Making Your Results Stick – Using General Likelihood Uncertainty Estimation (GLUE) In Water Quality Prediction From Coal Discard

Terry Harck

Solution[H+], Pretoria, South Africa, terry.harck@solutionhplus.com

Abstract

The GLUE methodology sets a "likelihood measure" in advance of modelling. Random realisations of uncertain values and input parameters are drawn from appropriate distributions during multiple model simulations. If the model results conform to the likelihood measure, the model is considered "behavioural", that is, one of a set of models that can replicate real-world observations.

We apply the GLUE methodology to a two-component ensemble geochemical model of a coal discard facility and one-dimensional transport in the underlying shallow groundwater system. The results indicate that our geochemical model was non-behavioural, that is, could not give an acceptable reproduction of observations, even with rather relaxed rejection criteria. This is probably due to the multitude of parameters required by the model.

This exercise illustrates that GLUE can be helpful in evaluating model results. However, random sampling can give rise to unrealistic sets of model inputs. These can generate noise in model results that may obscure whether the model is behavioural or not.

Keywords: ICARD, IMWA, MWD 2018, PHREEQC, geochemical model, coal discard, GLUE, uncertainty estimation

Introduction

Mine water management requires hydrological, hydrogeological, and geochemical models. These models are frequently "calibrated" against field and/or laboratory measurements. In general, agreement between the model and the observed measurements is considered an indicator that the model is a reasonable proxy for the system being considered. However, what constitutes "agreement" between observed and simulated data, and what does "agreement" actually mean in terms of simulating the system? Actually, the calibration process tells us nothing about whether the model is suitable for forecasting.

Uncertainty estimation in geochemical models is challenging when there are data available to evaluate model performance. The commonly-applied calibration approach can only be conditionally optimal as data errors will propagate through the model structure, which is itself a simplification of the reality generating the observations (Beven 2009). Therefore, there may be several different sets of parameters consistent with the data used for "calibration".

There are various statistical methods to address model optimisation including Bayesian and Monte Carlo methods. These are commonly applied with modern computing tools and technologies. However, these methods assume we have the correct model structure and only need to find the optimal input parameters. Beven (2009) considered the shortcomings of the calibration process and formulated the General Uncertainty Likelihood Estimation (GLUE) methodology as a means to provide more information on the suitability of a model for forecasting.

For any model there are many combinations of input parameters that may produce simulation results consistent with observations. However, there is no single "true" model. GLUE is an approach to model calibration that allows for the effects of model structural and data errors. "Model conditioning" is the descriptor used by Beven to describe a process that tries to find only those models that are acceptable from a theoretical set of all possible models.

The GLUE methodology is summarised as follows (from Beven 2009):

1. Before running the model, decide on a likelihood measure in evaluating each model run, including rejection criteria for non-behavioural model
2. Decide which model parameters and input variables are to be considered uncertain
3. Decide on prior distributions from which those uncertain parameters and variables can be sampled
4. Decide on a method of generating random realisations of models consistent with the assumptions in steps 1 and 2.

In this paper, we present the partial application of GLUE to an ensemble geochemical model of a coal discard heap. GLUE can provide insights into model structure not provided by other optimisation methods. While GLUE has been previously applied to hydrogeochemical models (eg. Zhang et al 2006), to our knowledge, it has not been applied to pyrite oxidation models.

Methods

The opportunity to apply the GLUE methodology arose while characterising a coal discard facility for environmental permitting. The facility is located in central Mpumalanga Province of South Africa. Satellite imagery (Google Earth) indicates the footprint topsoil was stripped during December 2007. Discard deposition commenced in January 2008.

Two boreholes "BH Shallow" (11 m) and "BH Deep" (30 m) monitor groundwater quality approximately 30 m downstream of the toe of the facility. The available monitoring data extends from November 2008 to June 2015. Groundwater quality monitoring indicates contamination by acid seepage from the discard in BH Shallow, while groundwater quality in BH Deep has remained relatively unimpacted (Figure 1). A marked increase in pH and decrease in sulphate concentration in early 2014 is considered anomalous.

Excluding the anomalous data, the shape of the monitored sulphate time series is reminiscent of a column breakthrough curve. The work in this paper attempts to model the groundwater quality at BH Shallow using the one-dimensional transport capability of PHREEQC (Parkhurst and Appelo 2013).

A two-component ensemble model was developed. The first component of the model simulated pore water quality due to sulphide oxidation in the coal discard using the relations of Williamson and Rimstidt (1994). Model parameters were determined from geochemical characterisation (Table 1) and physical characterisation of discard (Figure 2).

The second model component simulated one-dimensional transport in the shallow groundwater aquifer. The local hydrogeology consists of Karoo Sequence sedimentary units, including shaly sandstones and siltstones. BH Shallow samples groundwater in the weathered zone aquifer, while BH Deep is screened in the fractured zone aquifer. Groundwater velocity in the shallow aquifer, based on the data in Table 2, is 1.4 to 4.1×10^{-7} m/s.

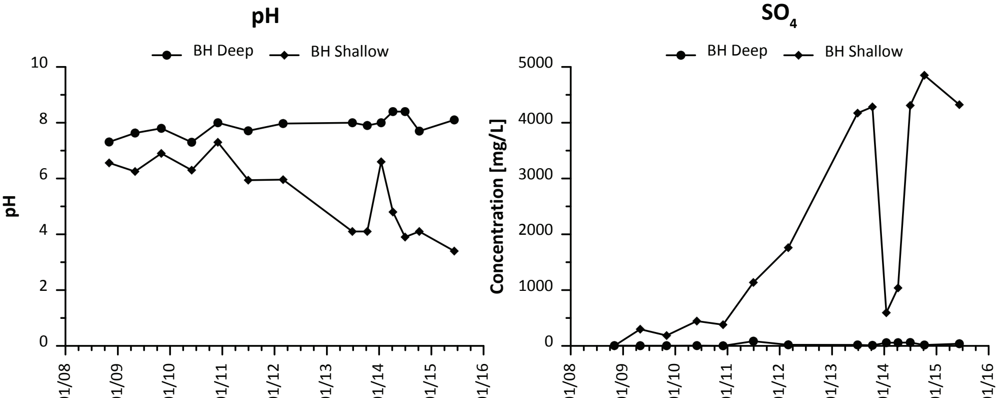

Figure 1 *Groundwater quality monitored downstream of discard.*

Table 1 Geochemical characterisation of coal discard.

Parameter	Unit	Value	Comment
Bulk density	kg/L	1.211	Mean of compacted bulk densities of three samples
Porosity	L/L	0.26	Estimated from particle size distribution
Pyrite	wt%	2.92	XRD on composite of three samples
Calcite	wt%	1.66	XRD on composite of three samples
Siderite	wt%	3.23	XRD on composite of three samples
Gypsum	wt%	--	not detected in sample
Particle density	kg/L	2.55	Calculated from mineralogy data
Porosity	L/L	0.48	Calculated from bulk density and particle size
Moisture content	wt%	48	Mean of three samples
Discard particle area	m^2/kg	30.12	Estimated from particle size distribution

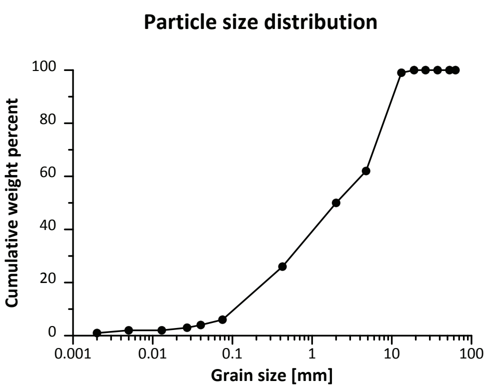

Figure 2 Particle size distribution of discard.

Table 2 Hydrogeology at coal discard facility.

Aquifer	Depth	Estimated K	Estimated porosity
Weathered zone	±10	0.044 – 0.155	3%
Fractured zone	>10	0.015 – 0.04	0.4%

Applying Step 1 of the GLUE methodology to the first model component, the model results were "acceptable" if simulated pH at any time step was less than observed pH and if simulated SO_4 was higher than observed SO_4 at any step.

For the second model component an arbitrary likelihood measure of 80% of simulation results within one standard deviation was selected. That is, if 80% of the simulation results were within the field defined by one standard deviation of the mean parameter value (pH and SO_4), and one standard deviation of the mean period between successive monitoring events, the run would be considered "successful".

Thirteen parameters/variables considered uncertain in the first model component were given prior distributions based on sample results, as indicated in Table 3. Table 4 shows eleven parameters/variables for the second model component. This comprised Steps 2 and 3 of the GLUE methodology.

For Step 4 of the methodology, one hundred repetitions of the ensemble model were run using a Python 33 module and the IPHREEQC module (Charlton and Parkhurst 2011) called by the PhreeqPy method (Müller 2013). For each repetition, the Python code randomly selected values for each parameter from the relevant distribution and wrote the values to the PHREEQC input file, which was then executed using PhreeqPy to call IPHREEQC.

Results

For the first model component, simulated pH in the discard pore water was lower than observed pH at BH Shallow. However, simulated SO_4 was only higher than observed SO_4 in 9 out of 100 simulations (Figure 3).

The "acceptable" model results from the first model component were applied in the second model component. The model was run beyond the available observations to predict future groundwater quality at BH Shallow (Figure 4). The simulated time series of pH and SO4 bear some similarity to the real-world observations. However, the simulated pH values generally do not meet the criteria for a behavioural model, while the simulated sulphate values do. The prediction suggests

Table 3 *Uncertain model input parameters and values and the assumed distributions for the first model component (discard pore water quality).*

Parameter/ variable	Unit	Distribution	Distribution parameters[A]	Comment
Initial pyrite	mol	Triangular	0.04, 0.15, 0.35	Range of XRD
Pyrite log A/V		Triangular	-0.94, -0.28, 0.1	Calculated (3 samples)
Initial pore water:				
pH	pH unit	Triangular	8.0, 8.1, 8.3	Range of leach tests
Al	mg/L	Triangular	0.021, 0.032, 0.051	Range of leach tests
Alkalinity	mg/L as $CaCO_3$	Triangular	28, 51, 64	Range of leach tests
Ca	mg/L	Triangular	41, 60, 95	Range of leach tests
Cl	mg/L	Triangular	3, 5, 6	Range of leach tests
Fe	mg/L	Triangular	0.064, 0.110, 0.199	Range of leach tests
K	mg/L	Triangular	6.0, 7.1, 8.9	Range of leach tests
Mg	mg/L	Triangular	10, 15, 25	Range of leach tests
Mn	mg/L	Triangular	0.046, 0.086, 0.158	Range of leach tests
Na	mg/L	Triangular	8, 10, 15	Range of leach tests
SO_4	mg/L	Triangular	100, 155, 264	Range of leach tests

[A] Left limit, mode, right limit

Table 4 *Uncertain model input parameters and values and the assumed distributions for the second model component (one-dimensional transport).*

Parameter/ variable	**Unit**	**Distribution**	**Distribution parameters[A]**	**Comment**
Background groundwater:				
pH	pH unit	Normal	7.6, 0.503	All distributions determined from 63 analyses of groundwater from several years of monitoring data
Al	mg/L	Lognormal	-1.98, 0.781	
Alkalinity	mg/L as $CaCO_3$	Normal	142, 50	
Ca	mg/L	Normal	22.7, 11.15	
Cl	mg/L	Normal	16.2, 6.632	
Fe	mg/L	Lognormal	-1.74, 1.500	
K	mg/L	Normal	1.63, 0.549	
Mg	mg/L	Normal	10.8, 4.492	
Mn	mg/L	Lognormal	-1.84, 1.012	
Na	mg/L	Normal	22.3, 7.516	
SO_4	mg/L	Lognormal	1.73, 0.299	

[A] Mean, standard deviation

that SO_4 concentrations will continue to increase.

Discussion

The results of this application of GLUE to a geochemical model highlight several advantages and disadvantages of the methodology.

First, the criteria for acceptability or "likelihood measure" could have been more carefully selected. Figure 4 shows that the pH and SO_4 concentration of the source were too low to correspond well with observed values. In fact, the one-dimensional transport model results skirt the lower boundary of the "acceptable" field. More rigorous criteria for the first model component may have avoided this issue.

A challenging aspect of geochemical models is the large number of variables. If likelihood measures are set for several output variables, a condition may arise, as in our coal discard example, in which one output

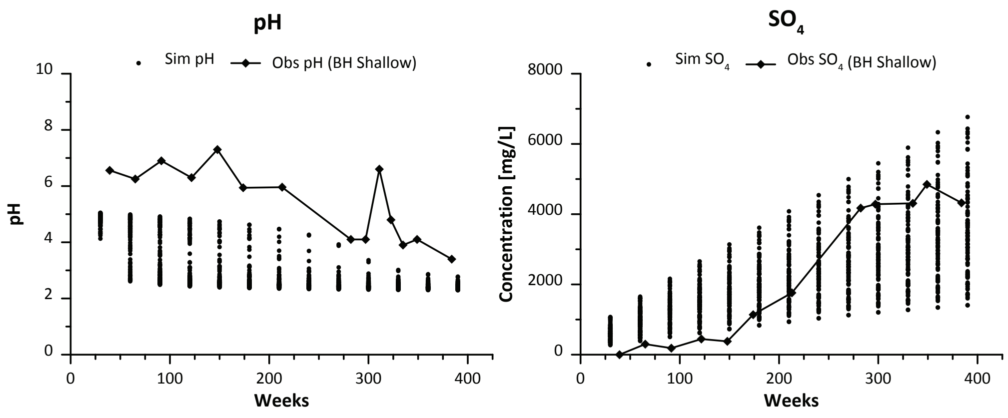

Figure 3 Results of 100 simulations of discard pore water.

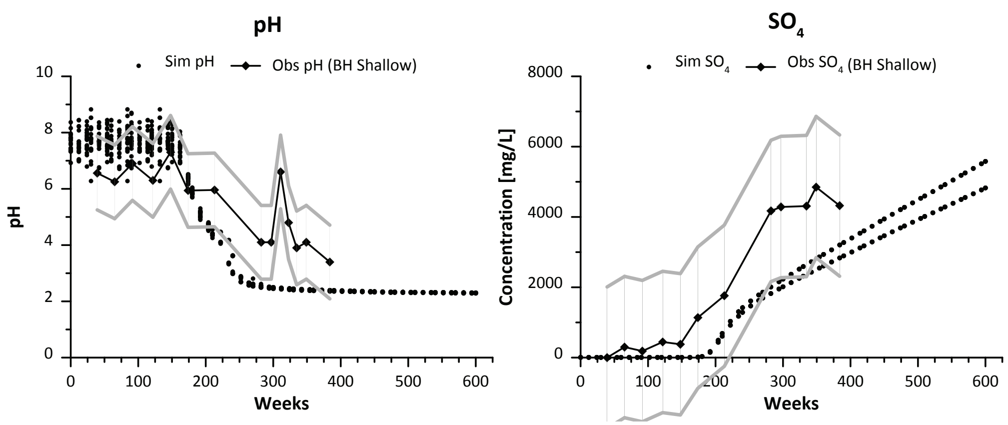

Figure 4 *Results of 100 simulations of one-dimensional transport. Grey lines indicate ±one standard deviation from observed values.*

variable indicates a behavioural model (SO_4 in this case), while another may not (pH in this case).

Second, and related to the above, the veracity of the discard pore water model component is called into question. Several possibly significant factors may have been excluded from the model, for example, a buffering mechanism resulting in a higher pH than modelled. Also, initial SO_4 concentrations in discard pore water were estimated from leach test results and did not consider evaporative concentration.

Third, the selection of uncertain parameters. Background groundwater concentrations were randomised in the one-dimensional transport model. However, the impact of this randomisation is only seen in the first 200 weeks of pH in Figure 4. Thereafter, the source concentrations appear to dominate the results. If other parameters were selected for random realisations, for example, groundwater velocity or dispersivity, the model results may have included more behavioural outcomes.

The above points may arise from the non-independence of model input variables. Therefore, random sampling from probability distributions can generate unrealistic sets of model inputs. These can generate "noise" in the model output, which may obscure useful results.

Last, development of this ensemble model was made possible with modern programming tools and techniques. However, the process was lengthy and complex as it required writing a custom code for the ensemble model. Multiple model simulations generate thousands of data points, which require extensive visualisation and interpretation.

Conclusions

The results indicate that our discard facility geochemical model was non-behavioural, that is, could not give an acceptable reproduction of observations, even with rather relaxed rejection criteria. This is probably due to the multitude of parameters required by the model, i.e. the model is probably too complex.

Selecting a likelihood measure is a critical aspect of the analysis. How closely should model results reproduce observations? Within one standard deviation of observed values seems reasonable given the uncertainties of geochemical and groundwater modelling. Rather than the wrong likelihood measure, it appears from this example, that the wrong model parameters and input variables were considered uncertain (eg. Background groundwater quality). Better results may have been obtained by considering variability in the pore water quality model.

The GLUE methodology, while somewhat cumbersome to implement, can deliver additional insight to the validity of geochemical models. Some form of filter is required for model inputs generated by random sampling to reduce "noise" in the output.

References

Beven K (2009) Environmental modelling: An uncertain future? Routledge, 310 pp

Charlton SR and Parkhurst DL (2011) Modules Based on the Geochemical Model PHREEQC for use in Scripting and Programming Languages. Computers and Geosciences, 37:1653-1663

Müller M (2013) PhreeqPy Documentation (Release 0.2). http://www.hydrocomputing.com

Parkhurst DL and Appelo CAJ (2013) User's Guide to PHREEQC (Version 3) – A computer program for speciation, batch reaction, one dimensional transport, and inverse geochemical calculations. U.S. Geological Survey Techniques and Methods, book 6, chap. A43, 497 p., available only at https://pubs.usgs.gov/tm/06/a43/

Williamson MA and Rimstidt JD (1994) The kinetics and electrochemical rate-determining step of aqueous pyrite oxidation. Geochimica et Cosmochimica Acta, 58:5443-5454

Zhang D, Beven K, and Mermoud A (2006) A comparison of nonlinear least square and GLUE for model calibration and uncertainty estimation for pesticide transport in soils. Advances in Water Research 29:1924-1933