ground water

Issue Paper/

Are Models Too Simple? Arguments for Increased Parameterization

by Randall J. Hunt[1], John Doherty[2], and Matthew J. Tonkin[3,4]

Abstract

The idea that models should be as simple as possible is often accepted without question. However, too much simplification and parsimony may degrade a model's utility. Models are often constructed to make predictions; yet, they are commonly parameterized with a focus on calibration, regardless of whether (1) the calibration data can constrain simulated predictions or (2) the number and type of calibration parameters are commensurate with the hydraulic property details on which key predictions may depend. Parameterization estimated through the calibration process is commonly limited by the necessity that the number of calibration parameters be smaller than the number of observations. This limitation largely stems from historical restrictions in calibration and computing capability; we argue here that better methods and computing capabilities are now available and should become more widely used. To make this case, two approaches to model calibration are contrasted: (1) a traditional approach based on a small number of homogeneous parameter zones defined by the modeler a priori and (2) regularized inversion, which includes many more parameters than the traditional approach. We discuss some advantages of regularized inversion, focusing on the increased insight that can be gained from calibration data. We present these issues using reasoning that we believe has a common sense appeal to modelers; knowledge of mathematics is not required to follow our arguments. We present equations in an Appendix, however, to illustrate the fundamental differences between traditional model calibration and a regularized inversion approach.

Introduction

Albert Einstein observed that our approach to problem solving should be "as simple as possible but not simpler." This philosophy appears to be a firm tenet for modeling environmental systems. How this philosophy should be applied in practice is not, however, universally agreed upon. Problems associated with too little complexity are well known, as evidenced by the popularity of complex simulation tools beyond simple analytical solutions. However, Anderson (1983) observed that modelers often invoke complexity beyond that which is warranted from a conceptual understanding gleaned from field data, such that often the "emperor has no clothes." That is, the model is not as good as it might appear, given its cost and reams of output. Freyberg (1988) noted that in a modeling class he taught, predicted system response was better simulated with more parsimonious but less well-calibrated models than with models calibrated using a large number of parameters to obtain a good fit (a phenomenon often referred to as "point calibration"). There appear to be diminishing returns whereby some level of parameter complexity improves our simulation capabilities, but too much leads to instability, nonuniqueness, long run times, and an increased potential for predictive error.

What is the optimal level of parameterization? The emphasis within the environmental modeling community is often "the simpler the better." For example, Hill (1998, 2006) lists parameter parsimony—essentially, the use of a small number of parameters—as the number one guideline for effective model calibration. An American

[1]Corresponding author: U.S. Geological Survey, 8505 Research Way, Middleton, WI 53562; (608) 828-9901; fax (608) 821-3817; rjhunt@usgs.gov

[2]Watermark Numerical Computing, 336 Cliveden Avenue, Corinda, 4075, Brisbane, Australia; johndoherty@ozemail.com.au

[3]Department of Civil Engineering, University of Queensland, St Lucia, Australia 4068.

[4]S.S. Papadopulos & Associates Inc., 7944 Wisconsin Avenue, Bethesda, MD 20814; matt@sspa.com

Received May 2006, accepted January 2007.

doi: 10.1111/j.1745-6584.2007.00316.x

Geophysical Union special session on model complexity concluded that "if models are kept in the context of their objective, we should feel comfortable resisting the siren of complexity and construct simpler, less encompassing models" (Hunt and Zheng 1999). This logic seems unassailable and can produce models that provide some level of insight. However, questions linger about the focus on a priori parsimony, such as the following: Did this approach get the most information from the calibration data? Does the model adequately reflect the response of the physical system, or have bias and/or uncertainty been introduced into model predictions as a result of "handcuffing" to a simplified parameter structure?

Clearly, it is critical to consider the modeling objective(s) when answering these questions. Predictions requiring "ballpark" estimates based on bulk system properties may be made using simple methods—even hand calculations (for example, Haitjema 1995, 2006). For other predictions such as contaminant transport, however, complex simulation capabilities and/or detailed parameterization may be required. There has been some previous work on the utility of model calibration under such circumstances, with some researchers arguing that the notion of the calibrated model has no place in environmental simulation since uniqueness is always achieved at the cost of error-inducing simplification(s). For example, Woodbury and Ulrich (2000), Gómez-Hernández et al. (2003), Gómez-Hernández (2006), and others argue that a large number of model runs should be used within a probabilistic Bayesian framework to explore the range of predictive possibilities rather than making a discrete prediction. Although these arguments have merit, for better or for worse calibrated models commonly form the basis of environmental decision making and it is within this context that our discussion takes place.

There appears to be a middle ground between model calibration using relatively few zones and very complex parameterization using stochastic or other methods. To illustrate this middle ground, we contrast a traditional approach to calibration based on zones of piecewise constancy with an approach based on regularized inversion. The discussion focuses on model calibration, but also touches upon the use of calibrated models to evaluate predictive error and assess the potential relative worth of future data collection. The discussion is presented in terms of words and concepts. For the reader interested in the underlying mathematics, an Appendix is included that contains mathematical formulations and references that explore these concepts in greater detail.

The Traditional Approach

In order to estimate the values of parameters used in a model, comparisons are made between real-world data (measurements of ground water elevations, baseflows, etc.) and simulated equivalents. This can be accomplished using manual trial-and-error or inverse codes that "automate" the trial-and-error process (Anderson and Woessner 1992). Although the value of intuition developed by manually manipulating model parameters should not be understated, Poeter and Hill (1997) among others suggest that automated methods are often superior. Our discussion focuses on automated calibration methods that seek a best fit by minimizing the weighted squared differences between measured data and their simulated equivalents.

For the purpose of this discussion, it is sufficient to understand that traditional automated approaches require that the calibration problem be "well posed" in order to obtain a unique set of parameters that provide the "best fit" between model outputs and observed equivalents (e.g., Draper and Smith 1998). This usually means that the number of estimated parameters is fewer than the number of observations plus the number of items of independent information about parameters (often referred to as "prior information"—see, for example, Cooley 1982). This requirement is dictated at least in part by (1) historical restrictions in computing capability; (2) the fact that observation data often contain redundant information for calibration purposes; and (3) the traditional use of direct matrix inversion to solve the linear system of equations that is required for estimation of parameters (see Appendix). The traditional restriction of small numbers of parameters often results in many important but subjective decisions during the model construction process.

In truth, the system to be modeled likely contains far more spatial variability in hydraulic properties than can possibly be estimated through the calibration process. To achieve stable and unique calibration using traditional methods, a priori, user-defined parameter "parsimony" is used to subdivide the model domain into a small number of zones with uniform parameter values inferred from geological knowledge about the system. In some cases, such zones simply represent areas where few or no data exist. If this a priori "lumping" leads to more parameters than observations and/or excessive parameter insensitivity and/or correlation, the modeler combines zones to reduce the number of estimated parameters and/or fixes some parameters at "reasonable" values. Calibration often ceases once an acceptable fit is achieved. In the case where data are plentiful, it may become apparent that the level of parsimony is too great (for example, if it is obvious that better fits between measured and simulated values can be obtained). New zones are then introduced in a somewhat ad hoc fashion to reduce the misfit incurred by the assumption of piecewise property uniformity on too broad a scale. Unless rigorous methods are employed to simultaneously estimate parameter values and structure (e.g., Sun and Yeh 1985; Eppstein and Dougherty 1996; Zheng and Wang 1996; Tsai et al. 2003), the modeler faces agonizing choices over which parameters to lump, which to fix, what values to assign to fixed parameters, when the fit is good enough, and when the fit is too good—often using arbitrary criteria that are hard to convey, document, or defend. All of these decisions are made while recognizing that every decision on what to fix or lump may introduce "hardwired error" into model predictions.

Problems with this approach extend beyond the difficulty of identifying the optimal level of parameter simplification. First, rigid a priori parameter parsimony based on an existing understanding of property variability can make it difficult to improve that understanding through

calibration, since the calibration data may be speaking to "ears" that are not listening. Second, a fundamental motivation for developing models—to test hypotheses that would be impossible to test in the real world—may be undermined if parsimony limits the capability of the model to analyze variations and alternatives. Third, property variability beyond that which can be uniquely inferred is ignored: Moore and Doherty (2005) show that this may undermine efforts to analyze model predictive uncertainty.

Many modelers believe that a well-calibrated model will naturally provide accurate predictions. However, Moore and Doherty (2006) demonstrate that transport predictions made by a model that calibrates perfectly to ground water elevation data can be 100% wrong as a consequence of the simplifications required to achieve a unique calibration. Thus, it is important that model predictions be accompanied by an estimate of their likely error. One way to evaluate the potential magnitude of predictive error is through implementing a constrained predictive maximization/minimization process through which a prediction of interest is "stretched" while ensuring that a suitably good fit between model outputs and field measurements is maintained (Vecchia and Cooley 1987). Unfortunately, when these calculations are based on the traditional use of a small number of parameters, they can be susceptible to errors incurred because simplifications required to obtain a unique calibration create a form of "structural noise" that commonly exceeds measurement noise. This structural noise can be evaluated using methods such as those developed by Cooley (2004) and Cooley and Christensen (2006), and this can, in turn, facilitate proper estimation of potential predictive error. These methods can be difficult to implement in complex modeling contexts, however, and do not directly address the problem of minimization of that error or help identify conditions that could lead to extreme values of a prediction—despite the fact this might be of primary interest to decision makers.

The ability of models to evaluate the utility of future data collection exemplifies the underlying issue. If structural noise is not accounted for, predictive error calculations used to compare the relative worth of different potential data gathering strategies (e.g., Tonkin et al. 2007) may be misleading. When using traditional methods, it follows that the fewer data available within a study area, the greater the need to acquire extra data. Yet, the fewer the data, the greater the level of parsimony required to ensure that the number of observations exceeds the number of parameters. Moreover, as detailed previously, the fewer the number of parameters, the greater the contribution to predictive error from structural noise. One could argue that, as a consequence, when the need for more data is most urgent, the use of precalibration parsimony is most questionable.

Potential weaknesses of precalibration parsimony may be most evident when extended from situations where very few data are available to the extreme case where none is available at all. In this case, should a model have no parameters? This contravenes common practice, where on occasions that predictions must be made with no calibration support, highly parameterized models are often used to account for parameter variability—often together with Monte Carlo and related techniques for exploration of predictive variability (e.g., U.S. EPA 2001). Why is it that when a limited number of data become available, methods that recognize the complexity inherent to real-world systems are disregarded in favor of methods based upon ad hoc simplification and parameter parsimony?

An Alternative to the Traditional Approach: Regularized Inversion

Let us be clear that there is no obvious formulation of model calibration that previous generations of modelers have overlooked. Model calibration is fundamentally nonunique because there is always more spatial property variability—and hence more parameters—than observation data can constrain. The fundamental tension remains: parameterization of true hydraulic property detail is impossible; yet, "calibration" implies that a unique parameter field is obtained one way or another. Therefore, the question is not whether a model—which is always a simplification of reality—should simplify parameter variability. Rather, it is how this simplification should be performed.

Advances in computing power, equation solution techniques, and techniques for formulating the inverse (i.e., parameter estimation) problem have allowed more sophisticated methodologies to be used in other fields such as medical imaging and geophysical prospecting than are typically used in ground water modeling. These capabilities present an opportunity to use parsimony to estimate parameters in a more rigorous way than is traditionally done using ad hoc parsimony described previously. Of these, "regularized inversion" (Engl et al. 2006) is described here. "Inversion" simply refers to the use of measured data (such as heads, fluxes, etc.) to estimate model parameters. The term "regularization" describes any process that makes a function more regular or smooth; it can be broadly interpreted as any method that helps provide an approximate and meaningful answer to an ill-posed problem. In this sense, traditional parsimony is an *informal* regularization strategy that reduces a variable world to a small number of model parameters. Hence, the need for regularization is not in question—rather the question is how should it be accomplished?

The basic tenet of regularized inversion described here is that the level of parameterization used in a model should not be unnecessarily restricted by ad hoc precalibration parsimony because that may contravene the original motivation for building the model. Thus, the calibration process should be flexible in order to be maximally responsive to information contained in the calibration data and in the assignment of parameters on the basis of that information. In doing so, regularized inversion provides a systematic and quantitative framework for achieving parameter simplifications, whereby the rationale for the simplification is formally constructed and decipherable. Regularized inversion incorporates two primary differences from the traditional approach described previously. These are now outlined.

The Level of Parameterization

The first difference is that more parameters can be assigned to a model because alternative techniques to direct matrix inversion are used to solve the system of equations formed during model calibration—techniques such as singular value decomposition. In practice, there are still practical limits to how many parameters are appropriate. Thus, rather than specifying a parameter for every model cell, for example, parameterization devices such as "pilot points" (de Marsily et al. 1984; RamaRao et al. 1995; Doherty 2003) are often used. In the pilot-point approach, parameter values are estimated at a number of discrete locations distributed throughout the model domain; cell-by-cell parameterization then takes place through spatial interpolation from the pilot points to the model grid or mesh. If necessary, pilot points can be grouped to represent geologic continuity where it is believed to exist and combined with the use of zones so that hydraulic property heterogeneity is preferentially expressed at zone boundaries. This approach provides a zonation scheme that it is not "hardwired" as in traditional approaches, and adding parameters can help the calibration process extract more information from the calibration data. The result, however, is many more parameters than is typical in traditional model calibration, which can lead to (1) parameter insensitivity and correlation, which in turn lead to solution nonuniqueness and (2) long run times. To overcome these issues, "regularization" is required. Some commonly used regularized inversion approaches are described below, which can be used by themselves or in combination.

Stabilizing the Calibration Process with Mathematical Regularization

In contrast to the more hardwired and static nature of the traditional approach, regularized inversion can enable knowledge of a study site to form a flexible mathematical regularization strategy that can be employed to attain parameter uniqueness and stability of the inversion process. Understanding of a site can enter into the calibration process through definition of a preferred system condition (for example "the hydraulic conductivity should have a value around 1 m/d" or "the hydraulic conductivity should be uniform in this area"). This condition is formally injected into the calibration process using "Tikhonov regularization" (Tikhonov and Arsenin 1977; Ory and Pratt 1995). In Tikhonov regularization, preferred system conditions are maintained if possible. However, minimized deviation from these conditions is acceptable if this is required in order to obtain a user-specified level of fit (Doherty 2003). Tikhonov regularization has been used in the calibration of a variety of ground water models (e.g., Skaggs and Kabala 1994; Liu and Ball 1999; Doherty 2003; van den Doel and Ascher 2006), achieving results that would not have been possible using traditional precalibration parsimony. The use of Tikhonov regularization can enable the estimation of a large number of parameters in a geologically reasonable manner. However, it does not relieve the computational burden of estimating many parameters, nor is it unconditionally numerically stable.

A second approach to stabilizing the estimation of a large number of parameters is to formally decompose the information on parameter sensitivities and correlations, and eliminate (effectively, fix) inestimable combinations of parameters that destabilize the inverse problem, while estimating the remaining parameter combinations. This approach to regularized inversion uses the truncated singular value decomposition (TSVD) matrix analysis technique to identify combinations of parameters that cannot be estimated using the available calibration data (insensitive parameters, for example, are included in this "calibration null space") and combinations of parameters that can be estimated on the basis of the available calibration data (these comprising the "calibration solution space"). The threshold for truncation is specified by the user. If too many combinations of parameters are estimated, the problem will still be numerically unstable; if too few parameters are estimated, the model fit may be unnecessarily poor and predictive error may be larger than an optimally parameterized model. While TSVD can provide stable and unique model calibration, it too does not alleviate the high computational burden incurred by the use of many parameters. Nor are parameter fields sometimes as "aesthetically pleasing" or geologically reasonable as they are for Tikhonov calibration where reasonableness is built into the regularization process through use of a preferred condition.

Tikhonov regularization can be combined with TSVD to simultaneously achieve the benefits of both. However, neither TSVD nor Tikhonov regularization, nor their use in combination, reduces the computational cost incurred by a highly parameterized model because sensitivities of model outputs must be calculated with respect to all parameters during each iteration of the calibration process regardless of whether they are sensitive or not. On some occasions, the use of adjoint sensitivity techniques (Townley and Wilson 1985; Clemo et al. 2003) can expedite the computation of these parameter sensitivities. However, if adjoint sensitivities are not available, the computational burden may be reduced through the use of a novel extension of subspace inversion techniques—hybrid Tikhonov-TSVD (Tonkin and Doherty 2005).

This regularization approach is a hybrid of Tikhonov and TSVD regularization that accelerates the process by solving the inverse problem in a parameter "subspace" computed using TSVD rather than solving the problem in true parameter space where the sensitivity of each parameter is required independently. Once estimable parameter combinations (referred to as "super parameters") are defined, derivatives of model outputs with respect to these parameter combinations are calculated through direct finite differencing of these combinations. Collectively, these combinations span the calibration solution space. In distinction from the traditional use of TSVD where this subdivision occurs in each iteration of the inverse process, the initial subdivision is retained throughout the inverse process. As the dimensionality of this space is normally smaller than that of the calibration null space

(comprising parameter combinations that remain unestimated and thus retain their original values), the number of model runs required per iteration for computation of derivatives can be reduced enormously. Unconditional numerical stability is also obtained, as insensitive and highly correlated parameter combinations are automatically excluded from the parameter solution space through the TSVD process through which super parameters are defined.

The appendix illustrates the form of the basic parameter estimation equations when Tikhonov (Equation A-6), TSVD (Equation A-8), and hybrid Tikhonov-TSVD (Equation A-9) regularization are employed. In many cases, solving the problem in a subspace of the important parameter combinations reduces the number of model runs required (often by more than an order of magnitude), removes the effects of insensitive parameters, and can obtain solutions to the inverse problem that are sufficiently close to the "true" solution to be useful.

In addition to extracting the more information from a calibration data set, regularized inversion as a calibration methodology allows quantification of potential prediction error using both linear (Moore and Doherty 2005) and nonlinear (Tonkin et al. in review) methods. The nonlinear approach employs an extension of the prediction minimization/maximization technique of Vecchia and Cooley (1987), in which constraints applied to parameters during the predictive analysis process ensure that model-to-measurement fits remain within expectations and that parameters remain geostatistically reasonable as predictions are maximized or minimized. Such analyses can be easily undertaken following model calibration using any of the regularized inversion methods discussed previously. Because many parameters are included in the predictive analysis process, errors that result from failure of the calibration process to "capture" system heterogeneity due to limitations in the calibration data are explicitly accounted for.

A final note regarding regularized inversion is that Moore and Doherty (2005, 2006) demonstrate that regardless of the method employed, the calibrated hydraulic property field is smoother or simpler than the true hydraulic property field. This is simply the cost of obtaining a unique solution to the inverse problem of model calibration. Nevertheless, this blurred image of the subsurface (e.g., McLaughlin and Townley 1996) so obtained often appears more geologically reasonable than an image comprising traditional zones of piecewise constancy since it lacks abrupt and artificial looking discontinuities. Furthermore, this "blurred" image of the subsurface encapsulates as much information as possible from the calibration data and is thus as representative a picture as can be expected given the observations available. This image can be improved upon, thereby making it sharper (and bringing hydraulic property detail into focus) by the acquisition of extra data. Put another way, regularized inversion (and model calibration in general) does *not* necessarily produce an accurate representation of real-world complexity. Rather, it reflects the complexity that is supported by the data—and commonly far more than can be inferred using traditional parsimony.

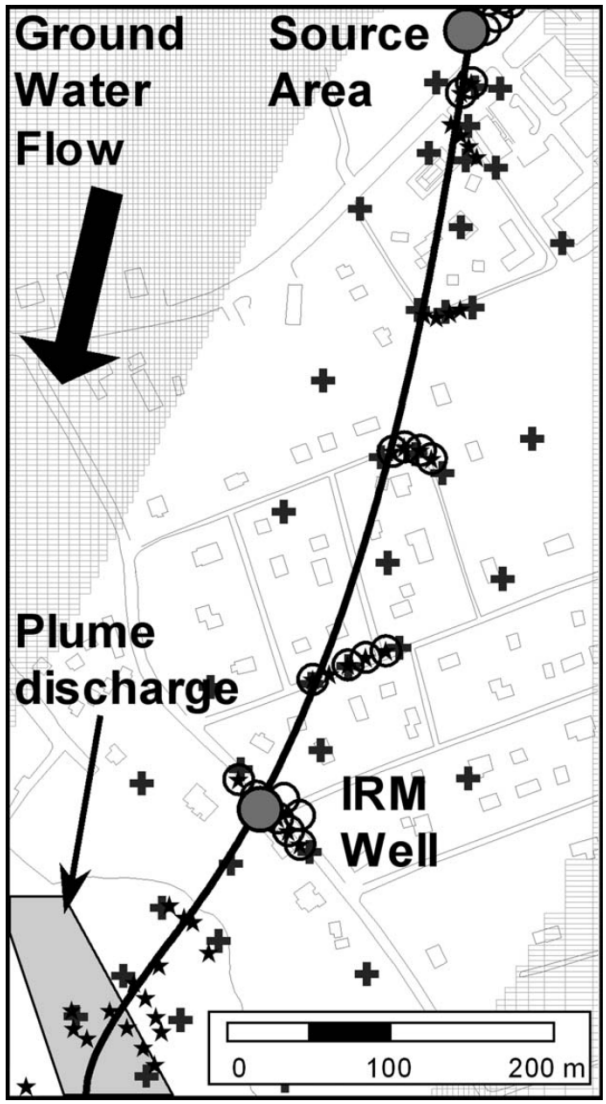

Figure 1. Active area of transport (white) from Tonkin and Doherty (2005) showing water level observation locations (open circles), MTBE observation locations (stars), Interim Remedial Measure (IRM) well (filled circle), and pilot points (crosses). A calibrated pathline from the regularized inversion is also shown (black line).

Comparison of Traditional and Regularized Inversion Approaches

To the authors' knowledge, only a small number of publications directly compare the two methods described previously. For example, Tonkin and Doherty (2005) compare a traditional zone–based parameterization with the hybrid TSVD regularized inversion, in the context of a ground water flow and transport model. A brief summary is provided here; the reader is directed to Tonkin and Doherty (2005) for more detailed discussion.

The problem was evaluated using a traditional approach using a single zone for each model layer and a regularized inversion approach that included 1195 base

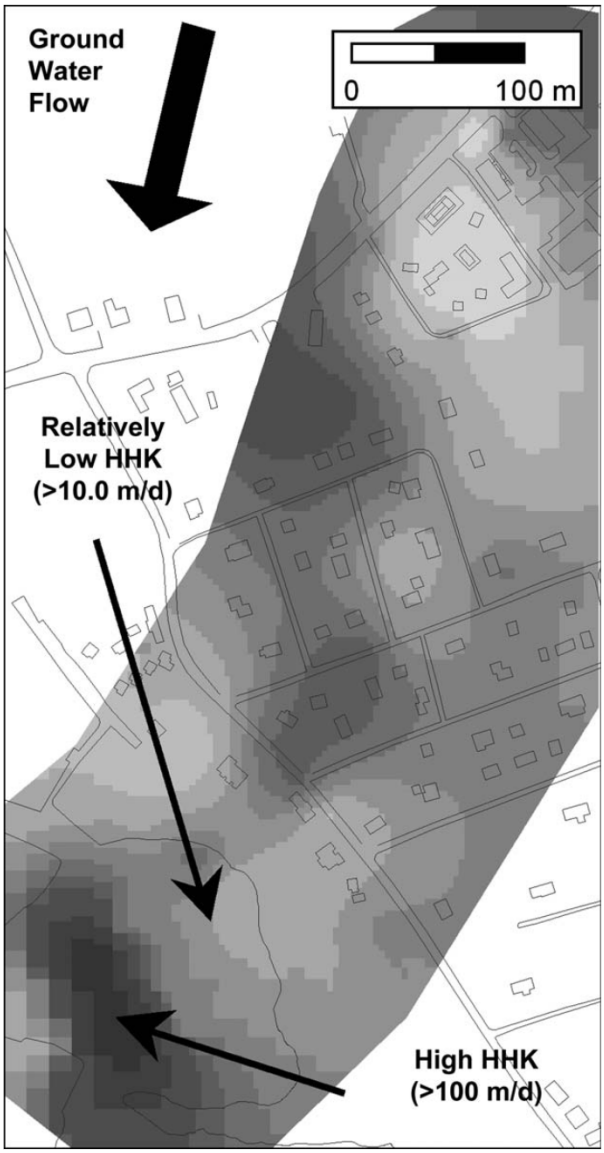

Figure 2. Regularized inversion calibrated horizontal hydraulic conductivity (HHK), model layer 1. For comparison, the traditional approach used one value of conductivity for the layer (Tonkin and Doherty 2005).

parameters, including horizontal and vertical hydraulic conductivity, recharge, porosity, boundary conditions, and parameters describing the contaminant source term. Each model was calibrated to ground water levels and contaminant concentrations collected from about 40 multilevel monitoring locations (Figure 1). Slug tests and boring logs indicate that there is geological heterogeneity present; however, the highly transmissive sand aquifer did not lend itself to a priori delineation of laterally and/or vertically contiguous parameter zones. Therefore, the results of the lumped parameter calibration were used as initial conditions for the regularized inversion.

The hydraulic conductivities estimated using regularized inversion were generally consistent with the field data and other independent sources of information (Figure 2), and produced an improved fit to the measured data (Table 1)—as might be expected. The most significant improvement in the objective function was achieved through improved fit of simulated and measured concentrations at monitoring wells. Plots of simulated and measured concentrations illustrate that the regularized inversion model reproduced sharp fronts better than the lumped model. This is most evident in "bubble plots" of simulated and observed methyl tert-butyl ether (MTBE) concentrations in monitoring wells (Figure 3). In these figures, the source area is located at right and the discharge area is located at left. The lumped parameter model produces a simulated plume that gives the appearance of simple spreading about a centerline; areas of low or high measured concentrations are not reflected in their simulated equivalents. The regularized inversion model (1) more accurately reproduces areas of higher and lower concentrations throughout the plume and (2) better matches the true location of the discharge of the plume to the surface water body.

Review of the parameters estimated through regularized inversion suggests that the improved fit was largely obtained by the introduction of spatial variability into hydraulic conductivity and porosity parameters. This is expected since using a large number of pilot points together with Tikhonov regularization enables the hybrid scheme to introduce such variability. Nonetheless, parameter values estimated in the inversion are generally within the range of values estimated from field testing. In particular, regularized inversion appeared to identify an area of low hydraulic conductivity that causes the plume to discharge further offshore than in the lumped calibration (Figure 2). Tonkin and Doherty (2005), and the authors of the current discussion paper, do not infer that regularized inversion identified the "true" parameter values and/or

Table 1
Summary of Traditional and Regularized Inversion Calibration (Tonkin and Doherty 2005)

Observation Group	Weighted Residuals from Lumped Calibration Using Traditional Methods	Weighed Residuals from Regularized Inversion Using Hybrid TSVD	Percent Reduction in Weighted Residual
MTBE[1] mass removal at interim remedial measure well	5.27×10^6	2.00×10^6	62
MTBE concentrations in observation wells	9.31×10^6	1.68×10^6	82
Water levels in observation wells	6.19×10^2	3.71×10^2	40
Composite objective function	1.46×10^7	3.68×10^6	74

[1]MTBE = methyl tert-butyl ether.

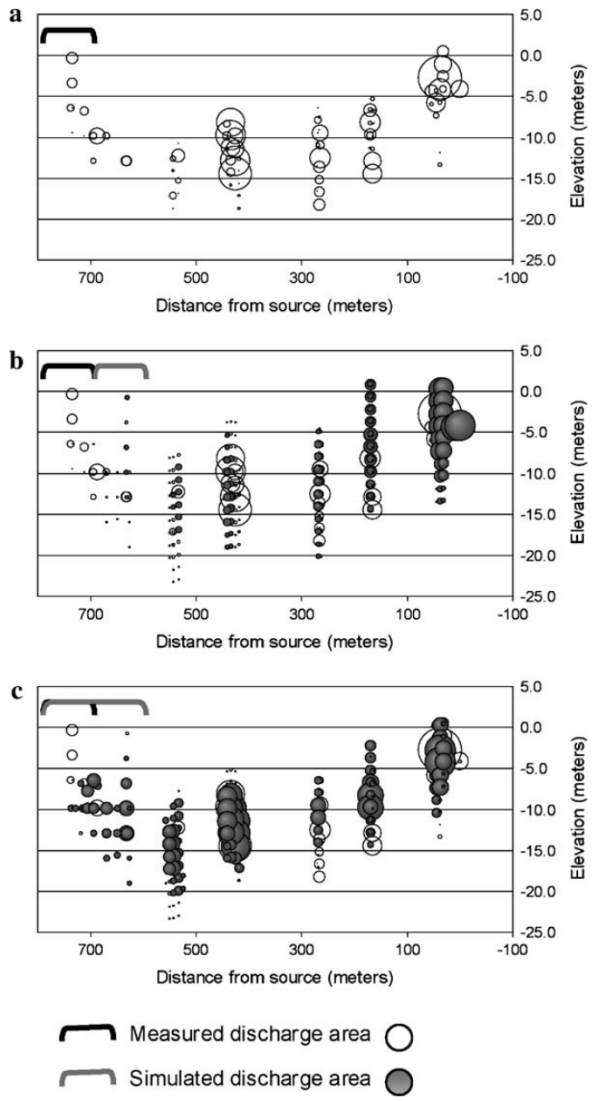

Figure 3. Profile of MTBE concentrations in wells from source area (right) to discharge area in bay (left), showing (a) measured MTBE, (b) measured MTBE and simulated MTBE from traditional zoned calibration, and (c) measured MTBE and simulated MTBE from hybrid TSVD regularized inversion (Tonkin and Doherty 2005).

their distribution—for reasons discussed by Menke (1989) among others. Regularized inversion, however, did lead to greatly improved fits (and hopefully predictions) and produced parameter distributions that are consistent with field and other independent data, and that could not reasonably be refuted without the collection of additional site-specific information.

Discussion and Conclusions

There is no universally correct way to parameterize and calibrate a ground water model or indeed any environmental model. Nevertheless, a case can be made that regularized inversion provides a more rigorous and less subjective mechanism for calibrating ground water models than do traditional zone–based approaches. We do not suggest that mapped geological, morphological, and other boundaries should not be respected where they are believed to exist; indeed, regularized inversion accommodates this. However, one must ask why traditional zone–based parsimony should be employed in contexts where parsimony can be flexibly and formally implemented using regularized inversion. This approach is employed as a matter of course in other industries such as geophysical data analysis and medical imaging. Imagine having to draw a zone where an anomaly *may* reside in your kidney or brain in order to interpret the wealth of information that is available in an image of that organ. Using the same reasoning, one might ask why it is still the standard approach in ground water modeling?

While many calibration strategies exist, we believe that regularized inversion is attractive since it is theoretically rigorous and sufficiently model-run efficient to be practical. It offers the potential to include parameter spatial variability in a model on a scale commensurate with model predictions, at the same time can encapsulate the modeler's understanding of a system through the use of preferred condition constraints. Regularized inversion helps maximize the insights gained from the field data and facilitates more encompassing evaluations of predictive uncertainty, while providing a mechanism for simulating conditions that may lead to extreme outcomes when testing the bounds of model predictive confidence—often the principal motivation for developing the model in the first place.

Why is regularized inversion not ubiquitous in ground water modeling? This may be due in part to the perception that it is difficult to do and computationally expensive to implement. However, open-source software that implements these approaches is freely available (for example, see Doherty 2003; Tonkin and Doherty 2005; Hunt et al. in review) and incorporated in several graphical user interfaces, and most current computational resources are more than sufficient to allow its use. Perhaps modelers have grown comfortable with the traditional approach; if so, we hope that this discussion will allow exploration of potentially better alternatives. While there is, indeed, no single way to calibrate a model, we believe that regularized inversion may help ensure that our models are not only "as simple as possible"—as with traditional parameter parsimony—but also fulfill the tenet "but not simpler."

Acknowledgments

We acknowledge the USGS Trout Lake Water, Energy, and Biogeochemical Budgets program for funding support, and we thank Allen Wylie, Daniel Feinstein, Mike Fienen, Charlie Andrews, and two anonymous reviewers for reviewing the article. We also thank Mary Anderson for soliciting and providing the soapbox for this article.

References

Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammerling, A. McKenney, and D. Sorenson. 1999. *LAPACK Users' Guide*, 3rd ed. Philadelphia, Pennsylvania: SIAM Publications.

Anderson, M.P. 1983. Ground-water modeling—The emperor has no clothes. *Ground Water* 21, no. 6: 666–669.

Anderson, M.P., and W.W. Woessner. 1992. *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*. New York: Academic Press Inc.

Aster, R.C., B. Borchers, and C.H. Thurber. 2005. *Parameter Estimation and Inverse Problems*. Burlington, Massachusetts: Elsevier Inc.

Clemo, T., P. Michaels, and R.M. Lehman. 2003. Transmissivity resolution obtained from the inversion of transient and pseudo-steady drawdown measurements. In *MODFLOW and More 2003: Understanding Through Modeling: Proceedings of the 5th International Conference of the International Ground Water Modeling Center*, ed. E.P. Poeter, C. Zheng, M.C. Hill, and J. Doherty, 629–633. Golden, Colorado: Colorado School of Mines.

Cooley, R.L. 2004. A theory for modeling ground-water flow in heterogeneous media. Professional Paper 1679. Reston, Virginia: USGS.

Cooley, R.L. 1982. Incorporation of prior information on parameters into nonlinear regression groundwater flow models: 1—Theory. *Water Resources Research* 18, no. 4: 965–976.

Cooley, R.L., and S. Christensen. 2006. Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media. *Advances in Water Resources* 29, no. 5: 639–656.

de Marsily, G., C. Lavedan, 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*, ed. G. Verly, M. David, A.G. Journel, and A. Marechal, 831–849. NATO ASI, Ser. C. 182 Norwell, Massachusetts: D. Reidel.

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

Draper, N.R., and H. Smith. 1998. *Applied Regression Analysis*, 3rd ed. New York: John Wiley and Sons Inc.

Engl, H.W., M. Hanke, and A. Neubauer. 1996. *Regularization of Inverse Problems*. Dordrecht, The Netherlands: Kluwer Academic.

Eppstein, M.J., and D.E. Dougherty. 1996. Simultaneous estimation of transmissivitiy values and zonation. *Water Resources Research* 32, no. 11: 3321–3336.

Freyberg, D.L. 1988. An exercise in ground-water model calibration and prediction. *Ground Water* 26, no. 3: 350–360.

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

Gómez-Hernández, J.J., H.J. Hendricks Franssen, and A. Sahuquillo. 2003. Stochastic conditional inverse modeling of subsurface mass transport: A brief review of the self-calibrating method. *Stochastic Environmental Research and Risk Assessment* 17, no. 5: 319–328.

Haitjema, H.M. 2006. The role of hand calculations in ground water flow modeling. *Ground Water* 44, no. 6: 786–791.

Haitjema, H.M., 1995. *Analytic Element Modeling of Groundwater Flow.* San Diego, California: Academic Press.

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

Hill, M.C. 1998. *Methods and Guidelines for Effective Model Calibration*. USGS Open File Report 98-4005. Reston, Virginia: USGS.

Hunt, R.J., and C. Zheng. 1999. Debating complexity in modeling. *Eos, Transactions of the American Geophysical Union* 80, no. 3: 29.

Jacobson, E.A. 1985. Estimation method using singular value decomposition with application to Avra Valley Aquifer in Southern Arizona. Ph.D. thesis, Department of Hydrology and Water Resources, The University of Arizona.

Lawson, C.L., and J.R. Hanson. 1995. *Solving Least Squares Problems, SIAM Classics in Mathematics*. Society for Industrial and Applied Mathematics: Philadelphia, Pennsylvania.

Liu, C., and W.P. Ball. 1999. Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware. *Water Resources Research* 35, no. 7: 1975–1985.

McLaughlin, D., and L.R. Townley. 1996. A reassessment of the groundwater inverse problem. *Water Resources Research*, 32, no. 5: 1131–1161.

Menke, W. 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, 2nd ed. New York: Academic Press Inc.

Moore, C., and J. Doherty. 2006. The cost of uniqueness in groundwater model calibration. *Advances in Water Resources* 29, no. 4: 605–623.

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

Ory, J., and R.G. Pratt. 1995. Are our parameter estimators biased? The significance of finite-difference regularization operators. *Inverse Problems* 11, no. 2: 397–424.

Poeter, E.P., and M.C. Hill. 1997. Inverse models: A necessary next step in ground-water modeling. *Ground Water* 35, no. 2: 250–260.

RamaRao, B.S., M. LaVenue, G. de Marsily, 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.

Skaggs, T.H., and Z.J. Kabala. 1994. Recovering the release history of a groundwater contaminant. *Water Resources Research* 30, no. 1: 71–79.

Sun, N.-Z., and W.W.-G. Yeh. 1985. Identification of parameter structure in groundwater inverse problem. *Water Resources Research* 21, no. 6: 869–883.

Tikhonov, A.N., and V.Y. Arsenin. 1977. *Solutions of Ill-Posed Problems*. New York: Wiley.

Tonkin, M.J., and J. Doherty. 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research* 41, W10412, doi:10.1029/2005WR003995.

Tonkin, M.J., J. Doherty, and C. Moore. Efficient non-linear predictive error variance for highly parameterized models. *Water Resources Research.* In review.

Tonkin, M.J., C.R. Tiedeman, D.M. Ely, and M.C. Hill. OPR-PPR, a computer program for assessing data importance to model predictions using linear statistics. Constructed using the JUPITER API. Chapter 2 of Book 6. *Modeling Techniques, Section E*, Reston, Virginia: USGS.

Townley, L.R., and J.L. Wilson. 1985. Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow. *Water Resources Research* 21, no. 12: 1851–1860.

Tsai, F.T.-C., N. Sun, and W.W.-G. Yeh. 2003. Global-local optimization for parameter structure identification in three-dimensional groundwater modeling. *Water Resources Research* 39, no. 2: 1043, doi:10.1029/2001WR001135.

U.S. EPA. 2001. Risk assessment guidance for superfund: Volume III (part A) Process for conducting probabilistic risk assessment. EPA/540/R/02/002. Cincinnati, Ohio: U.S. EPA.

van den Doel, K., and U.M. Ascher. 2006. On level set regularization for highly ill-posed distributed parameter estimation problems. *Journal of Computational Physics* 216, no. 2: 707–723.

Vecchia, A.V., and R.L. Cooley. 1987. Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. *Water Resources Research* 23, no. 7: 1237–1250.

Woodbury, A.D., and T.J. Ulrich. 2000. A full-bayesian approach to the groundwater inverse problem for steady state flow. *Water Resources Research* 36, no. 8: 2081–2093.

Zheng, C., and P.P. Wang. 1996. Parameter structure identification using tabu search and simulated annealing. *Advances in Water Resources*, 19, no. 4: 215–224.

Appendix

Mathematics of Traditional Calibration and Regularized Inversion

Model Calibration

Let the vector $\mathbf{p}$ represent the true values of a system's properties, and let $\mathbf{h}$ denote the n observations on which calibration is based. Let ϵ represent measurement noise. If the matrix $\mathbf{X}$ denotes the linear sensitivities of observations of model outputs corresponding to system state $\mathbf{h}$ to changes in system properties $\mathbf{p}$, the relationship between $\mathbf{h}$ and $\mathbf{p}$ can be represented as follows:

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \epsilon \tag{A-1}$$

Let $\hat{\mathbf{p}}$ represent values inferred for m model parameters through model calibration. If traditional model calibration is employed to limit the number of parameters in $\hat{\mathbf{p}}$ so that a well-posed inverse problem is formed, $\hat{\mathbf{p}}$ can be estimated using the Gauss-Marquardt-Levenberg method as follows:

$$\hat{\mathbf{p}} = (\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{h} = \mathbf{G}\mathbf{h} \tag{A-2}$$

where $\mathbf{G}$ denotes the (linear) relationship between $\mathbf{h}$ and $\hat{\mathbf{p}}$, and $\mathbf{Q}$ is the observation weight matrix. If $\mathbf{Q}$ is proportional to the inverse of the measurement error covariance matrix $C(\epsilon)$, the covariance matrix of the estimated parameters is as follows:

$$C(\hat{\mathbf{p}}) = \sigma_r^2(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1} \tag{A-3}$$

where σ_r^2, the reference variance that quantifies the level of measurement uncertainty, is given by:

$$\sigma_r^2 = \Phi/(n - m) \tag{A-4}$$

where Φ is the minimized sum of squared weighted residuals. Equations A-1 through A-4 apply where the model is linear. Nonlinear parameter estimation is achieved through extension of these concepts to an iterative solution process whereby a parameter change vector $\Delta\hat{\mathbf{p}}$ is determined on the basis of the current vector of residuals (i.e., model-to-measurement misfits) $\mathbf{r}$ using:

$$\Delta\hat{\mathbf{p}} = (\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{r} = \mathbf{G}\mathbf{r} \tag{A-5}$$

In traditional parameter estimation as represented by Equations A-2 and A-5, matrix $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ must be invertible to obtain a unique solution—i.e., to estimate the parameter values comprising $\hat{\mathbf{p}}$. This requires the number of observations be greater than or equal to the number of parameters—often much greater.

Two particular regularized inversion techniques are discussed in this article, together with a third that combines these into a hybrid regularization strategy. The first of these, Tikhonov regularization, supplements calibration data with information pertaining to parameters using regularization equations, the weights for which are determined during calibration. This "penalized least squares" approach sums measurement and regularization objective functions to form a global objective function. The expression for $\mathbf{G}$ when Tikhonov regularization is employed is as follows:

$$\mathbf{G} = (\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{T}^t\mathbf{S}\mathbf{T})^{-1}\mathbf{X}^t\mathbf{Q} \tag{A-6}$$

where $\mathbf{T}$ is the vector of Tikhonov regularization constraints on parameters, $\mathbf{S}$ is the regularization weight matrix, and β^2 is the regularization weight factor (determined though the calibration process as that which achieves a user-specified level of model-to-measurement fit).

The second regularization strategy is based upon TSVD (Anderson et al. 1999). In the special case of the square symmetric matrix $\mathbf{X}^t\mathbf{Q}\mathbf{X}$, TSVD decomposes $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ into:

$$\mathbf{X}^t\mathbf{Q}\mathbf{X} = \mathbf{V}\mathbf{E}\mathbf{V}^t \tag{A-7}$$

where $\mathbf{E}$ is diagonal and lists the m singular values of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$, while the m column vectors of $\mathbf{V}$ are the eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ (Lawson and Hanson 1995). $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ possesses m real-valued eigenvalues. In practice, a small number of these dominate, corresponding to calibration solution space eigenvectors (Aster et al. 2005). Stable inversion is achieved using regularization as a filter that estimates parameter combinations that reside in the calibration solution space, while ignoring parameter combinations that reside in the calibration null space. The expression for $\mathbf{G}$ when TSVD regularization is employed is as follows:

$$\mathbf{G} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q} \tag{A-8}$$

where $\mathbf{V}_1$ and $\mathbf{E}_1$ contain the eigenvectors and eigenvalues that are retained following the application of the regularization filter.

The hybrid regularization methodology of Tonkin and Doherty (2005) combines Tikhonov and TSVD regularization strategies to estimate a limited number of "super parameters," while enforcing Tikhonov constraints on base parameter values as actually employed by the model. Super parameters are scalar multipliers for the solution space eigenvectors (column vectors of $\mathbf{V}_1$); these are estimated using classical least squares in a reformulated inverse problem. Jacobson (1985) describes a similar approach to redefining the inverse problem on the basis of an eigenvalue-eigenvector decomposition referred to as "surrogate parameters." The expression for $\mathbf{G}$ for the hybrid Tikhonov-TSVD regularization process is as follows:

$$\mathbf{G} = \mathbf{V}_1(\mathbf{Z}^t\mathbf{Q}\mathbf{Z} + \beta^2\mathbf{T}^t\mathbf{S}\mathbf{T})^{-1}\mathbf{Z}^t\mathbf{Q} \tag{A-9}$$

where $\mathbf{Z}$ is the sensitivity matrix of model outputs with respect to "super parameters", and $\mathbf{T}$ is the matrix of regularization constraints on base parameters after projection of these constraints onto the calibration solution subspace.