Standard deviation of estimation for the synthetic test case shown in Figure 2. (left) Prior standard deviations of log conductivity, head, drawdown, and tracer concentration. (middle) Standard deviations in the conditional ensemble from the quasi-linear Kalman ensemble generator. (right) Same as the middle plots but for the quasi-linear geostatistical approach. Dots are the locations of measurements; contour lines are drawn at identical levels.



close to the mean value) and pretend the existence of a filter inbreeding problem. In order to overcome these effects, a total of 200 random data sets are used. The KEG is applied to each of them, and the desired accuracy measures are evaluated from statistics across all cases.

[87] The properties of linear estimators are well researched and rigorously defined and can be derived analytically. Of course, any method should approach the best unbiased linear estimator at the limit of linear dependence between data and parameters. The KEG fulfills this property, as follows directly from equation (14) at the limit of an infinite ensemble size and for linear  $\mathbf{f}(\mathbf{s})$ . The dependence of EnKF accuracy on the ensemble size has been investigated by *Chen and Zhang* [2006]. Because of the similarity in using ensemble statistics, an according convergence analysis for the KEG is not repeated here.

[88] For the more general nonlinear case, fewer options are available. A set of sensible and easy-to-check postulations used in the current study is that nonlinear geostatistical conditioning method (1) should not violate fundamental principles of information processing, (2) should assimilate any given data set while accurately accounting for its information and uncertainty, and (3) should honor the spatial structure of the random space function as prescribed by the prior model. These three postulations will be discussed in the following three subsections.

### 6.1. Principles of Information Processing

[89] Consistency with the fundamental principles of information processing can be tested in many ways. Here, a test based on the  $A$  measure of information in geostatistical estimation (W. Nowak, Measures of parameter uncertainty in geostatistical estimation and design, submitted to *Mathematical Geology*, 2008) is used.  $A$  is the average of all eigenvalues of  $\mathbf{G}_{\text{ss}|\mathbf{y}}$ , and also the spatial average of the estimation variance. For the test, evaluate the conditional ensemble estimation variance

$$\hat{\mathbf{s}} = \frac{1}{n_r} \sum_{i=1}^{n_r} \mathbf{s}_{ci} \quad (15)$$

$$\sigma_{\text{est}}^2 = \frac{1}{n_r - 1} \sum_{i=1}^{n_r} (\mathbf{s}_{ci} - \hat{\mathbf{s}})^2 \quad (16)$$

and assure that (1) the estimation variance is always equal or smaller than the prior variance and that (2) its spatial average  $A$  approaches zero at the limit of exhaustive sampling with exact measurements.

[90] The estimation variance stayed smaller than the prior variance of  $\sigma_Y^2 = 1$  in all but one in 200 cases. In the one exceptional case, a small spatial peak reached a value of 1.01, which is regarded as insignificant. For the current quantity and quality of data,  $A$  assumed an average value of  $A = 0.492$  over all test cases, with a very small coefficient of variation  $CV_A = 0.051$ . This indicates a highly accurate processing of information across all 200 synthetic data sets. In a series of test cases not shown here, the asymptotic approach  $A \rightarrow 0$  for increasing data quantity and quality was ensured.

[91] A more widespread and more visual test examines scatterplots of the synthetic true field versus the conditional

ensemble mean [e.g., *Chen and Zhang*, 2006; *Zhu and Yeh*, 2005; *Woodbury and Ulrych*, 2000] and then computes correlation coefficients. Nowak (submitted manuscript, 2008) showed that this intuitive approach is based on a hidden but less rigorous version of the  $A$  measure, so it is not pursued here.

### 6.2. Measurement-Simulation Mismatch

[92] The second postulation requires the normalized mismatch

$$\mathbf{r}_{c,i} = \mathbf{R}^{1/2} [\mathbf{y}_o - \mathbf{f}(\mathbf{s}_{c,i})] \quad (17)$$

to honor the assumed distribution  $p(\varepsilon) \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$  of the measurement error, normalized to a variance of unity.  $\mathbf{R}^{1/2}$  is an appropriate square root decomposition of  $\mathbf{R}$ . This postulation is explicitly enforced by the built-in acceptance/rejection scheme, so additional tests are redundant.

[93] Figure 4 compares the ensemble statistics of the measurement-simulation mismatch to the assumed distribution of measurement error. Each gray line is the normalized distribution of measurement-simulation mismatch of a particular data point, averaged over 200 conditional ensembles of 2000 realizations each. The resulting distributions should match the assumed normal distribution, normalized to a standard deviation of unity. The overall mean and two times the standard deviation (the 95% confidence interval) are indicated by the gray circles and cross marks, respectively.

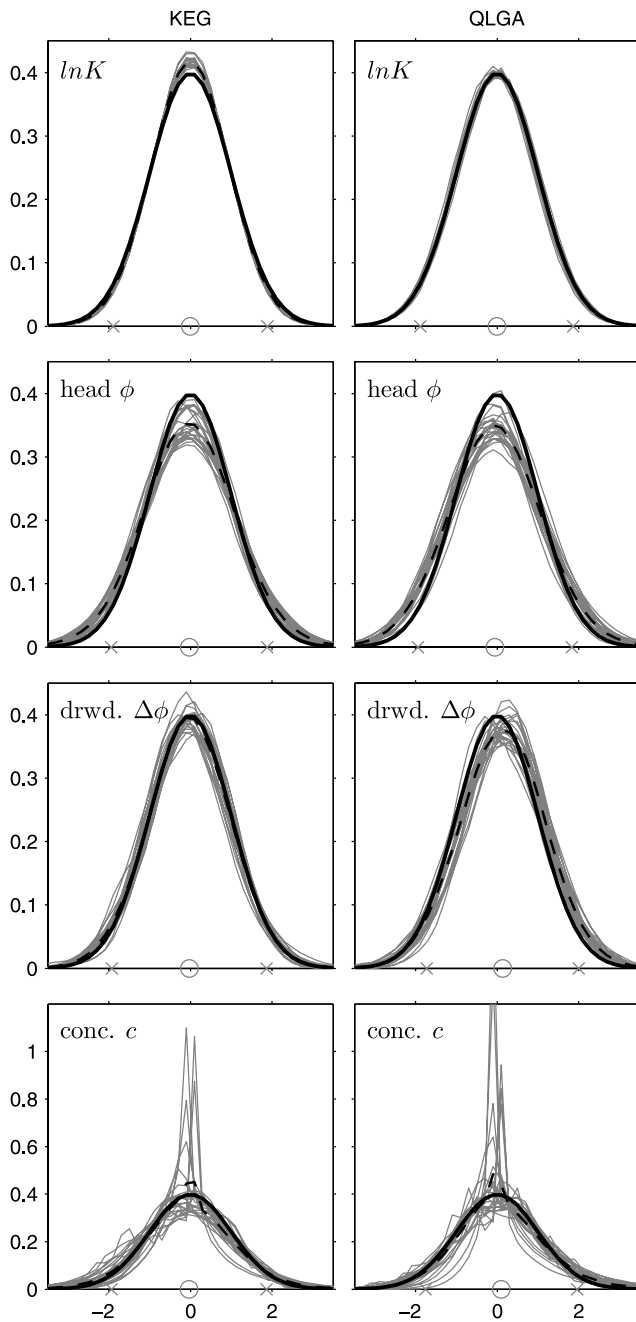
[94] The overall mean and standard deviation displays very accurate values for all four types of data. The histograms of log conductivity, heads and drawdown show an excellent fit, but some measurement locations of concentration fail to do so. The reason lies in the non-Gaussian distribution of concentration. Bounded quantities cannot assume arbitrary prescribed distributions  $p(\varepsilon)$  if the measured value is close to bounding values. Given the boundary conditions of the test case, heads and concentrations are bounded between zero and one, and drawdown is nonpositive.

[95] Theoretically, concentration follows the beta distribution [*Fiorotto and Caroni*, 2002; *Caroni and Fiorotto*, 2005] if measured in small support volumes, and approaches normality for larger support volumes [*Schwede et al.*, 2008]. Similar distributions have been found for hydraulic heads between two Dirichlet boundaries [*Nowak et al.*, 2008].

[96] In the test case, heads and drawdown appear unaffected because the measurements are not placed close to the boundaries. The locations of the concentration measurements that produce the worst fit are those where measured concentration is close to either zero or to one in most of the 200 synthetic cases. In conclusion, the data are assimilated as accurate as possible.

[97] The shown accuracy for bounded variables is an improvement over the original EnKF. *Zhou et al.* [2006, Figure 1] tested the ability of EnKFs to handle bounded state variables. Their conditional values exceeded the physically admissible range, even if the mean value was very accurate and statistical moments up to fourth order were quite acceptable. In restart versions of the EnKF and in the KEG, this could not happen because state variables are always evaluated according to the model equations, and





**Figure 4.** Ensemble statistics of measurement-simulation mismatch. Gray lines, distributions for each measurement location, averaged over 200 conditional ensembles of 2000 realizations each. Black solid lines, assumed normal distribution of measurement error. Black dashed lines, average over all distributions of one data type. All distributions are normalized by the respective standard deviation of measurement error. Gray circles and crosses, overall mean and double standard deviation for each data type.

never directly updated. The acceptance/rejection sampling further adds to the accuracy of conditional statistics.

[98] Unfortunately, Kalman filters and its derivatives are optimal only for multi-Gaussian distributions of all involved variables. Further improvement might be possible via suitable transforms to render the data approximately univariate Gaussian, which is a necessary but not sufficient condition

for multi-Gaussianity. A promising technique along these lines is the Box-Cox transform for concentration data, which includes the log transform as special case [Kitanidis and Shen, 1996]. The degree of possible improvement in conditional simulation should be a subject for further investigation.

### 6.3. Fidelity Toward Spatial Statistics

[99] Even if a random field is large enough to be ergodic, a data set of limited size (which is only a subset of the field) is not ergodic, so each data sets leaves its nonergodic fingerprint on the respective conditional ensemble. Therefore, to check the third requirement, the only rigorous option is to combine all conditional ensembles into one. Only the combined ensemble can be asked to match the prior distribution  $p(\mathbf{s})$ . This follows from

$$\begin{aligned} E_{\mathbf{y}}[p(\mathbf{s}|\mathbf{y})] &= \int p(\mathbf{s}|\mathbf{y})p(\mathbf{y}) d\mathbf{y} \\ &= \int p(\mathbf{s},\mathbf{y})d\mathbf{y} = p(\mathbf{s}). \end{aligned} \quad (18)$$

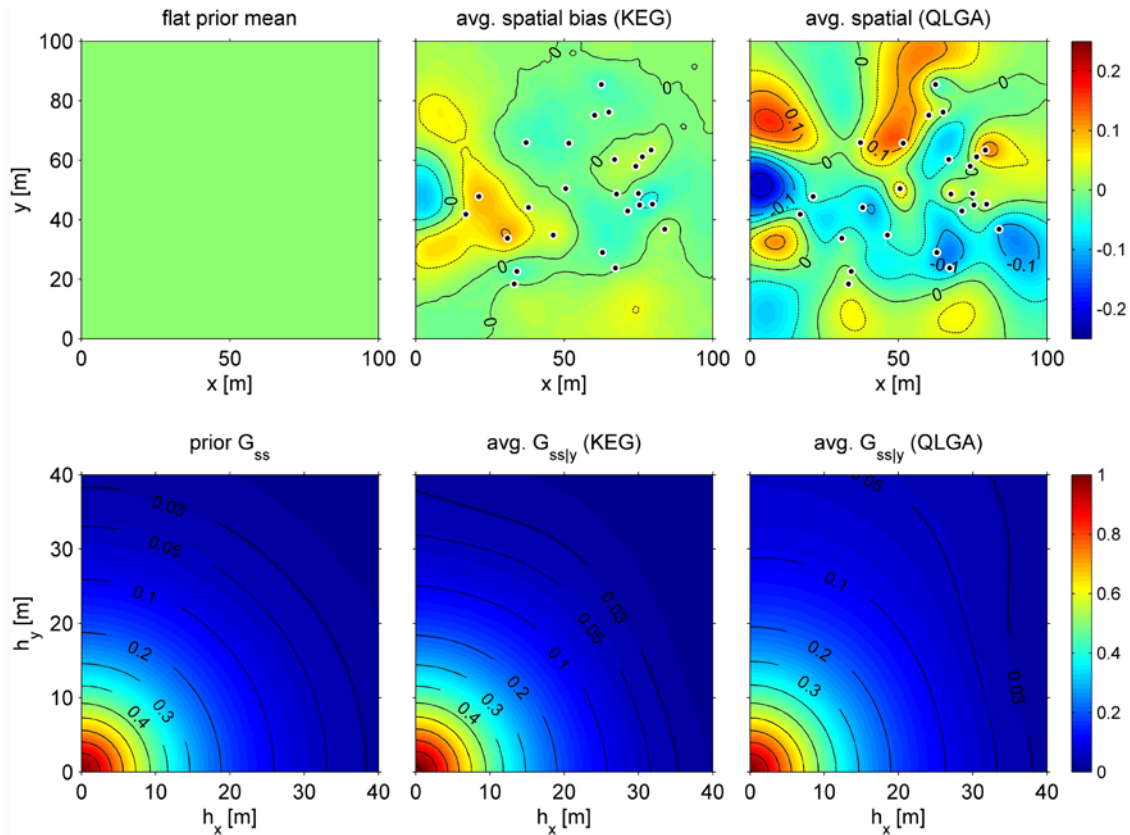
[100] When testing only with a single data set, the expected value over  $\mathbf{y}$  is not considered. In the current case, the overall combined ensemble has to exhibit the assumed prior mean and the prior covariance  $\mathbf{G}_{ss}$ .

[101] The difference  $b$  between the spatial average of  $\hat{\mathbf{s}}$  and the prior mean  $\ln K_g$  is an indication of filter bias. The average value of  $b$  over all test cases is  $\bar{b} = 0.015$ , corresponding to a 1.49% deviation from the true value of  $K_g$ . Given the variance  $\sigma_Y^2 = 1$  and the standard deviation of the filter bias  $\sigma_b^2 = 0.240$ ,  $\bar{b}$  is not significantly different from zero. Figure 5 (top) compares the flat prior mean of  $\ln K(x)$  with the conditional mean  $\ln K(x)$  field, averaged over all 200 cases. It resembles the flat prior mean with a low degree of spatial trends or fluctuations. In conclusion, this confirms the unbiasedness of the KEG.

[102] The case-averaged covariance function is compared to the assumed covariance model in Figure 5 (bottom). The overall structure has been preserved to a very high degree. Close comparisons of the contour lines reveal that the KEG introduced slightly smaller correlation over large distances, and a slightly increased overall variance (i.e., the covariance for zero separation distance) from  $\sigma_Y^2 = 1$  to  $\sigma_Y^2 = 1.019$ . Considering that  $\sigma_Y^2$  itself has a standard deviation between individual test cases of 0.192, this deviation is statistically not significant and could vanish at a higher number of test cases.

## 7. Computational Efficiency and Robustness

[103] The computational efficiency of EnKFs for large problems is well established. Their well-known main advantage lies in avoiding sensitivity analyses that require many calls of the numerical model [e.g., Evensen, 2003; Chen and Zhang, 2006], especially for large data sets. The same beneficial properties apply to the new KEG. In the previous section, 200 test cases were performed with 2,000 conditional realizations per test case (not counting the initial generation of the unconditional ensemble). The average computational cost in the test cases was 4,291 calls to the simulation code for the 2,000 conditional realizations in each case. These computational costs are significantly smaller than for alternative realization-based



**Figure 5.** Conservation of spatial structure by the quasi-linear Kalman ensemble generator. (top) Comparison of (left) assumed flat prior mean and ensemble means in the union of the 200 conditional ensembles for (middle) the KEG and (right) the QLGA. (bottom) Comparison of (left) assumed covariance model and covariance in the union of the 200 conditional ensembles for (middle) the KEG and (right) the QLGA.

methods. For the same ensemble size, the pilot point method would take 150,000 calls to obtain sensitivities for each indirect measurement (represented in this example by just one pilot point per measurement), plus additional calls for its iterative procedure.

[104] For fairness in comparison, two additional points should be mentioned. First, ensemble-based methods require a certain number of realizations in order to achieve accurate corrections of its individual realizations, whereas realization-based methods correct individual realizations. If only a smaller ensemble size is required for a given application (e.g., only 200 realizations), then the pilot point method would only take in the order of 15,000 simulation calls plus the iterative effort. Second, the 2,000 realizations chosen in the current test case are a relatively high number to exclude with certainty any effects of limited ensemble size. Typical numbers for ensemble-based methods range about 500 [e.g., *Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008b*].

[105] At only 2.15 calls to the simulation model per average realization, the computational costs are amazingly low. The coefficient of variation for computational costs was 0.2538. The average number of quasi-linear iteration steps was 2.165 with a *CV* of 0.2527. This is a striking demonstration for the computational efficiency, and documents the robustness of the iterative procedure.

[106] A second aspect of computational efficiency is that storage requirements of some methods may restrict the

freedom in choosing geostatistical models or the spatial resolution of  $\mathbf{s}$  [*Zimmerman et al., 1998*]. Traditional linearized error propagation via equation (12) quickly becomes cumbersome for large domains, up to the point where explicit storage of  $\mathbf{C}_{ss}$  exceeds the capability of arrays of modern hard disk drives [*Zimmerman, 1989; Nowak et al., 2003*]. When assuming stationarity or intrinsicity (or certain simple cases of nonstationarity) in conjunction with regular equispaced grids, FFT-based methods for error propagation [*Nowak et al., 2003; Círpka and Nowak, 2004*] avoid this problem. Extensions to irregular grids are offered by *Fritz et al. [2009]* and *Li and Círpka [2006]*.

[107] More flexibility, free of any such assumptions, is provided by the pilot point method of *RamaRao et al. [1995]*. It can use any arbitrarily complex geostatistical model, given an adequate field generator. The same flexibility is offered by the KEG, when replacing the deconvolution in equation (9) by the approximation in equation (10). Given an adequate field generator, the huge autocovariance matrix of the unknown parameters is obsolete, allowing for larger problems and finer resolutions.

## 8. Comparison to the Quasi-Linear Geostatistical Approach

[108] The QLGA is a classical representative of methods that use Jacobian-based linearizations, implemented with

adjoint state sensitivities in the current study. It is a well-researched and efficient method. At first sight, its results (Figures 2 and 3, right) may not seem to differ drastically from the results of the KEG.

[109] A closer look reveals that it is affected by the biasedness and nonminimal approximation error of Jacobian-based linearizations that were discussed in the introduction. For example, the conditional mean of concentration by the QLGA (Figure 2, bottom right) is less dispersive than the corresponding synthetic field (see, for example, the longer persistence of the inner isocontour lines), because it is computed with local dispersivities on the smooth conditional mean conductivity field. The conditional mean of concentration has to be more dispersive than the synthetic truth because of the remaining uncertainty within the conditional ensemble. As discussed in depth by *Nowak and Cirpka* [2006], the interpretation of tracer data with local dispersivities on smooth estimated fields is a discrepancy that inevitably leads to inaccurate data interpretation, scale inconsistencies and convergence problems.

[110] The QLGA allows generating conditional realizations on the basis of the conditional covariance of the estimated field. In the current study, this is done without further iteration on the basis of the linearization about the conditional mean. For each synthetic data set from the test case, 2000 conditional realizations were generated on the basis of this technique. The conditional standard deviation for dependent variables (Figure 3, right) is evaluated from such a conditional ensemble. The resulting histograms from all 200 cases are shown in the right plots of Figure 4. Apparent differences are a bias toward stronger drawdown, a bias toward higher concentrations, and a worse histogram fit for concentration data.

[111] For the QLGA, the overall filter bias  $b$  is negligible, but the conditional mean field averaged across all 200 test cases (Figure 5, top right) shows a pronounced spatial pattern of estimation bias compared to the results of the KEG. The case-averaged covariance function (Figure 5, bottom right) shows that the QLGA has introduced a significant increase of long-range correlation because of its weaker accuracy in handling nonlinear data.

[112] The author expects the bias to be more pronounced for weaker data sets, when the conditional covariance is larger and the linearization has to hold over larger intervals. Also, smaller dispersivities will further reduce the accuracy of the QLGA with respect to concentration data, since the discrepancy between ensemble dispersion and local dispersion increases with decreasing local dispersion.

[113] Thanks to the modified Levenberg-Marquardt algorithm, the QLGA converged in an average of 3.95 quasi-linear iteration steps, and required 407 calls to the simulation model on average. This is mostly for the adjoint state sensitivity analysis in each iteration step, where conductivity data require no simulation call, heads and drawdown data require one, and concentration data require two simulation calls each. Because of the inconsistency of tracer data interpretation on smooth estimated fields, the QLGA converged to estimates with mostly unacceptable statistics of measurement-simulation mismatch.

[114] When using numerical differentiation instead of adjoint state sensitivities, the computational costs of the QLGA would have risen by a factor of 100, far above the

computational costs of the KEG. The disadvantage of relying on adjoint state sensitivity analysis is that they require a certain freedom of modifying simulation codes. This is hardly possible for commercial software. The KEG does not rely on adjoint states and so is more compatible with arbitrary commercial simulation codes, while being computationally much more efficient than the QLGA with numerical differentiation. In summary, the computational costs of the QLGA are below those of the KEG, but the savings come at the price of four drawbacks: (1) bias and a lower accuracy, (2) inconsistent interpretation of tracer data on smooth estimated fields with local dispersivities, (3) an extensive list of methodical add ons is required compared to the lightweight implementation of the KEG, and (4) a lower flexibility in the choice of commercial simulation software due to the use of adjoint state sensitivities.

## 9. Summary and Conclusions

[115] This study has pursued and combined two investigations: (1) the concept of best unbiased ensemble linearization and (2) its application to generate ensembles of log conductivity conditional on head and tracer data.

[116] The concept of best unbiased ensemble linearization for the stochastic groundwater flow and transport equation (or any other stochastic partial differential equation) has been derived. It may be called best and unbiased because the error between exact function and linearized approximation is zero on average and minimal in the mean square sense. Its key properties are as follows.

[117] 1. The best unbiased ensemble linearization is a best fit secant to the original function over the entire population rather than a tangent.

[118] 2. The unbiasedness condition requires the secant to pass through the expected values of the dependent state variable at the expected value of the parameters. This is regardless of the slope.

[119] 3. The minimum error variance condition requires a slope such that the linearized cross covariance between parameters and modeled quantity meets their exact cross covariance, resembling an effective average slope rather than a traditional Jacobian matrix of model sensitivities.

[120] 4. Using a best unbiased ensemble linearization implies to use empirical mean values, covariances and variances from Monte Carlo analyses.

[121] Application of this principle to geostatistical conditioning problems lead to the new quasi-linear Kalman ensemble generator (KEG). The KEG is the transformation of the ensemble Kalman filter (EnKF) idea to geostatistical inverse problems. The EnKF is a state space method that updates time-dependent model predictions in real-time applications. The KEG, in contrast, is a parameter space method that generates an ensemble of log conductivity fields (and arbitrary additional uncertain parameters) conditional to arbitrary data. State variables are updated through simulation with updated parameters, which always ensures the physicalness of all states and improves the conditional statistics.

[122] The similarity between both methods is that they use ensemble covariances instead of first-order approximations for their conditioning steps. This makes implicit use of a best unbiased ensemble linearization. The derivation of ensemble Kalman filters has not been aware of the unbi-



asedness and minimum error variance property of its implicit linearization. The idea to use ensemble means and covariances has been supported with a new theoretical basis in the current study. The resulting conditional ensembles display a high accuracy of their desired statistics at intriguingly low computational costs.

[123] At the same time, known issues with solute dispersion in estimation problems are avoided. It is common to all estimation methods that estimated conductivity fields are inadequate for transport simulations with hydrodynamic dispersion coefficients. Since EnKFs and the KEG are conditional simulation tools rather than estimation tools, they avoid this inconsistency.

[124] Similar to the two-step approach of *Hendricks Franssen and Kinzelbach* [2008a], the KEG has been equipped with a two-stage updating scheme that first uses direct measurements of log conductivity. Indirect data with a nonlinear relation to log conductivity are processed in a separate second step. The current study improved the accuracy of the second step by a quasi-linear iteration scheme, combined with an acceptance/rejection sampling scheme.

[125] Further improvements may be achieved by suitable transforms of non-Gaussian data (such as concentrations) to almost-Gaussian forms. By ensuring at least univariate normality, the optimality of Kalman filters and for multi-Gaussian relations may be exploited to a larger extent. The degree of possible improvement should be subject for further investigations. The KEG is not meant to replace Monte Carlo Markov chain methods such as the one by *Zanini and Kitanidis* [2008]. Instead, it would be desirable to combine these two methods in later developments.

[126] The accuracy, robustness and efficiency of the KEG were positively assessed in a large-scale series of test cases. 200 synthetic data sets of log conductivity, hydraulic heads, drawdown and tracer data were generated. Each data set was used to condition an ensemble of 2,000 realizations, and the statistics of performance were discussed. When using only direct data, the KEG is a best unbiased simulator. When including hydraulic heads, drawdown data and tracer data, the accuracy of conditional simulations versus the measured data values was highly satisfactory across all test cases. As test for fidelity toward a prescribed geostatistical model, all 200 conditional ensembles were combined, and the overall mean and covariance function assessed via variogram analysis. The resulting covariance function did not differ significantly from the prescribed covariance model used for generation. Also, no significant filter bias or signs of filter inbreeding could be found.

[127] The computational efficiency is extremely high. On average over the 200 test cases, the KEG required only 2.15 calls to the simulation model per conditional realization, or 4,291 calls for conditioning an ensemble of 2000 realizations. The author expects that for sequential updating in time-dependent systems, the number of iterations for later data sets will decrease, because more information has already been absorbed in the ensemble. For the same ensemble size, the pilot point method would take about 150,000 calls (at one pilot point per indirect measurement), but the ratios depend of course on the desired ensemble sizes. Using ensemble statistics avoids storing or handling the autocovariance matrix of log conductivity, which has

been a limitation to the allowable problem size of many conditioning methods in the past.

[128] The same series of test cases was performed with the quasi-linear geostatistical approach (QLGA) by *Kitanidis* [1995] for direct comparison to Jacobian-based methods. Only when equipped with an extensive list of methodical upgrades, the QLGA is computationally more efficient than the KEG. For example, the QLGA can do without adjoint state sensitivities, but its computational effort would rise by 2 or more orders of magnitude, depending on the spatial resolution of the conductivity field. Compared to the QLGA and comparable Jacobian-based methods, the KEG overcomes a list of drawbacks.

[129] 1. It is unbiased with minimal approximation error.

[130] 2. It avoids the inconsistent interpretation of tracer data on smooth estimated fields.

[131] 3. Its implementation is easy and does not require an extensive list of methodical add ons.

[132] 4. It offers a maximum flexibility in the choice of commercial simulation software due to its nonintrusive Monte Carlo-like character, e.g., by avoiding adjoint state sensitivities.

[133] In combination, the suggested KEG method offers an accuracy that compares to expensive realization-based methods at a computational efficiency almost as low as quasi-linear estimation methods. The final conclusion is that the KEG (just like the ensemble Kalman filter) can be used with confidence for geostatistical applications. Further promising fields of application include geostatistical optimal design of site exploration W. Nowak and F. P. J. de Barros (Bayesian geostatistical design: Optimal site investigation when the geostatistical model is uncertain, manuscript in preparation, 2009) successfully apply the KEG for that purpose in a parallel study. At the same time, they extend the KEG to estimate covariance parameters from the data, similar to the quasi-linear geostatistical approach by *Kitanidis* [1995].

## Appendix A: Unbiasedness

[134] The error of linearization is defined by

$$\varepsilon = f(\mathbf{s}) - y_0 - \mathbf{H}(\mathbf{s} - \mathbf{s}_0). \quad (\text{A1})$$

Inserting into the unbiasedness condition (equation (2)) leads to

$$E[f(\mathbf{s}) - y_0 - \mathbf{H}(\mathbf{s} - \mathbf{s}_0)] = \bar{y} - y_0 - \mathbf{H}(\bar{\mathbf{s}} - \mathbf{s}_0) = 0, \quad (\text{A2})$$

where  $\bar{y} = E[f(\mathbf{s})]$  and  $\bar{\mathbf{s}} = E[\mathbf{s}]$ . Equation (A2) requires that, regardless of  $\mathbf{H}$ , the condition  $f(\bar{\mathbf{s}}) = \bar{y}$  has to be fulfilled. Without loss of generality, one may choose

$$\mathbf{s}_0 = \bar{\mathbf{s}} \quad (\text{A3})$$

$$y_0 = \bar{y}. \quad (\text{A4})$$

## Appendix B: Minimum Squared Error

[135] Together with equations (4) and (5), the linearization and its error become

$$\tilde{f}(\mathbf{s}) = \bar{y} + \mathbf{H}(\mathbf{s} - \bar{\mathbf{s}}) \quad (\text{B1})$$

$$\varepsilon = f(\mathbf{s}) - \bar{y} - H(\mathbf{s} - \bar{\mathbf{s}}). \quad (\text{B2})$$

For notational convenience, set  $f(\mathbf{s}) - \bar{y} = y'$  and  $\mathbf{s} - \bar{\mathbf{s}} = \mathbf{s}'$ . Then

$$E[\varepsilon^2] = E[(y' - \mathbf{H}\mathbf{s}')^T (y' - \mathbf{H}\mathbf{s}')^T] \\ = \sigma_y^2 - \mathbf{q}_{ys}\mathbf{H}^T - \mathbf{H}\mathbf{q}_{sy} + \mathbf{H}\mathbf{Q}_{ss}\mathbf{H}^T, \quad (\text{B3})$$

where  $\sigma_y^2 = E[(y')^2]$  is the variance of  $y$ ,  $\mathbf{q}_{ys} = \mathbf{q}_{sy}^T = E[y'(\mathbf{s}')^T]$  is the cross covariance between  $y$  and  $\mathbf{s}$ , and  $\mathbf{Q}_{ss} = E[(\mathbf{s} - \bar{\mathbf{s}})(\mathbf{s} - \bar{\mathbf{s}})^T]$  is the autocovariance of  $\mathbf{s}$ . Identical to the minimum estimation variance in kriging, condition (3) leads to normal equations:

$$E[\varepsilon^2] \rightarrow \min \Rightarrow \frac{\partial}{\partial \mathbf{H}} E[\varepsilon^2] = 0, \quad (\text{B4})$$

which is in this case

$$0 = \frac{\partial}{\partial \mathbf{H}} [\sigma_y^2 - \mathbf{q}_{ys}\mathbf{H}^T - \mathbf{H}\mathbf{q}_{sy} + \mathbf{H}\mathbf{Q}_{ss}\mathbf{H}^T] \\ = -2\mathbf{q}_{sy} + 2\mathbf{Q}_{ss}\mathbf{H}^T. \quad (\text{B5})$$

This results in

$$\mathbf{Q}_{ss}\mathbf{H}^T = \mathbf{q}_{sy} \\ \iff \mathbf{H} = (\mathbf{Q}_{ss}^{-1}\mathbf{q}_{sy})^T. \quad (\text{B6})$$

Inserting into equation (B3) yields

$$E[\varepsilon^2] = \sigma_y^2 - \mathbf{H}\mathbf{q}_{sy} = \sigma_y^2 - \mathbf{H}\mathbf{Q}\mathbf{H}^T. \quad (\text{B7})$$

[136] **Acknowledgments.** This study has been funded by the Deutsche Forschungsgemeinschaft (DFG) under grant NO 805/1-1. The author is indebted to Yoram Rubin for hosting and discussion and would like to thank Zepu Zhang for discussions about statistical topics and Felipe de Barros and Erika Bäcker for editorial comments. The constructive review comments of H.-J. Hendricks Franssen and two other reviewers helped to strengthen this manuscript.

## References

- Alcolea, A., J. Carrera, and A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Res.*, 29, 1678–1689.
- Burgers, G., P. J. V. Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126, 1719–1724.
- Capilla, J. E., J. J. Gómez-Hernández, and A. Sahuquillo (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 2, demonstration on a synthetic aquifer, *J. Hydrol.*, 203, 175–188.
- Caroni, E., and V. Fiorotto (2005), Analysis of concentration as sampled in natural aquifers, *Transp. Porous Media*, 59, 19–45, doi:10.107/s11,242-004-1119-x.
- Carrera, J., and L. Glorioso (1991), On geostatistical formulation of the groundwater flow inverse problem, *Adv. Water Resour.*, 14, 273–283.
- Chan, R. H., and M. K. Ng (1996), Conjugate gradient methods for Toeplitz systems, *SIAM Rev.*, 38(3), 427–482.
- Chen, Y., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv. Water Resour.*, 29, 1107–1122.
- Cirpka, O. A., and P. K. Kitanidis (2001), Sensitivity of temporal moments calculated by the adjoint-state method, and joint inverting of head and tracer data, *Adv. Water Resour.*, 24, 89–103.
- Cirpka, O. A., and W. Nowak (2004), First-order variance of travel time in nonstationary formations, *Water Resour. Res.*, 40, W03507, doi:10.1029/2003WR002851.
- Cirpka, O. A., C. M. Bürger, W. Nowak, and M. Finkel (2004), Uncertainty and data worth analysis for the hydraulic design of funnel-and-gate systems in heterogeneous aquifers, *Water Resour. Res.*, 40, W11502, doi:10.1029/2004WR003352.
- Davis, T. A. (2004), Algorithm 832: UMFPACK v.4.3—An unsymmetric-pattern multifrontal method, *ACM Trans. Math. Software*, 30(2), 196–199.
- Dietrich, C. R., and G. N. Newsam (1989), A stability analysis of the geostatistical approach to aquifer transmissivity identification, *Stochastic Hydrol. Hydraul.*, 3, 293–316.
- Dietrich, C. R., and G. N. Newsam (1993), A fast and exact method for multidimensional Gaussian stochastic simulations, *Water Resour. Res.*, 29(8), 2861–2869.
- Diggle, P. J., and P. J. Ribeiro (2007), *Model-Based Geostatistics*, Springer, New York.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99(C5), 10,143–10,162.
- Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, 53, 343–367, doi:10.1007/s10236-003-0036-9.
- Evensen, G. (2007), *Data Assimilation—The Ensemble Kalman Filter*, Springer, Berlin.
- Feyen, L., and S. M. Gorelick (2005), Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas, *Water Resour. Res.*, 41, W03019, doi:10.1029/2003WR002901.
- Fiorotto, V., and E. Caroni (2002), Solute concentration statistics in heterogeneous aquifers for finite Péclet values, *Transp. Porous Media*, 48, 331–351.
- Fletcher, C. A. J. (1996), *Computational Techniques for Fluid Dynamics*, vol. 1, *Fundamental and General Techniques*, 2nd ed., Springer, New York.
- Fritz, J., W. Nowak, and I. Neuweiler (2009), Application of FFT-based algorithms for large-scale universal kriging problems, *Math. Geol.*, in press.
- Gómez-Hernández, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory, *J. Hydrol.*, 203, 162–174.
- Hendricks Franssen, H. J., and W. Kinzelbach (2008a), 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 (2008b), Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems, *J. Hydrol.*, 365, 261–274, doi:10.1016/j.jhydrol.2008.11.033.
- Hendricks Franssen, H. J., J. J. Gómez-Hernández, and A. Sahuquillo (2003), Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data, *J. Hydrol.*, 281, 281–295.
- Herrera, G. S. (1998), Cost effective groundwater quality sampling network design, Ph.D. thesis, Univ. of Vermont, Burlington.
- Herrera, G. S., and G. F. Pinder (2005), Space-time optimization of groundwater quality sampling networks, *Water Resour. Res.*, 41, W12407, doi:10.1029/2004WR003626.
- Hughes, J. R. (1987), *Finite Element Method*, Prentice-Hall, Englewood Cliffs, N. J.
- Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, *Math. Sci. Eng.*, vol. 64, Academic, San Diego, Calif.
- Keidser, A., and D. Rosbjerg (1991), A comparison of four inverse approaches to groundwater flow and transport parameter identification, *Water Resour. Res.*, 27(9), 2219–2232.
- Kitanidis, P. K. (1986), Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, 22(4), 499–507.
- Kitanidis, P. K. (1993), Generalized covariance functions in estimation, *Math. Geol.*, 25(5), 525–540.
- Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inverting, *Water Resour. Res.*, 31(10), 2411–2419.
- Kitanidis, P. K. (1996), On the geostatistical approach to the inverse problem, *Adv. Water Resour.*, 19, 333–342.
- Kitanidis, P. K. (1997), *Introduction to Geostatistics*, Cambridge Univ. Press, Cambridge, U. K.
- Kitanidis, P. K., and R. W. Lane (1985), Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *J. Hydrol.*, 79, 53–71.

- Kitanidis, P. K., and K.-F. Shen (1996), Geostatistical interpolation of chemical concentration, *Adv. Water Res.*, 19, 369–378.
- Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resour. Res.*, 19(3), 677–690.
- LaVenue, A. M., B. S. RamaRao, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 2. Application, *Water Resour. Res.*, 31(3), 495–516.
- Li, W., and O. A. Cirpka (2006), Efficient geostatistical inverse methods for structured and unstructured grids, *Water Resour. Res.*, 42, W06402, doi:10.1029/2005WR004668.
- 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, 975–986.
- Matheron, G. (1971), *The Theory of Regionalized Variables and Its Applications*, Ecole de Mines, Fontainebleau, France.
- Nowak, W. (2005), Geostatistical methods for the identification of flow and transport parameters in subsurface flow, Ph.D. thesis, Inst. fuer Wasserbau, Univ. Stuttgart, Stuttgart, Germany. (Available at [http://elib.uni-stuttgart.de/opus/frontdoor.php?source\\_opus=2275](http://elib.uni-stuttgart.de/opus/frontdoor.php?source_opus=2275))
- Nowak, W., and O. A. Cirpka (2004), A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inverting, *Adv. Water Resour.*, 27, 737–750.
- Nowak, W., and O. A. Cirpka (2006), Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data, *Water Resour. Res.*, 42, W08416, doi:10.1029/2005WR004832.
- Nowak, W., S. Tenkleve, and O. A. 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, W08452, doi:10.1029/2007WR006383.
- RamaRao, B. S., A. M. LaVenue, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resour. Res.*, 31(3), 475–493.
- Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., John Wiley, New York.
- Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, Oxford, U. K.
- Rubin, Y., A. Sun, R. Maxwell, and A. Bellin (1999), The concept of block-effective macrodispersivity and a unified approach for grid-scale- and plume-scale-dependent transport, *J. Fluid Mech.*, 395, 161–180.
- Rubin, Y., A. Bellin, and A. E. Lawrence (2003), On the use of block-effective macrodispersion for numerical simulations of transport in heterogeneous formations, *Water Resour. Res.*, 39(9), 1242, doi:10.1029/2002WR001727.
- Schwede, R. L., O. A. 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.
- Schwepe, F. C. (1973), *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, N. J.
- Sykes, J. F., J. L. Wilson, and R. W. Andrews (1985), Sensitivity analysis for steady-state groundwater-flow using adjoint operators, *Water Resour. Res.*, 21(3), 359–371.
- Townley, L. R., and J. L. Wilson (1985), Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow, *Water Resour. Res.*, 21(12), 1851–1860.
- Vargas-Guzmán, J. A., and T.-C. J. Yeh (2002), The successive linear estimator: A revisit, *Adv. Water Resour.*, 25, 773–781.
- Wen, X.-H., and W. Chen (2006), Real-time reservoir updating using ensemble Kalman filter: The confirming approach, *Soc. Pet. Eng. J.*, 11(4), 431–442.
- Woodbury, A. D., and T. J. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, 36(8), 2081–2093.
- Yeh, T.-C. J., M. Jin, and S. Hanna (1996), An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, 32(1), 85–92.
- Zanini, A., and P. K. Kitanidis (2008), Geostatistical inverting for large-contrast transmissivity fields (online), *Stochastic Environ. Res. Risk Assess.*, doi:10.1007/s00477-008-0241-7.
- Zhang, D. (2002), *Stochastic Methods for Flow in Porous Media*, Academic, San Diego, Calif.
- Zhang, Y., G. F. Pinder, and G. S. Herrera (2005), Least cost design of groundwater quality monitoring networks, *Water Resour. Res.*, 41, W08412, doi:10.1029/2005WR003936.
- Zhou, Y., D. McLaughlin, and D. Entekhabi (2006), Assessing the performance of the ensemble kalman filter for land surface data assimilation, *Mon. Weather Rev.*, 134, 2128–2142.
- 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.
- Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, 34(6), 1373–1413.
- Zimmerman, D. L. (1989), Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models, *J. Stat. Comput. Simul.*, 32(1–2), 1–15.

---

W. Nowak, Institute of Hydraulic Engineering, Universität Stuttgart, D-70569 Stuttgart, Germany. (wolfgang.nowak@iws.uni-stuttgart.de)