Hydrology and Earth System Sciences Discussions

Increasing parameter certainty and data utility through multi-objective calibration of a spatially distributed temperature and solute model

C. Bandaragoda[1] and B. T. Neilson[2]

[1]Silver Tip Solutions, LLC., Mukilteo, WA, 98275, USA
[2]Utah Water Research Laboratory, Department of Civil and Environmental Engineering, Utah State University, Logan, UT, 84322, USA

Correspondence to: C. Bandaragoda (christina@silvertipsol.com)

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

When prediction in space and time is the goal of distributed hydrologic and instream models, the importance of basing model structure and parameterization on physical processes is fundamental. In this paper, we present a systematic approach to us-
ing various data types at spatially distributed locations to decrease parameter bounds sampled within calibration algorithms that ultimately provide information regarding the extent of individual processes represented within the model structure. Through the use of a simulation matrix, parameter sets are first locally optimized by fitting the respective data at two locations and then the best results are selected to resolve which param-
eter sets perform best at all locations, or globally. This approach is illustrated using the Two-Zone Temperature and Solute (TZTS) model for a case study in the Virgin River, Utah, USA, where temperature and solute tracer data were collected at multiple locations and zones within the river that represent the fate and transport of both heat and solute through the study reach. We found improved model performance over the
range of spatially distributed datasets relative to more common calibration approaches that use data at one location with multiple criteria objectives or at multiple locations with a single criteria objective. We also found that the global optimum is best defined by multiple spatially distributed local optima, which supports the hypothesis that there is a discrete and narrowly bounded parameter range that represents the processes
controlling dominant hydrologic responses. Further, we illustrate that the optimization process itself can be used to determine which observed responses and locations are most useful for estimating the parameters that result in a global fit to guide future data collection efforts.

1 Introduction

Typically, the calibration of models involves fitting simulations to either single or multiple variables, error measures at a single location, or combining information from multiple locations (Duan, 2003). Early calibration techniques were notorious for converging to
5 local optimal solutions and did not reliably find the global optimum (Schaake, 2003). Additionally, many existing hydrological modeling procedures do not make the best use of available information (Wagener et al., 2001). Current research on the calibration problem primarily focuses on uncertainty analysis and consideration of multiple objectives (Fu and Gomez-Hernandez, 2009; Blasone et al.,2008; Ajami et al., 2007; Duan
10 et al., 2007; and Vrugt and Robinson, 2007). Rather than selecting a single preferred parameter set, equifinality of models recognizes that there may be no single, correct set of parameter values for a given model and that different parameter sets may give acceptable model performance (Beven, 2001). All calibration algorithms have basic design requirements, including the selection of calibration parameters, objectives, and
15 the a priori space within which to search for an optimum solution or set of solutions. The measure of "acceptable" and "optimal" is left to the design of the optimization problem, the model application, and the modeler.

In this paper, we consider a global optimum as the solution where there is acceptable tradeoff between fitting the model at all locations there is data available; this can be ac-
20 complished by using a range of multiple local optima defined by a narrowly bounded global optima. Since a model is not an exact representation of reality, and observed data used to verify it is not perfect, the global optima of a physical model distributed in space and in time may be an unrealistic goal. However, a practical goal is to resolve the multiple local optima which simultaneously perform well on a local scale using all
25 available information to narrowly bound the region surrounding the theoretical global optimum. Performing well locally and globally, or glocalization, can be used to define an optimum in model calibration which bridges scales between local and global performance. A systematic approach to using various data types at spatially distributed

locations to decrease parameter bounds sampled within optimization algorithms is relevant to instream and hydrologic models ranging in applications from the stream reach to the watershed scale. In this application, our aim is to ascertain a narrow range of parameters that perform well based on multiple data types distributed throughout the spatial extent of the instream temperature and solute modeling application.

The Two-Zone Temperature and Solute (TZTS) model (Neilson et al., 2010a,b) was developed to capture the dominant instream processes associated with heat and solute fate and transport. The TZTS model separates transient storage (Bencala and Walters, 1983) into two zones, (1) dead zones or the surface transient storage (STS) zone that represents the eddies, recirculating zones, and side pockets of water and (2) subsurface or hyporheic transient storage (HTS) zone that represents the flow into or out of the stream substrate. As discussed in Neilson et al. (2010a), sources and sinks of heat include fluxes across the air-water interface, bed conduction, conduction between the bed and deeper ground substrate, HTS exchange, and STS exchange. Solute mass is primarily influenced by HTS and STS exchange (Neilson et al., 2010b). To account for each of these fluxes, the TZTS model calculates energy and mass balances on the main channel, the STS zone, and the HTS zone for each reach or control volume (further details are included in Neilson et al., 2010a,b).

To support TZTS model applications, simultaneous data collection of temperature and solute tracer data (referred to more simply as tracer data throughout the rest of the paper) in the main channel and storage zones distributed laterally at one location and longitudinally along a river segment, has created datasets that can be used to address the high dimensional problems associated with predicting heat and solute movement within streams and rivers. In recent studies beginning with Neilson et al. (2010a,b), the TZTS model was calibrated using the Multi-Objective Shuffled Complex Evolution Metropolis algorithm (MOSCEM; see Vrugt et al., 2003a for algorithm description) and used to predict solute concentrations and temperatures in the Virgin River, Utah, USA, in two storage zones at two different locations within the study reach. Using temperature and tracer observations at two different sites illustrated that using more spatially

Increasing parameter certainty and data utility

C. Bandaragoda and
B. T. Neilson

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

distributed information and two different environmental tracers (temperature and solute) in the optimization improves the overall performance of the model. Bingham et al. (2010) further illustrated how thermal imagery of the river system could be used to physically estimate the surface areas associated with the STS and, therefore, reduce the number of parameters being estimated in the calibration and consequently, reduce parameter uncertainty. Regardless, each of these studies found that even with the use of multi-objective calibration, many optimal parameter sets were indistinguishable based on the objective function, fairly broad parameter ranges resulted, and parameter uncertainty was still a concern.

In this paper, we address these issues by presenting a systematic approach to using various data types at spatially distributed locations to decrease parameter bounds sampled within optimization algorithms in the context of a case study. We developed a simulation matrix of data types and sites that is used first to locally optimize parameter sets by fitting the respective main channel data at two locations using both single and multi-objective calibration algorithms. These results were then used to resolve which parameter sets perform best at individual locations (distributed laterally and longitudinally) or have the best local fit, and which calibration data sets result in the best global fit. Our hypothesis is that there is a narrowly bounded parameter range that best represents the hydrologic processes controlling the system, which can be determined by using key data sets as multiple calibration objectives. Our goal in this paper is to select the key data sets and determine how useful they are at narrowing the parameter bounds to increase certainty about the model predictions and their role in calibration.

2 Study area and data

A highly managed portion of the Virgin River, Utah, USA (Fig. 1), is considered impaired due to elevated temperatures that have adversely affected two endangered fish species (Virgin River Chub – Gila seminuda and woundfin – Plagopterus argentissimus) and other native fishes unique to this river system. As described in detail in Bingham et

al. (2010), the 11.94 km study reach of the Virgin River used in this application was divided into two main sections on the basis of bed slope and stream substrate distribution identified from a previous mapping effort (Fig. 1). There are several inflows along the study reach; the largest is the return flow from Quail Creek Reservoir (0.6 $m^3 s^{-1}$).

To support the TZTS model population, calibration, and model testing, various data types were collected from 22–25 June 2007. The instream flow during the study period was found to be an average of 1.06 $m^3 s^{-1}$ at Site 1 and 1.96 $m^3 s^{-1}$ at Site 3. Information regarding lateral inflow rates and temperatures were also collected during the study. Groundwater exchanges were set according to Herbert (1995) with a total gain 0.17 $m^3 s^{-1}$ over the entire reach. Weather information (air temperature, solar radiation, wind speed, and relative humidity) was gathered at Site 1 using a Davis Wireless Vantage Pro (Hayward, CA) weather station to provide the data necessary to calculate the atmospheric fluxes. Similar to Neilson et al. (2010a,b), solute and temperature information were collected at Site 2 and Site 3 to support model calibration and testing. The data included solute tracer experiments resulting in main channel and STS concentrations at both Site 2 and Site 3. Simultaneous temperatures at Site 2 and Site 3 were also collected in the main channel (sensor 2), STS (sensor 1 and 3), and HTS (sensor 4, 5, and 6) (Fig. 2). The temperature sensors were Hobo[®] Water Temp ProV1 (Onset Corporation, Bourne, MA) with a ± 0.2 ˚C accuracy and resolution of 0.02 ˚C.

As with Neilson et al. (2010b), a 180 g instantaneous pulse of fluorescent Rhodamine WT dye was injected at 02:00:00 on 6 June 2007, at the head of a riffle just upstream of Site 1. A Self-Contained Underwater Fluorescence Apparatus (SCUFA) (Turner Designs, Sunnyvale, CA) was deployed in the main flow of the channel at both Site 2 and Site 3. Measurements were taken in situ every ten seconds for approximately seven hours at Site 2 and 6 h at Site 3. Grab samples were also collected at both Site 2 and 3 near the SCUFA to provide an independent measure in the main channel and in two representative STS locations. The grab samples were kept cool, stored in the dark in amber bottles with PTFE caps, and analyzed using a Turner Model 450 fluorometer (Turner Designs, Sunnyvale, CA). As discussed in Neilson et al. (2010b), loss

of Rhodamine WT due to sorption to streambed sediments (mineral and organic) was not a concern in this study because the organic matter content in the bed sediments was extremely low (averaging 0.05% at four sampling locations). Additionally, a recent sorption study within this portion of the Virgin River (Bingham, 2009) provided average K_d values of 1.5 mL/g, which is low based on other Rhodamine WT sorption studies (Bencala et al., 1983; Everts and Kanwar, 1994; Lin et al., 2003; Shiau et al., 1993).

3   Methods

3.1   Simulation matrix

With the overall goal of iteratively reducing the size of the global search space while simultaneously investigating the information content within the available data types, we established a simulation matrix (Table 1) to test the use of the most commonly collected main channel data sets used in calibration of instream temperature and solute models. Each row and column denotes a data type that represents both heat and tracer fate and transport at Site 2 and 3 along the study reach and within different zones at each location.

This matrix represents all possible combinations of single and two-objective calibrations that use the available main channel temperature and tracer data. The calibration tests were Tests 1 through 4, which are single-objective calibrations using main channel temperature and tracer at Site 2 and Site 3, and Tests 5 through 10 which are various combinations of data resulting in two-objective optimizations. The latter two-objective tests include the following combinations: main channel temperatures at Site 2 and Site 3 (Test 5), main channel tracer observations at Site 2 and Site 3 (Test 6), main channel temperature and tracer observations at Site 2 (Test 7), main channel temperature at Site 3 and tracer observations at Site 2 (Test 8), main channel temperature at Site 2 and tracer observations at Site 3 (Test 9), and main channel temperature and tracer observation at Site 3 (Test 10).

3.2 Calibration technique

Similar to previous TZTS calibration studies (Neilson et al., 2010a,b; Bingham et al., 2010), SCEM (for single-objective calibration) and MOSCEM (for multi-objective calibration) (Vrugt et al., 2003a,b) were the optimization algorithms utilized to evaluate
each model test. To ensure that we were adequately searching the parameter space, MOSCEM was run with a random sample of 300 parameter sets that evolved using two complexes for a total of 3000 model runs for each of the ten tests. We experimented with a range of sample and complex sizes (e.g., 400 samples and four complexes with a total of 10 000 model runs) and we found that an increase in the simulations and
complexes did not significantly improve calibration results.

In this application, measurements within the STS and HTS are withheld during calibration and used to assess the predictive capacity of these components as "ungauged" model outputs. As will be described in detail later, the STS data were used to assist in selecting globally acceptable parameter sets. The HTS data were reserved for corroboration and testing of the model calibration. Since temperature and tracer data in the main channel are the most commonly collected data sets, we needed to further understand whether model calibration to main channel temperature and tracer data results in realistic and representative STS and HTS predictions. Likewise, little was known about how single-objective model calibration at individual sites controlled the resulting parameterization at other site locations and for other data types. In addition to investigating how to narrow the optimization parameter space, our methods are designed to test how a priori choices in study and project design, as well as data availability, may affect the model calibration and resulting simulation performance.

3.3 Model parameters

The a priori uniform distribution of the feasible parameter space was determined primarily based on earlier work (Neilson et al., 2010; Bingham et al., 2010) (Table 2). The calibration parameters include: STS fraction of the total channel width

(β), cross-sectional area of the STS (m^2) ($A_{cs,\,STS}$), exchange between the main channel and the STS (m^2 d^{-1}) (α_{STS}), HTS advective transport coefficient (m^3 d^{-1}) (Q_{HTS}), and HTS depth (Y_{HTS}) for each of the two sections within the study reach (resulting in 10 parameters). The depth of the ground layer below the HTS (Y_{gr}) was also estimated, but was represented by one value for both sections and became the eleventh calibration parameter. The total width of the main channel (B_{tot}) and the Manning's roughness coefficient (n) were set based on the results of Bingham et al. (2010). The longitudinal dispersion (D) coefficient was set based on the methods described within Neilson et al. (2010a).

3.4 Calibration objectives

To evaluate local and global model performance, various types of statistical measures were investigated. Each of the ten tests shown in Table 1 were run using different statistical objectives including bias, Nash-Sutcliffe Efficiency (E), log error, and root-mean square error. Similar to Neilson et al. (2010a,b), we found that E (Eq. 1; Nash and Sutcliffe, 1970) provided the most consistent calibration results and we used this objective function throughout the remainder of the study and to quantify all local calibrations.

$$E = 1 - \frac{\sum\limits_{t=1}^{N} \left(T_o^t - T_m^t \right)^2}{\sum\limits_{t=1}^{N} \left(T_o^t - \overline{T}_o \right)^2} \tag{1}$$

where: T_o^t = observations, T_m^t = modeled simulations (at time t), and $\overline{T}_o$ is the mean of the observations. When used in calibration, the algorithm minimizes the result of $1 - E$, since the bounds of E are $[1, -\infty]$. The normalization of the difference in error by the difference between the observed and the mean of the observed, allows comparison of results when the observations at different locations have different scales of variability, as is the case of temperature and tracer information.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

To achieve an acceptable globally optimal calibration, we must consider the need to match all local data regardless of the differences in optimal parameter sets associated with individual single-objective optimizations. In this study, our local problem is that any unacceptable local optimization (i.e., the model does not represent one observed time series well) signifies a model failure to adequately reproduce the dominant processes affecting the heat or solute response within the modeled reach. Our global problem is that we have ten time series distributed in space, six temperature and four tracer datasets, with 11 different parameters that need to be estimated based on matching both the observed temperature and tracer data in all zones and at all locations. The six locations for temperature calibration or comparisons based on available data include: Site 2 main channel ($E_{MC2\ Temp}$), STS ($E_{STS2\ Temp}$), HTS ($E_{HTS2\ Temp}$); and, Site 3 main channel ($E_{MC3\ Temp}$), STS ($E_{STS3\ Temp}$), HTS ($E_{HTS3\ Temp}$). Note that each observed time series used in these E values for the STS and HTS consists of the average of temperatures observed within the two representative STS zones and the most representative HTS time series, respectively. The appropriate HTS time series was determined based on the calibrated Y_{HTS} values: when $Y_{HTS} < 3\,cm$, the $3\,cm$ HTS data were used, when $3\,cm < Y_{HTS} < 9\,cm$, an average of the 3 and $9\,cm$ HTS time series were used, when $9\,cm < Y_{HTS} < 20\,cm$, an average of the 9 and $20\,cm$ HTS time series were used; and when $Y_{HTS} > 20\,cm$, the $20\,cm$ HTS time series was used. The four local tracer data locations used for comparison or calibration include: Site 2 main channel ($E_{MC2\ Tr}$), STS ($E_{STS2\ Tr}$); and, Site 3 main channel ($E_{MC3\ Tr}$), STS ($E_{STS3\ Tr}$). The observed STS time series used in these calibrations are the average concentrations observed within the two representative STS zones.

The first step in our calibration method was to populate the simulation matrix (Table 1) based on available observations. We then identified the a priori parameter search bounds and the most appropriate statistical objective function, E. To compare the global calibration results (i.e., matching the observations at all ten locations) for each of the tests within the simulation matrix (Table 1), we then calculated the arithmetic average (AE) of various combinations of local E values (Eq. 2).

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►I

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

$$AE = \frac{1}{n}\sum_{i=1}^{n} E_i \qquad\qquad (2)$$

An AE that used only surface data (AE_S) was first defined and included the local E values for all tracer and temperature data collected in the main channel and STS, but did not include the HTS information.

3.5 Narrowing search bounds

Using the initial a priori bounds (Table 2), we defined Level 1 results as calibrated parameter sets from the single-objective optimizations (Tests 1–4). Level 2 results represent the parameter sets from the two-objective optimizations with these same a priori bounds (Tests 5–10). Level 1 and 2 results were then used to define more narrow search bounds for each parameter by selecting the pareto optimal parameter sets that met both local ($E > 0.8$ for all calibration data sets) and global ($AE_s > 0.7$) criteria we set within this study. The parameter sets which met both criteria were used to establish new, more focused parameter search ranges. All simulations in Table 1 were repeated using these narrower search bounds. Level 3 results represent the new parameter sets from all single-objective optimizations (Tests 1–4) and Level 4 represent the new two-objective simulation (Tests 5–10) results given the narrowed search range.

The last step was using Level 3 and 4 results to further test the model calibration. Similar to the new AE_s, a new AE_{all} was calculated for these simulations that used all of the data including the temperatures within the HTS. Together the AE_s and AE_{all} measures were used to summarize the spatially aggregated performance of model predictions of temperature and tracer at multiple locations, and determine the ability to predict the HTS temperatures if only surface data were available. This gives an indication of the added utility of collecting subsurface data and whether the model can be calibrated sufficiently in this watershed using only surface data collected at multiple locations and within different zones.

By comparing Levels 1 and 2, a wide parameter search space, to Levels 3 and 4, a narrow parameter search space, we can investigate the importance of a priori parameterization. In comparing Levels 1 and 3, single-objective calibrations, to Levels 2 and 4, two-objective calibration, we will learn information about how best to utilize available calibration algorithms and various types of spatially distributed information simultaneously.

4   Results

4.1   Level 1

The AE_S, and individual E values for each calibration location and data type are presented in Table 3 for the calibrations from the simulation matrix (Table 1). The ten rows correspond to model outputs by test and shaded boxes represent the data used from that location for calibration. All other observations were used as validation data sets.

Level 1 results (Table 3) provide initial information regarding how optimization at single locations can impact the model performance at ungauged locations. Of Tests 1–4, no tests with the main channel tracer data at Site 2 or Site 3 as the objective had results that met the selection criteria of $AE_S > 0.7$, with the best results $^2AE_s = 0.65$ and $^2E_{MC3,\ Temp} = 0.95$ and $^2AE_S = 0.6$ (preceding superscripts indicate Test numbers). Although the E for each of these tests meet the criteria of $E > 0.8$ and the calibration did well at fitting the dataset used as the objective, the calibration was not acceptable at other locations.

Figures 3 and 4 show the highest performing single-objective Level 1 results (Test 2) for each of the ten total data locations. The observed temperature and tracer data at Site 2 and Site 3 are shown as black circles (Figs. 3 and 4), and the E values for each location are shown within each subplot. The predicted values are shown in grey, and in this case there is a single line since a single objective calibration results in a single optimal parameter set. The calibrated Y_{HTS} value is also shown with the HTS subplots

(Fig. 3d and e) since this value is used to determine the most representative HTS temperature time series for calculating E_{HTS}. Although the temperature results seem to fit the observations well (Fig. 3), the tracer results (Fig. 4) show how the model optimized to temperature at Site 3 ($^2E_{MC3,\ Temp} = 0.95$) is not able to capture the timing and magnitude of the tracer pulse. This may be in part due to fixing the Manning's n parameter in calibration.

4.2 Level 2

Level 2 simulations were used to determine which parameter sets resulting from the two-objective optimizations (Tests 5–10) converge to the established criteria of $E > 0.8$ for all calibration data sets and $AE_S > 0.7$ (Table 3). The E values reported for the two-objective optimizations are based on the parameter set that represents the best compromise solution or the pareto solution with the smallest Euclidean distance from the origin. The best results are from Test 7 with values of $^7E_{MC2,\ Tr} = 0.94$, $^7E_{MC2,\ Temp} = 0.91$, and $AE_S > 0.81$. Figures 5 and 6 present Test 7 results where the uncertainty bounds resulting from pareto optimal parameter sets are shown. The uncertainty in the temperature predictions are less at Site 2 (Fig. 5) and there is a much better fit in terms of timing of the tracer curve at Site 2 (Fig. 6), but there are still relatively large bounds. It should also be noted that this calibration does not capture the tail of the tracer curve at Site 2, which is critical to understand the transient storage within the study reach (Bencala and Walters, 1983). Similar to what Neilson et al. (2010a) found, comparing Level 1 and 2 results (Table 3) illustrates the relative benefit of using two-objective optimization compared to single-objective optimizations. For Tests 5–10, Tests 6 and 10 did not meet the local criteria of $E > 0.8$ with tracer data used as a calibration objective, although Test 6 did meet the global criteria (Table 3).

Since Test 7 met the local and global criteria, all the acceptable parameter sets (i.e., the pareto optimal parameter sets that also met the local and global criteria) from this test were used to define the narrowed upper and lower bounds for a new round of calibrations using the simulation matrix (Table 1). The narrowed minimum and maximum

parameter range (Table 4) represent a parameter range reduction with a high of 67% for the $A_{cs, STS}$ in Sect. 1 and the least reduction of 4% for the β in Sect. 2. Comparing between sections, Sect. 1 had an average of 40% reduction in bounds while Sect. 2 had an average of 17% reduction. To visually compare the a priori parameter range and the narrowed parameter range derived from Test 7 results, each of the 11 calibrated parameters were normalized or scaled between the lower bound, 0, and the upper bound, 1 (Fig. 7). The thick black solid lines represents the parameter bounds if all pareto rank one sets resulting from the Test 7 calibrations are considered. The grey shaded area represents the narrowed parameter bounds for parameter sets that resulted in meeting both local and global criteria from the Test 7 optimizations.

4.3 Level 3 and Level 4

Similar to Level 1 results, Tests 1 through 4 all converged to $E > 0.9$ for the data used in calibration during the Level 3 calibrations (Table 5). However, model performance at other locations was poor with the exception of Test 3, which had better AE results than Level 1: $^{3}AE_{S} = 0.76$, and $^{3}AE_{all} = 0.62$. While these results are promising, it is important to note that only the tracer at Site 2 (the calibration objective) fit the observations well (not shown here for brevity).

Level 4 had improved results when compared to Levels 1–3. The AE_{all} and AE_{s} values increased for most tests (Tables 3 and 5), and the maximum value increased to 0.78 and 0.9 for AE_{all} and AE_{s}, respectively. Although Test 6 met the global and local criteria, the temperature simulations at Site 2 overestimated the high temperatures and underestimated the low temperatures by approximately $3\,^{\circ}C$ in the main channel, STS, and HTS zones. Figures 8 and 9 show the best overall result for Level 4 temperature and tracer predictions, Test 9: $^{9}AE_{S} = 0.9$ and $^{9}AE_{all} = 0.78$. Not only are the temperature predictions more representative, but the tracer responses are generally captured better in the tail of the tracer curves. As with the Level 2 calibrations, both temperature and tracer objectives at different locations seem to provide the information necessary to achieve an acceptable global calibration.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Figure 10 shows the parameter ranges resulting from the Test 9 optimization that met the local and global criteria and the bounds of all the pareto optimal sets. The dashed line shows the narrowed parameter range within the original a priori search range (normalized here [0,1]). The thick black line is the bounds of the pareto optimal parameter sets. The grey area is the parameter variability given the parameter sets which meet both local and global performance criteria.

5 Discussion

Comparing the results of the simulation matrix calibrations when using only the main channel temperatures or tracer concentrations as an objective (Test 1–4, Table 3), we see how the choice of a calibration objective effects the global performance of the model by comparing the AE_s and AE_{all} values. In general, the best temperature and tracer main channel result is from a single objective optimization of that parameter at that location, but the corresponding model results are generally inappropriate at other locations. Our results also show that when a temperature objective at one location results in reasonable predictions, the temperature at the other location will also be reasonable. However, this is not necessarily the case when using tracer data in single objective optimizations in this study.

The best Level 2 local results at Site 2 and Site 3 for tracer are $^8E_{MC2,\ Tr} = 0.98$ and $^6E_{MC3,\ Tr} = 0.99$ and for temperature are $^5E_{MC2,\ Temp} = 0.96$ and $^{10}E_{MC3,\ Temp} = 0.95$ (Table 3). It is interesting that the best fit for tracer at Site 3 uses tracer information at both Site 2 and 3 (Test 6), but the best fit at Site 2 uses tracer information at Site 2 and temperature information at Site 3 (Test 8). In this case, the tradeoff between solute at two sites is greater than the tradeoff between solute and temperature. For temperature, the best fit at Site 2 uses temperature data at both Site 2 and Site 3 (Test 5). However, the best temperature fit at Site 3 uses temperature and tracer data at Site 3 (Test 10). It should be noted that when temperature data at Site 3 and tracer data at Site 2 were used (Test 8), $^8E_{MC2,\ Temp} = 0.94$ which is not significantly different than Test 10. Having

both main channel temperature and tracer data at two different longitudinal locations provided more information about the system than just one data type.

While these local results give insight into the utility of calibration data, it is important to acknowledge how each of these calibrations perform globally. Given a broad parameter search range (Level 2), Test 7 had the best overall results with $AE_s = 0.81$ and provided some corroboration of the model representing the dominant processes through an $AE_{all} = 0.75$. Most Level 2 AE_s and AE_{all} values were higher than Level 1 values. This is consistent with the findings of Neilson et al. (2010a) where they found two-objective calibrations performed better at locations not used in model calibration than did single objective calibrations. While Test 7 had the best global value, the individual results were not nearly as good as the best fits at each location for each data type. It did, however, provide the necessary information to narrow the search bounds for the Level 3 and 4 simulations.

With this initial understanding of the importance of single versus two-objective calibration and various data types in model calibration to narrow the search space, Level 3 and 4 results provide a more complete picture of how the system is functioning (Table 5). The majority of the Level 3 single-objective optimizations have AE_s and AE_{all} values that are higher than those in the Level 1 simulations. The actual E values for the location being used in the calibration are also higher with the exception of Test 1. This suggests that the more narrow search range was appropriate.

The best Level 4 results at Site 2 and Site 3 for tracer are $^8E_{MC2, Tr} = {}^6E_{MC2, Tr} = 0.98$ and $^{10}E_{MC3, Tr} = 0.99$ and for temperature are $^7E_{MC2, Temp} = 0.95$ and both $^5E_{MC3, Temp} = 0.94$ and $^{10}E_{MC3, Temp} = 0.94$ (Table 5). The best tracer results at Site 2 are consistent with the Level 2 results where tracer information at Site 2 and temperature information at Site 3 is most appropriate (Test 8). The best Site 3 tracer results now suggest that both temperature and tracer data at Site 3 (Test 10) is better than tracer data at Site 2 and Site 3 (Test 6). Within the narrow search bounds, the best tracer results rely on temperature information at some location.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

For Level 4 temperature results, the best fit at Site 2 uses temperature and tracer data at Site 2 (Test 7), however the Test 5 results are quite similar. The best temperature fit at Site 3 still uses temperature and tracer data at Site 3 (Test 10), but the results for Test 5 (which uses Site 2 and 3 temperatures) has the same E. These results demonstrate the need to use both temperature and solute data in two-objective TZTS calibration. The Level 4 results also showed a marked improvement in most AE_s and AE_{all} values from Level 1–3 simulations. This improvement can be related to the increased parameter certainty when comparing Level 2, Test 7 (Fig. 7) with Level 4, Test 9 (Fig. 10). These figures show the usefulness of using more information, or local data, to define a narrow range bounding the global optimum. They also highlight the importance of multi-objective calibrations to capture the spatial heterogeneity within streams and rivers and the need to determine the appropriate optimization parameter ranges.

To further incorporate important processes and continue advancing our predictive capabilities, there is a need for a connected cycle of inquiry that includes model development and refinement, identification of data types and scales of measurement required to support modeling, and establishing the most effective approach for calibration based on the application of interest. Inclusion of all available site specific data in model calibration assists these efforts by providing information that decreases the number and range of parameters, provides information about model certainty, can guide the incorporation of processes missing in the conceptual model, and will assist in prioritization of future supporting data collection efforts. Future work varying additional parameters or holding others constant may improve overall results. Expanding the simulation matrix to examine the use of STS temperature and tracer observations in the calibration would further highlight the utility of these datasets.

6   Conclusions

With the overall goal of iteratively reducing the size of the global search space while simultaneously investigating the information content within the available data types, we established a simulation matrix to test the use of the most commonly collected main channel data sets used for model calibration of instream temperature and solute models. This systematic approach to using multiple types of distributed information allowed us to examine the application of both single and multi-objective optimization algorithms to the TZTS model using both temperature and solute data available within the main channel and transient storage zones (STS and HTS).

In the context of a case study in the Virgin River, Utah, USA, our global problem was that we have ten time series distributed in space. Our local problem was that any unacceptable local optimization (i.e., the model does not represent one observed time series well) signified a failure to adequately reproduce the dominant processes affecting both the heat and solute response at that location. Using data representing the effects of both main channel and transient storage processes, we found that two-objective calibrations consistently performed better at all locations where data were available within the study reach for corroboration than did single objective calibrations. However, we also found that neither single objective results nor multiple objective pareto optimal results alone were able to produce acceptable global calibrations or appropriately match data at both locations and within each zone. This led to using parameter sets from initial calibration efforts (Level 1 and 2) to narrow parameter ranges resulting in a reduction of bounds in the upstream section of the river by an average of 40%, and in the downstream section by an average of 17%. In doing this, Level 3 and 4 calibrations, which used these narrow parameter bounds, led to improved predictions of instream temperatures and tracer concentrations at multiple locations and zones in the study area. This global fit resulted a better representation of the dominant processes.

Another key finding was that, in general, using both main channel temperature and solute data in calibration provided better global results. Therefore, we suggest that both

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

data types be collected at different locations, for example, solute at one calibration site and temperature at another. Based on the results of this study, and the need to use resources associated with data collection more efficiently, future data collection could focus on collecting a single tracer observation time series in the main channel with temperatures are collected simultaneously in multiple locations and zones to be used in model calibration and testing.

References

Ajami, N. K., Duan, Q., and Sorooshian, S.: An Integrated Hydrologic Bayesian Multi-Model Combination Framework: Confronting Input, Parameter and Model Structural Uncertainty in Hydrologic Prediction, Water Resour. Res., 43, W01403, doi:10.1029/2005WR004745, 2007.

Bencala, K. E. and Walters, R. A.: Simulation of solute transport in a mountain pool-and-riffle stream: a transient storage model, Water Resour. Res., 19, 718–724, 1983.

Beven, K.: Rainfall-Runoff Modelling: The Primer, John Wiley & Sons, LTD, Chichester, England, 2001.

Beven, K. and Binley, A.: The Future of Distributed Models: Model Calibration and Uncertainty Prediction, Hydrol. Process., 6, 279–298, 1992.

Bingham, Q. G.: Data Collection and Analysis Methods for Two-Zone Temperature and Solute Model Parameter Estimation and Corroboration, M. S. Thesis, Utah State University, http://digitalcommons.usu.edu/etd/564, last access: 13 October 2010, Logan, USA, 2010,.

Bingham, Q. G., Neilson, B. T., Neale, C. M. U., and Cardenas, M. B.: Delineation of dead zones in rivers using remotely-sensed data and their utility in improving two-zone temperature and solute transport model performance, J. Hydrol., in review, 2010.

Increasing parameter certainty and data utility

C. Bandaragoda and B. T. Neilson

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., and, Zyvoloski, G. A.: Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling, Adv. Water. Resour., 31, 630–648, 2008.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward Improved Calibration of Hydrologic Models: Combining the Strengths of Manual and Automatic Methods, Water Resour. Res., 36, 3663–3674, 2000.

Duan, Q.: Global Optimization for Watershed Model Calibration, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., American Geophysical Union, Washington, DC, 2003.

Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S.: Multi-Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging, Adv. Water Resour., 30, 1371–1386, 2007.

Everts, C. J. and Kanwar, R. S.: Evaluation of Rhodamine WT as an absorbed tracer in an agricultural soil, J. Hydrol., 153, 53–70, 1994.

Fu, J. and Gómez-Hernández, J.: Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method, J. Hydrol., 364, 328–341, 2009.

Gupta, H. V., Bastidas, L. A., Vrugt, J. A., and Sorooshian, S.: Multiple Criteria Global Optimization for Watershed Model Calibration, in: Calibration of Watershed Models, editec by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., American Geophysical Union, Washington, DC, 2003.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward Improved Calibration of Hydrologic Models: Multiple and Noncommensurable Measures of Information, Water Resour. Res., 34, 751–776, 1998.

Herbert, L. R.: Seepage Study of the Virgin River from Ash Creek to Harrisburg Dome, Washington County, Utah, Rep. Technical Publication no. 106, United States Geological Survey/State of Utah Department of Natural Resources, 1995.

Hill, M. C.: Methods and Guidelines for Effective Model Calibration, Report 98-4005, USGS Water Resources Investigations, 1998.

Lin, A. Y., Debroux, J.-F., Cunningham, J. A., and Reinhard, M.: Comparison of rhodamine WT and bromide in the determination of hydraulic characteristics of constructed wetlands, Ecol. Eng., 20, 75–88, 2003.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Madsen, H. and Kristensen, M.: A Multi-Objective Calibration Framework for Parameter estimation in the MIKE SHE Integrated Hydrologic Modelling System, in: Proceedings of the 4th International Conference on Calibration and Reliability in Groundwater Modeling, Prague, 2002.

5 Madsen, H., Rosbjerg, D., and Harremoes, P.: Application of the Bayesian Approach in Regional Analysis of Extreme Rainfalls, Stoch. Hydrol. Hydraul., 9, 77–88, 1995.

Nash, J. E. and Sutcliffe, J. V.: River Flow Forecasting through Conceptual Models Part 1 – a Discussion of Principles, J. Hydrol., 10, 282–290, 1970.

Neilson, B. T., Chapra, S. C., Stevens, D. K., and Bandaragoda, C. J.: Two-zone transient
10 storage modeling using temperature and solute data with multi-objective calibration: Part 1 Temperature, Water Resour. Res., in press, doi:10.1029/2009WR008756, 2010a.

Neilson, B. T., Stevens, D. K., Chapra, S. C., and Bandaragoda, C. J.: Two-zone transient storage modeling using temperature and solute data with multi-objective calibration: Part 2 Temperature and Solute, Water Resour. Res., in press, doi:10.1029/2009WR008759, 2010b.

15 Refsgaard, J. C. and Henriksen, H. J.: Modeling Guidelines, Terminology, and Guiding Principles, Adv. Water Resour., 27, 71–82, 2004.

Schaake, J.: Introduction, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., American Geophysical Union, Washington, DC, 2003.

20 Schoups, G., Lee Addams, C., and Gorelick, S. M.: Multi-objective calibration of a surface water-groundwater flow model in an irrigated agricultural region: Yaqui Valley, Sonora, Mexico, Hydrol. Earth Syst. Sci., 9, 549–568, doi:10.5194/hess-9-549-2005, 2005a.

Schoups, G., Hopmans, J. W., Young, C. A., Vrugt, J. A., and Wallender, W. W.: Multi-Criteria Optimization of a Regional Spatially-Distributed Subsurface Water Flow Model, J. Hydrol.,
25 311, 20–48, 2005b.

Shiau, B.-J., Sabatini, D. A., and Harwell, J. H.: Influence of rhodamine WT properties on sorption and transport in subsurface media, Ground Water, 31, 913–920, 1993.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and Efficient Algorithm for Multiobjective Optimization of Hydrologic Models, Water Resour. Res.,
30 39, 1214, doi:10.1029/2002WR001746, 2003a.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameters, Water Resour. Res., 39, 1201, 2003b.

Increasing parameter certainty and data utility

C. Bandaragoda and B. T. Neilson

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Vrugt, J. A. and Robinson, B. A.: Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, Water Resour. Res., 43, W01411, doi:10.1029/2005WR004838, 2007.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, Hydrol. Earth Syst. Sci., 5, 13–26, doi:10.5194/hess-5-13-2001, 2001.

Increasing parameter certainty and data utility

C. Bandaragoda and
B. T. Neilson

Increasing parameter certainty and data utility

C. Bandaragoda and B. T. Neilson

Table 1. Simulation matrix of ten single (1–4) and multi-objective (5–10) calibrations combining main channel temperature and tracer observations at two locations (Site 2 and Site 3).

	Temperature Null	Temperature Site 2	Temperature Site 3	Temperature Site 2 and Site 3
Tracer Null		1. Temp Site 2	2. Temp Site 3	5. Temp Site 2 Temp Site 3
Tracer Site 2	3. Tracer Site 2	7. Temp Site 2 Tracer Site 2	8. Temp Site 3 Tracer Site 2	
Tracer Site 3	4. Tracer Site 3	9. Temp Site 2 Tracer Site 3	10. Temp Site 3 Tracer Site 3	
Tracer Site 2 and Site 3	6. Tracer Site 2 Tracer Site 3			

Table 2. A priori parameter range and calibrated parameter list for the TZTS model.

Parameter Description	Parameter Name	Parameter Range	
		Lower Bound	Upper Bound
STS Width (% Total Channel Width)	β	5	35
STS CS Area (m^2)	$A_{cs,\,STS}$	0.5	3
STS Exchange Coefficient (m^2/d)	α_{STS}	1.7×10^4	8.5×10^4
Hyporheic Storage Advective Transport Coefficient (m^3/d)	Q_{HTS}	86	864
Depth of Hyporheic Storage (m)	Y_{HTS}	0.01	1
Depth of Ground Conduction (m)	Y_{gr}	0.1	1.0

Table 3. Results for single objective (SO, Level 1) and multi-objective (MO, Level 2) calibration tests. Including HTS data gives the AE_{all} result shown in Column 1, excluding HTS and using only main channel (MC) and STS data resulted in AE_s shown in Column 2. Following the AE_{all} and AE_s results are the E results for each test in the simulation matrix. E and AE_s were used to determine the best models using parameter sets that meet both local ($E > 0.8$) and global ($AE_S > 0.7$, bolded) criteria. AE_{all} was included for comparison to Level 3 and 4 calibrations. Shown in grey shading are the Site 2 and Site 3 locations in the main channel used for a calibration objective; unshaded boxes are locations where data was withheld during the calibration.

	AE_{all}	AE_s	Site 2 Temp MC	Site 3 Temp MC	Site 2 Tracer MC	Site 3 Tracer MC
Level 1						
1 – SO Temp 2	0.30	0.36	0.95	0.87	0.32	−0.10
2 – SO Temp 3	0.60	0.65	0.93	0.95	0.23	0.72
3 – SO Tr 2	0.34	0.50	0.72	0.91	0.96	−0.42
4 – SO Tr 3	0.16	0.42	0.89	0.92	−0.70	0.96
Level 2						
5 – MO Temp 2 Temp 3	0.42	0.46	0.96	0.93	0.36	0.11
6 – MO Tr 2 Tr 3	0.61	**0.76**	0.83	0.93	0.35	0.99
7 – MO Temp2 Tr 2	**0.75**	**0.81**	0.91	0.88	0.94	0.62
8 – MO Temp 3 Tr 2	0.39	0.57	0.86	0.94	0.98	−0.17
9 – MO Temp 2 Tr 3	0.47	0.58	0.91	0.93	−0.16	0.92
10 – MO Temp 3 Tr 3	0.65	0.68	0.91	0.95	0.94	0.12

Table 4. The 11 calibration parameters distributed between two sites, the narrowed upper and lower parameter bounds, and associated percent reduction in parameter range compared to the a priori values shown in Table 2. The a priori range was the same for each section, but, the narrowed bounds resulting from calibration varied between Sects. 1 and 2.

Parameter Description	Parameter Name	Sect.	Narrow Lower Bound	Narrow Upper Bound	Bound reduction (%)
STS Width	β_1	1	19	35	47%
(% Total Channel Width)	β_2	2	6	30	4%
STS CS Area	$A_{cs,\ STS1}$	1	0.8	1.3	67%
(m^2)	$A_{cs,\ STS2}$	2	1.0	2.4	44%
STS Exchange Coefficient	α_{STS1}	1	3.8×10^4	8.1×10^4	38%
(m^2/d)	α_{STS2}	2	2.2×10^4	8.1×10^4	15%
Hyporheic Storage Advective	Q_{HTS1}	1	86	415	58%
Transport Coefficient (m^3/d)	Q_{HTS2}	2	173	786	21%
Depth of Hyporheic Storage (m)	Y_{HTS1}	1	0.04	0.82	21%
	Y_{HTS2}	2	0.06	0.92	9%
Depth of Ground Conduction (m)	Y_{gr}	1 and 2	0.2	1.0	11%

C. Bandaragoda and
B. T. Neilson

Table 5. Results for single objective (SO, Level 3, Tests 1–4) and multi-objective (MO, Level 4, Test 5–10) calibration tests using E and AE_s to determine the best model results using parameter sets that meet both local ($E > 0.8$) and global ($AE_S > 0.7$, bolded) criteria. Including HTS data gives the AE_{all} result shown in Column 1. Following the AE_{all} and AE_s results are the E results for each test in the simulation matrix. Shown in grey shading are the Site 2 (S2) and Site 3 (S3) main channel (MC) information used as the temperature (Temp) and solute tracer (Tr) calibration objectives; unshaded boxes are locations where data was withheld during the calibration.

	AE_{all}	AE_s	Site 2 Temp MC	Site 3 Temp MC	Site 2 Tracer MC	Site 3 Tracer MC
Level 3						
1 – SO Temp S2	0.34	0.45	0.94	0.81	0.35	0.04
2 – SO Temp S3	0.64	**0.7**	0.91	0.95	0.81	0.33
3 – SO Tr S2	0.62	**0.76**	0.79	0.84	0.98	0.61
4 – SO Tr S3	0.64	0.69	0.92	0.94	0.06	0.99
Level 4						
5 – MO Temp S2 Temp S3	**0.73**	**0.76**	0.94	0.94	0.59	0.71
6 – MO Tr S2 Tr S3	**0.72**	**0.9**	0.79	0.91	0.98	0.97
7 – MO Temp S2 Tr S2	0.41	0.48	0.95	0.83	0.53	−0.10
8 – MO Temp S3 Tr S2	0.66	**0.79**	0.79	0.83	0.98	0.72
9 – MO Temp S2 Tr S3	**0.78**	**0.9**	0.82	0.92	0.90	0.98
10 – MO Temp S3 Tr S3	0.67	**0.75**	0.89	0.94	0.26	0.99
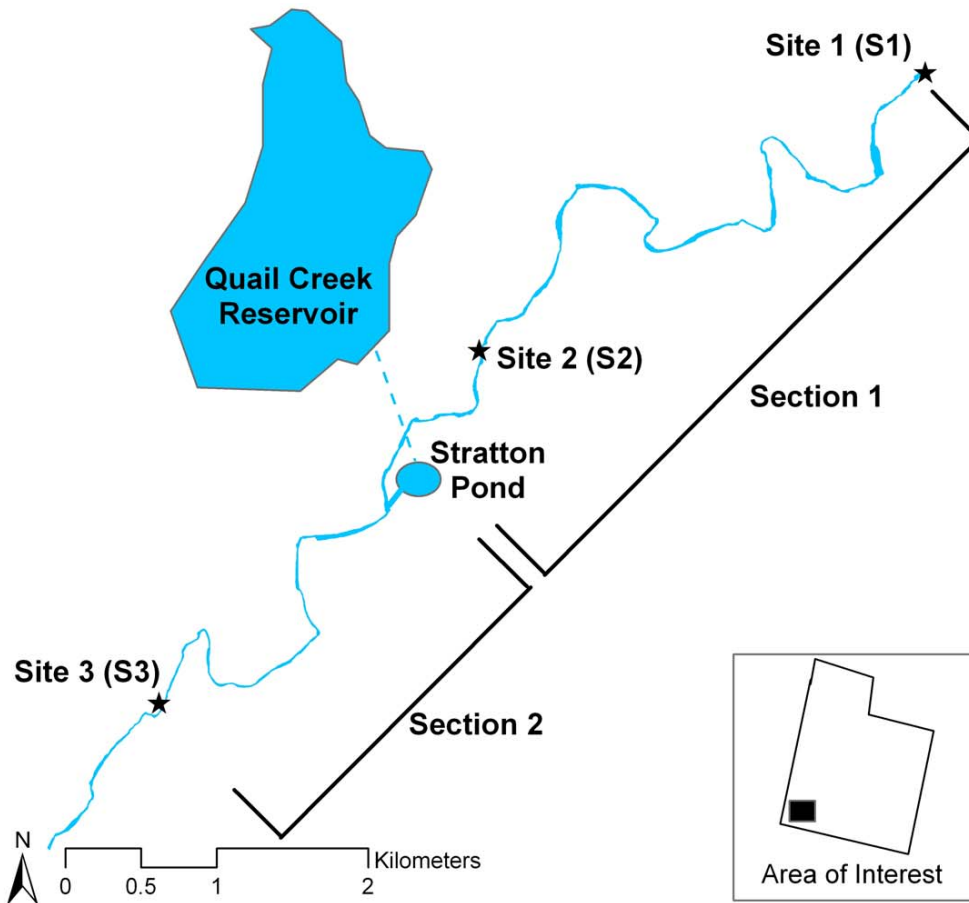
Fig. 1. Study reach layout including data collection locations. Inset map shows the state of Utah, USA, with the study area shown highlighted in black. (Taken directly from Bingham et al., 2010.)

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►I

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
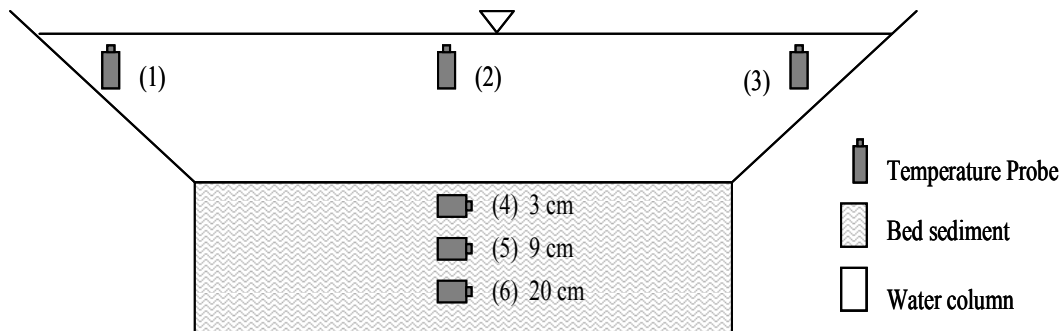
Fig. 2. Locations of temperature probes at Sites 1 and 3 within the study reach. (Taken directly from Bingham et al., 2010).
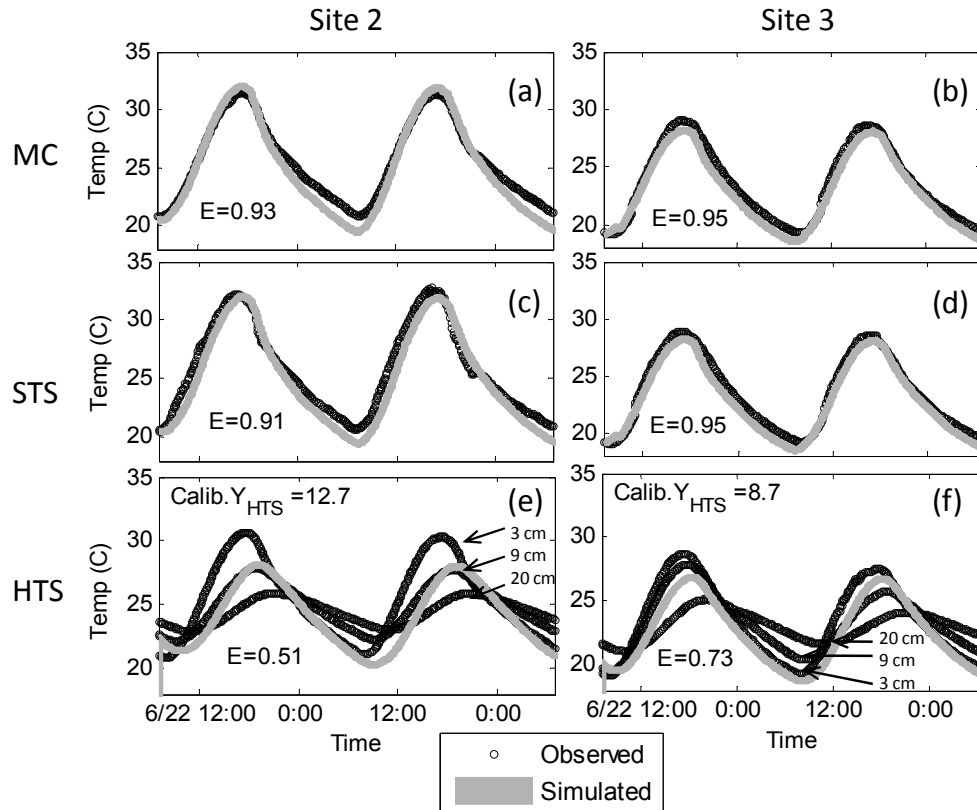
Fig. 3. Test 2 (Level 1) plots of temperature data for Site 2 and Site 3 in the main channel (MC) **(a, b)**, STS **(c, d)**, and HTS **(e, f)**. Test 2 met the local criteria ($E > 0.8$), but not the global criteria ($AE_s > 0.7$). E for each location is shown in each subplot. The calibrated hyphoreic storage sediment depth (Y_{HTS} in cm) is shown in the HTS (e, f) with the observations at three depths labeled (3, 9 and 20 cm). The temperature data sets closest to this Y_{HTS} are used to calculate the E_{HTS} since observations at multiple depths were available.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
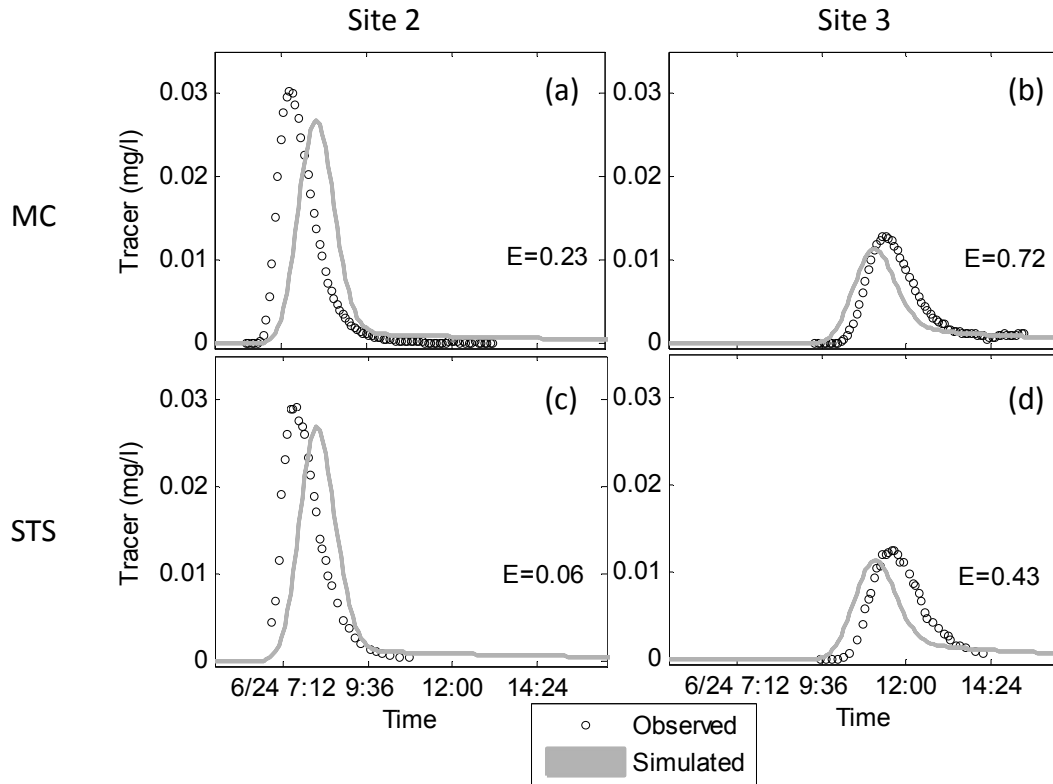
Fig. 4. Test 2 (Level 1) plots of tracer data with results at Site 2 and Site 3 in the main channel (MC) **(a, b)**, and in the STS **(c, d)**. E, the performance at each location, is shown in each subplot, observations are shown as a dotted line, and the model simulations are in grey.
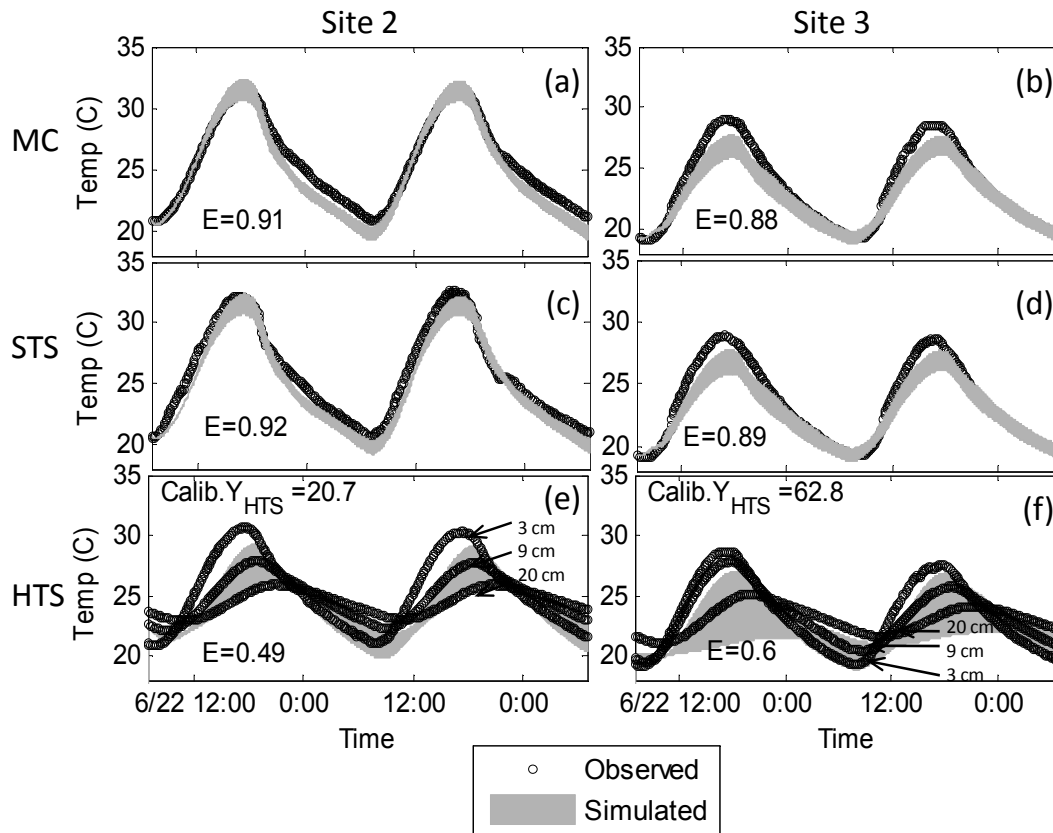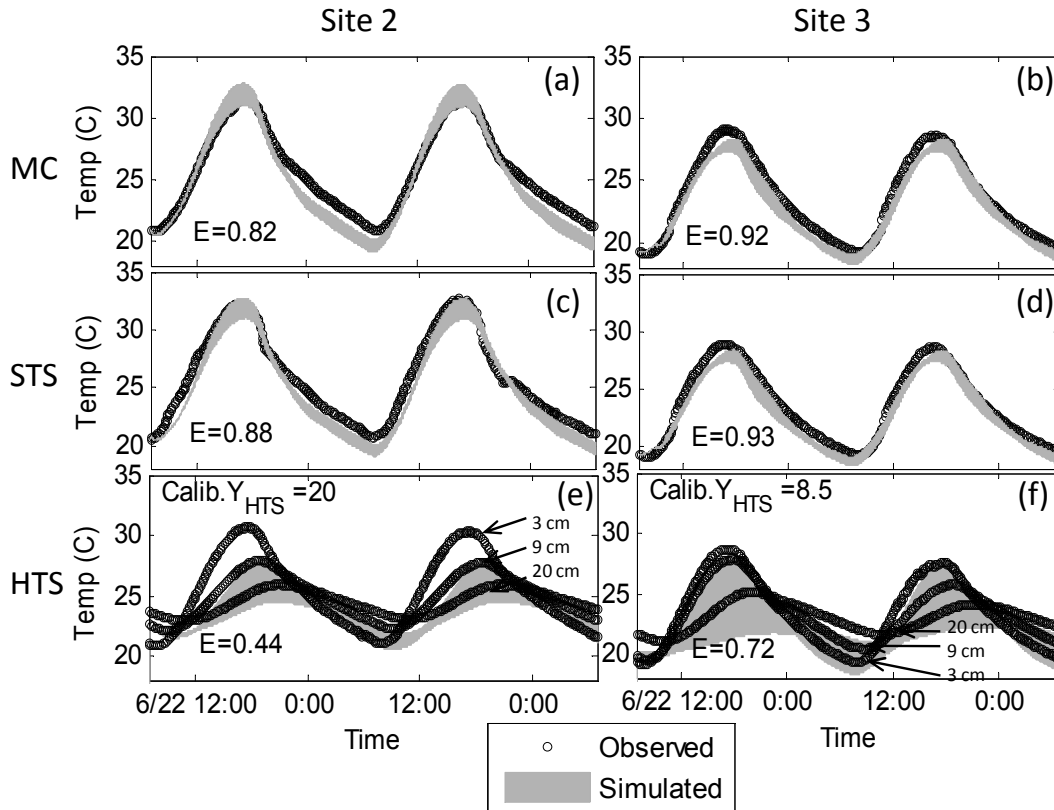
Fig. 5. Test 7 (Level 2) plots of temperature data for Site 2 and Site 3 in the main channel (MC) **(a, b)**, STS **(c, d)**, and HTS **(e, f)**. E, the performance at each location, is shown in each subplot. The optimized hyphoreic storage sediment depth (Y_{HTS} in cm) is shown for the HTS zone with the observations at three depths (3, 9 and 20 cm).

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►◄

◄ | ►

Back | Close

Full Screen / Esc

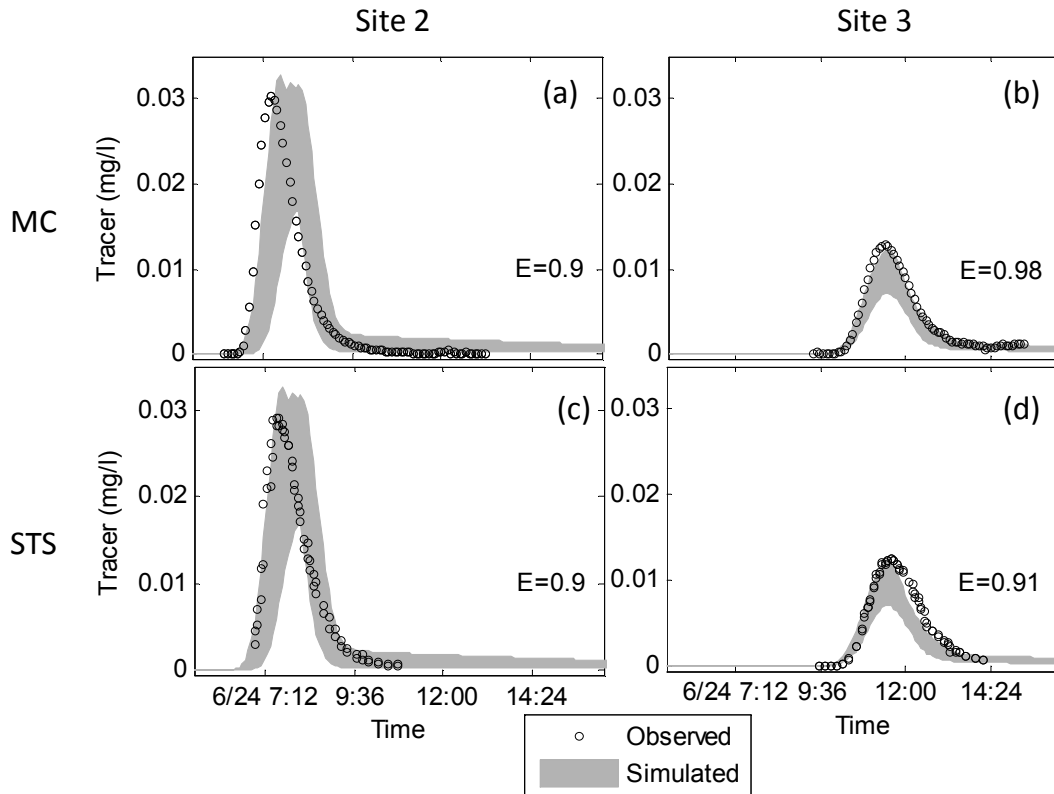Printer-friendly Version

Interactive Discussion
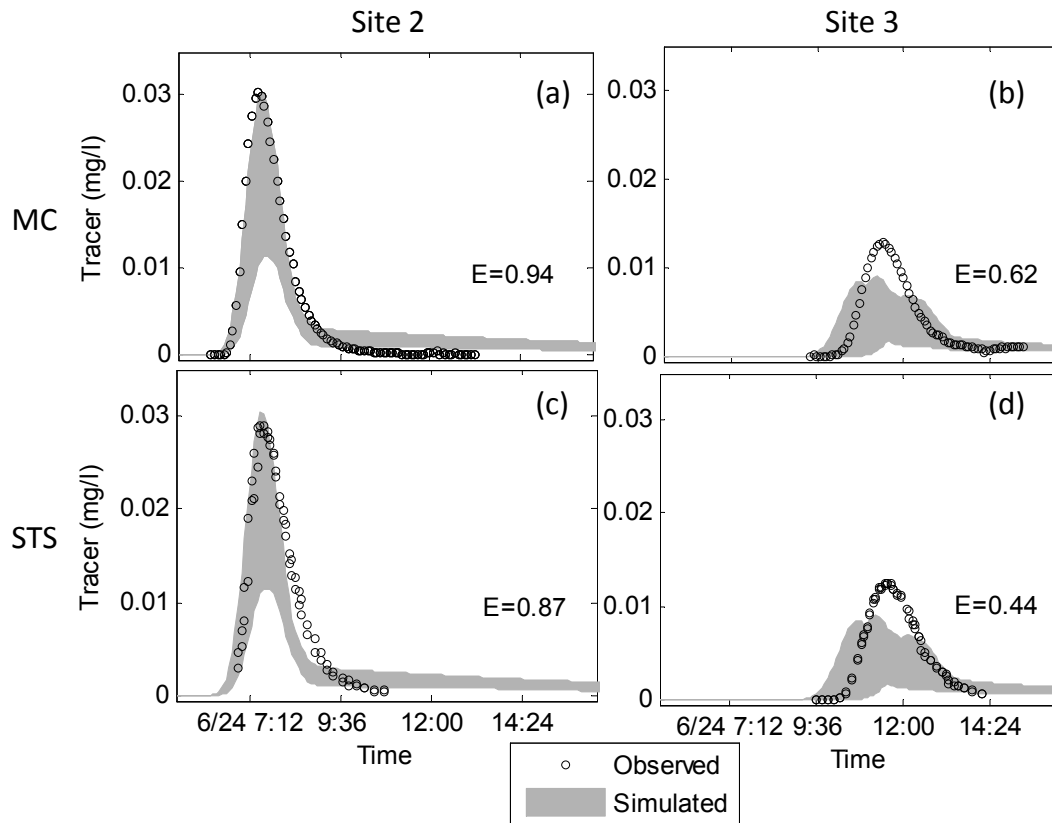
Fig. 6. Test 7 (Level 2) plots of tracer data with results at Site 2 and Site 3 in the main channel (MC) **(a, b)**, and STS **(c, d)**. E, the performance at each location, is shown in each subplot.

C. Bandaragoda and
B. T. Neilson

Fig. 7. The parameter bounds for 11 calibrated parameters within the normalized a priori search space [0,1]. The parameter sets which met the global and local performance criteria for single objective and multi-objective tests, Levels 1 and 2, are used to define a narrowed search space (the grey shaded area) for the Level 3 and 4 calibrations.

Fig. 8. Test 9 (Level 4) plots of temperature data for Site 2 and Site 3 in the main channel (MC) **(a, b)**, STS **(c, d)**, and HTS **(e, f)**, where the observations at three depths are labeled (3, 9 and 20 cm). E, the performance at each location, is shown in each subplot.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Fig. 9. Test 9 (Level 4) plots of tracer data with results at Site 2 and Site 3 in the main channel (MC) **(a, b)**, and STS **(c, d)**. E, the performance at each location, is shown in each subplot.
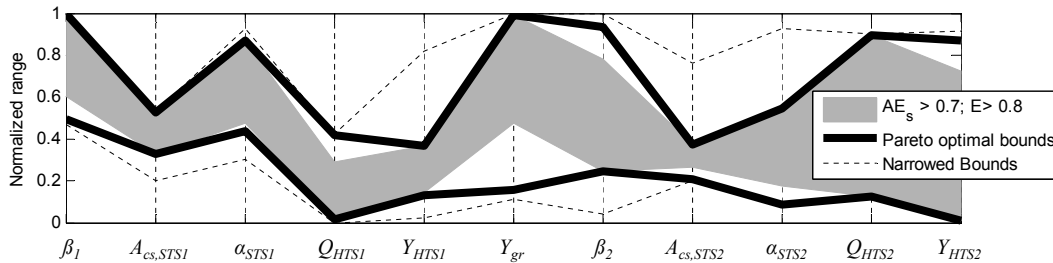
Fig. 10. The parameter sets which met the global and local performance criteria for multiple objective tests in Level 4, Test 9, are shown in grey within the bounds of all of the Pareto optimal parameter sets from Test 9 (black lines). The narrowed search space for the Level 3 and Level 4 calibrations, derived from the Level 1 and Level 2 results, is shown with the dashed line, the a priori search space is the [0,1] bounds of the normalized parameter.