5-28-2019

Using Satellite-Based Hydro-Climate Variables And Machine Learning For Streamflow Modeling At Various Scales In The Upper Mississippi River Basin

Dongjae Kwon
Illinois State University, djkwon85@gmail.com

USING SATELLITE-BASED HYDRO-CLIMATE VARIABLES AND MACHINE LEARNING FOR STREAMFLOW MODELING AT VARIOUS SCALES IN THE UPPER MISSISSIPPI RIVER BASIN

DONGJAE KWON

61 Pages

Streamflow data are essential to study the hydrologic cycle and to attain appropriate water resource management policies. However, the availability of gauge data is limited due to various reasons such as economic, political, instrumental malfunctioning, and poor spatial distribution. Although streamflow can be simulated by process-based and machine learning approaches, applicability is limited due to intensive modeling effort, or its black-box nature, respectively. Here, we introduce a machine learning (Boosted Regression Tree (BRT)) approach based on remote sensing data to simulate monthly streamflow for three of varying sizes watersheds in the Upper Mississippi River Basin (UMRB). By integrating spatial land surface and climate variables that describe the subwatersheds in a basin as an input dataset and streamflow as an output learning dataset in a machine learning model (MLM), relationships between watershed characteristics and streamflow are established. The testing results of NSE with UMRB, IRW, and RRW of 0.8042, 0.7593, and 0.6856, respectively showed the remote sensing-based MLM can be effectively applied to streamflow prediction and has advantages for large basins compared with the performances of process-based approaches. Further, Predictor Importance (PI) analysis revealed the most important remote sensing variables and the most representative subwatersheds.

KEYWORDS: Streamflow; Machine Learning; GRACE; Remote Sensing; Upper Mississippi

USING SATELLITE-BASED HYDRO-CLIMATE VARIABLES AND MACHINE LEARNING FOR STREAMFLOW MODELING AT VARIOUS SCALES IN THE UPPER MISSISSIPPI RIVER BASIN

DONGJAE KWON

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Geography, Geology, and the Environment

ILLINOIS STATE UNIVERSITY

2019

USING SATELLITE-BASED HYDRO-CLIMATE VARIABLES AND MACHINE LEARNING FOR STREAMFLOW MODELING AT VARIOUS SCALES IN THE UPPER MISSISSIPPI RIVER BASIN

DONGJAE KWON

COMMITTEE MEMBERS:

Wondwosen Mekonnen Seyoum, Chair

Dagmar Budikova

Eric Peterson

ACKNOWLEDGMENTS

It is hard to believe that I have almost finished my master's course in ISU. If I close my eyes, I still can see and feel the first moment when I came here. I was worried and nervous. The sunshine, breeze, trees, rains, and atmosphere, everything was unfamiliar. However, the worries were gone after I met the people at ISU who have been unbelievably kind. I could do my best work because of all the support I have received from the Department of Geography, Geology, and the Environment at Illinois State University.

I cannot say enough how much I appreciate people I met here in ISU because of my poor English. I thank my advisor, Dr. Wondwosen Seyoum for his encouragement, guidance, and patience. I could overcome the difficulties of the research and my life based on his advice and dedicated help. I can say the same of the rest of the Geology faculty, who share these same virtues. Dr. Seyoum and the other two members of my committee, Dr. Eric Peterson and Dr. Dagmar Budikova, have been supportive and posed challenging questions to assure my best work. Thank you for your patience, encouragement, and supporting.

Lastly, a thank you to my fellow students. It was a very good experience with you, going to the field, sharing ideas, working together, participating in the symposium, party, etc. Thank you for my officemates, Prince, Luis, Teddy, Gare, and Joe. Thank you a lot for your support. I will not forget your help, especially when I was sick, you gave me food and taught me how to make my diet better. Those will be remembered for my lifetime.

D. K.

CONTENTS

TABLES

FIGURES

CHAPTER I: INTRODUCTION

Managing streamflow has been considered one of the most important challenges. Globally, we have faced severe droughts and floods due to unexpected climate impacts (Alexander et al., 2006). Furthermore, in the future, 10-40% increase in streamflow is expected throughout eastern equatorial Africa, the La Plata basin, and high-latitude North America and Eurasia, while 10-30% of decrease in streamflow is expected in the southern part of Africa, southern Europe, the Middle East, and mid-latitude western North America (Milly et al., 2005). Governments around the world will face serious challenges regarding water resources management strategies. Therefore, estimation of watershed responses to various climate states is essential, and this can be accomplished by appropriate hydrologic modeling techniques. Streamflow measurement is very important in hydrologic modeling tasks because it is the only phase of the hydrological cycle that can be measured accurately in well-defined and confined channels (Herschy, 2014). However, *in-situ* streamflow data are not fully available globally due to poor distributions of the gauging stations, economic reasons, political issues, and restricted data sharing (Beven, 2011). Even in developed countries, malfunctioning of gauging stations is an inevitable task. Fortunately, the limitations of available *in-situ* streamflow data can be supplemented by rainfall-runoff/streamflow modeling.

Streamflow modeling methods are generally classified into two categories: process-based models (i.e., physics-based-models) and empirical models (i.e., black box models) (Bourdin et al., 2012; Chiew et al., 1993; Minns and Hall, 1996) Process-based models such as Soil and Water Assessment Tool (SWAT), MIKE 11, and the Precipitation-Runoff Modeling System (PRMS), are based on water balance equations they can compute streamflow by simulating the contributions of hydrologic reservoirs such as soil, snow pack, canopy water, and groundwater and climatic factors

such as evaporation, transpiration, temperature and solar radiation (DHI, 2003; Markstrom et al., 2015; Neitsch et al., 2011). Some of them can also be used for water quality assessments. Process-based models require many physical parameters including elevation, land use, soil type, tile drainage, and climate data to obtain precise estimation of streamflow. Process-based approaches are data intensive, as they require many kinds of dataset to define parameters and characterize the basins (Beven, 2011; Seyoum and Milewski, 2016; Tokar and Johnson, 1999).Since there are many poorly studied or insufficiently gauged basins around the globe (Blöschl, 2005), accomplishing processed-based approach everywhere is burdensome, especially, for a large area.

Empirical models, also called black-box models, are data-driven approaches that require as input fewer basin characteristics. The empirical approaches have two categories: conventional statistical approaches and machine learning (Bourdin et al., 2012). Statistical approaches are based on regression of the relationships between input (e.g. rainfall) and output (e.g. runoff) data and give us mathematical representation of physical hydrological processes (Bourdin et al., 2012). Some of them are univariate methods that are directly established without any detailed physical information of a basin, while the others (e.g., principle component regression, and Auto-Regressive with exogenous variables) consider several catchment variables (Beven, 2011; Bourdin et al., 2012; Chiew et al., 1993; Khosravi et al., 2013). The conventional statistical approaches are relatively easy to develop and use, however, uncertainties are larger than machine learning approaches because the streamflow process is highly nonlinear and the conventional methods typically assume linear relationships (Bourdin et al., 2012; Hsu et al., 1995).

Machine learning approaches such as artificial neural network (ANN) are beginning to receive attention as computing power and techniques are developing. They provide promising ways to infer a complex, perhaps non-linear relationships among input variables (e.g. watershed

characteristics) and output variable (e.g. streamflow) ofwatersheds (Tanty and Desmukh, 2015). Early research has shown the effectiveness of machine learning techniques to study streamflow responses (Dawson and Wilby, 1998; Hsu et al., 1995; Minns and Hall, 1996; Mutlu et al., 2008; Riad et al., 2004; Tokar and Johnson, 1999). However, the works were done for relatively small watersheds (< 500 km^2; except Hsu et al., 1995: 2781 km^2), and mainly considered precipitation data as the sole input variable. To estimate streamflow for a larger scale watershed, it is important to include variables other than the precipitation, such as water-budget-related variables including evapotranspiration (ET), memory effect (antecedent precipitation), and snow water melting. During the last few decades, extensive research has been conducted by applying more developed machine learning techniques (Chen et al., 2015; Kratzert et al., 2018; Seo et al., 2018; Taormina et al., 2015; Yaseen et al., 2017) and/or assigning additional variables such as land surface and meteorological data (Bajwa and Vibhava, 2009; Chang and Chen, 2018; Deo and Şahin, 2016; Kratzert et al., 2018; Rasouli et al., 2012; Seyoum et al., 2019; Seyoum and Milewski, 2017) to increase applicability of MLM for hydrologic studies.

Regardless of the specific MLM methods, the most important tasks for precise streamflow estimations can be summarized as (1) accounting for the effects of baseflow and antecedent precipitation, (2) including the accurate water-budget variables information such as ET, and (3) minimizing data demand. These tasks can be supplemented by remote sensing methods. Today, numerous remote sensing datasets are publicly available and are collected by various instruments, such as satellite and airborne systems. The advantages of utilizing satellite data is its availability in time and space, especially in data sparse regions, and large area coverage. Hence, many hydrology-related studies have shown the effectiveness of combining remote sensing data with conventional *in-situ* data (Ahmad et al., 2010; Boegh et al., 2009; Chen

et al., 2005; Melesse and Graham, 2004; Seyoum, 2018; Seyoum et al., 2015). As space technology evolves, the quality of remotely sensed data will increase as wil the aspect of spatial and/or temporal coverage and resolution. Available satellite data with global coverage and relatively high temporal resolution includes Land Surface Temperature (LST), precipitation, vegetation index, soil moisture, canopy water, terrestrial water storage, these have been utilized in many hydrological studies (Brakenridge et al., 2005; Hong et al., 2007; Liu et al., 2012; Mahmoud, 2014).

Machine-learning methods combined with remote sensing data may provide a great opportunity to investigate watershed response to streamflow. The Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS) anomaly (TWSA) data can provide monthly terrestrial water storage. TWS is related to groundwater storage and baseflow, that cannot be easily estimated by the surface information of a basin. However, one of the limitations of GRACE data is its coarse spatial resolution (100 km × 100 km). Several studies in hydrology showed MLM could overcome this limitation. Seyoum and Milewski (2016) have shown downscaling of the GRACE TWSA by the combination of machine learning, remote sensing (e.g., LST, soil moisture, precipitation, vegetation index, and GRACE TWSA), and streamflow. Irreversibly, GRACE TWSA can be used for streamflow estimation based on MLM and the other remote sensing-based variables.The objectives of this study are: (1) establish streamflow prediction models for the UMRB and its subwatersheds based on the remote sensing-based MLM, (2) investigate optimal input variables for the training dataset of MLM, which are valid for streamflow prediction, (3) evaluate the predictor variable importance, and (4) assess the efficiency of the remote sensing-based MLM. By integrating satellite-based spatial land surface and climate data describing the watersheds as an input dataset and in-situ streamflow data as an

output learning dataset, relationships between watershed characteristics and streamflow will be established using MLM. Boosted Regression Tree (BRT) method is used for MLM as it provides a better understanding on its intrinsic structure and interpretation than ANN method which is commonly used in hydrological studies (Bourdin et al., 2012; Friedman et al., 2001). The results are tested using *in-situ* streamflow data independent of the training data. The effectiveness of the method developed in this study is evaluated by comparing the statistical performance metrics of this study with results from previous studies conducted in the same study areas using process-based modeling approach. The result from this study opens up a new avenue of using spatio-temporal remote sensing data in streamflow prediction.

CHAPTER II: STUDY AREA

Three different-sized basins located in the Upper Mississippi River Basin were used to test the approach developed in this study. These are the Raccoon River Watershed, Illinois River Watershed, and the Upper Mississippi River Basin. The Upper Mississippi River Basin covers an area of 492,000 km^2 while the Illinois River Watershed and the Raccoon River Watersheds cover area of 74,677 km^2–and 9,400 km^2, respectively (Figure 1). The study sites were chosen by considering the availability of previous hydrological modeling research conducted using process-based models. One of the objectives of this study is to assess the efficiency of the MLM by comparing it with results of previous research. In addition, various sized basin/watersheds are selected to explore if size of a watershed affects the accuracy of MLM. Higher accuracy of streamflow estimation from the MLM is expected for larger basins/watersheds as the effectiveness of MLM is expected to be be limited in smaller basins/watersheds due to shorter time of concentration in smaller basins/watersheds.

Figure 1. Map of the study area containing Upper Mississippi River Basin (Top) and its subwatersheds (Bottom left and right). Stars indicate gauging stations utilized in this study. Note that the subbasin and the subwatershed at downstream of the gauging stations are excluded (light colored parts in UMRB, IRW and RRW). The subwatershed IDs are labelled according to USGS Watershed Boundary Dataset (WBD) Hydrologic Units (HU).

Upper Mississippi River Basin (UMRB)

UMRB is one of the major sub-basins of the Mississippi River Basin (MRB) which is the largest river basin in North America. UMRB includes large parts of Illinois, Iowa, Minnesota, Missouri, and Wisconsin and small parts of Indiana, Michigan, and South Dakota with underlying glacial aquifer system. More than 30 million residents live in this region and rely on river water and discharge that has significantly increased due to land cover/land use changes from cultivation such as an expansion of soybean/corn fields (NRCS, 2010; Schilling et al., 2010; Srinivasan et al., 2010). Understanding and quantifying the factors affecting streamflow is important to ecological and agricultural aspects because it is highly related with nutrient delivery processes (Schilling et al., 2010). In the UMRB, land use type consists of deciduous forest (19.4 %), corn-soybean (33.9%), hay (11.5 %), developed area (8.4 %), the other cultivated crop (7.5 %), pasture (4.9 %), open water (2.8 %), and grassland herbaceous (2.8 %), (Srinivasan et al., 2010). Soil leaching potential and soil runoff potential varies spatially according to various soil type and surface slope (NRCS, 2010). Table 1 shows estimated annual precipitation and land surface temperature (LST) of UMRB for 5 years (from Jan. 2002 to Dec. 2016) from remote sensing data (TRMM and MODIS LST) according to its sub-basin's hydrologic unit code (HUC) of United States Geological Survey (USGS) Watershed Boundary Dataset Hydrologic Units (WBDHU). The annual precipitation is from 980 to 1150 mm and average land surface temperature (LST) is from 9°C to 13°C for the 5-year period.

Table 1

Land surface temperature (LST) and annual precipitation based on remote sensing data according to 15 of 6-digit level sub-basins (HUC6) in UMRB (from Jan. 2002 to Dec. 2016)

HUC6	NAME	area [km^2]	avg. LST [°C]	max. LST [°C]	min. LST [°C]	P [mm/y]
70600	Upper Mississippi-Maquoketa-Plum	22258	8.65	26.57	-16.25	1072
70802	Iowa	32709	9.48	26.31	-16.73	1040
71000	Des Moines	37442	10.28	27.01	-15.33	972
71300	Lower Illinois	46354	12.04	27.92	-11.95	1107
70801	Upper Mississippi-Skunk-Wapsipinicon	26694	10.64	26.94	-15.21	1062
71200	Upper Illinois	28314	10.91	27.55	-15.95	1059
71100	Upper Mississippi-Salt	26110	12.92	29.87	-10.87	1122
70900	Rock	28276	9.10	27.12	-18.87	1067
70101	Mississippi Headwaters	29973	4.73	23.54	-18.73	749
70200	Minnesota	44051	7.97	25.82	-17.38	756
70400	Upper Mississippi-Black-Root	27870	7.47	25.39	-18.39	983
70500	Chippewa	24706	5.76	23.42	-17.01	1010
70700	Wisconsin	30906	6.59	25.10	-19.27	1017
70300	St. Croix	19995	5.81	24.05	-17.25	930
70102	Upper Mississippi-Crow-Rum	22113	6.80	25.50	-18.65	885

Note: Wabash (51201) at downstream of the gauging station was excluded.

Due to geographical importance of the MRB and UMRB, many studies have been conducted in the region, including large projects such as World Climate Research Programme's Global Energy and Water Cycle Experiment (GEWEX) and Continental-Scale International Project (GCIP) for the long term goal of demonstrating skill in predicting changes in water

resources on timescales up to seasonal, annual, and inter-annual (Maurer and Lettenmaier, 2003). Previous studies are mainly based on process-based modeling approaches and use the soil and water assessment tool model (SWAT) (Arnold et al., 2000; Gassman et al., 2006; Jha et al., 2006; Jha et al., 2004; Srinivasan et al., 2010). Specifically, Arnold et al. (2000) used SWAT to estimate baseflow and groundwater recharge; Jha et al. (2004) and Jha et al. (2006) used it to conduct climate change sensitivity assessment of streamflow, while Srinivasan et al. (2010) estimated hydrological budget and crop yield prediction in ungauged perspective by using SWAT with spatial data (e.g., DEM, land use). Several conceptual and empirical approaches have been conducted using the rainfall-runoff model (Liston et al., 1994; Maurer and Lettenmaier, 2003; Perrin et al., 2007).

Illinois River Watershed (IRW)

The Illinois River Watershed (Figure 1 bottom right) has a drainage area of approximately 75,000 km^2, which includes 2,800 km^2 in Wisconsin and 7,900 km^2 in Indiana. IRW is the most important watershed in Illinois; 44 percent of the state's land is included by the watershed, while 46 percent of agricultural land, 28 percent of forest, 37 percent of surface waters, and 95 percent of urban areas are included (USACE, 2006). Due to extensive human development in this region, most of prairies and forests have disappeared and recently, the largest land use in IRW is agriculture (64 %) and the rest are grassland (17 %), forest (10 %), urban (5 %), and water and wetland (4 %) (Demissie et al., 2006; USACE, 2006). Annual air temperature is approximately 11.5°C and precipitation is 1050 mm during last 10 years with warm (23°C ~ 24°C) and wet (90 mm ~ 130 mm) summers (June, July, and August), and cold (-4°C ~ 1°C) and relatively dry (40 mm ~ 120 mm) winters (December, January, and February) (Illinois Climate Network, 2015). Table 2 shows detailed precipitation and LST records for the

subwatersheds. The dominant soil types are Mollisols and Alfisols with some of Entisols and Inceptisols, which are underlined by glacial aquifer system. Land use changes and widening urban areas in this region caused more rapid streamflow responses in storm events, increasing erosive force, and decreasing baseflow (USACE, 2006) that may bring out the difficulties of streamflow and ecological management.

Table 2

Land surface temperature (LST) and annual precipitation based on remote sensing data according to 18 of 8-digit level subwatersheds (HUC8) in IRW (from Jan. 2002 to Dec. 2016)

HUC8	NAME	area [km^2]	avg. LST [°C]	max. LST [°C]	min. LST [°C]	P [mm/y]
7120002	Iroquois	5537	11.41	26.61	-18.91	1116
7130001	Lower Illinois-Senachwine Lake	5076	10.89	27.07	-13.27	1039
7130007	South Fork Sangamon	3030	12.80	28.15	-9.80	1124
7130002	Vermilion	3453	11.64	26.96	-13.76	1094
7120001	Kankakee	7846	11.04	26.64	-18.66	1079
7130010	La Moine	3495	11.94	27.94	-13.37	1091
7120004	Des Plaines	3770	11.91	29.75	-14.33	1017
7120003	Chicago	1699	12.27	29.91	-12.28	988
7130009	Salt	4836	12.11	27.80	-14.14	1158
7120005	Upper Illinois	2606	11.21	27.18	-19.59	1042
7130004	Mackinaw	2976	11.72	27.72	-13.13	1147
7130011	Lower Illinois	5887	12.88	28.84	-10.37	1098
7130006	Upper Sangamon	3732	12.22	27.93	-14.04	1123
7130008	Lower Sangamon	2311	12.80	28.64	-11.73	1135
7130003	Lower Illinois-Lake Chautauqua	4203	11.96	28.48	-17.44	1139
7130005	Spoon	4831	11.41	27.16	-15.03	1079
7120007	Lower Fox	2857	10.55	27.19	-17.03	1028
7120006	Upper Fox	3999	9.05	27.93	-19.93	1055

Note: Macoupin (7130012) at downstream of the gauging station was excluded.

Raccoon River Watershed (RRW)

The Raccoon River Watershed (Figure 1) is located in the SW section of the UMRB and encompasses approximately 9,400 km^2 of prime agricultural land in west-central Iowa, which consist of cropland (75.3%), grassland (16.3%), forest (4.4%), and urban (4.0%) areas (Jha et al., 2007). As with the most part of the agricultural Midwest, land use in the watershed has significantly changed, which can affect streamflow responses (Schilling et al., 2008). Annual precipitation ranges between 860 and 1070 mm and mean surface temperature falls between 9.5°C and 11°C (Table 3).

Table 3

Land surface temperature (LST) and annual precipitation based on remote sensing data according to 24 of 10-digit subwatersheds (HUC10) in RRW (from Jan. 2002 to Dec. 2016).

HUC10	NAME	area [km^2]	avg. LST [°C]	max. LST [°C]	min. LST [°C]	P [mm/y]
710000704	Upper South Raccoon River	315	10.64	27.61	-15.63	1040
710000707	Middle South Raccoon River	290	10.97	28.14	-14.99	1063
710000703	Brushy Creek	368	10.42	27.42	-14.78	997
710000615	Swan Lake Branch-North Raccoon River	482	10.94	27.29	-14.03	1054
710000705	Mosquito Creek	297	10.79	27.70	-15.12	1044
710000606	Lake Creek	331	9.97	27.84	-14.96	908
710000604	Indian Creek	226	10.08	27.64	-14.18	904
710000709	Lower South Raccoon River	239	10.94	28.71	-15.02	1067
710000611	East Buttrick Creek	193	10.45	27.38	-16.07	974

(Table Continues)

HUC10	NAME	area [km^2]	avg. LST [°C]	max. LST [°C]	min. LST [°C]	P [mm/y]
710000612	Buttrick Creek	352	10.51	27.74	-15.67	957
710000614	Otter Creek-North Raccoon River	392	10.57	27.32	-15.04	982
710000601	Little Cedar Creek	218	9.71	27.72	-14.80	868
710000602	Prairie Creek-Cedar Creek	679	9.82	27.78	-15.14	874
710000603	Headwaters North Raccoon River	900	9.66	27.36	-14.85	862
710000706	Lower Middle Raccoon River	301	10.63	28.22	-15.17	1051
710000605	Camp Creek	381	10.05	27.99	-15.33	905
710000608	Elk Run-North Raccoon River	541	10.37	27.68	-14.42	925
710000607	Purgatory Creek	187	10.34	27.72	-14.67	932
710000613	Greenbrier Creek	183	10.68	27.71	-15.85	1026
710000702	Upper Middle Raccoon River	645	10.49	27.49	-14.44	957
710000610	Hardin Creek	443	10.48	27.55	-15.08	952
710000701	Willow Creek	316	10.58	27.46	-15.52	1005
710000708	Panther Creek	173	11.06	27.25	-14.03	1063
710000609	Welshs Slough-Cedar Creek	419	10.32	27.55	-15.28	934

Note: Walnut Creek (0710000616) and Raccoon River (0710000617) at downstream of the gauging station were excluded.

CHAPTER III: METHODS AND DATA

Data Source and Processing

Streamflow is governed by the water balance equation (Equation 1). Thus, streamflow can be calculated, if we have information about the other components of the water balance, such as evapotranspiration and groundwater storage. In terms of processes, streamflow emerging out of a given watershed is influenced by various watershed characteristics such as amount and type of vegetations, climate conditions (precipitation, temperature, snow melt, wind speed, humidity), soil type, topography (surface slope), and land cover type (Beven, 2011; Schilling et al., 2010). Like distributed models, it is possible to characterize the streamflow in a given watershed using information from watershed characteristics listed above. These watershed characteristics can be obtained from satellite data. Therefore, considering data availability and spatial resolution, various remote sensing-based and other spatial datasets that control watershed's streamflow responses were collected for this study.

$$P = Q + ET + \Delta S \quad (1)$$

where P is precipitation, ET is evapotranspiration, Q is streamflow, and ΔS is the change in storage

First, the study used 14 variables including Terrestrial Water Storage (TWSA), Land Surface Temperature (LST), Monthly LST change (ΔLST), Normalized Difference Vegetation Index (NDVI), plant canopy water, soil moisture, snow water equivalent, humidity, wind speed, precipitation in current month (P), precipitation in previous month (P_{M-1}), and fraction of amount of precipitation of wet condition, (P > 2.5 mm), extreme condition (P > 90%), and very extreme condition (P > 99%) representing each watershed. Variable importance analysis and cross-correlation test (APPENDIX A) showed that variables such as plant canopy water, soil moisture,

snow water equivalent, humidity, and wind speed have insignificant roles in simulating streamflow. Thus, the remaining nine important variables were used as input variables in the MLM to simulate streamflow (Figure 2). The selected remote sensing data were resampled according to the HUs (Table 1, 2, and 3) of the study sites (please see the model design section) and assigned as predictors in MLM. The total study period is bounded by the availability of GRACE data, which is from April 2002 to July 2016. This range is divided into training (October 2004 to July 2016, 142 samples) and testing (April 2002 to September 2004, 30 samples). Detailed descriptions for each remote sensing data are provided below.

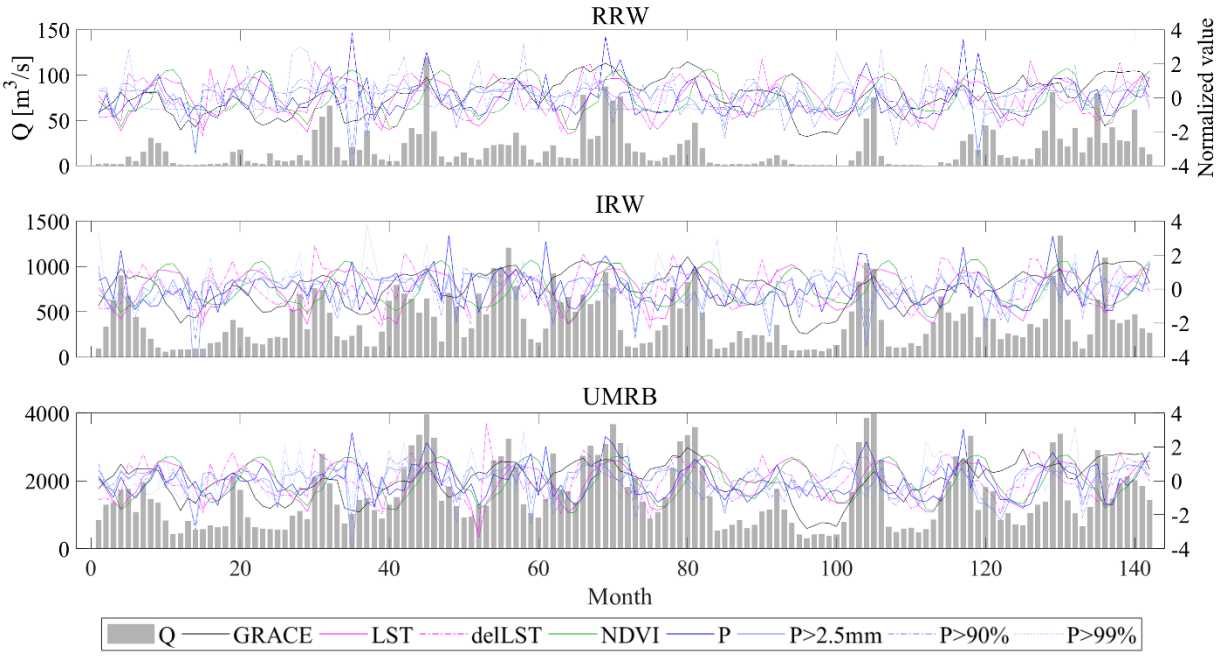


Figure 2. Monthly time series (from October 2004 to July 2016; training period) plots of the selected remote sensing-based data (GRACE TWSA, LST, ΔLST, P, P > 2.5 mm, P > 90%, P > 99%, and NDVI; right axis) and streamflow (left axis) from gauging stations (outlets). Please note that the remote sensing data are showing averaged-and-normalized values to visualize here. For more details, please see the appendix C, D and E. *P_{M-1} is not included since it is just a time-lagged repetition of P.

Terrestrial Water Storage Anomaly (TWSA)

The Gravity Recovery and Climate Experiment (GRACE) TWSA dataset was used to explain contributions of groundwater (baseflow) to streamflow. GRACE TWSA can provide firsthand information of water storage anomaly directly linked to the water balance of a hydrologic system (Seyoum and Milewski, 2016). GRACE mission consists of two identical satellites that have 500 km orbit altitude and separated 220 km each other (Steitz et al., 2002). The K-band ranging system provides precise (within 1 micron, or the width of a human hair) measurements of the distance change between the two satellites, which can be calculated to fluctuations in Earth's gravity field (Steitz et al., 2002). Most of the GRACE TWSA is related to the fluctuations of TWS after atmospheric and oceanic effects are removed (Landerer and Swenson, 2012). Three solutions of the RL-05 gridded (1 ° × 1 °; ~ 100 km × 100 km) level-3 GRACE data from the processing centers (the Center for Space Research at the University of Texas, Austin; CSR, the Jet Propulsion Laboratory; JPL, and the GeoforschungsZentrum Potsdam; GFZ) were downloaded, restored by multiplying the scaling factor, and ensembled (averaged) to ensure the highest level of accuracy (Landerer and Swenson, 2012). The data are provided by the NASA MEaSUREs Program (URL: https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/ ; date accessed: 1 September 2017).

Precipitation (P)

Tropical Rainfall Measurement Mission (TRMM) 3B43 and 3B42 products were used for precipitation data. TRMM provides global (from the equator to mid-latitudes; 50°N-50°S) monthly (3B43) and daily (3B42) rainfall estimation products based on precipitation rate retrieved by spaceborne sensors such as microwave imager, precipitation radar, and visible-

infrared scanner (Kummerow et al., 1998). Cumulative monthly [mm/month] and daily [mm/day] precipitation data were used in the analysis. The spatial resolution of the data is 0.25° × 0.25° (~ 27.8 km × 27.8 km). Various precipitation indices recommended by Expert Team on Climate Change Detection and Indices (ETCCDI) and others (Zhang et al., 2011), such as fraction of total monthly precipitation calculated using number of days greater than median precipitation (P ~ 2.5 mm), very wet days (P > 90%), and extremely wet days (P > 99%) condition in a month. In addition, one month-lagged precipitation (PM-1) was assigned as one of a training variable in the MLMs in order to include the effect of antecedent precipitation on streamflow. TRMM is available at NASA Giovanni (Geospatial Interactive Online Visualization ANd aNalysis Infrastructure) service (URL: https://giovanni.gsfc.nasa.gov/giovanni/ ; date accessed: 26 September 2017).

Land Surface Temperature (LST)

MODIS MOD11C3 (MODIS/Terra Land Surface Temperature and Emissivity Monthly L3 Global 0.05Deg CMG) product was used for watershed temperature information is expected to be related to ET and unspecified seasonal effect. The product is monthly composited average derived from the MOD11C1 daily LST with 0.05° × 0.05° (~ 5.6 km × 5.6 km) resolution (Wan et al., 2015). The latest version is 5 (V5), however, version 4 products (V4 and V41) were used in this study because V5 may have underestimation when in heavy aerosol condition due to its algorithm is based on longer wavelength bands (Hulley and Hook, 2009; Wan, 2008). Night and day images are averaged to represent the overall monthly temperature of a watershed. Additionally, the monthly change of LSTs (ΔLST) was calculated by subtracting the previous month's LST from the current month's LST. The data are available from NASA Earthdata Search. (URL: https://search.earthdata.nasa.gov/ ;date accessed: 31 August 2017).

Vegetation Index (VI)

Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation index product (MOD13A3: MODIS/Terra Vegetation Indices Monthly L3 Global 1 km SIN Grid V006) was selected for VI would be related to ET. MODIS VI product provides a 16-day composite of monthly normalized vegetation index (NDVI) and covers global in 1 km × 1 km spatial resolution (Didan, 2015). Since the coverage of its single scene is not large enough for the largest study area (UMRB), four scenes (h10v04, h10v05, h11v04, h11v05) were used. Data are available from NASA Earth Data Search. (URL: https://search.earthdata.nasa.gov/ ; date accessed: 28 August 2017).

Streamflow (Q)

The streamflow gauging data from the outlets of study areas (USGS 0587450, USGS 05586100 and USGS 05484500) were collected from the National Water Information System (NWIS). The unit of data is converted from [ft^3/s] to [m^3/s] to be consistent with the SI unit. The data were downloaded from USGS NWIS website (URL: https://waterdata.usgs.gov/nwis ; date accessed: 9 November 2018). Streamflow data were used for both training and testing stage.

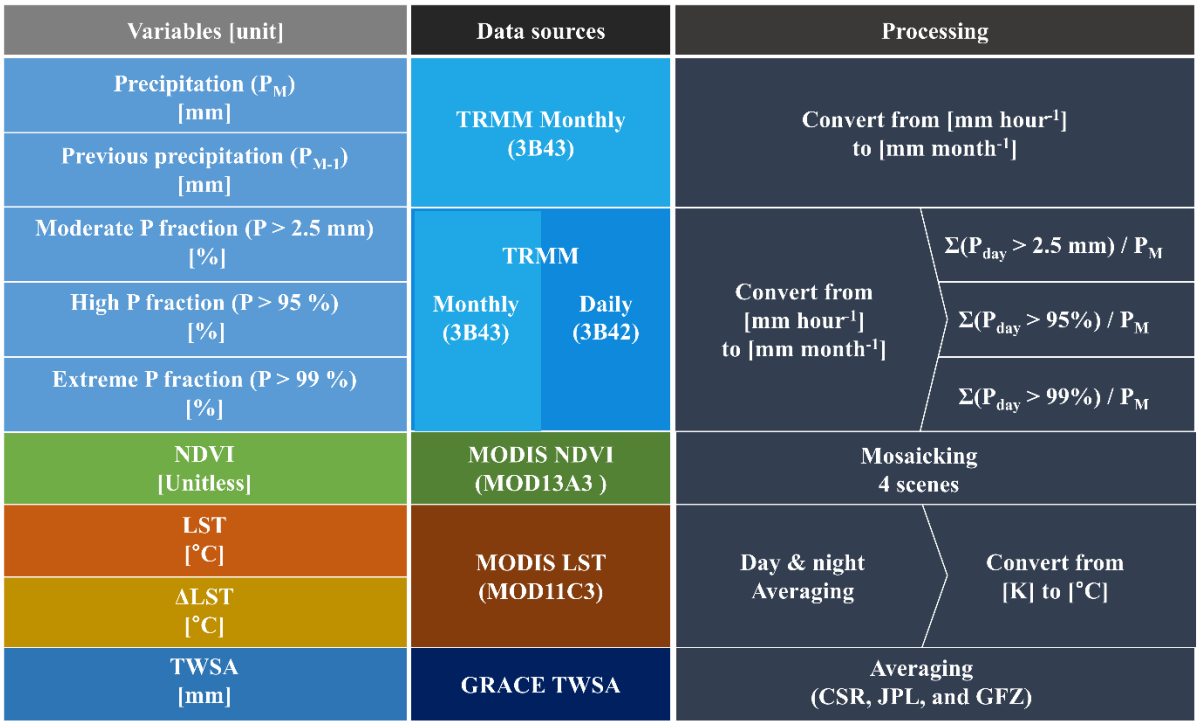


Figure 3. Summary of data sources and processing applied on the dataset used in the MLM. Processed variables are resampled according to HUs of each study site.

Model Design

Boosted Regression Tree (BRT)

The estimation of monthly streamflow and evaluation of the predictor importance were accomplished using BRT, also known as gradient boosting. BRT is based on a summation of many decision trees partitioning the covariant space by successive binary partitions (Breiman et al., 1984; Friedman et al., 2001). Building a decision tree is a repetitive work to find the best split variable and its split value, based on residual errors.

For example, if there is a multivariable and nonlinear relationship (Figure 4a), a decision tree can be constructed by the following sequences: the best location to divide the surface into the most distinguishable two partitions is X1 = 15 and it will be assigned as the first spit node of

the decision tree (Figure 4b and 4d). Mathematically, this means the residual error between the original surface and newly create surface (Figure 4b) is minimum. And then, the next split variable and its value can be found by considering both parts (X1 > 15 and X1 ≦ 15) of the original surface. The same iterative works would be conducted until predefined residual error level is accomplished or the maximum number of split nodes is reached (Figure 4c and 4d). Note that the final fitting surface may still have a residual error, that can be reduced by the boosting process.

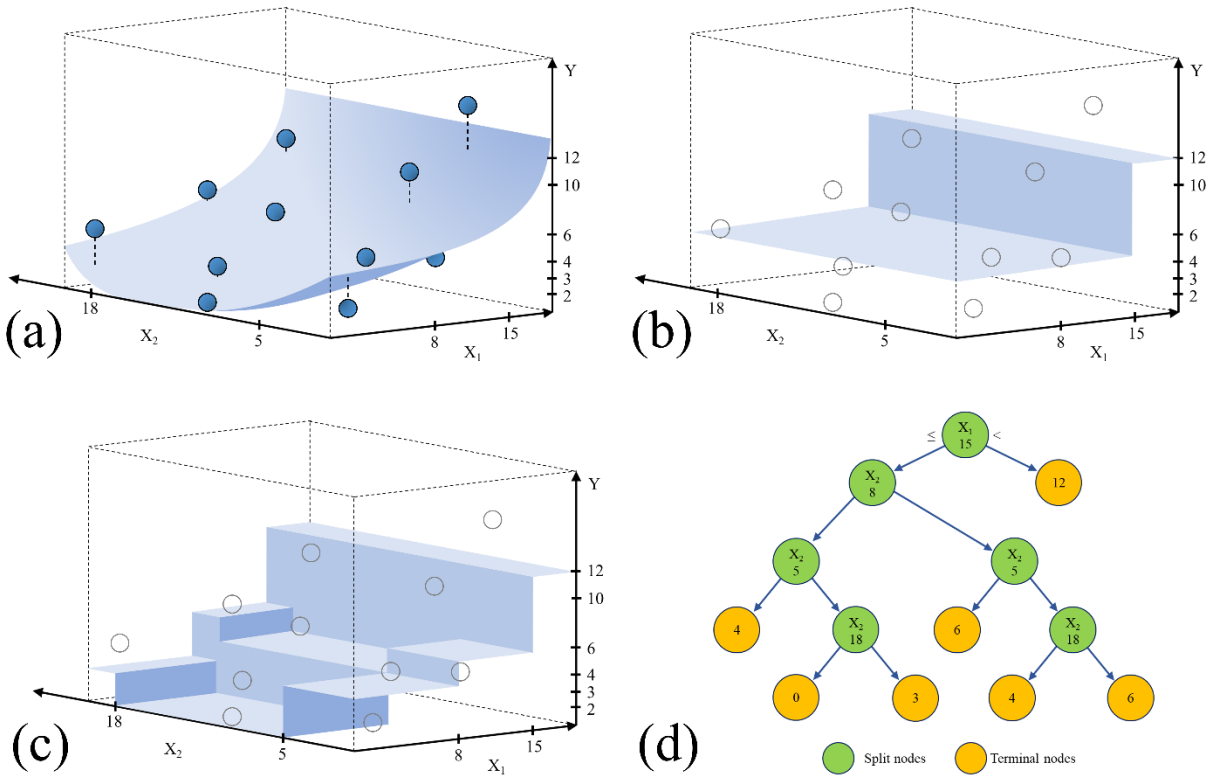


Figure 4. Decision tree processes (a) showing an example of covariant space where Y (predictand or target) is explained by X_1 and X_2 (predictors). The covariant space can be approximated by rectangular spaces (b) and (c), each representing the first node and entire nodes of the decision tree (d).

Boosting is based on the observation that finding many rough rules can be a lot easier than finding a single, highly accurate prediction rule (Schapire, 2003). In boosting, weak models (decision trees) are fitted iteratively to the residual of training data from the previous models, thereby producing a sequence of weak classifiers (Elith et al., 2008; Friedman et al., 2001). Figure 5 shows a schematic diagram of the boosting process in the BRT. The number of decision trees in boosting is determined by predefined target (goal) residual error and learning rate (residual decrement in each decision tree step).

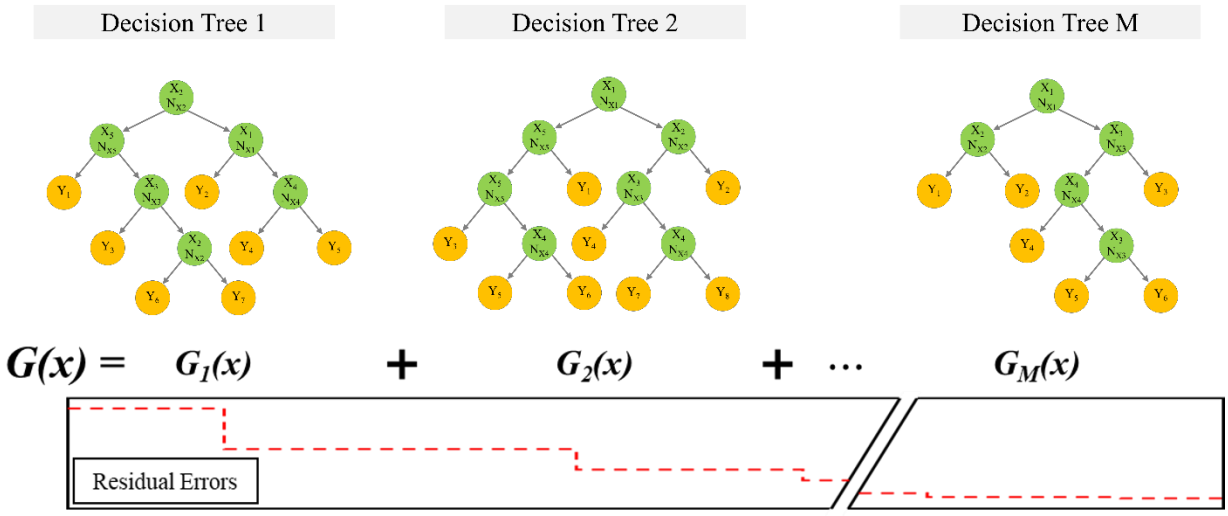
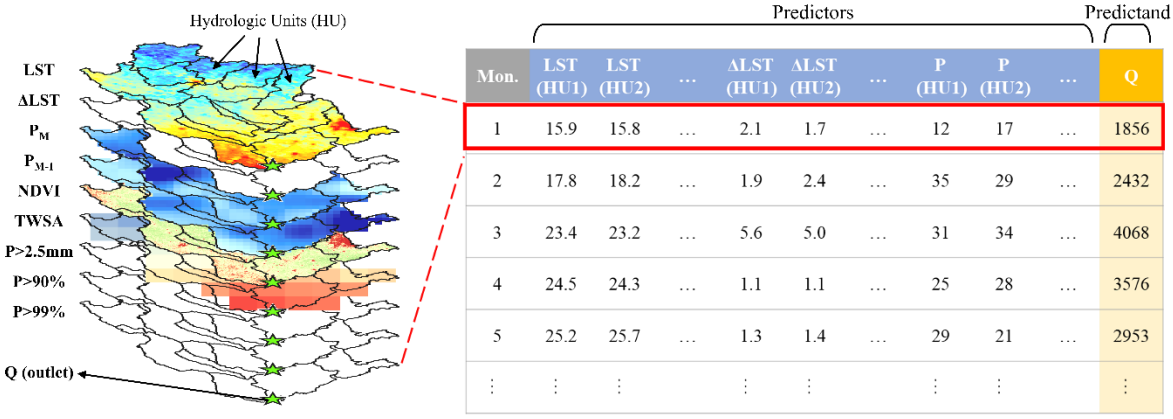


Figure 5. A schematic diagram of the boosting process in the Boosted Regression Tree (BRT) method.

Training Design

To mimic the relationship between watershed characteristics and streamflow, remote sensing data were aggregated for each subwatershed (smaller hydrologic units (HU)) in the basin and used as predictors (Figure 6). For example, if a watershed has 15 HUs, data for nine predictor variables (see Figure 3) were extracted for each HUs, the total number of predictor variables for that watershed is the number of HUs multiplied by the number of variables (15 × 9).

The scale of HUs in each watershed were chosen appropriately considering the size of the watershed (basin) (Figure 1; Table 1, 2, and 3).



Mon.	LST (HU1)	LST (HU2)	…	ΔLST (HU1)	ΔLST (HU2)	…	P (HU1)	P (HU2)	…	Q
1	15.9	15.8	…	2.1	1.7	…	12	17	…	1856
2	17.8	18.2	…	1.9	2.4	…	35	29	…	2432
3	23.4	23.2	…	5.6	5.0	…	31	34	…	4068
4	24.5	24.3	…	1.1	1.1	…	25	28	…	3576
5	25.2	25.7	…	1.3	1.4	…	29	21	…	2953
⋮	⋮	⋮		⋮	⋮		⋮	⋮		⋮

Figure 6. A conceptual diagram of the design of training data.

In this study, LS Boost (least-squares regression) algorithm was used to assess residual errors while boosting and K-fold method was used for cross-validation in BRT modeling. The K-fold method is a cross-validation method (including holdout and Leave-one-out) used to evaluate the performance of a newly-trained model by considering new dataset during the modeling process. The K-fold cross-validation method is a better choice when the size of the training dataset is small. This method randomly divides a dataset into [K] partitions and the individual partitions are used [K-1] times as a part of training data, [1] time as a cross-validation data (Figure 7). Consequently, the final model can cover the entire provided observations in the training data. To find the best models for each study sites, a total of 315 (105 iterations for each watershed (basin)) modeling iterations were run by changing the K-fold factor from 2 to 36.

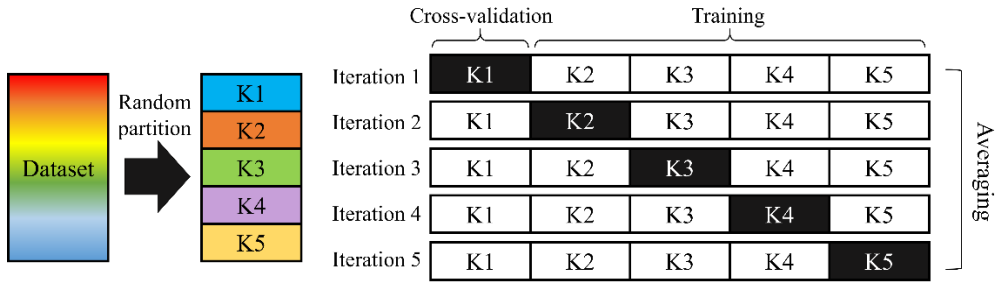


Figure 7. A schematic diagram of K-fold method.

Figure 8 shows a summary of workflow of this study. After the training stage, the best model for each study sites were selected and tested by independent remote sensing and streamflow data. Performances of MLM, implications of the results, and its applicability are discussed in the next section.

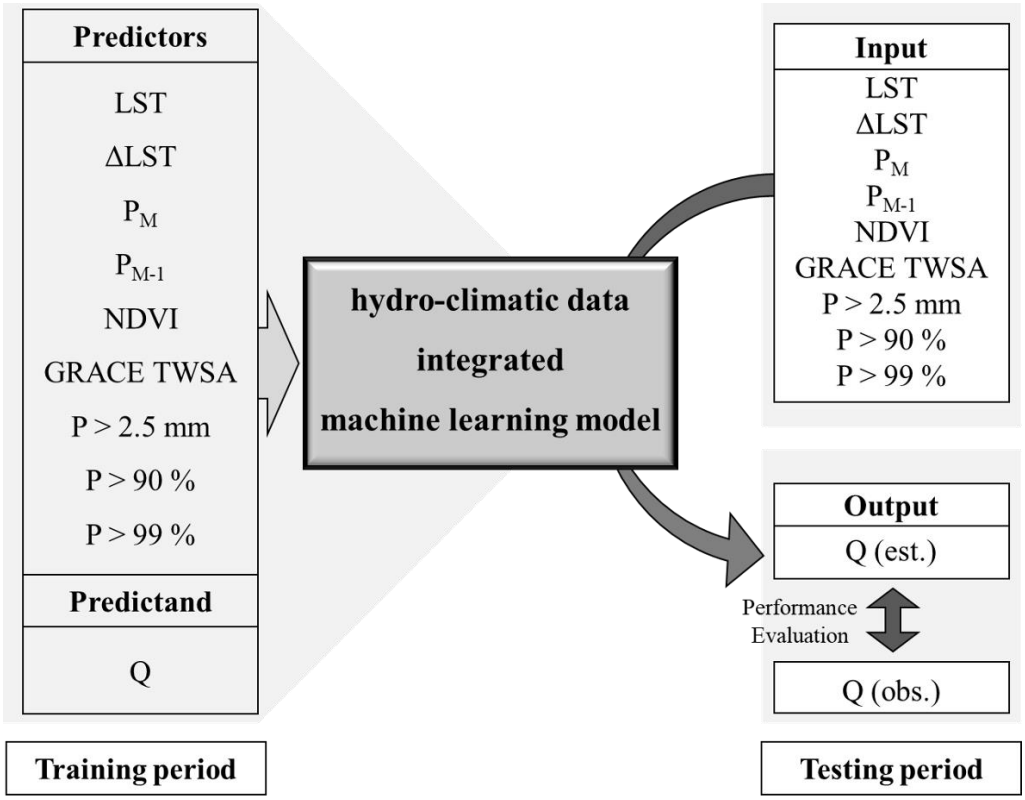


Figure 8. Workflow and data requirements for training and testing in this study.

Performance Evaluation

The performance of the MLM was evaluated statistically using coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE), percent bias (PBIAS), and mean absolute error (MAE) by comparing model estimated with the observed streamflow data. For the equations below, y_t is the observed value at time t, f_t is the simulated (model estimated) value at time t, $\bar{y}$ is the mean of observed values, and $\bar{f}$ is the mean of simulated values for the entire evaluation period (T).The R^2 is computed as shown in Equation 2; it indicates the portion of the variance of data that can be explained by the model, its value ranges from 0 (no explanation) to 1 (the model explains 100 % of the observed data).

$$R^2 = \left\{ \frac{\sum_{t=1}^{T}(y_t - \bar{y})(f_t - \bar{f})}{[\sum_{t=1}^{T}(y_t - \bar{y})^2]^{0.5}\left[\sum_{t=1}^{T}(f_t - \bar{f})^2\right]^{0.5}} \right\}^2 \quad (2)$$

The NSE (Equation 3) indicates how a scatterplot of observed versus simulated data well fits to the 1:1 line; its value ranges from - ∞ to 1, and NSE values close to 1 denote good model performance (Nash and Sutcliffe, 1970).

$$NSE = 1.0 - \sum_{t=1}^{T} \frac{(y_t - f_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \quad (3)$$

The PBIAS (Equation 4) measures the tendency to have larger or smaller model estimation than the observed data (Moriasi et al., 2007). The optimal value is 0.0, negative value indicates overestimation, and positive value indicates underestimation by the model.

$$PBIAS = \left\{\frac{\sum_{t=1}^{T}(y_t - f_t)}{\sum_{t=1}^{T}(y_t)}\right\} * 100 \quad (4)$$

The MAE is average of difference between simulated and observed values. MAE uses the same unit of the data being used; hence, it helps to understand the scale of error directly.

$$MAE = \frac{\sum_{t=1}^{T}(y_t - f_t)}{T} \quad (5)$$

The performance metrics from this study using MLM were compared with performance metrics of previous studies conducted in the watersheds using process-based hydrologic modeling approaches. For the Upper Mississippi Basin, a SWAT model based study conducted by Jha et al. (2006) was used. In the study, they modelled UMRB streamflow at daily, monthly, and annual scale using land use, soil type, topography (digital elevation model), daily precipitation, maximum/minimum air temperature, solar radiation, wind speed, and relative humidity data according to 8-digit HU level. They tested the model using data from 1988 to 1997, the test results for monthly streamflow data showed R^2, NSE, and PBIAS of 0.82, 0.81, and 3.9 %, respectively.

Similarly, for the Illinois River Watershed, a SWAT model and Hydrologic and water Quality System (HAWQS) based study conducted by Yen et al. (2016) was used to assess the performance of MLM. In their study, IRW was modelled based on climate, land use, reservoirs, soil type, topography, and water usage data from HAWQS to estimate monthly streamflow, sediment yield, and total nitrogen. Their testing results using data from 1990 to 2001 demonstrated NSE and PBIAS values of 0.72 and 13.91 %, respectively.

Lastly, for the Raccoon River Watershed, a SWAT model based study conducted by Jha et al. (2007) was used. They estimated monthly and annual streamflow, sediment yield, and nitrate of the RRW based on daily precipitation, maximum/minimum air temperature, solar radiation, wind speed, relative humidity, and topography, soil type, fertilizer application rate, land use, and livestock distributions. The testing results using monthly streamflow data from 1993 to 2003 showed R^2 and NSE values of 0.89 and 0.88, respectively.

CHAPTER IV: RESULTS

ML-based streamflow model performances are discussed for both the training and testing periods. The model estimated streamflow fits the observed data well during the training and testing period, with slight overestimation or underestimation of streamflow in the testing period. Generally, the smaller the size of the watershed is the better the model fit the observed data during the training period, however, in the testing period, the larger watershed (e.g. UMRB) showed better performance. Predictor importance analysis that shows the efficiency of each predictor variable varies according to the watershed size. Overall, the applicability of the remote sensing-based MLM proposed in this study is good and comparable, perhaps better in some instances, compared to the process-based approaches.

The Effect of Training Data Partitioning

To find the optimum value that provides the least error, for K-fold cross-validation, a total of 105 iterative modeling (3 iterations for each K number) were run by changing K value from 2 to 36 for each study site. NSE, R^2, PBIAS, and MAE were calculated and used to find the optimum K number. Figure 9 shows the effect of K-fold numbers for each study site in terms of NSE. Overall, the effect of K-fold numbers seem to be not significant on the performance of streamflow modeling. However, there is a relationship between the pattern of NSE, K-fold numbers and the size of the watersheds (basins). For example, in the RRW MLMs, NSE values for training period are not stable compared to IRW and UMRB (Figure 9; blue semi-dotted lines). The larger fluctuations of NSE values for training in RRW implies the variables used are not sufficiently representing the behaviour of the streamflow at the outlet. This may due to the size of the RRW where the efficiency of GRACE TWSA predicting streamflow is much less due to its coarse spatial resolution. In addition, the rapid streamflow responses of RRW to short-term

rainfall events may not be well captured at a monthly scale (the temporal resolution of this study). Further, the contributions of the ET-related variables could become smaller since time of concentration of the RRW is short which could limit the effect of ET on streamflow.

The decreasing trends of mean testing NSE (magenta solid lines in Figure 9) show over-fitting effect due to fragmented training data in high K numbers. The higher K-fold number means more fragmented training and cross-validation dataset, and this can force the MLM to explain detailed pattern in shorter time-period. This can let the model contains pseudo-relationship which is not valid for external period (testing period). Rather than the K-fold number, the effect of appropriately mingled training data based on random partitioning seems to be more important to obtain a better testing result. For example, in the IRW and UMRB, the best NSE is accomplished when K = 18 and K= 35, respectively, even though the mean NSE shows a decreasing trend.

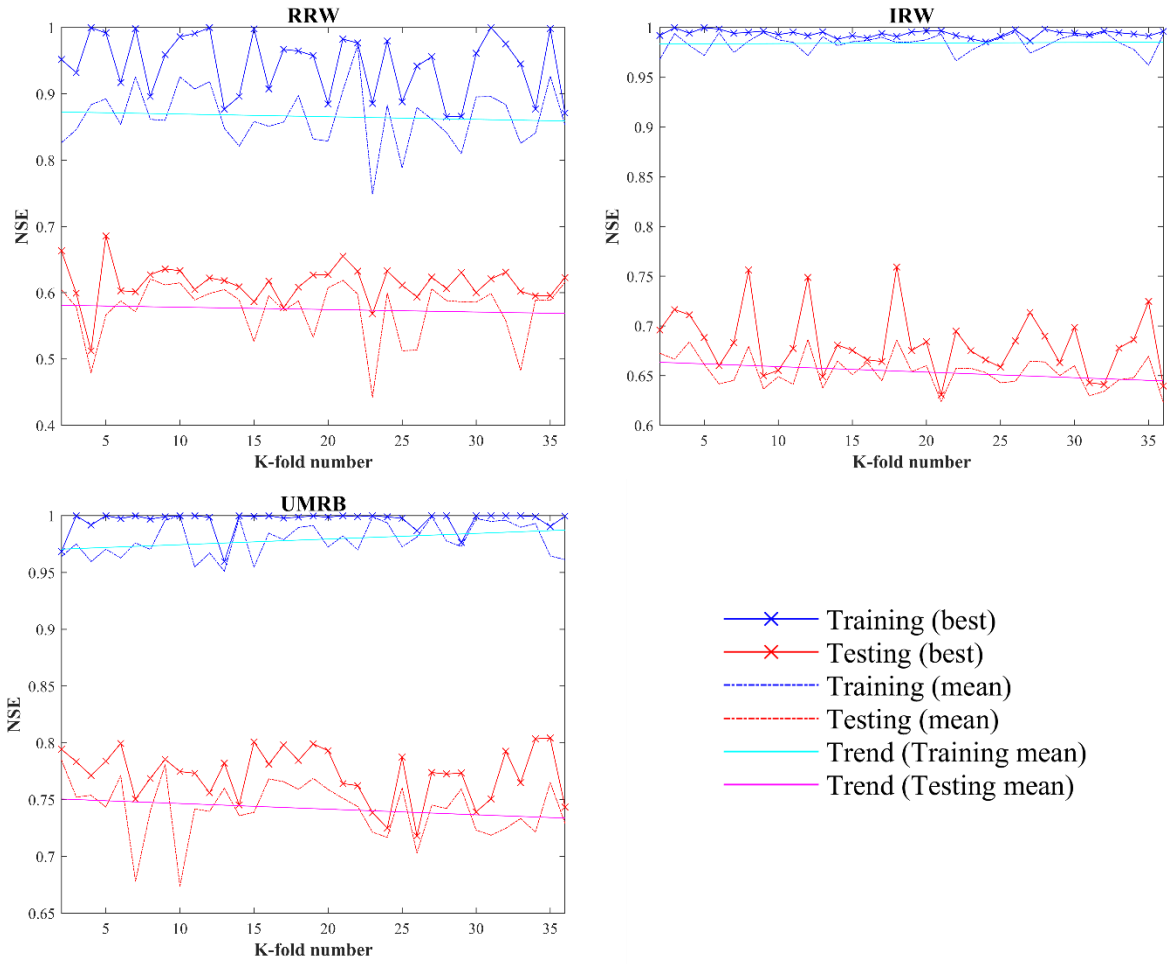


Figure 9. Streamflow model performances according to K-fold numbers. Blue and red lines indicate NSE of training period and testing period, respectively (solid line: best model; semi-dotted line: mean of three models). Cyan and magenta lines indicate trend line of mean NSE in the training and testing periods.

Streamflow Modeling

The best three MLMs representing the study sites were selected based on testing performances from the total of 315 trained models (Table 4). Overall, the MLMs simulated streamflow well fitted the observed streamflow during training period. Figure 10 demonstrates scatterplots of observed streamflow versus simulated streamflow for the RRW, IRW, and UMRB in the training period and the testing period. For the training period, the scatterplots show the

model simulated streamflow explained the observed data better for the small-sized watershed (the RRW) compared to the larger basin (the UMRB). All the data are plotted close to the 1:1 line for both low flow and high flow conditions. However, for the testing period, watershed size has an opposite effect on model performance (Figure 10b, 10d, and 10f). The MLMs simulated well for larger watersheds such as the UMRB. We presumed this is due to the limitations of time scale (monthly) and spatial resolution (e.g., GRACE TWSA) of the predictor variables at small-sized watershed. Furthermore, in the smaller watersheds, the effect of precipitation could not be well-captured at a monthly time scale due to its rapid streamflow responses to precipitation.

Table 4

The performance metrics for training and testing.

			Training				Testing			
Site	Mean Q [m^3/s]	K-fold	NSE	R^2	PBIAS [%]	MAE [m^3/s]	NSE	R^2	PBIAS [%]	MAE [m^3/s]
RRW	71.16	5	0.9919	0.9935	5.56×10^{-5}	5.48	0.6856	0.7130	5.81	23.67
IRW	760.60	18	0.9796	0.9813	1.75×10^{-4}	61.45	0.7593	0.7996	-10.96	186.28
UMRB	3791.16	35	0.9575	0.9601	2.05×10^{-8}	383.46	0.8042	0.8238	-9.28	773.92

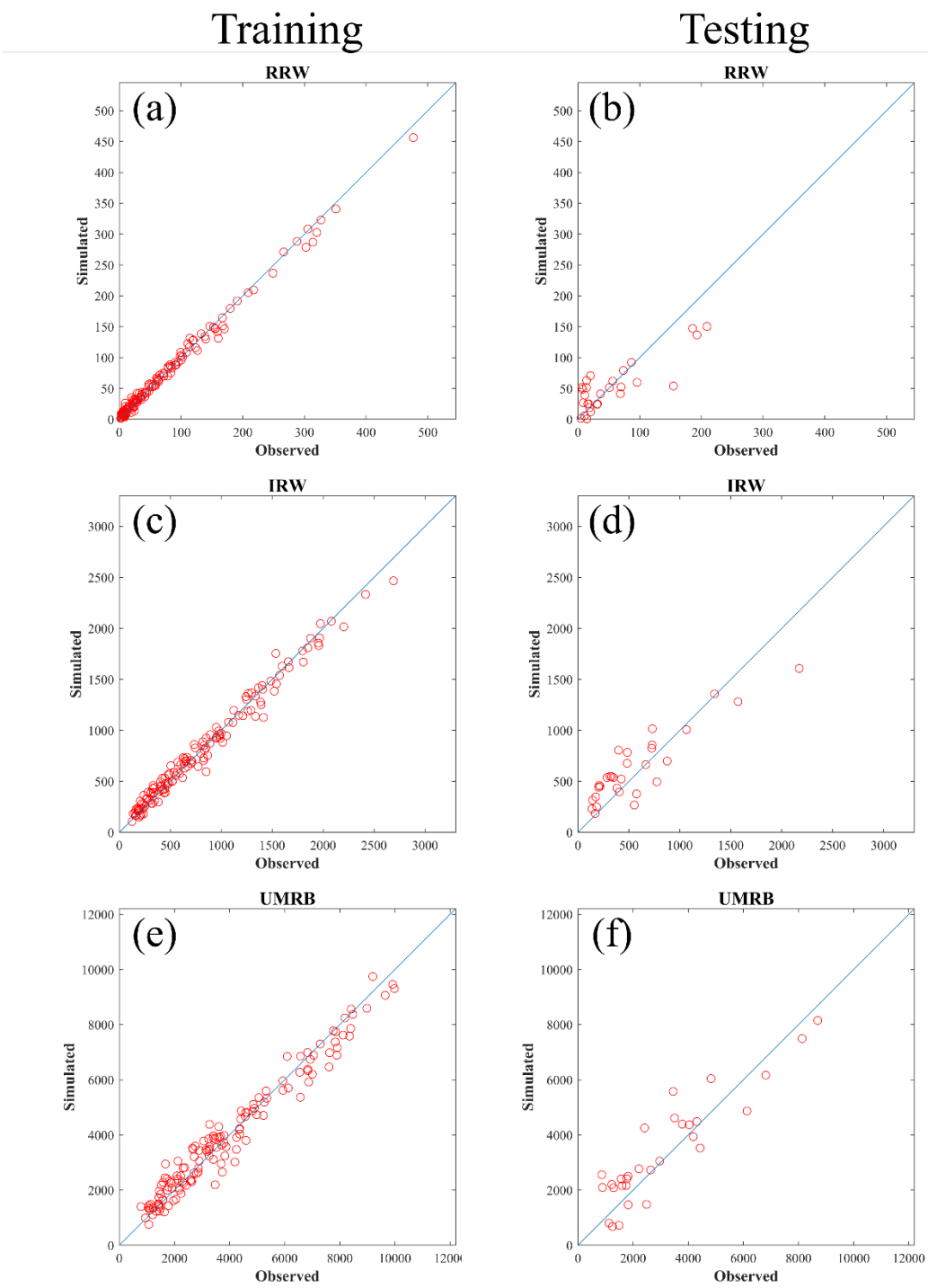


Figure 10. Scatterplots of model simulated vs. observed streamflow for training (a) RRW, (c) IRW, and (e) UMRB and testing (b) RRW, (d) IRW, and (f) UMRB.

Figure 11 shows the timeseries plots of observed vs. simulated streamflow for each watershed. MLM for the smaller watershed (RRW) estimates streamflow in the training period perfectly well, however, in the testing period; there are relatively large underestimations and overestimations with the poorly predicted pick flow. This implies MLM of RRW based on monthly variables couldn't capture some short-term events that extremely affect streamflow. Contrary, the MLM of UMRB shows relatively constant patterns between simulated and observed streamflow throughout the study period. However, it also has slight overestimations under low streamflow conditions (winter) while underestimations are occurred under high streamflow conditions (spring-summer).

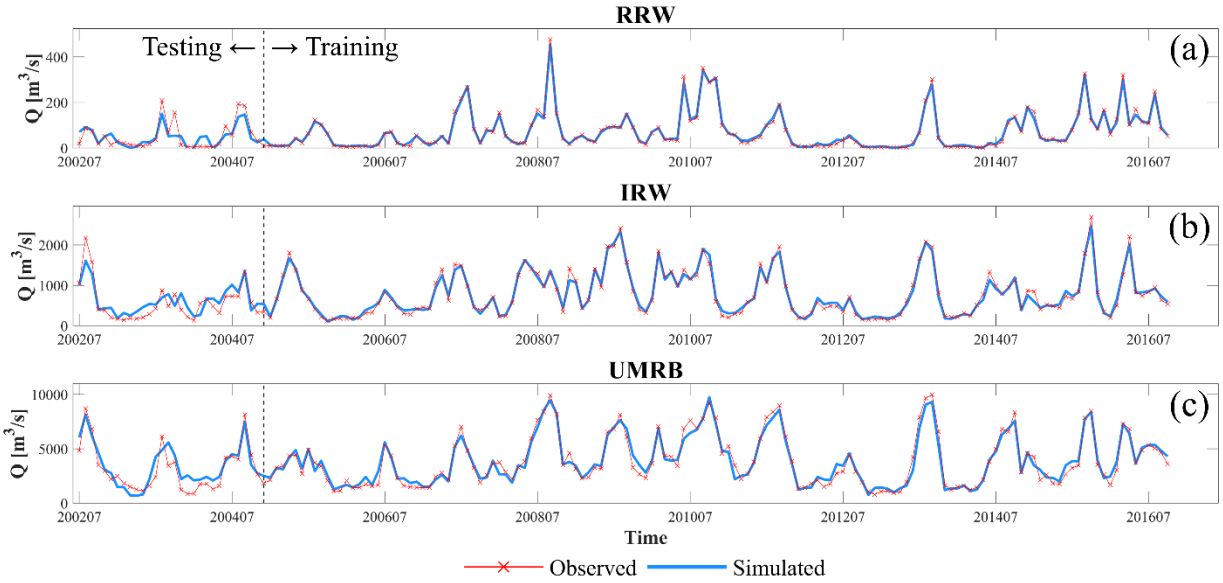


Figure 11. Time series plots of model simulated streamflow and observed streamflow for the entire study period for (a) UMRB, (b) IRW, and (c) RRW.

Variable Contributions

The BRT model provides Predictor Importance (PI) analysis in terms of mean squared error (MSE) by considering the role of individual predictor variable in splitting and trees in the BRT model. In order to bring the calculated PI to the same scale, a relative PI was calculated and

plotted as percentage. Figure 12 shows relative PI of each watershed (basin) as a percentile. In the RRW, the most important predictor was P (42.8 %), and the other predictors were GRACE TWSA (28.3 %), P_{M-1} (12.2 %), ΔLST (5.1 %), P > 90% (4.7 %), P > 2.5 mm (3.9 %), P > 99 % (1.5 %), NDVI (1.3 %), and LST (0.2 %), in that order. In the IRW, the most important predictor was GRACE TWSA (51.2 %), and the others were P_{M-1} (27.5 %), P (11.8 %), ΔLST (6.2 %), NDVI (1.8 %), P > 90 % (0.6 %), P > 2.5 mm (0.5 %), LST (0.2 %), and P > 99 % (0.1 %). In the UMRB, the most important predictor was GRACE TWSA (39.6 %), and the others were P_{M-1} (23.6 %), ΔLST (20.0 %), P (15.2 %), NDVI (0.6 %), P > 2.5 mm (0.3 %), P > 90 % (0.3 %), LST (0.2 %), and P > 99 % (0.1 %). Overall, the relative PI demonstrated TWSA, P, and P_{M-1} are the most important variables.

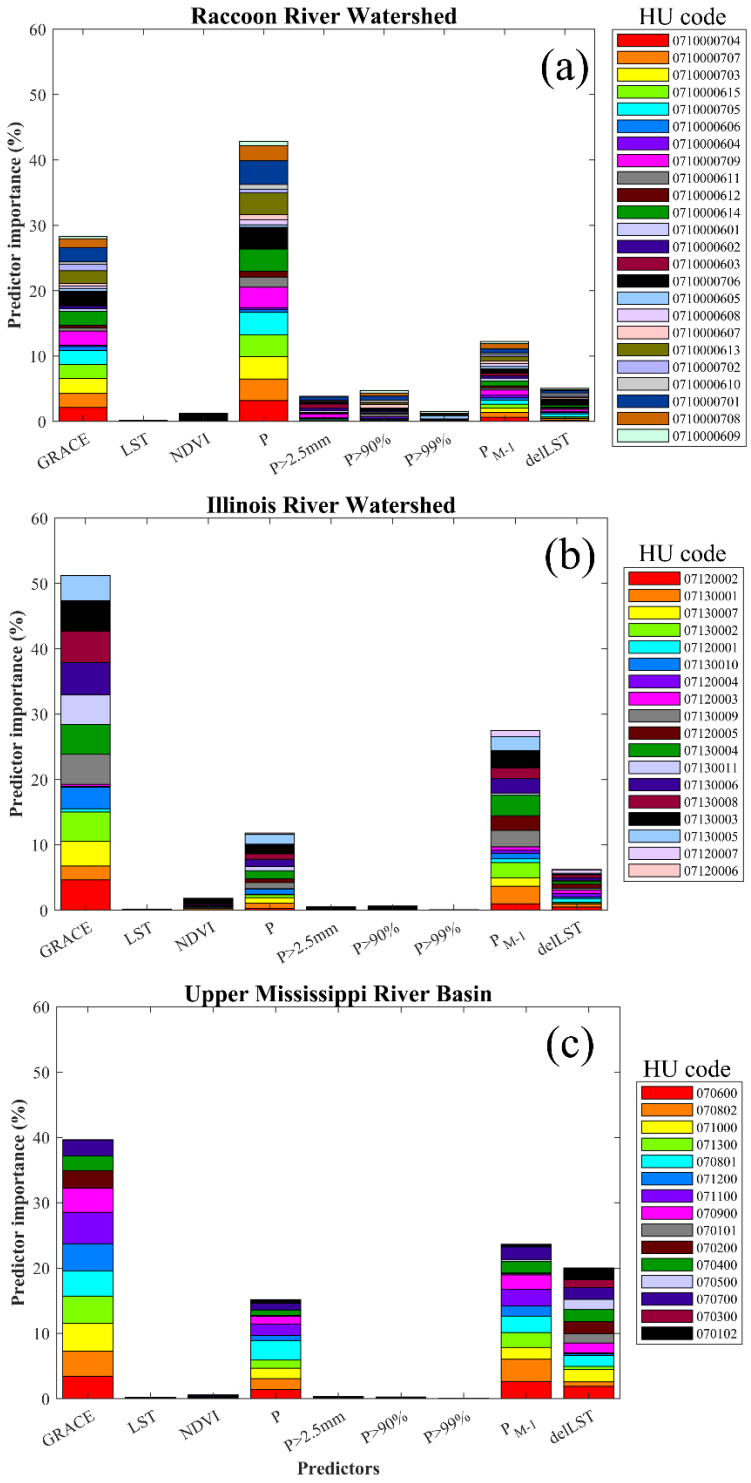


Figure 12. Relative importance of predictor variables for each watershed (a) RRW, (b) IRW, and (c) UMRB. Each colored part in the bars indicates the contributions of the hydrologic units (HU).

Comparison with Process-based Models

The testing performances of the MLM in this study were compared with process-based studies conducted on the same study sites and gauging stations (Table 5). Since the studies have variable testing periods, effort was made to match the testing period of this study (from April 2002 to September 2004; 30 months) with the previous studies. Even though a direct comparison is impossible, some cons and pros between MLM and process-based approach could be outlined from the performance comparison. The comparison indicates that the process-based approach conducted in a small watershed (e.g., RRW) achieves better performance metrics compared to the MLM in this study. Process-based models are more cost effective and manageable for small watersheds, thus, uncertainties from the input data, processing, and calibration are relatively low. However, for larger watersheds (basins) (e.g., IRW and UMRB), the environment is highly heterogeneous and process-based modeling is less manageable, making it more difficult to build an efficient process-based model. This is due to the effort of data collection that may extensively increase and due to uncertainties from data, processing, and calibration that become considerable. Conversely, remote sensing data-based MLM introduced in this study showed better performances in the large watershed (basin), which was developed using publicly accessible data, and a cost-effective and manageable method. Considering the streamflow model performance of the IRW and UMRB, the remote sensing-based MLM looks competitive against the conventional process-based approaches.

Table 5

Comparison of streamflow estimation performances between process-based model (SWAT) and MLM for each study sites.

		Process-based model				MLM (this study)		
Site	Author	Modeling method	Testing period	NSE	PBIAS	Testing period	NSE	PBIAS
RRW	Jha et al., 2007	SWAT	1993-2003	0.88	-	2002-2004	0.69	5.81 %
IRW	Yen et al., 2016	SWAT	1990-2001	0.72	13.91 %	2002-2004	0.76	-10.96 %
UMRB	Jha et al., 2006	SWAT	1988-1997	0.81	3.9 %	2002-2004	0.80	-9.28 %

Note: The exact testing period of MLM is from April 2002 to September 2004 (30 months).

CHAPTER V: DISCUSSION AND CONCLUSIONS

Discussion

Monthly streamflow has modeled using a remote sensing-based MLM approach in this study. The MLM was applied to different-sized watersheds (the RRW and the IRW) and basin (the UMRB). For the model training phase, semi-distributed approach was applied to capture the contributions of each subwatershed, and the iterative K-fold cross-validation was conducted to find the optimal K number. K-fold number analysis showed the K number is insignificant in the performance of streamflow modeling.

In the testing period, the RRW model poorly predicted pick flow and showed larger under/overestimations, while the IRW and the UMRB models showed better fits to the observed streamflow. Regarding the testing results, the remote sensing-based MLM perform better for both the IRW and UMRB compared to the smaller watershed-the RRW. This could be due to the coarse spatial resolution of the GRACE data and coarse temporal resolution (monthly) used in this study. This is especially true for small watershed like RRW where time of concentration of runoff may be shorter. The size of subwatersheds (HUs) in the RRW are much smaller than a GRACE pixel. Thus, many HUs in the RRW have the same value in a month and TWSA contributions of each HUs cannot be well captured during modeling process. Moreover, TWSA may contaminated by the signals from out of the watershed, which may become significant for the smaller watersheds. Monthly datasets such as GRACE TWSA and TRMM may not sufficiently capture storm events that are important to streamflow responses, this also may decrease streamflow estimation performance of the MLM.

Relative predictor importance (PI) results showed GRACE TWSA, P, and P_{M-1} are the most important predictor variables. GRACE TWSA plays the most important role in the IRW

and the UMRB, and the second most important role in the RRW. This emphasizes the effectiveness of GRACE data for streamflow modeling in terms of baseflow estimation. The relative importance between these two variables depends on the watershed (basin) size. For example, GRACE TWSA plays the most important role in the IRW and UMRB models, however, P is the most important in the RRW model. This is because GRACE TWSA has a coarse spatial resolution. Thus, the limitation of the spatial resolution of GRACE TWSA could be one of the reasons that the MLM performance in small watershed (e.g., RRW) is relatively weaker. The contribution of the previous month's precipitation (P_{M-1}) tends to increase as the watershed (basin) size increases. This is consistent with expected longer time of concentration in larger watershed (basin) where precipitation in the upstream area will take longer time to reach the outlet as the watershed size increases. A relatively higher predictor importance of derivative P data such as fraction of moderate, high, and extreme precipitation ($P > 2.5$ mm, $P > 90$ %, and $P > 99$ %) in the RRW implies streamflow of small watershed is more sensitive to the magnitude of extreme precipitation events as well as the total amount of precipitation.

LST and ΔLST were expected to explain the timing of snow melting, ET processes, and other variables (e.g., soil moisture, wind speed, humidity) indirectly. The relative PI shows ΔLST seems more important predictor variable than LST. ΔLST is significant in that large magnitude of ΔLST value indicates season change. This is important in simulating seasonally induced streamflow such as the increase in discharge in early spring driven by snow melting or low flow in late summer caused by high ET. Furthermore, ΔLST overwhelms NDVI, which is expected to account for the ET processes, an important component of water balance. The low relative PIs of LST and NDVI indicate the variables have minimum effect on streamflow, or they

had no chance to be used in a decision tree because the other predictors such as ΔLST explain most of streamflow responses.

Location of subwatersheds (HUs) influence streamflow simulation by the MLMs. The relative PI predictors aggregated over the subwatersheds (HUs) depicted a few subwatersheds with high relative PI values (Figure 13). Generally, HUs located close to the trunk stream and in the middle part of the watershed (basin) tend to have larger relative importance compared to the HUs located upstream or tributaries. For example, Otter Creek-North Raccoon River (0710000614) in RRW, Lower Illinois-Senachwine Lake (07130001) in IRW, and Upper Mississippi-Maquoketa-Plum (070600) in UMRB. This indicates some HUs are to have variables that estimate streamflow pattern at the outlet more than others. Therefore, HUs with low relative PI imply that the remote sensing variables in the area had little or no chance of being selected to explain streamflow responses compared to the dominant variables of the more representative HUs during regression tree construction. For example, in the UMRB, the Lower Illinois (071300) has very low relative PI because most of the outlet's streamflow pattern is explained by the Upper Illinois (071200). Likewise, in the RRW, the Upper South Raccoon River (0710000704) and its adjacent subwatersheds (0710000703 and 0710000707) have high relative PI because those belong to the South Raccoon River Watershed, which is one of the main tributary of the Raccoon River and it is separated from the main stream of the RRW.

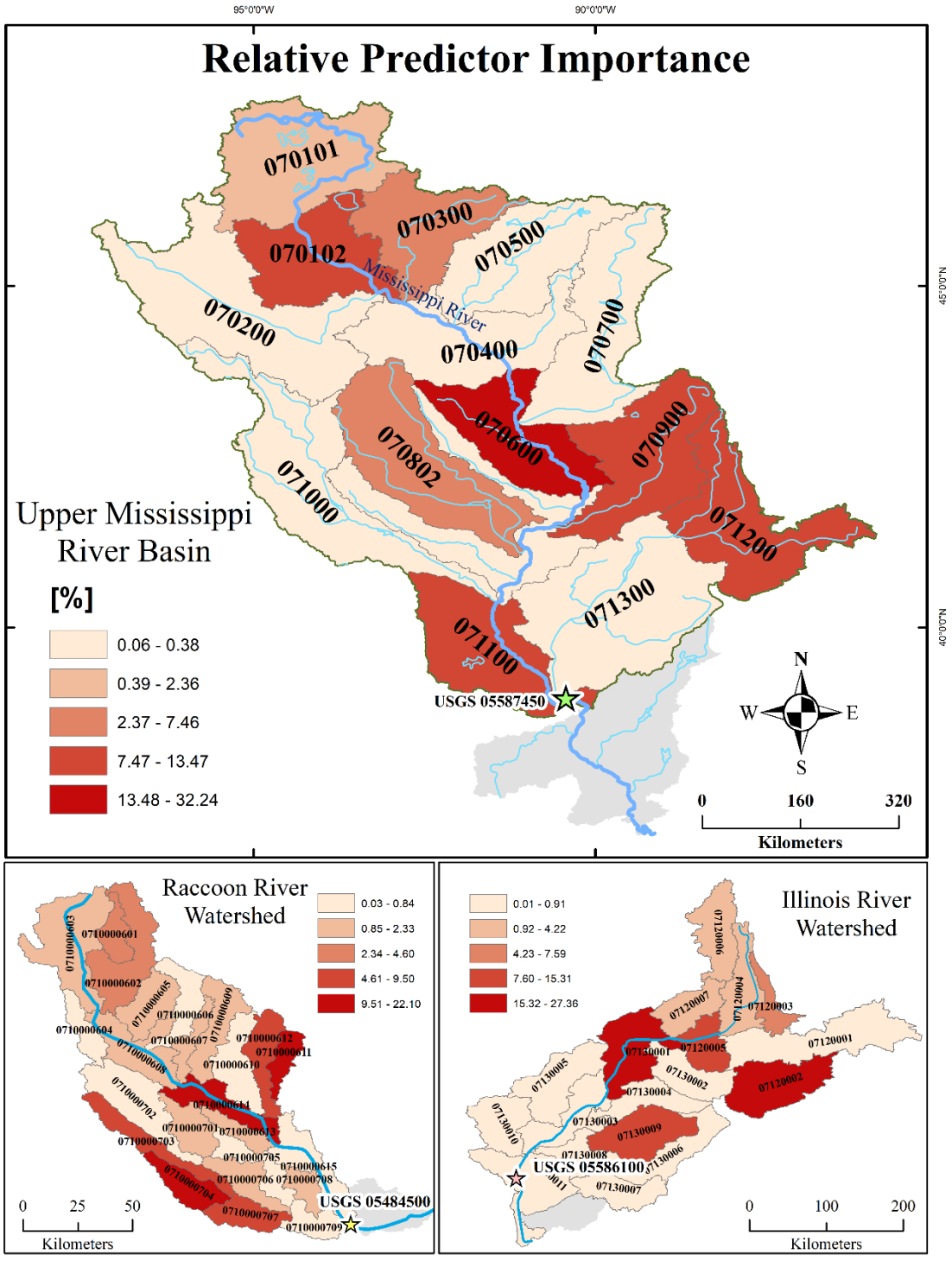


Figure 13. Map showing the relative contribution of sub-watersheds in simulating the MLMs for each basin.

The scales of subwatersheds (HU-level) used in this study for UMRB, IRW, and RRW with HU-6, HU-8, and HU-10, respectively, are larger than those in process-based studies. The HU scales were chosen to keep appropriate numbers of predictor variables for each watershed, however, this may one of the reasons that the MLMs of UMRB and RRW have lower testing performance than the process-based models. Thus, training of MLM based on the finer level of HUs need to be tested. Moreover, comparison of testing performances between process-based approaches and MLMs in this study is not based on the same period, this also need to be improved.

The main objective of this study is to evaluate the effectiveness of the remote sensing-based MLM. The results demonstrate the approach could be effectively applied to hydrologic studies to provide streamflow estimation. The method has advantages to simulate streamflow for large watersheds or basins without any in-situ-based surface information. However, some limitations should be considered before applying the remote sensing-based MLM approach to other areas: (1) in-situ streamflow data are necessary for training dataset, (2) temporal applicability is limited according to available remote sensing data such as GRACE, and (3) the PI of MLM is not the reflection of the real nature of a watershed.

The remote sensing-based MLM also have a potential to be applied to flood predictions. However, some limitations need to be improved: (1) variables in finer temporal resolution such as daily data should be used while model construction since streamflow responses to extreme precipitation would take short time, and (2) GRACE TWSA may be replaced by antecedent precipitation datasets, which are expected to improve streamflow estimation performance of MLM in terms of both temporal and spatial resolution.

Conclusion

The remote sensing-based MLM demonstrated it performs well to estimate monthly streamflow. The effectiveness of remote sensing-based MLM depends on the watershed size and can be limited in a small watershed due to the limited spatial resolution of remote sensing data such as GRACE. However, as Seyoum et al. (2019) showed, downscaling of GRACE data is possible and better results can be expected in small watersheds if the downscaled TWSA is used.

The following important conclusions are drawn from this study:

- The remote sensing-based MLM has advantage in estimating streamflow for large (> 75,000 km^2) watersheds (basin) compared to process-based models in terms of data acquisition, computing resources, and testing performance.
- Distributed approach and Predictor Importance (PI) analysis can demonstrate the most representative subwatersheds in the streamflow responses. This information can be used to establish an efficient water resource management policy in terms of land use and water usage.
- The efficiency of the MLM is dominated by TWSA and the higher performance can be achieved if better quality of TWSA or TWSA-related data is provided.
- The importance of P_{M-1}, which is one of the TWSA-related variables, implies the aggregated antecedent precipitation conditions have the potential to be used as a supplement of GRACE TWSA. This supports the most recent machine learning-based streamflow modeling approaches such as Long Short-Term Memory (LSTM) network (Kratzert et al., 2018).

- ΔLST can be used in MLM to reflect indirectly seasonal changes (e.g., snow melting) and ET-related climatic effects (humidity and wind speed) which cannot be easily acquired by remote sensing.
- BRT model can be trained based on a limited amount of data (142 months) and performs well for monthly streamflow simulations. Thus, the model may be applicable for the basins where streamflow gauge data is limited.

The remote sensing-based MLM introduced in this study has the potential to be a supplementary and popular approach in estimating streamflow. Overall performance is comparable with process-based approaches, but with significantly less modeling effort. Remote sensing-based MLM has the potential to be a very attractive tool in estimating streamflow for watersheds that have not been sufficiently studied in hydrologic and geologic manner.

REFERENCES

Ahmad, S., Kalra, A., and Stephen, H., 2010, Estimating soil moisture using remote sensing data: A machine learning approach: Advances in Water Resources, v. 33, no. 1, p. 69-80.

Alexander, L., Zhang, X., Peterson, T., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., and Rahimzadeh, F., 2006, Global observed changes in daily climate extremes of temperature and precipitation: Journal of Geophysical Research: Atmospheres, v. 111, no. D5.

Arnold, J. G., Muttiah, R. S., Srinivasan, R., and Allen, P. M., 2000, Regional estimation of base flow and groundwater recharge in the Upper Mississippi river basin: Journal of Hydrology, v. 227, no. 1-4, p. 21-40.

Bajwa, S., and Vibhava, V., 2009, A Distributed Artificial Neural Network Model for Watershed-Scale Rainfall-Runoff Modeling: Transactions of the ASABE, v. 52, no. 3, p. 813-823.

Beven, K. J., 2011, Rainfall-runoff modelling: the primer, John Wiley & Sons.

Blöschl, G., 2005, Rainfall-runoff modeling of ungauged catchments, Wiley Online Library.

Boegh, E., Poulsen, R., Butts, M., Abrahamsen, P., Dellwik, E., Hansen, S., Hasager, C. B., Ibrom, A., Loerup, J.-K., and Pilegaard, K., 2009, Remote sensing based evapotranspiration and runoff modeling of agricultural, forest and urban flux sites in Denmark: From field to macro-scale: Journal of Hydrology, v. 377, no. 3, p. 300-316.

Bourdin, D. R., Fleming, S. W., and Stull, R. B., 2012, Streamflow modelling: a primer on applications, approaches and challenges: Atmosphere-Ocean, v. 50, no. 4, p. 507-536.

Brakenridge, G. R., Nghiem, S. V., Anderson, E., and Chien, S., 2005, Space-based measurement of river runoff: EOS, Transactions American Geophysical Union, v. 86, no. 19, p. 185-188.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. J., 1984, Classification and regression trees.

Chang, W., and Chen, X., 2018, Monthly Rainfall-Runoff Modeling at Watershed Scale: A Comparative Study of Data-Driven and Theory-Driven Approaches: Water, v. 10, no. 9, p. 1116.

Chen, J. M., Chen, X., Ju, W., and Geng, X., 2005, Distributed hydrological model for mapping evapotranspiration using remote sensing inputs: Journal of Hydrology, v. 305, no. 1, p. 15-39.

Chen, X.-y., Chau, K.-w., and Wang, W.-c., 2015, A novel hybrid neural network based on continuity equation and fuzzy pattern-recognition for downstream daily river discharge forecasting: Journal of hydroinformatics, v. 17, no. 5, p. 733-744.

Chiew, F., Stewardson, M., and McMahon, T., 1993, Comparison of six rainfall-runoff modelling approaches: Journal of Hydrology, v. 147, no. 1-4, p. 1-36.

Dawson, C. W., and Wilby, R., 1998, An artificial neural network approach to rainfall-runoff modelling: Hydrological Sciences Journal, v. 43, no. 1, p. 47-66.

Demissie, M., Keefer, L., Slowikowski, J., and Stevenson, K., 2006, Evaluating the effectiveness of the Illinois River Conservation Reserve Enhancement Program in reducing sediment delivery: IAHS PUBLICATION, v. 306, p. 295.

Deo, R. C., and Şahin, M., 2016, An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland: Environmental monitoring and assessment, v. 188, no. 2, p. 90.

DHI, D., 2003, Mike-11: a modelling system for rivers and channels, reference manual: DHI–Water and Development, Horsholm, Denmark.

Didan, K., 2015, MOD13A3: MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006: NASA EOSDIS Land Processes DAAC, doi, v. 10.

Elith, J., Leathwick, J. R., and Hastie, T., 2008, A working guide to boosted regression trees: Journal of Animal Ecology, v. 77, no. 4, p. 802-813.

Friedman, J., Hastie, T., and Tibshirani, R., 2001, The elements of statistical learning, Springer series in statistics New York.

Gassman, P. W., Secchi, S., Jha, M., and Kurkalova, L., 2006, Upper Mississippi River Basin modeling system part 1: SWAT input data requirements and issues: Coastal hydrology and processes, p. 103-115.

Herschy, R. W., 2014, Streamflow measurement, CRC Press.

Hong, Y., Adler, R. F., Hossain, F., Curtis, S., and Huffman, G. J., 2007, A first approach to global runoff simulation using satellite rainfall estimation: Water Resources Research, v. 43, no. 8.

Hsu, K. l., Gupta, H. V., and Sorooshian, S., 1995, Artificial neural network modeling of the rainfall-runoff process: Water resources research, v. 31, no. 10, p. 2517-2530.

Hulley, G. C., and Hook, S. J., 2009, Intercomparison of versions 4, 4.1 and 5 of the MODIS Land Surface Temperature and Emissivity products and validation with laboratory measurements of sand samples from the Namib desert, Namibia: Remote Sensing of Environment, v. 113, no. 6, p. 1313-1318.

Illinois Climate Network, 2015, Water and Atmospheric Resources Monitoring Program: Illinois State Water Survey, 2204 Griffith Drive, Champaign, IL 61820-7495.

Jha, M., Arnold, J. G., Gassman, P. W., Giorgi, F., and Gu, R. R., 2006, CLIMATE CHHANGE SENSITIVITY ASSESSMENT ON UPPER MISSISSIPPI RIVER BASIN STREAMFLOWS USING SWAT 1: JAWRA Journal of the American Water Resources Association, v. 42, no. 4, p. 997-1015.

Jha, M., Pan, Z., Takle, E. S., and Gu, R., 2004, Impacts of climate change on streamflow in the Upper Mississippi River Basin: A regional climate model perspective: Journal of Geophysical Research: Atmospheres, v. 109, no. D9.

Jha, M. K., Gassman, P. W., and Arnold, J. G., 2007, Water quality modeling for the Raccoon River watershed using SWAT: Transactions of the ASABE, v. 50, no. 2, p. 479-493.

Khosravi, K., Mirzai, H., and Saleh, I., 2013, Assessment of Empirical Methods of Runoff Estimation by Statistical test (Case study: BanadakSadat Watershed, Yazd Province): CASRP Publishing Company.

Kratzert, F., Klotz, D., Brenner, C., and Schulz, K., 2018, Rainfall-runoff modelling using long short-term memory (lstm) networks.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J., 1998, The tropical rainfall measuring mission (TRMM) sensor package: Journal of atmospheric and oceanic technology, v. 15, no. 3, p. 809-817.

Landerer, F. W., and Swenson, S. C., 2012, Accuracy of scaled GRACE terrestrial water storage estimates: Water Resources Research, v. 48, no. 4.

Liston, G., Sud, Y., and Wood, E., 1994, Evaluating GCM land surface hydrology parameterizations by computing river discharges using a runoff routing model: Application to the Mississippi basin: Journal of Applied Meteorology, v. 33, no. 3, p. 394-405.

Liu, T., Willems, P., Feng, X. W., Li, Q., Huang, Y., Bao, A. M., Chen, X., Veroustraete, F., and Dong, Q. H., 2012, On the usefulness of remote sensing input data for spatially distributed hydrological modelling: case of the Tarim River basin in China: Hydrological Processes, v. 26, no. 3, p. 335-344.

Mahmoud, S. H., 2014, Investigation of rainfall–runoff modeling for Egypt by using remote sensing and GIS integration: Catena, v. 120, p. 111-121.

Markstrom, S. L., Regan, R. S., Hay, L. E., Viger, R. J., Webb, R. M., Payn, R. A., and LaFontaine, J. H., 2015, PRMS-IV, the precipitation-runoff modeling system, version 4: US Geological Survey Techniques and Methods, no. 6-B7.

Maurer, E. P., and Lettenmaier, D. P., 2003, Predictability of seasonal runoff in the Mississippi River basin: Journal of Geophysical Research: Atmospheres, v. 108, no. D16.

Melesse, A. M., and Graham, W. D., 2004, Storm runoff prediction based on a spatially distributed travel time method utilizing remote sensing and GIS: JAWRA Journal of the American Water Resources Association, v. 40, no. 4, p. 863-879.

Milly, P. C., Dunne, K. A., and Vecchia, A. V., 2005, Global pattern of trends in streamflow and water availability in a changing climate: Nature, v. 438, no. 7066, p. 347-350.

Minns, A., and Hall, M., 1996, Artificial neural networks as rainfall-runoff models: Hydrological sciences journal, v. 41, no. 3, p. 399-417.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L., 2007, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations: Transactions of the ASABE, v. 50, no. 3, p. 885-900.

Mutlu, E., Chaubey, I., Hexmoor, H., and Bajwa, S., 2008, Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed: Hydrological processes, v. 22, no. 26, p. 5097-5106.

Nash, J. E., and Sutcliffe, J. V., 1970, River flow forecasting through conceptual models part I—A discussion of principles: Journal of hydrology, v. 10, no. 3, p. 282-290.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., and Williams, J. R., 2011, Soil and water assessment tool theoretical documentation version 2009: Texas Water Resources Institute.

NRCS, 2010, Assessment of the effects of conservation practices on cultivated cropland in the upper Mississippi River basin: US Department of Agriculture, Natural Resources Conservation Service.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., and Mathevet, T., 2007, Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models: Hydrological sciences journal, v. 52, no. 1, p. 131-151.

Rasouli, K., Hsieh, W. W., and Cannon, A. J., 2012, Daily streamflow forecasting by machine learning methods with weather and climate inputs: Journal of Hydrology, v. 414, p. 284-293.

Riad, S., Mania, J., Bouchaou, L., and Najjar, Y., 2004, Rainfall-runoff model usingan artificial neural network approach: Mathematical and Computer Modelling, v. 40, no. 7-8, p. 839-846.

Schapire, R. E., 2003, The boosting approach to machine learning: An overview, Nonlinear estimation and classification, Springer, p. 149-171.

Schilling, K. E., Chan, K.-S., Liu, H., and Zhang, Y.-K., 2010, Quantifying the effect of land use land cover change on increasing discharge in the Upper Mississippi River: Journal of Hydrology, v. 387, no. 3-4, p. 343-345.

Schilling, K. E., Jha, M. K., Zhang, Y. K., Gassman, P. W., and Wolter, C. F., 2008, Impact of land use and land cover change on the water balance of a large agricultural watershed: Historical effects and future directions: Water Resources Research, v. 44, no. 7.

Seo, Y., Kim, S., and Singh, V., 2018, Machine learning models coupled with variational mode decomposition: A new approach for modeling daily rainfall-runoff: Atmosphere, v. 9, no. 7, p. 251.

Seyoum, W. M., 2018, Characterizing water storage trends and regional climate influence using GRACE observation and satellite altimetry data in the Upper Blue Nile River Basin: Journal of Hydrology, v. 566, p. 274-284.

Seyoum, W. M., Kwon, D., and Milewski, A. M., 2019, Downscaling GRACE TWSA Data into High-Resolution Groundwater Level Anomaly Using Machine Learning-Based Models in a Glacial Aquifer System: Remote Sensing, v. 11, no. 7, p. 824.

Seyoum, W. M., and Milewski, A. M., 2016, Monitoring and comparison of terrestrial water storage changes in the northern high plains using GRACE and in-situ based integrated hydrologic model estimates: Advances in Water Resources, v. 94, p. 31-44.

Seyoum, W. M., and Milewski, A. M., 2017, Improved methods for estimating local terrestrial water dynamics from GRACE in the Northern High Plains: Advances in Water Resources, v. 110, p. 279-290.

Seyoum, W. M., Milewski, A. M., and Durham, M. C., 2015, Understanding the relative impacts of natural processes and human activities on the hydrology of the Central Rift Valley lakes, East Africa: Hydrological Processes, v. 29, no. 19, p. 4312-4324.

Srinivasan, R., Zhang, X., and Arnold, J., 2010, SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin: Transactions of the ASABE, v. 53, no. 5, p. 1533-1546.

Steitz, D., O'Donnell, F., Buis, A., Chandler, L., Baguio, M., and Weber, V., 2002, GRACE Launch, Press Kit: DE Steitz, F. O'Donnell, A. Buis, L. Chandler, M. Baguio, V. Weber.

Tanty, R., and Desmukh, T. S., 2015, Application of artificial neural network in hydrology—A review: Int. J. Eng. Technol. Res, v. 4, p. 184-188.

Taormina, R., Chau, K.-W., and Sivakumar, B., 2015, Neural network river forecasting through baseflow separation and binary-coded swarm optimization: Journal of Hydrology, v. 529, p. 1788-1797.

Tokar, A. S., and Johnson, P. A., 1999, Rainfall-runoff modeling using artificial neural networks: Journal of Hydrologic Engineering, v. 4, no. 3, p. 232-239.

USACE, 2006, (U.S. Army Corps of Engineers) Illinois River Basin Restoration Comprehensive Plan with Integrated Environmental Assessment, Public Review Draft, February 2006, Rock Island District, USACE: Rock Island, IL.

Wan, Z., 2008, New refinements and validation of the MODIS land-surface temperature/emissivity products: Remote sensing of Environment, v. 112, no. 1, p. 59-74.

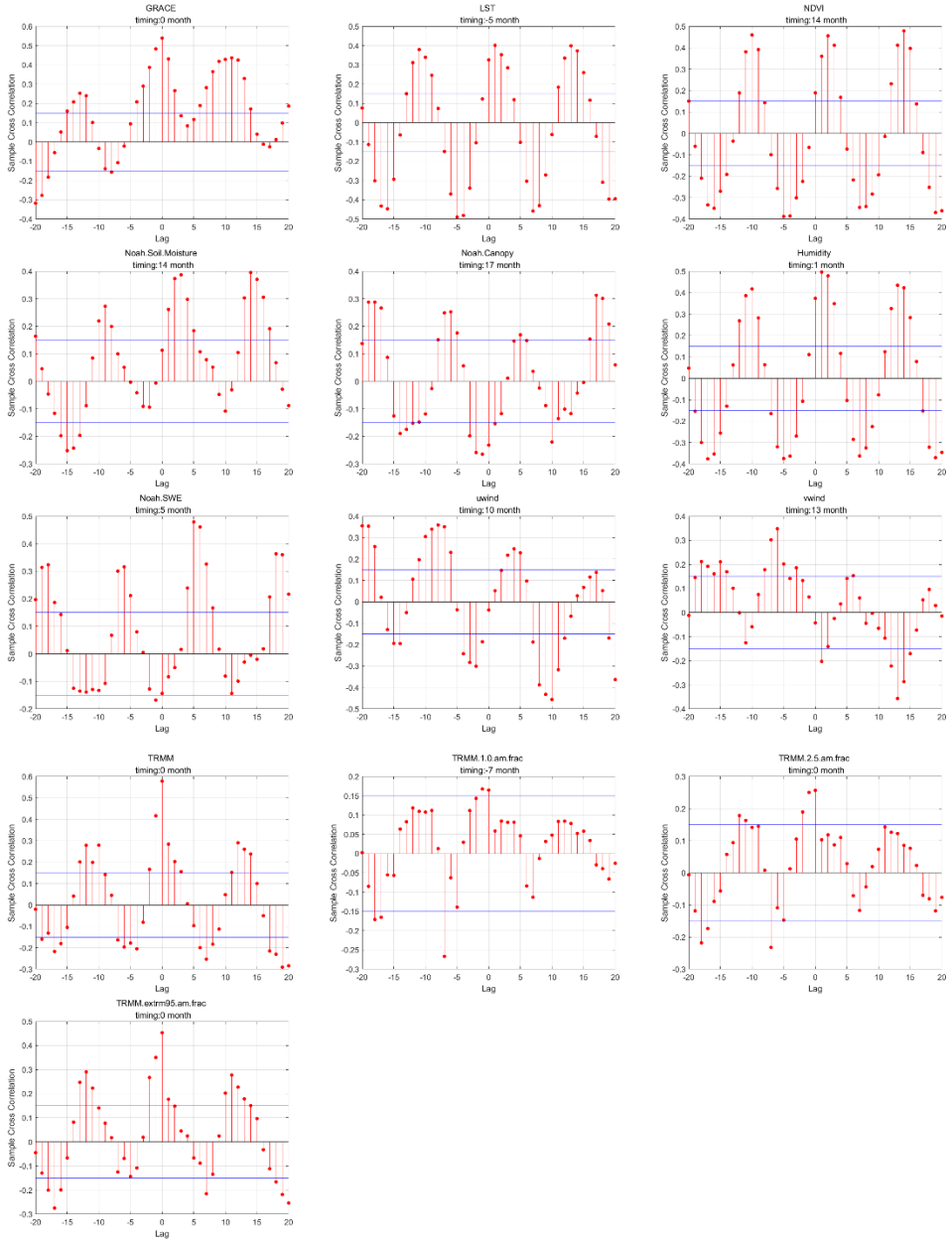
Wan, Z., Hook, S., and Hulley, G., 2015, MOD11C3 MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 0.05 Deg CMG V006.

Yaseen, Z. M., Ebtehaj, I., Bonakdari, H., Deo, R. C., Mehr, A. D., Mohtar, W. H. M. W., Diop, L., El-Shafie, A., and Singh, V. P., 2017, Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model: Journal of Hydrology, v. 554, p. 263-276.

Yen, H., Daggupati, P., White, M., Srinivasan, R., Gossel, A., Wells, D., and Arnold, J., 2016, Application of Large-Scale, Multi-Resolution Watershed Modeling Framework Using the Hydrologic and Water Quality System (HAWQS): Water, v. 8, no. 12.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W., 2011, Indices for monitoring changes in extremes based on daily temperature and precipitation data: Wiley Interdisciplinary Reviews: Climate Change, v. 2, no. 6, p. 851-870.

APPENDIX A: CROSS-CORRELATION TEST OF THE POTENTIAL INPUT-VARIABLES

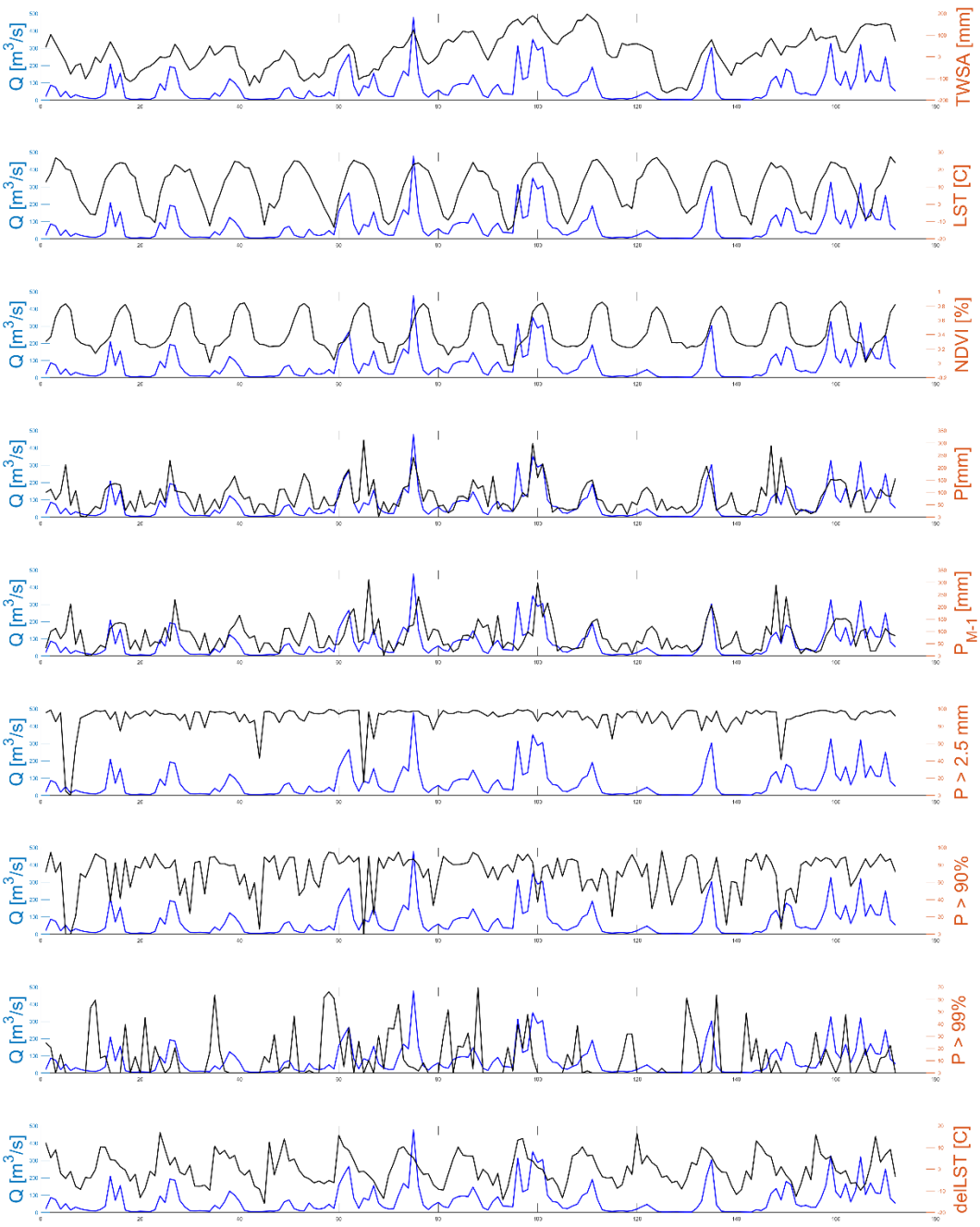


Cross-correlation of potential input-variables (from the top left: GRACE TWSA, LST, NDVI, soil moisture, plant canopy water, humidity, SWE, W-E wind speed, S-N wind speed, TRMM, TRMM > 1.0 mm fraction, TRMM > 2.5 mm fraction, and TRMM > 95% fraction) versus streamflow of the outlet (USGS 587450). Blue lines indicate significance value. Note that all the variables are mean value of RRW.

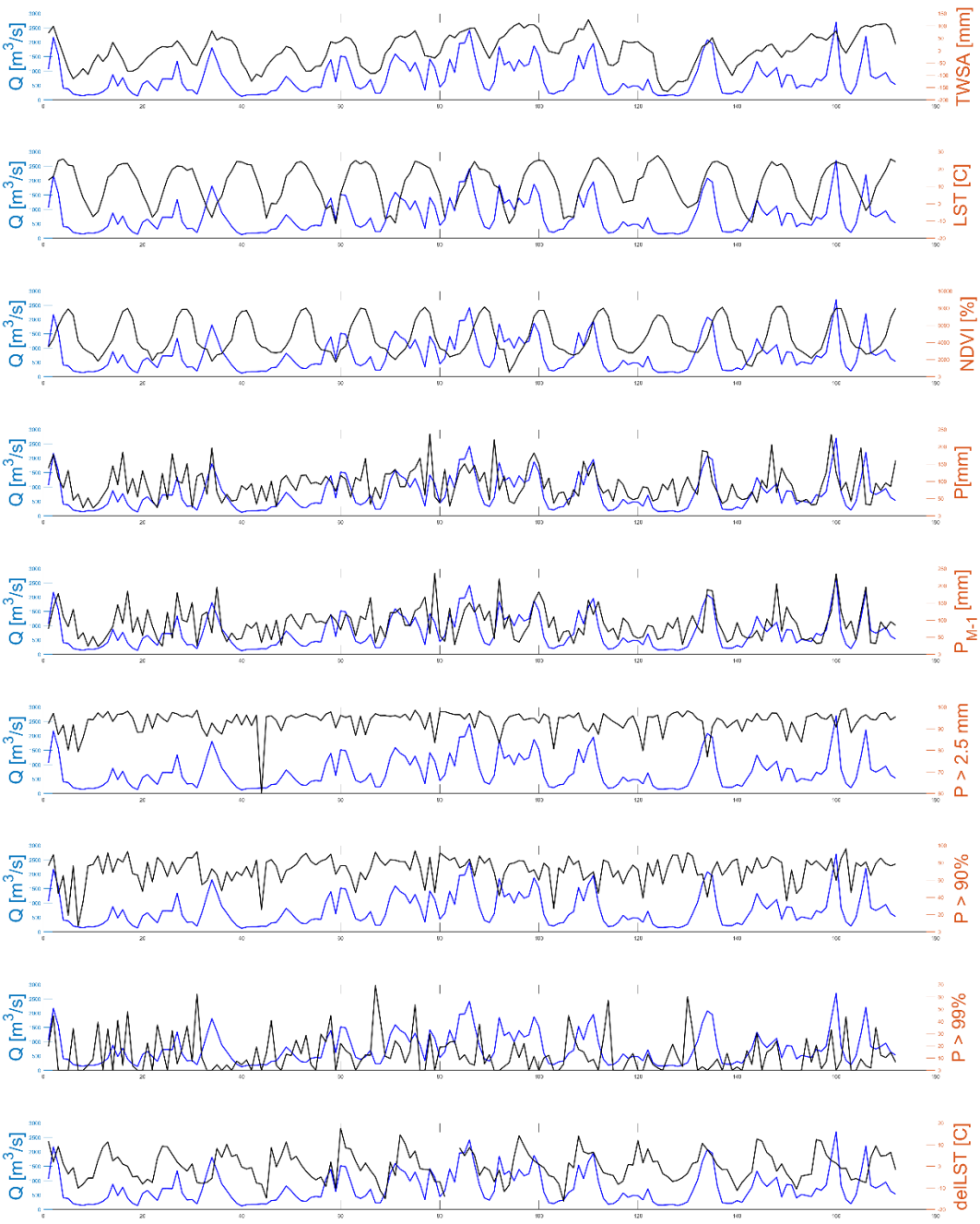
APPENDIX B: ACRONYMS

ANN	Artificial Neural Network
BRT	Boosted Regression Tree
GRACE	The Gravity Recovery and Climate Experiment
HU	Hydrologic Unit
HUC	Hydrologic Unit Code
IRW	Illinois River Watershed
LST	Land Surface Temperature
MAE	Mean Absolute Error
MLM	Machine Learning Model
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NSE	Nash-Sutcliffe Efficiency
PBIAS	Percent BIAS
PI	Predictor Importance
R^2	Coefficient of determination
RRW	Raccoon River Watershed
SWAT	Soil and Water Assessment Tool
SWE	Snow Water Equivalent
TWS	Terrestrial Water Storage
TWSA	Terrestrial Water Storage Anomaly
UMRB	Upper Mississippi River Basin
WBDHU	Watershed Boundary Dataset Hydrologic Units

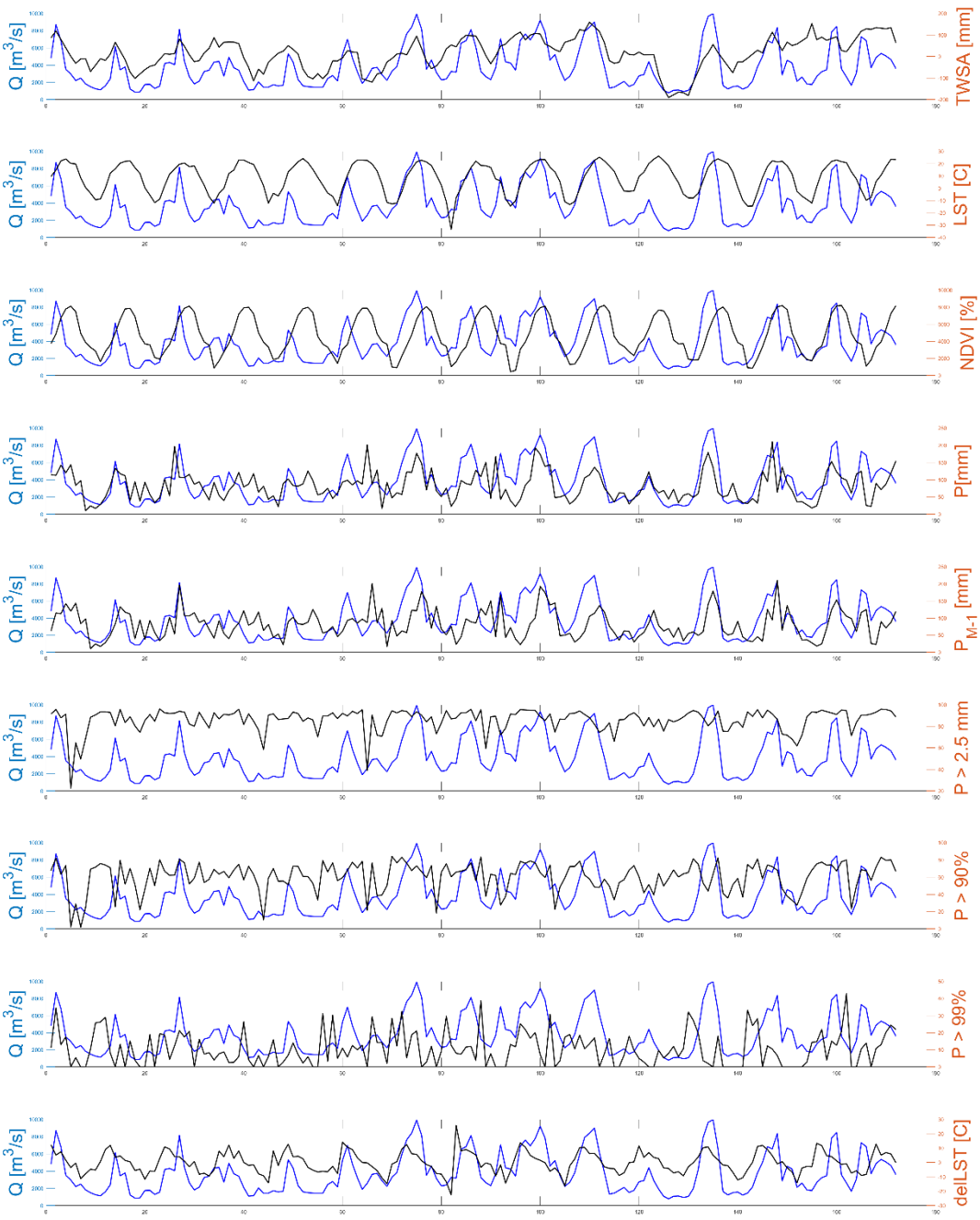
APPENDIX C: STREAMFLOW (Q) VS. VARIABLES – RRW



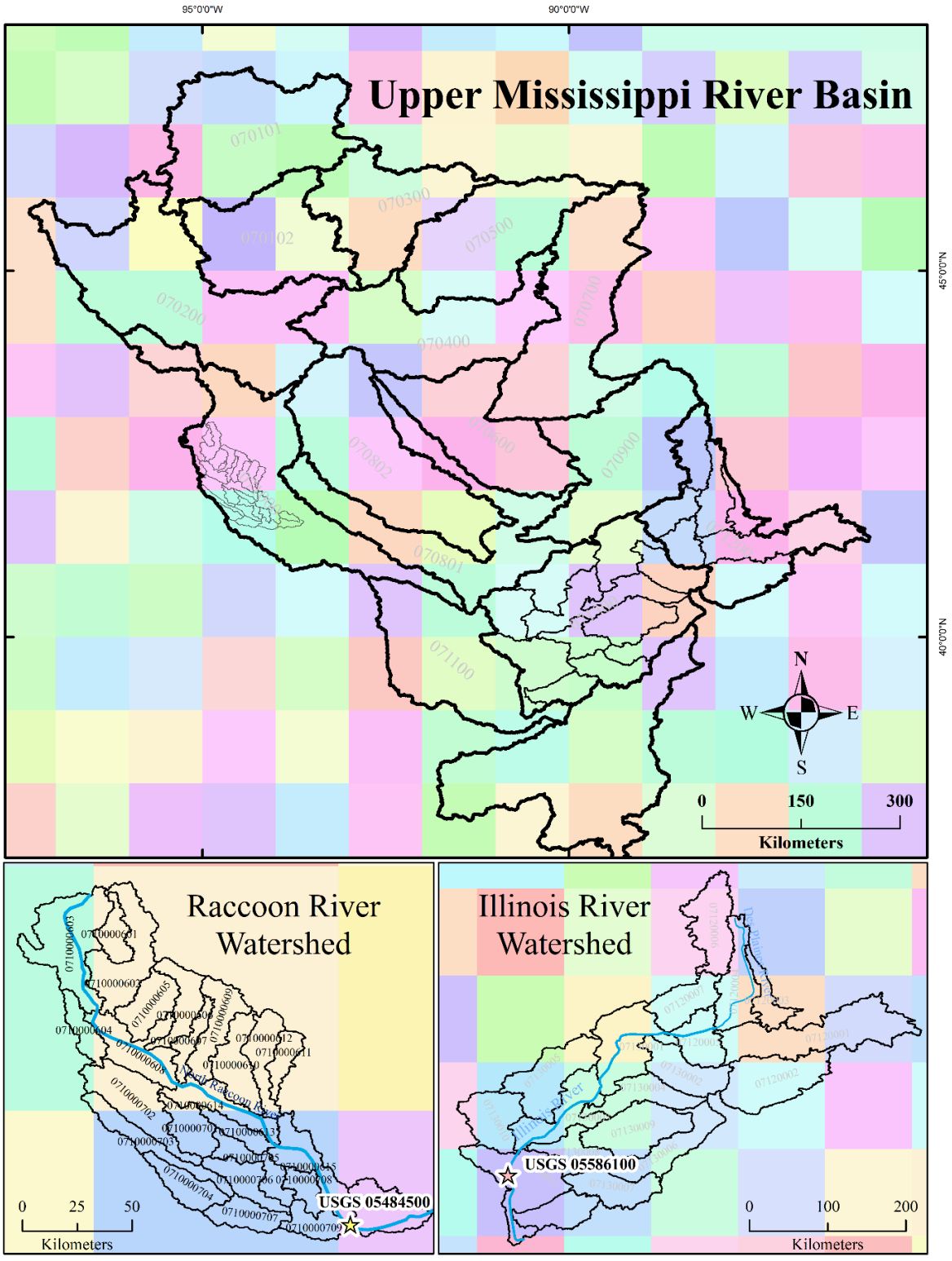
APPENDIX D: STREAMFLOW (Q) VS. VARIABLES – IRW



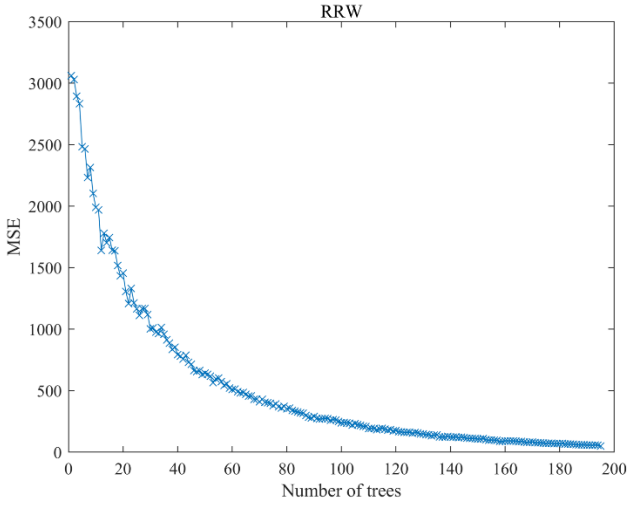
APPENDIX E: STREAMFLOW (Q) VS. VARIABLES – UMRB

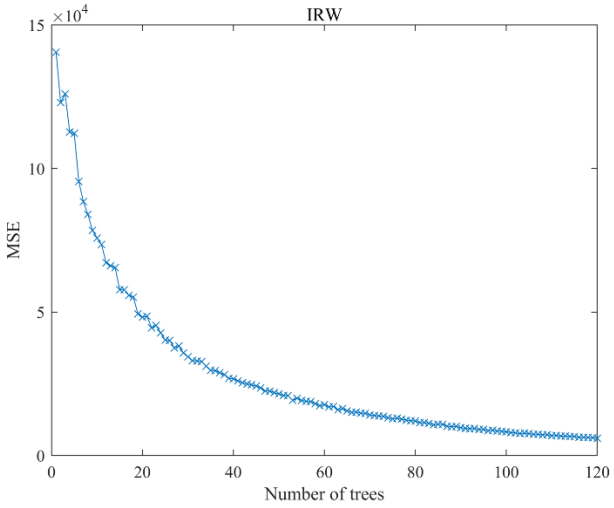


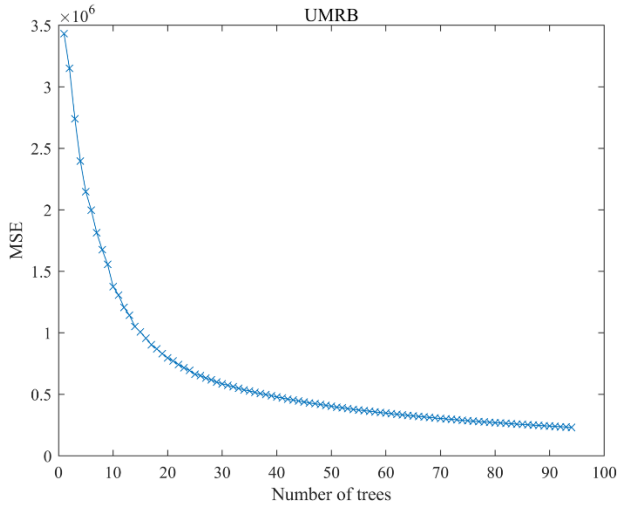
APPENDIX F: COMPARISON BETWEEN GRACE PIXELS AND WATERSHEDS



APPENDIX G: MSE VERSUS ADDITIVE TREES

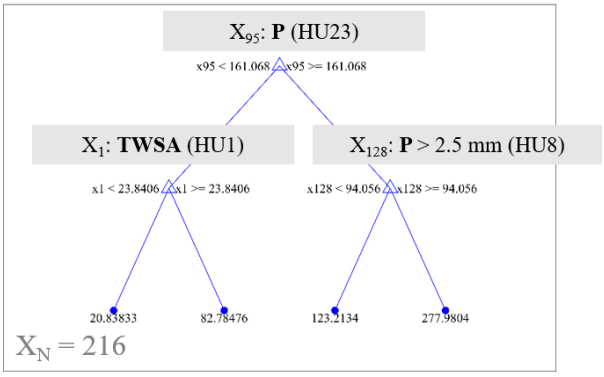






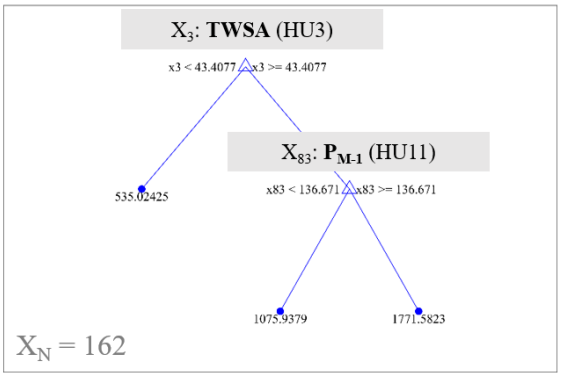
APPENDIX H: EXAMPLE OF TRAINED TREES (1ST TREES)

RRW



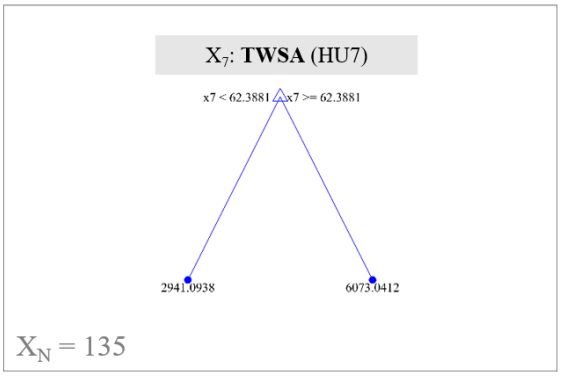
Total 195 trees

IRW



Total 120 trees

UMRB



Total 94 trees