Hydrol. Earth Syst. Sci. Discuss., 7, C5054–C5066, 2011
www.hydrol-earth-syst-sci-discuss.net/7/C5054/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Increasing parameter certainty and data utility through multi-objective calibration of a spatially distributed temperature and solute model" *by* C. Bandaragoda and B. T. Neilson

C. Bandaragoda and B. T. Neilson

christina@silvertipsol.com

Received and published: 10 February 2011

Based on the reviews, it is clear that we need to improve the wording that describes the goals of the paper. 'Prediction in space and time' as currently written in the first sentence of the abstract, is also too broad in scope. Likewise, improving the 'performance of the optimization procedure and the parameter uncertainty' has not succinctly been described in such a way as to clearly elucidate the contributions of our approach. To be clear, the objectives of this paper were to: 1.) develop a systematic, effective,

and reproducible approach to determining an appropriate parameter space from which to sample during calibration that would result in a fit of ten different data types that were distributed in space and time; 2.) test the utility of single and multi objective optimizations in matching information at many locations not used in calibration; and 3.) determine which datasets contain the most information and therefore should be used in calibration to assist in meeting the objective of fitting all the observation types in all locations. The result was a narrowed parameter space and increased parameter certainty which based on our results, would not have been as successful if only single objective algorithms were used in this effort. Instead we found that the single objective results typically only do well for the data set used in calibration and do not provide information regarding parameter ranges or sets of parameters that simultaneously do well everywhere data are available. While one reviewer believes that our approach has a large computational overhead and similar information could be extracted using more common local sensitivity methods, we believe our approach is novel by providing a framework for multiple objective calibration that is not dimensionally limited, nor does it force the modeler to subjectively choose only one dataset as an optimization goal. From our understanding, use of an optimization matrix to simultaneously combine, search, and systematically narrow the search space has not previously been described and validated using such an extensive dataset as in our case study. We expect the method and results to be applicable a broad class of hydrologic and water quality models.

Individual responses are given for each comment in as follows.

Reviewer 1: The paper largely follows the methods described by previous works (Neilson et al.2010, which are correctly cited), and in particular adopts the same mathematical model for heat/solute transport and an improved version of the calibration procedure. The novel aspects introduced by this work is the use of a narrowed parameter space to improve the performance of the optimization procedure and the parameter uncertainty. In my opinion, the authors should make more explicit that this is the novelty of the work, e.g. in the abstract and in the introduction.

Response: Agreed. We have added some clearer statements regarding the paper objectives in the Introduction and Abstract.

Reviewer 1: More importantly, they should also provide more details on how the optimization algorithm works, since this aspect is intimately related to the reason for the remarkable reduction of parameter uncertainty when narrower parameter bounds are used. Apparently, the algorithm is converging to a local minimum at level 1-2, and providing a better first guess of the parameter values enhances the convergence toward the global minimum. However, it is difficult to understand it from the brief description presented in section 3.2 (first paragraph). While a complete description of the optimization algorithm would be excessive, some additional explanations would be helpful, in particular since the improved calibration technique is the novel contribution of the manuscript.

Response: We agree that the novelty of this manuscript is finding a way to narrow the parameter space of a high dimensional problem. We have added clarifying language about how we used a MOSCEM to achieve this aim. See response to comment about this below.

Reviewer 1: The model equations are not presented in the paper, which makes difficult to understand the number of parameters involved in the modeling as well as the precise meaning of the calibrated ones. They should be included in the methods section, or at least in an appendix.

Response: We have added the equations associated with both the temperature and solute portions of the model.

Reviewer 1: The significance of the calibration depends on the interplay between the timescales of in-stream transport and transient storage. I suggest to evaluate values of DaI to verify the amount of information about storage processes carried by in-stream

breakthrough curves.

Response: We agree that this information could be very useful. However, we did not include the DaI values for the two zone calibration due to a lack of guidance regarding the interpretation of these values. Within Neilson et al. 2010 the equations that convert the parameters used within the TZTS modeling framework to those used with typical TSM applications (see Part 1: Temperature p.12) are provided. With this information and the DaI equation provided by Wagner and Harvey (1997), information regarding the DaI for STS and HTS could be determined. Although DaI values have been calculated for two-storage zone models in a few cases (e.g. Harvey, Saiers, and Newlin, 2005), there is not yet the same type of guidance for interpreting the results, and following other recent papers (e.g. Briggs et al. 2009), we do not report DaI values for our two-storage zone calibrations. Further investigation of these values are certainly needed.

Reviewer 1: The reason for adopting a priori values (see p.8317, l.6-7) for roughness, n, and channel width, Btot, instead of calibrating them is unclear. The authors do not report the values, and only give a reference to a manuscript (Bingham et al - also cited later in this work) which is still under review. This reference cannot be used unless it has been accepted. Clearly, the larger the number of parameter to calibrate the more delicate is the calibration. The uncertainty of estimated parameter values also increases with parameters, and given that in their previous work (Neilson et al, 2010b) the authors considered 15 parameters they could do the same here. Otherwise, they should explain the reason for keeping n and Btot constant. Moreover, the temperature of deep sediments layers below 'active' hyporheic exchange zone (T_gr) appears in the model equations. Please specify how this value has been estimated. Response: These average channel width estimates were based on high resolution 3-band and thermal imagery which provides a significantly better estimate the channel width than calibration of this parameter would. Therefore, setting this value provides the ability to focus on the other parameters that are generally not measureable. The channel roughness was established using these widths and ensuring appropriate travel times based on the tracer curves. We anticipated that Bingham et al. would be published at this point, however, due to editorial problems, the publication process has been significantly delayed. We have referred to the thesis instead of the paper and clarification of the justification has been added to the methods. The temperature of the deep sediments was estimated based on an average annual air temperature since actual deep measurements were not available.

In the discussion section, it would be interesting to report the ranges of calibrated values (which are now only given in scaled form in fig.10) and comment them in the context of stream transport processes (e.g. which are the dominant exchange processes in the study reach?).

Response: We agree that this type of discussion would be interesting, but we don't think that a table of values provides more information than what is already shown in Figure 10. Additionally, we feel that more information (e.g., estimates of % fluxes) would need to be added to the paper in order to talk about dominant exchange processes. Given the current length and the primary focus of the paper, we think that this type of discussion would potentially be a detraction.

Reviewer 1: p.8311, l.21: Shouldn't be 'global optimum' instead of 'global optima' here? Please check also in the following lines

Response: The global optimum, a single point in the parameter space, is a theoretical location we don't expect to find given uncertainty in data. However, we do think we can come close, surrounding the global optimum with multiple sets of parameters, to define the global optima. Given this understanding of the point, we did not change the text.

Reviewer 1: p.8312, l.21: 'distributed laterally ... and longitudinally...' This sentence is unclear. Please explain better.

Response: Changed to "...distributed laterally (e.g., within the main channel, HTS, and STS) at one location and longitudinally along a river segment...."

Reviewer 1: p.8314, l.9: Why is the gw exchange important?

Response: Groundwater exchange can be extremely important in accounting for heat sources and sinks along rivers. It can also influence instream solute concentrations. In this case, the groundwater inflows are very small and do not significantly influence instream temperatures or concentrations. However, we included them here to be thorough.

Reviewer 1: p.8314, l.12-13: 'calculate atmospheric fluxes': please explain which flux is referred to here, and how it has been determined.

Response: We have listed which surface fluxes that are estimated within the model equations that have been added to the manuscript.

Reviewer 1: p.8316, l.1-10: this is a point that could be expanded to better clarify how the optimization algorithm works. For instance, the meaning of 'parameter sets ... evolved using two complexes' is not clear.

Response: We have added the following paragraph:

"Each parameter set consists of a different combination of parameter values for each of the 11 parameters that were calibrated. A complex is a group of parameter sets within which the MOSCEM algorithm compares objective function results. The parameter sets with the best results from each complex are selected, new randomly selected parameter sets are added, and the complexes are shuffled with each search iteration."

Reviewer 1: p.8316, l.25: the choice of the a priori parameter space seems very important for the convergence of the optimization. Please explain how this choice was made in the cited work.

Response: This was clarified in the text. A Latin Hypercube sensitivity analysis was conducted to establish the a priori parameter space as described in Neilson et al. 2010a. Additionally, preliminary calibration results which converged at the extents of the parameter bounds were used to expand those corresponding parameter bounds.

Reviewer 1: p.8316, l.5: why a single value is assumed for Y_gr?

Response: We assumed that the depth to a constant temperature would be consistent throughout the entire study reach.

Reviewer 1: p.8319, l.9: the definition of the narrower bound is a central point. Please clarify better how this operation is performed.

Response: As stated on page 8319 lines 9-13, the local criteria (E > 0.8) and global criteria (AEÂňs >0.7) were calculated for each parameter set for each test run in the matrix. A minimum and maximum for each parameter was assessed from the array which met both criteria, and used to define the narrower bounds. We have added some clarifying language here.

Reviewer 1: p.8321, l.14: AE_s is actually equal (not greater than) 0.81

Response: Correct! Thank you for catching this detail.

Reviewer 1: p.8322, l.8: which are the 'pareto rank one sets'? In general, the manuscript lacks a figure showing the pareto curve and the 'best' result of the multiobjective optimization, which would be helpful for the reader. The authors could consider to include it or to provide a reference where such a figure is presented.

Response: We added a reference to the original MOSCEM work, later work by Gupta (1998), and to Neilson et al. 2010a. These references provide this type of information.

Reviewer 1: p.8324, l.7-10: the fact that using two observation series instead of one improves the calibration performance is not unexpected. It is clear that adding more information (assuming they are not flawed by errors) reduces the uncertainty of calibrated values.

Response: This is true in part, but occasionally the tradeoff between two different objectives is so extreme that the performance is significantly degraded at both locations. This may be due to errors in data or model structure. In some applications (not

process-based models), the goal may be to have the best performance at one location without regard to the performance at other locations, in which case, a single objective calibration and one observations series would be appropriate and likely give the best performance. The point that these calibrations performed better at locations not used in the calibration is the significant detail we are highlighting here.

Reviewer 2: As stated in lines 10-23 on p. 8313 the goals of the manuscript are to use the proposed approach to gain insights regarding 1) parameter sensitivity and parameter bounds, 2) importance/worth of data used for calibration, and 3) prediction uncertainty. It seems that one of the key findings of the proposed approach is that improved parameter estimates (as measured by the extent of the plausible range of parameter values) can be obtained by including multiple data sets. These results should not be surprising and is well established in the literature on inverse modeling. In particular, I am wondering how many of these insights listed above could be obtained by conventional (and computationally more efficient) local sensitivity analysis, which has been used for decades to make similar assessments. For example, the results in Figure 10 would correspond to parameters that showed higher sensitivity (that is, parameters with greater sensitivity typically would have narrower bounds). To include the effects of parameter correlation, one could look at the parameter variance, as obtained by the variance-covariance matrix of the parameter sensitivities. Parameters with lower variance have corresponding narrower parameter ranges. To assess the worth of the various data types, plots of the sensitivity to each parameter could be examined. For a composite measure, integrated metrics like CookâĂ§s D or similar influence measures indicate which data have the most influence on calibrated parameter values. Furthermore, examination of the residuals after calibration would indicate which data sets were better fit than others. To assess the importance of data on prediction uncertainty, measures such as OPR and PREDUNC available in inverse modeling software UCODE and PEST, respectively, can provide these insights. Specifically, they measure the effect of data (current or potential new data) on the variance of predictions of interest. Finally, many of these local sensitivity techniques are described by Hill, which is listed in the references, but not actually cited in the body of the manuscript. This makes me wonder if this reference, or other local sensitivity analysis literature, was consulted. In summary, this is an interesting study and will be of interest to users of hydrologic models, but I am concerned about the insights gained via the proposed approach. Specifically, I am wondering if the major insights can be obtained using standard measures based on local sensitivity analyses, which require much less computational overhead as the proposed method. Lack of discussion of local sensitivity methods and how this work relates to those methods is a shortcoming of this manuscript.

Response: Our original statement of the objectives was obviously not clearly articulated in the manuscript and we believe this was misleading. A clearer statement of our objectives are to: 1.) develop a systematic, effective, and reproducible approach to determining an appropriate parameter space from which to sample during calibration that would fit 10 different data types that were distributed in space and time; 2.) test the utility of single and multi objective optimizations in matching information at many locations not used in calibration; and 3.) determine which datasets contain the most information and therefore should be used in multi-objective calibration to assist in fitting all the observed data types in all locations. These have been clarified and restated within the paper. Since we are not focusing on quantitatively determining parameter sensitivity or uncertainty, we have been careful with the use of these terms within the manuscript.

Reviewer 2: Both UCODE and PEST are single objective search algorithms and we believe that one of the primary contributions of this manuscript is to highlight the benefit of multi-objective approaches and the information provided when considering the tradeoffs associated with trying to fit two data sets simultaneously. In the first paragraph of our Introduction, we refer readers to more background on why this is the case, specifically, Schaake, 2003. Single objective methods converge to a single point in the parameter space, and cannot capture the parameter region within which a compromise solution may be a better fit to multiple streams of data. In Neilson et al. 2010

(Part 1:Temperature), we found in the context of that study (which is similar in nature to this one), a two-objective calibration resulted in a better fit everywhere (or global fit) than did a single objective calibration. Consistent with those findings, within this study we found that the use of single objective optimizations were not as successful as the multi-objective optimizations at obtaining global fits to data not used in calibration.

Response: We clearly understand the point the reviewer is trying to make and we agree that some of the information presented in the manuscript could have resulted from using standard local sensitivity methods. However, we also feel that there was a slight misunderstanding of the what was done within this study and how information was used. Our first goal was to narrow the parameter bounds so that the calibration using two data sets results in a fit of all ten datasets. Simultaneously, we wanted to determine which datasets contain the most information, that is, which two could safely be used to meet the objective of fitting all ten. We found that the apriori search bounds mattered and we established a method to use multi-objective calibration to meet optimization objectives and narrow parameter bounds. Further investigation into parameter sensitivity could be completed using various methods mentioned by the reviewer, however, we do not believe this type of analysis belongs within this paper.

Reviewer 2: P. 8311, lines 19-20 – Tradeoff between what? Implies that at least 2 factors need consideration, but it is unclear which factors. Response: In this context, we are talking about the tradeoff between matching one observed data set at the cost of not matching other data sets. We have clarified this within the text.

Reviewer 2: P. 8311, line 23 – Unclear. Global optimum does not exist or finding it is unrealistic? Response: The global optimum exists in theory, but it is unrealistic to expect to find it since the data used as the objective and model structure cannot be expected to be a perfect representation of reality. We want to narrowly bound the global optimum to a useful region where we have good results for data distributed throughout the system – in our case, 10 locations.

Reviewer 2: P. 8311, line 24 – There is no guarantee that the local optimum “bound” the global. Note that on p. 8318, line 65-8 local vs global is defined. Perhaps this should be moved here (p. 8311) to make clear.

Response: We moved this sentence to the introduction.

"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."

Reviewer 2: P. 8314, lines 5-10 – Q1 = 1.06; Q3=1.96; Qgain = 0.17; This does not add up. Is the rest coming from the pond? Response: Yes. Thank you for taking the time to read our paper so carefully. We moved the statement about the point inflows near the text about groundwater inflows to make this more obvious.

Reviewer 2: P.8315, lines 1-5 – How about calculating a percent mass recovered. That would be a direct measure of loss. Response: The point of presenting these numbers are to show that the sorption of RhWT (which can be a big concern) is less of a concern in this study reach. We did not present mass recoveries as justification of minimal sorption because of the limited representation of the tails in the tracer data used in this paper. However, we have done a number of tracer studies during different times of year in this reach. For those tracer curves that include reasonable representation of the tails, the mass recoveries are all >90% and upwards of 102%. However, many of the response curves used in these mass recoveries still have incomplete tails.

Reviewer 2: P. 8317, line 4 – 11 parameters (not 10) Response: There are five parameters in two sections, summarized in Line 4. Line 4-6, the next sentence, describes the eleventh parameter.

Reviewer 2: P. 8317, eqn. 1 – What is "NâĂ§? Number of data of each type? Also, how do you calculate the mean when the data have different units? Some additional details on the use of this equation should be provided. Response: N is timesteps, which is the same value as observations, the number of observations, one for each timestep. In line 18 added, “where for N timesteps, . . .”

Reviewer 2: P. 8318, line 5 – to do this you would actually have to compare different calibration algorithms, the results of which were not reported in this study. Were different algorithms compared?

Response: The way this sentence was originally worded did not clearly state what we intended. It seems that our prior version led the reader to believe that the optimization algorithm has failed to find the optimum. What we were intending to say is that " In this study, our local problem is that any unacceptable parameter set signifies a model failure to adequately reproduce the dominant processes affecting the heat or solute response within the modeled reach." In other words, we need to find the optimal parameters to meet each individual or local data set. We did not compare search algorithms. Based on findings in Tang et al., 2006 (see ref below), we were not convinced the algorithm would make a significant difference in where the convergence occurred, but rather when, an issue which was not of concern to this study. In other words, they all have similar performance, but some may be faster than others. Second, the algorithm used in this study always converged to a good result for the location where data was used in the calibration, but did not always produce good results at the locations not used in calibration. For example, any parameter set producing results with ten good local fits (meaning that it fit all data types at all locations), by definition, was acceptable on a global scale. The aim of our criteria was to filter out any parameter set that could perform well at one or two locations but be totally off the mark everywhere else. The narrowed parameter bounds where determined with the selection of models which performed well (based on our criteria) everywhere we had data. Tang, Y., P. Reed, and T. Wagener. 2006. How effective and efficient are

multiobjective evolutionary algorithms at hydrologic model calibration? Hydrol. Earth Syst. Sci. 10(2): 289âĂŘ307.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/7/C5054/2011/hessd-7-C5054-2011-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 8309, 2010.