Check for updates

Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR031458

An Approach to Include Hydrogeologic Barriers With Unknown Geometric Properties in Groundwater Model Inversions

S. K. Marshall[1,2], P. G. Cook[1], C. T. Simmons[1], L. F. Konikow[3], and S. Dogramaci[4]

[1]National Centre for Groundwater Research and Training (NCGRT), College of Science and Engineering, Flinders University, Adelaide, SA, Australia, [2]CSIRO Land and Water, Glen Osmond, SA, Australia, [3]Leonard Konikow Hydrogeologist, Reston, VA, USA, [4]Rio Tinto Iron Ore, Perth, WA, Australia

Key Points:

- A method for including hydrogeological barriers in model inversion is developed using a regularization device we call phantom structures
- Narrow, low-permeability structures such as faults and dykes are included in model inversions with no prior knowledge of their geometries
- Inverting models with phantom structures provide improved results compared to using pilot points for a wide range of conditions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. K. Marshall,
sarah.marshall@flinders.edu.au

Citation:

Marshall, S. K., Cook, P. G., Simmons, C. T., Konikow, L. F., & Dogramaci, S. (2022). An approach to include hydrogeologic barriers with unknown geometric properties in groundwater model inversions. *Water Resources Research*, *58*, e2021WR031458. https://doi.org/10.1029/2021WR031458

Received 21 OCT 2021
Accepted 24 JUN 2022

Author Contributions:

Conceptualization: S. K. Marshall, P. G. Cook, C. T. Simmons, L. F. Konikow
Funding acquisition: P. G. Cook, C. T. Simmons
Methodology: S. K. Marshall, P. G. Cook, C. T. Simmons, L. F. Konikow
Project Administration: P. G. Cook, C. T. Simmons
Supervision: P. G. Cook, L. F. Konikow
Writing – original draft: S. K. Marshall, P. G. Cook
Writing – review & editing: S. K. Marshall, P. G. Cook, C. T. Simmons, L. F. Konikow

© 2022. American Geophysical Union. All Rights Reserved.

Abstract Hydrogeological barriers can significantly impact groundwater model predictions. They are, however, often excluded from, or misrepresented in, groundwater models if their presence is unknown or their properties are poorly constrained. Here we show that sharp barriers can be included in groundwater model inversion, even where their presence is uncertain. We describe an approach utilizing "phantom structures"—randomly located, linear groups of model cells assigned a unique hydraulic conductivity value—to improve identifiability of barriers. Algorithmic parameter estimation using model-independent parameter estimation code (PEST) is implemented to determine model structures that best match the hydraulic head and groundwater age observation data from a hypothetical aquifer. Our results show that for a series of case studies, this method was successful in inferring the appropriate location and properties of hydrogeological barriers, when that barrier was not aligned with the dominant flow direction. We compare these results to model inversion using traditional pilot points. The phantom structures approach shows promise in identifying hydrogeological structures and in reproducing groundwater flow across a model domain. Our results demonstrate that the geometric properties of geological structures can remain flexible in a model inversion. This is a step toward reducing conceptual model uncertainty where the presence and properties of hydrogeologic barriers are undefined.

Plain Language Summary Understanding the geological characteristics of a region can be very challenging. This is particularly true if the region contains narrow, linear structures known as faults and dykes. However, these structures can be important when it comes to defining the flow of groundwater. Many faults and dykes obstruct groundwater flow, so may inhibit groundwater availability to rivers and ecosystems or to other groundwater users. In this paper, a methodology is introduced that uses commonly collected hydrogeological datasets (known as hydraulic head and groundwater age) to discover the location, shape and physical properties of certain geological structures in the subsurface. The method relies on "inverse approaches", where data observed in the field is compared to a modeled equivalent. This method of incorporating narrow, linear structures in groundwater models results in a better match between field and modeled data when compared to another approach that relies on representing faults and dykes as broader, flatter, smoother geological anomalies.

1. Introduction

Numerical model development to predict or describe groundwater flow and transport is based on a practitioner's understanding of an aquifer system, known as its conceptualization. Yet a large amount of uncertainty can exist in groundwater system understanding related to physical processes and the subsurface geological framework, amongst other things (e.g., Enemark et al., 2019). Model inversion, such as history matching (or model calibration), is one procedure whereby expert knowledge and direct measurements can be combined to estimate groundwater system parameters. The most common parameter estimated is hydraulic conductivity, *K*. Many examples demonstrate the use of hydraulic head and groundwater age for the estimation of aquifer properties (Knowling et al., 2019; Konikow et al., 2008; Leray et al., 2012; Sanford, 2011; Zell et al., 2018). Embedding knowledge about aquifer structures helps to ensure that an inversion is well-posed but can also misrepresent salient aquifer features (Carrera & Neuman, 1986). Uncertainty in model structures can result in a "perfectly calibrated" model that remains incorrect with respect to the prediction of interest. This may be due to overfitting, where the model domain has an unrealistically high spatial variability yet is unable to capture the true aquifer geometry (Moore & Doherty, 2005).

Inverse methods can be used to estimate aquifer structures where no prior knowledge of their geometric properties exist. The use of pilot points—assigning hydraulic properties to specific positions within the model domain rather than directly to the whole model grid, allows for flexibility in determining the location and shape of model parameters (Doherty, 2003). However, this requires a form of spatial averaging, resulting in a smoothed parameter field, which may not be suitable where sharp changes in hydraulic properties are known to exist. The use of training images has also shown to be successful in understanding changes between high and low permeability units (Chen & Rubin, 2003; McCallum, Herckenrath, & Simmons, 2014). In these examples, many conditioning points, or considerable knowledge of existing aquifer properties, are required at the outset to interpolate other aquifer structures. However, there may be just one (or a few) relatively narrow low permeability barriers in an otherwise higher-permeability aquifer. These are often difficult to detect, due to limited data intersecting and directly measuring their properties.

Across faults and dykes, sharp changes in K can occur between a higher-permeability host rock and lower-permeability barriers. These structures may also form conduits to groundwater flow or complex conduit-barrier systems (Caine et al., 1996). Geologic structures that act as hydrogeologic barriers can form important controls on groundwater flow (Bense et al., 2003; Cilona et al., 2015), chemistry (Gumm et al., 2016; Mayer et al., 2007), groundwater age (Castro & Goblet, 2005; Marshall et al., 2020), discharge processes (Bense & Kooi, 2004; Gleeson & Novakowski, 2009), and drawdown (Hadley et al., 2019; Marshall et al., 2019). Locating and characterizing hydrogeologic barriers can be difficult and costly. Wells intersecting or surrounding them are uncommon and challenges exist in scaling local-scale structural data to a regional area (Bense et al., 2013).

Model inversions have been used to estimate the properties of hydrogeological barriers (Comte et al., 2017; Hadley et al., 2019; Marler & Ge, 2003; Mayer et al., 2007; Ochoa-González et al., 2015). In all of these examples, however, the locations and geometric properties of structures (such as width, strike and dip) were assumed to be known prior to the inversion. Advances have been made to develop methods for including structures where their locations and geometric properties are not known in advance, for example, using non-linear model inversion (Fienen, Hunt, et al., 2009; Harp et al., 2008; Le Goc et al., 2010), shape-based inversions (Kilmer et al., 2003), and level set inversion methods (Cardiff & Kitanidis, 2009; Chou et al., 2016). Knowledge of transmissive zones have also been derived from inverse methods utilizing pumping tests (Tiedeman et al., 1995) or hydraulic tomography data (Illman et al., 2009; Somogyvári et al., 2017). Few methodologies, however, have been demonstrated in regional-scale (i.e., tens of kilometers or more) systems where the geometric properties of low-permeability structures are estimated based on little prior knowledge and potentially sparse datasets.

In this study, our objective is to assess the importance of, and potential to include, hydrogeological barriers in a regional-scale groundwater model where their presence is suspected but unknown. A method is proposed that builds on previous inverse approaches but with the addition of thin near-vertical low-permeability structures, such as faults and dykes, without knowing their physical or geometric properties in advance. These structures are identified using hydraulic head and groundwater age data that are stationary in time. The aim is to demonstrate the application of existing regional datasets rather than a data set explicitly designed to identify these features.

The target application of this approach is to better understand aquifers with structural complexity, such as fault zones. The method utilizes what we have labeled "phantom structures"— randomly located, linear groups of model cells that are assigned unique representative properties during model inversion. Algorithmic parameter estimation is used to determine the smallest number of these structures that produce the best match to the hydraulic head and/or groundwater age observation data. The outcome is not necessarily to reproduce one optimal depiction of the subsurface but rather to quantify the likely locations and properties of subsurface linear low-permeability structures, if they exist. Our method aims to avoid conceptual model "surprise" and unforeseen consequences of incorrect model structures in the implementation of numerical models (Bredehoeft, 2005). This study utilizes tools and techniques already widely available in the groundwater industry and is intended as a prototype upon which further studies utilizing field examples can be based.

2. Modeling Methods

2.1. Hypothetical Model

The study is designed to evaluate the use of advanced groundwater modeling and parameter estimation methods to help identify linear features in an aquifer. We create a hypothetical aquifer system containing an embedded

Table 1
Model Dimensions and Universal Parameters for the Real Model

Parameter	Symbol	Value
MODFLOW-NWT		
Grid size (x and y directions) [m]	Δx, Δy	100
Layer thickness [m]	Δz	300
Specific yield	Sy	0.1
Recharge rate [m/d]	R	1.37e−5
Hydraulic conductivity—aquifer [m/d]	K_a	1
Hydraulic conductivity—barrier [m/d]	K_b	1e−3
MT3D-USGS		
Effective porosity	η_e	0.1
Zeroth order reaction rate [yrs/d]		1/365.25
Longitudinal dispersivity [m]	α_L	1.5
Transverse dispersivity [m]	α_T	0.15
Effective molecular diffusion coefficient, water [m^2/d]	D^*	9e−6

linear low-permeability feature, which provides a control. A simple hypothetical example is first tested, and this is followed with alternative barrier conceptualizations. The range of plausible barrier geometries that could be tested is unlimited and the approach can theoretically be used for any given aquifer containing one or more structures where the flow system is perturbed. We assume the properties of the barrier are unknown and use numerical methods to try to identify and characterize the suspected feature given varying assumptions about the availability of known head and age data. Here we attempt to emulate the challenges faced in a real system with some, but limited, data available.

The method is initially set up with a known location of a hydrogeological barrier, herein referred to as the *real* model. This is a simple model that includes only two zones of differing hydraulic conductivity. The hydraulic conductivity of the aquifer, K_a, is homogeneous, except for a single barrier with a value of hydraulic conductivity, K_b, three orders of magnitude lower than K_a. Barrier structures can often have complex permeability and anisotropy structures, for example, a lower permeability core surrounded by higher permeability conduits. This study does not investigate barrier-conduit features but conceptually these could be investigated using a similar process. The application of this method to more complex structure types is an opportunity for future research.

Groundwater flow was simulated using MODFLOW-NWT (Niswonger et al., 2011). The groundwater flow equation was solved with the Newton solver using a head change and residual convergence criterion of 1e−6 m. Groundwater age was calculated using MT3D-USGS (Bedekar et al., 2016) with direct age simulation (Goode, 1996). Age was simulated as a nonreactive solute species where water age increases by one unit of time per unit time. This approach allows age mixing (dispersion) by applying the advection-dispersion transport equation, which is not the case with purely advective approaches. A value of longitudinal dispersivity of 1.5 m, representing transport over approximately 100s of meters (Gelhar et al., 1992) was adopted. This, as well as other flow and transport parameters (Table 1) were held constant. However, actual parameter values are not highly relevant as the purpose is to demonstrate a new approach for model inversion. A range of dispersivity coefficients were tested and showed that the calculation of age distributions is relatively insensitive to the dispersivity values except for longitudinal dispersivity greater than 200 m. This is because, unlike in solute-plume problems, age has a "source" everywhere and increases consistently across the domain with no effect on the age gradient. A standard finite-difference method was used to solve the transport equation, with a convergence criterion of 1e−5 in terms of relative concentration (age).

Dimensions of the real model are 20 km by 10 km with a cell size of 100 × 100 m (Figure 1). It is a one-layer model and grid size was held constant between the forward and inverse simulations. A 2D rather than a 3D model was used for simplicity where structures are represented as vertical features. The left and right model sides both have general head boundaries, and a regional head difference of 100 m across the domain was imposed (an

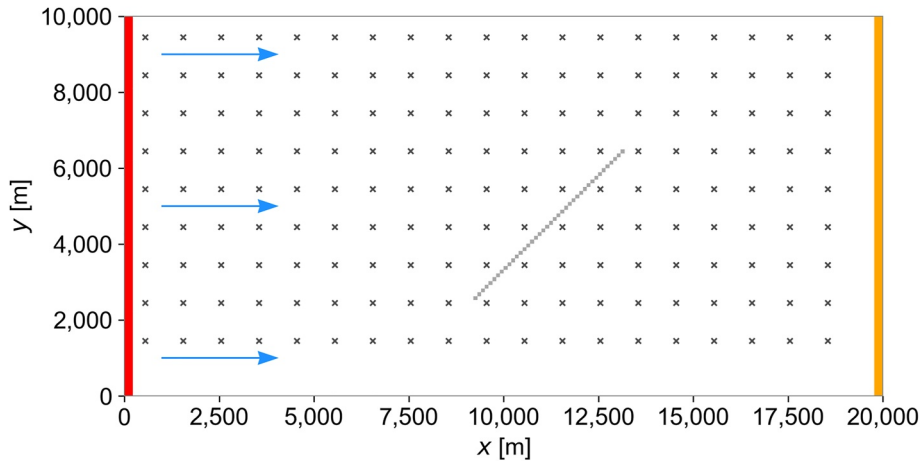

Figure 1. Schematic, 2D plan-view representation of the simulation domain showing the dimensions and boundary conditions of the real model. The slanted gray line shows the location of the low permeability barrier (where K_b = 1e−3 m/d) and the white background is the remaining aquifer (where K_a = 1 m/d). The red and orange lines show general head boundaries and *gray crosses* represent the position of the pilot points for the inversions without phantom structures. Top and bottom boundaries are no-flow. The *blue arrows* show the principal direction of groundwater flow. Recharge is added uniformly to the surface of the model.

overall hydraulic gradient of 0.005 m/m). The two lateral sides and base of the model are no-flow boundaries and the upper surface is simulated as a free water-table boundary. Recharge is applied uniformly across the aquifer at a rate of 1.367e−5 m/d (5 mm/yr). This synthetic example was constructed to represent plausible real-world examples in arid climate regions and at the same time allow for rigorous examination of the robustness of this inversion technique with respect to the known aquifer structure. It is based on aquifer systems in the arid Pilbara region of Western Australia, as this work forms part of a larger project assessing groundwater resources in this region (Cook et al., 2016; Underwood et al., 2018). It is not, however, intended to directly model this system and is purely hypothetical with no calibration to field data undertaken.

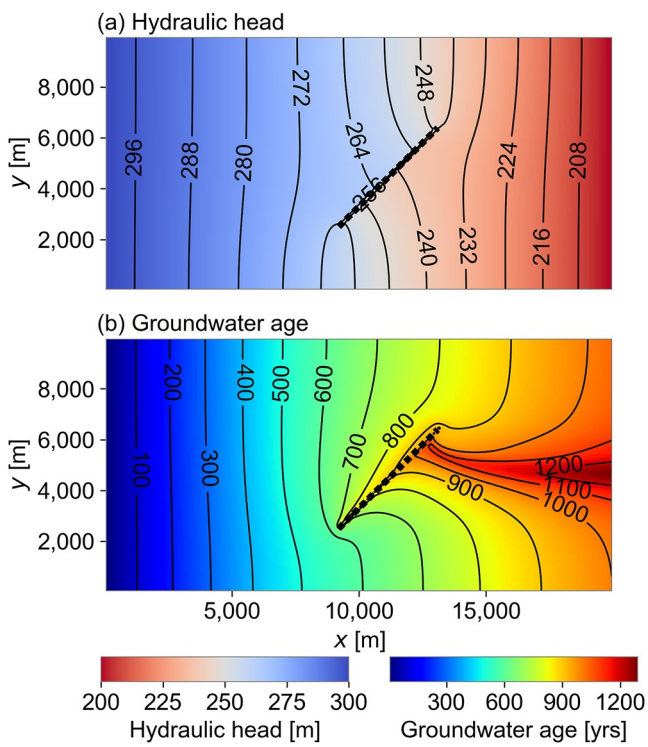

Figure 2. Model-calculated distributions under steady-state conditions of (a) hydraulic head, and (b) groundwater age for the real model; hydraulic barrier location denoted by *dashed black line*.

The hydraulic conductivity of the aquifer, K_a, is 1 m/d in all cells except those at the location of the barrier. Barrier cells have a hydraulic conductivity of 1e−3 m/d and are located in the central part of the aquifer (Figure 1). The barrier location was randomly assigned and its length and width are again based on what is known of barrier characteristics in the Pilbara field region. The length of the barrier is 5,515 m and it is one cell width wide (100 m). The models were set up and post-processed using the FloPy package for Python (Bakker et al., 2016) and all scripts are publicly available. Both the flow and transport models were executed as steady-state simulations.

2.2. Hydraulic Head and Groundwater Age Data

Results of running the model with parameters listed in Table 1 show that hydraulic heads gradually decrease in the direction of groundwater flow except across the barrier, where it sharply declines (Figure 2a). Groundwater age increases gradually in the direction of groundwater flow except downstream of the barrier, where a tail of older groundwater extends from the barrier toward the downstream end of the aquifer (Figure 2b), these results are similar to those observed in Marshall et al. (2020).

Randomly selected points (cells) across the model grid are designated as *observation wells* and calculated hydraulic head and groundwater age data at these locations in the real model are used as observation data to estimate hydraulic conductivity values in all inversions. Observation wells were

randomly sampled within the grid, excluding cells where phantom structures are present, using the random package in Python. No measurement error was added to the model-produced "observation data" in most cases, except for a single example where error was included in one inversion as a comparison (Section 3.1.2). An inversion without age data and only hydraulic head data was also conducted. Inversions were completed with 500, 250, 100, 50, 25, 10, or 5 observation wells, equating to 2.5, 1.25, 0.5, 0.25, 0.125, 0.05 and 0.025 wells per km^2. Well density across regional areas varies and in the Pilbara region of Western Australia, mining areas have a density of approximately 1 well per km^2 and prospective sites have approximately 1 well per 3 km^2 (Rio Tinto, pers comms 2022). Steady-state hydraulic head and groundwater age data are used to emulate systems where time-series water level data and aquifer testing are not commonly available.

For each observation well density, three different configurations of observation wells (OW1, OW2 and OW3) were tested to ensure that the density of wells was tested, rather than the specific placement of particular wells. Head distributions in these wells help to show when a barrier may possibly be identified via visual inspection only (Figure 3). Groundwater ages at observation wells can be found in Supporting Information S1. For each well configuration, initially the largest number of wells (500) was randomly allocated and each lower well density is a subset of that original well distribution. The number of wells was varied for numerical experimentation and sensitivity purposes, and we do not imply that as many as 500 wells are expected or needed in a real field analysis.

2.3. Phantom Structures

Phantom structures are introduced in this study as groups of model cells assigned a single fitting parameter in the inversion, in this case barrier hydraulic conductivity (K_b). Their name stems from the fact that they are a representation of something abstract or ideal in that they can be "seen" in the inversion, when they are in fact not (necessarily) real. Phantom structures are a spatial parameterization device to help guide the solution toward convergence, which is similar to the use of pilot points. However, unlike pilot points, they are not intended to be the basis for spatial interpolation.

One hundred phantom structures were included in each inversion. The location, orientation and length of each structure were randomly assigned. For most examples shown, broad prior constraints were placed on the characteristics for each phantom structure. They had to occur on the grid with a length between 100 and 10,000 m. Orientation was assigned randomly between 0 and 180°. In one example, the results from one inversion were used to refine the prior constraints on the phantom structure configuration for a subsequent inversion (Supporting Information S1). In general, there is a slight skew toward including barriers of a shorter length and to the right-hand side of the domain, with more in the 100–1,000 m bracket than the 9,000–10,000 m length bracket for example, This is because the starting location and length were assigned randomly, and where structures extended beyond the model domain, they were cut off (and the length re-calculated based on the final phantom structure within the aquifer).

Tests were conducted for including fewer or more phantom structures. Ideally, as many as possible would be included—this increases the possibility that one phantom structure would occur close to the position of the real structure. However, as these structures are fixed, not every possible plausible structure is tested, so the chance that a phantom structure exists that completely matches the field structure is unlikely. Instead, three configurations of phantom structures, herein referred to as Phantom Structures 1, 2 and 3 (PS1, PS2 and PS3), were tested for each observation well configuration (Figure 4) to improve the likelihood that a phantom structure with similar geometric properties as the real structure occurred. Multiple configurations assist in identifying the location and properties of structures likely to exist. Local-scale field studies, such as geophysics or aquifer testing, would be advised to gain confidence in the physical properties of barriers and these could be used as prior information to better inform the distribution of phantom structures.

2.4. Model Inversion

We used 100 phantom structure parameters (K_b) and one parameter for the remaining aquifer (K_a). The model was calibrated to the hydraulic head and groundwater age observation data using the model-independent parameter estimation code (PEST) (Doherty, 2010) and executed using pyEMU (White et al., 2016). Within PEST, the Gauss-Levenberg-Marquardt algorithm is used to iteratively solve a measurement objective function ϕ_m, calculated from the weighted sum of squares of the residuals between observed (measured) and modeled data:

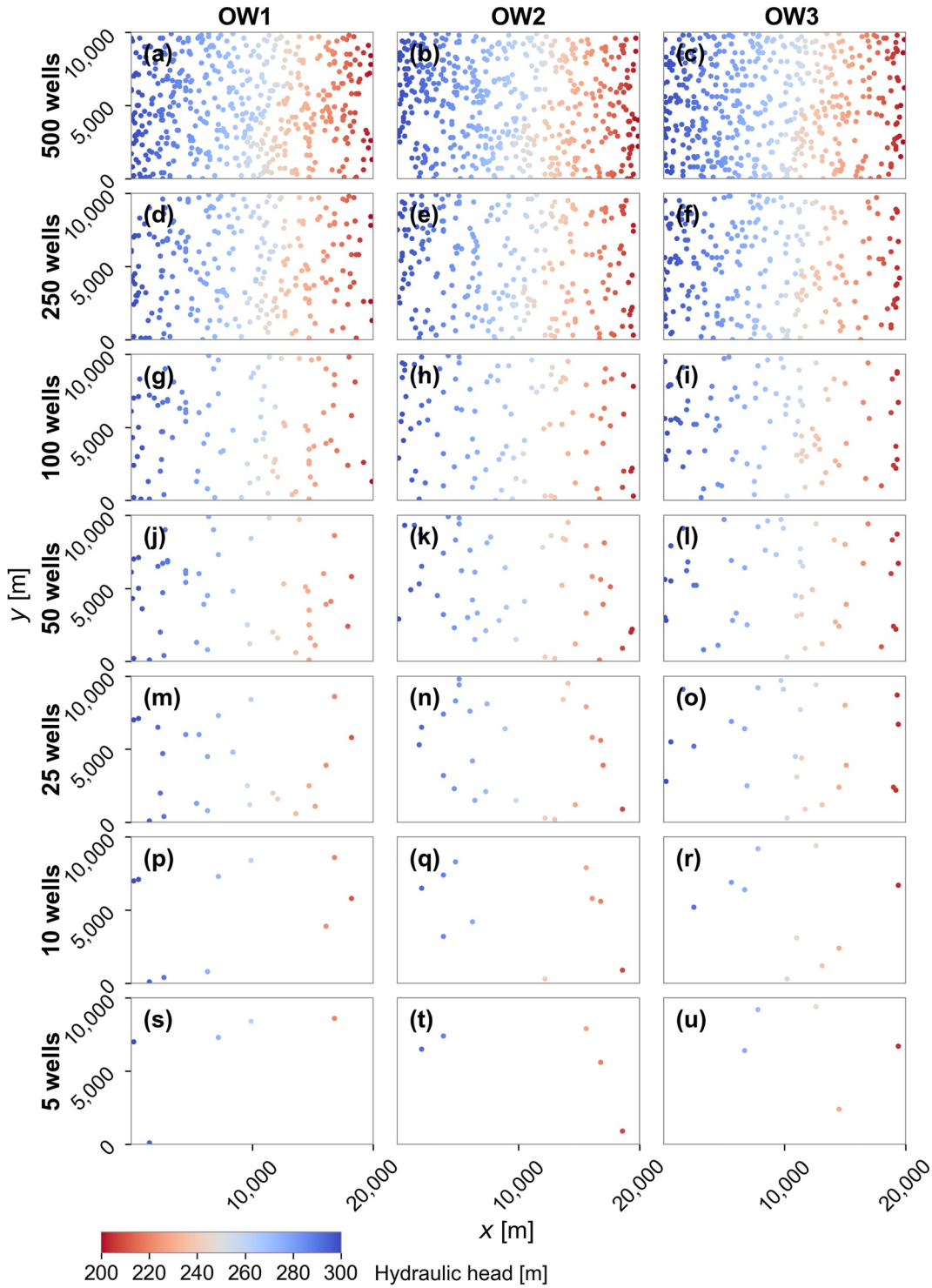

Figure 3. Observation well locations, shown by the *colored markers*, for the three different observation well configurations, OW1, OW2 and OW3, represented each by a column. The marker color corresponds to their observed value of hydraulic head from the real model. Each row represents a different observation well density.

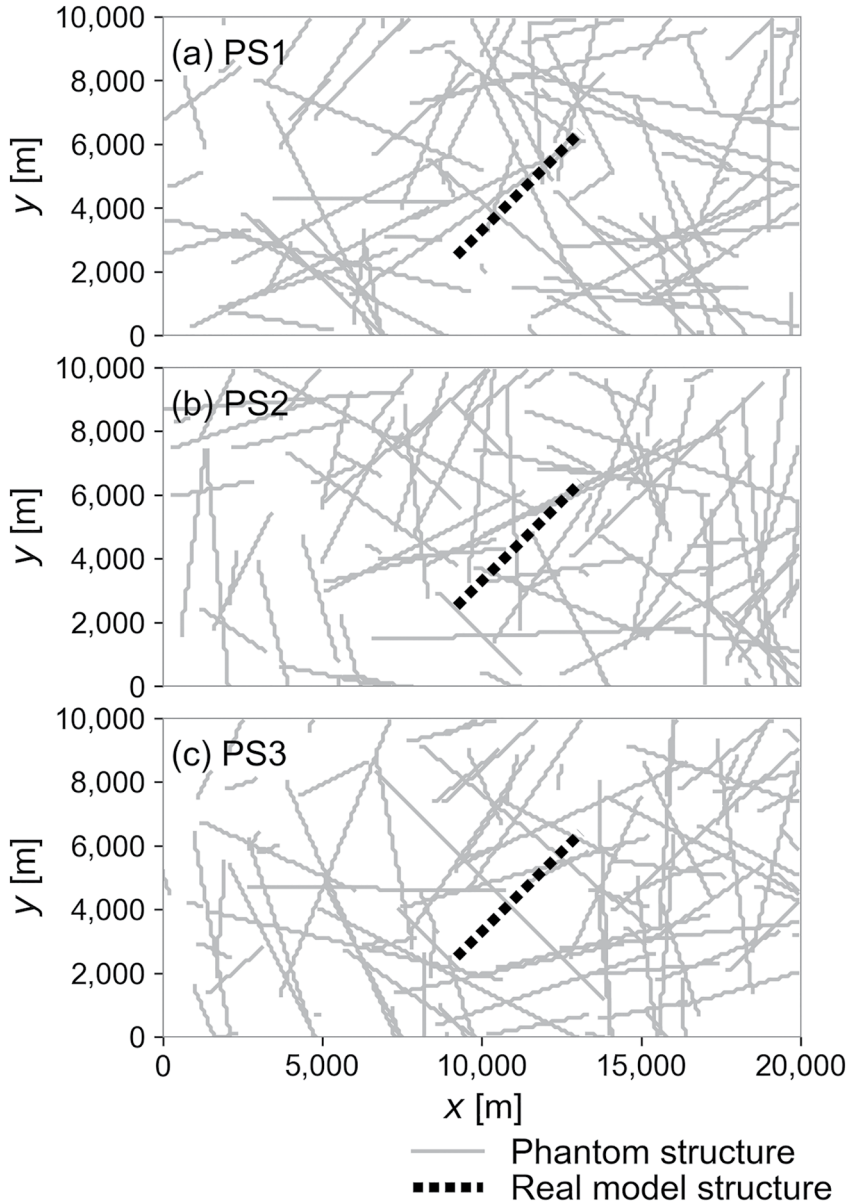

Figure 4. Locations of phantom structures represented by *solid gray lines* for each phantom structure configuration, (a) PS1, (b) PS2, and (c) PS3. The real structure is shown as a *dashed black line*.

$$\phi_m = \sum_{i=1}^{NP} \left[w_{\mathrm{hi}} \left(\hat{h}_i - h_i\right)\right]^2 + \sum_{i=1}^{NP} \left[w_{\mathrm{ai}} \left(\hat{a}_i - a_i\right)\right]^2, \quad (1)$$

where h and a are observed values (hydraulic head and groundwater age, respectively), $\hat{h}$ and $\hat{a}$ are the equivalent modeled values, NP is the number of parameters, and w_{h} and w_{a} are the weights applied to the hydraulic head and groundwater age measurements, respectively. The objective is to minimize ϕ_{m} to achieve an acceptable level of model-to-measurement fit.

Hydraulic head and groundwater age data were simultaneously included in a single measurement objective function. The age data have a larger misfit in the inversion when compared to the head data, so weights of $w_{\mathrm{h}} = 25$

and $w_a = 1$ were used so that the two datasets had approximately the same contribution to the total objective function (Doherty & Hunt, 2010). The limits for the posterior values of hydraulic conductivity were set between 1e−10 m/d and 1 m/d. That is, the structures were only permitted to form barriers to groundwater flow and not to be conduits.

For each inversion, PEST was run in regularization mode. This means that a total objective function, ϕ_t, was minimized. This includes a term for the regularization objective function, ϕ_r, as well as for ϕ_m:

$$\phi_t = \phi_m + \mu^2 \phi_r, \tag{2}$$

where μ is a regularization weighting factor determined by PEST in the inversion. In this study, Tikhonov regularization was employed, which allows for prior information (i.e., expert knowledge) to be included as a constraint to the inversion (Doherty, 2015; Hill & Tiedeman, 2007). To achieve this, a regularization objective function that penalizes any difference between K_a and K_b was applied to every cell that constitutes a phantom structure:

$$\phi_r = \sum_{i=1}^{n} w_{ri} e^{K_a - K_{bi}}, \tag{3}$$

where n is the number of phantom structures and w_r is the regularization weight used (of $w_r = 1$). We used an equation for ϕ_r that exerts a preference for parsimony. That is, phantom structures only "appear", or only have $K_b \neq K_a$, if they provide a large enough reduction in ϕ_m to offset the increase in ϕ_r.

For each configuration of phantom structures and each well configuration tested, two model inversions were run. In the first, the target measurement objective function, ϕ_m^t, was set extremely low at 1e−10, meaning that the inversion results could be considered as overfitted with little or no regularization employed. The resulting value of ϕ_m from this inversion was then multiplied by 1.1 and used as the value of ϕ_m^t with exactly the same model set-up. By using a value of ϕ_m^t 10% higher than what was achieved with little or no regularization, it helps to ensure that the algorithm is as closely constrained by observation data as possible, without overfitting. This is a compromise between reducing observation well data residuals while also accommodating regularization. Using two stages of the parameter estimation in this way was based on a similar approach used by Fienen, Muffels, and Hunt (2009). Only those for the second inversion are shown in these results.

We compared the results from the phantom faults approach to one using pilot points. In this test, with phantom structures implemented, no pilot points are used and K_a and K_{bi} are estimated. Here, ϕ_r is set up to minimize the exponential difference between the K_a and K_{bi}. In the pilot point approach, no phantom structures are used and the estimated values of K are assigned to 171 discrete points across the model domain (see Figure 1). These values are used as the basis of a kriging algorithm (performed using the PEST suite and pyEMU) to assign K to the remaining model cells. In this case, ϕ_r was set up to minimize the (linear) difference between values of K assigned to pilot points and a prior K value of 1 m/d. This regularization approach uses a prior K equal to the known value of the hypothetical aquifer as a "preferred value", where deviation from this value is penalized. Another option would be to use a geostatistical approach, where the regularization penalty stems from a variogram-weighted spatial difference between each pilot point, or a "preferred homogeneity" condition (Doherty, 2003; Fienen, Muffels, & Hunt, 2009). This approach would be less idealized because no assumed K value of the aquifer is required. However, it would also potentially result in a smoother K field that is less able to distinguish sharp features.

2.5. Sensitivity Analysis

Composite sensitivity for each head and age observation was calculated for each inversion by PEST. Composite sensitivity is a measure of the sensitivity of an observation to all parameters involved in the parameter estimation process (Doherty, 2015). This highlights observations that are sensitive to many parameters. One limitation is that it does not distinguish if there is another (or several other) observations, for example, in the same area, that also have high composite sensitivities. Recording and comparing sensitivity data for observation data is valued because it provides insight into the locations of observation wells that are most useful in estimating a hydraulic conductivity field with a barrier present.

2.6. Metrics for Evaluating Inversion Results

2.6.1. Assessing Fit Using Data From Observation Wells

As there are 63 model inversion set ups (3 phantom structure configurations and 21 observation well location scenarios), it is beneficial to have methods for comparing the results from each inversion. The first step was to analyze the model-to-measurement misfit. We use a simple measure of root mean square error (*RMSE*) calculated for the hydraulic head and groundwater age data, which is larger where the misfit is greater:

$$\mathrm{RMSE_h} = \left[\frac{1}{NO}\sum_{i=1}^{NO}\left(\hat{h}-h\right)_i^2\right]^{0.5}, \tag{4a}$$

$$\mathrm{RMSE_a} = \left[\frac{1}{NO}\sum_{i=1}^{NO}\left(\hat{a}-a\right)_i^2\right]^{0.5}, \tag{4b}$$

where NO is the number of observation wells, $\hat{h}$ and $\hat{a}$ are the posterior values of hydraulic head and direct age determined in the inversion and h and a are the values of hydraulic head and direct age extracted from the real model (representing "measured" observations). A combined error term, $RMSE_c$ was also determined:

$$\mathrm{RMSE}_c = \sum_{i=1}^{NO} w_\mathrm{h}\mathrm{RMSE}_{\mathrm{h}i} + w_\mathrm{a}\mathrm{RMSE}_{\mathrm{a}i}, \tag{5}$$

using the same values of w_h and w_a of 25 and 1 (respectively) as used in the model inversions.

2.6.2. Key Phantom Structures

We want to know if all one hundred phantom structures are required to best represent the model, or if there is a subset of one or more that are the most significant given the observation data. The first step was to distinguish phantom structures with $K_b < 0.99K_a$, in order that only structures that are notably different from the aquifer are assessed. Phantom structures with $K_b \approx K_a$ effectively disappear from the model and therefore are discarded as not useful to the results. The second step was to see which phantom structures had a high value of identifiability.

Identifiability is a measure to assess which of the parameters (i.e., which phantom structures) can be uniquely estimated given the observation data (Doherty & Hunt, 2009). It is based on the sensitivity statistics of each model output with respect to all adjustable parameters. Identifiability values for each phantom structure were calculated with a solution space cut-off of 10 singular values using the IDENPAR executable provided with the PEST software suite (Doherty, 2010). The value of identifiability for a given parameter lies between zero (complete non-identifiability based on the observation data) and one (complete identifiability based on the observation data). An arbitrary value of 0.8 was used as a cut-off to determine which phantom structures are identifiable. The phantom structures that then remained after removing those with $K_b \approx K_a$ or with identifiability <0.8 are denoted *key* structures in the inversion results and were ranked from high to low identifiability.

To isolate the smallest number of phantom structures that are most worthwhile for representing our real model, we calculated how the value of *RMSE* changed with different combinations of the key structures. First, $RMSE_h$, $RMSE_a$ and $RMSE_c$ were calculated for a model with no structures (completely homogeneous), then they were calculated for a model with the first ranked structure (that with the highest identifiability) only, then for a model with the first ranked structure and the second ranked structure, then for the first three ranked structures and so forth until all of the key structures had been evaluated. This shows how many of the structures are required to achieve a considerable drop in *RMSE*.

2.6.3. Assessing Fit Using Real Model Cell Values

Because we do know the real model in this analysis, we made a final assessment of inversion results based on comparing the real model's head and age distribution with each inversion's posterior head and age distribution. To do this, both hydraulic head and groundwater age distributions from the model results were filtered to avoid observation wells, phantom structures and model boundaries. This was done to select only cells not active in the model inversion, or which could be subject to boundary effects, and is referred to as the *filtered model* (or the

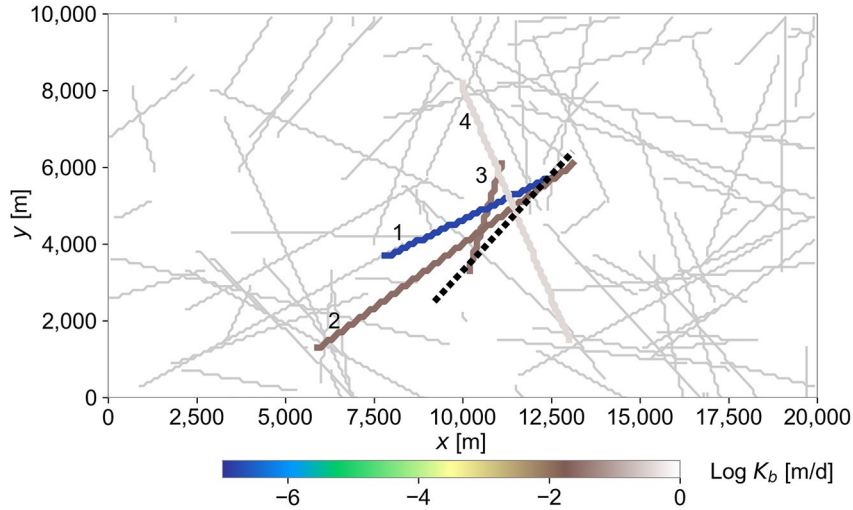

Figure 5. Results for a single inversion showing the posterior value of log K_b for each *key* phantom structure. The real structure is shown as a *dashed line*. Phantom structures not deemed as *key* are shown as *light gray lines*.

filtered cells). For all filtered cells, the modeled hydraulic head and groundwater age values were compared to those from the real model. Again, $RMSE_h$, $RMSE_a$ and $RMSE_c$ were calculated as per Equations 4a, 4b and 5 but with a value of *NO* equivalent to the number of filtered cells.

3. Results

The results are first presented for a single model to illustrate the process. We then show results for all observation well and phantom structure configurations. For the single inversion, results are shown for a case with PS1 (Figure 4a) and OW1 with 50 observation wells or 1 well per 4 km^2 (Figure 3j).

3.1. Single Inversion Example

After calibration of the data set of 50 hydraulic head and 50 groundwater age observations with 100 phantom structures, the key phantom structures were determined using the approach outlined above. The hydraulic conductivity of the four identified key structures, when compared to the real model (Figure 5) show that key phantom structures are located in the vicinity of the real barrier and display a substantially reduced K_b. Those structures that are approximately the same orientation as the real structure (Structure 1 and 2) have K_b values of 1.6e−7 m/d and 2.6e−2 m/d, whereas the real structure has a K_b of 1e−3 m/d (Table 2). Structure four is the closest in distance to the real structure (at the center point of both structures) and has a K_b value of 3.8e−1 m/d, which is the most different from the real structure. It also has an orientation almost orthogonal to the real structure. The posterior value for K_a was 0.99 m/d, within 1% of the real K_a of 1 m/d. The identifiabilities of the key structures that are close to the location of the real structure with a similar orientation have the higher values of identifiability. Structures 1, 2 and 3 all have identifiability values >0.99. Structure 4 has a marginally lower value of identifiability of 0.948. All four of the key structures intersect the real structure.

Table 2
Parameter Values for Key Phantom Structures in the Single Inversion With 50 Wells and the PS1 Configuration

Phantom structure	K_b [m/d]	Identifiability	Distance from real structure [m]	Orientation [°]	Length [m]
1	1.6e−7	1.00	1,020	66	5,239
2	2.6e−2	1.00	1,879	56	8,653
3	3.3e−2	1.00	632	18	2,941
4	3.8e−1	0.948	424	156	7,341

Note. The real barrier parameters are: K_b = 1e−3 m/d; orientation = 45°; and length = 5,515 m. Note that the order of phantom structures is based on identifiability values, which are shown to three significant figures.

The difference in hydraulic head and groundwater age distributions for a model with phantom structures and those for the real model show that over the entire model domain, the maximum difference in hydraulic head is −16 m (Figure 6a). Overall, the area of greatest hydraulic head difference is confined to a narrow zone adjacent to the location of the real barrier. The remainder of the model has been relatively well reproduced with negligible error. The maximum difference in groundwater age is 915 years in a few cells located adjacent to the real structure location. However, elsewhere the only

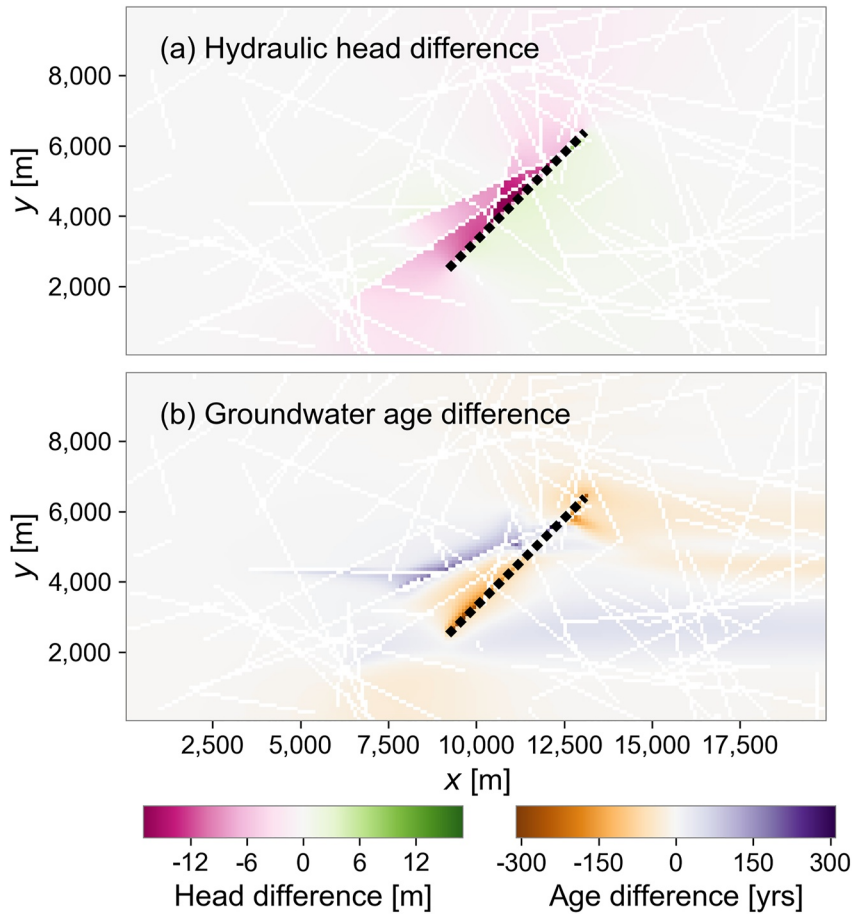

Figure 6. Difference between real model and the posterior parameterization with phantom structures in single inversion example for (a) hydraulic head, and (b) groundwater age. The difference represents posterior single example results minus known (real model) values. For groundwater age, the color scale has been capped at 310 years even though at a few cells the age difference reaches up to 915 years. The real structure is shown as a *dashed line*.

significant age difference is in the order of 50–100 years in zones both upstream and downstream the real barrier where the age gradients are relatively steep (Figure 6b). Once again, the differences are limited to small areas near the real barrier. For the remainder of the model domain the groundwater age distribution has been reproduced well, including the tail of older groundwater extending downstream of the real barrier. As the groundwater age in the vicinity of the real barrier is on the order of 700–1,000 years, a difference of 300 years (43–30%) is significant, though highly localized.

The values of $RMSE_h$, $RMSE_a$ and $RMSE_c$ all follow a similar pattern of decline with each key phantom structure added subsequently in order of identifiability (Figure 7). Initially, without any structures in the system, $RMSE_h$, $RMSE_a$ and $RMSE_c$ are equal to 2.6 m, 86.2 years, and 150.4, respectively. These values drop considerably (by 1.8 m, 51.3 years and 97.1 respectively) with the additions of Structures 1 and 2, then again with Structures 3 and 4. These drops show that Structure 1 and 2, with high values of identifiability, at close distances to the real structure, represent a good fit with the observation data. After the first four structures, no further large decrease in $RMSE_h$, $RMSE_a$ and $RMSE_c$ occurs with the addition of the other 96 phantom structures that were not key because they either have a value of K_b that approaches K_a or have very low identifiabilities (i.e., they are not sensitive to the observation data).

Evaluating sensitivity of the observation well data shows the sampling locations that are most useful for defining K_b values of the phantom structures. Sensitivity data for hydraulic head is highest for observation wells located up to 5 km downstream of the real structure (Figure 8a). The sensitivity of groundwater age observation

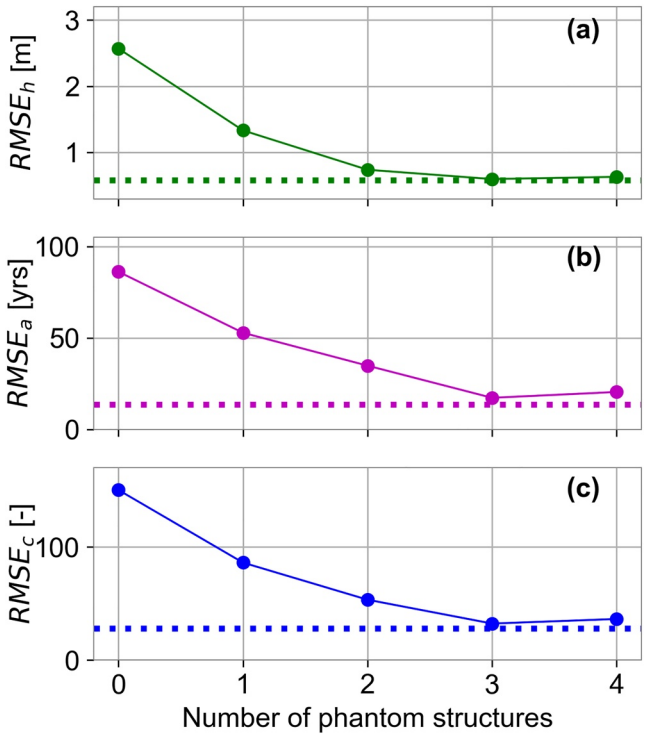

Figure 7. The root mean square error (*RMSE*) values (showing misfit at observation wells) for the single inversion example with each *key* phantom structure, added cumulatively in order of identifiability for (a) hydraulic head, (b) groundwater age, and (c) hydraulic head and groundwater age combined. The *dotted lines* show the *RMSE* value with all 100 phantom structures.

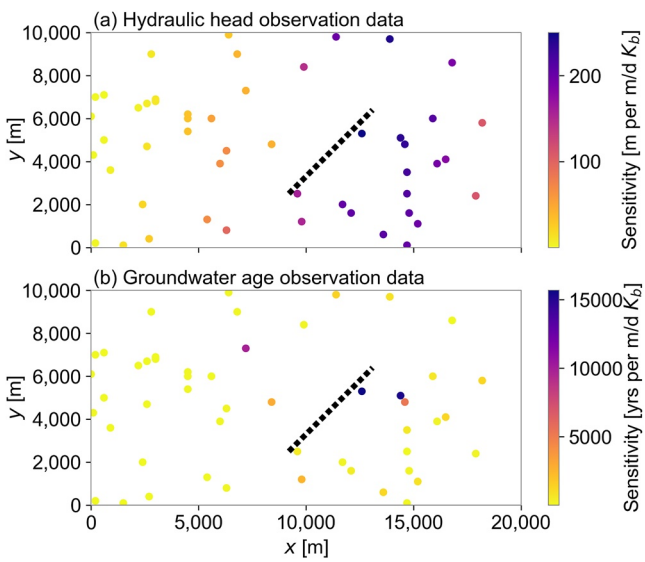

Figure 8. Composite sensitivity of observation well data for (a) hydraulic head, and (b) groundwater age for the single inversion example. The real structure is shown as a *dashed line*.

illustrates that data collected within the tail of old water extending downstream of the real structure have higher sensitivities than elsewhere in the model (Figure 8b).

3.1.1. Comparison of Head and Age Data

The inversion was also conducted with only hydraulic head data with no age data and no observation error (note there was no error in the single inversion example shown above). This helps to show whether or not the age data improves the prediction. The results without age (Figure 9a) data are not as reliable as those with age data (Figure 5). This is because an additional structure (Structure 8) has been identified, which is not as similar in location or orientation as the structures identified with age data (Table 3). Nevertheless, one of the identified structures with head data only, Structure 2, is in a similar location with a similar orientation to the real structure. The estimated K_b (2.6e−2 m/d) is also very close to that of the real structure. This shows that an inversion with head data only is still helpful in ascertaining the region and properties of the real barrier, and that the age data refines the results to be closer to the real structure (when compared to Figure 5).

3.1.2. Addition of Measurement Error

Measurement error is commonplace when collecting hydrogeologic data. Thus, the single inversion example was replicated with realistically representative observation data error. To the value observed at each observation well, an error term was added. Head error was sampled from a Gaussian distribution N(0, 0.5) and was a value in meters. Age error was also sampled from a Gaussian distribution N(0, 5) but it was a relative term in percentage change from the observed value. When error is added, the K_b values of the key structures (Figure 9b) are similar to the original model (Figure 5). Three of the same structures are identified—Structures 1, 2 and 4. Structure one is in a similar location to the real barrier, but in this case has a K_b value three orders of magnitude lower than the real structure (Table 3). Several other structures have been picked up as key to the solution but with K_b values not significantly different from the aquifer, these are structures 5, 6 and 7.

3.1.3. Alternative Structure Locations

Inversions were run for several other case studies with structures that pose challenges for identification using randomly located, steady-state, field data. This includes aquifers with:

- A barrier predominantly in line with the flow direction;
- A small barrier relative to the domain size;
- Multiple barriers; and
- A barrier that exists on the edge of the model domain.

In reality, there are an infinite number of aquifer and barrier geometries that could be tested. The case studies presented here were chosen to demonstrate some endmembers in the application of the method. In each case study, results are presented for an inversion with hydraulic head and groundwater age observation data sampled from 50 wells across the aquifer (distribution shown in Figure 3j) with PS1. All real structures have a K_b value of 1e−3 m/d as per the other single inversion examples. With the barrier aligned with the groundwater flow direction, it does not cause a significant perturbation in hydraulic head (Figure 10a) and only one key phantom structure has been identified. This structure (Figure 10b) is 4,250 m from the real structure and has a K_b value of 6.6e−2 m/d. Where the phantom structure is on the edge of

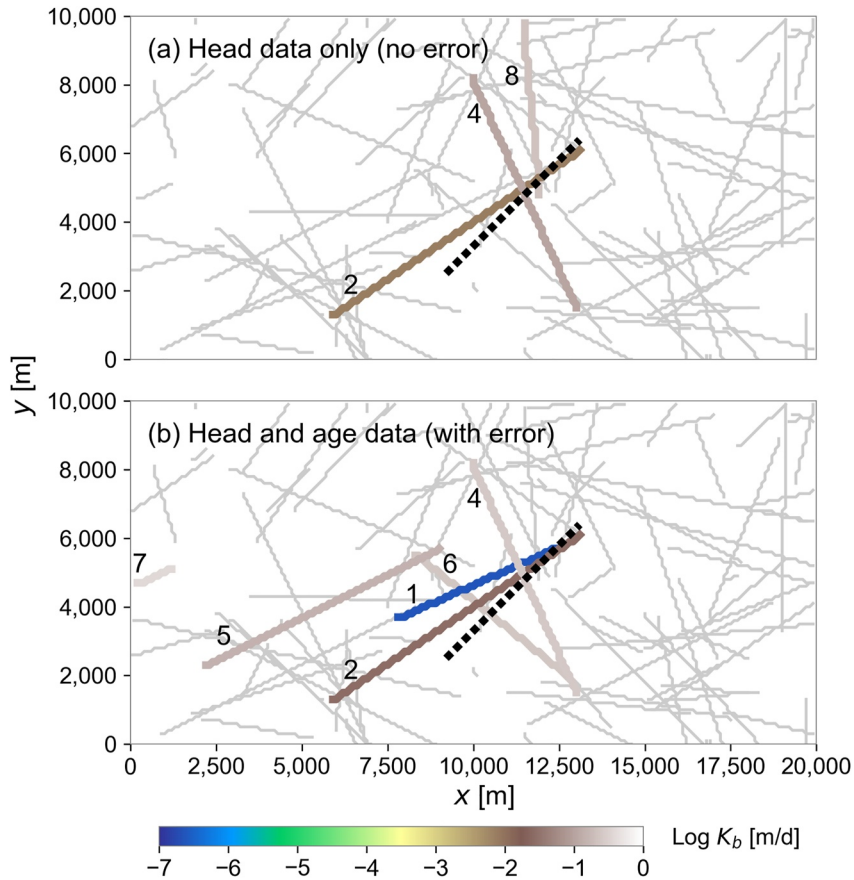

Figure 9. Results for a single inversion with (a) only hydraulic head data (no observation error), and (b) observation error added, showing the posterior value of log K_b for each *key* phantom structure. Phantom structures not deemed as *key* are shown as *light gray lines*. The real structure is shown as a *dashed line*.

Table 3
Comparison Between Parameter Values for Key Phantom Structures in the Single Inversion With 50 Wells and PS1 With: The Original Inversion That Included no Observation Error, an Inversion With Only Hydraulic Head Data and no Observation Error, and an Inversion That Included Error on Head and Age Observations

	K_b [m/d]		
Phantom structure	Original	Head data only	Including observation error
1	1.6e−2	–	2.3e−7
2	2.6e−2	7.5e−3	2.7e−2
3	3.3e−2	–	–
4	3.8e−1	1.1e−1	2.5e−1
5	–	–	1.5e−1
6	–	–	2.4e−1
7	–	–	3.8e−1
8	–	2.1e−1	–

Note. Dash value means it is not a key structure identified for that inversion.

the domain (Figure 10c), it can still be readily identified—perhaps because there is still at least one observation well on the upstream side of the barrier and several on the downstream side. In this case, two phantom structures have been identified that are both located less than 2 km from the real structure (Figure 10d). These also have similar orientation and have K_b values of 1e−1 m/d and 1.1e−3 m/d. When the barrier is approximately one third the length of the original single inversion example, it is more difficult to identify as it is less likely to have wells located in its vicinity (Figure 10e). One key structure was identified from the inversion (Figure 10f) with a location that is 1,100 m from the real structure, a similar orientation, and a K_b of 2.9e−1 m/d. In the last case, three barriers were included: two that overlap and one that is located approximately 10 km from these structures (Figure 10g). Here, four key structures were identified, two groups of two that center around the separate locations of the real barriers (Figure 10h).

3.2. Variation in Well Density

3.2.1. Key Phantom Structures

As the observation well density and phantom structure configurations vary, so does the posterior value of K_b for each phantom structure. For each observation well configuration, the K_b results with the same number of wells were averaged, and show (Figure 11) that as the number of wells increases, the key

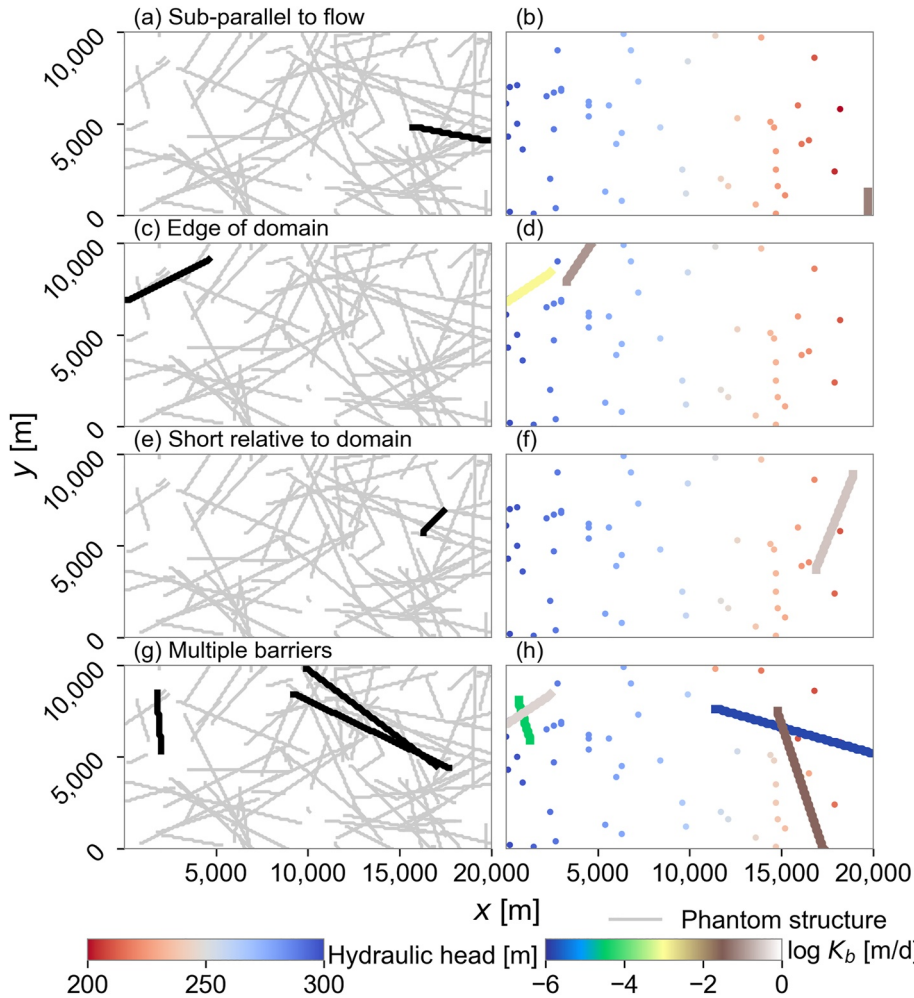

Figure 10. Alternative inversion set ups for PS1 with 50 wells showing (a, c, e and g) "real" structure locations (*black lines*) and phantom structure locations (*gray lines*); and (b, d, f and h) observation well locations (displaying hydraulic head) and posterior log K_b values for identified key structures.

phantom structures are more centrally located around the site of the real structure. This can be particularly seen for PS1 where, as the number of wells increases to 100 or more, all phantom structures are located within a short distance of the real structure (Figures 11a, 11d and 11g). This trend is similar but not as pronounced with PS2 and PS3. This is likely because the randomly allocated phantom structures did not include structures as close to the real structure as for PS1. In all cases, the K_b of the key structures are mostly less than 1e−3 m/d.

As with the single inversion example (Figure 5), we can assess how many of the key structures contribute to the fit of modeled-to-measured misfit at the observation well locations. These results (Figure 12) show $RMSE_c$ for the key phantom structures determined for each inversion then averaged (across the nine inversions) per well density. $RMSE_c$ increases as the number of wells increases, up until 100 wells. Generally, one would expect that more observation wells would result in a *lowering* of $RMSE_c$, not the opposite that is seen here. In this methodology, the location of the phantom structures is fixed. Therefore, as more observation wells are added to the inversion, it becomes difficult for PEST to find a solution that fits all of these wells because the structure is less flexible than, say, a pilot points approach. With fewer wells, there are fewer misfits to minimize and therefore it is easier to fit a K distribution for those wells only given the embedded phantom structures. Among 100, 250 and 500 wells there is no large difference in $RMSE_c$. As the number of key phantom structures increases from 0 to 1, in all cases the $RMSE_c$ decreases. This is also usually the case for addition of a second and a third phantom structure. With 5–100 wells, the change in $RMSE_c$ is more erratic and does not follow a smooth downward trend.

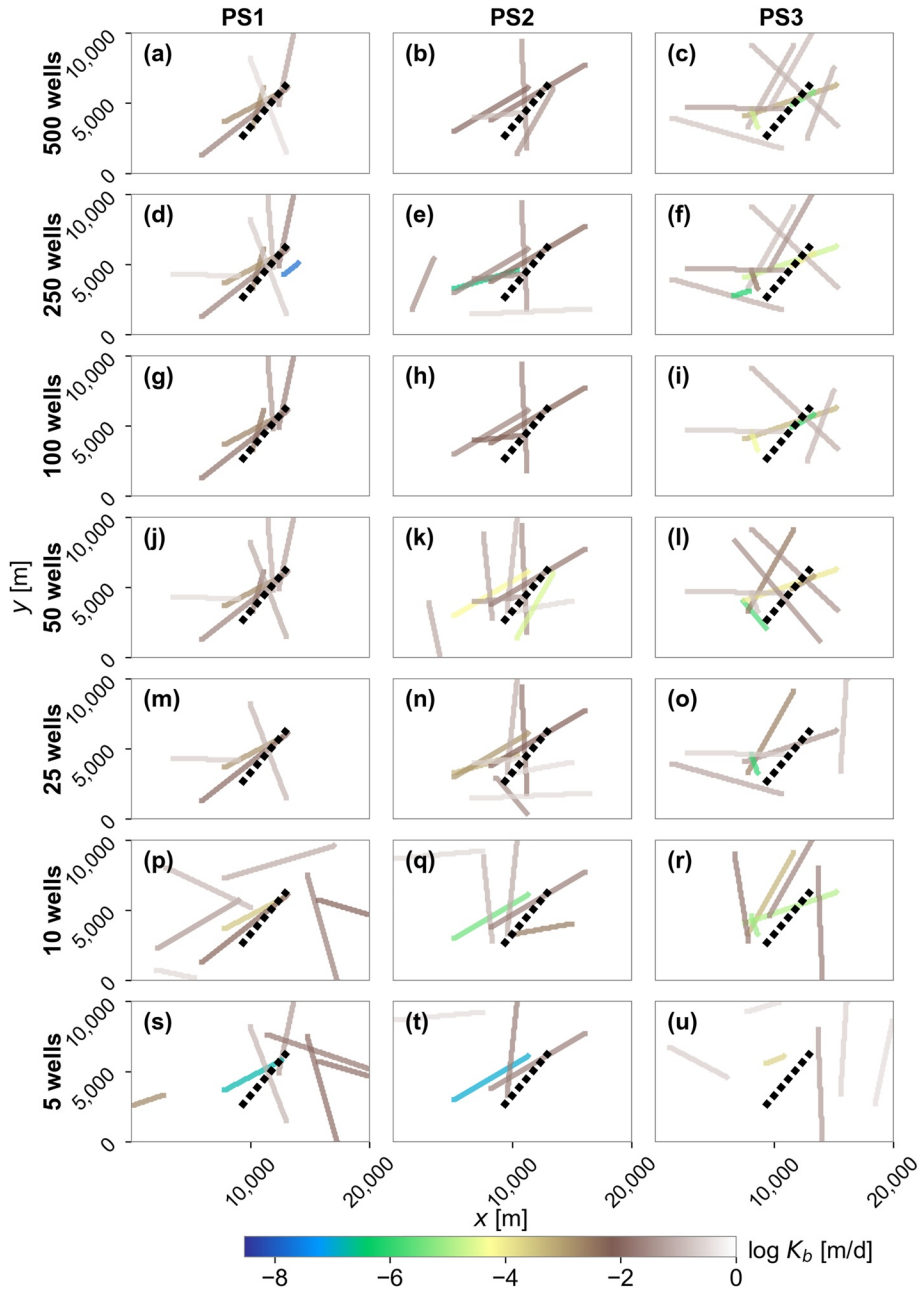

Figure 11. The mean values of K_b for each key phantom structure, where columns represent different phantom structure configurations, and rows represent different observation well densities. The mean is taken for each well configuration with the same number of observation wells. Only the *key* structures are shown.

We compared the $RMSE_c$ calculated using data from observation wells to that calculated using data from the other cells in the model, excluding the observation wells (the filtered model) (Figures 13a–13c). The $RMSE_c$ calculated with the observation wells represents the model fit that would be determined in the usual situation where the true properties of the real model are not known. This increases as the number of wells increases until 100 wells, then it does not change significantly between 100, 250 and 500 wells. $RMSE_c$ values for the filtered models (Figures 13a–13c) follow an opposite trend. As the number of observation wells increase, the $RMSE_c$ decreases. The average misfit is higher overall for the cells of the filtered model than for the observation wells themselves. This suggests that the actual distribution of hydraulic head and groundwater age across the model becomes more

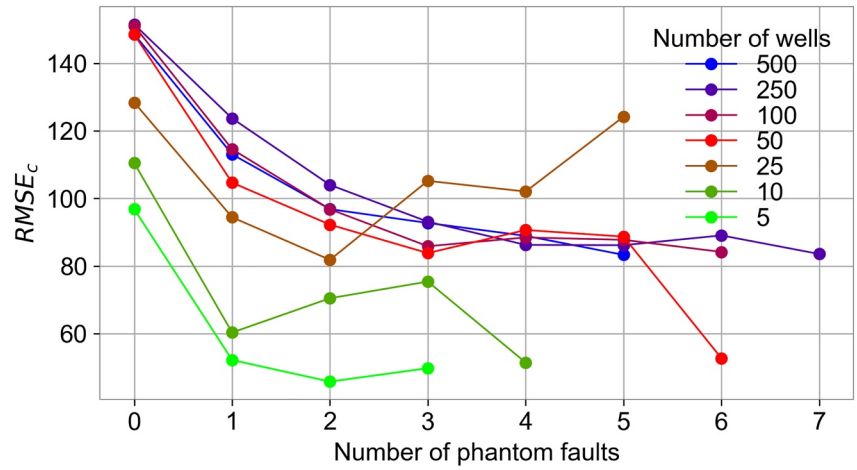

Figure 12. The $RMSE_c$ (showing misfit at observation wells) for an average of nine inversions (3 observation well configurations, 3 phantom structures) taken as a mean across the three different observation well scenarios for models with *key* phantom structures.

divergent from the real model with fewer observation wells included in the inversion. For our model, 50 wells are required for a reasonable fit across the model domain.

3.3. Pilot Points

We again compared the $RMSE_c$ at the observation wells to the filtered model for the pilot point method (Figure 13d), which does not include the use of phantom structures. The $RMSE_c$ calculated with the observation wells shows a similar trend to that of the phantom structures as it increases as the number of wells increase. For the filtered model, across most of the observation well range, the filtered model $RMSE_c$ values are higher with pilot points than with phantom structures.

The posterior K field from the pilot point inversions with 500 and 250 wells (Figures 14a and 14b, respectively), show a zone of low K approximately at the same location as the real structure with a similar trending orientation. In the 100 and 50 well cases (Figures 14c and 14d, respectively), we also visualize low K fields at approximately the same location as the real structure, but with less continuity. They consist of smaller, partially connected patches of low K rather than a single elongated low K zone. With 25 wells (Figure 14e), two small circular zones of low K are present on the upgradient side of the real structure. The cases with 10 and 5 wells (Figures 14f and 14g, respectively) show almost homogeneous K fields with a similar K to the real model K_a of 1 m/d. In all cases, and in contrast to using phantom structures, the K values are not estimated to be as low as the real structure K_b (of 1e−3 m/d).

The use of pilot points is shown to be particularly helpful for the cases with high well densities (>100 wells), where the orientation of the low-permeability structure could be ascertained from the kriged K distributions. However, when the case with 50 wells (Figure 14g) is contrasted to the single inversion example with phantom structures (Figure 5), it is clear the use of pilot points does not provide information about structure length or orientation.

4. Discussion

4.1. Utility of the Phantom Structures Approach

This study demonstrates that groundwater model inversion can be used to conceptualize systems with hydrogeological barriers where no prior knowledge of the barriers exists. However, having some prior knowledge of low-permeability structures, either from other areas of the studied aquifer or analogous aquifers, can be helpful in constraining the inversions. Innovative methods have previously been developed to explicitly represent certain model structures in aquifer conceptualizations (e.g., Harp et al., 2008; Illman et al., 2009; Koohbor et al., 2019;

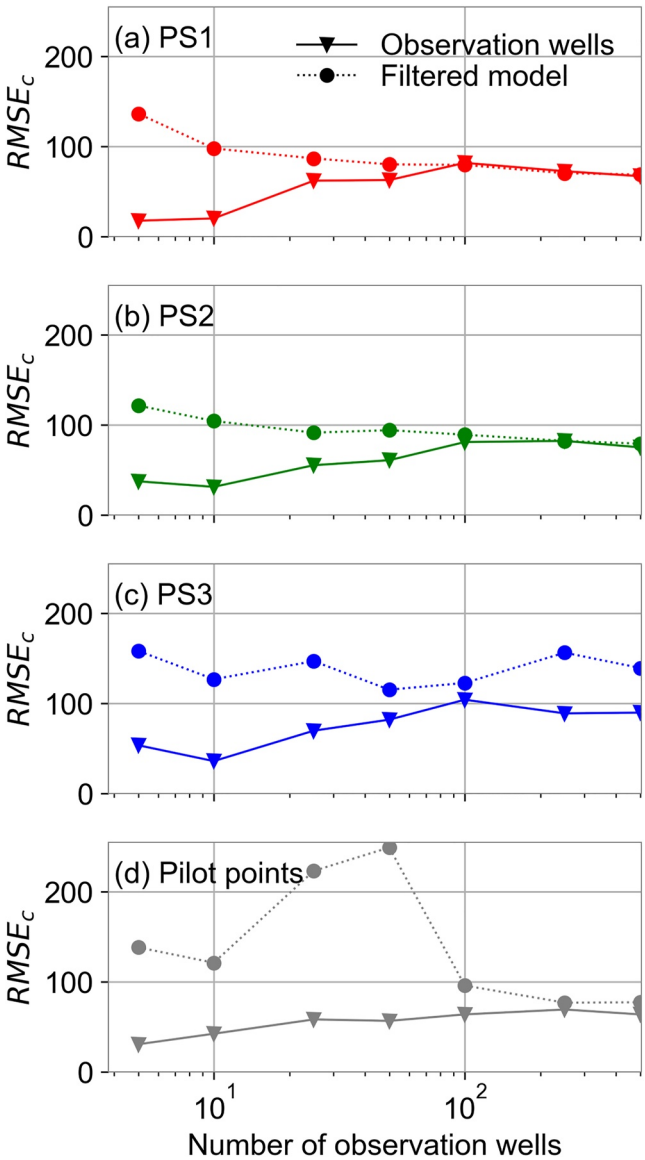

Figure 13. $RMSE_c$ values from model inversions using phantom structure configurations PS1 (a), PS2 (b) and PS3 (c), and from the model inversions using pilot points (d). $RMSE_c$ values are shown when calculated for observation wells only (*solid lines with triangle markers*) and for all cells in the filtered model that excludes observation wells (*dotted lines with circle markers*). Each *marker* represents the mean result of the three observation well configurations for that number of observation wells.

Le Goc et al., 2010; Somogyvári et al., 2017). However, few are applicable to large-scale structures such as faults and dykes, which only comprise a minor spatial extent of an aquifer. This study introduced a novel method for identifying if a barrier is present, then characterizing its geometric (i.e., location, orientation and length) and physical (i.e., K) properties. In these results we see that, for most observation well scenarios, the phantom structures located closest to the real barrier are parameterized with a low K_b similar to that of the real barrier. We compared the results with phantom structures to that with pilot points (Doherty, 2003). For the filtered model, *RMSE* is poorer with pilot points than with the phantom structures over much of the observation well range. Phantom structures result in a better representation of hydraulic head and groundwater age across the model, likely because the sharp changes in hydraulic head and groundwater age are difficult to reproduce with the smoothed hydraulic conductivity profile generated using pilot points. However, the decision to use phantom structures over an approach such as pilot points should depend on the model purpose and the impact of hydrogeological barriers on specific model predictions.

The phantom structures method is not intended to reveal the exact properties of a natural system's barrier(s). Rather, the results indicate locations and geometries where barriers are likely to exist. In the absence of further information, this provides a groundwater modeler with possible alternative hydrogeological conceptualizations and thus some assessment of possible model structural noise (Doherty & Welter, 2010). This is perhaps the most useful application of our technique—that it be run in combination with pilot point approaches to provide some information on structural uncertainty. The results also provide knowledge of the most probable locations of barriers, which can assist with more targeted investigations. An inversion with phantom structures could be followed by more direct, local-scale field studies, such as classical hydraulic testing (e.g., Hadley et al., 2019), geophysical exploration techniques (e.g., Comte et al., 2017), or hydraulic tomography (e.g., Hao et al., 2008). However, in regional aquifer systems, data collection can be logistically challenging and expensive. As such, it is useful to have as much information about the estimated location of a structure prior to conducting a well test or geophysical survey and for this reason the phantom structures approach provides considerable utility in estimating structure locations and properties.

4.2. Important Study Considerations

The degree to which this methodology will be effective depends on the nature of the hydraulic head and groundwater age perturbation caused by the hydrogeologic barrier—and indeed if this perturbation is sensed by the observation wells. There are several important study considerations for using a phantom structure inversion. This method has not been shown to be useful in detecting barrier features that are aligned with the groundwater flow direction. Other structures likely to be difficult to detect include those with a hydraulic conductivity close to that of the aquifer, that are very short, or that do not penetrate the entire vertical extent of the aquifer. However, such structures are unlikely to greatly perturb the regional flow system, and so their identification and incorporation in a regional groundwater model may not be necessary.

The density and locations of observation wells is a critical factor in the success of this inversion methodology. We show that with a density of one well per 4 km^2, a good fit is achieved between the real and estimated structure properties. However, the optimal well density depends on many factors including the barrier length, hydraulic conductivity, and other factors related to the flow field. The sensitivity of the observation well data show that

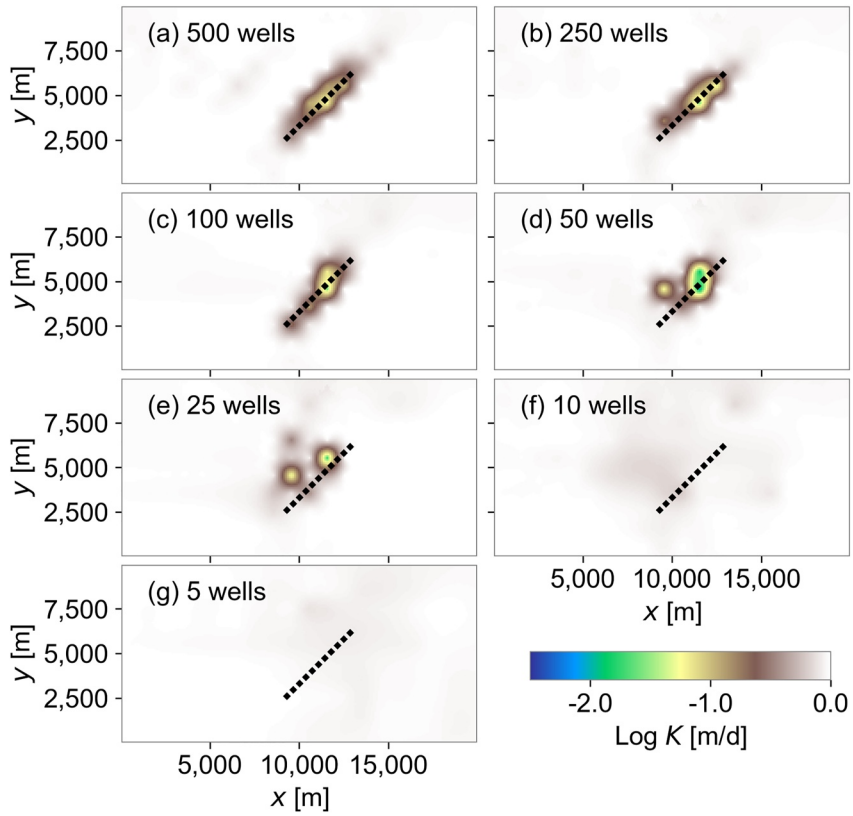

Figure 14. Inversion results for log *K* when using the pilot point approach with no phantom structures and (a) 500, (b) 250, (c) 100, (d) 50, (e) 25, (f) 10, and (g) 5 wells. Results shown are the mean results of the three different well configurations for that number of observation wells.

it is useful for hydraulic head data to be collected downstream of the barrier. Steep hydraulic head gradients across barriers have been observed in other studies and often can be used to determine their properties (Bense et al., 2003; Bense & Van Balen, 2004; Fienen, Hunt, et al., 2009). In this study, inversions with only few observation wells close to the real barrier yielded a good match with the real structure, showing that sparse wells can still guide the detection of structures using phantom structures.

The type of observation data will be integral to the success of this approach. We analyzed hydraulic head and groundwater age data. Incorporating groundwater age data in model inversions can shed light on groundwater system processes or parameterizations that may not otherwise be evident (Schilling et al., 2019). In our results age data strengthened the inversion outcome. Other novel datasets to characterize regional systems could also be used, such as river stage data (Wang et al., 2017), or hydraulic testing data (Tiedeman et al., 1995). When error was added to the observation data in the inversion, the final structure distribution did not resemble the real model as closely. Observation error is relatively commonplace in the collection of hydraulic head (Post and von Asmuth, 2013) and groundwater age (Cook & Solomon, 1997) data and so care should be taken to examine the observation data that exert the most control on the phantom structure location (e.g., using a sensitivity analysis) and take repeat measurements at these locations where possible.

The geometry of the phantom structures generated in a given realization is a critical factor in determining the outcome of the inversions. We saw that out of the three configurations, PS1 provided structures that most closely resembled the real structure and so resulted in the best fitting key structures. We see this approach as highly beneficial for identifying likely locations and properties of structures. However, we do not suggest the solution should be taken as an accurate representation of the subsurface and used for a generating a single predictive model without any further field investigations and uncertainty analysis. One of the reasons for this is because the real structure is not definitively present as one of the phantom structures. Our results show that as the number of observation wells increase, the predicted properties and locations of barriers more accurately resemble the real

structure (Figure 11). However, the exact geometric properties of the real structure were not converged upon, even in the case with 500 wells, because the location of phantom structures remained fixed during the model inversion. To increase the chance that the inversion outcome replicates true subsurface structure, the prior constraints on the phantom structure distribution in terms of their location, orientation, and length characteristics can be refined. These prior constraints may be based on field knowledge of structure properties, or they may be obtained by a previous model inversion. In this study, the results from one inversion were used to provide prior phantom structure constraints for a second inversion (Supporting Information S1) and the true subsurface structure is thus more closely replicated by estimated phantom structures. The better informed one is in generating the phantom structures to begin with, the more successful this approach will be. Depending on the grid resolution, more phantom structures could be included, which would increase the likelihood of estimating the true structure properties. Phantom structures could even be curved or kinked if these properties are commonly exhibited in a field region.

4.3. Limitations and Extensions

This study has several limitations. First, the real model was controlled with fixed recharge and fixed aquifer hydraulic conductivity in areas without barriers. We know that recharge can influence hydraulic head and groundwater age distributions across barriers, as can the geometry of the barrier itself (Marshall et al., 2020). Future studies could include additional parameters in the model inversion, such as heterogeneous aquifer hydraulic conductivity, recharge, and more complex barrier geometries. In real systems, the physical dimensions of faults and dykes vary with depth. Three-dimensional models would allow more realistic barrier geometries. A follow on from this work would be to test many different case studies and to quantify the likelihood of success in detecting a barrier and the number and location of observation wells given certain phantom structure distributions. This likelihood would depend on the size of the barrier relative to the study extent and to the grid spacing, the size of the barrier compared to the well density, the aquifer and barrier hydraulic conductivity values, and the hydraulic gradient, amongst other things.

A recommended area of future research would be to compare the use of direct age simulation with the explicit modeling of tracer concentrations. When using age as an observation, age is not directly measured or observed (as head is), but rather is a derived (indirect) estimate based on interpretive analyses of the concentrations of specific environmental tracers in water samples. Using a direct age simulation is helpful for showing the potential for age data in this type of model inversion. The use of environmental tracers for the determination of groundwater age is complex, and no single tracer is accurate over a broad range of ages. For the ages shown in this hypothetical model set up, where age had a mean value of 620 years, age tracers of ^{39}Ar and He are likely to be the most beneficial (Cook & Böhlke, 2000). Clearly, age tracers that will be sensitive over the age range within the aquifer need to be used and using a suite of multiple tracers that span the age range is always beneficial for characterizing a complex system (McCallum, Cook, et al., 2014).

An extension to this approach would be to investigate new methods to address flexibility in the parameterization of the structures during the inversion. This could include research into approaches that allow the barrier locations and properties to freely move during the inversion, rather than being fixed. Methods to achieve this may include non-linear or transdimensional inversion approaches, where a broader and more flexible ensemble of hydrogeological barrier and aquifer configurations could be included in the inversion (e.g., Athens & Caers, 2021; Fienen, Hunt, et al., 2009; Harp et al., 2008; Khambhammettu et al., 2020; Somogyvári et al., 2017). However, the data and resource intensity of these approaches could be a challenge for structurally complex regions. More flexible model structures may be possible using many more pilot points in the interpolation, but overfitting is also a concern (Moore & Doherty, 2005). Using pilot points in combination with ellipse or ellipsoid shaped interpolation algorithms may also be a simple solution to reduce parameterization demands and still estimate locations and properties of linear-type structures in an inversion (e.g., Kilmer et al., 2003; Tiedeman et al., 1995).

5. Conclusions

This study demonstrates that groundwater model inversion can be used to assess the possible influence of hydrogeological barriers where no prior knowledge of the barriers exists. Phantom structures are presented as a regularization device to allow sharp structures, such as faults and dykes, to be included in a model inversion. These are used for identifying the geometries and hydraulic conductivities of internal low-permeability barriers in models.

The results show that two to three phantom structures can reasonably and successfully reproduce hydraulic head and groundwater age distributions for a hypothetical, "real" aquifer that has one internal hydrogeologic barrier. We found good agreement between the lengths, orientations and hydraulic conductivities of the key phantom structures and the real hydrogeologic barrier. For the hypothetical case shown, the location and geometric properties of phantom structures approach the real structure when 50 observation wells or more were used in an inversion. They also indicate that hydraulic head data are most beneficial within approximately 2 km of the barrier. Groundwater age observations can also be useful when located at great distances downstream of the barrier, where groundwater age is older than otherwise expected due to the constricting presence of the barrier on groundwater flow.

Model inversions using phantom structures are compared to using more traditional pilot points. As the number of observation wells decreases, both approaches become less successful in replicating the hydraulic head and groundwater age distribution across the model grid. Yet we found that the fit was overall better with phantom structures than with pilot points over much of the observation well range. This shows that the method is a viable one and is successful when compared to pilot points for replicating groundwater flow resulting from narrow structures such as faults and dykes. The best approach to use, however, will depend on the purpose of a specific model inversion study. The results of applying the phantom structure approach identify hypothetical structures that affect the groundwater flow field; their identification provides a basis for refining a conceptual model and for focusing a field exploration program with limited resources.

Data Availability Statement

All codes for model inversion pre- and post-processing can be found at https://github.com/sarahkmarshall/uncertainty_barriers.

Acknowledgments
This study contributes to a collaborative project between the National Centre for Groundwater Research and Training (NCGRT) and Rio Tinto Iron Ore. Funding was received from the Australian Government Research Training Program, the Australian Research Council, through Linkage Grants LP150100395 and LP180101153, and by Rio Tinto Iron Ore. We appreciate advice from James McCallum, John Doherty, and Rui Hugman. We appreciate feedback on the manuscript from Paul Hsieh. For insightful reviews, we thank Mike Fienen, Daniel Hadley, and anonymous reviewers. The corresponding author would like to thank the U.S Geological Survey for hosting her as an exchange scholar. There were no samples or observational experimental data generated as a part of this study.

References

Athens, N., & Caers, J. (2021). Stochastic inversion of gravity data accounting for structural uncertainty. *Mathematical Geosciences*, *54*(2), 413–436. https://doi.org/10.1007/s11004-021-09978-2

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

Bedekar, V., Morway, E. D., Langevin, C. D., & Tonkin, M. (2016). *MT3D-USGS version 1: A U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW. Techniques and Methods 6–A53*. U.S. Geological Survey.

Bense, V. F., Gleeson, T., Loveless, S. E., Bour, O., & Scibek, J. (2013). Fault zone hydrogeology. *Earth-Science Reviews*, *127*, 171–192. https://doi.org/10.1016/j.earscirev.2013.09.008

Bense, V. F., & Kooi, H. (2004). Temporal and spatial variations of shallow subsurface temperature as a record of lateral variations in groundwater flow. *Journal of Geophysical Research*, *109*(B4), B04103. https://doi.org/10.1029/2003JB002782

Bense, V. F., & Van Balen, R. (2004). The effect of fault relay and clay smearing on groundwater flow patterns in the Lower Rhine Embayment. *Basin Research*, *16*(3), 397–411. https://doi.org/10.1111/j.1365-2117.2004.00238.x

Bense, V. F., Van Balen, R. T., & De Vries, J. J. (2003). The impact of faults on the hydrogeological conditions in the Roer Valley Rift System: An overview. *Netherlands Journal of Geosciences*, *82*(1), 41–54. https://doi.org/10.1017/S0016774600022782

Bredehoeft, J. (2005). The conceptualization model problem—Surprise. *Hydrogeology Journal*, *13*(1), 37–46. https://doi.org/10.1007/s10040-004-0430-5

Caine, J. S., Evans, J. P., & Forster, C. B. (1996). Fault zone architecture and permeability structure. *Geology*, *24*(11), 1025–1028. https://doi.org/10.1130/0091-7613(1996)024<1025:fzaaps>2.3.co;2

Cardiff, M., & Kitanidis, P. K. (2009). Bayesian inversion for facies detection: An extensible level set framework. *Water Resources Research*, *45*(10), W10416. https://doi.org/10.1029/2008WR007675

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research*, *22*(2), 199–210. https://doi.org/10.1029/WR022i002p00199

Castro, M. C., & Goblet, P. (2005). Calculation of ground water ages—A comparative analysis. *Groundwater*, *43*(3), 368–380. https://doi.org/10.1111/j.1745-6584.2005.0046.x

Chen, J., & Rubin, Y. (2003). An effective Bayesian model for lithofacies estimation using geophysical data. *Water Resources Research*, *39*(5), 1118. https://doi.org/10.1029/2002WR001666

Chou, T.-K., Chouteau, M., & Dubé, J.-S. (2016). Estimation of saturated hydraulic conductivity during infiltration test with the aid of ERT and level-set method. *Vadose Zone Journal*, *15*(7), vzj2015.05.0082. https://doi.org/10.2136/vzj2015.05.0082

Cilona, A., Aydin, A., & Johnson, N. M. (2015). Permeability of a fault zone crosscutting a sequence of sandstones and shales and its influence on hydraulic head distribution in the Chatsworth Formation, California, USA. *Hydrogeology Journal*, *23*(2), 405–419. https://doi.org/10.1007/s10040-014-1206-1

Comte, J.-C., Wilson, C., Ofterdinger, U., & González-Quirós, A. (2017). Effect of volcanic dykes on coastal groundwater flow and saltwater intrusion: A field-scale multiphysics approach and parameter evaluation. *Water Resources Research*, *53*(3), 2171–2198. https://doi.org/10.1002/2016WR019480

Cook, P. G., & Böhlke, J. K. (2000). Determining timescales for groundwater flow and solute transport. In P. G. Cook & A. L. Herczeg (Eds.), *Environmental tracers in subsurface hydrology* (pp. 1–30). Kluwer.

Cook, P. G., Dogramaci, S., McCallum, J. L., & Hedley, J. (2016). Groundwater age, mixing and flow rates in the vicinity of large open pit mines, Pilbara region, northwestern Australia. *Hydrogeology Journal*, *25*(1), 39–53. https://doi.org/10.1007/s10040-016-1467-y

Cook, P. G., & Solomon, D. K. (1997). Recent advances in dating young groundwater: Chlorofluorocarbons, $^{3}H/^{3}He$ and ^{85}Kr. *Journal of Hydrology*, *191*(1–4), 245–265. https://doi.org/10.1016/S0022-1694(96)03051-X

Doherty, J., & Welter, D. (2010). A short exploration of structural noise. *Water Resources Research*, *46*(5), W05525. https://doi.org/10.1029/2009WR008377

Doherty, J. E. (2003). Ground water model calibration using pilot points and regularization. *Groundwater*, *41*(2), 170–177. https://doi.org/10.1111/j.1745-6584.2003.tb02580.x

Doherty, J. E. (2010). *PEST Model-Independent Parameter Estimation User Manual*. Watermark Numerical Computing.

Doherty, J. E. (2015). *Calibration and uncertainty analysis for complex environmental models. PEST: Complete theory and what it means for modeling in the real world*. Watermark Numerical Computing.

Doherty, J. E., & Hunt, R. J. (2009). Two statistics for evaluating parameter identifiability and error reduction. *Journal of Hydrology*, *366*(1–4), 119–127. https://doi.org/10.1016/j.jhydrol.2008.12.018

Doherty, J. E., & Hunt, R. J. (2010). Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration. In *Scientific investigations report 2010–5169*. U.S. Geological Survey.

Enemark, T., Peeters, L. J. M., Mallants, D., & Batelaan, O. (2019). Hydrogeological conceptual model building and testing: A review. *Journal of Hydrology*, *569*, 310–329. ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2018.12.007

Fienen, M. N., Hunt, R., Krabbenhoft, D., & Clemo, T. (2009). Obtaining parsimonious hydraulic conductivity field using head and transport observations: A Bayesian geostatistical parameter estimation approach. *Water Resources Research*, *45*(8), 1–23. https://doi.org/10.1029/2008WR007431

Fienen, M. N., Muffels, C. T., & Hunt, R. J. (2009). On constraining pilot point calibration with regularization in PEST. *Groundwater*, *47*(6), 835–844. https://doi.org/10.1111/j.1745-6584.2009.00579.x

Gelhar, L. W., Welty, C., & Rehfeldt, K. R. (1992). A critical review of data on field-scale dispersion in aquifers. *Water Resources Research*, *28*(7), 1955–1974. https://doi.org/10.1029/92wr00607

Gleeson, T., & Novakowski, K. (2009). Identifying watershed-scale barriers to groundwater flow: Lineaments in the Canadian Shield. *GSA Bulletin*, *121*(3/4), 333–347. https://doi.org/10.1130/B26241.1

Goode, D. J. (1996). Direct simulation of groundwater age. *Water Resources Research*, *32*(2), 289–296. https://doi.org/10.1029/95wr03401

Gumm, L. P., Bense, V. F., Dennis, P. F., Hiscock, K. M., Cremer, N., & Simon, S. (2016). Dissolved noble gases and stable isotopes as tracers of preferential fluid flow along faults in the Lower Rhine Embayment, Germany. *Hydrogeology Journal*, *24*(1), 99–108. https://doi.org/10.1007/s10040-015-1321-7

Hadley, D. R., Abrams, D. B., & Roadcap, G. S. (2019). Modeling a large-scale historic aquifer test: Insight into the hydrogeology of a regional fault zone. *Groundwater*, *58*(3), 453–463. https://doi.org/10.1111/gwat.12922

Hao, Y., Yeh, T. C. J., Xiang, J., Illman, W. A., Ando, K., Hsu, K. C., & Lee, C. H. (2008). Hydraulic tomography for detecting fracture zone connectivity. *Groundwater*, *46*(2), 183–192. https://doi.org/10.1111/j.1745-6584.2007.00388.x

Harp, D. R., Dai, Z., Wolfsberg, A. V., Vrugt, J. A., Robinson, B. A., & Vesselinov, V. V. (2008). Aquifer structure identification using stochastic inversion. *Geophysical Research Letters*, *35*(8), L08404. https://doi.org/10.1029/2008GL033585

Hill, M. C., & Tiedeman, C. R. (2007). *Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty*. Wiley-Interscience.

Illman, W. A., Liu, X., Takeuchi, S., Yeh, T.-C. J., Ando, K., & Saegusa, H. (2009). Hydraulic tomography in fractured granite: Mizunami Underground Research site, Japan. *Water Resources Research*, *45*(1), W01406. https://doi.org/10.1029/2007WR006715

Khambhammettu, P., Renard, P., & Doherty, J. (2020). The Traveling Pilot Point method. A novel approach to parameterize the inverse problem for categorical fields. *Advances in Water Resources*, *138*, 103556. https://doi.org/10.1016/j.advwatres.2020.103556

Kilmer, M. E., Miller, E. L., Barbaro, A., & Boas, D. (2003). Three-dimensional shape-based imaging of absorption perturbation for diffuse optical tomography. *Applied Optics*, *42*(16), 3129–3144. https://doi.org/10.1364/AO.42.003129

Knowling, M. J., White, J. T., Moore, C. R., Rakowski, P., & Hayley, K. (2019). On the assimilation of environmental tracer observation for model-based decision support. *Hydrology and Earth System Sciences*, *24*(4), 1677–1689. https://doi.org/10.5194/hess-24-1677-2020

Konikow, L. F., Hornberger, G. Z., Putnam, L. D., Shapiro, A. M., & Zinn, B. A. (2008). The use of groundwater age as a calibration target. In J. C. Refsgaard, K. Kovar, E. Haarder, & E. Nygaard (Eds.), *Calibration and reliability in groundwater modeling: Credibility of modeling* (Vol. 320, pp. 250–256). IAHS Publication.

Koohbor, B., Fahs, M., Ataie-Ashtiani, B., Belfort, B., Simmons, C. T., & Younes, A. (2019). Uncertainty analysis for seawater intrusion in fractured coastal aquifers: Effects of fracture location, aperture, density and hydrodynamic parameters. *Journal of Hydrology*, *571*, 159–177. https://doi.org/10.1016/j.jhydrol.2019.01.0

Le Goc, R., de Dreuzy, J.-R., & Davy, P. (2010). An inverse problem methodology to identify flow channels in fractured media using synthetic steady-state head and geometrical data. *Advances in Water Resources*, *33*(7), 782–800. https://doi.org/10.1016/j.advwatres.2010.04.011

Leray, S., de Dreuzy, J.-R., Bour, O., Labasque, T., & Aquilina, L. (2012). Contribution of age data to the characterization of complex aquifers. *Journal of Hydrology*, *464*(465), 54–68. https://doi.org/10.1016/j.jhydrol.2012.06.052

Marler, J., & Ge, S. (2003). The permeability of the Elkhorn fault zone, South Park, Colorado. *Groundwater*, *41*(3), 321–332. https://doi.org/10.1111/j.1745-6584.2003.tb02601.x

Marshall, S. K., Cook, P. G., Konikow, L. F., Simmons, C. T., & Dogramaci, S. (2020). Conjoint use of hydraulic head and groundwater age data to detect hydrogeologic barriers. *Hydrogeology Journal*, *28*(3), 1003–1019. https://doi.org/10.1007/s10040-019-02095-9

Marshall, S. K., Cook, P. G., Miller, A. D., Simmons, C. T., & Dogramaci, S. (2019). The effect of undetected barriers on groundwater drawdown and recovery. *Groundwater*, *57*(5), 718–726. https://doi.org/10.1111/gwat.12856

Mayer, A., May, W., Lukkarila, C., & Diehl, J. (2007). Estimation of fault-zone conductance by calibration of a regional groundwater flow model: Desert Hot Springs, California. *Hydrogeology Journal*, *15*(6), 1093–1106. https://doi.org/10.1007/s10040-007-0158-0

McCallum, J. L., Cook, P. G., Simmons, C. T., & Werner, A. D. (2014). Bias of apparent tracer ages in heterogeneous environments. *Groundwater*, *52*(2), 239–250. https://doi.org/10.1111/gwat.12052

McCallum, J. L., Herckenrath, D., & Simmons, C. T. (2014). Impact of data density and geostatistical simulation technique on the estimation of residence times in a synthetic two-dimensional aquifer. *Mathematical Geosciences*, *46*(5), 539–560. https://doi.org/10.1007/s11004-013-9518-6

Moore, C., & Doherty, J. (2005). Role of the calibration process in reducing model predictive error. *Water Resources Research*, *41*(5), W05020. https://doi.org/10.1029/2004WR003501

Niswonger, R. G., Panday, S., & Ibaraki, M. (2011). *MODFLOW-NWT, A Newton formulation for MODFLOW-2005. Techniques and Methods 6–A37*. U.S. Geological Survey.

Ochoa-González, G. H., Carreón-Freyre, D., Cerca, M., & López-Martínez, M. (2015). Assessment of groundwater flow in volcanic faulted areas. A study case in Queretaro, Mexico. *Geofísica Internacional*, *54*(3), 199–220. https://doi.org/10.1016/j.gi.2015.04.016

Post, V. E., & von Asmuth, J. R. (2013). Hydraulic head measurements—New technologies, classic pitfalls. *Hydrogeology Journal*, *21*(4), 737–750. https://doi.org/10.1007/s10040-013-0969-0

Sanford, W. (2011). Calibration of models using groundwater age. *Hydrogeology Journal*, *19*(1), 13–16. https://doi.org/10.1007/s10040-010-0637-6

Schilling, O. S., Cook, P. G., & Brunner, P. (2019). Beyond classical observations in hydrogeology: The advantages of including exchange flux, temperature, tracer concentration, residence time, and soil moisture observations in groundwater model calibration. *Reviews of Geophysics*, *57*(1), 146–182. https://doi.org/10.1029/2018RG000619

Somogyvári, M., Jalali, M., Jimenez Parras, S., & Bayer, P. (2017). Synthetic fracture network characterization with transdimensional inversion. *Water Resources Research*, *53*(6), 5104–5123. https://doi.org/10.1002/2016WR020293

Tiedeman, C. R., Hsieh, P. A., & Christian, S. B. (1995). Characterization of a high-transmissivity zone by well test analysis: Steady state case. *Water Resources Research*, *31*(1), 27–37. https://doi.org/10.1029/94wr01965

Underwood, S. C., McCallum, J. L., Cook, P. G., Simmons, C. T., Dogramaci, S., Purtschert, R., et al. (2018). Physical and chemical controls on the simultaneous occurrence of young and old groundwater inferred from multiple age tracers. *Water Resources Research*, *54*(11), 9514–9532. https://doi.org/10.1029/2018WR022800

Wang, Y. L., Yeh, T. C. J., Wen, J. C., Huang, S. Y., Zha, Y., Tsai, J. P., et al. (2017). Characterizing subsurface hydraulic heterogeneity of alluvial fan using riverstage fluctuations. *Journal of Hydrology*, *547*, 650–663. https://doi.org/10.1016/j.jhydrol.2017.02.032

White, J. T., Fienen, M. N., & Doherty, J. E. (2016). A Python framework for environmental model uncertainty analysis. *Environmental Modelling & Software*, *85*, 217–228. https://doi.org/10.1016/j.envsoft.2016.08.017

Zell, W. O., Culver, T. B., & Sanford, W. E. (2018). Prediction uncertainty and data worth assessment for groundwater transport times in an agricultural catchment. *Journal of Hydrology*, *561*, 1019–1036. https://doi.org/10.1016/j.jhydrol.2018.02.0