Journal of Hydrology 380 (2010) 489–496

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Discussion

Response to Comment on "Two statistics for evaluating parameter identifiability and error reduction"

John Doherty [a,b], Randall J. Hunt [c,*]

[a] Watermark Numerical Computing, Brisbane, Australia
[b] National Centre for Groundwater Research and Training, Flinders University, Australia
[c] US Geological Survey, 8505 Research Way, Middleton, WI 53562, United States

Introduction

We appreciate the interest in the identifiability and relative parameter error variance reduction statistics that we describe in Doherty and Hunt (2009). Dr. Hill's comments, referred to here as Hill (2010), include concerns specific to our paper, but also examples and issues that extend beyond our paper to the work of others. As such, they extend to broader methodological issues which underpin our general approach to model calibration and uncertainty analysis. In this regard they are similar to those raised in a recent paper (Hill, 2008a) and in a number of scientific presentations (Hill, 2006, 2007, 2008b, 2009). We consider a discussion of these issues vital to achieving changes in approaches to parameter estimation and uncertainty analysis that we believe will substantially enhance the ability of the modeling community to address societally relevant problems. We hope that discussion of the topics raised in Hill (2010) will increase the understanding of the approach that we are trying to advance, and are therefore pleased to engage in such a discussion.

The material covered in Hill (2010), and therefore our response, encompasses the larger fundamental difference in approach between older, traditional methods developed for problems with few parameters, and new methodologies for parameter estimation and uncertainty analysis based on solution of highly parameterized, ill-posed inverse problems that seek to redress the deficiencies of these older methods. Hunt et al. (2007) summarize the underlying issues; they define a "traditional" approach to model calibration as one which relies on pre-calibration user-intervention to formulate an over-determined (more useful observations than parameters solved for), well-posed (mathematically solvable) inverse problem through pre-emptive parameter simplification. In distributed hydrologic models, traditional parameter simplification is most commonly accomplished using devices such as zones of piecewise constancy, and may be accompanied by the introduction of prior information pertaining to one or a number of elements comprising the simplified parameter set. A nagging concern with these and similar approaches, however, is that the configuration of such zones, and the concomitant use of prior information assigned to such zones, may induce hard-to-quantify structural artifacts that hobble the ability of the calibration process to extract information from the calibration dataset. This is because such over-simplified and subjective parameterization devices rarely form the ideal receptacles for that information. Furthermore, simplifications inherent to parameterization of a model (and indeed to modeling itself) will always impart some "structural noise" on simulation results due to the model's inability to perfectly represent the natural world. The amount of this structural noise will depend on model construction and parameterization details. Overly simple parameterization may induce considerable structural noise, and may thus introduce undesirable artifacts to predictions of interest made by a model (Moore and Doherty, 2005).

Sophisticated and practical alternatives to this traditional approach to model calibration based on regularized inversion are widely applied in other fields. Hunt et al. (2007) describe the application of these methods to calibration of groundwater models. This consists of: (1) introducing parameter flexibility at the beginning of the model calibration process by including larger numbers of parameters than could be accommodated by the traditional approach; and (2) constraining the increased parameter set with mathematical regularization encapsulating soft knowledge of the system. A regularized inversion approach to model calibration seeks to maximize the ability of observation data to "speak" through the calibration process without being hindered by pre-calibration simplification undertaken by the modeler. Model parameterization can include such devices as pilot points (distributed liberally throughout a model domain) in place of (or together with) any number of zones of piecewise constancy.

Regardless of the parameterization device employed, mathematical regularization attempts to constrain the calibration process to estimate the *simplest* parameter field that is compatible with the data (e.g., Hunt et al., 2007; Fienen et al., 2009a). Such an approach attains parameter uniqueness while yielding a data-informed, estimated parameter field which approaches that of minimum error variance. Meanwhile the use of many parameters grants the regularized calibration process freedom to introduce complexity if and where it is needed, thus helping to reduce artifacts resulting from over-simplification of parameter structure. Thus, *both* traditional and regularized inversion approaches to model calibration have a common pursuit of parsimony; the difference between them lies in the manner in which parsimony is de-

DOI of original article: 10.1016/j.jhydrol.2009.10.011
* Corresponding author. Tel.: +1 608 828 9901; fax: +1 608 821 3817.
E-mail addresses: johndoherty@ozemail.com.au (J. Doherty), rjhunt@usgs.gov (R.J. Hunt).

0022-1694/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jhydrol.2009.10.012

fined and attained. Traditional methods rely heavily on subjective manual simplification. Regularized inversion, on the other hand, defines optimal simplicity as that which minimizes parameter and predictive error variance; it then achieves optimal simplification, defined in this way, as part of the calibration process itself, within the context of the dataset that it attempts to fit. Moreover, as pointed out by Hunt et al. (2007), the true intent of parsimony is better realized in a regularized inversion framework because the final result honors the precept "as simple as possible" while also honoring the complementary precept "*but not simpler*".

The use of regularized inversion is not restricted to calibration of groundwater models. Doherty and Johnston (2003), Gallagher and Doherty (2007), Tonkin and Doherty (2009) and Doherty and Hunt (2009) demonstrate its applicability to calibration of surface-water models, and in doing so address calibration problems which are commonly considered to be under-determined and thus difficult for traditional methods. A full description of the regularized inversion approach to calibration of environmental models is beyond the scope of this response, but can be found in Moore and Doherty (2005, 2006), Tonkin and Doherty (2005), Gallagher and Doherty (2007), Tonkin et al. (2007), and Hunt et al. (2007). The mathematical basis of the statistics presented by Doherty and Hunt (2009) is discussed most fully in Moore and Doherty (2005). This overarching approach to calibration and uncertainty analysis is rooted in a large body of well-established theory described extensively in works such as Menke (1984) and Aster et al. (2005). Notwithstanding the wide usage of these methods in other scientific fields, Hill (2010) does not constitute the first occasion on which the concepts underpinning regularized inversion, and their applicability to environmental model calibration, have been questioned by Dr. Hill. Issues pertaining to pilot-point parameterization and concomitant mathematical regularization formed the heart of a previous critique of the regularized inversion approach (see Hill, 2008a); concerns raised in that commentary were addressed by Fienen et al. (2009b).

In summary, both older traditional calibration methods and regularized inversion attempt to address the same larger question of how to best represent with integrity as much of the complexity of the natural world in a numerical model as can be inferred from scientific expertise, site characterization studies, and the calibration dataset. Because the information content of all of these is limited, the parameter field employed by a calibrated model must necessarily constitute a simplified version of reality. Because natural systems are so complex, all model calibration starts out as an attempt to solve a highly parameterized, non-unique, inverse problem. Thus, the question is not *whether* some type of regularization strategy is needed; rather, it is *which* regularization strategy is best suited for the site and modeling objective at hand. Traditional zonation and other *a priori*, user-specified parameter simplification schemes can be thought of as simple regularization devices that are designed to make this ill-posed inverse problem sufficiently well-posed as to be solvable using traditional calibration methodologies – methodologies which are numerically incapable of solving ill-posed problems through mathematical means. On the other hand, approaches to calibration based on regularized inversion rigorously pursue the minimum error variance solution to the necessarily ill-posed problem of model calibration. In doing so, they are mathematically more robust and have greater general applicability to the continuum of hydrologic problems than traditional, pre-emptive simplification calibration methodologies. We therefore maintain that regularized inversion is superior to older traditional approaches that rely on extensive, subjective, manual intervention in order to attempt to achieve the same end as newer calibration methods, but offer no mathematical guarantee that such an end has in fact been attained.

We now respond to the five issues raised in Hill (2010) in the order in which they were presented.

Joint use of composite scaled sensitivities (CSS) and parameter correlation coefficients (PCC)

Hill (2010) agrees with Doherty and Hunt (2009) that composite scaled sensitivities do not reflect parameter correlation. The remainder of this section of Hill (2010) suggests that additional PCC calculations, in concert with CSS analyses, can give considerable information about parameter identifiability. As described on page 120 of Doherty and Hunt (2009), the statistics discussed in our original paper are not the only ones that might be used for elucidation of parameter identifiability. Indeed, we direct the reader to two other potentially useful statistics, thereby underscoring our view that a reader should not be compelled to limit his/her analysis to one method or another. Thus, we believe that the suggestion by Hill (2010) that concomitant PCC analysis be performed where possible is within the spirit of our original presentation.

We noted, however, in our original paper that the utility of our identifiability and relative error variance reduction statistics lies in the fact that they are easily calculated, and can handle large numbers of parameters and observations typical of problems arising out of both groundwater and surface-water model calibration, irrespective of whether the calibration problem is well-posed or ill-posed. In contrast, PCC cannot be robustly calculated when an inverse problem is ill-posed (as can occur because of the excessive parameter correlation that PCC attempts to characterize) because the "normal matrix" through which the parameter covariance matrix must be computed cannot be inverted. Although Hill and Østerby (2003) present a workaround that allows computation of PCC where an inverse problem is almost, but not quite, ill-posed, this approach is not applicable to more general highly parameterized calibration contexts characterized by null spaces of high dimensionality. Fortunately, as shown in Doherty and Hunt (2009) and elsewhere, identification of parameter combinations that can be estimated in concert, but not individually, is readily and robustly forthcoming through singular value decomposition of the normal matrix prior to its inversion, regardless of the degree of ill-posedness of the inverse problem.

Parameter identifiability statistic and composite scaled sensitivity

Hill (2010) suggests that the parameter identifiability statistic of Doherty and Hunt (2009) will "often do a poor job of parameter identifiability". While our original work acknowledged some limitations on the use of these statistics arising from assumptions of differentiability and linearity (page 120), we do not believe that the issues raised by Hill (2010) invalidate our original assertion that the statistics have utility in providing "qualitative rather than quantitative insights into relative parameter estimability".

Hill (2009) asserts that the CSS statistic can (with the four intervening steps outlined at the end of Section 2.1) yield the same insights as our parameter identifiability statistic. This assertion does not recognize the vital role played by singular values in linking orthogonal combinations of observations to orthogonal combinations of parameters that they uniquely and exclusively inform; see, for example, Lines and Treitel (1984) for a discussion of this issue. In our original paper, **S** represents the matrix of singular values defined through singular value decomposition using the equation:

$$\mathbf{Q}^{1/2}\mathbf{X} = \mathbf{USV}^{t} \tag{1}$$

Our **Q** matrix is the same as **ω** employed by Hill (2010). From (1) and the fact that:

$$\mathbf{U}^t\mathbf{U} = \mathbf{I} \tag{2}$$

it follows that:

$$\mathbf{X}^t\mathbf{Q}\mathbf{X} = \mathbf{V}\mathbf{S}^2\mathbf{V}^t \tag{3}$$

where "t" indicates matrix transpose. This is similar to Eq. (1b) of Hill (2010). The diagonal elements of $\mathbf{V}\mathbf{S}^2\mathbf{V}^t$ are related to parameter composite scaled sensitivities as Hill (2010) states. Meanwhile the diagonal elements of $\mathbf{V}_1\mathbf{V}_1^t$ are our identifiabilities, where the subscript "1" defines the calibration solution space. Hill (2010) asserts that inclusion or exclusion of the $\mathbf{S}^2$ matrix (together with its zero and near-zero elements) makes little difference to computation of an identifiability statistic, and then suggests that our identifiability statistic is therefore similar to CSS. However, indifference regarding the presence or absence of the $\mathbf{S}^2$ matrix is highly problematic, for it is a vital component of the characterization of the inverse problem that singular value decomposition provides. Singular values constitute the "response function" of the inverse problem as currently posed, not unlike the frequency spectrum of an electrical or optical filter. In particular, as Lines and Treitel (1984) point out, each element of the diagonal **S** matrix records precisely the relationship between a particular combination of observations and the particular combination of parameters that it uniquely and completely informs. Combinations of observations comprise columns of the **U** matrix of Eq. (1), while informed combinations of parameters comprise corresponding columns of the $\mathbf{V}_1$ submatrix of the **V** matrix of this same equation. Columns of $\mathbf{V}_1$ comprise axes of the solution subspace of parameter space referred to above.

Implicit inclusion of the $\mathbf{S}^2$ matrix in the definition of CSS yields a statistic that, while useful for some purposes, contains "too much information" for other purposes. In particular, it loses its ability to serve as a means of readily identifying estimability of individual parameters. To see why this is so, consider the $\mathbf{V}_1$ submatrix of the **V** matrix of Eq. (1). This is comprised of columns of **V** for which corresponding diagonal elements of **S** are significantly non-zero; hence, it contains parameter combinations that can be uniquely estimated through model calibration. Through consideration of this matrix in isolation from the **S** matrix, nothing can be said of what those estimated values are – only that parameter combinations comprising the columns of $\mathbf{V}_1$ are estimable. Meanwhile the $\mathbf{V}_1\mathbf{V}_1^t$ matrix presents this estimability information in an easily digestible form as a set of direction cosines, as is explained in Doherty and Hunt (2009). On the other hand, through inclusion of $\mathbf{S}^2$, the $\mathbf{V}\mathbf{S}^2\mathbf{V}^t$ matrix of Eq. (3) supplements this identifiability information with information on how the values of estimable parameter combinations can be computed. However, in doing this it does not actually provide estimates for those parameters, for it does not include the data on which parameter estimation is necessarily based. Instead, it provides a statistic which is by no means uninformative, but on the other hand is not clearly informative either, as it is not directed at a particular property of the parameters employed by the model.

In considering this issue, it is important to understand the role of the "calibration null space" in the overall process of model calibration. The calibration null space can be conceptualized as being comprised of all differences that exist between the necessarily simplified parameter field represented in a model and the limitless complexity of the actual natural world. Thus, in its broadest sense, the null space consists not only of parameters specified but not estimated, but also of the world of possible parameters not specified and thus not estimated. Orthogonal to this is the "calibration solution space," which is comprised of combinations of parameters that are uniquely inferable on the basis of a given calibration dataset.

In a stricter sense, the calibration null space is comprised of combinations of parameters whose effect on model outputs employed in the calibration process is zero or negligible; these combinations of parameters cannot therefore be inferred through calibration. In seeking the simplest solution to the inverse problem of model calibration, these parameter combinations are effectively unchanged from their initial values; solution simplicity follows from the fact that these combinations often pertain to fine-scale hydraulic property detail. In contrast, solution-space parameter combinations mostly refer to broad scale features such as aerially averaged hydraulic properties.

Note use of the term "combinations of parameters" in the above description. This is unavoidable, for the decomposition of parameter space into solution and null subspaces effectively defines a new set of axes for parameter space which, in the context of a given calibration dataset, constitutes a "more natural" set of axes than those comprised of individual parameters. Some of these axes (those comprised of estimable parameter combinations) span the calibration solution space; the remainder span the calibration null space. Hill (2010) present a partial portrait of the null space – one characterized only by individual parameters removed from the model prior to calibration and subsequently re-introduced for the purpose of calculating prediction uncertainty. This would only comprise an accurate depiction of the null space if each model parameter had an identifiability of exactly one or exactly zero. In practice, parameters commonly have identifiabilities other than zero or one because the observation dataset allows inference of some combinations of parameters (generally broad-scale system behaviour and hence parameter sums), while not allowing inference of other combinations (generally system detail and hence specific parameter differences). Thus, an individual parameter may be wholly estimable, wholly inestimable, or partly estimable and partly not. In the latter case "partly" is defined by the relative sizes of its projections onto the calibration solution and null subspaces. (Recall that the projection of a vector onto an arbitrary subspace is rather like the shadow that it casts onto that subspace when illuminated by a light that is directly orthogonal to that subspace.)

The second issue in this section of Hill (2010) involves a simple groundwater model test case. We agree that differences between CSS and our identifiability statistic will be smaller when applied to simplified, over-determined problems such as the synthetic groundwater example discussed in Hill (2010). Thus, it is not surprising that when an appropriate singular value cutoff is used – as illustrated in the upper panels of Fig. 1 of Hill (2010) – insights gained from use of our identifiability statistic are similar to those gained from use of CSS (after the inverse problem is enhanced with user-suppled prior information) shown in Fig. 2 of Hill (2010). However, these types of simple problems were not the subject of our original paper; nor do they represent the context where most real-world model calibration takes place. Indeed, when the inverse problem of model calibration becomes poorly posed, use of CSS alone may lead one astray. For example, the CSS statistic computed for the surface-water model presented in our original paper suggests that all 18 parameters are sensitive (page 125), and thus may be considered estimable – a very different result from that provided by our identifiability statistic which, consistent with experience calibrating surface-water models, quite properly characterizes the inverse problem as ill-posed. This underscores the difficulties that inevitably attend extrapolation of older methods developed for over-determined problems (where, by definition, all parameters are identifiable) to the highly parameterized, under-determined problems that characterize real-world environ-

mental model calibration, which were the subject of our original paper.

The third issue in this section of Hill (2010) is not so much an issue as a suggestion to replace the term "identifiability" with the term "SVD parameter composition statistic". We selected "identifiability" because it reflects both the meaning and intended use of this statistic. There are limits to the information that a single statistic can convey, especially one that relies on local model linearization. However, there is something very intuitive about a projection operation. If the projection of a parameter onto the calibration solution space is zero (thereby implying that its CSS is zero), the parameter cannot be estimated because the calibration dataset contains no information to constrain it. Thus it is unidentifiable, and its identifiability (as computed using the methodology provided in our paper) equals zero. On the other hand, if the projection of a parameter onto the calibration solution space is equal to itself because it lies entirely within that space, it has an identifiability equal to one, and can therefore be uniquely estimated. As we point out in our original paper, this does not mean that it can be estimated without error. It means that the entirety of the error associated with its estimation is inherited from noise associated with the calibration dataset on which basis its estimation takes place; no error arises from inseparable correlation with one or more other parameters as evinced by a non-zero projection onto the calibration null space, and hence a non-unity identifiability. Such a clear and straightforward visualization of the proportion of a parameter that lies in the calibration solution space, and the proportion that does not, is, in our opinion, better captured by the term "identifiability" than by the term "SVD parameter composition statistic".

In light of the above, we do not believe the issues laid out in Hill (2010) support the statement that statistics of Doherty and Hunt (2009) will often do a poor job of characterizing parameter identifiability. We do agree with Hill (2010) that there are some similarities between our identifiability statistic and CSS; in fact, if a parameter is totally insensitive, then both its CSS and its identifiability will be zero (this negating the assertion by Hill (2010) that "when parameter insensitivity plays a significant role...conclusions drawn from the parameter identifiability statistic are likely to be incorrect"). However, the differences between our identifiability statistic and CSS are important. To summarize and further clarify these differences, let us consider two parameters which are highly sensitive, but that cannot be individually estimated on the basis of the current calibration dataset. Let us consider that only their sum can be estimated, but that their difference lies in the null space. (This is a common occurrence in environmental modeling where broad-scale system properties are often estimable, but detail is not.) In this case, each parameter will have an identifiability of about 1/sqrt(2). The fact that their identifiabilities are non-zero conveys to the modeler that the calibration dataset has something to say about each parameter (it can estimate their sum). The fact that their identifiabilities are less than unity indicates that the calibration dataset does not contain enough information to allow estimation of unique values for each (albeit with error inherited from the calibration dataset). In contrast, each of these parameters will have a high CSS (there is no inherent upper or lower limit to CSS). The fact that these parameters are inseparable as far as their estimation is concerned is not conveyed through CSS.

Truncation point for estimated SVD parameters

We agree with Hill (2010) that the number of pre-truncation singular values to employ in solution of an ill-posed inverse problem is an important issue. Indeed, inappropriate truncation, for example through inclusion of too many singular values in solution of the inverse problem of model calibration such as is demonstrated in the lower panels of Fig. 1 of Hill (2010), can lead to unrepresentative results – a point discussed at length on page 123–124 of Doherty and Hunt (2009). However, we differ with Hill (2010) in regard to the rest of this comment. How the trade-off in predictive error variance depends on shrinking the calibration null space by attempting to estimate more parameters on the one hand, but on increasing the potential for estimation error based on amplification of measurement noise through expansion of the solution space on the other hand, is extensively discussed and demonstrated in Moore and Doherty (2005), Hunt and Doherty (2006), Gallagher and Doherty (2007), and in our original paper. From the description provided, we cannot discern why the Hill (2010) implementation of the Moore and Doherty (2005) methodology does not agree with other previously published examples. One possibility is that Hill (2010) may have confused the outcomes of one realization presented in Moore and Doherty (2005) with the overall envelope of many realizations that is the subject of the theory.

The relative error variance reduction statistic and perturbed calculated variance statistics

Hill (2010) proposes a categorization scheme in which the statistics of Doherty and Hunt (2009) are grouped with others developed by Dr. Hill and her co-workers. While we agree with Hill (2010) that our relative error reduction statistic will likely be very useful in practice, we consider the regularized inversion approach that underpins our statistic to have broader applicability to today's environmental problems than the other statistics listed in Hill (2010).

The idea that data worth can be assessed in terms of its ability to reduce parameter and/or predictive uncertainty is not new; see, for example, Beven (1993) and references cited therein. We did not cite these earlier articles in our original manuscript though perhaps we should have, given the discussion in this section of the comment. We did, however, recognize that the overall intent of our statistics was not unique to our problem, and acknowledged the work of others. We included work based on traditional approaches in our citation of Tonkin et al. (2007), which we considered a "tap root" for many of the references listed in Hill (2010). We believed this constituted an appropriate acknowledgement of previous work, while setting an appropriate context for our own work given the different methodological and philosophical foundations that exist between previous approaches to data-worth assessment and those based on highly parameterized inversion. The theory on which previous approaches were based, as provided in the literature and in user documentation of software which calculates many of the statistics listed in Hill (2010), is rooted in concepts that are appropriate for over-determined inverse problems. Application of these concepts, in our view, introduces confounding artifacts to a model's parameterization scheme that limits its utility for simulating complex, real-world hydrogeological and hydrologic systems that were the subject of our original paper. In contrast, the statistics of Doherty and Hunt (2009), and other related statistics available through the software referred to in that paper, are based on regularized inversion theory which is specifically formulated and tested to address ill-posed problems such as those resulting from highly parameterized models, and are designed to include approaches to calibration and uncertainty analysis that readily accommodate real-world complexity. Thus, they are designed "from the ground up" to inform data collection that effectively characterizes that complexity.

OPR and PPR Statistics

Hill (2010) suggests that a traditional approach to post-calibration uncertainty analysis can account for the contribution made to predictive uncertainty by the parameter simplification required for model calibration. This is accomplished by activating parameters that were frozen during calibration due to lack of individual estimability. Hill (2010) equates the limited number of new parameters introduced for the purpose of uncertainty assessment to the null space described by Moore and Doherty (2005). As described in our response to Section 2 of Hill (2010) above, we believe that such a characterization of the null space is not complete. In fact, as stated above, the null space consists not only of parameters specified but not estimated, but also of the world of possible parameters not specified and thus not estimated. Furthermore, for reasons discussed above, the null space is more accurately described in terms of *combinations* of parameters than in terms of individual parameters.

There will obviously be limits to the extent to which any model can represent the true null space as it attempts to characterize the contribution that this space makes to the uncertainty associated with model predictions. However, highly parameterized methods where the null space is computed using devices such as singular value decomposition, which are designed for this purpose, are intrinsically better suited to characterization of the null space than traditional methodologies which were designed for use in user-simplified parameter contexts specifically for the purpose of eradicating the null space. In particular, as documented in Tonkin et al. (2007), the OPR and PPR statistics which are discussed in detail by Hill (2010) were developed for use in conjunction with sparsely parameterized models where parameters can be estimated using traditional methodologies based on well-posed inverse problems. In the environmental modeling context, a well-posed inverse problem can be achieved if a modeler preemptively removes from the calibration process any parameters that cannot be uniquely estimated through that process. Paradoxically, such simplification prior to calibration compromises the use of statistics such as OPR/PPR. This is because the extra information that resides in data that is worth collecting, but that has not yet been collected, should illuminate complexity which is presently beyond the purview of the calibration process. This, however, is the very complexity that must be eliminated if traditional methods are employed to calculate parameter and predictive uncertainty – both as it presently exists and as it would exist with acquisition of further data.

In recognition of this conflict, Hill (2010) describes a previously unpublished modification of the OPR/PPR methodology that attempts to overcome this problem. Although the description of the revised methodology provided by Hill (2010) includes a number of equations, the material presented does not constitute proof of the validity of the proposed approach, but only a description of how it could be implemented. In doing so, mathematical justification is replaced with an assertion that OPR/PPR enhancements incorporate use of a $\mathbf{C}(\mathbf{p})$ covariance matrix of innate parameter variability, as is done in theory presented by Moore and Doherty (2005). It is contended that the two approaches are therefore equivalent, or at least interchangeable. We do not agree.

Because the calibrated model constitutes, of necessity, a simplified version of reality, predictions made by such a model will be in error. In general, the greater the degree to which a prediction depends on parameterization detail that cannot be supported by the calibrated parameter field, the greater its potential for error. Ideally, this potential falls within the range of predictive confidence that is forthcoming from post-calibration uncertainty analysis. A fundamental difference between traditional methods of model parameterization and newer methods based on regularized inversion is that the latter methods provide the user with a quantitative description of how simplification has had to take place in order to accommodate the information inadequacies of the calibration dataset. This information is normally provided through the resolution matrix. The substantial contribution that parameter simplification makes to predictive uncertainty is therefore quantifiable through post-calibration uncertainty analysis. In contrast, where simplification is subjective and ad-hoc (as it must be when undertaken manually in traditional approaches), no such information is available. The cost of not having this information on over-simplification is a loss of ability to quantify the potential for error in model parameters and predictions incurred by parameter simplification.

The theory provided by Moore and Doherty (2005), which Hill (2010) does not appear to dispute, can be readily deployed to compute parameter/predictive uncertainty pursuant to use of manual regularization implemented, for example, through use of zones of piecewise constancy. Using equations provided by Moore and Doherty (2005) as a foundation, mathematical underpinnings can be provided for the approach recommended by Hill (2010) for expanding the use of OPR/PPR to the highly parameterized context, including appropriate use of the $\mathbf{C}(\mathbf{p})$ matrix of innate parameter variability. We have included this derivation in the Appendix to this response. It is evident that it leads to a different set of equations from those listed in Hill (2010).

In the derived expansion of the OPR methodology to the highly parameterized context presented in the Appendix, it is of interest to note that one possible formulation of the expanded OPR methodology presented by Hill (2010) is mathematically equivalent to a methodology documented in Christensen and Doherty (2008) and implemented in the PEST PREDUNC ("Prediction Uncertainty") utility suite referenced in Hill (2010). The PREDUNC methodology is based on Bayes theorem, and is theoretically rigorous if innate parameter variability and measurement noise both satisfy multinormal distributions. In implementing this methodology, however, the PREDUNC formulation is numerically more stable and efficient where parameter numbers are high and observation numbers are low (as is often the case where issues of data-acquisition optimization must be addressed).

It should be noted that if simplified parameterization devices such as zones of piecewise constancy that are commonly employed in traditional model calibration exercises are then employed in data-worth analysis, their use can confound the insights that might otherwise be gained through such analysis. For example, how can the subtle information contained in a proposed head measurement here, or pumping test there, be heard above the clamor of misinformation encapsulated in the requirement that large parts of a model domain possess spatially invariant properties, that the boundaries between these parts are at exactly known locations, and that these boundaries are marked by abrupt hydraulic property changes? Our experience, as well as that of Fienen et al. (submitted for publication), suggests that the outcomes of data-worth analysis in such contexts are more reflective of parameter simplification devices than of the true information content of data comprising an existing or posited calibration dataset. A regularized inversion approach to data-worth analysis, on the other hand, ensures that the spurious effects of over-simplified parameterization schemes are reduced to a minimum.

We welcome all work in this area, including efforts such as the extension of the OPR methodology. However, we remain convinced that a highly parameterized context for data-worth analysis should be considered mandatory and not optional, and that extension of the OPR methodology to that context be in accordance with Eq. (A15) of the Appendix, rather than equations listed in Hill (2010). Finally, it is worth pointing out that if the full utility of data-worth

analysis is to be realized, it is incumbent on the modeler to ensure that the promised reductions in predictive uncertainty forthcoming from such analysis are actually realized. This is more likely to occur if model re-calibration is undertaken with the tools and flexibility inherent to the regularized inversion/highly parameterized context, the same context wherein the worth of data is best assessed. For it is in this context that the ability of this data to raise the identifiability level of prediction-sensitive parameters will be realized, as prediction-informative parameter combinations that were previously in the null space cross the identifiability threshold into the solution space. It is only through a highly parameterized approach to calibration and uncertainty analysis that these spaces can be identified, and that the calibration null space can be reduced to the maximum extent possible, given the updated information content of a revised calibration dataset.

Conclusions

The comment of Hill (2010) and our associated response have explored issues which are fundamental to environmental model calibration and ancillary uncertainty and data-worth analysis. Hill (2010) asserts that methods based on calibration of sparsely parameterized models are readily extensible to highly parameterized contexts, and further suggests that such extensions provide the same or better utility and insights as do mathematically rigorous methods based on regularized inversion. As discussed in this response, we remain unconvinced. We do not believe that such extensions to traditional approaches can be generally and robustly applied to the highly parameterized and ill-posed problems that were the subject of our original paper, and that comprise the context in which most environmental modeling must take place. Instead, we suggest that such modeling is better served through use of mathematically rigorous approaches that are designed specifically to accommodate the complexity of natural systems, approaches whose ubiquitous usage in other fields has demonstrated their general applicability and utility in complex settings.

Acknowledgements

Support for this response was provided by the USGS Office of Ground Water. The authors thank Matthew J. Tonkin (S.S. Papadopulos & Assoc., Inc.) and Michael N. Fienen (US Geological Survey) for their technical review of early drafts of the response and appendix, and Demetris Koutsoyiannis for review comments on a later draft.

Appendix A

In this Appendix we demonstrate a methodology for expanding the use of OPR/PPR to the highly parameterized context. We do not consider this to be an optimal method for exploring observation worth, as it requires explicit or implicit calibration under user-specified simplified conditions before adding post-calibration complexity for the purpose of data-worth assessment. This need for pre-calibration simplification erodes the ability of the explicit or implicit calibration process implemented in OPR/PPR to extract maximum information from existing data. It therefore can be expected to provide a poor basis for assessing the enhanced information content of new data. However, if this path is followed (which we do not recommend), we believe it should be based on the theory explained below rather than on the methodology proposed in Hill (2010).

Let $\mathbf{h}$ represent a set of observations of system state (for example, groundwater heads). Let the parameters pertaining to that system be denoted by the vector $\mathbf{p}$. Let the noise associated with $\mathbf{h}$ be encapsulated in the random vector $\boldsymbol{\varepsilon}$. Assuming linear conditions (and ignoring parameter and predictive offsets for notational efficiency thus treating the parameters $\mathbf{p}$ as deviations from expected system properties):

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon} \tag{A1}$$

where $\mathbf{X}$ is the (linearized) model operator. Suppose that the system is under-determined (as is always the case). Suppose that, for the sake of using traditional parameter estimation methods, pre-calibration simplification is achieved through estimation of a subset $\mathbf{p}_1$ of $\mathbf{p}$ while holding the other subset $\mathbf{p}_2$ of $\mathbf{p}$ fixed at its most likely value. (Other forms of simplification such as parameter lumping are also readily accommodated, as is discussed below). If $\mathbf{p}_1$ and $\mathbf{p}_2$ are represented explicitly, Eq. (A1) becomes:

$$\mathbf{h} = [\mathbf{X}_1 \quad \mathbf{X}_2]\begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\end{bmatrix} + \boldsymbol{\varepsilon} = \mathbf{X}_1\mathbf{p}_1 + \mathbf{X}_2\mathbf{p}_2 + \boldsymbol{\varepsilon} \tag{A2}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are defined through appropriate partitioning of $\mathbf{X}$. Ignoring parameter offsets is equivalent to assuming that the expected values of all parameters (including those comprising $\mathbf{p}_2$) are zero. Thus in assigning the elements of $\mathbf{p}_2$ a fixed value during (notional) OPR/PPR calibration, this value must be $\mathbf{0}$. $\mathbf{p}_1$ is then estimated through the traditional over-determined parameter estimation equation:

$$\underline{\mathbf{p}}_1 = (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{h} \tag{A3}$$

where $\underline{\mathbf{p}}_1$ signifies the estimated value of $\mathbf{p}_1$ and $\mathbf{Q}$ is a weight matrix chosen (if possible) to be such that:

$$\mathbf{Q} = \mathrm{C}^{-1}(\boldsymbol{\varepsilon}) \tag{A4}$$

where $\mathrm{C}(\boldsymbol{\varepsilon})$ is the covariance matrix of measurement noise. Parameter error is then given by:

$$\underline{\mathbf{p}} - \mathbf{p} = [\underline{\mathbf{p}}_1 - \mathbf{p}_1 \;\; \underline{\mathbf{p}}_2 - \mathbf{p}_2]^t = [(\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{h} - \mathbf{p}_1 \;\; - \mathbf{p}_2]^t \tag{A5}$$

Let us look at the first element of the parameter error vector in detail. If (A2) is substituted for $\mathbf{h}$ this first element becomes:

$$(\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{h} - \mathbf{p}_1 = (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}(\mathbf{X}_1\mathbf{p}_1 + \mathbf{X}_2\mathbf{p}_2 + \boldsymbol{\varepsilon}) - \mathbf{p}_1 \tag{A6}$$

which becomes:

$$(\mathbf{X}_1^t\mathbf{Q}\mathbf{X}1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{h} - \mathbf{p}_1 = (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_2\mathbf{p}_2 + (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\boldsymbol{\varepsilon} \tag{A7}$$

Combining this first term with the second term of (A5) results in:

$$\underline{\mathbf{p}} - \mathbf{p} = \begin{bmatrix}0 & (\mathbf{X}_1\mathbf{Q}\mathbf{X}_1^t)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_2\\ \mathbf{0} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\end{bmatrix} + \begin{bmatrix}(\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\boldsymbol{\varepsilon}\\ \mathbf{0}\end{bmatrix} \tag{A8}$$

from which it is seen that estimates $\underline{\mathbf{p}}_1$ of $\mathbf{p}_1$ are contaminated by $\mathbf{p}_2$. This is born of the need to fix the elements of this vector for the sake of achieving actual or notional model calibration. Actually, parameter contamination is an unavoidable outcome of calibration, no matter how it is undertaken. It can be expressed using the resolution matrix $\mathbf{R}$, defined as the relationship between what can be estimated and what exists in reality; see Moore and Doherty (2005) or Menke (1984) and Aster et al. (2005), among others, for further details. For regularization achieved through fixing of parameters $\mathbf{p}_2$, $\mathbf{R}$ is given by:

$$\mathbf{R} = \begin{bmatrix}\mathbf{I} & (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_2\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{A9}$$

As many of its diagonal elements are zero, this is a far-from-optimal resolution matrix. (The "ideal" resolution matrix is the identity matrix **I**.)

From (A8), the covariance matrix of parameter error is given by:

$$\mathrm{C}(\underline{\mathbf{p}}-\mathbf{p}) = \begin{bmatrix} 0 & (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_2 \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \mathrm{C}\left(\begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\end{bmatrix}\right) \begin{bmatrix} 0 & 0 \\ \mathbf{X}_2^t\mathbf{Q}\mathbf{X}_1(\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1} & -\mathbf{I} \end{bmatrix} + \begin{bmatrix} (\mathbf{X}_1^t\mathbf{Q}\mathbf{X}_1)^{-1} & 0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{A10}$$

where (A4) is used to simplify the second term and $\mathrm{C}\left(\begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\end{bmatrix}\right)$ is the covariance matrix of innate variability of $\begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\end{bmatrix}$, i.e., $\mathrm{C}(\mathbf{p})$. Let s be a prediction of interest, and let its dependence on parameters $\mathbf{p}$ be given by:

$$s = \mathbf{y}^t\mathbf{p} \tag{A11}$$

The variance of predictive error is then given by:

$$\sigma^2_{\underline{s}-s} = \mathbf{y}^t\mathrm{C}(\underline{\mathbf{p}}-\mathbf{p})\mathbf{y} \tag{A12}$$

with $\mathrm{C}(\underline{\mathbf{p}}-\mathbf{p})$ given as above. Eqs. (A12) and (A10) then form the basis for linear computation of predictive error variance. More importantly, they must also form the basis for computation of changes to predictive error variance accrued through adding or subtracting observations from the calibration dataset **h**, and hence for data-worth assessment. Although this is not among equations listed by Hill (2010) in the suggested enhanced OPR/PPR methodology, it is readily derivable from the Moore and Doherty (2005) error variance Eq. (26a):

$$\sigma^2_{\underline{s}-s} = \mathbf{y}^t(\mathbf{I}-\mathbf{R})\mathrm{C}(\mathbf{p})(\mathbf{I}-\mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}\mathrm{C}(\boldsymbol{\varepsilon})\mathbf{G}^t\mathbf{y} \tag{A13}$$

through substitution of (A9) for **R** and $(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}$ for **G**.

The methodology is easily extended to cases where notional calibration undertaken prior to data-worth assessment involves lumping of parameters rather than elimination of parameters for the purpose of achieving well-posedness of the inverse problem. Post-calibration subdivision of these lumped parameters into many for the purpose of data-worth analysis must then take place in implementing the expanded methodology suggested by Hill (2010). The **R** matrix pertinent to this procedure is readily calculated in the manner described by Moore and Doherty (2005).

The above methodology for analysis of data worth is based on the two-step process of calibration followed by computation of parameter and predictive error variance. A better alternative is to follow a Bayesian approach in which computation of parameter and predictive error variance is replaced by computation of parameter and predictive uncertainty. Predictive uncertainty (and reduction of predictive uncertainty accrued through acquisition of further data) can then be calculated using the following equation:

$$\sigma_s^2 = \mathbf{y}^t\mathrm{C}(\mathbf{p})\mathbf{y} - \mathbf{y}^t\mathrm{C}(\mathbf{p})\mathbf{X}^t[\mathbf{X}\mathrm{C}(\mathbf{p})\mathbf{X}^t + \mathrm{C}(\boldsymbol{\varepsilon})]^{-1}\mathbf{X}\mathrm{C}(\mathbf{p})\mathbf{y} \tag{A14}$$

This equation is derived in Doherty (2009) and Christensen and Doherty (2008). It was used by Dausman et al. (2009) and Dausman et al. (submitted for publication) in analysis of data-acquisition optimization in the context of saltwater intrusion modeling. Reductions in predictive uncertainty accrued by introduction of extra data to a calibration dataset (or through modification of $\mathrm{C}(\mathbf{p})$ to accommodate direct measurements of hydraulic properties) are easily calculated using this equation.

Use of Eq. (A14) is particularly efficient where parameter numbers are large, and observation numbers are small (a common context for data-worth analysis) as the dimensionality of the $[\mathbf{X}\mathrm{C}(\mathbf{p})\mathbf{X}^t + \mathrm{C}(\boldsymbol{\varepsilon})]$ matrix requiring inversion in this equation is equal to the number of observations comprising the calibration dataset. With some mathematical manipulation, Eq. (A14) can be reformulated as:

$$\sigma_s^2 = \mathbf{y}^t[\mathbf{X}^t\mathrm{C}^{-1}(\boldsymbol{\varepsilon})\mathbf{X} + \mathrm{C}^{-1}(\mathbf{p})]^{-1}\mathbf{y} \tag{A15}$$

which is more easily used where observation numbers are large and parameter numbers are small, as the dimensionality of the matrix which must be inverted is now equal to that of the number of parameters rather than the number of observations. The steps through which (A15) can be derived from (A14) are not provided herein, but can be found in Fienen et al. (submitted for publication). However the interested reader may like to note that linkage between these two equations relies on the following matrix identity:

$$(\mathbf{B}^t\mathbf{D}^{-1}\mathbf{B} + \mathbf{A}^{-1})^{-1}\mathbf{B}^t\mathbf{D}^{-1} = \mathbf{A}\mathbf{B}^t(\mathbf{B}\mathbf{A}\mathbf{B}^t + \mathbf{D})^{-1} \tag{A16}$$

References

Aster, R.C., Borchers, B., Thurber, C., 2005. Parameter Estimation and Inverse Problems. Elsevier Academic Press, Amsterdam. 301 pp.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources 16, 41–51.

Christensen, S., Doherty, J., 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. Advances in Water Resources 31 (4), 674–700.

Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2008. Quantifying data contributions toward reducing predictive uncertainty in a variable-density flow and solute/heat transport model. In: MODFLOW and More: Ground Water and Public Policy, Proceedings of the International Conference of the International Ground Water Modeling Center. Colorado School of Mines, Golden, CO, pp. 320–324.

Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., submitted for publication, Quantifying data worth through reducing predictive uncertainty. Ground Water.

Doherty, J., 2009. PEST: Model-independent Parameter Estimation. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. Journal of Hydrology 366, 119–127.

Doherty, J., Johnston, J.M., 2003. Methodologies for calibration and predictive analysis of a watershed model. Journal of American Water Resources Association 39 (2), 251–265.

Fienen, M.N., Hunt, R.J., Krabbenhoft, D.P., Clemo, T., 2009a. Obtaining parsimonious hydraulic conductivity fields using head and transport observations: a Bayesian geostatistical parameter estimation approach. Water Resource Research 45, W08405. doi:10.1029/2008WR007431.

Fienen, M.N., Muffles, C.T., Hunt, R.J., 2009b. On constraining pilot point calibration with regularization in PEST. Ground Water. doi:10.1111/j.1745-6584.2009.00579.

Fienen, M.N., Doherty, J., Hunt, R.J., and Reeves, H.W., submitted for publication. Using prediction uncertainty analysis to design hydrologic monitoring networks: Example applications from the Great Lakes Water Availability Pilot Project. US Geological Survey Scientific Investigation Report in preparation.

Gallagher, M., Doherty, J., 2007. Predictive error analysis for a water resource management model. Journal of Hydrology 34 (3–4), 513–533.

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

Hill, M.C., 2007. Quantifying groundwater model uncertainty, EOS Transactions AGU Fall 2007 meeting. <http://65.216.151.13/meetings/fm07/fm07-sessions/fm07_H21K.html>.

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

Hill, M.C., 2008b. But the fit is so good, EOS Transactions AGU Fall Meeting Fall 2008, <http://adsabs.harvard.edu/abs/2008AGUFM.H41H.05H>.

Hill, M.C., 2009. Good fit, error-based observation weights, and sensitivity analysis. National Ground Water Summit, Tucson, AZ, April 21, 2009.

Hill, M.C., 2010. Comment on "Two statistics for evaluating parameter identifiability and error reduction" by John Doherty and Randall J. Hunt. Journal of Hydrology 380 (3–4), 481–488.

Hill, M.C., Østerby, O., 2003. Determining extreme parameter correlation in ground water models. Ground water 41 (4), 420–430.

Hunt, R.J., Doherty, J., 2006. A strategy of constructing models to minimize prediction uncertainty. In: MODFLOW and More 2006 – Managing Ground Water Systems: Proceedings of the 7th International Conference of the International Ground Water Modeling Center. Colorado School of Mines, Golden, CO, pp. 56–60.

Hunt, R.J., Doherty, J., Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterization. Ground Water 45 (3), 254–262.

Lines, L.R., Treitel, S., 1984. Tutorial a review of least-squares inversion and its application to geophysical problems. Geophysical Prospecting 32 (2), 159–186.
Menke, W., 1984. Geophysical Data Analysis: Discrete Inverse Theory. Academic Press Inc.
Moore, C., Doherty, J., 2005. The role of the calibration process in reducing model predictive error. Water Resources Research 41 (5), W05050.
Moore, C., Doherty, J., 2006. The cost of uniqueness in groundwater model calibration. Advances in Water Resources 29 (4), 605–623.
Tonkin, M.J., Doherty, J., 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. Water Resources Research 41, W10412. doi:10.1029/2005WR003995.
Tonkin, M.J., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resources Research 45, W00B10. doi:10.1029/2007WR006678.
Tonkin, M., Doherty, J., Moore, C., 2007. Efficient nonlinear predictive error variance for highly parameterized models. Water Resources Research 43, W07429. doi:10.1029/2006WR005348.
Tonkin, M.J., Tiedeman, C.R., Ely, D.M., Hill, M.C., 2007. OPR-PPR, a computer program for assessing data importance to model predictions using linear statistics. US Geological Survey Techniques and Methods Report, Book 6, pp. 115.