Groundwater

Extending Data Worth Analyses to Select Multiple Observations Targeting Multiple Forecasts

by Troels N. Vilhelmsen[1,2] and Ty P.A. Ferré[2]

Abstract

Hydrological models are often set up to provide specific forecasts of interest. Owing to the inherent uncertainty in data used to derive model structure and used to constrain parameter variations, the model forecasts will be uncertain. Additional data collection is often performed to minimize this forecast uncertainty. Given our common financial restrictions, it is critical that we identify data with maximal information content with respect to forecast of interest. In practice, this often devolves to qualitative decisions based on expert opinion. However, there is no assurance that this will lead to optimal design, especially for complex hydrogeological problems. Specifically, these complexities include considerations of multiple forecasts, shared information among potential observations, information content of existing data, and the assumptions and simplifications underlying model construction. In the present study, we extend previous data worth analyses to include: simultaneous selection of multiple new measurements and consideration of multiple forecasts of interest. We show how the suggested approach can be used to optimize data collection. This can be used in a manner that suggests specific measurement sets or that produces probability maps indicating areas likely to be informative for specific forecasts. Moreover, we provide examples documenting that sequential measurement election approaches often lead to suboptimal designs and that estimates of data covariance should be included when selecting future measurement sets.

Introduction

In hydrological science, we are moving toward a basic framework when developing groundwater flow models for water resource management. Stripped down, this framework includes four steps. First, existing data are retrieved to build an initial version of the groundwater flow model, once analyzed, the second step is to collected additional information about the system. This often includes the collection of data, which can provide structural information on the subsurface (e.g., boreholes or geophysical data). This step also includes collecting information on state variables, which can be used to constrain model parameters in an inversion scheme. Such data could include hydraulic heads, streamflow discharges, chemical concentrations, specific types of geophysical data, etc. The third step is refining the numerical flow model based on the new evidence. Subsequently, the model responses are fitted to hydrological data by adjusting model parameters, and the estimated parameters are evaluated for their relation to the physical properties they represent (e.g., Anderson et al. 2015; Doherty 2015; Hill and Tiedeman 2007). Ideally (but often overlooked!), the calibrated model is then subject to detailed uncertainty analyses, which result in a model prognosis presented with quantitative uncertainties (e.g., Tonkin and Doherty 2009). The third and final step is applying the results from the model in a decision/management framework. Here the model prognosis is used together with other supplementary information either to gain insight or to inform regulatory or treatment actions. The challenge, especially given the relatively limited budgets associated with most hydrologic investigations, is to determine the optimal types, timing, and locations for acquiring new data. In practice, this is most often done based on the expert knowledge of the practitioner, meaning that it is qualitative and subjective. It is generally not reasonable to expect that intuition alone can lead to optimal data collection, especially given that the hydrologist must consider the information content of existing data, often comprised of many data types, and when the goal of the modeling task is to optimize for multiple forecast scenarios.

[1]Corresponding author: Department of Geoscience, Aarhus University, Aarhus, Denmark; troels.norvin@geo.au.dk

[2]Department of Hydrology and Atmospheric Science, University of Arizona, Tucson, AZ 85721.

Article impact statement: Using models to select multiple future measurements to collect is difficult, not impossible, but worth the effort.

Received February 2017, accepted August 2017.

doi: 10.1111/gwat.12595

When considering more complex questions, such as multi-objective decision support, the task of selecting new data may require approaches that are collectively referred to as optimal design (OD). In relation to hydrological studies, OD has been investigated for over 20 years. For complete references to early OD work, we refer to the review studies by Dixon and Chiswell (1996) and/or Loaiciga et al. (1992). We also need to distinguish between different goals of the OD analysis. In relation to contaminant hydrology, several studies advocate that OD should be related to economic benefits (Freeze et al. 1992; James and Gorelick 1994; Criminisi et al. 1997; Feyen and Gorelick 2005; Norberg and Rosen 2006). This type of analysis requires a specific definition of "the price of being wrong," which may not always be easy to quantify (e.g., how do you quantify the price of a spring being depleted due to groundwater abstraction?). Alternatively, OD can be related to the optimal distribution of a fixed number of measurements based on their ability to reduce the uncertainty of a specific forecast (e.g., Kikuchi et al. 2015; Wöhling et al. 2016). This sort of OD design typically requires that the number of observations to include into the analysis is predefined. Including the number of observations as a variable would require a trade-off function to the overall value of the measurement, for example, the price or effort required to make the measurement. Examples of such trade-offs in relation to groundwater remediation can be seen from the classic reference of James and Gorelick (1994). In the present analysis, we have not adopted this price trade-off. Focus is therefore instead to select the optimal spatial distribution of a fixed number of measurements to obtain the maximum decrease in the predictive uncertainty.

OD studies can also be subdivided into two main categories: those applying nonlinear Monte-Carlo (MC) based methods to estimate predictive uncertainties (e.g., Nowak 2010; Leube et al. 2012; Kikuchi et al. 2015), and those applying linear approximations (Dausman et al. 2010; Fienen et al. 2010; Engelhardt et al. 2013; Hill et al. 2013; Wallis et al. 2014; Wöhling et al. 2016). Nonlinear MC methods may be necessary for problems including processes or parameter interactions that lead to highly nonlinear responses. The linear methods rely purely on forecast and parameter sensitivities. That is that the relation between model responses and model parameters can be approximated as linear. The dependency on forward model runtime is therefore limited, because this linearization often only has to be performed once. From a practical standpoint such a linearization can be performed using inversion software such as PEST (Doherty 2016) or UCODE (Poeter et al. 2014). OD relates closely to parameter identifiability (Doherty and Hunt 2009), namely that we seek to identify model parameters with high uncertainty, that have high sensitivity to model forecasts. Once identified, we can seek to find the new data leading to maximum reduction in uncertainty of these parameters. Formally, this can be derived in a Bayesian context as done by Christensen and Doherty (2008). In relation to OD, this estimate was initially applied to select the optimal location of one additional observation well to constrain the forecast of stream flow reduction due to groundwater abstraction (Fienen et al. 2010). Dausman et al. (2010) used the method to estimate important observation locations of temperature and salinity concentrations to constrain uncertainty after a change in stress on the system. Hill et al. (2013) used the observation-prediction (OPR) statistic (Tonkin et al. 2007) to determine the optimal locations for an additional head measurement in the Death Valley regional groundwater flow model, and found multiple places for optimal placement. Common among these three studies, they evaluated the contributions to the reduction in predictive uncertainty due to the addition of a single new observation. Wallis et al. (2014) extended this approach to selecting multiple observations in an OD of a tracer test experiment. However, they did not optimize for configurations and were therefore limited to testing a narrow predefined number of observation sets. Wöhling et al. (2016) extended the OD analysis to optimize for multiple new observations of two different types (hydraulic conductivity and heads) to decrease the predictive uncertainty of mean travel times for hyporheic exchange fluxes. Wöhling et al. (2016) optimized the observation locations using a genetic algorithm (GA) they developed based on the work of Goldberg (1989). They could thereby make an effective search to find the optimal combination given a predefined number of measurement locations. As a method to calculate the DW, their study was built on the existing PREDUNC program build into the PEST suite of utilities (Doherty 2010; Doherty 2013). Similar to previous studies, their approach was also limited to optimizing for a single forecast. In practice, multiple targets are often of interest in OD, especially when multiple stakeholders have competing interests in a basin. Linear as well as nonlinear uncertainty analysis is also available through the pyemu framework (White et al. 2016). The study also included a section about OD analysis. Here they used a selection approach, where proposed observations were added sequentially. They also assumed that new locations were uncorrelated, why the optimal combination of observation locations were grouped within a small area, which is likely not going to be the optimal solution. Methods to perform linear analysis are also available directly in model analysis/inversion software PEST++ (Welter et al. 2015). Such availability will hopefully lead to a more general application in the future.

Based on the existing studies, we find that a number of topics remain to be investigated. The first is linked to the data redundancy often present in observation data sets. This redundancy originates from observations having the same or similar information content with respect to either the parameters or the forecast of interest. With respect to groundwater hydrology, the obvious example would be two proximate hydraulic head measurements in the same aquifer. Such observations would not have an unique information content, and one of them could likely be excluded from a hydrological analysis without deteriorating model performance. Formally, this can be estimated by knowing or approximating the correlations in the observation data set. Due to lack of information about the covariance structure of the new and existing observation data set, this is for better or worse most often ignored. However, because this is an inherent part of the analysis, the potential bias introduced, using an assumed known covariance structure could degrade the quality of monitoring designs. Given that data collection, especially in relation to large distributed groundwater models, often has multiple objectives, we also believe that the OD methodology should be extended to include multiple forecasts. Where multiple forecast could be different model outcomes of interest, for example, reduction of stream base flow due to groundwater abstraction, risk of groundwater pollution, or resulting aquifer drawdown.

Methods

In the following, we provide details of the underlying assumptions on which the proposed methodology lies. We then derive the equations needed to estimate the linear approximation of predictive uncertainty, relate this equation to the data worth (DW) obtained by collecting additional data, and use the DW concept to estimate a new measure of assessing the value of information, the retrieval of information index (RII). The RII is used to evaluate bias introduced into the analysis by performing the analysis using an erroneous covariance function. Using the DW concept, the analysis is extended to optimize for multiple forecast scenarios, and it is shown how this concept can be used to derive probability maps that identify locations that are likely to be important for multiple forecasts of interest. The entire analysis is illustrated using a synthetic case study.

Uncertainty Analysis Using Linear Estimates

Assuming model linearity, the predictive uncertainty variance of a given forecast (s) can be determined as

$$\sigma_s^2 = \boldsymbol{y}^T \boldsymbol{C}_{pp} \boldsymbol{y} - \boldsymbol{y}^T \boldsymbol{C}_{pp} \boldsymbol{X}^T \left(\boldsymbol{X} \boldsymbol{C}_{pp} \boldsymbol{X}^T + \boldsymbol{C}_{\varepsilon\varepsilon} \right)^{-1} \boldsymbol{X} \boldsymbol{C}_{pp} \boldsymbol{y}. \quad (1)$$

Here $\boldsymbol{X}$ is the actions of a mathematical model used to simulate a set of observation data. Similarly, $\boldsymbol{y}$ is the action of the model to produce the forecast. $\boldsymbol{C}_{pp}$ and $\boldsymbol{C}_{\varepsilon\varepsilon}$ are the prior parameter and the data covariance matrix, respectively. These matrices hold information encapsulating out imperfect knowledge of the system we are analyzing. Both matrices are square and diagonal. This means that the rows and the columns with the same indices represent the same parameter or observation. The diagonal elements represent the individual parameter or observation variances. The larger the diagonal elements are the larger is the uncertainty. Elements on the off diagonal represent the interdependency between observations or parameters. In a model calibration context where a specific off diagonal element in $\boldsymbol{C}_{pp}$ approximates the product of the standard deviation of their pertaining parameters, it means, that the parameters cannot be estimated individually. For the diagonal elements in $\boldsymbol{C}_{pp}$, this means that the specific parameter can only be estimated with great uncertainty. With respect to $\boldsymbol{C}_{\varepsilon\varepsilon}$ high values in the diagonal of this matrix would mean that the specific observations are highly uncertain, and high off diagonal elements would mean that the observations are potentially redundant, thus providing similar information about the system analyzed. Thus, if $\boldsymbol{C}_{\varepsilon\varepsilon}$ is diagonal, the underlying assumption is that all observations contain unique information. This is an obvious flawed assumption for proximate observations.

Equation 1 builds on assumptions of model linearity. Most hydrological models are, however nonlinear, and the elements of $\boldsymbol{X}$ will thereby be dependent on the parameter values ($\boldsymbol{p}$) imposed on the model. $\boldsymbol{X}$ can be approximated by the Jacobian/sensitivity matrix. In the following, $\boldsymbol{X}$ will therefore be used as symbol for the Jacobian matrix and $\boldsymbol{y}$ will be used to describe the forecast sensitivities to changes in the model parameters. Additional elaboration on how (Equation 1) can be derived can be found in Appendix S1, Supporting Information, and a detailed and insightful discussion of the applicability of Equation 1 and the individual terms it contains can be found in Anderson et al. (2015, 461–465).

The Concept of DW

Because Equation 1 only holds parameter sensitivities to modeled targets (observations and forecasts) and not absolute data values, it can be used to evaluate the value of existing or new data sets based on their ability to reduce the uncertainty of a given forecast of interest. This type of analysis has been documented previously in the literature (Dausman et al. 2010; Fienen et al. 2010; Brunner et al. 2012; Wallis et al. 2014; Wöhling et al. 2016). Equation 1 can thereby be used to evaluate the value of existing and yet to be collected data sets. For this purpose, we have adopted the DW term defined as

$$\mathrm{DW} = \frac{\sigma_{\mathrm{dec}}^2}{\sigma_{\mathrm{base}}^2}, \quad (2)$$

where DW ranges between 0 and 1, hereby expressing the relative value of individual or combinations of data points. The term σ_{dec}^2 is the decrease in predictive uncertainty obtained by adding new data to the set of observations, and σ_{base}^2 is the base uncertainty corresponding to existing data only. Both terms are determined using Equation 1, but with different elements.

DW estimates can often be obtained at a low computational burden because the Jacobian matrix containing forecast sensitivities, existing data, and new data often only needs to be calculated once. But even for cases needing multiple evaluations of the Jacobian, the computational burden is limited compared to nonlinear methods. The complete Jacobian matrix can be exemplified as

$$X_{\text{all}} = \begin{bmatrix} (y_1)^T \\ \vdots \\ (y_n)^T \\ X_{\text{base}} \\ X_{\text{nd}} \end{bmatrix}, \tag{3}$$

where X_{all} is the complete Jacobian matrix containing forecast sensitivities to one or more forecasts of interest $[(y_1)^T$ to $(y_n)^T]$. The term X_{base} is the base Jacobian matrix containing the existing data set used for model calibration, and X_{nd} is the Jacobian matrix for the yet to be collected new data containing one line for each potential new observation (in total equal to nd). In the present analysis, X_{all} was estimated using PEST. By extracting a given forecast sensitivity vector (y_i), corresponding to forecast i, combined with X_{base}, σ^2_{base} can be calculated using Equation 1 (Dausman et al. 2010). To estimate the DW of a single new observation, X_{base} needs to be extended with the row from X_{nd} representing the new data point and σ^2_{dec} can be calculated. $C_{\varepsilon\varepsilon}$ must be adjusted accordingly. This process is repeated one at a time for each potential new observation to assess the DW of the all-potential measurements.

DW Concept: Single Forecast, Multiple New Observations

Compared to previously published work (Dausman et al. 2010; Fienen et al. 2010; Brunner et al. 2012; White et al. 2016), the purpose of the present analysis is to find optimal combinations of new, yet to be collected, observations given a single or multiple forecasts of interest (the latter also distinguishing it from the study by Kikuchi et al. (2015) and Wöhling et al. (2016)). We do this usng the methodology outlined in Figure 1, namely by adding combinations of potential observations sequentially, and subsequently evaluating their combined DW. Instead of taking each line of $\mathbf{X}_{\text{nd}}$ individually, we extract combinations of observations, by selecting their corresponding lines from $\mathbf{X}_{\text{nd}}$ (Figure 1, box 3). Because each line in $\mathbf{X}_{\text{nd}}$ holds the sensitivity of a given new observation to the parameters, selecting multiple lines will facilitate the DW of combinations of new data. The number of lines to extract is equal to the number of observations to be collected. Unfortunately, exploring the complete set of possible combinations is rarely possible because these expand to a maximum of $\frac{\text{nnewmax}!}{(\text{nnewmax}-\text{nnew})!}$ combinations, where nnewmax is the number of potential observations locations, and nnew is the maximum number of new observations to collect. To make a subsample of the complete set of possible solutions, we introduce what we call the random selection matrix ($\mathbf{M}_{\text{rs}}$), also referred to as the pool (see Figure 1). $\mathbf{M}_{\text{rs}}$ holds integer values each corresponding to a specific row of $\mathbf{X}_{\text{nd}}$. The number of rows in $\mathbf{M}_{\text{rs}}$ is equal to the combinations of measurements to analyze and the number of columns is equal to the maximum number of new observations to be collected (nnew). The purpose of $\mathbf{M}_{\text{rs}}$ is twofold. First, we needed a representative set of measurement combinations that down samples the complete set of potential combinations. $\mathbf{M}_{\text{rs}}$ is therefore generated at random, by combining indices of potential observation locations (see Figure 1, box 3), linked to rows in $\mathbf{X}_{\text{nd}}$. Here, the only restriction is that each row of $\mathbf{M}_{\text{rs}}$ only holds unique values. Second, by repeated use of $\mathbf{M}_{\text{rs}}$ for estimating the DW of multiple forecasts ($\mathbf{y}_1$ to $\mathbf{y}_n$), the value of the same combinations of observations can be evaluated for different forecasts. This is an efficient approach to addressing multi-objective problems which will be described in the following.

DW Concept: Multiple Forecasts, Multiple New Observations

The DW of a single or combinations of new observations will change depending on the forecast of interest because each forecast will depend upon the uncertainty of different aspects of a model, thereby showing different sensitivities to each combination of proposed observations. It is often the case that multiple targets are of interest when conducting a hydrogeological investigation. This is especially the case if multiple stakeholders have competing interests as to the water resources. For water resource evaluations, forecasts of interest related to water abstraction may include effects on stream discharge, aquifer drawdown, or contaminant migration. For each new set of measurements, the DW for all forecasts of interest can be calculated (Figure 1, box 2). One way to combine multiple objectives is through a priority-weighting (e.g., based on estimated cost or regulatory priority). This can be achieved through the calculation of a value index (VI) using the following equation:

$$\text{VI}_j = \sum_{i=1}^{n} w_i \text{DW}_{i,j}, \tag{4}$$

where VI_j is the value index of observation set j to the n analyzed forecasts, w_i is the weight given to forecast i, and $\text{DW}_{i,j}$ is the DW of observation set j to forecast i. The weighting is introduced to prioritize different forecasts. Weighting could be based on economic valuation of a set of predictions for different impacts and to different stakeholders. Or the weighting could be based on qualitative prioritization for a single stakeholder when designing a management plan. Assuming that these weights can be defined quantitatively, the combination of proposed new observations that provides the maximum VI would be deemed optimal. Alternatively, a continuum approach can be used, which examines the dependence of the choice of optimal measurements on the weights given to different objectives. That is, a manager or decision maker can choose measurement sets that provide an acceptable trade-off among several prediction targets based on results that better communicate the risks and benefits of different proposed observation plans. Using the

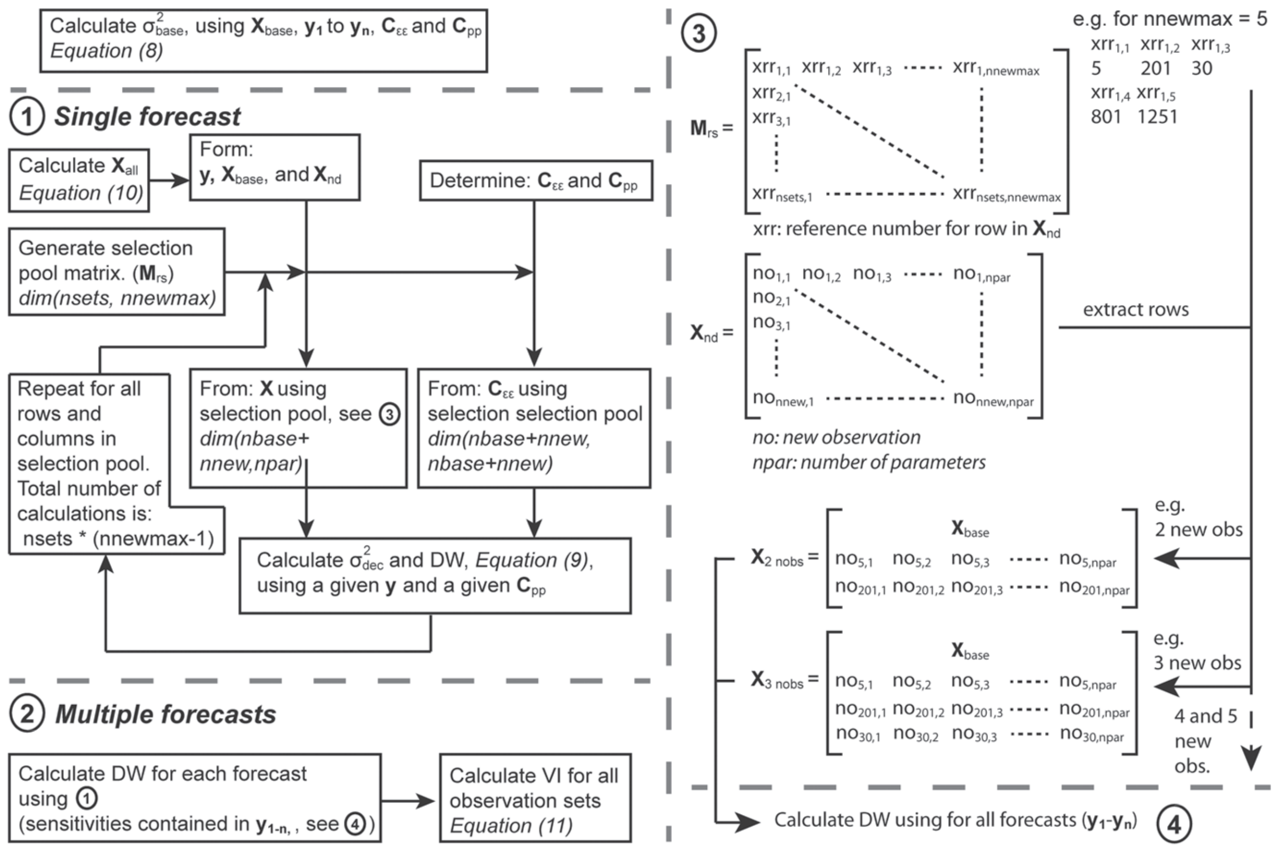

Figure 1. Selection framework for single or multiple forecasts. Dim refers to the dimension of the matrix. nsets is the number of observation combinations to test for. nnewmax is the maximum number of potential new observation, nbase is the number of base/existing observations. nnew is the number of observations in the present step of the analysis.

selection matrix concept allows us to apply OD efficiently enough that it can be used for these examinations of the interactions between prediction priorities and observation network design.

Estimating the Validity of New Sampling Locations

Because the VI in the present study is calculated for all combinations of measurements indexed in $\boldsymbol{M}_{\mathrm{rs}}$, it can be used to identify the optimal set of measurements for any defined number of observations. It can also be used to look across measurement set sizes to identify the expected benefit of each additional observation. This can be repeated for different predictions of interest. Then, the results can be used to find combinations of measurements that are optimal as a function of the importance weighting of the prediction targets. Alternatively, it can be used to identify observations that are generally likely to be important to the forecast or forecasts of interest because they tend to be selected into many optimal observation sets. This is another common way to consider the selection of new observations. Essentially, asking the question: “If I were to add an observation at this location, how likely is it that it will be useful to support model-informed decisions?” This approach is less discrete and more qualitative than selecting specific observation sets; but, in many cases it may be more informative for hydrogeologists and stakeholders alike. Some arbitrary choices are necessary to implement this approach. In the present case, for example, we chose to identify the combinations of observations that were within 2.5% of the most informative observation sets for each forecast of interest and observation set size. Then, we determined what fraction of these observation sets included any given observation as a measure of the probability that the observation would be informative. This result can be plotted as a probability contour map to give a more general indication of the distribution of DW in a multiple-observation, multiple-decision context. In addition, such a map will show how accurate the field campaign needs to be with respect to measurement locations, which can have important practical implications when planning real field campaigns. The map can also be used to identify alternative locations that are likely to inform the forecasts acceptably well if the highest value locations must be discounted based on consideration of access (permission or cost) and other practical aspects of field data collection.

DW Concept: Using a Global Search Genetic Algorithm

Despite the benefits of being able to estimate probability maps, the limitation of the proposed methodology is the relatively ineffective random search procedure applied. Other studies have applied GAs in OD studies (Zhang et al. 2005; Park et al. 2006; Chadalavada and Datta 2008; Wöhling et al. 2016). A more general overview of GA used in water resource studies can be obtained from Nicklow et al. (2010). GA is an optimization approach, which means that the outcome of the analysis will be an OD, which in the present case will be the combination of measurements that maximize the weighted VI. The random search approach presented previously has therefore been compared to the results of a global search GA similar to that applied by Wöhling et al. (2016), which builds on the methods documented by Goldberg (1989). The GA was implemented in Python for the present study. To make the results comparable to the random search procedure, we optimized for the VI instead of minimizing the predictive uncertainty as done by Wöhling et al. (2016). In the GA, we used the same standard GA selection scheme as Wöhling et al. (2016), namely selection, mutation, and crossover. In the present study, we retained 40% of the population, we used a mutation probability of 5% that we allowed to increase, in the case we obtained a too uniform population, and we allowed a 15% chance of selecting outside elite candidates. This should most likely be optimized and adapted to future applications.

Estimating the RII

One of the challenges of the DW concept as outlined in this study is the estimation of $\boldsymbol{C}_{pp}$ and $\boldsymbol{C}_{\varepsilon\varepsilon}$. A proxy for the $\boldsymbol{C}_{pp}$ can be estimated by combining numerical analysis of the model and geological understanding of the area investigated. However, for most practical cases of groundwater model analysis, $\boldsymbol{C}_{\varepsilon\varepsilon}$ is assumed diagonal, effectively ignoring potential observation correlations. As mentioned in the introduction, this can be problematic for DW analyses. This problem is avoided in sequential OD approaches, wherein one data point is collected after each analysis and the sensitivities are recalculated before collecting more data. However, this is impractical for real hydrologic investigations. To analyze this potential problem, we have decided to add two additional indexes. The first is used to estimate the bias that we introduced in the measurement selection when applying an incorrect covariance matrix in Equation 1. This is called the RII. The RII is a normalized factor used to describe the bias introduced into the analysis by performing the analysis using an erroneous covariance function, with 0 indicating that the bias overwhelms the value of the added data and 1 indicating no influence of this bias. The RII is thus calculated by selecting the optimal set of measurements based on an incorrect covariance matrix. The VI of these points is subsequently calculated using the true covariance matrix. By normalizing the biased estimate of VI with the optimal VI obtained by selecting the same number of new observations using the true covariance function, an estimate is obtained, which ranges between 0 and 1. A RII of 0 means that the selected points pose a zero VI when evaluated using the true covariance function, and a RII of 1 will mean that the points selected pose the same VI as the points selected using the true covariance function. Alternatively, as shown in the results section, we can use this index to determine if increasing the number of new data points collected can compensate for an incorrect covariance structure; we call this second analysis RII2. The RII2 index is calculated using the largest sample with the true covariance matrix for reference. For real cases, where the covariance structure will be unknown, the RII and the RII2 can be calculated using the best estimate of the covariance structure as reference. This allows for calculating the sensitivity of the optimization problem to the choice of covariance matrix.

Synthetic Test Case

To test the performance of the proposed method, we have set up a synthetic test model such that we have control and perfect knowledge of the covariance of the input structure as well as the covariance of the proposed new observations. The model is setup in MODFLOW-2005 (Harbaugh 2005) and is presented in Figure 2. The model is two-dimensional, with a domain size of 1000×1000 m, subdivided into 100 rows and 100 columns. There is one layer with a thickness of 100 m. The model has a constant head BC in the lower left corner of the domain with a value of 100 m, and a head dependent BC in the upper right corner, simulated using the river package. This BC has a head value of 90 m, a bottom elevation of 85 m and a riverbed conductance of 1×10^{-3} m^2/s. A uniform recharge is distributed over the area with a value of 1.1574×10^{-8} m/s. The true hydraulic conductivity field was generated by interpolating from a set of 81 pilot points (e.g., Certes and de Marsily 1991; Lavenue and de Marsily 2001; Doherty 2003) distributed on a regular 9×9 grid using an exponential variogram with a sill variance of 0.7544 and a range of 250. The hydraulic conductivity at each pilot point was simulated based on the same variogram structure using a $\log(k)$ mean value of -4. The data set used as base data (forming $\boldsymbol{X}_{\text{base}}$) was generated from this true model structure by simulating the steady state heads at the 10 locations marked with blue dots in Figure 2. This data set also includes the streamflow discharge, simulated in the river. The head values were perturbed with a standard deviation of 1, and the stream discharge was perturbed with a standard deviation of 20% of the true streamflow. These data are then used in the subsequent analysis.

The model was calibrated using PEST by applying singular value decomposition (SVD) as a regularization methodology. The SVD was set up such that the truncation threshold between the largest and the smallest eigenvalue to include in the analysis was 5×10^{-7}, for full description of this, see Doherty and Hunt (2010). Assuming perfect model linearity, this calibration step is not necessary; however, we choose to include it due to the slight nonlinearity in our model setup, caused by the head dependent river BC. The initial value for each pilot point was assumed to be 1×10^{-4}, equal to the mean of the true field. We estimate the uniform recharge as a multiplication factor on the true recharge distribution. The initial multiplication factor was set to one. The initial recharge estimate going into the inversion was thereby equal to the true value. The potential new measurements

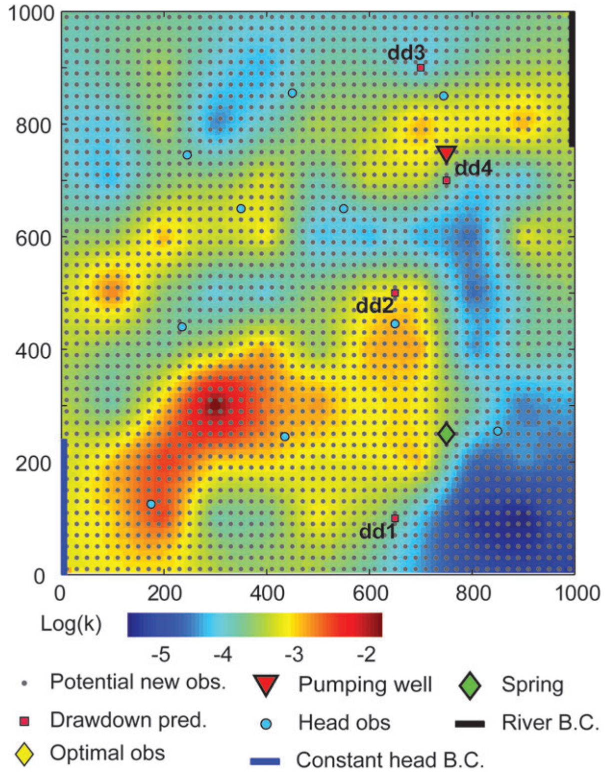

Figure 2. (A) Model setup used in the study. The underlying *k*-field is the true distribution of hydraulic conductivities used to generate calibration data set.

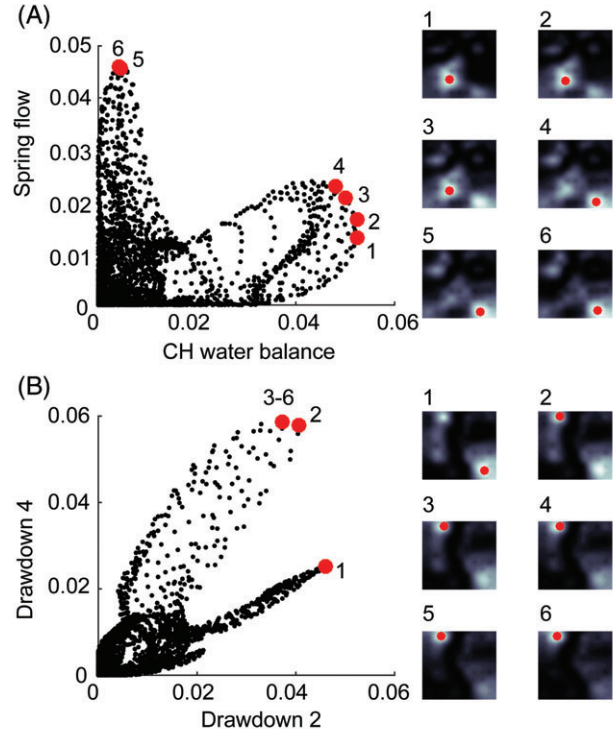

Figure 3. The figure show the DW of a single new observation, optimized for two new forecasts. Each point in the cloud in the left plot represents the DW of a new observation to two new forecasts. The right-hand side plots show how the optimal sampling location moves through space when changing the weight in the VI according to Equation 4.

comprise estimates of hydraulic conductivity at the 2500 potential sampling sites shown in Figure 2. We chose hydraulic conductivity estimates as our data because the true covariance function of these is known and equal to that used to generate the true model structure.

In total, six forecast scenarios are included in the analysis. All of these are related to the establishment of a pumping well (located at the red triangle in Figure 2). The forecasts comprised six targets: four drawdown locations; the water balance on the constant head BC; and the water balance on the spring located at the green triangle in Figure 2. This well was included in the model after the calibration step. Subsequently, the parameter sensitivities to existing and potential new data as well as the forecast sensitivities were calculated. The sensitivities to the new data were estimated by calculating how much the hydraulic conductivity changed at a specific location when the value of the (pilot point) parameter was altered.

The model presented in Figure 2 was used for the majority of the analyses presented. However, for the analysis presenting the sensitivity to the bias in the covariance structure (the RII and the RII2), we generated an ensemble of 10 reference model realizations, and repeated the analysis on each of these structures. Each of these realizations had a different "true" structure determined using the same methodology outlined above, and using the same variogram function. This will be further elaborated in the result section.

Results

The results section below contains only results related to optimizing for multiple forecasts. We have also performed the analysis for single forecasts, and single new observations. Readers interested in this analysis are referred to the online supplementary material.

Multiple Forecasts, Single New Observation

To approach the challenge of optimizing multiple new observations to multiple new forecasts, we initiate with the case where we optimize the collection of a single new observation to multiple forecasts. The results of this analysis are shown in Figure 3. The DW for each of the 2500 potential new measurements is shown as a cloud plot. Each point in the cloud represents a single new observation, and the axes show the DW for each of the two target forecasts. The right hand side of each plot shows the contoured VI for the two forecasts, given a specific weight distribution in Equation 4. The contour plots are marked with 1 to 6. Subplot 1 shows the case where all of the weight is given to the forecasts on the *x* axis, and subplot 6 shows the case where the entire weight is given to the forecast on the *y* axis. From subplots 2 to 5 the weight is shifted by 20% going from 80% weight to observation on the *x* axis to 20% weight to that observation. The optimal new observation in each subplot is marked by a red circle, and their location in the cloud can be seen from left hand plots. By tracking the locations in the cloud as

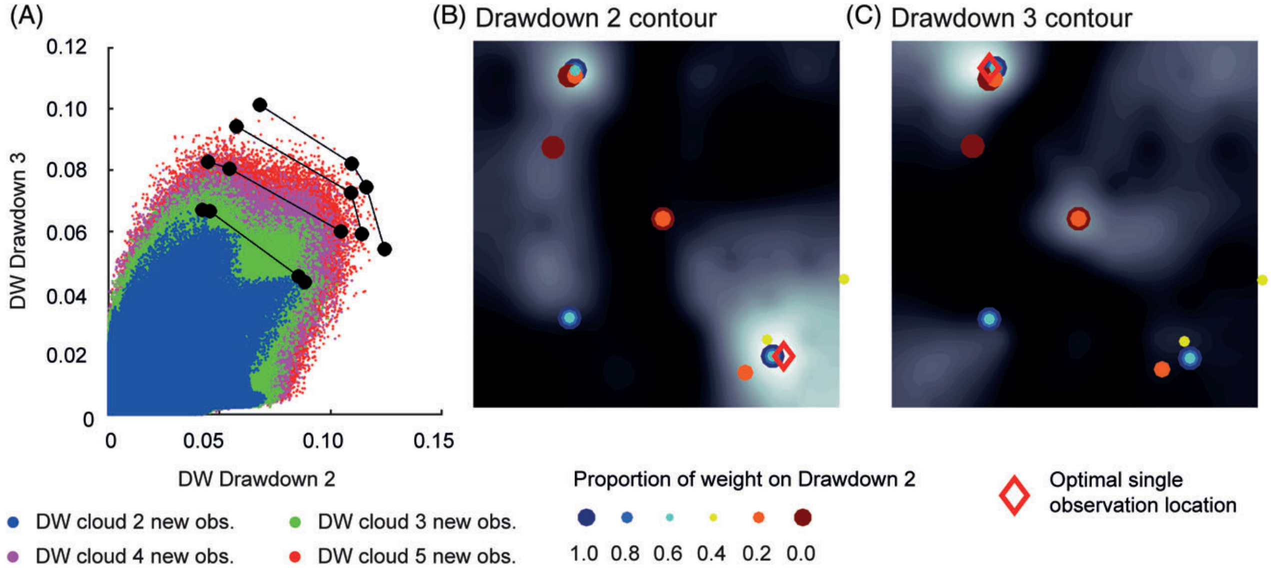

Figure 4. (A) The DW cloud for two forecasts given multiple sampling locations. (B) The optimal sampling locations for three new observations by changing the forecast weight in Equation 4, combined with the DW contour for the drawdown 2 forecast. (C) The optimal sampling locations for three new observations with the drawdown 3 DW contour as background. The diamond shows the optimal single observation location for each of the forecasts.

a function of relative weighting of the two forecasts of interest, we can visualize the trade-off that is necessary to satisfy both objectives with different levels of importance. Figure 3A shows the DW cloud for the CH water balance and the spring flow forecast. Following the cloud from point 1 to 4 it can be seen that the sampling site can be optimized slightly for the spring flow forecast by a marginal deterioration (shift to the left) of the DW for the CH water balance. In Figure 3B the cloud is plotted for the DW of drawdown 2 and 4. Adding 20% weighting to the drawdown 4 results in a very large increase in DW for this forecast (shift upward) with a relatively small loss in DW for drawdown 3. Additional weighting on drawdown 4 adds very little to its DW with slightly greater reduction in the DW for drawdown 2.

Multiple Forecasts, Multiple Observations

The results for optimizing single-point data collection to support multiple predictions suggest that it may be beneficial to include multiple new measurement locations when considering multiple objectives. This is especially true for mutually exclusive predictions, such as the example in Figure 3A. To optimize for multiple new forecasts given multiple new observations, we can perform a new analysis that is similar to the one presented in Figure 3. However, now each point in the DW cloud represents one combination of measurements taken from the selection matrix. These results are presented for a forecast scenarios in Figure 4. The cloud of blue points in Figure 4A is the representation of two new measurements, the cloud of green points shows combinations of three new observations, the cloud of magenta points is for four and the red is for five new measurements. Similar to the one observation case, we can find the combinations of measurements that outline the cloud by adjusting the weights given to the individual DW terms in the determination of the VI. For each cloud, these are marked with black dots. All series track from right to left as the weighting of drawdown 2 is reduced from 100% to 0%.

For the specific case with three new observations, we have elaborated the analysis slightly (Figures 4B and 4C). Here the background contour shows the DW contour for a single new observation and a single forecast, and the optimal single observation is marked with a red diamond. Figures 4B and 4C also show a number of points in different colors. Each set of points with the same color show the location of three observations, given a specific weight distribution in Equation 1. To be able to compare with the DW contour for each forecast, the same selected points are shown on both Figures 4B and 4C. In Figure 4B, the optimal three points considering only the drawdown 2 forecast is marked with dark blue. Similarly, in Figure 4C, the optimal three measurement locations for the drawdown 3 forecast is marked with brown. By adjusting the color scale to indicate the weight placed on each prediction, we can see how the distribution of optimal observations changes. An example of the importance of the forecast weighting can be seen by comparing the locations of the brown and the orange dots. As mentioned, the brown dots mark the optimal locations for the drawdown 3 forecast; however, by reducing the weight of drawdown 3 to 80%, one of the sample sites is moved close to the optimal sampling site for the drawdown 2 forecast (marked with the red diamond in Figure 4B).

Retrieval of Information Index

The results presented above have all been derived assuming perfect knowledge of the true covariance function describing the subsurface heterogeneity. Unfortunately, such knowledge is rarely available in reality, and it

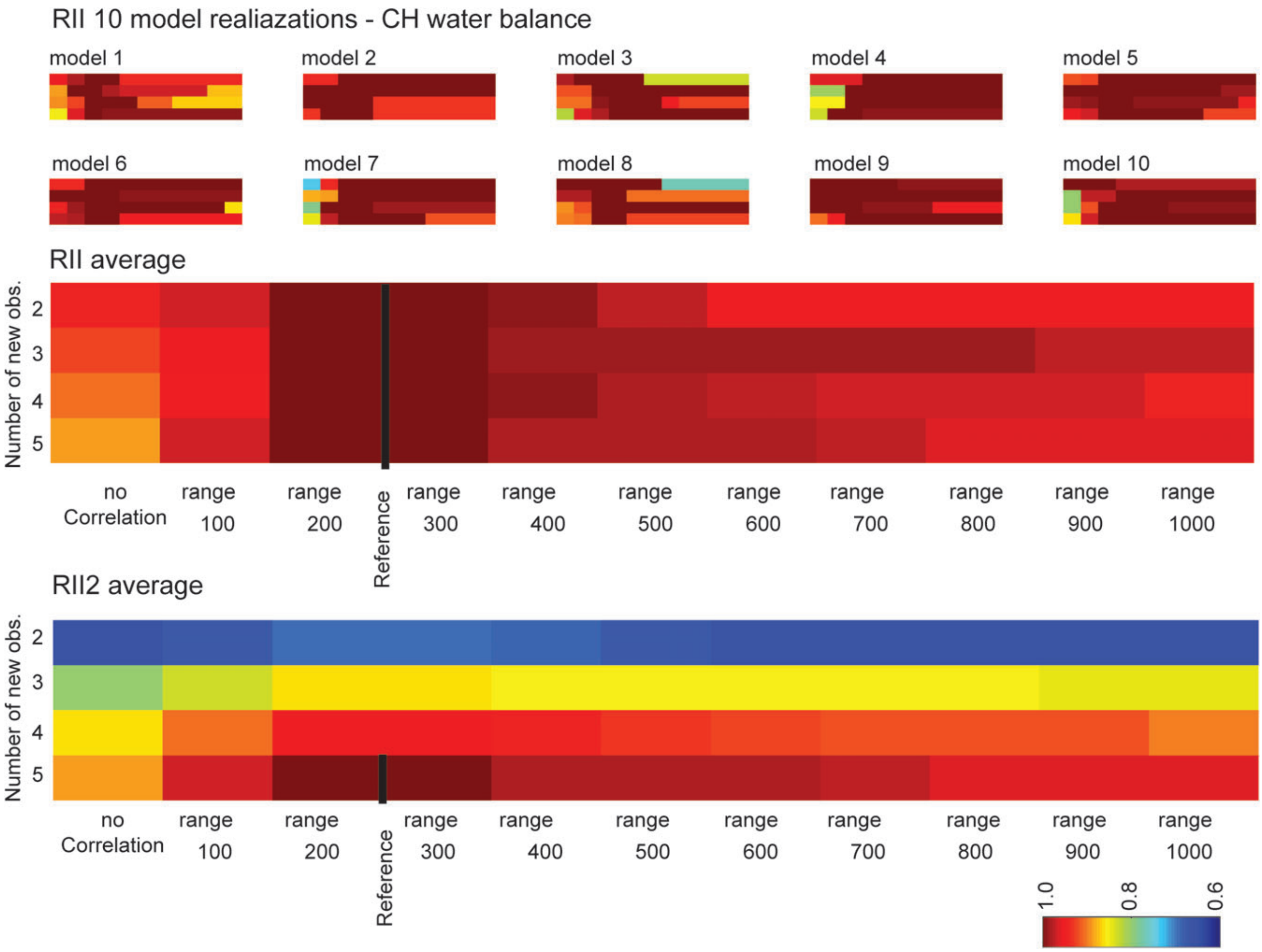

Figure 5. This figure show the RII for the CH water balance forecast for 10 model realizations created using the same variograms as the model shown in Figure 2. Based on these individual models, the average RII is calculated and presented below. Finally, the RII2 index is shown in the bottom by averaging over the 10 model realizations. The 10 small plots have the same axes labels as the large plots.

must therefore be estimated based on prior knowledge or from existing field data. As part of the present study, we have therefore estimated the reduction in DW expected due to selecting data using an incorrect covariance function. This has been done by rerunning the selection algorithm using covariance functions with increasing degrees of bias and calculating RII and RII2. This bias was introduced by calculating covariance matrices with a variogram model, where the range was gradually changed away from the true value. The analysis was performed on 10 model realizations and then averaged to get a mean estimate of the performance. We did this for two different forecasts to be able to evaluate the consistency of the results (the second scenario is presented in the online supplementary material). The results for the CH water balance forecast are shown in Figure 5. The figure is vertically subdivided into three parts. The upper section contains the RII for the 10 individual model realizations. By comparing the RII for the individual realizations, it is apparent that the effect of a biased covariance function is highly dependent on the analyzed model. The averaged result is presented in the central section of Figure 5. Here the black vertical line shows the range of the true covariance function. The upper row of this subplot is the estimated RII for two observations. As expected, when we approach the true covariance function (range of 250), the points selected result in RII values close to 1, meaning that the observations selected have the same DW (but not necessarily the same location) as the observations selected using the true covariance function. As more bias is introduced, the RII deteriorates, thereby resulting in lower DW of the selected points than calculated. However, even with a highly biased covariance function (in the range of 1000). The RII approaches 0.9 meaning that the DW obtained with the points selected using a highly biased covariance function is still close to the DW obtained using the true covariance function. In this plot, special notice should be given to the case where new data points are assumed uncorrelated. This is the case where a diagonal data covariance matrix is implemented in Equation 1. Using the case with two new observations (the upper row of the RII average figure) we see that the RII obtained with the assumption of no correlation has similar performance as the highly biased covariance functions. However, the bias introduced by

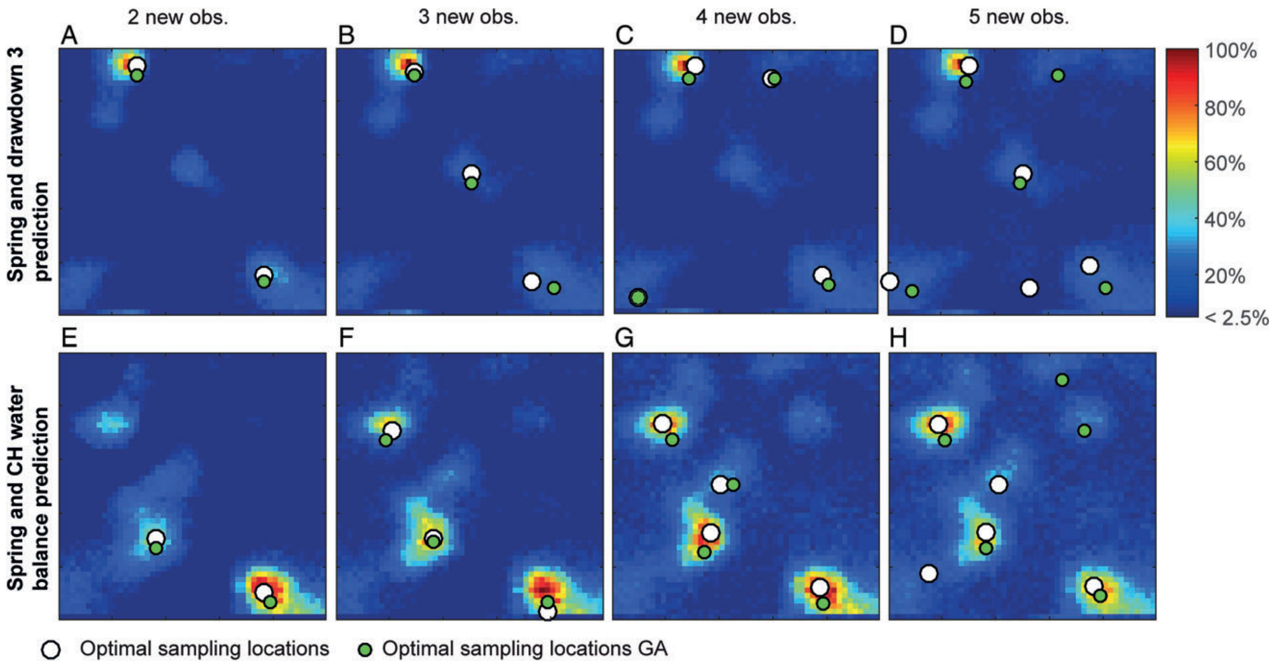

Figure 6. Probability contour maps together with the optimal sampling locations for random selection approach as well as points selected using the GA. The contour plots show the probability that a given observation location is part of the optimal 2.5% sampling locations given two to five new observation locations. The upper row shows a case where sequential selection would not result in optimal sampling, and the lower plot show a case where sequential sampling is expected to give the same result as the methodology suggested here.

assuming uncorrelated observations gradually increases, as more observations are included into the analysis. The same pattern cannot be observed for the remaining covariance structures. Based on this, it is evident that, as the size of the new data set increases, it becomes more important to include some representation of the covariance structure, even if this representation is somewhat biased.

The lower subplot of Figure 5 shows the average RII2 value. Compared to RII we, have now normalized with the reference scenario of five new data points selected using the true covariance function. Based on this, we can determine if increasing the number of observations collected can compensate for a biased covariance structure. This is indeed the case, and it can be seen by comparing the results for four new observations chosen with a covariance function with low bias (range 200 and range 300). The RII2 has approximately the same value as that calculated with five observations selected with a high bias (800 and above). Again, here we should notice that this is not the case when assuming new observations to be uncorrelated. Selecting five observations that are assumed uncorrelated have only slightly higher RII2 than three observations selected using a range close to the true covariance structure.

New Observation Probability Maps

Based on the DW clouds such as that shown in Figure 4 we can derive the optimal locations for acquiring new data. This is done by calculating the VI for all combinations of measurements using Equation 4 with a proper weight distribution depending on the focus of the study. For the VI of the spring flow and the drawdown 3 forecast, given an equal weight of 50% to each, the optimal sampling sites can be derived from Figures 6A–6D for two to five new observations. Similarly, the optimal sampling sites using the same weight distribution are shown for the combined forecast of the spring and CH water balance (Figures 6E–6H). For reference, the selected points are compared to those selected using the GA. For two to four new observations, there is a clear consistency between the points selected using the random selection approach and the points selected using the GA. This is not the case for five new observations. The explanation for this can be twofold. First, the explanation may lie in an incomplete search of the set of possible solutions. Second, it can be explained by an equivalence, where multiple combinations of measurements have equal or close to equal VI. For the spring flow and drawdown 3 forecast, this is due to the first of these explanations: the VI for the five points selected with the GA is 0.20 while the VI for the random approach is 0.18. For four new observations, the VI values are 0.18 for the GA and 0.17 for the random search. The same VI is thereby obtained for four observations selected with the GA as for five points selected with the random search. Moreover, we see that the VI does not increase significantly going from four to five new measurements. For the spring and CH water balance forecast, the VI for five new observations is 0.21 for the GA and 0.18 for the random search approach. Corresponding VI values for four new observations identified using the GA and random search approaches are 0.18 and 0.17. Again, we get the same VI for four observations selected with the GA as we do for five points selected with the random search. However, based on the random search method we can determine how likely it is that a given observation is part of the optimal set. This is shown by the background contours. These probabilities are derived from the DW cloud similar to those seen in Figure 4. Here we selected the 2.5% of the combinations of measurements with the highest VI from the cloud of two to five measurements and located these observations in space. The probabilities were then estimated by calculating the chance that a given observation would be included in this optimized subset. Based on these results, it is possible to identify areas where new observations are likely to contribute to an increased VI. Because these contour plots are held relative to the number of new observations in the set, they will change with the number of new observations. Based on these probability contours it is evident that the optimal combinations of measurements are most often located in areas likely to have a positive effect on the VI. The only exception is one of the randomly selected observations (Figure 6D). As mentioned, this is caused by an incomplete search of the all possible observation sets. It would be worthwhile expanding this investigation to more problems, including problems with extensive existing data sets, to see if this result is generally true. If so, it could point to a cost effective approach to determining the optimal locations for measurements and, to some degree, the optimal number of measurements to add.

Discussion

In the present study, a framework is presented for assessing the value of yet to be collected data with the purpose of maximizing the DW for multiple forecasts of interest, and analyzing the sensitivity to their absolute locations. As mentioned in several of the studies cited, both the strength and the limitations of the proposed method lie in the assumption of model linearity. The two main advantages of this are: we do not need to know or to make assumptions for the actual value of the data (that have not been collected yet); and the computational burden is limited because the results do not rely on a high number of potentially computationally expensive forward model runs. Fienen et al. (2010) showed the need to have a dense pilot point parameterization when performing linear DW analysis. For field applications of the methodology proposed in this study, the same would apply. We did not include this analysis in the proposed analysis, since we designed our model such that we, in principle, did not have any representation from structural uncertainty.

The proposed method ignores potential conceptual model uncertainty. Other studies (Neuman et al. 2012; Xue et al. 2014) have shown that forecast uncertainty can be biased and/or affected by an ignoring conceptual model error. The proposed methodology and other studies would therefore benefit from an extension to account for multi-model realizations. This would require an analysis of multiple model realizations, which is beyond the scope of the present study, but could be a target for expansion of the proposed methodology.

Compared to other studies (e.g., Dausman et al. 2010; Wallis et al. 2014; Wöhling et al. 2016) we have limited our search to a single type of measurement. This allows for analyzing the importance of knowing the correlation of the potential new measurements. Moreover, by restricting our search to measurements of hydraulic conductivity, we could use the true covariance function on which the model was built as a reference. By gradually introducing increasing levels of bias in the covariance structure we could analyze how this effected the selected measurement locations. Doing this, we found that the selected measurement locations had limited sensitivity to the chosen covariance structure. Even with highly biased covariance structures we still obtained a RII of close to 0.9. Moreover, we also found that the largest bias was introduced by assuming that new measurements were uncorrelated. The negative effect of assuming uncorrelated measurements also increased with the number of measurements, which can be explained easily with the increasing redundancy introduced as more observations are included. Unfortunately, the true covariance function can rarely be determined in practice. We therefore analyzed if a biased covariance structure could be compensated by an increase in sample size. We found that the RII2 obtained from a sample size of 4 using the true covariance function was similar to the RII2 with a sample size of 5 using a highly biased covariance function. Again, it would be useful to extend this investigation to more measurement types and hydrogeologic conditions. But, if it is found to be generally true, then it confirms the general expectation that we can compensate a biased assumption of correlation with an increased number of samples. For the special case where new data was assumed uncorrelated, this is not the case. Increasing the sample size from four new observations to five could not compensate for the assumption of uncorrelated measurements. Therefore, when performing OD studies, we suggest always to include some assumption of observation correlation. Unfortunately, measurements of hydraulic conductivity as used here, are not commonly collected in hydrological studies. Here the dominant observation types are hydraulic head data and streamflow measurements. Correlation structures for hydraulic head measurements can be assessed from the data sets directly, for example, using variogram models fitted to observation data or prior knowledge. Correlations and/or cross-correlations between head and flow measurements can also be evaluated using the numerical model using first-order second-moment methods. A methodology to do this is outlined in Kunstmann et al. (2002).

In the present case, we had knowledge of the true correlation between potential new measurements. As mentioned, this will rarely be the case in a practice. To analyze the importance of the assumed correlation structure, the RII and RII2 indexes can be used to evaluate the bias introduced by different assumptions on measurement correlations, and an estimate can be obtained on the expected bias introduced. In practice, this is done by selecting the best approximation to the true covariance function (either based on expert knowledge, from existing data or through numerical analysis). The RII and the RII2 indexes are then calculated, using this reference, by gradually diverging from this (e.g., by changing the assumed correlation length of the variogram). For cases with multiple different types of measurements (e.g., heads and hydraulic conductivity measurements), this becomes a more challenging problem. However, methods to estimate the correlation between different measurement types were suggested in the previous section.

Most of the present study builds on a random search procedure to find the optimal combination of potential new measurements. The benefit of this method is that it allowed us to estimate a probability map showing how likely it would be that a given location would increase the VI. From a practical application standpoint, such probability maps are valuable because measurements in the field rarely can be made at the exact location proposed by the optimization procedure. The probability maps can thereby be seen as a guide to determine best alternatives. However, in cases where the number of potential measurement locations is high, the random selection method becomes inadequate due to the poor scaling as the number of new measurement locations increase. Wöhling et al. (2016), acknowledged this by utilizing a GA global optimization scheme similar to the one applied here. By comparing the optimal data points selected using the GA and the ones selected using the random search approach, we conclude that for a sample size of 2500 potential new measurements, a population of 1.000.000 samples was insufficient to locate the optimal combination of measurements when selecting five potential new locations. For selection of four potential new measurements, both routines selected approximately the same points, or at least selected measurement sets with very similar VI values. Analyzing the probability contour maps, it should, however, be noted that these do indicate the points selected with the GA as potentially valuable. This shows that the probability maps are more robust than the selection of optimal locations. To further document this, a video sequence showing how the probability maps behave with a reduced sample size is provided in the supplementary material. For the present case, it is shown that approximately the same maps are obtained down to a sample size of 200.000 combinations (see Video S1). The effectiveness of the random search procedure will be highly affected by the number of potential new measurements. In the present case, we had a sample size of 2500 potential locations. This will often be much more than what is proposed in a field case. Reducing this size will render the random search much more effective. However, we do suggest comparing the probability maps with points selected using the GA.

The performance of the method could be greatly increased if new observation locations could be selected sequentially, as done by White et al. (2016). However, as pointed out several times in this study, the new locations will be correlated, thereby making sequential selection inadequate (this was already acknowledged by Carrera et al. (1984)). However, the performance of sequential selection will be dependent on the problem analyzed. By comparing the two forecast scenarios in Figure 9 it is evident that the sequential selection will work better for the forecast of constant head and spring flow than for the forecast of spring flow and drawdown 3. Analyzing these types of forecasts, it is evident, that the both forecasts in the former case are predominantly flux driven, where in the latter case, they are flux (spring flow) and transmissivity (drawdown 3) driven, respectively. This could therefore indicate that sequential selection is sufficient, when the same processes drive the multiple forecasts. Another particular case where sequential selection could be preferred is in the presence of forecast nonlinearity. Here forecast nonlinearity is defined as the scenario where the forecast may be dependent on the information gained by making new measurements. The obvious case is where the assumed model structure is biased, and the new observation corrects this bias. This could be exemplified by a borehole leading to an update of an otherwise fixed model structure. Depending on the prior availability of data and on the model, such an update will often be necessary after data collection. It could therefore be preferred to make a combined analysis, where the selection of new data points is limited in sample size. After data collection, an evaluation should be performed to determine if the data collection should result in model updating. After the potential model update, a new series of potential measurement locations could be selected, based on a second round of OD analysis. Ultimately, this will rely on experience and judgment of the hydrogeologist and it will be constrained by time and budgetary considerations.

The suggested methodology could be extended to include both multiple types of measurements, and a monetary trade-off. Such a trade-off should be included in cases where different types of measurements are considered that have different costs (and, potentially, different uncertainties and information contents). An example of this could be a comparison between the value of indirect geophysical measurements versus direct measurements in the form of boreholes. The indirect measurements can often be acquired at a reduced cost, but potentially also with a higher uncertainty. By including the trade-off between price and uncertainty into the analysis, the optimal combination of measurement types can be determined. A similar approach could be used to determine the optimal number of new data of a single type in a cost/benefit framework.

Conclusions

In the present study, we have extended an existing method for OD of data collection to reduce uncertainties in hydrological model forecasts. Compared to previous studies, we have allowed for optimization for multiple forecasts, which we believe to be of real practical benefit. Here the main finding shows that an optimal compromise between multiple targets can often be found in a way that maximizes the value of the data with respect to the forecasts. We have also included a detailed analysis of the importance of knowing the true correlation between candidate measurement locations. Based on this analysis, we show that an assumption of uncorrelated measurements will have a negative impact on the value of the selected candidate locations. Moreover, this negative impact will increase with the total number of new measurements selected. Based on these findings, we therefore advocate that estimates of measurement correlations should always be included in the analysis when selecting multiple new measurements. This is the case even if this can only be based on expert judgment or limited field data.

In this study, we optimized data sampling using both a random search procedure that considered the entire set of possible measurement combinations and a GA approach to search through measurement set space. In the current study, we included 2500 potential new measurement locations, and considered 1 million combinations of these observations. Using this sample size, we were unable to find the optimal set of measurements for sets larger than four measurements with the random search approach. The size of the pool of potential measurement sets will have to be increased for more complex problems, but would be decreased if fewer potential measurements were considered. In contrast, for our investigation, the GA approach was capable of finding optimal data sets efficiently. Based on this, we suggest that the GA approach be used first, to identify important observations to include in the random search analysis. Then a random search analysis should be conducted to generate maps showing the probability of a given observation to be part of the optimal set. Such maps can be highly valuable in a field case, because they allow a hydrogeologist to find alternative optimal locations in the case where the elite candidate locations are inaccessible. Ultimately, it would be beneficial to derive the probability maps or direct estimates of a Pareto front showing trade-off among different forecasts from the results of the GA. This should therefore be a target for further investigations.

Acknowledgments

The authors would like to acknowledge the valuable comments and suggestions given by Editors Henk M. Haitjema and Yu-Feng Lin and two anonymous reviewers. The comments provided by Randall Hunt were especially very detailed and constructive and greatly improved the quality of our work. The first author was financed by the HyGEM project "Integrating geophysics, geology, and hydrology for improved groundwater and environmental management," project no. 11-116763. The funding for HyGEM is provided by The Danish Council for Strategic Research.

Appendix

All calculations related to the OD in the present analysis have been implemented in Python. This includes both the random search procedure, and the GA. These scripts can be obtained on request from the first author. When this study was completed, the authors became aware of the existing Python package DEAP (Fortin et al. 2012) which incorporated GA methods similar to the ones developed in this study. In future applications, we would consider using this framework instead of the one developed as a part of this study.

Supporting Information

Additional Supporting Information may be found in the online version of this article. Supporting Information is generally *not* peer reviewed.

Appendix S1. Theoretical background and additional results.

Video S1. Robustness of probability maps with decreasing sample size

References

Anderson, M.P., W.W. Woessner, and R.J. Hunt. 2015. *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*. London: Elsevier.

Brunner, P., J. Doherty, and C.T. Simmons. 2012. Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models. *Water Resources Research* 48: W07513.

Carrera, J., E. Usunoff, and F. Szidarovszky. 1984. A method for optimal observation network design for groundwater-management. *Journal of Hydrology* 73: 147–163.

Certes, C., and G. de Marsily. 1991. Application of the pilot point method to the identification of aquifer transmissivities. *Advances in Water Resources* 14: 284–300.

Chadalavada, S., and B. Datta. 2008. Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers. *Water Resources Management* 22: 651–670.

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

Criminisis, A., T. Tucciarelli, and G.P. Karatzas. 1997. A methodology to determine optimal transmissivity measurement locations in groundwater quality management models with scarce field information. *Water Resources Research* 33: 1265–1274.

Dausman, A.M., J. Doherty, C.D. Langevin, and M.C. Sukop. 2010. Quantifying data worth toward reducing predictive uncertainty. *Groundwater* 48: 729–740.

Dixon, W., and B. Chiswell. 1996. Review of aquatic monitoring program design. *Water Research* 30: 1935–1948.

Doherty, J. 2016. *PEST, Model-Independent Parameter Estimation, User Manual Part I: PEST, SENSAN and Global Optimisers*. Brisbane, Australia: Watermark Numerical Computing.

Doherty, J. 2015. *Calibration and Uncertainty Analysis for Complex Environmental Models*. Brisbane, Australia: Watermark Numerical Computing, Brisbane, Australia.

Doherty, J. 2013. *Addendum to the PEST Manual*, 293. Brisbane, Australia: Watermark Numerical Computing.

Doherty, J. 2010. *PEST Model-Independent Parameter Estimation User Manual*, 5th ed. Brisbane, Australia: Watermark Numerical Computing.
Doherty, J. 2003. Ground water model calibration using pilot points and regularization. *Groundwater* 41: 170–177.
Doherty, J.E., and Hunt, R.J., 2010, Approaches to highly parameterized inversion—A guide to using PEST for groundwater-model calibration. Scientific Investigations Report 2010–5169. Reston, Virginia: USGS, 59.
Doherty, J., and R. Hunt. 2009. Two statistics for evaluating parameter identifiability and error reduction. *Journal of Hydrology* 366: 119–127.
Engelhardt, I., H. Prommer, C. Moore, M. Schulz, C. Schueth, and T.A. Ternes. 2013. Suitability of temperature, hydraulic heads, and acesulfame to quantify wastewater-related fluxes in the hyporheic and riparian zone. *Water Resources Research* 49: 426–440.
Feyen, L., and S. Gorelick. 2005. Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas. *Water Resources Research* 41: W03019.
Fienen, M.N., Doherty, J.E., Hunt, R.J., and Reeves, H.W., 2010, Using prediction uncertainty analysis to design hydrologic monitoring networks: Example applications from the Great Lakes water availability pilot project. Scientific Investigations Report 2010–5159. Reston, Virginia: USGS, 44.
Fortin, F., F. De Rainville, M. Gardner, M. Parizeau, and C. Gagne. 2012. DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research* 13: 2171–2175.
Freeze, R.A., B. James, J. Massmann, T. Sperling, and L. Smith. 1992. Hydrogeological decision analysis: 4. The concept of data worth and its use in the development of site investigation strategies. *Groundwater* 30: 574–588.
Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Co. Inc, Boston, MA, USA.
Harbaugh, A.W., 2005, MODFLOW 2005, the U.S. Geological Survey modular ground-water model—The ground-water flow process. Techniques and Methods A16. Reston, Virginia: USGS.
Hill, M.C., C.C. Faunt, W.R. Belcher, D.S. Sweetkind, C.R. Tiedeman, and D. Kavetski. 2013. Knowledge, transparency, and refutability in groundwater models, an example from the Death Valley regional groundwater flow system. *Physics and Chemistry of the Earth* 64: 105–116.
Hill, M.C., and C.R. Tiedeman. 2007. *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*, Vol. 18, 455. Hoboken, New Jersey: Wiley-Interscience.
James, B.R., and S.M. Gorelick. 1994. When enough is enough—The worth of monitoring data in aquifer remediation design. *Water Resources Research* 30: 3499–3513.
Kikuchi, C.P., T.P.A. Ferre, and J.A. Vrugt. 2015. On the optimal design of experiments for conceptual and predictive discrimination of hydrologic system models. *Water Resources Research* 51: 4454–4481.
Kunstmann, H., W. Kinzelbach, and T. Siegfried. 2002. Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling. *Water Resources Research* 38, no. 4: 1035. https://doi.org/10.1029/2000WR000022
Lavenue, M., and G. de Marsily. 2001. Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method. *Water Resources Research* 37: 2659–2675.
Leube, P., A. Geiges, and W. Nowak. 2012. Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. *Water Resources Research* 48: W02501. https://doi.org/10.1029/2010WR01
Loaiciga, H.A., R.J. Charbeneau, L.G. Everett, G.E. Fogg, B.F. Hobbs, and S. Rouhani. 1992. Review of groundwater quality monitoring network design. *Journal of Hydraulic Engineering-Asce* 118: 11–37.
Neuman, S., L. Xue, M. Ye, and D. Lu. 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. *Advances in Water Resources* 36: 75–85.
Nicklow, J., P. Reed, D. Savic, T. Dessalegne, L. Harrell, A. Chan-Hilton, M. Karamouz, B. Minsker, A. Ostfeld, A. Singh, E. Zechman, and A.T.C. Evolutionary. 2010. State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management-ASCE* 136: 412–432.
Norberg, T., and L. Rosen. 2006. Calculating the optimal number of contaminant samples by means of data worth analysis. *Environmetrics* 17: 705–719.
Nowak, W. 2010. Measures of parameter uncertainty in geostatistical estimation and geostatistical optimal design. *Mathematical Geosciences* 42: 199–221.
Park, S., J. Choi, S. Wang, and S. Park. 2006. Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modelling* 199: 289–297.
Poeter, E.P., Hill, M.C., Lu, D, Tiedeman C.R., and Mehl, S., 2014, UCODE_2014, with new capabilities to define parameters unique to predictions, calculate weights using simulated values, estimate parameters with SVD, evaluate uncertainty with MCMC, and more. Integrated Groundwater Modeling Center Report Number GWMI 2014-02, Golden, CO: Integrated GroundWater Modeling Center, Colorado School of Mines.
Tonkin, M., and J. Doherty. 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. *Water Resources Research* 45: 1–17.
Tonkin M.J., Tiedeman C.R., Ely D.M., and Hill M.C., 2007, OPR-PPR, a computer program for assessing data importance to model predictions using linear statistics. Techniques and Methods Report TM-6E2. Reston Virginia: USGS, 115.
Wallis, I., C. Moore, V. Post, L. Wolf, E. Martens, and H. Promrner. 2014. Using predictive uncertainty analysis to optimise tracer test design and data acquisition. *Journal of Hydrology* 515: 191–204.
Welter, D.E., White, J.T., Hunt, R.J., and Doherty, J.E., 2015, Approaches in highly parameterized inversion—PEST++ version 3, a parameter estimation and uncertainty analysis software suite optimized for large environmental models. Techniques and Methods. Reston, Virginia: USGS, book 7, chap. C12, 54. https://doi.org/10.3133/tm7C12
White, J., M. Fienen, and J. Doherty. 2016. A python framework for environmental model uncertainty analysis. *Environmental Modelling & Software* 85: 217–228.
Wöhling, T., A. Geigers, and W. Nowak. 2016. Optimal design of multitype groundwater monitoring networks using easily accessible tools. *Groundwater* 54, no. 6: 861–870.
Xue, L., D. Zhang, A. Guadagnini, and S. Neuman. 2014. Multimodel Bayesian analysis of groundwater data worth. *Water Resources Research* 50: 8481–8496.
Zhang, Y., G. Pinder, and G. Herrera. 2005. Least cost design of groundwater quality monitoring networks. *Water Resources Research* 41, no. 8: W08412. https://doi.org/10.1029/2005WR003936

Authors' Note

The authors do not have any conflicts of interest or financial disclosures to report.