



# EPA Public Access

Author manuscript

*J Hydrol (Amst)*. Author manuscript; available in PMC 2019 August 08.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*J Hydrol (Amst)*. 2018 ; 567: 668–683. doi:10.1016/j.jhydrol.2018.10.024.

## Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters

Adnan Rajib<sup>a</sup>, Grey R. Evenson<sup>b</sup>, Heather E. Golden<sup>c,\*</sup>, Charles R. Lane<sup>c</sup>

<sup>a</sup>Oak Ridge Institute for Science and Education, US Environmental Protection Agency, Office of Research and Development, Cincinnati, OH, USA

<sup>b</sup>Department of Food, Agricultural and Biological Engineering, Ohio State University, Columbus, OH, USA

<sup>c</sup>US Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Cincinnati, OH, USA

### Abstract

A hydrologic model, calibrated using only streamflow data, can produce acceptable streamflow simulation at the watershed outlet yet unrealistic representations of water balance across the landscape. Recent studies have demonstrated the potential of multi-objective calibration using remotely sensed evapotranspiration (ET) and gaged streamflow data to spatially improve the water balance. However, methodological clarity on how to “best” integrate ET data and model parameters in multi-objective model calibration to improve simulations is lacking. To address these limitations, we assessed how a spatially explicit, distributed calibration approach that uses (1) remotely sensed ET data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and (2) frequently overlooked biophysical parameters can improve the overall predictability of two key components of the water balance: streamflow and ET at different locations throughout the watershed. We used the Soil and Water Assessment Tool (SWAT), previously modified to represent hydrologic transport and filling-spilling of landscape depressions, in a large watershed of the Prairie Pothole Region, United States. We employed a novel stepwise series of calibration experiments to isolate the effects (on streamflow and simulated ET) of integrating biophysical parameters and spatially explicit remotely sensed ET data into model calibration. Results suggest that the inclusion of biophysical parameters involving vegetation dynamics and energy utilization mechanisms tend to increase model accuracy. Furthermore, we found that using a lumped, versus a spatially explicit, approach for integrating ET into model calibration produces a sub-optimal model state with no potential improvement in model performance across large spatial scales. However, when we utilized the same MODIS ET datasets but calibrated each sub-basin in the spatially explicit approach, water yield prediction uncertainty decreased, including a distinct improvement in the temporal and spatial accuracy of simulated ET and streamflow. This further resulted in a more realistic simulation of vegetation growth when compared to MODIS Leaf-

---

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\*Corresponding author at: 26 W. Martin Luther King Dr., MS-592, Cincinnati, OH 45268, USA., [golden.heather@epa.gov](mailto:golden.heather@epa.gov) (H.E. Golden).

Area Index data. These findings afford critical insights into the efficient integration of remotely sensed “big data” into hydrologic modeling and associated watershed management decisions. Our approach can be generalized and potentially replicated using other hydrologic models and remotely sensed data resources – and in different geophysical settings of the globe.

## Keywords

MODIS; Multi-objective; Calibration; Prairie Pothole Region; SWAT; Uncertainty

---

## 1. Introduction

Calibration of a hydrologic model involves constraining the solution space, i.e., the range of parameter combinations, to identify the most optimal parameter set that ostensibly represents watershed physics. The traditional approach of calibration with gaged streamflow data may produce a pseudo-accurate model, showing “acceptable” skill for streamflow simulation at target locations while incorrectly representing internal watershed processes (Maneta et al., 2017; Rajib et al., 2016a, 2018a). Such a model, giving right answers for wrong reasons (a characteristic of equifinality; Beven, 2012; Beven and Freer, 2001; Favis-Mortlock, 2004), might lead to incorrect interpretation of a hypothesis or an unsuitable selection of land management alternatives for future scenarios. The potential availability of remotely sensed streamflow estimates at ungaged locations is alluring (NASA, 2016), but the information contained in streamflow time series may not sufficiently capture how vertical fluxes evolve at different spatial and temporal scales within the watershed (Birkel et al., 2014; Li et al., 2018). Against this backdrop, using spatially distributed remotely sensed estimates of water balance components (e.g., soil moisture and evapotranspiration) affords multi-scale, multi-objective calibration of hydrologic models (Bai et al., 2018; Fatichi et al., 2016; Li et al., 2016), which may help to remedy this equifinality, or pseudo-accuracy, issue.

While remotely sensed surface soil moisture is being increasingly used in model calibrations (e.g., Milzow et al., 2011; Li et al., 2018; Parajka et al., 2009; Sutanudjaja et al., 2014; Wanders et al., 2014), its influence on the simulation of root zone moisture, evapotranspiration and streamflow could be limited depending on model structure (Chen et al., 2011, Han et al., 2012). Several studies (e.g., Brocca et al., 2012; Rajib et al., 2016a; Silvestro et al., 2015) have recommended constraining models with remotely sensed estimate(s) that can represent the entire/majority portion of the root zone, biotic (e.g., vegetation growth) and abiotic (e.g., energy exchange) processes. This is where remotely sensed actual evapotranspiration (ET) emerge as an effective data resource (e.g., Herman et al., 2018; Immerzeel and Droogers; 2008; Kunnath-Poovakka et al., 2016; Rientjes et al., 2013; Vervoort et al., 2014; Winsemius et al., 2008).

Currently, two primary issues exist related to the use of remotely sensed ET in model calibration. First, reproducible methods for handling spatially distributed and widely available “big data”, such as remotely sensed ET, are not prominent in the literature. For example, inefficient methods for using remotely sensed ET data as part of a multi-objective calibration of hydrologic models across a large number of sub-basins (or any other spatial

unit, e.g., grid-cells) may have resulted in the current perception that this approach improves ET but may have mixed or limited influence on streamflow simulations (e.g., López López et al., 2017; Tobin and Bennett, 2017). Second, most studies that have used remotely sensed ET data for model calibration focus only on calibrating surface, sub-surface and river routing parameters (e.g., Immerzeel and Droogers; 2008; Kunnath-Poovakka et al., 2016; Rientjes et al., 2013; Tobin and Bennett, 2017). This practice does not leverage the potential for biophysical parameters within a hydrologic model to improve the overall representation of its physical processes (Yang and Zhang, 2016; Zhu and Zhuang, 2015). It is yet to be explored whether biophysical parameters, such as those associated with vegetation dynamics, land-atmosphere interaction, and energy utilization, affect model calibration performance – with or without the use of remotely sensed ET data in calibration process. Our study fills these knowledge gaps using a modified version of the process-based semi-distributed hydrologic model Soil and Water Assessment Tool (SWAT) (Arnold et al., 2012; Gassman et al., 2007; Neitsch et al., 2011).

Through a series of calibration configurations, we investigate (i) the degree to which adding biophysical parameters not traditionally considered in a streamflow-only calibration effects hydrologic simulation, (ii) the difference in simulated hydrologic responses due to the inclusion of remotely sensed ET in a multi-objective calibration with gaged streamflow data, and (iii) the additional change in hydrologic responses if remotely sensed ET is utilized in a spatially distributed calibration approach. While this study methodologically benchmarks remote sensing-enabled SWAT calibration in a North American watershed, insights provided here should be applicable for other hydrologic models and geophysical settings as well.

## 2. Study area

We conducted this study on the ~1670 km<sup>2</sup> Pipestem Creek watershed in North Dakota, United States (Fig. 1). Located in the Prairie Pothole Region in the northern Great Plains, Pipestem Creek watershed landform consists of two ecoregions – the high-elevation Missouri Coteau in the west, and the Drift Plains in the eastern part of the watershed (EPA, 2011). The Coteau landscape in particular is pockmarked with thousands of wetland depressions, created by glacial retreat during the Wisconsinan period (Phillips et al., 2005; Winter and Rosenberry, 1995). These wetlands have been variously termed Prairie Potholes, including non-floodplain wetlands (NFWs) (Lane et al., 2018) amongst others (Tiner, 2003; Leibowitz, 2015; Mushet et al., 2015). Here, we will use the term surface depressions.

The surface depressions in the Prairie Pothole Region may be permanently inundated, ephemeral, or completely dry year-to-year (LaBaugh et al., 1998; Winter and Rosenberry, 1998). Primary water input in these surface depressions is the spring snowmelt and direct precipitation; as such water levels progressively drop through the summer months (Kantrud et al., 1989). This oscillatory dry-wet cycle makes ET a dominant driver of water, which responds to – and also affects – the “fluctuations and successions” of vegetation (Euliss et al., 2004; Gleason et al., 2008; van der Valk, 1981, 2005; van der Valk and Mushet, 2016; van der Kamp et al., 2016). With all these features, Pipestem Creek watershed is an ideal watershed for hydrologic modeling experiments.

### 3. Methodology

We constructed four calibration configurations of a modified SWAT model (Evenson et al., 2016) with increasing number of parameters and a progressively discretized way of optimizing them using remotely sensed ET data. The four configurations include:

- i. M1[Basic]: calibration using only streamflow data (Evenson et al., 2016);
- ii. M2[+Biophysics]: another application of M1, incorporating biophysical parameters;
- iii. M3[Lumped ET]: multi-objective calibration of M2 using both remotely sensed ET and streamflow data; parameters were optimized in a lumped approach commonly followed in many previous studies;
- iv. M4[Spatial ET]: the same as M3, except the parameters were optimized in a spatially distributed approach.

Based on the outcome of these calibration experiments, we quantified the effects of biophysical parameters and remotely sensed ET data on simulated water balance components (specifically, ET and streamflow). To further reveal the nature of change in prediction uncertainty, temporal and spatial variability of streamflow, and vegetation growth, we have conducted additional assessments using two configurations. In the following, we first outline the model and the data used, then provide a detailed description of the model configurations.

#### 3.1. A modified SWAT model

The major modeling components in SWAT include hydrology, vegetation dynamics, landscape and in-stream nutrient loading, stream temperature, and land management (Gassman et al., 2007). SWAT divides a watershed into multiple sub-basins, which are then further discretized into Hydrologic Response Units (HRUs) consisting of homogeneous land use, management, topographical, and soil characteristics (Neitsch et al., 2011). The modified SWAT model used in this study was originally developed by Evenson et al. (2016) with improved spatial representation and hydrologic functions. Evenson et al. (2016) conducted a series of novel modifications in the existing SWAT source code (Neitsch et al., 2011) to (i) enable a spatially explicit representation of surface depressions and hydrologic transport between surface depressions and uplands, (ii) improve seepage and sub-surface inflow simulation, and (iii) facilitate the simulation of fill-spill hydrologic relationships. There have been several applications of the SWAT model to specifically understand the hydrologic effects of surface depressions (e.g., Almendinger et al., 2011; Evenson et al., 2015, 2018; Liu et al., 2008; Rahman et al., 2016; Wang et al., 2008). SWAT has also been widely used for hydrological modeling in various geophysical settings worldwide, including major river basins in the United States (Rajib and Merwade, 2017), Europe (Abbaspour et al., 2015), Africa (Schuol et al., 2008) and Asia (Islam et al., 2017). Thus, findings from this study may be pertinent to worldwide SWAT model applications and at spatial scales much larger than the Pipestem Creek watershed.

### 3.2. Geospatial inputs and weather forcing

The 30-m National Elevation Dataset (USGS-NED, 2015), 30-m 2011 National Land Cover Database (NLCD) (Homer et al., 2015), and 1:250,000 State Soil Geographic Database (STATSGO) (NRCS, 2015; available within SWAT database) were the primary geospatial inputs in the modified SWAT model. To have a spatially explicit representation of surface depressions in the SWAT model, Evenson et al. (2016) first identified surface depressions using the National Wetland Inventory (NWI; <http://www.fws.gov/wetlands/>) and National Hydrography dataset (NHD; <https://nhd.usgs.gov/>) following Lane et al. (2012), and then estimated their storage capacities using a digital terrain constructed from 10-m NED (e.g., Lane and D'Amico (2010)). This approach provided unprecedented geospatial specificity to model a watershed as discretely as possible. Contrasting the 1217 HRUs that a SWAT model would produce for the Pipestem Creek watershed (based upon NED, NLCD, and STATSGO data), modified spatial representation by Evenson et al. (2016) had 62,721 HRUs within the same number of sub-basins ( $n = 29$ ).

We used the Daymet daily total precipitation and average maximum-minimum temperature as the primary weather inputs (Thornton et al., 2014; gridded data available at 1-km spatial resolution), while creating ancillary weather inputs (i.e., solar radiation, wind speed and relative humidity) using the SWAT model's built-in weather generator (WGN) (Neitsch et al., 2011). As we maintained identical weather inputs in all the four calibration configurations, possible propagation error resulted from Daymet and WGN data may have equivalent effect on the simulation results.

### 3.3. Mechanism for simulating ET and vegetation dynamics

We let SWAT estimate Potential Evapotranspiration (PET) using the Penman-Monteith (P-M) method such that ET simulations were compatible with MODIS (Moderate Resolution Imaging Spectroradiometer) algorithm (Section 3.4.2). Taking this PET as the maximum permissible limit, SWAT simulates an evaporative demand ( $ET_d$ ) that is spatially (across HRUs) and temporally (along daily time-steps) variable. SWAT's P-M equation to estimate plant demand (the transpiration component of  $ET_d$ ) requires Leaf-Area Index (LAI). LAI, representing the spatio-temporal dynamics of vegetation growth, is simulated via a biome-specific semi-empirical leaf development curve with several user-defined biophysical parameters like radiation-use efficiency and maximum canopy height. Soil demand or the evaporation component of  $ET_d$  follows an exponential relationship with above-ground vegetation, also simulated by leveraging similar parameters (Rajib et al., 2018a). Eventually, SWAT uses the existing moisture content in soil profile to scale  $ET_d$  into ET (Rajib and Merwade, 2016). This built-in architecture to simulate LAI and ET was retained in our modified SWAT model.

### 3.4. Reference data for model calibration and verification

**3.4.1. Streamflow**—For calibration and verification of the modified SWAT model, we used streamflow data from the only available US Geological Survey (USGS) gaging station at the watershed outlet (Fig. 1), but varied how this data was included in the multi-objective calibration depending on the model configuration (see Section 3.5). This gage had 41% estimated values over a 5-year period (753 of 1826 data points between 2009 and 2013),

especially prevalent during the winter months when icy conditions prevent gage reading. To avoid uncertainties implicit in statistical estimation of streamflow time-series (Chokmani et al., 2008; Hamilton, 2004), we only used the observed data in model calibration following Evenson et al. (2018 and 2016). However, given the large inter-annual hydrologic variability in the watershed (and the Prairie Pothole Region as a whole; van der Valk, 2005; van der Valk and Mushet, 2016; van der Kamp et al., 2016), it was deemed necessary to perform a long-term verification of the calibrated model leveraging the best available data. To facilitate this, we included both estimated and observed streamflow data for model verification purposes.

**3.4.2. Remotely sensed evapotranspiration and vegetation**—The 8-day total, ~1 km gridded ET data from MODIS (MOD16A2 collection 5; Mu et al., 2013) was geo-referenced and spatially aggregated into each of the 29 sub-basins of the Pipestem Creek watershed using a semi-automatic web-based tool (Rajib et al., 2018b). While dynamically accounting for the heterogeneity in size, shape and locations of the sub-basins, the tool puts area-weighted average value of 8-day total ET (mm H<sub>2</sub>O) from encompassing and/or intersecting MODIS grid-cells onto each of the sub-basins. Despite potential inaccuracies due to spatial aggregation (Zhang et al., 2016), this areawighted averaging is a reasonable scheme to create spatially explicit time-series of remotely sensed data for hydrologic modeling purposes (Ershadi et al., 2013; Rajib et al., 2018a).

To enable SWAT calibration at a daily time-scale, we temporally disaggregated the sub-basin level 8-day total MODIS ET data using the corresponding daily estimates from the North American Land Data Assimilation System (NLDAS- phase 2; NASA, 2017) (Eq. (1)):

$$ET_{i(M)} = ET_{T(M)} \times \left[ \frac{ET_{i(N)}}{ET_{T(N)}} \right] \quad (1)$$

here,  $M$  and  $N$  represent MODIS and NLDAS respectively;  $T$  refers to the total 8-day value; the index  $i$  denotes a particular day within that 8-day segment ( $T$ ). We applied the same data-processor tool to create subbasin level NLDAS time-series from its original ~12 km gridded estimates. Both MODIS and NLDAS estimates are based on the Penman-Monteith (P-M) method, however, assumptions on aerodynamic/surface resistances as well as the sources of land use and meteorological input data in their respective algorithms are different. Differences between the absolute values of MODIS and NLDAS ET did not affect  $ET_{i(M)}$  as we used NLDAS estimate in Eq. (1) only as a temporal scaling factor, without altering the total amount of ET estimated by MODIS in each 8-day window ( $ET_{T(M)}$ ). Therefore, we labeled the MODIS-NLDAS hybrid estimate ( $ET_{i(M)}$ ) simply as MODIS ET.

Logic suggests that simulated ET will follow the spatio-temporal trend of reference ET that constrains a hydrologic model during the calibration process (MODIS data in our study). Therefore, to build multiple lines of evidence on the overall improvement of the model in representing watershed characteristics (and particularly the energy balance), we used LAI as an independent reference variable not included in the calibration process. Our LAI based evaluation is novel because it is the first to use vegetation dynamics for cross-verification when remotely sensed ET is applied in model calibration. Previous studies

on model calibration using remotely sensed ET measured performance improvements based solely on simulated ET outputs (e.g., Herman et al., 2018; Immerzeel and Droogers; 2008; Rientjes et al., 2013; Tobin and Bennett, 2017) – no other variables were considered. To compare LAI simulations across model configurations, we obtained watershed-average LAI data from a 4-day total MODIS product (MCD15A3H version 6; Myneni et al., 2015). Creation of watershed-average LAI time-series from the original ~500 m gridded data again required employing the semi-automatic data-processor tool. However, no temporal re-scaling was necessary in this case as we computed 4-day total values from SWAT’s daily outputs to match temporal resolution of the original MODIS data.

With the advances in satellite sensor technologies and data assimilation techniques, various multi-source, robust ET products have emerged in recent years. Some of these relatively new ET estimates are available at a daily time-step, not requiring any temporal re-scaling/disaggregation (e.g., Zhang et al., 2016, 2010). Yet we preferred using MODIS ET and LAI data because these are open-source global datasets with long-term and continuous availability, reasonably good spatiotemporal resolution, and more importantly, well-documented quality assessments (e.g., Mu et al., 2011, 2013; Ruhoff et al., 2013; Yan et al., 2016; Yang et al., 2006; Velpuri et al., 2013). Because uncertainties are common in any remotely sensed product, none of these datasets should be considered as actual observations.

### 3.5. Calibration configurations

Fig. 2 schematically outlines the modeling architecture involved in the four calibration configurations, namely M1[Basic], M2[+Biophysics], M3[Lumped ET], and M4[Spatial ET], all explained below in individual subsections. All calibrations were conducted at daily time-step using the Sequential Uncertainty Fitting algorithm (SUFI-version 2) in the SWAT-CUP platform (Abbaspour, 2015; Abbaspour et al., 2007). To be consistent with Evenson et al. (2018 and 2016), each configuration went through a 3-year initialization (2006–2008) followed by a 3-year calibration (2009–2011). The iterations were divided into two successive batches. SUFI-2 uses Latin Hypercube Sampling (LHS) to set potential parameter combinations based on the specified initial range for every calibration parameter. Using an inverse optimization approach, SUFI-2 then finds a narrower range of parameter values at the end of the preliminary batch to initiate the next batch for further iterations. We selected Kling-Gupta Efficiency (KGE; Gupta et al., 2009; Kling et al., 2012) as the objective function which was maximized (for streamflow and ET, depending on the model configuration) to find the most optimal parameter set. KGE decomposes Nash-Sutcliffe Efficiency (NSE) and Mean Squared Error (MSE) into a three-dimensional criteria space and finds out a Pareto front in terms of the shortest Euclidean distance (ED):

$$KGE = 1 - ED = 1 - \sqrt{\{(r - 1)^2 + (\mu_s/\mu_o - 1)^2 + (\sigma_s/\sigma_o - 1)^2\}} \quad (2)$$

where  $r$  represents the correlation,  $\mu_s/\mu_o$  and  $\sigma_s/\sigma_o$  respectively represent bias and variability ratio between the simulated and observed variable,  $\mu$  and  $\sigma$  are the mean and standard deviation of the variable; the indices  $s$  and  $o$  denote simulation and observed data,

respectively. KGE ranges from  $-\infty$  to 1, with a value closer to 1 indicates a relatively accurate model.

As noted above, our goal was to keep the total length of the simulation consistent with previously published research on Pipestem Creek watershed using the same modified model (e.g., Evenson et al., 2018, 2016). However, the need for long-term verification of model outputs and the limited availability of observed streamflow data (see Section 3.4.1) made it challenging to follow the conventional split-sample approach where model calibration and verification is conducted over two non-overlapping “independent” time-periods. We therefore evaluated (verified) our calibrated model configurations over a 5-year period (2009–2013). Despite this partial overlap with the calibration period (2009–2011), our approach resulted in an acceptable degree of independence among the calibration and verification datasets for three reasons: (1) we excluded estimated streamflow data from the model calibration but included them in the verification stage, (2) our model verification with streamflow and ET data included two more years (2012–2013) beyond the calibration period, and (3) we used LAI data as a measure of cross-verification. To facilitate further independent evaluation, and assess how a multi-objective use of ET and streamflow in SWAT calibration might change the model’s response during extreme flood events, we also excluded a specific portion of observed stream-flow record (August 2011) from the calibration period and used it for model verification.

**3.5.1. Baseline model (M1)**—M1[Basic] involved calibration only with streamflow data at the watershed outlet, and included 21 parameters related to surface/subsurface runoff, channel routing, and snow accumulation processes in the calibration (Fig. 3 and Table A1 in Appendix A). All of these parameters, their respective initial range, and the M1 configuration as a whole, were identical with the calibration performed by Evenson et al. (2016), differing in this study only in the management of snow parameters. We excluded snow parameters after the first batch of iterations to be able to compare with M4 outputs (see below in Section 3.5.4). M1 was our baseline to measure the degree of model improvement in subsequent configurations.

**3.5.2. Effect of biophysical parameters (M2)**—M2[+ Biophysics] was the same as M1[Basic], except with the addition of 10 parameters related to biophysical processes (see Fig. 3). There are more than 20 such parameters in SWAT geodatabase (Neitsch et al., 2011); however, they are often underutilized and assigned default values. The parameters we selected collectively represent a range of processes including vegetation growth, land-atmosphere feedback, and energy use mechanisms. Having more parameters would push the computational overhead of our hyper-resolution model (~63,000 HRUs) beyond a workable limit. Initial ranges for the 10 additional parameters were set following the suggestions from SWAT developers (Abbaspour, 2015).

Most of the contemporary SWAT applications, including those on watersheds with abundant surface depressions (e.g., Evenson et al., 2016; Rahman et al., 2016; Ikenberry et al., 2017), do not consider biophysical parameters in model calibration. The underlying notion is that biophysical parameters are least sensitive for hydrologic simulation compared to those representing runoff, routing and snow accumulation processes. However, due to the heavily



conceptualized equations that SWAT employs to capture biophysical interactions, it is possible that including the relevant parameters within a simple streamflow-only calibration can make notable difference in the overall spatial hydrologic response of the model. We performed the built-in global sensitivity analysis scheme in SUFI-2 (Abbaspour, 2015; Rajib and Merwade, 2016) to substantiate this notion.

Optimizing 31 parameters through the first batch of iterations, the second batch in M2[+ Biophysics] excluded the snow parameters which made it comparable with M1[Basic] (see Fig. 2). Thus, differences in outputs between these two configurations identified the effect of biophysical parameters on model predictability.

**3.5.3. Use of remotely sensed ET data in model calibration (M3)**—The most optimal model resulting from the first batch of iterations in M2[+ Biophysics] became the starting point for M3[Lumped ET] (Fig. 2). The difference between these two configurations was solely that M3[Lumped ET] included MODIS ET as an additional model constraint in the second (and final) batch of iterations. The optimization algorithm (SUFI-2) used 30 constraints together - 29 time-series of MODIS ET (one for each sub-basin) and a single time-series of streamflow data (watershed outlet). The 30 objective functions (essentially 30 KGEs), as shown in Eq. (3), were given equal weighting factor to form an aggregated value (KGE’):

$$KGE' = \sum_{i=1}^{n_f} w_{f_i}(KGE_{f_i}) + \sum_{j=1}^{n_e} w_{e_j}(KGE_{e_j}) \quad (3)$$

here,  $n$  and  $w$  are the number of objective variables (reference datasets) and the weight assigned to each of them, respectively; the values  $f$  and  $e$  stand for streamflow and ET, respectively. Also,  $i$  denotes the streamflow gaging stations and  $j$  denotes sub-basins with MODIS ET data. The parameter set producing the highest KGE’ was the most optimal one.

Unlike the runoff/routing parameters in SWAT, the majority of its biophysical parameters cannot be defined at HRU or sub-basin level. Instead, these parameters are associated with specific plant/biome type, and it is possible that the same biome type may exist in multiple HRUs or sub-basins. Because of this inconsistency, the M3 configuration restricts every parameter so that it can only be changed “in a unique way” irrespective of the HRU/sub-basin. More specifically, in successive iterations during the calibration, a particular parameter must have an identical amount/rate of change across the entire watershed, and that amount/rate is pre-selected from a Latin Hypercube Sampling of its initial range of values. This is essentially a lumped approach notwithstanding the inclusion of spatially distributed MODIS ET data in the calibration process (hereafter, referred as M3[Lumped ET]). Previous studies indiscriminately applied this lumped approach even though biophysical parameters were not included in their respective calibrations; consequently, it became common to lump all the available ET time-series into a single calibration target (Eq. (3)) (e.g., Immerzeel and Droogers; 2008; Rientjes et al., 2013; Tobin and Bennett, 2017).

Regardless of these limitations, M3[Lumped ET] configuration provided new insights. As noted, considering that the use of MODIS ET in M3[Lumped ET] was the only difference

from M2[+ Biophysics], comparison of these two configurations can distinguish the effect of remotely sensed data from that of biophysical parameters which had not been addressed in any previous study.

**3.5.4. Value of spatially distributed calibration using remotely sensed ET (M4)**—M3[Lumped ET] and M4[Spatial ET] were identical, except a “spatially distributed” calibration approach was implemented in the M4 configuration. The goal of M4[Spatial ET] was to leverage the maximum potential of spatially distributed MODIS ET data and allow highly discretized parameter optimization by calibrating 29 sub-basins individually. Each sub-basin had two constraints to find optimal parameters - a representative time-series of MODIS ET specific to the subbasin and a time-series of streamflow data at the watershed outlet.

Lack of spatial heterogeneity in snow accumulation processes is a limitation in the current SWAT modeling framework. All the snow parameters in SWAT can only be defined at watershed level (Neitsch et al., 2011). That is why iteration of snow parameters was technically impossible in a sub-basin level calibration approach such as M4[Spatial ET]. In order to have an even comparison with M4[Spatial ET], snow parameters were not adjusted after the first batch of iterations in any of the four configurations. This problem is very specific to SWAT; continued optimization of snow parameters throughout the calibration process, and that in an M4[Spatial ET]-type setup, might be doable in the case of other hydrologic models.

## 4. Results

### 4.1. Model predictability across the spectrum of calibration configurations

Both biophysical parameters and remotely sensed ET data influenced model predictability. This is evidenced by the relative change of model performance among the four configurations, in terms of KGE for simulated daily ET at sub-basin level and streamflow at the watershed outlet (Fig. 4a–c). Considering the improved spatial representation, seepage/sub-surface inflow, and fill-spill hydrologic simulation of surface depressions (Evenson et al., 2016), the baseline configuration M1 should exhibit minimal inaccuracy in reproducing hydrologic processes. As expected, the high KGE score (0.90) in M1[Basic] for the most optimal streamflow simulation during the calibration period fit the traditional SWAT evaluation rubric of a “very good” hydrologic model (Moriassi et al., 2015). But pseudo-accuracy was evident based on the low watershed-average KGE (~0.60) for ET, demonstrating potential limitations of simulated vertical water fluxes in the M1[Basic] configuration. This may, in part, be an underlying reason that the streamflow simulation in M1[Basic] did not perform well in the verification phase (see Fig. 4c). The pseudo-accuracy was intriguing as it indicated the need for supplementing spatially explicit hydrologic transport between surface depressions and fill-spill mechanisms in a hydrologic model with measures that potentially account for other process uncertainties in the model. One way to address this issue was to focus on biophysical parameters. Inclusion of biophysical parameters in model calibration is appropriate because of the coupled relationships the

vegetation and energy related processes have with the partitioning of water at different scales of a watershed.

The same model (M1[Basic]) showed a different, and slightly improved (see Fig. 4c), model performance simply when additional bio-physical parameters were included in a streamflow-only calibration (M2[+ Biophysics]). Compared to the variable spatial nature of ET simulation in the M1[Basic] configuration with sub-basins having very high (0.75) to very low (0.40) KGEs, M2[+ Biophysics] suggested an improved spatial accuracy via more consistent KGEs (0.50–0.70) across the watershed (Fig. 4b). This seemingly improved model state in M2[+ Biophysics] resonated in the streamflow verification showing relatively higher KGE than M1[Basic] (Fig. 4c). While these results confirmed the sensitivity of biophysical parameters to mediate hydro-logic processes in a watershed with abundant surface depressions and fill-spill phenomena (Fig. 3), one can expect that these parameters would affect hydrologic simulations in other geophysical settings too including agricultural and forested landscapes (e.g., Quinton et al., 2010; Yang and Zhang, 2016). However, it is expected that the calibration of biophysical parameters is more effective in the presence of constraint datasets “relevant” to their physical processes (e.g., ET and vegetation growth).

In the next configuration with MODIS ET as an additional calibration target (M3[Lumped ET]), we found considerable improvement in SWAT’s ET simulation (Fig. 4a and b). Hypothetically, improved ET implies improved physical representation of the model’s water balance components. As such, other hydrologic processes including streamflow should reflect traces of improvement too. This was evident in M3[Lumped ET] through a distinctively increased streamflow KGE during the model verification, although streamflow KGE decreased in the calibration period compared to M1[Basic] and M2[+ Biophysics] (Fig. 4c).

The goal of the M4[Spatial ET] configuration was to evaluate whether using the same calibration targets of M3[Lumped ET] in a spatially distributed approach can further improve model performance. Results from M4[Spatial ET] were the most telling in terms of KGE scores for both ET and streamflow simulations. The watershed-average KGE for ET was 0.70 in M4[Spatial ET], particularly contrasting to 0.60 in M1[Basic] and the highest among the four configurations. As we moved through our calibration experiments from one level to the next (M1 to M4), we identified increasing gain in the overall spatial accuracy of ET simulations (Fig. 4a and b). We interpreted this drift in spatial accuracy (“pulling of models”: Efstratiadis and Koutsoyiannis, 2010) by looking onto the spread and skewness of sub-basin level KGE distributions. For instance, the best spatial accuracy in M4[Spatial ET] (Fig. 4a) was reflected through the smallest spread (range of KGEs) and the maximum skewness (toward high KGEs) (Fig. 4b). This evolution of model performance quantitatively showed how the state of a model can transitionally improve in predictability and shrink in equifinality with a spatially explicit ET calibration.

In line with its spatial accuracy in ET simulation, M4[Spatial ET] showed remarkably improved streamflow simulation skill during the model verification (KGE = 0.75; Fig. 4c). While these results collectively made M4[Spatial ET] the superior configuration, another key aspect made marked difference in the efficacy of M4’s spatially distributed

calibration approach. Unlike M3[Lumped ET], M4[Spatial ET] did not show a drastic drop in streamflow calibration performance. It rather retained nearly the same calibration performance as M1[Basic] and M2[+ Biophysics], despite a multi-objective optimization with the same total volume of ET data used to calibrate M3[Lumped ET]. In this way, M4[Spatial ET] fully complied with the theoretical cause-and-effect expectancy that an improved representation of water and energy balance should result in improved streamflow simulation as well.

## 4.2. Additional insights on hydrologic representation

To further explore the overall improvements achievable through the M4[Spatial ET] configuration, below we present a holistic evaluation of its outputs focusing on three different perspectives: (i) uncertainty in predicted water balance, (ii) temporal and spatial variability of streamflow, and (iii) vegetation growth. Here, M4[Spatial ET] is contrasted only against M1[Basic] given these are the two extreme configurations in terms of model accuracy.

**4.2.1. Uncertainty in predicted water balance**—M4[Spatial ET] showed substantially reduced uncertainty in simulated water yields compared to M1 (Fig. 5). We chose water yield as a holistic index of water balance because it is the amount of water contributed to stream after the model accounts for fill-spill hydrology and all hydrological losses (e.g., to evapotranspiration and soil infiltration; Neitsch et al., 2011). The relative uncertainty mapped in Fig. 5a reflects the average 95% prediction uncertainty (95ppu) band width in calibrated water yield outputs at the sub-basin level, normalized with respect to the minimum and maximum band widths across all the subbasins (irrespective of a particular configuration). Here, our simulations showed consistently smaller prediction uncertainty in M4[Spatial ET] irrespective of the location and timing of high flow events within the watershed (Fig. 5b).

**4.2.2. Temporal and spatial variability of streamflow**—We compared the temporal variability of streamflow between M1[Basic] and M4[Spatial ET] in terms of their flow exceedance probability at the watershed outlet (Fig. 6). Irrespective of the flow regime (high flow, medium flow, and low flow), the most optimal streamflow output from M4[Spatial ET] generally showed better conformity with USGS data compared to M1[Basic]. Taking into account fill-spill processes during extreme events, improved high flow simulations in M4[Spatial ET] indicate enhanced functioning of the modified spatial representation and hydrologic transport proposed by Evenson et al. (2016) (Fig. 6). However, given the high inter-annual variability of ponded area in the Prairie Pothole Region (van der Valk and Mushet, 2016), it is not surprising that hydrologic models would struggle in reproducing average-condition water balance in the watershed. Accordingly, M1[Basic] and M4[Spatial ET] both resulted in large departures from USGS data during medium flow conditions, however, the difference (USGS *minus* SWAT) was consistently smaller in the case of M4[Spatial ET]. Furthermore, simulated streamflow in M4[Spatial ET] was in closer proximity with USGS data 90% of the time during the 2009–2013 period, indicated by the low flow conditions in Fig. 6. This was expected given the influence of ET on the diffusivity of soil moisture and mobility of groundwater when there is no surface runoff

(Tobin and Bennett, 2017; Szilagyi et al., 2007). These results also afford temporal insight to the variation of streamflow KGE values shown in Fig. 4c.

In addition to temporal dynamics, we found prominent differences between M1[Basic] and M4[Spatial ET] in the spatial variability of streamflow even when both configurations produced peak flows at the same time-stamp as the USGS data (Fig. 7). Apart from this identical temporal response particularly in the two specific flood dates shown here, we also noted the peak magnitudes which were distinctively more accurate in the case of M4[Spatial ET]. The combined impression from this duo (i.e., timing and magnitude) suggest that the spatial variability captured in M1[Basic] may have unrealistic representation of landscape response.

**4.2.3. Model accuracy in simulating vegetation growth**—The traditional way of gaging the level of physical realism in a hydrologic model is to assess improvements in water balance components. Therefore, models are often not assessed for accuracy in energy balance components, although the energy balance can considerably influence water balance (Hurkmans et al., 2008; Rajib et al., 2018a; Troch et al., 2009). We considered watershed-average vegetation growth (LAI) as an indicator that implicitly represents the collective effects of improved water and energy balances in the model (Long et al., 2014).

M1[Basic] produced unrealistically high LAI values compared to M4[Spatial ET] and MODIS data (Fig. 8). The lack of biophysical parameters in the M1[Basic] calibration, and therefore the use of default values that, for the most part, stayed spatially constant across the watershed, undermined the ability of the model to simulate vegetation dynamics. The inaccurate LAI in M1[Basic] indicated another persistent limitation of the model suggesting that a modified water retention/discharge (fill-spill) mechanism (Evenson et al., 2016) may only have limited influence on other hydrologic processes in a watershed. Although SWAT's default LAI development formulation was not modified in any of our model configurations (based on SWAT's internal structure; see Section 3.3), the parameter values defining the shape and inflection points of the LAI curve could be varied spatially in the M4[Spatial ET] calibration, and that using spatially explicit ET data. Therefore, the M4 LAI values compare better with the MODIS LAI. This improved representation of vegetation dynamics in M4[Spatial ET] compared to M1[Basic] was a reasonable indicator that the model was efficiently reproducing both water and energy partitioning, which may affect future biogeochemical applications of the model including carbon storage and nutrient cycling.

## 5. Discussion

### 5.1. Managing potential limitations involving remotely sensed data for calibration

Using remotely sensed data to calibrate a hydrologic model will inevitably produce different outputs compared to the same model that uses gaged streamflow data as the calibration objective. It is questionable, however, whether the differences in the remote sensing-integrated model emerge for right reasons. To enable physically meaningful application of remotely sensed ET in model calibration, and therefore get the right changes in model predictability for the right reasons, we took measures to overcome the following challenges.

**5.1.1. Is multi-objective calibration with ET and streamflow more effective, compared to using ET alone?**—Simultaneous use of streamflow and ET in hydrologic model calibration is recommended as a “best practice” (e.g., Kunnath-Poovakka et al., 2016; Rientjes et al., 2013; Vervoort et al., 2014). Studies using remotely sensed soil moisture preferred the same (Li et al., 2018; Rajib et al., 2016a; Wanders et al., 2014). A model calibrated against only ET or soil moisture would produce too little or too much of a particular vertical water flux (to the atmosphere or through soil horizons) unless the horizontal water-routing in the model are also simultaneously adjusted. The outcome would be a highly equifinal model having good accuracy for a specific vertical flux but adverse or no change in streamflow simulation accuracy (e.g., López López et al., 2017; Tobin and Bennett, 2017; Wanders et al., 2014). For similar reasons, decoupling the horizontal water-routing component is a potential reason why some current-generation hybrid land surface-river routing models exhibit acknowledged timing problems in vertical fluxes as well as in streamflow simulations (e.g., Lin et al., 2018). Therefore, it is rational to accept that watershed hydrologic responses can be best captured when both the vertical and the horizontal flux components in the model are constrained with relevant calibration targets, such as ET and streamflow.

The amount of uncertainty in remotely sensed ET is purportedly higher than that in gaged streamflow data (Vervoort et al., 2014). In our initial exploration (not shown), we found that MODIS ET, if used from the very beginning of the calibration process, inserts propagating error so that the accuracy of simulated streamflow (with respect to gage station data) cannot be improved beyond a certain level. We also found that this model state is irreversible and may not be reinstated without changing the sequence of introducing MODIS ET data into the calibration process. Therefore, we hypothesized that calibrating the model first with gaged streamflow, a relatively well-verified in-situ dataset, produces a reasonable water balance that could serve as a baseline for further adjustment of hydrologic fluxes through a multi-objective approach involving both streamflow *and* ET. This two-step calibration is also favorable for applying other remotely sensed datasets including soil moisture for which “systematic bias” is a deterring factor (Draper et al., 2009; Reichle and Koster, 2004). Simulated soil moisture output from a model that is pre-calibrated against streamflow data can be used to reduce systematic bias in the remotely sensed soil moisture before it is integrated with the model through a data assimilation and/or a recalibration process (e.g., López López et al., 2017; Rajib et al., 2016a; Wanders et al., 2014).

**5.1.2. Is it justified to calibrate ET-related parameters separately?**—Related to the two-step calibration explained above, it may not be physically coherent to calibrate biophysical or ET-related parameters separately. Herman et al. (2018) applied remotely sensed ET data to adjust only four additional parameters in a SWAT model, excluding all the other surface, subsurface, and routing parameters that were initially calibrated using streamflow data in a preceding step. As we performed global sensitivity analysis, several parameters with apparent high sensitivity in M1[Basic] (e.g., main channel Manning’s  $n$ ) were ultimately found to be less sensitive in the presence of biophysical parameters (see Fig. 3). This provides supportive evidence that parameters in processbased hydrologic models (e.g., SWAT) or coupled atmospheric-land surface models (e.g., WRF-Hydro, Gochis et al.,

2015) represent an integrated natural system. As such, they have an unknown degree of mutual relationships that may not be explicitly evident from their mathematical equations (Sivakumar and Singh, 2012). Therefore, not using a common set of parameters throughout the calibration process convolutes the effect of a new objective function (remotely sensed data) with that of the additional parameters initially not considered in preceding iterations. Such a calibration protocol may not be ideal, because it cannot identify whether the improvement in model simulation is solely due to the use of remotely sensed data.

## 5.2. Benchmarking a calibration strategy using spatially distributed remotely sensed data

The substantial disparity of simulation accuracy in M4[Spatial ET], compared to M1[Basic], M2[+ Biophysics] or M3[Lumped ET], ratio-nalizes the “spatially distributed use” of remotely sensed ET data in hydrologic model calibration. While this corroborates suggestions made in previous literature (e.g., Immerzeel and Droogers, 2008; Rientjes et al., 2013; Tobin and Bennett, 2017), assessments in these studies were mostly based on an M3[Lumped ET]-type configuration. Our study is the first to identify that M3[Lumped ET] produces an ostensibly good but actually sub-optimal model state. It is therefore worthwhile to review what factors undermine M3[Lumped ET] so that one can benchmark M4[Spatial ET] as a better approach for model calibrations, especially on large spatial scales.

The sub-optimality of M3[Lumped ET] was apparent mainly through the distinctive decrease in its streamflow calibration performance, notwithstanding the improvement it offers in ET simulation (Section 4.1). Importantly, Herman et al. (2018) showed similar tendency with a number of optimization algorithms including Non-dominated Sorted Genetic Algorithm II and Monte Carlo Simulations. It could also be a problem specific to the source of our data (MODIS ET; Mu et al., 2011, 2013), considering the uncertainty commonly seen in all remotely sensed estimates. However, previous studies reported similar results calibrating respective hydrologic models with ET data from the Surface Energy Balance System (SEBS) (Rientjes et al., 2013), Atmosphere-Land Exchange Inverse (ALEXI) model (Herman et al., 2018), and Global Land Evaporation-Amsterdam Model (GLEAM) (López López et al., 2017; Tobin and Bennett, 2017). This was evident when models were calibrated with soil moisture data as well (Li et al., 2018; Rajib et al., 2016a; Wanders et al., 2014). Given the above, the sub-optimality problem does not seem fundamental to the data, model or optimization algorithm.

We suggest that the sub-optimality in M3[Lumped ET] was, in part, an implementation problem in this configuration’s set-up that requires the optimization algorithm to handle too many objectives simultaneously. In the context of this study, the SUFI-2 parameter identification program resulted in an “averaging effect” while searching for a parameter combination that was optimal for all the 30 objectives (or constraints; Section 3.5.3; Eq. (3)). What had not been addressed in the existing literature is the inefficient use of emerging earth data resources and/or inappropriate handling of model parameters (e.g., Herman et al., 2018; López López et al., 2017; Tobin and Bennett, 2017; see Section 5.1), which are also viable factors affecting model performance in general.

The potential explanations for the limited performance or sub-optimality in M3[Lumped ET] have contributed to an evolving misperception that calibration with remotely sensed ET or

soil moisture data may not improve streamflow simulations. As a result, there is a growing number of remote sensing studies that discourage the use of “lowered streamflow calibration performance” as a generic goodness metric to label a model’s overall streamflow prediction skills (Kunnath-Poovakka et al., 2016; Rajib et al., 2016a; Rientjes et al., 2013; Wanders et al., 2014). Li et al. (2018), in a multi-objective calibration using soil moisture and streamflow, reported substantially increased streamflow accuracy when their most optimal model was verified at locations that were not included in the calibration process, despite an apparently poor performance in target calibration locations. This was analogous to our M3[Lumped ET] results, which had a higher KGE value than M1[Basic] and M2[+ Biophysics] in the streamflow verification process (Fig. 4c), indicating model improvement in the right direction. Nonetheless, continued applications of M3[Lumped ET], given what our results suggest, could produce acute problems in large scale hydrologic modeling (e.g., Emerton et al., 2016; Lin et al., 2018). With more objectives to target and sub-basins to calibrate (i.e., a very large number of streamflow and ET time-series), M3[Lumped ET] might result in diminishing returns with no net improvement in water balance.

The M4[Spatial ET] configuration moves beyond the “averaging effect” issue, even though it uses the same total volume of MODIS datasets as M3[Lumped ET]. We therefore considered M4[Spatial ET] an improved approach compared to M3[Lumped ET]. However, it is computationally expensive to operate an M4[Spatial ET]-type setup and harness the maximum capacity of spatially distributed data by calibrating each of the sub-basins individually. At this point, we could argue that one can still adopt M3[Lumped ET] and make a tradeoff by accepting a calibrated model that is generally representative of internal watershed processes but relatively poor in streamflow simulation. Yet this argument lacks scientific integrity because it gives an impression that models with persistent problems are acceptable. The tradeoff argument also weakens the value of remote sensing integrated hydrologic modeling amidst the new trend of large scale streamflow/flood forecasting initiatives world-wide (e.g., Li et al., 2016; Lin et al., 2018). With the recent advancements in open-access, cyber-enabled, high performance calibration platforms such as SWATShare (Rajib et al., 2016b) and their potential interoperability with observational data resources (Horsburgh et al., 2016; Morsy et al., 2017; Tarboton et al., 2014), developing a computationally efficient reproducible workflow for the M4[Spatial ET]-type calibration configuration is no more impossible.

## 6. Summary and conclusion

It is not surprising that a hydrologic model may misrepresent the true state of water and energy balances given the complexity involved in simulating watershed hydrology correctly. One approach that has been increasingly used to improve simulation accuracy, particularly in watersheds with abundant surface depressions, is the modification of existing model frameworks to improve spatially explicit hydrologic transport between and among surface depressions and other water bodies, as well as retention-discharge (fill-spill) mechanisms. In this study, we hypothesized that the accuracy of such a modified model can be further improved by including biophysical parameters and remotely sensed ET data in the model calibration process. Specifically, we constructed four calibration configurations of a modified SWAT model, namely M1[Basic], M2[+ Biophysics], M3[Lumped ET] and



M4[Spatial ET], with increasing number of parameters and a progressively discretized way of optimizing them using MODIS ET data. Based on the outcome of this experiment, the following conclusions were drawn.

1. Despite modifying the hydrologic transport and water retention/discharge (fill-spill) mechanism, calibration of a SWAT model involving the traditionally used runoff-routing parameters and using only streamflow data (M1[Basic]) was less accurate in its simulated water balance components (namely ET and streamflow) than the other configurations with improved calibration methods (M2-M4).
2. The same model showed slightly improved ET and streamflow simulations when biophysical parameters were included in the calibration process (M2[+ Biophysics]). While this finding confirms the sensitivity of biophysical parameters, we suggest that their calibration would be effective (and meaningful) if the model is constrained with datasets that are more relevant to vegetation dynamics, land-atmosphere interaction, and energy utilization processes (e.g., ET and/or LAI).
3. Regardless of the lumped nature of calibration, we found that the M3[Lumped ET] (i.e., the M2[+ Biophysics model calibrated with sub-basin MODIS ET data and gaged streamflow aggregated as a single optimization objective) noticeably increased the accuracy of SWAT's ET simulation compared to M1[Basic] and M2[+ Biophysics]. This configuration also resulted in improved streamflow simulations during model verification, indicating a transition of the model state towards increased predictability and reduced equifinality. Despite the seemingly improved water and energy balance, streamflow calibration performance in M3[Lumped ET] was distinctively lower than M1[Basic] and M2[+ Biophysics] configurations. We argued that this sub-optimality problem is not specific to a model, remotely sensed dataset or optimization algorithm. Instead, it is an implementation problem due to the aggregation of too many objectives, which imparts an "averaging effect" on the optimized solution. This may end up producing no net improvement of the model in large spatial scales.
4. Our results confirmed the superiority of M4[Spatial ET] (i.e., the model configuration with a spatially distributed approach such that each sub-basin was calibrated individually) over M3[Lumped ET] for overall model accuracy, although both configurations used the same total volume of remotely sensed data. Further, a one-to-one comparison between the two extreme configurations (i.e., M4[Spatial ET] versus M1[Basic]) indicated that an M4-type setup may exhibit reduced uncertainty in the simulation of landscape's water yield, with distinctly different temporal and spatial variability of streamflow and more accurate representations of vegetation growth with respect to MODIS LAI data. We therefore conclude that modifying certain process descriptions of a hydrologic model (e.g., hydrologic transport between surface depressions, fill-spill mechanism) (M1[Basic]) may not sufficiently improve overall partitioning of water and energy in a watershed unless the model is calibrated with remotely sensed ET data, and that including biophysical parameters (M4[Spatial ET]).

The substantial disparity between the two extreme configurations (i.e., M4[Spatial ET] versus M1[Basic]) bears immense practical implications. The choice of model configurations can impact how scientists and managers quantify and interpret the hydrological and biogeochemical influence of surface depressions under different climate, land use, and potential depression loss/restoration scenarios. Using M1[Basic] for modeling landscape hydrological and nutrient hotspots/hot moments may lead to inefficient watershed management decisions. Therefore, we suggest that the spatially distributed calibration approach using remotely sensed ET data and including important biophysical parameters (M4[Spatial ET]) will lead to a more accurately calibrated model that does improve the representation of a watershed's water and energy balance. While our methodological insights benchmarked efficient utilization of remotely sensed big data, the findings further provided scientific evidence on the evolution of a hydrologic model in the continuum of equifinality.

## Acknowledgements

This paper has been reviewed in accordance with the US Environmental Protection Agency's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Statements in this publication reflect the authors' professional views and opinions and should not be construed to represent any determination or policy of the US Environmental Protection Agency. We appreciate helpful suggestions from Muluken Muche, Sumathy Sinnathamby, and anonymous journal reviewers.

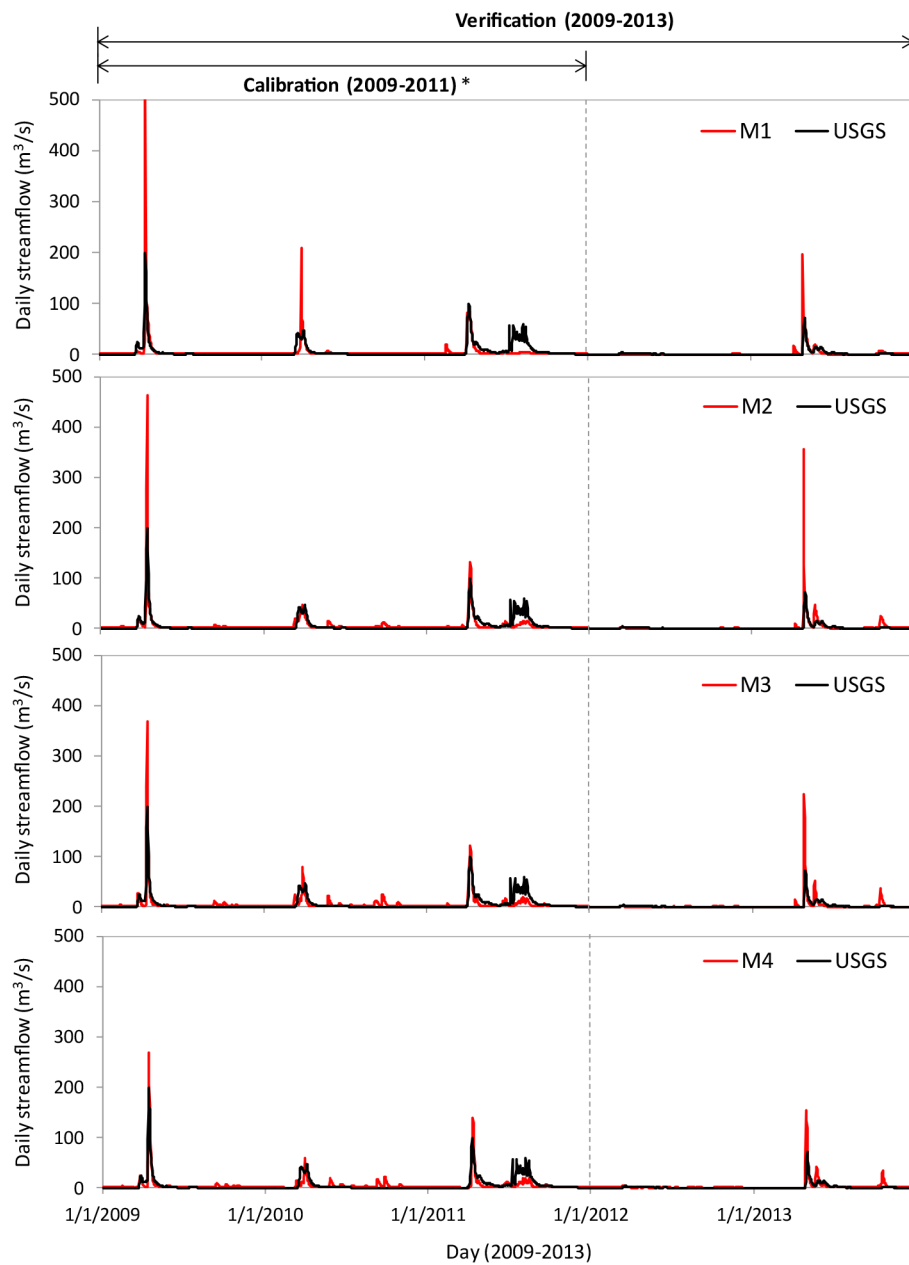
## Appendix A

**Table A1**

Calibration parameters, their respective initial ranges and the spatial scale of iteration. All parameters, except CN2 and SOL\_AWC, were varied such that the original parameter value gets replaced by another value within the assigned range. CN2 and SOL\_AWC were adjusted with a multiplying factor relative to the original value. Calibration of snow parameters was discontinued after the first batch of iterations. Biophysical parameters were not calibrated in the M1[Basic] configuration. See Section 3.5 and Fig. 2 for details.

No.	Parameter	Name	Spatial scale	Initial range
<i>Surface/sub-surface runoff</i>				
1	GIW_K	Hydraulic conductivity of GIW subsurface (mm/h)	HRU	0–3.6
2	GW_DELAY	Groundwater delay (days)	HRU	0–500
3	GW_REVAP	Groundwater 'revap' coefficient	HRU	0.02–0.2
4	RCHRG_DP	Deep aquifer percolation fraction	HRU	0–1
5	CN2	Curve Number	HRU	–0.15–0.15
6	ALPHA_BF	Baseflow recession constant	HRU	0–1
7	SOLAWC	Soil available water capacity (mm/mm)	HRU	–0.2–0.2
8	EPCO	Plant uptake compensation factor	Basin	0–1
9	ESCO	Soil evaporation compensation factor	Basin	0–1
10	REVAPMN	Threshold for 'revap' from shallow aquifer (mm)	HRU	0–500
11	GWQMN	Threshold for return flow from shallow aquifer (mm)	HRU	0–5000
<i>Channel routing</i>				

No.	Parameter	Name	Spatial scale	Initial range
12	CH_N2	Manning's n for main channel	Sub-basin/Reach	0.01–0.3
13	CH_N1	Manning's n for tributary channel	Sub-basin/Reach	0.01–0.3
14	CH_K2	Effective hydraulic conductivity in main channel (mm/h)	Sub-basin/Reach	0–150
15	CH_K1	Effective hydraulic conductivity in tributaries (mm/h)	Sub-basin/Reach	0–150
<i>Snow accumulation</i>				
16	SFTMP	Snowfall temperature (deg C)	Basin	–1.5–1
17	SNOCVMX	Min snow water content at 100% snow cover (mm)	Basin	5–35
18	SMFMN	Melt factor (mm/deg C/day)	Basin	1.4–6.9
19	SMTMP	Snowmelt base temperature (deg C)	Basin	0–3
20	SNO50COV	Snow water content at 50% snow cover (mm)	Basin	0.05–0.35
21	TIMP	Snow pack temperature lag factor	Basin	0–1
<i>Biophysical parameters</i>				
22	BLAI	Maximum potential leaf area index	Biome-specific	0.5–10
23	DLAI	Fraction of growing season when leaf area declines	Biome-specific	0.15–1
24	LAIMX1	1st point of leaf area development curve	Biome-specific	0.001–1
25	LAIMX2	2nd point of leaf area development curve	Biome-specific	0.001–1
26	SOL_ALB	Soil albedo	HRU	0.001–0.25
27	CHTMX	Max canopy height (m)	Biome-specific	0.1–20
28	RDMX	Max root depth (m)	Biome-specific	0.001–3
29	FRGMX	Fraction of stomatal conductance	Biome-specific	0.001–1
30	VPDFR	Vapor pressure deficit (kPa)	Biome-specific	1.5–6
31	BIO_E	Radiation use efficiency ((kg/ha)/(MJ/m <sup>2</sup> ))	Biome-specific	10–90



**Fig. A1.** Comparison of simulated streamflow with USGS data at the watershed outlet during 2009–2013: \* the “estimated” portion of the available USGS data were not included in calibration (see Sections 3.4.1 and 3.5 for details). Simulated hydrograph in a calibration configuration (M1-M4) corresponds to respective most optimal parameter combination. Goodness measures of model performance are shown in Fig. 4c.

## References

- Abbaspour KC, 2015. SWAT-CUP 2012: SWAT calibration and uncertainty programs - a user manual. Available online at: < [http://swat.tamu.edu/media/114860/usermanual\\_swatcup.pdf](http://swat.tamu.edu/media/114860/usermanual_swatcup.pdf) > (last cited on March 21, 2017).
- Abbaspour KC, Rouholahnejad E, Vaghefi S, Srinivasan R, Yang H, Kløve B, 2015. A continental-scale hydrology and water quality model for Europe: calibration and uncertainty of a high-resolution large-scale SWAT model. *J. Hydrol* 524, 733–752. 10.1016/j.jhydrol.2015.03.027.
- Abbaspour KC, Yang J, Maximov I, Siber R, Bogner K, Mieleitner J, Zobrist J, Srinivasan R, 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *J. Hydrol* 333, 413–430. 10.1016/j.jhydrol.2006.09.014.
- Almendinger JE, Murphy MS, Ulrich JS, 2011. Use of the Soil and Water Assessment Tool to scale sediment delivery from field to watershed in an agricultural landscape with topographic depressions. *J. Environ. Qual* 43 (1), 9–17. 10.2134/jeq2011.0340.
- Arnold J, Moriasi D, Gassman P, Abbaspour K, White M, Srinivasan R, Santhi C, Harmel RD, Van Griensven A, 2012. SWAT: model use, calibration, and validation. *Trans. ASABE* 55, 1491–1508.
- Bai P, Liu X, Liu C, 2018. Improving hydrological simulations by incorporating GRACE data for model calibration. *J. Hydrol* 557, 291–304. 10.1016/j.jhydrol.2017.12.025.
- Beven K, 2012. *Rainfall-Runoff Modelling: The Primer*, second ed. John Wiley & Sons Ltd ISBN 978-0-470-71459-1.
- Beven K, Freer J, 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol* 249, 11–29. 10.1016/S0022-1694(01)00421-8.
- Birkel C, Soulsby C, Tetzlaff D, 2014. Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. *Water Resour. Res* 50, 3481–3501. 10.1002/2013WR014925.
- Brocca L, Moramarco T, Melone F, Wagner W, Hasenauer S, Hahn S, 2012. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. *IEEE Trans. Geosci. Rem. Sens* 50, 2542–2555. 10.1109/TGRS.2011.2177468.
- Chen F, Crow WT, Starks PJ, Moriasi DN, 2011. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Adv. Water Resour* 34, 526–536. 10.1016/j.advwatres.2011.01.011.
- Chokmani K, Ouara T, Hamilton S, Ghedira MH, Gingras H, 2008. Comparison of ice-affected streamflow estimates computed using artificial neural networks and multiple regression techniques. *J. Hydrol* 349, 383–396.
- Draper CS, Walker JP, Steinle PJ, de Jeu R, Holmes TRH, 2009. An evaluation of AMSR-E derived soil moisture over Australia. *Rem. Sens. Environ* 113, 703–710. <https://doi.org/10.1016/j.rse.2008.11.011>.
- Efstratiadis A, Koutsoyiannis D, 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrol. Sci. J* 55, 58–78.
- Emerton R, Stephens E, Pappenberger F, Pagano T, Weerts A, Wood A, et al. , 2016. Continental and global scale flood forecasting systems. *WIREs Water* 3 (3), 391–418. 10.1002/wat2.1137.
- EPA, 2011. Level III and IV ecoregions of the continental United States. U.S Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Corvallis, Oregon, Map scale 1:3,000,000. Available online at: < <https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states> > (last cited on February 18, 2018).
- EPA, 2007. *An Approach for Using Load Duration Curves in the Development of TMDLs*. U.S Environmental Protection Agency, Washington DC EPA 841-B-07-006.
- Ershadi A, McCabe M, Evans JP, Walker JP, 2013. Effects of spatial aggregation on the multi-scale estimation of evapotranspiration. *Rem. Sens. Environ* 131, 51–62.
- Evenson GR, Golden HE, Lane C, McLaughlin DL, D'Amico E, 2018. Depressional wetlands affect watershed hydrological, biogeochemical, and ecological functions. *Ecol. Appl* 10.1002/eap.1701.

- Evenson GR, Golden HE, Lane C, D'Amico E, 2016. An improved representation of geographically isolated wetlands in a watershed-scale hydrologic model. *Hydrol. Process* 30, 4168–4184. 10.1002/hyp.10930.
- Evenson GR, Golden HE, Lane C, D'Amico E, 2015. Geographically isolated wet-lands and watershed hydrology: a modified model analysis. *J. Hydrol* 529, 240–256. 10.1016/j.jhydrol.2015.07.039.
- Euliss NH, LaBaugh JW, Fredrickson LH, Mushet D, Laubhan MK, et al. , 2004. The wetland continuum: a conceptual framework for interpreting biological studies. *Wetlands* 24 (2), 448–458. 10.1672/0277-5212(2004)024[0448:TWACAF]2.0.CO;2.
- Fatichi S, Vivoni ER, Ogden FL, Ivanov VY, Mirus B, et al. , 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *J. Hydrol* 537, 45–60.
- Favis-Mortlock D, 2004. Self-organization and cellular automata models. In: Wainright Muiligan (Eds.), *Environmental Modelling: Finding Simplicity in Complexity*. John Wiley & Sons Ltd. ISBN 10-0-471-49618-9.
- Gassman PW, Reyes MR, Green CH, Arnold JG, 2007. The soil and water assessment tool: historical development, applications, and future research directions. *Trans. ASABE* 50, 1211–1250.
- Gleason RA, Laubhan MK, Euliss NH Jr. 2008. Ecosystem services derived from wetland conservation practices in the United States Prairie Pothole Region with an emphasis on the U.S. Department of Agriculture conservation reserve and wetlands reserve programs. U.S. Geological Professional Paper 1745. Available online at: < <https://pubs.usgs.gov/pp/1745/> > (last cited on March 9, 2018).
- Gochis DJ, Yu W, Yates DN, 2015. The WRF-Hydro model technical description and user's guide, version 3.0. NCAR Technical Document. 123 p. Available online at: < [https://ral.ucar.edu/projects/wrf\\_hydro/technical-description-user-guide](https://ral.ucar.edu/projects/wrf_hydro/technical-description-user-guide) > (last cited on February 9, 2018).
- Gupta HV, Kling H, Yilmaz KK, Martinez GF, 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol* 377, 80–91.
- Hamilton S, 2004. Winter streamflow as a source of uncertainty in water balance calculations. Proceedings of the workshop on Northern Research Basins Water Balance, Victoria, Canada. International Association of Hydrological Sciences (IAHS) publication no. 290, pp. 249–256.
- Han E, Merwade V, Heathman GC, 2012. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *J. Hydrol* 416–417, 98–117. 10.1016/j.jhydrol.2011.11.039.
- Herman M, Nejadhashemi A, Abouali M, Hernandez-Suarez J, Daneshvar F, et al. , 2018. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. *J. Hydrol* 556, 39–49.
- Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, et al. , 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogramm. Eng. Rem. Sens* 81 (5), 345–354.
- Horsburgh JS, Aufdenkampe AK, Mayorga E, Lehnert KA, Hsu L, et al. , 2016. Observations Data Model 2: a community information model for spatially discrete Earth observations. *Environ. Modell. Software* 79, 55–74.
- Hurkmans RT, de Moel H, Aerts JC, Troch PA, 2008. Water balance versus land surface model in the simulation of Rhine river discharges. *Water Resour. Res* 44, W01418. 10.1029/2007WR006168.
- Ikenberry CD, Crumpton W, Arnold JG, Soupier ML, Gassman PW, 2017. Evaluation of existing and modified wetland equations in the SWAT model. *J. Am. Water Resour. Assoc* 53 (6), 1267–1280. 10.1111/1752-1688.12570.
- Immerzeel W, Droogers P, 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. *J. Hydrol* 349, 411–424. 10.1016/j.jhydrol.2007.11.017.
- Islam AKMS, Paul S, Mohammed K, Billah M, Fahad MGR, Hasan A, et al. , 2017. Hydrological response to climate change of the Brahmaputra basin using CMIP5 general circulation model ensemble. *J. Water Clim. Change* 9 (3) 434–338.
- Kantrud HA, Krapu GL, Swanson GA 1989. Prairie basin wetlands of the Dakotas: a community profile. U.S. Fish and Wildlife Service Biological Report 85(7.28), 116. Available online at: < <https://www.nwrc.usgs.gov/techrpt/85-7-28.pdf> > (last cited on March 9, 2018).

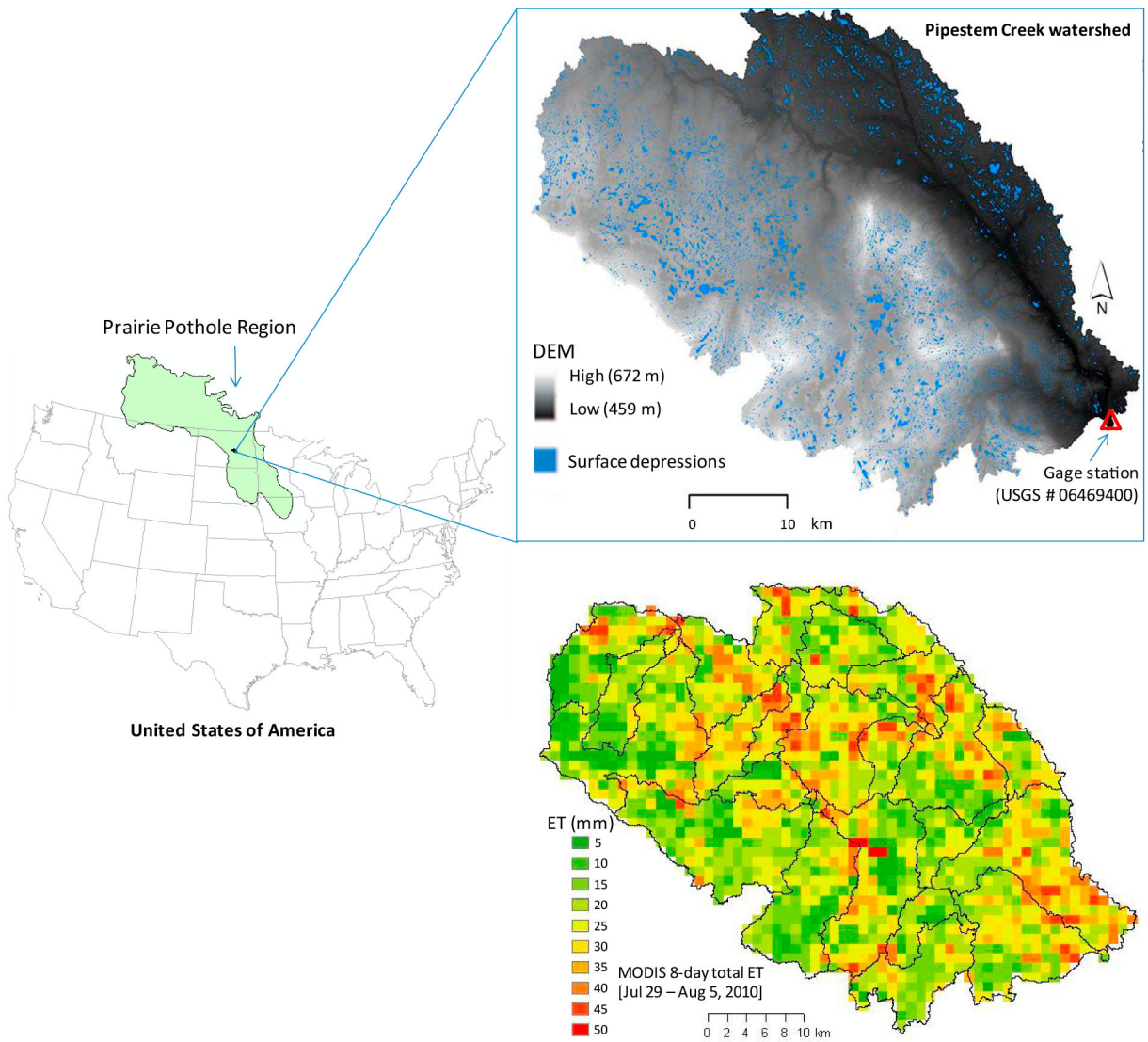
- Kling H, Fuchs M, Paulin M, 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol* 424–425, 264–277.
- Kunnath-Poovakka A, Ryu D, Renzullo LJ, George B, 2016. The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction. *J. Hydrol* 535, 509–524.
- LaBaugh JW, Winter TC, Rosenberry DO, 1998. Hydrologic functions of prairie wetlands. *Great Plains Res.* 8, 17–37.
- Lane C, Leibowitz S, Autrey B, LeDuc S, Alexander L, 2018. Hydrological, physical, and chemical functions and connectivity of non-floodplain wetlands to downstream waters: a review. *J. Am. Water Resour. Assoc* 10.1111/1752-1688.12633.
- Lane C, D'Amico E, Autrey B, 2012. Isolated wetlands of the southeastern united states: abundance and expected condition. *Wetlands* 32 (4), 753–767.
- Lane C, D'Amico E, 2010. Calculating the ecosystem service of water storage in isolated wetlands using LiDAR in North Central Florida, USA. *Wetlands* 30, 967–977.
- Leibowitz S, 2015. Geographically Isolated Wetlands: why we should keep the term. *Wetlands* 35 (5), 997–1003. 10.1007/s13157-015-0691-x.
- Li Y, Grimaldi S, Pauwels V, Walker J, 2018. Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: the impact on predictions at gauged and ungauged locations. *J. Hydrol* 557, 897–909.
- Li Y, Grimaldi S, Walker J, Pauwels V, 2016. Application of remote sensing data to constrain operational rainfall-driven flood forecasting: a review. *Rem. Sens* 8 (6), 456. 10.3390/rs8060456.
- Lin P, Rajib MA, Yang Z-L, Somos-Valenzuela M, Merwade V, Maidment D, et al. , 2018. Spatiotemporal evaluation of simulated evapotranspiration and streamflow over Texas using the WRF-Hydro-RAPID modeling framework. *J. Am. Water Resour. Assoc* 54 (1), 40–54. 10.1111/1752-1688.12585.
- Liu Y, Yang W, Wang X, 2008. Development of a SWAT extension module to simulate riparian wetland hydrologic processes at a watershed scale. *Hydrol. Process* 22, 2901–2915. 10.1002/hyp.6874.
- Long D, Longuevergne L, Scanlon B, 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res* 50, 1131–1151. 10.1002/2013WR014581.
- López López P, Sutanudjaja EH, Schellekens J, Sterk G, Bierkens MFP, 2017. Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrol. Earth Syst. Sci* 21, 3125–3144. 10.5194/hess-21-3125-2017.
- Maneta MP, Pasternack GB, Wallender WW, Jetten V, Schnabel S, 2007. Temporal instability of parameters in an event-based distributed hydrologic model applied to a small semiarid catchment. *J. Hydrol* 341, 207–221.
- Moriasi DN, Gitau MW, Pai N, Daggupati P, 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. *Trans. ASABE* 58 (6), 1763–1785.
- Morsy M, Goodall J, Castronova A, Dash P, Merwade V, 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. *Environ. Modell. Software* 93, 13–28. 10.1016/j.envsoft.2017.02.028.
- Milzow C, Krogh PE, Bauer-Gottwein P, 2011. Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment. *Hydrol. Earth Syst. Sci* 15, 1729–1743. 10.5194/hess-15-1729-2011.
- Mu Q, Zhao M, Running S, 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Rem. Sens. Environ* 115, 1781–1800.
- Mu Q, Zhao M, Running S, 2013. MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3): algorithm theoretical basis document collection 5 (last cited on February 18, 2018).
- Mushet D, Calhoun AJK, Alexander L, Cohen MJ, DeKeyser ES, et al. , 2015. Geographically Isolated Wetlands: rethinking a misnomer. *Wetlands* 35, 423–431. 10.1007/s13157-015-0631-9.

- Myneni R, Knyazikhin Y, Park T, 2015. MCD15A3H MODIS/Terra + Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [data set]. NASA EOSDIS Land Processes DAAC. DOI:10.5067/MODIS/MCD15A3H.006.
- NASA, 2017. README document for North American Land Data Assimilation System Phase 2 products, Goddard Earth Sciences Data and Information Services Center. Available online at: < <https://hydro1.gesdisc.eosdis.nasa.gov/data/NLDAS/README.NLDAS2.pdf> > (last cited on November 1, 2017).
- NASA, 2016. Surface Water and Ocean Topography Mission Project Science Requirements Document, Rev A, JPL D-61923. Available online at: < [https://swot.jpl.nasa.gov/docs/D-61923\\_SRD\\_Rev\\_A\\_20160318.pdf](https://swot.jpl.nasa.gov/docs/D-61923_SRD_Rev_A_20160318.pdf) > (last cited on March 8, 2018).
- Neitsch SL, Arnold J, Kiniry J, Williams J, 2011. Soil & Water Assessment Tool Theoretical Documentation Version 2009. Texas A&M University System, College Station, TX, USA.
- NRCS, 2015. Soil Survey Staff, Natural Resources Conservation Service. U.S. Department of Agriculture. Web Soil Survey Available online at < <https://websoilsurvey.nrcs.usda.gov/> >.
- Parajka J, Naeimi V, Blöschl G, Komma J, 2009. Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria. *Hydrol. Earth Syst. Sci* 13 (2), 259–271.
- Phillips RL, Beeri O, DeKeyser ES, 2005. Remote wetland assessment for Missouri Coteau prairie glacial basins. *Wetlands* 25 (2), 335–349.
- Quinton J, Govers G, Van Oost K, Bardgett RD, 2010. The impact of agricultural soil erosion on biogeochemical cycling. *Nat. Geosci* 3, 311–314. 10.1038/ngeo838.
- Rahman M, Thompson J, Flower R, 2016. An enhanced SWAT wetland module to quantify hydraulic interactions between riparian depressional wetlands, rivers and aquifers. *Environ. Modell. Software* 84, 263–289. 10.1016/j.envsoft.2016.07.003.
- Rajib MA, Merwade V, Yu Z, 2018a. Rationale and efficacy of assimilating remotely sensed potential evapotranspiration for reduced uncertainty of hydrologic models. *Water Resour. Res* 54. 10.1029/2017WR021147.
- Rajib A, Merwade V, Zhao L, Shin J, Smith J, Song C, 2018b. HydroGlobe Tool. Available online at: < <https://mygeohub.org/resources/hydroglobetool> >.
- Rajib A, Merwade V, 2017. Hydrologic response to future land use change in the Upper Mississippi River Basin by the end of 21st century. *Hydrol. Process* 31, 3645–3661. 10.1002/hyp.11282.
- Rajib MA, Merwade V, Yu Z, 2016a. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *J. Hydrol* 536, 192–207. 10.1016/j.jhydrol.2016.02.037.
- Rajib MA, Merwade V, Kim IL, Zhao L, Song C, Zhe S, 2016b. SWATShare - a web platform for collaborative research and education through online sharing, simulation and visualization of SWAT models. *Environ. Model. Softw* 75, 498–512. 10.1016/j.envsoft.2015.10.032.
- Rajib MA, Merwade V, 2016. Improving soil moisture accounting and streamflow prediction in SWAT by incorporating a modified time-dependent Curve Number method. *Hydrol. Process* 30, 603–624. 10.1002/hyp.10639.
- Reichle RH, Koster RD, 2004. Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett* 31. 10.1029/2004GL020938.
- Rientjes T, Muthuwatta L, Bos M, Booij M, Bhatti H, 2013. Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellitebased evapotranspiration. *J. Hydrol* 505, 276–290.
- Ruhoff AL, Paz AR, Aragao L, Mu Q, Malhi Y, et al. , 2013. Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin. *Hydrol. Sci. J* 58 (8), 1658–1676.
- Schuel J, Abbaspour KC, Srinivasan R, Yang H, 2008. Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model. *J. Hydrol* 352 (1), 30–49.
- Silvestro F, Gabellani S, Rudari R, Delogu F, Laiolo P, Boni G, 2015. Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrol. Earth Syst. Sci* 19 (4), 1727–1751.

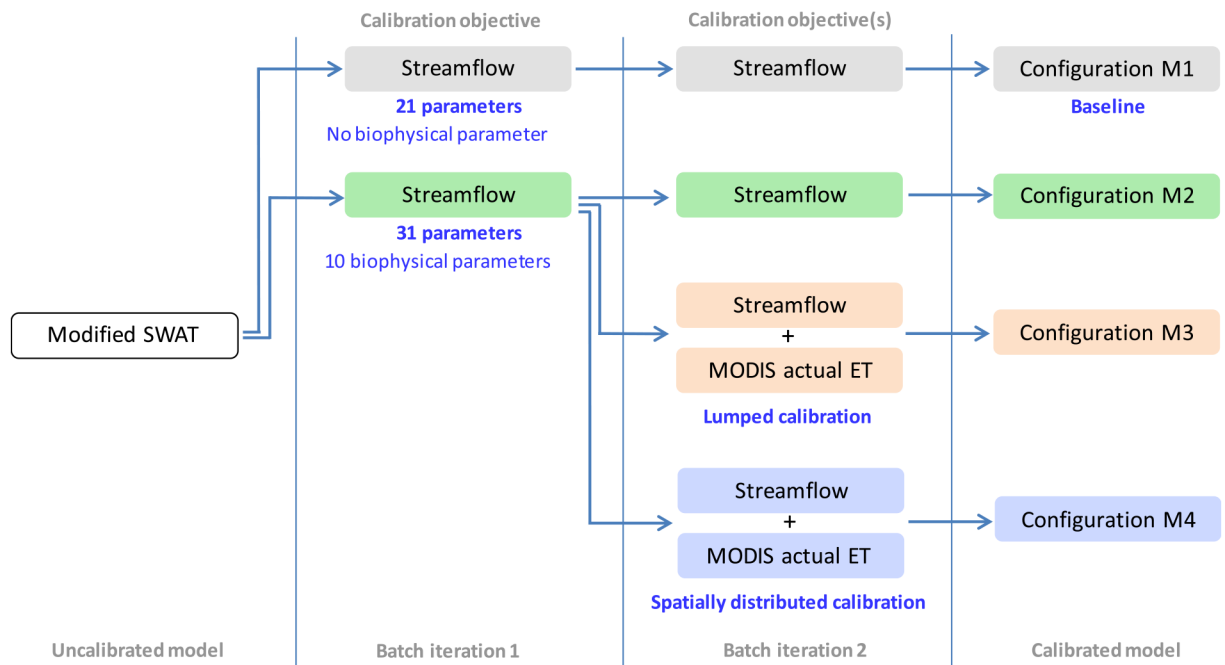


- Sivakumar B, Singh VP, 2012. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. *Hydrol. Earth Syst. Sci* 16, 4119–4131. 10.5194/hess-16-4119-2012.
- Sutanudjaja EH, Van Beek LPH, De Jong SM, Van Geer FC, Bierkens MFP, 2014. Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data. *Water Resour. Res* 50, 687–705. 10.1002/2013WR013807.
- Szilagyi J, Gribovszki Z, Kalicz P, 2007. Estimation of catchment-scale evapotranspiration from baseflow recession data: numerical model and practical application results. *J. Hydrol* 336, 206–217.
- Tarboton D, Idaszak R, Horsburgh JS, Heard J, Ames D, et al. , 2014. HydroShare: advancing collaboration through hydrologic data and model sharing. Proceedings of the 7th International Congress on Environmental Modelling and Software. Available online at: < <https://scholarsarchive.byu.edu/iemssconference/2014/Stream-A/7> >.
- Thornton PE, Thornton M, Mayer B, Wilhelmi N, Wei Y, et al., 2014. Daymet: daily surface weather data on a 1-km grid for North America, edited. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA.
- Tiner RW, 2003. Geographically Isolated Wetlands of the United States. *Wetlands* 23, 494–516. 10.1672/0277-5212(2003)023.
- Tobin K, Bennett M, 2017. Constraining SWAT calibration with remotely sensed evapotranspiration data. *J. Am. Water Resour. Assoc* 53 (3), 593–604. 10.1111/1752-1688.12516.
- Troch PA, Martinez GF, Pauwels VRN, Durcik M, Sivapalan M, Harman C, Brooks PD, Gupta H, Huxman T, 2009. Climate and vegetation water use efficiency at catchment scales. *Hydrol. Process* 23, 2409–2414. 10.1002/hyp.7358.
- USGS-NED, 2015. National Elevation Dataset: United States Geological Survey National Map Viewer. Available at: < <http://viewer.nationalmap.gov/viewer/> > (accessed 10 March, 2015)
- van der Kamp G, Hayashi M, Bedard-Haughn A, Pennock D, 2016. Prairie pothole wetlands - suggestions for practical and objective definitions and terminology. *Wetlands* 36 (Suppl 2), 229–235. 10.1007/s13157-016-0809-9.
- van der Valk AG, Mushet DM, 2016. Interannual water-level fluctuations and the vegetation of prairie potholes: potential impacts of climate change wetlands. *Wetlands* 36 (Suppl 2), 397–406. 10.1007/s13157-016-0850-8.
- van der Valk AG, 2005. Water-level fluctuations in North American prairie wetlands. *Hydrobiologia* 539, 171–188. 10.1007/s10750-004-4866-3.
- van der Valk AG, 1981. Succession in wetlands: a gleasonian approach. *Ecology* 62, 688–696. 10.2307/1937737.
- Velpuri NM, Senay GB, Singh RK, Bohms S, Verdin J, 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. *Rem. Sens. Environ* 139, 35–49.
- Vervoort RW, Mielchels SF, van Ogtrop FF, Guillaume J, 2014. Remotely sensed evapotranspiration to calibrate a lumped conceptual model: Pitfalls and opportunities. *J. Hydrol* 519, 3223–3236.
- Wanders N, Bierkens MFP, De Jong SM, de Roo A, Karssen D, 2014. The benefits of using remotely sensed soilmoisture in parameter identification of largescale hydrologicalmodels. *Water Resour. Res* 50, 6874–6891. 10.1002/2013WR014639.
- Wang X, Yang W, Melesse AM, 2008. Using hydrologic equivalent wetland concept within SWAT to estimate streamflow in watersheds with numerous wetlands. *Trans. ASABE* 51 (1), 55–72. 10.13031/2013.24227.
- Winsemius HC, Savenije HHG, Bastiaansen WGM, 2008. Constraining model parameters on remote sensed evaporation: justification for distribution in ungauged basins? *Hydrol. Earth Syst. Sci* 12, 1403–1413.
- Winter TC, Rosenberry DO, 1998. Hydrology of prairie pothole wetlands during drought and deluge: A 17-year study of the Cottonwood Lake wetland complex in North Dakota in the perspective of longer term measured and proxy hydrologic records. *Clim. Change* 40 (2), 189–209.
- Winter TC, Rosenberry DO, 1995. The interaction of ground water with prairie pot-hole wetlands in the Cottonwood Lake area, east-central North Dakota, 1979–1990. *Wetlands* 15 (3), 193–211.

- Wu Q, Lane CR, 2016. Delineation and quantification of wetland depressions in the Prairie Pothole Region of North Dakota. *Wetlands* 36, 215–227. 10.1007/s13157-015-0731-6.
- Yan K, Park T, Yan G, Chen C, Yang B, Liu Z, Nemani RR, Knyazikhin Y, Myneni RB, 2016. Evaluation of MODIS LAI/FPAR Product Collection 6. Part 1: Consistency and Improvements. *Rem. Sens* 8 (5), 359. 10.3390/rs8050359.
- Yang Q, Zhang X, 2016. Improving SWAT for simulating water and carbon fluxes of forest ecosystems. *Sci. Total Environ* 569–570, 1478–1488.
- Yang W, Tan B, Huang D, Rautiainen M, Shabanov NV, et al. , 2006. MODIS leaf area index products: from validation to algorithm improvement. *IEEE Trans. Geosci. Rem. Sens* 44, 1885–1898.
- Zhang K, Kimball JS, Running SW, 2016. A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdiscipl. Rev.: Water* 3 (6), 834–853. 10.1002/wat2.1168.
- Zhang K, Kimball JS, Ramakrishna N, Running SW, 2010. A continuous satellitederived global record of land surface evapotranspiration from 1983 to 2006. *Water Resour. Res* 46. 10.1029/2009WR008800.
- Zhu Q, Zhuang Q, 2015. Ecosystem biogeochemistry model parameterization: do more flux data result in a better model in predicting carbon flux? *Ecosphere* 6 (12), 283. 10.1890/ES15-00259.1.



**Fig. 1.** The Pipestem Creek watershed in Pingree, North Dakota, United States. The topographic map on top shows ~25,000 potential surface depressions; however, more than 50% of the watershed constitutes cropland and rangeland (Wu and Lane, 2016). The map at the bottom shows 29 sub-basins overlaid on 1 km MODIS grid; the example MODIS data represents total ET (mm) in an 8-day period (July 29 to August 5, 2010).



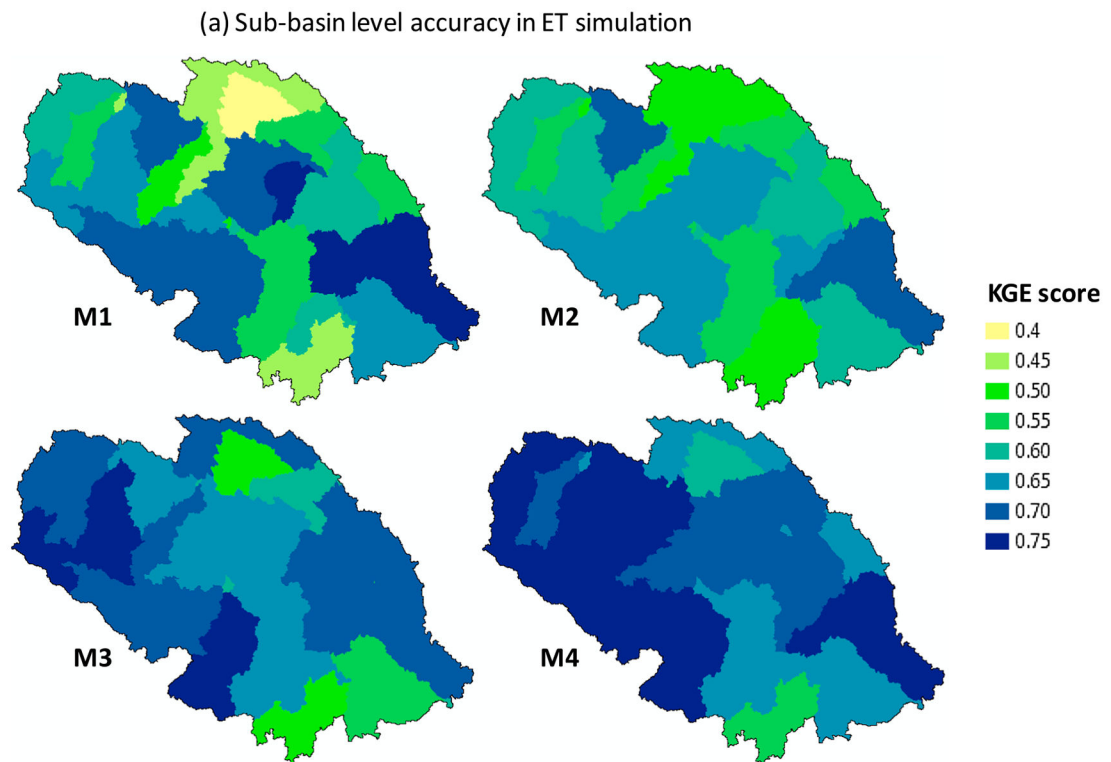
**Fig. 2.**

A schematic representation of calibration architecture and associated workflow involved in the four configurations. We reproduced the modified SWAT model developed by Evenson et al. (2016). Iteration 1 involved a simplistic sensitivity analysis to identify how inclusion of biophysical parameters might have affected the order of sensitivity (see Fig. 3). The most optimal model obtained through Iteration 1 became the starting point for Iteration 2. Calibration of snow parameters were discontinued after Iteration 1 in all cases (see Section 3.5.4 for details). Each batch of iterations (1 and 2) contains 250 simulations, however, the number is subjective.

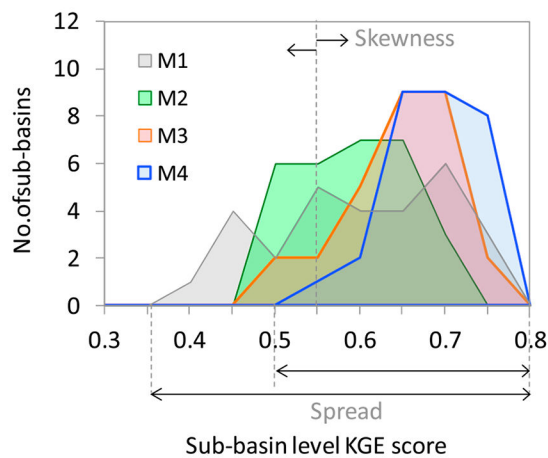
Surface/sub-surface runoff			Order of parameter sensitivity	
Parameter	Order	Description	M1	M2
GIW_K	1	Hydraulic conductivity of GIW subsurface (mm/h)	20	22
GW_DELAY	2	Groundwater delay (days)	10	25
GW_REVAP	3	Groundwater 'revap' coefficient	12	14
RCHRG_DP	4	Deep aquifer percolation fraction	14	28
CN2	5	Curve Number	21	6
ALPHA_BF	6	Baseflow recession constant	2	21
SOL_AWC	7	Soil available water capacity (mm/mm)	7	20
EPCO	8	Plant uptake compensation factor	18	26
ESCO	9	Soil evaporation compensation factor	3	31
REVAPMN	10	Threshold for 'revap' from shallow aquifer (mm)	13	29
GWQMN	11	Threshold for return flow from shallow aquifer (mm)	8	10
<b>Channel routing</b>			17	16
CH_N2	12	Manning's n for main channel	5	7
CH_N1	13	Manning's n for tributary channel	19	5
CH_K2	14	Effective hydraulic conductivity in main channel (mm/h)	4	23
CH_K1	15	Effective hydraulic conductivity in tributaries (mm/h)	1	30
<b>Snow accumulation **</b>			6	4
SFTMP	16	Snowfall temperature (deg C)	11	8
SNOCOVMX	17	Min snow water content at 100% snow cover (mm)	9	2
SMFMN	18	Melt factor (mm /deg C /day)	16	1
SMTMP	19	Snowmelt base temperature (deg C)	15	27
SNO50COV	20	Snow water content at 50% snow cover (mm)		9
TIMP	21	Snow pack temperature lag factor		3
<b>Biophysical parameters ***</b>				18
BLAI	22	Max potential leaf area index		13
DLAI	23	Fraction of growing season when leaf area declines		12
LAIMX1	24	1st point of leaf area development curve		15
LAIMX2	25	2nd point of leaf area development curve		17
SOL_ALB	26	Soil albedo		24
CHTMX	27	Max canopy height (m)		19
RDMX	28	Max root depth (m)		11
FRGMX	29	Fraction of stomatal conductance		
VPDFR	30	Vapor pressure deficit (kPa)		
BIO_E	31	Radiation use efficiency ((kg/ha)/(MJ/m <sup>2</sup> ))		

Higher sensitivity

**Fig. 3.** Calibration parameters and their order of sensitivity with and without the involvement of biophysical parameters. Respective initial ranges and the spatial scale of iteration are shown in Table A1 (Appendix A). See Neitsch et al. (2011) for parameter definitions, except GIW\_K; GIW\_K is a new SWAT parameter for surface depressions (Evenson et al., 2016). \*\* Calibration was discontinued after the first batch of iterations; \*\*\* these parameters were not calibrated in M1[Basic].



(b) Spatial distribution of KGE for ET simulation

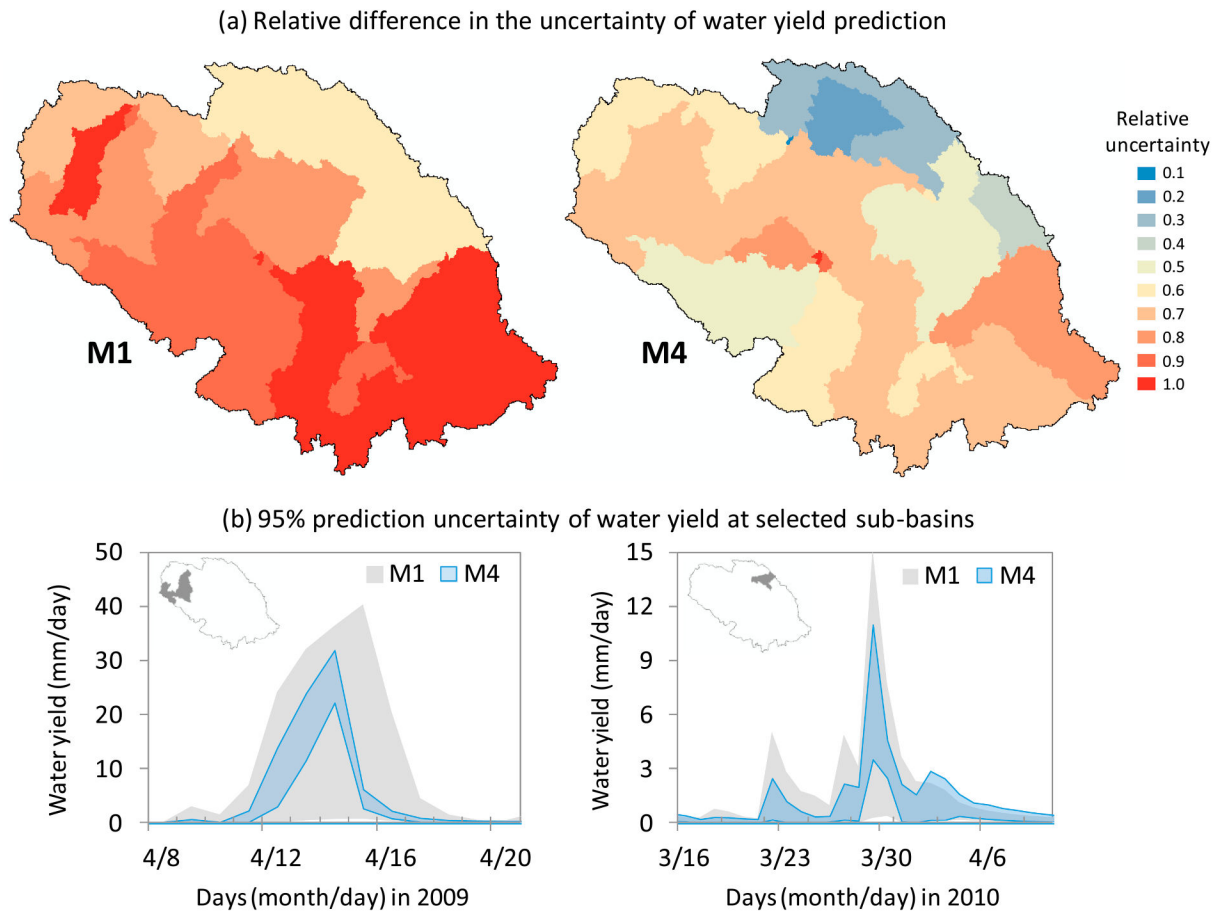


(c) KGE for streamflow simulation

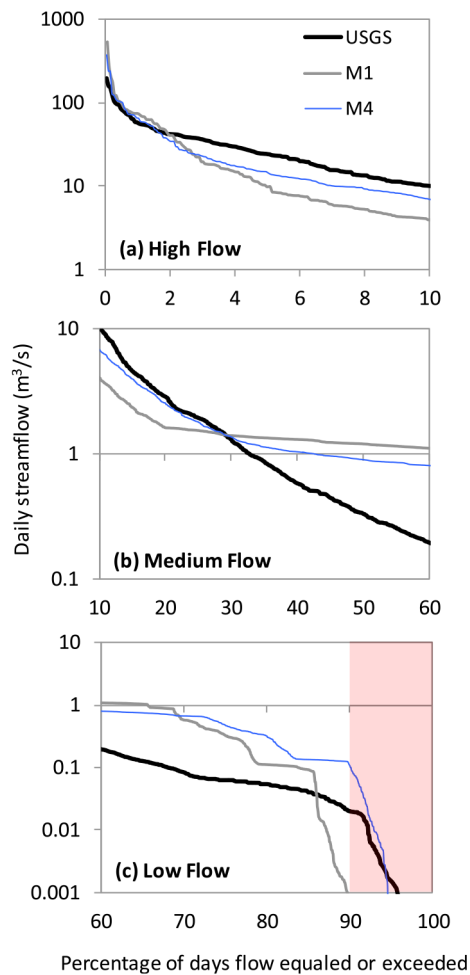
Configuration	Calibration* (2009-2011)	Verification** (2009-2013)
M1	0.91	0.50
M2	0.91	0.63
M3	0.82	0.70
M4	0.90	0.75

**Fig. 4.**

Goodness of model performance based on the most optimal parameter combination. KGE values shown here measure the agreement of the model with respect to MODIS ET at sub-basin level and USGS streamflow at the watershed outlet. All assessments are for daily simulation. While ET simulation performance shown here correspond only to the calibration period (2009–2011; Fig. 4a–b), streamflow simulation performance was reported both for calibration and verification (2009–2013; Fig. 4c): \* using only the “observed” portion of the available USGS data; \*\* using both “observed” and “estimated” data (see Fig. A1 in appendix A for streamflow hydrographs).

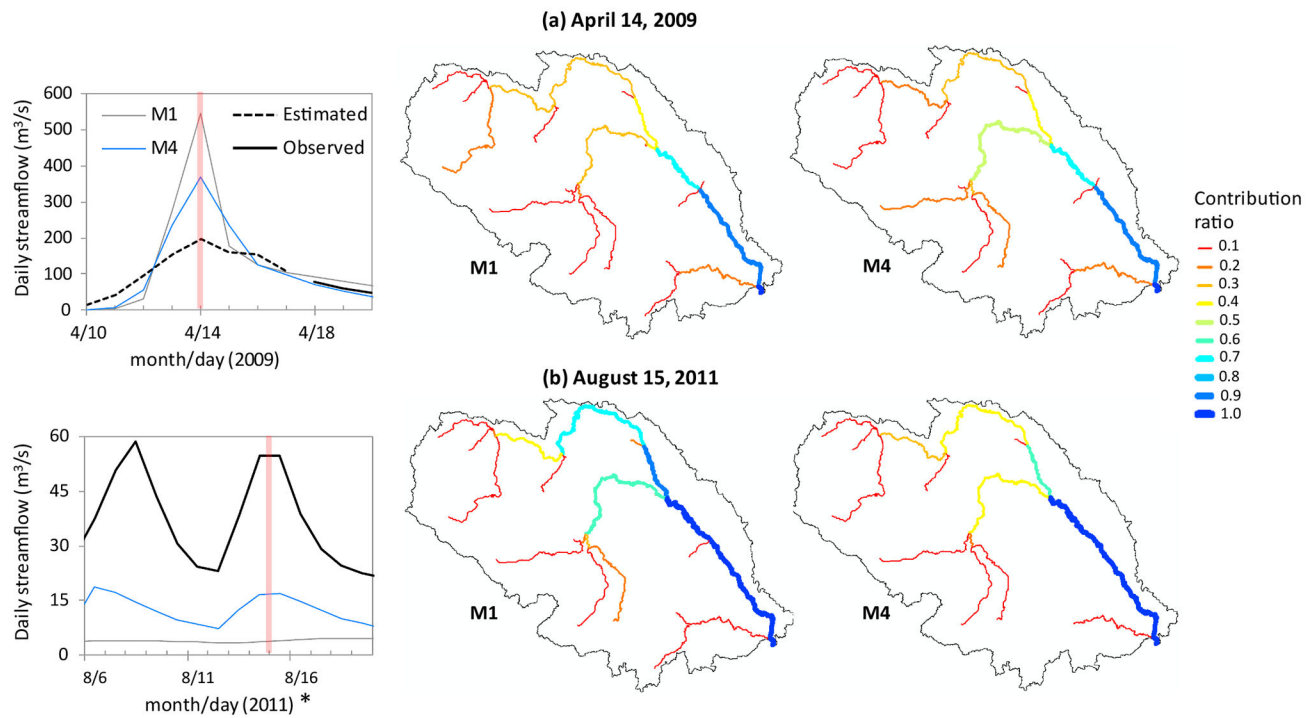


**Fig. 5.** Sub-basin level assessment of water yield prediction uncertainty. The dimensionless relative uncertainty values mapped in (a) indicate average band width of 95% prediction uncertainty over a 3-year calibration period, normalized across space and irrespective of the configuration. Because of this normalization, these band width values measure the “relative” difference in prediction uncertainty between M1 and M4 (1 meaning the highest uncertainty). The band hydrographs in (b) compare 95% prediction uncertainty of water yield (mm/day) during two different peak flow events and two different physiographic regions within the watershed (high wetland density (left) and low wetland density (right)). We chose two different events to evaluate whether the reduced uncertainty in M4 was consistent regardless of the antecedent wetness condition and the peak flow magnitude.

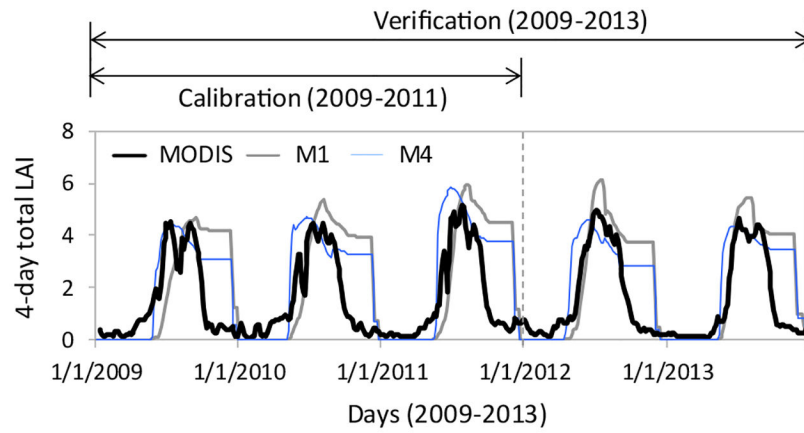


**Fig. 6.** Model verification with USGS observed/estimated data for different exceedance probability of simulated streamflow at the watershed outlet during 2009–2013. Five different flow conditions (high, moist, mid-range, dry, and low) as defined in EPA (2007) were aggregated into three categories here: (a) high, (b) medium, and (c) low. Y-axis is in  $\log_{10}$  scale. We compare the baseline model (M1) with our final model (M4) to convey the maximum contrast. Across all flow conditions, M4 simulations follow USGS data closer than M1, with exception of flows that are exceeded 20% of the time in the low flow condition (i.e., between 70 and 90% exceedance range; Fig. 6c). The highlighted portion in (c) indicates the prevalent watershed condition (~90% of the time during the 5-year simulation).





**Fig. 7.** Relative contribution of tributaries and the main stem to the streamflow at the watershed outlet, comparing the M1 and M4 configurations. At a particular date, the so-called “Contribution Ratio” was calculated as the ratio of streamflow in a given stream segment to the streamflow at the outlet (e.g., 0.3 for a tributary indicates that 30% of the streamflow simulated at the watershed outlet is contributed by the particular tributary). To show consistently different landscape response in M4, we chose two different peak flow events: (a) a snow melt event on April 14, 2009 and (b) a rainfall event on August 15, 2011. In the respective events, peak flows in M1 and M4 configurations occurred at the same date (highlighted). In (b), \* indicates a recorded flood event. USGS observed data from this particular date range was excluded from model calibration, but compared here with model simulations for cross-verification of the models’ goodness in responding to extreme events.



**Fig. 8.** Comparison of simulated watershed-average LAI with corresponding MODIS 4-day total estimates. This is a cross-verification of the model because MODIS LAI was not part of our calibrations. LAI in M4 simulation follows MODIS data closer than M1.