4-5-2019

WATERSHED CHARACTERIZATION AND STREAMFLOW FORECASTING USING REMOTE SENSING AND MACHINE LEARNING

Dongjae Kwon
Illinois State University, dkwon@ilstu.edu

Watershed Characterization and Streamflow Estimation Using Remote Sensing and Machine Learning

Department of
Geography, Geology,
and the Environment
Illinois State University

Dongjae Kwon[1] and Wondwosen M. Seyoum[1]
[1] Department of Geography, Geology, and the Environment, Illinois State University, Normal, IL, USA

Water **RES**ource and **RE**mote **S**ensing **RES**earch **L**aboratory — WRESL

Abstract

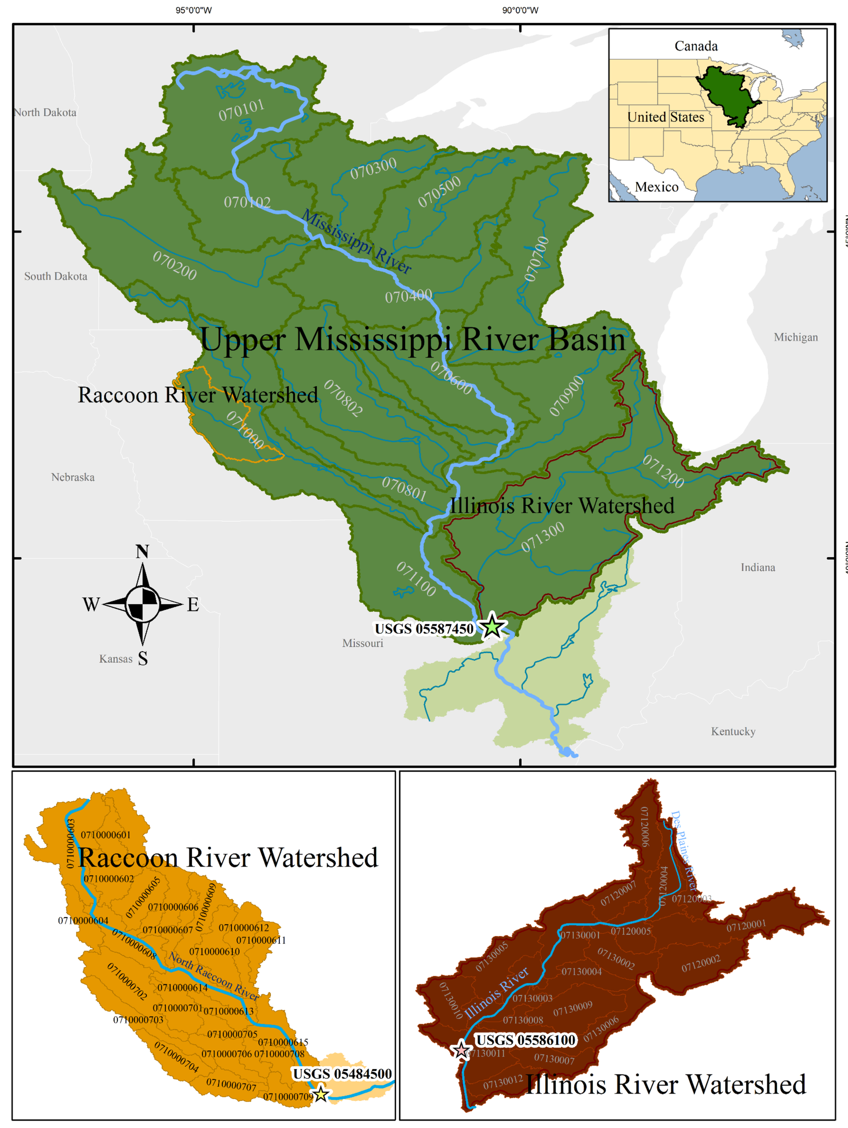
Streamflow data are essential to study the hydrologic cycle and to attain appropriate water resource management policies. Unfortunately, the availability of gauging data is limited due to instrumental malfunction and its poor spatial distribution. Since streamflow is governed by the water balance equation, it can be estimated by the other water-budget variables. Many studies have done to estimate streamflow using conventional approaches such as process-based and empirical modeling. However, the approaches have limitations: require many input data so, relatively expansive (process-based modelings), have relatively low performances (regression equations and machine learning models for Rainfall-Runoff), and need gauging data from upstream (machine learning). Here, we introduce a machine learning approach based on remotely sensed hydro-climatic variables to estimate monthly streamflow for three different-sized hydrologic units. By integrating spatial land surface and climate data that describe a watershed as an input dataset in a machine learning model (MLM), and streamflow data for an output learning dataset, relationships between watershed characteristics and streamflow are established. The results are validated using gauged streamflow data and compared with results from process-based modeling studies. The testing result shows relatively good statistics (UMRB: R-0.9066, NSE-0.7926; IRW: R-0.9122, NSE-0.7666; Raccoon River Watershed: R-0.8443, NSE-0.6856). The overall performance of the models shows the hydro-climatic data integrated MLM could be effectively applied to streamflow estimations.

Objectives

The main objective of this study is to characterize a watershed and evaluate the effectiveness of a machine learning-based hydrologic model in simulating the water cycle using remote sensing data that potentially aid the conventional watershed analyses.

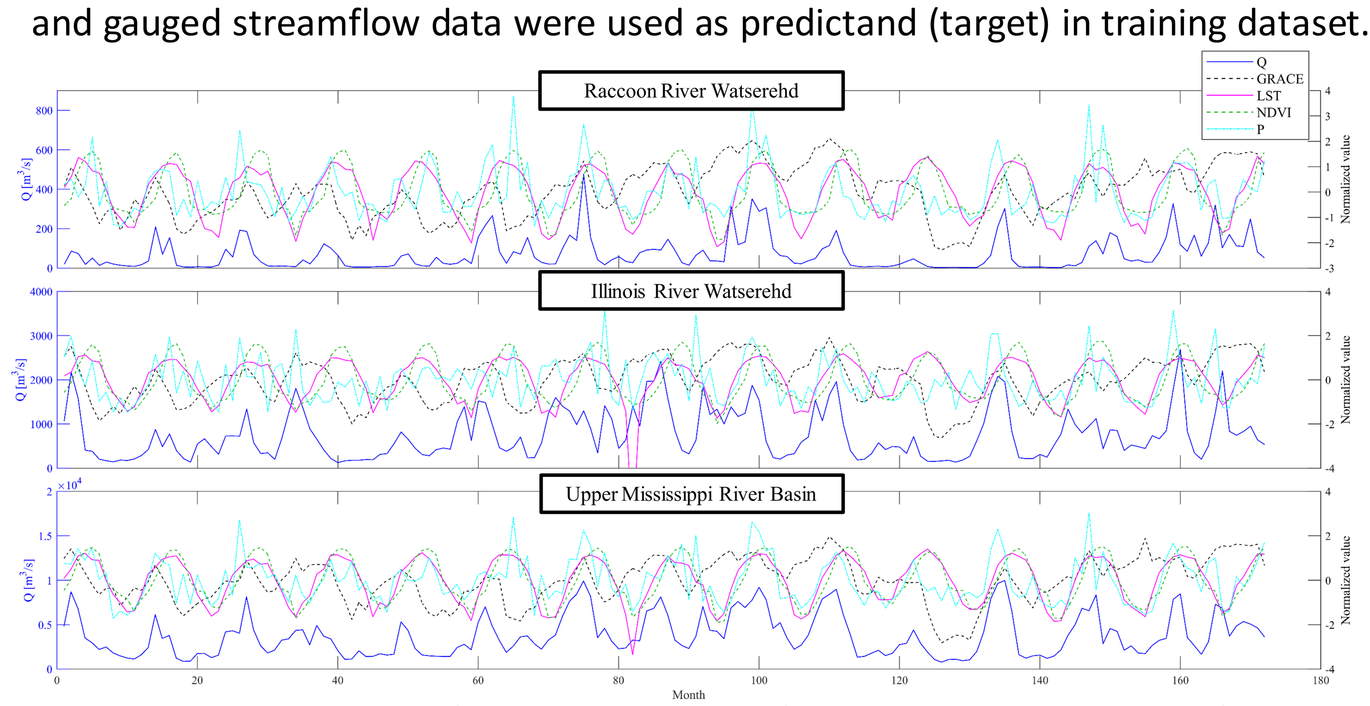
Study Area

- Upper Mississippi River Basin (UMRB; 492,000 km^2) is one of the major sub-basins of Mississippi River Basin and more than 30 million people rely on streamflow.
- Illinois River Watershed (IRW; 74,677 km^2) and Raccoon River Watershed (9,400 km^2) are part of UMRB.
- Landcover of UMRB: mostly cropland (~40%) and forest (~19.4%).
- River discharge has significantly increased due to the expansion of soybean fields.



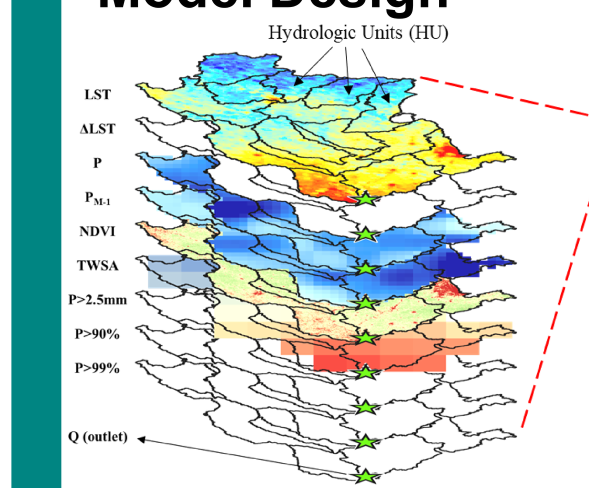
Methods

Various remote sensing data were collected and tested which could be related to streamflow responses of watersheds by considering available period and spatial resolution. Monthly time series of GRACE Terrestrial Water Storage (TWS), MODIS Land Surface Temperature (LST) and ΔLST, TRMM Precipitation (P) and detailed statistics (P > 2.5mm; 90%; 99%) derived from daily TRMM, and MODIS Normalized Difference Vegetation Index (NDVI) were used as predictor variables and gauged streamflow data were used as predictand (target) in training dataset.



Above: Monthly time series (Apr. 2002 to Dec. 2016) plots of remote sensing data (GRACE TWS anomaly, LST, P, NDVI; averaged in a watershed and normalized; right axis) and streamflow (left axis) according to watershed and its gauging station (outlet).

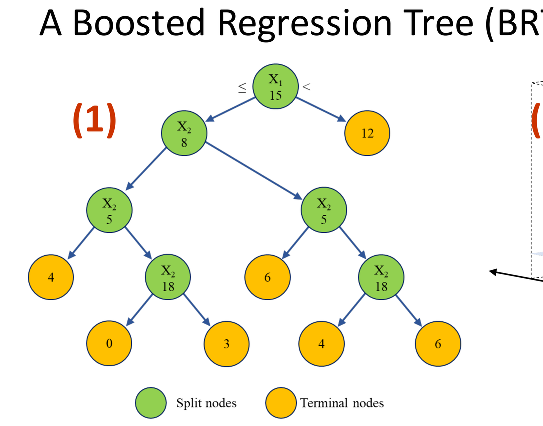
Model Design



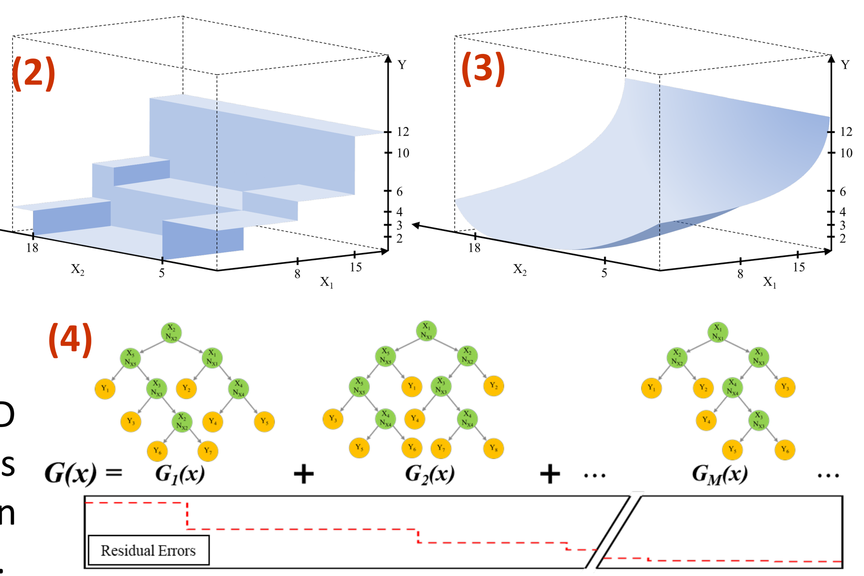
	Predictors									Predictand
Mon.	LST (HU1)	LST (HU2)	…	ΔLST (HU1)	ΔLST (HU2)	…	P (HU1)	P (HU2)	…	Q
1	15.9	15.8	…	2.1	1.7	…	12	17	…	1856
2	17.8	18.2	…	1.9	2.4	…	35	29	…	2432
3	23.4	23.2	…	5.6	5.0	…	31	34	…	4068
4	24.5	24.3	…	1.1	1.1	…	25	28	…	3576
5	25.2	25.7	…	1.3	1.4	…	29	21	…	2953
⋮	⋮	⋮		⋮	⋮		⋮	⋮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

To mimic the relationship between watershed characteristics and streamflow, remote sensing data are averaged according to smaller hydrologic units (subwatersheds) and all of them were assigned as predictors. This approach could effectively reflect the contribution of subwatersheds to streamflow responses. Total 142 (from July 2004 to July 2016) observations are used for training and 30 observations are used for testing (from April 2002 to June 2004). A Boosted Regression Tree (BRT) method is used in this study (below).

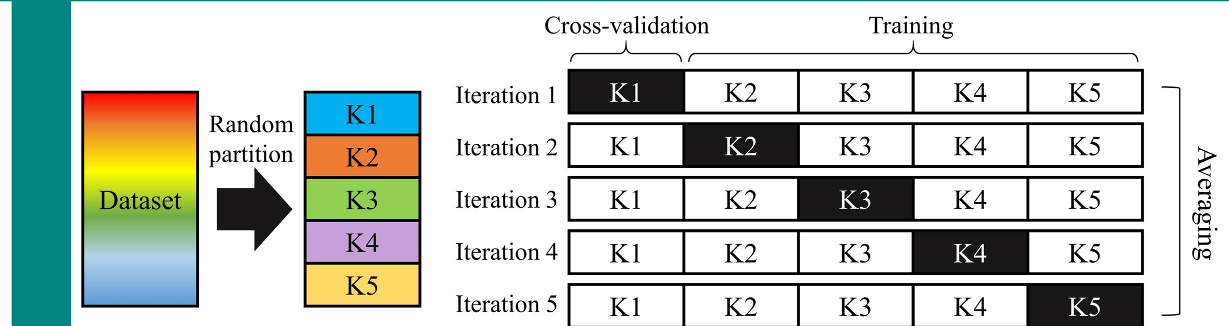
(1)



(2) (3) (4)



A decision tree[1] simulates 3-D covariant space[2]. Boosting[4] is a series of decision trees can reduce errors dramatically[3&4].



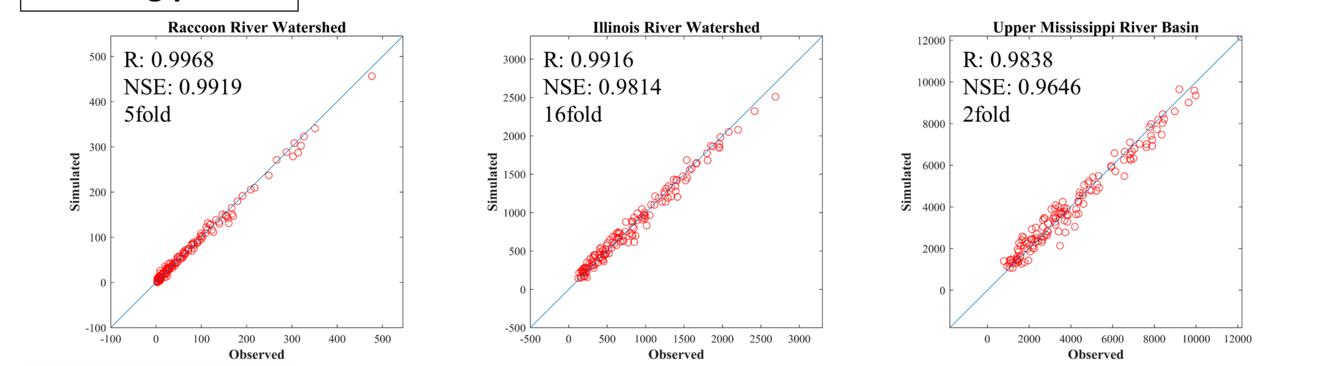
Left: K-fold method is used for training.

The method is better than the random partition method if a size of training data is small. Total 309 (103x3) modeling processes were run while the K-fold factor is changing from 2 to 36.

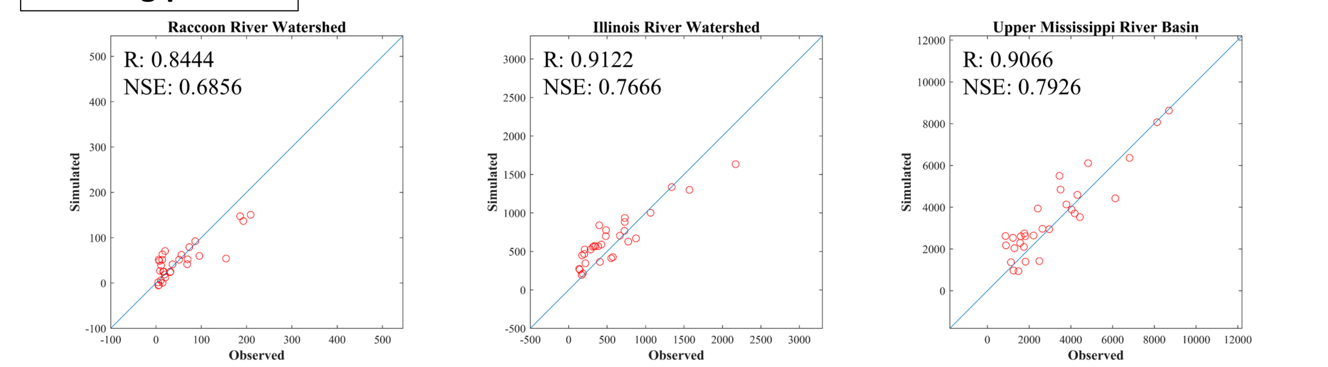
Result

Streamflow estimations: training and testing

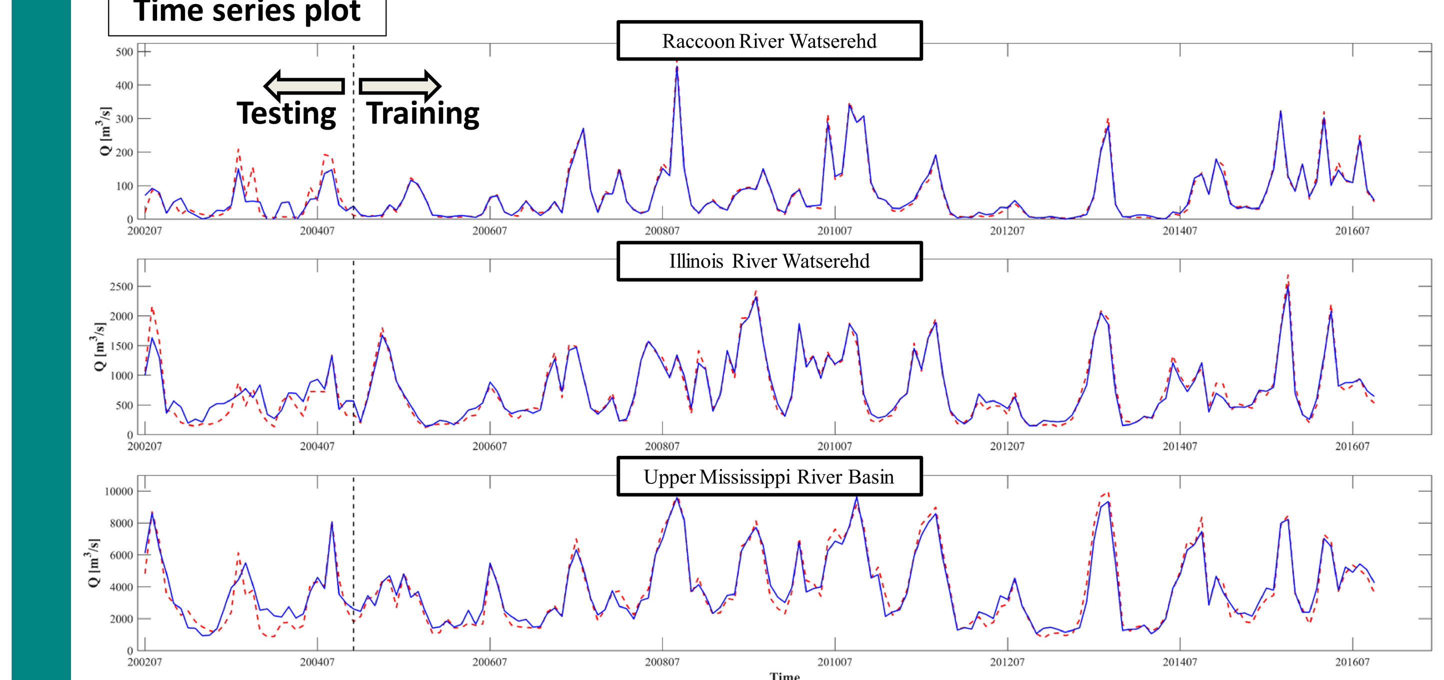
Training period



Testing period



Time series plot

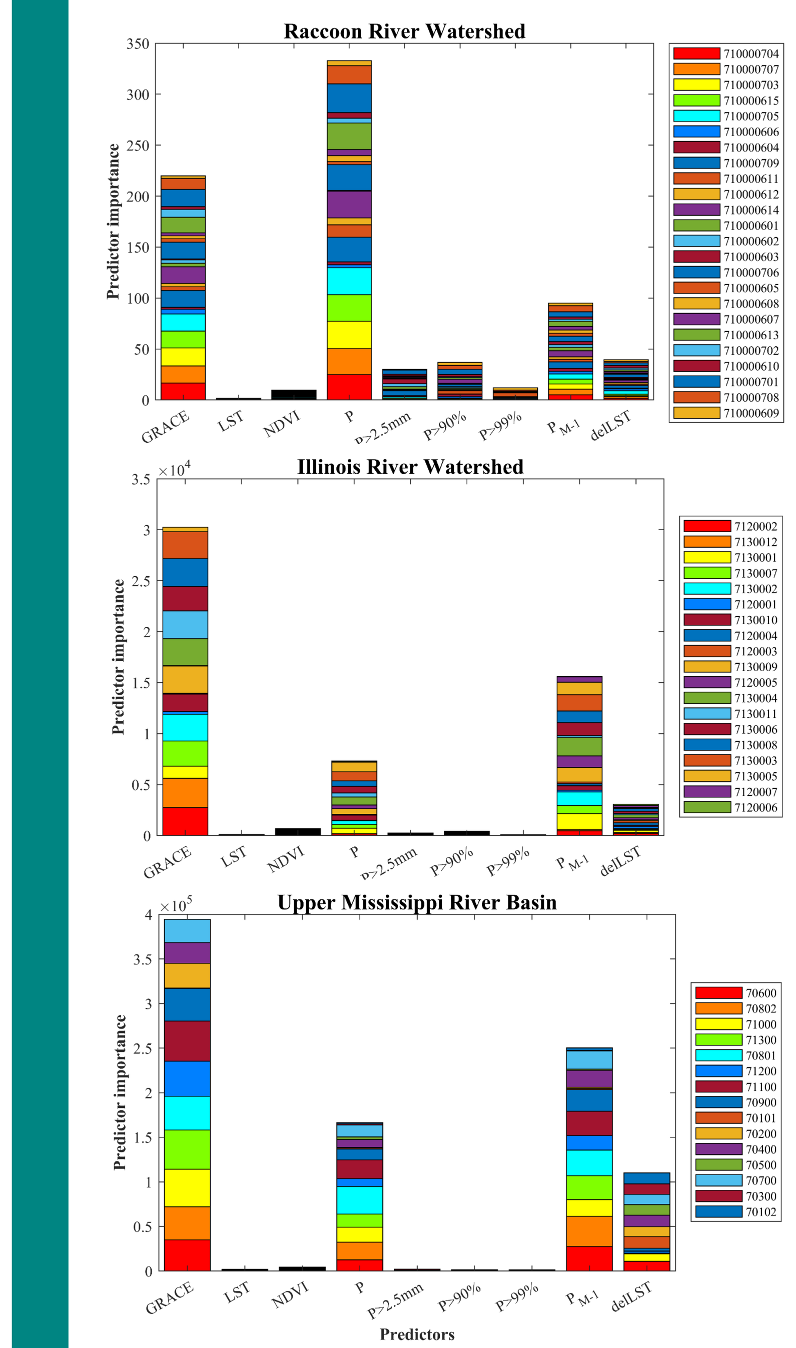


Process-based modeling vs. MLM

Watershed	Gauging station	Author	Modeling method	Testing period	Performance (NSE)	This study (NSE)
Raccoon	USGS 05587450	Jah et al., 2007	SWAT	1993-2003	0.88	0.6856
IRW	USGS 05586100	Yen et al., 2016	SWAT	1990-2001	0.72	0.7666
UMRB	USGS 05484500	Srinivasan et al., 2010	SWAT	1980-1988	0.69	0.7926

- The testing performances of this study are comparable with previous process-based modeling studies (**table above**). However, smaller watershed (e.g., Raccoon) shows higher uncertainties such as underestimation in high streamflow conditions and overestimation in low streamflow conditions.

Predictor Importance



Predictor importance (**left**) shows different characteristics of watersheds according to size

- GRACE TWS anomaly plays the most important role in except Raccoon River Watershed. This is because GRACE has coarse spatial resolution (approx. 100 km x 100 km).
- The contributions of the previous month's precipitation (P_{M-1}) have a tendency to increases as watershed size is increased. Precipitation in the upstream area will take a long time to reach outlet as watershed size increases
- The higher predictor importance of detailed P statistics (P > 2.5mm, P > 90%, and P > 99%) in smaller watersheds also implies the nature of them have rapid and sensitive streamflow responses from precipitation.
- ΔLST may explains not only water from snow-melting but also seasonal changes which could be indirectly related with climatic effect such as humidity, and wind speed. The low PI of LST and NDVI indicates the variable have small effect to streamflow or they had no chance because most of streamflow responses were explained by the other predictors.

Conclusion

The following conclusions were drawn from this research:

- Monthly streamflow can be effectively estimated using MLM and remote sensing. MLMs for each basin and watersheds show reasonable performances. Even though MLM is considered as 'Black box' model, BRT can tell us its implications through predictor importance analysis.
- The effectiveness of remote sensing-integrated MLM depends on the watershed scale due to the spatial resolution of RS data such as GRACE. If better spatial resolution can be used, the better result is expected.
- To have higher modeling performance in smaller watersheds, the temporal resolution to analysis needs to be improved. Monthly time-scale seems not good enough to small watersheds such as Raccoon River Watershed where streamflow response is immediate and sensitive.

Remote sensing-based MLM introduced in this study has the potential to be a supplementary approach to estimate streamflow. Overall performance is comparable with process-based approaches. Especially, the method could be a very attractive tool to study areas out of the US such as developing countries because lack of gauging streamflow data is serious and crucial.