ground water

Methods Note/

On Constraining Pilot Point Calibration with Regularization in PEST

by Michael N. Fienen[1], Christopher T. Muffels[2], and Randall J. Hunt[3]

Abstract

Ground water model calibration has made great advances in recent years with practical tools such as PEST being instrumental for making the latest techniques available to practitioners. As models and calibration tools get more sophisticated, however, the power of these tools can be misapplied, resulting in poor parameter estimates and/or nonoptimally calibrated models that do not suit their intended purpose. Here, we focus on an increasingly common technique for calibrating highly parameterized numerical models—pilot point parameterization with Tikhonov regularization. Pilot points are a popular method for spatially parameterizing complex hydrogeologic systems; however, additional flexibility offered by pilot points can become problematic if not constrained by Tikhonov regularization. The objective of this work is to explain and illustrate the specific roles played by control variables in the PEST software for Tikhonov regularization applied to pilot points. A recent study encountered difficulties implementing this approach, but through examination of that analysis, insight into underlying sources of potential misapplication can be gained and some guidelines for overcoming them developed.

Introduction

The hydrogeologic environment is inherently complex. As a result, hydrogeologists can never fully represent true subsurface variability, and environmental models are always a simplification of reality. In a broad sense, two endmember approaches were developed to represent true system variability in models.

1. Consolidate system variability by "lumping" real-world variability into a set of piecewise constant (homogeneous) zones to have fewer zones than observations, thus maintaining a well-posed problem with a single representative set of optimal parameters.
2. Represent the stochastic characteristics of the subsurface using multiple realizations rather than a single representative parameter set. The stochastic technique requires many more parameters to be "distributed" rather than lumped.

The strengths and weaknesses of the two approaches were described by Hill (2006a) and Gomez-Hernandez (2006), respectively. An intermediate "regularized inversion" approach has been recently reviewed (Hunt et al. 2007) that revisits the overall goal of parameter estimation (see Neuman 1973; Carrera and Neuman 1986, among others). The objective of regularized inversion was to include many parameters like the stochastic approach, but mathematical regularization is employed to obtain a single solution like in the lumped approach. Hunt et al. (2007) advocate the regularization approach as a "middle ground" because it reduces structural error due to a priori zonation but produces a single model for use by decision makers rather than multiple realizations. They further suggest that wider use of regularized inversion techniques is possible with currently available practical tools such as PEST (Doherty 2008a). PEST is the most widely used software

[1]Corresponding author: U.S. Geological Survey, Wisconsin Water Science Center, Middleton, WI 53562; (608) 821-3894; fax (608) 821-3817; mnfienen@usgs.gov

[2]S. S. Papadopulos and Associates Inc., Bethesda, MD 20814

[3]U.S. Geological Survey, Wisconsin Water Science Center, Middleton, WI 53562

[Correction added after online publication April 21, 2009: on page 9, a reviewer's name was incorrectly listed as "Stefen Mehl", and should have been listed as "Steffen Mehl". We apologize for this error.]

Received August 2008, accepted March 2009.

doi: 10.1111/j.1745-6584.2009.00579.x

suite of its kind (Ginn et al. 2007) and has the capability to handle large numbers of parameters with mathematical regularization.

This work focuses on using pilot points (e.g., de Marsily et al. 1984) with Tikhonov regularization (Tikhonov 1963a, 1963b) that together represent the most common path between the traditional lumped approach and distributed approaches representing parameters on the individual model-cell scale. Pilot points are a discrete subset of points in the model domain at which values are estimated during parameter estimation. The values estimated at the pilot points are interpolated to the remaining cells (typically through kriging). This approach provides greater parameter flexibility than using piecewise homogeneous zones but avoids the computational expense and mathematical challenges of trying to estimate a parameter value at every node in a model. Although sequential simulation at the interpolation step to generate multiple realizations of the parameter fields can also be used to move beyond the traditional approach (see, e.g., Ramarao et al. 1995; Alcolea et al. 2006), these algorithms are not currently available in the PEST software suite, thus are not considered further in this Methods Note. General guidelines for pilot points and Tikhonov regularization are available both in the PEST software documentation in Doherty (2008a) and in Doherty (2003). Nonetheless, Hill (2008) described potential challenges remaining in the practical understanding of controlling regularization in PEST; specifically, with increased flexibility comes a danger of "overfitting," in which unreasonable parameter values are estimated to force the model results to correspond closely to observations. Moreover, Alcolea et al. (2006) note that the location and density of pilot points can also be an important consideration for avoiding overfitting. However, overfitting is not inherent to the method itself but stems from improper specification of pilot point conceptualization and PEST control variables by modelers applying the method.

The goal of this Methods Note is to provide guidelines for specifying control variable values for Tikhonov regularization in PEST to avoid overfitting. Specifically, the PEST control variable PHIMLIM needs to be carefully considered to obtain geologically reasonable optimized parameters. PHIMLIM corresponds to the target measurement objective function (Φ_m^l in Doherty [2003]) that can be expected given a model conceptualization. This control variable dictates the tradeoff between fitting model outputs to observations and enforcement of the regularization constraints (e.g., deviations from adherence to preferred homogeneity or preferred value conditions). It can also determine the parameter estimation termination and interact with the Marquardt parameter. Doherty (2003) discussed the theory and an example application of pilot points with Tikhonov regularization in detail; thus, our focus is on application of the use of Tikhonov and Marquardt parameters in PEST. A complete list of the PEST control variables (and suggested default values) controlling both the regularization and the Gauss-Levenberg-Marquardt method is included in Table S1 of the Supporting Information, along with suggested default values.

In the remainder of this Methods Note, we discuss the control variables introduced previously using the synthetic problem of Hill (2008) to illustrate the behavior of the regularization and Marquardt controls. Using this example, we evaluate how control variable specification can be used to address the challenges for pilot point calibration discussed by Hill (2008); links to supporting information are given so that the reader can access the model input and output files discussed here.

Controls on Regularization

In this section, we briefly discuss the controls on regularization and the formulation of the objective function for parameter estimation. Appendix S1 in the Supporting Information and the references provide further details. One criterion for evaluating parameter estimation results is the comparison of model-simulated results to measured observations (misfit), represented by an objective function. In parameter estimation without regularization, parameters are systematically adjusted until the best fit between simulated and measured is achieved (i.e., the objective function reaches a minimum) or other termination criteria are met. When Tikhonov regularization (Tikhonov and Arsenin 1977) is applied, the objective function is supplemented with a term quantifying the degree of regularization on the parameter field. These two objective function components can offset one another, and the optimal parameter set is determined by the minimum of both in combination. Thus, the parameter estimation process can be thought of as a balance between less regularized (rougher) solutions that have more freedom to decrease the misfit and more regularized (smoother) solutions that sacrifice fit. This is most easily seen in a preferred homogeneity regularization strategy in which more regularization indicates greater adherence to a preferred condition of homogeneity of the parameter field. The discussion could be easily extended to other strategies for regularization, such as more general preferred difference constraints, or preferred value in which more regularization indicates parameter values forced to be closer to preferred (user specified) values. Balancing misfit against the requirement of geologically reasonable parameters forms the fundamental basis for selecting an optimal solution from a family of possible solutions.

The PEST formulation of this balancing quantitatively ties the two objective function components to a physically meaningful control variable, the target measurement objective function (PHIMLIM, Φ_m^l). It is never appropriate to obtain zero misfit due to uncertainties in observed values, model assumptions, simplified model structure, and other sources of error. The target objective function accounts for this total uncertainty and provides a mechanism to explore the effect of different levels of misfit even though we may never perfectly know the appropriate level of uncertainty. With a rougher solution, misfit may be minimized, but this may not be optimal for several reasons. First, the best fit (rough solution) often includes geologically or physically unreasonable values. Mathematical regularization provides a flexible approach to "rein in" unreasonable values while accepting a specified increase in misfit. Second, unregularized solutions are commonly nonunique; when parameter values are allowed to vary too freely, an infinite number of combinations can result in a similar low level of misfit. Finally, models are an imperfect representation of reality so some sacrifice in fit is expected and the pursuit of perfect fit is unreasonable.

The Control Variable PHIMLIM

For most problems, the single most important control variable controlling regularization strength in PEST is PHIMLIM (Φ_m^l in Doherty [2003] and Appendix S1 in the Supporting Information). A low value of PHIMLIM weakens the regularization constraint and enables PEST to use rough solutions to obtain a lower misfit for a given conceptualization. When PHIMLIM is set too low, pursuit of a best fit can negate the effect of the preferred condition and result in unrealistically large parameter ranges and/or parameter "bull's eyes" near observations. This artifact is defined here as "overfitting," which indicates the parameter estimation process is chasing noise in the measurements or errors in model conceptualization. Much criticism of highly parameterized models stems from results that show overfitting, but it is clear that overfitting is the result of user input rather than an inherent artifact of the method.

A high value of PHIMLIM enforces a stronger level of smoothing, which increases misfit. Extremely high values can result in underfitting where the parameter estimation process is not fully extracting the information provided by the observations. In other words, too much regularization prevents the parameter estimation from reaching a reasonable level of misfit. As a result, the goal was to find a balance between overfitting and underfitting conditions.

The regularization weight factor (μ) is the parameter that governs the tradeoff between over- and underfitting by controlling the balance between the fit and the regularization components of the objective function (Appendix S1 in the Supporting Information). In practice, this factor is difficult to interpret physically and, as implemented in PEST, is internally calculated and not constant throughout the parameter estimation. Calculation of μ by PEST is based on PHIMLIM, a variable that gives some physical meaning to the relation between misfit and regularization (as it can be thought of in terms of expected error of the observations and model structure). Thus, it is the user-specified basis of controlling the strength of the regularization. In the maximum likelihood context, it is possible to interpret μ in terms of the parameter covariance matrix, but in many practical applications, it is more convenient to consider terms of fitting criteria.

Others have suggested a more systematic approach for identifying an optimal balancing between misfit and regularization (see, e.g., Carrera and Neuman 1986; Alcolea et al. 2006), but this is currently not implemented in PEST.

Setting the Control Variable PHIMLIM

How does one set PHIMLIM appropriately? This falls, in part, under the "art" of modeling, but some guidelines are given subsequently. The modeler's insights and "soft" knowledge about the modeled system, such as the geologic setting in which the model resides, are critical for deciding what to consider realistic and unrealistic parameter sets. A key difference is that in regularized inversion, this decision is made after the field data have directly informed the optimal parameter set rather than using a priori zonation to define what is realistic. These calibration data, however, cannot decide what is the reasonable tradeoff; thus, in common practice, setting, PHIMLIM is best accomplished through exploration of the behavior of the inversion using a range of values of PHIMLIM.

Although hard and fast guidelines do not exist for setting PHIMLIM, one starting point applies when the weights assigned to the observations equal the inverse of the measurement standard deviation (which has advantages when interpreting regression results using linear statistics [Hill and Tiedeman 2007]) and the structural error is negligible; in this case, a reasonable starting PHIMLIM value is:

$$\text{PHIMLIM} = \text{NOBS}, \tag{1}$$

where NOBS is the number of observations (Appendix S2 in the Supporting Information).

This will likely result in overfitting because measurement standard deviation neglects other sources of uncertainty, model structural uncertainty paramount among other sources. In other words, expecting model results to correspond to actual observations assumes the only source of error is the innate variability of the measurements. All models are imperfect, however, so there are other sources of error that prevent such a close fit from being achieved. Nonetheless, this allows the modeler to see a potential "endmember" best fit that can be expected for a given conceptualization. PHIMLIM values higher than the resulting PEST-reported (overfit and unrealistic) optimized objective function value can then be used as a starting point for subsequent parameter estimation. Too low a value of PHIMLIM rarely, if ever, results in a geologically reasonable parameter set, and the final values of PHIMLIM should be revised upward from this value.

In addition to its principal role in controlling the strength of regularization, PHIMLIM is also a criterion below which PEST terminates the estimation. Thus, PHIMLIM can also be thought of as the user-specified definition of "good enough" for the misfit. In most cases, PHIMLIM fulfills both roles, but in other cases, control of regularization strength and estimation termination can be separated.

Tuning PHIMLIM with PHIMACCEPT

The control variable PHIMACCEPT helps account for deviations from the linearity assumption used in this linearized approach—in most cases, a process that is less critical for the parameter estimation. Possible deviations are accounted for by redefining an acceptable measurement objective function target, which is recommended to be set slightly higher (about 5% or 10%) than PHIMLIM Doherty (2008a, p. 7-7). This enables flexibility when choosing the best parameter set during each iteration. If the measurement objective function is lower than PHIMACCEPT for any parameter set calculated up to that point, PEST will select the candidate parameter set with the lowest value for the regularization objective function within that subset. PHIMACCEPT also allows some flexibility in the calculation of the Marquardt parameter by the algorithm because it relinearizes the problem. According to Doherty (2008a, p. 7-7), if the parameter estimation process is behaving well, PHIMACCEPT can remain close to PHIMLIM. If it is slow to converge, increasing PHIMACCEPT can accelerate convergence.

The Control Variable FRACPHIM

When PHIMLIM is not well known, the modeler may still like to have some regularization applied regardless of his or her a priori estimate of the measurement objective function. FRACPHIM can be used when the modeler is unsure of what value to use for PHIMLIM for regularization. A nonzero value for FRACPHIM causes PHIMLIM to be set at the greater of the user-specified PHIMLIM value or the product of FRACPHIM with the current measurement objective function value. PEST will then "aim for" a measurement objective function lower than the current one, but regularization strength will be enforced as though the user had specified a higher level of PHIMLIM until the calculated PHIMLIM value is reduced to the user-specified value. Termination of the algorithm is still controlled by the user-specified value of PHIMLIM.

FRACPHIM can range from 0.0 (equivalent to omitting the control variable) to 1.0, with optimal values commonly between 0.1 and 0.3 (Doherty 2008a) in cases where the objective function of the initial model is appreciably higher than expected from the optimal model parameters; a higher value would be more suited for cases where the initial and final optimal objective function are closer, as would be expected if starting values were derived from optimal values from an earlier parameter estimation. In addition to adjusting the value of PHIMLIM during every optimization iteration, PEST also adjusts the value of PHIMACCEPT when FRACPHIM is specified so that the ratio of PHIMACCEPT to PHIMLIM is the same as that specified by the user in the PEST control file.

Using Marquardt Lambda with Tikhonov Regularization

The algorithm PEST uses to solve the parameter estimation problem in regularization mode is a modified Gauss-Newton method, assisted by a Levenberg-Marquardt (LM) formulation (Doherty 2008a, chapter 2). LM is a means to an end; it stabilizes the nonlinear regression by augmenting the approximated Hessian matrix diagonal to ensure it is nonsingular (invertible). Using LM guides the optimal parameter search closer to a steepest-descent formulation, which is certain to converge, but may repeatedly undershoot the objective function minimum causing "hemstitching" or retarded progress toward the objective function minimum and slower convergence (Aster et al. 2005, p. 176).

Caution is required when using LM with regularization. A highly regularized problem is no longer underdetermined, so LM may not be needed and can impede the estimation by reducing the size of parameter value updates so their improvements become negligible. This result, in turn, may lead the algorithm to prematurely conclude that it has converged on optimal parameters based on the termination criteria. Descriptions of the PEST control variables guiding the LM process are provided in Doherty (2008a, 2008b).

When LM interferes with the regularization, it can be turned off by setting RLAMBDA1 to 0 and NUMLAM to 1. Alternatively, by setting RLAMFAC to a negative number, a starting initial LM value near 0 can be allowed to grow rapidly if needed. This latter approach may be of most general use as it recognizes the value of starting with low Marquardt parameter values (often appropriate when regularization is used) but giving flexibility to the selection of an optimal LM (Doherty 2008b). The end result is that the algorithm can move from a too restrictively high lambda value when the algorithm is bogged down. The calculation of the LM can be further tuned using these and additional PEST control variables. However, such a discussion is outside the scope of this work, and the reader is referred to Doherty (2008a, 2008b) for additional information.

Illustrations with a Synthetic Problem

A synthetic problem illustrates the effects of control variables on a pilot point calibration and is described in detail by Hill and Tiedeman (2007, p. 21); the parameterization, including the pilot point configuration, was initially described by Hill (2006b) and expanded by Hill (2008).

Problem Description

The example consists of a steady-state flow problem with two horizontal aquifers separated by a confining bed (Figure 1) with vertical hydraulic conductivity specified by parameter VK_CB. The upper aquifer is homogeneous (horizontal hydraulic conductivity parameter HK_1), and the lower aquifer has "a mild degree of heterogeneity" (Hill and Tiedeman 2007, p. 21) in the form of a linear increase in horizontal hydraulic conductivity from 4.4×10^{-5} m/s to 3.95×10^{-4} m/s across the domain. This heterogeneity is imposed through a MODFLOW-2000 (Harbaugh et al. 2000) multiplier array with a set

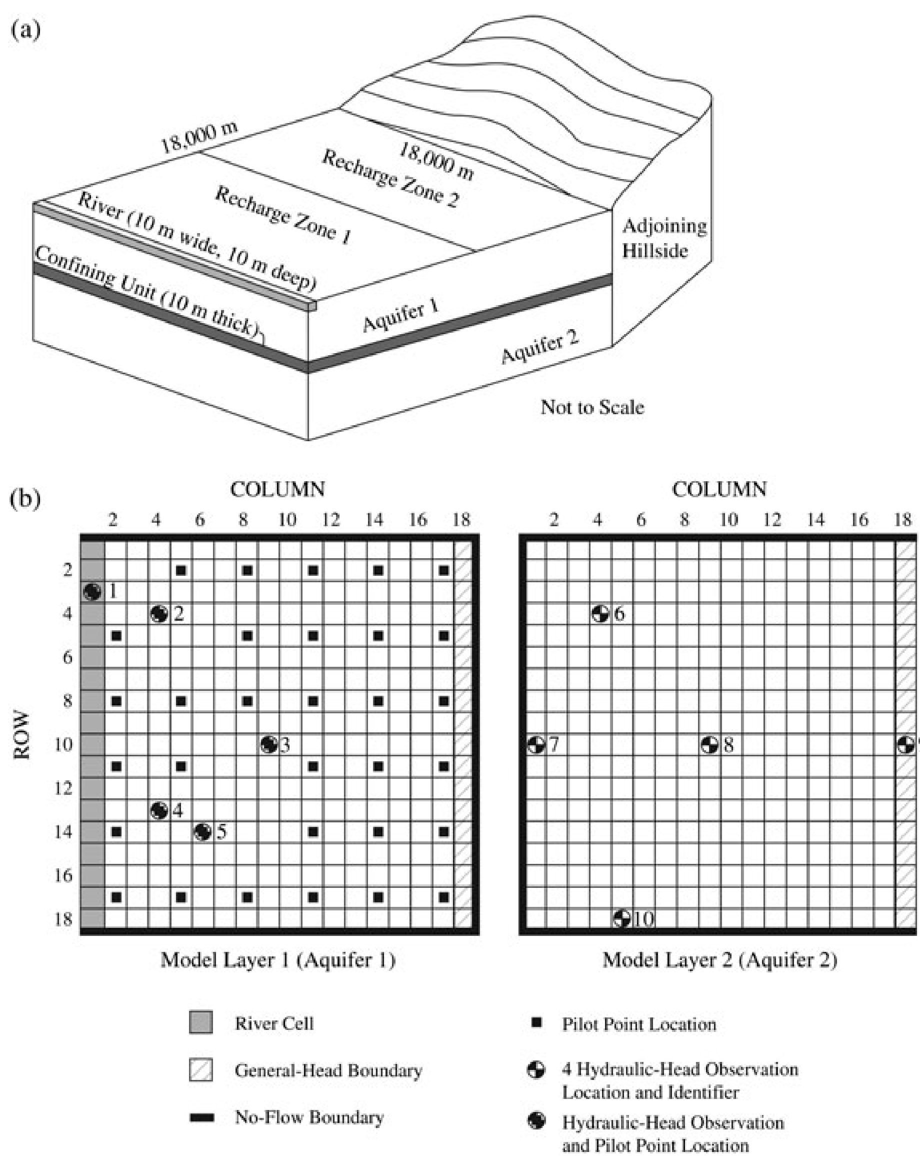

Figure 1. Synthetic problem described in Hill and Tiedeman (2007, p. 21) and used in this Methods Note. Figure modified from Hill (2008).

value specified for each model node that reflects the hydraulic conductivity trend. This relative gradient in hydraulic conductivity is thus fixed, and hydraulic conductivity at every node in the layer is controlled by a single parameter value (HK_2).

The synthetic domain is bounded on one side by a hillslope divide with a head value of about 177 m and on the other by a river with a head value of 100 m (hydraulic conductivity parameter K_RB). The other two boundaries and the bottom of the model are no-flow boundaries. Two recharge zones (parameters RCH_1 and RCH_2) provide water into the model from the top; a general head boundary provides influx from the hillslope; and in the uppermost model layer 1, a river opposite from the hillslope allows discharge (Figure 1). Because the riverbed conductance and aquifer vertical K parameters are correlated and cannot be simultaneously estimated using the observation data set, Hill (2008) used prior information on VK_CB and K_RB (set using their true values) to constrain their estimated value and obtain a unique set of optimal parameters.

Pilot points (36) are placed in the upper model layer 1 to represent HK_1 and regularized using a "preferred homogeneity" strategy. The 36 pilot points, two recharge values, and the K_RB, VK_CB, and HK_2 combined result in 41 parameters to be estimated. The 10 head measurements and the single river flow measurement provide a total of 11 observations. Similar to Hill (2008), this version of the model, including pilot points, is used to explain the use and behavior of the control variables discussed previously both with and without measurement error on the observations. PEST and MODFLOW input and output files for all simulations described here are provided in Supporting Information.

This synthetic model allows comparisons to be made between the estimated parameter values and the known true values. This comparison is quantified as "parameter error" following the convention of Hill (2008):

$$PE = 100\% \frac{(p_{\mathrm{true}} - p_{\mathrm{estimated}})}{p_{\mathrm{true}}} \tag{2}$$

where p_{true} is the true parameter value and $p_{\mathrm{estimated}}$ is the estimated parameter value.

Challenges Encountered in Previous Work

In this work, we reexamine both a test case with no errors on observations and a second case with observations blurred with measurement noise from Hill (2006b, 2008). This reexamination was motivated by challenges reported in Hill (2008). Specifically, significant deviation from the true homogeneous case was observed due to overfitting in a case where observations were considered error free, and significant errors were reported in estimating hydrogeologic parameters when observations included error.

In the case with no observation noise, Hill (2008) reported parameter errors ranging from −57% to 86% when using default Marquardt control variable settings, concluding that interference between LM and regularization were responsible for the errors. In a synthetic case like this, it is possible to generate observations that are truly without noise by running the forward model with the known correct parameters and using the results from that forward model run as observations. This was the clear intention of Hill (2008), but in fact, there was a small difference between the observations used by Hill (2008) and the results obtained by running the model with the correct parameters. The difference is small but significant in this synthetic case because it is possible to set PHIMLIM at a level assumed to be representative only of roundoff error that is actually lower than the misfit that would be obtained even when running the forward model with the correct parameters. We show below that proper adjustment of regularization control variables and a *truly* error free set of observations (in other words, observations that would be obtained perfectly with the correct forward model parameters) enables us to obtain results with parameter percent errors less than 0.03% even when using Marquardt lambda.

In the second case, two issues with the synthetic model setup colored the reported results. First, normally distributed errors with zero mean and standard deviation of 1.001 m were applied to the observations, which resulted in flipping the vertical gradient from upward to downward (and vice versa) in both well locations that have observation values in both the upper and lower layers (well pair 3 and 8 and well pair 2 and 6). It is conceivable that such high measurement error is reasonable in the horizontal context. However, it is likely unreasonable for the vertical context as in many cases the direction of the vertical gradient is known even if the magnitude is not, if for no other reason than systematic error (e.g., surveying error) would apply nearly equally at the same well nest and would be removed by the differencing. Thus, it was unreasonable to expect a correctly estimated parameter set when the only observed flow directions in the model domain were entered into the problem as the reverse of the true vertical flow directions. To overcome this artifact, we kept the spirit of the error case but considered an alternative set of measurement errors. That is, we applied normally distributed errors with a mean of 0 and a standard deviation of 0.3 m. This is still an appreciable level of measurement error for head observations, but it does not violate the conceptual model setup and allows the observed vertical gradients to have the same direction (though not magnitude) as the true flow direction.

Second, in all examples provided in Hill and Tiedeman (2007) and Hill (2008), the true recharge in zone 2 should exceed the recharge in zone 1, but the initial values specified have zone 1 exceeding zone 2. This reversal of recharge relative magnitude imposes a level of correlated structural error, which is typically defined as error due to formulation of the problem that cannot be reduced simply by changing parameter values. Correlated structural error is challenging for a regression algorithm because a fundamental assumption of regression-based parameter estimation methods is uncorrelated, random measurement errors. Incorrect zone relative rank (the magnitude of one vs. the other) results in correlated errors from the start that are extremely robust such that the regression often encounters significant difficulty overcoming the correlation of the errors. Structural error is typically discussed in the context of the model structure and not by parameter values. However, the reversal of recharge rank in such a simple framework amounts to routing of water in a totally different way than reality with a similar impact as routing water using a no-flow barrier or other structural device. In this case, therefore, the results of severely misspecified parameter values are manifest in the same way as a structural error with a similar impact on the model calibration problem. In the case including measurement error in Hill (2008), the reversal of gradients in the observations discussed in the previous paragraph prevents correct relative recharge magnitudes regardless of starting values. The combined adverse effects of reversed recharge magnitudes in starting values and the gradient reversal errors due to measurement errors are not discussed in Hill (2008); however, it should be noted that issues in the "optimal" parameter estimates reported therein are artifacts of the errant vertical gradients and from reversed recharge starting value ranks.

An alternative approach is starting uncertain parameters at expected ballpark average values and to evaluate parameter relative magnitudes after parameter estimation has been completed rather than presupposing them at the beginning of the process. During this work, it became apparent that many of the performance issues presented in Hill (2008) could be mitigated by this approach to starting values. Nonetheless, we explore the synthetic problem here using the reversed magnitude starting values as in the original case to highlight the flexibility and power of the Tikhonov regularization for constraining poorly posed problems.

Example with No Observation Error

Four scenarios were examined in the error-free case and are discussed here. Error-free observations are obtained by running the forward model and recording the observation values. Five additional scenarios and all files necessary to run the simulations are presented in the Supporting Information. Table 1 shows the starting values and three permutations of regularization controls that we discuss in this Methods Note. Three major characteristics were considered in all eight possible permutations in the subcases: Marquardt parameter (LM), inclusion of prior information (PI), and regularization (Regul). Table 2 shows the results for each subcase.

In cases 1e and 1h, regularization is used without the LM method. In both cases, with and without prior information, homogeneous results are found for HK_1 and all other parameters are estimated with parameter error of less than 0.32%. These cases converged in less than 50 iterations. In case 1g, using both the regularization and the LM method greatly slows convergence to the point that, using generally accepted stopping criteria, the PEST algorithm terminates due to slow convergence before reaching a good solution. As a special case, case 1g was modified as case 1g_long to remove restrictions on convergence rate, and using regularization, LM and prior information, the exact parameter values were estimated. This highlights that LM used in conjunction with regularization is stable and can provide the correct parameter estimates but may take an inordinate number of parameter estimation iterations (125 in this case vs. 10 iterations for case 1g).

Example Including Measurement Error

An example case with measurements blurred by error is included to illustrate the behavior of regularization of the pilot points for HK_1 (upper aquifer horizontal hydraulic conductivity) and the influence of control variable PHIMLIM. Results are shown in Table 3. For this section, the observation weights were set at the inverse of the measurement standard deviation. As a result, PHIMLIM equal to NOBS should be a reasonable value according to the discussion previously. In general, however, most estimates of σ are optimistic rather than conservative, so as a guideline, PHIMLIM = NOBS is a good starting point. The starting values were the same as in the error-free case (Table 1).

In these examples, only HK_1 and the recharge values are unknown and the impact of specifying PHIMLIM is clear. When PHIMLIM is set too low, the pilot points are too free to vary in an effort to fit the noisy measurements. This is the classic case of overfitting in which heterogeneity is imposed in the HK_1 parameter field to chase the noise in the measurements even though a homogeneous solution is more appropriate. Using PHIMLIM = NOBS (11 in this case), a homogeneous answer is found as expected with the HK_1 value within 0.35% of the true value and recharge within 0.56% of the true values.

Summary and Conclusions

Pilot points with Tikhonov regularization provide a great increase in the power to calibrate models using observations and soft knowledge of the system. However, with great power comes great responsibility. The objective of this Methods Note was to provide guidance for using regularized parameter estimation in the context of pilot points as implemented in PEST.

The most important control variable related to the strength of regularization is PHIMLIM, serving the dual purposes of controlling the strength of regularization and informing the algorithm of the user's expected level of agreement between observations and corresponding model results (misfit). Excessively high values of PHIMLIM result in more regularized solutions with inadequate reduction of misfit, whereas low values may reduce misfit dramatically by overfitting, resulting in geologically unreasonable parameters. PHIMACCEPT and FRACPHIM play secondary fine-tuning roles by adjusting the way PHIMLIM influences the PEST algorithm.

When regularization is used, conflicts can arise with respect to the LM process. When measurement uncertainty is small and regularization is used, it is often important to increase the Marquardt parameter range and lower its minimum value to avoid dramatic slowing of the algorithm.

Through a synthetic problem, we illustrated the behavior of pilot points implementation and made recommendations to overcome pitfalls that can be encountered

Table 1
Definitions of Scenarios Used in the Synthetic Examples for the Error-Free Case

	Starting Parameter Values								
Case	K_RB	VK_CB	HK_2	RCH_1	RCH_2	HK_1	LM?	PI?	Regul?
Truth	1.0E−03	2.0E−07	4.4E−05	31.526	47.304	4.0E−04	—	—	—
Case 1e							No	No	Yes
Case 1g							Yes	Yes	Yes
Case 1g_long	1.2E−03	1.0E−07	4.0E−05	63.072	31.536	3.0E−04	Yes	Yes	Yes
Case 1h							No	Yes	Yes

Table 2
Results of Parameter Estimation When Measurements Are Known Perfectly (error-free case)

Parameter / Case	Calibrated Parameter Set: HK_1 ($\times 10^{-4}$) Minimum	Maximum	Average	SD	K_RB	VK_CB	HK_2	RCH_1	RCH_2
Truth		4.000		—	1.00E−03	2.00E−07	4.40E−05	31.526	47.304
Case 1e	4.002	4.002	4.002	0.000	1.00E−03	2.01E−07	4.41E−05	31.538	47.368
Case 1g	1.451	6.198	2.495	1.250	9.99E−04	1.99E−07	3.43E−05	54.876	24.127
Case 1g_long	4.000	4.000	4.000	0.000	1.00E−03	2.00E−07	4.40E−05	31.548	47.291
Case 1h	4.004	4.004	4.004	0.000	1.00E−03	2.00E−07	4.40E−05	31.588	47.319
Parameter / Case	**Percent Error Results: HK_1 Minimum**	**Maximum**	**Average**	**SD**	**K_RB**	**VK_CB**	**HK_2**	**RCH_1**	**RCH_2**
Case 1e	−0.06	−0.06	−0.06	—	−0.05	−0.32	−0.26	−0.04	−0.14
Case 1g	63.73	−54.96	37.62	—	0.07	0.32	22.02	−74.06	48.99
Case 1g_long	−0.01	−0.01	−0.01	—	0.00	0.02	0.08	−0.07	0.03
Case 1h	−0.09	−0.09	−0.09	—	0.00	0.00	0.04	−0.20	−0.03

Note: Summary statistics of the estimated parameter fields are presented to evaluate the level of heterogeneity including minimum, maximum, mean, and SD.

when using pilot points with regularization. The interaction between LM and regularization can be important to consider, though the artifacts of specifying the LM value too high shown here are most likely to occur when the problem is very well conditioned and measurement and model error are small. This set of conditions is not expected to be the norm in a highly parameterized model of a real-world setting. In the presence of noise, avoiding overfitting is paramount and it is incumbent on the modeler to set the control variable PHIMLIM so that the optimal parameter set reflects a geologically and physically realistic conceptualization of the system. General guidelines are provided for selecting and adjusting PHIMLIM and related control variables.

Finally, we emphasize the importance of a reasonable starting parameter set when moving from model calibration using simple zonation to a more highly parameterized model such as those using regularized inversion approaches like pilot points. The flexibility afforded can be constrained using appropriate regularization, but the

Table 3
Results of Parameter Estimation with Measurements Blurred by Errors with a Mean of Zero and a Standard Deviation of 0.3 m

Case	Parameter: PHIMLIM	Calibrated Parameter Set: HK_1 ($\times 10^{-4}$) Minimum	Maximum	Average	SD	RCH_1	RCH_2
Truth	—		4.000		—	31.526	47.304
Very low	0.11	0.216	13.016	3.140	2.418	47.876	31.034
Low	5.5	3.847	4.530	4.027	0.117	31.481	47.552
NOBS	11	4.012	4.013	4.012	0.000	31.541	47.498
Case	**Parameter: PHIMLIM**	**Percent Error Results: HK_1 Minimum**	**Maximum**	**Average**	**SD**	**RCH_1**	**RCH_2**
Very low	0.11	94.59	−225.39	21.50	—	−51.86	34.40
Low	5.5	3.82	−13.25	−0.67	—	0.14	−0.52
NOBS	11	−0.30	−0.32	−0.31	—	0.56	−0.41

parameter estimation process will be quicker and more robust if the starting values reflect good "ballpark" estimates for the site area. Moreover, most inverse algorithms are predicated on the assumptions that the problem can be linearized and that the errors in observations are independent and uncorrelated. As a result, the linearization is only reasonable when the parameters are reasonably close to their true values. Perhaps more importantly, incorrectly specified starting values can introduce a level of correlated structural noise that eclipses measurement noise. As a result, it can be difficult or impossible for any gradient-based parameter estimation method to estimate reasonable parameters when starting so far out of the ballpark. In cases such as that illustrated in the synthetic example of this work, where ranking of parameters (recharge in this case) can have such a dramatic impact on the structural noise of the model, we recommend either setting the ranked values equal or trying the inverse method with each potential ranking.

Acknowledgments

The authors wish to thank Mary Hill of USGS for bringing the challenges associated with this synthetic example to our attention and for graciously sharing her input files. We also thank Matt Tonkin of S. S. Papadopulos and Associates Inc., David Dahlstrom of Barr Engineering, and John Doherty of Watermark Numerical Computing for their suggestions. We also acknowledge reviews by Steffen Mehl, Andreas Alcolea, Mary Anderson, and one anonymous reviewer, which also improved this manuscript. The work was funded by the USGS Office of Ground Water, the National Academies Postdoctoral Research Associateship, and S. S. Papadopulos and Associates, Inc. The use of trade, product, or firm names in this report is for descriptive purposes only and does not imply endorsement by the U.S. government.

Supporting Information

Additional supporting information may be found in the online version of this article.

The supporting information includes expanded discussion of the synthetic case studies examined in this Methods Note, appendices discussing aspects of the mathematics in further detail, figures and tables related to the synthetic case studies, and a complete set of input files for all synthetic cases evaluated in this work for readers to experiment with.

Appendix S1. Review of objective function calculation.

Appendix S2. Calculating a target PHIMLIM.

Table S1. Control variables controlling regularization and the Gauss-Levenberg-Marquardt process. Descriptions of these variables and default values are from Doherty (2008a, 2008b). The reader is urged to consult Doherty (2008a, 2008b) for definitive definitions and descriptions. Control variables in italics are discussed in some detail in the text of this article.

Table S2. Definitions of scenarios used in the synthetic examples.

Table S3. Results of parameter estimation when measurements are known perfectly (error-free case).

Table S4. Results of parameter estimation with measurements blurred by errors.

Figure S1. Synthetic problem described in Hill and Tiedeman (2007, p. 21) and used in this Methods Note. Figure modified from Hill (2008).

Figure S2. HK fields in model layer 1 for cases B, D, F, and G (clockwise from upper left) of the error-free test cases.

Figure S3. HK fields in model layer 1 for cases PPVL, PPL, PPRVL, and PPRL (clockwise from upper left) of the test cases blurred with measurement errors.

References

Alcolea, A., J. Carrera, and A. Medina. 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources* 29, no. 11: 1678–1689.

Aster, R.C., B. Borchers, and C.H. Thurber. 2005. *Parameter Estimation and Inverse Problems, International Geophysics Series*, vol. 90. Amsterdam, The Netherlands: Elsevier Academic Press.

Carrera, J., and S.P. Neuman. 1986. Estimation of aquifer parameters under transient and steady state conditions. 1. Maximum likelihood method incorporating prior information. *Water Resources Research* 22, no. 2: 199–210.

de Marsily, G., C. Lavendan, M. Boucher, and G. Fasanino. 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In *Geostatistics for Natural Resources Characterization*, vol. series C 182, ed. G. Verly, M. David, A.G. Journel, and A. Marechal, 831–849. Boston, Massachusetts: NATO.

Doherty, J. 2008a. *PEST, Model Independent Parameter Estimation*. User manual: 5th ed. Brisbane, Australia: Watermark Numerical Computing. http://www.sspa.com/PEST/ (accessed October 2, 2008).

Doherty, J. 2008b. *PEST, Model Independent Parameter Estimation*. Addendum to user manual: 5th ed. Brisbane, Australia: Watermark Numerical Computing. http://www.sspa.com/PEST/ (accessed October 2, 2008).

Doherty, J. 2003. Ground water model calibration using pilot points and regularization. *Ground Water* 41, no. 2: 170–177.

Ginn, T.R., T.D. Scheibe, H. Haeri, and C.N. McClain. 2007. Paper 40–1: An expanded survey of groundwater modeling practitioners about how they quantify uncertainty: Which tools they use, why, and why not. *Geological Society of America Abstracts With Programs*, 39, no. 6: 40–1.

Gomez-Hernandez, J.J. 2006. Complexity. *Ground Water* 44, no. 6: 782–785.

Harbaugh, A.W., E.W. Banta, M. Hill, and M.G. McDonald. 2000. MODFLOW-2000, the U.S. Geological Survey modular ground-water model—User guide to modularization concepts and the ground-water flow process. USGS Open File Report 00–92. Reston, Virginia: USGS.

Hill, M.C. 2008. Data error and highly parameterized groundwater models. In *Calibration and Reliability in Groundwater Modelling: Credibility of Modelling*, IAHS Publ. 320, ed. J. C. Refsgaard, K. Kovar, E. Haarder, and E. Nygaard. 316–321. Copenhagen, Denmark: IAHS.

Hill, M.C. 2006a. The practical use of simplicity in developing ground water models. *Ground Water* 44, no. 6: 775–781.

Hill, M.C. 2006b. Observation errors and highly parameterized models. *Eos Transactions American Geophysical Union* 87, no. 52:. Fall Meeting Supplement, Abstract H31F–06.

Hill, M.C., and C.R. Tiedeman. 2007. *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*. Hoboken, New York: Wiley-Interscience.

Hunt, R.J., J. Doherty, and M.J. Tonkin. 2007. Are models too simple? Arguments for increased parameterization. *Ground Water* 45, no. 3: 254–262.

Neuman, S.P. 1973. Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty. *Water Resources Research* 9, no. 4: 1006–1021.

Ramarao, B.S., A.M. Lavenue, G. Demarsily, and M.G. Marietta. 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1.Theory and computational experiments. *Water Resources Research* 31, no. 3: 475–493.

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

Tikhonov, A.N. 1963b. Regularization of incorrectly posed problems. *Soviet Mathematics Doklady* 4, 1624–1637.

Tikhonov, A.N., and V.I.A. Arsenin. 1977. *Solutions of Ill-Posed Problems*, Scripta Series in Mathematics, Winston. Washington, New York: Halsted Press.