



Towards an effective application of parameter estimation and uncertainty analysis to mathematical groundwater models

Paulo A. Herrera¹ · Miguel Angel Marazuela¹ · Giovanni Formentin¹ · Thilo Hofmann¹

Received: 15 March 2022 / Accepted: 20 June 2022

Published online: 08 July 2022

© The Author(s) 2022

Abstract

Groundwater models serve as support tools to among others: assess water resources, evaluate management strategies, design remediation systems and optimize monitoring networks. Thus, the assimilation of information from observations into models is crucial to improve forecasts and reduce uncertainty of their results. As more information is collected routinely due to the use of automatic sensors, data loggers and real time transmission systems; groundwater modelers are becoming increasingly aware of the importance of using sophisticated tools to perform model calibration in combination with sensitivity and uncertainty analysis. Despite their usefulness, available approaches to perform this kind of analyses still present some challenges such as non-unique solution for the parameter estimation problem, high computational burden and a need of a deep understanding of the theoretical basis for the correct interpretation and use of their results, in particular the ones related to uncertainty analysis. We present a brief derivation of the main equations that serve as basis for this kind of analysis. We demonstrate how to use them to estimate parameters, assess the sensitivity and quantify the uncertainty of the model results using an example inspired by a real world setting. We analyze some of the main pitfalls that can occur when performing such kind of analyses and comment on practical approaches to overcome them. We also demonstrate that including groundwater flow estimations, although helpful in constraining the solution of the inverse problem as shown previously, may be difficult to apply in practice and, in some cases, may not provide enough information to significantly constrain the set of potential solutions. Therefore, this article can serve as a practitioner-oriented introduction for the application of parameter estimation and uncertainty analysis to groundwater models.

Article highlights

- We review the theoretical background that supports parameter estimation and uncertainty analysis techniques applied to groundwater models.
- We demonstrate the application of parameter estimation and uncertainty analysis to mathematical groundwater models through an example.
- We comment on the main challenges that can be faced when applying parameter estimation and uncertainty analysis to groundwater models.

Keywords Calibration · Uncertainty quantification · Numerical modeling · Groundwater hydrology

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42452-022-05086-w>.

✉ Miguel Angel Marazuela, miguel.angel.marazuela@univie.ac.at | ¹Environmental Geosciences, Centre for Microbiology and Environmental Systems Sciences, University of Vienna, Althanstrasse 14, UZA 2, 1090 Vienna, Austria.



1 Introduction

Groundwater models serve as support tools to assess water resources, evaluate management strategies, design remediation systems and optimize monitoring networks [1–4]. As such, results of numerical groundwater models guide decision makers on practical issues of public concern [5]. Therefore, the correct development of groundwater models and their ability to simulate the functioning of real groundwater systems are important beyond the realm of particular engineering or environmental projects. As a minimum, it is expected that models are able to reproduce observed data and that their results can be used as reasonable approximations to evaluate future or hypothetical scenarios for groundwater systems. This requires the integration of observations and models through calibration [6–9]. Furthermore, it is generally also important to quantify the uncertainty of the model results in order to include it into the risk associated to decisions supported by them [10, 11].

The increase in computational power and availability of new software have made possible the development of complex mathematical models to simulate groundwater systems [11–16]. Until recently, such models were developed, calibrated and analyzed in ad hoc ways, e.g. using trial and error for calibration and the manual variation of selected parameters for sensitivity analysis. However, the level of complexity and the number of parameters included in new models have made necessary the use of more systematic approaches [11, 13, 17, 18]. Methodological frameworks for these purposes have become available during the past decades; some were brought to groundwater modelling from other fields, such as the oil industry or climate science [10, 19, 20]. However, the correct application of a systematic approach to develop, calibrate and analyze the results of such mathematical models is challenging, because it requires understanding and applying concepts that are far beyond traditional groundwater hydrology from disciplines such as linear algebra, calculus, and geostatistics. There are books, technical documents and detailed review articles about the subject [7–11, 13, 17, 20–22]. However, there is still much need for a brief, simple and practitioner oriented introductory discussion that explains with the use of only a few equations the functioning of sophisticated software tools that modelers apply nowadays.

Previous articles [16, 23] used simple groundwater models to discuss problems that may arise when trying to calibrate them and assess the uncertainty of their results. Freyberg [23] reported on the results of an exercise performed by graduate students who used trial and error (manual calibration) to calibrate a synthetic

groundwater model using only a set of observed piezometric head values. The main findings of that exercise were that groups of students used very different strategies to accomplish the calibration and that their results were in some cases quite different. Moreover, the quality of the forecasts performed with the calibrated models, assuming slightly different boundary conditions to the ones considered for calibration, was wide-ranging and, surprisingly, it did not correlated well with the quality of the calibration. Calibrated models that had low residuals between simulated and observed piezometric heads did not provide better forecasts than calibrated models that had larger residuals. The group of students that reported the worst prediction used different values of hydraulic conductivity for each cell of the model. Moreover, the quality of the forecasts did not directly correlate with the ability of the modelers to find the correct values of the hydraulic conductivity used in the synthetic model adopted as virtual reality. Freyberg [23] explained some of the results of the exercise based on the theoretical understanding about parameter estimation for groundwater models available at the time. The solution of the inverse problem to estimate parameters of groundwater models, even discounting differences in their conceptualization, admits multiple solutions [7]. The non-uniqueness of the solution can be due to a larger number of parameters than available observations or to correlation between parameters [16]. For example, the problem reported by [23] considered 1389 unknown values of hydraulic conductivity (parameters), but only 22 piezometric head measurements. The findings of [23] established reservations with respect to the end result of the calibration of groundwater models and its usefulness to reduce the uncertainty of forecasts.

Recently, [16] revisited the original problem presented by [23], this time applying the latest available automatic parameter estimation and uncertainty analysis techniques, including mathematical regularization techniques based on subspace methods to decrease the effective number of parameters included in the inverse problem [10, 13]. Their main findings were that automatic calibration of even highly parameterized models could be constrained, i.e. the potential large number of multiple solutions could be reduced to a much lower number, by including a single groundwater flow estimation as additional information to the available piezometric heads. [16] found that the use of automatic frameworks for parameter estimation resulted in better-calibrated models that produced improved forecasts when compared to the original results presented by [23], which demonstrated the potential usefulness of using recently developed methods. Then, contrary to one of the conclusions of [23], highly parameterized

models with many parameters performed better than simple models with fewer parameters. The solution of the inverse problem for highly parameterized models, which demands a large number of model runs, is well suited for increasingly available parallel platforms. Thus, the calibration of this type of models is becoming more accessible in practical applications. On the other hand, [16] also concluded that good calibration does not necessarily imply good forecasts assuming slightly different forcing terms, e.g. boundary conditions, to the scenario considered for calibration.

The main objective of this article is to explain, using an example, the main concepts required to calibrate a physically based mathematical model to simulate a groundwater system, assess its sensitivity and estimate the uncertainty of its results. We explain the main steps that are required and comment on some of the main pitfalls that can arise. We use as example a simplified model of a real natural setting previously studied as part of a long-term environmental study [24, 25]. To promote the use of simple analytical calculations to gain understanding that can be useful to interpret the results of more complex numerical models and to overcome some numerical challenges for obtaining reliable numerical solutions for this particular setting, we use an analytical model instead of a numerical groundwater model [26]. Therefore, this article aims to address the needs of modelers interested in applying methods for parameter estimation and uncertainty analysis. It also attempts to be useful for decision makers who base their decisions on the results of groundwater models and need to understand their potential and limitations. Finally, since this article provides a practitioner-oriented introduction

to the topic, it can also serve as material for an advanced graduate level groundwater modeling course.

This article is structured in four main sections, including this introduction. In the next section we describe the Example problem that we use throughout the article. Next, we present a mathematical analysis of the application of parameter estimation and uncertainty analysis for the Example problem. Finally, we provide some concluding remarks based on the results presented in the other sections.

2 Example description

We will consider as example the estimation of the position of the phreatic surface (water table) for an aquifer located between two rivers as shown in Fig. 1. Both rivers are separated by a distance L [L], the aquifer receives a constant recharge rate R [L/T], and we assume that there is only one observed value of the piezometric head h [L], h_w , at the position of an existing well x_w [L]. This example corresponds to an idealization of a real groundwater system previously analyzed as part of an environmental study [24, 25] and, hence, it provides a realistic case problem. The numerical solution for this problem is challenging because it requires simulating the position of the water table, which for this case is very sensitive to the grid resolution due to the potential occurrence of high hydraulic gradients as result of the combination of recharge and low hydraulic conductivity. See the Supplementary Material for details about the difficulties to obtain reliable numerical solutions for this example.

This example provides a reasonable case of study of practical interest that is similar to many other cases

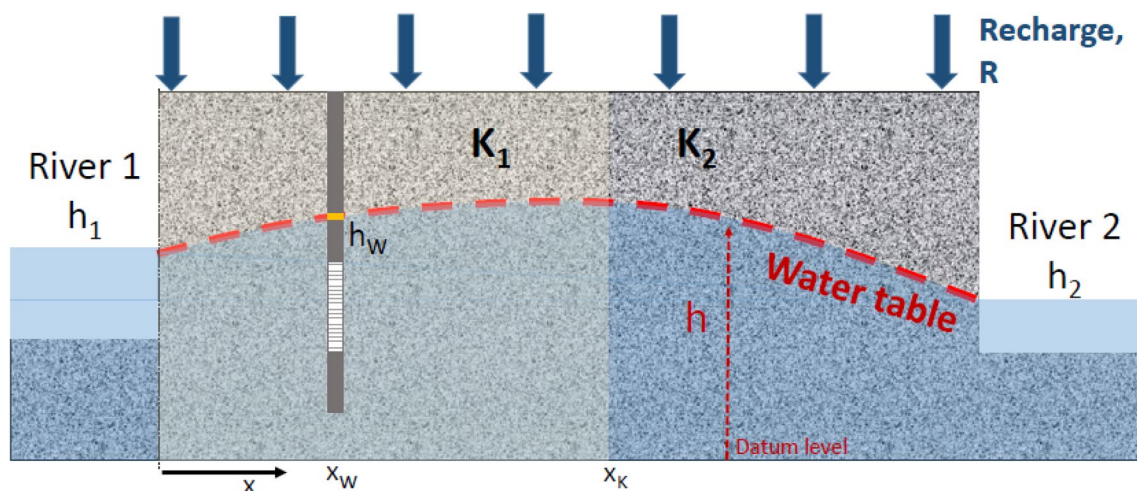


Fig. 1 Schematic of conceptual problem: Water table position for a phreatic aquifer located between two rivers that are connected to the groundwater system, so that they act as prescribed piezometric head boundary conditions

(i.e. unconfined aquifer with recharge) that occur often in practice, while it includes only a limited number of parameters. Thus, the detailed explanation of the application of the methods for parameter estimation and uncertainty analysis can be kept brief. The development of a mathematical model to simulate a groundwater system similar to this example includes three main parts:

- (a) *Conceptual model* We consider only the part of the subsurface that is completely saturated and that the aquifer behaves as unconfined, so that all recharge R enters directly and immediately to the aquifer. Furthermore, the two rivers are connected to the aquifer and can be modeled as prescribed head boundary conditions, h_1 and h_2 . The water elevation in both rivers and the recharge rate are relatively stable, so that we can model the system as steady-state. Given the different magnitude of the flow discharge in both rivers, it is likely that the shallow aquifer contains at least two materials of different grain size with an interface located at position x_K [L]. Based on experience and comparison with similar sites, it is estimated that the values of $K_1 \in k_1[0.1, 10]$ [L/T] and that $K_2 \in k_2[0.1, 10]$ [L/T], where k_1 and k_2 represent a scale of the magnitude of K for both materials. Therefore, the range of variability of K for each hydrogeological unit is two orders of magnitude, which for hydraulic conductivity is a reasonable estimate without additional information [27]. We use dimensionless variables to define parameters that have the same units: $f_K = k_2/k_1$, $f_R = R/K_1$ and $f_h = (h_2 - h_1)/L$. The parameters are set such that recharge is never higher than hydraulic conductivity.
- (b) *Mathematical model* Darcy Law describes the saturated flow system, which combined with the mass balance equation results in a partial differential equation. We assume that the Dupuit approximation is valid, i.e. flow is preferentially horizontal, so that vertical piezometric gradients and velocity component are negligible [28]. Furthermore, the thickness of the saturated aquifer is equal to the piezometric head or water table elevation, so that the datum level considered for computing the energy of the flow coincides with the horizontal bottom of the aquifer.
- (c) *Evaluating solution* Given these assumptions, the piezometric head as a function of the position measured from River 1, x , can be calculated from

$$h^2(x) = \begin{cases} a_1x^2 + b_1x + c_1, & x \leq x_K \\ a_2(x - x_K)^2 + b_2(x - x_K) + c_2 & x > x_K \end{cases} \quad (1)$$

The analytical solution depends upon the hydraulic conductivity K and piezometric head h . The constant coefficients of both polynomials can be evaluated from both boundary conditions, $h(x = 0) = h_1$ and $h(x = L) = h_2$, and imposing continuity of the water table and flow through the interface, i.e. $h(x_K^-) = h(x_K^+)$, and $-K_1 \left[h \frac{\partial h}{\partial x} \right]_{x_K^-} = -K_2 \left[h \frac{\partial h}{\partial x} \right]_{x_K^+}$.

The model can be written most concisely as $h(x) = h(\{K_1, K_2\}; \{h_1, h_2, R\}, x)$, where we can identify the following sets of output variables $\mathbf{o} = \{h\}$, input parameters to be estimated $\mathbf{e} = \{K_1, K_2\}$, and forcing input variables $\mathbf{f} = \{h_1, h_2, R\}$. We have left R as a forcing parameter for reasons that will become clear later.

Although we perform this analysis based on an analytical solution, the same type of reasoning applies to a numerical model that solves similar partial differential equations, since the numerical solution should converge to the analytical solution. The comparison between the analytical and numerical solutions presented in the Supplementary Material, demonstrates that despite the presence of vertical gradients in this case, the analytical solution based on the Dupuit approximation provides an adequate approximation. For this example, the solution (analytical or numerical) for the position of the water table at a given location in the aquifer can be highly non-linear with respect to the input parameters (K_1 and K_2) depending on the settings of the problem as shown in the Supplementary Material. The non-linear behavior of the solution can represent additional challenges for the application of automatic techniques to estimate parameters and for assessing the uncertainty of the model results.

3 Mathematical analysis

3.1 Water table

Considering a homogeneous aquifer with a single material, i.e. single K value, the mathematical solution reduces to

$$h^2(x) = ax^2 + bx + c \quad (2)$$

where $a = -R/K$, $b = (h_2^2 - h_1^2 - aL^2)/L$ and $c = h_1^2$. Thus, different values of recharge R and hydraulic conductivity K that result in the same ratio R/K produce the exact same value of $h(x)$, i.e. K and R are strongly correlated. We can only estimate either K or R , or the combination of both R/K , which sometimes can be used as a surrogate or *modified parameter* for calibration. Therefore, we excluded R from the set of parameters that must be estimated. It is also direct from (2) that piezometric head depends in a

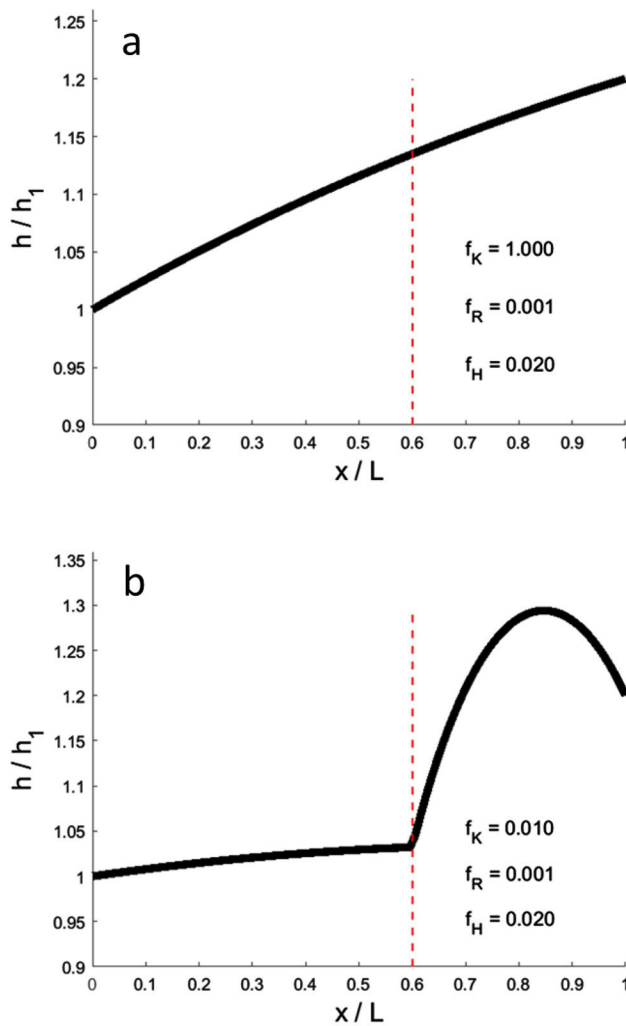


Fig. 2 Water table as a function of position between both rivers for two ratios of hydraulic conductivity, $f_K = K_2/K_1$. **a** Homogeneous and **b** heterogeneous aquifer. Dashed red line shows position of interface between materials x_K

non-linear fashion with the input parameters of the problem, K and R .

Figure 2 shows simulated water table elevations for the homogeneous case ($f_K = 1.00$). The water table has a parabolic shape that is the result of the difference between the prescribed head values at both boundaries and recharge that produces the mounding of the water table between both rivers. For this case, the magnitude of the maximum mounding due to recharge is proportional to R/K [26]. For $f_R = 1 \times 10^{-3}$ mounding is limited and all recharged water exits the model domain through the boundary with lower prescribed head on the left side, while water also enters the domain through the right boundary with higher h . Increasing the recharge rate, while keeping the values of hydraulic conductivity, i.e. larger f_R , would result in larger mounding.

Figure 2 also shows the solution for heterogeneous case ($f_K = 0.01$). Since we assume that K_2 is a couple of orders of magnitude smaller than K_1 , we observe that the mounding effect within material 2 is significantly larger than in material 1, and a flow divide occurs within material 2. Then, the only water that enters and exits the domain comes from recharge, i.e. the total flow is $Q = RL [L^2/T]$, which exits through the right and left boundaries.

3.2 Sensitivity analysis: Jacobian matrix

We numerically compute the Jacobian, \mathbf{J} , or *sensitivity matrix* to quantify the local sensitivity of the model, i.e. the magnitude of the changes of the results due to changes in the input parameters. The element of \mathbf{J} that corresponds to the result r and input parameter p , is defined as

$$J_{rp} = \frac{\partial r}{\partial p} \quad (3)$$

Then, \mathbf{J} can be evaluated using a first-order forward finite-difference approximation, to obtain

$$\mathbf{J}(\{K_1, K_2\} | \mathbf{f}, \mathbf{x}) = \left[\frac{\Delta h}{\Delta K_1}, \frac{\Delta h}{\Delta K_2} \right] \quad (4)$$

Figure 3 shows normalized values of the components of the sensitivity matrix as function of position along the model domain. The fact that the components of \mathbf{J} are equal to 0 at both boundaries indicates that the results of the model do not change at those positions independently of the values of the parameters, as it would be expected when applying prescribed head values at both boundaries. Two interesting results derive from this. First, values of head near the boundaries are less sensitive to changes in parameters than values located toward the center of the domain. Second, observations of wells located near both extremes are *less relevant* or provide less meaningful information for the calibration of K than observations collected near the position of the interface between both materials, since the simulated head at those positions does depend upon the values of the parameters [10, 17, 20]. This last conclusion has importance for real-world models that consider observations measured at tens or hundreds of different locations within a model domain. Moreover, if we had to decide where we should collect data or drill a new borehole to measure piezometric levels, we would select a location near the center between both rivers, since the variability of the results is greatest at that location.

It is important to notice the drastic change in the shape of J_{12} between the homogeneous and heterogeneous cases. In the latter, the simulated water table within material 1 is almost completely independent of the value of K_2 . Hence, any measurement of piezometric

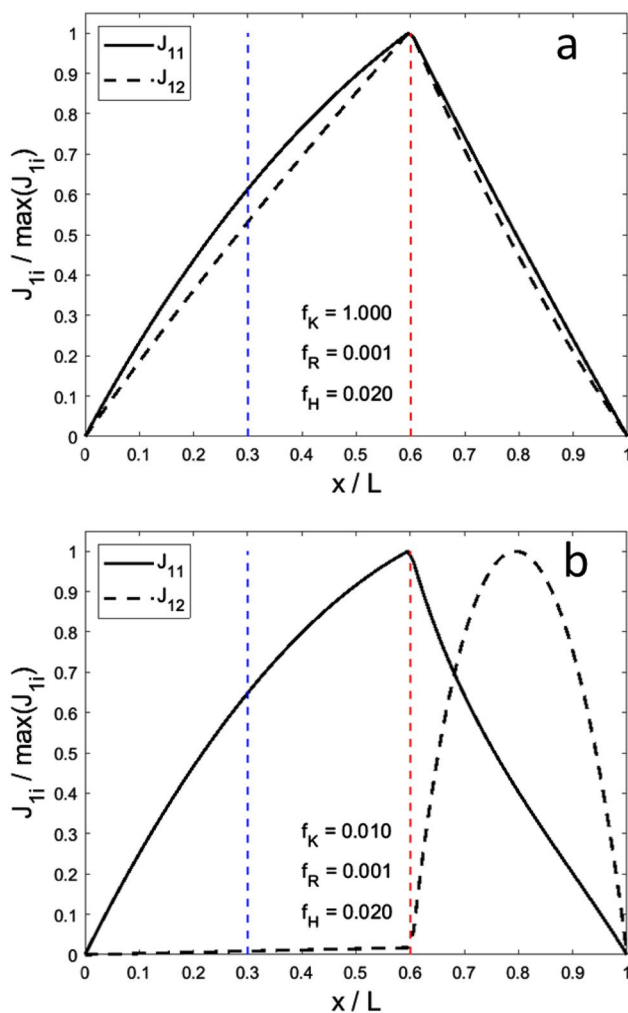


Fig. 3 First component of sensitivity matrix or Jacobian (\mathbf{J} , defined in Eq. 3) as function of the position of the observation well. **a** Homogeneous and **b** heterogeneous aquifer. Dashed red line shows position of interface between materials x_k and blue line shows position of well with measured head x_w

levels taken from a borehole located within material 1 would provide no relevant information for the estimation of K_2 . This result can vary between both extremes presented here, homogeneous and heterogeneous case, depending upon the specific value of f_K .

The fact that the components of the sensitivity matrix, \mathbf{J} , change with x and with the ratio between both values of K in a non-linear fashion, confirms that it only provides a measure of the *local-sensitivity* of the model around a given set of input parameters. This information may not be enough to evaluate the overall performance of complex models that include non-linear relations between parameters and results and many parameters for other material distributions or locations.

3.3 Calibration: inverse problem

We use the single measurement of h that is available to constrain or *condition* the values of K_1 and K_2 , i.e. to calibrate the parameters of the model. Once we have evaluated the Jacobian of the model at the position of the well with observed piezometric head h_w , we can write a first-order Taylor series approximation to get

$$h_w(\{K_1 + \Delta K_1, K_2 + \Delta K_2\} | \mathbf{f}, x_w) - h_w(\{K_1, K_2\} | \mathbf{f}, x_w) = \mathbf{J} \cdot \Delta \mathbf{K} + \epsilon \tag{5}$$

where $\Delta \mathbf{K} = [\Delta K_1, \Delta K_2]$ and ϵ corresponds to errors of the approximation that can be neglected. The last expression corresponds to Newton’s method. Assuming that h_w^i represents the solution of the model for an initial set of variables $\{K_1, K_2\}$ and h_w^o corresponds to the observed value at the well, we can write

$$h_w^o - h_w^i = \mathbf{J} \cdot \Delta \mathbf{K} \tag{6}$$

The solution of the linear system provides an expression to compute the magnitude of the variation of parameters K_i that required so that the results of the model are similar to the observed data. From linear algebra, we know that \mathbf{J} has an inverse only if it is square and no singular, i.e. $\det(\mathbf{J}) \neq 0$. For the general case of rectangular matrices with more rows than columns, as happens when there are more observations than independent parameters, the solution of (6) can be found by using methods to solve over-determined linear systems, e.g. least squares [29]. When there are more parameters than observations, there is only a reduced number of parameter combinations that can be estimated [10, 13]. The combination of parameters that cannot be uniquely estimated, because they do not have effect on the output of the model, correspond to the *null-space* of \mathbf{J} [30], i.e. all vectors \mathbf{x} such that $\mathbf{J} \cdot \mathbf{x} = 0$ [29].

To understand what happens when there are more parameters than observations, as in this example and in most real world applications, we can rewrite (6) as a system of algebraic equations, which for this case contains only one equation,

$$\Delta h_w = J_{11} \Delta K_1 + J_{12} \Delta K_2 \tag{7}$$

where $\Delta h_w = h_w^o - h_w^i$. Then, given an arbitrary variation of K_1 , we can compute the corresponding variation of K_2 needed to obtain a model that reproduces the observed data, as

$$\Delta K_2 = \frac{\Delta h_w - J_{11} \Delta K_1}{J_{22}} \tag{8}$$

Therefore, for this case we can find infinite solutions to the inverse problem by selecting an arbitrary variation of K_1 and computing the corresponding variation of K_2 from

the last equation. In summary, this example with 2 parameters and 1 single observation is equivalent to an algebraic system with 2 unknown variables and a single equation, which admits an infinite number of solutions. This is a simple and intuitive demonstration of the *non-uniqueness* of the solution for the parameter estimation problem when there are more parameters than observations, as it happens often for groundwater models due to the heterogeneity of the parameters that they include, e.g. hydraulic conductivity, recharge rates, etc.

An important step after computing the solution of the linear system, is to verify that the linear approximation is valid by evaluating the model using the estimated set of parameters and comparing the result with the observed measurements. In most cases, models are non-linear, so that differences between observations and the results provided by the calibrated model based on the linear approximation are expected. Moreover, in real cases, measurement errors affect observations; while simplifications in model structure and parameter distribution, referred to as structural errors, influence the results of simulations [10]. For simplicity, we consider that such errors do not affect our synthetic example.

The set of parameters that fit the observation, i.e. $\mathbf{e}^* = \{K_1, K_2\}$, found by solving (8) or by directly applying (6) is computed through an iterative process, since they are based on the local linear approximation (5), which is only valid for a small region around the set of parameters that are used to evaluate it. We opted to use (8), since it is straightforward to implement according to the following steps:

1. Select an initial set $\{K_1^0, K_2^0\}$.
2. Evaluate model to compute initial result, h_w^i and check if $\Delta h_w > \text{TOL}$, where TOL is a prescribed tolerance to stop iterations, which should be chosen such that it is compatible with the potential magnitude of measurement errors and the overall precision level of the model. If the last condition is true, continue to next step, else stop iterations.
3. Evaluate components of the Jacobian around the initial parameter set. For this model with two parameters, it requires evaluating the model two additional times.
4. Define a change in K_1 , ΔK_1 , and apply (8) to calculate a corresponding ΔK_2 . For example, ΔK_1 can be specified as a fixed percentage of the current value of K_1 .
5. Calculate new values $K_1 = K_1^i + \Delta K_1$ and $K_2 = K_2^i + \Delta K_2$.
6. Check if new values are within ranges defined during the conceptual model stage. If they are not, stop iterating and report failure to find a satisfactory solution. Then, only a limited number of initial parameters may result in a satisfactory solution. In practice, this

can result in a large computational overhead due to attempting to solve many cases that end up in failure.

7. Evaluate model for K_1 and K_2 to compute h_w and check residual versus tolerance. If residual is still too big, set last parameter values as initial set and go back to Step 1.

We applied the previous algorithm to estimate combinations of K_1 and K_2 that fits the observed head h_w . The true or target value \mathbf{e}^* for this synthetic example is known, since we use it to generate the observed head h_w . We used as initial guesses random combinations of both variables generated assuming a uniform distribution of $\ln(K)$ within the specified ranges of K for each material. Figure 4 shows

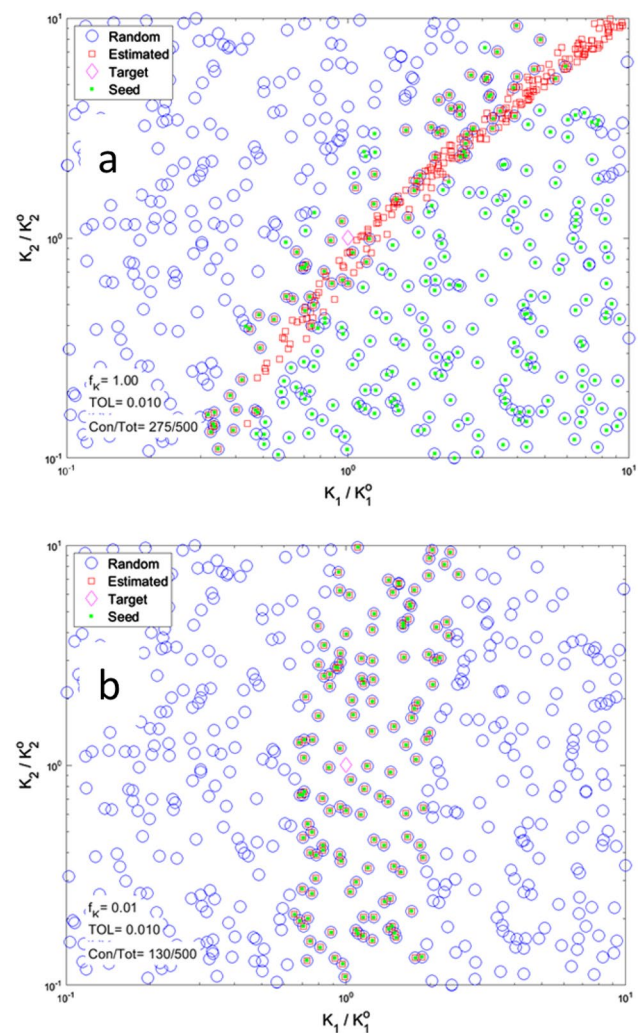


Fig. 4 Combinations of K_1 and K_2 assuming two different ratios, $f_K = 1.00, 0.01$; for randomly selected values within specified range (blue dots) and after parameter estimation (red squares). Green dots show initial parameter values (seed) for the cases for which convergence criteria was satisfied. **a** Homogeneous and **b** heterogeneous aquifer

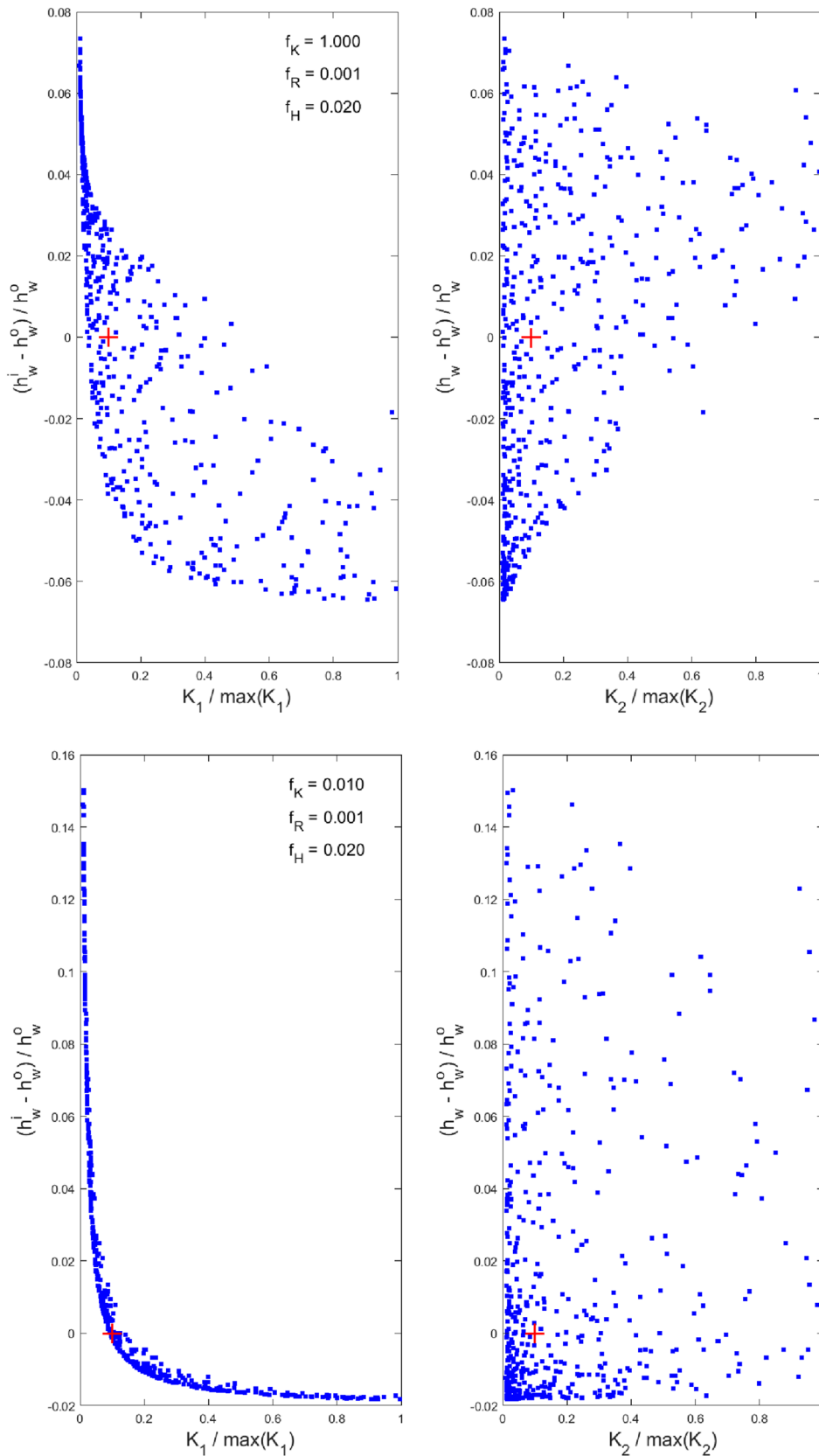


Fig. 5 Normalized residual (i.e. difference between simulated and observed head) at position of observation well versus normalized values of hydraulic conductivity for randomly generated parameter sets. Red cross indicates values for the adopted virtual reality site. **a** Homogeneous and **b** heterogeneous aquifer

the initial and estimated combinations of parameters for the homogeneous and heterogeneous cases. While the initial sets cover the full range of each parameter, the ones that result in satisfactory combination of parameters define a much smaller region that coincides with an almost straight line defined by (8), which passes through the target or true parameter set. For the heterogeneous case, the estimated parameters define an almost vertical rectangular region in the plane $K_1 - K_2$, which indicates that the simulated piezometric head at the well is much less sensitive to the value of K_2 than to the value of K_1 . This is a consequence of: (1) the location of the well with observed data within material 1, and (2) the ratio between f_K considered for the heterogeneous case (lower K_2). It also agrees with the estimation of the local sensitivity of the model shown in Fig. 3. Figure 4 also shows the initial sets of parameters for which the iterative process finished with a solution that satisfied the convergence criteria fixed at 1.5% of the observed head h_w . Note that for the heterogeneous case only 130 out of 500 initial sets of parameters resulted in satisfactory results.

As for most iterative algorithms, the convergence rate and the estimated solution depend on: (1) the initial set of parameters, (2) the tolerance or threshold value defined as criterion to stop the iterative process, TOL; and (3) the size of the potential range defined for each parameter during the conceptual model stage, since larger ranges would potentially result in more iterations. Therefore, finding acceptable solutions for the parameter estimation problem by using this type of algorithms involves a trade-off between the tolerance criteria that is set and the amount of computational effort considered acceptable. More relaxed criteria result in more parameter combinations that fit the observed data, but at the cost of larger residuals. This trade-off is important to consider for reducing the uncertainty of the calibrated model results, as we will show next.

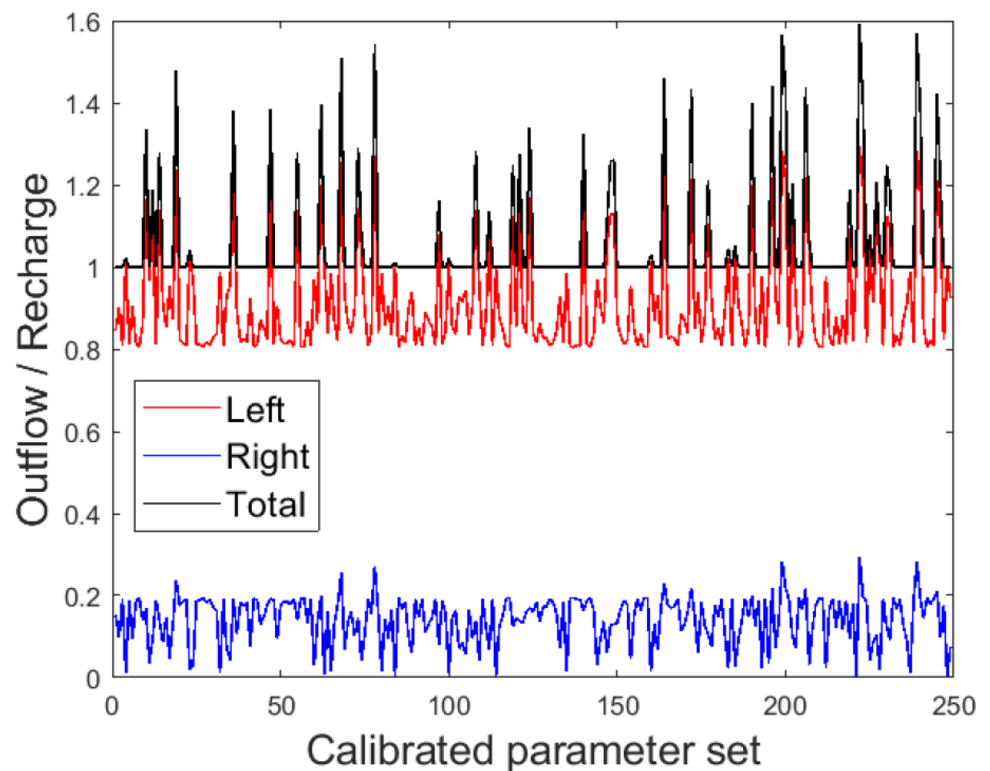
Figure 5 shows normalized residuals of piezometric head for the observation well versus normalized randomly generated values of K_1 and K_2 . For the homogeneous case, there is a wide range of values of K_1 and K_2 (from the minimum value to almost 60% of the maximum value) that can result in simulated head at the well that is similar to the observed or target value (indicated by a red cross). However, for the heterogeneous case ($f_K = 0.01$) the range of potential values of K_1 that result in a simulated value similar to the observed one is much reduced, while the range for K_2 covers the full potential range of values considered as part of the conceptual model. For the heterogeneous case, the relation between residuals and K_1 is well represented by a single decreasing line from low to high values of K_1 .

Previous studies have found that introducing a few estimates of groundwater flows as additional observations can reduce the number of solutions for the inverse problem [16]. Figure 6 shows computed left and right outflows from the model considering only calibrated parameter sets for $f_K = 0.01$. For a large number of estimated parameter sets, the water table includes a water divide within material 2 as shown in Fig. 2, so that water can flow to the left or the right depending on the location. Thus, the sum of both outflows (right and left) is equal to the total recharge. For other cases, when the total flow is larger than the total recharge, water can also enter the domain through the right boundary, thus it can only flow from right to left at any given location of the domain. This means that groundwater flow at some locations can even have different direction between calibrated models. Figure 6 shows that all calibrated parameter sets result in outflows through the left boundary that vary between $\pm 20\%$ of the total recharge. This means that for this particular example, any flow estimation that could be useful to discriminate between different calibrated parameter sets should have a level of confidence higher than 20% of the total flow, which is seldom achievable in real world projects. This could be a major limitation for the use of flow estimations to constrain the solution of the inverse problem for highly parameterized models, since groundwater flows cannot be measured except in very few cases and only with large margins of error.

Note that for this simple problem, the algorithm requires up to five iterations to find an acceptable solution, and that each iteration requires evaluating the sensitivity matrix, which means running the model $N_e + 1$ times. In addition, only 25% of the initial guesses provided solutions that were considered acceptable according to the set tolerance, which was relatively relaxed for this demonstration. Then, many iterations and computations did not produce a satisfactory solution. The method that we used is very basic and it does not incorporate some relatively easy improvements that have demonstrated to accelerate convergence, such as using transformed variables or using a modified Newton method by including an acceleration factor as in the Levenberg–Marquardt method implemented in software platforms for parameter estimation [13, 20]. However, these findings point to the importance of using *good* initial guesses, and that the computational cost of applying this type of algorithms can grow very quickly with the number of parameters and the number of initial sets considered.

Sometimes the problem of finding suitable parameter values expressed in (5) is recast into an optimization problem that aims to reduce the value of an objective function to assess the goodness of fit between observed and

Fig. 6 Outflows through left and right boundaries divided by total recharge for calibrated parameter sets for the heterogeneous aquifer, $f_k = 0.01$



simulated variables. This is, in most cases, equivalent to making the residuals in the left side of (5) close to zero [10]. The main advantages of this formulation is that using an objective function allows assigning different weights for the residuals of observed variables at different locations and, also, introducing additional penalty terms to further force the calibrated model. For example, penalty terms can be included to enforce estimated flow discharges, or to force the calibrated parameters to be close to preferred values assigned during the conceptual model stage. Nevertheless, the problems discussed previously such as non-uniqueness of the solution and dependency on the convergence criteria still apply.

3.4 Uncertainty analysis

There are three main sources of uncertainty that can affect model results, which we review next.

a. Uncertainty due to parameter distribution

We use the model to evaluate the water table for different random combinations of K_1 and K_2 that are within the ranges defined as part of the conceptual model, which is a form of *global sensitivity analysis*. Figure 7 shows simulated water table for 500 different combinations of K_1 and K_2 generated using a uniform distribution of $\ln(K)$ within the specified ranges.

For the homogeneous aquifer case ($f_k = 1.00$), the simulated water table differs from the true solution by up to 15% of the prescribed head at the left boundary, h_1 ; however the difference increases up to more than 100% for $f_k = 0.01$ (heterogeneous aquifer). Those few cases represent extreme situations that are the result of less likely combination of parameters, however, since all values of K are within the specified ranges, it is not possible to discard any of them without additional information. Furthermore, the variability between the different realizations is higher at the center of the simulated domain for the homogeneous case, while it is maximum at around the middle point of the area occupied by material 2 (lowest K) in the heterogeneous case. The magnitude of the variability of the simulated water table provides an indication of the *prior* uncertainty of the model results, which is only constrained by *soft-knowledge* before calibration. Despite the advantages and the valuable information that this type of analysis provides, it is seldom performed in practice, because it may still be excessively time consuming for models with hundreds or thousands of parameters, which require several hours or days to complete [31].

b. Uncertainty due to parameter estimation

By including additional information about observed measurements, we can decrease the uncertainty of the model results at least at the locations where data is

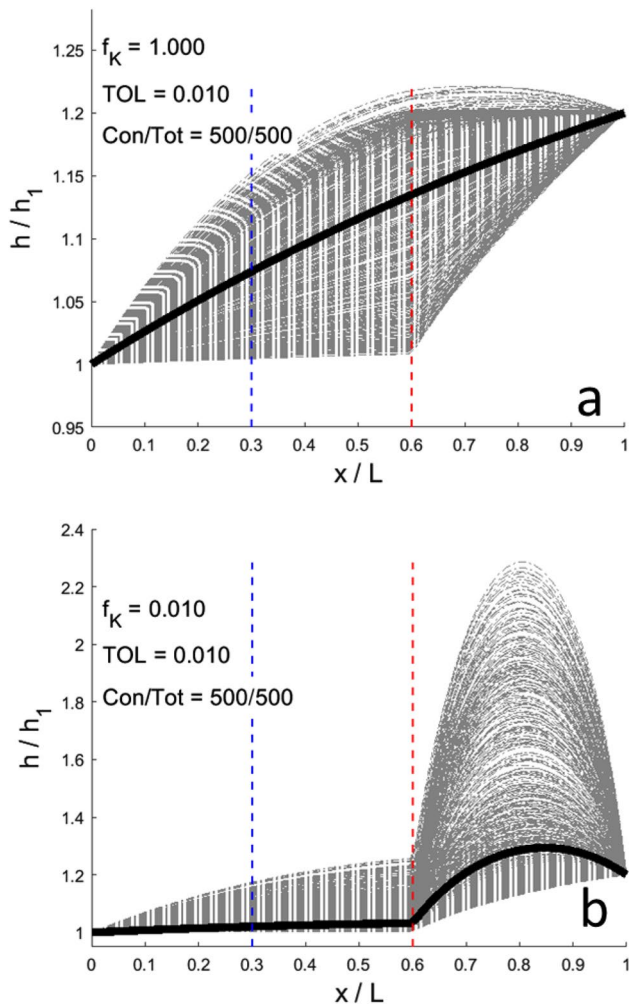


Fig. 7 Water table for different combinations of randomly generated values of K_1 and K_2 . Thick black line shows the water table for the combination of parameters that represents the virtual reality site. Dashed red line shows position of interface between materials x_k and blue line shows position of well with measured head x_w . **a** Homogeneous and **b** heterogeneous aquifer. Note: f_k corresponds to ratio between both scales k_1 and k_2 , thus for $f_k = 1.0$, K_1/K_2 ranges between $[0.01, 100.0]$

available. Since we created a synthetic base scenario or *virtual reality* by evaluating the model for a given set of parameters to generate the observed value of head at x_w , we can also compare the results of the model at other locations between both rivers. Figure 8 shows the simulated water table considering only sets of calibrated parameters, i.e. the ones that result in a simulated h_w similar to the observed value. A small tolerance criteria or threshold limit, TOL, set during the parameter estimation stage, results in a better match between simulated and observed h_w , while adopting a larger limit for the tolerance results in the opposite:

larger differences between simulated and observed h_w , together with more combinations of parameters that satisfy the converge criteria. Therefore, the tolerance threshold must be selected with care considering a trade-off between the need for reducing uncertainty of the model results and the computational cost.

Figure 8 shows that the selection of the convergence criteria must also consider the possibility that for some problems, even restrictive convergence criteria used to evaluate the difference between simulated and observed values at a few locations (e.g. at the well in this case) may not imply a significant reduction of the uncertainty of the model results at other locations. It is also immediate from Fig. 8 that all simulations satisfy the convergence criteria imposed at the well, but that depending on the ratio between the hydraulic conductivity of both materials f_k , the difference between the results of the model and the adopted virtual reality case can be quite large at other locations. Figure 8 also shows in a simple way that the non-uniqueness of the parameter estimation problem is equivalent to having a large number of alternative parameter sets that reproduce the observed data with similar precision and thus, are equally likely to be correct. A consequence of this is that, ideally, every calibrated model should be reported with at least a few sets of parameters that are equally *likeable* in the sense that they reproduce observations. A *null-space* Monte Carlo approach, which reuses some of the results computed as part of the solution of the inverse problem, can be an economical and practical way to produce multiple alternative sets of parameters that explain observations [10].

To quantify uncertainty, it is customary to identify the range of possible outcomes and the probability associated to intervals of those ranges. Histograms as the ones shown in Figs. 9 and 10 summarize well this type of information. Figures 9 and 10 show results for the heterogeneous aquifer ($f_k = 0.01$), while similar results for the homogeneous aquifer are included in the Supplementary Material. Figure 9 shows the probability of simulated heads at the well h_w versus the normalized simulated value h/h_0 . We used 5000 initial different parameter sets to obtain reasonable accurate estimations of the probabilities of the simulated heads, which allows quantifying the inherent uncertainty of the model result without conditioning them to observed data. In a Bayesian framework, the resulting simulated heads would represent the prior probability distribution of the variable [10]. This particular set of initial guesses resulted in only 1216 sets of estimated parameters that matched the observed h_w . The piezometric head values simulated consider-

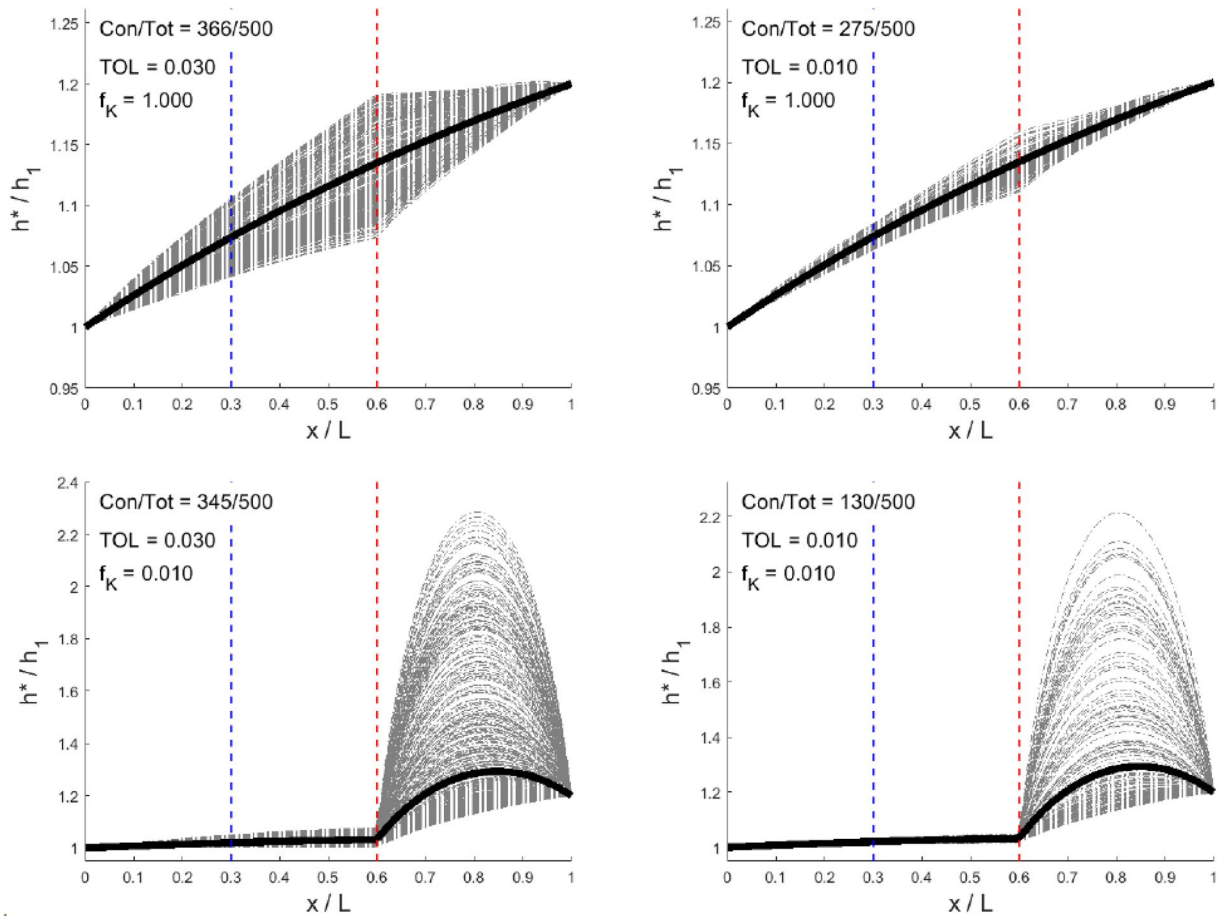


Fig. 8 Simulated piezometric head h^* using estimated parameters K^* for different tolerance threshold (TOL). Thick black line shows the water table for the combination of parameters that represents the virtual reality site. Dashed red line shows position of interface

between materials x_K and blue line shows position of well with measured head x_w . Note: f_K corresponds to ratio between both scales of hydraulic conductivity, k_1 and k_2 , thus for $f_K = 1.0$, K_1/K_2 ranges between $[0.01, 100.0]$

ing only the estimated parameter sets describe the posterior probability distribution. Additional analysis considering different sets of randomly generated initial parameter sets demonstrated that this number of realizations provided probabilities that are significant and are consistent between different groups of random parameter sets.

When randomly generated sets of parameters are used to compute the probabilities, the range of potential outcomes of the model is relatively large (about 20% of the true value), but most of the simulations produce values that are close to the true value, i.e. values of h/h_0 near 1 have highest probability. The shape of the probability distribution depends on the model features (see computed probabilities for the homogeneous aquifer included in Supplemental Material), but also on the sampling strategy used to define the input

parameter sets. For this example, the true parameter set, i.e. the one that defines the adopted virtual reality, was selected as centered within the ranges, hence the probability is higher for values of h similar to the true values. If we consider only calibrated sets, then the range of potential outcomes of the model for h_w decreases to less than 4%. Even though the estimated set of parameters results in a reduction of the uncertainty of the model results around the single location with information, x_w , the uncertainty of the results for other locations may not be much reduced. For example, Fig. 10 shows the probability of the simulated head at approximately the middle point of the area occupied by material 2 ($x_2 = 0.8L$). The range of potential values of h computed for x_2 is similar for the random and estimated parameter sets, however the probability of each interval is slightly different.

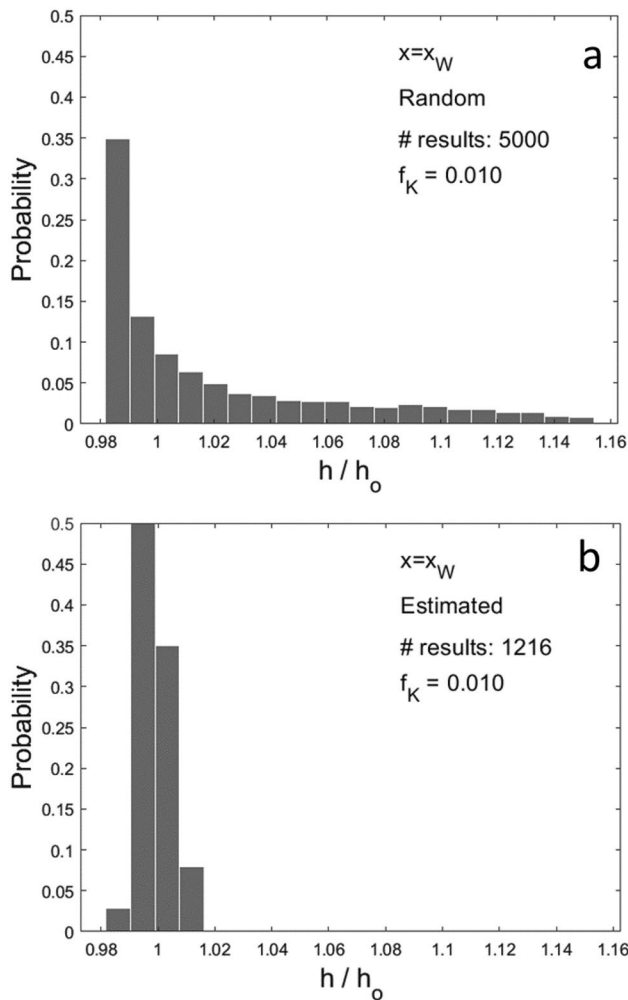


Fig. 9 Uncertainty of predicted piezometric head at the well position x_w using **a** randomly generated sets of parameters K_1 and K_2 and **b** after parameter estimation. Normalized simulated head expressed as function of the head value assigned to right boundary

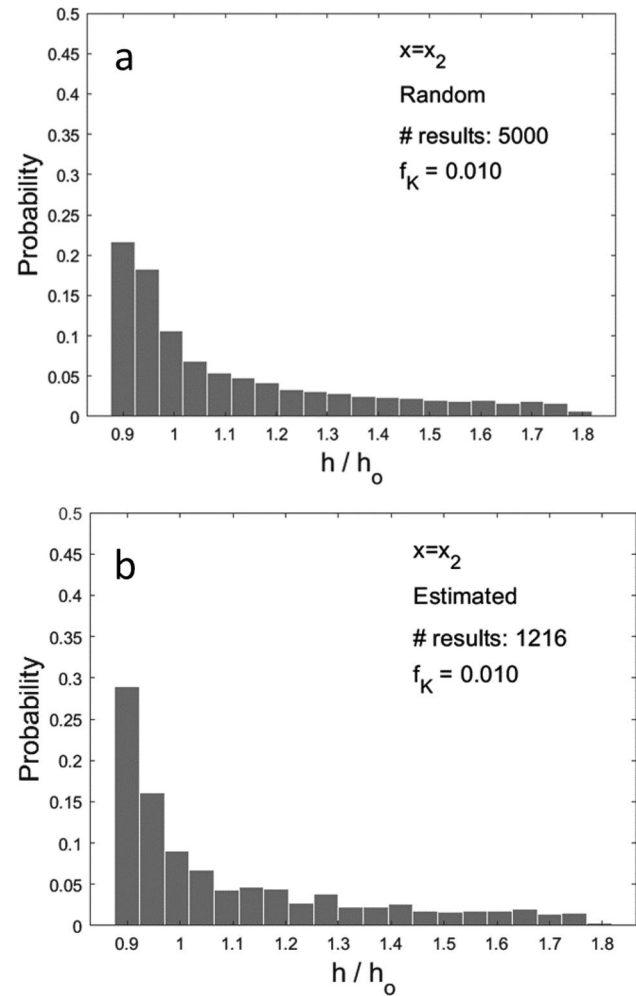


Fig. 10 Uncertainty of predicted piezometric head at position $x_2 = 0.8L$ using **a** randomly generated sets of parameters K_1 and K_2 and **b** after parameter estimation. Normalized simulated head expressed as function of the head value assigned to right boundary

c. Uncertainty due to forcing variables

The standard procedure to quantify this uncertainty is similar to the one used to quantify the uncertainty of the model due to estimated parameters: generate many equally likeable or probable series of forcing parameters, e.g. recharge series for this example, and use them to run the model and collect its results to generate probability distributions of the possible outcomes. Then, this analysis also requires defining possible ranges for forcing variables and selecting mathematical models to generate plausible time series. The rest of the analysis is similar to the one described previously (Figs. 9, 10). Some forcing variables, e.g. boundary conditions, can sometimes be considered as part of the model parameters as part of the uncertainty

analysis. However, it is better to keep them separate from the real parameters of the model that must be estimated, because, in general, the uncertainty related to their values have different origin and magnitude.

4 Conclusions

The fact that for practical applications, models always include more parameters than available observations, leads to an inverse problem that has multiple solutions. Automatic parameter estimation methods make explicit the existence of these multiple solutions and facilitate

to find them. Thereby, they help to explore the potential variability of the model results that is coherent or supported by collected observations. We demonstrated through a synthetic example inspired by a real-world setting, that adding estimates of groundwater flow as previously proposed [16] to reduce the number of potential solutions for the inverse problem, may be rather difficult and/or ineffective. Hence, it may be a better strategy to accept that the calibration of numerical models that include a large number of parameters will always lead to a large number of multiple solutions.

It is likely that the full quantification of the uncertainty associated to groundwater models will remain reserved for limited applications for some time due to the computational effort needed. The computational effort for models that include hundreds or even thousands of parameters may be prohibitively expensive even with the computational resources available today. Meanwhile, calibrated models used to make forecasts should at least include a sensitivity analysis that attempts to bound its potential results and help identifying the parameters that have greatest importance. Although this may be simple for linear models with few parameters, it may still be rather difficult for non-linear models that depend on multiple parameters with wide ranges of potential values, as it is the case of most groundwater models. Alternatively, linear uncertainty quantification, which requires little additional computational effort, can provide results that are informative for ranges of parameters for which models behave quasi-linearly, but must be analyzed considering their limitations [10, 11].

Author contributions PH: defined topic and structure of manuscript, prepared and used scripts, and prepared draft. MAM: revised draft and proposed improvements. GF: revised draft and proposed improvements. TH: revised draft and proposed improvements.

Funding Not applicable.

Availability of data and materials The scripts used to produce figures are available upon request.

Declarations

Conflict of interest The authors state that there is no conflict of interest.

Ethical approval This work has not been published nor submitted elsewhere.

Consent to participate All authors are aware of this submission.

Consent to publish All authors agree to publish this manuscript in Water Resources Management.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bredehoeft JD (2002) The water budget myth revisited: why hydrogeologists model. *Groundwater* 40:340–345. <https://doi.org/10.1111/j.1745-6584.2002.tb02511.x>
2. Zhou Y, Li W (2011) A review of regional groundwater flow modeling. *Geosci Front* 2:205–214. <https://doi.org/10.1016/j.gsf.2011.03.003>
3. Hrozencik RA, Manning DT, Suter JF, Goemans C, Bailey RT (2017) The heterogeneous impacts of groundwater management policies in the republican river basin of Colorado. *Water Resour Res* 53:10757–10778. <https://doi.org/10.1002/2017WR020927>
4. Noël PH, Cai X (2017) On the role of individuals in models of coupled human and natural systems: lessons from a case study in the Republican River Basin. *Environ Model Softw* 92:1–16. <https://doi.org/10.1016/j.envsoft.2017.02.010>
5. Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* 263:641–646. <https://doi.org/10.1126/science.263.5147.641>
6. Herrera PA, Marazuela MA, Hofmann T (2022) Parameter estimation and uncertainty analysis in hydrological modeling. *WIREs Water* 9:e1569. <https://doi.org/10.1002/wat2.1569>
7. Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ (2005) Inverse problem in hydrogeology. *Hydrogeol J* 13:206–222. <https://doi.org/10.1007/s10040-004-0404-7>
8. Vrugt JA, Stauffer PH, Wöhling T, Robinson BA, Vesselinov VV (2008) Inverse modeling of subsurface flow and transport properties: a review with new developments. *Vadose Zo J* 7:843–864. <https://doi.org/10.2136/vzj2007.0078>
9. Zhou H, Gómez-Hernández JJ, Li L (2014) Inverse methods in hydrogeology: evolution and recent trends. *Adv Water Resour* 63:22–37. <https://doi.org/10.1016/j.advwatres.2013.10.014>
10. Doherty J (2015) Calibration and uncertainty analysis for complex environmental models. *Groundwater* 56:673–674
11. White JT, Hunt RJ, Fienen MN, Doherty JE (2020) Approaches to highly parameterized inversion: PEST++ Version 5, a software suite for parameter estimation, uncertainty analysis, management optimization and sensitivity analysis. Reston, VA
12. Hunt RJ, Doherty J, Tonkin MJ (2007) Are models too simple? Arguments for increased parameterization. *Ground Water* 45:254–262. <https://doi.org/10.1111/j.1745-6584.2007.00316.x>
13. Doherty JE, Hunt RJ (2010) Approaches to highly parameterized inversion-A guide to using PEST for groundwater-model calibration
14. Mirus BB, Nimmo JR (2013) Balancing practicality and hydrologic realism: a parsimonious approach for simulating rapid groundwater recharge via unsaturated-zone preferential flow. *Water Resour Res* 49:1458–1465. <https://doi.org/10.1002/wrcr.20141>

15. Fatichi S, Vivoni ER, Ogden FL, Ivanov VY, Mirus B, Gochis D, Downer CW, Camporese M, Davison JH, Ebel B, Jones N, Kim J, Mascaro G, Niswonger R, Restrepo P, Rigon R, Shen C, Sulis M, Tarboton D (2016) An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *J Hydrol* 537:45–60. <https://doi.org/10.1016/j.jhydrol.2016.03.026>
16. Hunt RJ, Fienen MN, White JT (2020) Revisiting “an exercise in groundwater model calibration and prediction” after 30 years: insights and new directions. *Groundwater* 58:168–182. <https://doi.org/10.1111/gwat.12907>
17. Hill MC (2006) The practical use of simplicity in developing ground water models
18. Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LTH, Norton JP, Perrin C, Pierce SA, Robson B, Seppelt R, Voinov AA, Fath BD, Andreassian V (2013) Characterising performance of environmental models. *Environ Model Softw* 40:1–20. <https://doi.org/10.1016/j.envsoft.2012.09.011>
19. Hill MC (1998) Methods and guidelines for effective model calibration; with application to UCODE, a computer code for universal inverse modeling, and MODFLOW, a computer code for inverse modeling with MODFLOW
20. Poeter EP, Hill MC (1999) UCODE, a computer code for universal inverse modeling 1 Code available at http://water.usgs.gov/software/ground_water.html1. *Comput Geosci* 25:457–462. [https://doi.org/10.1016/S0098-3004\(98\)00149-6](https://doi.org/10.1016/S0098-3004(98)00149-6)
21. Zimmerman DA, de Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, Bras RL, Carrera J, Dagan G, Davies PB, Gallegos DP, Galli A, Gómez-Hernández J, Grindrod P, Gutjahr AL, Kitanidis PK, Lavenue AM, McLaughlin D, Neuman SP, RamaRao BS, Ravenne C, Rubin Y (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resour Res* 34:1373–1413. <https://doi.org/10.1029/98WR00003>
22. Anderson MP, Woessner WW, Hunt RJ (2015) Chapter 9: model calibration: assessing performance. In: Anderson MP, Woessner WW, Hunt RJ (eds) *Applied groundwater modeling*, 2nd edn. Academic Press, San Diego, pp 375–441
23. Freyberg DL (1988) An exercise in ground-water model calibration and prediction. *Groundwater* 26:350–360. <https://doi.org/10.1111/j.1745-6584.1988.tb00399.x>
24. Bichler A, Muellegger C, Brünjes R, Hofmann T (2016) Quantification of river water infiltration in shallow aquifers using acesulfame and anthropogenic gadolinium. *Hydrol Process* 30:1742–1756. <https://doi.org/10.1002/hyp.10735>
25. Brünjes R, Bichler A, Hoehn P, Lange FT, Brauch H-J, Hofmann T (2016) Anthropogenic gadolinium as a transient tracer for investigating river bank filtration. *Sci Total Environ* 571:1432–1440. <https://doi.org/10.1016/j.scitotenv.2016.06.105>
26. Haitjema H (2006) The role of hand calculations in ground water flow modeling. *Groundwater* 44:786–791. <https://doi.org/10.1111/j.1745-6584.2006.00189.x>
27. Sudicky EA (1986) A natural gradient experiment on solute transport in a sand aquifer: spatial variability of hydraulic conductivity and its role in the dispersion process. *Water Resour Res* 22:2069–2082. <https://doi.org/10.1029/WR022i013p02069>
28. Bear J (1979) *Hydraulics of groundwater*. McGraw-Hill, New York
29. Strang G (2019) *Introduction to linear algebra*. Wellesley-Cambridge Press
30. Aster RC, Borchers B, Thurber CH (2013) Chapter one: introduction. In: Aster RC, Borchers B, Thurber CH (eds) *Parameter estimation and inverse problems*, 2nd edn. Academic Press, Boston, pp 1–23
31. Hunt RJ, White JT, Duncan LL, Haugh CJ, Doherty J (2021) Evaluating lower computational burden approaches for calibration of large environmental models. *Ground Water* 59:788–798. <https://doi.org/10.1111/gwat.13106>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.