hess-2015-189

Joint Inference of Groundwater-Recharge and Hydraulic-Conductivity Fields from Head Data using the Ensemble-Kalman Filter.

D. Erdal and O. A. Cirpka

Revision highlights

- The text is altered such that it is clearer that the importance of the prior, as exemplified in the manuscript, is strongly related to the non-uniqueness of the data with respect to the estimated parameters.
- New texts included to highlight that the addition of informative data to the estimation process would alleviate the need for a strong prior.
- New figure with the ensemble variance included for the joint estimation simulations.
- Numerous small changes to clear misunderstandings and improve and clarify our argumentation.

In the following, specific answers to all comments from both reviewers will be presented. The reply is structured as follows:

Reviewer text

Author reply as published on HESSD

Specific reply describing to the actual changes applied to the manuscript

Reviewer #1

MAIN COMMENT

The paper discusses the interesting issue of the difficulty to identify simultaneously hydraulic conductivity and recharge from piezometric head observations. For this purpose the authors use the ensemble Kalman filter and build a synthetic experiment to conclude that unless the prior information used to generate the initial set of realizations is "correct" it is impossible to identify simultaneously both hydraulic conductivity and recharge.

Although the authors prove nicely their point, the arguments given to justify the final results are flawed: the emphasis should not be put in the prior information but rather in the need of extra information to be able to single out the combination of hydraulic conductivity and recharge that is correct out of the many combinations that are coherent with the observed piezometric head.

It is wrong to say that the prior information used to generate the initial set of ensemble realizations determines what the final estimates will be (after data assimilation) and that you cannot generate realizations outside the prior random function model. The final ensemble of realizations can largely depart from the initial one, when the observations are inconsistent with the prior model. The sequential set of ensembles that are obtained after each assimilation step can be interpreted as a Markov chain that will "forget" the structure built in the initial ensemble after some assimilation steps. More so, the updating of the ensemble realizations is solely based on the covariance structure of the ensemble, and for this reason, there is a tendency for the final ensemble of estimates to converge towards realizations drawn from a multiGaussian function, even when the initial ensemble is far from being multiGaussian. Enough observation data can change completely the random function of the final ensemble with regard to the one used to generate the initial ensemble.

What happens in the example presented by the authors is that the observations are consistent with all prior models used. In fact, the reference could be the one used, or it could be a realization generated with the "wrong" model, the results would be the same. Therefore, you need some additional piece of information, to discriminate among the different prior models which one is the one consistent with your unknown reality. It is not that the EnKF does not work when the prior model is incorrect. Knowing which is the orientation of the deposition will draw to prefer a prior model over another.

My request is that the paper be rewritten removing all these comments about the importance about the prior information, and the influence that this prior information has in the final ensemble, and replacing it by talking about the importance of having additional information that would allow you to discriminate among alternative models. I contend that assimilating data on fluxes or concentrations would also alleviate the problem of the prior model. And I insist that the prior model information will fade away as time passes and data are assimilated, and could vanish if the prior model is inconsistent with the observational data.

ADDITIONAL COMMENT

I do not think there is any need to present the extended Kalman filter equations, especially when they are not used to justify the use of the ensemble Kalman filter. In this respect, there are some conceptual misunderstandings about the ensemble Kalman filter that must be corrected. First of all, there is no need to make any multi-Gaussian assumption to get to the ensemble Kalman filter equations, like there is no need to make any multi-Gaussian assumption to get to the cokriging equations; it is true that under the multi-Gaussian assumption the ensemble mean and ensemble covariance would be the mean and covariance of the conditional distribution given the obsevations; however, from the point of view of optimal estimate in the a least-square sense, the Kalman filter equations do not need any multi-Gaussian assumption.

We would like to thank the reviewer for his/her extended review, and we are happy that she/he overall has a positive image of it. In this reply we will start by addressing the major comment together with the first of the additional comments, which we believe are strongly related to each other.

The reviewer's points may be summarized as follows: 1) The reviewer is convinced that formal priors, as considered in the manuscript, are not essential, they are only used to initialize the filter and given enough informative data, the prior knowledge will fade away, 2) the reviewer points out that Gaussian assumptions are not needed to use the EnKF, there are enough informative data, 3) following 1 and 2, (s)he concludes that we obviously have too little data to perform our EnKF analysis and should thus focus on evaluating which additional data we would need to overcome the issues exemplified in the manuscript rather than focusing on the importance of prior knowledge.

Obviously, we and the reviewer have a fundamentally different view on parameter estimation in general, which roots in a decades-old debate on the validity of the Bayesian paradigm, known as 'Bayesian vs Frequentist debate'. Honestly, we don't think it would be beneficial for the manuscript to continue this debate. By this we don't want to imply that the frequentists' view is necessarily wrong, but we clearly take the Bayesian standpoint, and will not leave it. We find, among other reasons, the Bayesian framework to be refreshingly transparent when it comes to which assumptions are being made throughout the estimation process (in difference to making more subjective assumptions, such as selecting a specific objective function). As our consideration in this manuscript is the importance of relevant prior, the Bayesian framework is elegant, as it offers a way to compensate for non-unique data by the use of a strong prior.

The major point to be made clear is: groundwater-head data, which we consider in this study, are non-unique with respect to unknown conductivity and unknown recharge. This is derived and exemplified in Section 2 of the manuscript and this non-uniqueness is also the core of our manuscript. We do not disagree with the reviewer that different types of data would help improve the parameter estimation. On the contrary, we share those views. However, in subsurface hydrology the data scarcity is usually a great problem. Fluxes, as suggested by the reviewer, cannot be measured as such; conductivity measurements are, if existing and trusted, very local; tracer tests are time consuming; age tracers may be costly and require very long simulation times. Head measurements are common and trustworthy measurement, and also in the literature (as cited in the Introduction) the observation setup used by us is not uncommon. Hence, our example is rather realistic for a real world scenario in which we want to estimate aquifer parameters! If anything, we have unrealistically many observation points. The main difference between the Bayesian standpoint and the standpoint suggested by the reviewer is that without a formal inclusion of the prior, we require fully informative data to perform parameter estimation. This is, as rightly pointed out by the reviewer, not what we are doing.

The reviewer points out that "need of extra information to be able to single out the combination of hydraulic conductivity and recharge that is correct" is important. However, this is also a form of prior. If we have this extra information we are, in the view of our manuscript, in the Good (correct) prior case. As such, we do not see that there is a large difference, in practice, between the suggestions of the reviewer and the approach and aims stated in the manuscript. The bottom line of the manuscript is that we need to be careful to have correct prior information. How to do this is not the topic of the current manuscript, nor do we wish it to be.

Many practitioners would argue that we should simplify the model, that is, assume fixed zones of uniform hydraulic conductivity and recharge. While we agree that this can "fix" the problem of non-uniqueness, restricting the solution space by hard-wiring the structure of the subsurface is the most extreme prior to be thought of, as it does not allow any uncertainty in the identified structure itself. That is, the problem does not really vanish, it's just hidden.

Concerning the statement that the prior caries over, we again agree with the reviewer. If the data is fully informative, we can also estimate values outside of our initial sample. Similarly, the prior could in such cases vanish as more data come in. As pointed out above, the data is rarely fully informative in the subsurface-flow applications. However, the original manuscript was probably not clearly enough formulated that we throughout consider the case of non-unique data (hence, using head observations). In the revised manuscript this will be revised and made clearer.

To summarize, we see the requests made by the reviewer as a wish to alter our philosophy of parameter estimation. We believe, by contrast, that the Bayesian framework is a legitimate way of introducing and interpreting Ensemble-Kalman approaches. Within this framework, studying the importance of the prior is highly relevant as the prior is at the very heart of Bayesian analysis. Thus, we will not rewrite the manuscript to coincide with philosophical viewings of the reviewer, who obviously thinks otherwise. However, in the revised manuscript we will take great care to make clear that the data considered are non-unique, and will surely point out that other types of data could help resolving the issue of the incorrect prior.

The manuscript is updated on numerous points in sections 3-5 in order to make it clear that (1) the data considered is non-unique with respect to the joint estimation of conductivity and recharge, and this is the full point with the paper, and (2) other types of data that are informative to our estimation (and hence solves the non-uniqueness) would make the prior sampling a lot less important. The main changes to the manuscript in this respect can be found on page 9 (lines 10-18), pages 13-14 and page 20-21 (lines 19-13). We would also like to draw the reviewers' attention to the track-changes version of the manuscript attached to this reply letter, where also the smaller changes can be detected. We hope that the reviewer finds the changes to her/his liking.

I do not quite understand the last paragraph in page 5576 when the authors say that the EnKF is a linearized estimate that is alleviated by the repeated application over many time steps. The authors should understand that the EnKF captures the linear relationship that there is between the parameter and state variables through the experimental covariance –that's all–, the fact that you apply the updating equations over many time steps does not "alleviate the effects of non-linearity". The reason why the EnKF works and the extended Kalman filter did not is because the covariances are computed on parameters and states which have been obtained by solving the state equation through an ensemble of realizations, and therefore are much closer to the "true" covariance than the one obtained by propagating the initial covariance in time through a linearization of the state equation.

The section referred to was not particularly well formulated. What we really mean is that when we have many observations in time we make good use of the filter dampening. This slows the filter down (hence we may need more temporal observations to reach a stable result) but also avoids making too large jumps in the parameter space. This can be beneficial when the non-linear relations between parameters and observations cases erroneous updates, and as such it is helping to alleviate the effects of the linearizations. We see that this was not clear within the text and in the revised manuscript the full section will be thoroughly revised.

Section revised as describe above (page 13, lines 14-20).

Please revise your presentation and discussion of the EnKF.

While we are willing to explain steps more clearly, we will keep the EnKF within a Bayesian framework, with all the consequences.

OTHER COMMENTS

Since the state (piezometric heads) is not updated, you should explain how the state is computed after each assimilation step. Is the model rerun from time zero with the updated parameters?

We do not understand why the reviewer believes, that we do not update heads, which, in fact, we do.

As we do not understand the nature of this comment, no alterations has been possible to perform in the manuscript.

Page 5577, line 2, the original prior knowledge is smeared out after the first assimilation step by the Kalman gain, it carries over at the beginning but it will eventually disappear.

See answer to general comments.

Page 5577, line 8, if the prior knowledge is erroneous and there are sufficient observation data inconsistent with that prior model the estimates will converge to the "truth".

With enough informative data we agree with the reviewer. This was maybe not clear in the original submission and will be improved in the revision. However, for the data in question (groundwater heads), we cannot expect to recover any pattern that are not part of the prior. For a better reasoning on this question, please consider the reply to the major comment.

Text is updated to make clear that other data could lead to parameters outside the prior, but that this is not the case with the non-unique head data used in our study (pages 13-15, lines 21-9).

Page 5578, why is NRMSE only computed at the observation wells and not over all the aquifer, since you have the reference information?

The NRMSE could, as the reviewer suggests, also be calculated for all grid cells in the model. The approach to calculate the error given the observation locations available is chosen to represent the type of information available in a real case, for which one would of course not know the head value in each cell. In the end we also conclude that the NRMSE on its own is not a sufficient metric to judge which EnKF setups are good or bad, which we find an illustrative example of a problem that can occur in any real parameter estimation setup.

A clarification of why we chose this setup has been added to the revised manuscript (pages 15-16, lines 26-4).

Page 5578, it is unclear what is the denominator of the equation, is it the ensemble variance? or is it the prior measurement uncertainty (in which case it is a constant value)?

Here we mean the uncertainty of the measurement (which indeed is considered a constant). This missing information will be added to the revised manuscript.

Information added (page 15, line 18-20).

It would have been nice to see some variance maps along the ensemble mean maps.

This is a useful suggestion that we intend to pick up in the revision.

New figure with variance maps for the joint estimation is added (Figure 6 in revised manuscript).

Top of 5583, in real applications you need information to discern among alternative combinations of conductivity and recharge, this information could help you in choosing the prior model, or it could be other types of data (such as fluxes).

Yes, we do agree with the reviewer. This will be better highlighted to the revised manuscript.

Text is updated (pages 20-21)

End of 5583. No, multiple-point statistics will not help here, that is, it will not allow you to discriminate between a good and a wrong model as long as the observation data are consistent with those models. Besides, it has been proven that the EnKF will filter out the non multi-Gaussian characteristics of the initial ensemble of realizations.

To a certain extent this is right: if we have the wrong training images, we will be back to square one again (wrong initial sample). However, what we wanted to express was that the use of training images is a slightly more advanced way of introducing prior information than generating multi-Gaussian fields.

The text about MPS has been updated to: "The combination of these approaches could prove a possible way to better include a more correct prior information and, hence, to improve the performance of the joint estimation of conductivity and recharge fields by lowering the risk of conductivity-to-recharge aliasing due to wrong prior knowledge" (pages 21-22,lines 27-2).

SUMMARY

I liked the paper and I think it should be published, but only after the emphasis in the conclusions is shifted.

We are glad that the reviewer likes our work; the focus of the paper has been discussed above.

Reviewer #2

General comments

Ensemble Kalman filter is used here to jointly infer hydraulic conductivities and recharge by assimilating head data. The authors evaluated effect of the prior model on success of the method and concluded that a correct prior model is critical, which is consistent with previous research. My main concern includes two items in the "specific comments": item 4 and 8, in which the former is related to the mathematical model of EnKF and the latter the example to illustrate the compensation between hydraulic conductivity and recharge. Please revise the manuscript accordingly.

> We would like to thank the reviewer for the thorough review. All comments are addressed below.
>
> > *For ease of reading, references to changes made in the manuscript will be printed in brown italic, while blue texts are still identical to the previously posted replies.*

Specific comments

1. Page 5568 line 20-25: The authors cited work by Hendricks Franssen et al. (2004) who jointly estimated hydraulic conductivity and groundwater recharge. What is the difference between this work and the one in the submission? What is the improvement here? Can the authors comment it?

> A major difference is that Hendricks-Franssen et al (2004) considered only spatially (blockwise) uniform recharge while the current manuscript considers spatially random fields also for the recharge. A new sentence will be added to the manuscript in order to clarify this point "The authors [*here meaning Hendricks Franssen et al. (2004)*] considered the problem of a well-capture zone, in which they estimated hydraulic conductivity as continuously varying spatial field, whereas recharge was parameterized by zones with uniform values." Another difference is the choice of the estimation method (as already in the manuscript) and the actual purpose of the investigation.
>
> > *Manuscript altered as already described (page 4, lines 22-24).*

2. Page 5572 line 1-5: Recharge rate, as a boundary condition, is determined based on such as precipitation, infiltration and geographic conditions (rivers and pumping wells). However the recharge here seems more likely the specific flux at the interface of blocks (discretized for numerical simulation) according to the authors "recharge depends on the gradient of the original transmissivity field". Please clarify the meaning of "recharge" that is inferred from head data in this study.

> Recharge in the context of this work is a flux into each cell across the top boundary; potentially variable in both space and time. In the citation above, the reviewer misses that the cited text talks about **apparent** recharge, which is a mathematical construct. Seen from the flow model it is also a flux, in the same way as the real recharge, but its value is calculated as a function of the (erroneous) conductivity field. This is the essence of Section 2.
>
> > *That recharge is based on geo-features is discussed on page 21, lines 13-16. That recharge is a flux is evident from Eq. 1 and the explaining texts. These texts are not new to the revision, but as we believe that the missing apparent recharge in the reviewers*

text is just a small mistake, no additional changes has been performed to the manuscript following this comment. Please also see reply to Comment 7 below.

3. Page 5572 on Fig.1: What are the initial and boundary conditions (except the recharge at the center) of this example?

Boundary conditions (west 50m, east 8m, north and south no flow) will be added to the figure in the revised manuscript. As this is a steady-state problem, the question about initial condition is not applicable.

Added on page 8, lines 12-14.

4. Section 3.1 Kalman filter on page 5573 through 5574: Vector Xt consists of two elements, heads ht and parameters (recharge and log-conductivity) while the head ht can be simulated as ft(ht-1, Xt), that is ht = ft(ht-1, Xt). If this is correct there would be mistake in the objective function W(x) (equation 6) since the Xt contains head vector ht that is considered again separately in the second term (ft(ht-1, Xt) - Yt)T R-1(ft(ht-1, Xt) – Yt). Either the head vectors can be excluded from the vector Xt or the second term of the objective function should be removed. The corresponding comments and following equations should be revised accordingly, i.e., equations 7-13.

This is a bit of a misunderstanding. The term ft(ht-1,Xt) does not denote the simulated head, but the observation operator. Hence, ft(ht-1,Xt) represents the simulated observations and the full term is the difference between the observed heads and the simulated ones. The two terms of the objective function are the prior (first term) and the likelihood of observations (second term), hence a standard objective function formulation. In the revised manuscript we will consider reformulating/renaming the terms to avoid any misunderstanding.

The following sentence has been added: "It should be noted that f here, hence, denotes both the forward model and the observation operator" (page 10, lines 10-11).

5. Page 5575 equation 11-13: The series of equations are used to calculate the covariance between parameters and/or simulations. The denominator should be n-1 rather than n, that is, 1/(n-1) instead of 1/n in these equations.

The reviewer is right; this error will be corrected.

Error is corrected.

6. Page 5577 line 6-7: What do the authors mean by "combined patterns of hydraulic conductivity and recharge"?

We meant: [*the prior knowledge determines what*] "patterns of conductivity and patterns of recharge that can be jointly inferred by the scheme". The sentence will be reformulated in the revised manuscript.

The sentence in question is replaced by the one in the answer (page 14, lines 5-8)

7. Page 5577 line 20: Here it says "the conductivity and recharge fields are uncorrelated". On the contrary the authors stated earlier that "the apparent recharge depends on the gradient of the original transmissivity field" (page 5572 line 1-2), which indicates a close correlation. Please clarify this.

We disagree that the two statements contradict each other. The first (page 5572) concerns the mathematical calculation of the apparent recharge that can be used to replace a false conductivity field with. The second (page 5577) on the other hand concerns the setup of the virtual experiment and simply states that the model input fields are generated without any correlation between the two fields. The apparent recharge is dependent on the false conductivity field, but this has no relation to the virtual experiment input data. Please also see the answer to Comment 2 above.

The first sentence is updated such that it is clearer that this concerns the generation of the reference fields. We now write "It should be noted that here the reference conductivity and the reference recharge fields are generated as fields that are uncorrelated to each other" (page 14, lines 21-23).

8. Page 5581 line 16-20: "it is always possible to compensate a missing or wrong conductivity with a recharge, and this is also clearly seen in the last column of Fig.5: the estimated recharge shows remarkable similarity with the reference conductivity field." In Fig.5 the estimated recharge does show similarity with the reference conductivity field because the wrong recharge prior is sampled using the true conductivity field model NOT because they can compensate each other. They are two different things in my opinion. The authors need find another example to illustrate the compensation effect between conductivity and recharge derived in Section 2.

We see the reviewer's point. The text was not well formulated in this section. The compensation effect does not cause the estimated fields, which are enforced by the erroneously sampled prior parameter distributions (as rightly pointed out by the reviewer). However, if the data were uniquely informative and no compensation was possible, the effect of the erroneously sampled prior would have disappeared after 300 days of data assimilation and the estimated fields would have altered into structurally (more) correct fields! Because the head data are not unique, compensation between conductivity and recharge is possible, and wrong prior assumptions prevail in data-assimilation and parameter estimation practically forever. The combination of wrong conductivity and recharge fields reproduce the head observations quite well, and the filter algorithm sees no need to change the (erroneous) fields.

We see that the choice of wording was suboptimal and in the revised manuscript we will make it clear that what we see is that the compensation effect sustains the erroneous fields. We believe, however, that the example is illustrative and assess that adding another experiment would unnecessarily complicate the manuscript.

The section referred to by the reviewer has been reformulated to better highlight the sustaining effect of the compensation mechanism. In the revised manuscript the section in question is found on page 18, lines 17-27.

9. Page 5595 Fig.2: The last plot in Fig.2 shows the spatial distribution of recharge over the domain. The recharge is time-varying with a seasonal trend (page 5577 line 13 as well as shown in the "river stage" plot of Fig.2). So the question is which time does this recharge plot show? Also please add title for the X axis in the plot of "river stage" (it should be time I guess).

The actual recharge is calculated by multiplying the spatially uniform, temporal trend parameter (at the right day) with the shown (relative) recharge field. The spatial field shown is for the case that temporal trend parameter has the value of unity. The sentence "The actual recharge is calculated by multiplying the trend parameter at any given time with the shown recharge field" will be added to the revised manuscript to make this setup clearer.

The above mentioned sentence is added (page 14, lines 15-16). Missing axis title is also added.

Technical corrections

All technical corrections corrected as already described.

1. Page 5570 line 3: "...was worse then..." should be "...was worse than..."
   Corrected.

2. Page 5570 line 5-10, page 5578 line4, page 5581 line 2 and 16: The authors mentioned "Section 2", "Sect.2", "Section 3", "Sect.4" and "Sect.5". It would be better to keep consistent.
   Corrected.

3. Page 5576 line 9-10: "...parameter that it is primarily required..." remove "it"
   Corrected.

4. Page 5577 line 19: ". it should be..." use capital letter in "it"
   Corrected.

5. Page 5579 line 7-8: "We have also conducted successful assimilations also estimating the trend parameter." Too many "also"
   Corrected.

6. Page 5580 line 20: "smaller errors in predicting heads then the...". Correct "then"to "than"
   Corrected.

7. Page 5581 line 28: "...it would to be difficult to..." modify this sentence.
   Sentence is altered and split in two.

8. Page 5582 line 4-6: revise this sentence.
   Sentence revised and shortened.

Manuscript prepared for Hydrol. Earth Syst. Sci. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 18 August 2015

Joint inference of groundwater-recharge and hydraulic-conductivity fields from head data using the Ensemble-Kalman filter

D. Erdal and O. A. Cirpka

Center for Applied Geoscience, University of Tübingen, 72074 Tübingen, Germany

Correspondence to: D. Erdal (daniel.erdal@uni-tuebingen.de)

Abstract

Regional groundwater flow strongly depends on groundwater recharge and hydraulic conductivity. Both are spatially variable fields, and their estimation is an ongoing topic in groundwater research and practice. In this study, we use the Ensemble Kalman filter as an inversion method to jointly estimate spatially variable recharge and conductivity fields from head observations. The success of the approach strongly depends on the assumed prior knowledge. If the structural assumptions underlying the initial ensemble of the parameter fields are correct, both estimated fields resemble the true ones. However, erroneous prior knowledge may not be corrected by the head data. In the worst case, the estimated recharge field resembles the true conductivity field, resulting in a model that meets the observations but has very poor predictive power. The study exemplifies the importance of prior knowledge in the joint estimation of parameters from ambiguous measurements.

1 Introduction

Regional groundwater flow depends on spatially variable properties of the subsurface, notably the hydraulic conductivity field, and boundary conditions such as groundwater recharge. In practical groundwater-modeling applications, parameters of both aquifer properties and boundary conditions are estimated from measurements of hydraulic heads at a limited number of observation locations (e.g. Hill and Tiedeman, 2007). While many theoretical studies on parameter estimation in aquifers have concentrated on the assessment of the spatially variable hydraulic-conductivity field, also groundwater recharge is known to be highly variable in both time and space (e.g. de Vries and Simmers, 2002). Among the different techniques of estimating recharge reviewed by Scanlon et al. (2002), we consider here numerical approaches in which measured time series of hydraulic head are used to estimate groundwater recharge. The key question to be addressed in the present study is under which conditions it is possible to infer both the recharge field (a space-time function) and the spatial distribution of hydraulic conductivity from the same data set of hydraulic-head measurements.

In engineering practice, the model domain is typically subdivided into a small number of zones with given geometry, and uniform values of the material properties are assigned to each zone. Likewise, the land-surface is subdivided into zones with uniform recharge values, reflecting land use, soil types, and local climate variability. As an alternative, parameter values may be estimated at a limited number of points and interpolated in between (e.g. Doherty, 2003). By construction, these approaches can only determine spatial structures of the parameter fields meeting the prescribed shapes. A particular difficulty of this approach is that the variability within the given zones may be bigger than between the zones, while the internal variability is completely neglected in the parameter estimation.

The estimation of hydraulic conductivity as a continuous field has been intensively investigated in the past (see for example the reviews of Sanchez-Vila et al., 2006; Vrugt et al., 2008 and recently Zhou et al., 2014). In these approaches discretization of the domain leads to a formal number of parameters to be estimated that is identical to the number of cells or grid points. Typical 2-D applications result in $\mathcal{O}$ (10^4) parameters, whereas 3-D numerical domains may easily be made of $\mathcal{O}$ (10^6) cells. As the number of measurement points is by orders of magnitude smaller, this inverse problem is inherently ill-posed without additional constraints. Some authors therefore rely on flexible sets of shapes, such as polynomial trends or Voronoi polygons (e.g. Tsai et al., 2003a, b) rather than estimating $\mathcal{O}$ (10^4–10^6) parameter values. In standard geophysical inversion, Tikhonov regularization is the common approach to estimate distributed parameter fields from a limited set of measurements. Here, the parameters are assumed to be continuous spatial functions, but large gradients, curvatures, or deviations from prior values are penalized (applications to subsurface hydrology are given by Doherty and Johnston, 2003; Tonkin and Doherty, 2005; Doherty and Skahill, 2006, among others). In subsurface hydrology, however, the geostatistical framework is more common. Kitanidis (1997) and independently Maurer et al. (1998) showed that the two approaches are mathematically equivalent to each other.

In geostatistical inversion, the parameter field to be estimated is assumed to be an autocorrelated random space function. This prior knowledge is used in Bayesian inference, where the statistical distribution of the parameters is conditioned on the measurements of dependent quantities, such as hydraulic heads. A variety of schemes targets a single smooth spatial distribution approximating the conditional mean of the parameter field using Gauss-Newton- or conjugate-gradient-type of estimation schemes (e.g. Yeh and Yoon, 1981; Kitanidis and Lane, 1985; Zou et al., 1993; Li and Elsworth, 1995; Kitanidis, 1995; Yeh et al., 1996; Aschenbrenner and Ostin, 1995; McLaughlin and Townley, 1996; Spedicato and Huang, 1997; Loke and Dahlin, 2002). These methods can be extended to the generation of multiple conditional realizations by the method of smallest modification (e.g. RamaRao et al., 1995; Gómez-Hernández et al., 1997). However, the computational costs to obtain a single conditional realization is identical to that of the smooth best estimate. Also, the Gauss-Newton method requires the evaluation of the sensitivity of each measurement with respect to all parameter values, involving the solution of as many adjoint problems as there are measurements, which may become unbearable in case of many measurements, such as those obtained from transient processes. In the context of the present study it may be noteworthy that many geostatistical approaches have focused on the exclusive estimation of hydraulic conductivity, some include storativity (e.g. Gómez-Hernández et al., 1997; Kuhlman et al., 2008; Li et al., 2007), but most assume that the boundary conditions are deterministic. An exception is Hendricks Franssen et al. (2004) who used the geostatistical approach of sequential self calibration to jointly estimate the fields of hydraulic conductivity and groundwater recharge from head measurements. The authors considered the problem of a well-capture zone, in which they estimated hydraulic conductivity as continuously varying spatial field, whereas recharge was parameterized by zones with uniform values.

In groundwater hydrology, sequential data assimilation and Kalman filter methods have been used since long (e.g. Ferraresi et al., 1996; Eppstein and Dougherty, 1996; Hantush and Mariño, 1997). Particularly, and increasingly, popular is the Ensemble Kalman filter (EnKF) (Evensen, 1994) or versions thereof. Although the EnKF was primarily constructed to update model-state variables, in subsurface hydrology it is commonly used to estimate hydraulic conductivity. For this purpose Hendricks Franssen and Kinzelbach (2008), Drécourt et al. (2006), Tong et al. (2010, 2011), Xu et al. (2013a, b), Panzeri et al. (2015) all showed that the use of head observations in an EnKF framework can help improving the conductivity estimates, while Crestani et al. (2013) and Tong et al. (2013), among others, considered tracer tests for the same purpose. Most parameter estimations used 2-D models, as these are conceptually simpler, faster and easier to constrain and display. However, EnKF has also successfully been applied to infer 3-D hydraulic-conductivity fields (e.g. Schöniger et al., 2012).

An important step in setting up an EnKF to estimate parameters is the choice of initial ensemble. This choice is the most straight forward way of allowing prior information, such as ideas about correlation lengths, mean values or spatial pattern, to influence the filter process. From a technical point of view, the issue of initial sampling is how to represent the prior knowledge in an ensemble that is as small as possible, by, for example, adding ensemble subspace restriction and requirements on the sampling (e.g, Evensen, 2004; Oliver and Chen, 2008). From a practical point of view, especially in subsurface modeling, the issue is that our prior knowledge of the parameters, their mean values, deterministic trends, and spatial correlation structure is often limited. This may be seen as a more severe problem than choosing a sufficiently large ensemble size to actually capture the assumed variability by the ensemble. To overcome the limited knowledge about true parameters values, the use of synthetic test cases for methods testing and evaluation is very common in subsurface hydrology (e.g. Schlüter et al., 2012; Schelle et al., 2013). Here, the prior knowledge is only limited to what the modeler considers a reasonable assumption and it is not uncommon in the groundwater-EnKF context that the synthetic true parameter field is a single realization generated the same way as the initial ensemble (e.g. Huang et al., 2008; Tong et al., 2011, 2013; Vogt et al., 2012; Panzeri et al., 2014; Zhou et al., 2014). Hence, perfect knowledge about the statistics of the estimated parameters is implicitly assumed, which is a highly unrealistic assumption. The impact of the prior assumptions in groundwater modeling were considered, for example, by Li et al. (2012) who concluded that it was possible to estimate reasonable log-conductivity fields using the EnKF despite wrong priors, although the result was worse ~~then~~ than when using correct information.

In this work we study the impact of the prior knowledge when jointly estimating conductivity and recharge. We use an EnKF setup in which the initial ensemble is drawn using different assumptions of the spatial pattern of the parameters. ~~Section~~Sect. 2 discusses why the conductivity and the recharge are so difficult to estimate jointly if only pressure-head data is available. ~~Section~~Sect. 3 explains the Ensemble Kalman filter and the synthetic example used throughout this paper, while results and discussions are found in Sect. 4. We end with conclusions in Sect. 5.

2 Theory

In regional-scale groundwater-flow problems, we typically rely on the validity of the Dupuit assumption, stating that variations in hydraulic head and groundwater velocity are restricted to the horizontal directions. Under this condition, the depth-averaged, two-dimensional groundwater-flow equation for a phreatic aquifer reads as:

$$S_y \frac{\partial h}{\partial t} - \nabla \cdot \left(K\left(h - z_0\right) \nabla h\right) = R \tag{1}$$

subject to initial and lateral boundary conditions. $S_y(\boldsymbol{x})$ [–] is the specific-yield field, which is the drainage-effective porosity of the formation, $K(\boldsymbol{x})$ [L T^{-1}] denotes the depth-averaged hydraulic-conductivity field, $R(\boldsymbol{x}, t)$ [L T^{-1}] is the field of groundwater recharge, $z_0(\boldsymbol{x})$ [L] denotes the geodetic height of the aquifer bottom, $h(\boldsymbol{x}, t)$ [L] is the hydraulic-head field to be simulated, t [T] is time, and $\boldsymbol{x}$ [L] is the vector of horizontal spatial coordinates.

The term $K(h - z_0)$ may be interpreted as a transmissivity field $T(\boldsymbol{x}, t)$ [L^2 T^{-1}], varying in space and time. We now consider a confined surrogate aquifer with an assumed transmissivity field $T_{\text{ass}}(\boldsymbol{x})$ [L^2 T^{-1}] that differs from the true one (e.g. an incorrectly estimated transmissivity field). The logarithm of the scaling factor between the two transmissivities is denoted $f(\boldsymbol{x}, t)$ [–]:

$$f = \ln\left(\frac{K \times (h - z_0)}{T_{\text{ass}}}\right). \tag{2}$$

Substituting Eq. (2) into Eq. (1) yields:

$$S_y \frac{\partial h}{\partial t} - \nabla \cdot (T_{\text{ass}} \exp(f) \nabla h) = R. \tag{3}$$

Applying the chain-rule of differentiation to the divergence in Eq. (3), the product rule of differentiation to $\nabla \exp(f)$, and dividing by $\exp(f)$ results in:

$$\underbrace{\exp(-f) S_y}_{:=S_{\text{app}}} \frac{\partial h}{\partial t} - \nabla \cdot (T_{\text{ass}} \nabla h) = \underbrace{\exp(-f) R + \nabla f \cdot \nabla h T_{\text{ass}}}_{:=R_{\text{app}}} \tag{4}$$

$$\Rightarrow S_{\text{app}} \frac{\partial h}{\partial t} - \nabla \cdot (T_{\text{ass}} \nabla h) = R_{\text{app}} \tag{5}$$

subject to the same initial and lateral boundary conditions as above. In Eq. (5), $S_{\text{app}}(\boldsymbol{x}, t)$ [–] and $R_{\text{app}}(\boldsymbol{x}, t)$ [L T^{-1}] are apparent specific-yield and groundwater-recharge fields. Equation (5) results in exactly the same hydraulic-head distribution as the original groundwater-flow Eq. (1), even though the transmissivity field is different. Note that $\exp(-f)$ is positive, so that the apparent specific yield S_{app} remains positive, whereas no sign restrictions apply to $\nabla f \cdot \nabla h$, resulting in both positive and negative R_{app} values. In case of a phreatic aquifer, the true transmissivity varies with hydraulic head, so that the apparent parameters change with time. If the water-filled thickness of the true aquifer does not change with time, which is the case for confined aquifers, the apparent fields are time-invariant.

The derivation given above exemplifies that the same hydraulic-head field can be obtained with different hydraulic-conductivity fields by modifying recharge and, in the case of transient flow, the specific yield. Noteworthy is that the apparent recharge depends on the gradient of the original transmissivity field. Hence, a large – positive or negative – apparent recharge is expected at locations where the transmissivity changes drastically. Though we have shown that modifications of recharge and specific yield can always replace the conductivity, the opposite case is not guaranteed, because the conductivity has clear physical limitations, notably it cannot be negative.

The fact that conductivity variation can be exchanged by recharge and specific-yield variations renders the joint estimation of hydraulic conductivity, recharge (and specific yield) an inherently ill-posed problem even when the hydraulic-head field is known at every point in the domain (and every time point).

We may illustrate the problem by the example of an unconfined aquifer at steady state, shown in Fig. 1. The original simulation (left column in Fig. 1) exhibits a square-shaped inclusion of low permeability in an otherwise uniform high permeability field (first row; two orders of magnitude difference in K) ~~,~~ and a constant low recharge rate (second row)~~and~~ . As boundaries, we employ a significant head drop from west ~~to east~~(50 m) to east (8 m) and no flow boundaries on the north and south sides. The resulting head field is shown in the third row of Fig. 1, and the corresponding field of Darcy velocities in the fourth row of Fig. 1.

If the inclusion is removed, and the recharge remains the same, the system shows a perfectly homogeneous behavior (middle column of Fig. 1). The third column in Fig. 1, on the other hand, shows exactly the same hydraulic-head field as the original simulation, but the permeability field is uniform, whereas the recharge field shows strong fluctuation. From Fig. 1 we can note that, in accordance with Eq. (4), the strong positive and negative recharge rates are introduced at the interface of the removed inclusion. Also, while the head fields of the original and surrogate models are identical, the velocity fields are quite different, because the conductivities are different. The latter implies that transport would be strongly different between the two cases. It becomes also clear that, without additional constraints, a unique joint estimation of both recharge and conductivity fields is strictly impossible.

In classical model calibration, the ambiguity between transmissivity and groundwater recharge may cause problems of ill-posedness, but assuming presumably known zones with block-wise uniform parameter values restricts the solution of the inverse problem. As example, the strong positive and negative recharge values of the surrogate model in Fig. 1 would most likely not be obtained in standard model calibration because the recharge zones would hardly be chosen as embedded rectangular frames. In shape-free inversion, using either Tikhonov regularization or geostatistical methods, by contrast, the solution space is much less restricted and chances that unresolved transmissivity variations are traded for recharge fluctuations are in principle fairly high. The question thus arises under which conditions the estimated fields are reasonable despite the ambiguity of aquifer properties and boundary conditions.

3 Methods

3.1 Kalman filter

In the following we briefly repeat the basic assumptions of deriving the Ensemble Kalman Filter (EnKF) within a Bayesian framework. While it is possible to have a much more pragmatic view on EnKF as an extended least-square estimator, we believe that the transparency of the Bayesian framework with respect to the underlying assumptions is beneficial. In particular, the Bayesian framework explains the choice of initial ensemble as prior knowledge and the conceptual importance of the prior knowledge in the estimation procedure, while a frequentist's standpoint of view is in contrast to making use of prior knowledge altogether. For further transparency, we first explain the extended Kalman filter (see for example similar derivations by Evensen (2009)).

We denote the vector of all parameters (recharge values and log-hydraulic conductivities of all cells) Φ. Prior to considering measurements, they are assumed to be random functions following a multi-Gaussian distribution, which is fully characterized by the prior mean $\boldsymbol{\mu}'_{\Phi}$ and covariance matrix $\mathbf{P}'_{\Phi\Phi}$. If we assume that the covariance function $P'_{\Phi\Phi}(h)$ is stationary with the distance vector $\boldsymbol{h}$ and known structural parameters (variance, correlation lengths, rotation angles), the element $(i,\, j)$ of the covariance matrix $\mathbf{P}'_{\Phi\Phi}$ is $P'_{\Phi\Phi}(x_2 - x_1)$. The full matrix is constructed by all grid points.

The vector of simulated hydraulic heads $\boldsymbol{h}_t$ at time level t depends on the heads $\boldsymbol{h}_{t-1}$ at the previous time level and on the parameters Φ. Because the old heads $\boldsymbol{h}_{t-1}$ depend on Φ, they are random variables, too. In the combination of data assimilation and parameter estimation applied here, the vector of all simulates states (the heads $\boldsymbol{h}_t$ in all cells) and the vector of all parameters Φ are concatenated to a single vector $\boldsymbol{x}_t$ of states and parameters, assumed to be random multi-Gaussian functions with unconditional mean $\boldsymbol{\mu}'_x$ and covariance matrix $\mathbf{P}'_{xx}$, in which the prior statistics of $\boldsymbol{h}_t$ are obtained by linearized uncertainty propagation of the statistics of $\boldsymbol{h}_{t-1}$ and Φ.

For convenience, we denote running the model and simulating the observations (which is here just picking the heads at the observation locations) as $\boldsymbol{f}_t(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$. It should be noted that $\boldsymbol{f}$ here, hence, denotes both the forward model and the observation operator. This model outcome is contrasted to the measurements of heads at time level t, here denoted $\boldsymbol{y}_t$. The true (unknown) heads at the measurement locations are considered to be a vector of random variables with a multi-Gaussian distribution, characterized by the measurement vector $\boldsymbol{y}_t$ as mean and the covariance matrix $\mathbf{R}$, reflecting measurement error.

Since we assume multi-Gaussian distributions, finding the best conditional estimate $\boldsymbol{\mu}''_x$, of the entire head field at the new time level and the parameters by application of Bayes' theorem results in minimizing the following objective function $W(\boldsymbol{x}_t)$:

$$W(\boldsymbol{x}_t) = \left(\boldsymbol{x}_t - \boldsymbol{\mu}'_{x_t}\right)^T \mathbf{P}'^{-1}_{x_t x_t} \left(\boldsymbol{x}_t - \boldsymbol{\mu}'_{x_t}\right) + \left(\boldsymbol{f}_t(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t) - \boldsymbol{y}_t\right)^T \mathbf{R}^{-1} \left(\boldsymbol{f}_t(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t) - \boldsymbol{y}_t\right) \quad (6)$$

which is done by setting the derivative of $W(\boldsymbol{x})$ to zero. In the linearized version, $\boldsymbol{f}_t(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$ is linearized about the prior mean $\boldsymbol{\mu}'_{x_t}$, and the linearized conditional covariance matrix $\mathbf{P}''_{x_t x_t}$ of $\boldsymbol{x}_t$ is obtained by inverting the Hessian of $W(\boldsymbol{x}_t)$, using the same linearization. Kalman filtering is based on these approximations. Here, the data are successively accounted for, considering one time level after the other. Then, the posterior mean $\boldsymbol{\mu}''_{x_t}$ and covariance matrix $\mathbf{P}''_{x_t x_t}$ of time level t are propagated to the next time level $t+1$ to obtain the corresponding prior mean and covariance matrix.

By applying rules of matrix identities it can be shown that linearization about the prior mean $\boldsymbol{\mu}'_{x_t}$ leads to the following expression for the conditional mean and covariance matrix:

$$\boldsymbol{\mu}''_{x_t} = \boldsymbol{\mu}'_{x_t} + \mathbf{P}'_{x_t y_t} \left(\mathbf{P}'_{y_t y_t} + \mathbf{R}\right)^{-1} \left(\boldsymbol{y}_t - \boldsymbol{f}_t\left(\boldsymbol{\mu}_{h_{t-1}}, \boldsymbol{\mu}'_{x_t}\right)\right) \tag{7}$$

$$\mathbf{P}''_{x_t x_t} = \mathbf{P}'_{x_t x_t} - \mathbf{P}'_{x_t y_t} \left(\mathbf{P}'_{y_t y_t} + \mathbf{R}\right)^{-1} \mathbf{P}'_{y_t x_t} \tag{8}$$

in which $\mathbf{P}'_{y_t x_t} = \mathbf{J}\,\mathbf{P}'_{x_t x_t}$ is the cross-covariance matrix between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$, $\mathbf{P}'_{x_t y_t} = \mathbf{P}'T_{y_t x_t}$, and $\mathbf{P}'_{y_t y_t} = \mathbf{J}\,\mathbf{P}'_{x_t x_t}\,\mathbf{J}^T$ is the propagated covariance matrix of $\boldsymbol{y}_t$, expressing the uncertainty of $\boldsymbol{y}_t$ caused by the uncertainty of $\boldsymbol{x}_t$. $\mathbf{J}$ denotes the sensitivity matrix of $\boldsymbol{f}_t$ with respect to $\boldsymbol{x}_t$, derived about the prior mean.

The scheme described so far is known as extended Kalman filter. It relies on linearization about the prior mean and has the disadvantages that the full sensitivity matrix $\mathbf{J}$ must be evaluated, which can be computationally very costly. Also, already slight nonlinearities in $\boldsymbol{f}_t(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$ imply that the propagated covariance matrices are not correct.

A popular alternative to the original Kalman filter is the Ensemble Kalman filter (EnKF) (Evensen, 1994), in which the linearization is performed about an entire ensemble of state and parameter values, and no sensitivity matrices are computed. The prior statistics are given by:

$$\boldsymbol{\mu}'_{x_t} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_t'^{(i)} \tag{9}$$

$$\boldsymbol{\mu}'_{y_t} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{f}_t\left(\boldsymbol{h}_{t-1}''^{(i)}, \boldsymbol{x}_t'^{(i)}\right) \tag{10}$$

$$\mathbf{P}'_{x_t x_t} = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_t'^{(i)} - \boldsymbol{\mu}'_{x_t}\right)\otimes\left(\boldsymbol{x}_t'^{(i)} - \boldsymbol{\mu}'_{x_t}\right) \tag{11}$$

$$\mathbf{P}'_{x_t y_t} = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_t'^{(i)} - \boldsymbol{\mu}'_{x_t}\right)\otimes\left(\boldsymbol{f}_t\left(\boldsymbol{h}_{t-1}''^{(i)}, \boldsymbol{x}_t''^{(i)}\right) - \boldsymbol{\mu}'_{y_t}\right) \tag{12}$$

$$\mathbf{P}'_{y_t y_t} = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{f}_t\left(\boldsymbol{h}_{t-1}''^{(i)}, \boldsymbol{x}_t'^{(i)}\right) - \boldsymbol{\mu}'_{y_t}\right)\otimes\left(\boldsymbol{f}_t\left(\boldsymbol{h}_{t-1}''^{(i)}, \boldsymbol{x}_t'^{(i)}\right) - \boldsymbol{\mu}'_{y_t}\right) \tag{13}$$

in which n is the number of ensemble members, the superscript (i) denotes the ith member, and $\boldsymbol{a}\otimes\boldsymbol{b}$ is the tensor product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. As before, the prior values are denoted by a single prime, and the posterior by a double prime. Upon initialization, the original ensemble members $\boldsymbol{x}_0^{(i)}$ are drawn from the unconditioned multi-Gaussian distribution of $\boldsymbol{x}$, whereas the updating of the individual ensemble members follows the procedure outlined above:

$$\boldsymbol{x}_t''^{(i)} = \boldsymbol{x}_t'^{(i)} + b\mathbf{P}'_{x_t y_t}\left(\mathbf{P}'_{y_t y_t} + \mathbf{R}\right)^{-1}\left(\boldsymbol{y}_t + \boldsymbol{\varepsilon}^{(i)} - \boldsymbol{f}_t\left(\boldsymbol{h}_{t-1}''^{(i)}, \boldsymbol{x}_t'^{(i)}\right)\right) \tag{14}$$

in which $\boldsymbol{\varepsilon}^{(i)}$ is a vector of random observation noise drawn from a multi-Gaussian distribution with zero mean and covariance matrix $\mathbf{R}$. The factor b is the so called damping parameter (e.g. Hendricks Franssen and Kinzelbach, 2008) which serves to slow down the update of states and parameters. It is an ad-hoc tuning parameter that ~~it~~ is primarily required for small ensemble sizes; few guidelines exist on how to select it. In this work, the damping is set to 0.6 for the updates of the head values and 0.05 for the parameter update, though since the ensemble size is large and there are many temporal observations (see below), the choice is not crucial in any sense. For a more in-depth description of the filter algorithm, the interested reader can consult Evensen (2003) or Burgers et al. (1998) for general filter details or Erdal et al. (2014) and Erdal (2014) for in-depth details on the actual implementation used in this study.

It should be noted that the ensemble Kalman filter still relies on the same assumptions as the original Kalman filter. Notably, the combined vector of states, parameters, and observations is assumed to be a multi-Gaussian random variable, which means that $\boldsymbol{x}_t$ is multi-Gaussian, the model $\boldsymbol{f}_t$ depends linearly on $\boldsymbol{x}_t$, and the measurement error is multi-Gaussian, too. These conditions are not strictly met, so that the EnKF solution is only a linearized estimate. However, ~~the repeated application over many time steps as well as the~~ in contrast to the extended Kalman Filter, in EnKF the linearization is performed by considering an entire ensemble rather than by taking derivatives at a single point (e.g. Nowak, 2009) . The large ensemble sizes used in this work as well as the repeated application over many time steps alleviates the effects of nonlinearity to some extent, by allowing a generous use of the dampening factor. Hence the filter is slowed down and the possible erroneous updates resulting from the linearization have a less strong effect on the update. Further, the model considered is only weakly nonlinear, so that in total the effects of the linearlizations are likely small compared to other sources of errors (e.g. prior uncertainties, as discussed later).

~~A second~~ An important constraint is that the scheme, like any other Bayesian method, depends on the choice of the unconditional mean and covariance structure of the parameters Φ. ~~While~~ It is important to keep in mind that our application (estimating spatial patterns of both hydraulic conductivity and recharge from hydraulic-head data) is based on, at least partially, ambiguous data, as outlined in Sect. 2. Bayesian parameter-estimation schemes are well posed even in the presence of non-informative or ambiguous data due to the prior information. Thus, while the updating procedure leads to modifications of the parameters, the original prior knowledge carries over. Spatial patterns that are in contradiction to the prior knowledge cannot be recovered by the scheme. This would of course be different if the observations were in strong contradiction to the prior. If so, we could see a departure from the prior, both in terms of absolute values as well as in terms of structure. This point will be discussed more in Sect. 5. In our application, however, Φ contains parameters describing both aquifer properties and boundary conditions and, as we have shown above, the effects of these two types of parameters on the measured heads can be similar. Hence, the ~~prior knowledge determines which combined patterns of hydraulic conductivity and recharge are~~ data can be non-unique with respect to the parameters and the prior knowledge may determine which patterns of conductivity and which patterns of recharge that can be jointly inferred by the scheme. If the prior knowledge is erroneous, the estimated fields may also be erroneous.

3.2 Setup of a synthetic experiment

For testing the possibilities and limitations in jointly estimating conductivity and recharge, we have set up a synthetic 2-D example of transient flow in an unconfined aquifer. The model setup is shown in Fig. 2 and consists of spatially variable recharge with a temporal seasonal trend, spatially variable conductivity, a temporally variable southern boundary corresponding to a river, as well as 5 pumping wells. The actual recharge is calculated by multiplying the trend parameter with the shown recharge field. More technical details about the setup is found in Table 1. Observations of groundwater heads are taken daily at 45 observation wells spread throughout the domain during a 1 year simulation and assuming an observation error of 1 cm. The recharge and log-conductivity fields are both sampled as random fields with anisotropic, exponential covariance functions and strong rotation of the principal directions of anisotropy (Table 2). ~~it~~ It should be noted that here the ~~conductivity and~~ reference conductivity and the reference recharge fields are ~~uncorrelated~~generated as fields that are uncorrelated to each other. This could, for example, represent a scenario in which the recharge is primarily controlled by variable land use and vegetations while the conductivity is a constant material property.

For the estimation of the recharge and conductivity fields, we apply the Ensemble Kalman filter using an ensemble of 2000 members. As this work aims at exploring which prior knowledge is required for the estimation process, three different cases of prior knowledge are considered. In the first, the initial ensemble members are drawn from the same (hence correct) distribution as the reference (true) field. The second case is identical to the first apart from the rotation angle of the anisotropy being randomly chosen for each ensemble member. In the third case, the rotation angle is fixed but wrong. Here, the recharge is sampled using the rotation angle and correlation lengths of the true conductivity field and vise versa, creating a rather problematic initial ensemble. A plot of the three correlation structures can be found in the bottom of Fig. 3 in Sect. 4 where the three initial ensembles are called the "good", "random" and "wrong" ones. Please note that the correlation plot for the random initial is only meant as an illustration of the fact that each ensemble has a unique rotation angle and does not show the actual angles considered.

The goodness of the resulting fields are judged in two ways. First, the ensemble mean of the fields are visually compared to the reference fields and subjectively judged to be similar or not. Second, the normalized root mean square error of the simulated heads in the 45 observation wells is computed by:

$$\mathrm{NRMSE} = \sqrt{\frac{1}{n_\mathrm{t} n_\mathrm{obs}} \sum_{t=t_1}^{t_2} \sum_{i=1}^{n_\mathrm{obs}} \frac{\left(h_\mathrm{true}(i,t) - \overline{h(i,t)}\right)^2}{\sigma_\mathrm{h}^2}} \tag{15}$$

where n_t is the number of temporal observations between t_1 and t_2, n_obs the number of observation locations (here 45), $\overline{h(i,t)}$ is the ensemble mean head observation at position i and time t, h_true is the corresponding true value, and σ_h is the measurement uncertainty of hydraulic-head observations, hence here a fixed value decided prior to the EnKF simulations. This gives a quantitative metric of judging the actual performance of the estimated model. We assimilate head observations from day 50 to day 300, while the remaining 65 days of the one-year data is used to test the model's predictive capabilities. This results in an assimilation error for judging how well the assimilation went and a prediction error for judging the models predictive powers. It should be noted that to properly asses the predictive power of the model in a scenario different to the one used for the assimilation, one of the four wells shown in Fig. 2 only starts pumping at day 301. As this is a virtual experiment, one could also consider calculating the NRMSE based on the head error in all grid cells, but in a real-world application such data would not exist. In this work, we prefer considering only the head-values at observation points, as it also gives an idea of the usefulness of using available observations as a means to estimate the goodness of the result.

We have combined the three different prior distributions with three different estimation problems, namely the estimation of (a) recharge alone, (b) hydraulic conductivity alone, and (c) recharge and hydraulic conductivity together, leading to a total of nine different scenarios. In the stand alone scenarios, all other parameters and settings are assumed known and, hence, set to their true values. As can be seen from Fig. 2, the recharge not only shows a strong spatial pattern but also a temporal trend. In the estimations shown below, this temporal trend is assumed known. We have also conducted successful assimilations ~~also~~ estimating the trend parameter. However, as the absolute recharge values of these tests may vary with the absolute value of the scaling parameter, the results are less intuitive to display and therefore only the assimilations with known trend function are shown.

4 Results and discussion

4.1 Stand-alone estimation of recharge or conductivity

The simplest of the estimation problems presented in this study is the stand-alone estimation of recharge, since the hydraulic heads depend linearly on recharge. This is reflected in the estimated recharge fields shown in Fig. 3. As expected, the best results are achieved with the best initial estimate (second column). However, also the estimates using the covariance functions with the random and wrong orientations of anisotropy show in large the right pattern. Table 3 quantitatively confirms these qualitative findings by low values of the normalized root mean square error of predicted heads. From the last column in Fig. 3

we see that, although the filter manages to produce a reasonable ensemble mean of the recharge field, the similarity with the covariance function used to create the initial ensemble is still very prominent. This is especially so if one starts considering individual ensemble members (not shown), and it demonstrates how sensitive the EnKF method is to the initial guess, even in this linear problem.

It is important to keep in mind that the ensemble size is large so that the plots of the ensemble means shown in Fig. 3 are smoothed. It is not expected that the smooth ensemble estimate exhibits the same extreme values as those seen in the true parameter distribution, whereas individual ensemble members should show the same variability as the (unknown) reference field.

In comparison to estimating the recharge fields, the estimation of conductivity fields alone is more complicated. Here, the nonlinearities of Eq. (1) affects the estimation. More importantly, the orientation of the anisotropy of heterogeneity plays a vital role in the behavior of groundwater flow. This is also seen in the final estimates of the conductivity fields, shown in Fig. 4, where the only reasonable result is achieved if the right pattern is assumed in the prior knowledge (second column) or if the prior pattern is random (third column). The reasonable performance of the prior distribution with diffuse knowledge about the anisotropy orientation may be explained by the large initial ensemble containing some members with reasonable patterns and decent behavior. In the case that the orientation of anisotropy is assumed erroneously in the prior knowledge (fourth column), the filter completely fails to produce any result similar to the truth. This finding does not depend on the ensemble size. The prediction errors listed in Table 3 clearly confirm the visual impression.

The prediction errors listed in Table 3 emphasize that estimating recharge leads to smaller errors in predicting heads ~~then~~ than the estimation of the hydraulic-conductivity field. This could indicate that improvements of the estimated conductivities are more important for lowering the prediction error, which would follow the findings of Hendricks Franssen et al. (2004). As pointed out above, the higher errors when estimating conductivities are likely related to the head value in a cell depending not only on the conductivity of that cell but to the macroscopic anisotropy of hydraulic conductivity in the entire aquifer.

4.2 Joint estimation of recharge and conductivity

As derived in Sect. 2, joint estimation of recharge and conductivity fields is impossible without prior knowledge about either of the two quantities. In Bayesian inversion methods, however, prior knowledge is assumed anyway. In the EnKF method, the prior information is conveyed by the initial ensemble drawn from the prior distribution. By this, the jointly estimated recharge and conductivity fields are unique and reproducible in a statistical sense. The remaining question is whether these estimates also resemble the true fields and whether they are good for prediction purposes.

Figure 5 shows the results of the joint estimation using the three different initial ensembles and Figure 6 shows the corresponding spatial distributions of the estimation variance. If the initial ensemble is good, that is the reference fields are drawn from the same statistical distribution as the initial ensemble, it is possible to estimate both conductivity and recharge with reasonable precision, given the number and accuracy of observations (second column). When the initial ensemble is poor, however, the result is rather poor for the recharge and more blurry for the conductivity (third column), or we infer fields that look good but are wrong (last column).

As shown theoretically in Sect. 2, it is always possible to compensate a missing or wrong conductivity with a recharge~~, and this~~. An effect of this compensation is also clearly seen in the last column of Fig. 5: even after 250 days of data assimilation, the estimated recharge shows remarkable similarity with the reference conductivity field. The long assimilation time is important, since, if there would have been no compensation, the estimated fields would not retain their erroneous structures for so many filter updates. This shows that the issue of trading one quantity for the other is not only a theoretical issue, but also relevant in practice. It should be noted here that the cause of the original poor estimations is not the compensation mechanism described in this paper, but the false prior sampling. However, the compensation mechanism sustains the poor estimates when the observations are, as in this work, non-unique.

The lacking ability of the random and wrong initial ensemble estimates with respect to predicting heads under conditions not encountered in the calibration period are documented in Table 3, where the prediction errors caused by the poorly estimated fields are often an order of magnitude larger then those resulting from a good estimation. It is interesting to note that the error obtained throughout the assimilation, shown in Table 4, is not a good indicator for the predictive capabilities of the various models, as quantified by the prediction errors listed in Table 3. ~~There~~ Although there are differences in the assimilation error, both within and between the different estimation setups, ~~but it would to~~ it would be difficult to ~~foresee that~~ predict any model performance from these errors. That the joint estimation is performing much better with the good prior compared to the poorer ones is only obvious if the full table is available. The same behavior is illustrated with an example of two observations wells in Fig. 7, from which it is clearly shown that all approaches has a good fit during assimilation but that the wrong prior deviates during the predictions. From a practical standpoint of view this highlights that it is important have relevant validation data to test the predictive power of a model when ~~performing data assimilationwith parameter update by EnKF (or any other approach)~~the parameters are inferred using sequential data assimilation.

Like in the scenarios in which only recharge or only conductivity were estimated, the mean joint estimate lack the extreme values of the reference fields. As discussed above, such behavior is expected for the smooth best estimate even in cases where the scheme works perfectly fine. Individual ensemble members show significantly stronger variability, as can be seen also from the maps of the estimation variance in Fig. 6. We consider the results from the good initial ensemble as good, since they capture the main patterns of the parameter fields well and have, overall seen, reasonable absolute parameter values. For purposes of transport predictions, we would recommend using the entire ensemble rather than the ensemble mean. In case of the estimates using the wrong prior knowledge, in particular where the orientation of anisotropy is chosen randomly, the fluctuations cannot be aligned well in the right direction, and averaging over features oriented in all directions lead to particularly smooth estimates of the mean.

5 Conclusions

In the present study we have shown that it is possible to jointly estimate reasonable fields of hydraulic conductivity (or its logarithm) and recharge as spatially fluctuating fields from pure head observations provided that the statistics of the true fields are fairly well understood. Starting with wrong assumptions about conductivity and recharge patterns can lead to aliasing, in which not detected features of hydraulic conductivity are traded for erroneous fluctuations in recharge.

In real-case applications, the prerequisite of a good prior can pose a severe problem because the true spatial patterns may be widely unknown. From a more technical standpoint of view it may be noteworthy that a rather common way of setting up a synthetic groundwater-EnKF test is to generate a large ensemble of realizations and use one of them as the truth and the rest as the initial ensemble. By this it is guaranteed that the statistics of the initial ensemble is perfect and, as shown here, a good result can be expected. Unfortunately, in real-world applications the geostatistics of (log)-hydraulic conductivity are typically quite uncertain so that the good performance of a scheme, involving both the measurement strategy and the inverse method, in an overly optimistic test case regarding prior knowledge may not be transferable. We thus highly recommend to design realistic test cases that include potential bias in prior knowledge.

In the present work, we only used head data for data assimilation and parameter estimation~~, while in reality probably at least a vague idea of conductivity values couldbe available from the bore holes required for the observations, and the patterns of recharge should~~ . With respect to unknown conductivity and unknown recharge, the drawback with head observations, as shown in Sect. 2, is that the observations are non-unique. Including more head observations will, due to the non-uniqueness, not alleviate this drawback. Even a perfectly observed groundwater system could be non-unique, since exactly same head field can be achieved with different combinations of conductivity and recharge. Therefore, the joint estimation is difficult and the goodness of the prior information becomes important. Other types of observations could, of course, also be considered. Ideally we would have (plenty of) observations of subsurface fluxes or of conductivity. In this case, the total data set would become highly informative and the prior would be significantly less important. However, this is not realistic for applications in subsurface hydrology. Fluxes cannot be measured as such and conductivity measurements are, if existing and trusted, very local. Further observations could be tracer tests, which are time consuming or age tracers that may be costly and require very long simulation times. Head observations are, in this respect, common and trustworthy measurement. Hence, our example can be considered rather realistic for a real world scenario of estimating aquifer parameters.

In real-world applications, vague guesses of the hydraulic conductivity distribution may exist from drilling logs, slug tests, and pumping tests (e.g. Dietrich et al., 2008; Lessoff et al., 2010) . All of these tests are independent of recharge so that making use of this information may alleviate the problem of non-uniqueness outlined in this paper to some extent. Vague guesses of K can find their way into parameter estimation either by means of an improved prior of K or by explicitly accounting for the additional measurement types in the EnKF procedure, including the full observation operator. For recharge, the patterns should in principle reflect land use and soil types, which are accessible information. ~~Spatially~~ Further, spatially variable recharge may also be constrained by the use of remote sensing information (Brunner et al., 2006; Hendricks Franssen et al., 2008). These type of data could either be used as direct observations in the assimilation ~~or~~ (if we trust them) or considered as prior information and used to condition the initial ensemble (Sun et al., 2009; Panzeri et al., 2013). The latter could also be seen as a way of discarding initial samples that contain unfeasible conductivity-recharge combinations. This would create a much more appropriate initial ensemble. Hence, as shown in this work, the filter would have an increased change of successfully estimating the parameters when the prior is good. The idea of improving the initial ensemble can also be related to the popular method of multiple-point statistics~~,~~ ~~where~~ . Here, the use of training images which should represent relevant spatial correlation patterns have been used to condition conductivity fields (see Okabe and Blunt, 2004; Hu and Chugunova, 2008). The combination of assimilating head data and the use of training images to condition the ensembles has also been tested with promising results (Li et al., 2013). The combination of these approaches could prove a possible way to ~~perform~~ achieve a more correct prior sample and, hence, to improve the performance of the joint estimation of conductivity and recharge fields ~~with a lowered~~ by lowering the risk of conductivity-to-recharge aliasing due to wrong prior knowledge.

Acknowledgements. Financial support from the Deutsche Forschungsgemeinschaft (DFG) under CI 26/13-1 in the framework of research unit FOR 2131 "Data Assimilation for Improved Characterization of Fluxes across Compartmental Interfaces" is gratefully acknowledged.

References

Aschenbrenner, F. and Ostin, A.: Automatic parameter estimation applied on a groundwater model: The problem of structure identification, Environ. Geol., 25, 205–210, doi:10.1007/BF00768550, 1995.

Brunner, P., Hendricks Franssen, H.-J., Kgotlhang, L., Bauer-Gottwein, P., and Kinzelbach, W.: How can remote sensing contribute in groundwater modeling?, Hydrogeol. J., 15, 5–18, doi:10.1007/s10040-006-0127-z, 2006.

Burgers, G., van Leeuwen, P. V., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.

Crestani, E., Camporese, M., Baú, D., and Salandin, P.: Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation, Hydrol. Earth Syst. Sci., 17, 1517–1531, doi:10.5194/hess-17-1517-2013, 2013.

de Vries, J. J. and Simmers, I.: Groundwater recharge: an overview of processes and challenges, Hydrogeol. J., 10, 5–17, doi:10.1007/s10040-001-0171-7, 2002.

Dietrich, P., Butler, J. J., and Faiss, K.: A rapid method for hydraulic profiling in unconsolidated formations, Ground Water, 46, 323–328, doi:10.1111/j.1745-6584.2007.00377.x, 2008.

Doherty, J.: Ground water model calibration using pilot points and regularization, Ground Water, 41, 170–177, 2003.

Doherty, J. and Johnston, J. M.: Methodologies for calibration and predictive analysis of a watershed model, J. Am. Water Resour. Assoc., 39, 251–265, doi:10.1111/j.1752-1688.2003.tb04381.x, 2003.

Doherty, J. and Skahill, B. E.: An advanced regularization methodology for use in watershed model calibration, J. Hydrol., 327, 564–577, doi:10.1016/j.jhydrol.2005.11.058, 2006.

Drécourt, J.-P., Madsen, H., and Rosbjerg, D.: Calibration framework for a Kalman filter applied to a groundwater model, Adv. Water Resour., 29, 719–734, doi:10.1016/j.advwatres.2005.07.007, 2006.

Eppstein, M. J. and Dougherty, D. E.: Simultaneous Estimation of Transmissivity Values and Zonation, Water Resour. Res., 32, 3321–3336, doi:10.1029/96WR02283, 1996.

Erdal, D.: Bias correction for compensating unresolved subsurface structure in unsaturated flow modelling, PhD thesis, Gottfried Wilhelm Leibniz Universität Hannover, Hannover, 2014.

Erdal, D., Neuweiler, I., and Wollschläger, U.: Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model, Water Resour. Res., 50, 123–147, doi:10.1002/2012WR013443, 2014.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 143–162, 1994.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF, Ocean Dynam., 54, 539–560, doi:10.1007/s10236-004-0099-2, 2004.

Evensen, G.: The Ensemble Kalman Filter for Combined State and Parameter Estimation, IEEE Control Syst. Mag., 29, 83–104, doi:10.1109/MCS.2009.932223, 2009.

Ferraresi, M., Todini, E., and Vignoli, R.: A solution to the inverse problem in groundwater hydrology based on Kalman filtering, J. Hydrol., 175, 567–581, doi:10.1016/S0022-1694(96)80025-4, 1996.

Gómez-Hernández, J., Sahuquillo, A., and Capilla, J.: Stochastic simulation of transmissivity fields conditional to both transmissivity and storativity, 1. Theory, J. Hydrol., 203, 162–174, 1997.

Hantush, M. M. and Mariño, M. A.: Estimation of Spatially Variable Aquifer Hydraulic Properties Using Kalman Filtering, J. Hydraul. Eng., 123, 1027–1035, 1997.

Hendricks Franssen, H. J. and Kinzelbach, W.: 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, 2008.

Hendricks Franssen, H. J., Stauffer, F., and Kinzelbach, W.: Joint estimation of transmissivities and recharges – Application: Stochastic characterization of well capture zones, J. Hydrol., 294, 87–102, doi:10.1016/j.jhydrol.2003.10.021, 2004.

Hendricks Franssen, H. J., Brunner, P., Makobo, P., and Kinzelbach, W.: Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images, Water Resour. Res., 44, W01419, doi:10.1029/2007WR006097, 2008.

Hill, M. C. and Tiedeman, C. R.: Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty, 1st Edn., John Wiley & Sons, New York, 2007.

Hu, L. Y. and Chugunova, T.: Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, Water Resour. Res., 44, 1–14, doi:10.1029/2008WR006993, 2008.

Huang, C., Hu, B. X., Li, X., and Ye, M.: Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source, Stoch. Environ. Res. Risk A., 23, 1155–1167, doi:10.1007/s00477-008-0289-4, 2008.

Kitanidis, P. K.: Quasi-linear geostatistical theory for inversing, Water Resour. Res., 31, 2411–2419, 1995.

Kitanidis, P. K.: The minimum structure solution to the inverse problem, Water Resour. Res., 33, 2263–2272, doi:10.1029/97WR01619, 1997.

Kitanidis, P. K. and Lane, R. W.: Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, J. Hydrol., 79, 53–71, doi:10.1016/0022-1694(85)90181-7, 1985.

Kuhlman, K. L., Hinnell, A. C., Mishra, P. K., and Yeh, T.-C. J.: Basin-scale transmissivity and storativity estimation using hydraulic tomography, Ground Water, 46, 706–715, doi:10.1111/j.1745-6584.2008.00455.x, 2010.

Lessoff, S. C., Schneidewind, U., Leven, C., Blum, P., Dietrich, P., and Dagan, G.: Spatial characterization of the hydraulic conductivity using direct-push injection logging , Water Resour. Res., 46, W12502, doi:10.1029/2009WR008949, 2008.

Li, J. and Elsworth, D.: A modified Gauss-Newton method for aquifer parameter identification, Ground Water, 33, 662–668, 1995.

Li, L., Zhou, H., Hendricks Franssen, H. J., and Gómez-Hernández, J. J.: Groundwater flow inverse modeling in non-MultiGaussian media: performance assessment of the normal-score Ensemble Kalman Filter, Hydrol. Earth Syst. Sci., 16, 573–590, doi:10.5194/hess-16-573-2012, 2012.

Li, L., Srinivasan, S., Zhou, H., and Gómez-Hernández, J. J.: Simultaneous Estimation of Geologic and Reservoir State Variables Within an Ensemble-Based Multiple-Point Statistic Framework, Math. Geosci., 46, 597–623, doi:10.1007/s11004-013-9504-z, 2013.

Li, W., Englert, A., Cirpka, O. A., Vanderborght, J., and Vereecken, H.: 2-D characterization of hydraulic heterogeneity by multiple pumping tests, Water Resour. Res., 43, W04433, doi:10.1029/2006WR005333, 2007.

Loke, M. H. and Dahlin, T.: A comparison of the Gauss-Newton and quasi-Newton methods in resistivity imaging inversion, J. Appl. Geophys., 49, 149–162, doi:10.1016/S0926-9851(01)00106-9, 2002.

Maurer, H., Holliger, K., and Boerner, D. E.: Stochastic regularization: Smoothness or similarity?, Geophys. Res. Lett., 25, 2889–2892, doi:10.1029/98GL02183, 1998.

McLaughlin, D. and Townley, L. R.: A reassessment of the groundwater inverse problem, Water Resour. Res., 32, 1131–1161, doi:10.1029/96WR00160, 1996.

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

Okabe, H. and Blunt, M.: Prediction of permeability for porous media reconstructed using multiple-point statistics, Phys. Rev., 70, 066135, doi:10.1103/PhysRevE.70.066135, 2004.

Oliver, D. S. and Chen, Y.: Improved initial sampling for the ensemble Kalman filter, Comput. Geosci., 13, 13–27, doi:10.1007/s10596-008-9101-2, 2008.

Panzeri, M., Riva, M., Guadagnini, A., and Neuman, S. P.: Data assimilation and parameter estimation via ensemble Kalman filter coupled with stochastic moment equations of transient groundwater flow, Water Resour. Res., 49, 1334–1344, doi:10.1002/wrcr.20113, 2013.

Panzeri, M., Riva, M., Guadagnini, A., and Neuman, S. P.: Comparison of Ensemble Kalman Filter groundwater-data assimilation methods based on stochastic moment equations and Monte Carlo simulation, Adv. Water Resour., 66, 8–18, doi:10.1016/j.advwatres.2014.01.007, 2014.

Panzeri, M., Riva, M., Guadagnini, A., and Neuman, S. P.: EnKF coupled with groundwater flow moment equations applied to Lauswiesen aquifer, Germany, J. Hydrol., 521, 205–216, doi:10.1016/j.jhydrol.2014.11.057, 2015.

RamaRao, B. S., LaVenue, A. M., De Marsily, G., and Marietta, M. G.: Pilot Point Methodology for Automated Calibration of an Ensemble of conditionally Simulated Transmissivity Fields: 1. Theory and Computational Experiments, Water Resour. Res., 31, 475–493, doi:10.1029/94WR02258, 1995.

Sanchez-Vila, X., Guadagnini, A., and Carrera, J.: Representative hydraulic conductivities in saturated groundwater flow, Rev. Geophys., 44, 1–46, doi:10.1029/2005RG000169, 2006.
Scanlon, B., Healy, R., and Cook, P.: Choosing appropriate techniques for quantifying groundwater recharge, Hydrogeol. J., 10, 18–39, doi:10.1007/s10040-001-0176-2, 2002.
Schelle, H., Durner, W., Schlüter, S., Vogel, H.-J., and Vanderborght, J.: Virtual Soils: Moisture Measurements and Their Interpretation by Inverse Modeling, Vadose Zo.ne J., 12, 3, doi:10.2136/vzj2012.0168, 2013.
Schlüter, S., Vogel, H.-J., Ippisch, O., Bastian, P., Roth, K., Schelle, H., Durner, W., Kasteel, R., and Vanderborght, J.: Virtual Soils: Assessment of the Effects of Soil Structure on the Hydraulic Behavior of Cultivated Soils, Vadose Zone J., 11, 3, doi:10.2136/vzj2011.0174, 2012.
Schöniger, A., Nowak, W., and Hendricks Franssen, H.-J.: Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography, Water Resour. Res., 48, W04502, doi:10.1029/2011WR010462, 2012.
Spedicato, E. and Huang, Z.: Numerical experience with newton-like methods for nonlinear algebraic systems, Computing, 58, 69–89, doi:10.1007/BF02684472, 1997.
Sun, A. Y., Morris, A., and Mohanty, S.: Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data, Adv. Water Resour., 32, 280–292, doi:10.1016/j.advwatres.2008.11.006, 2009.
Tong, J., Hu, B. X., and Yang, J.: Using data assimilation method to calibrate a heterogeneous conductivity field conditioning on transient flow test data, Stoch. Environ. Res. Risk A., 24, 1211–1223, doi:10.1007/s00477-010-0392-1, 2010.
Tong, J., Hu, B. X., and Yang, J.: Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field, Stoch. Environ. Res. Risk A., 26, 467–478, doi:10.1007/s00477-011-0534-0, 2011.
Tong, J., Hu, B. X., and Yang, J.: Data assimilation methods for estimating a heterogeneous conductivity field by assimilating transient solute transport data via ensemble Kalman filter, Hydrol. Process., 27, 3873–3884, doi:10.1002/hyp.9523, 2013.
Tonkin, M. J. and Doherty, J.: A hybrid regularized inversion methodology for highly parameterized environmental models, Water Resour. Res., 41, W10412, doi:10.1029/2005WR003995, 2005.
Tsai, F. T. C., Sun, N. Z., and Yeh, W. W. G.: A combinatorial optimization scheme for parameter structure identification in ground water modeling, Ground Water, 41, 156–169, doi:10.1111/j.1745-6584.2003.tb02579.x, 2003a.

Tsai, F. T.-C., Sun, N.-Z., and Yeh, W. W.-G.: Global-local optimization for parameter structure identification in three-dimensional groundwater modeling, Water Resour. Res., 39, 1043, doi:10.1029/2001WR001135, 2003b.

Vogt, C., Marquart, G., Kosack, C., Wolf, A., and Clauser, C.: Estimating the permeability distribution and its uncertainty at the EGS demonstration reservoir Soultz-sous-Forts using the ensemble Kalman filter, Water Resour. Res., 48, W08517, doi:10.1029/2011WR011673, 2012.

Vrugt, J. A., Stauffer, P. H., Wöhling, T., Robinson, B. A., and Vesselinov, V. V.: Inverse Modeling of Subsurface Flow and Transport Properties: A Review with New Developments, Vadose Zone J., 7, 843–864, doi:10.2136/vzj2007.0078, 2008.

Xu, T., Gómez-Hernández, J., Li, L., and Zhou, H.: Parallelized ensemble Kalman filter for hydraulic conductivity characterization, Comput. Geosci., 52, 42–49, doi:10.1016/j.cageo.2012.10.007, 2013a.

Xu, T., Gómez-Hernández, J., Zhou, H., and Li, L.: The power of transient piezometric head data in inverse modeling: An application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field, Adv. Water Resour., 54, 100–118, doi:10.1016/j.advwatres.2013.01.006, 2013b.

Yeh, T.-C. J., Jin, M., and Hanna, S.: An iterative stochastic inverse method: conditional effective transmissivity and hydraulic head fields, Water Resour. Res., 32, 85–92, 1996.

Yeh, W. W.-G. and Yoon, Y. S.: Aquifer parameter identification with optimum dimension in parameterization, Water Resour. Res., 17, 664–672, doi:10.1029/WR017i003p00664, 1981.

Zhou, H., Gómez-Hernández, J. J., and Li, L.: Inverse methods in hydrogeology: Evolution and recent trends, Adv. Water Resour., 63, 22–37, doi:10.1016/j.advwatres.2013.10.014, 2014.

Zou, X., Navon, I. M., Berger, M., Phua, K. H., Schlick, T., and Le Dimet, F. X.: Numerical Experience with Limited-Memory Quasi-Newton and Truncated Newton Methods, SIAM J. Optim., 3, 582–608, doi:10.1137/0803029, 1993.

Table 1. Pumping rates and general model setup*.

Pump	1	2	3	4	5
Rate ($m^3\,h^{-1}$)	9	18	90	0.09	0.9
Start (day)	20	300	200	0	0
Stop (day)	150	365	360	370	300
Model setup	Δx (m)	Δy (m)	dt (h)	z_0 (m)	poro (–)
	50	50	6	0	0.4

* Pumps are numberd as in Fig. 2, z_0 and poro are the homogeneous bedrock elevation and porosity.

Table 2. Parameters and properties used for the generation of the synthetic examples conductivity and recharge fields*.

	$\ln(K)$ $\ln$ (m s^{-1})	R (mm day^{-1})
μ	-8.5	-0.7
σ	1.7	0.1
α (°)	291	17
l_x (m)	2000	5000
l_y (m)	600	500

* μ is the mean, σ the variance, α the rotation angle and l_x and l_y are the correlation lenghts in x and y direction, respectively.

Table 3. Normalized root mean square error for the prediction period*.

	Good	Random	Wrong
R	1.3	1.6	1.9
K	2.6	3.1	17.4
R & K	6.0	13.5	15.0

* According to Eq. (15) for three setups of prior knowledge (good, random, wrong) to estimate recharge alone (R), conductivity alone (K) and to jointly estimate conductivity and recharge (R & K).

Table 4. Normalized root mean square error for the assimilation period.

	Good	Random	Wrong
R	0.3	0.4	0.5
K	1.2	0.9	3.7
R & K	2.2	2.4	3.7

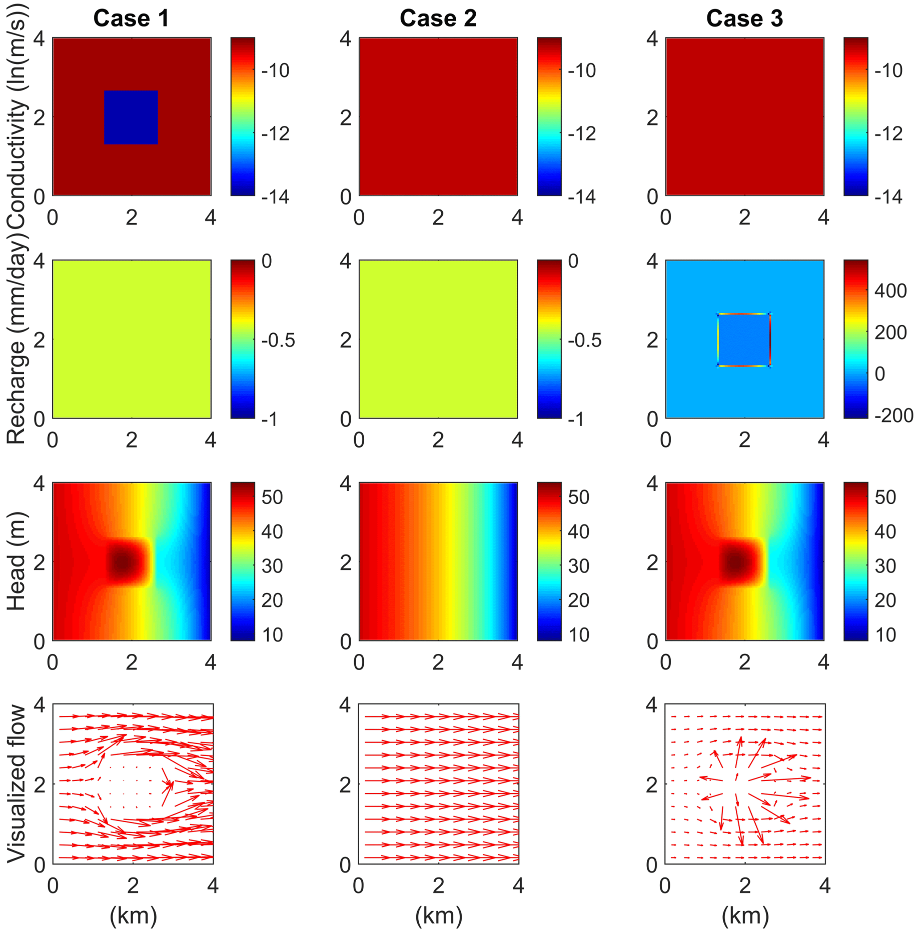

Figure 1. Illustrative example of replacing a heterogeneous conductivity field (left column panels) with a homogeneous conductivity and an effective recharge (right column panels). Please note the different scale on the third recharge plot.

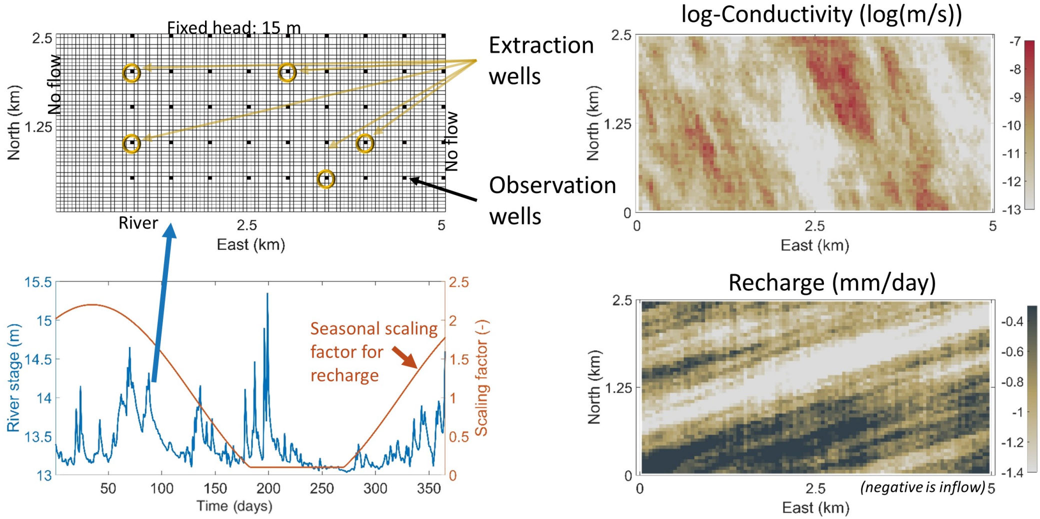

Figure 2. Setup of the synthetic test case used for the parameter field estimations.

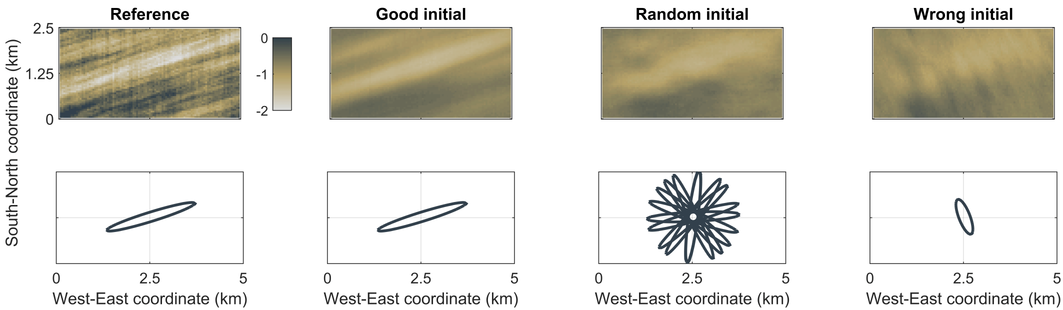

Figure 3. Estimation of stand-alone recharge. Upper panels show the ensemble mean and lower plots the covariance function used to generate the initial ensemble. Please note that the random covariance functions imply drawing the rotation angle from a uniform distribution between 0 and 2π, whereas only a few illustrative examples are shown.

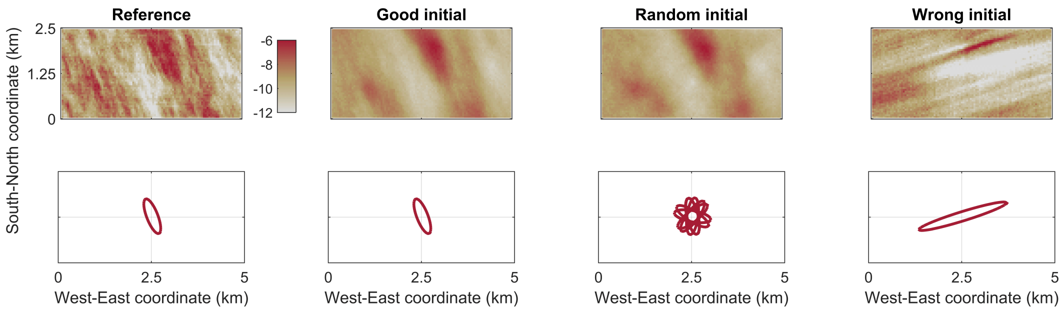

Figure 4. Estimation of stand-alone conductivity. Upper panels show the ensemble mean and lower plots the covariance function used to generate the initial ensemble. Please note that only a few illustrative examples of the random orientation angle of anisotropy are shown.

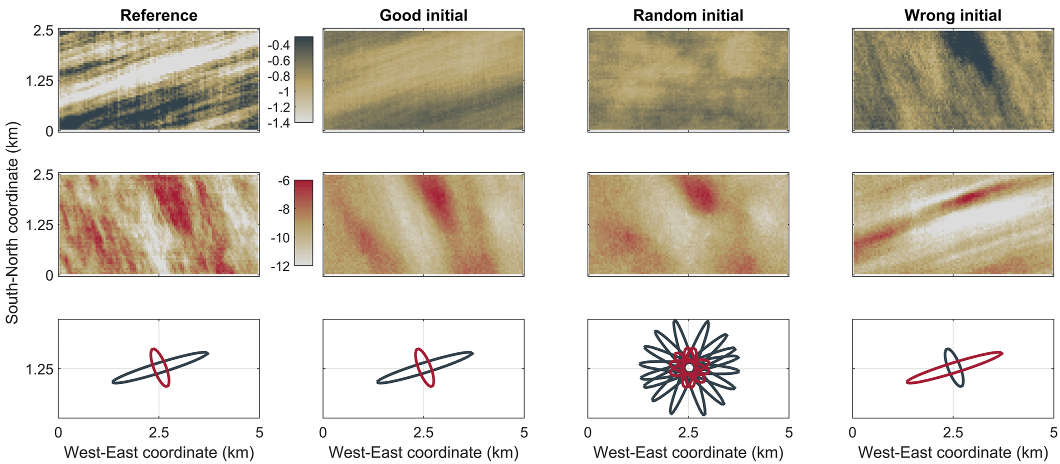

Figure 5. Joint estimation of recharge (top row panels) and conductivity (middle row panels). Shown is the ensemble mean and the covariance functions used to generate the initial ensembles (bottom row panels).

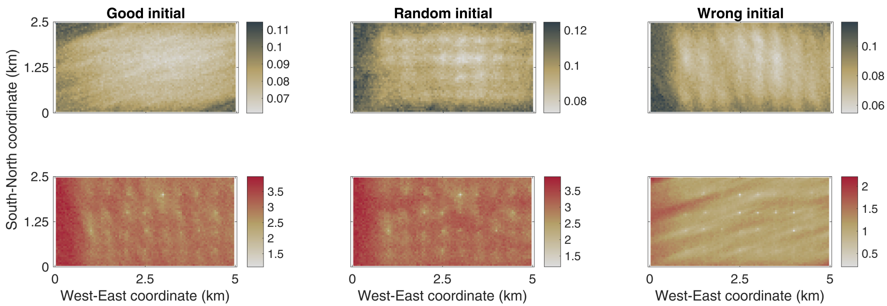

Figure 6. Joint estimation of recharge (top row panels) and conductivity (bottom row panels). Shown is the ensemble variance.

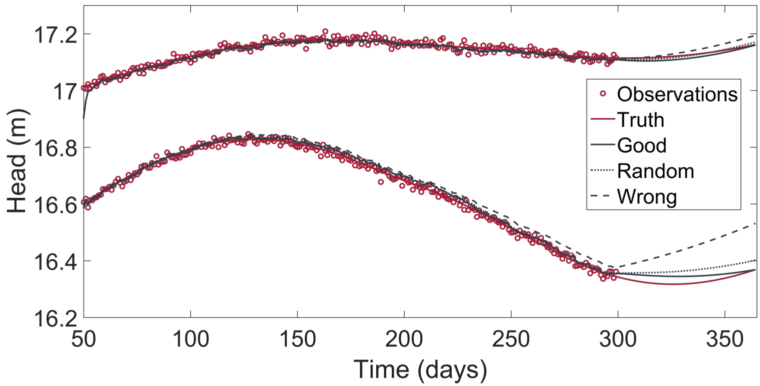

Figure 7. Two head observations plotted over time for the joint estimation of recharge and conductivity. Shown is the ensemble mean. Assimilation is performed from day 50 to day 300 while the remaining days are considered for prediction.