Journal of Hydrology 556 (2018) 39–49

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability

Matthew R. Herman [a], A. Pouyan Nejadhashemi [a,*], Mohammad Abouali [a], Juan Sebastian Hernandez-Suarez [a], Fariborz Daneshvar [a], Zhen Zhang [b], Martha C. Anderson [c], Ali M. Sadeghi [c], Christopher R. Hain [d], Amirreza Sharifi [e]

[a] Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA
[b] Physical Sciences Division, Department of Statistics, University of Chicago, Chicago, IL 60637, USA
[c] Hydrology and Remote Sensing Laboratory, US Department of Agriculture-Agricultural Research Service, Beltsville, MD 20705, USA
[d] Marshall Space Flight Center, US National Aeronautics and Space Administration, Huntsville, AL 35811, USA
[e] Department of Energy & Environment, Government of the District of Columbia, Washington, DC 20002, USA

A R T I C L E I N F O

Article history:
Received 30 April 2017
Received in revised form 5 November 2017
Accepted 6 November 2017
Available online 9 November 2017
This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Saeid Eslamian, Associate Editor

Keywords:
Evapotranspiration
Streamflow
Monte Carlo simulation
NSGA-II
SSEBop
ALEXI

A B S T R A C T

As the global demands for the use of freshwater resources continues to rise, it has become increasingly important to insure the sustainability of this resources. This is accomplished through the use of management strategies that often utilize monitoring and the use of hydrological models. However, monitoring at large scales is not feasible and therefore model applications are becoming challenging, especially when spatially distributed datasets, such as evapotranspiration, are needed to understand the model performances. Due to these limitations, most of the hydrological models are only calibrated for data obtained from site/point observations, such as streamflow. Therefore, the main focus of this paper is to examine whether the incorporation of remotely sensed and spatially distributed datasets can improve the overall performance of the model. In this study, actual evapotranspiration (ETa) data was obtained from the two different sets of satellite based remote sensing data. One dataset estimates ETa based on the Simplified Surface Energy Balance (SSEBop) model while the other one estimates ETa based on the Atmosphere-Land Exchange Inverse (ALEXI) model. The hydrological model used in this study is the Soil and Water Assessment Tool (SWAT), which was calibrated against spatially distributed ETa and single point streamflow records for the Honeyoey Creek-Pine Creek Watershed, located in Michigan, USA. Two different techniques, multi-variable and genetic algorithm, were used to calibrate the SWAT model. Using the aforementioned datasets, the performance of the hydrological model in estimating ETa was improved using both calibration techniques by achieving Nash-Sutcliffe efficiency (NSE) values >0.5 (0.73–0.85), percent bias (PBIAS) values within ±25% (±21.73%), and root mean squared error – observations standard deviation ratio (RSR) values <0.7 (0.39–0.52). However, the genetic algorithm technique was more effective with the ETa calibration while significantly reducing the model performance for estimating the streamflow (NSE: 0.32–0.52, PBIAS: ±32.73%, and RSR: 0.63–0.82). Meanwhile, using the multi-variable technique, the model performance for estimating the streamflow was maintained with a high level of accuracy (NSE: 0.59–0.61, PBIAS: ±13.70%, and RSR: 0.63–0.64) while the evapotranspiration estimations

Abbreviations: ALEXI, Atmosphere-Land Exchange Inverse; ALPHA_BF, Baseflow recession constant; BIOMIX, Biological mixing efficiency; CANMX, Maximum canopy storage; CANMX, Maximum canopy storage; CH_K2, Effective hydraulic conductivity of channel; CH_N2, Manning's n value for the main channel; CN2, Moisture condition II curve number; CO2, Carbon dioxide concentration; EPA, Environmental Protection Agency; EPCO, Plant uptake compensation factor; ESCO, Soil evaporation compensation coefficient; ET, evapotranspiration; ETa, Actual evapotranspiration; FRGMAX, Fraction of maximum stomatal conductance corresponding to the second point on the stomatal conductance curve; GA, Genetic algorithm; GOES, Geostationary Operational Environmental Satellites; GSI, Maximum stomatal conductance; GW_DELAY, Delay time for aquifer recharge; GW_REAP, Revap coefficient; GWQMN, Threshold water level in shallow aquifer for base flow; IPET, Potential evapotranspiration method; MAX TEMP, daily maximum temperature; MIN TEMP, daily minimum temperature; MODIS, Moderate Resolution Imaging Spectroradiometer; NASA, National Aeronautics and Space Administration; NASS, National Agricultural Statistics Service; NCDC, National Climatic Data Center; NED, National Elevation Dataset; NHDPlus, National Hydrology Dataset plus; NRCS, Natural Resources Conservation Service; NSE, Nash-Sutcliffe efficiency; NSGA-II, Nondominated Sorted Genetic Algorithm II; OF, objective function; PBIAS, percent bias; RCHRG_DP, Aquifer percolation coefficient; REVAPMN, Threshold water level in shallow aquifer for revap; RSME, Root mean squared error; RSR, root mean squared error -observations standard deviation ratio; SOL_AWC, Available water capacity; SSEBop, Simplified Surface Energy Balance; SURLAG, Surface runoff lag coefficient; SWAT, Soil and Water Assessment Tool; USDA, United States Department of Agriculture; USGS, United States Geological Survey; VPDFR, Vapor pressure deficit corresponding to the fraction given by FRGMAX; WND_SP, Daily wind speed.

* Corresponding author at: Department of Biosystems & Agricultural Engineering, Farrall Agriculture Engineering Hall, 524 S. Shaw Lane, Room 225, Michigan State University, East Lansing, MI 48824, USA.

E-mail address: pouyan@msu.edu (A.P. Nejadhashemi).

https://doi.org/10.1016/j.jhydrol.2017.11.009

were improved. Results from this assessment shows that incorporation of remotely sensed and spatially distributed data can improve the hydrological model performance if it is coupled with a right calibration technique.

1. Introduction

As extreme climate conditions and anthropogenic activities continue to impact environmental systems, mitigation and restoration related projects have become common. Furthermore, environmental systems, such as watersheds, are very complex with many relationships and interlocking processes (Sivakumar and Singh, 2012; Guerrero et al., 2013). Therefore, it can be challenging to determine which management solution(s) should be selected and implemented (Herman et al., 2015; Sabbaghian et al., 2016). This has led to the development of many different modeling techniques that can simulate a variety of options and identify the best solution (s), based on the criteria put forth mostly by stakeholders and policy makers (Chen et al., 2012; Beven and Smith, 2014; Giri et al., 2016).

Meanwhile, the first step in a model implementation is parameter calibration. Parameter calibration in model applications is used to adjust model performance to better simulate the natural systems they are trying to describe (Guerrero et al., 2013; Zhan et al., 2013; Rajib et al., 2016). While parameter calibration improves the ability of models to more accurately represent natural systems, models' performances are still limited by the quality and quantity of input data and their availabilities (Nejadhashemi et al., 2011). Today, most hydrological studies rely on data collected at monitoring stations across the world. In fact, the United States Geological Survey (USGS) has about 1.5 million monitoring sites from which data can be obtained (USGS, 2016a). However, even with the existence of all these monitoring sites, there are times where higher spatial resolutions are needed by researchers, stakeholders, and policy makers to more precisely evaluate the hydrologic conditions and to determine the best place to implement mitigation and restoration projects (Wanders et al., 2014). One way to address this issue is the use of remotely sensed data. Remote sensing is defined as the science of identifying, observing, and measuring an object without physical contact (Graham, 1999). With the advancements in satellite technology, remotely sensed satellite data has become a source of consistent monitoring for the entire globe, with applications ranging from crop yields to water resources assessments (Graham, 1999; Long et al., 2014).

In order to model water resources more accurately, it is important to examining different components of the hydrologic cycle, including water movement processes (e.g. evaporation and streamflow) and water storage (e.g. soil moisture, water vapor, groundwater, and surface water bodies). While hydrological models simulate all components of the hydrological cycle, streamflow is often the only component that the model outputs are compared against during the calibration process since it can be measured more accurately than the other components (Immerzeel and Droogers, 2008; Wanders et al., 2014; Rajib et al., 2016). This can result in poor simulations of other hydrologic components, which ultimately lowers the model performance (Wanders et al., 2014; Rajib et al., 2016). Therefore, including additional hydrological components in the parameter calibration process could allow the model to better represent all process occurring in the environment (Crow et al., 2003). In particular, evapotranspiration (ET) could be considered an important hydrological component added to the calibration process since it describes the moisture lost to the atmosphere from both biotic (e.g. plants) and abiotic (e.g. soils) sources (Hanson, 1991; USGS, 2016b). Meanwhile, ET plays a major role in the cycling of water from land and ocean surface sources into the atmosphere, which in turn drives precipitation (Pan et al., 2015). Furthermore, Immerzeel and Droogers (2008) found that calibrating a hydrological model for ET significantly improved ET simulations; and that ET simulation values were more sensitive to groundwater and meteorological parameters compared to soil and landuse parameters.

This indicates that including additional parameters in a model calibration can improve the overall model performance. However, the applicably of different calibration techniques has not been explored when both remotely sensed ET and streamflow data are involved. In addition, this study is unique in a sense that the performance of a hydrologic model for estimating streamflow was evaluated using different remotely sensed ET products. Therefore, the objectives for this paper are to (1) determine the performance of a calibrated hydrologic model in estimating ET against spatially distributed time series ET products obtained from remote sensing; (2) determine the impact of ET parameter calibration on streamflow estimation; and (3) evaluate the performances of two different genetic algorithm-based calibration techniques for streamflow and ET estimations.

2. Materials and methods

2.1. Study area

The study area is the Honeyoey Creek-Pine Creek Watershed (Hydrologic Unit Code 0408020203), which is located within the Saginaw Bay Watershed in Michigan's Lower Peninsula (Fig. 1). The US Environmental Protection Agency (EPA) identified the Saginaw Bay Watershed as an area of concern due to the presence of contaminated soils and degradation of fisheries within the region (EPA, 2017). These conditions were caused by the addition of both point and non-point source pollutants from a variety of sources such as industrial waste and agricultural runoff (EPA, 2016). The final outlet for this watershed is Lake Huron via the Saginaw River. Out of the approximately 1100 km^2 within the Honeyoey Watershed, agriculture is the dominant landuse (~52%) followed by forests (~23%), wetlands (~17%) and pasturelands (~5%). The remaining land is classified as urban (~3%). The Honeyoey Creek-Pine Creek watershed has been significantly altered by anthropogenic activities as evidenced by the landuse change (agricultural lands and urban area are dominant in the region), which in turn impacts the natural environment, especially water quality and quantity.

2.2. Data collection

2.2.1. Physiographic data

Several spatial and temporal input datasets were needed to describe the study area in a hydrological model. These datasets describe characteristics such as topography, landuse, soil properties, climate, and crop management practices. Data from the USGS were obtained to represent the topography of the region through the use of their 30 m spatial resolution National Elevation Dataset

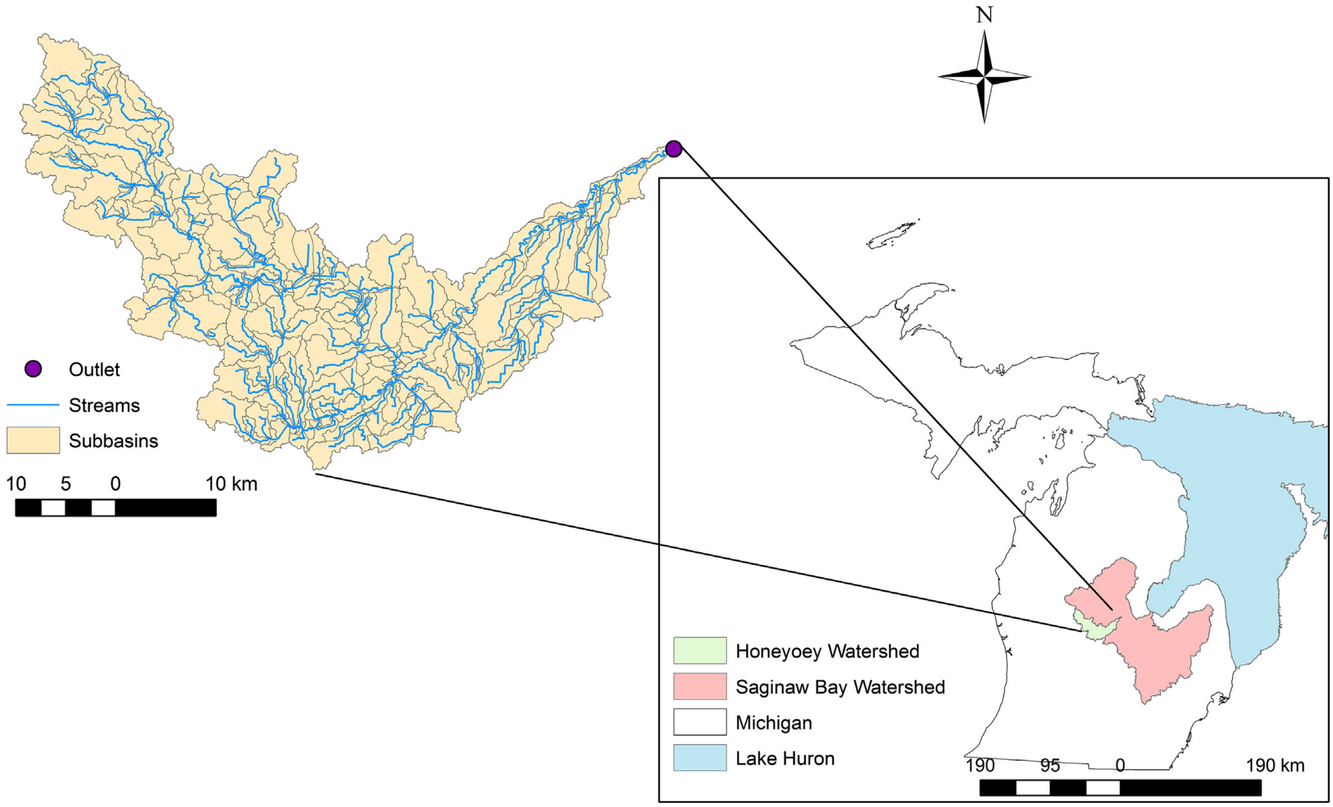

Fig. 1. The study area (Honeyoey Creek-Pine Creek watershed).

(NED, 2014). Landuse information was acquired from the 30 m spatial resolution Cropland Data Layer developed by the United States Department of Agriculture-National Agricultural Statistics Service (USDA-NASS) (NASS, 2012). The Natural Resources Conservation Service (NRCS) Soil Survey Geographic (SSURGO) Database was used to describe the soil properties for the region at a scale of 1:250,000 (NRCS, 2014). National Climatic Data Center (NCDC) weather stations (two precipitation stations and two temperature stations) were used to obtain daily precipitation and temperature data for the time span of 2003–2014. A widely used stochastic weather generator called WXGEN was employed (Sharpley and Williams, 1990; Wallis and Griffiths, 1995), which is embedded in the Soil Water Assessment Tool (SWAT), to create climate time series for other climatological records (e.g. relative humidity, solar radiation, and wind speed) that are required for SWAT to operate (Neitsch et al., 2011). Predefined crop management operations, schedules, and rotations were adopted from previous studies performed in the same region (Love and Nejadhashemi, 2011; Giri et al., 2015). Due to limitation of SWAT in simulating up to 250 different landuse, the subwatershed map that was provided by the National Hydrology Dataset plus (NHDPlus) and the Michigan Institute for Fisheries Research at a scale of 1:24,000 were modified to accommodate this limitation (Einheuser et al., 2013).

2.2.2. Remote sensing data

In order to evaluate the role of ET remote sensing data in improving a hydrologic model predictability, two satellite based ET datasets were obtained for the period of 2003–2014 for the study area. One dataset was created based on the Simplified Surface Energy Balance (SSEBop) model while the other one was based on the Atmosphere-Land Exchange Inverse (ALEXI) model.

The USGS dataset reported monthly actual evapotranspiration (ETa) using the SSEBop model (Senay et al., 2013). ETa is limited by the amount of water present at a site since it refers to the actual amount of water that is lost through both evaporation and transpiration (NOAA, 2017). This model utilizes ET fractions derived from 1 km Moderate Resolution Imaging Spectroradiometer (MODIS) thermal imagery collected every 8 days to develop a 1 km monthly ETa dataset for the Conterminous U.S. (Senay et al., 2013; Velpuri et al., 2013). Data were obtained from this dataset for each subwatershed in the study area. In order to provide an overall ETa for each subwatershed, all SSEBop's ETa pixels within each subwatershed were averaged with respect to area to generate the overall area weighted ETa average values for each month (USGS, 2016c).

The second ETa dataset is created based on the ALEXI model, which was sponsored by the USDA and US National Aeronautics and Space Administration (NASA). The ALEXI model utilizes remotely sensed morning land surface temperatures to determine ETa by relating the observed change in temperature to changes in surface moisture and ETa (Anderson et al., 1997, 2007). For this study, 4 km thermal images were obtained from Geostationary Operational Environmental Satellites (GOES) and used as to develop a daily 4 km ETa dataset for the Conterminous U.S. (Hain et al., 2015). In order to make the second set of ETa data comparable to the first set, the daily ETa values from the ALEXI model were averaged to create monthly ETa values. Next, these values were averaged for each subwatershed with respect to area.

Concerning the quality of ETa datasets for the region of study, this cannot be performed since there is no observation points for ETa in the Honeyoey Creek-Pine Creek Watershed. However, both sets of ETa datasets have been validated in many regions. For both SSEBop and ALEXI models, the average validation results were reported in terms of a root mean squared error (RSME) about 1 mm/day (Singh and Senay, 2015; Yang et al., 2017). However, based on the location of study and time of year, the RMSE values can vary.

2.3. Hydrological model: SWAT

The ETa outputs of both the ALEXI and SSEBop models were used for evaluation of SWAT models for the study region. SWAT is a widely used, continuous-time, semi-distributed, hydrological model that was developed by the USDA Agricultural Research Service (USDA-ARS) and Texas A&M AgriLife Research (Texas A&M University, 2017). By taking into account different spatiotemporal layers of information (Section 2.2.1), such as topography, landuse, and climate, SWAT models are able to simulate a variety of hydrological process, such as runoff, sediment transport, and ET (Gassman et al., 2007). This makes it a very useful tool for both researchers and policy makers.

2.4. Calibration approaches

For this study, all of the collected physiographic data was incorporated into a SWAT model. However, there are many default parameters in a SWAT model that represent an average or more probable condition that may or may not be true for the region of study (Arnold et al., 2012). Therefore, the SWAT model used in this study underwent a series of calibration and validation processes. To do this, all observed time series data were divided into calibration (2003–2008) and validation (2009–2014) periods. This process is simply referred to as calibration in the rest of the paper.

Three different types of model calibration were used in this study with the goal of understanding whether: 1) calibrating the model for only streamflow can help with improvement of ETa predictability; 2) calibrating the model for both streamflow and ETa improves model predictabilities concerning both streamflow and ETa; 3) calibrating the model for only ETa can help with improvement of streamflow predictability.

The first approach was a streamflow calibration in which individual SWAT parameters that influence the streamflow calculations were tested to find their near-optimal value through the comparison of simulated streamflows to observed streamflows. Observed streamflow data was obtained from a USGS streamflow station on the Pine River at the outlet of the study area (USGS, 2016d). The next two calibration approaches, multi-variable and genetic algorithm, were used to improve the ETa estimation for the study region. For these sets of calibrations, SWAT parameters used in ETa calculations at the subwatershed level were altered to replicate the values obtained from the ALEXI and SSEBop ETa datasets. In order to examine the role of these remotely sensed data on the performance of SWAT for estimating ETa, a multi-variable calibration approach was selected to determine the impact of adding ETa calibration on the SWAT model performance for both ETa and streamflow estimation. Meanwhile, the genetic algorithm approach was used since it is able to optimize the system for a single variable. Detailed descriptions of these calibration approaches are provided below.

2.4.1. SWAT parameters

As mentioned above, during the SWAT model calibration, the SWAT parameter values were altered. The selection of these variables was done through the use of literature review and sensitivity analysis (Woznicki and Nejadhashemi, 2012). With respect to streamflow, 15 SWAT parameters were identified and altered during the calibration process including: baseflow recession constant (ALPHA_BF), biological mixing efficiency (BIOMIX), maximum canopy storage (CANMX), effective hydraulic conductivity of channel (CH_K2), Manning's n value for the main channel (CH_N2), moisture condition II curve number (CN2), plant uptake compensation factor (EPCO), soil evaporation compensation coefficient (ESCO), delay time for aquifer recharge (GW_DELAY), revap coefficient (GW_REAP), threshold water level in shallow aquifer for base flow (GWQMN), aquifer percolation coefficient (RCHRG_DP), threshold water level in shallow aquifer for revap (REVAPMN), available water capacity (SOL_AWC), and surface runoff lag coefficient (SURLAG). These parameters were selected based on the information provided by the SWAT developer (Arnold et al., 2012). Table 1 presents the minimum, maximum, default, and calibrated values for all of these parameters for the Honeyoey watershed.

Table 1
Initial streamflow calibration parameters used in this study.

Parameter	Minimum	Maximum	Default	Calibrated
ALPHA_BF	0	1	0.048	0.55
BIOMIX	0	1	0.2	0.01
CANMX	0	100	0	1
CH_k2	−0.01	500	0	65
CH_N2	−0.01	0.3	0.014	0.025
CN2	−25%	25%	NA	−0.22%
EPCO	0	1	1	0.37
ESCO	0	1	0.95	0.97
GW_DELAY	0	500	31	9
GW_REVAP	0.02	0.2	0.02	0.055
GWQMN	0	5000	1000	1000
RCHRG_DP	0	1	0.05	0.35
REVAPMN	0	1000	750	900
SOL_AWC	0	1	NA	20%
SURLAG	1	24	4	1

In regards to the ETa calibration, another set of 10 SWAT parameters was identified as being influential to the ETa calculations (Neitsch et al., 2011). These included: maximum canopy storage (CANMX), carbon dioxide concentration (CO_2), soil evaporation compensation coefficient (ESCO), fraction of maximum stomatal conductance corresponding to the second point on the stomatal conductance curve (FRGMAX), maximum stomatal conductance (GSI), potential evapotranspiration method (IPET), daily maximum temperature (MAX TEMP), daily minimum temperature (MIN TEMP), vapor pressure deficit corresponding to the fraction given by FRGMAX (VPDFR), and daily wind speed (WND_SP). However, some of these parameters could not be altered since they were provided by either observed data or the weather generator used in this study, including MAX TEMP, MIN TEMP, and WND_SP. In addition, since climate change was not a factor for this study, CO_2 was also not altered.

The common practice in hydrologic modeling is to calibrate a model only for streamflow. However, it is interesting to know how much additional improvement in model predictability can be gained by tweaking the already calibrated model (for the streamflow). This can be achieved by only changing the parameters that are not used for streamflow calibration but relevant to the ETa calibration. Therefore, any SWAT parameters already used in the streamflow calibration, CANMX and ESCO, were also not used during the ETa calibration process. This reduced the initial set of ETa parameters from 10 to four. Of this set of four parameters, three are crop properties and have ranges of 0.001–0.1 for GSI, 0–1 for FRGMAX, and 1.5–6 for VPDFR. The last parameter used in this study, IPET, indicates which method to use when calculating potential evapotranspiration (ETp). Within SWAT three different ETp methods are available: namely the Penman-Monteith method, the Priestley-Taylor method, and the Hargreaves method (Neitsch et al., 2011). All three methods were included in the ETa calibration process; however, it was found that the Penman-Monteith method produced the best results for the study area. The list of SWAT parameters that were used for the calibration procedures are presented in Table S1 in the Supplementary Material section.

2.4.2. Initial streamflow calibration

A streamflow calibration was performed to generate a base condition to which the ETa calibrations could be compared. In this section, it was hypothesized that calibrating the model for only streamflow can help with improvement of ETa predictability. Therefore, only parameters that were commonly used for the streamflow calibration were altered. In order to evaluate the performance of a hydrological model, three statistical criteria that were suggested by Moriasi et al. (2007), were used in this study. These criteria include: 1) Nash-Sutcliffe efficiency (NSE) representing the ratio of residual variance and observed data variance (Nash and Sutcliffe, 1970); 2) Percent bias (PBIAS) evaluating how much larger/smaller simulated data are than their corresponding observed data (Gupta et al., 1999); and 3) Root mean squared error (RMSE)-observations standard deviation ratio (RSR), reporting the ratio of RMSE and standard deviation of measured data (Legates and McCabe, 1999). For evaluating the performance of a hydrologic model on simulating monthly streamflow values, NSE values above 0.5, PBIAS values within ± 25%, and RSR values below 0.7 are considered as satisfactory (Moriasi et al., 2007). In addition, we also reported RMSE to examine the error associated with the simulated data in which lower values represent the better model performance.

2.4.3. Multi-variable calibration

A multi-variable calibration procedure, based on Monte Carlo simulation and an evolutionary algorithm, was applied to the SWAT model using both remotely sensed ETa datasets and observed streamflow from the study area. The aim of the procedure was to identify the Pareto optimal frontier and the best trade-off solution.

A solution is classified as Pareto optimal (also known as nondominated) when the value of any objective function cannot be improved without decreasing the performance of at least one other objective function (Chankong and Haimes, 1993; Tang et al., 2005). In multi-variable calibration, there is at least one objective function per observed variable. For this study, the minimization objective function (OF) for each variable (i.e. ETa and streamflow) was based on the NSE.

$$OF = 1 - NSE$$

The objective function for ETa was computed using the area weighted average of the monthly simulated from the hydrologic model and satellite-based ETa time series for each subwatershed, which was determined as follows:

$$\overline{ET}_j = \frac{1}{A_T}\sum_{i=1}^{n} A_i ET_{ij}$$

where, $\overline{ET}_j$ is the average ETa for month j; A_T is the total surface area of the watershed; A_i is the surface area of subwatershed i; ET_{ij} is the ETa for subwatershed i and month j; and n is the number of subwatersheds. Therefore, one pair of simulated-observed ETa series for the whole watershed was obtained to determine a unique NSE for this variable. This process was not employed for streamflow since there is only one gauging station at the outlet of the study area (Fig. 1).

The general outline of the multi-variable calibration, which is further explained in the following sections, is as follows: A Monte Carlo simulation is performed to understand the SWAT model performance for ETa and streamflow with respect to the selected calibration parameters. Thus, 5000 parameter sets were randomly generated via uniform sampling, which were then evaluated by executing the SWAT model for each generated parameter set. The results were used to define, if possible, narrower calibration parameter ranges, and to obtain multi-objective scatter plots to identify preliminarily Pareto Optimal solutions. The next step consists of the application of a multi-objective evolutionary algorithm known as the Nondominated Sorted Genetic Algorithm II (NSGA-II) (Deb et al., 2002) to determine the optimal Pareto population. Finally, the decision-making method known as the Compromise Programming (Deb, 2001), using a Euclidean distance metric, was employed to select the final optimal trade-off solution from the resulting Pareto Optimal population.

2.4.3.1. Monte Carlo simulation. A total of 5000 runs for Monte Carlo simulation were performed using MATLAB®, with randomly generated corresponding parameter sets selected from uniform distributions. Ranges for calibration parameters were defined as follows: 0.001–0.1 for GSI, 0–1 for FRGMAX, and 1.5–6 for VPDFR. A SWAT model run was executed for each parameter combination, computing NSE for both ETa and streamflow. Dotty plots relating each OF with parameter values were obtained to analyze parameter identifiability, and if possible, narrower calibration ranges to be explored with the NSGA-II algorithm. Likewise, multi-objective plots relating ETa and streamflow OF values were generated for preliminary Pareto frontiers identification.

2.4.3.2. Multi-objective evolutionary algorithm: NSGA-II. The NSGA-II is a multi-objective genetic algorithm that has been widely used in various disciplines and has been successfully implemented in other SWAT applications (Zhang et al., 2010, 2016; Lu et al., 2014). The NSGA-II is a population-based algorithm that is comprised of a nondominated ranking process, a crowded distance calculation, an elitist selection method, and offspring reproduction operations (Deb, 2001). For this study, a real-coded NSGA-II with simulated binary crossover (SBX) and polynomial mutation (Baskar et al., 2015) was applied, requiring the prior definition of distribution indexes for each operation (defined as 20 for crossover and mutation each). Other input parameters include the population size (defined as 100), the maximum number of generations as stopping criteria (defined as 50), and the mutation probability (defined as the reciprocal of the number of calibration parameters).

2.4.3.3. Compromise programing approach. The compromise programing approach using the l_2 metric (which becomes the Euclidean distance metric) is used to select the optimal Pareto population member that is closest to a reference point (Deb, 2001). In this case, the ideal point, which is unfeasible and is not located on the Pareto frontier, is selected as the reference point and it is comprised by the best objective function values (Deb, 2001). Before computing the distance between each Pareto point and the ideal point, the objective function values are normalized employing a Euclidian non-dimensionalization (Sayyaadi and Mehrabipour, 2012):

$$OF_{ij}^{n} = \frac{OF_{ij}}{\sqrt{\sum_{i=1}^{m} OF_{ij}^2}}$$

where, i is the index for each point in the Pareto frontier, j is the index for each OF, m is the total number of the Pareto population, and n superscript refers to "non-dimensional". The distance d_i between each Pareto point and the ideal point, which is the l_2 metric, is calculated as follows:

$$l_2 = \sqrt{\sum_{j=1}^{N} \left(OF_{ij} - OF_{ij}^{ideal}\right)^2}$$

where, N denotes the total number of objective functions.

In the compromise programming approach, the point with the minimum distance metric value is chosen as the best trade-off solution.

2.4.4. Genetic algorithm calibration

The other approach used to calibrate the SWAT models with respect to the ETa datasets was a genetic algorithm (GA). A GA is an optimization technique that imitates biological process to refine a population of potential solutions to identify the best final or set of final solutions (Goldberg, 1989; Conn et al., 1991, 1997). For this study, a GA was used to guide ETa calibrations by changing the values of three parameters within the SWAT model, namely GSI, FRGMAX, and VPDFR. In this section, it was hypothesized that calibrating the model for only ETa can help with improvement of streamflow predictability. Therefore, only parameters that were commonly used for the ETa calibration were altered. These are the same parameters that were modified in the multi-variable optimization approach, and thus the same ranges were used for this optimization. With each successive set of parameter values, a series of MATLAB® codes were used to update and run the SWAT model (Abouali, 2017). First, the parameter values were accepted by the code, which checked the values to the defined ranges and then applied the values to all subwatersheds within the region. After this was completed, the code executed the SWAT model and stored the outputs for further analysis. In summary, the SWAT model was run 904,900 times. While executing these runs will not necessarily develop an ideal model, it will generate a landscape of how ET changes for each subwatershed based on the specified parameters. For each set of parameter values, the SWAT ETa outputs were compared to the ALEXI and SSEBop datasets and NSE and RMSE were calculated for each subwatershed. The parameter set that had the largest NSE was considered to be the best and the lowest RMSE was used as the tiebreaker. This allowed for the identification of the best parameter values for each subwatershed, which then used to parametrizes the best model that maximizes the ETa calibration. It should be noted that this is only possible based on the assumption that the ETa calculation for one subwatershed is not affected by the ETa calculation for another subwatershed, otherwise it would not be possible to create the mosaic landscape of parameter values used in the best model, which to the best of our knowledge has not been done in other SWAT studies. Furthermore, after the best parameters for each subwatershed were identified and applied within the SWAT models, the simulated ETa values were area averaged to produce a single ETa value for the entire watershed. This set of ETa values was then used to calculate the NSE, PBIAS, RSR, and RSME for the entire region, just like was done in the multi-variable calibration. This was done to allow for a watershed level evaluation of the calibration approaches.

Table 2
Calibration and validation criteria based on the initial streamflow calibration.

	NSE	PBIAS (%)	RSR	RSME
Overall (2003–2014)	0.612	−0.965	0.623	6.522
Calibration (2003–2008)	0.611	4.303	0.624	5.996
Validation (2009–2014)	0.613	−5.856	0.622	7.009

2.5. Statistical analysis

To further evaluate the streamflow and ETa outputs from the calibrated models and ETa datasets, a mixed-effects model was used to compare the mean difference between each of the outputs (Kuznetsova et al., 2015). This process was performed twice, once for the streamflow datasets (observed, initial streamflow calibrated model, ALEXI multi-variable calibrated model, ALEXI genetic algorithm calibrated model, SSEBop multi-variable calibrated model, and SSEBop genetic algorithm calibrated model) and once for the ETa datasets (ALEXI, SSEBop, ALEXI multi-variable calibrated model, ALEXI genetic algorithm calibrated model, SSEBop multi-variable calibrated model, and SSEBop genetic algorithm calibrated model). This allowed for the determination of significant mean differences between the datasets with a 95% confidence level.

3. Results and discussion

3.1. Initial streamflow calibration

Daily streamflow was calibrated and validated for a 12-year period (6 years calibration and 6 years validation) from 2003 to 2014 for the region. Table 2 shows the NSE, Pbias, RSR, and RSME values achieved for the calibrated model. As shown in the table, all criteria (NSE, PBIAS, and RSR) are in their respective satisfactory ranges (Moriasi et al., 2007) indicating that the model was successfully calibrated and can be used to simulate streamflow values for the region. Furthermore, while the overall RSME was 6.522, the calibration period had a smaller RSME compared to the validation period, indicating a better model fit during the calibration period than the validation period.

The temporal variability of observed and simulated streamflow is also presented in Fig. 2. Overall, the SWAT model represents observed flow variations very accurately.

The results of this section present the performance of the SWAT model in replicating the spatially distributed ETa data obtained

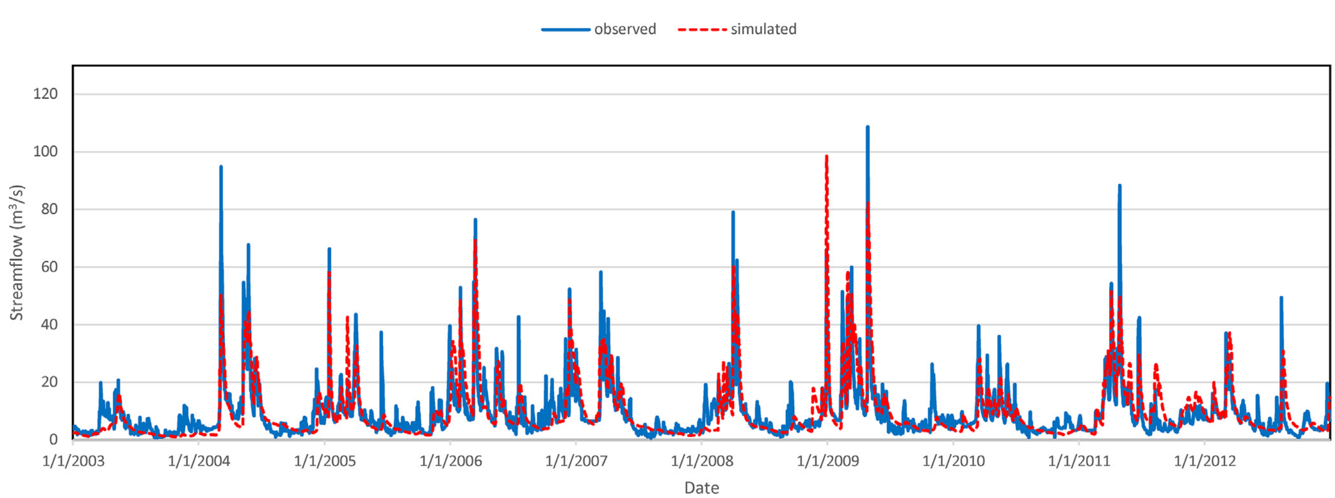

Fig. 2. Comparison of observed and simulated streamflow based the initial streamflow calibration.

Table 3
Statistical criteria ETa when the results from initial streamflow calibrated SWAT model was used.

Period	Variable/ Dataset	Statistical measure			
		NSE	PBIAS (%)	RSR	RMSE
Overall (2003–2014)	ALEXI ETa	0.62	27.82	0.62	21.79
	SSEBop ETa	0.81	−10.12	0.44	18.28
Calibration (2003–2008)	ALEXI ETa	0.62	27.83	0.62	21.48
	SSEBop ETa	0.81	−8.46	0.44	18.13
Validation (2009–2014)	ALEXI ETa	0.62	27.80	0.61	22.10
	SSEBop ETa	0.80	−11.78	0.44	18.42

from two remote sensing products (SSEBop and ALEXI datasets). Table 3 shows the SWAT model performance for the overall, calibration, and validation periods based on NSE, PBIAS, RSR, and RMSE of the ETa for the condition in which only the streamflow calibration was performed. These calculations followed the same procedure that was discussed in the multi-variable and GA calibration sections, in which ETa values were area averaged across the watershed and then used to calculate watershed level statistical criteria. When considering the entire time period, the streamflow calibrated SWAT model was able to replicate the SSEBop ETa dataset more accurately than the ALEXI ETa dataset. This can be seen by the fact that the statistical criteria for the SSEBop ETa were better than those for the ALEXI ETa. Similar results were see for the calibration and validation periods. Overall, this shows that the SWAT model can better replicate the SSEBop ETa data compared to the ALEXI data.

3.2. Multi-variable calibration

A combination of 5000 Monte Carlo simulations and a NSGA-II evolutionary algorithm were used to identify the Pareto frontiers for the SWAT model calibrations for both the ALEXI and SSEBop ETa datasets. Fig. 3 shows both the entire Monte Carlo population as well as the Pareto frontiers identified by the NSGA-II evolutionary algorithm for each ETa dataset. This shows that Pareto frontiers were able to be identified from the Monte Carlo simulations run for each ETa datasets, which indicates the first phase of the multi-variable optimization was successful for both datasets. However, the SSEBop Pareto frontier was able to further minimize streamflow and ETa OFs compared to the ALEXI Pareto frontier. Therefore, calibrating the SWAT model using the SSEBop ETa data was able to produce a more accurate model performance. This can be seen more clearly in Fig. 4, which shows the Pareto frontiers for both the SSEBop and ALEXI datasets. This figure also highlights the optimal Pareto population member selected by the compromise programing method, which shows the optimal model calibration for each dataset. This reinforces the conclusions that the SSEBop dataset performed better than the ALEXI dataset and achieved a model calibration that was able to better simulate both streamflow and ET values for the entire region. In addition, the results showed that the multi-variable calibration was able to identify a final calibrated model for each dataset that improved both streamflow and ET simulations.

Table 4 shows the NSE, PBIAS, RSR, and RMSE values achieved for both final calibrated models. All values presented in the table fall within the satisfactory ranges and indicate that the models were successfully calibrated. Furthermore, comparison of these values with the base model simulations showed that with respect to ET there was an improvement in the statistical criteria. For example, when considering overall NSE the ALEXI calibrated model had a value of 0.73 compared to the 0.62 for the base model and the SSEBop calibrated model had a value of 0.85 compared to the 0.81 for the base model. This indicates that the newly calibrated models are better able to simulate ETa data. However, with respect to streamflow, all statistical criteria remain within the satisfactory ranges and often similar to the base model statistical criteria, suggesting that the streamflow simulations were not heavily impacted by the addition of the ET calibration. Overall, the results show that this calibration approach was successful at improving the models' performances while maintaining the current streamflow accuracies.

3.3. Genetic algorithm calibration

In addition to the multi-variable approach, a GA optimization was also performed. Unlike the multi-variable approach, this approach focused on only improving the ETa estimations for two remotely sensed datasets (ALEXI and SSEBop) without considering the streamflow calibration. After hundreds of runs for each subwatershed, the GA was able to identify the optimal parameters values for each subwatershed and the ETa datasets. These final parameter values were used to develop SWAT models that represented the optimal ETa calibration for each subwatershed. Table 5 shows the NSE, PBIAS, RSR, and RMSE values achieved for both final calibrated models. All of the ETa statistical criteria values presented in the table fall within the satisfactory ranges and indicate that the models were successfully calibrated with respect to ET. When compared to the base model, it can be seen that the ETa calibration

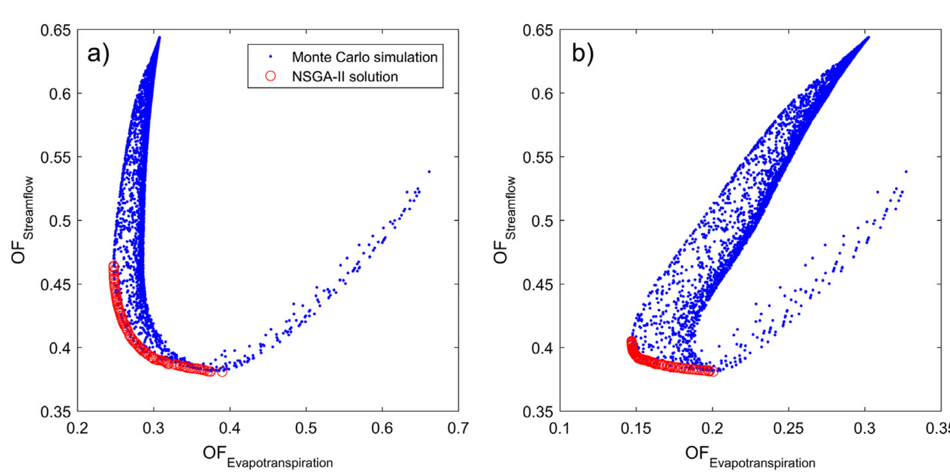

Fig. 3. Monte Carlo populations and Pareto frontiers for a) ALEXI and b) SSEBop datasets.

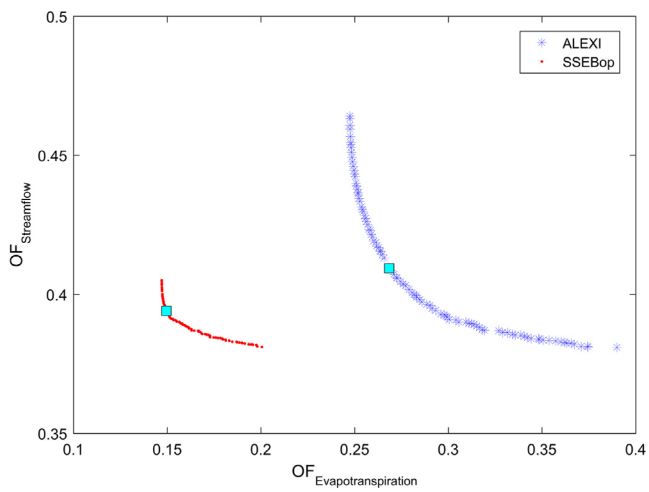

Fig. 4. Pareto frontiers and optimal Pareto population members for both ALEXI and SSEBop datasets.

performed here was able to improve the simulation of ETa values for both the ALEXI and SSEBop datasets. For example when considering the overall NSE, the ALEXI calibrated model had a value of 0.75 compared to the 0.62 for the base model and the SSEBop calibrated model had a value of 0.84 compared to the 0.81 for the base model. However, when considering the streamflow calibration, most of the statistical values have fallen outside the satisfactory ranges (NSE > 0.5, PBIAS ± 25%, and RSR < 0.7) for each criteria. This indicates that while this process was able to improve the ET simulations, it was done at the cost of compromising streamflow simulations. This seems logical, knowing that this approach did not consider the streamflow calibration during the ETa calibration process. However, this does indicate that this approach would be unsuitable for calibrating models that require accurate streamflow values.

3.4. Statistical significance

The results of the statistical analysis of the mean difference between each of the datasets are presented for streamflow and ETa in Tables 6 and 7, respectively. Linear mixed-effects models were employed to account for the spatiotemporal effects that cause sample correlation violating the independence assumption for the usual paired *t*-test (Esfahanian et al., 2017). With regard to the streamflow datasets, all comparisons were found to be significantly different from each other except for the comparison of the observed dataset with the initial streamflow calibrated model. This indicates that the initial calibration was able to closely replicate the observed data to the point where statistically there is no difference between them. However, the significant difference observed for all other models compared to the observed data indicates that those models are not as accurate when simulating streamflow. This seems logical for the models calibrated via the genetic algorithm approach since there was a noticeable decrease in the statistical criteria for the streamflow calibration in these models. However, we did not expect this for the models calibrated using the multi-variable approach, since these models showed little to no change in the calibration criteria for streamflow. These results indicate that even though the calibration process was able to satisfactorily calibrate streamflow, there exist more inconsistencies within the final simulated streamflow when compared to the observed data. When considering the comparison of streamflow simulations between the initial model and the other four models tested, the significant difference makes sense and indicates that the addition of the ETa calibration influenced the streamflow calibration to an extent. Furthermore, since all of these the p-values were negative, the ETa calibrated models all underestimated the streamflow when compared to both the observed dataset and the initial streamflow model. This indicates that regardless of the calibration method used or the impact seen on the statistical criteria, the ETa calibrated models produced lower streamflow values on average. Finally, the comparisons between the four ET calibrated models also showed significant difference, which seems

Table 4
Statistical criteria for the optimal multi-variable calibration models.

Period	ET dataset	Statistical criteria							
		NSE		PBIAS (%)		RSR		RMSE	
		ET	Streamflow	ET	Streamflow	ET	Streamflow	ET	Streamflow
Overall (2003–2014)	ALEXI	0.73	0.59	21.73	13.70	0.52	0.64	18.32	6.70
	SSEBop	0.85	0.61	−16.05	8.20	0.39	0.63	16.05	6.57
Calibration (2003–2008)	ALEXI	0.72	0.59	22.01	18.52	0.53	0.64	18.37	6.19
	SSEBop	0.85	0.61	−14.03	12.94	0.38	0.63	15.85	6.01
Validation (2009–2014)	ALEXI	0.74	0.59	21.46	9.22	0.51	0.64	18.28	7.18
	SSEBop	0.85	0.60	−18.07	3.79	0.39	0.63	16.25	7.09

Table 5
Statistical criteria for the optimal genetic algorithm calibrated models.

Period	ET dataset	Statistical criteria							
		NSE		PBIAS (%)		RSR		RMSE	
		ET	Streamflow	ET	Streamflow	ET	Streamflow	ET	Streamflow
Overall (2003–2014)	ALEXI	0.75	0.32	14.34	32.73	0.50	0.82	17.84	8.61
	SSEBop	0.84	0.52	−17.42	10.69	0.39	0.69	16/35	7.28
Calibration (2003–2008)	ALEXI	0.74	0.22	14.89	39.24	0.51	0.88	17.82	8.50
	SSEBop	0.85	0.50	−15.25	16.22	0.39	0.71	16.19	6.80
Validation (2009–2014)	ALEXI	0.75	0.40	13.80	26.67	0.50	0.77	17.86	8.71
	SSEBop	0.84	0.53	−19.59	5.55	0.40	0.69	16.51	7.73

Table 6
Mean differences and p-values from mixed-effects model for comparison of the different streamflow datasets used in this study. Bolded values indicate significant difference at the 0.05 level.

Streamflow datasets*	Streamflow datasets*					
	A	B	C	D	E	F
A						
B	0.08 (0.75)					
C	**−3.18 (0.00)**	**−3.26 (0.00)**				
D	**−1.04 (0.00)**	**−1.13 (0.00)**	**2.13 (0.00)**			
E	**−1.34 (0.00)**	**−1.42 (0.00)**	**1.84 (0.00)**	**−0.29 (0.01)**		
F	**−0.80 (0.00)**	**−0.89 (0.00)**	**2.37 (0.00)**	**0.24 (0.03)**	**0.53 (0.00)**	

* A = Observed Streamflow, B = Initial Streamflow Calibrated Model, C = ALEXI Genetic Algorithm Calibrated Model, D = SSEBop Genetic Algorithm Calibrated Model, E = ALEXI Multi-Variable Calibrated Model, and F = SSEBop Multi-Variable Calibrated Model.

Table 7
Mean differences and p-values from mixed-effects model for comparison of the different ETa datasets used in this study. Bolded values indicate significant difference at the 0.05 level.

ET datasets*	ET datasets*						
	A	B	C	D	E	F	G
A							
B	**20.10 (0.00)**						
C	2.75 (0.09)	**−17.35 (0.00)**					
D	**11.07 (0.00)**	**−9.03 (0.00)**	**8.32 (0.00)**				
E	**6.69 (0.00)**	**−13.41 (0.00)**	**3.94 (0.00)**	**−4.38 (0.00)**			
F	**6.97 (0.00)**	**−13.13 (0.00)**	**4.22 (0.00)**	**−4.10 (0.00)**	0.28 (0.23)		
G	**5.67 (0.00)**	**−14.43 (0.00)**	**2.92 (0.00)**	**−5.40 (0.00)**	**−1.02 (0.00)**	**−1.30 (0.00)**	

* A = SSEBop, B = ALEXI, C = Initial Streamflow Calibrated Model, D = ALEXI Genetic Algorithm Calibrated Model, E = SSEBop Genetic Algorithm Calibrated Model, F = ALEXI Multi-Variable Calibrated Model, and G = SSEBop Multi-Variable Calibrated Model.

understandable given the use of different ET datasets and calibration process used in this study.

With regards to the ETa datasets (Table 7), almost all comparisons among datasets showed significant differences except for the SSEBop dataset versus the initial streamflow calibrated model (column A and row C) and the SSEBop genetic algorithm calibrated model versus the ALEXI multi-variable calibrated model (column E and row F). These two cases are rather interesting since the first comparison (SSEBop versus the initial streamflow calibrated model) indicates that by only calibrating for streamflow it was possible to simulate ETa so that it is not statistically different from the remotely sensed data. Meanwhile the second case (SSEBop genetic algorithm calibrated model versus ALEXI multi-variable calibrated model) indicates that in some cases, calibrated SWAT model can generate comparable ETa time series even though different approaches and datasets were used. Considering all of the other significant differences, the comparison between the ALEXI and SSEBop data (column A and row B in Table 7) made the most logical sense since different methodologies were used to calculate these datasets. Furthermore, similar results to the streamflow were also seen when comparing the ETa calibrated models to the remotely sensed ETa datasets. These observations confirm that even though these models were able to satisfactorily simulate ETa values, the SWAT simulated ETa was statistically different from the remotely sensed data used to calibrate them, and thus could not accurately replicate the remotely sensed data. However, while the streamflow comparisons showed that all of the ETa calibrated SWAT models underestimated streamflow, here it can be seen that the SSEBop calibrated SWAT models overestimated ETa values while the ALEXI calibrated SWAT models underestimated the ETa values when compared to the SSEBop and ALEXI datasets, respectively. In addition, similar to the streamflow comparisons, the four ETa calibrated models were significantly different from the initial streamflow calibrated model, which makes sense since all of the ETa calibrated models had an increase in the statistical criteria for ETa calibration compared to the initial streamflow calibrated model. Finally, the comparisons between the four ETa calibrated models (columns D through F and rows E through G) showed significant difference from each other except for the case of the SSEBop genetic algorithm calibrated model versus the ALEXI multi-variable calibrated model discussed previously. This is reasonable since different calibration approaches and ETa datasets were used.

3.5. Comparison of the multi-variable and genetic algorithm calibrations

Based on the information provided in Tables 4 and 5, it can be concluded that the multi-variable approach used in this study was able to generate better overall SWAT models compared to the GA approach. However, if the goal of the model is to generate more accurate ETa data, the GA approach was able to outperform the multi-variable approach. This shows that depending on the purpose of the model applications, different calibration techniques should be used. Furthermore, it is to be noted that for both approaches the models built using the SSEBop data were able to achieve higher performances in simulating both streamflow and ETa data than the models made based on the ALEXI data.

4. Conclusions

In this study, two different ETa calibration techniques were used to evaluate the impact of adding spatially distributed and remotely sensed ETa datasets to the traditional streamflow calibration used in hydrological models. Both techniques, multi-variable and GA, were able to successfully improve the ETa calibration for the hydrological model using both remotely sensed ETa datasets. The GA technique was able to produce better ETa calibrations and thus better ETa simulations; however, this was achieved at the cost of lowering the streamflow calibrations. Meanwhile the multi-variable technique was able to improve the ETa calibration while maintaining the streamflow calibration. Therefore, future use of these approaches should be driven by the needs of the research. For example, if a study is focused solely on better ETa estimation, the GA approach is the better option; meanwhile studies focused on better simulating the entire hydrological cycle for a region should use the multi-variable approach. Concerning the ETa datasets used in this study, the calibrations performed with the SSEBop dataset resulted better ETa estimations compared to the calibrations based on the ALEXI dataset for this study area. Therefore, it is recommended that future studies should perform this analysis in other regions to better understand how these datasets compare to each other as well as evaluating the impacts of different climate variabilities (e.g., snow cover).

Statistical analysis of the streamflow and ETa showed that the remotely sensed ETa datasets were significantly different from each other, which was expected. Moreover, except for one exception, all of the streamflow and ETa datasets produced by the ETa calibrated SWAT models were also significantly different from each other. However, all four ETa calibrated models were also significantly different when compared to the remotely sensed data. This indicated that while the overall model calibration was successful it was unable to closely replicate the remotely sensed data, showing that there still could be additional improvements in the both in calibration process and the SWAT model simulations.

It is to be noted that the ETa calibration processes used in this study only altered three parameters within the SWAT model. This was due to temporal and computational limitations. However, the addition of other parameters to the calibration process, such as the soil evaporation compensation factor (ESCO), could result in even better model calibrations and thus better model outputs and should be the focus of future studies. In addition, while adding ETa calibration to the overall model calibration process was successful in this study, future studies should consider additional hydrological cycle components, such as remotely sensed soil moisture datasets. This would allow for the development of even more realistic models and thus more accurate results for stakeholders and policy makers who rely on model outputs for managing freshwater resources.

Acknowledgments

Authors would like to thank Dr. Wade Crow from USDA-ARS Hydrology and Remote Sensing Laboratory at Beltsville, Maryland for his help with editing the paper. This work is supported by the USDA National Institute of Food and Agriculture, Hatch project MICL02359.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jhydrol.2017.11.009.

References

Abouali M., (2017) SWATUtilities. <https://github.com/maboualidev/SWATUtilities/tree/master/MATLAB/MCode> (accessed 4.26.17).

Anderson, M.C., Norman, J.M., Diak, G.R., Kustas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. Remote Sens. Environ. 60, 195–216.

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. J. Geophys. Res. 112, D10117.

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. T. ASABE 55, 1491–1508.

Baskar S., Tamilselvi S., Varshini P.R., 2015. MATLAB code for Constrained NSGA II - Dr.S.Baskar, S. Tamilselvi and P.R.Varshini - File Exchange - MATLAB Central. <http://www.mathworks.com/matlabcentral/fileexchange/49806-matlab-code-for-constrained-nsga-ii-dr-s-baskar–s-tamilselvi-and-p-r-varshini> (accessed 4.26.17).

Beven, K., Smith, P., 2014. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. J. Hydrol. Eng. 20, A4014010.

Chankong, V., Haimes, Y.Y., 1993. Multi-objective optimization: Pareto optimality. In: Concise Encyclopedia of Environmental Systems. Pergamon Press, UK, pp. 387–396.

Chen, H., Xu, C.-Y., Guo, S., 2012. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. J. Hydrol. 434, 36–45.

Conn, A., Gould, N., Toint, P., 1997. A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. Math. Comput. Am. Math. Soc. 66, 261–288.

Conn, A.R., Gould, N.I., Toint, P., 1991. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numer. Anal. 28, 545–572.

Crow, W.T., Wood, E.F., Pan, M., 2003. Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals. J. Geophys. Res.-Atmos., 108

Deb, K., 2001. Multi-Objective Optimization Using Evolutionary Algorithms. Chichester, John-Wiley.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE T. Evolut. Comput. 6, 182–197.

Einheuser, M.D., Nejadhashemi, A.P., Woznicki, S.A., 2013. Simulating stream health sensitivity to landscape changes due to bioenergy crops expansion. Biomass Bioenergy 58, 198–209.

EPA, 2016. About Saginaw River and Bay AOC. <https://www.epa.gov/saginaw-river-bay-aoc/about-saginaw-river-and-bay-aoc> (accessed 4.26.17).

EPA, 2017. Saginaw River and Bay Area of Concern <https://www.epa.gov/saginaw-river-bay-aoc> accessed 4.26.17.

Esfahanian, E., Nejadhashemi, A.P., Abouali, M., Adhikari, U., Zhang, Z., Daneshvar, F., Herman, M.R., 2017. Development and evaluation of a comprehensive drought index. J. Environ. Manage. 185, 31–43.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Invited review series. T. ASABE 50, 1211–1250.

Giri, S., Nejadhashemi, A.P., Woznicki, S.A., 2016. Regulators' and stakeholders' perspectives in a framework for bioenergy development. Land Use Policy 59, 143–153.

Giri, S., Nejadhashemi, A.P., Zhang, Z., Woznicki, S.A., 2015. Integrating statistical and hydrological models to identify implementation sites for agricultural conservation practices. Environ. Model. Softw. 72, 327–340.

Goldberg, D.E., 1989. Genetic algorithms in search, optimization, and machine learning, addison-wesley, reading, ma. Reading, MA.

Graham S., 1999. Remote Sensing : Feature Articles. <https://earthobservatory.nasa.gov/Features/RemoteSensing/> (accessed 4.26.17).

Guerrero, J.-L., Westerberg, I.K., Halldin, S., Lundin, L.-C., Xu, C.-Y., 2013. Exploring the hydrological robustness of model-parameter values with alpha shapes. Water Resour. Res. 49, 6700–6715.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. J. Hydrol. Eng. 4, 135–143.

Hain, C.R., Crow, W.T., Anderson, M.C., Yilmaz, M.T., 2015. Diagnosing neglected soil moisture source-sink processes via a thermal infrared-based two-source energy balance model. J. Hydrometeorol. 16, 1070–1086.

Hanson, R.L., 1991. Evapotranspiration and droughts. Geol. Survey Water-Supply Paper 2375, 99–104.

Herman, M.R., Nejadhashemi, A.P., Daneshvar, F., Ross, D.M., Woznicki, S.A., Zhang, Z., Esfahanian, A.-H., 2015. Optimization of conservation practice implementation strategies in the context of stream health. Ecol. Eng. 84, 1–12.

Immerzeel, W.W., Droogers, P., 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. J. Hydrol. 349, 411–424.

Kuznetsova A., Brockhoff P.B., Christensen R.H.B., 2015. Package "lmerTest." R package version 2.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241.

Long, D., Longuevergne, L., Scanlon, B.R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour. Res. 50, 1131–1151.

Love, B., Nejadhashemi, A.P., 2011. Environmental impact analysis of biofuel crops expansion in the Saginaw River watershed. J. Biobased Mater. Bio. 5, 30–54.

Lu, S., Kayastha, N., Thodsen, H., Van Griensven, A., Andersen, H.E., 2014. Multiobjective calibration for comparing channel sediment routing models in the soil and water assessment tool. J. Environ. Qual. 43, 110–120.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. T. ASABE 50, 885–900.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. J. Hydrol. 10, 282–290.

NASS, 2012. CropScape - Cropland Data Layer <https://nassgeodata.gmu.edu/CropScape/> accessed 4.26.17.

NED, 2014. The National Map: Elevation <https://nationalmap.gov/elevation.html> accessed 4.26.17.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute.
Nejadhashemi, A.P., Woznicki, S.A., Douglas-Mankin, K.R., 2011. Comparison of four models (STEPL, PLOAD, L-THIA, and SWAT) in simulating sediment, nitrogen, and phosphorus loads and pollutant source areas. T. ASABE 54, 875–890.
NOAA, 2017. Potential Evapotranspiration <https://www.ncdc.noaa.gov/monitoring-references/dyk/potential-evapotranspiration> accessed 4.26.17.
NRCS, 2014. Web Soil Survey <https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx> accessed 4.26.17.
Pan, S., Tian, H., Dangal, S.R., Yang, Q., Yang, J., Lu, C., Tao, B., Ren, W., Ouyang, Z., 2015. Responses of global terrestrial evapotranspiration to climate change and increasing atmospheric CO2 in the 21st century. Earth's Future 3, 15–35.
Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. J. Hydrol. 536, 192–207.
Sabbaghian, R.J., Zarghami, M., Nejadhashemi, A.P., Sharifi, M.B., Herman, M.R., Daneshvar, F., 2016. Application of risk-based multiple criteria decision analysis for selection of the best agricultural scenario for effective watershed management. J. Environ. Manage. 168, 260–272.
Sayyaadi, H., Mehrabipour, R., 2012. Efficiency enhancement of a gas turbine cycle using an optimized tubular recuperative heat exchanger. Energy 38, 362–375.
Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational evapotranspiration mapping using remote sensing and weather datasets: a new parameterization for the SSEB approach. JAWRA J. Am. Water Resour. Assoc. 49, 577–591.
Sharpley, A.N., Williams, J.R., 1990. EPIC. Erosion/Productivity Impact Calculator, Technical bulletin (USA).
Singh, R.K., Senay, G.B., 2015. Comparison of four different energy balance models for estimating evapotranspiration in the Midwestern United States. Water 8, 1–19.
Sivakumar, B., Singh, V.P., 2012. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. Hydrol. Earth Sys. Sc. 16, 4119.
Tang, Y., Reed, P., Wagener, T., 2005. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrol. Earth Sys. Sc. 2, 2465–2520.
Texas A&M University, 2017. SWAT: Soil and Water Assessment Tool <http://swat.tamu.edu/> 2017 accessed 4.26.17.
USGS, 2016. USGS Water Data for the Nation <https://waterdata.usgs.gov/nwis> accessed 4.26.17.
USGS, 2016. Evapotranspiration - The Water Cycle, from USGS Water-Science School <https://water.usgs.gov/edu/watercycleevapotranspiration.html> 2016 accessed 4.26.17.
USGS, 2016. USGS Geo Data Portal <https://cida.usgs.gov/gdp/client/#!catalog/gdp/dataset/54dd5d21e4b08de9379b38b6> accessed 4.26.17.
USGS,2016. USGS Current Conditions for USGS 04155500 Pine River Near Midland, MI <https://waterdata.usgs.gov/mi/nwis/dv?referred_module=sw&site_no=04155500> 2016 accessed 4.26.17.
Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49.
Wallis, T.W., Griffiths, J.F., 1995. An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). Agr. Forest Meteorol. 73, 115–133.
Wanders, N., Bierkens, M.F., de Jong, S.M., de Roo, A., Karssenberg, D., 2014. The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. Water Resour. Res. 50, 6874–6891.
Woznicki, S.A., Nejadhashemi, A.P., 2012. Sensitivity analysis of best management practices under climate change scenarios. J. Am. Water Resour. As. 48, 90–112.
Yang, Y., Anderson, M.C., Gao, F., Hain, C.R., Semmens, K.A., Kustas, W.P., Noormets, A., Wynne, R.H., Thomas, V.A., Sun, G., 2017. Daily Landsat-scale evapotranspiration estimation over a forested landscape in North Carolina, USA, using multi-satellite data fusion. Hydrol. Earth Syst. Sci. 21, 1017–1037.
Zhan, C.-S., Song, X.-M., Xia, J., Tong, C., 2013. An efficient integrated approach for global sensitivity analysis of hydrological model parameters. Environ. Model. Software 41, 39–52.
Zhang, X., 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. Process. 24, 955–969.
Zhang, Y., Shao, Q., Taylor, J.A., 2016. A balanced calibration of water quantity and quality by multi-objective optimization for integrated water system model. J. Hydrol. 538, 802–816.