Improved Stream Temperature Simulations in SWAT Using NSGA-II for Automatic Multi-Site Calibration

B. L. Barnhart, G. W. Whittaker, D. L. Ficklin

Abstract. *Stream temperature is one of the most influential parameters impacting the survival, growth rates, distribution, and migration patterns of many aquatic organisms. Distributed stream temperature models are crucial for providing insights into variations of stream temperature for regions and time periods for which observed data do not exist. This study uses a relatively new stream temperature model incorporated into a modified version of the Soil and Water Assessment Tool (SWAT) in order to simulate stream temperatures at five sites on the Calapooia River within the Calapooia basin in northwest Oregon. The Nondominated Sorting Genetic Algorithm II (NSGA-II) is used to calibrate flow at a single outlet and stream temperatures at five sites. Few studies have calibrated this stream temperature model for different basins, and this is the first demonstration of an automatic, subbasin-level calibration for stream temperature at multiple sites. The subbasin calibration is shown to better match the observed data than the original SWAT temperature model as well as the modified temperature model calibrated basinwide. In addition to providing improved stream temperature simulations for the Calapooia River, the subbasin-level automatic calibration technique extends the applicability of the model, especially for complex basins with large spatial variability of topography, land use, and soil type.*

Keywords. *Genetic algorithms, Model calibration, NSGA-II, Stream temperature, SWAT.*

Throughout the western U.S., rising stream temperatures are becoming an increasing problem, affecting coldwater aquatic habitats and fish species. Section 303(d) of the Clean Water Act requires states to develop lists of impaired waters that do not meet water quality standards for a variety of environmental variables. In Oregon, the most commonly listed impairment of rivers is due to high temperatures (USEPA, 2003). This is highly relevant, since increasing stream temperatures can affect the survival, growth rates, distribution, and migration patterns of many aquatic organisms, including salmon and trout (USEPA, 2003).

To determine whether streams are 303(d) impaired due to high temperatures, states must facilitate studies to track stream fluctuations amid natural and anthropogenic changes, typically by installing and monitoring *in situ* measurement devices at discrete locations along stream segments. However, the spatial and temporal resolution of *in situ* measurements is inherently limited. In order to address these limitations, models are often used to provide insights into variations of stream temperature for regions and time periods for which observed data do not exist. A large number of stream temperature models exist, ranging in complexity, computational efficiency, and applicability. For a recent overview of available stream temperature models, see Ficklin et al. (2012). This investigation focuses on one such model, introduced by Ficklin et al. (2012), that utilizes the Soil and Water Assessment Tool (SWAT) to produce daily stream temperature simulations.

SWAT is a mechanistic, or process-based, model that divides a basin into a number of subbasins and hydrologic response units (HRU) in order to model agricultural practices, fate and transport of soil and chemicals, and hydrology at small scales (Arnold et al., 2012; Neitsch et al., 2005). SWAT has been utilized for distributed hydrologic modeling in basins throughout the world and is especially useful for modeling the environmental impacts associated with different land use and management practices (Chaplot et al., 2004; Fohrer et al., 2005; Li et al., 2009; Vache et al., 2002).

Currently, SWAT calculates stream temperatures using a simple linear model that directly relates stream temperatures to air temperature, as demonstrated by Stefan and Preud'homme (1993). Ficklin et al. (2012) modified the original linear stream temperature model within SWAT. The modified model utilizes the stream network within SWAT to account for surface and lateral flow, groundwater and snowmelt contributions in each subbasin, as well as travel time and ambient air temperatures. The modified model better matched observed data for seven coastal and mountainous basins in the western U.S. (Ficklin et al., 2012). While the modified model was an invaluable improvement for simulating stream temperatures using SWAT, it used a manual calibration technique. In order for

Submitted for review in October 2013 as manuscript number SW 10472; approved for publication by the Soil & Water Division of ASABE in February 2014.

Mention of company or trade names is for description only and does not imply endorsement by the USDA. The USDA is an equal opportunity provider and employer.

The authors are **Bradley L. Barnhart,** Research Physical Scientist, and **Gerald W. Whittaker,** Research Hydrologist, USDA-ARS Forage Seed and Cereal Research Unit, Corvallis, Oregon; **Darren L. Ficklin,** Professor, Department of Geography, Indiana University, Bloomington, Indiana. **Corresponding author:** Bradley Barnhart, USDA-ARS, 3450 Campus Way, Corvallis, OR 97331-8539; phone: 541-738-4182; e-mail: bradleybarnhart@gmail.com.

the model to be applicable to large basins, where stream temperatures are simulated within different regions containing different topography, soil types, and land use, a more applicable calibration routine must be adopted.

Automatic calibration techniques have been widely used for environmental models, including SWAT. In particular, SWAT has been calibrated at basins throughout the world using single or multiple sites and objectives for variables such as streamflow, sediment, and nutrients (N and P) (Bekele and Nicklow, 2007; Santhi et al., 2001; Santhi et al., 2008; White and Chaubey, 2005; Zhang et al., 2009b; Zhang et al., 2010). Automatic methods are especially attractive because they are objective and relatively easy to use, in contrast to subjective and labor-intensive manual techniques (Zhang et al., 2009a).

Among the most successful automatic calibration techniques are numerous evolutionary search algorithms. These algorithms seek the fittest sets of calibration parameters by minimizing one or more objective functions that represent the differences between observed and simulated data. There are a wide variety of evolutionary algorithms, differentiated by their methods for seeking solutions within the decision space and calculating fitness. Some examples include the Nondominated Sorting Genetic Algorithm II (NSGA-II) introduced by Deb et al. (2002), the Strength Pareto Evolutionary Algorithm 2 (SPEA2) introduced by Zitzler et al. (2001), the S-Metric Selection Evolutionary Multiobjective Optimization Algorithm (SMS-EMOA) introduced by Beume et al. (2007), particle swarm optimization (PSO) introduced by Kennedy and Eberhart (1995), and the Shuffled Complex Evolution Algorithm (SCE-UA) introduced by Duan et al. (1992). While the choice of the ideal algorithm depends on the application, there are a number of studies that attempt to examine the relative performance of evolutionary algorithms (Tang et al., 2006; Zhang et al., 2009a). Beume et al. (2007) demonstrated that SMS-EMOA, which calculates fitness based on a dominated hypervolume within the solution space, provides better results than NSGA-II and SPEA2 for some select optimization problems (Beume et al., 2007; Naujoks et al., 2005). For hydrologic model calibration, Tang et al. (2006) suggested that while SPEA2 and NSGA-II demonstrate comparable results, SCE-UA is only suitable for calibration of small parameter sets.

Dealing specifically with the Calapooia River, Confesor and Whittaker (2007) used NSGA-II to calibrate surface and baseflow using two separate objectives. In addition, Whittaker et al. (2010) used NSGA-II to calibrate streamflow and baseflow using over 4,000 parameters within the Blue River basin. Whittaker et al. (2010) demonstrated that NSGA-II successfully calibrated this large set of parameters while avoiding overparameterization and overfitting of the data.

For this study, the modified stream temperature model requires four calibration parameters. Additionally, the parameters can be set to vary for different periods of time to represent seasonal effects. We specify four seasons; therefore, the model requires 16 total parameters for basinwide calibration and 16 parameters for each subbasin for subbasin calibration. With even a modest number of subbasins, the total number of parameters to be calibrated will be large. Therefore, we chose to utilize NSGA-II and a methodology similar to that of Whittaker et al. (2010) to calibrate flow at the watershed outlet and stream temperature at five sites within the Calapooia basin. Stream temperature using the modified model will be calibrated at both the basin and subbasin scales. Both the basinwide and subbasin-level calibrated stream temperatures will be compared to existing temperature data sets for five different sites on the Calapooia River as well as the original SWAT stream temperature model by Stefan and Preud'homme (1993). The benefits of using an automatic, multi-site, subbasin-level calibration technique will be demonstrated.

To our knowledge, only a few studies have calibrated the Ficklin et al. (2012) modified stream temperature model in SWAT (Ficklin et al., 2012; Ficklin et al., 2013; Luo et al., 2013). None of these studies successfully calibrated the model for multiple sites simultaneously, nor at subbasin scales. Therefore, the application of NSGA-II to calibrate the modified stream temperature model at subbasin scales will increase the model's applicability to other basins, especially those that contain a large amount of spatial variability (e.g., topography, land use, and soil type) between subbasins.

The resulting calibrated model will simulate daily stream temperatures for five sites along the Calapooia River from 2009 to 2012. The ability to produce continuous, daily time series simulations of stream temperatures can be extremely useful for accurately modeling conditions and dynamics of aquatic species and habitats, especially in 303(d) impaired rivers throughout the western U.S.

MATERIALS AND METHODS

STUDY AREA

The Calapooia basin (USGS 10-digit HUC 1709000303) is located in northwest Oregon and has a drainage area of approximately 950 km^2, as shown in figure 1. Its longest river at 120 km long, the Calapooia River, is a gravel-bed tributary of the Willamette River and flows northwest from the western Cascade Mountains to its outlet at Albany, Oregon, where it joins the Willamette River (Hoag et al., 2012). The elevation of the basin ranges from approximately 1,576 m in the Cascade Mountains to 56 m in the valley (Hoag et al., 2012).

The basin consists mostly of two distinct regions. The upper region east of Holley, Oregon, in the western Cascade Mountains is characterized by steep slopes and forested land, while the low region of the valley, northwest of Holley, is characterized by flat agricultural land, hay and pasture land, and rural residential areas. Annual precipitation, 50% of which occurs between December and February, varies from less than 1,000 mm (39 in.) at low elevations to more than 2,000 mm (79 in.) in the foothills of the western Cascade Mountains (Hoag et al., 2012).

SOIL AND WATER ASSESSMENT TOOL

The Soil and Water Assessment Tool (SWAT) is a widely known, mechanistic, semi-distributed simulation

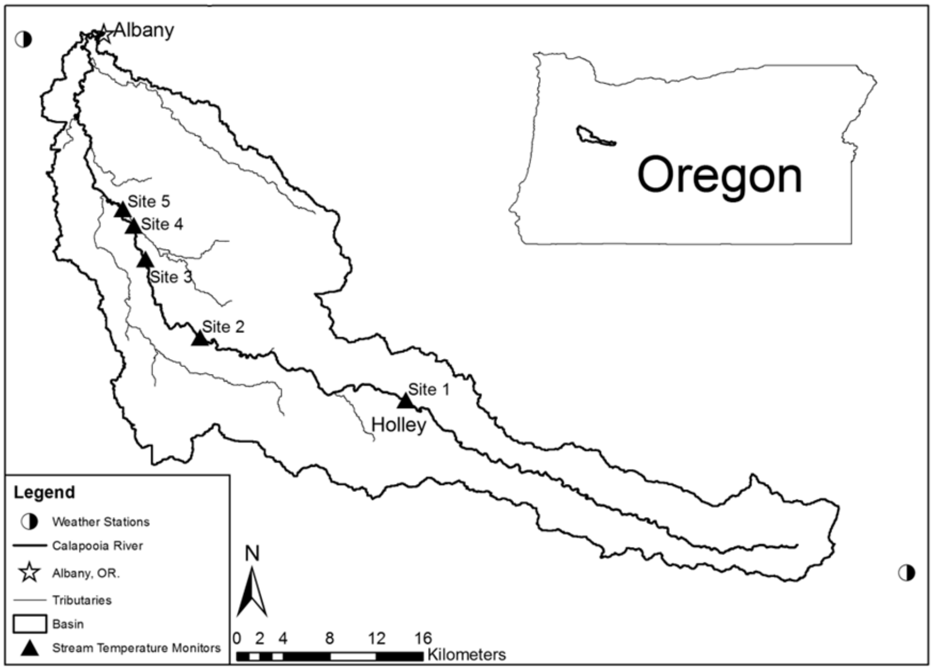

Figure 1. The Calapooia basin in northwest Oregon.

model that can be used to quantify the impact of different land management practices on soil and water quality (Arnold et al., 2012; Neitsch et al., 2005). It is a comprehensive model that includes components of hydrology, weather, sedimentation, erosion, soil and water temperature, plant growth, nutrients, pesticides, and agricultural management practices (Arnold et al., 2012; Neitsch et al., 2005).

For this investigation, we used ArcSwat 2009.10.1 along with ArcGIS 10 to delineate the Calapooia basin using a 1 arc-second National Elevation Dataset (~30 m) available from the U.S. Geological Survey (Gesch et al., 2002). The Calapooia basin was divided into 27 separate subbasins and a total of 176 hydrologic response units (HRUs) characterized by different combinations of land use, land cover, soil type, and slope. The state soil geographic (STATSGO) database for Oregon was used to specify soil types, and land use was determined from the USGS National Water-Quality Assessment (NAWQA) program (https://water.usgs.gov/nawqa/). Nine categories of land use were characterized. Generally, these included the forests of the upper Calapooia basin (~41%), the agricultural (~34%) and pasture (~11%) lands of the lower Calapooia, and the remaining wetlands (~10%), urban areas (~3%), and water bodies (<1%).

We acknowledge that this setup is a fairly simplistic model of the basin. It was specifically chosen to disregard small, intermittent tributaries and model stream temperature for the Calapooia River only.

Two weather stations were chosen for precipitation and temperature inputs from the two main topographical regions: the H.J. Andrews Experimental Forest meteorological site (44° 12′ 42″ N, 122° 15′ 21″ W, 482 m elev.), and the Hyslop weather station (44° 38′ 03″ N, 123° 11′ 24″ W, 70 m elev.) in Corvallis, Oregon. Available daily weather data from 1970 through 2012 were used. In addition, observed daily streamflow and stream temperature data were utilized for calibration and validation of SWAT. A USGS streamflow gauge provided daily discharge measurements for the Calapooia River at Albany, Oregon, from 1970 through 1981. The gauge was removed in 1981, preventing more recent data from being obtained. Measured stream temperature data were obtained from Joe Ebersole of the U.S. EPA Western Ecological Division for five sites on the mainstem Calapooia River from 2009 through 2012 (J. Ebersole, personal communication). Dates of the available data, split into calibration and validation regions, are shown in figure 2. Daily averages were aggregated from the initial 15 min sampling times. Stream temperatures were measured in °C with an accuracy of 0.3°C and precision of 0.2° (J. Ebersole, personal communication).

MODIFIED STREAM TEMPERATURE MODEL

The original stream temperature model within SWAT was introduced by Stefan and Preud'homme (1993) and assumes that stream temperature is linearly related to air temperature, as shown in equation 1:

$$T_{\text{stream}} = 5.0 + 0.75 T_{\text{air}} \tag{1}$$

This model is highly simplistic, does not take into account the overall stream network, and often overestimates stream temperatures when compared to observed data (Ficklin et al., 2012). Therefore, Ficklin et al. (2012) developed a modified stream temperature model within SWAT that utilizes the distributed hydrologic network. The modified model requires additional Fortran files to accompany the standard SWAT distribution, which are available upon request from Darren Ficklin at dficklin@indiana.edu. The model has been tested with SWAT2009 rev. 591. The modified model uses water temperatures from upstream, surface runoff, groundwater flow, and snowmelt, as well as travel

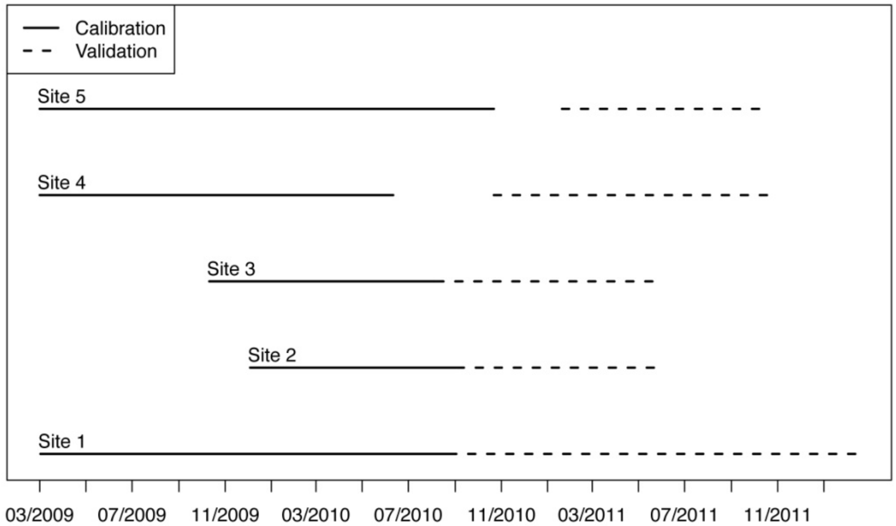

Figure 2. Time periods of available stream temperature data for five different sites on the Calapooia River.

time and exposure to the ambient air. Equations 2 through 4 demonstrate explicitly the stream temperature calculations for each subbasin. First, equation 2 shows the basic mixing model to estimate local water temperature:

$$T_{\mathrm{w,local}} = \frac{\left[T_{\mathrm{snow}} Q_{\mathrm{snow,sub}}\right] + \left[T_{\mathrm{GW}} Q_{\mathrm{GW,sub}}\right] + \lambda\left[Q_{\mathrm{surf,sub}} + Q_{\mathrm{lat,sub}}\right] T_{\mathrm{air,lag}}}{Q_{\mathrm{sub,yld}}} \quad (2)$$

where $T_{\mathrm{w,local}}$ is the local water temperature, T_{snow} and $Q_{\mathrm{snow,sub}}$ are the snowmelt temperature and flow contribution within the subbasin, T_{GW} and $Q_{\mathrm{GW,sub}}$ are the groundwater temperature and flow contribution within the subbasin, λ is a calibration coefficient, $T_{\mathrm{air,lag}}$ is the air temperature calculated with a user-specified lag, $Q_{\mathrm{surf,sub}}$ is the surface water runoff contribution within the subbasin, $Q_{\mathrm{lat,sub}}$ is the soil water lateral flow contribution within the subbasin, and $Q_{\mathrm{sub,yld}}$ is the total water yield within the subbasin. Temperatures are given in °C, flows are given in m^3 day^{-1}, the calibration coefficient is dimensionless, and the air temperature is the average air temperature in °C as averaged over a user-specified integer of days, or lag.

The local stream temperature calculation for the subbasin is then adjusted in order to include contributions from upstream subbasins, as shown in equation 3:

$$T_{\mathrm{w,adj}} = \frac{T_{\mathrm{w,upstream}}\left[Q_{\mathrm{outlet}} - Q_{\mathrm{sub,yld}}\right] + T_{\mathrm{w,local}} Q_{\mathrm{sub,yld}}}{Q_{\mathrm{outlet}}} \quad (3)$$

where $T_{\mathrm{w,adj}}$ is the adjusted water temperature. The first term in the numerator of equation 3 is the upstream contribution to the stream temperature, and the second term in the numerator is the local contribution to the stream temperature. These contributions are divided by the total flow leaving the subbasin (Q_{outlet}) to achieve the adjusted stream temperature.

Finally, the stream temperatures are allowed to interact with the surrounding air temperature for some duration or travel time as the stream travels through the subbasin. This is demonstrated in equation 4:

$$\begin{aligned} T_{\mathrm{w}} &= T_{\mathrm{w,adj}} + K\left[T_{\mathrm{air}} - T_{\mathrm{w,adj}}\right] TT && \text{if } T_{\mathrm{air}} > 0 \\ T_{\mathrm{w}} &= T_{\mathrm{w,adj}} + K\left[\left(T_{\mathrm{air}} + \varepsilon\right) - T_{\mathrm{w,adj}}\right] TT && \text{if } T_{\mathrm{air}} < 0 \end{aligned} \quad (4)$$

Essentially equation 4 describes one-dimensional heat transfer between the air and stream, where TT is the travel time of the stream (h), and K is a bulk heat transfer coefficient (h^{-1}). TT is calculated internally by the SWAT model, while K is a calibration parameter ultimately decided by the user. The calibration parameter ε is required to adequately model water temperature pulses when T_{air} is below but near 0°C (Ficklin et al., 2012).

The above modified stream temperature model in equations 2 through 4 uses four calibration parameters to adjust the model to match observed measurements of stream temperature. Ficklin et al. (2012) showed that these coefficients can achieve better calibration results when they are allowed to vary seasonally. Therefore, the modified model specifies four calibration parameters for each of four time durations, totaling 16 parameters for basin calibration. While Ficklin et al. (2012) introduced and tested the modified stream temperature model within SWAT, the model has only been calibrated at the basin level, using a single-site, manual calibration method. In this study, we extend the applicability of this stream temperature model by demonstrating a new automatic calibration technique that calibrates the parameters at the subbasin scale using NSGA-II. That is, the 16 parameters will be adjusted automatically for each subbasin within the basin. First, an overview of the genetic algorithm will be described.

NONDOMINATED SORTING GENETIC ALGORITHM II

Genetic algorithms (GA) use the idea of evolution's competitive selection as a computational concept to achieve solutions to optimization problems (Deb et al., 2002). Essentially, possible solutions are treated as individuals within a population. New solutions are found via a simulated reproduction within the population, and the fittest of the population are selectively chosen for the next generation, where fit is determined by the best optimization of a chosen objective function.

The Nondominated Sorting Genetic Algorithm II (NSGA-II) is a genetic algorithm that relies on a nondominated sorting algorithm, an elitist selection method, and a crowding distance parameter to determine fit for multiple objectives (Deb et al., 2002). The method finds a set of individuals that approximate the Pareto optimal front, or solutions that are completely nondominated by any other solutions. This means that no additional solution can be found that further optimizes one objective without diminishing the optimization of another objective (Deb et al., 2002). NSGA-II is fast and efficient for a small number of objectives and has been widely used for multi-objective calibration and validation of hydrological models (Bekele and Nicklow, 2007; Confesor and Whittaker, 2007; Whittaker et al., 2010; Zhang et al., 2010). As mentioned previously Whittaker et al. (2010) demonstrated that NSGA-II can efficiently and effectively find optimal parameter sets to calibrate SWAT using large sets of calibration parameters without suffering from overparameterization and overfitting. Therefore, we adopted their methodology in order to utilize NSGA-II to calibrate SWAT for streamflow at the basin outlet and temperature at five sites on the Calapooia River.

ALGORITHM SETUP AND IMPLEMENTATION

An overview of the calibration setup utilizing NSGA-II and SWAT is shown in figure 3. Basically, *N* individuals are generated where each individual is comprised of a set of calibration parameters. The individuals are initially chosen randomly within their specified upper and lower limits. Parallel computing enables each individual to be evaluated simultaneously on a separate processor. This means that each individual, or set of parameters, is sent to SWAT as input and run on each node simultaneously. The number of individuals able to be processed is limited by the number of parallel nodes (computer processors) available to the user. The calibration techniques described here were coded in R and called SWAT as a Fortran subroutine from within R using the ".Fortran" command. For computation, we used 100 nodes of a 206-node Beowulf cluster built by Gerald Whittaker at the USDA-ARS in Corvallis, Oregon (Whittaker, 2004).

The outputs from the individual SWAT runs are received from each of the nodes and are used along with measured data sets to calculate one or more objective functions. NSGA-II sorting is then used to rank the individuals in terms of their success at optimizing the objective functions, and crossover and mutation computations are performed to produce the next generation of individuals. The process is then repeated on the new population for several hundred or thousand generations until stopping criteria are met. Different stopping criteria can be specified to ensure convergence. Typically, the algorithm is run until changes in the objective functions are below a specified threshold. For this study, we ran the algorithm for 1,000 generations to ensure that no further optimal solutions could be found. It should be mentioned that the use of optimization algorithms for calibration introduces complexity to the standard SWAT model. This is manifested by the computational time required to achieve an optimal set of calibration parameters. For our application, computational time was not significant and did not hinder model calculation.

The calibration techniques for streamflow and temperature are similar; however, the parameters and the objective functions differ. The modified stream temperature model is highly dependent on the flow values at each subbasin. Therefore, we first calibrated SWAT for streamflow before calibrating for stream temperature. Note that the hydrological parameters, and resulting hydrology, are not dependent on the stream temperature model. This optimization routine is capable of simultaneously calibrating streamflows and temperatures. However, the observed data available for flow at Albany exists 30 years prior to the observed stream temperature data. Rather than running 30-year SWAT simulations, we calibrated the flow and temperature separately. The calibration parameters for streamflow are shown in table 1.

The streamflow parameters include nine basinwide parameters, 13 subbasin-level parameters, and 18 HRU-level parameters. The parameters, as well as the upper and lower limits, were specified following suggestions made by Whittaker et al. (2010). For computing the objective functions, daily flow measurements at the basin outlet at Albany, Oregon, were used. Flow measurements were not available concurrently with stream temperature measurements, and therefore the calibration (1970-1975) and validation (1976-

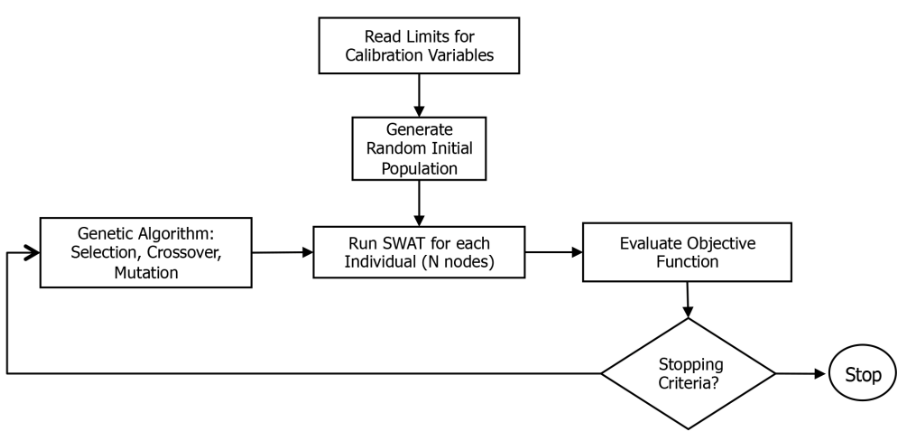

Figure 3. Flowchart of automatic calibration of SWAT using a genetic algorithm.

Table 1. Parameters used for discharge calibration using the SWAT model.

Parameter	Definition	Limits[a]	Resolution[b]
PHU	Heat units to bring plant to maturity	0.0 to 1500	HRU
SOL_K	Saturated hydraulic conductivity (mm h^{-1})	0.75 to 1.1*	HRU-S
SOL_AWC	Available water capacity (mm mm^{-1})	0.75 to 1.1*	HRU-S
SOL_CRK	Crack volume potential	0.0 to 0.3	HRU
CH_N2	Manning's n for main channel	0.014 to 0.075	SUB
OV_N2	Manning's n for overland flow	0.1 to 4.0	HRU
CANMX	Maximum canopy storage (mm)	0.0 to 5.0	HRU
ESCO	Soil evaporation compensation factor	0.1 to 1.0	HRU/W
EPCO	Plant water uptake compensation factor	0.1 to 1.0	HRU/W
REVAPMN	Threshold depth (mm)	0.001 to 200	HRU
ALPHA_BF	Alpha factor for baseflow (days)	0.04 to 1.0	HRU
GW_DELAY	Groundwater delay time (days)	0.0 to 60.0	HRU
GW_REVAP	Groundwater revap coefficient (days)	0.02 to 0.20	HRU
SURLAG	Surface runoff lag coefficient	1.0 to 21.0	W
MSK_CO1	Calibration coefficient	0.0 to 3.0	W
MSK_CO2	Calibration coefficient	0.0 to 5.0	W
MSK_X	Weighting factor	0.0 to 0.50	W
TRNSRCH	Transmission loss	0.1 to 0.90	HRU
EVRCH	Reach evaporation adjustment factor	0.1 to 0.90	HRU
SLSUBBSN	Average slope length (m)	0.75 to 1.25*	HRU
SLSOIL	Slope length for lateral slope length (m)	0.0 to 30.0	W
HRU_SLP	Average slope steepness (m m^{-1})	0.75 to 1.25*	W
TIMP	Snow pack temperature lag factor	0.01 to 1.00	W
SMFMN	Minimum melt factor for snow (mm °C^{-1} day^{-1})	1.4 to 6.9	SUB
SMFMX	Maximum melt factor for snow (mm °C^{-1} day^{-1})	1.4 to 6.9	SUB
CHL_1	Longest tributary channel length in subbasin	0.75 to 1.25*	SUB
CH_S1	Average slope of tributary channels	0.75 to 1.25*	SUB
CH_W1	Average width of tributary channels	0.75 to 1.25*	SUB
CH_N1	Manning's n for the tributary channels	0.75 to 1.25*	SUB
CH_K1	Effective hydraulic conductivity of tributary	0.025 to 10.0	SUB
CH_L2	Length of main channel	0.75 to 1.25*	SUB
CH_S2	Average slope of main channel	0.75 to 1.25*	SUB
CH_W2	Average width of main channel	0.75 to 1.25*	SUB
CH_D	Average depth of main channel	0.75 to 1.25*	SUB
CH_K2	Effective hydraulic conductivity of main channel	0.025 to 10.0	SUB
CH_WDR	Channel width to depth ratio	0.75 to 1.25*	SUB
ALPHA_BNK	Alpha factor for bank storage recession curve	0.001 to 0.99	SUB
GWQMN	Threshold depth of water in shallow aquifer	0.0 to 200.0	HRU
RCHRG_DP	Recharge to deep aquifer	0.0 to 1.0	HRU
GW_SPYLD	Specific yield for shallow aquifer	0.001 to 0.009	HRU

[a] Parameters marked with an asterisk (*) are multipliers of the default value in the SWAT model setup.
[b] HRU = hydrologic response unit, HRU-S = HRU soil layer, SUB = subbasin, and W = basin.

1981) durations of the flow data were limited to the most recent periods when continuous data was available. In addition, an automated baseflow filter was used to separate flow into event-driven flow and baseflow (Arnold et al., 1995). Studies have pointed out the benefit of separating objective functions into baseflow and streamflow components (Whittaker et al., 2010; Zhang et al., 2010). The main reason is that calibration parameters will differ when trying to optimize either peakflow or baseflow conditions. The genetic algorithm was set up to have two objectives: to minimize the root mean square error (RMSE) between the measured and observed values for event-driven flow and baseflow separately. The genetic algorithm is then able to find the approximation to the Pareto front, or set of optimal solutions, which invariably consist of trade-offs between the two objectives. The user must then select an individual from the Pareto front to use as the optimal set of calibration parameters.

Once the flow was calibrated within SWAT, the stream temperature could be simulated and calibrated using the flow-calibrated model. Stream temperatures were calibrated separately at the basin and subbasin levels and will be compared. Table 2 shows the temperature calibration parameters used for both setups. The four calibration parameters for stream temperature were allowed to vary for four separate seasonal time periods, making a total of 16 parameters to be calibrated at the basin level (Ficklin et al., 2012). For the subbasin calibration, the total number of parameters to be calibrated was 432, or 16 multiplied by the 27 subbasins in the basin. Both the basin and subbasin level calibration techniques utilize a single objective function: the total sum of RMSE values between measured and simulated stream temperatures for all five stream temperature monitoring sites. The choice and number of objective functions is discussed further in the Discussion section.

Due to the sparse stream temperature data available, the calibration and validation time periods were limited. Figure 2 shows the time periods of observed stream temperature used for calibration and validation between 2009 and

Table 2. Parameters used for stream temperature calibration.

Parameter	Description	Limits	Resolution
λ	Calibration multiplier	0.0 to 1.0	W or SUB
K	Bulk heat transfer coefficient (h^{-1})	0.0 to 1.0	W or SUB
ε	Snowmelt adjustment factor (°C)	0.0 to 10	W or SUB
Lag	Integer of days over which to average air temperature	0 to 14	W or SUB

2012. Efforts were made to evenly divide the calibration and validation periods while maintaining continuous data sets.

RESULTS

STREAMFLOW CALIBRATION AND VALIDATION

As mentioned previously, streamflow was calibrated at the USGS site in Albany, Oregon, using parameters suggested by Whittaker et al. (2010). A total of 1,000 generations were calculated, resulting in a population of nondominated individuals, or sets of calibrated parameters, that optimally and jointly minimized the RMSE between simulated and observed flow for both baseflow and event-driven flow. Since all individuals are equally optimal, the individual chosen is decided by the user according to the particular modeling needs. For example, a researcher interested in sediment analysis may prefer a streamflow simulation that weighs the event-driven objective more heavily than baseflow (Confesor and Whittaker, 2007). For our purposes, we chose the individual in the middle of the front that weighs each objective equally. The resulting flow for the calibration and validation periods is plotted in figure 4.

Table 3 shows the metrics used to determine the goodness-of-fit between the simulated flow and observed flow. SWAT has been shown to provide remarkably reliable results when predicting streamflow and crop yield, even when uncalibrated (Srinivasan et al., 2010). However, it is clear that the calibrated SWAT flows are superior to the uncalibrated values. For example, the mean error (ME), mean average error (MAE), root mean squared error (RMSE), and percent bias (PBIAS%) were all significantly reduced for both the calibration and validation periods using the genetic algorithm. In addition, three efficiency criteria were used to determine the goodness-of-fit, including the Nash-Sutcliffe (NSE), Kling-Gupta (KGE), and volumetric (VE) efficiencies (Gupta et al., 2009; Krause et al., 2005; Nash and Sutcliffe, 1970). All metrics show a clear increase in fit values when the genetic algorithm calibration technique is utilized.

Table 3. Streamflow calibration and validation statistics.

	Calibration (1970-1975)		Validation (1976-1981)	
Statistic	No Calib.	GA Calib.	No Calib.	GA Calib.
ME	6.97	1.52	4.91	-2.42
MAE	15.71	7.33	10.81	6.01
RMSE	37.71	18.12	25.89	14.28
BIAS%	24.3	5.6	26.1	-12.9
NSE	0.41	0.84	0.43	0.83
KGE	0.58	0.87	0.57	0.8
VE	0.45	0.73	0.43	0.68

To demonstrate this further, figure 5 shows kernel density estimates calculated for the differences between simulated and observed flows for both uncalibrated and calibrated outputs. Figure 5 clearly shows that the calibrated flows demonstrate more narrowly peaked density functions, indicating a closer match to the observed data sets.

TEMPERATURE CALIBRATION AND VALIDATION (BASIN VS. SUBBASIN SCALES)

After calibrated streamflow parameters were available, these parameters were used within the original and modified stream temperature models. The modified stream temperature model, by Ficklin et al. (2012), was calibrated at five stream temperature monitoring sites using the automatic calibration technique described previously. The routine was performed to calibrate at the basin and subbasin scales, separately. Calibration and validation time periods were defined as shown in figure 2. Each calibration used the same set of optimized flow parameters and was run for 1,000 generations.

Figures 6 and 7 compare the resulting simulated daily stream temperatures between the original SWAT model and the subbasin-calibrated modified model. Table 4 shows the goodness-of-fit metrics between the three simulations and the observed stream temperatures. As expected, the

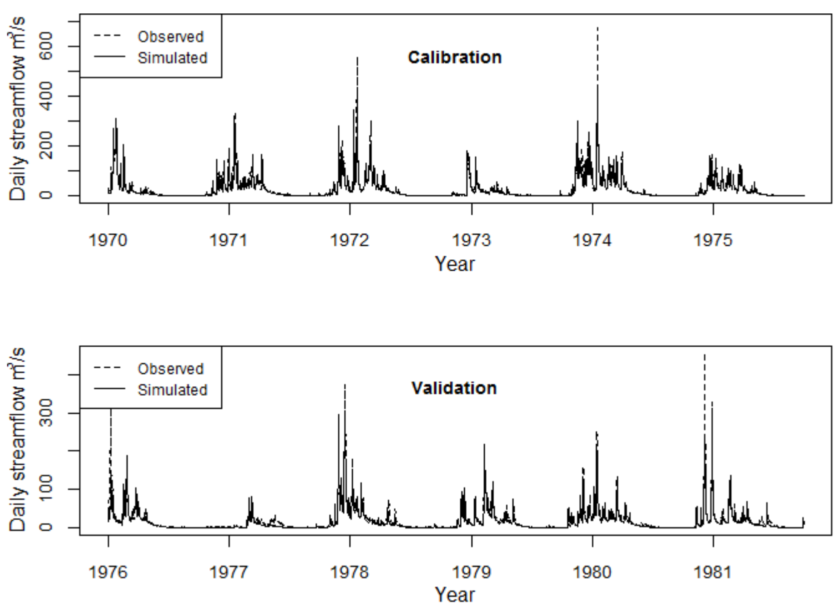

Figure 4. Observed and simulated daily streamflow from SWAT for the calibration (1970-1975) and validation (1976-1981) periods. The optimal set of calibration parameters that weighed each objective equally was used to produce the simulated output in SWAT.

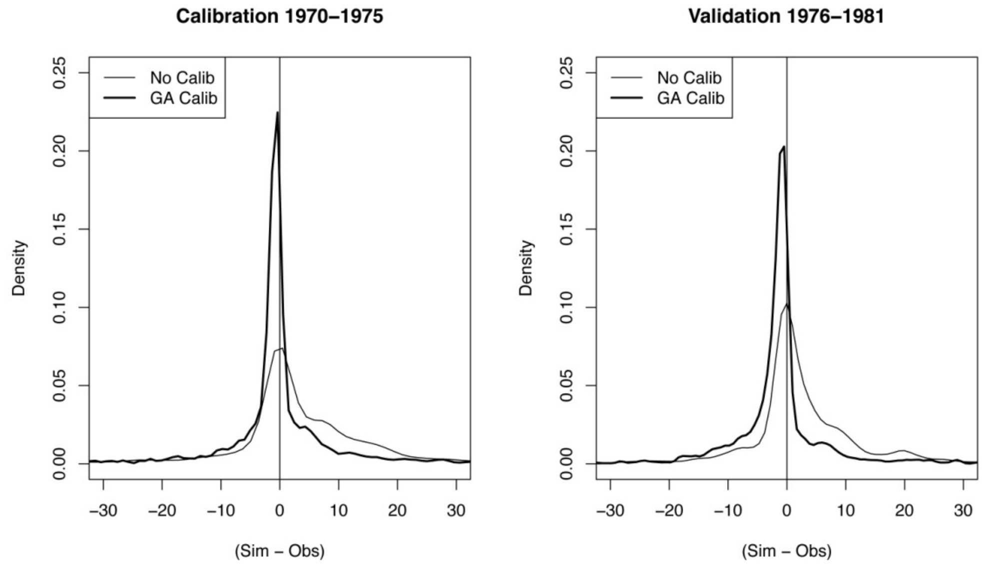

Figure 5. Kernel density estimation for the differences between daily simulated and observed streamflow at Albany, Oregon, as calculated by the uncalibrated SWAT model and the SWAT model calibrated with the genetic algorithm. The calibration (1970-1975) and validation (1976-1981) periods are shown.

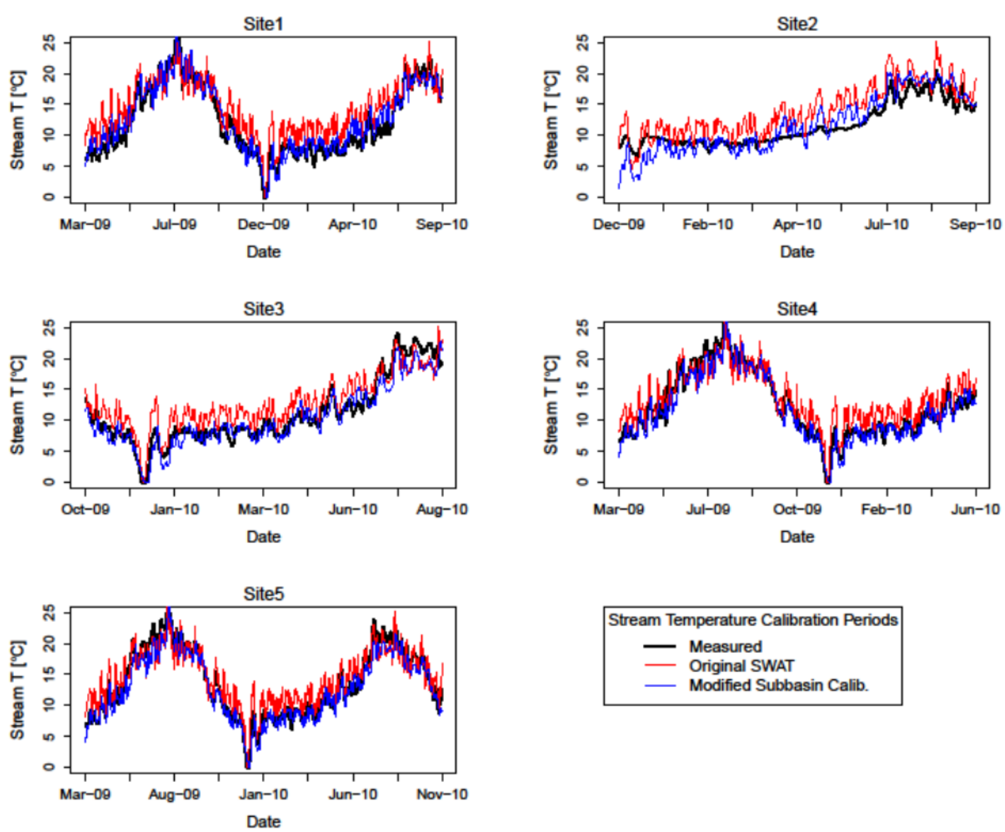

Figure 6. Observed and simulated daily temperature values for five different sites on the Calapooia River during calibration.

modified temperature model (at both calibrated scales) more closely matched the observed data than the original SWAT model. Note that all three simulations used the same flow-calibrated parameters. When averaging the stream temperature metrics for all sites, the average calibration period PBIAS% was 13.62, -5.92, and -2.92 for the original SWAT model, the modified model with basin calibration, and the modified model with subbasin calibration, respectively. The validation period showed similar PBIAS% values of 15.7, -6, and -4.08, respectively. Therefore, it is clear that the modified SWAT stream temperature model consistently reduced the overestimation of stream temperatures seen in the original model.

For the modified model, the subbasin calibration consistently improved the results compared to basin calibration. From table 4, the RMSE averaged for all sites was

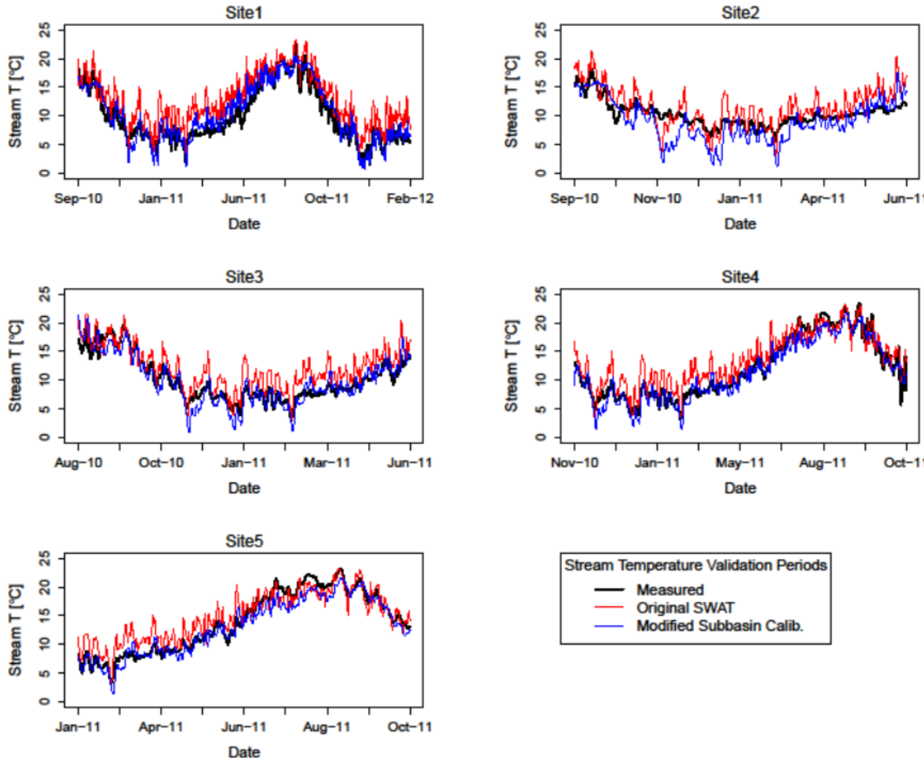

Figure 7. Observed and simulated daily temperature values for five different sites on the Calapooia River during validation.

Table 4. Stream temperature calibration and validation statistics at five sites on the Calapooia River.

Period	Statistic	Model	Site 1	Site 2	Site 3	Site 4	Site 5
Calibration	RMSE	Original	2.94	2.93	2.59	2.25	2.18
		Basin	2.17	2.08	1.9	2.08	2.09
		Subbasin	1.71	2.03	1.77	1.64	1.7
	PBIAS%	Original	16.9	20.5	15.2	9.8	5.7
		Basin	-5.3	1.4	-8.3	-8.7	-8.7
		Subbasin	2.5	1.5	-6	-5.7	-6.9
	R^2	Original	0.88	0.83	0.88	0.89	0.9
		Basin	0.87	0.88	0.91	0.89	0.91
		Subbasin	0.91	0.85	0.91	0.92	0.94
	NSE	Original	0.73	0.24	0.76	0.82	0.85
		Basin	0.85	0.62	0.87	0.85	0.86
		Subbasin	0.91	0.64	0.89	0.91	0.91
Validation	RMSE	Original	3.18	2.44	2.59	2.3	2.14
		Basin	2.17	2.63	1.98	2.07	1.79
		Subbasin	1.79	2.51	1.76	1.69	1.62
	PBIAS%	Original	27.3	13.4	20	11.3	6.5
		Basin	-5.1	-12.5	-2.3	-5.2	-4.9
		Subbasin	4.2	-11	-2.7	-5.1	-5.8
	R^2	Original	0.88	0.67	0.83	0.9	0.89
		Basin	0.81	0.57	0.79	0.86	0.91
		Subbasin	0.88	0.64	0.83	0.91	0.93
	NSE	Original	0.54	-0.29	0.55	0.82	0.84
		Basin	0.79	-0.49	0.74	0.85	0.89
		Subbasin	0.86	-0.37	0.79	0.9	0.91

2.58, 2.06, and 1.77, respectively, for the three models for the calibration period and 2.53, 2.13, and 1.87, respectively, for the validation period. Each subbasin had its own set of calibration parameters, which greatly increased the model's ability to more accurately model stream temperatures throughout the basin.

Site 1 shows the most dramatic improvement in metrics between the basin calibration and the subbasin calibration. The NSE values were 0.73, 0.85, and 0.91, respectively, for the three models for the calibration period and 0.54, 0.79, and 0.86, respectively, for the validation period. In addition, the PBIAS% decreased from 16.9 to -5.3 to 2.5, respectively, for the three models during the calibration period and from 27.3 to -5.1 to 4.2, respectively, during the validation period. Site 1 is closest to the western Cascade Mountains and is characterized by very different land use, topography, and soil types from the remaining sites, which are situated within the valley dominated by agriculture. Therefore, it is evident that a different set of calibration parameters should apply. The basin calibration, while performing better than the original SWAT model, was unable to account for these differences between subbasins.

The improvement with the subbasin calibration of the modified model is again evident in figures 8 and 9, which show kernel density estimates for the differences between the simulated and observed stream temperatures as compared between the original SWAT model and the modified calibrated models. The original SWAT model clearly consistently overestimated the observed stream temperatures for all sites. In addition, the subbasin-calibrated modified model temperatures were better than the basin-level for all sites except site 2. Site 2 was the shortest data set available and exhibited the worst metrics for all three models tested. Sites 3, 4, and 5 showed incremental improvements from the basin to the subbasin calibrations. However, site 1 again showed dramatic improvement of the subbasin calibration over the basin calibration.

The calibration parameters selected by the genetic algorithm for the basin-level and subbasin-level methods are shown in tables 5 and 6. For reference, note that Ficklin et al. (2012) performed a sensitivity analysis for these four parameters within five different basins. They concluded that λ was the most sensitive parameter, followed by K and ε, whereas Lag was shown to be insensitive.

The parameter λ adjusts the effect of surface runoff and lateral soil flow on stream temperatures, as shown in equation 2. Values closer to one indicate that surface runoff and lateral soil flows are closer to air temperature, which are usually the result of shallow soils or steeper slopes. The basin calibration gave values near 1.0 for all time regions. However, the subbasin calibration showed highly variable λ values for the different sites and different time regions. Since this variable is highly sensitive, according to Ficklin et al. (2012), its variability may explain why the subbasin calibration was able to more accurately simulate heterogeneous stream temperature fluctuations between the sites.

The parameter K is the heat transfer coefficient, shown in equation 4, and represents the transfer of heat from the air to streams during travel time. For basinwide calibration, the values are near zero for all time regions. Again, the subbasin calibration method exhibited a wide variety of K values, indicating diversity in the exchange of heat from air to streams for the different sites and throughout the year.

The parameter ε accounts for snowmelt pulses when the stream temperature is close to 0°C and the air temperature is less than 0°C, as shown in equation 4. Larger values of this parameter typically exist at higher elevations or in colder conditions where snowmelt occurs. Days 1 to 65 show that site 1 has the largest value; however, this is not consistent for all time periods.

Overall, allowing each subbasin to utilize its own calibration variables has been shown to improve stream temperature simulations, as shown in figure 10. Figure 10 shows the resulting modified stream temperature simulations for 2009-2012 for five sites on the Calapooia River. Vertical lines indicate the separation between the calibration, or training set, data and the validation data used. No vertical lines are shown for sites 4 and 5, where the calibration and validation sets were chosen based on natural breaks in the measured data. Clearly, the subbasin calibra-

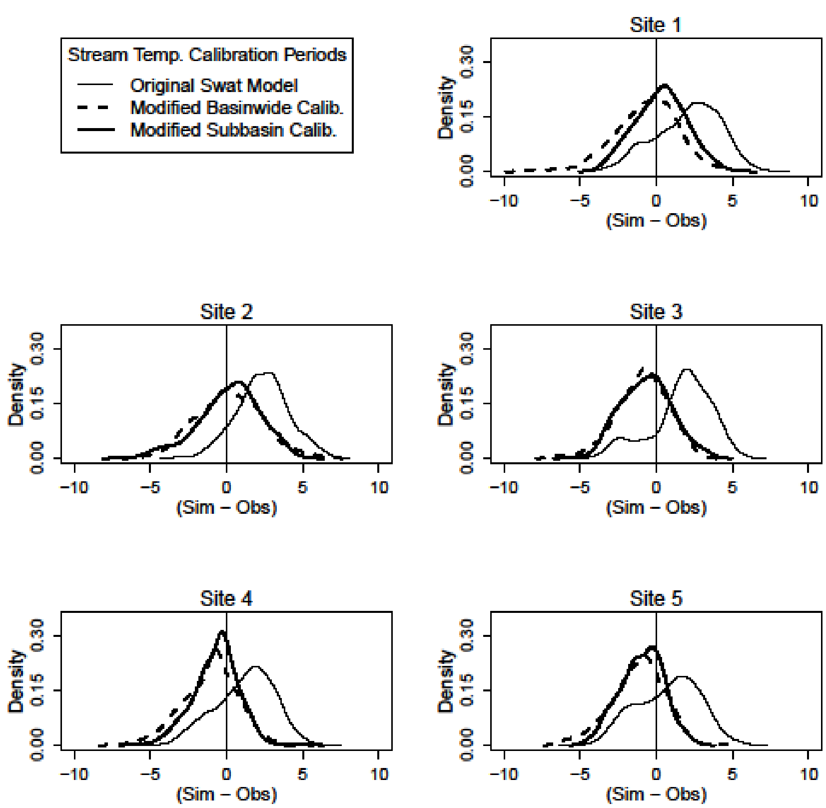

Figure 8. Kernel density estimation for differences between measured and simulated stream temperature data for five sites on the Calapooia River during calibration.

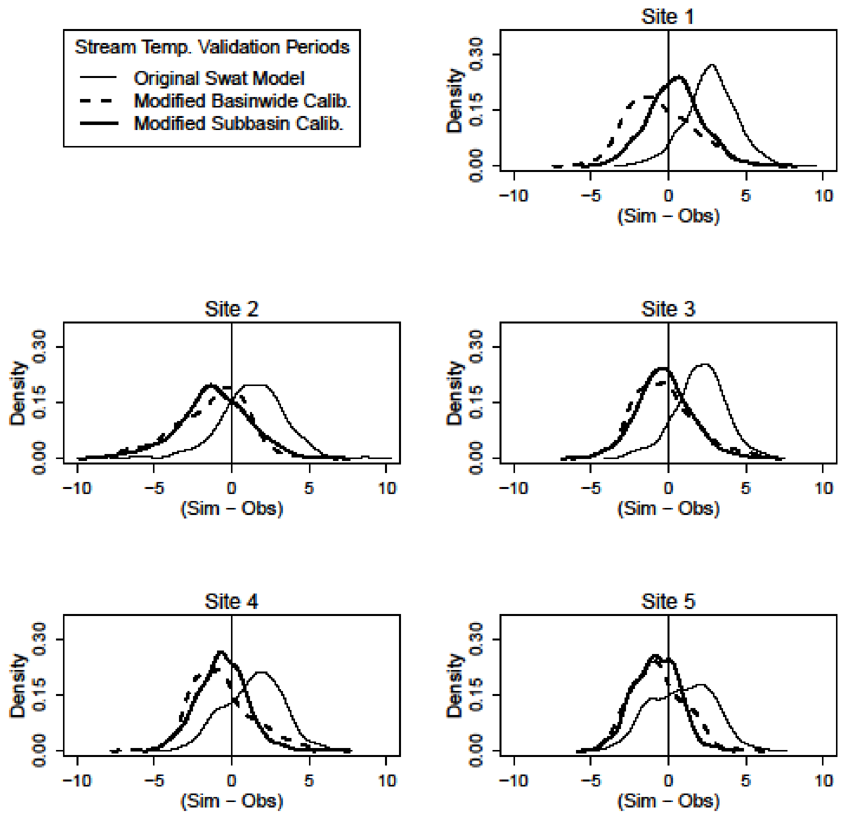

Figure 9. Kernel density estimation for differences between measured and simulated stream temperature data for five sites on the Calapooia River during validation.

tion of the modified model improved the stream temperature simulations for the Calapooia River. Therefore, this model better captured the variability of stream temperatures on the Calapooia River.

Table 5. Optimally chosen parameters for basin-level calibration.

	λ	*K*	ε	Lag
Days 1-65	0.999	0.004	5.214	3
Days 66-125	0.998	0.008	4.017	1
Days 126-285	0.923	0.051	5.862	14
Days 286-366	0.999	0.017	0.034	4

Table 6. Optimally chosen parameters for subbasin-level calibration.

	Site 1	Site 2	Site 3	Site 4	Site 5
Days 1-65					
λ	0.81	0.80	0.90	0.92	0.71
K	0.62	0.04	0.02	0.04	0.01
ε	7.31	1.58	2.09	2.85	4.34
Lag	1.00	13.00	2.00	13.00	11.00
Days 66-125					
λ	0.60	0.23	0.33	0.87	0.25
K	0.53	0.68	0.59	0.64	0.50
ε	4.83	2.58	4.57	4.20	8.75
Lag	5.00	8.00	9.00	9.00	5.00
Days 126-285					
λ	0.74	0.71	0.47	0.42	0.56
K	0.95	0.49	0.43	0.43	0.55
ε	4.72	4.23	5.15	1.92	4.36
Lag	13.00	6.00	13.00	3.00	12.00
Days 286-366					
λ	0.12	0.51	0.18	0.95	0.54
K	0.47	0.70	0.51	0.62	0.46
ε	2.73	6.13	3.36	6.28	4.37
Lag	9.00	6.00	9.00	7.00	9.00

DISCUSSION

Stream temperature simulations at five sites in the Calapooia River using the Ficklin et al. (2012) model within SWAT were greatly improved by using an automatic, subbasin-level calibration method. This method, therefore, can be extremely useful when simulating stream temperatures at multiple locations within large, diverse basins. A drawback of this particular investigation was the lack of observed streamflow data for recent years as well as the short durations of stream temperature data available for the Calapooia River. Longer periods of measured data would allow for more accurate calibration and would improve the validation results.

As described earlier, our study was limited to a single objective function: minimizing the sum of all RMS errors for all the measurement sites. In fact, the calibration routine was tested using each of the five sites as separate objectives with NSGA-II; however, the five-objective calibration results were inferior to the single objective where all sites were superposed. This is contrary to findings of other studies, which insist that tension among multiple objectives improves the agreement between simulated and observed data (Bekele and Nicklow, 2007; Cao et al., 2006; Zhang et al., 2008; Zhang et al., 2010). One explanation for our result could be the relatively high number of objectives used. While NSGA-II is robust for multi-objective optimization, it may encounter problems in finding optimal solutions when a large number of objectives are specified (Beume et al., 2007). The use of more advanced genetic algorithms,

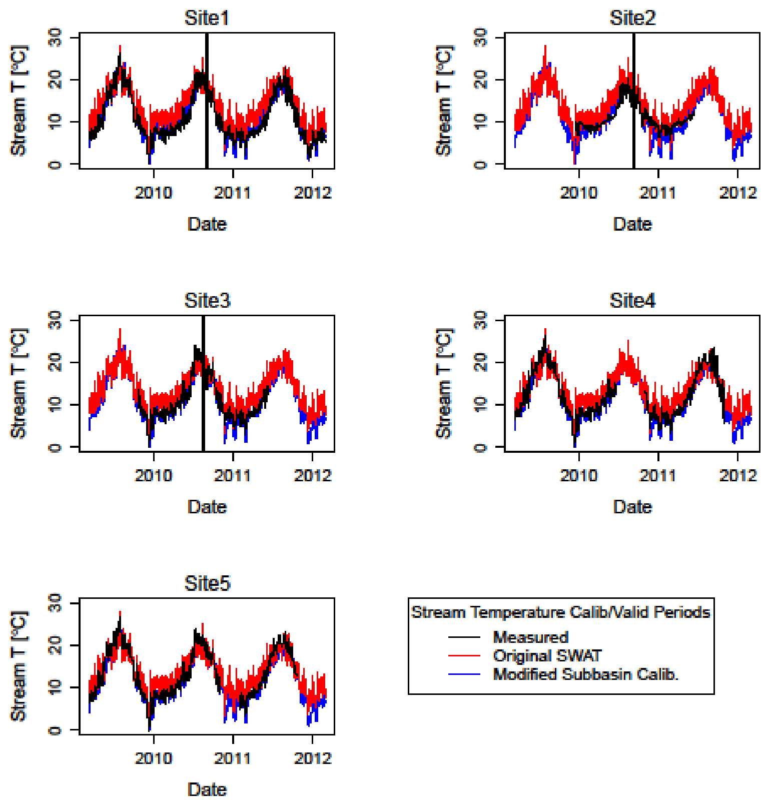

Figure 10. Simulated stream temperatures from 2009-2012 compared with observed data. Measured data (black) are shown along with the original linear model simulations (red) and the Ficklin et al. (2012) modified model simulations calibrated at the subbasin-level (blue).

such as SMS-EMOA, which calculates selection based on a dominated hypervolume, may allow for optimization of a larger number of objectives, for example, one objective for each site, while still keeping computational time low. This may achieve more accurate simulation results and allow for simultaneous calibration of streamflow and temperature simulations using one algorithm, assuming that simultaneous observed data are available.

In terms of describing the resulting optimal set of calibration parameters, this study does not show a consistent physical interpretation for all the calibration parameters. Further study is needed to analyze the potential relationships between these calibrated parameters and particular land uses, soil types, or topology. With a more detailed study, and systematic subbasin delineation, perhaps the calibration parameters could be tied to physical parameters, such as land use, soil depth, canopy covering, or riparian barrier width.

Finally, the modified stream temperature model takes into account different temporal regimes to capture the different impacts that hydrology and air temperature have on stream temperatures throughout the year. However, seasonal periods were selected before conducting the genetic algorithm calibration method. An improvement in the method would be to allow for fluctuating temporal scales, each containing its own set of calibration parameters within each subbasin. This may increase the computational time, but it would ensure that seasonality is correctly, and objectively, accounted for within the stream temperature model.

Conclusion

This investigation used a modified stream temperature model within SWAT, introduced by Ficklin et al. (2012), to simulate stream temperatures at five different sites along the Calapooia River in northwest Oregon. We used NSGA-II to automatically parameterize the models in order to accurately simulate both streamflow and temperature at multiple sites within the Calapooia basin. We are not aware of any other study that automatically calibrates the Ficklin et al. (2012) model for multiple sites and at the subbasin scale. The results showed that the Ficklin et al. (2012) model, whether calibrated for basinwide or subbasin levels, gave better results than the original SWAT stream temperature model for five sites on the Calapooia River. The results also showed that subbasin calibration improved stream temperature simulations compared to basinwide calibration. These results will be useful for other studies attempting to calibrate and simulate stream temperatures using the Ficklin et al. (2012) model. They will especially be pertinent to large, highly diverse basins exhibiting dramatic changes in topography, land use, and soil type.

References

Arnold, J. G., Allen, P. M., Muttiah, R., & Bernhardt, G. (1995). Automated base-flow separation and recession analysis techniques. *Ground Water, 33*(6), 1010-1018. http://dx.doi.org/10.1111/j.1745-6584.1995.tb00046.x.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van

Griensven, A., Van Liew, M. W., Kannan, N., & Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1491-1508. http://dx.doi.org/10.13031/2013.42256.

Bekele, E. G., & Nicklow, J. W. (2007). Multi-objective automatic calibration of SWAT using NSGA-II. *J. Hydrol., 341*(3-4), 165-176. http://dx.doi.org/10.1016/j.jhydrol.2007.05.014.

Beume, N., Naujoks, B., & Emmerich, M. (2007). Multiobjective selection based on dominated hypervolume. *European J. Oper. Res., 181*(3), 1653-1669. http://dx.doi.org/10.1016/j.ejor.2006.08.008.

Cao, W. Z., Bowden, W. B., Davie, T., & Fenemor, A. (2006). Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrol. Proc., 20*(5), 1057-1073. http://dx.doi.org/10.1002/hyp.5933.

Chaplot, V., Saleh, A., Jaynes, D. B., & Arnold, J. (2004). Predicting water, sediment, and NO_3-N loads under scenarios of land-use and management practices in a flat watershed. *Water Air Soil Poll., 154*(1-4), 271-293. http://dx.doi.org/10.1023/B:WATE.0000022973.60928.30.

Confesor, R. B., & Whittaker, G. W. (2007). Automatic calibration of hydrologic models with multi-objective evolutionary algorithm and Pareto optimization. *J. American Water Resources Assoc., 43*(4), 981-989. http://dx.doi.org/10.1111/j.1752-1688.2007.00080.x.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Computation, 6*(2), 182-197. http://dx.doi.org/10.1109/4235.996017.

Duan, Q. Y., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Res., 28*(4), 1015-1031. http://dx.doi.org/10.1029/91WR02985.

Ficklin, D. L., Luo, Y. Z., Stewart, I. T., & Maurer, E. P. (2012). Development and application of a hydroclimatological stream temperature model within the Soil and Water Assessment Tool. *Water Resources Res., 48*(1), W01511. http://dx.doi.org/10.1029/2011WR011256.

Ficklin, D. L., Stewart, E. P., & Maurer, E. P. (2013). Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada mountain range in California. *Water Resources Res, 49*(5), 2765-2782. http://dx.doi.org/10.1002/wrcr.20248.

Fohrer, N., Haverkamp, S., & Frede, H. G. (2005). Assessment of the effects of land use patterns on hydrologic landscape functions: Development of sustainable land use concepts for low mountain range areas. *Hydrol. Proc., 19*(3), 659-672. http://dx.doi.org/10.1002/hyp.5623.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The national elevation dataset. *Photogram. Eng. and Remote Sensing, 68*(1), 5-32.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol., 377*(1-2), 80-91. http://dx.doi.org/10.1016/j.jhydrol.2009.08.003.

Hoag, D. L., Giannico, G. R., Li, J., Garcia, T., Gerth, W., Mellbye, M., Mueller-Warrant, G., Griffith, S., Osmond, D. L., Jennings, G. D., Arabi, M., & Meals, D. W. (2012). Chapter 19: Calapooia watershed, Oregon: National Institute of Food and Agriculture—Conservation Effects Assessment Program. In *How to Build Better Agricultural Conservation Programs to Protect Water Quality: The National Institute of Food and Agriculture—Conservation Effects Assessment Project Experience*, 327-341. D. L. Osmond, D. W. Meals, D. L. Hoag, & M. Arabi, eds. Ankeny, Iowa: Soil and Water Conservation Society.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proc. IEEE Intl. Conf. Neural Networks*, vol. 4: 1942-1948. Piscataway, N.J.: IEEE.

Krause, P., Boyle, D., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci., 5*(5), 89-97. http://dx.doi.org/10.5194/adgeo-5-89-2005.

Li, Z., Liu, W.-Z., Zhang, X.-C., & Zheng, F.-l. (2009). Impacts of land use change and climate variability on hydrology in an agricultural catchment on the Loess Plateau of China. *J. Hydrol., 377*(1-2), 35-42. http://dx.doi.org/10.1016/j.jhydrol.2009.08.007.

Luo, Y., Ficklin, D. L., Liu, X., & Zhang, M. (2013). Assessment of climate change impacts on hydrology and water quality with a watershed modeling approach. *Sci. Total Environ., 450-451*, 72-82. http://dx.doi.org/10.1016/j.scitotenv.2013.02.004.

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models: Part I. A discussion of principles. *J. Hydrol., 10*(3), 282-290. http://dx.doi.org/10.1016/0022-1694(70)90255-6.

Naujoks, B., Beume, N., & Emmerich, M. (2005). Multi-objective optimization using S-metric selection: Application to three-dimensional solution spaces. In *Proc. 2005 IEEE Congress on Evolutionary Computation* , vol. 2: 1282-1289. Piscataway, N.J.: IEEE. http://dx.doi.org/10.1109/CEC.2005.1554838.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Williams, J. R., & King, K. W. (2005). *Soil and Water Assessment Tool Theoretical Documentation: Version 2005.* College Station, Tex.: Texas Water Resources Institute.

Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R., & Hauck, L. M. (2001). Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. American Water Resources Assoc., 37*(5), 1169-1188. http://dx.doi.org/10.1111/j.1752-1688.2001.tb03630.x.

Santhi, C., Kannan, N., Arnold, J. G., & Luzio, M. D. (2008). Spatial calibration and temporal validation of flow for regional-scale hydrologic modeling. *J. American Water Resources Assoc., 44*(4), 829-846. http://dx.doi.org/10.1111/j.1752-1688.2008.00207.x.

Srinivasan, R., Zhang, X., & Arnold, J. (2010). SWAT ungauged: Hydrological budget and crop yield predictions in the upper Mississippi River basin. *Trans. ASABE, 53*(5), 1533-1546. http://dx.doi.org/10.13031/2013.34903.

Stefan, H. G., & Preud'homme, E. B. (1993). Stream temperature estimation from air temperature. *Water Resources Bull, 29*(1), 27-45. http://dx.doi.org/10.1111/j.1752-1688.1993.tb01502.x.

Tang, Y., Reed, P., & Wagener, T. (2006). How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrol. Earth Syst. Sci., 10*(2), 289-307. http://dx.doi.org/10.5194/hess-10-289-2006.

USEPA. (2003). National management measures to control of nonpoint-source pollution from agriculture. EPA 841-B-03-004. Washington, D.C.: U.S. EPA, Office of Water. Retrieved from http://water.epa.gov/polwaste/nps/agriculture/agmm_index.cfm.

Vache, K. B., Eilers, J. M., & Santelmann, M. V. (2002). Water quality modeling of alternative agricultural scenarios in the U.S. corn belt. *J. American Water Resources Assoc., 38*(3), 773-787. http://dx.doi.org/10.1111/j.1752-1688.2002.tb00996.x.

White, K. L., & Chaubey, I. (2005). Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *J. American Water Resources Assoc., 41*(5), 1077-1089. http://dx.doi.org/10.1111/j.1752-1688.2005.tb03786.x.

Whittaker, G. (2004). Use of a Beowulf cluster for estimation of risk using SWAT. *Agron. J., 96*(5), 1495-1497. http://dx.doi.org/10.2134/agronj2004.1495.

Whittaker, G., Confesor, R., Luzio, M. D., & Arnold, J. G. (2010). Detection of overparameterization and overfitting in an automatic calibration of SWAT. *Trans. ASABE, 53*(5), 1487-1499. http://dx.doi.org/10.13031/2013.34909.

Zhang, X., Srinivasan, R., & Liew, M. V. (2008). Multi-site calibration of the SWAT model for hydrologic modeling. *Trans. ASABE, 51*(6), 2039-2049. http://dx.doi.org/10.13031/2013.25407.

Zhang, X., Srinivasan, R., Zhao, K., & Liew, M. V. (2009a). Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrol. Proc., 23*(3), 430-441. http://dx.doi.org/10.1002/hyp.7152.

Zhang, X. S., Srinivasan, R., & Bosch, D. (2009b). Calibration and uncertainty analysis of the SWAT model using genetic algorithms and Bayesian model averaging. *J. Hydrol., 374*(3-4), 307-317. http://dx.doi.org/10.1016/j.jhydrol.2009.06.023.

Zhang, X. S., Srinivasan, R., & Liew, M. V. (2010). On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrol. Proc., 24*(8), 955-969. http://dx.doi.org/10.1002/hyp.7528.

Zitzler, E., Laumanns, M., & Thiele, L. (2001). SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Technical report 103. Zurich, Switzerland: Swiss Federal Institute of Technology (ETH), Computer Engineering and Networks Laboratory (TIK).