On the evaluation of both spatial and temporal performance of distributed hydrological models

Tam Van Nguyen1, Dang An Tran2, Bhumika Uniyal3, Thi Bich Thuc Phan4

1Department of Hydrogeology, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany.

2Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Viet Nam.

3Professorship of Ecological Services, Bayreuth Center of Ecology and Environmental Research, Bayreuth University, 95447 Bayreuth, Germany.

4Institute of Applied Mechanics and Informatics-Vietnam Academy of Science and Technology, 291 Dien Bien Phu Street, Ward 7, District 3, Ho Chi Minh City, Vietnam.

ABSTRACT

Evaluating the spatial and temporal model performance of distributed hydrological models is necessary to ensure that the simulated spatial and temporal patterns are meaningful. In recent years, spatial and temporal remote sensing data have been increasingly used for model performance evaluation. Previous studies, however, have focused on either the temporal or spatial model performance evaluation. In addition, temporal (or spatial) model performance evaluation is often done in a spatially (or temporally) lumped approach. Here, we evaluated (1) the temporal model performance evaluation in a spatially distributed approach (spatiotemporal) and (2) the spatial model performance in a temporally distributed approach (temporospatial) model performance evaluation. This study demonstrated that both spatiotemporal and temporospatial model performance evaluations are necessary since they provide different aspects of the model performance. For example, spatiotemporal model performance evaluation helps in detecting the areas with an issue in the simulated temporal patterns. However, temporospatial model performance evaluation helps in detecting the time with an issue in the simulated spatial patterns. The results also show that an increase in the spatiotemporal model performance will not necessarily lead to an increase in the temporospatial model performance and vice versa, depending on the evaluation statistics. Overall, this study has highlighted the necessity of a joint spatiotemporal and temporospatial model performance evaluation to understand/improve spatial and temporal model behavior/performance.

***Keywords:*** *distributed hydrological models, model performance evaluation, spatiotemporal; temporospatial, remote-sensing data*

# 1. Introduction

Spatial heterogeneity in catchment characteristics (land use, soil type, and slope) and spatial and temporal variations of hydrological forcing (rainfall, wind speed, and solar radiation) are often observed in nature (Klingler et al., 2021). These factors cause spatial and temporal variations of hydrological processes/responses, e.g, soil moisture (Brocca et al., 2012; Wilson et al., 2004), evapotranspiration (Thomas, 2000), and groundwater recharge (Edmunds et al., 2002). Spatially and temporally distributed hydrological models (herein referred to as distributed models) have been proven as effective tools for modeling spatial and temporal hydrological processes, e.g., the Soil and Water Assessment Tool (SWAT; Gassman et al., 2007; Neitsch et al., 2011), the TOPography based hydrological MODEL (TOPMODEL; Beven & Kirkby, 1979), and the Variable Infitration Capacity (VIC; Liang et al., 1994) model, the mesoscale Hydrologic Model (mHM; Kumar et al., 2013; Samaniego et al., 2010).

To ensure that these models provide meaningful spatial and temporal patterns of the interested processes, these spatial and temporal patterns should be evaluated against the reference spatiotemporal data. In recent years, remote-sensing data of high spatial and temporal resolution have been increasingly used for model evaluation, especially in data-scarce regions (Campo et al., 2006; Dembélé et al., 2020; Herman et al., 2018; Immerzeel & Droogers, 2008; Jiang et al., 2020; Koch et al., 2018; Mendiguren et al., 2017; Nguyen et al., 2020; Odusanya et al., 2019; Rajib et al., 2018; Rientjes et al., 2013; Stisen et al., 2018). However, evaluating both spatial and temporal model performance of distributed models has not been properly addressed. The aforementioned studies have focused on evaluating either the temporal or spatial model performance. In addition, while temporal model performance is evaluated, the evaluation is often done a spatially lumped approach, e.g., at the basin or subbasin level (Nguyen et al., 2020; Rajib et al., 2018; Rientjes et al., 2013). Similarly, while the spatial model performance is evaluated, the evaluation is often done in a temporally lumped approach, e.g., at monthly, yearly, or the entire simulation time (e.g., Koch et al., 2018). Up to now, no study has evaluated (1) the temporal model performance in a spatially distributed approach (hereinafter referred to as the spatiotemporal model performance evaluation) and (2) spatial model performance in both temporally distributed approaches (hereinafter referred to as the temporospatial model performance evaluation).

Evaluating the spatiotemporal and temporospatial model performance are the two different aspects of the model evaluation. Spatiotemporal model performance evaluation quantifies the temporal pattern matching between the simulated and reference time-series data at each spatial unit (e.g., pixel) of the model domain. This can be done by using different temporal performance indices, e.g., the Nash-Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970), the Kling-Gupta efficiency (KGE; Gupta et al., 2009), percentage bias (PBIAS), the root mean square error (RMSE), and the ratio of the RMSE to the standard deviation of measured data (RSR). Spatiotemporal model performance evaluation, therefore, provides information about the locations where the temporal model performance is not good (spatial model issue).

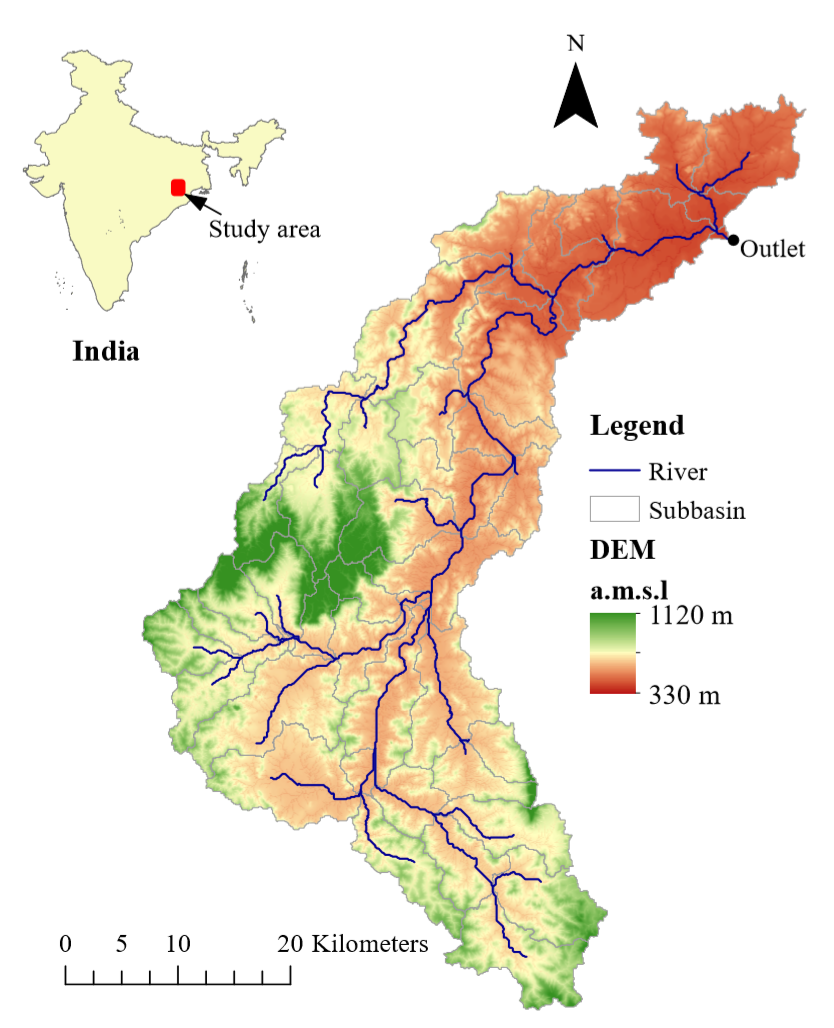
In contrast, temporospatial model performance evaluation quantifies the matching between the simulated and reference spatial patterns at each model time step. This can be done by using different spatial performance indices. Some of the aforementioned temporal performance metrics could be used (e.g., RSR, BIAS). Some performance metrics were specially developed for evaluating the spatial pattern matching, e.g., the SPAtial EFficiency metric (SPAEF; Koch et al., 2018), fractions skill score (Roberts & Lean, 2008), and others (Koch et al., 2015). Among these spatial performance matrices, the SPAEF metric, which consists of three equally weighted components (correlation, coefficient of variation, and histogram overlap) was demonstrated as a robust statistical index for evaluating the spatial pattern matching (Koch et al., 2018). Temporospatial model performance evaluation provides information about the timing when the spatial model performance is not good (temporal model issue).

From the above discussion, it can be said that spatiotemporal and temporospatial model evaluation (1) should be used together to detect both spatial and temporal model issues, and (2) are the two different aspects of model evaluation. This statement, however, has not been validated. The main objective of this study is to validate the two aforementioned arguments. For these objectives, the Soil and Water Assessment Tool (SWAT; Neitsch et al., 2011) was set up for the Upper Baitarani catchment located in a data-scare region in India. Satellite-derived ETa data from multiple sources were used for model evaluation.

# 2. Methodology

## 2.1. Study area and data

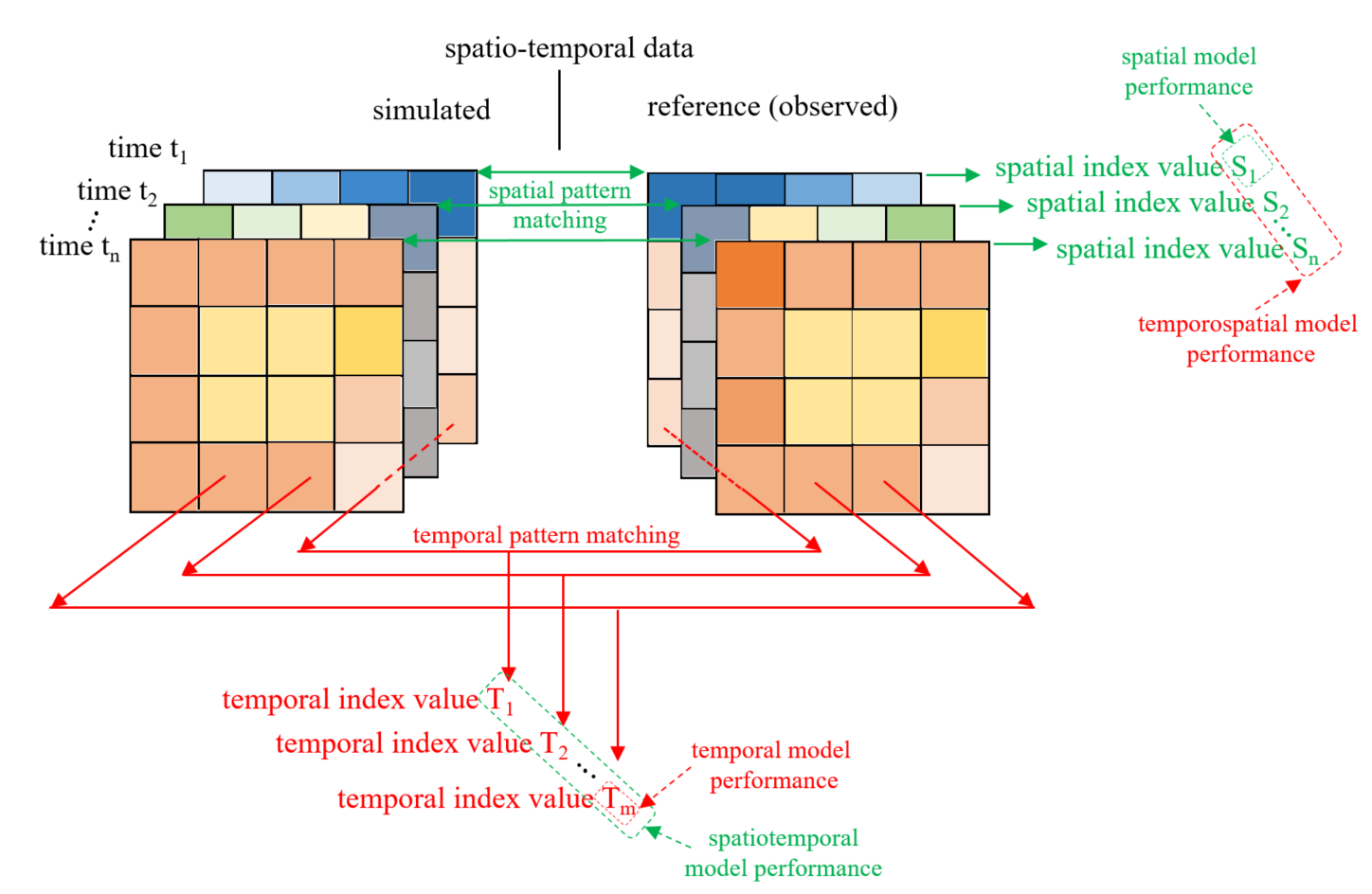
The Upper Baitarani catchment is located in India with an area of about 1711 km2 (Figure 1). The digital elevation model (DEM) from the Shuttle Radar Topography Missionshows that elevation of the catchment varies in a wide range, from 330 to 1120 m above mean sea level (a.m.s.l). Despite the area of the study area covers a relatively large area with high variation in topography, there is only one weather station located in the area. Observed rainfall shows that the average annual rainfall of 1165 mm with high temporal variation, more than 80% of the rainfall occurs between June and October (Uniyal et al., 2019). The average monthly minimum and maximum temperature in the area are 11℃ and 34℃, respectively. The dominant land uses/land covers in the area are agricultural, forest, and range grass, accounting for about 36%, 32%, and 25% of the study area, respectively. The main soil type in the area is sandy clay loam with low water holding capacity (Uniyal et al., 2019).



**Figure 1**. Location of the Upper Baitarani catchment and the Digital Elevation Model (DEM).

## 2.2. Spatiotemporal and temporospatial model performance evaluation

Evaluation of the spatial and temporal performance is the evaluation of the simulated spatiotemporal data against the reference (or observed) spatiotemporal data. Here, we consider the models with output data that need to be evaluated are time-series data at every pixel (with an equal area) of the modeling domain (Figure 2).



**Figure 2**. Approaches for evaluating spatiotemporal and temporospatial model performance.

In the spatiotemporal approach (Figure 2), the simulated time-series data at every pixel is first evaluated against the reference time-series data at the corresponding pixel using a temporal performance index T (e.g., NSE, KGE, PBIAS, RMSE). The spatial variation of the temporal index T provides information on the spatiotemporal model performance. In the temporospatial approach (Figure 2), the simulated spatial patterns are evaluated against the reference spatial patterns at every time step using the spatial index S (e.g., SPAEF, RMSE). Information about the temporospatial model performance could be obtained from the temporal variation of the spatial index S.

## 2.3. Reference spatiotemporal ETa

In this study, global satellite-based ETa products from the Moderate Resolution Imaging Spectroradiometer (MOD16 A2; Mu et al., 2013), the operational Simplified Surface Energy Balance model (SSEBop; Senay et al., 2013), and the TerraClimate were used. ETa from these products was estimated based on different techniques and are available at different spatial and temporal resolutions. For example, MOD16 A2 ETa was derived using the Penman*–*Monteith approach with daily meteorological reanalysis data (air pressure, temperature, humidity, solar radiation) and 8-day remote-sensing vegetation indices (leaf area index, albedo, and fraction of photosynthetically active radiation). MOD16 A2 ETa is available at 0.5-km spatial resolution and 8-day interval. SSEBop ETa was calculated with the simplified energy balance model (Senay et al., 2013). This model combines the estimated ET fraction based on MODIS thermal imagery and grass-reference potential ET. SSEBop ETa is available at 1-km resolution and monthly timestep. TerraClimate ETa was estimated based on the one-dimensional modified Thornthwaite-Mather climatic water-balance model with (1) monthly climate data from the WorldClim (Fick & Hijmans, 2017; Hijmans et al., 2005), the Climate Research Unit (Harris et al., 2014), and the Japanese 55-year Reanalysis (Kobayashi et al., 2015), and (2) water storage capacity from the root zone storage capacity (Wang-Erlandsson et al., 2016). Global TerraClimate ETa is available at 4-km spatial resolution and monthly time step.

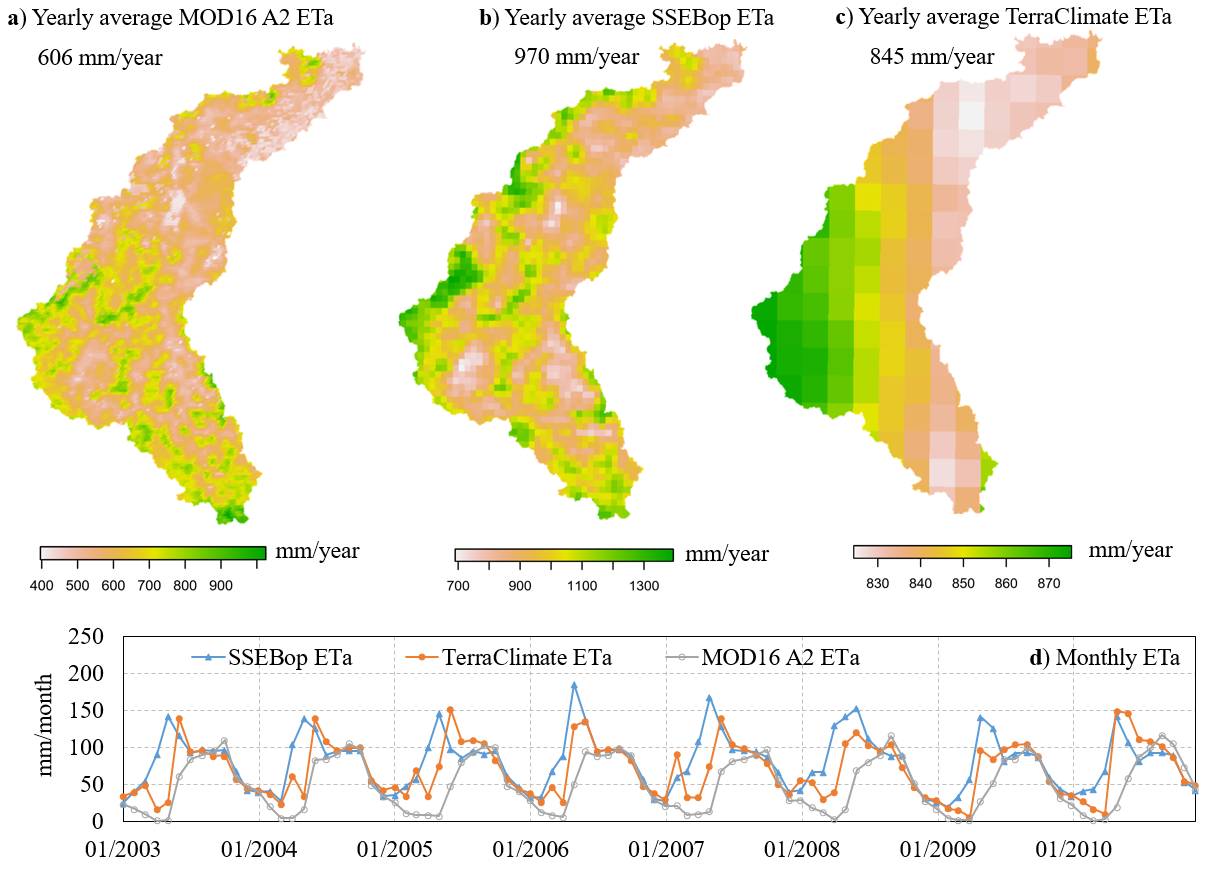


Figure 3: (a,b,c) Spatial and (d) temporal variations of ETa from MOD16 A2, SSEBop, and TerraClimate in the study area during the period from 2003 to 2010.

Due to different techniques as well as data used for the estimation of ETa as aforementioned, differences in the absolute ETa between these products are expected. The annual average ETa during the period 2003-2010 in the study area is 606 mm (MOD16 A2), 970 mm (SSEBop), and 845 mm (TerraClimate). The spatial and monthly ETa patterns also show their dissimilarities (Figure 3). However, there are some common spatial and temporal ETa patterns among these products, e.g., low ETa near the catchment outlet, high ETa in the western part of the study area, and high seasonal variation. The accuracy of these products in the study area is not verified due to the lack of observed data. In this study, all of these ETa products were used for model evaluation to (1) explore the model performance compared with different ETa products, and (2) account for the uncertainty in the estimated ETa from different ETa products.

## 2.4. The SWAT model

SWAT is a distributed hydrological used for evaluating the impacts of land use management practices on water, sediment, and nutrients yields (Neitsch et al., 2011). SWAT has been widely used and tested in various catchments worldwide (Arnold & Fohrer, 2005). In SWAT, a catchment is divided into subcatchments which are further divided into hydrologic response units (HRUs). HRU is a fraction of land with a unique combination of land use, soil type, and slope within a subcatchment. SWAT simulates two phases of the hydrological cycle, the land phase, and the routing phase. The land phase includes HRU-related processes (e.g., evapotranspiration, surface runoff, infiltration, lateral flow, groundwater recharge, and based flow). The routing phase simulates flow in water bodies (e.g., river, reservoir, and pond). SWAT provides outputs at different spatial levels, e.g., HRU, subcatchment, or catchment. Outputs at the HRU levels could be mapped to the HRU raster file, which contains spatial information about HRU, created during model setup (e.g., Nguyen & Dietrich, 2018).

## 2.5 Model setup, parameter variation, and model evaluation for ETa

### **2.5.1. Model setup**

Based on land use, soil type, and slope, the study area is divided into 45 catchments, which are further divided into 875 hydrologic response units (HRUs). The model was set up to simulate for ten years with two years of warm-up (2001-2002) and eight years of model evaluation (2003-2010). Daily simulated ETa is mapped to the HRU raster map of 90 m resolution that was created during the model setup with ArcSWAT (e.g., Kim et al., 2008; Nguyen & Dietrich, 2018).

### **2.5.2. Parameter variation and model performance evaluation scenarios**

The objectives of parameter variation are to 1) search for a global optimal solution, 2) find a relationship between spatiotemporal and temporospatial model performance in different simulations. The parameters selected for variation (Table 1) were based on a literature review of the most common parameters for ETa calibration (Abiodun et al., 2018; Kannan et al., 2007; Nguyen et al., 2020; Odusanya et al., 2019; Rajib et al., 2018). In this study, some of the parameters were changed in a spatial approach according to land use and soil type (Table 1). Latin Hypercube Sampling (LHS) was used to generate 2000 parameter sets in a search for a global optimum solution. LHS has been demonstrated as an efficient sampling approach for searching global optimal solution for high dimensional problems (e.g., Nguyen et al., 2021).

**Table 1**. List of model parameters and their ranges for parameter optimization. The prefixes “v” and “r” indicate “replace” and “relative” change compared to the original values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nr. | Parameters | Description | Parameter Range | |
|  |  |  | Min | Max |
| 1 | r\_CN2FRSD | SCS runoff curve number of forest (FRSD), agriculture (AGRL), and range grass (RNGE) lands | -0.2 | 0.2 |
| 2 | r\_CN2AGRL | -0.2 | 0.2 |
| 3 | r\_CN2RNGE | -0.2 | 0.2 |
| 4 | v\_ESCOFRSD | Soil evaporation compensation factor of forest agriculture, and range grass lands | 0 | 1 |
| 5 | v\_ESCOAGRL | 0 | 1 |
| 6 | v\_ESCORNGE | 0 | 1 |
| 7 | v\_EPCOFRSD | Plant uptake compensation factor of forest agriculture, and range grass lands | 0 | 1 |
| 8 | v\_EPCOAGRL | 0 | 1 |
| 9 | v\_EPCORNGE | 0 | 1 |
| 10 | v\_GWQMN | Groundwater baseflow thereshold (mm) | 0 | 2000 |
| 11 | v\_GW\_REVAP | Groundwater "revap" coefficient | 0.02 | 0.2 |
| 12 | v\_REVAPMN | Groundwater "revap" threshold (mm) | 0 | 500 |
| 13 | r\_SOL\_AWCSOIL1 | Soil available water content of soil classes 1 and 2 | -0.2 | 0.2 |
| 14 | r\_SOL\_AWCSOIL2 | -0.2 | 0.2 |
| 15 | r\_SOL\_KSOIL1 | Soil hydraulic conductivity of soil classes 1 and 2 (mm/h) | -0.2 | 0.2 |
| 16 | r\_SOL\_KSOIL2 | -0.2 | 0.2 |
| 17 | v\_CANMX | Maximum canopy storage (mm) | 0 | 5 |

In this study, the optimal solution was used as an example for demonstrating (1) how spatiotemporal and temporospatial model performance evaluation could be done, and (2) the benefits of both spatiotemporal and temporospatial model performance. The following multi-objective function (OF) was used for selecting the optimal solution:

(1)

(2)

(3)

where *TS* and *ST* are the spatiotemporal and temporospatial model performance indices, respectively, and are the statistical indices for temporal and spatial model performance, respectively, for actual evapotranspiration (ETa) at the pixel *j* and time step *k*, *npixels* is the number of pixels, and k is the number of evaluation time steps. For this analysis, we select MOD16 A2 ETa product with spatial index *S* is -*SPAEF* (Equation 5) and the temporal index *T* is –*NSE* (Equation 6) as an example.

To evaluate the relation between spatiotemporal and temporospatial model performance, we used different spatiotemporal (*ST*) and temporospatial (*TS*) performance indices (Table 2). For this evaluation, we used different ETa products (MOD16 A2, SEEBop, and TerraClimate) to draw a more reliable conclusion. The relation between spatiotemporal and temporospatial model performance was analyzed base on 2000 model runs.

Table 2: List of spatial and temporal model performance statistics used in this study. Numbers in bold indicate the ideal values.

|  |  |  |  |
| --- | --- | --- | --- |
| Spatiotemporal statistic ST (Equation 2) | | Temporospatial statistic TS (Equation 3) | |
| Notation | Range and ideal value | Notation | Range and ideal value |
| *S-NSE* | [-**1**, ∞) | *T-SPAEF* | [-**1**, ∞) |
| *S-NSE* | [-**1**, ∞) | *T-NSE* | [-**1**, ∞) |
| *SRMSE* | [**0**, ∞) | *TRMSE* | [**0**, ∞) |
| *SRSR* | [**0**, ∞) | *TRSR* | [**0**, ∞) |
| *SaBIAS* | [**0**, ∞) | *TaBIAS* | [**0**, ∞) |

The spatiotemporal (*ST*) and temporospatial (*TS*) performance statistics (Table 2) are calculated using equations (2) and (3), respectively, while the temporal *T* (e.g., *SPAEF, NSE, RMSE, RSE, aBIAS*) and spatial *S* (e.g., *KGE, NSE, RMSE, RSR, aBIAS*) performance are calculated as follows:

(4)

(5)

(6)

(7)

(8)

(9)

with:

(10)

(11)

(12)

(13)

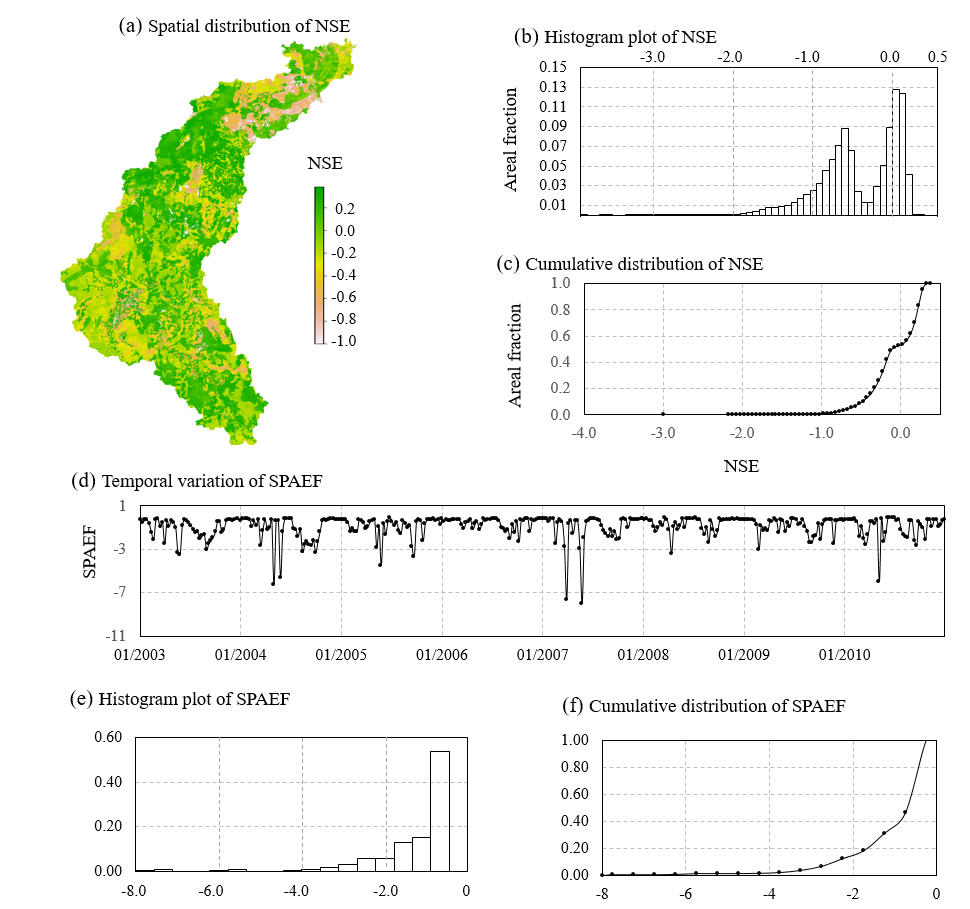
where and are the simulated and observed values (in this study is the simulated ETa and from satellite-derived ETa), is the mean of observed , and L are the histograms of observed and simulated variables, m is the number of bins (Koch et al., 2018). is the absolute sign.

# 3. Results and discussion

## 3.1. Spatiotemporal and temporospatial model performance

Figure 4 shows the best model spatiotemporal and temporospatial model performance for ETa (with MOD16 A2 ETa as the reference dataset). It is seen that the spatial distribution of the NSE is far from homogeneous (Figure 4a), which could not be seen in a spatially lumped approach. From the spatial distribution of NSE, the areas with a relatively poor temporal model performance compared to others could be easily identified. This information could be used to (1) help further improving model performance and (1) detecting the reason for a poor temporal model performance in these areas. The histogram and cumulative plots provide useful statistical information about the temporal model performance (Figure 4b-c). The histogram (or cumulative) plot shows the fraction of the study area that has the NSE within a certain range (or below a certain value). The histogram plot is seen to be highly skewed toward the optimal NSE value (optimal NSE = 1). This is the desired model performance, in other words, a higher skewness of the NSE distribution toward the optimal point indicates a better the model performance.

Despite the NSE is a temporal performance index, it does not provide information about the time when the model poorly performs. This, however, only can be seen in the temporal variation of the SPAEF index (Figure 3d). For example, during April 2004 and April-May 2007, the simulated patterns are seen to be poorly matched with the reference data while in other periods the matching is much better. The distribution of SPAEF indices is also highly skewed toward the optimal value with about 40% of the time step the SPAEF indices are less than -1 (Figure 4e,f).



**Figure 4**. Spatiotemporal and temporospatial and model performance represented by (a) the spatial distribution of the NSE and (b, c) the histogram and cumulative plots of NSE, (d) the temporal variation of SPAEF, (e, f) the histogram and cumulative plots of SPAEF. The ideal value of NSE and SPAEF is 1.

## 3.2. The relation between spatiotemporal and temporospatial model performance

Figure 5 shows that the relation between spatiotemporal (ST) and temporospatial (TS) model performance varies depending on the reference ETa data and the statistical indices used for evaluation. For example, when *KGE* and *SPAEF* were used to derive the spatiotemporal (*S-KGE*) and temporospatial (T-SPAEF) mode performance statistics, *S-KGE* and *T-SPAEF* could be highly positive correlated (Figure 5a), highly negative correlated (Figure 5b), or uncorrelated (Figure 5c). A high positive correlation between *S-KGE* and *T-SPAEF* indicates that an increase in the spatiotemporal model performance will likely lead to an increase in temporospatial model performance. However, a high negative correlation between S-KGE and T-SPAEF means that an increase in spatiotemporal model performance will likely result in a decrease in temporospatial model performance and vice versa. An uncorrelated between S-KGE and T-SPAEF indicates that the spatiotemporal model performance cannot be inferred from the temporospatial model performance and vice versa. The results show that even the same statistical index was used for spatiotemporal and temporospatial model evaluation (e.g., *S-NSE* and *T-NSE*, Figure 5a) it does not always guarantee that an increase in the spatiotemporal model performance will always lead to an increase in the temporospatial model performance (Figure 5a-c). Overall, the results show that both spatiotemporal and temporospatial model performance evaluation are needed.

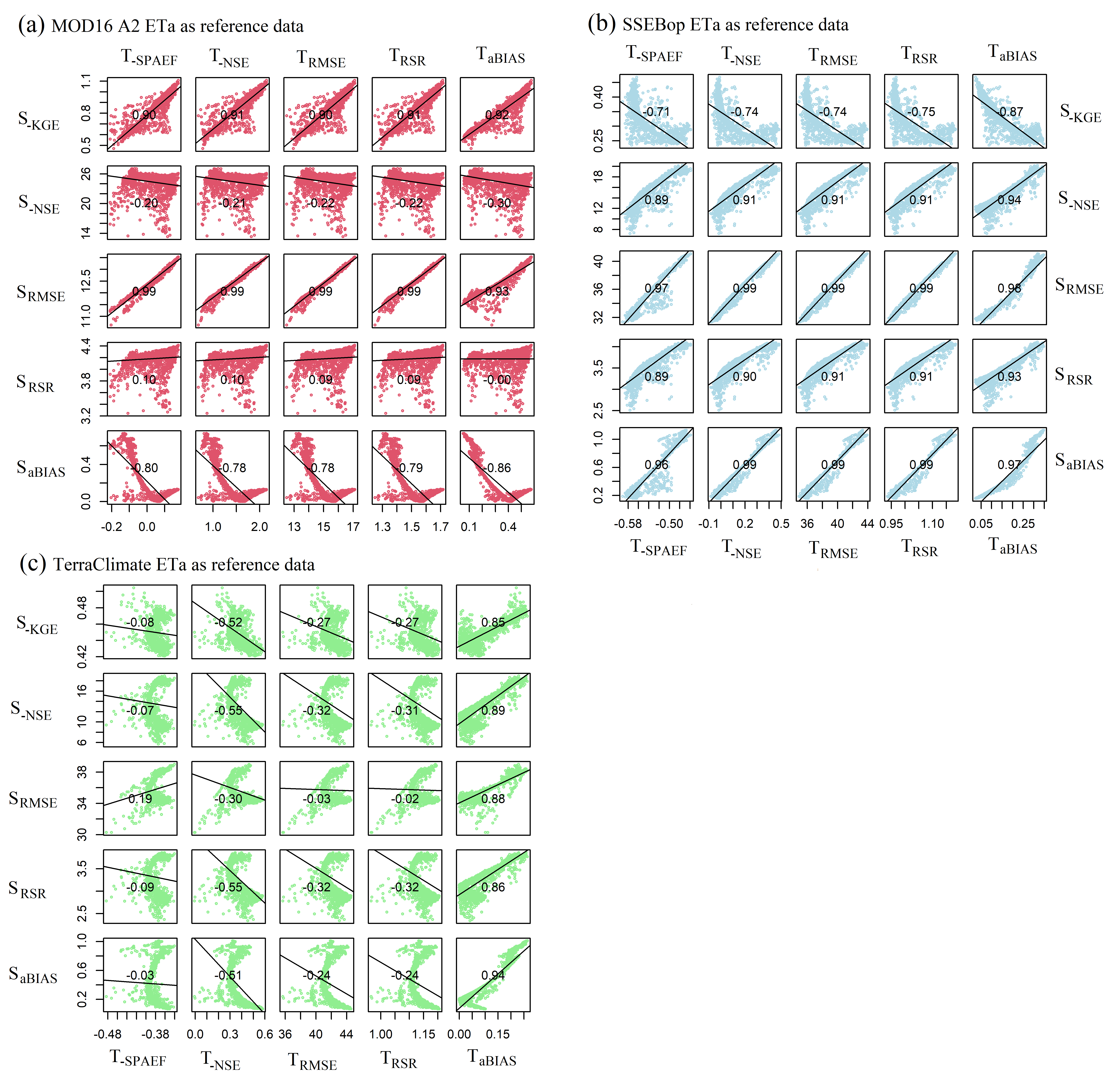


Figure 5: Spatiotemporal and temporospatial model performance evaluation against different ETa product (a) MOD16 A2 ETa, (b) SSEBop, and (c) TerraClimate. Results are from 2000 model runs, lines are the regression lines and numbers inside the plot are the correlation values. In all correlation plots, the points located closer to the lower-left corner indicate a better model performance.

# 4. Conclusions

Distributed hydrological models have long been used for understanding the spatial and temporal hydrological responses of a catchment. To ensure that the model can provide meaningful spatial and temporal information, the simulated spatiotemporal data from these models should be evaluated against the reference data. However, studies have either focused on temporal or spatial model evaluation. In addition, temporal (or spatial) model evaluation is often done in a spatially (or temporally) lumped approach. The terms “spatiotemporal” and “temporospatial” model performance evaluations introduced in this study refer to the evaluation of (1) the temporal model performance in a spatially distributed approach and (2) the spatial model performance in a temporally distributed approach, respectively. Here, we demonstrated that spatiotemporal and temporospatial model performance evaluations are the two different aspects of model evaluation. The spatiotemporal (or temporospatial) model performance evaluation could help in detecting the locations (or the time) where (or when) the temporal (or spatial) patterns are poorly represented by the model. The results show that an increase in the spatiotemporal model performance will not necessarily lead to an increase in the temporospatial model performance and vice versa. This further suggesting that both spatiotemporal and temporospatial model performance evaluation are needed. Overall, the approach proposed in this provides graphical and statistical indicators for understanding spatiotemporal and temporospatial model behavior, therefore, helping in improving the model performance.

# Data availability statement

The MOD16 A2, SSEbop, TerraClimate data used for model evaluation are freely available at <https://modis.gsfc.nasa.gov/data/dataprod/mod16.php>, and <https://earlywarning.usgs.gov/fews/product/460>, and <http://www.climatologylab.org/terraclimate.html>, respectively. The R code and SWAT model used in this study are available from the corresponding author on request.

# Acknowledgments

# References

Abiodun, O. O., Guan, H., Post, V. E. A., & Batelaan, O. (2018). Comparison of MODIS and SWAT evapotranspiration over a complex terrain at different spatial scales. *Hydrology and Earth System Sciences*, *22*(5). https://doi.org/10.5194/hess-22-2775-2018

Arnold, J. G., & Fohrer, N. (2005). SWAT2000: Current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes*, *19*(3). https://doi.org/10.1002/hyp.5611

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1). https://doi.org/10.1080/02626667909491834

Brocca, L., Tullo, T., Melone, F., Moramarco, T., & Morbidelli, R. (2012). Catchment scale soil moisture spatial-temporal variability. *Journal of Hydrology*, *422*–*423*. https://doi.org/10.1016/j.jhydrol.2011.12.039

Campo, L., Caparrini, F., & Castelli, F. (2006). Use of multi-platform, multi-temporal remote-sensing data for calibration of a distributed hydrological model: An application in the Arno basin, Italy. *Hydrological Processes*, *20*(13). https://doi.org/10.1002/hyp.6061

Dembélé, M., Ceperley, N., Zwart, S. J., Salvadore, E., Mariethoz, G., & Schaefli, B. (2020). Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Advances in Water Resources*, *143*. https://doi.org/10.1016/j.advwatres.2020.103667

Edmunds, W. M., Fellman, E., Goni, I. B., & Prudhomme, C. (2002). Spatial and temporal distribution of groundwater recharge in northern Nigeria. *Hydrogeology Journal*, *10*(1). https://doi.org/10.1007/s10040-001-0179-z

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12). https://doi.org/10.1002/joc.5086

Gassman, P. W., Reyes, M. R., Green, C. H., & Arnold, J. G. (2007). The soil and water assessment tool: Historical development, applications, and future research directions. In *Transactions of the ASABE* (Vol. 50, Issue 4).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2). https://doi.org/10.1016/j.jhydrol.2009.08.003

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3). https://doi.org/10.1002/joc.3711

Herman, M. R., Nejadhashemi, A. P., Abouali, M., Hernandez-Suarez, J. S., Daneshvar, F., Zhang, Z., Anderson, M. C., Sadeghi, A. M., Hain, C. R., & Sharifi, A. (2018). Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. *Journal of Hydrology*, *556*. https://doi.org/10.1016/j.jhydrol.2017.11.009

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15). https://doi.org/10.1002/joc.1276

Immerzeel, W. W., & Droogers, P. (2008). Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology*, *349*(3–4). https://doi.org/10.1016/j.jhydrol.2007.11.017

Jiang, L., Wu, H., Tao, J., Kimball, J. S., Alfieri, L., & Chen, X. (2020). Satellite-based evapotranspiration in hydrological model calibration. *Remote Sensing*, *12*(3). https://doi.org/10.3390/rs12030428

Kannan, N., White, S. M., Worrall, F., & Whelan, M. J. (2007). Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000. *Journal of Hydrology*, *332*(3–4). https://doi.org/10.1016/j.jhydrol.2006.08.001

Kim, N. W., Chung, I. M., Won, Y. S., & Arnold, J. G. (2008). Development and application of the integrated SWAT-MODFLOW model. *Journal of Hydrology*, *356*(1–2). https://doi.org/10.1016/j.jhydrol.2008.02.024

Klingler, C., Schulz, K., & Herrnegger, M. (2021). Large-Sample Data for Hydrology and Environmental Sciences for Central Europe. *Earth Science Data*, *March*.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Kiyotoshi, T. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1). https://doi.org/10.2151/jmsj.2015-001

Koch, J., Demirel, M. C., & Stisen, S. (2018). The SPAtial EFficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development*, *11*(5). https://doi.org/10.5194/gmd-11-1873-2018

Koch, J., Jensen, K. H., & Stisen, S. (2015). Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. *Water Resources Research*, *51*(2). https://doi.org/10.1002/2014WR016607

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*. https://doi.org/10.1029/2012WR012195

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7). https://doi.org/10.1029/94jd00483

Mendiguren, G., Koch, J., & Stisen, S. (2017). Spatial pattern evaluation of a calibrated national hydrological model - A remote-sensing-based diagnostic approach. *Hydrology and Earth System Sciences*, *21*(12). https://doi.org/10.5194/hess-21-5987-2017

Mu, Q., Zhao, M., & Running, S. W. (2013). MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3). *Algorithm Theoretical Basis Document*, *Collection*.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3). https://doi.org/10.1016/0022-1694(70)90255-6

Neitsch, S. ., Arnold, J. ., Kiniry, J. ., & Williams, J. . (2011). Soil & Water Assessment Tool Theoretical Documentation Version 2009. *Texas Water Resources Institute*. https://doi.org/10.1016/j.scitotenv.2015.11.063

Nguyen, V.T., & Dietrich, J. (2018). Modification of the SWAT model to simulate regional groundwater flow using a multicell aquifer. *Hydrological Processes*, *32*(7). https://doi.org/10.1002/hyp.11466

Nguyen, V.T., Dietrich, J., & Uniyal, B. (2020). Modeling interbasin groundwater flow in karst areas: Model development, application, and calibration strategy. *Environmental Modelling and Software*, *124*. https://doi.org/10.1016/j.envsoft.2019.104606

Nguyen, T. V., Kumar, R., Lutz, S. R., Musolff, A., Yang, J., & Fleckenstein, J. H. (2021). Modeling Nitrate Export From a Mesoscale Catchment Using StorAge Selection Functions. *Water Resources Research*. https://doi.org/10.1029/2020wr028490

Odusanya, A. E., Mehdi, B., Schürz, C., Oke, A. O., Awokola, O. S., Awomeso, J. A., Adejuwon, J. O., & Schulz, K. (2019). Multi-site calibration and validation of SWAT with satellite-based evapotranspiration in a data-sparse catchment in southwestern Nigeria. *Hydrology and Earth System Sciences*, *23*(2). https://doi.org/10.5194/hess-23-1113-2019

Rajib, A., Evenson, G. R., Golden, H. E., & Lane, C. R. (2018). Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. *Journal of Hydrology*, *567*. https://doi.org/10.1016/j.jhydrol.2018.10.024

Rientjes, T. H. M., Muthuwatta, L. P., Bos, M. G., Booij, M. J., & Bhatti, H. A. (2013). Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration. *Journal of Hydrology*, *505*. https://doi.org/10.1016/j.jhydrol.2013.10.006

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*(1). https://doi.org/10.1175/2007MWR2123.1

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), 1–25. https://doi.org/10.1029/2008WR007327

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., & Verdin, J. P. (2013). Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. *Journal of the American Water Resources Association*, *49*(3). https://doi.org/10.1111/jawr.12057

Stisen, S., Koch, J., Sonnenborg, T. O., Refsgaard, J. C., Bircher, S., Ringgaard, R., & Jensen, K. H. (2018). Moving beyond run-off calibration—Multivariable optimization of a surface–subsurface–atmosphere model. *Hydrological Processes*, *32*(17). https://doi.org/10.1002/hyp.13177

Thomas, A. (2000). Spatial and temporal characteristics of potential evapotranspiration trends over China. *International Journal of Climatology*, *20*(4). https://doi.org/10.1002/(SICI)1097-0088(20000330)20:4<381::AID-JOC477>3.0.CO;2-K

Uniyal, B., Dietrich, J., Vu, N. Q., Jha, M. K., & Arumí, J. L. (2019). Simulation of regional irrigation requirement with SWAT in different agro-climatic zones driven by observed climate and two reanalysis datasets. *Science of the Total Environment*, *649*. https://doi.org/10.1016/j.scitotenv.2018.08.248

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., Van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., & Savenije, H. H. G. (2016). Global root zone storage capacity from satellite-based evaporation. *Hydrology and Earth System Sciences*, *20*(4). https://doi.org/10.5194/hess-20-1459-2016

Wilson, D. J., Western, A. W., & Grayson, R. B. (2004). Identifying and quantifying sources of variability in temporal and spatial soil moisture observations. *Water Resources Research*, *40*(2). https://doi.org/10.1029/2003WR002306