“Qualitative Model Predictive Error Analysis” - a Protocol for a Simple Practical Application

Piotr A. Rzepecki
Senior Ground Water Modeller, AECOM, Perth, Australia
E-mail: piotr.rzepecki@aecom.com

Abstract

Post project audits reveal that a large portion of the predictions generated by groundwater models are at significant error. Since these models are often used by environmental managers for making decisions, the cost of such predictive errors may be substantial. This represents a considerable problem, and sensitivity analysis alone is not considered to be an acceptable approach to address it. Some emerging approaches and technologies, like a “Null Space Monte Carlo” (or “Calibration-Constrained Monte Carlo”), indicate the direction for future developments. However, many practitioners, who are currently constrained by time and budget, find those technologies rather difficult to implement. To overcome this, a simple protocol termed “qualitative model predictive error analysis” is proposed. This paper presents the application of this protocol using a simple, synthetic model as an example. The outcome of the analysis is compared to the results of a more systematic procedure implementing the concept of “Calibration-Constrained Monte Carlo”.

1. INTRODUCTION

Predictions generated by groundwater models are often at significant error. Predictive errors can be attributed to a variety of factors, including: insufficient data for characterizing hydraulic properties (over a large enough area), or system’s states / responses to stress (over sufficient range of values over sufficient time) on which the model predictions depend (Hunt et al, 2007); inaccurate conceptual models / assumptions (Stewart & Langevin, 1999; Roadcap & Wilson, 2002); inadequate geometrical representation of a complex system and its heterogeneities; errors resulting from spatial interpolations (Tonkin et al, 2005); errors in measurements; poor test data collection designs and inadequate interpretation of the collected data (Andersen & Silong, 2003); not representing relevant processes; limitations of models and numerical methods used; and finally, unpredictable natural and human factors (Konikow, 1986).

Model users are increasingly aware of the need to evaluate uncertainties; and a variety of approaches have been proposed to address this. Prominent examples include sensitivity analysis, which is a prevailing approach (ASTM, 2008); “Bayesian model averaging” applied to multiple conceptual models and multiple parameter estimation methods (Li & Tsai, 2008); parallel testing of several viable conceptual models, combined with parametric uncertainty analysis carried out for each conceptual model (Ye et al, 2010); ‘the use of “pilot points” in conjunction with nonlinear parameter estimation software that incorporates advanced regularization functionality’; ‘Calibration-Constrained Monte-Carlo’, also called ‘Null Space Monte Carlo’ (Doherty, 2003); ‘Subspace Monte Carlo’ that allows calibration-constrained random heterogeneity and/or related approaches (Tonkin & Doherty, 2005; Tonkin et al, 2005; Tonkin & Doherty, 2008).

Sensitivity analysis alone is not considered to be a substitute approach, as varying the values of model parameters often results in a significant model “de-calibration”, and de-calibrated models should not be used for predictions. However, sensitivity analysis is still considered an important part of uncertainty analysis. In turn, some of the other, more sophisticated approaches can be quite involved and practitioners often find them difficult to implement.

This paper outlines a qualitative error analysis procedure and presents the results of a numerical experiment carried out to test it and compare it with a more quantitative approach using the ‘Calibration-Constrained Monte-Carlo’ method. Both the proposed approach and the Monte-Carlo procedure, can be used to evaluate parametric uncertainty. Such uncertainty is related to what is called a non-uniqueness of the solution to ill-posed inverse problems (Ne-Zheng Sun, 1999). Given a set of targets (derived from field measurements), an infinite number of calibrated models can be produced – each generating a somewhat different prediction.

2. PROCEDURE FOR QUALITATIVE MODEL PREDICTIVE ERROR ANALYSIS

The proposed procedure was developed in an attempt to find an approach to parametric uncertainty analysis that is both practical and relatively easy to implement. It is referred to as a "qualitative model predictive error analysis" and consists of the following steps:

- Conduct preliminary, manual model calibration
- Conduct a systematic sensitivity analysis
- Conduct automatic model calibration to obtain a "baseline model version"
- Develop several model versions by fixing the most sensitive parameters at their lower and upper value in a range of values derived using "previous knowledge" or professional judgment
- Calibrate each model version – if fixing the parameter value results in an inability to calibrate the model, adjust that value stepwise, until the calibration can be achieved
- Run predictive simulations using each model version and tabulate the results

The reminder of this article documents a numerical experiment designed and carried out to test this proposed procedure and summarizes its results.

3. A SYNTHETIC MODEL CREATED FOR COMPARATIVE ERROR ANALYSIS

A simple, synthetic, two-dimensional, steady-state numeric finite-difference MODFLOW model was established (see Figure 1 and Table 1) in order to carry out comparative error analysis: qualitative analysis (described in Section 4.1) vs. semi-quantitative analysis (described in Section 4.2). The model was intended to represent a hypothetical alluvial aquifer located between two lakes (east and west of the aquifer) and bound by impermeable bedrock (present south and north of the aquifer and at the base of the aquifer).

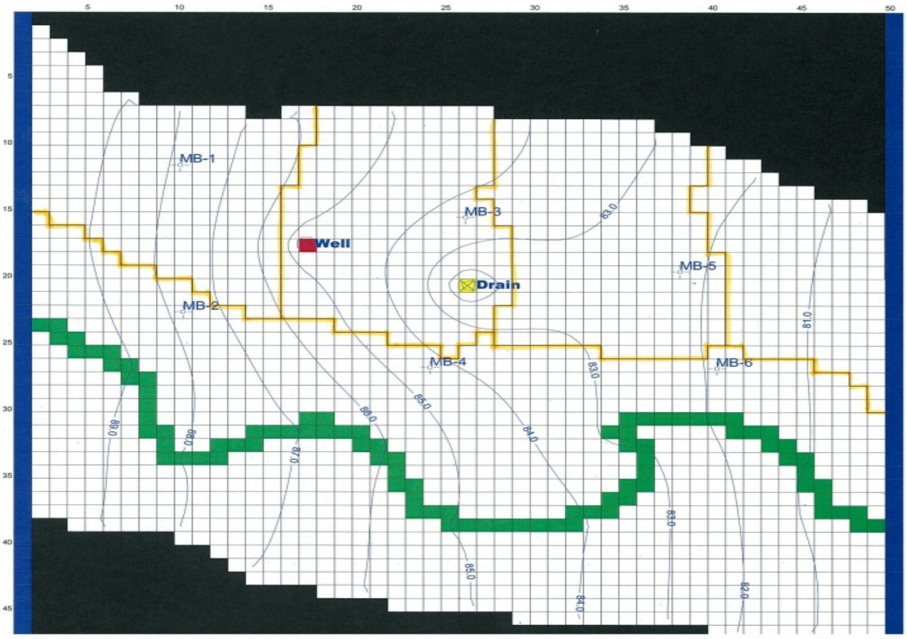

Figure 1 MODFLOW Model Set-up

The model grid consists of one layer, fifty rows and fifty columns, defining 2,500 cells, each 50 m x 50 m in size. Elevation of the grid base was set at 0 m and the top at 100 m. The lakes are represented using constant head boundaries, set at 90 m and 80 m on the west and east sides of the model domain, respectively. A 20 m wide river connecting the lakes is represented in the model with the stage changing linearly from cell to cell, from 90 m in the west to 80 m in the east. The 1 m thick river bed was assigned a 5 m/day vertical hydraulic conductivity. One well was set in the central-western part of the model domain, pumping at a rate of 3,000 m^3/day. A uniform recharge over the entire model domain was set to 0.00067 m/day (see Table 1).

Five hydraulic conductivity zones were delineated, with initial values as shown in Table 1. Six head targets (water level observation points), MB-1 through MB-6, were placed in various parts of the model's domain for use in the model calibration, with assigned target values of 88.9, 88.3, 86.0, 86.1, 83.2 and 82.5 m.

Table 1. Model Parameters, Calibration and Predictions

	Unit	Qualitative Analysis Model Versions							Semi-Quantitative Analysis Model Versions Producing Extreme Predictions			
Parameters:		Baseline Model	High Recharge	Low Recharge	High K_{xy} Zone 3	Low K_{xy} Zone 3	High K_{xy} Zone 4	Low Kxy Zone 4	Minimum Prediction		Maximum Prediction	
									Monte Carlo Realization Number			
									Re.: 97	Re.: 56	Re.: 102	Re.: 14
K_{xy} Zone 1	m/day	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
K_{xy} Zone 2	m/day	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
K_{xy} Zone 3	m/day	7.37	14.74	3.69	14.74	3.68	13.75	4.19	4.44	13.73	13.59	4.45
K_{xy} Zone 4	m/day	12.23	24.46	6.12	20.07	7.32	24.47	6.12	11.37	12.13	19.82	13.65
K_{xy} Zone 5	m/day	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
Recharge	m/day	0.00067	0.00135	0.00034	0.00103	0.00043	0.00110	0.00041	0.00041	0.00067	0.00114	0.00069
		Measures of Calibration										
RSS	m	0.12	0.68	0.39	0.14	0.31	0.18	0.32	0.67	0.50	0.34	0.54
RMS	m	0.14	0.34	0.25	0.15	0.23	0.17	0.23	0.34	0.29	0.23	0.30
RMS / Range of Target Observations	%	2.2%	5.3%	3.9%	2.3%	3.6%	2.7%	3.6%	5.3%	4.5%	3.6%	4.7%
		Model Predictions										
Inflow into Drain	m^3/day	4995	5772	4532	5326	4684	5366	4675	4559		5550	
Drawdown at MB-3	m	2.3	2.1	2.6	2.0	2.7	2.0	2.6		2.0		2.7

Note: RSS and RMS are commonly used measures of a model calibration. RSS signifies Residual Sum of Squares; RMS signifies the Root Mean Squared Error. The information presented in Table 1 will be discussed in the following sections of this paper.

4. MODEL PREDICTIVE ERROR ANALYSIS

4.1. Qualitative Model Predictive Error Analysis

The model described in Section 3 was calibrated manually by varying the values of recharge and hydraulic conductivity. The level of calibration accomplished is characterized by RSS and RMS values of 0.15 m and 0.16 m, respectively.

Systematic model sensitivity analysis was carried out by varying values of the seven model parameters: namely hydraulic conductivity for each of the five zones, vertical hydraulic conductivity of the river's bed, and recharge. It was essential to calculate the model sensitivity (and also, later, to calibrate the model) using the model-generated values that are related (directly or, as a minimum, indirectly) to the category of prediction required. For example, to predict the change in water level at a point, the sensitivity should be calculated as a "change in calibration" calculated from the differences between the model-calculated and measured groundwater levels. It is also important to note that the value of such analysis diminishes, as the model-predicted value and/or location of the point of prediction move further from the range and/or area of measurements available for model calibration.

The model was run eleven times for each of the seven parameters, giving a total of 77 runs, with the parameter values changed for each run using the following ratios (with regard to the parameter's baseline value): 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.3, 1.5, 1.7 and 2.0. The simulations were carried out using the "Automatic Sensitivity Analysis" routine provided by the Groundwater Vistas platform (Rumbaugh & Rumbaugh, 2011). The three most sensitive parameters were found to be: recharge, and hydraulic conductivities for zones 4 and 3 (see Table 2). A general note: it is often the case that some parameters can vary over a much larger range of values than others, like hydraulic conductivity vs. recharge. Then, the sequence of values used in a systematic sensitivity analysis should be calculated (consistently for all parameters subject to analysis) as varying percentages of viable ranges (or varying fractions of standard deviations, in case normal distributions are assumed), rather than ratios of base values, as used in this experiment (following a common approach).

The model was then calibrated using the Parameter ESTimation (PEST) program by varying the three most sensitive parameters. The range of values for each of those parameters and distributional assumptions are provided in Table 3. The range of values adopted for model parameters were set to represent an assumed prior knowledge (such as a range of values calculated from aquifer test data) and/or professional judgment. The resulting automatically calibrated model was considered a baseline model with parameters as presented in Table 1. The RSS and RMS of the baseline model are 0.12 m and 0.14 m, respectively.

Table 2. Results of Model Sensitivity Analysis

	Sum of Squared Residuals						
Parameters:	K_{xy} Zone 1	K_{xy} Zone 2	K_{xy} Zone 3	K_{xy} Zone 4	K_{xy} Zone 5	Recharge	K_v of River Bed
Parameter Multiplier							
0.5	0.287	0.143	1.040	1.592	0.273	1.550	0.155
0.6	0.201	0.120	0.695	0.991	0.216	0.984	0.155
0.7	0.155	0.116	0.463	0.603	0.183	0.561	0.155
0.8	0.136	0.124	0.309	0.362	0.165	0.283	0.155
0.9	0.137	0.137	0.211	0.222	0.157	0.148	0.155
1	0.155	0.155	0.155	0.155	0.155	0.155	0.155
1.1	0.184	0.174	0.129	0.140	0.157	0.304	0.155
1.3	0.268	0.215	0.139	0.212	0.169	1.025	0.155
1.5	0.375	0.256	0.202	0.364	0.186	2.305	0.155
1.7	0.495	0.295	0.294	0.558	0.205	4.141	0.155
2	0.688	0.348	0.459	0.879	0.234	7.923	0.155
Total of Sum of Squared Residuals:	3.080	2.085	4.096	6.077	2.099	19.378	1.703
Ranking of Parameter Sensitivity	**4**	**6**	**3**	**2**	**5**	**1**	**7**

Table 3. Set-Up for PEST and Stochastic MODFLOW Runs

				PEST		Stochastic MODFLOW	
		Parameter Bounds		Transformation	Limit	Std. Dev.	Distribution
	Unit	Lower	Upper		(PARCHGLIM)		
Parameter							
K_{xy} Zone 3	m/day	0.00034	0.00136	None	Relat.	0.00034	Uniform
K_{xy} Zone 4	m/day	3.5	14	Log	Factor	3.5	Normal
Recharge	m/day	6	24	Log	Factor	6	Normal

In the next phase of this analysis, versions of a baseline model were developed (in addition to a baseline model). Each version was set for PEST calibration with the value of one of the three most sensitive parameters fixed at the lower or upper end of the range of values used during the automatic calibration of the baseline model. The values of parameters and measures of calibration for all seven models (baseline model and its six versions) are presented in Table 1. Those models can be considered a sample drawn from a population of an infinite number of calibrated models that can be developed, given a set of targets.

Finally, all the seven model versions were modified for carrying out predictive simulations. A drain element was added to one of the model cells, about mid distance between MB-3 and MB-4, to simulate mine (or construction) dewatering to elevation of 80 m (set as drain stage).

The one meter thick drain bed was assigned a vertical hydraulic conductivity of 10 m/day. Each of the seven models was then run, and two results from each of the predictive runs were noted, as shown in Table 1: the rate of groundwater inflow into the drain and drawdown at a nearby observation point, MB-3. The results of the predictive simulations are shown in Table 1 (the bottom rows) and represent a subset of a theoretically infinite number of predictions obtained with the use of an infinite number of calibrated models that can be developed, given a set of targets.

4.2. Semi-Quantitative Model Predictive Error Analysis using the "Calibration-Constrained Monte Carlo Procedure"

The baseline model was then used to set-up and carry out what can be called a "semi-quantitative model predictive error analysis". Such analysis represents a more quantitative, thorough and systematic approach, compared with the "qualitative" analysis outlined in the previous section. The term "semi-quantitative", rather than "quantitative", is used as several important uncertainties (about this more systematic "uncertainty" analysis) still remain, including distributional assumptions and other uncertainty and predictive error generating factors.

The semi-quantitative analysis described below, utilizes the concept of "Calibration-Constrained Monte Carlo" (Doherty, 2003) and was implemented with the use of the Stochastic MODFLOW program (Ruskauff, 1998) executed from within the Groundwater Vistas platform (Rumbaugh & Rumbaugh, 2011). This approach can also be implemented using advanced functionalities of the PEST suite (like PREDUNC program), which compared to Stochastic MODFLOW, offers more sophistication, flexibility and efficiency, but the set-up is more involved.

Stochastic MODFLOW generates a large number of models, some of them calibrated, and some not calibrated. Thus, using Stochastic MODFLOW necessitates one extra step – the screening (conditioning) of the resulting models to retain only the calibrated models. The baseline model was set for stochastic MODFLOW simulations by allowing only the three most sensitive parameters (determined during sensitivity analysis described previously) to be subject to "random sampling". The distributional assumptions and lower and upper bounds were set the same as for model calibration during the qualitative analysis described in section 4.1 (see Table 3).

The Stochastic MODFLOW was set to run 300 simulations (realizations). Figure 2 shows the "cumulative average of sums of squares". This is one of the tools that can be used to determine if the number of realizations is sufficient.

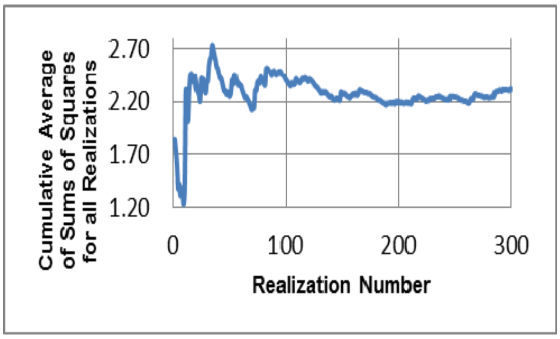

Figure 2 Cumulative Average of Sums of Squares for all Realizations

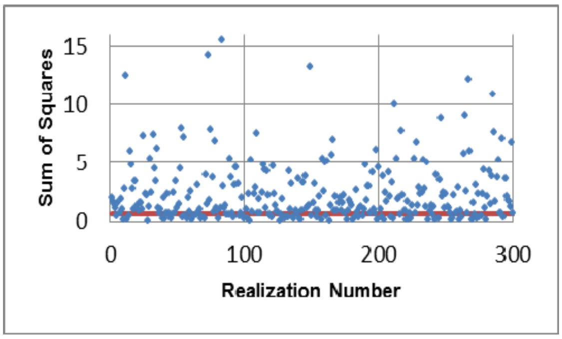

Figure 3 Sum of Squares (SSRs) for All Realizations

In the next phase of this analysis, a cut-off Sum of Squares (SSR) value of 0.675 was selected to separate the calibrated models from the non-calibrated models (see Figure 3). An SSR of 0.675 equals the largest SSR calculated for all seven alternative models developed as part of the qualitative error predictive analysis described in Section 4.1. It corresponds to about 5.3% of the range of groundwater levels set for the model's six targets.

Applying the "0.675 SSR filter" resulted in retaining 76 models, out of original 300 models, that were judged to be calibrated. Those models were then prepared for predictive simulations (by adding a drain) to calculate the groundwater inflow into the drain and subsequent drawdown at MB-3.

4.3. Comparison between the Results of Qualitative and Semi-Quantitative Error Analysis

The combined results of predictive model simulations (using both qualitative and semi-quantitative analysis) are plotted as cumulative distribution functions in Figures 4 and 5. The plots demonstrate that the model predictions are similar for both approaches to predictive error analysis – qualitative and semi-quantitative.

The inflow and drawdown predictions generated by the semi-quantitative analysis were then ranked into classes of predictions and plotted as histograms shown in Figures 6 and 7. These figures also include the normal distribution curves that were fitted to the data (by minimizing the sum of squares of differences between the values of the theoretical curve and the values of the rank classes).

The values of mean and standard deviation of the fitted normal distributions were used to calculate the probability that the responses of the hypothetical-real system (subject to modelling) can be smaller / larger than the minimum/maximum model predictions derived from the qualitative and semi-quantitative analysis. The results of such calculations, when using 5% significance level, are shown in Table 4.

Table 4, together with Figures 6 and 7, are provided to summarize results of the comparing the two methods – qualitative vs. semi-quantitative. However, considering all the remaining uncertainties, it is safer not to attribute statistical significance to the results of such simulations/analysis and simply report ranges of predictions.

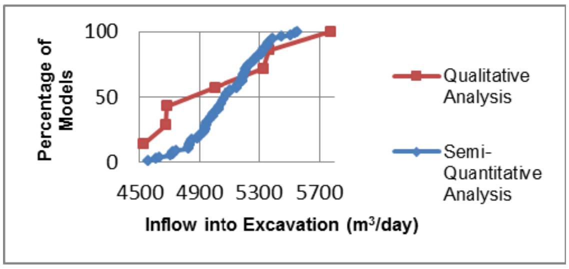

Figure 4 Percentage of Models Predicting a Given Groundwater Inflows into Drain

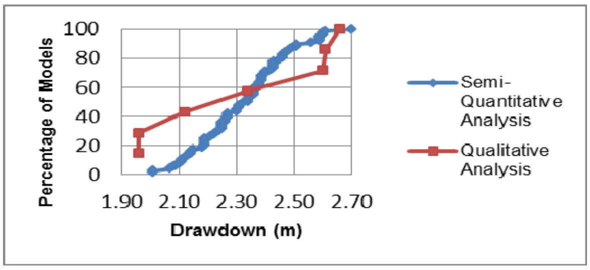

Figure 5 Percentage of Models Predicting a Given Drawdown at MB-3

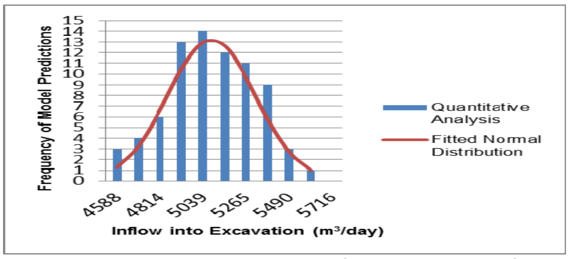

Figure 6 Histogram of Model Inflow Predictions and Fitted Normal Distribution

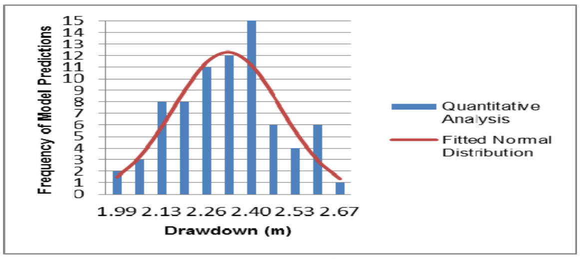

Figure 7 Histogram of Model Drawdown Predictions and Fitted Normal Distribution

Table 4. Probabilities that Inflow into Drain and Drawdown at MB-3 is Outside of Predicted Ranges

Fitted Normal Distributions:				
Inflow Mean = 5,087 m³/day, St. Dev. = 229 m³/day				
Drawdown Mean = 2.33 m, St. Dev. = 0.162 m			Probability	Distance from Mean Measured in Std. Devs.
Qualitative Analysis				
Probability that Inflow Will Be Smaller than:	4532	m^3/day	0.7%	2.4
Probability that Inflow Will Be Larger than :	5772	m^3/day	0.1%	3.0
Probability that Drawdown Will Be Smaller than:	1.96	m	0.9%	2.3
Probability that Drawdown Will Be Larger than :	2.66	m	2.0%	2.0
Semi-Quantitative Analysis				
Probability that Inflow Will Be Smaller than:	4559	m^3/day	0.9%	2.3
Probability that Inflow Will Be Larger than :	5550	m^3/day	2.0%	2.0
Probability that Drawdown Will Be Smaller than:	2.01	m	2.0%	2.0
Probability that Drawdown Will Be Larger than :	2.70	m	0.9%	2.3

5. CONCLUSIONS

The results of this numerical experiment indicate that the proposed procedure for “qualitative model predictive error analysis” may be a promising, less time and budget consuming alternative to the more systematic and resource intensive “Calibration Constrained Monte Carlo” procedure. However, confirmation of such conclusion would require more similar experiments conducted on more complex, “real-world” models.

It is important to stress, again, that the proposed procedure addresses only parametric uncertainty. There are other kinds of uncertainty that also can be analysed. For instance, where several conceptual models are plausible, a more complete analysis would require repeating this qualitative procedure for each conceptual model.

Among the many limitations of the proposed qualitative procedure is the fact that it is applied to models using rigid ‘zones of piecewise parameter uniformity’. Such an approach may be acceptable when applied to many water resources problems, but is likely to be less effective when tackling some fate and transport modelling problems. Model predictions regarding fate and transport are often very sensitive to small scale heterogeneities.

Finally, in case time and budget limitations are strong factors, which they are most of the time, an even more simplified version of the qualitative analysis could be implemented, using just three parallel models (derived using the upper and lower bounds of the one or two most sensitive parameters), instead of seven, as described in this paper. Such analysis would still represent a significant improvement over reporting just a single prediction, qualified by the results of a (“de-calibrating”) sensitivity analysis.

6. ACKNOWLEDGMENTS

I would like to say thank you to Ian Rowbottom for his thorough review of the manuscript of this article. I also appreciate the helpful comments and edits offered by Graham Hawkes.

7. REFERENCES

Andersen P.F. and Silong L. (2003), *A Post Audit of a Model-Designed Ground Water Extraction System*, Ground Water, 41 (2), 212-218.

ASTM (2008), *Standard Guide for Conducting a Sensitivity Analysis for a Ground-Water Flow Model Application*, ASTM Designation: D 5611 – 94.

Doherty J. (2003), *Ground Water Model Calibration Using Pilot Points and Regularization*, Ground Water, 41 (2), 170-177).

Hunt R.J., Doherty J. and Tonkin M.J. (2007), *Are Models Too Simple? Arguments for Increased Parametrization*, Ground Water, 45 (3), 254-262.

Konikow L.F. (1986), *Predictive Accuracy of a Ground-Water Model – Lessons from a Postaudit*. Ground Water, 24 (2), 173-184).

Li X. and Tsai F.T. (2008), *Groundwater head prediction and uncertainty propagation using Bayesian multi-model multi-method*, MODFLOW and More 2008: Ground Water and Public Policy – Conference Proceedings, 300-304.

Ne-Zheng Sun. (1999), *Inverse Problems in Groundwater Modelling, Theory and Applications of Transport in Porous Media*, Kluwer Academic Publishers.

Roadcap G.S. and Wilson S.D. (2002), *Groundwater Modelling of the Mahomet Aquifer: From Conceptual Models to Post-Audits*, Midwest Focus Ground Water Conference, April 11-12, 2002, Chicago.

Rumbaugh J.O and Rumbaugh D.B. (2011), *Guide to Using Groundwater Vistas, Vestion 6,* Environmental Solutions, Inc.

Ruskauff G.J. (1998), *Guide to Using Stochastic MODFLOW for Monte Carlo Simulations,* Environmental Solutions, Inc.

Stewart M. and Langevin C. (1999), *Post Audit of a Numerical Prediction of Wellfield Drawdown in a Semiconfined Aquifer System,* Ground Water, 37 (2), 245-252.

Tonkin, M. and Doherty, J. (2005), *A hybrid regularised inversion methodology for highly parameterised models, Water Resources Research*, 41 (W10412)

Tonkin M., Hass J., Trego D. and Muffels C. (2005), *Predictive and Post-Audit Mass flux Estimation*, National Ground Water Association, Abstract Book of the 2005 Ground Water Summit Program, April 17-20, 2005, 222.

Tonkin, M. and Doherty, J. (2008), *Calibration-constrained Monte Carlo analysis of highly-parameterized models using subspace techniques*, Water Resources Research, 45 (W00B10).
Ye M., Pohlmann F., Chapman J.B., Pohll G.M. and Reeves D.M. (2010). *A Model-Averaging Method for Assessing Groundwater Conceptual Model Uncertainty*, Ground Water, 48 (5), 716-728.

Ye M., Pohlmann F., Chapman J.B., Pohll G.M. and Reeves D.M. (2010), *A Model-Averaging Method for Assessing Groundwater Conceptual Model Uncertainty*, Ground Water, 48 (5), 716-728.