Bayesian evidential learning: a field validation using push-pull tests

Thomas Hermans, Ghent University, Department of Geology

Nolwenn Lesparre, Strasbourg University, Laboratory of Hydrology and Geochemistry; previously at Liege University, Department of Urban and Environmental Engineering.

Guillaume De Schepper, Aquale SPRL, R&D Department, Noville-les-Bois, Belgium.

Tanguy Robert, Liege University, Department of Urban and Environmental Engineering; F.R.S.-FNRS (Fonds de la Recherche Scientifique); previously @ Aquale SPRL, Department of R&D, Noville-les-Bois, Belgium

Corresponding author

Thomas Hermans, Ghent University, Department of Geology, Krijgslaan 281, 9000 Gent, Belgium, thomas.hermans@ugent.be

Keywords: Bayesian evidential learning, push-pull tests, tracer tests, heterogeneity, uncertainty

"This is a post-peer-review, pre-copyedit version of an article published in Hydrogeology Journal. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10040-019-01962-9"

Abstract

Recent developments in uncertainty quantification show that a full inversion of model parameters is not always necessary to forecast the range of uncertainty of a specific prediction in Earth Sciences. Instead, Bayesian evidential learning (BEL) uses a set of prior models to derive a direct relationship between data and prediction. This recent technique has been mostly demonstrated for synthetic cases. This paper demonstrates the ability of BEL to predict the posterior distribution of temperature in an alluvial aquifer during a cyclic heat tracer push-pull test. The data set corresponds to another push-pull experiment with different characteristics (amplitude, duration, number of cycles). This experiment constitutes the first demonstration of BEL on real data in a hydrogeological context. It should open the range of future applications of the framework for both scientists and practitioners.

1. Introduction

The ability of researchers and decision makers to anticipate the consequences of external events, or their actions in complex environments, depends on the predictive capacity of science, and in particular the reliance on models. For future generations this predictive ability will impact the management of groundwater resources, including climate-change effects (e.g., Aquilina et al., 2015), environmental issues (e.g., MacDonald et al., 2016), and the transition to sustainable energy (e.g., Kammen and Sunter, 2016).

Researchers and decision makers are grappling with very complex models to enhance these models' predictive abilities. The very nature of the subsurface is so complex that any prediction is subject to large uncertainties. It is clear that a prediction alone is not sufficient, but an entire uncertainty quantification, reflecting all possible outcomes, is required for a proper risk analysis and subsequent decision making (Scheidt et al., 2018).

Recent advances show that predicting the outcomes of subsurface models does not necessarily require solving an inverse problem and generating model(s) fitting the data (Scheidt et al., 2018). Instead, Bayesian evidential learning (BEL) proposes to use an ensemble of prior realizations to learn a direct relationship between data and prediction variables. Those prior models are samples of the prior distribution of model parameters, reflecting the range of uncertainty before data acquisition. The derived relationship between data and prediction enables one to directly forecast the predictions corresponding to the field observed data and their associated uncertainty (Scheidt et al. 2018; Hermans, 2017). This process does not require a full explicit model inversion, making it computationally less expensive than standard inversion methods.

It must be stressed that BEL is fundamentally different from surrogate-based approaches (see Razavi et al., 2012, for a review). Surrogate approaches are seeking an approximation of the physical forward model to speed up the simulation process and make Markov chain Monte Carlo methods more efficient (e.g., Chen et al., 2018). In BEL, the physics of the processes are fully accounted for. The derivation of a direct relationship between data and prediction, made possible by the use of dimension reduction techniques, eliminates the need to run any additional forward simulations.

The initial idea behind BEL was first introduced by Scheidt et al. (2015b) and Satija and Caers (2015) with synthetic examples for predicting the arrival of a contaminant in a well using monitoring data collected in three upstream locations. It was then extended by Hermans et al. (2016) for estimating aquifer properties using time-lapse geophysical data, and by Satija et al. (2017) for history matching of petroleum reservoirs. Those two studies investigated complex heterogeneous reservoirs inspired by real conditions, but still with synthetic cases.

Although the number of real field applications is still limited, BEL has recently been illustrated for real case studies in relation to oil resources, groundwater resources, shallow geothermal energy and contamination problems (Scheidt et al., 2018). By definition, predictions from subsurface models generally concern the future behavior of the system with different stress factors corresponding to alternative management strategies. Therefore, there is almost always a lack of available data to verify the solution in real case studies. The prior uncertainty in such contexts is often very large, and a demonstration of the applicability of BEL in a complex field case is still missing. In a recent study, Hermans et al. (2018) used time-lapse electrical resistivity tomography data collected during a heat tracing experiment to estimate the heat storage capacity of an alluvial aquifer. They illustrated the approach for the estimation of spatially distributed temperature using time-lapse geophysical data. However, their ground-truth data were limited to two point measurements. Moreover, the application to geophysical data means that data and prediction are co-located in time and space, a favorable situation for the prediction.

In this paper, it is proposed to validate BEL as an accurate prediction framework using two independent hydrogeological field experiments, namely push-pull tests. Push-pull tests are informative, single-well experiments that do not require extensive monitoring networks or heavy field campaigns (Haggerty et al., 1998). They are therefore particularly suited to poorly equipped sites and in absence of extensive prior information, to derive both flow and transport behaviors (e.g., Vandenbohede et al., 2009; Paradis et al., 2018). In the following, the second experiment is considered as the target prediction, and the field observations are used to assess the consistency of the posterior distribution. Although a validation of the framework in the Bayesian sense would require more repetitions, which is not possible in the context of this field experiment, it will be shown that the calculated posterior cannot be falsified by the data. This demonstrates that BEL, upon a realistic characterization of the prior uncertainty, can be used to realistically forecast the desired prediction in real field applications. In this contribution, the term validation should thus be interpreted in that broader sense.

2. Methods

2.1. Bayesian evidential learning

The objective of the paper is the application of BEL in field conditions and the assessment of the consistency of BEL predictions. Therefore, the framework itself is only shortly described, following the description provided by Hermans et al. (2018), where an exhaustive description can be found. Although some technical details and choices (sensitivity analysis, dimension reduction techniques) are highlighted, BEL is a general framework and can be applied using other techniques (Scheidt et al., 2018).

BEL can be usually divided into 4 main steps (Fig. 1). The first step consists of the definition of the prior model, i.e. the range of variations of the model parameters (hydraulic conductivity, porosity), stress factors (boundary conditions, pumping rates) and aquifer structure (geological scenarios, spatial heterogeneity) based on the current knowledge, before any new data acquisition. This step is extremely important because ignoring some prior uncertainty component bears the risk of artificially reducing the uncertainty in the prediction. This prior model is then sampled to generate a representative set of model realizations or prior samples. The two experiments corresponding to data and prediction variables are simulated using a forward groundwater flow and transport model. BEL allows using a relatively limited number of models even for large prior uncertainty, because it is driven by the complexity (often limited) of the prediction (Hermans et al., 2018) and not by the model parameterization. In this study, 500 prior samples are used.

In a second step, BEL proceeds to data-worth assessment. Using a global sensitivity approach based on the prior samples' response, it identifies the most sensitive parameters for data and prediction variables. If both are sensitive to the same parameters, then the data are likely informative for the prediction. If not, an alternative data set can be proposed. Here, distance-based global sensitivity analysis (DGSA) was used to identify the most sensitive parameters (Park et al., 2016; Fenwick et al., 2014). It is worth noticing that these 2 first steps in BEL are field data independent, i.e. they can be performed before data acquisition, for example, for experimental design (Hermans, 2017).

The third step is prior falsification. Once field data are collected, it is crucial to verify that the observed data can be predicted by the prior. Otherwise, a risk exists for the prediction to be erroneous. Indeed, BEL, as with any Bayesian method, requires the posterior distribution to be part of the prior span (Hou and Rubin, 2005). If the prior model is falsified (inconsistency with the data), a revision of the latter is mandatory. As will be seen in section '*Prior model falsification'*, for simple data sets, falsification can be performed by simple visualization of the prior samples' response and field data response. For more complex data sets, dimension reduction techniques might be needed to visually assess the consistency of the prior model (e.g., Hermans et al., 2018).

Finally, a prediction-focused approach is used to generate the posterior distribution of the prediction given the observed data. A direct relationship between data and prediction variables is sought using the responses of the prior samples. Given the generally high dimensionality of data and prediction variables, this objective is achieved through statistical and/or machine learning techniques in a reduced dimension space. Once such a relationship is found, it is possible to forecast the prediction based on field data. Many technical solutions can be implemented (e.g., Scheidt et al., 2018). Here, a combination of principal component analysis (PCA, see e.g., Krzanowski (2000)) to reduce the dimensionality of data and prediction variables, canonical correlation analysis (CCA, see e.g., Krzanowski (2000)) to linearize the relationship between both variable types, and kernel density estimation (KDE, e.g., Bowman and Azzalini, (1997)) to estimate the distribution corresponding to field data, were used. Kernel density requires definition of the bandwidth of the kernel for estimation. An automatic choice can be implemented based on the density of samples, but the choice can also be adapted depending on local conditions. (e.g., Bowman and Azzalini, 1997)

2.2. Field site

The studied field site is located in Hermalle-sous-Argenteau (Belgium), in the alluvial aquifer of the Meuse River. The area of interest has already been investigated using hydrogeological and geophysical experiments (Brouyère, 2001; Wildemeersch et al., 2014; Hermans et al., 2015a, 2015b, 2018; Hermans and Irving, 2017; Klepikova et al., 2016; Lesparre et al., 2019). It consists of three main layers: a first (top) layer composed of unsaturated loam and loamy to clayey sands, 3 m thick; the first aquifer layer composed of sandy gravel, about 4 m thick; and then a more hydraulically conductive layer composed of clean coarse gravel, about 3 m thick. Below, the Carboniferous bedrock (shale) constitutes a low-permeability layer and the base of the alluvial aquifer. The water level is located at around 3 m depth, coincident with the boundary between the loam and sandy gravel layer (Fig. 2).

In this paper, two single-well experiments carried out in well Pz15sup are considered. This well is drilled down to the middle of the sandy gravel layer and screened between 4 and 5 meters below ground surface (mbgs) (Fig. 2). The interested reader can refer to the above-mentioned references for details on the Hermalle-sous-Argenteau site and to the H^+ database for access to the data (Réseau National de Sites Hydrogéologiques , 2019).

2.3. Field experiments

The two considered experiments correspond to push-pull tests carried out in October 2016 and February 2017, respectively. A push-pull test consists of three phases: 1) an injection phase (push) during which a tracer is injected into a single monitoring well, 2) an optional storage or resting phase during which the tracer is subjected to natural conditions, and 3) a pumping phase (pull) during which water is extracted from the aquifer and the tracer recovery curve is analyzed (e.g. Haggerty et al., 1998). For both experiments, the tracer was heated water. During the whole experiment, the temperature in the well was continuously monitored using a CTD diver. The water used for injection was pumped from a well located downstream at a distance satisfactory enough to avoid any significant influence on the hydraulic heads, and subsequently heated using a mobile water heater before use as the tracer water. Recorded drawdowns/rises in both wells were found to be limited to +/-1 cm; nevertheless, Jamin and Brouyère (2018) have shown that a limited pumping rate still influences the fluxes in the aquifer. The pumping well is thus explicitly represented in the hydrogeological model.

During the first experiment, heated water was injected in the well at the rate of 3 m^3/h with an average temperature difference (ΔT) of 28 K during 6 h at the outlet of the water heater. At the end of the injection period, due to a technical problem with the water heater, cold water (ΔT = 0 K) was injected for 20 minutes. The storage phase lasted for 91 h, after which water was extracted from the well at the rate of 5 m^3/h during 15.5 h. To minimize the influence of the injection of cold water on the process, the first 36 hours of the storage phase are disregarded from the dataset (Fig. 3a). Indeed, after the injection of cold water, a rebound is observed (temperature increases in the well). However, during that phase, the temperature in the well and in the aquifer are not at equilibrium. Such a discrepancy exists at any moment, but is more significant after the injection of cold water. For the same reason, the temperatures recorded during the injection phase are not representative of the temperature in the aquifer and are removed from the dataset. Note that the injection of cold water is still numerically modeled. More details on this experiment can be found in Lesparre et al. (2019).

The second experiment is the target prediction of the study. It also consisted of a push-pull test with a storage phase, but was made of two successive cycles. Each cycle corresponded to an injection phase of 5 h at a 3 m³/h rate, a storage phase of 19 h, and a pumping phase at a rate of 5 m³/h for 5 h. The temperature difference was ΔT = 30 K and ΔT = 35 K for the first and second cycles, respectively. During both cycles, another storage phase of 19 h took place (Fig. 3b).

3. Results

3.1. Definition of the prior model

The prior model should be defined based on current knowledge of the site, which is relatively well documented (see section '*Field site*'). However, it is rare to have such a large amount of information and field data for real-world case studies. To avoid any bias in the validation process, the range of uncertainty of the parameters was broadened to a more realistic situation in terms of real-world applications, as if the experiments were performed on a largely unknown site.

Spatial heterogeneity in the hydraulic conductivity of the sandy gravel layer is generated by means of sequential Gaussian simulations (Goovaerts, 1997) using a spherical variogram model. The range, the mean, the variance, the anisotropy and the orientation of the spatial random field are all considered uncertain. In particular, the mean hydraulic conductivity and its variance have large prior ranges, ignoring prior information on the site. Such values can generate high and low conductive environments, as well as almost homogeneous to highly heterogeneous models. Similarly, the porosity (indirectly affecting the bulk thermal properties) and the natural gradient in the aquifer are uncertain. The considered ranges of variation of those parameters in the prior are shown in Table 1. Each parameter is independently and randomly sampled from a uniform distribution to generate a unique prior realization. In total, 500 independent realizations are used.

In each model, the first soil layer is unsaturated and considered as a confining layer, whereas the third layer (clean gravel) is simulated using an average value of hydraulic conductivity of 0.05 m/s. This is justified because the aquifer response is not very sensitive to those parameters.

Parameter	**Range of uncertainty**
Mean of $\log_{10} K$ (m/s)	U[-4 to -1]
Variance $\log_{10} K$ (m/s)	U[0.05 to 2]
Range (m)	U[1 to 10]
Anisotropy ratio	U[0.1 to 0.5]
Orientation	U[0 to π]
Porosity	U[0.05 to 0.30]
Gradient (%)	U[0.083 to 0.167]

Table 1. Range of variation of the parameters in the prior. U means that a uniform distribution with specified range is assumed.

The control volume finite-element code HydroGeoSphere (Therrien et al., 2010) is used to simulate the field experiments. The model is oriented along the direction of flow identified in previous studies (Wildemeersch et al., 2014). The saturated part of the aquifer is modeled using 14 layers, 0.5 m thick, with 8 in the sandy layer and 6 in the clean gravel. The grid is centered on the injection well with an extension of 40 m in the direction perpendicular to flow and 60 m in the direction of flow. The grid is refined around the well with cell size starting at 2.5 cm and increasing with a multiplying factor of 1.15 up to a maximum value of 2.5 m. In the direction perpendicular to flow, the size of the cells is further limited to 0.25 cm within 3 m around the well in order to accommodate the presence of other monitoring wells, although they are not used in this study.

No-flow boundary conditions are used everywhere, except at the boundaries perpendicular to the direction of flow where the gradient is imposed based on the prior range (Table 1). Boundary conditions for heat transport assume fixed temperature equal to the initial temperature (T = 10.5°C) during the whole duration of both experiments.

3.2. Sensitivity analysis

A global sensitivity analysis both on data and prediction variables is carried out using DGSA. DGSA is based on the distance between the responses from pairs of models within the 500 prior models. The Euclidean distance is used between the time-dependent temperature curves at the well (Fig. 2). Based on the distance, a map of the models in a reduced dimension space is produced and classified using clusters. In this case, three clusters are a good compromise between the number of clusters and the number of models within clusters. It clearly identified curves with low, intermediate and high temperature (Fig. 3). In DGSA, the sensitivity of a parameter depends on the distribution of model parameters within those clusters compared to the initial distribution. A similar approach can be used to analyze interactions between parameters. To analyze the effect of parameter B on parameter A, the model responses are simply grouped in bins depending on their parameter B values. Then the sensitivity analysis for parameter A is repeated for each bin. If the response between bins is different, then a conditional effect or interaction is identified (Park et al., 2016).

The result of the sensitivity analysis for the two experiments shows similar sensitivity patterns (Fig. 4a and 4b). The most sensitive parameters are the mean and variance of the hydraulic conductivity distribution. Hydraulic conductivity influences the flow patterns in the aquifer and the advection velocity in particular. The variance is an indication of the heterogeneity of the medium (high variance means high heterogeneity), so that spatial heterogeneity also plays a role in the range of observed responses. The gradient, the range, and the anisotropy are also sensitive parameters but to a lesser extent. The influence of the gradient is expected to also influence advective fluxes. The gradient is not highly sensitive, probably because the prior range is relatively narrow compared to the range of variation of hydraulic conductivity (several orders of magnitude). The ranges of the variogram and the anisotropy ratio are parameters related to the spatial distribution of hydraulic conductivity. In combination with the variance, they control the degree of heterogeneity around the well and significantly influence the temperature curves. The porosity is not a sensitive parameter in the response of the aquifer to the two tests, although it has some direct influence on the bulk thermal parameters and advection velocity. Note that the results of the sensitivity are dominated by the mean hydraulic conductivity and its variance, which have the larger prior range of uncertainty. It is thus expected that they dominate the aquifer response in terms of sensitivity. Narrowing the range of prior uncertainty (see section '*Discussion*') would slightly reduce the observed difference between the parameters. However, the relative position of the parameters would remain the same and the same conclusions could be drawn (not shown in Fig 4).

The interaction between the parameters is related to the distance of their respective bubble in the interaction plot. Since the distances are relative, there is no unit on those plots. Fig. 4c and 4d show that the interaction between parameters is limited, except between the mean value of hydraulic conductivity and its variance. This probably indicates that the heterogeneity in the hydraulic conductivity distribution has a significant effect on the response of the aquifer to the push-pull tests. The result of this sensitivity analysis confirms that the standard experiment is somewhat informative in predicting the cyclic experiment, as the same sensitivity patterns are observed for both variables. In this case, the patterns are almost exactly the same, which is a favorable factor. However, it is not a requirement to apply BEL; only some overlapping is required (see Hermans et al., 2018). The global sensitivity analysis can also be used at an early stage to identify which parameters must be accounted for, and therefore reduce the complexity of the prior model by dropping insensitive parameters (Scheidt et al., 2018).

3.3. Prior model falsification

In BEL, prior model falsification is a crucial step. Indeed, the two first steps are field data independent. One can draw first conclusions about the usefulness of a specific experiment for a given prior model without the acquisition of any field data. However, the pre-conclusions are only valid if the prior model can be considered as consistent with the data. If the prior model is falsified, then the whole process might be influenced and the results of the sensitivity analysis might not hold for another prior model.

The prior model consistency is verified for both the data and the prediction. In most studies, only the data can be used because the prediction is not available yet. Both consist of temperature distribution through time at the injection well. Therefore, it is relatively easy to verify that the response's ensemble encompasses the observed data in terms of amplitude (maximum/minimum temperature changes) and temporal behavior (global trend, location of maximum/minimum, etc.).

Fig. 3 shows the data and prediction variables for the 500 prior samples and the field data. In the first experiment, the storage phase shows slowly decreasing temperature as heat diffuses and moves away from the injection well. The decrease in temperature speeds up once pumping begins and heat is recovered from the aquifer. At the end of the pumping phase, temperature stabilizes with residual heat stored in the medium matrix (Fig. 3a). The same phases are repeated twice in the cyclic experiment (Fig. 3b).

In this specific case, the prior model cannot be falsified based on data or prediction (Fig. 3). The prior model covers a wide range of possible outputs, with rapid or slow decrease of temperature during the pumping and storage phases of both experiments. The field data and predictions are located within the range observed in the prior samples' responses and have similar temporal behavior to most of the prior samples. For the first experiment, the effect of cold-water injection is still visible for models displaying temperature changes above 15°C, 2 days after the beginning of injection (the inflection point in the breakthrough curve after the rebound has not been reached yet).

For more complex data/prediction, a direct visualization of the prior span might not be easy. In such a case, it is useful to apply a dimension reduction technique to visualize the position of observed data compared to prior models in a 2D or 3D space (e.g., Hermans et al., 2015a). In this case, PCA is applied, as it will be later used in the prediction-focused step of the framework (Fig. 5). 500 temperature curves from the prior model and the field curve are simultaneously considered, and these are analyzed to determine whether the latter is encompassed in the prior span in the PCA-score space. For the standard test, almost 99% of the variance is explained by the first dimension. For the cyclic test, the two first dimensions explain 87.2 and 9.2% of the variance respectively. It is interesting to observe that the cyclic experiment seems to convey more variability than the standard test. Therefore, the standard test might not be sufficient to predict all the variability observed in the cyclic. Again, the prior model cannot be falsified on this basis (Fig. 5).

Interestingly, the field observation for the standard data set lies in the middle of the distribution while most models are concentrated at the borders. Those “extreme” models correspond to rapid or slow temperature decrease during the storage and pumping phases, while the field data show a rather intermediate behavior. Also relatively similar, the two maps for data and prediction are not the same, showing that the two responses share some components but also have differences. At this step, one could assess prior assumptions and update the prior model according to the falsification procedure (see section *‘Discussion’*). A thorough analysis of the mapping in Fig. 5 can reveal which range of parameters is more likely to generate data responses close to the observed one (e.g., Scheidt et al., 2015a).

3.4. Prediction

Following the logical path of BEL, it is shown that the data are likely informative for the prediction and that the prior is consistent with the data. Therefore, one can seek a direct relationship between the data and the prediction. This is done using the reduced dimensions after PCA. Three dimensions are kept for the data (more than 99.5% of the variance) and two dimensions for the prediction (96 % of the variance). The choice of two dimensions is guided by a compromise: it is desirable to keep as much variance as possible while reducing the dimensionality of the problem at maximum. Attempts to predict more dimensions in the prediction showed that the data are not informative on the higher dimensions of the prediction. Trying to explain more variance in the prediction is thus useless. CCA is then applied to the reduced data and prediction sets to generate independent linear relationships between reduced data and prediction (Fig. 6). Note that CCA is reversible if more dimensions are used for the data than for the prediction.

The direct relationship obtained after CCA is not simple. For the first dimension, the obtained relationship is not strictly linear (Fig. 6a). For the second dimension, CCA fails to find a unique linear relationship, but two different trends exist (Fig. 6b). The models aligned along $d_2^c = 0$ (d^c refers to the data variable in the low dimensional CCA space; 2 refers to its second dimension) correspond to models with very rapid temperature decrease during storage and do not follow the same trend as the others. Those models also correspond to the cluster around $d_1^c = 2$ in the first dimension of the CCA space. This behavior is further analyzed in the *'Discussion'* section.

The conditions to estimate the posterior distribution by linear regression are not met (linearity and Gaussianity). Therefore, one cannot estimate the posterior distribution analytically; it is instead estimated using KDE with a Gaussian kernel (Bowman and Azzalini, 1997). The latter is simply based on the distribution of prior samples in the CCA space. Note that it is still useful to apply CCA to derive the most linear relationship between data and prediction variables. Working in the PCA space would not ensure any relationship. The posterior distribution of the prediction in the CCA space is computed given the observed data (Fig. 6c and 6d). In this case, a reduced kernel bandwidth was used to avoid too much effect of the samples aligned along $d_2^c = 0$, explaining the peaks observed in the posterior (Fig. 6d). This parameter can be easily adapted based on the density of points in the CCA space.

Once the posterior distribution of the prediction in the reduced dimension space is known (Fig. 6c and 6d), it can be easily sampled and back transformed in the original space where the posterior distribution of the prediction can be displayed (Fig. 7). The predicted samples encompass the real observation, showing that BEL is successful in forecasting the desired prediction. However, the behavior during the storage and pulling phase is clearly different. During the pulling phase, BEL is able to predict with a very narrow range of uncertainty (~1°C) the temperature decrease of the extracted water. This is very satisfactory as this would be a typical prediction in applications such as aquifer thermal energy storage systems (Hermans et al., 2018). For the storage phase however, the uncertainty is wider. BEL tends to predict a relatively linear decrease of temperature as observed for the prior models with the highest temperature, while the real observation has an exponential decrease. Only a few predictions reproduce this trend, but the real prediction is still within the span of the posterior and therefore coherent with the uncertainty quantification.

4. Discussion

The larger uncertainty observed during the storage phase can probably be related to the design of the experiment. The standard test suffered from a technical problem of the mobile water flow heater resulting in the injection of cold water. It affected the whole storage phase, weakening the ability to predict the same phase for the cyclic test. In contrast to the pulling phase, during which water is extracted from the aquifer, the storage phase might suffer from a discrepancy in temperature between the water of the aquifer and in the well (loss of energy towards the atmosphere).

A few posterior models (blue lines in Fig. 7) display an unexpected behavior during the storage phase: after a rapid decrease in temperature, a rebound is generated followed by an almost constant temperature. This behavior is not physically plausible and constitutes one of the limitations of BEL. Indeed, since the prediction is generated on a statistical basis, it is never ensured that the sampled values are actually observed within the prior. In some cases, it can yield unrealistic solutions as observed here. Those solutions can be easily filtered out if needed. In this case, they seem to originate from the influence of the series of prior samples' response displaying a sharp temperature decrease during the beginning of the storage phase as shown by their low predicted temperature at the end of injection. This study investigates their influence on the results by removing them from the prior.

The results of the global sensitivity analysis are used and the 300 models corresponding to the most distant cluster (models at the extreme right in Fig. 5a) are removed from the prior realizations. Fig. 8 shows the distribution of model parameters in the removed samples and in the reduced prior model. Those samples generally correspond to a large average value of the hydraulic conductivity with large variance. For other parameters, the difference in the distribution is smaller. Those results are thus in agreement with the sensitivity analysis, showing that the hydraulic conductivity distribution is the main factor affecting the model response. It also indicates that the prior range is too large in terms of hydraulic conductivity. Values greater than 10^{-2} m/s are not realistic for the sandy gravel, but are characteristic of the underlying clean gravel layer. Similarly, extremely heterogeneous models with very large variance are not consistent with the data. Remember that the prior model was purposely enlarged compared to the actual knowledge of the site.

As shown by Fig. 9, removing those prior samples improves the capacity of CCA to derive a linear relationship between data and prediction. However, the conditions to calculate an analytical solution by linear regression are still not met. Therefore, KDE was also used. The effect on the posterior distribution however is limited (Fig. 10a). The posterior samples with unrealistic behaviors are successfully removed, confirming that their occurrence was correctly identified. The uncertainty during injection phases is also strongly reduced. However, the “new prior model” barely has an effect on the range of generated predictions during the storage phase. The real observation is still at the extreme limit of the posterior.

The reason for the slight overestimation of the temperature during the storage phase can be elucidated in the CCA space (Fig. 9). The black square indicates the real value of the prediction in the low dimension space. Generally, this value is unknown, but this study case has access to the reduced dimension of the prediction. For the second dimension, the real model lies in a densely populated zone of the space. However, for the first dimension, it lies at the extreme limit of the distribution. One of the prior samples is in the close vicinity of the real observation, but they both lie outside the main trend. Therefore, the prior model is able to produce data-prediction pairs similar to the observed one. However, the sampling of the cumulative distribution function will logically generate more samples in the denser area around $h_1^c \approx -5$, leading to higher temperature predictions. In short, given the observed data, the probability to get higher temperatures than observed, in reality, is high.

The predicted probability density function (pdf) of the first dimension has a mean value of – 5.11 (Fig. 9c) while the real prediction is – 16.65. If the pdf was corrected to have a mean value equal to the observed value, one would obtain the posterior distribution of Fig. 10b. On the latter, the posterior distribution is more centered on the real prediction, especially during the first cycle. This observation is further illustrated by the distribution of the scores in the CCA space (Fig. 11). It shows that the true prediction is located at the edge of the prior distribution, which makes it a difficult target for prediction (Satija and Caers, 2015; Hermans et al., 2016). In consequence, it is also in the edge of the posterior distribution.

The latter analysis indicates that BEL performs relatively well although it presents a challenging situation. The posterior distribution of the temperature curve is correctly estimated during both the pulling and the storage phases. During the storage phase, the real observation is within the posterior, although it lies at its extremity.

These observations can be related to the variability of the prior model, considering the large uncertainty in this case. There are not many models in the vicinity of the prediction, which is not a favorable condition to make a prediction. One possibility could be to generate more samples in this vicinity by identifying model parameters responsible for similar predictions. This can be done, for example, through advanced falsification approaches (Hermans et al., 2015a; Scheidt et al., 2015a, 2018).

However, one cannot disregard a possible discrepancy linked to the difference between field conditions and numerical simulations. As an example, the temperature measured in the well is likely not quite at equilibrium with the aquifer as simulated by the numerical model. It was also considered that the porosity is constant within the aquifer, which might be an oversimplification. However, those limitations are not inherent to BEL, but related to numerical tools.

5. Conclusion

This paper demonstrates that Bayesian evidential learning (BEL) is a successful framework for prediction and uncertainty quantification in subsurface reservoirs. The ability of BEL to predict a cyclic push-pull test using another single-well experiment with different signal amplitudes and durations is illustrated. The whole process is decomposed in 4 steps, relatively simple to implement: definition and sampling of the prior model, global sensitivity analysis, prior model falsification and prediction. Every step is illustrated using the reported field experiment.

Although the framework is stochastic, it does not require heavy computations. Indeed, BEL is based on the analysis of model responses (data and prediction) using a limited number of prior realizations. Data and prediction being relatively simple, the number of models is limited to 500 in this case. This signifies that only 1000 forward groundwater flow and heat transport runs are necessary to successfully assess the posterior distribution. All the models are independent, avoiding any time-consuming procedure as encountered in deterministic calibration or stochastic inversion, but allowing for parallelization.

The key for a successful application of BEL is the definition of the prior model. It should encompass all information available on the study site to derive realistic ranges of uncertainty for each sensitive parameter. On one hand, ignoring components of uncertainty might yield unrealistic uncertainty estimation. On the other hand, an unrealistic large uncertainty range might complicate the data-prediction relationship and reduce its accuracy. The prior model falsification and the prediction steps use tools allowing one to easily diagnose such kind of problems, as illustrated by this case study.

Those characteristics make BEL an ideal candidate for the introduction of uncertainty quantification in real-life applications and within practitioners. The demonstration of the ability of the framework to work in real field conditions should open a new range of perspectives and applications of the method.

Acknowledgement

We thank Thomas Kremer, Maxime Evrard and Solomon Eshioke for their precious help in the field and Frédéric Nguyen and Jef Caers for fruitful discussions. Field experiments were possible thanks to the F.R.S.-FNRS research credit 4D Thermography, grant number J.0045.16. N.

Lesparre and G. De Schepper were supported by the project SUITE4D from the BEWARE Fellowships Academia Program (contract No. 1510466) and the SMARTMODEL project from the BEWARE Fellowships Industry Program (contract No. 1610056), respectively. Both programs are co-financed by the department of Research Programs of the Wallonia-Brussels Federation and Marie Skłodowska-Curie COFUND program of the European Union. We thank the associated editor and 4 anonymous reviewers for their constructive comments.

References

Aquilina, L., Vergnaud-Ayraud, V., Les Landes, A.A., Pauwels, H., Davy, P., Pételet-Giraud, E., Labasque, T., Roques, C., Chatton, E., Bour, O., Ben Maamar, S., Dufresne, A., Khaska, M., La Salle, C.L.G., Barbecot, F. (2015). Impact of climate changes during the last 5 million years on groundwater in basement aquifers. Scientific Reports 5:14132. https://doi.org/10.1038/srep14132

Bowman, A.W., Azzalini, A. (1997). Applied smoothing techniques for data analysis, Oxford Statistical Science Series. Oxford University Press, New York.

Brouyère, S. (2001). Etude et modélisation du transport et du piégeage des solutés en milieu souterrain variablement saturé (study and modelling of transport and retardation of solutes in variably saturated media) (PhD Thesis). University of Liege, Liege.

Chen, M., Izady, A., Abdalla, O.A., Amerjeed, M. (2018). A surrogate-based sensitivity quantification and Bayesian inversion of a regional groundwater flow model. Journal of Hydrology 557:826–837. https://doi.org/10.1016/j.jhydrol.2017.12.071

Fenwick, D., Scheidt, C., Caers, J. (2014). Quantifying Asymmetric Parameter Interactions in Sensitivity Analysis: Application to Reservoir Modeling. Mathematical Geosciences 46:493–511. https://doi.org/10.1007/s11004-014-9530-5

Goovaerts, P. (1997). Geostatistics for natural resources evaluation, Applied geostatistics series. Oxford University Press, New York.

Haggerty, R., Schroth, M.H., Istok, J.D. (1998). Simplified method of “push-pull” test data analysis for determining in situ reaction rate coefficients. Ground Water 36:314–324. https://doi.org/10.1111/j.1745-6584.1998.tb01097.x

Hermans, T. (2017). Prediction-Focused Approaches: An Opportunity for Hydrology. Groundwater 55:683–687.

Hermans, T., Irving, J. (2017). Facies discrimination with ERT using a probabilistic methodology: effect of sensitivity and regularization. Near Surface Geophysics 15:13–25.

Hermans, T., Nguyen, F., Caers, J. (2015a). Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study. Water Resources Research 51:5332–5352. https://doi.org/10.1002/2014WR016460

Hermans, T., Wildemeersch, S., Jamin, P., Orban, P., Brouyère, S., Dassargues, A., Nguyen, F. (2015b). Quantitative temperature monitoring of a heat tracing experiment using cross-borehole ERT. Geothermics 53:14–26. https://doi.org/10.1016/j.geothermics.2014.03.013

Hermans, T., Oware, E.K., Caers, J. (2016). Direct prediction of spatially and temporally varying physical properties from time-lapse electrical resistance data. Water Resources Research 52:7262–7283.

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A., Caers, J. (2018). Uncertainty Quantification of Medium-Term Heat Storage From Short-Term Geophysical Experiments Using Bayesian Evidential Learning. Water Resources Research 54:2931–2948. https://doi.org/10.1002/2017WR022135

Hou, Z., Rubin, Y. (2005). On minimum relative entropy concepts and prior compatibility issues in vadose zone inverse and forward modeling. Water Resources Research, 41, WR004082. https://doi.org/10.1029/2005WR004082

Jamin, P., Brouyère, S. (2018). Monitoring transient groundwater fluxes using the finite volume point dilution method. Journal of Contaminant Hydrology, in press. https://doi.org/10.1016/j.jconhyd.2018.07.005

Kammen, D.M., Sunter, D.A. (2016). City-integrated renewable energy for urban sustainability. Science 352:922–928.

Klepikova, M., Wildemeersch, S., Hermans, T., Jamin, P., Orban, P., Nguyen, F., Brouyère, S., Dassargues, A. (2016). Heat tracer test in an alluvial aquifer: Field experiment and inverse modelling. Journal of Hydrology 540:812–823. https://doi.org/10.1016/j.jhydrol.2016.06.066

Lesparre, N., Robert, T., Nguyen, F., Boyle, A., Hermans, T. (2019). 4D electrical resistivity tomography (ERT) for aquifer thermal energy storage monitoring. Geothermics 77:368–382.MacDonald, A.M., Bonsor, H.C., Ahmed, K.M., Burgess, W.G., Basharat, M., Calow, R.C., Dixit, A., Foster, S.S.D., Gopal, K., Lapworth, D.J., Lark, R.M., Moench, M., Mukherjee, A., Rao, M.S., Shamsudduha, M., Smith, L., Taylor, R.G., Tucker, J., van Steenbergen, F., Yadav, S.K. (2016). Groundwater quality and depletion in the Indo-Gangetic Basin mapped from in situ observations. Nature Geoscience, 9:762-766. https://doi.org/10.1038/ngeo2791

Paradis, C.J., McKay, L.D., Perfect, E., Istok, J.D., Hazen, T.C. (2018). Push-pull tests for estimating effective porosity: expanded analytical solution and in situ application. Hydrogeology Journal 26:381–393. https://doi.org/10.1007/s10040-017-1672-3

Park, J., Yang, G., Satija, A., Scheidt, C., Caers, J. (2016). DGSA: A Matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments. Computers & Geosciences 97:15–29. https://doi.org/10.1016/j.cageo.2016.08.021

Razavi, S., Tolson, B.A., Burn, D.H. (2012). Review of surrogate modeling in water resources. Water Resources Research 48:W07401. https://doi.org/10.1029/2011WR011527

Réseau National de Sites Hydrogéologiques. 2019. Network of hydrogeological research sites – Enigma – Data Hermalle. http://hplus.ore.fr/en/enigma/data-hermalle . Accessed 30 June 2018.

Satija, A., Caers, J. (2015). Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. Advances in Water Resources 77:69–81. https://doi.org/10.1016/j.advwatres.2015.01.002

Satija, A., Scheidt, C., Li, L., Caers, J. (2017). Direct forecasting of reservoir performance using production data without history matching. Computational Geosciences 21:315–333. https://doi.org/10.1007/s10596-017-9614-7

Scheidt, C., Jeong, C., Mukerji, T., Caers, J. (2015a). Probabilistic falsification of prior geologic uncertainty with seismic amplitude data: Application to a turbidite reservoir case. Geophysics 80:M89–M100. https://doi.org/10.1190/geo2015-0084.1

Scheidt, C., Renard, P., Caers, J. (2015b). Prediction-Focused Subsurface Modeling: Investigating the Need for Accuracy in Flow-Based Inverse Modeling. Mathematical Geosciences 47:173–191. https://doi.org/10.1007/s11004-014-9521-6

Scheidt, C., LI, L., Caers, J. (2018). Quantifying Uncertainty in Subsurface Systems, Geophysical Monograph Series. John Wiley and Sos & American Geophysical Union, Hoboken, NJ & Washington D.C.

Therrien, R., McLaren, R., Sudicky, E., Panday, S. (2010). HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater Simulation Group, Waterloo, ON, Canada.

Vandenbohede, A., Louwyck, A., Lebbe, L. (2009). Conservative Solute Versus Heat Transport in Porous Media During Push-pull Tests. Transport in Porous Media 76:265–287. https://doi.org/10.1007/s11242-008-9246-4

Wildemeersch, S., Jamin, P., Orban, P., Hermans, T., Klepikova, M., Nguyen, F., Brouyère, S., Dassargues, A., (2014). Coupling heat and chemical tracer experiments for estimating heat transfer parameters in shallow alluvial aquifers. Journal of Contaminant Hydrology 169:90–99. https://doi.org/10.1016/j.jconhyd.2014.08.001

Figure

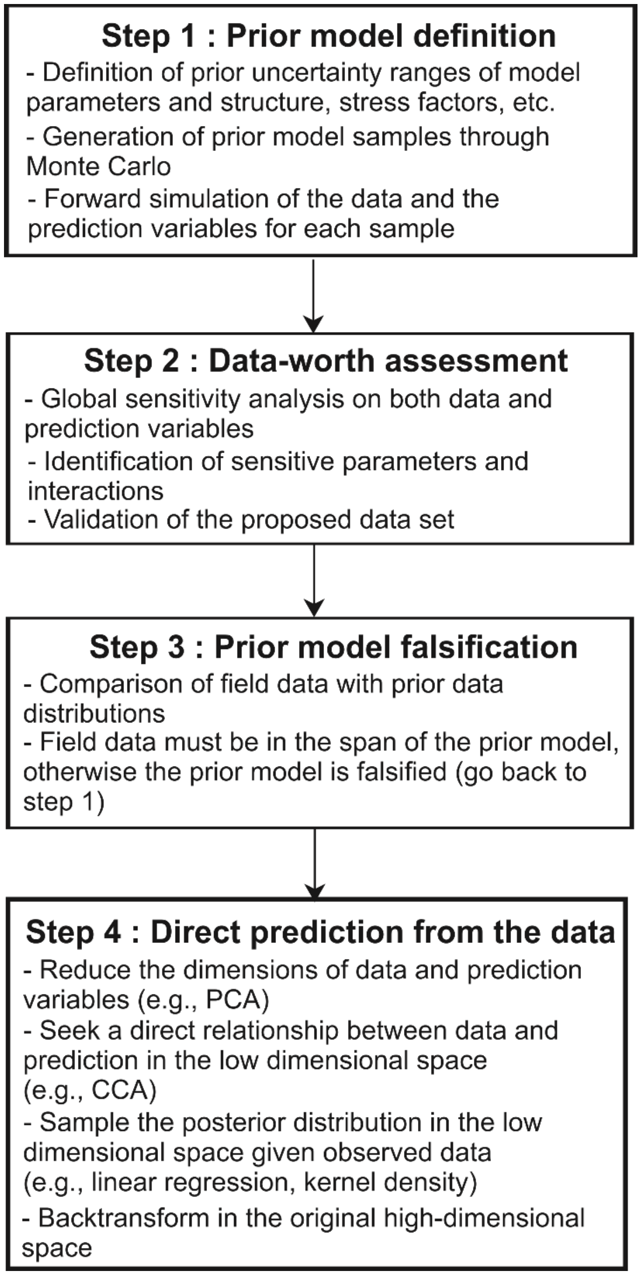


Figure 1. Flowchart of Bayesian evidential learning (BEL) framework as applied in this case study.

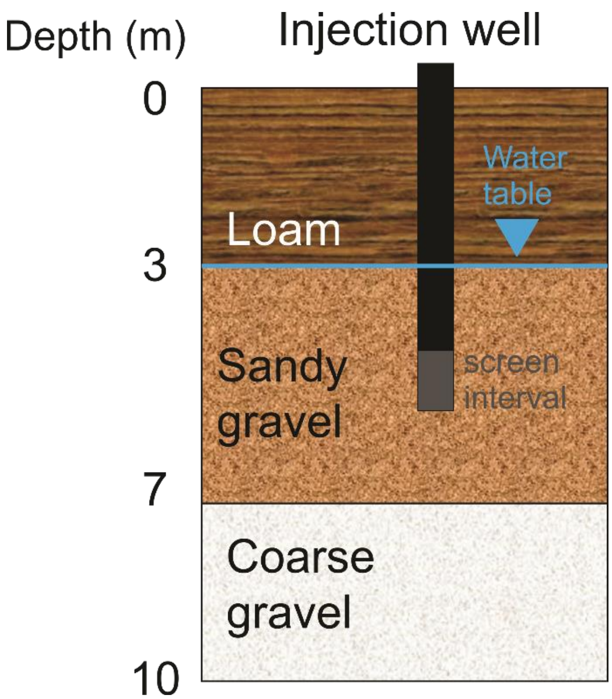


Figure 2. Hydrostratigraphic description of the study site located in the alluvial aquifer of the Meuse River in Hermalle-sous-Argenteau, Belgium.

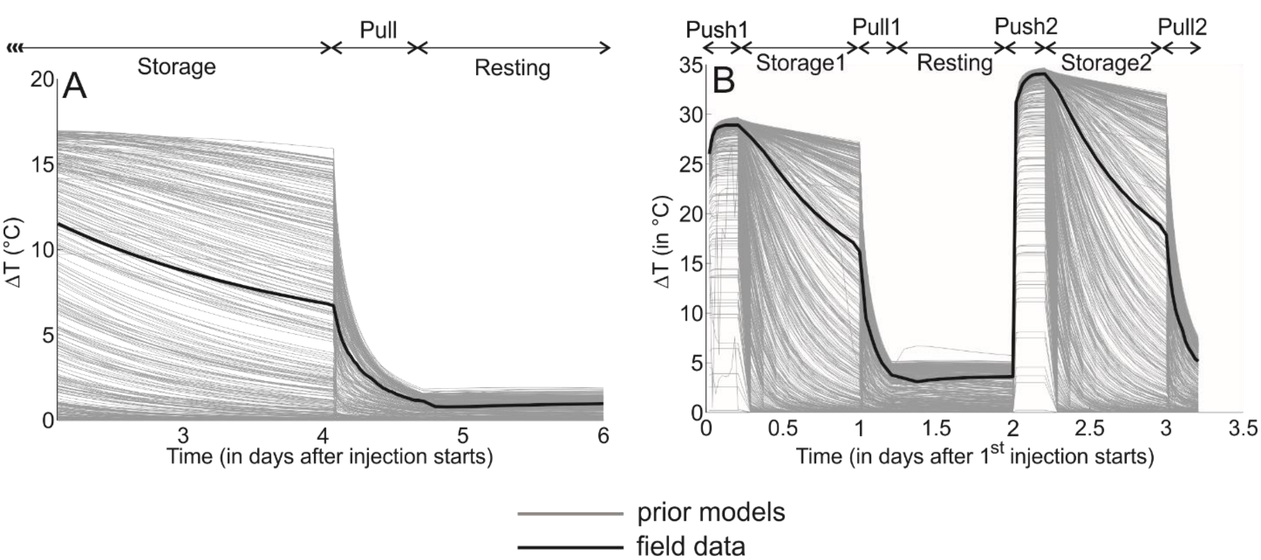
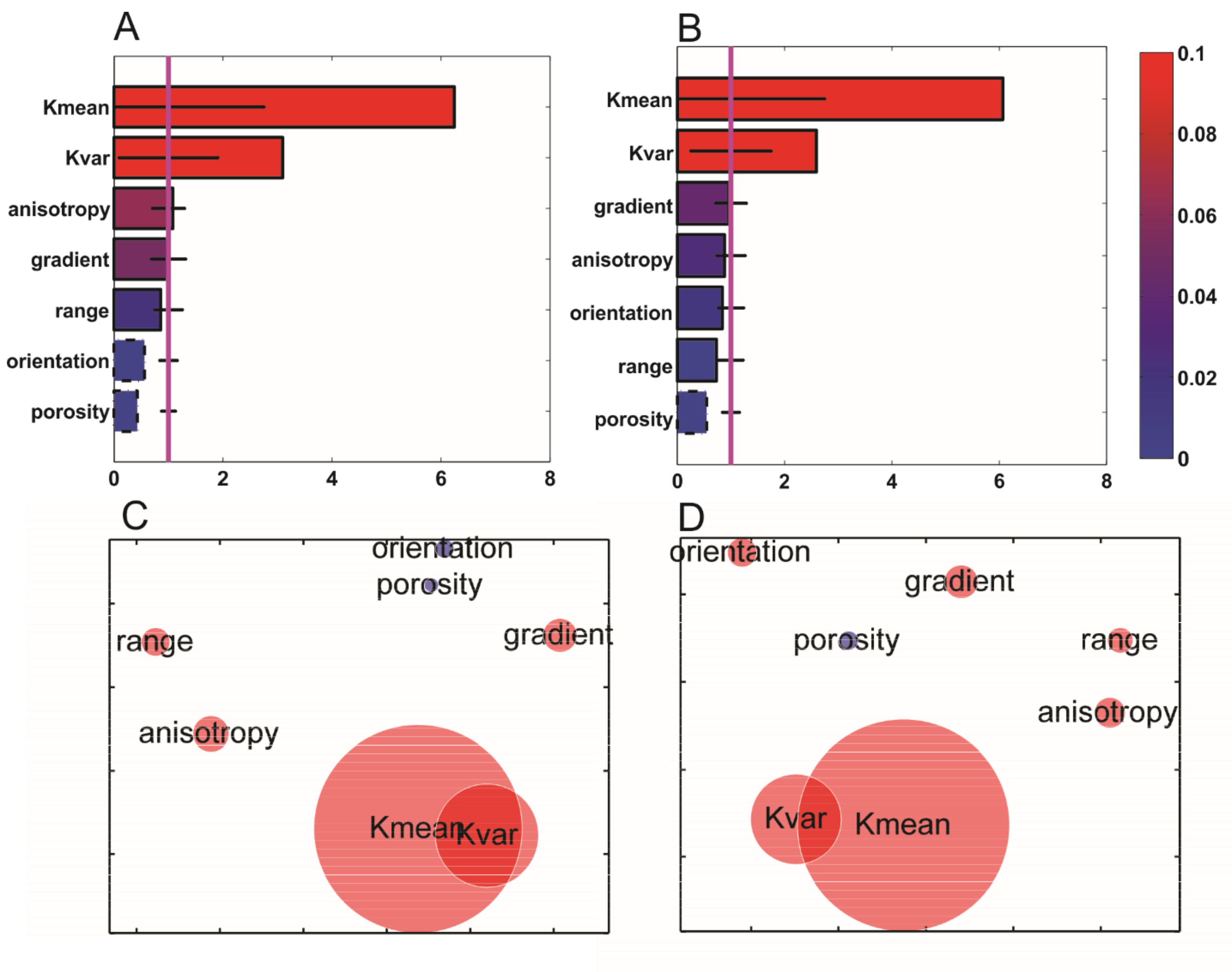


Figure 3. Prior model falsification for (a) the data and (b) the prediction. The observed curves are within the span of the prior, meaning that the prior is not falsified by the data.



573

574 Figure 4. Sensitivity analysis. The standardized sensitivity is similar for the (a) standard and (b)
575 cyclic experiment. The most sensitive parameters are the mean and variance of the hydraulic
576 conductivity. The respective interaction plots (c and d) also show similar patterns with an
577 interaction between mean and variance of the hydraulic conductivity. The closer the individual
578 bubbles are, the larger their interaction. The size of the bubble corresponds to the total effect (a
579 and b). On the interactions plot, red and blue colors correspond to globally sensitive and
580 insensitive parameters, respectively.

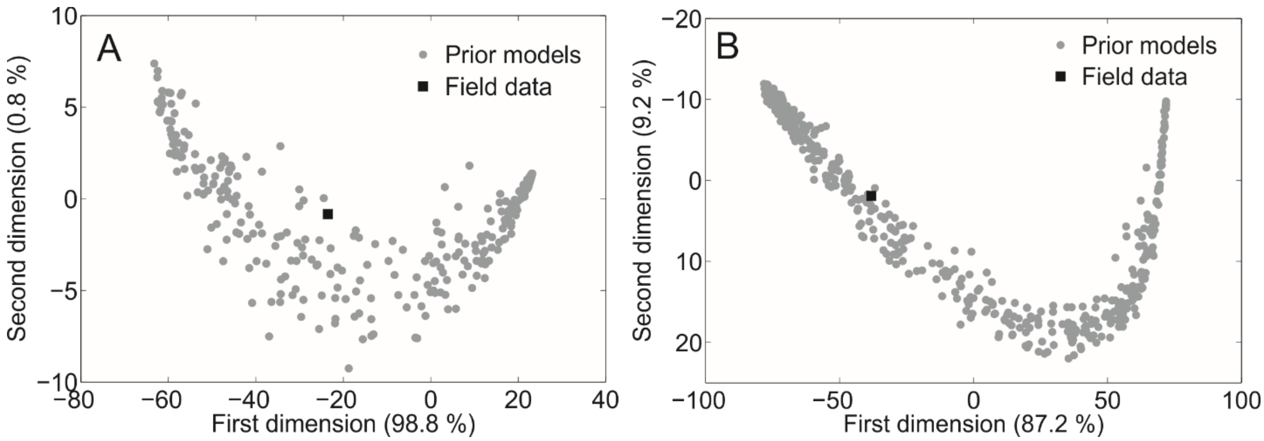


Figure 5. Prior model falsification in the reduced dimension space (PCA) for (a) the data and (b) the prediction. The field observations are within the span of the prior samples' responses, meaning that the prior model is not falsified.

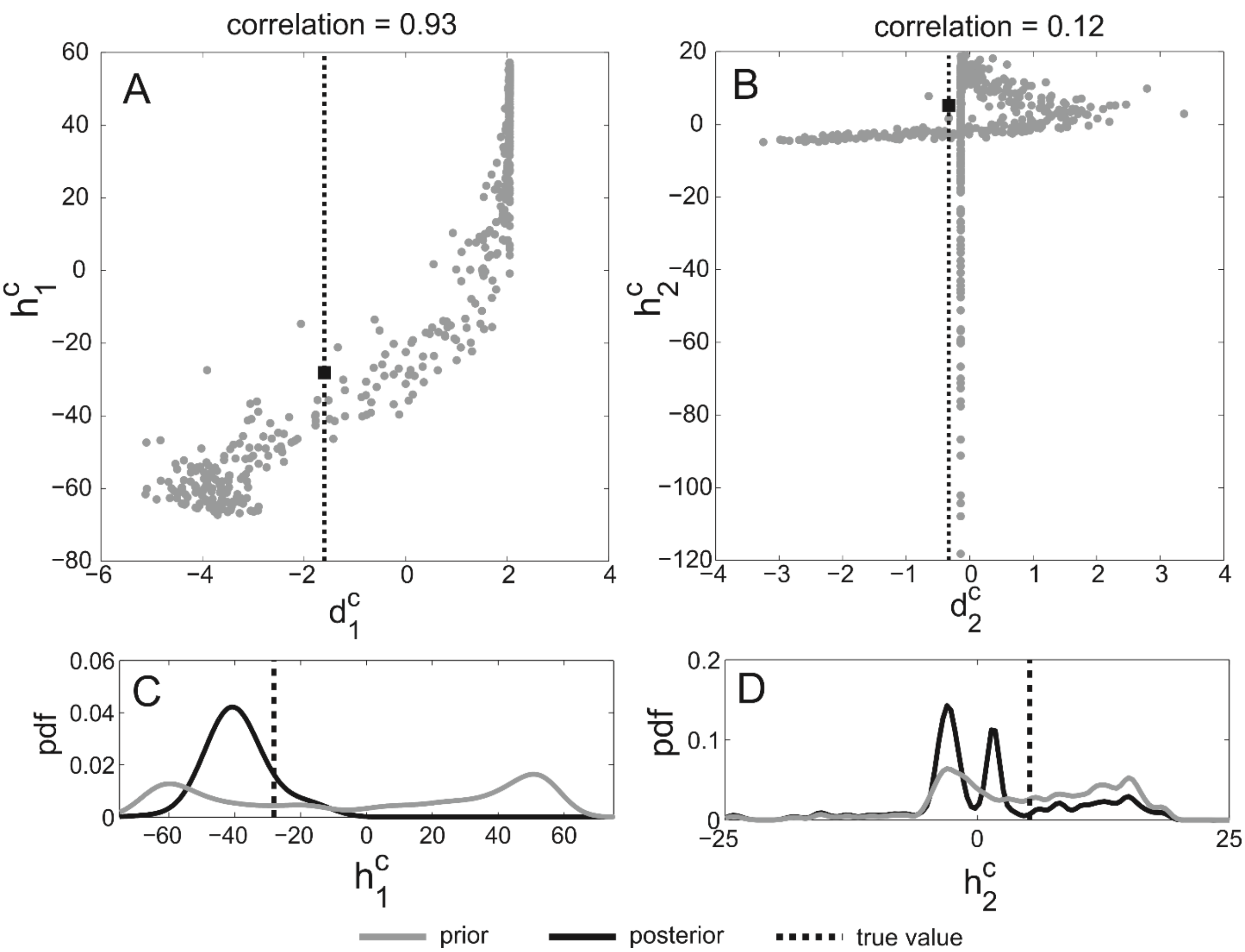


Figure 6. Canonical correlation analysis for forecasting the first two dimensions of the prediction (a, first dimension $\boldsymbol{h}_{\mathbf{1}}^{c}$; b, second dimension $\boldsymbol{h}_{\mathbf{2}}^{c}$) using the reduced data ($\boldsymbol{d}_{\mathbf{1}}^{c}$ and $\boldsymbol{d}_{\mathbf{2}}^{c}$). Grey points correspond to prior models, the black line to the field observation. Prior and posterior probability density function (pdf) of the (c) first and (d) second dimensions of the prediction.

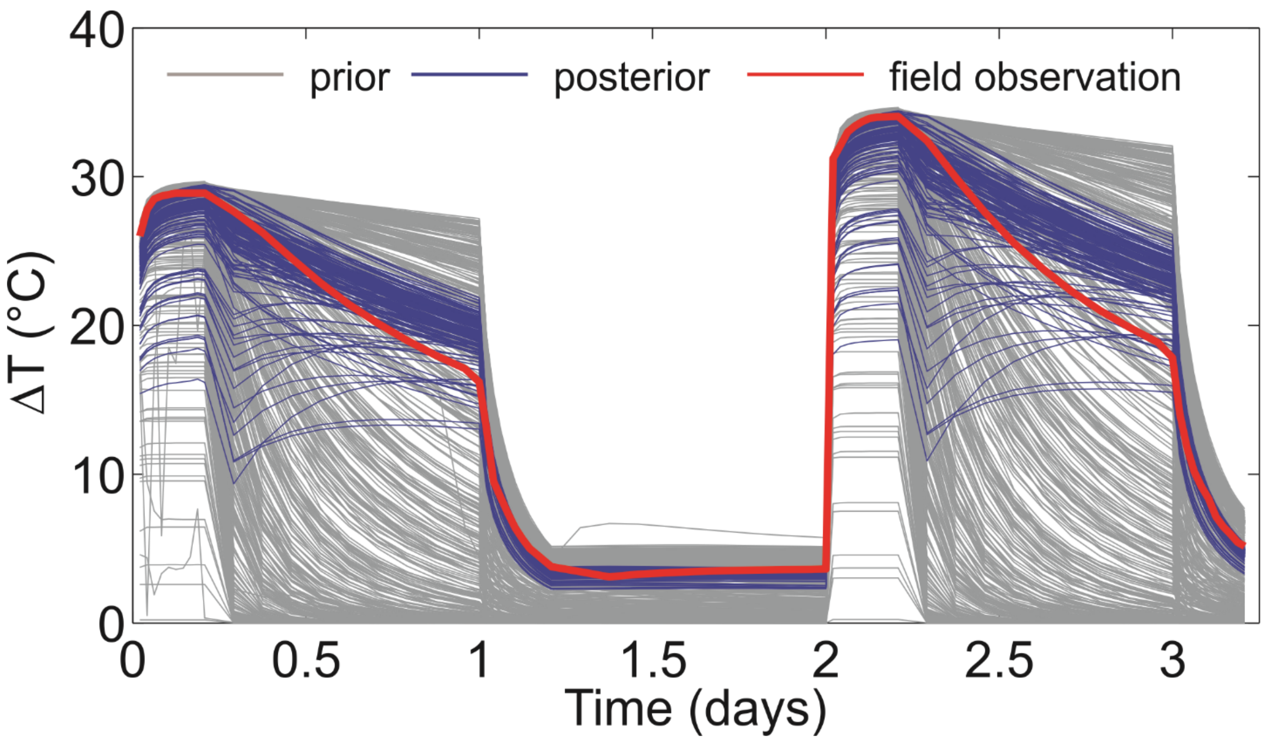


Figure 7. The posterior distribution of the prediction encompasses the field observation, demonstrating the ability of BEL to forecast the response of the aquifer for another solicitation.

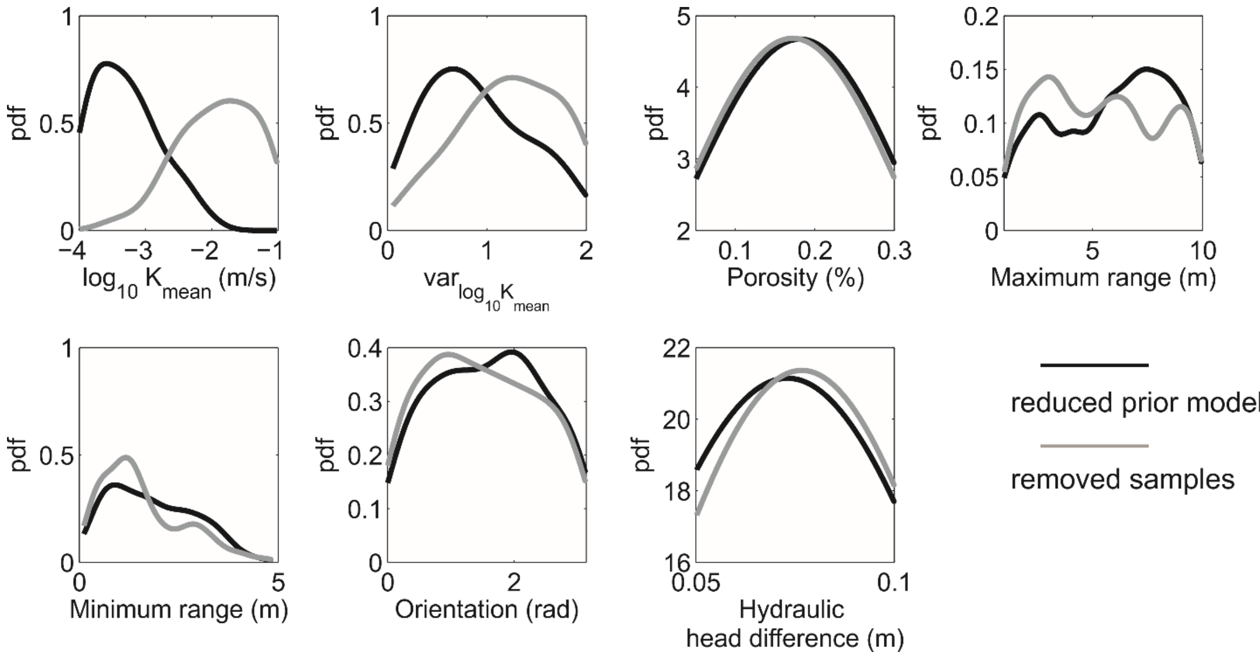


Figure 8. Distribution of model parameters in the reduced prior model and in the removed samples. Only the mean and variance of the hydraulic conductivity are significantly different.

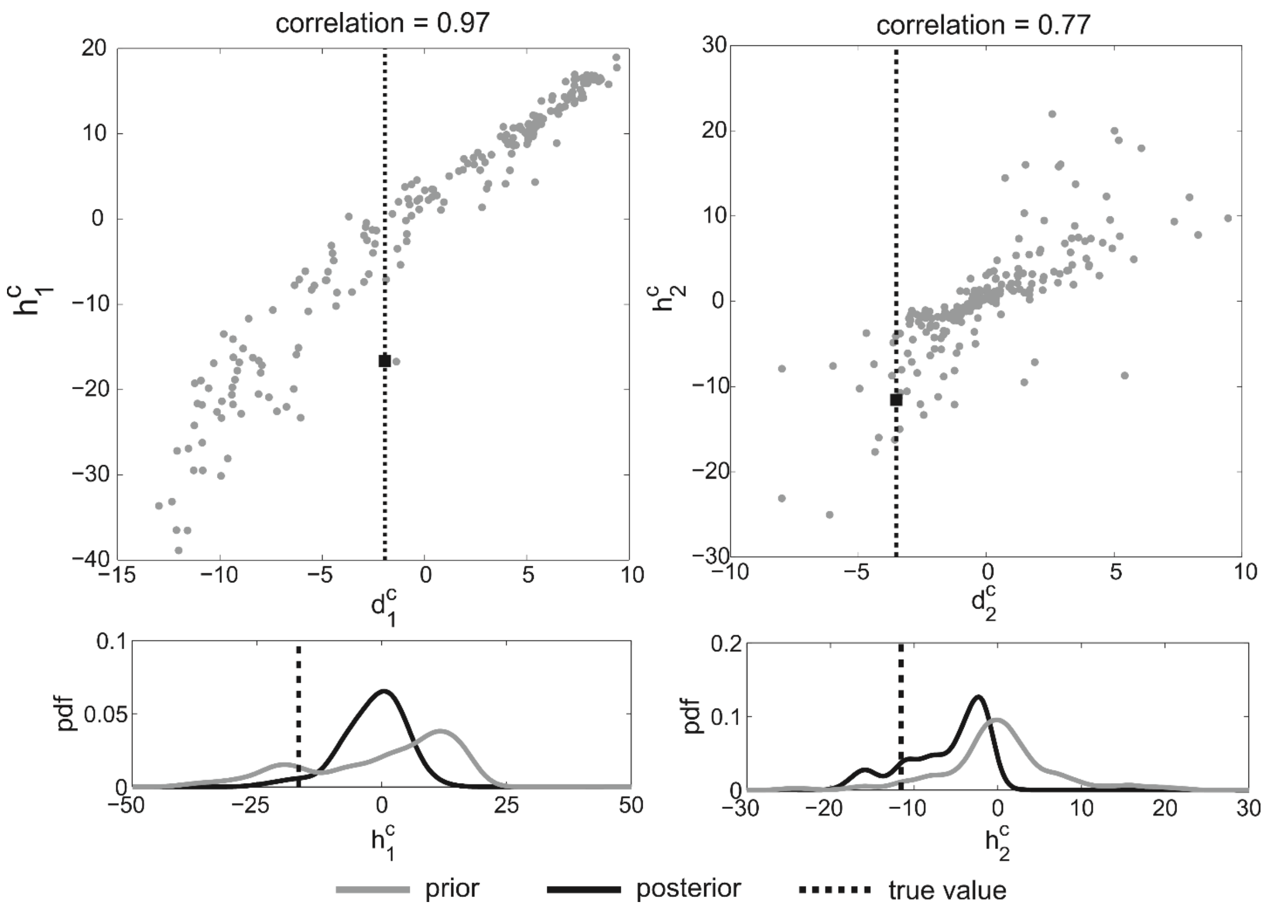
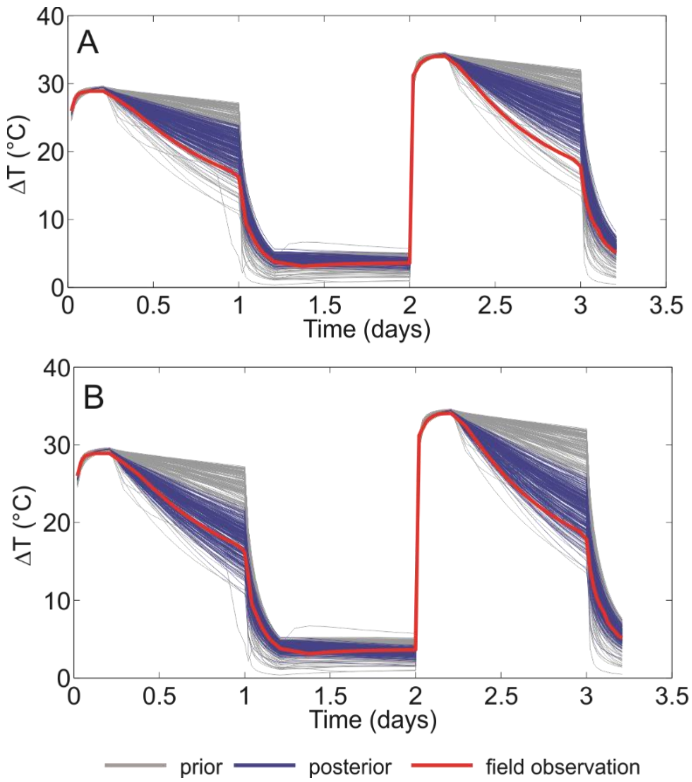


Figure 9. Canonical correlation analysis for forecasting the first two dimensions of the prediction (a, first dimension $\boldsymbol{h_1^c}$; b, second dimension $\boldsymbol{h_2^c}$) using the reduced data ($\boldsymbol{d_1^c}$ and $\boldsymbol{d_2^c}$). Grey points correspond to prior models, the black line to the field observation. The black square indicates the value of the true prediction. Prior and posterior probability density function (pdf) of the (c) first and (d) second dimensions of the prediction.



603

604 Figure 10. Posterior distribution of the prediction with (a) a reduced prior and (b) after correcting

605 the mean of the first dimension of the prediction.

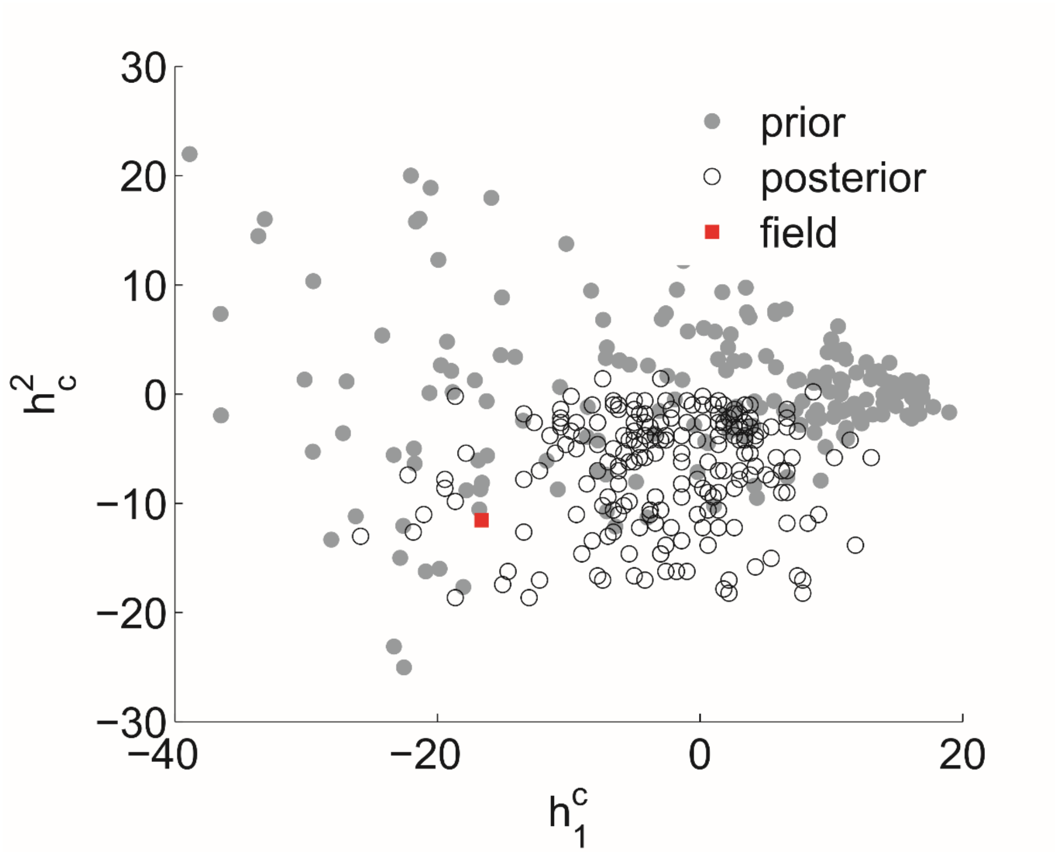


Figure 11. Prior and posterior score distributions in the CCA space. The field prediction is located in an area poorly sampled by prior realizations.