Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 321 (2006) 286–296
Journal of Hydrology
www.elsevier.com/locate/jhydrol

The importance of diverse data types to calibrate a watershed model of the Trout Lake Basin, Northern Wisconsin, USA

Randall J. Hunt[a,*], Daniel T. Feinstein[b], Christine D. Pint[1,c], Mary P. Anderson[c]

[a] *US Geological Survey, 8505 Research Way, Middleton, WI 53562, USA*
[b] *US Geological Survey, Milwaukee, WI, USA*
[c] *University of Wisconsin-Madison, Madison, WI, USA*

Received 19 April 2005; revised 13 July 2005; accepted 1 August 2005

Abstract

As part of the USGS Water, Energy, and Biogeochemical Budgets project and the NSF Long-Term Ecological Research work, a parameter estimation code was used to calibrate a deterministic groundwater flow model of the Trout Lake Basin in northern Wisconsin. Observations included traditional calibration targets (head, lake stage, and baseflow observations) as well as unconventional targets such as groundwater flows to and from lakes, depth of a lake water plume, and time of travel. The unconventional data types were important for parameter estimation convergence and allowed the development of a more detailed parameterization capable of resolving model objectives with well-constrained parameter values. Independent estimates of groundwater inflow to lakes were most important for constraining lakebed leakance and the depth of the lake water plume was important for determining hydraulic conductivity and conceptual aquifer layering. The most important target overall, however, was a conventional regional baseflow target that led to correct distribution of flow between sub-basins and the regional system during model calibration. The use of an automated parameter estimation code: (1) facilitated the calibration process by providing a quantitative assessment of the model's ability to match disparate observed data types; and (2) allowed assessment of the influence of observed targets on the calibration process. The model calibration required the use of a 'universal' parameter estimation code in order to include all types of observations in the objective function. The methods described in this paper help address issues of watershed complexity and non-uniqueness common to deterministic watershed models.

Keywords: Watershed; Numerical model; Parameter Estimation; Calibration; Lakes

1. Introduction

In order to have confidence in a model's interpretive or predictive capability, we need to assess how well the simplified system represented in the model simulates the natural system observed in the field. This is typically done through calibration

* Corresponding author. Tel.: +1 608 828 9901; fax: +1 608 821 3817.

E-mail addresses: rjhunt@usgs.gov (R.J. Hunt), dtfeinst@usgs.gov (D.T. Feinstein), cpint@barr.com (C.D. Pint), andy@geology.wisc.edu (M.P. Anderson).

[1] Now with Barr Engineering, Minneapolis, MN, USA.

0022-1694/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.08.005

whereby simulated heads and fluxes are compared to field measurements. Inclusion of flux targets, in addition to heads, is important because it overcomes parameter correlation between hydraulic conductivity (K) and recharge, and facilitates unique calibrations (Poeter and Hill, 1997). Head and flux targets may not be sufficient, however, for constraining some model parameters such as streambed conductance (Hunt, 2002). Hence, other insight (e.g. Seibert and McDonnell, 2002), or field data in addition to heads and fluxes, can be important for improved calibration and system understanding. Previous work has demonstrated that solute distributions (Christensen et al., 1995; Anderman et al., 1996), isotopes (Krabbenhoft et al., 1990a; Poeter and Gaylord, 1990), and temperature distributions (Bravo et al., 2002) are useful in calibrating a flow model. In the Trout Lake Basin, a variety of data types have been collected that are potentially useful in model calibration, including commonly collected head and flux data. Less commonly collected data available for this basin include water isotopes (Ackerman, 1992; Walker et al., 2003), and groundwater age dating (Walker et al., in review). Travel times and groundwater and surface-water interaction have been previously simulated with three-dimensional groundwater flow models. In this paper, we describe how a combination of conventional and unconventional information can improve the calibration of a watershed-scale flow model.

2. Site description and previous modeling

The Trout Lake Basin (Fig. 1) is home to the North Temperate Lakes Long Term Ecological Research (NTL-LTER) site and the US Geological Survey's Northern Temperate Lakes Water, Energy, and Biogeochemical Budgets (WEBB) site. The system is groundwater dominated, with groundwater derived baseflow accounting for over 90% of total streamflow (USGS, unpublished data). The aquifer consists of 40–60 m of unconsolidated Pleistocene glacial deposits, mostly glacial outwash sands and gravel (Attig, 1985). Horizontal hydraulic conductivities are estimated to be about 10 m/d (Okwueze, 1983; Hunt et al., 1998), with localized zones of higher conductivity in the near-surface sediments around Sparkling and Crystal Lakes (K4, K5 in Fig. 1) based on field evidence and smaller-

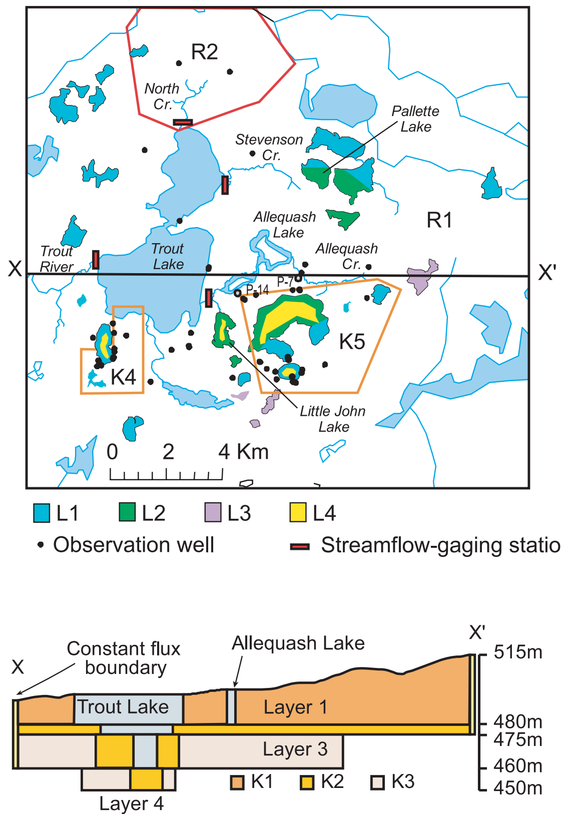

Fig. 1. Parameter zonation, model design, and target locations for the Trout Lake watershed model. K, hydraulic conductivity zones; R, recharge zones; and L, lakebed leakance zones.

scale modeling (Krabbenhoft et al., 1990b; Hunt et al., 1998; Kim et al., 1999). Vertical anisotropy in hydraulic conductivity is relatively small, with the ratio of horizontal to vertical conductivity ranging from 4:1 to 8:1 at a scale of a couple of meters (Kenoyer, 1988). The lakes occupy depressions in the glacial deposits that may penetrate more than 80% of the aquifer. Trout Lake, the largest lake in the basin with an area of 11 km^2, is drained by the Trout River (Fig. 1) and is fed by four streams. Annual precipitation averages about 79 cm/yr (Cheng, 1994) and average terrestrial groundwater recharge is estimated to be 27 cm/yr (Hunt et al., 1998), with slightly higher rates in areas composed of a higher percentage of conifer trees (Dripps et al., 2006). Annual evaporation from the lakes is about 54 cm/yr (Krabbenhoft et al., 1990; Wentz and Rose, 1991); thus, net precipitation on the lakes is about 25 cm/yr. Lakes are well connected to the groundwater system and many lakes are flow-through lakes with respect to groundwater.

The Trout Lake basin has been the focus of several modeling studies (Cheng, 1994; Hunt et al., 1998; Champion, 1998; Pint, 2002) that represent stages in the development and refinement of a regional groundwater model, which will be used in future studies to address a variety of research problems including the effects of climate change.

3. Methods

3.1. Model design

A steady-state model was constructed using MODFLOW2000 (Harbaugh et al., 2000) and MODPATH (Pollock, 1994). The three-dimensional model applied a uniform horizontal nodal spacing of 75 m and four layers (Fig. 1). The bottom three layers ranged in thickness from 5 to 15 m while the upper layer was relatively thick, with a saturated thickness between 8 and 35 m, to minimize the possibility of nodes drying during calibration and during transient simulations. A two-dimensional analytic element (AE) model using GFLOW (Haitjema, 1995a) was modified from an existing regional model of the Trout Lake area (Hunt et al., 1998) and was used to derive boundary conditions for the finite difference model according to the methodology of Hunt et al. (1998). The model included areas outside the Trout Lake basin to allow groundwater divides to move as the model was calibrated. This is important in this study because the groundwatershed and surface watershed are not aligned (Hunt et al., 1998) and the groundwatershed is estimated to be more than 40% larger than the surface watershed. The MODFLOW grid was extended beyond the groundwatershed boundaries; groundwater fluxes calculated at the boundaries of the MODFLOW grid by the AE model were distributed to the upper three layers of the finite difference model proportional to layer transmissivity and input to MODFLOW's well package. Crystalline bedrock underlies the glacial deposits and is assumed to act as an impervious bottom boundary of the model. Recharge flux was specified across the water table, which formed the upper boundary.

Thirty lakes within the Trout Lake basin or near its boundary were simulated using the LAK3 Lake Package (Merritt and Konikow, 2000), which calculates lake stages based on volumetric water budgets. Simulating lakes stages within the model is superior to specifying lake stages using constant head nodes because it helps ensure that heads are not overly specified in the immediate area of interest. Similarly, streams located within the Trout Lake basin were simulated using the Streamflow Routing Package (Prudic et al., 2004), thereby allowing calculation of stream stage and flow. For convenience, other lakes and streams distant from the area of interest were represented as head dependent flux boundaries using the River Package (McDonald and Harbaugh, 1988). The streambed sediments were assumed to have a uniform thickness of 1 m and a vertical hydraulic conductivity of 8.63 m/day; though it should be noted this parameter was relatively insensitive over the range of reasonable values in this watershed (Hunt, 2002). All aquifer hydraulic conductivity zones were assumed to have a vertical anisotropy ratio (K_x/K_z) equal to four (Kenoyer, 1988) within a given model layer. Effective porosity, used in particle tracking, was set equal to 0.29 (Krabbenhoft and Babiarz, 1992).

3.2. Calibration approach

Calibration was automated using the non-linear regression parameter estimation code UCODE (Poeter and Hill, 1998). Eleven model parameters (two recharge zones, five hydraulic conductivity zones, and four lakebed leakance zones—Fig. 1) were allowed to vary during model calibrations. UCODE adjusts the squared model residual by a weight (variance of the measurement^{-1}), resulting in dimensionless residuals. This formulation allows different target types to be evaluated in the same objective function. Taking advantage of this capability, five types of targets were used. The first two types are typically used in groundwater models and referred to as 'traditional targets'; these included water levels from lakes and wells (head targets) and the groundwater component of streamflow (baseflow targets). Five Long-Term Ecological Research (LTER) lakes had measured stages; 20 additional lake stages were estimated from topographic maps. The LTER lake targets were given a relatively high importance, thus a relatively small standard deviation in UCODE—0.5 m for lakes without surface water outflows (seepage lakes) and 0.25 m for drainage lakes based on the 17 year (1984–2001) measured range. A range of ±2 standard deviations represents the approximate 95% confidence interval around the observed value. Lake stages obtained from topographic maps were given less importance by using a standard deviation equal to 1 m to reflect the increased uncertainty of their vertical elevation and uncertainty associated with how well the stage reported on the topographic map represents a long-term average. Groundwater level measurements from 58 wells measured during July 2001—a near average period (Pint, 2002)–were used as head targets. The UCODE weight assigned to all head targets (standard deviation equal to 0.3 m) corresponds to a representative variation determined using wells with long-term data sets. The 10-year (1991–2000) mean baseflows at the four stream gaging stations were calculated using the methods of White and Sloto (1990) and used as baseflow targets. These discharge records are of relatively high quality and were given a coefficient of variation (CV) ranging from 0.02 to 0.05. A CV of 0.02 represents an approximate 95% confidence interval of ±4% around the observed mean baseflow at a given station.

In addition to traditional head and baseflow targets, three types of 'unconventional' data were also used in the parameter-estimation objective function (the sum of squares of weighted residuals) as discussed below.

- Groundwater fluxes (m^3/d) to and from 11 selected lakes in the basin—These targets were obtained using a stable-isotope mass balance (Ackerman, 1992) and water budget analysis. Groundwater inflow rates are considered to have less uncertainty than the groundwater outflow rates for the lakes; thus, inflow rates were given a CV of 0.3 and the outflow rates were given a CV of 0.7. The simulated values were obtained from the LAK3 package (Merritt and Konikow, 2000) output.
- Elevation of the top of the Big Muskellunge lake water plume—Elevation targets also were included in the objective function at two piezometer nest locations (labeled P7 and P14 in Fig. 1). Using stable isotopes of water, the interface between terrestrial-derived and lake-derived water sources was found to be approximately 11 and 16 m below the water table at P7 and P14, respectively. These targets were given a standard deviation of 0.5 m to account for the uncertainty that the plume may lie between two vertically spaced sampling points. The simulated lake plume elevations were obtained from MODPATH flowpaths that traveled from Big Muskellunge Lake to the piezometers; two MODPATH automatic termination zones were used to obtain elevation results at the piezometer nest locations.
- Time of travel to one well nest—Travel time from Big Muskellunge Lake to P7 (Fig. 1) was estimated using CFC and tritium sampling (Walker et al., in review). The approximate flowpath and sampling depth were identified using the analytic element model of Hunt et al. (1998). A standard deviation of 1 year was assigned to the target value (8 years), reflecting the expected uncertainty in aquifer porosity. The simulated result was obtained from the MODPATH travel time output for the automatic termination zone specified at P7. It should be noted that this data class would have limited utility without adequate control on lateral and vertical flowpaths. Moreover, the use of groundwater samples for measuring groundwater age has recently been questioned (e.g. Bethke and Johnson, 2002; Pint et al., 2003).

3.3. Statistical analysis of target influence

One way to assess the influence of a given observation target on the parameter estimation regression is to compare the results using all observations with the results when the observation in question is omitted (Hadi, 1992). This method assigns increased influence to observations whose deletion from the parameter-estimation process has a relatively large effect on the overall measure of model error (the residual sum of squares) and whose simulated values are relatively sensitive to small changes in parameter values. However, influence is greatest for parameter-sensitive observations that are spatially isolated, that is, not clustered with other observations within a property zone of the model. Fig. 2 shows a simple linear regression model with two observations that are difficult to match and whose simulated values are sensitive to small parameter changes. The parameter estimation results are more

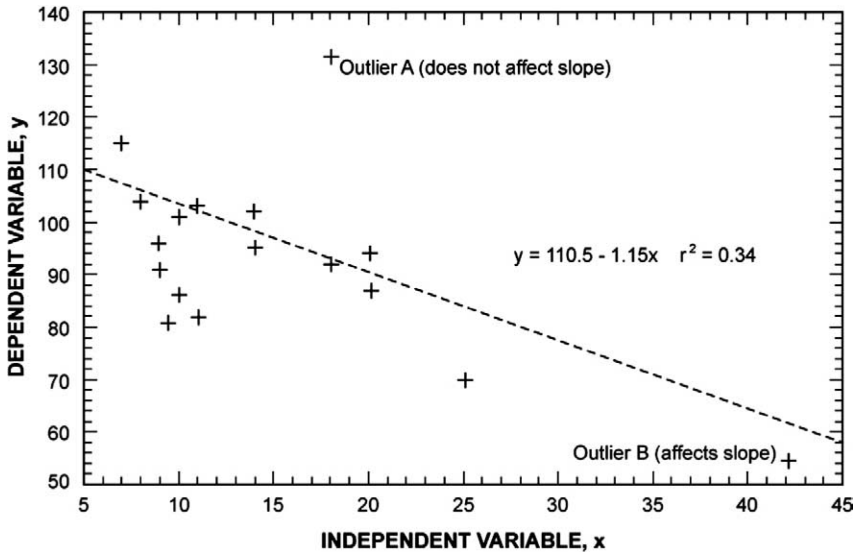

Fig. 2. Hypothetical linear regression results from Cook and Weisberg (1982) showing outliers with little influence on the regression (Outlier A) and with appreciable influence on the regression ,Outlier B).

strongly influenced by outlier B than by outlier A. Observation A is near other observations which offsets potential influence; observation B uniquely represents one part of the data range and its absence would most strongly affect the slope of the solution. To the extent that this observation is reliable, its influence is desirable (Yager, 1998).

'Cook's D' statistic ranks the influence of a single observation on the set of model parameters as a whole. The statistic measures the relative distance between the center of the confidence ellipsoids for the parameters based on the full data and the center of the confidence ellipsoids when a given observation is omitted. (Cook, 1977; Hadi, 1992). The expression for Cook's D combines a measure of the degree to which the residual sum of squares is changed by omitting the single observation with the 'leverage' that the observation exerts on the estimation process. High leverage observations are those whose simulated values are not only sensitive to perturbations in parameter values, but are also relatively isolated. The 'DFBETAS' statistic measures the influence of a single observation on the estimation of a single parameter. The two statistics can, along with parameter sensitivity and correlation, give insight into controls on the parameter estimation process. Using these additional statistics, the model structure (e.g. Yager, 1998) and observation weights can be adjusted to reduce the influence of unreliable observations. In this work we use these statistics to determine the value of competing data types. The Cook's D and DFBETAS algorithm of MODFLOW2000 (RESANP, Harbaugh et al., 2000) was modified to accept UCODE output for selected simulations in the form of (a) the weighed residuals of each target (from the *._ws UCODE file), (b) the parameter covariance matrix relating the interdependence of estimated parameters (from the *._tp UCODE file), and the sensitivity matrix relating the change in simulated results at each target to a small perturbation in each parameter value (also from the *._tp UCODE FILE). Results from the algorithm adapted to UCODE output compared well to MODFLOW2000-derived Cook's D and DFBETAS when checked using a synthetic model containing head and baseflow targets.

4. Results and discussion

The optimized model matched all five data classes well by varying the 11 parameters in the optimization. The optimized hydraulic conductivity parameters are near previously estimated values and support the presence of a layer of a slightly lower conductivity (K2 in layer 2, Fig. 1) hypothesized to be present as a result of a period of relatively fine-sand deposition associated with distant ice during late Wisconsinian glaciation (Attig, 1985). The root mean squared errors for the calibrated model were 0.56 m for head (for measured heads spanning 9.47 m), 0.14 m for lake stage of the five principal LTER lakes, and 0.90 m for 20 other lakes. Head calibration statistics are similar to those reported by Hunt et al. (1998), but are considered superior because of the larger number of head calibration targets (58 versus 31) and a wider distribution through the model domain. The simulated lake stages were within the 95% confidence interval around the measured values calculated using the expected standard deviation. Simulated base flows were within 1 percent of measured flows. The calibrated model also closely reproduced the unconventional data. Groundwater inflows to the lake were commonly within the expected uncertainty (Fig. 3), especially at the higher values (>10,000 m^3/d) where a regional model is expected to be more representative. Top elevations of the lake water plume were close (within 0.3 and 1.1 m of observed elevations) at locations 1 km and 2 km away from Big Muskellunge Lake. Finally, the simulated time of travel to P7 was within 10% of the 8-year travel time estimated using age dating. Additional information on calibration of the model is given by Pint (2002).

The inclusion of unconventional types of calibration targets was important to constrain the optimization of the watershed model. Most telling,

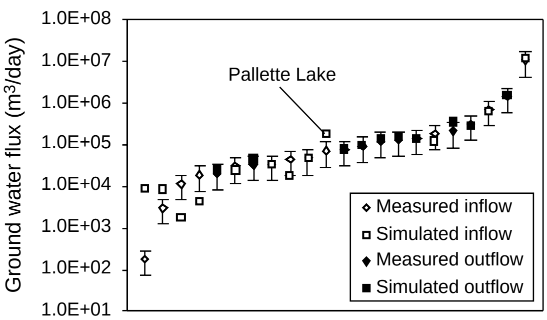

Fig. 3. Lake groundwater inflow and outflow fluxes. Ninety-five percent confidence intervals for measured inflow from Ackerman (1992) are also shown.

the parameter estimation regression would not converge if only head and flux targets were used, demonstrating that the traditional targets alone could not be used to estimate the number of parameters needed for accurate simulation of groundwater/surface-water interactions. Therefore, if the target data had consisted of only heads and flows, it would have been necessary to reduce the number of model parameters, or combine parameters and/or include prior information.

Observations with absolute values of Cook's D greater than $4/n$ (>0.036 for 112 observations) are considered globally influential (Cook and Weisberg, 1982). Of the 112 traditional and unconventional targets, a Cooks' D statistical analysis showed 10 targets had global influence on the parameter set. Of these, one half were unconventional data types (Table 1). Absolute values of DFBETAS greater than $2/n^{1/2}$ (>0.189) are considered influential with respect to a particular parameter. A DFBETAS analysis showed that an additional 10 targets had influence for a subset of the parameters (Table 1). Eight of the 20 influential DFBETAS targets were unconventional targets (bold font, Table 1). In this application, the number of globally influential observations is relatively small (10 out of 112 observations, Fig. 4(a)). Joint consideration of scaled residuals and leverage (Fig. 4(b)) shows that some observation groups lack influence (heads, topographic lake stages), while other groups with few members are especially important (baseflows, plume depth). The statistics single out observations that either have a strong relation to some portion of the model (e.g., measures of groundwater inflow at lakes) or integrate the overall model behavior (e.g., measurements of baseflow, particularly at the basin outlet). The influence of specific data types is considered in more detail below.

Of the conventional head and flow targets, three of the four baseflow targets and 8 of 58 head targets were influential. The relative importance of baseflow targets for groundwater flow modeling has been noted previously (Yager, 1998; Hunt, 2002), though the reasons are not well documented. All stream gage locations are not equally influential, however, as the DFBETAS analysis (Table 1) illustrated the importance of having conventional flow targets at multiple locations within the flow system for

Table 1
DFBETAS for influential observations arranged in order of decreasing influence

Observation name	K1	K2	K3	K4	K5	R1	R2	L1	L2	L3	L4	DFBETAS		Cook's D?	Calibration target type
												Count	Sum		
Trout river	2.5	7.0	7.5	6.7	3.5	3.7	0.6	2.4	11.5	1.9	0.3	11	47.7	Yes	baseflow
Depth@P-7	3.1	0.3	3.3	1.1	1.1	1.1	–	1.6	1.0	–	0.4	**9**	**12.9**	**Yes**	**lk plume depth**
Trout Lk stage	0.5	0.8	0.3	–	–	0.4	0.3	0.3	1.6	8.1	0.3	9	12.6	Yes	lake stage
Little John Lk gw-in	–	0.5	–	0.5	0.4	0.3	–	0.9	0.6	0.2	0.3	**8**	**3.6**	**Yes**	**lake inflow**
Allequash Ck	1.0	1.3	2.1	1.7	0.7	4.6	–	–	1.0	–	–	7	12.3	Yes	baseflow
Pallette Lk gw-in	0.7	1.5	–	0.2	–	0.4	–	0.8	–	0.7	0.9	**7**	**5.3**	**Yes**	**lake inflow**
Firefly Lk gw-in	0.2	0.4	–	–	–	0.2	–	0.2	0.2	0.2	1.5	**7**	**3.1**	**Yes**	**lake inflow**
Well CC	0.7	0.4	–	–	0.2	0.4	–	0.4	–	0.2	–	6	2.3	Yes	head
Well HS	0.7	0.4	–	–	0.3	0.5	–	0.3	–	0.2	–	6	2.3	Yes	head
Depth@P-14	–	–	0.6	0.4	0.5	–	–	0.3	–	–	–	**4**	**1.8**	**Yes**	**lk plume depth**
Day Lk gw-in	–	–	–	0.2	–	–	–	0.2	–	–	0.4	**3**	**0.8**	**No**	**lake inflow**
Diamond Lk gw-in	–	–	–	0.2	–	0.3	–	–	0.2	–	–	**3**	**0.7**	**No**	**lake inflow**
North Ck	–	–	–	–	–	–	0.4	–	0.3	–	–	2	0.6	No	base inflow
Well K59	–	–	–	0.3	0.3	–	–	–	–	–	–	2	0.6	No	head
Well P12s	0.3	–	–	–	–	0.2	–	–	–	–	–	2	0.6	No	head
Well P12d	0.3	–	–	–	–	0.2	–	–	–	–	–	2	0.5	No	head
Well K58	–	–	–	0.2	0.2	–	–	–	–	–	–	2	0.5	No	head
Crystal Lake gw-in	–	–	–	–	–	–	–	–	–	–	0.3	**1**	**0.3**	**No**	**lake inflow**
Well P5d	0.2	–	–	–	–	–	–	–	–	–	–	1	0.2	No	head
Well P7d	0.2	–	–	–	–	–	–	–	–	–	–	1	0.2	No	head

–, DFBETAS value less than threshold. Unconventional target types are highlighted.

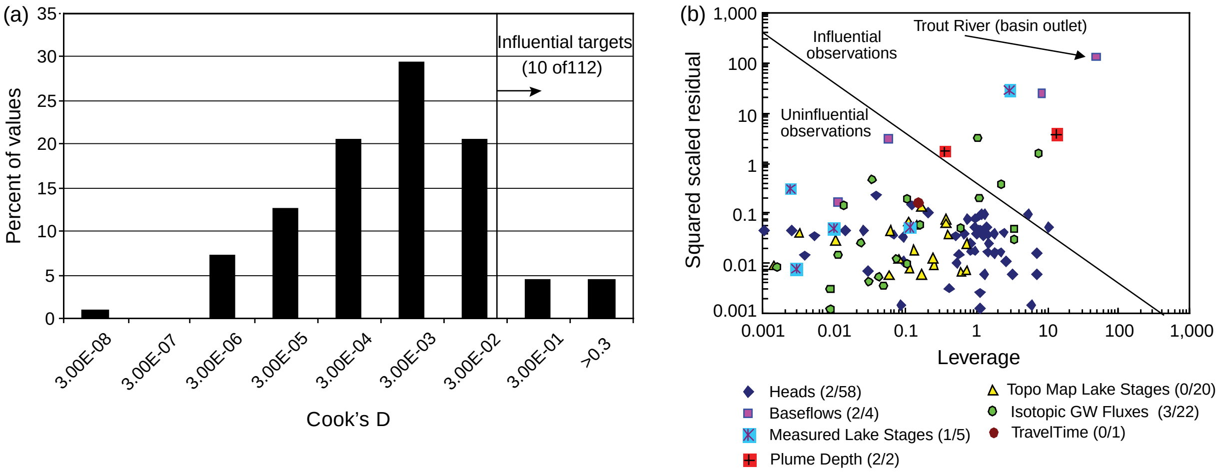

Fig. 4. (a) Histogram of target Cook's D values. Ten of 112 targets had global significance for model calibration. (b) A plot of squared scale residual versus leverage shows that 10 of 112 targets were influential due to a combination of residual and leverage. A line representing the threshold value of Cook's D = 4/*n* is also shown. Ratios in parentheses refer to Cook's D influential targets compared to total number of targets in the given type.

distributing flow between sub-basins and the larger system. The analysis underlined the importance of the flow target from the Trout River, which is the ultimate drain for the area; this target was the most influential of all, being influential for all 11 parameters estimated. The watershed outlet appears to be important not only for ensuring the correct watershed mass balance, but it is also important for quantifying partially penetrating conditions for surface water bodies upgradient of the Trout Lake regional sink. Correctly distributing flow between underflow and capture by lakes and streams is expected to be critical for subwatershed mass balance near a strong regional discharge feature such as Trout Lake. The flow at North Creek (Fig. 1), on the other hand, did not have Cook's D global influence, but was locally influential for constraining the local recharge zone that surrounded its subwatershed (as shown by its DFBETAS statistics in Table 1). Stevenson Creek, the only baseflow target not to have influence, is exceptional because it did not have a local recharge zone. This illustrates how initial model parameterization can affect observation influence.

Whereas head targets were occasionally influential, only one lake stage target (Trout Lake) had influence across the parameters. This is expected given the large areal extent of the lake (Fig. 1). Lake inflows, on the other hand, had more influence and comprised 6 of the 20 DFBETAS influential targets. Because the DFBETAS values for lakebed leakance were generally larger than for other parameters, these target types were primarily important for constraining lakebed leakance values. The existence of target types that can facilitate leakance estimates of surface water features is notable because it is often difficult to estimate this parameter using head and flux targets (Hunt, 2002).

The depth of the lake water plume was most important for constraining hydraulic conductivity (higher DFBETAS values in Table 1), though it had some influence on other parameter types as well. This increased constraint on hydraulic conductivity allowed estimation of layered zones in the aquifer. Whereas the regression could not converge in earlier models when only head and flux targets were used for calibration, the addition of lake water plume data allowed parameters for a layered conceptual model to be estimated. It is expected that this more realistic conceptualization of aquifer layering will improve simulated flowpaths, time of travel, and residence time, within the basin (Haitjema, 1995b; Bethke and Johnson, 2002).

Time of travel within the aquifer was not influential for estimating any parameter. This result was not unwelcome because there are concerns about how representative age estimates are when measured using samples collected from wells. For example, there are known difficulties in interpreting CFC dates of lake-derived recharge (Walker et al. in review), estimating representative groundwater ages given multiple sources of water (Pint et al., 2003), and mixing of waters during sampling (e.g. Bethke and Johnson, 2002). This increased uncertainty was not included in the weight used in the regression, and if it were, the observation weight and related importance of the travel time observation would be further reduced.

It is interesting to note that the importance of the unconventional data types is not easily identified using a common measure of parameter sensitivity - dimensionless scaled sensitivity (DSS-Poeter and Hill, 1998). The DSS of unconventional targets (Fig. 5) are generally of the same rank as those of conventional targets (heads, lake stage, and stream baseflows). Anderman et al. (1996) also noted that the addition of advective flowpath information did not appreciably increase the parameter sensitivity over what was obtained by using head and flow data. However, parameter sensitivity measures such as DSS simply measure the response of the simulated target values to a parameter perturbation, while the Cook's D and DFBETAS analyses evaluate changes to the regression residuals that would occur if the observation was removed from the regression. The latter approach provides information on influence rather than simple sensitivity.

Throughout this exercise, an automated parameter estimation code proved to be very useful for a model

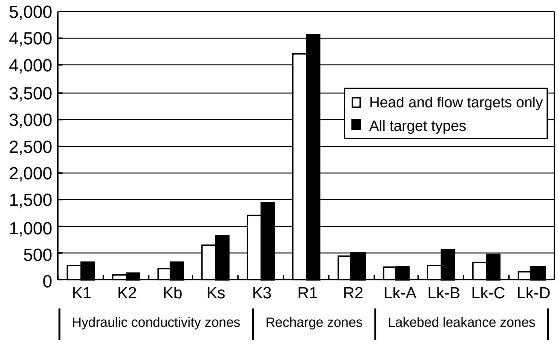

Fig. 5. Dimensionless scaled sensitivities (DSS) for traditional (head and streamflow targets only) are similar to DSS calculated using all target types (head, streamflow, lake flux, plume depth, and time of travel). This illustrates that common measures of parameter sensitivities such as DSS may not convey the influence of different target types.

where many targets using different types of data were used. Trial-and-error calibration to the head and flux data used in a 1998 modeling effort of this basin 'proved to be very difficult' (Champion, 1998) even though the targets were fewer in number and included only heads and fluxes. It should be noted that this application needed a 'universal' parameter estimation code because disparate data needed to be considered in a single objective function; other codes that have internal parameter estimation routines limit the types of data that can be included in the regression.

The use of influence statistics from the parameter estimation process also allows the modeler to focus on the quality of specific targets that drive the parameter estimation regression. In the Trout Lake model, two influential observations (Little John Lake and Pallette Lake groundwater inflows—Table 1, Figs. 1 and 3) were not previously thought to be important for model calibration. Pallette Lake is near the basin divide and well constrained by other surface water features; Little John Lake has very small groundwater flow thus was not considered to be of high interest for groundwater modeling. More importantly, the observed values for these targets are expected to have higher uncertainty than other targets because of method insensitivity for low groundwater flows (Little John Lake) and concerns about inputs to the isotopic method used to estimate groundwater flow (Pallette Lake). Given the influence of these observations and the expected high levels of uncertainty, future calibration should more critically evaluate whether these measured values are representative and their importance to the calibration should be adjusted accordingly.

5. Summary and conclusions

There are five primary findings from this work:

- Not all observed data had equal importance for model calibration. Of the 112 measured targets, 10 had global influence (Cook's D analyses) and an additional 10 targets had influence for a subset of the parameters (DFBETAS analysis). Of the conventional head and flow targets, three of the four (75%) baseflow targets and 8 of 58 head targets (12%) were influential. The most influential observation was baseflow at the basin outlet.

- Different types of data were important for calibrating a groundwater-based watershed model. In this application unconventional data types (independent estimates of groundwater inflow to lakes, and elevation of a lake isotope plume) comprised half of the observations with global (Cook's D) influence; unconventional data types comprised 40% of the observations with global and local influence (DFBETAS analysis). The regression did not converge when the unconventional data types were removed from the parameter estimation process.
- Some types of observed targets lack influence (heads, topographic lake stages) despite having a relatively large number of observations, while other groups with few members were especially important (baseflows, plume depth). Important observations are those that either have a strong relation to some portion of the model (e.g., measures of groundwater inflow at lakes) or those that integrate the overall model behavior (e.g., measurements of baseflow, particularly at the basin outlet).
- Independent estimates of groundwater inflow to lakes were most important for constraining lakebed leakance (a parameter difficult to estimate using only head and flux targets). Depth of a lake water plume was important for constraining estimates of hydraulic conductivity and a conceptual model of layering within the aquifer.
- The use of an automated parameter estimation code provided a quantitative assessment of the model's ability to simulate disparate data types. The watershed model calibration discussed here required the use of a 'universal' parameter estimation code in order to include all types of observations in the objective function. The evaluation of influence allowed a critical reassessment of observed values and their importance to the parameter estimation regression.

Deterministic groundwater flow models aimed at simulating groundwater/surface-water interactions commonly either: (1) oversimplify the watershed and, therefore, limit their ability to attain model objectives; or (2) over-parameterize the input and become subject to non-uniqueness and large uncertainty in parameter estimates. We suggest that these problems can be overcome by incorporating a rigorous calibration approach using diverse types of calibration data. This, in turn, will facilitate a better understanding of the importance of groundwater flows at the watershed scale.

Acknowledgements

Richard Yager is thanked for his help in revising the RESANP code, insight into the use of influence statistics, and review of the manuscript. Keith Halford, John Walker, and two anonymous reviewers are also thanked for review of the manuscript. This work was funded by the Wisconsin Groundwater Coordinating Council, the US Geological Survey's Northern Temperate Lakes Water, Energy, and Biogeochemical Budgets (WEBB) program, in cooperation with NSF's North Temperate Lakes LTER program (DEB-9632853).

References

Ackerman, J. A., 1992. Extending the isotope based (δ^{18}O) mass budget technique for lakes and comparison with solute based lake budgets, Masters thesis, University of Wisconsin-Madison.

Anderman, E.R., Hill, M.C., Poeter, E.P., 1996. Two-dimensional advective transport in ground-water flow parameter estimation. Ground Water 34 (6), 1001–1009.

Attig,, J.W., 1985. Pleistocene Geology of Vilas County, Wisconsin Geological and Natural History Survey Information Circular, vol. 50. Wisconsin. 32 p..

Bethke, C.M., Johnson, T.M., 2002. Paradox of Groundwater Age. Geology 30 (2), 107–110.

Bravo, H.R., Jiang, F., Hunt, R.J., 2002. Using groundwater temperature data to constrain parameter estimation of groundwater flow models that include wetlands. Water Resources Research 38 (8). doi:10.1029/2000WR000172.

Champion, G., 1998. Transient and steady-state flow models of a ground-water and lake system: Trout Lake Basin, Northern Wisconsin. Masters thesis, University of Wisconsin-Madison.

Cheng, X., 1994. *Numerical analysis of groundwater and lake systems with application to the Trout River Basin, Vilas County, Wisconsin*, Ph.D. dissertation, University of Wisconsin-Madison, 191 p.

Christensen, H., Hill, M.C., Rosbjerg, D., Jensen, K.H., 1995. Three-dimensional inverse modeling using heads and concentration at a Danish landfill. In: Wagner, B., Illangesekare, T. (Eds.), Proceedings of IAHS-IUGG XXI General Assembly, pp. 167–175 (IAHS Pub. No. 27).

Cook, R.D., 1977. Detection of influential observations in linear regression. Technometrics 19, 15–18.

Cook, R.D., Weisberg, S., 1982. Residuals and influence in regression, Monograph Stat. Appl. Probability, vol. 18. Chapman and Hall, New York, NY, USA.

Dripps, W.R., Hunt, R.J., Anderson, M.P., 2006. Estimating recharge rates with analytic element models and parameter estimation. Ground Water 44 (1).

Hadi, A.S., 1992. A new measure of overall potential influence in linear regression. Journal of Computational Analysis and Data Analysis 14, 1–27.

Haitjema, H.M., 1995a. On the residence time distribution in idealized groundwatersheds. Journal of Hydrology 172, 127–146.

Haitjema, H.M., 1995b. Analytic Element Modeling of Groundwater Flow. Academic Press, San Diego. 394 p.

Harbaugh, A. W., Banta, E. R., Hill, M. C., McDonald, M. G., 2000. MODFLOW-2000, The U.S. Geological Survey modular ground-water model - User guide to modularization concepts and the ground-water flow process. U.S. Geological Survey Open-File Report 00-92. 121 p.

Hunt, R.J., 2002. Evaluating the importance of future data collection sites using parameter estimation and analytic element groundwater flow models. Proceedings from the XIV International Conference on Computational Methods in Water Resources Conference. Delft, The Netherlands, p. 755 p. 755–762.

Hunt, R.J., Anderson, M.P., Kelson, V.A., 1998. Improving a complex finite-difference ground water flow model through the use of an analytic element screening model. Ground Water 36 (6), 1011–1017.

Kenoyer, G.J., 1988. Tracer test analysis of anisotropy in hydraulic conductivity of granular aquifers. Ground Water Monitoring Review 8 (3), 67–70.

Kim, K., Anderson, M.P., Bowser, C.J., 1999. Model calibration with multiple targets: A case study. Ground Water 37 (3), 345–351.

Krabbenhoft, D.P., Babiarz, C.L., 1992. Role of groundwater transport in aquatic mercury cycling. Water Resources Research 28 (12), 3119–3128.

Krabbenhoft, D.P., Bowser, C.J., Anderson, M.P., Valley, J.W., 1990a. Estimating groundwater exchange with lakes, 1: The stable isotope mass balance method. Water Resources Research 26 (10), 2445–2453.

Krabbenhoft, D.P., Anderson, M.P., Bowser, C.J., 1990b. Estimating groundwater exchange with lakes, 2. Calibration of a three dimensional, solute transport model to a stable isotope plume. Water Resources Research 26 (10), 2455–2462.

McDonald, M. G., Harbaugh, A. W., A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, in:. (Eds.), U.S. Geological Survey Techniques of Water-Resources Investigations Report,, 1988,. (Book 6, Ch. A1) 576 p.

Merritt, M. L., Konikow, L. F., 2000. Documentation of a computer program to simulate lake-aquifer interaction using the MODFLOW ground-water flow model and the MOC3D solute-transport model. US Geological Survey Water-Resources Investigations Report 00-4167. 146 p.

Okwueze, E., 1983. Geophysical investigation of the bedrock and groundwater-lake flow system in the Trout Lake region of Vilas County, northern Wisconsin. Ph.D. dissertation, Department of Geology and Geophysics, University of Wisconsin-Madison.

Pint, C. D., 2002. A Groundwater Flow Model of the Trout Lake Basin, Wisconsin: Calibration and Lake Capture Zone Analysis. Masters thesis, Department of Geology and Geophysics, University of Wisconsin-Madison.

Pint, C.D., Hunt, R.J., Anderson, M.P., 2003. Flowpath delineation and ground water age, Allequash Basin, Wisconsin. Ground Water 41 (7), 895–902.

Poeter, E.P., Gaylord, D.R., 1990. Influence of aquifer heterogeneity on contaminant transport at the Hanford Site. Ground Water 28 (6), 900–909.

Poeter, E.P., Hill, M.C., 1997. Inverse models: A necessary next step in ground-water flow modeling. Ground Water 35 (2), 250–260.

Poeter, E. P., Hill, M. C., 1998. Documentation of UCODE, a computer code for universal inverse modeling. U.S. Geological Survey Water-Resources Investigation Report 98-4080, 116 p.

Pollock, D. W., 1994. User's guide for MODPATH/MODPATH-PLOT, Version3: A particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference ground-water flow model, U.S. Geological Survey Open-File Report 94-464, 249 p.

Prudic, D. E., Konikow, L. F., Banta, E. R., 2004. A new streamflow-routing (SFR1) package to simulate stream-aquifer interaction with MODFLOW-2000. U.S. Geological Survey Open-File Report 2004-1042, 95 p.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resources Research 38 (11), 1241. doi:10.1029/2001WR000978.

Walker, J.F., Hunt, R.J., Bullen, T.D., Kendall, C., Krabbenhoft, D.P., 2003. Spatial and temporal variability of isotope and major ion chemistry in the Allequash Creek basin, northern Wisconsin. Ground Water 41 (7), 883–894.

Walker, J. F., Saad, D. A., Hunt R. J., in review. Factors affecting CFCs in ground-water recharge derived from Northern Wisconsin lakes. Water Resources Research.

Wentz, D. A., Rose, W. J., 1991. Hydrology of lakes Clara and Vandercook in north-central Wisconsin. U.S. Geological Survey Water Resources Investigations Report 89-4204. 24 p.

White, K. E., Sloto, R. A., 1990. Base-flow-frequency characteristics of selected Pennsylvania streams. U.S. Geological Survey Water-Resources Investigation Report 90-4160, 67 p.

Yager, R.M., 1998. Detecting influential observations in nonlinear regression modeling of groundwater flow. Water Resources Research 34 (7), 1623–1633.