Appendix B.
Parameter estimation algorithm

The parameter estimation process was achieved through history matching, using observations of both steady-state and transient conditions, as described in the Data Sources section, and using the Gauss-Levenberg-Marquardt algorithm implemented in the PEST software suite (Doherty, 2014) and SVD-assist (Tonkin and Doherty, 2005). In this section we briefly summarize the technique which is described in greater detail in Doherty (2014), Doherty and Hunt (2010), and references therein.

The Gauss-Levenberg-Marquardt method is a gradient-based search algorithm that adjusts parameters in pursuit of the minimum value of an objective function. The objective function is the weighted sum of squared errors comparing field observations with forecasts at the same time and place made by the model. The objective function is referred to as

$$\mathbf{\Phi} = \left(\mathbf{y} - g(\mathbf{p})\right)^{T}\mathbf{Q}^{-\mathbf{1}}\left(\mathbf{y} - g(\mathbf{p})\right)$$

where:

$\mathbf{\Phi}$ is the objective function;

y is a vector of observations;

g(**p**) is a vector of modeled values collocated in time and space with the observations, evaluated at parameter values **p**;

Q is a matrix of observation weights (in this work, **Q** is a diagonal matrix indicating no correlation among observation errors is assumed). The units of **Q** are 1/(units of **y**);

$(\cdot)^{T}$ indicates a vector transpose; and

$(\cdot)^{-1}$ indicates a matrix inversion.

The term $\left(\mathbf{y} - g(\mathbf{p})\right)$ is the vector of residuals—also called errors—comparing measured and modeled results. The observation weights play important roles both in enforcing an appropriate level of fit (correspondence between observed and modeled results) and in balancing $\mathbf{\Phi}$ such that observations of various types all contribute to the objective function. The Data Sources section describes the assignment and adjustment of weights in more detail. The use of variable observation weights acknowledges that a perfect match between modeled and observed values is unattainable and, in fact, undesirable. The many reasons for this include error in the observations, the necessary fact that the model is a simplification of the true physical system, and the smoothing of time signals by the model. The flexibility that this imparts, however, means that many (in fact, infinite) arrangements of parameters can result in the same value of $\mathbf{\Phi}$. This non-uniqueness motivates the need to incorporate expert ("soft") knowledge to arrive at a set of parameters that satisfies both the desired level of fit and conforms to expert understanding of reasonable values for the parameters.

The introduction of qualitative expert prior information is made through several avenues including 1) enforcing the level of fit desired through the assignment of weights, 2) normalizing observation group weight contributions through weight adjustment as discussed in the Parameter Estimation section, 3) inclusion of a penalty to the objective function for parameters deviating from a preferred condition through regularization, and 4) the singular value decomposition algorithm. Decisions made regarding all these aspects of prior information are inherently subjective. However, it is an important way for human understanding of the groundwater flow system to play a role in the process beyond blind trust of the algorithm, leading to more meaningful results (e.g. Fienen, 2013).

The penalty to the objective function is assigned as a form of Tikhonov regularization (Tikhonov 1963a, b) through an additional term in the objective function that penalizes deviation from a preferred condition—in this case, preferred homogeneity of spatially distributed parameters.

$$\mathbf{\Phi} = \left(\mathbf{y} - g(\mathbf{p})\right)^T \mathbf{Q}^{-1}\left(\mathbf{y} - g(\mathbf{p})\right) + \boldsymbol{\beta}\mathbf{p}^T\mathbf{Z}\mathbf{p}$$

where:

$\boldsymbol{\beta}$ is a weight-balancing regularization with fit as a function of $\mathbf{\Phi}_{MLIM}$;

$\mathbf{\Phi}_{MLIM}$ controls the strength of regularization (Doherty 2003; Fienen and others, 2009); and

Z is a matrix of Kriging weights relating parameters to one another based on an exponential variogram and the distances between them. The Kriging weights are set at the beginning of the process and remain constant. However, the strength with which the regularization is enforced changes throughout the iterations of PEST.

The variable $\mathbf{\Phi}_{MLIM}$ was set to a value the same order of magnitude as the number of observations, as suggested by Fienen and others (2009) to balance the level of fit with the importance of the prior information.

Singular value decomposition (SVD) was also used to enhance solution stability and provide a secondary level of regularization. In SVD, the sensitivity of observations to parameters is transformed to align with principal orientations of maximum information. This transformed space can then be divided into the calibration space and the null space. The calibration space is the region in which information from observations is meaningfully projected onto parameters while the null space represents a space where variability in parameters has little or no impact on model outputs of interest. This division between solution and null space is controlled by the stability of the sensitivity matrix and the settings recommended by Doherty and Hunt (2010) was adopted.

Null-Space Monte Carlo details

When using null-space Monte Carlo, a parameter called the "singular value cutoff" defines how much parameters are allowed to vary as they are sampled. As discussed above, a stability criterion was used to control the singular value decomposition cutoff during history matching. As a result, the singular value cutoff of 40 was selected to be representative of the stability criterion derived cutoff in the history matching. The parameter distributions sampled are derived from the posterior Schur complement after history matching which further informs the parameter estimates based on the history matching performance (see White and others, 2016, for more details). This choice of distributions enforces correlation among the parameters making each sample consistent with the Bayesian posterior covariance. The implication of this is that parameters which are more informed in the history matching process are more tightly constrained in the sampling process than parameters which are less constrained by history matching.

The corresponding prior covariance was set as bounds such that, for hydraulic conductivity, 95% of the normally distributed samples should fall within plus or minus one order of magnitude. For recharge, a more restrictive prior parameter bound was assigned such that 95% of samples should fall within plus or minus 10% of the estimated recharge multiplier from history matching. These prior assumptions are subjective and based on expert knowledge. The posterior distributions incorporate formal information from the history matching process.

References

Doherty, J., 2003, Ground water model calibration using pilot points and regularization: Ground Water, v. 41, p. 170–177, http://doi.org/dsp9zv.

Doherty, J., 2014, PEST, model-independent parameter estimation: User manual, 5th ed.: Brisbane, Australia, Watermark Numerical Computing, [variously paged].

Doherty, J.E., and Hunt, R.J., 2010, Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration: U.S. Geological Survey Scientific Investigations Report 2010-5169, 59 p.

Fienen, M.N., 2013, We speak for the data: Ground Water, v. 51, p. 157, http://doi.org/c48c.

Fienen, M.N., Muffels, C.T., and Hunt, R.J., 2009, On constraining pilot point calibration with regularization in PEST: Ground Water, v. 47, p. 835–844, http://doi.org/bk6z24.

Tikhonov, A.N., 1963a, Solution of incorrectly formulated problems and the regularization method: Soviet Mathematics Doklady, v. 4, p. 1035–1038.

Tikhonov, A.N., 1963b, Regularization of incorrectly posed problems: Soviet Mathematics Doklady, v. 4, p. 1624–1627.

Tonkin, M.J., and Doherty, J., 2005, A hybrid regularized inversion methodology for highly parameterized models: Water Resources Research, v. 41, no. 10, 16 p.

White, J.T., Fienen, M.N., and Doherty, J.E., 2016, A python framework for environmental model uncertainty analysis: Environmental Modelling & Software, v. 85, p. 217–228.