**PAPER**

# Comparison of different pilot point parameterization strategies when measurements are unevenly distributed in space

Ryma Aissat[1] · Alexandre Pryet[2] · Marc Saltel[3] · Alain Dupuy[2]

## Abstract

The parameterization of spatially distributed hydraulic properties is one of the most crucial steps in groundwater modeling. A common approach is to estimate hydraulic properties at a set of pilot points and interpolate the values at each model cell. Despite the popularity of this method, several questions remain about the optimum number and distribution of pilot points, which are determining factors for the efficiency of the method. This study proposes a strategy for optimal pilot point parameterization that minimizes the number of parameters while maximizing the assimilation of an observed dataset unevenly distributed in space. The performance of different pilot point distributions has been compared with a synthetic groundwater model, considering regular grids of pilot points with different spacings and adaptive grids with different refinement criteria. This work considered both prior and iterative refinements, with a parameter estimation step between successive refinements. The parameter estimation was conducted with the Gauss–Levenberg–Marquardt algorithm, and the strategies were ranked according to the number of model calls to reach the target objective function. The strategy leading to the best fit with the measurement dataset at the minimum computational burden is an adaptive grid of pilot points with prior refinement based on measurement density. This strategy was successfully implemented on a regional, multilayered groundwater flow model in the south-western geological basin of France.

**Keywords** Groundwater modeling · Parameter estimation · Pilot points · Heterogeneity · Hydraulic properties

## Introduction

During the last decades, numerical models have been extensively used to gain insights into aquifer system behavior. The interest of water managers in model-based decision support is still growing, and models are more than ever expected to guide the definition of sustainable management policies

✉ Ryma Aissat
r.aissat@brgm.fr

Alexandre Pryet
alexandre.pryet@bordeaux-inp.fr

Marc Saltel
m.saltel@brgm.fr

Alain Dupuy
alain.dupuy@bordeaux-inp.fr

[1] BRGM French Geological Survey, Water Dept. 3, avenue C. Guillemin, 45100 Orleans, France

[2] Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France

[3] BRGM French Geological Survey, 24 Av. Léonard de Vinci, 33600 Pessac, France

(Elshall et al. 2020). Groundwater models are necessarily a simplification of the geology, and hydraulic properties are most often heterogeneous and poorly constrained by direct measurements (Anderson et al. 2015). In this context, the precision and reliability of model-predicted values is a critical issue that justifies the need for practical and robust parameter estimation methods (Delottier et al. 2017).

A number of parameter estimation methods have been used in the last few decades (Doherty 2015; Zhou et al. 2014). Manual trial-and-error calibration that prevailed historically is gradually replaced by algorithmic methods such as the Gauss-Levenberg-Marquardt approach, which iteratively seeks the minimum error variance solution to the inverse problem from local finite-difference linearization of the model (Doherty 2015; Poeter et al. 2014). Such methods are widely used as they are easy to implement and effective for regularized inversion (Anderson et al. 2015); yet, they suffer from the computational burden associated with the estimation of the Jacobian matrix and the linear assumption is a limitation for uncertainty quantification. More recently, ensemble-based methods such as the Iterative-Ensemble-Smoother (IES) have been proposed as an alternative (White 2018). They also rely

on strong assumptions but present the advantage of joining the parameter estimation and uncertainty quantification steps. The relevance of these algorithms largely depends on the context of the application, data availability, and the purpose of the modeling exercise; however, in a large majority of cases, it is neither relevant nor practical to consider the value of hydraulic properties at each model cell as an independent, adjustable parameter. Therefore, a common and crucial step is choosing an appropriate parameterization method, which consists of the definition of a mapping between model properties and adjustable parameters to be estimated by the algorithm (Cui et al. 2021).

Zonation and 'pilot points' methods are the most common approaches for the parameterization of spatially distributed parameters. The former divides the model domain into a finite number of zones where parameters are assumed to be constant (Neuman 1973; Carrera 1986). Although some algorithms allow identifying the number of zones and the positions of the discontinuities (Hayek et al. 2009), this approach is still conditioned by prior knowledge of geological information and stratigraphy (Klaas and Imteaz 2017). As stated by Zhou et al. (2014), such parameterization may cause a structural error by introducing unrealistic, arbitrary discontinuities between zones. As an alternative to zones, the pilot points method introduced by de Marsily (1978) estimates the parameter values at a predetermined number of fixed points distributed in the model domain. Values at model cells are obtained by a smooth interpolation technique such as kriging (de Marsily 1984). A brief overview and history of the method has been recently proposed by White and Lavenue (2023).

Pilot points have been widely used for groundwater model calibration, and several practical concerns for implementing this approach have been widely discussed (Certes and de Marsily 1991; Doherty 2003; Doherty et al. 2010; Fienen et al. 2009). Pilot points generally lead to smooth isotropic fields, but recent studies have focused on the representation of hydraulic connectivity (Schilling et al. 2022; Trabucchi et al. 2022). Despite these advances, some issues are not fully addressed, particularly regarding the distribution of pilot points in the model domain, which can lead to subjective choices and diminish the feasibility or efficiency of the method. Choosing a small number of pilot points can lead to satisfactory identification of pilot-point parameter values with limited computational resources, but it decreases the level of detail and the description of heterogeneities and therefore misses the opportunity to explore some structural errors that could be made when conceptualizing the model (LaVenue and Pickens 1992). In contrast, using numerous pilot points can improve the description of heterogeneities, although this increases the need for computational resources. Additionally, using too many pilot points can lead to overparameterization where the model becomes too complex and

sensitive to noise in the data, which can result in numerical instability and difficulty in obtaining a unique solution to the inverse problem. Regularization can avoid the latter (Alcolea et al. 2006; Fienen et al. 2009); however, the numerical burden remains an issue. To date, a limited series of recommendations are available to define the appropriate number of pilot points and their suitable location. Gómez-Hernánez et al. (1997) recommend placing pilot points in a regular grid with a spacing of one-third of the variogram range. Wen et al. (2006) reported that placing pilot points with random methods yields better results. Doherty (2003) and Alcolea et al. (2006) suggested including as many pilot points as possible with the appropriate mathematical regularization and subspace methods (Christensen and Doherty 2008). It is not often mentioned that for real-world studies, measurements tend to be heterogeneously distributed in the model domain, so pilot points may be unevenly constrained if they are homogeneously distributed. Few authors discussed the relationship between pilot point placement and measurement availability, except Doherty et al. (2010), who recommended placing pilot points in areas with the highest data density (typically between wells in the groundwater flow direction). Klaas and Imteaz (2017) investigated the effects of pilot point density in a regular grid and discussed the interest in placing pilot points based on the experimental groundwater head contours. More recently, Baalousha et al. (2019) highlighted the effect of pilot point numbers and configurations on calibrated parameters, and Kapoor and Kashyap (2021) proposed a hybrid placement method based on both hydraulic head measurements and transmissivity estimates derived from pumping tests.

Despite the studies already mentioned, configuring pilot points remains challenging and mainly constrained by practical concerns and manual placement. This paper aims to define a reproducible strategy that would minimize the number of pilot points, save computational resources, and maximize the fit to measured data for the heterogeneities to be described at best, given measurements unevenly distributed in space.

This study investigates the efficiency of parameterization for both regular and adaptive grids of pilot points, where the density of pilot points is increased in the function of refinement criteria. The following questions arise: What is the optimum spacing for regular grids of pilot points? Can irregular, adaptive grids of pilot points improve the calibration process? If so, what are the optimum refinement criteria for these grids? And eventually, should iterative refinements of pilot points be considered? This paper addresses these issues by performing a series of tests on a synthetic model. The best approach is eventually implemented in a real-world case study: the multilayered groundwater model of the south-western geological basin of France.

## Methods

### Parameter estimation

The calibration has been conducted with the Gauss–Levenberg–Marquardt Algorithm (GLMA) as implemented in PEST_HP (Doherty 2020). The Jacobian matrix is computed at each iteration to form a linearized version of the model, containing parameter sensitivities to each measured observation. Sensitivities can be estimated with a classical 2-point finite difference approach or a more robust 3-point parabolic approximation (Doherty 2015). The former approach requires $m+1$ model calls, while the second, more demanding method requires $(2m+1)$ model calls for each iteration, with $m$ being the number of adjustable parameters. The computational burden associated with the computation of the Jacobian matrix increases with the number of adjustable parameters, which constrains the application of the GLMA to highly parameterized models.

At each iteration of the calibration process, the algorithm upgrades the parameter vector to reduce the misfit between model outputs and the measurements by minimizing the measurement objective function corresponding to the sum of weighted squared residuals, $\phi_m = \sum_i \left( w_i r_i \right)^2$, where the residual $r_i$ refers to the difference between the $i$th model output and its observed counterparts and the weights ($w_i$) are assigned based on the inverse of the standard deviation of measurement error. With this weighting scheme and assuming a perfect model, the target value for $\phi_m$ is the number of measurements. For real-world cases, the occurrence of model structural error entails that $\phi_m$ converges to higher values (Doherty 2015). The closer the final objective

function value is to the target objective function value, the better the calibration is in terms of fitting of measured data.

This study employed the truncated singular value decomposition (SVD) and Tikhonov regularizations to stabilize the inversion of the underdetermined problem and guarantee the convergence to a unique solution (Tonkin and Doherty 2005). Zero-order (preferred value) and first-order (preferred homogeneity) Tikhonov regularizations were both employed (Tikhonov and Arsenin 1977; Doherty 2015). The former attempts to draw estimated pilot point values toward their initial (prior) value. The latter promotes homogeneity between neighboring pilot point values by introducing the weighted difference between pilot point values in the regularization objective function. The termination criteria were chosen for the algorithm to stop when the reduction in the objective function is less than 1% for the last three iterations, or when the total number of iterations exceeds 30.

### Pilot point parameterization

In this approach, pilot points are placed according to a grid, which is necessarily coarser than the model grid and should not be confused with it. Two different strategies for pilot point placement have been considered: (1) regular grids with uniform spacings and (2) adaptive grids with local refinements (Fig. 1). In addition, an iterative strategy was explored, involving the estimation of parameters between successive refinements of the pilot point grids. The refinement is conducted following a 2D quadtree style, which allows cells to be subdivided repeatedly into four child cells. It should be noted that the parent pilot points are kept throughout refinement steps to facilitate the implementation of the iterative approach. Four
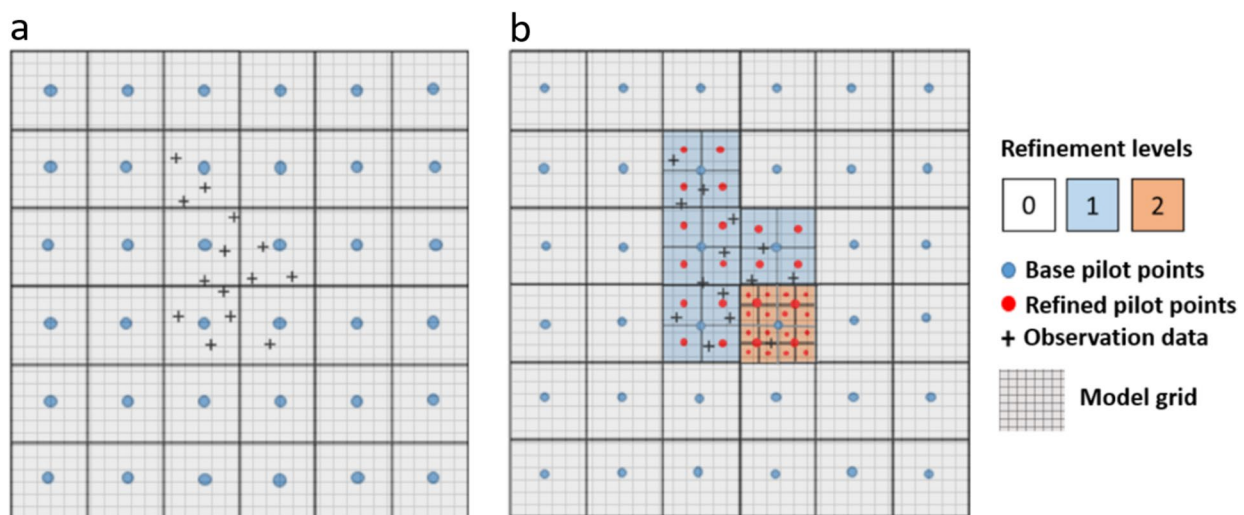


**Fig. 1 a** Base pilot-points distribution according to a regular grid and **b** adaptive parameterization where pilot points constrained by measurements are refined following the quadtree grid method. Cells with black bold outlines are for the grid of pilot points, and cells with thin grey outlines are for the model grid

refinement criteria were evaluated for each pilot point: the composite scaled sensitivity, the component of the gradient of the objective function corresponding to the pilot point, the parameter identifiability, and the density of the measurement data, which corresponds to the number of measures in the cell of the pilot points grid (Fig. 1). The positioning of the pilot points was achieved using Python scripts.

The composite scaled sensitivities (CSS) describe the intensity of the control provided by a measured dataset over a given parameter (Hill and Tiedeman 2006):

$$\mathrm{css}_i = \left( \left( \frac{\left(\mathbf{J^t Q J}\right)_{ii}^{\frac{1}{2}}}{n} \right) \right) \times | \mathbf{b}_i | \tag{1}$$

where $\mathbf{J}$ is the Jacobian matrix, $\mathbf{Q}$ is the measurement weight matrix, $\mathbf{b}$ is the vector of (transformed) parameter values, and $n$ is the number of nonzero-weighted observations (Doherty 2015). The component of the gradient of the objective function pertaining to the $i$-th parameter can be expressed as follows (Doherty 2015):

$$\frac{\partial \phi}{\partial b_i} = 2 \left[ \mathbf{J^t Q r} \right]_i \tag{2}$$

where $r$ is the vector of the model to measurement residuals. $\frac{\partial \phi}{\partial b_i}$ is the partial derivative of the objective function $\phi$ with respect to a parameter $b_i$.

The parameter identifiability, $f_i$ corresponds to the cosine of the angle between the vector pointing in the parameter's direction and its projection onto the calibration solution space (Doherty and Hunt 2009). It ranges between 0 for complete nonidentifiability and 1 for complete identifiability. It can be used to describe the capability of a calibration dataset to constrain a parameter value as it accounts for both parameter sensitivity and correlation. The computation of $f_i$ is conducted after singular value decomposition (SVD) of the weighted Jacobian matrix and subdivision of the parameter space into a "solution" and a "null" space. The dimension of the solution space is considered optimal when the introduction of additional eigenvectors leads to an increase (rather than a decrease) in the associated error variance (Doherty 2015). This was performed with the PEST SUPCALC utility (Doherty 2019).

The adaptive approach is tested with refinements of the pilot points grid conducted prior to the calibration process, and in an iterative manner (Fig. 2). The iterative refinement procedure starts with the initial, regular grid of pilot points until the convergence of the objective function. The parameterization is then updated by refining the pilot points satisfying the refinement criteria. The calibration is then resumed with the new parameterization and the procedure continues until the objective function reaches the target value.
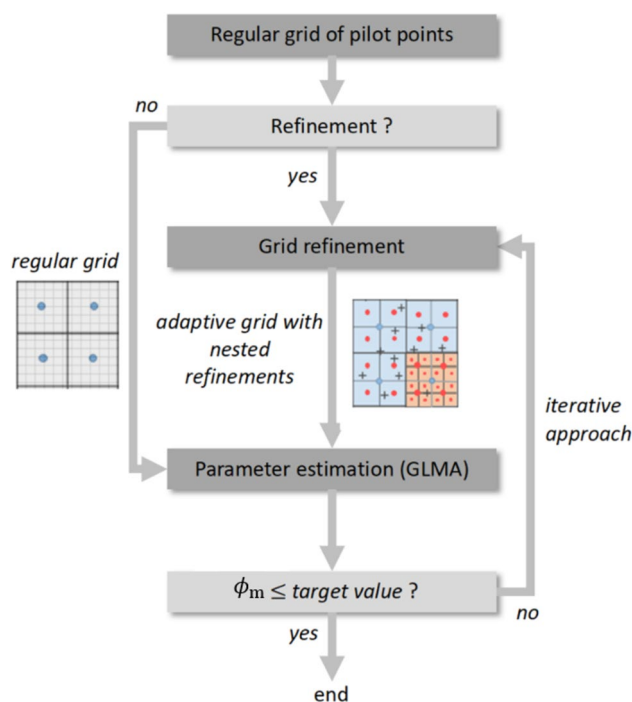


**Fig. 2** Summary of tested strategies for optimizing the pilot points configuration

The calibration efficiency has been evaluated for both regular and adaptive parameterization approaches. The comparison is conducted by analyzing the evolution of the objective function with respect to the number of model calls. The best configuration presents the fastest decrease of the objective function to the target value at the minimum calculation costs.

The interpolation of hydraulic conductivity at model cells was undertaken by kriging from pilot point values as implemented in the PyEMU Python package (White et al. 2016). For all the configurations considered in this study, the kriging factors were computed with Python considering an exponential variogram with a range of twice the largest spacing between pilot points. This as a ratio avoids the "bulls-eyes" effect in the interpolated field. The exponential variogram model is recommended when describing the heterogeneity of hydraulic properties because it avoids the occurrence of parasitic values between closely spaced pilot points of very different values (Doherty et al. 2010).

## Synthetic model

### Model description

Investigations are conducted on a synthetic model derived from Moore and Doherty (2005), which considers steady-state flow in a confined aquifer described with a single 10-m-thick layer over a domain of 500 m × 800 m, discretized

with a regular grid of 10 m square cells (Fig. 3). A fixed inflow of 1 m$^3$ day$^{-1}$ is imposed at the model upper boundary, and heads are fixed at 0 m along the lower boundary.

The reference hydraulic conductivity field was generated by Gaussian sequential simulation with Gstat (Pebesma 2004) using a log exponential variogram with a range of 100 m and a sill of 0.3. Flow within the domain was simulated using the finite-difference MARTHE model (Thiery 2015). Simulated heads were extracted at 26 observation wells unevenly distributed over the model domain to reflect real-world configurations. Indeed, observation networks typically present clusters with high measurement density and areas deprived of measurements. Gaussian noise was added to the simulated heads with a standard deviation of 0.01 m to introduce measurement error. Following the weighing strategy described in section 'Parameter estimation', an equal weight of 100 was assigned to each measure, corresponding to the inverse of the standard deviation of measurement error. In such conditions, the target measurement objective function corresponds to the number of measured data (here, 26).
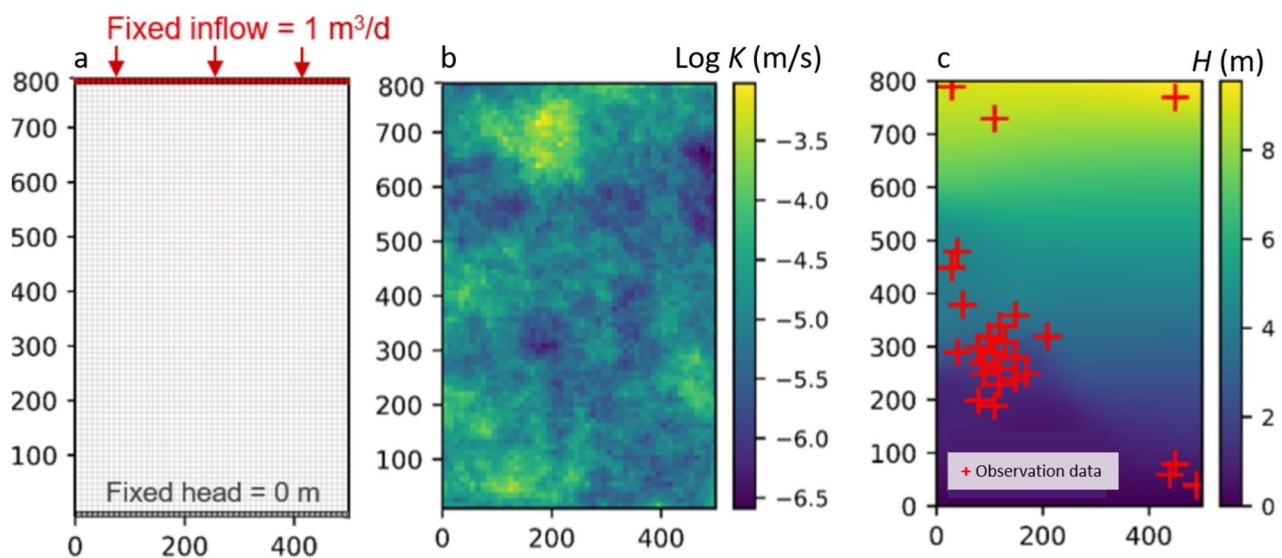
## Pilot points

The initial exploration involved regular grids of pilot points, covering spacings ranging from 24 to 3 model cells (Fig. 4). This resulted in varying numbers of pilot points, starting at 12 for the largest spacing (240 m) and reaching 456 for the shortest spacing (50 m).

Adaptive grids of pilot points were obtained by local refinements of the coarse regular grid with a spacing of 24 model cells with 12 pilot points (referred as "REG 24"). Different refinement strategies were tested with 4 different criteria

(refer to section 'Parameter estimation' for a description of these criteria). In all, 30% of the pilot points presenting the highest criteria values were subsequently refined to reach a resolution twice finer than the original grid. The number of pilot points after the refinement is 52, except for the refinement based on the measurement availability, where the total number is 40. There are fewer pilot points for this criterion due to successive refinements of the same points, which are discarded. The adaptive parameterization was tested from the beginning of the calibration (Fig. 5) and iteratively (Fig. 6). In the first approach, the grid of pilot points is refined according to the criteria inferred from the initial parameter values. In the second approach, the first calibration is conducted with a regular grid of pilot points, and the refinement criteria are then computed from the Jacobian matrix with updated parameter values. As for consequence, the set of pilot points to be refined may differ between the two approaches.

## Results

The calibration efficiency for the configurations of pilot points is evaluated by comparing the evolution of the measurement objective function with the number of model runs. The discussion begins with the results achieved using regular grids of pilot points (Fig. 7a). Large spacings between pilot points with 24 and 12 model cells (REG 24 and REG 12) led to fast convergence of the objective function value but did not wholly assimilate the measurements dataset. The objective function target value was reached with small spacings but with increased calculation times (REG 6 and REG 3). The best convergence of the objective function at minimum calculation costs is obtained with a spacing of 6 (REG 6) model



**Fig. 3** The synthetic model: **a** model domain and boundary conditions, **b** True hydraulic conductivity distribution, and **c** Simulated heads and locations of observation wells (red crosses)

**Fig. 4** Regular grids of pilot points with four different spacing (**a** REG 24, **b** REG 12, **c** REG 16, **d** REG 3 model cells). Measurements are depicted with red crosses and pilot points with black points. *N* is the total number of pilot points
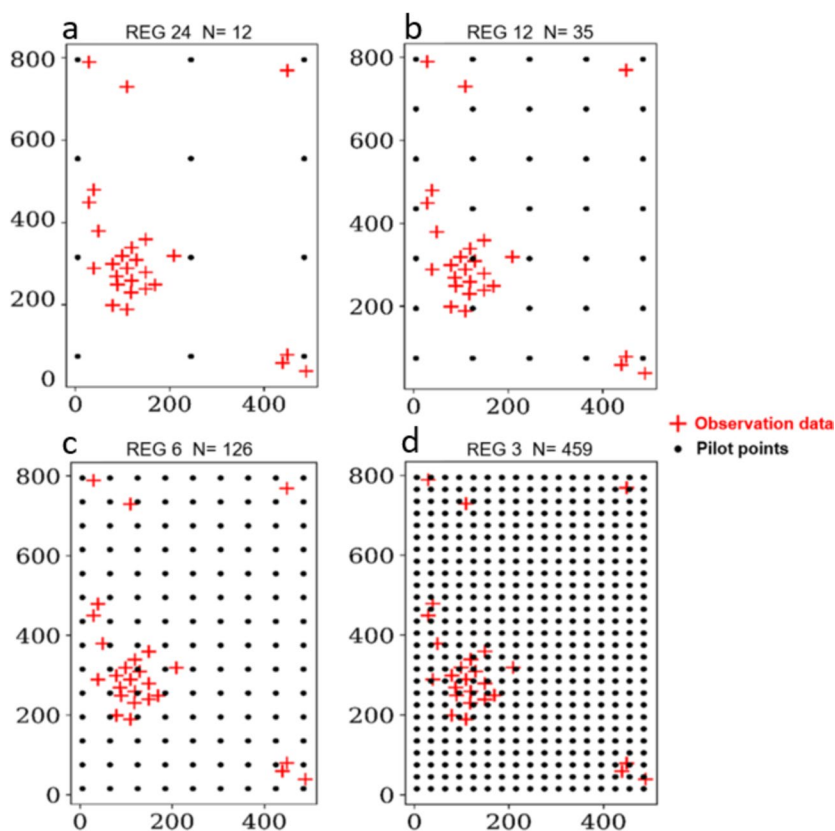


cells. Configurations with spacings of 3 (REG 3) model cells provide the same level of fit but require almost double model runs. This demonstrates the effect of pilot point numbers on fitting the measurements and the importance of choosing adequate spacing for the regular grid. Though it may be of interest for uncertainty analysis, including as many pilot points as possible for the calibration step increases the computational burden, which can be a limiting factor when dealing with highly parameterized models with long computation time.

The size of the solution space for the configurations with regular grids increases with the number of adjustable parameters (pilot points) but reaches a plateau beyond 126 pilot points (spacing of 6 grid cells; Fig. 8). Thus, when seeking to enhance the fit of measurements, it is unnecessary to increase the number of parameters beyond a certain threshold.

The performance of the initial adaptive configurations appears to be largely dependent on the choice of the refinement criterion (Fig. 7b). Refinements based on measurement availability and parameter identifiability lead to better performance than regular grids. In contrast, refinements based on CSS and the gradient of the objective function did not reach the target objective function and required more than 4,400 model runs for convergence.

For the adaptive approach (Fig. 7c), the refinements of the pilot points grid with the different criteria were conducted after calibration with the regular grid with a spacing of 24 model cells. From these results, the interest of the iterative approach

is not salient compared to the initial adaptive approach. Except for the refinement based on the number of measurements, the performance of the iterative adaptive strategy is poor compared to the best configuration with the regular grid.

The results for the three different approaches for pilot point placement (regular grid, initial adaptive grid, iterative adaptive grid) are summarized in Fig. 9. The best configurations with low objective function and a small number of model runs (bottom left portion of the plot) correspond to the initial adaptive approach based on measurement availability ($N=40$ pilot points). The regular grid with a spacing of 6 model cells also reaches the target objective function value, but at a much greater cost in terms of model runs. The hydraulic conductivity fields for these two "best" configurations are compared to the "reference" field in Fig. 10. As expected, only large-scale heterogeneities can be described with a better resolution where the measured dataset is dense. In areas poorly constrained by measurements, the pilot points of the regular grid present similar values, which leads to an outcome similar to the more parsimonious adaptive approach.

## Application to a regional flow model

The findings obtained with the synthetic model have been applied to estimate the hydraulic properties of the regional groundwater "MOdel of North Aquitania" (MONA). The

**Fig. 5** Initial adaptive grid with refinement of pilot points based on **a** measurements availability, **b** sensitivity, **c** identifiability, and **d** the gradient objective function. "Parent" pilot points (pp) are displayed as large grey points, "child" pilot points as small black points, and measurements as red crosses
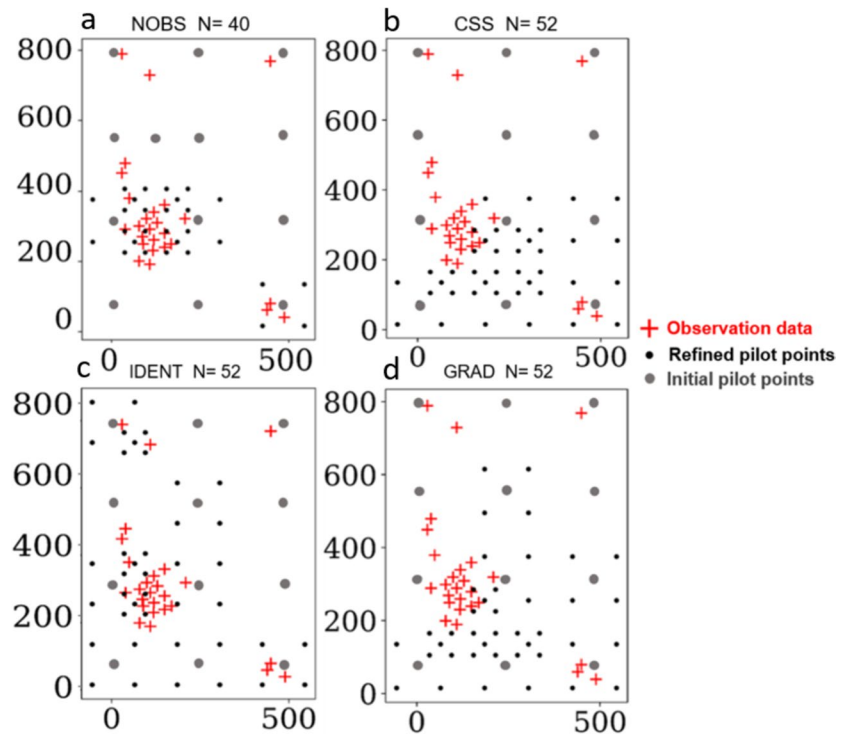


model was developed by the French Geological Survey (BRGM) to simulate flow and investigate the impact of pumping in the extensive unconfined aquifers supplying the city of Bordeaux. The model covers the northern part (46,032 km²) of the French south-west sedimentary basin (Fig. 11). It has

15 aquifers interbedded by aquitards discretized with a regular grid of 2 × 2 km from Plio-Quaternary down to Jurassic units (Thiéry et al. 2011). The model does not explicitly simulate flow in the aquitards but accounts for vertical flows adjusted by a conductance parameter (pseudo-3-D assumption). The

**Fig. 6** Iterative adaptive grids. "Parent" pilot points (pp) are displayed as large grey points, "child" pilot points as small black points, and measurements as red crosses. 30% of the initial pilot points are refined two successive times based on **a** measurements availability, **b** sensitivity, **c** identifiability, and **d** the highest value of the objective function's gradient
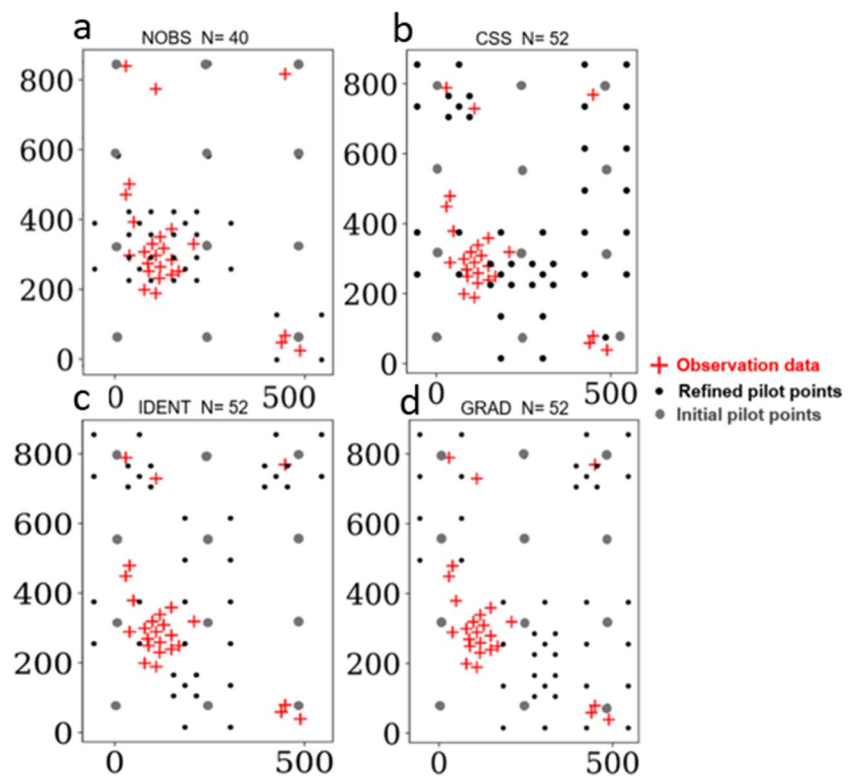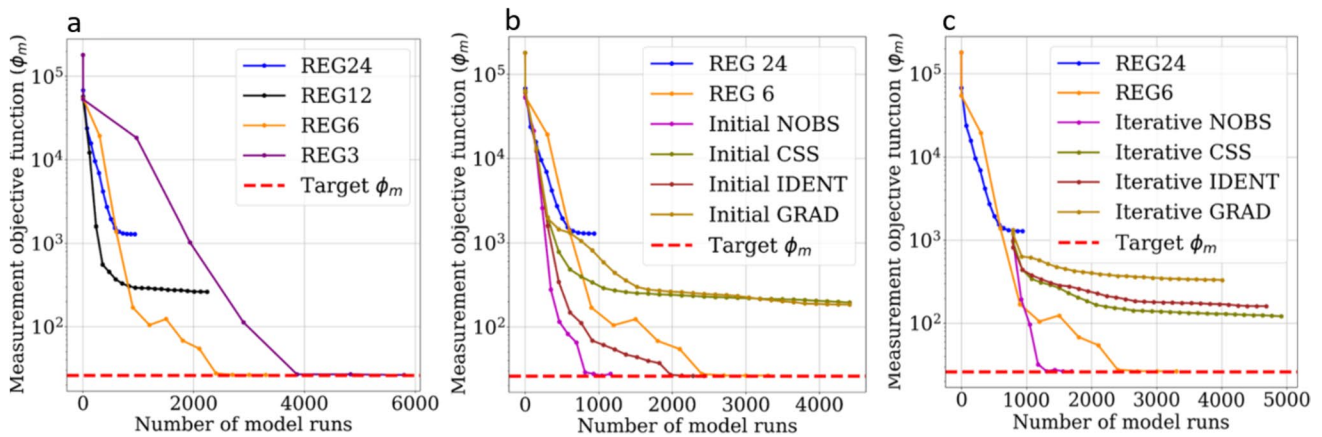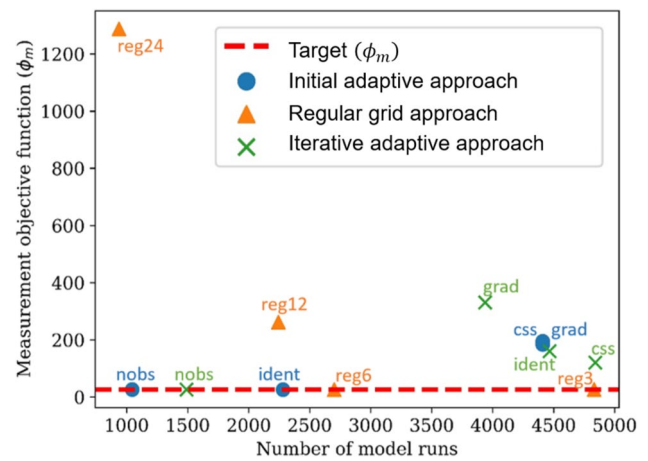
**Fig. 7** Evolution of the measurement objective function with the number of model runs for different pilot point parameterizations: **a** regular grids, **b** initial adaptive grids, and **c** iterative adaptive grids. Results of regular grids are also presented to facilitate the comparison (**b**–**c**)
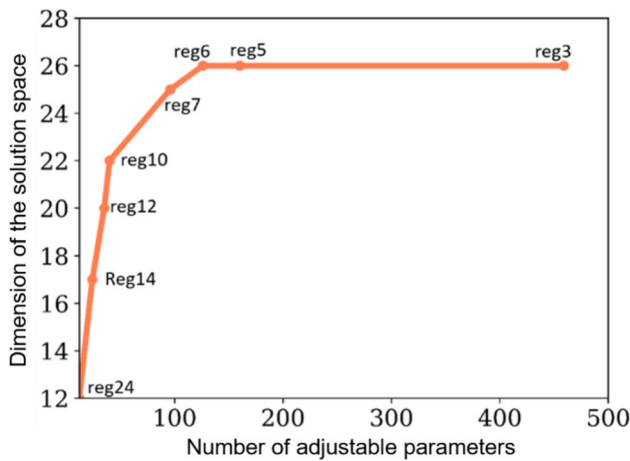


**Fig. 8** Evolution of the dimension of the solution space for the regular grids of pilot points with different spacing



**Fig. 9** Values of measurement objective functions against the total number of model calls at convergence for the regular, initial adaptive, and iterative adaptive parameterizations. The best parameterization is obtained with the initial adaptive approach (bottom left), which presents the lower number of model runs to reach the target objective function. "nobs" stands for the number of observations, "ident" represents identifiability criteria, "css" denotes the composite scaled sensitivity, "grad" signifies the gradient of the objective function, and "reg" refers to the regular grid, followed by the corresponding spacing. For example, "reg12" indicates that the spacing for the regular grid is 12

domain is bounded by Cretaceous and Jurassic outcrops to the east and north, the Atlantic Ocean and the Gironde Estuary to the west. Hydraulic heads are imposed on the western boundary, accounting for seawater level along the Atlantic Coast. Heads are also prescribed along the Garonne River and its estuary, and no-flow boundaries are assumed at the southern limit, which corresponds with the separation from the southern part of the Aquitaine basin (Buscarlet et al. 2019). Recharge is estimated annually by an empirical formula using climatic data (precipitation and evapotranspiration) from a series of weather stations (Pédron and Platel 2005). The pumping database includes 6,235 wells distributed within the 15 geological formations (Saltel et al. 2016). The diffusivity equation is solved at the annual time step in a finite volume scheme with MARTHE (Thiery 2015).

The purpose of this calibration exercise is to use historical head measurements to improve the predictive

capacity of the model for simulating heads in different prospective management scenarios over the next decades. The model was calibrated in the transient state over the 1972–2011 period using 423 observation wells. To assess the measurement noise and assign weights, two types of uncertainties were considered: the measurement uncertainty, $\sigma_m$, and the uncertainty raised from the aggregation of incomplete daily data at the annual time step, $\sigma_a$. Assuming a standard deviation of ±3 m at the 68% confidence level for measurement uncertainty, arising from all

potential errors in the measurements (groundwater depth and borehole leveling), the aggregation uncertainty was determined for each individual annual value by calculating the standard error of the mean (Hughes and Hase 2013).

The parameter estimation procedure focused on the distributed hydraulic properties: the horizontal hydraulic conductivity of aquifers, $K_h$, the vertical hydraulic conductivity of aquitards, $K_v$, the unconfined specific yield $\omega$, and the specific storage, $S_s$. Prior information on these parameters was first collected for each layer of the model from the French geological database (BSS) and several local studies (Moussié 1972; Hosteins 1982; Larroque 2004). These values were used as a starting point for the GLMA and were considered the preferred value for 0-order Tikhonov regularization.

## Parameterization

Both pilot points (PP) and zones of piecewise constancy (ZPC) were used for the parameterization of this regional model, which is summarized in Table 1. The specific yield ($\omega$) was parameterized with pilot points for the first layer (QUAT) with regular grids of pilot points with a spacing of 20 model cells (40 km). ZPC were used for layers 2–15 since most parts of these aquifers remain confined and the annual time step reduces the sensitivity of this parameter in unconfined areas. For the specific storage ($S_s$), pilot points were placed with a spacing of 40 model cells (80 km) in the permanently confined parts of the layers, and ZPC were used for the permanently unconfined parts, where $S_s$ are insensitive. The vertical hydraulic conductivity of aquitards was parameterized with a regular grid of pilot points with a spacing of 40 model cells (80 km). Maps with the locations

of pilot points are provided in all the items in the Appendix Figs. 14, 15, 16, 17, 18, 19 and Table 2.

Several configurations were considered for the parameterization of horizontal hydraulic conductivity ($K_h$). A coarse regular grid with a 20-model-cell spacing (40 km) and a finer regular grid with a 5-model-cell spacing (10 km) were considered. An adaptive grid of pilot points was also taken into account, employing measurement density as the refinement criterion. Pilot points with a minimum of 1 measurement within their neighboring values were refined twice, resulting in a 5-model-cell spacing (10 km) between pilot points in areas with a high measurement density and 20-model-cell spacing (40 km) in areas with sparse measurements. Both the initial and iterative adaptive refinement strategies were tested. In areas with high measurement density, the adaptive grids present the same spacing as the fine regular grid. It should be noted that spacings between pilot points are large (10–80 km), even with the local refinements. With such parameterizations, only the most salient heterogeneities impacting the regional groundwater flow can be described. This can be appropriate for the simulation of the long-term regional flow dynamics; however, this approach would not be relevant to describe small-scale local heterogeneities in the hydraulic property fields, which would require a much denser measurement dataset.

## Results

The calibration was conducted with three different strategies for the parameterization of hydraulic conductivities with pilot points: the coarse regular grid, the initial adaptive grid, and the adaptive grid with an iterative approach. Each model run took ~ 20 min of CPU time and calculations were
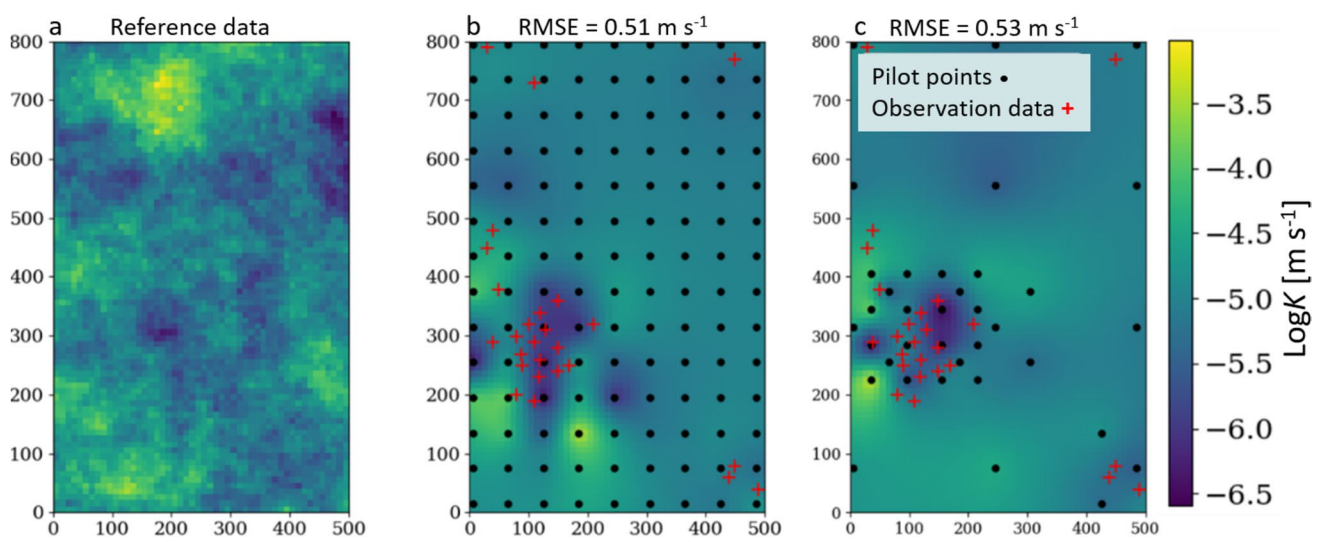


**Fig. 10** Comparison of the **a** "true" hydraulic conductivity field with **b** its estimated counterparts obtained with the best regular grid (pilot point spacing of six model cells), and **c** the adaptive grid refined according to the measurement density
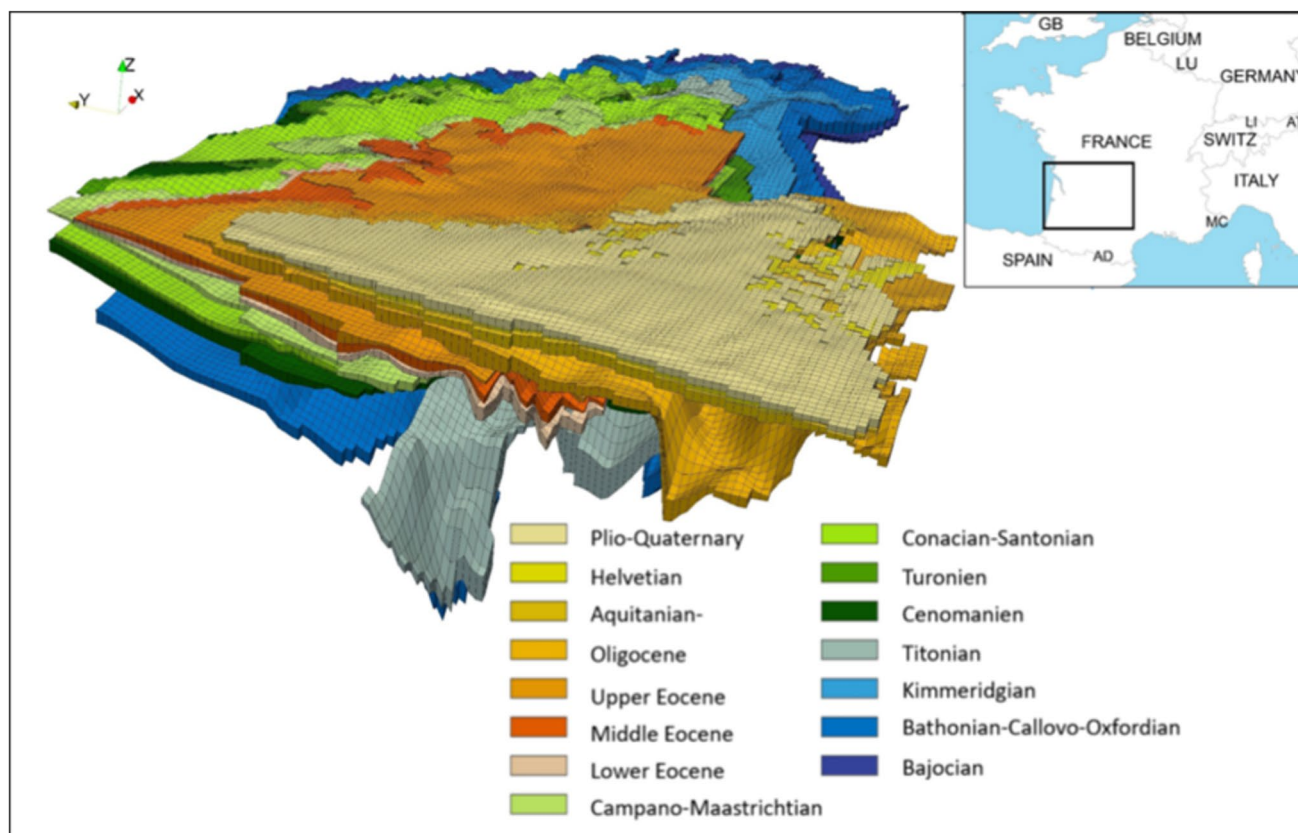
**Fig. 11** Representation of the aquifers considered in the MONA. The insert shows the location of the Aquitaine region in France

parallelized over 114 CPU cores. The fine regular grid of pilot points resulted in an excessive number of parameters and prohibitive computation time, so it could not be tested.

The evolution of the objective function for the tested configurations is presented in Fig. 12. As expected, the configuration with the coarse grid of pilot points converged to a relatively high value of the objective function, illustrating the incapacity of this parameterization to represent sufficiently fine details in the hydraulic conductivity field to satisfy the measurement dataset. The iterative adaptive approach performed well but with results close to the initial adaptive approach, which is easier to implement. The initial adaptive approach, which took a week and 56,393 model runs to converge, can therefore be considered the best approach.

The post-calibration hydraulic conductivity fields obtained with the "best" initial adaptive approach are provided in Fig. 13. The other estimated fields and performance statistics are provided in the Appendix Figs. 14, 15, 16, 17, 18, 19 and Table 2. Overall, simulated values reproduce their observed counterparts with a root mean squared error (RMSE) of 6 m and a mean bias of −0.16

m. The goodness of fit varies substantially between layers. Errors are the largest for the deep layers, for which the density of measurements is low. The misfit is more heterogeneous in those layers that include some outliers with high error values that affect the average. The outliers are mostly related to insufficient or inconsistent data to constrain the calibration; therefore, the algorithm is not focusing on their fits since the weight assigned is low.

## Discussion

The present study confirmed the importance of pilot point distribution on calibration efficiency and provides guidelines for optimizing the placement of pilot points when the computational burden matters, such as for highly parameterized regional models with long execution times. The purpose was to identify the configuration leading to the fastest convergence of the objective function to the target value, indicative of appropriate assimilation of the measured dataset.

Investigations were conducted on a synthetic model with both regular and adaptive grids of pilot points. The

**Table 1** Summary of the parameterization of the regional model

| Layers | | $K_h$ (m/s) | | | | $K_v$ (m/s) | | $S_s$ (m$^{-1}$) | | $\omega$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Par. | Spacing for coarse reg. grid | Spacing for fine reg. grid | Spacings for adapt. grid | Par. | Spacing | Par. | Spacing | Par. | Spacing |
| 1 | QUAT | PP | 20 | 5 | 20/5 | - | - | - | - | PP | 20 |
| 2 | HELV | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 3 | AQUI | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 4 | OLNP | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 5 | EOCS | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 6 | EOM | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 7 | EOCI | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 8 | CAMP | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 9 | COST | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 10 | TURO | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 11 | CENO | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 12 | TITH | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 13 | KIMM | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 14 | BACX | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |
| 15 | BAJO | PP | 20 | 5 | 20/5 | PP | 40 | PP/ZPC | 40 | ZPC | - |

*Par.* refers to the type of parameterization, *PP* for pilot points and *ZPC* for zones of piecewise constancy. Spacings are expressed in terms of model cells. Three configurations have been considered for the parameterization of hydraulic conductivity: a coarse regular grid, a fine regular grid, and an adaptive refinement based on measurement density. For the latter, both the maximum/minimum values of pilot point spacing are provided

first analysis with regular grids revealed the importance of choosing an appropriate spacing between pilot points. Large spacings lead to fast convergence of the objective function but do not allow complete assimilation of the measured dataset. In contrast, including as many pilot points as possible for calibration, as suggested by Alcolea et al. (2006), leads to a good fit of observed data but increases the computational burden. The optimum spacing, leading to the fastest convergence of the objective function to the target value was obtained with a pilot point spacing of 60 m and a variogram range of 120 m, which is close to the variogram used for the generation of the synthetic hydraulic conductivity field (100 m). Such a result is in line with the expectations for a synthetic case, but further questions remain when dealing with unknown, real-world hydraulic conductivity fields, for which the structure of heterogeneities that matter for the modeling exercise is not precisely known. An option can be to use the calibration dataset to investigate the size of solution space and get insights on the optimum number of pilot points by truncated singular value decomposition. This requires the computation of a Jacobian matrix for each configuration which may be computationally intensive. Furthermore, the size of the solution space suggested by this approach is often too large because its computations fail to account for the contribution made to measurement uncertainty by structural noise of unknown covariance structure (Doherty et al. 2010).

Results of the synthetic case also revealed that an adaptive approach with refinement based on measurement density leads to better performances than the optimum
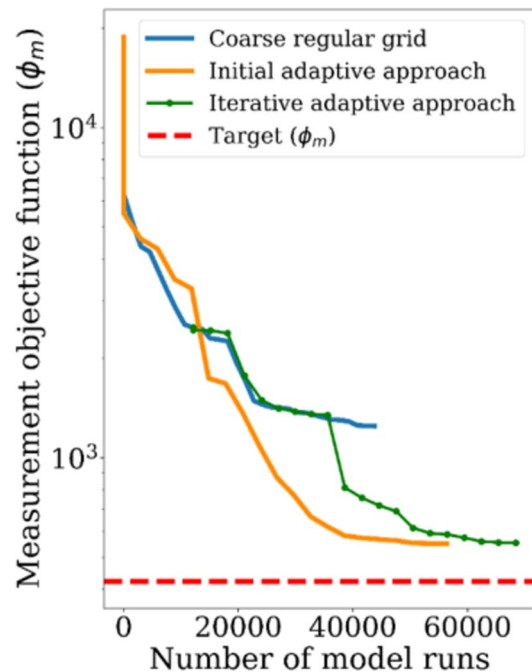


**Fig. 12** Summary of objective function vs. the number of model runs for regular and adaptive approaches
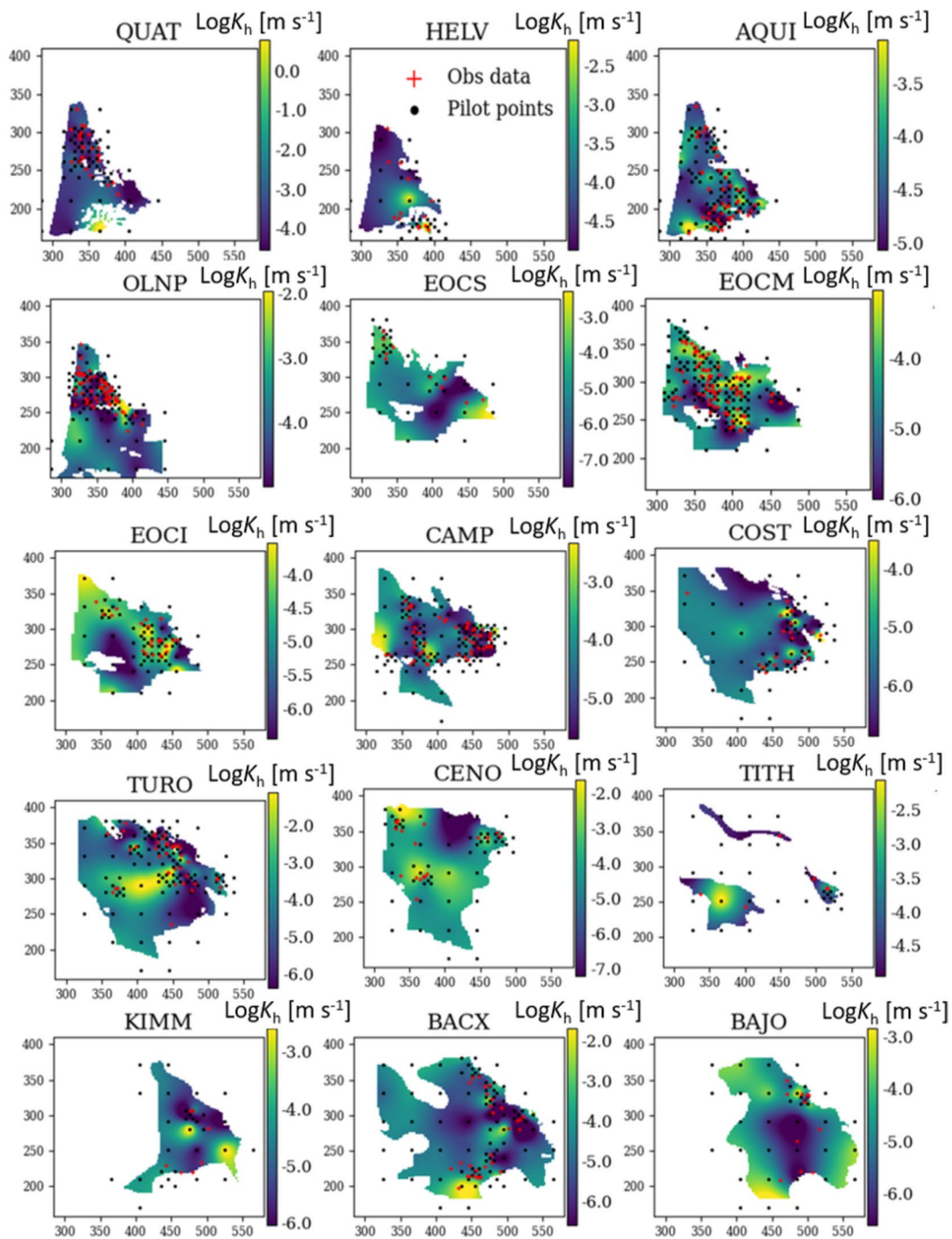
**Fig. 13** Post-calibration horizontal hydraulic conductivities ($K_h$) obtained with the initial adaptive approach

regular grid of pilot points. Both configurations reached the target objective function and a similar description of the heterogeneity, but the adaptive approach required much fewer model calls. This stems from the fact that measurements were heterogeneously distributed, as is often the case in the real world. The refinement criteria based on the measurement density outperformed the criteria derived from parameter sensitivities (CSS, identifiability, gradient of the objective function). This is quite unexpected given previous studies on this topic (Ackerer et al. 2014; RamaRao et al. 1995) but can be interpreted as an effect of the local and approximate nature of model derivatives

obtained by perturbation of prior (initial) parameter values in this study. Apart from its efficiency, a great advantage of the criteria based on measurement density is that it does not require any model runs to be evaluated. In contrast, the iterative adaptive approach, which involves parameter calibration before the refinement of the pilot points grid, yields disappointing results. This may be explained by the characteristics of the GLMA, which could remain "trapped" in a local optimum in the parameter space and prevent further descent of the objective function after refinement. Another tuning factor of the adaptive strategy is the proportion of pilot points to be refined. This threshold was set to 30% in this study, and those pilot points were refined two successive times, leading to the best results in this case but may be subject to further investigation and adjustments in other case studies.

The best strategy identified with the synthetic model (adaptive grid of pilot points refined according to measurement density) was tested on a real-world, regional groundwater model together with two other configurations for comparative purposes: a coarse grid and an iterative adaptive approach. As expected, the objective function converged to a high value with the coarse grid. The iterative adaptive approach gave satisfactory results but was outperformed by the adaptive approach, which is easier to implement. This approach can therefore be recommended for similar configurations, which is a step forward, but a series of improvements and perspectives can be outlined.

While it can significantly facilitate parameter estimation of highly parameterized models, the main limitation of the presented approach is its implication for uncertainty quantification. The proposed refinement strategy focuses on the heterogeneity that can be described by the assimilation of the measured dataset, not necessarily on the heterogeneity that matters for the predictions of interest. As for consequence, heterogeneities in areas with no data may be poorly described, and the uncertainty of related predictions can be underestimated. Measurements are usually conducted where they are supposed to be of greater interest, but this is not always the case. The proposed approach could be extended with refinement criteria not only based on the measurement dataset but also criteria derived from prediction sensitivities, which would allow better consideration of heterogeneities that matter for predictions of interest for decision-making.

The presented approach is optimum in the sense that it maximizes data assimilation while minimizing the number of parameters, and consequently, the computational burden of parameter estimation. This effort on pilot point parameterization may yet be insufficient for parameter estimation to become tractable. It should be accompanied by optimizing the numerical solver and model complexity level that may reduce computation times (Doherty and Moore 2020; Guthke 2017). Apart from reducing the number of parameters and model run times, more frugal parameter estimation algorithms may also be considered. Randomized Jacobian estimates (Doherty 2020) can reduce the need for model calls, and ensemble-based approaches such as IES (White 2018) can also be considered. Most likely, a smart combination of these options may be relevant to address impractical inverse problems.

## Conclusion

The parameter estimation of highly parameterized models with long run times, such as regional multilayered groundwater models, is often associated with challenges that threaten its practicality. In such models, most parameters are representative of parameter values at pilot points. A series of configurations was explored to optimize their distribution and reduce their number. The strategy leading to the best data assimilation, while minimizing the computational burden, is an adaptive grid of pilot points with a refinement based on measurement density. In this approach, the grid of pilot points was refined in areas with the largest number of measurements. This strategy was successfully implemented on a regional flow model, illustrating its efficiency; however, the best parameterization approach to fit the available measurements may not be optimum for uncertainty quantification. To this purpose, the current approach could be extended with pilot point refinement criteria accounting for the sensitivity of predictions of interest. Evaluating these options is a topic for future work.

# Appendix
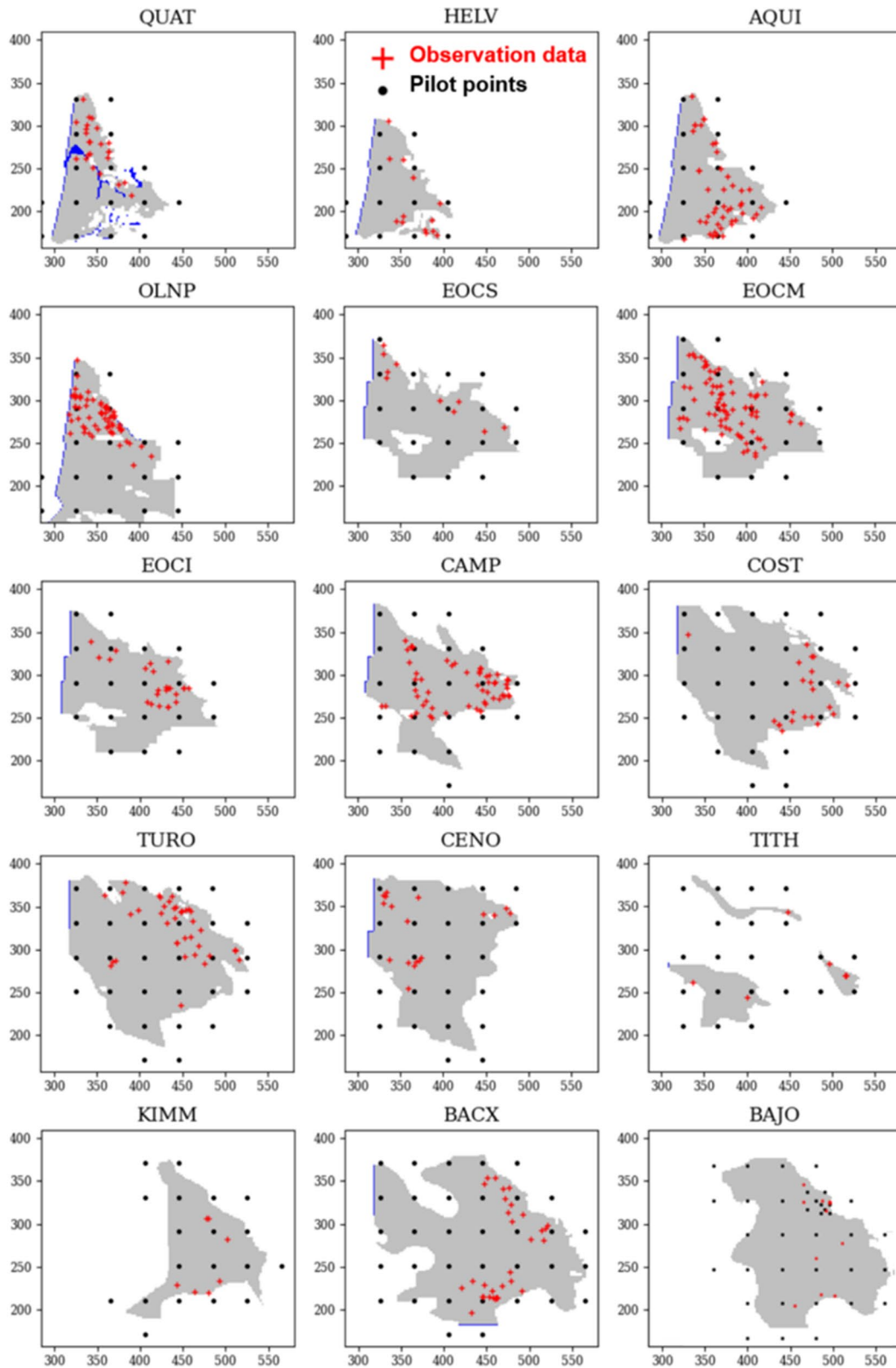


**Fig. 14** Pilot points distribution with the coarse regular grid with a spacing of 20 model cells between pilot points (black points). Measurement/observation data are represented by red crosses
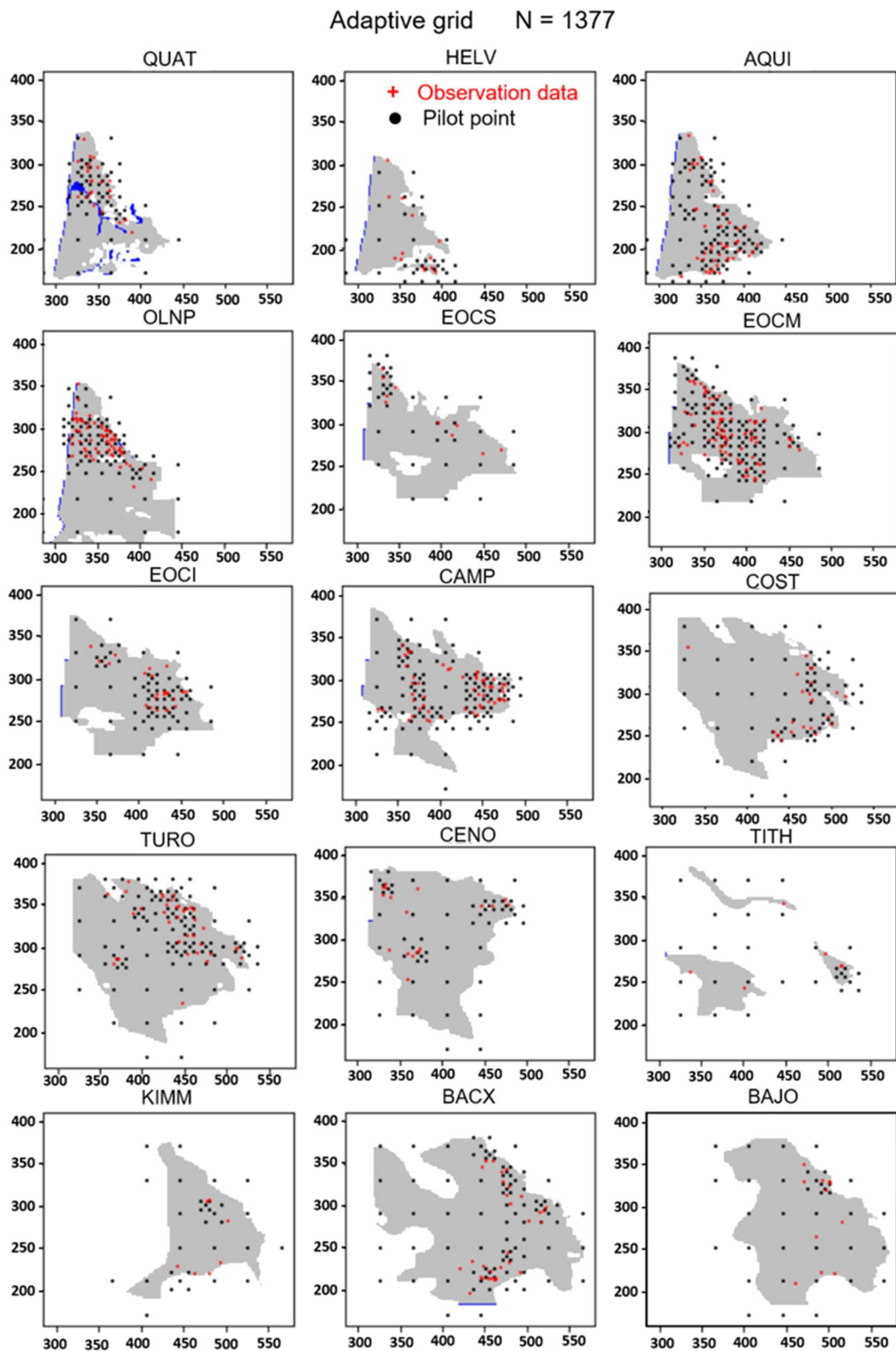
**Fig 15** Adaptive pilot points distribution for all MONA layers. The pilot points were initially distributed using a coarse grid of 20 cell spacing and then refined two successive times based on measurement data availability
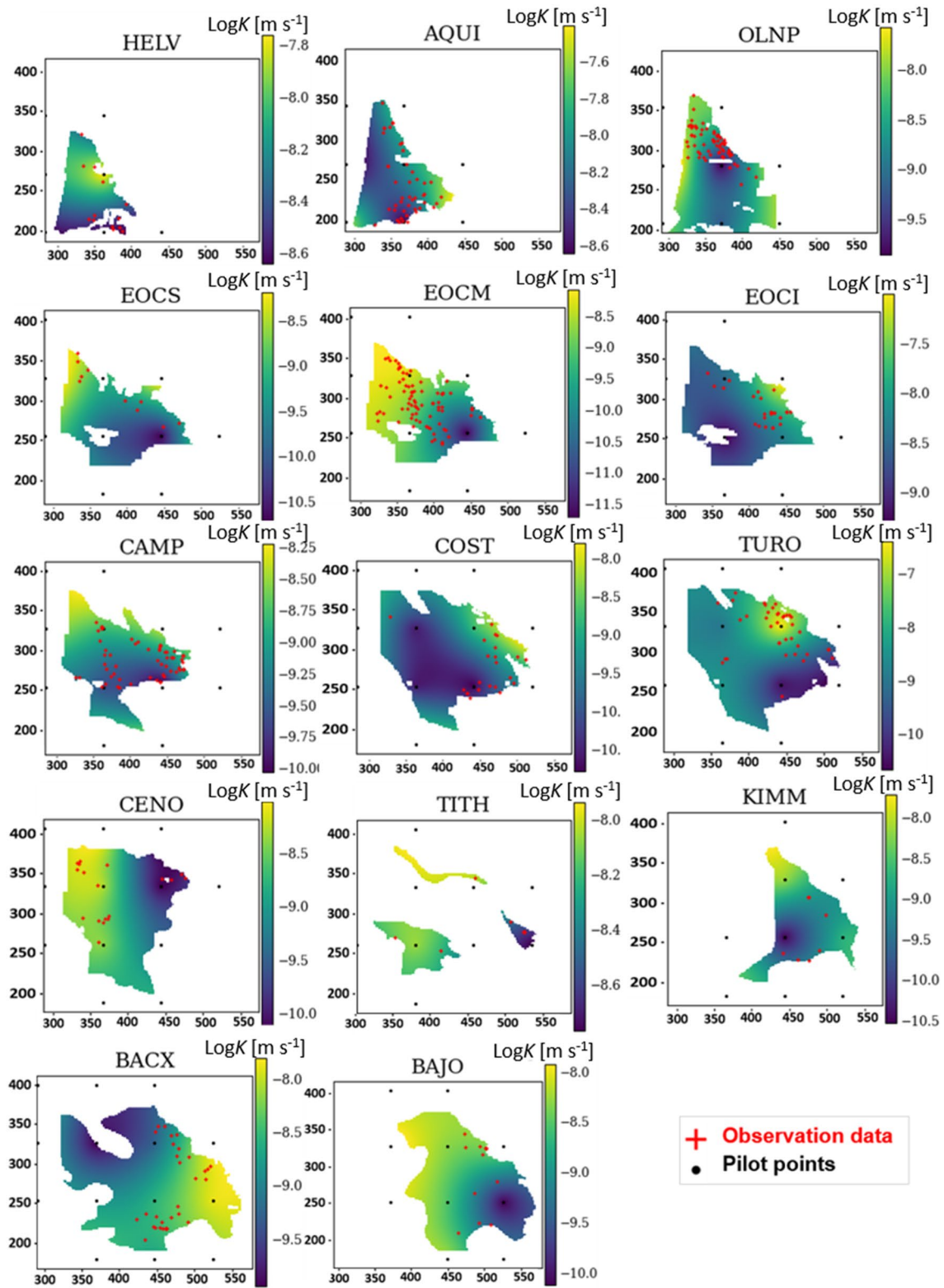
**Fig. 16** Post-calibration vertical hydraulic conductivities ($K_v$) for all aquitard layers
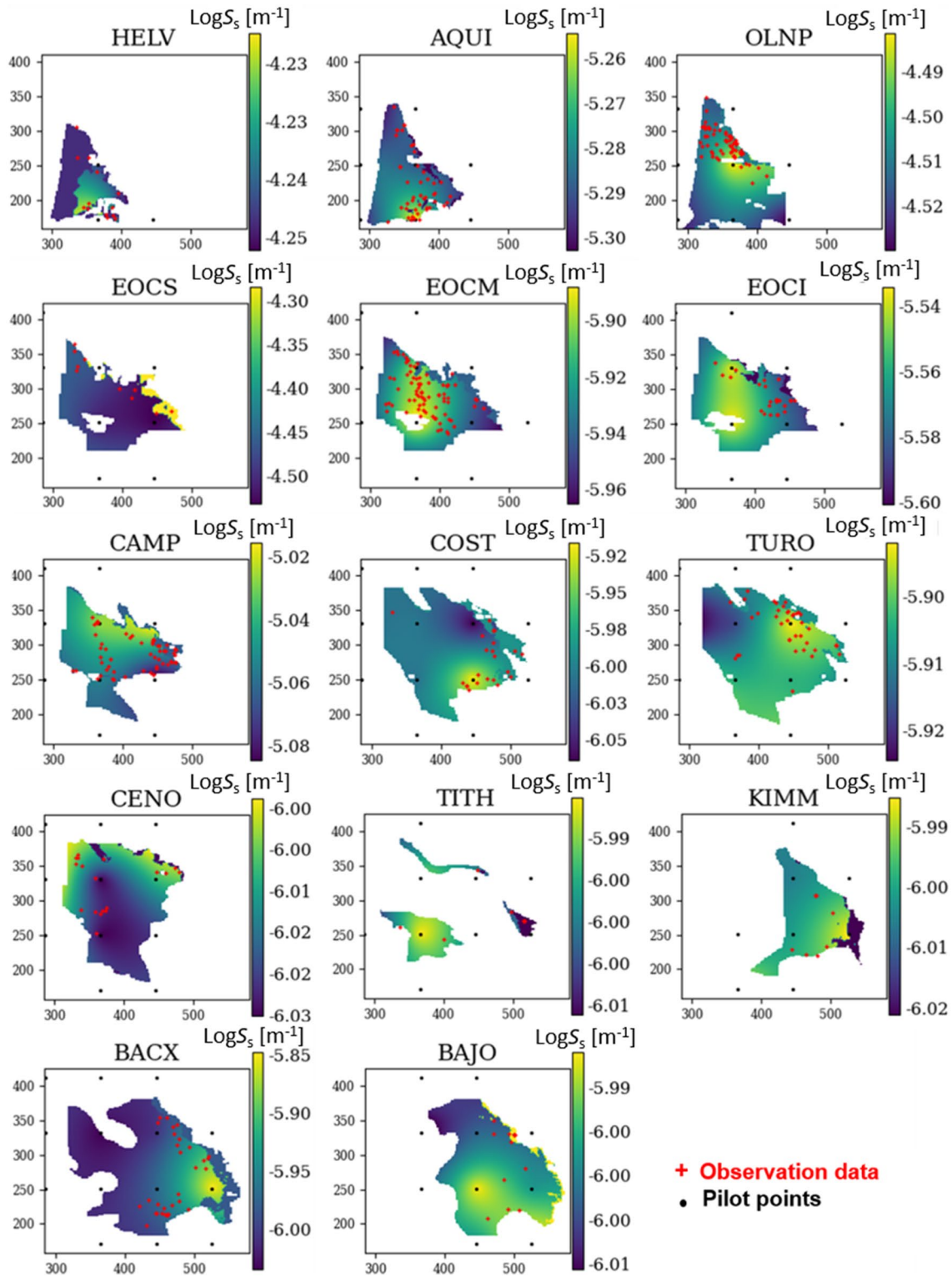
**Fig. 17** Post-calibration specific storage ($S_s$) for all aquifer layers
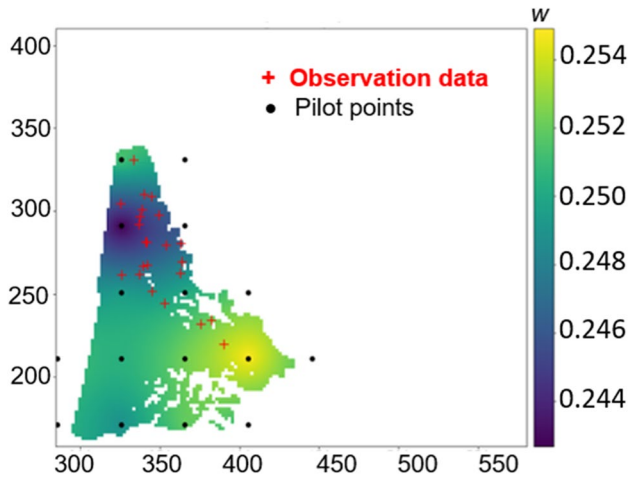
**Fig. 18** Post-calibration specific yield, *ω*, for the 1st aquifer

**Table 2** Summary statistics for calibration: root mean squared error (RMSE) and bias calculated for each MONA layer. Mean values are in *italic* font. *obs* observations

| Layer | | No. of obs | RMSE (m) | BIAS (m) |
|---|---|---|---|---|
| 1 | QUAT | 296 | 1.77 | –0.30 |
| 2 | HELV | 258 | 1.94 | 0.86 |
| 3 | AQUI | 741 | 1.95 | –0.14 |
| 4 | OLNP | 1,250 | 2.77 | –0.09 |
| 5 | EOCS | 129 | 2.93 | –0.98 |
| 6 | EOCM | 1,800 | 2.13 | –0.32 |
| 7 | EOCI | 416 | 2.49 | –0.01 |
| 8 | CAMP | 720 | 10.76 | 0.99 |
| 9 | COST | 265 | 5.23 | –0.48 |
| 10 | TURO | 427 | 7.04 | 0.63 |
| 11 | CENO | 219 | 3.69 | –0.75 |
| 12 | TITH | 61 | 11.21 | –1.90 |
| 13 | KIMM | 48 | 13.10 | 0.51 |
| 14 | BACX | 273 | 7.75 | –0.78 |
| 15 | BAJO | 97 | 3.85 | 0.30 |
| Mean | | *466* | *6.0* | *–0.16* |
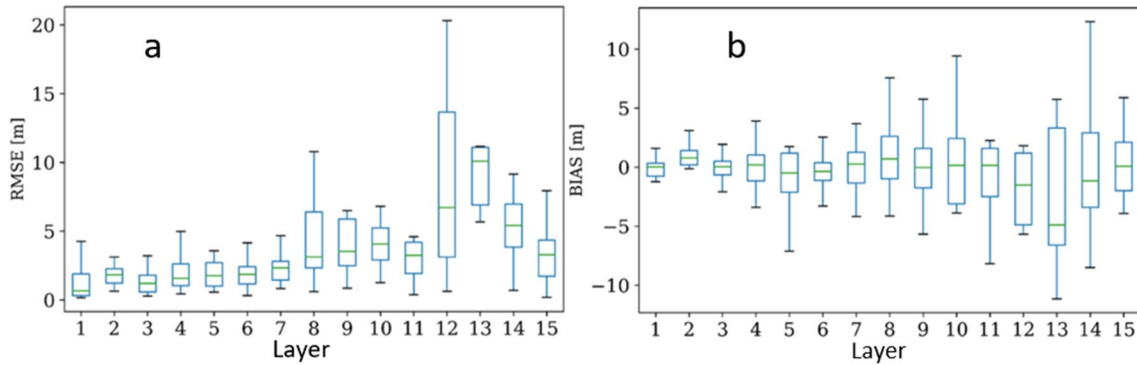


**Fig. 19** Boxplot distribution comparing **a** the root mean squared error (RMSE) and **b** the bias of each layer. The lowest RMSE and bias values indicate better calibration performance for the first eight layers. The boxes span from the 25th to the 75th percentile; green horizon-tal lines denote the median, and whiskers span from 1.5 of the inter-quartile (IQR) range below the low quartile to the 1.5 IQR above the upper quartile

## Declarations

## References

Ackerer P, Trottier N, Delay F (2014) Flow in double-porosity aquifers: parameter estimation using an adaptive multiscale method. Adv Water Resour 73:108–122. https://doi.org/10.1016/j.advwatres.2014.07.001

Alcolea A, Carrera J, Medina A (2006) Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Adv Water Resour 29:1678–1689. https://doi.org/10.1016/j.advwatres.2005.12.009

Anderson MP, Woessner WW, Hunt RJ (2015) Applied groundwater modeling: simulation of flow and advective transport. Academic. https://doi.org/10.1016/C2009-0-21563-7.

Baalousha HM, Fahs M, Ramasomanana F, Younes A (2019) Effect of pilot-points location on model calibration: application to the northern karst aquifer of Qatar. Water 11:679. https://doi.org/10.3390/w11040679

Buscarlet E, Cabaret O, Saltel M (2019) Gestion des eaux souterraines en Région Aquitaine - Développements et maintenance du Modèle Nord-Aquitain de gestion des nappes - Modules 1.1 & 1.2 - Année 2. Rapport final. BRGM/RP-68863-FR, 57 p., 33 ill., 7 tabl., 2 ann. [Groundwater management in the Aquitaine Region: development and maintenance of the North-Aquitain model of groundwater management—Module 1.1, Year 2. Final report. BRGM/RP-68863-FR, 57 p., 33 illustrations, 7 tables. 2 annals.]

Carrera J (1986) Estimation of aquifer parameters under transient and steady state conditions: 1. maximum likelihood method incorporating prior information. https://doi.org/10.1029/WR022i002p00199

Certes C, de Marsily G (1991) Application of the pilot point method to the identification of aquifer transmissivities. Adv Water Resour 14(5):284–300. https://doi.org/10.1016/0309-1708(91)90040-U

Christensen S, Doherty J (2008) Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. Adv Water Resour 31:674–700. https://doi.org/10.1016/j.advwatres.2008.01.003

Cui T, Sreekanth J, Pickett T, Rassam D, Gilfedder M, Barrett D (2021) Impact of model parameterization on predictive uncertainty of regional groundwater models in the context of environmental impact assessment. Environ Impact Assess Rev 90:106620. https://doi.org/10.1016/j.eiar.2021.106620

de Marsily Gh (1978) De l'identification des systèmes hydrogéologiques [On the identification of hydrogeological systems]. PhD Thesis, University Paris VI, France

de Marsily Gh (1984) Spatial variability of properties in porous media: a stochastic approach. In: Bear J, Corapcioglu MY (eds) Fundamentals of transport phenomena in porous media. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-6175-3_15

Delottier H, Pryet A, Dupuy A (2017) Why should practitioners be concerned about predictive uncertainty of groundwater management models? Water Resour Manag 31:61–73. https://doi.org/10.1007/s11269-016-1508-2

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

Doherty J (2015) Calibration and uncertainty analysis for complex environmental models. Watermark, Brisbane

Doherty JE (2019) Model-independent parameter estimation user manual part II: PEST utility support software. Watermark, Brisbane

Doherty JE (2020) PEST_HP, PEST for highly parallelized computing environments. Watermark, Brisbane

Doherty J, Hunt RJ (2009) Two statistics for evaluating parameter identifiability and error reduction. J Hydrol 366:119–127. https://doi.org/10.1016/j.jhydrol.2008.12.018

Doherty J, Moore C (2020) Decision support modeling: data assimilation, uncertainty quantification, and strategic abstraction. Groundwater 58:327–337. https://doi.org/10.1111/gwat.12969

Doherty J, Fienen MN, Hunt RJ (2010) Approaches to highly parameterized inversion: pilot-point theory, guidelines, and research directions. US Geol Surv Sci Invest Rep 2010-5168, 36 pp. https://doi.org/10.3133/sir20105168.

Elshall AS, Arik AD, El-Kadi AI, Pierce S, Ye M, Burnett KM, Wada CA, Bremer LL, Chun G (2020) Groundwater sustainability: a review of the interactions between science and policy. Environ Res Lett 15:093004. https://doi.org/10.1088/1748-9326/ab8e8c

Fienen MN, Muffels CT, Hunt RJ (2009) On constraining pilot point calibration with regularization in PEST. Ground Water 47:835–844. https://doi.org/10.1111/j.1745-6584.2009.00579.x

Gómez-Hernánez J, Sahuquillo A, Capilla JE (1997) Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. theory. J Hydrol 203:162–174. https://doi.org/10.1016/S0022-1694(97)00098-X

Guthke A (2017) Defensible model complexity: a call for data-based and goal-oriented model choice. Groundwater 55:646–650. https://doi.org/10.1111/gwat.12554

Hayek M, Ackerer P, Sonnendrücker É (2009) A new refinement indicator for adaptive parameterization: application to the estimation of the diffusion coefficient in an elliptic problem. J Comput Appl Math 224:307–319. https://doi.org/10.1016/j.cam.2008.05.006

Hill MC, Tiedemann CR (2006) Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty. Wiley, London

Hosteins L (1982) Étude hydrogéologique du réservoir Oligocène en Aquitaine occidentale: gestion et conservation de la ressource de cette nappe dans la région de Bordeaux. [Hydrogeological study of the Oligocene reservoir in western Aquitaine: management and conservation of this aquifer resource in the Bordeaux region]. PhD Thesis, University of Bordeaux 1, Bordeaux

Hughes IG, Hase TPA (2013) Measurements and their uncertainties: a practical guide to modern error analysis. OUP Oxford, Oxford

Kapoor A, Kashyap D (2021) Parameterization of pilot point methodology for supplementing sparse transmissivity data. Water 13:2082. https://doi.org/10.3390/w13152082

Klaas DKSY, Imteaz MA (2017) Investigating the impact of the properties of pilot points on the calibration of groundwater models: a case study of a karst catchment in Rote Island, Indonesia. Hydrogeol J 25:1703–1719. https://doi.org/10.1007/s10040-017-1590-4

Larroque F (2004) Gestion globale d'un système aquifère complexe Application à l'ensemble aquifère multicouche médocain [Comprehensive management of a complex aquifer system: application to the multilayered Médocian aquifer]. PhD Thesis, University of Michel de Montaigne-Bordeaux III, Bordeaux

LaVenue AM, Pickens JF (1992) Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resour Res 28:1543–1569. https://doi.org/10.1029/92WR00208

Moore C, Doherty J (2005) The role of the calibration process in reducing model predictive error. Water Resour Res 41. https://doi.org/10.1029/2004WR003501

Moussié B (1972) Le système aquifère de l'Éocène moyen et supérieur du bassin nord aquitain - Influence du cadre géologique sur les modalités de circulation. [The aquifer system of the Middle and Upper Eocene in the North Aquitaine Basin: influence of the geological framework on circulation modalities]. PhD Thesis, University of Bordeaux, France

Neuman SP (1973) Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty. Water Resour Res 9:1006–1021. https://doi.org/10.1029/WR009i004p01006

Pebesma EJ (2004) Multivariate geostatistics in S: the Gstat package. Comput Geosci 30:683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pédron N, Platel JP (2005) Gestion des eaux souterraines en Région Aquitaine. Développements et maintenance du Modèle Nord Aquitain de gestion des nappes: Module 4, Année 2. [Groundwater management in the Aquitaine Region. Development and maintenance of the North Aquitaine groundwater management model: Module 4, Year 2]. BRGM/RP -53659- FR, BRGM, Orléans

Poeter E, Hill M, Lu D, Tiedeman C, Mehl S (2014) UCODE_2014, with new capabilities to define parameters unique to predictions, calculate weights using simulated values, estimate parameters with SVD, evaluate uncertainty with MCMC, and more. Report GWMI 2014-02, Integrated Groundwater Modeling Center. http://pubs.er.usgs.gov/publication/70159674. Accessed Oct 2023

RamaRao BS, LaVenue AM, De Marsily G, Marietta MG (1995) Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. Water Resour Res 31:475–493. https://doi.org/10.1029/94WR02258

Saltel M, Wuilleumier A, Cabaret O (2016) Gestion des eaux souterraines en Région Aquitaine - Développements et maintenance du Modèle Nord-Aquitain de gestion des nappes. Module 1 - Année 5 - Convention 2008-2013. [Groundwater management in the Aquitaine Region: development and maintenance of the North-Aquitain model of groundwater management—Module 1, Year 5, Convention 2008-2013]. Final report BRGM/RP-65039-FR, BRGM, Orléans, 82 pp

Schilling OS, Partington DJ, Doherty J, Kipfer R, Hunkeler D, Brunner P (2022) Buried paleo-channel detection with a groundwater model, tracer-based observations, and spatially varying, preferred anisotropy pilot point calibration. Geophys Res Lett 49:e2022GL098944. https://doi.org/10.1029/2022GL098944

Thiery D (2015) Code de calcul MARTHE - Modélisation 3D des écoulements dans les hydrosystèmes -Notice d'utilisation de la version 7.5. [MARTHE calculation code: 3D modeling of flow in hydrosystems—user manual for version 7.5]. BRGM/RP-64554-FR report, BRGM, Orléans, France, 306 pp

Thiéry D, Amraoui N, Gomez E, Pédron N, Seguin J-J (2011) Regional model of groundwater management in the North Aquitania Aquifer System: water resources optimization and implementation of prospective scenarios taking into account climate change. https://doi.org/10.1007/978-94-007-1623-0_19.

Tikhonov A, Arsenin V (1977) Solutions of ill-posed problems. Winston, Great Falls

Tonkin MJ, Doherty J (2005) A hybrid regularized inversion methodology for highly parameterized environmental models. Water Resour Res 41:W10412. https://doi.org/10.1029/2005WR003995

Trabucchi M, Fernàndez-Garcia D, Carrera J (2022) The worth of stochastic inversion for identifying connectivity by means of a long-lasting large-scale hydraulic test: the Salar de Atacama case study. Water Resour Res 58:e2021WR030676. https://doi.org/10.1029/2021WR030676

Wen X-H, Lee S, Yu T (2006) Simultaneous integration of pressure, water cut, and 4-D seismic data in geostatistical reservoir modeling. Math Geol 38:301–325. https://doi.org/10.1007/s11004-005-9016-6

White JT (2018) A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. Environ Model Softw 109:191–201. https://doi.org/10.1016/j.envsoft.2018.06.009

White J, Lavenue M (2023) Advances in the pilot point inverse method: where are we now? C R Géosci 355(S1):1–9. https://doi.org/10.5802/crgeos.161

White JT, Fienen MN, Doherty JE (2016) A python framework for environmental model uncertainty analysis. Environ Model Softw 85:217–228. https://doi.org/10.1016/j.envsoft.2016.08.017

Zhou H, Gómez-Hernández JJ, Li L (2014) Inverse methods in hydrogeology: evolution and recent trends. Adv Water Resour 63:22–37. https://doi.org/10.1016/j.advwatres.2013.10.014