

## Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models

Philip Brunner,<sup>1</sup> J. Doherty,<sup>2</sup> and Craig T. Simmons<sup>2,3</sup>

Received 1 September 2011; revised 30 May 2012; accepted 1 June 2012; published 17 July 2012.

[1] The data set used for calibration of regional numerical models which simulate groundwater flow and vadose zone processes is often dominated by head observations. It is to be expected therefore, that parameters describing vadose zone processes are poorly constrained. A number of studies on small spatial scales explored how additional data types used in calibration constrain vadose zone parameters or reduce predictive uncertainty. However, available studies focused on subsets of observation types and did not jointly account for different measurement accuracies or different hydrologic conditions. In this study, parameter identifiability and predictive uncertainty are quantified in simulation of a 1-D vadose zone soil system driven by infiltration, evaporation and transpiration. The worth of different types of observation data (employed individually, in combination, and with different measurement accuracies) is evaluated by using a linear methodology and a nonlinear Pareto-based methodology under different hydrological conditions. Our main conclusions are (1) Linear analysis provides valuable information on comparative parameter and predictive uncertainty reduction accrued through acquisition of different data types. Its use can be supplemented by nonlinear methods. (2) Measurements of water table elevation can support future water table predictions, even if such measurements inform the individual parameters of vadose zone models to only a small degree. (3) The benefits of including ET and soil moisture observations in the calibration data set are heavily dependent on depth to groundwater. (4) Measurements of groundwater levels, measurements of vadose ET or soil moisture poorly constrain regional groundwater system forcing functions.

**Citation:** Brunner, P. J. Doherty, and C. T. Simmons (2012), Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models, *Water Resour. Res.*, 48, W07513, doi:10.1029/2011WR011342.

### 1. Introduction

[2] With the development of physically based models such as PARFLOW [Maxwell and Miller, 2005], HydroGeoSphere [Therrien *et al.*, 2006], MikeShe (DHI Water and Environment, MikeShe, Integrated catchment modeling, 2012, available at <http://www.dhigroup.com>), or Hydrus 3d [Simunek *et al.*, 2011], the use of numerical models simulating saturated groundwater flow, unsaturated flow and vadose zone processes such as evapotranspiration is becoming more common. The vadose zone controls infiltration and evapotranspiration, and therefore exerts considerable influence on groundwater recharge. However, by including vadose zone processes in regional groundwater

models, the number of overall model parameters rises considerably. Given the increase in the number of parameters, the calibration of such models would be expected to benefit from an increase in the amount and type of observation data employed in their calibration. However, this does not seem to be the case in general groundwater modeling practice. Instead, parameters for hydrological models which include vadose zone processes are often constrained using head measurements, sometimes supplemented with measurements of river base flow to maintain the overall water balance. A comprehensive analysis of the extent to which vadose zone parameters of a regional scale model can be estimated using head observations alone has not yet been carried out; however, it is natural to suspect that vadose zone parameters are constrained by head measurements to only a small degree.

[3] New measurement technologies, such as remote sensing of the topsoil moisture content, and of the spatial distribution of evapotranspiration, are now becoming available for inclusion in regional scale calibration datasets. Koren *et al.* [2008] analyzed to what extent soil moisture observations could improve the estimation of parameters employed by a lumped catchment model in simulating the hydrograph at a basin outlet. Another recent example is the work of Li *et al.* [2009] who used remotely sensed patterns of phreatic evaporation, in addition to some observations of

<sup>1</sup>Centre d'Hydrogéologie et de Géothermie, University of Neuchâtel, Neuchâtel, Switzerland.

<sup>2</sup>National Centre for Groundwater Research and Training, Flinders University, Adelaide, South Australia, Australia.

<sup>3</sup>School of the Environment, Flinders University, Adelaide, South Australia, Australia.

Corresponding author: P. Brunner, Centre d'Hydrogéologie et de Géothermie, University of Neuchâtel, Rue Emil Argand 11, 2000 Neuchâtel, Switzerland. (philip.brunner@unine.ch)

hydraulic heads, to calibrate a groundwater flow model. *Bauer et al.* [2006] included flooding patterns obtained through remote sensing techniques in the calibration process of a coupled surface water groundwater model. *Hendricks Franssen et al.* [2008] obtained equally likely solutions to the inverse problem of groundwater model calibration by constraining these solutions using remotely sensed patterns of recharge. In calibrating the IHM (Integrated Hydrological Modeling) model at a study site in central Florida, *Zhang et al.* [2010] attempted to reproduce observations of ET, infiltration rates and soil moisture content. *Li, Q., et al.* [2008] used observations of streamflow (another spatially integrating observation) to estimate parameters controlling evapotranspiration (Leaf Area Index and root zone depth) in the vadose zone component of a fully coupled HydroGeoSphere model.

[4] A common conclusion from the abovementioned (and other) studies is that, despite the inclusion of additional, regionally applicable observations in the model calibration data set, many model predictions of interest are nevertheless characterized by relatively large uncertainties. From these studies it is thus apparent that augmentation of traditional calibration datasets with observations that appear to be directly informative of vadose zone properties has not yielded benefits that may have been expected.

[5] The matter of parameter uncertainty reduction accrued through data acquisition as applied to plot and laboratory-scale vadose zone modeling has received considerable attention in the literature, see for example *Wind* [1986], *Romano and Santini* [1999], *Zelege and Si* [2005], *Schwarzel et al.* [2006] and references cited therein. In a study that is pertinent to that described herein, *Montzka et al.* [2011] explored the extent to which remotely sensed observations of soil moisture (obtained through the SMOS and ALOS missions) could be used to estimate soil hydraulic parameters in synthetic 1-D models using data assimilation techniques. They concluded that including remotely sensed moisture observations in the calibration data set was beneficial, but pointed out that these benefits were dependent on soil type. *Ines and Mohanty* [2008a] and *Ines and Mohanty* [2009] estimated soil hydraulic properties using data assimilation techniques and soil moisture indicators obtained through different remote sensing platforms. In *Ines and Mohanty* [2008b] the influence of different climatic conditions on near-surface soil moisture assimilation for quantifying soil hydraulic properties were tested.

[6] A number of studies on small spatial scales (lysimeter of soil core scale) have focused specifically on the comparative information contents of different types of data in the vadose zone model calibration process. *Friedel* [2005] was one of the first to compare the reduction of predictive uncertainty incurred by inclusion of different combinations of observations in the calibration data set of a vadose zone model. The calibration data set variously comprised pressure head, temperature and/or the concentration of a solute. Similar investigations, though in a slightly different context, were undertaken by *Schneider-Zapp et al.* [2010]. They developed a model which computed evaporation through coupled simulation of boundary layer processes and unsaturated flow. Observations of matrix potential and evaporation were used to estimate boundary layer resistance, hydraulic conductivity, the two shape parameters of

the Van Genuchten equations and saturated water content for different soil types. A key conclusion of their study was that evaporation flux data provides a poor basis for estimation of saturated hydraulic conductivity. However, they found that estimation of these quantities can be significantly improved by including measurements of matrix potential in the calibration data set. *Ines and Droogers* [2002] included observations of transpiration, evaporation, soil moisture and evapotranspiration in their calibration datasets, both individually and in combination. *Jhorar et al.* [2002] estimated soil hydraulic properties in a numerical model (SWAP) using observations of evapotranspiration and soil moisture content, both individually and in combination. They concluded that if the root water extraction functions are known, frequent observations of ET can allow a model to reproduce the hydraulic behavior of the soil. However, they note that while it is possible to constrain parameters in order to achieve this end, the application of these constraints does not result in uniqueness of parameter values.

[7] Studies such as the above which specifically explore the effects of different types of data (used alone or in combination) on predictive uncertainty can assist in the design of field data acquisition campaigns. However, more studies of this type are needed, for analyses to date have limited their focus to only a small subset of the many observations types that are potentially available for regional model calibration. Further studies should include different data types in their analyses (particularly those pertaining to the regional scale), and address other issues such as the effect of measurement accuracy on the ability of observation data to reduce the uncertainties of predictions. The matter of measurement accuracy can be of considerable importance in the setting of regional models, particularly as remotely sensed measurement of, for example, evapotranspiration or soil moisture are likely to be of considerably less accuracy than those available in the laboratory.

[8] In this paper we attempt to address some of these knowledge gaps. We systematically compare the worth of several types of observation data in calibrating a 1-D vadose zone model. The amount of data types we cover significantly exceeds those employed in previous work of which we are aware, focusing on those that are available in regional studies. Data worth is assessed in terms of both its capacity to reduce predictive uncertainty and its capacity to increase parameter identifiability, the latter quantifying the level of parameter uniqueness forthcoming from the calibration process. Simulated vadose zone processes include infiltration, transpiration and evaporation. We focus on predictions of future water table elevations. Datasets employed for model calibration are initially only composed of water table elevations alone. The calibration data set is then supplemented with measurements of soil moisture (both throughout the profile or simply at the top of the soil column), evapotranspiration, transpiration and evaporation, and different combinations thereof. We investigate the extent to which noise associated with these different datasets degrades their ability to reduce model predictive uncertainty. Finally, by repeating the analysis for different hydraulic conditions we quantify how the worth of different types of data can alter as these conditions change. At the same time we investigate the differential effects of the depth to water table on the comparative worth of different

types of data. In undertaking this study we employ both linear and nonlinear methods, and compare their relative speed and efficacy.

[9] The remainder of this paper is organized as follows. First we briefly review available sources of data that could potentially help to reduce the uncertainties of predictions required by regional-scale groundwater/vadose zone models. We then introduce the mathematical foundations of the methods that we employ for uncertainty quantification, parameter identifiability analysis and data worth assessment. The model which forms the basis of our analysis is then introduced. Outcomes of our analyses are then presented and discussed. Some conclusions are then drawn.

## 2. Data Types

[10] Recent technological developments provide a range of data types covering various spatial and temporal scales that can potentially be used to supplement the use of heads in the calibration of regional groundwater/vadose zone models. In particular, large-scale patterns of soil moisture can be obtained using satellite data made available through various missions, for example the ALOS/PALSAR mission [Takada *et al.*, 2009], the SMOS mission [Kerr *et al.*, 2001], and the ASCAT mission [Brocca *et al.*, 2010], to mention but a few. Examples of soil moisture data obtained from airborne platforms are given by Saleh *et al.* [2009]. The methods discussed by these authors achieve root-mean-square errors of measured water content ranging from  $0.034 \text{ m}^3 \text{ m}^{-3}$  to  $0.054 \text{ m}^3 \text{ m}^{-3}$ , depending on the vegetation present. Recently developed ground based methods employed to measure soil moisture include ground based radiometry [Grant *et al.*, 2007], cosmic ray sensors [Zreda *et al.*, 2008] and ground penetrating radar [Weihermüller *et al.*, 2007]. Ground based estimates of soil moisture are typically more accurate than estimates obtained using air or spaceborne platforms. For example, Grant *et al.* [2007] reported root-mean-square errors of around  $0.004 \text{ m}^3 \text{ m}^{-3}$  for ground based radiometry. However their coverage is necessarily more limited.

[11] Estimation of the spatial distribution of ET using spaceborne multispectral data is now a standard technique. Discussions of available data sources and processing algorithms, as well as their application in hydrological sciences, can be found in, for example, Brunner *et al.* [2007], Timmermans *et al.* [2007] and Gowda *et al.* [2008]. In the work of Brunner *et al.* [2008] a method is proposed for splitting remotely sensed, imaged ET into its transpiration and phreatic evaporation components. In the review of Gowda *et al.* [2008], estimates of ET were between 67–97% of the actual ET, this demonstrating that while they may show some bias, their accuracy is such as to warrant inclusion in the calibration process, possibly following bias correction.

## 3. Discussion of Uncertainty

[12] In the present study we assume for simplicity that the numerical model we employ is realistic enough for its predictive uncertainty to be dominated by the uncertainty of its parameters, and not by uncertainties associated with prevailing environmental processes. These parameters represent the hydraulic properties of the system under investigation. In harmony with the complexity of natural systems, the

complexity of the system considered in the present study is such that the number of parameters required to characterize system processes is high.

[13] Let the vector  $\mathbf{p}$  designate parameters employed by the model. Let the prior probability distribution of these parameters be designated as  $P(\mathbf{p})$ , this encapsulating expert knowledge conditioned perhaps by direct measurements of system properties at one or a number of discrete locations. Let the vector  $\mathbf{h}$  comprise measurements of system state at discrete points in time, for example moisture content, evapotranspiration rate, elevation of the phreatic surface, etc. Bayes theorem states that

$$P(\mathbf{p}|\mathbf{h}) \propto P(\mathbf{h}|\mathbf{p})P(\mathbf{p}), \quad (1)$$

where  $P(\mathbf{p}|\mathbf{h})$  is the posterior probability distribution of model parameters.  $P(\mathbf{h}|\mathbf{p})$  is the so-called “likelihood function”; this increases with the extent to which the observation data set is matched by model outputs. Normally, a perfect fit between model outputs and field measurements is unattainable (or even sought) because of the presence of measurement/structural noise (which we characterize by the vector  $\boldsymbol{\varepsilon}$  in this paper) in the measurement data set  $\mathbf{h}$ . “History matching” or “calibration” thus becomes a filtering process through which the prior probability distribution of model parameters is narrowed through assignment of low posterior probabilities to parameter sets which do not allow the model to match the data well, and higher probabilities to those that do. Because model predictions of future system behavior are functions of parameters employed by the model, their uncertainty will hopefully be narrowed through the calibration/history-matching process in conjunction with the reduction of parameter uncertainty.

[14] Equation (1) thus assumes that the action of the model in computing outcomes corresponding to measurements of the calibration data set  $\mathbf{h}$  is described by an equation such as:

$$\mathbf{h} = \mathbf{Z}(\mathbf{p}) + \boldsymbol{\varepsilon}, \quad (2)$$

where  $\mathbf{Z}$  signifies the model as it operates on parameters  $\mathbf{p}$ .

[15] Direct implementation of Bayes equation in computing posterior parameter and predictive probability distributions can be a numerically intensive procedure, especially if a model is complex and its run time is large. Nevertheless, a number of methodologies for doing just this have indeed been implemented in conjunction with environmental models. Examples include the Markov chain Monte Carlo methodology as implemented by Gallagher and Doherty [2007], Vrugt and Robinson [2007], and others cited in these papers; see also the calibration-constrained predictive maximization/minimization methodology as implemented by Cooley and Christensen [2006], Tonkin *et al.* [2007], and Vecchia and Cooley [1987]. Other methodologies, though less mathematically rigorous from a Bayesian perspective, can obtain great efficiency gains through use of subspace techniques in computing samples from the posterior parameter distribution; see, for example Tonkin and Doherty [2009].

[16] Where a model’s outputs are linear with respect to its parameters, where the prior parameter probability distribution is Gaussian, and where measurements are contaminated

by noise which is also Gaussian in nature, equation (1) can be written as [see *Christensen and Doherty*, 2008]:

$$C(\mathbf{p}|\mathbf{h}) = C(\mathbf{p}) - C(\mathbf{p})\mathbf{Z}^t[\mathbf{Z}C(\mathbf{p})\mathbf{Z}^t + C(\boldsymbol{\varepsilon})]^{-1}\mathbf{Z}C(\mathbf{p}), \quad (3)$$

where  $C(\mathbf{p})$  is the covariance matrix associated with the prior parameter probability distribution,  $\mathbf{Z}$  is a matrix which characterizes the action of the (now linear) model on its parameters  $\mathbf{p}$ ,  $C(\boldsymbol{\varepsilon})$  is the covariance matrix of measurement noise, and  $C(\mathbf{p}|\mathbf{h})$  is the posterior covariance matrix of parameters, or in other words the matrix of parameter variability conditioned on the measurement data set  $\mathbf{h}$ . Derivation of equation (3) is based on the linearized equivalent of equation (2), which is

$$\mathbf{h} = \mathbf{Z}\mathbf{p} + \boldsymbol{\varepsilon}. \quad (4)$$

[17] Let the scalar  $s$  signify a prediction of interest. With the continued assumption of model linearity, the prediction  $s$  can be calculated from model parameters  $\mathbf{p}$  using the relationship

$$s = \mathbf{y}^t\mathbf{p}, \quad (5)$$

where  $\mathbf{y}$  is the vector of sensitivities of the model output  $s$  to parameters  $\mathbf{p}$  of the model. The posterior variance (square of standard deviation) associated with the posterior probability distribution of the prediction  $s$  can then be computed as

$$\sigma_{(s|\mathbf{h})}^2 = \mathbf{y}^tC(\mathbf{p})\mathbf{y} - \mathbf{y}^tC(\mathbf{p})\mathbf{Z}^t[\mathbf{Z}C(\mathbf{p})\mathbf{Z}^t + C(\boldsymbol{\varepsilon})]^{-1}\mathbf{Z}C(\mathbf{p})\mathbf{y}. \quad (6)$$

[18] The first term on the right of equation (6) specifies the precalibration uncertainty associated with the prediction  $s$  of interest. The second term characterizes the reduction in this uncertainty that is accrued through calibration against the data set  $\mathbf{h}$ .

[19] A feature of equation (6) is that while it includes sensitivities of model outputs to model parameters under both calibration and predictive conditions (through  $\mathbf{Z}$  and  $\mathbf{y}$ , respectively), it does not include actual parameter values, nor the model outputs associated with these parameter values. Hence, as *Dausman et al.* [2010], *James et al.* [2009], and *Moore* [2007] demonstrate, it can be easily used as a method for assessing the relative worth of different data acquisition strategies based on the premise that the worth of data increases with the extent to which acquisition of this data reduces the uncertainties of key model predictions. All that is required when using equation (6) to make such an assessment are the sensitivities of model outputs corresponding to new data to parameters employed by the model, as well as an estimate of the noise that is likely to be associated with this data. As actual data values are not required, the worth of data that has yet to be collected is thereby easily assessed. Data worth assessment based on equation (6) is available through the PEST suite of software; see *Doherty* [2010].

[20] Where models employ many parameters (as normally follows from physically based simulation of environmental processes), the model matrix  $\mathbf{Z}$  is likely to be characterized by a high-dimensional null space. Mathematically, the null

space is defined as that subspace of parameter space that is occupied by all vectors  $\delta\mathbf{p}$  for which:

$$\mathbf{0} = \mathbf{Z}\delta\mathbf{p}. \quad (7)$$

[21] If a set of parameters  $\underline{\mathbf{p}}$  can be found which calibrates a model, many other sets  $\underline{\mathbf{p}} + \delta\mathbf{p}$  which calibrate the model just as well as  $\underline{\mathbf{p}}$  can also be found, for combining (4) and (7) yields the equation:

$$\mathbf{h} = \mathbf{Z}(\underline{\mathbf{p}} + \delta\mathbf{p}). \quad (8)$$

[22] The null space is thus composed of individual parameters, or combinations of parameters, which are not estimable on the basis of the calibration data set. Hence their uncertainty is not reduced through the history-matching process. To the extent that a prediction depends on any such parameter or parameter combination, its uncertainty also accrues no reduction through the calibration process from that which could be made on the basis of expert knowledge alone. *Moore and Doherty* [2005] demonstrate this using a simple groundwater model, while *Gallagher and Doherty* [2007] demonstrate the dominant contribution that the null space makes to the uncertainty of many predictions of environmental interest in the groundwater modeling context, even where the calibration data set appears to be rich in information. In general, the greater the extent to which a prediction depends on broad-scale parameter combinations (for example parameter combinations which represent spatially averaged hydraulic properties), the greater the reduction in its uncertainty that is likely to be accrued through the calibration process. This is an outcome of the fact that broad-scale parameter combinations tend to be estimable, and hence do not lie within the calibration null space. In contrast, to the extent that a prediction is sensitive to parameterization detail, its uncertainty is less likely to be reduced through the model calibration process. This is an outcome of the fact that parameter combinations which express hydraulic property detail tend to lie within the calibration null space, and are hence inestimable. The concepts of parameter nonuniqueness and null space are closely related to that of “equifinality” as discussed extensively in papers such as *Beven* [2006] and references cited therein. A useful characterization of the “estimability status” of any parameter is the direction cosine of its projection from parameter space into the “solution subspace”. The solution space is the orthogonal complement of the null space. This direction cosine is referred to by *Doherty and Hunt* [2009] as the “identifiability” of that parameter. Parameter identifiability ranges between zero and one. If it is zero, the current calibration data set is completely uninformative of the parameter. If it is one, the opposite is the case, and uncertainty in estimating the value of this parameter arises only from measurement noise associated with the calibration data set. Alternatively, if the identifiability of a parameter lies somewhere between zero and one, then its value cannot be estimated uniquely, for the information content of the calibration data set pertaining to this parameter is inextricably mixed with that pertaining to at least one other parameter. In short, particular combinations of parameters whose identifiability is between 0 and 1 are more easily estimated

than *individual* parameters whose identifiability is less than 1.0 but greater than 0.0. Examples of parameters whose identifiabilities are in this range will be provided later in the paper.

[23] As stated above, nonlinear methods of gauging precalibration and postcalibration uncertainty are generally more difficult to implement than linear methods because of the requirement for a much greater number of model runs to be carried out. However where the uncertainty associated with only one or a small number of predictions must be explored, Pareto concepts offer a convenient and relatively model-run efficient, nonlinear methodology, even where a model employs many parameters.

[24] The “Pareto front” is a concept that is often employed to explore the trade-off between two or more objective functions that cannot all be simultaneously minimized. In the model calibration context these objective functions normally define model-to-measurement misfit pertaining to subsets of the total calibration data set composed of data of different types. The Pareto front is then defined as the locus of points comprising a hypersurface in objective function space over which it is not possible to lower one objective function without raising at least one of the others. Where differing objective functions measure a model’s ability (or lack thereof) to fit data of different types, possibly gathered at different locations, much can be learned of a model’s strengths and weaknesses (and of ways to reduce or eradicate the latter) through exploration of parameter sets comprising this front. See, for example, *Gupta et al.* [1998], *Boyle et al.* [2000], *Madsen* [2000, 2003], *Deb et al.* [2002], *Vrugt et al.* [2003] and *Vrugt and Robinson* [2007] for further details.

[25] *Moore et al.* [2010] demonstrated that the Pareto concept can also be applied to exploration of the posterior probability distribution of a prediction of interest. Only two objective functions are employed in this case. One of these is composed of the traditional calibration objective function, possibly supplemented with direct “observations” of parameter values based on expert knowledge. If weights assigned to these observations and articles of prior information are in accordance with measurement noise on the one hand and the prior probability distribution of parameters on the other hand, minimization of this objective function corresponds to finding the postcalibration expected value of the model parameter set. Here “expected value” is defined in terms of the posterior probability distribution of Bayes equation. We refer to this first objective function herein as the “calibration objective function.”

[26] The second objective function is defined as the squared difference between the user-specified value for a specific prediction, and the model-computed value of this same prediction. The user-specified value thus constitutes a kind of “attractor prediction”; as this component of the objective function is minimized, the model-generated counterpart to the prediction approaches the user-specified prediction value. *Moore et al.* [2010] show that a traversal of the Pareto front defined in this manner constitutes the solution to a series of constrained optimization problems in which, for a given value of the Bayes-equation-defined objective function (i.e., the first of the objective functions described above), the prediction is either maximized or minimized subject to the constraint that the Bayes objective

function is respected. Using statistics presented by *Vecchia and Cooley* [1987], the posterior probability distribution of the prediction of interest can therefore be constructed through traversal of the Pareto front. Alternatively, through consideration of only a few points along the Pareto front, the level of confidence associated with a certain predictive value not being exceeded or undercut can be readily assessed.

[27] The theory and concepts of this methodology are fully described by *Moore et al.* [2010]. In their study these authors explored the Pareto front using a modification of the efficient CMAES global optimization method of *Hansen and Ostermeier* [2001] and *Hansen et al.* [2003]. In some circumstances, more efficient traversal of a two dimensional Pareto front constructed to explore postcalibration predictive uncertainty can be implemented using gradient-based methods such as the Gauss-Marquardt-Levenberg method, or truncated singular value decomposition. The latter are employed by PEST [*Doherty*, 2010], and used in the study described herein, to effect an efficient traversal of the Pareto front through undertaking a series sequential recalibration exercises with greater and greater weight being placed on the attractor prediction. See documentation supplied with the PEST suite for full implementation details.

[28] In the analysis that we present below, the calibration objective function has a number of different components, this number varying with the specifics of each numerical experiment that we undertook. It includes one or a number of the different types of data employed in various calibration datasets, as well as deviations of parameters from their precalibration preferred values. In all cases, weights assigned to observation and parameter residuals were equated to the inverse of the standard deviations of associated measurement noise or prior parameter uncertainties. Hence a rising objective function infers decreasing posterior parameter probability in accordance with Bayes equation. As we assumed normality for both measurement noise and prior parameter distributions, an objective function of 1.0 marks one standard deviation of removal of a prediction from its postcalibration expected value. The difference between maximized and minimized predictive values constrained by an objective function of 1.0 in each case thus defines the 67% two-sided confidence interval of the prediction. In the discussion which follows, we loosely refer to this as the “uncertainty” of the prediction. Through repeating an analysis of this type with various observations (or even all observations) included/excluded from the calibration data set, the dependency of the uncertainty of the prediction on the composition of the calibration data set can be gauged. Through this mechanism the “worth” of different data types in reducing the uncertainty of the prediction can thus be calculated. This provides a nonlinear alternative to the linear data worth analysis methodology discussed above; however, notwithstanding the efficiency of the Pareto methodology, it is a far more numerically expensive alternative to linear data worth analysis.

## 4. Modeling Context

### 4.1. Conceptual Formulation

[29] We use the numerical model HydroGeoSphere [*Therrien et al.*, 2006] for our simulations. We briefly present the governing flow equations and the conceptualization

of evapotranspiration (ET) (following the HGS manual); many of the parameters defined in these equations will subsequently be subject to adjustment during calibration of the model used in our study. More details, together with discussions of the governing equations, are given by *Therrien et al.* [2006] and *Maneta et al.* [2008].

[30] In HGS subsurface flow is calculated using the Richards equation:

$$-\nabla(q) + \sum \Gamma_{ex} \pm Q = \frac{\partial}{\partial t}(\theta_s S_w). \quad (9)$$

[31] In equation (9),  $S_w$  is the degree of water saturation [-],  $\theta_s$  is the saturated water content (porosity);  $Q$  ( $L^3 L^{-3} T^{-1}$ ) represents a fluid exchange from boundary conditions;  $\Gamma_{ex}$  represents the volumetric fluid exchange rate ( $L^3 L^{-3} T^{-1}$ ) between the subsurface domain and all other types of domains supported by the model expressed per unit volume of the other domain types. The fluid flux  $q$  ( $L T^{-1}$ ) is calculated as

$$q = -Kk_r \nabla(\gamma + z), \quad (10)$$

where  $K$  is the saturated hydraulic conductivity ( $L T^{-1}$ ),  $k_r$  represents the relative permeability of the medium [-] and is a function of the degree of water saturation,  $\gamma$  is the pressure head (L), and  $z$  is the elevation head [L]. HydroGeoSphere implements a number of different characterizations of the relationship between pressure, saturation, and hydraulic conductivity. For the present study we use the formulation suggested by *Van Genuchten* [1980]:

$$k_r = S_e^{(l_p)} \left[ 1 - \left( 1 - S_e^{\frac{1}{\nu}} \right) \right], \quad (11)$$

and

$$S_e = \frac{(1 - S_{wr}) \left[ 1 + |\alpha \gamma|^{-\beta} \right]^{-\nu}}{(1 - S_{wr})}, \quad (12)$$

where

$$\nu = \left( 1 - \frac{1}{\beta} \right). \quad (13)$$

[32] In the above equations  $l_p$  (L) is the pore connectivity parameter,  $\gamma$  is the pressure head (L),  $S_e$  (-) is the effective saturation,  $S_{wr}$  is the residual water saturation,  $\alpha$  ( $L^{-1}$ ) is the inverse of the air-entry pressure head and  $\beta$  (-) is the pore size distribution index. Transpiration from vegetation occurs within the root zone of the subsurface and is a function of the leaf area index (LAI) [-], nodal water content  $\theta$  [-] and a root distribution function (RDF) over a prescribed extinction depth. The rate of transpiration ( $T_p$ ) is estimated using the following relationship [*Kristensen and Jensen*, 1975]:

$$T_p = f_1(LAI) f_2(\theta) RDF(E_{pot} - E_{can}), \quad (14)$$

where  $E_{pot}$  is the reference evapotranspiration, [ $L T^{-1}$ ] and  $E_{can}$  is the tree canopy evaporation [ $L T^{-1}$ ]. Because

interception and the related evaporation from canopy is not considered in this study,  $E_{can}$  is not discussed further. The vegetation function ( $f_1$ ) relates the transpiration ( $T_p$ ) to the leaf area index (LAI) in a linear fashion and is expressed as:

$$f_1(LAI) = \max \{0, \min [1, C_2 + C_1 LAI]\}, \quad (15)$$

where  $C_1$  and  $C_2$  are dimensionless fitting parameters. The root zone distribution function (RDF) is given by:

$$RDF = \frac{\int_{z_1}^{z_2} r_f(z') dz'}{\int_0^{L_r} r_f(z') dz'}, \quad (16)$$

[33] In equation (16)  $L_r$  is the effective root length [L],  $z$  is the depth beneath the soil surface [L], while  $r_f(z')$  is the root extraction function [ $L^3 T^{-1}$ ]. The moisture content function ( $f_2$ ) in equation (17) relates  $T_p$  to the moisture state of the roots and is expressed as:

$$f_2(\theta) = \begin{cases} 0 & \text{for } 0 \leq \theta \leq \theta_{wp} \\ f_3 & \text{for } \theta_{wp} \leq \theta \leq \theta_{fc} \\ 1 & \text{for } \theta_{fc} \leq \theta \leq \theta_{oc} \\ f_4 & \text{for } \theta_{oc} \leq \theta \leq \theta_{an} \\ 0 & \text{for } \theta_{an} \leq \theta \end{cases}, \quad (17)$$

$$f_3(\theta) = 1 - \left( \frac{\theta_{fc} - \theta}{\theta_{fc} - \theta_{wp}} \right)^{C_3}, \quad (18)$$

$$f_4(\theta) = 1 - \left( \frac{\theta_{an} - \theta}{\theta_{an} - \theta_{oc}} \right)^{C_3}, \quad (19)$$

where  $C_3$  is a dimensionless fitting parameter,  $\theta_{fc}$  is the moisture content at field capacity,  $\theta_{wp}$  is the moisture content at the wilting point,  $\theta_0$  is the moisture content at the oxalic limit,  $\theta_{an}$  is the moisture content at the anoxic limit, and  $r_f(z)$  is the root extraction function [ $L^3 T^{-1}$ ]. Below the wilting-point moisture content, transpiration is zero; transpiration then increases with moisture content to a maximum at the field-capacity moisture content. This maximum is maintained up to the oxalic moisture content, beyond which the transpiration decreases to zero at the anoxic moisture content. When the available moisture content exceeds the anoxic moisture content, the roots become inactive due to lack of aeration.

[34] In HGS, evaporation from the soil surface and subsurface soil layers is a function of nodal water content and an evaporation distribution function (EDF) over a prescribed extinction depth. The model assumes that evaporation occurs together with transpiration, this resulting from energy that penetrates the vegetation cover. It is expressed as

$$E_s = \alpha^* (E_p - E_{can} - T_p) EDF. \quad (20)$$

[35] The wetness factor ( $\alpha^*$ ) is given by

$$\alpha^* = \begin{cases} \frac{\theta - \theta_{e2}}{\theta_{e1} - \theta_{e2}} & \text{for } \theta_{e2} \leq \theta \leq \theta_{e1} \\ 1 & \text{for } \theta > \theta_{e1} \\ 0 & \text{for } \theta < \theta_{e2} \end{cases}, \quad (21)$$

where  $\theta_{e1}$  is the moisture content at the end of the energy-limiting stage (above which full evaporation can occur) and  $\theta_{e2}$  is the limiting moisture content below which evaporation is zero.

[36] EDF is an evaporation density function defined by the user. It is assumed that the amount of available energy for evaporation decreases with increasing depth. Here, we have chosen a linear function to describe the rate of decrease between the soil surface and the extinction depth  $L_E$  (L).

[37] The rate of transpiration for a give node  $i$  ( $T_{pi}$ ) can be estimated by substituting  $\theta$  in equation (21) with the nodal water content  $\theta_i$  featured in equations (6)–(11). The total transpiration rate is then calculated using:

$$T_P = \sum_{i=1}^{n_R} T_{pi}, \quad (22)$$

where  $n_R$  is the number of nodes that lie within the depth interval  $0 \leq z \leq L_r$ . The rate of evaporation for node  $i$  can then be estimated by substituting the nodal water content  $i$  and nodal evaporation distribution function  $EDF_i$  into equations (14)–(19).

## 4.2. Modeling Strategy

[38] Our analysis is based on a finely vertically discretized 1-D column filled with soil. We simulate evapotranspiration (ET) as well as two infiltration events to generate synthetic observations of different types. (The details of the observation types generated for use in calibrating the model, as well as the assumed standard deviations of noise associated with these measurements, are described in section 4.3.) To generate these synthetic observations, homogenous hydraulic properties were imposed throughout the soil column, with each hydraulic parameter being assigned a value equal to its mean (i.e., its expected value from the standpoint of its prior probability distribution). In this way model parameters and corresponding model outputs (and therefore synthetic observations used in the calibration process) thus represent precalibration minimum error variance estimates of these quantities. The observations were generated daily for the first 30 days of the model run. Different combinations of these observations were then used to calibrate the model parameters. In parameterizing the model for the purpose of model calibration, soil heterogeneity is represented by 10 zones with a vertical extension of 0.5 m for each. The hydraulic parameters assigned to these zones are treated as uncertain quantities whose values are informed by expert knowledge, and then refined through matching model outputs with observations through the calibration process.

[39] The model, with calibration-constrained parameters, was subsequently employed to predict the hydraulic head at a point in time 37 days into the future. The uncertainty of this prediction, as well as the identifiabilities of all calibrated parameters, were evaluated for individual as well as

combinations of observation data types comprising the calibration data set. This allowed quantification of the extent to which observations of different types, and in different combinations, affect the identifiability of individual parameters, as well as the uncertainty associated with the prediction of interest. To account for different hydrological conditions, we repeated the entire procedure for three different initial conditions, these representing shallow, medium, and deep water table conditions. Finally, the weights associated with the observations were varied in order to explore the relationship between measurement accuracy and predictive uncertainty.

[40] As discussed above, quantification of parameter identifiability and predictive uncertainty is readily achieved through linear analysis. However, as the processes affecting movement of water through the unsaturated zone are in fact highly nonlinear, the validity of linear analysis in this context must be established. We did this by addressing issues of predictive uncertainty and data worth using both linear analysis and the more numerically intensive nonlinear analysis procedure based on Pareto concepts; we then compared results of both of these analyses.

## 4.3. Model Setup, Parameterization and Generation of Synthetic Observation Data

[41] We now describe the synthetic model that forms the basis of our analysis. Like any model, it is built on a domain of a certain size. However, the modeling exercise (if not the model itself) is intended to allow conclusions to be drawn that pertain to a variety of modeling scales. The issue of scale is implied in the types of measurements that we examine, in the errors that are assumed to pertain to those measurements, and in conclusions that we draw on data worth, and algorithmic relevance to small and regional model domains.

[42] We analyze movement of water within a column of unit area. The total height  $z$  of the column is 5 m, while the vertical numerical grid discretization is 2.5 cm. In the simulation representing shallow conditions, the initial head was set to 5 m for all nodes in the model domain; therefore the entire column was initially saturated. For the medium water table conditions, the initial head was set to 4.25 m for all nodes in the model domain. For the deep conditions, the initial head was set to 3.5 m for all nodes in the model domain. The side and bottom of the model domain are no flow boundaries. We applied two different types of forcing functions to the top boundary, these being infiltration and potential evaporation. The former is composed of two events, while the latter is continuous and invariant. All soil properties and forcing functions (including intensity and duration of the rain events) used to generate the synthetic observation data set are specified in Table 1.

[43] The timing of the rain events was the same for medium and deep conditions, namely on days 20 and 35. However, for the shallow water table conditions the precipitation events were applied earlier, this ensuring that observations used in the parameter estimation process all represent shallow conditions. Note, however, that the intensity and duration of infiltration were the same in all cases. In Figure 1 the response of hydraulic head to the employed forcing functions is shown for the three conditions.

**Table 1.** Notation, Units of Relevant Model Parameters, and Forcing Functions<sup>a</sup>

	Symbol	Value	Lower Limit	Upper Limit	Unit
<i>Model Parameter</i>					
Hydraulic conductivity	$K$	4	1	20	$\text{m d}^{-1}$
Porosity (Saturated water content)	$\theta_s$	0.3	0.2	0.45	–
Residual water saturation	$S_{wr}$	0.05	0.005	0.15	–
Van Genuchten $\alpha$	$\alpha$	4	2	10	$\text{m}^{-1}$
Van Genuchten $\beta$	$\beta$	2	1.01	4	–
Evaporation extinction depth	$L_E$	0.5	0.1	4	m
Evaporation limiting saturation (min)	$\theta_{e1}$	0.25	–	–	–
Evaporation limiting saturation (max)	$\theta_{e2}$	0.5	0.1	4	–
Transpiration extinction depth (Root depth)	$L_T$	1	0.1	4	m
Leaf area index	(LAI)	1	0.001	5	–
Transpiration fitting parameter	$C_1$	0.5	0.001	1	–
Transpiration fitting parameter	$C_2$	0.1	0.001	1	–
Transpiration fitting parameter	$C_3$	1	0.001	1	–
Transpiration limiting saturation (wilting point)	$\theta_{wp}$	0.1	0.0001	0.2	–
Transpiration limiting saturation (field capacity)	$\theta_{fc}$	0.3	0.201	4	–
Transpiration limiting saturation (oxic limit)	$\theta_0$	0.5	0.401	0.7	–
Transpiration limiting saturation (anoxic limit)	$\theta_{an}$	0.8	0.701	1	–
<i>Forcing Function</i>					
Potential evaporation	$E_{pot}$	0.01	0.007	0.013	m
Rain event 1 deep and medium water table (days 20/21)	$RE1$	0.02	0.016	0.024	$\text{m d}^{-1}$
Rain event 2 deep and medium water table (day 35)	$RE2$	0.05	0.046	0.054	$\text{m d}^{-1}$
Rain event 1 shallow water table (days 5/6)	$RE1$	0.02	0.016	0.024	$\text{m d}^{-1}$
Rain event 2 shallow water table (days 20/21)	$RE2$	0.05	0.046	0.054	$\text{m d}^{-1}$

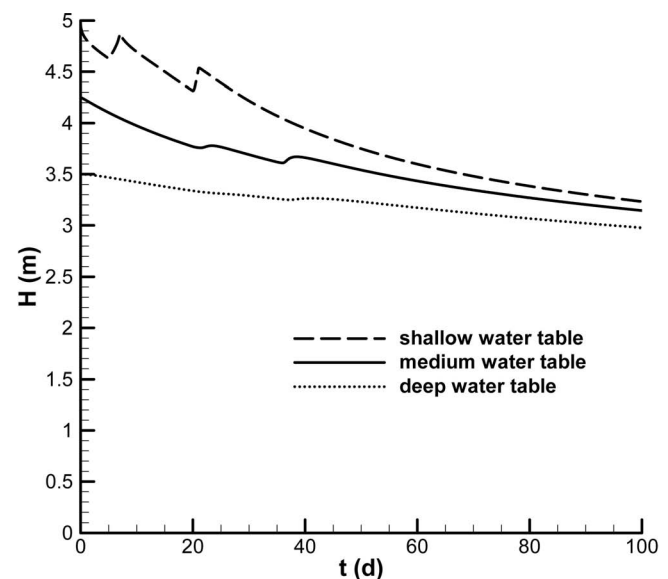
<sup>a</sup>The column “Value” shows the data that were used in the original model to generate sets of observation data. The upper and lower range represent the given upper and lower bounds used during the calibration process. The hydraulic parameters are the same throughout the column.

[44] For all three hydrological conditions, synthetic observations were obtained once a day (and subsequently used in model calibration) for the first 30 days of the simulation. Every observation was given a certain weight in the calibration process, this weight being the inverse of the assumed measurement standard deviation. We generated the following observations: hydraulic head, soil moisture content at the top of the column measured with a low accuracy (a measurement that could be based on remote sensing data), soil moisture content at the top of the soil with a high accuracy (representing a measurement using ground based methods such as radiometry), a profile of soil moisture content, as well as observations of transpiration, evaporation and evapotranspiration. An overview of the synthetic observations generated in this manner is provided in Table 2. Table 2 also shows the depths at which the observations were taken, as well as the weights used in calibrating the model. To explore the influence of measurement accuracy on the outcomes, we varied these weights in the final part of the analysis (these altered weights are not shown in Table 2).

[45] The observations discussed above formed the basis on which all model parameters were constrained during calibration. In total, 65 parameters were adjusted in every calibration run. In every layer (10 in total) of the column hydraulic conductivity, porosity, residual water saturation, as well as Van Genuchten parameters  $\alpha$  and  $\beta$  were adjusted (50 parameters). Additionally, the 12 parameters related to the simulation of ET, absolute values of potential evaporation, and the magnitude of the two rain events, were adjusted. (The timing of the rain events was not adjusted.) The upper and lower limits for all parameters imposed during the calibration process are listed in Table 1.

[46] Clearly, the inverse problem defined by the calibration process as thus described is ill-posed as all of the abovementioned parameters cannot be estimated uniquely

on the basis of the synthetic data set. In fact a unique solution to the inverse problem of model calibration is not being sought. To the extent that calibration is undertaken in any environmental modeling context it must attain uniqueness through seeking the posterior expected value of each



**Figure 1.** Model response (hydraulic head measured at  $z = 0$ ) to two rain events and continuous evapotranspiration for three water table depth scenarios. The key difference between the medium and deep water table conditions are the initial conditions. The timing of the rain events is the same for the medium and large depths to groundwater while the rain events are applied earlier for the shallow water table case.

**Table 2.** Observation Types, Abbreviation, Measurement Location, Their Weight During the Calibration and the Corresponding SD of the Measurement<sup>a</sup>

Observation Type	Abbreviation	Measured at	Weight (Provided Used During Calibration)	Measurement SD
Hydraulic head	H	$z = 0$	10	0.1 m
Saturation profile	S	$z = 4.0; 4.25; 4.5; 4.75; 5$	5	0.2 (-)
Saturation at the top	St	$z = 5$	5	0.2 (-)
Saturation top, high accuracy	Sth	$z = 5$	50	0.02 (-)
Evapotranspiration	ET	Integral from $z = 0$ to $z = 5$	500	0.002 m
Evaporation	E	Integral from $z = 0$ to $z = 5$	500	0.002 m
Transpiration	T	Integral from $z = 0$ to $z = 5$	500	0.002 m

<sup>a</sup>All observations are obtained daily from the model as described above, for a time period of 30 days. The standard deviation of the measurement is the inverse of the weight.

parameter, i.e., the expected value of each parameter as calculated from the posterior parameter probability density function. In contrast, the present study seeks to explore the extent of parameter variability that is in accordance with the posterior parameter probability density function, and the consequential extent of predictive variability. The high-dimensional null space fosters a high degree of posterior uncertainty for parameters, and for predictions which are sensitive to them. Exploration and reduction of this uncertainty is the principle aim of our study. The large number of parameters whose values are not exactly known, and are hence adjustable in the uncertainty analysis process, reflects the fact that natural soil properties are highly heterogeneous. It follows from recognition of this fact that the parameterization scheme on which uncertainty analysis is based should reflect this heterogeneity.

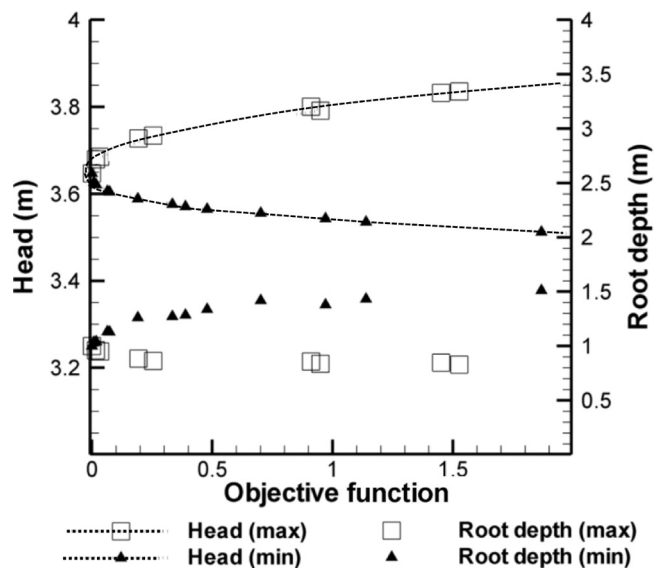
[47] In assigning prior probabilities to all adjustable parameters, normality and statistical independence of parameters was assumed. Each parameter was assigned a standard deviation equal to a quarter of the distance between its upper and lower bounds. Clearly, the assumption of statistical independence is simplistic. It is reasonable to suggest that our study would have been more realistic if spatial correlation between parameters of the same type, and statistical correlation between parameters of different types at the same location, were assumed. Such an assumption would have affected neither the implementation nor the performance of either of the linear or nonlinear methodologies, a different covariance matrix would simply have been employed to characterize prior parameter probability. However, the assumption of statistical independence between parameters allowed us to obtain a better sense of the intrinsic worth of different types of observation data, as reductions in the uncertainty of one parameter acquired through application of calibration constraints on another parameter was thus prevented. In other words, the purposeful omission of expert knowledge implied by the existence of off-diagonal terms of the precalibration covariance matrix, allowed us to gain a clearer picture of the worth of different data types, as it is this data alone that is able to effect parameter, and hence predictive, uncertainty reduction through the calibration process.

## 5. Results

### 5.1. Pareto Method

[48] As has already been discussed, in implementing uncertainty analysis using Pareto concepts, a user monitors the growth in the calibration objective function (this measuring

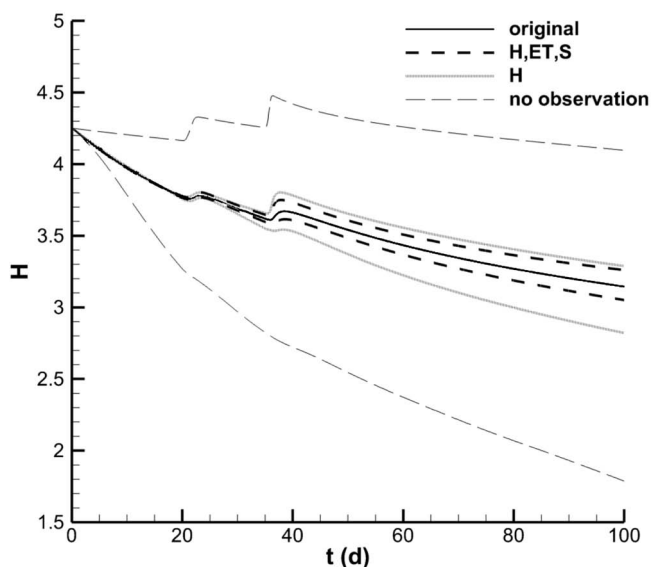
misfit between field measurements and corresponding model outputs as well as between parameters and their precalibration expected values) as the value of a model prediction slowly changes upward or downward while traversing the Pareto front. The Pareto front then provides the lowest value of the calibration objective function that is compatible with any value of the prediction. Figure 2 shows an example. In this plot, the variation of the objective function as the prediction rises (indicated as “Head(max)” in Figure 2) and falls (indicated as “Head(min)” in Figure 2) is demonstrated. The prediction is the hydraulic head at 37 days. Also shown is the variation of one of the model parameters (the root depth) with distance along the Pareto front. It is apparent that predictions of lower head are associated with an increase in root depth, as would be expected. Note that a total of 3700 model runs were required for generation of Figure 2; as has already been discussed, Pareto analysis was undertaken using PEST; see *Doherty* [2010].



**Figure 2.** Plot of predicted hydraulic head (left axis) as a function of minimum objective function required to achieve that prediction. The second y axis shows root depth as a function of minimized objective function. The simulations are based on medium-depth water table conditions. Note that for an objective function of zero, parameter values, and simulated heads are those expected on the basis of the prior parameter probability distribution.

[49] Every point plotted in Figure 2 corresponds to a point along the Pareto front, and hence to a set of parameter values that can be considered to (almost) calibrate the model. In Figure 3, the variation of heads with time, calculated using parameters corresponding to an objective function of 1.0, are shown; that is, the head at 37 days is maximized/minimized subject to the constraining objective function value of 1.0. (Note that the objective functions obtained through PEST's implementation of the Pareto method may never exactly correspond to 1.0; hence linear interpolation between the heads corresponding to the closest two objective functions must be used to infer the head corresponding to an objective function of 1.0.) The data shown in Figure 3 pertain to the medium-depth water table conditions. Three pairs of minimum/maximum predictions are shown for three different combinations of observations, namely no observations, heads only, and heads + ET + observations of the saturation profile. Head is plotted against time also for the precalibration expected values of the parameters, these being employed by the model when it was used to generate the "observations" comprising the calibration data set, as discussed above. Plots are labeled as "no observation" (where the prediction is constrained only by the prior distribution of model parameters) "H" (where in addition to the prior distribution of the model parameters hydraulic heads comprise the calibration data set and hence constrain the parameters), and "H,ET,S" (for constraints imposed by hydraulic heads, a saturation profile, as well as observations of ET in addition to the prior distribution of parameters).

[50] As stated above, because the Pareto method allows a user to calculate the uncertainty associated with a particular prediction conditioned by constraints imposed by different sets of observations, it thereby allows the user to quantify



**Figure 3.** Plot of hydraulic head versus time where the prediction is constrained by differently constituted calibration datasets. The solid line in the center represents the original parameter set used to generate the calibration data set; this is composed of precalibration expected parameter values (medium-depth water table conditions).

the relative worth of different calibration datasets, with "worth" being calculated as the decrease in uncertainty accrued through use of that observation data set. We define the "relative uncertainty reduction"  $\sigma_r$  of a prediction accrued by use of a certain calibration data set, through the equation:

$$\sigma_r = (\sigma_0 - \sigma_c) / \sigma_0, \quad (23)$$

where  $\sigma_0$  is the precalibration uncertainty of that prediction, and  $\sigma_c$  is the postcalibration uncertainty of the prediction conditioned by the calibration data set under consideration.

[51] The predictive uncertainty decrease was calculated in this manner for 4 different combinations of observations, these being (1) heads only; (2) heads and saturation at the top of the profile; (3) heads and saturation at the top of the profile measured with a high level of accuracy; and (4) heads, ET and a profile of saturation measurements.

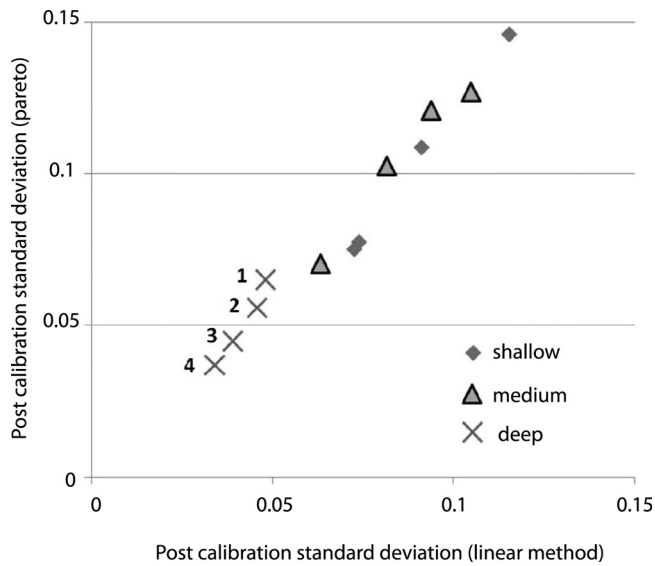
[52] This was repeated for shallow, medium, and deep water table conditions. Not surprisingly, the computational requirements of this undertaking were considerable. If a linearity assumption can be tolerated (i.e., linearity of model outputs with respect to model parameters) these same calculations can be undertaken with greatly reduced numerical burden using equation (6).

## 5.2. Linear Analysis

[53] The computational cost of using equation (6) for quantification of data worth is far smaller than that of Pareto analysis because it requires that only one model run be conducted per parameter, these being required for filling of the  $\mathbf{Z}$  matrix which constitutes linearization of the model under calibration conditions, and the  $\mathbf{y}$  vector of predictive sensitivities to model parameters. (Filling of a column of  $\mathbf{Z}$  and an element of  $\mathbf{y}$  pertaining to a particular parameter required only one model run in the present case.)

[54] In order to assess the legitimacy of linear analysis when applied to our nonlinear model, postcalibration predictive standard deviations were calculated using both the linear and the Pareto methodologies; the results are shown in Figure 4. Uncertainty was obtained using the Pareto method by calculating the minimum and maximum head predictions corresponding to an objective function of 1.0; compare with Figure 2. (As previously mentioned, the objective functions obtained through the Pareto method never exactly correspond to 1.0. Hence linear interpolation between the heads corresponding to the closest two objective functions was used to infer the head corresponding to an objective function of 1.0.) The post calibration standard deviation of predicted head was calculated in this way for different combinations of observations, for the three different hydrological conditions (shallow, medium, and deep water tables) employed in our study.

[55] The comparison of Pareto and linear estimates of predictive uncertainty strongly suggests that the linear method's assessment of post calibration predictive uncertainty is reliable for this particular model setup. This finding allows us to employ the linear method to further analyze the relationships between measurement type and predictive accuracy under the range of conditions simulated by our model.



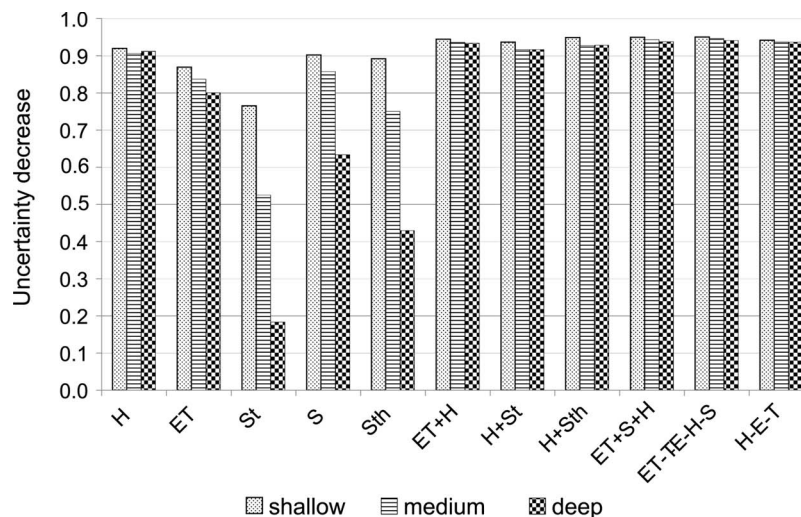
**Figure 4.** Comparison of postcalibration standard deviations obtained through the Pareto method and those obtained through linear analysis for the three different depths to groundwater employed in our study. The analysis was carried out for the following combinations of observations: heads only (indicated with “1” at the deep water table condition), heads and saturation at the top (labeled as “2”), heads and saturation at the top of the profile with high accuracy (labeled as “3”), and heads, ET and profile of saturation (labeled as “4”). The ordering of points in the figure for the medium and shallow water table conductions is the same as that for the deep water table conditions.

[56] First, the relative predictive uncertainty reduction gained through use of a wide range of combinations of observation is calculated and compared. This comparison (Figure 5) is instructive. It is apparent that the observation type of highest worth in terms of its capacity to reduce the

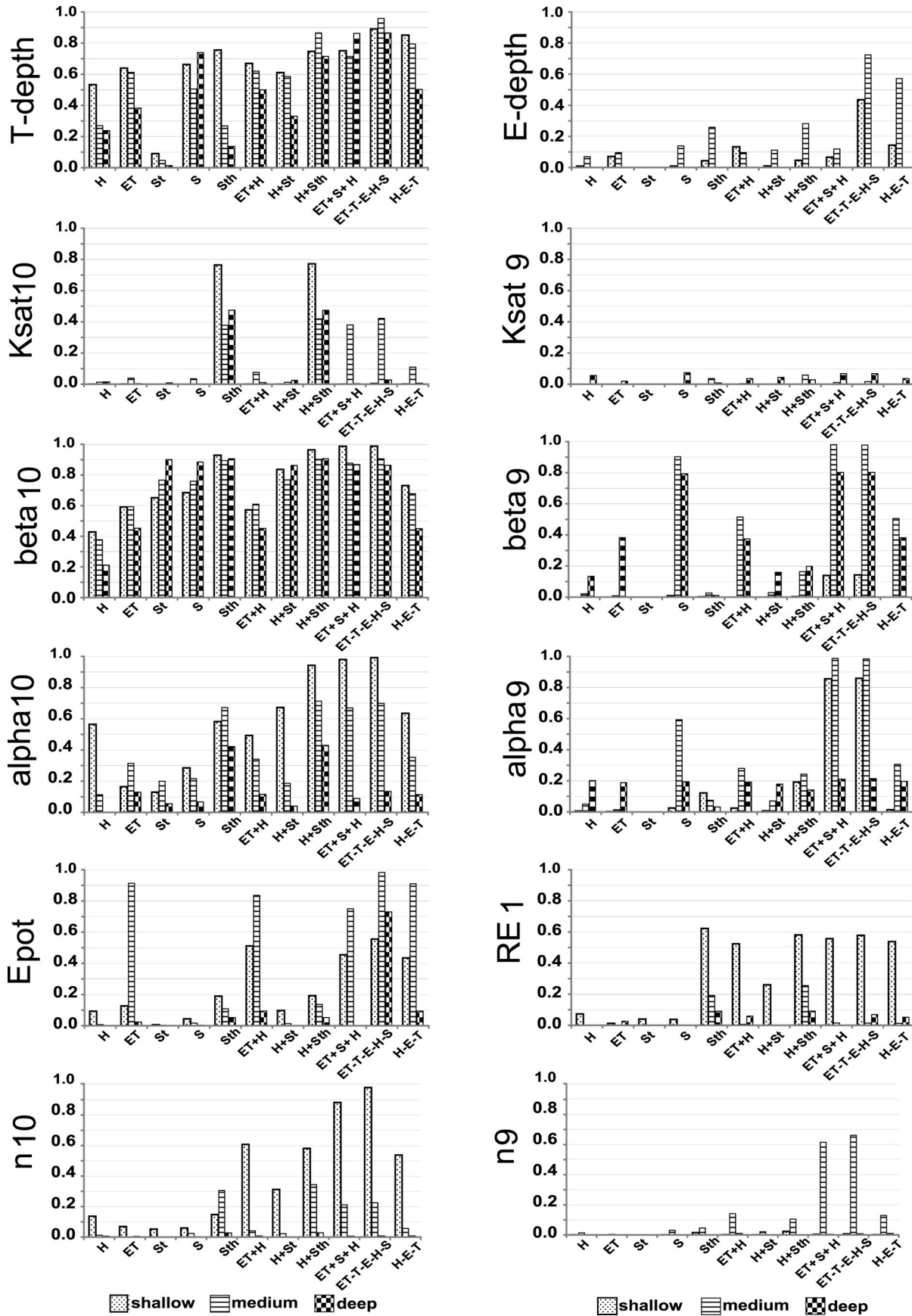
uncertainty of predicted head at 37 days is the hydraulic head observation type (i.e., H). (Recall that the calibration period extends from day 1 to day 30, and that the head prediction follows from a rain event that is not included in the calibration data set.) The heads-only calibration data set performs equally well for all hydraulic scenarios. Sole usage of evapotranspiration (ET) observations also promulgates a large reduction in predictive uncertainty. (Note the slight decrease in uncertainty reduction with increasing depth to groundwater.) The worth of saturation measured at the top of the soil column (St) is highly dependent on the depth to groundwater. While for shallow conditions a measurement of soil moisture at the top of the column can reduce predictive uncertainty significantly, it is of little use when the water table is deep. The same dependency of observation worth on depth to the water table can be seen for observations of soil moisture down the profile (S), as well as for a saturation measurement at the top of the soil column with high accuracy (Sth).

[57] From Figure 5 it is apparent that use of heads in combination with different types of other observations aids uncertainty reduction. However, because the use of hydraulic heads alone already promulgates a relative uncertainty reduction of around 0.9 in the prediction of future heads, the additional decrease in uncertainty gained through inclusion of other types of data in the calibration data set is relatively insignificant.

[58] Figure 6 shows the identifiability of several individual model parameters for a wide range of observation combinations for the three hydraulic conditions. Recall that, like relative parameter uncertainty reduction, identifiability ranges between zero and one. It provides a measure of solution and null space dimensionality as a function of calibration data set constitution, and of the relationship of individual parameters to these two subspaces. Note that in theory, if the identifiability of a parameter is less than one, then its identifiability will either increase or stay the same as the calibration data set is expanded. In practice, small drops in identifiability are seen for some parameters in Figure 6 as



**Figure 5.** Uncertainty decrease for a prediction of the hydraulic head at  $t = 37$  for the three different hydraulic scenarios and for different combinations of observation data. An explanation of the different observations employed for model calibration is given in Table 2.

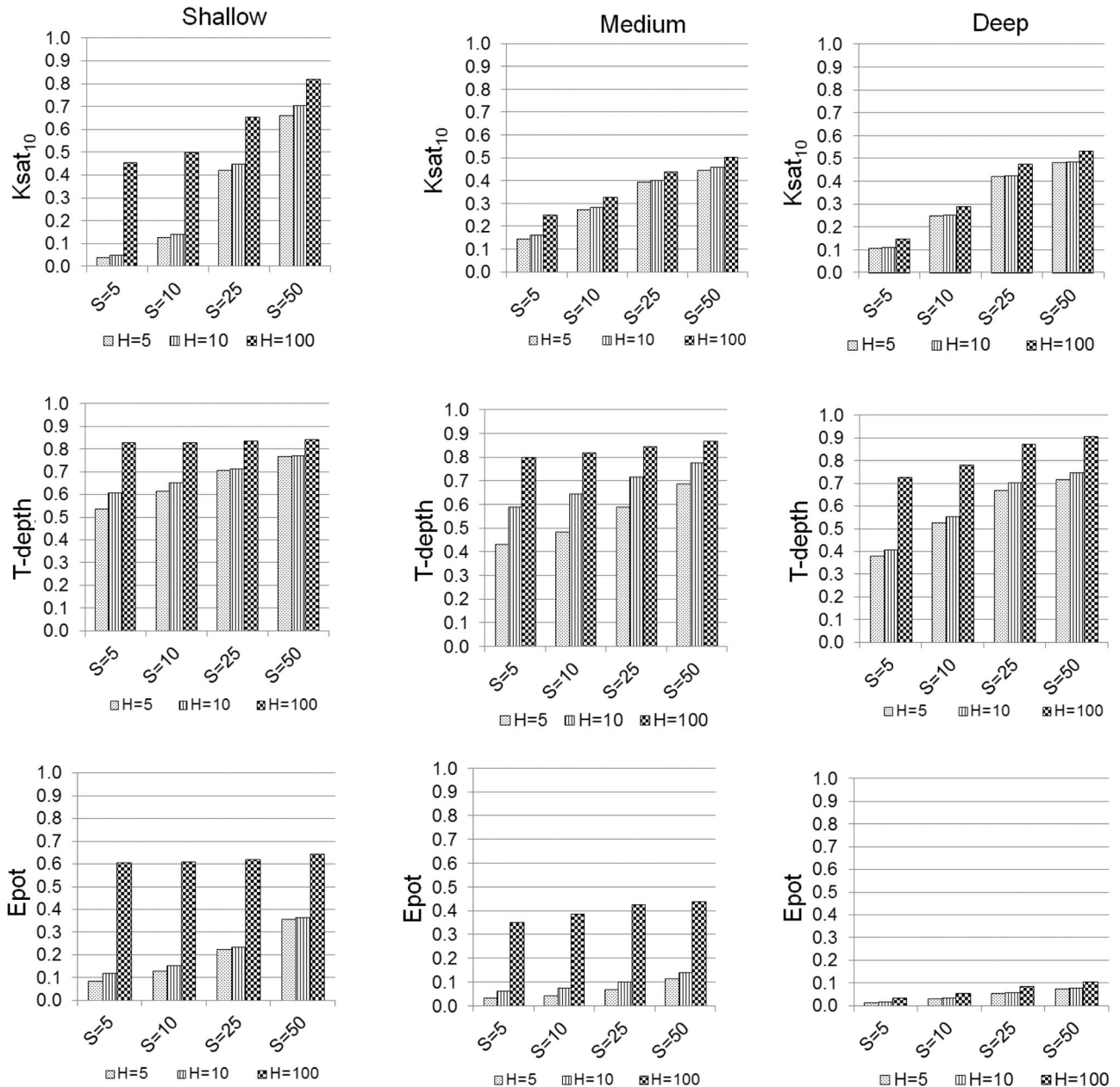


**Figure 6.** Identifiability of a range of calibration parameters for different combinations of observations as well as all simulated hydraulic conditions. The numbers associated with the parameters indicate the zone in the model (e.g., n10 corresponds to  $z = 4.5$  to  $5$  m, n9 to  $z = 4.0$  to  $4.5$  m).

the data set is expanded, this being an outcome of the fact that a small amount of “numerical granularity” is sometimes associated with calculation of identifiability. This follows from the fact that the dimensionality of each of the solution and null subspaces must be an integer. In computing the dimensionality of these two subspaces the method documented by *Doherty* [2010], and encapsulated in the PEST SUPCALC utility, was employed. This method computes the uncertainty of each eigencomponent of the weighted  $\mathbf{Z}$  matrix of equation (4). The calibration solution space is extended to

include this eigencomponent if its uncertainty is thereby reduced; however if its uncertainty of estimation is greater than its prior uncertainty, this eigencomponent is relegated to the null space.

[59] In Figure 7, the relationship between measurement accuracy and estimation uncertainty reduction is illustrated for selected parameters, different combinations of measurement accuracies and the three hydraulic conditions. Only observations of saturation at the top of the column and hydraulic heads are considered. Relative parameter uncertainty



**Figure 7.** Relative uncertainty reduction for selected parameters and assumed measurement accuracies for (left) shallow, (middle) medium, and (right) deep water table conditions. The numbers associated with the H and S observation types in this figures indicate the weights associated with these data types during uncertainty analysis based on equation (6); the standard deviation of measurement noise as featured as the square root of the diagonal elements of the  $C(\epsilon)$  term of equation (6) is the inverse of each measurement weight.

reduction is defined in the same way as relative prediction uncertainty reduction using equation (24). As stated above it is a number that, like identifiability, varies between zero and one. The closer it approaches one (implying reduction of parameter/predictive uncertainty to zero), the greater is the information content of the calibration data set with respect to that parameter or prediction.

## 6. Discussion

[60] Aspects of the methodologies that we applied in our study, and insights into data worth that are suggested by it are now discussed. However, because our analysis is based on a simple 1d approach, some of our findings may not be applicable in more complex spatial settings. Data worth details for these settings can be explored through usage of the methodologies described herein in conjunction with a simulator that is appropriate for that context. We believe, however, that our study does indeed yield a number of conclusions that are applicable across a range of spatial settings.

### 6.1. Methodology

[61] Our study provided an opportunity to deploy and demonstrate the Pareto method in the context of uncertainty assessment. To our knowledge, the only previous documented use of Pareto concepts for this purpose was that of Moore *et al.* [2010]. The present application achieves greater model-run efficiency than in that study by using a gradient-based methodology to traverse this front. This implementation of the method offers a greater degree of flexibility than the previous implementation by including the possibility of saving model runs through faster traversal of the Pareto front. The use of fewer points to define the Pareto front has little or no drawbacks in most uncertainty analysis contexts. This is because the relationship between particular values of a model prediction and minimized objective function values associated with these predictive values could be obtained through interpolation between a more limited set of points that are used to define the curve. In the present case we wished to obtain parameter and prediction values corresponding to a specific objective function (namely 1.0), and so implemented denser sampling of the curve than would be required under normal circumstances. However, even though the computational cost of the Pareto method as employed here is comparatively small compared to other nonlinear uncertainty analysis alternatives such as Markov chain Monte Carlo, the computational cost significantly exceeds that of the linear method.

[62] The attractiveness of the linear method lies not only in its speed. As pointed out earlier in this paper, a significant advantage of this method is that a user can assess the worth of a specific observation even before the value of the observation has been made available through measurement. The obvious disadvantage of using a linear method to assess predictive uncertainty is that environmental processes are generally not linear. In fact, processes that take place within the vadose zone can be characterized by an extremely high degree of nonlinearity. It follows that, in spite of their ease of use and high level of model run efficiency, conclusions on data worth drawn from linear analysis must be used with caution. However in defense of these methods, it must be recalled that when used in the context of data worth analy-

sis, neither the value of a prediction nor a quantitative evaluation of the uncertainty of a prediction is required. Instead, data worth analysis requires only an assessment of the *relative* uncertainties of the same prediction with, and without, the inclusion of different types and amounts of information in the calibration data set. The requirements for computational accuracy are therefore somewhat reduced.

[63] If a modeler is concerned that linear data worth analysis may lead to conclusions of doubtful veracity, it is important and recommended to repeat the comparison between the linear and nonlinear approaches. He/she could also repeat the analysis using a number of widely different realizations of parameter values for calculation of the sensitivities that underpin linear analysis. This was done by Dausman *et al.* [2010] for a saltwater intrusion model, this being another modeling context that is characterized by a high degree of nonlinearity. These authors found that, in their case at least, assessment of data worth was impaired to only a small degree by nonlinearity of their model.

### 6.2. Data Worth in Terms of Uncertainty Reduction of Hydraulic Heads

[64] A (perhaps unsurprising) conclusion to be drawn from the analyses presented herein is that in this particular setup the data type that is most informative of future groundwater levels is, in fact, past groundwater levels. Indeed, past head data are so informative of future head data that acquisition of other types of data in addition to historical heads does little to further reduce the uncertainty of future head predictions. We expect that even in significantly more complex settings, the information contained in hydraulic heads will exceed the information content of other observations when the purpose of a model is to predict future heads. It is of interest to note that the large reduction in future water level predictive uncertainty gained through acquisition of historical water level data is not achieved through definitive estimation of the value of any particular model parameter. It is highly unlikely that this finding will change for more complex spatial modeling settings, as an increase in the complexity of a conceptual model cannot increase the information content of a particular observation.

[65] The large reduction in water level predictive uncertainty accrued through use of head measurements in the calibration data set is achieved through estimation of values for *combinations* of parameters on which predictions of future water levels depend. The values of these combinations are not actually seen in the calibrated model, for all model parameters are still uncertain, notwithstanding the “calibrated state” of the model. Nevertheless, model parameter uncertainty is reduced—constrained by the fact that the values of certain model parameters are no longer statistically independent of the values of other model parameters. If these parameters vary, they must vary collectively in such a way that the values of these parameter combinations are maintained.

[66] This has repercussions for the type of model used for simulating recharge processes on a regional scale. A complex model based on Richards equation, calibrated against historical water level variations, may indeed be used in this role. However, if its use in conjunction with an appropriate saturated zone model allows the latter to

adequately reproduce past water level variations and predict future ones, this may not be because the many parameters required by such a model are particularly well known. Instead, simulation accuracy may follow from similarities which exist between the predictions required of the model and the data set against which it is calibrated. *Doherty and Welter* [2010] and *Doherty and Christensen* [2012] show that where this kind of similarity exists, the conceptual basis of the model is of far less importance than its ability to track the movement of the system state that it is meant to predict. The present study illustrates this same point. It demonstrates that if the complex recharge model were to be replaced by a much simpler model (but a model which is nevertheless complicated enough to replicate historical and future water level variations), calibration of that simpler model would endow it with the ability to make reasonably accurate predictions of the recharge on which future groundwater head predictions depend. However, the simpler model would presumably possess fewer (and probably lumped) parameters than the complex, Richards equation model that it replaces—possibly few enough parameters for its calibration to represent a well-posed inverse problem. It is likely that at least some of these lumped parameters would serve the same purpose as combinations of parameters employed by the complex Richards equation model that are uniquely estimable on the basis of the calibration data set.

[67] It is apparent from the work documented herein that ET data has high worth in the calibration of our simple 1-D model. Unfortunately, however, it is also apparent that the information that it contains largely reproduces that available through water level measurements. We speculate that this applies to other settings where variations in water table occur predominantly in response to infiltration or ET processes. Our results further show that the worth of ET data is heavily dependent on depth to the water table. This implies that its worth may be less in different spatial settings. Nevertheless, the availability of ET data over large areas from satellite observations may compensate for local variations in its worth. It may thus provide a useful supplement to borehole water-level measurements where the latter are sporadic in space and in time, and where the depth to groundwater is not large. However, where an assessment of its worth in a particular study area is required, it is suggested that the methodological approaches employed herein be employed. Such a study should take account of the accuracy with which ET measurements are made. Because remote sensing product integrates over the spatial scale of a pixel, the subpixel heterogeneity should be considered in this context. A discussion of subpixel heterogeneity in the context of modeling ET is given by *Li, H.T., et al.* [2008].

[68] Our study suggests that measurements of soil moisture, especially if restricted to the surface (as are measurements of soil moisture obtained through remote sensing), are of less use than evapotranspiration (and of course head) measurements. This is in accordance with previous studies. Furthermore, because ET is a process that occurs over a range of depths, it is not surprising that the information content of ET measurements is greater than that of surficial soil moisture measurements. Our study further suggests that (as for observations of ET) the usefulness of soil moisture data is strongly dependent on the depth to groundwater, this appearing to be a more important factor in

determination of their worth than the accuracy with which soil moisture measurements are made. The potential use of soil moisture data in predicting future hydraulic heads therefore appears to be limited, as is therefore their potential to compensate for a deficit of hydraulic head measurements. In contrast, as will be discussed in section 6.3, soil moisture measurements are the observations of greatest importance as far as increasing the identifiability of individual soil hydraulic parameters is concerned.

### 6.3. Parameter Identifiability and Data Worth

[69] While measurements of hydraulic head are effective in reducing the uncertainties of future head predictions, they do not significantly increase the identifiability of most parameters. Parameters which describe upper soil layer saturated hydraulic conductivity and porosity, total evaporation depth, and climatic forcing functions are informed to only a small degree by measured heads. On the other hand, the identifiability of parameters describing the retention curve are significantly increased by observations of soil moisture, but only if these parameters are in close spatial proximity to the actual observation; See, for example, Figure 6, where observations of soil moisture (St and Sth) inform only  $\beta$  in the top layer (beta10 in Figure 6), and not in layer 9. This suggests that observations of soil moisture are sensitive to soil heterogeneity, and that the information which they carry may be more local than regional.

[70] Observations of soil moisture do not significantly increase the identifiability of potential ET, nor the identifiability of the magnitude of a rain event. It appears that only soil moisture observations at the top of a soil column obtained through highly accurate, ground based methods under shallow water table conditions increases the identifiability of rain event 1 (RE1).

[71] Our analysis further suggests that the identifiability of parameters related to evaporation and transpiration processes is low in the absence of observations of transpiration, evaporation and evapotranspiration. Even though not explicitly demonstrated in this study, it is expected that without such observations, predictions of soil moisture content or ET-processes will therefore be highly uncertain.

[72] Finally, we address the relationship between measurement accuracy and the uncertainty of estimated parameters. The uncertainty reductions depicted in Figure 7 illustrate that the ability of increasing measurement accuracy to promulgate reductions in estimated parameter uncertainty is a function of the depth to groundwater. It also demonstrates the intuitive notion that the information potential of a certain measurement type is more fully realized when measurements of that type are made with a high degree of accuracy. It also demonstrates that upper limits for this potential exist. This is exemplified in several panels of Figure 7. If, for example, heads are measured with high accuracy ( $H = 100$ ), then increasing the measurement accuracy of soil moisture does not further reduce the uncertainty of estimated transpiration depth in shallow water conditions.

## 7. Conclusions

[73] In the numerical study documented in this paper, parameter identifiability and predictive uncertainty were quantified for a 1-D vadose zone soil system in which

movement of water is driven by infiltration, evaporation and transpiration. The worth of different types of observation data (when employed both individually as well as in combination) was calculated using both linear and nonlinear methodologies. In doing so, the numerical efficiency of the linear and Pareto methodologies was demonstrated. Our study differs from previous work in this area in that (1) the range of observation data types considered in our analysis was broader than in previous studies, and (2) our study attempts to quantify data worth through its effect on both parameter identifiability as well as on predictive uncertainty. The influences of changing hydrological conditions and of measurement accuracy were also considered.

[74] A comparison between linear and nonlinear approaches to data worth analysis undertaken in our study suggests that linear analysis can be useful despite the high degree of nonlinearity that is associated with many environmental processes. If, however, linear analysis is applied in more complex settings than those examined herein, its credibility should be assessed by comparing some of its outcomes with concomitant nonlinear analysis, in the same manner as that which was accomplished herein. Given the high cost of the acquisition of much environmental data, such an analysis is warranted both as a means of assessing the worth of existing data, and to provide a scientific basis for investment in acquisition of further data.

[75] Some of our findings confirm conclusions drawn from previous studies. For example, our analyses have demonstrated that for our particular case (and we suggest in many other contexts) the principal source of enabling information for making predictions of future groundwater head behavior is in fact historical groundwater head behavior. Another finding is that observations of evapotranspiration contain more information that is pertinent to the prediction of future water table fluctuations than observations of soil moisture. However, once different measurement accuracies and hydrological conditions are taken into account, the relationship between additional observations and predictive uncertainty becomes less obvious.

[76] Given the simplified setup of our model, the absolute values of parameter identifiability and the predictive uncertainty presented here are expected to change for more complex systems, especially in situations where changes of the water table are dominated by groundwater flow (and not by vadose zone processes). However, the extension of data worth analysis to other contexts should not be too difficult. Despite the simplified geometry of our model, several findings have emerged from our study that are relevant for data acquisition on all spatial scales. We have shown that a different depth to groundwater affects data worth of observations both in terms of uncertainty reduction as well as parameter identifiability, and differently so, for the different parameters. This implies that data worth is a function of both space and time. To the best of our knowledge this has so far not been demonstrated quantitatively. Furthermore, this conclusion suggests that future work on data worth should account for different hydrologic conditions.

[77] Our study has demonstrated that, with the exception of ET measurements for one particular groundwater depth scenario, no observation type, nor any combination of observations contain all of the information necessary to significantly increase the identifiability of climatic forcing

functions. This reinforces the need to accurately measure forcing functions such as precipitation and potential ET. Another finding that is particularly relevant to the calibration of regional-scale models which include simulation of vadose zone processes is that observations of hydraulic heads do not allow unique identification of any vadose zone parameter. Where spatial models of increased complexity are employed, the information deficit of head measurements with respect to individual vadose zone properties is likely to be even more apparent. It follows that if evaporation, transpiration, or the exact soil moisture content of a soil profile are of primary interest in a large-scale model, the low identifiability of parameters which govern these processes can lead to a high degree of uncertainty associated with predictive outcomes of these processes. Our study does indicate, however, that through acquisition of pertinent data (such as combinations of soil moisture, transpiration rates or ET rates), this uncertainty can be reduced.

[78] The above conclusion raises some interesting questions concerning the level of complexity required for simulation of recharge processes in regional groundwater models. Calibration of most models of environmental interest leads to ill-posed inverse problems, and therefore to nonunique parameter sets. Our study strongly suggests that the capacity of a recharge model to reduce the uncertainty of future groundwater level predictions does not rely on accurate estimation of all hydraulic properties affecting unsaturated zone water movement. Instead it relies on good estimation of only certain combinations of these properties. It follows that the recharge component of a regional water resource model can be “well calibrated” while each of its parameters may yet possess a high degree of uncertainty. The key feature of the calibration process in this case is that it endows the model with a good ability to make the predictions required of it, even though the constraints imposed on parameters that endow it with this ability pertain to combinations of parameters rather than to individual parameters. It follows that the model may lose little if it were simplified in ways suggested by the calibration process. Significant gains in model run efficiency could thereby be gained resulting in a more tractable calibration process.

[79] **Acknowledgments.** This work was funded by the Swiss National Foundation, Ambizione Grant PZ00P2\_126415 and the National Center for Groundwater Research and Training, Adelaide, Australia. We thank an anonymous reviewer and the AE for fair and well-articulated comments.

## References

- Bauer, P., T. Gumbrecht, and W. Kinzelbach (2006), A regional coupled surface water/groundwater model of the Okavango Delta, Botswana, *Water Resour. Res.*, *42*, W04403, doi:10.1029/2005WR004234.
- Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*(1–2), 18–36.
- Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*, 3663–3674.
- Brocca, L., F. Melone, T. Moramarco, W. Wagner, and S. Hasenauer (2010), ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy, *Remote Sensing Environ.*, *114*(11), 2745–2755.
- Brunner, P., H. J. H. Franssen, L. Kgotlhang, P. Bauer-Gottwein, and W. Kinzelbach (2007), How can remote sensing contribute in groundwater modeling?, *Hydrogeol. J.*, *15*(1), 5–18.

- Brunner, P., H. T. Li, W. Kinzelbach, W. P. Li, and X. G. Dong (2008), Extracting phreatic evaporation from remotely sensed maps of evapotranspiration, *Water Resour. Res.*, *44*, W08428, doi:10.1029/2007WR006063.
- Christensen, S., and J. Doherty (2008), Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration, *Adv. Water Resour.*, *31*(4), 674–700.
- Cooley, R. L., and S. Christensen (2006), Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media, *Adv. Water Resour.*, *29*(5), 639–656.
- Dausman, A. M., J. Doherty, C. D. Langevin, and M. C. Sukop (2010), Quantifying data worth toward reducing predictive uncertainty, *Ground Water*, *48*(5), 729–740.
- Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002), A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Trans. Evolut. Comput.*, *6*, 182–197.
- Doherty, J. (2010), *PEST: Model independent parameter estimation, User Manual*, Watermark Numer. Comput., Brisbane, Australia.
- Doherty, J., and S. Christensen (2012), Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, *Water Resour. Res.*, *47*, W12534, doi:10.1029/2011WR010763.
- Doherty, J., and R. J. Hunt (2009), Two statistics for evaluating parameter identifiability and error reduction, *J. Hydrol.*, *366*(1–4), 119–127.
- Doherty, J. and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, *46*, W05525, doi:10.1029/2009WR008377.
- Friedel, M. J. (2005), Coupled inverse modeling of vadose zone water, heat, and solute transport: Calibration constraints, parameter nonuniqueness, and predictive uncertainty, *J. Hydrol.*, *312*(1–4), 148–175.
- Gallagher, M., and J. Doherty (2007), Predictive error analysis for a water resource management model, *J. Hydrol.*, *334*(3–4), 513–533.
- Gowda, P. H., et al. (2008), ET mapping for agricultural water management: Present status and challenges, *Irrigation Sci.*, *26*(3), 223–237.
- Grant, J. P., et al. (2007), A field experiment on microwave forest radiometry: L-band signal behaviour for varying conditions of surface wetness, *Remote Sensing Environ.*, *109*(1), 10–19.
- Gupta, H. V., Sorooshian, S., and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–764.
- Hansen, N., and A. Ostermeier (2001), Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.*, *9*, 159–195.
- Hansen, N., S. D. Muller, and P. Koumoutsakos (2003), Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evol. Comput.*, *9*, 159–195.
- Hendricks Franssen, H. J., P. Brunner, P. Makobo, and W. Kinzelbach (2008), Equally likely inverse solutions of a groundwater flow problem including pattern information from remote sensing images, *Water Resour. Res.*, *44*, W01419, doi:10.1029/2007WR006097.
- Ines, A. V. M. I., and P. Droogers (2002), Inverse modelling in estimating soil hydraulic functions: A genetic algorithm approach, *Hydrol. Earth Syst. Sci.*, *6*(1), 49–66.
- Ines, A. V. M., and B. P. Mohanty (2008a), Parameter conditioning with a noisy Monte Carlo genetic algorithm for estimating effective soil hydraulic properties from space, *Water Resour. Res.*, *44*, W08441, doi:10.1029/2007WR006125.
- Ines, A. V. M., and B. P. Mohanty (2008b), Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties under different hydroclimatic conditions, *Vadose Zone J.*, *7*, 39–52, doi:10.2136/vzj2007.0048.
- Ines, A. V. M., and B. P. Mohanty (2009), Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithms: 2. Using airborne remote sensing during SGP97 and SMEX02, *Water Resour. Res.*, *45*, W01408, doi:10.1029/2008WR007022.
- James, S. C., J. E. Doherty, and A.-A. Eddebarh (2009), Practical postcalibration uncertainty analysis: Yucca Mountain, Nevada, *Ground Water*, *47*(6), 851–869.
- Jhorar, R. K., W. G. M. Bastiaanssen, R. A. Feddes, and J. C. Van Dam (2002), Inversely estimating soil hydraulic functions using evapotranspiration fluxes, *J. Hydrol.*, *258*(1–4), 198–213.
- Kerr, Y. H., et al. (2001), Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sensing*, *39*(8), 1729–1735.
- Koren, V., F. Moreda, and M. Smith (2008), Use of soil moisture observations to improve parameter consistency in watershed calibration, *Phys. Chem. Earth Parts A/B/C*, *33*(17–18), 1068–1080.
- Kristensen, K. J., and S. E. Jensen (1975), A model for estimating actual evapotranspiration from potential evapotranspiration, *Nord. Hydrol.*, *6*(3), 170–188.
- Li, H. T., W. Kinzelbach, P. Brunner, W. P. Li, and X. G. Dong (2008), Topography representation methods for improving evaporation simulation in groundwater modeling, *J. Hydrol.*, *356*(1–2), 199–208.
- Li, H. T., P. Brunner, W. Kinzelbach, W. P. Li, and X. G. Dong (2009), Calibration of a groundwater model using pattern information from remote sensing data, *J. Hydrol.*, *377*(1–2), 120–130.
- Li, Q., A. J. A. Unger, E. A. Sudicky, D. Kassenaar, E. J. Wexler, and S. Shikaze (2008), Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model, *J. Hydrol.*, *357*(3–4), 317–336.
- Madsen, H. (2000), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, *235*, 267–288.
- Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*, 205–216.
- Maneta, M. P., S. Schnabel, W. W. Wallender, S. Panday, and V. Jetten (2008), Calibration of an evapotranspiration model to simulate soil water dynamics in a semiarid rangeland, *Hydrol. Processes*, *22*(24), 4655–4669.
- Maxwell, R. M., and N. L. Miller (2005), Development of a coupled land surface and groundwater model, *J. Hydrometeorol.*, *6*(3), 233–247.
- Montzka, C., et al. (2011), Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, *J. Hydrol.*, *399*(3–4), 410–421.
- Moore, C., and J. Doherty (2005), Role of the calibration process in reducing model predictive error, *Water Resour. Res.*, *41*, W05020, doi:10.1029/2004WR003501.
- Moore, C., T. Wöhling, and J. Doherty (2010), Efficient regularization and uncertainty analysis using a global optimization methodology, *Water Resour. Res.*, *46*, W08527, doi:10.1029/2009WR008627.
- Romano, N., and A. Santini (1999), Determining soil hydraulic functions from evaporation experiments by a parameter estimation approach: Experimental verifications and numerical studies, *Water Resour. Res.*, *35*(11), 3343–3359.
- Saleh, K., et al. (2009), Soil moisture retrievals at L-band using a two-step inversion approach (COSMOS/NAFE'05 Experiment), *Remote Sensing Environ.*, *113*(6), 1304–1312.
- Schneider-Zapp, K., O. Ippisch, and K. Roth (2010), Numerical study of the evaporation process and parameter estimation analysis of an evaporation experiment, *Hydrol. Earth Syst. Sci.*, *14*(5), 765–781.
- Schwarzel, K., J. Simunek, H. Stoffregen, G. Wessolek, and M. T. Van Genuchten (2006), Estimation of the unsaturated hydraulic conductivity of peat soils: Laboratory versus field data, *Vadose Zone J.*, *5*(2), 628–640.
- Šimunek, J., M. Th. van Genuchten, and M. Šejna (2011), *The HYDRUS software package for simulating two- and three-dimensional movement of water, heat, and multiple solutes in variably-saturated media, technical manual, Version 2.0*, pp. 258, PC Progress, Prague, Czech Republic.
- Takada, M., Y. Mishima, and S. Natsume (2009), Estimation of surface soil properties in peatland using ALOS/PALSAR, *Landsc. Ecol. Eng.*, *5*(1), 45–58.
- Therrien, R., R. G. McLaren, E. A. Sudicky, and S. M. Panday (2006), *Hydrogeosphere, Groundwater Simul. Group*, Univ. of Waterloo, (Link to software: <http://hydrogeosphere.org/>)
- Timmermans, W. J., W. P. Kustas, M. C. Anderson, and A. N. French (2007), An intercomparison of the Surface Energy Balance Algorithm for Land (SEBAL) and the Two-Source Energy Balance (TSEB) modeling schemes, *Remote Sensing Environ.*, *108*(4), 369–384.
- Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, *45*, W00B10, doi:10.1029/2007WR006678.
- Tonkin, M., J. Doherty, and C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resour. Res.*, *43*, W07429, doi:10.1029/2006WR005348.
- Van Genuchten, M. T. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Am. J.*, *44*, 892–898.
- 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 Resour. Res.*, *23*(7), 1237–1250.
- Vrugt, J. A., and B. A. Robinson (2007), Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci.*, *104*, 708–711.
- Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization

- of hydrologic models, *Water Resour. Res.*, 39(8), 1214, doi:10.1029/2002WR001746.
- Weihermüller, L., J. A. Huisman, S. Lambot, M. Herbst, and H. Vereecken (2007), Mapping the spatial variation of soil water content at the field scale with different ground penetrating radar techniques, *J. Hydrol.*, 340(3–4), 205–216.
- Wind, G. P. (1986), Capillary conductivity data estimated by a simple method, in *Water in the Unsaturated Zone*, Vol. 1, edited by P. E. Rijtema and H. Wassink, pp. 181–191, Int. Assoc. of Sci. Hydrol., Gentbrugge, Belgium.
- Zekele, T. B., and B. C. Si (2005), Parameter estimation using the falling head infiltration model: Simulation and field experiment, *Water Resour. Res.*, 41, W02027, doi:10.1029/2004WR003407.
- Zhang, J., M. A. Ross, J. S. Geurink, and H. L. Gong (2010), Modelling vadose zone moisture dynamics in shallow water table settings with the Integrated Hydrologic Model: Field-scale testing and application, *Water Environ. J.*, 24(1), 9–20.
- Zreda, M., D. Desilets, T. P. A. Ferre, and R. L. Scott (2008), Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, *Geophys. Res. Lett.*, 35, L21402, doi:10.1029/2008GL035655.