Hydrol. Earth Syst. Sci. Discuss., 7, 2121–2155, 2010 www.hydrol-earth-syst-sci-discuss.net/7/2121/2010/ © Author(s) 2010. This work is distributed under the Creative Commons Attribution 3.0 License.



This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Multiobjective calibration of the MESH hydrological model on the Reynolds Creek Experimental Watershed



. MacLean, B. A. Tolson, F. R. Seglenieks, and E. Soulis

Department of Civil Engineering, University of Waterloo, Waterloo ON Canada N2L 3G1, Canada

Received: 25 January 2010 - Accepted: 8 February 2010 - Published: 26 March 2010

Correspondence to: B. A. Tolson (btolson@civmail.uwaterloo.ca)

Published by Copernicus Publications on behalf of the European Geosciences Union.



Abstract

The spatially distributed MESH hydrologic model (Pietroniro et al., 2007) was successfully calibrated and then validated for the prediction of snow water equivalent (SWE) and streamflow in the Reynolds Creek Experimental Watershed in Idaho, USA. The tradeoff between fitting to SWE versus streamflow data was assessed and showed that both could be simultaneously predicted with good quality by the MESH model. Not surprisingly, calibrating to only one objective (e.g. SWE) yielded poor simulation results for the other objective (e.g. streamflow). The multiobjective calibration problem in this study was efficiently solved via a simple weighted objective function approach and analyses showed that the approach yielded a balanced solution between the objectives. Our approach therefore eliminated the need to rely on a potentially more computationally intensive evolutionary multiobjective algorithm to approximate the entire tradeoff surface between objectives. Additional calibration experiments showed

that for our calibration computational budget (2000 model evaluations), the autocali-¹⁵ bration procedure would fail without being initialized to a model parameter set carefully determined for this specific case study. This study serves as a benchmark for MESH model simulation accuracy which can be compared with future versions of MESH.

1 Introduction

Almost all hydrologic models contain effective physical and/or conceptual model parameters that are either difficult or impossible to measure directly. The same is true for more complex hydrologic models coupling land-surface and hydrologic schemes to better represent both the energy and water balance. Therefore, when possible, application of these models require that model parameters are adjusted so that move predictions reasonably replicate the observed environmental system response data.

²⁵ The process of model parameter conditioning to historical system response data is called calibration. Hydrologic model calibration in cold regions can be a difficult task

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model





because modellers usually wish to calibrate the model to multiple observed time series such as distributed point-scale snow water equivalent (SWE) measurements and one or more streamflow measurements that aggregate the upstream basin response.

- The traditional or original approach to model calibration has been to calibrate the model manually by trial and error. Hydrologic modellers soon realized the benefits of posing the model calibration problem as a formal optimization problem that can be solved with a variety of numerical optimization algorithms. Solving a calibration problem this way is often referred to as automatic calibration. Early automatic calibration studies utilized local optimizers that find locally optimal solutions (lbbitt, 1970; Nash and
- Sutcliffe, 1970). Studies followed that recognized the need to employ global optimizers to better approximate the global optimum (e.g. Price, 1978; Masri et al., 1980). Early global optimizers were then replaced with evolutionary optimization algorithms such as the Genetic Algorithm (Wang, 1991) and the Shuffled Complex Evolution algorithm (Duan et al., 1992) that solve the optimization problem by mimicing the evolution of biological populations over time.
- In addition to the above progress in single objective automatic calibration, the model calibration literature has progressed to more advanced approaches that pose and solve multiobjective calibration problems (the focus of this paper) as well as uncertaintybased calibration problems (e.g. Beven and Binley, 1992). Multiobjective model cali-²⁰ bration methods have generally been derived from the field of multiobjective optimization. In multiobjective optimization, there are multiple objectives to simulataneously optimize that are conflicting to the words, the optimal solution for one objective does
- not yield the optimal solution for other objectives that are conflicting. Instead of a single optimum, in multiobjective optimization, there is a set of solutions that define the
- tradeoff between objectives such that each solution is non-dominated with respect to one another. A solution is non-dominated (also called Pareto Optimal) when no feasible solution exists that improves upon one of the objectives without degrading the value of at least one of the other objectives. The general goal of multiobjective optimization is to either locate one non-dominated solution or approximate the entire set of

HESSD 7, 2121–2155, 2010 Multiobjective calibration of the MESH hydrological model A. J. MacLean et al.





non-dominated solutions.

Since hydrologic model development in cold regions typically involves measures of snow processes (i.e. snowmelt and snow accumulation) and streamflow over time, the calibration of models in cold regions could generally be thought of as a multiobjective

- ⁵ problem. Multiobjective model calibration has been solved in a number of ways but the majority of studies have solved the problem in one of two different approaches. The first approach is a classic approach in multiobjective optimization and aggregates different calibration objectives into a single objective function (e.g. a weighted objective function) such that single objective optimizers can be used to solve the calibration prob-
- ¹⁰ Iem. Some examples of aggregating objective functions into a single objective in multiobjective hydrologic model calibration include Madsen (2000, 2003); van Griensven and Bauwens (2003) and Tolson and Shoemaker (2007). The aggregated objective approach can be solved to yield a single non-dominated solution which the modeller would have to deem an acceptable calibration result (although the search history can
- ¹⁵ certainly be post-processed to search for more solutions). The aggregated objective approach requires modellers to specify preferences about objectives prior to optimization. Alternatively, the iterative application of such an approach with a systematic variation in preferences (i.e. systematic variation of the weights in a weighted objective function) can produce an approximation of the non-dominated set of solutions.
- The other approach to multiobjective model calibration which is perhaps now more common than the aggregation of objectives approach is to apply advanced algorithms to approximate the entire non-dominated set such that the modeller can then evaluate the resultant tradeoffs and select a solution they deem to best balance all objectives. The advanced algorithms often utilized for this multiobjective procedure are Evolution-
- ary Multiobjective Optimizers or EMOs (see for example Tang et al., 2006). Other EMOs applied to hydrologic model calibration are demonstrated in Yapo et al. (1998); Vrugt et al. (2003); Fenicia et al. (2007) and Khu et al. (2008). EMOs are much like their single objective evolutionary algorithm predecessors except that they evolve a population of solutions towards the non-dominated front or set of solutions. Deb (2001)

HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model





refers to this multiobjective optimization approach as an ideal approach. Given enough computation time, this is certainly the ideal multiobjective model calibration approach since the hydrologic modeller would be able to examine the tradeoffs in objectives from the results of the multiobjective algorithm and then select a calibration solution that yielded their preferred balance between calibration objectives.

5

10

Unfortunately, with complicated distributed hydrologic models, runtimes are often prohibitive and provide for limited total model evaluations during calibration. Furthermore, van Griensven and Bauwens (2003) report that EMOs such as MOCOM (Yapo et al., 1998) can require very high computation times. Higher computation times relative to single objective procedures might be expected because EMOs are attempting to recover a set of solutions instead of searching for only the best solution. Therefore, this study solves a computationally intensive multiobjective calibration study using

a weighted objective function approach and performs some simple checks to ensure the resultant weighted solutions are appropriately balanced solutions. Specifically, we compare the independent single objective calibration results to the solution from our weighted objective function approach.

This study focuses on evaluating the spatially distributed MESH hydrologic model (Pietroniro et al., 2007) performance for snow water equivalent (SWE) and streamflow predictions in the Reynolds Creek Experimental Watershed in Idaho, USA. MESH and

- its earlier versions have been previously applied in cold regions (Davison et al., 2006; Soulis and Seglenieks, 2007; Dornes et al., 2008a, b). Of these studies, Davison et al. (2006) and Dornes et al. (2008) calibrated MESH for SWE or snow covered area (SCA) and streamflow while Pietroniro et al. (2007) assessed MESH model predictions of SWE and streamflow after very little calibration. Despite the consideration of mul-
- tiple objectives in a number of MESH modelling studies, none of the above studies have compared the calibrated solution to an approximate set of non-dominated solutions. In other words, the tradeoffs between MESH calibration objectives has yet to be assessed. In order to evaluate the quality of the MESH simulations, we use MESH to model the extremely well-instrumented Reynolds Creek Experimental Watershed.

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model





This study therefore also serves as a benchmark for MESH model simulation accuracy which can be compared with future versions of MESH.

2 Methodology

2.1 Case study: Reynolds Creek Experimental Watershed

The Reynolds Creek Experimental Watershed is located in the Owyhee Mountains of south western Idaho, approximately 80 km west of Boise, Idaho in the United States of America. The basin was set up as a research basin by the United States Department of Agriculture (USDA) in the mid 1960's to address the issues of water supply, seasonal snow, soil freezing, water quality, and rangeland hydrology (Slaughter et al., 2001).
This study focuses on the headwater study area of Reynolds Creek, referred to as the Tollgate weir. This area of the watershed was selected because it receives the most annual precipitation, contains all of the snow study sites, and is considered an un altered watershed (i.e. water is not diverted for irrigation).

The Tollgate sub-watershed ranges in elevation from 1410–2241 m a.s.l. and covers
an area of approximately 54.5 km² (Peirson et al., 2001). The majority of the precipitation for this sub-watershed is received as snow during the months of December to May (Marks, 2001). There is a large range in annual precipitation within the sub-watershed, with annual precipitation varying from 500 mm/year to 1100 mm/year (Hanson, 2000). Many other studies have used the Reynolds Creek Expirmental Watershed for the simulation of SWE and/or streamflow with models other than MESH (e.g. Bathurst and Cooley, 1996; Franz et al., 2008; Zhang et al., 2008).

2.1.1 MESH Model description

25

Environment Canada's Modelling Environment Community Surface Hydrological model (Pietroniro et al., 2007), MESH, is the successor to the regional domain mesoscale hydrometeorolgy model developed under the Mackenzie GEWEX study (MAGS) (Soulis



HESSD



and Seglenieks, 2007) when the model was named WATCLASS. MESH combines the Canadian Land Surface Scheme (CLASS) developed at Environment Canada (Vershegy, 1991) for the vertical energy and water balance and the routing code from WAT-FLOOD (Kouwen and Mousavi, 2002), a distributed hydrological model developed at

⁵ the University of Waterloo. Thus, MESH has combined the strengths of a sophisticated energy and water balance model in CLASS with the ability to compare runoff to measured streamflow using the routing code from WATFLOOD.

Previous studies using WATCLASS (a predecessor of MESH) and MESH have been performed using arctic case studies (Davison et al., 2006; Dornes et al., 2008a, b) and regional scale case studies (Pietroniro et al., 2007). The studies by Davison et al. (2006), Dornes et al. (2008a, b) examined approaches to improve simulations in artic watersheds, while Pietroniro et al. (2007) applied the MESH model to regional scale hydrological forecasting for the Laurentian Great Lakes.

10

The MESH model uses the grouped response approach developed by Kouwen et al. (1993) for modeling subgrid variability in addition to the CLASS approach of subgrid variability that allows the GRUs to be subdivided based on land surface characteristics (i.e. vegetation, slope or aspect) into "tiles". The GRU approach was developed for the WATFLOOD hydrological model to deal with basin heterogeneity in a computationally efficient manner by combining areas of similar hydrological behavior.

In the GRU approach, a grouping of all areas with similar land cover (or other attribute) such that a grid square will contain a limited number of distinct GRUs. Runoff generated from the different groups of GRUs are then summed together and routed to the stream and river systems. Two GRUs with the same percentages of land cover types, rainfall, and initial conditions will produce the same amount of runoff regard-

²⁵ less of how these land cover classes are distibuted. The major advantage of the GRU approach is that it can incorporate necessary physics while retaining simplicity of operation (Kouwen and Mousavi, 2002).

This study used MESH version 1.3 with some slight modifications for compilation on an HP Linux operating system. For the modeling work done in this study using MESH,

7, 2121–2155, 2010 Multiobjective calibration of the MESH hydrological

HESSD

A. J. MacLean et al.

model





only one level of subgrid representation was applied so that multiple GRUs only have one vegetation class and there was no further subdivision of the GRUs into tiles. Under this GRU strategy, MESH computes fluxes (e.g. overland flow) and tracks state or prognostic variables (e.g. SWE, soil water content) for each GRU in every grid cell.
 Streamflow for each grid cell outlet is computed from the total (or area weighted average) overland flow, interflow and baseflow of all GRUs in the grid cell. Similarly, fluxes into the atmosphere for each grid cell are calculated as the area weighted averages from the GRUs.

2.2 Model input development

- ¹⁰ The MESH model was configured using meteorological data, geographic information and observed state values. The model was initialized to start on 2 September 1986. Meteorological data (shortwave solar radiation, relative humidity, ambient air temperature, dew-point temperature, wind speed and direction, atmospheric pressure and vapour pressure) were collected at three stations within the Reynolds Creek Water-
- shed. One of these stations is located within the Tollgate sub-watershed, in the headwaters of Reynolds Creek. The three stations were used to interpolate the meteorological data for the Tollgate watershed. Precipitation in Reynolds Creek is monitored extensively and in 1986 there were seventeen dual precipitation gauges active within the watershed, and seven of them were located within the Tollgate sub-watershed. Any
- ²⁰ required meteorological data that were not measured, such as long wave radiation were interpolated using measured data and standard methods.

Reynolds Creek was distributed into six consolidated GRUs based on the vegetative land cover data, mapped by the USDA. Five of these land cover types are present within the tollgate sub-watershed and are presented in 1. The dominant vegetative land

²⁵ cover at Reynolds Creek is a shrub species, commonly know as sagebrush. Details on the breakdown of the land areas within Reynolds Creek are presented in Table 1.

HESSD 7, 2121-2155, 2010 **Multiobjective** calibration of the **MESH** hydrological model A. J. MacLean et al. **Title Page** Introduction Abstract Conclusions References **Figures Tables** 14 Back Close Full Screen / Esc **Printer-friendly Version** Interactive Discussion

2.2.1 Calibration data

At the Tollgate weir, flows are measured using drop-box V-notch weir. The mean annual discharge for the site is 0.424 m^3 /s with the peak flow typically occurring after the spring snowmelt period (Pierson et al., 2001).

- The USDA established seven initial snow course sites in 1961, and one additional snow course was added in 1970 (Marks et al., 2001). A snow pillow was also installed near a snow course site in 1983 to record daily SWE readings that could be validated using data collected from the snow course site. Two snow course sites plus one snow pillow site were selected as SWE calibration locations as shown in Fig. 2. The three sites range in elevation from 1743 ma.s.l. to 2061 ma.s.l. The average SWE for the sites is 225 mm, 325 mm and 500 mm for 155×54, 167×07 and the snow pillow, respectively. The calibration sites were selected based on the ability to match vegetation descriptions of each site with modelled vegetation as well as the conditions at each site. For example, a snow course site known to have significant snow redistribution
- was not included for calibration since MESH does not currently simulate this process.

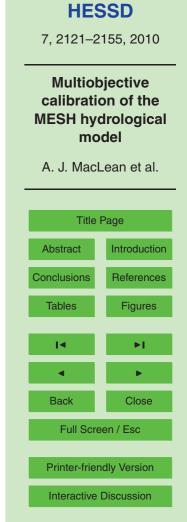
3 Model calibration and validation

20

All MESH model simulations were performed on SHARCNET (Shared Hierarchical Academic Research Computing Network), a parallel computing facility. This was done to decrease total computational time by taking advantage of the thousands of processors available through SHARCNET. The executable files and binary files for the models were compiled on SHARCNET (an HP LINUX operating system).

Model calibration is the perturbation of model parameters within reasonable ranges to improve the agreement between simulated model predictions and measured data for the system being modelled. Calibration may involve manual methods such as trial

²⁵ and error, or automatic calibration procedures that use an optimization algorithm. This study used an automatic calibration procedure.





10

25

5

 $NS = 1 - \frac{\sum_{i=1}^{N} (S_i - O_i)^2}{\sum_{i=1}^{N} (O_i - O_{avg})^2}$

is calculated as follows:

¹⁵ Where S_i is the simulated value at time step *i*, O_i is the observed value at time step *i*, and O_{avg} is the average observed value for the *N* time steps.

The Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007) was selected as the automatic calibration tool for this study. DDS is well suited for optimization problems with a large number of calibration parameters, such as a dis-

Model calibration was focused on improving the agreement between simulated and measured streamflow and/or SWE time series data. The objective function selected

to quantify the quality of agreement between the simulated and measured SWE and

streamflow was the Nash-Sutcliffe efficiency measure (Nash and Sutcliffe, 1970). It was selected because it is a common measure of model performance in hydrology that

can be applied to both streamflow and SWE time series and because it normalizes

the calibration objectives to the same scale. The Nash-Sutcliffe measure is a sum of squares based efficiency measure with a maximum value of 1.0 resulting from a

perfect fit of the simulation and negative values indicating very poor model predictive

guality. It generally measures simulation guality in terms of the shape and volume of

the hydrograph; however it does place a large emphasis on peak events. The Nash-Sutcliffe (NS) value compares a simulated and observed time series over N time steps

tributed watershed model. DDS was designed specifically for automatic calibration and the algorithm is able to rapidly converge to a good calibration solution and easily avoids poor local optima (Tolson and Shoemaker, 2007).

The model calibration experiments are described in detail in Sect. 4.1 to go along with all the results and the experimental design is only outlined here. Three separate calibration problems or experiments were solved:

HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model





1. A single objective calibration to streamflow at the Tollgate weir.

5

10

- A single objective calibration to calibrate SWE for each of the three SWE calibration locations. Sites were each located on different GRU types. For comparative purposes with the streamflow objective, the overall SWE calibration results are summarized as an overall average NS value with each SWE site given equal importance.
- 3. A multiobjective calibration to optimize both the overall average SWE *NS* value and the streamflow *NS* value. To calibrate both objective functions (streamflow and SWE) simultaneously, the objective function was defined as the average of the streamflow *NS* value and the overall average SWE *NS* such that streamflow and overall SWE performance were weighted equally. The goal of calibrating both objectives with equal importance was to try to yield a balanced solution between the streamflow and SWE calibration objectives.

Since DDS, like most global optimizers, is stochastic due to the use of random numbers, optimization results can vary between optimization trials. As such, each of the calibration problems above were solved with five different optimization trials. This was done to minimize the impacts of randomness on the resulting comparison of calibration results between the three calibration problems.

The simulation period for Reynolds Creek began on 2 September 1986 and ran until 31 December 1988. Because the model was initialized with measured data, a relatively short spin-up period was used to reduce the computational time of the model. Model spin-up took place from 2 September to 31 December 1986. Model calibration took place during the 1987 and 1988 calendar years. Each calibration period simulation required approximately five minutes to complete on the SHARCNET computing system.

Although more data were available for calibration, the extreme computational burden associated with the tens of thousands of model runs required in our experiments prohibited using more than a two year long calibration period. The model validation period (post-calibration performance) was assessed for the 1990 to 1992 calendar years.

HESSD					
7, 2121–2155, 2010					
Multiobjective calibration of the MESH hydrological model					
A. J. Macl	A. J. MacLean et al.				
Title	Page				
Abstract	Introduction				
Conclusions	References				
Tables	Figures				
I	۶I				
•	•				
Back	Close				
Full Scre	Full Screen / Esc				
Printer-friendly Version					
Interactive Discussion					



The model parameters selected for calibration are described in detail in MacLean (2009) and are summarized here. Only the model parameters in the two largest GRUs in the Tollgate watershed or those in the GRUs where the SWE calibration data were located were selected for calibration (calibrated GRUs covered approximately 90% of the basin). Parameters in each GRU were calibrated independently instead of using a single calibration multiplier. Model parameters calibrated included those describing vegetation properties (e.g. rooting depth), soil composition percentages and hydrologic conveyance their initial values is their initial values is their impact on the calibrated simultaneously in the calibration experiments.

4 Results and discussion

4.1 Uncalibrated results

The initial parameter values used for the calibration of the MESH model were assigned based on case study specific measured values where possible. Coefficients and other parameters that were not measured were assigned values based on recommended values from CLASS or MESH model documentation or previous MESH calibration studies. Some preliminary MESH runs were utilized to identify initial values of some model parameters.

Using the initiat or uncalibrated parameter values for the different land classes, the resulting *NS* values for both streamflow (0.19) and the snow course sites (-0.46 and -1.13) were quite poor, while the NS value at the snow pillow site (0.40) was somewhat better. In the case of streamflow, the simulation resulted in an earlier spring melt as well as an unrealistically low flow during the summer (Fig. 3). The timing of the snowmelt at the snow pillow site (Fig. 4) and the snow course sites (Figs. 5 and 6) was far too early. As well, the bias values showed that the simulated values were much less than

HESSD

7, 2121-2155, 2010

Multiobjective calibration of the MESH hydrological model



the measured values for both streamflow and SWE (Table 2).

4.2 Streamflow calibration

A calibration was then performed solely using the *NS* value for streamflow as the objective function. The calibration problem was solved with five different DDS optimization trials of 2000 model simulations each where each trial used a different random number seed. Each DDS optimization trial was initialized to the initial model parameter values in Sect. 4.1. Only parameters in the two largest GRUs were calibrated in this problem (approximately 80% of the watershed) yielding a total of 62 calibrated parameters.

The *NS* value for streamflow ranged from 0.80 to 0.90 with an average of 0.86. Both the timing and magnitude of the simulated streamflow was much improved in all the calibrated runs as reflected in their very high *NS* value. Table 2 summarizes the bias and *NS* for streamflow (as well as SWE) for the best streamflow calibration result. Although it was not an explicit calibration objective, it was encouraging that the overall streamflow bias improved to -1% showing that the overall simulated volume of runoff was very close to the measured volume.

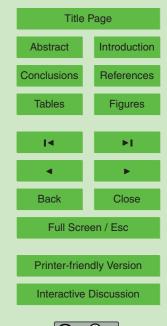
The calibration of the streamflow parameters only impacted one snow course site, 155×54 , due to the distribution of the snow courses in the GRUs. Thus, the other snow course site, 167×07 , and the snow pillow site showed the same performance metrics (bias and *NS*) relative to the uncalibrated results in Table 2. However the snow course site impacted by the streamflow calibration, 155×54 , showed a drastic decrease in model performance with the calibration of streamflow (Table 2).

The process of determining a representative initial parameter values (see Sect. 4.1) can be a difficult or time consuming part of the model set up for the hydrologist and typically requires a strong working knowledge of the model. As an alternative to this process, a randomly generated set of parameter values confined by appropriate ranges could also be used to initialize the automatic calibration algorithm to approach when global optimization algorithms are used for automatic calibration (Duan 1992; Tolson and Shoemaker, 2007) and thus a potentially time saving alternative

HESSD

7, 2121-2155, 2010

Multiobjective calibration of the MESH hydrological model



alternative was tested for the calibration of the model to streamflow and a second set of autocalibration results under the same conditions as above (except utilizing a random initial solution) showed severely degraded algorithm performance with the average *NS* at approximately 0.0 instead of 0.86.

5 4.3 SWE calibration

10

25

A set of five DDS calibration trials of 2000 model simulations were then performed using the average of the *NS* values at the three SWE sites as the objective function. The calibration problem was solved with five different DDS optimization trials of 2000 model simulations each, where each trial used a different random number seed. Each DDS optimization trial was initialized to the initial model parameter values in Sect. 4.1.

Only parameters that impacted SWE in the three GRUs with SWE calibration data were calibrated (39 parameters).

The resulting values of *NS* and bias for the SWE sites (Table 2) show that there were parameter sets that generated good quality simulations of SWE. The performance of the model at the 155×54 snow course site was degraded relative to the other two SWE sites. The simulation of the SWE at the snow pillow was very close the measured values with a *NS* value of 0.97 and a bias of only 1%.

However, the streamflow results were very poor when using SWE as the objective function. The -2.29 value of *NS* was much worse than the uncalibrated results and the bias of 75% shows that the parameter sets greatly overestimated the simulated flow at

²⁰ bias of 75% shows that the parameter sets greatly overestimated the simulated flow at the Tollgate weir.

4.4 Calibration of both streamflow and SWE

The previous calibration experiments (using only SWE and only streamflow for the objective function) were bounding cases to be used to help assess the quality and balance between objectives of our main or overall strategy to calibrate both objectives simultaneously. The multiobjective calibration of SWE and streamflow was performed

HESSD 7, 2121-2155, 2010 Multiobjective calibration of the **MESH** hydrological model A. J. MacLean et al. **Title Page** Introduction Abstract Conclusions References **Figures Tables** Back Close Full Screen / Esc **Printer-friendly Version** Interactive Discussion



by maximizing the average of the streamflow *NS* and overall SWE *NS* values. A total of five DDS optimization trials were performed in the same manner as the previous calibrations.

- Results for this multiobjective calibration are shown in Table 2 for the best of the five calibration trials. Compared to the bounding calibrations, the multiobjective calibration parameter set reflected a reasonable compromise or balanced solution between the calibration objectives. For example, for the best multiobjective calibration result (highest average *NS*), the corresponding *NS* value for streamflow (0.89) was only slightly less than the best streamflow *NS* when the calibration objective was only streamflow (0.90). Similarly the bias value degraded only slightly (from -1% to -8%). Considering SWE
- Similarly, the bias value degraded only slightly (from –1% to –8%). Considering SWE, the best multiobjective calibration result (highest average *NS*) also showed fairly small degradations in performance relative to the case where only SWE was the calibration objective (the *NS* value was degraded by 0.02, 0.18 and 0.23 for sites 167×07, snow pillow and 155×54, respectively).
- Since we consider the best multiobjective calibration result (highest average *NS*) as our calibrated model, the corresponding calibration hydrographs as well as the calibration measured and modelled SWE time series' are shown in Figs. 3–6. In general, comparing the uncalibrated to calibrated simulation results in Figs. 3–6 shows that improvements in model predictions due to calibration were substantial. The hydrographs
 of the calibrated and observed streamflow in Fig. 3 show that the timing and magnitude
 - of the peaks matched quite well.

The time series comparison at the snow pillow site (Fig. 4) shows that in general the simulated maximum SWE value was notably less than the measured value, particularly during the 1987–1988 snow season. Although the simulated accumulation of SWE

tracked well with the measured values in the early part of winter, the simulated result shows the snowpack disappeared almost a month later than the measured snowpack. The snow course data from site 167×07 (Fig. 5) shows that the simulation of SWE was very close to measured snow course data and in particular, the model did a better job simulating the date the snowpack disappeared in comparison with the snow pillow. The

HESSD 7, 2121-2155, 2010 **Multiobjective** calibration of the **MESH** hydrological model A. J. MacLean et al. **Title Page** Introduction Abstract Conclusions References **Figures Tables** 14 Back Close Full Screen / Esc

Printer-friendly Version Interactive Discussion



snow course simulated at site 155×54 (Fig. 6), showed the lowest agreement with the measured data (*NS* value of 0.45) as some substantial deviations in the simulated and measured trends were apparent around the peak measured SWE values. The SWE peaks at site 155×54 were also under-predicted.

5 4.5 Validation

20

The model performance with the calibrated model parameter set was assessed relative to the measured data for the 1990–1992 validation period with a model spin-up period from 2 September 1989 to 31 December 1989.

The performance metrics (bias and *NS*) for the validation period are also included
 in Table 2 and show that the MESH model performed reasonably well. The *NS* value for streamflow was 0.73 with a bias of 17% (both of which were somewhat degraded relative to the calibration period). For the snow pillow and snow course site 167×07, results were similarly degraded. However, at the location with the worst calibration period performance metrics (155×54) the bias and NS both improved notably in validation
 (e.g. *NS* went from 0.45 in calibration to 0.85 in validation).

Figure 7 compares the time series of the simulated and measured streamflow for the validation period and shows that the timing of the simulated streamflow was generally very good with some peak flow magnitudes having notable errors at times during the simulation. Results in Fig. 7 show that although the model closely simulated the largest flow event in the validation period, in each year between March and May the model

failed to predict a few other large events.

The simulation of SWE at the snow pillow site in Fig. 8 showed the same pattern of lower SWE values and late melt that was present in the calibrated simulations. The simulation results for site 167×07 (Fig. 9) showed better melt timing than the snow rillow while simulation results at 155 - 54 (Fig. 10) were clearly improved relative to the

pillow while simulation results at 155×54 (Fig. 10) were clearly improved relative to the calibration period.

HESSD

7, 2121-2155, 2010

Multiobjective calibration of the MESH hydrological model



4.6 Discussion

Overall, the multiobjective calibration results were promising. However, more investigation is needed to determine the source of the snowmelt timing problems evident at the snow pillow site. When the model was calibrated only to SWE, these timing issues

at the snow pillow site were not present. It appears that simulating high quality hydrographs with MESH comes at the expense of slower than observed snowmelt rates at the snow pillow site (as well as site 155×54).

Figure 11 shows the results of the three calibration experiments (previously discussed and analyzed in Sect. 4) in calibration objective space. All calibration experiment optimization trials (five for each of the three experiments) are depicted and the solutions or points on the graph that are non-dominated are denoted with a large open circle. Note that each of the five streamflow calibration results are just slightly dominated by one of the multiobjective solutions. The reason for this result is that the multiobjective calibration involved calibrating more parameters (more land classes) than the streamflow calibration experiments. As such, the autocalibration algorithm was able to

generate a solution in one multiobjective run with a slightly better *NS* for streamflow (0.914 versus 0.904).

Figure 11 demonstrates that the weighted objective approach utilized for calibration yielded balanced solutions that are close to the ideal but unreachable calibration result (NS=1 for both objectives). Clearly, in this problem, the non-dominated solutions or tradeoff curve has a kink or knee-point very close to the ideal point. We believe that given a complete set of non-dominated solutions (or tradeoff curve) like this one between two objectives, the vast majority of hydrologic modellers would quickly select a calibration from this kinked region of the tradeoff. In such cases (a kinked

tradeoff close to the ideal calibration solution), so long as the objectives are described in units that are comparable to the hydrologist or they are appropriately normalized (for example with Nash-Sutcliffe or with the method proposed in van Griensven and Bauwens, 2003), the weighted objective approach to multiobjective calibration can find

HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model





these balanced solutions. Madsen (2003) presents another example of a biobjective tradeoff between calibration objectives with a well defined kink towards the ideal calibration solution. As a result, if modellers have concerns about using Evolutionary Multiobjective Optimizers (EMOs) due to excessive model evaluations in combination

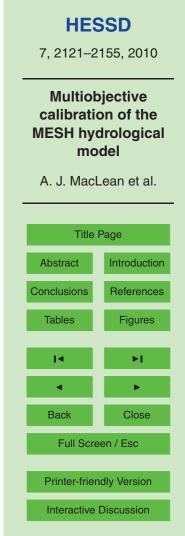
 with excessive model run times, the weighted objective approach may be an acceptable alternative in certain instances. Ahmadi and Arabi (2009) also indicated that in certain multiobjective calibration problems, single objective optimizers can be used in place of more complex EMOs. However, if computation time is not a concern, applying an EMO to approximate the objective between tradeoffs, and thus completely inform
 the hydrologic modeller, is definitely preferred over a weighted objective approach.

With respect to the MESH model, the tradeoff in Fig. 11 icates MESH can simultaneously simulate SWE and streamflow with fairly high quality. Consider if, instead of the results in Fig. 11, the non-dominated simultaneous streamflow and SWE calibration points in Fig. 11 fell close to the straight line joining the top-left and bottom-right

- ¹⁵ corners of Fig. 11 (e.g. at the point [0.0, 0.0]). Such a result would strongly indicate (at least for Reynolds Creek) that MESH was not able to simultaneously simulate these two quantities with reasonable quality. Considering the previous MESH model calibration studies reviewed in Sect. 1.0, this study is the first to depict the tradeoff between calibration objectives. As a result, we have shown a case study with the MESH model
 where drastic improvements in one objective are possible with only small changes in
- the other objective.

5 Conclusions

The spatially distributed MESH hydrologic model (Pietroniro et al., 2007) was successfully calibrated and then validated for the prediction of snow water equivalent (SWE) and streamflow in the Reynolds Creek Experimental Watershed in Idaho, USA. The tradeoff between fitting to SWE versus streamflow data was assessed and showed that both could be simultaneously predicted with MESH. Not surprisingly, calibrating





to only one objective (e.g. SWE) yielded poor simulation results for the other objective (e.g. streamflow). In this case study, the multiobjective calibration problem was efficiently solved via a simple weighted objective function approach to yield a balanced solution between the objectives and thus eliminated the need to rely on a potentially

- ⁵ more computationally intensive evolutionary multiobjective algorithm to approximate the entire tradeoff surface between objectives. The single objective DDS algorithm (Tolson and Shoemaker, 2007) was applied to solve the 88 parameter calibration problem using a total of 2000 MESH model evaluations per optimization trial. A reasonable quality initial model parameter set carefully determined for this specific case study was
- absolutely critical to achieve satisfactory autocalibration results on our computational scale of interest. Therefore, MESH hydrologic modellers should carefully determine an appropriate and case study specific initial parameter set to inform the autocalibration procedure.
- Given the Reynolds Creek Experimental Watershed is so extremely wellinstrumented, this study therefore also serves as a benchmark for MESH model simulation accuracy which can be compared with future versions of MESH. Future work is being conducted to improve MESH efficiency and continue calibration experiments for additional sites within Reynolds Creek.

Acknowledgements. The authors thank Danny Marks of the USDA for his help in not only
 acquiring the data from the Reynolds Creek Experimental Watershed, but also in gaining an understanding of the watershed. The authors also acknowledge the financial support of the Canadian Foundation for Climate and Atmospheric Sciences through the Improving Processes and Parameterization for Prediction in Cold Regions Network. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:
 www.sharcnet.ca) and Compute/Calcul Canada.

HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model



References

- Ahmadi, M. and Arabi, M.: Efficiency of Evolutionary Algorithms for Calibration of Watershed Models, Eos Trans. AGU, 90(52), Fall Meet. Suppl., Abstract H44A-05, 2009.
- Bathurst, J. C. and Cooley, K. R.: Use of the SHE hydrological modeling system to investigate
- basin response to snow melt at Reynolds Creek, Idaho, J. Hydrol., 175(1996) 181–211, 1996.
 - Beven, K. and Binley, A.: The Future Of Distributed Models Model Calibration And Uncertainty Prediction, Hydrol. Process., 6, 279–298, 1992.
- Davison, B., Pohl, S., Marsh, P., Pietroniro, A., and MacKay, M.: Characterizing Snowmelt Variability in a Land-Surface-Hydrological, Model. Atmosphere-Ocean, 44(3) 2009, 271–287, 2006.
 - Deb, K.: Multi-objective optimization using evolutionary algorithms, Wiley, New York, USA, 2001.

Dornes, P., Tolson, B., Davison, B., Pietroniro, A., Pomeroy, J., and Marsh, P.: Regionalization

- of land surface hydrological model parameters in subarctic and arctic environments, Phys. Chem. Earth, 33(17–18), 1081–1089, 2008a.
 - Dornes, P. F., Pomeroy, J. W., Pietroniro, A., and Verseghy, D. L.: Effects of Spatial Aggregation of Initial Conditions and Forcing Data on Modeling Snowmelt Using a Land Surface Scheme, J. Hydrometeorol., 9, 789–803, 2008b.
- ²⁰ Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective And Efficient Global Optimization For Conceptual Rainfall-Runoff Models, Water Resour. Res., 28, 1015–1031, 1992.
 - Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resour. Res., 43, W03434, doi:10.1029/2006WR005098, 2007.
- Franz, K. J., Hogue, T. S., and Sorooshian, S.: Operational snow modelling: Addressing the challenges of an energy balance model for National Weather Service forecasts, J. Hydrol., 360, 48–66, 2008.
 - Hanson, C. L.: Precipitation Monitoring at the Reynolds Creek Experimental Watershed, Idaho, USA, ARS Technical Bulletin NWRC-2000-4, 2000.
- ³⁰ Ibbitt, R. P.: Systematic parameter fitting for conceptual models of catchment hydrology, University of London, London, UK, 1970.

Khu, S.-T., Madsen, H., and di Pierro, F.: Incorporating multiple observations for distributed

HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model

Title Page						
Abstract	Introduction					
Conclusions	References					
Tables	Figures					
I	۶I					
•	•					
Back	Close					
Full Scre	Full Screen / Esc					
Printer-friendly Version						
Interactive Discussion						

2141

hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, Adv. Water Resour., 31, 1387–1398, 2008.

Kouwen, N., Soulis, E.D., Pietroniro, A., Donald, J., and Harrington, R.A.: Grouped Response Units for Distributed Hydrologic Modelling, J. Water Res PI.-ASCE, 119, 289–305, 1993.

Kouwen, N. and S-F. Mousavi: WATFLOOD/SPL9 hydrological and flood forecasting system, Chapter 15 in Mathematical Models of Large Watershed Hydrology, edited by: Singh, V. and Frevert, D. K., Water Resources Publications, Highlands Research, Colorado, 649–686, 2002.

MacLean, A.: Calibration and analysis of the MESH hydrological model applied to cold regions,

- ¹⁰ MASc Thesis, Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, 2009.
 - Madsen, H.: Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, J. Hydrol., 235, 276–288, 2000.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26, 205–216, 2003.

Marks, D., Cooley, K., Robertson, D., and Winstral, A.: Long-term snow database, Reynolds Creek Experimental Watershed, Idaho, United States, Water Resour. Res., 37, 11, 2835– 2838, 2001.

Masri, S. F., Bekey, G. A., and Safford, F. B.: A global optimization algorithm using adaptive random search, Appl. Math. Comput., 7, 353–375, 1980.

- Nash J. E. and Sutcliffe J. V.: River flow forecasting though conceptual models. Part I: a discussion of principles, J. Hydrol., 10, 282–290, 1970.
- Pierson, F., Slaughter, C., and Cram, Z.: Long-term stream discharge and suspended-sediment database, Reynolds Creek Experimental Watershed, Idaho, United States, Water Resou. Res., 37, 11, 2857–2861, 2001.
- Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N., and Pellerin, P.: Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale, Hydrol. Earth Syst. Sci., 11, 1279–1294, 2007,
- ³⁰ http://www.hydrol-earth-syst-sci.net/11/1279/2007/.

15

20

25

Price, W. L.: A controlled random search procedure for global optimization, in: Towards Global Optimization 2, edited by: Dixon, C. W. L. and Szego, G. P., North-Holland, Amsterdam, 71–84, 1978.

HESSD

7, 2121-2155, 2010

Multiobjective calibration of the MESH hydrological model



Printer-friendly Version Interactive Discussion

- Slaughter, C., Marks, D., Flerchinger, G., Van Vactor, S. S., and Burgess, M.: Thirty-five years of research data collection at the Reynolds Creek Experimental Watershed, Idaho, United States, Water Resour. Res., 37, 11, 2819–2823, 2001.
- Soulis, E. D. and Seglenieks, F. R.: The MAGS Integrated Modeling System. Cold Regions Atmospheric and Hydrologic Studies: the Mackenzie GEWEX Experience, Vol. II: Hydrologic 5 Processes, edited by: Woo, M. K., Springer-Verlag, Berlin, Heidelberg, Germany, 445–474, 2007.
 - Tang, Y., Reed, P., and Wagener, T.: How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, Hydrol. Earth Syst. Sci., 10, 289–307, 2006, http://www.hydrol-earth-syst-sci.net/10/289/2006/.

10

20

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resour. Res., 43, W01413, doi:10.1029/2005wr004723.2007.

van Griensven, A. and Bauwens, W.: Multiobjective autocalibration for semidistributed water guality models, Water Resour. Res., 39(12), 1348, doi:10.1029/2003WR002284. 15

- Verseghy, D. L.: CLASS A Canadian land surface scheme for GCMS. I. Soil model, Int. J. Climatol., 11, 111–133, 1991.
 - Vrugt, J., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39, 1214, doi:1210.1029/2002WR001746, 2003.
- Wang, Q. J.: The Genetic Algorithm and its application to calibrating rainfall-runoff models, Water Resour. Res., 27, 2467–2471, 1991.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, J. Hydrol., 204, 83–97, 1998.
- ²⁵ Zhang, X., Srivivasan, R., and Van Liew, M.: Multi-Site Calibration of the SWAT Model for Hydrologic Modeling, T. ASABE, 51, 6, 2039–2049, 2008.

7, 2121-2155, 2010

Multiobjective calibration of the **MESH hydrological** model

A. J. MacLean et al.

Title Page

Abstract

Conclusions

Tables

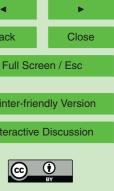
14

Back

Introduction

References

Figures



HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model

A. J. MacLean et al.



Table 1. Land class descriptions and percent area for Tollgate subwatershed.

GRU Number	CLASS code	Area (%)	CLASS Descriptions	Vegetation Description (USDA)
1	2	64%	Broad Leaf	Low Sagebrush Mountain Sagebrush-Snowberry
2	2	15%	Broad Leaf	Wyoming Sagebrush Wyoming Sagebrush-Bitterbrush
3	1	0%	Needle Leaf	Greasewood
4	2	9%	Broad Leaf	Quaking Aspen
5	1	8%	Needle Leaf	Conifer
6	3	4%	Crops	Cultivated Other Vegetation

HESSD

7, 2121–2155, 2010

Multiobjective calibration of the MESH hydrological model

A. J. MacLean et al.

Title Page					
Abstract	Introduction				
Conclusions	References				
Tables	Figures				
I.	۶I				
•	•				
Back	Close				
Full Screen / Esc					
Printer-friendly Version					
Interactive Discussion					

ВУ

Table 2. Model performance metrics across all three calibration problems (only the best calibration result for each problem is shown).

			Streamflow SWE			Multiobjective				
Locations	Uncalibrated (1987–1988)		Calibration (1987–1988)		Calibration (1987–1988)		Calibration (1987–1988)		Validation (1990–1992)	
	Bias	NS	Bias	NS	Bias	NS	Bias	NS	Bias	NS
Streamflow	-30%	0.19	-1%	0.90	75%	-2.29	-8%	0.89	17%	0.73
Snow Pillow	-44%	0.40	-44%	0.40	1%	0.97	-20%	0.79	-26%	0.75
167×07	-49%	-0.46	-49%	-0.46	0%	0.83	0%	0.81	-13%	0.68
15×54	-64%	-1.13	-76%	-1.87	-20%	0.68	-29%	0.45	-2%	0.85

	7, 2121–2	155, 2010
	Multiob calibratio MESH hyo mo A. J. MacL	on of the drological del
	Title I	Page
	Abstract	Introduction
	Conclusions	References
Low Sagebrush Wyoming Sagebrush Quaking Aspen	Tables	Figures
Conifer Other Vegetation	14	►I ►

Fig. 1. Consolidated land classes simulated in MESH model for the southern portion of Reynolds Creek (Tollgate subwatershed), Idaho, USA.



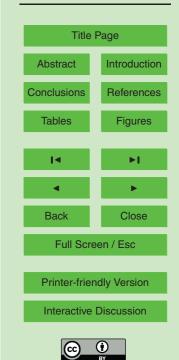
HESSD

HESSD

7, 2121-2155, 2010

Multiobjective calibration of the MESH hydrological model

A. J. MacLean et al.



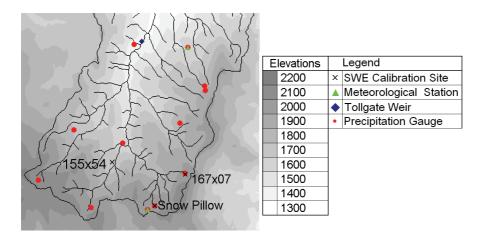
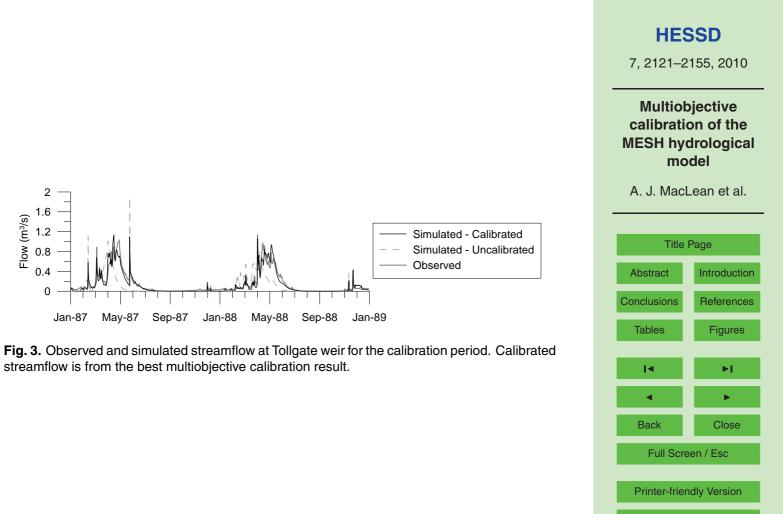


Fig. 2. Data monitoring locations used in the modelling of the Tollgate subwatershed in Reynolds Creek, Idaho, USA.



Interactive Discussion



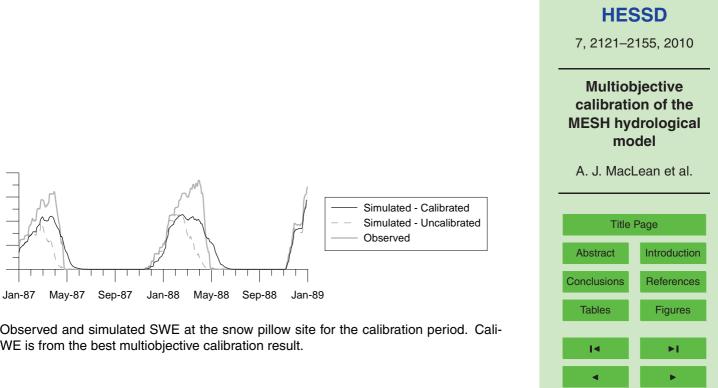


Fig. 4. Observed and simulated SWE at the snow pillow site for the calibration period. Calibrated SWE is from the best multiobjective calibration result.

400 300

200

100

0

SWE (mm)



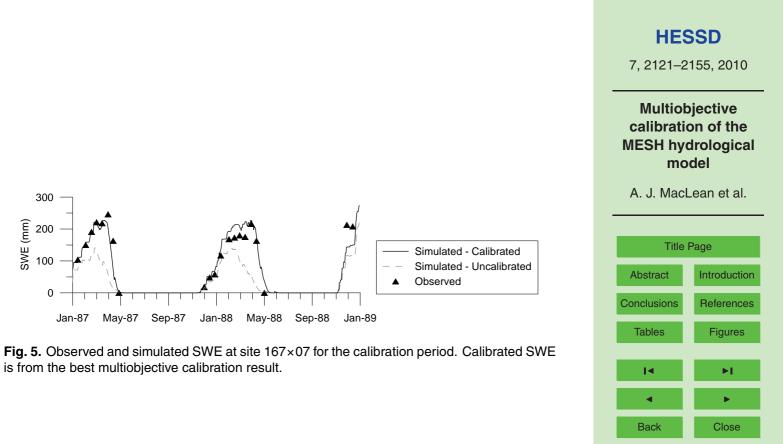
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Close

Back

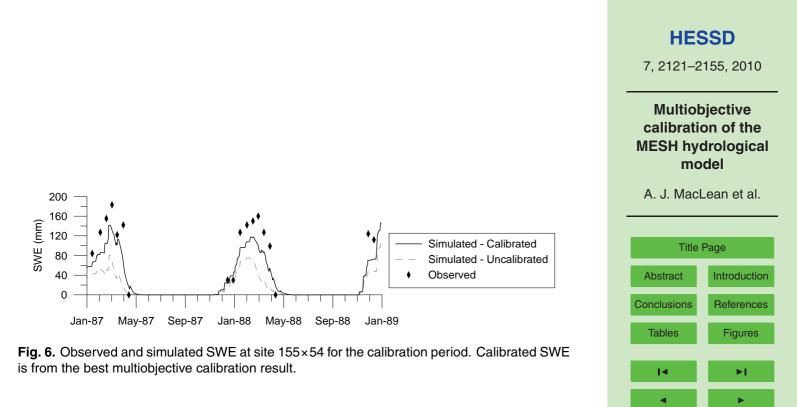




Printer-friendly Version

Interactive Discussion







Close

Back

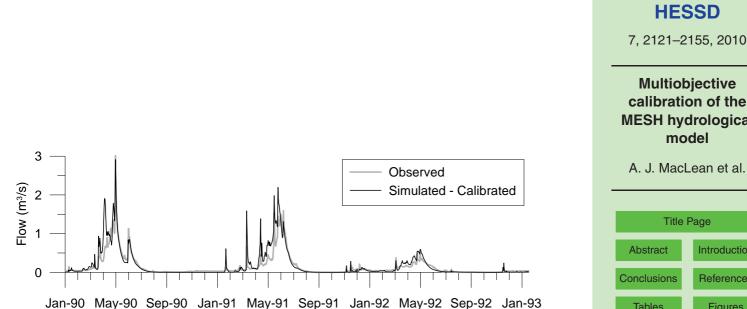


Fig. 7. Observed and simulated streamflow at the Tollgate weir for the validation period. Calibrated streamflow is from the best multiobjective calibration result.



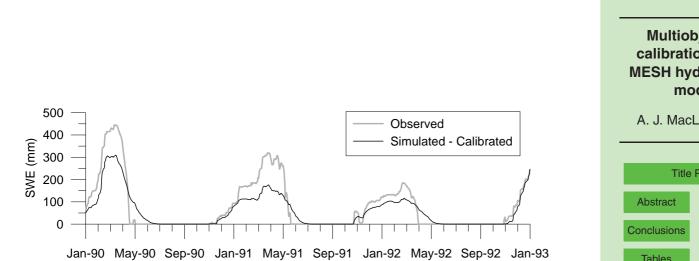


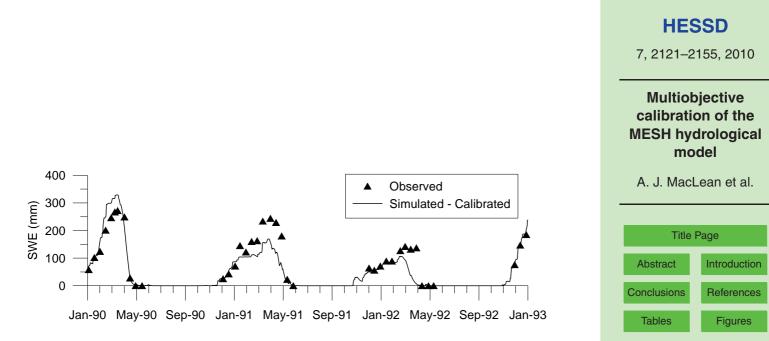
Fig. 8. Observed and simulated SWE at snow pillow site for the validation period. Calibrated SWE is from the best multiobjective calibration result.

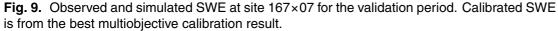
HESSD

7, 2121-2155, 2010

Multiobjective calibration of the **MESH hydrological** model







Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Þ١

Close

.∎∢

Back

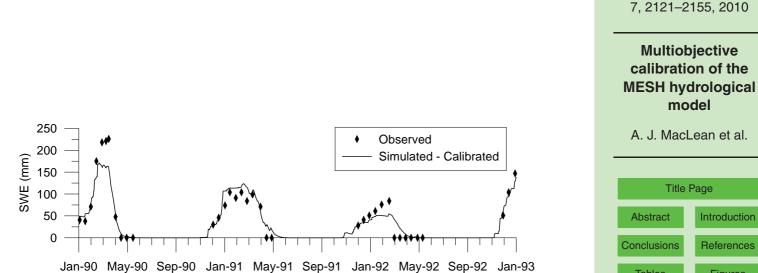


Fig. 10. Observed and simulated SWE at site 155×54 for the validation period. Calibrated SWE is from the best multiobjective calibration result.



HESSD

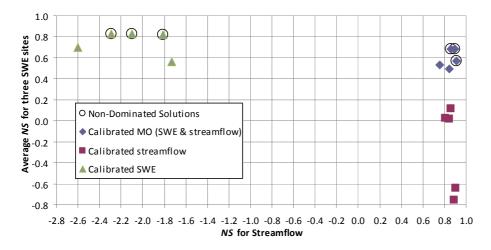


Fig. 11. Calibration experiment results in terms of the streamflow NS and overall average SWE NS for all three calibration experiments (five optimization trials per experiment).

HESSD

7, 2121-2155, 2010

Multiobjective calibration of the MESH hydrological model

