Final Project Report - ABE69100

Topic: Sensitivity of Streamflow Parameters

Rishitha Bikkumalla, Lyles School of Civil Engineering, Purdue University

Date of Completion: 04/30/2024

Introduction:

Publication synopsis

A comprehensive study on the impact of varying parameters on streamflow estimation accuracy using HEC-HMS.

Publication description

This research delves into a systematic sensitivity analysis of critical parameters in the HEC-HMS for estimating streamflow. By utilizing datasets like precipitation, streamflow, etc, the study seeks to understand how different parameters influence the accuracy of streamflow predictions, aiming to contribute significantly to water resources management and flood forecasting.

Tags

Precipitation, Streamflow, Watershed, Modeling (1952, 4316), prediction, sensitivity analysis

Source data overview:

Provenance of the source data

The research employs data from multiple sources. All the data is for the Upper Wabash Basin. The rainfall data was obtained by an ex-group student from the NASA Land Data Assimilation System, where hourly gridded rainfall data were aggregated over the entire watershed to derive a single rainfall time series. Hourly rainfall data is available only for the event from 06/09/1981 to 06/11/1981.

Daily streamflow data is sourced from USGS for 1964-2003, covering the Upper Wabash watershed monitored by four-gauge stations. Only the timeframe of interest, 1980-1982, was studied.

Streamflow Nodes	Site Number	Guage Station Name	Longitude	Latitude
1	03342500	BUSSERON CREEK NEAR CARLISLE, IN	-87.4258	38.9742
28	03342300	BUSSERON CR NR SULLIVAN IN	-87.3864	39.0758

54	03342219	BIG BRANCH TRIB 4.4 MILES NE OF DUGGER, IN	-87.2783	39.1308
59	03342100	BUSSERON CREEK NEAR HYMERA, IN	-87.3114	39.215

Additional datasets include DEM, SSURGO soil data, and Land-use land cover (LULC). DEM is obtained from the Earth Explorer of USGS: SRTM data from 2014 was used. SSURGO, a comprehensive database on soil characteristics, is accessed from the Natural Resources Conservation Service (USDA) of 2023. LULC data is downloaded from the National Land Cover Database 2021(NLCD).

Furthermore, streamflow data was predicted through HEC-HMS for various combinations of the parameters: Time of Concentration (Tc) and Storage Coefficient(S). These parameters were selected from a uniform distribution. This predicted streamflow was compared with the observed streamflow, and Nash-Sutcliffe efficiency (NSE) was calculated for each simulation.

Data Source:

Daily Streamflow data: https://waterdata.usgs.gov/nwis/sw

Rainfall Data: https://ldas.gsfc.nasa.gov/nldas/v2/forcing

DEM: https://earthexplorer.usgs.gov/

SSURGO: https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo

LULC: https://www.usgs.gov/centers/eros/science/national-land-cover-database

Format of the source data

Data	Format	Description
Precipitation	.txt	Space delimited hourly rainfall data in mm from 06/09/1981_2200 to 06/11/1981_0000
Streamflow	.txt	Space delimited daily streamflow in cfs from 6/1/1964This was originally extracted from to 9/30/2003 for 4 node points.
DEM	.tif	NAD 1983 UTM Zone 16N DEM over the Wabash River in .tif format.
Soil Data	.gdb	Soil data, containing Hydrologic Group in letters and other details.
LULC data	.tif	Land Use and Land Cover (LULC) data, containing numbers corresponding to their type. They were

		reclassified into 4 types: Urban, Agricultural, Forest, and Water
Predicted Streamflow	.csv	A comma-separated values file containing predicted streamflow data was generated by combining multiple HEC HMS results.

Methods:

Overview of Processing Completed as Dataset Preparation

The data preparation involved aggregating rainfall and precipitation data and checking their gaps or abnormalities. A specific rainfall event was then selected, and the streamflow corresponding to it was extracted for all the nodes. Next, SSURGO and LULC data were used to find the Curve Number of the basin, using which an empirical basin model was prepared from the DEM. This model is then automated through Python by varying the Time of Concentration (Tc) and Storage Coefficient(S) to predict the streamflow.

Programs/Scripts used

ArcGIS was employed for spatial data analysis and preparation. It helped understand the behavior of the river in the basin. HEC-HMS was used to perform continuous hydrologic simulations, model streamflow, and conduct sensitivity analysis. Python scripts facilitated data aggregation, preprocessing, quality checking, automating HEC-HMS, and graphically analyzing the data and results.

Processes

Graphical Data Analysis, Data Quality Checking and Statistics, and Metrics Calculations were performed to understand the data and further the process.

Graphical data analysis:

Overview of what graphical data analysis was conducted and a summary of what was found

Graphical analysis was conducted to visualize the source data and the impact of different parameters on streamflow estimation. Key findings include understanding the parameter sensitivity range and it's impact on model accuracy.

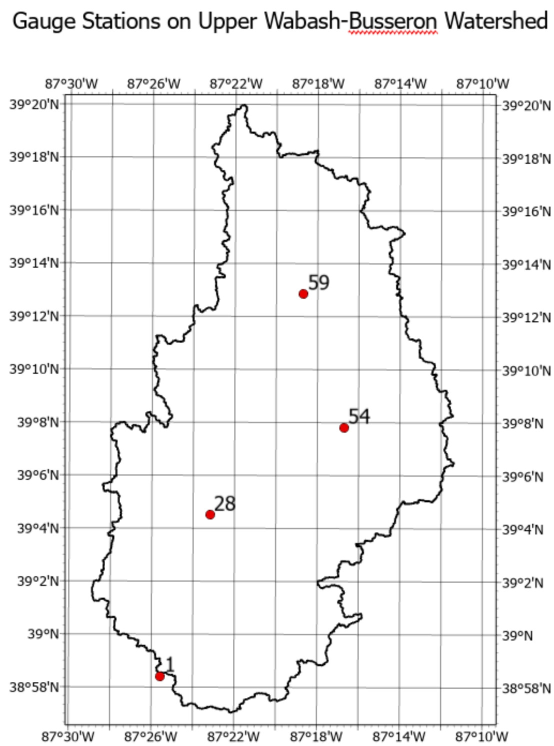

Figure 1: Upper Wabash-Busseron watershed, HUC 0512011115, with 4 different streamflow nodes.

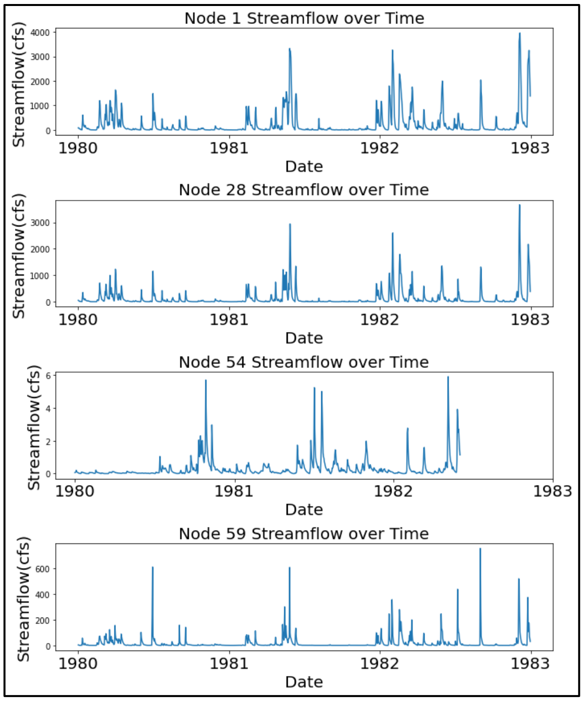

Figure 2: Streamflow of 4 gauge stations of Upper Wabash watershed for 3 years. It can be seen that only nodes 1,28, and 59 have significant streamflow. Node 54 is a very small tributary with less streamflow.

Further, the data of the selected rainfall event was analyzed. **Event rainfall:** Hourly from 06/09/1981_22:00 to 06/11/1981_00:00, **Nodewise Streamflow:** Daily from 6/1/1966 to 9/30/2003.

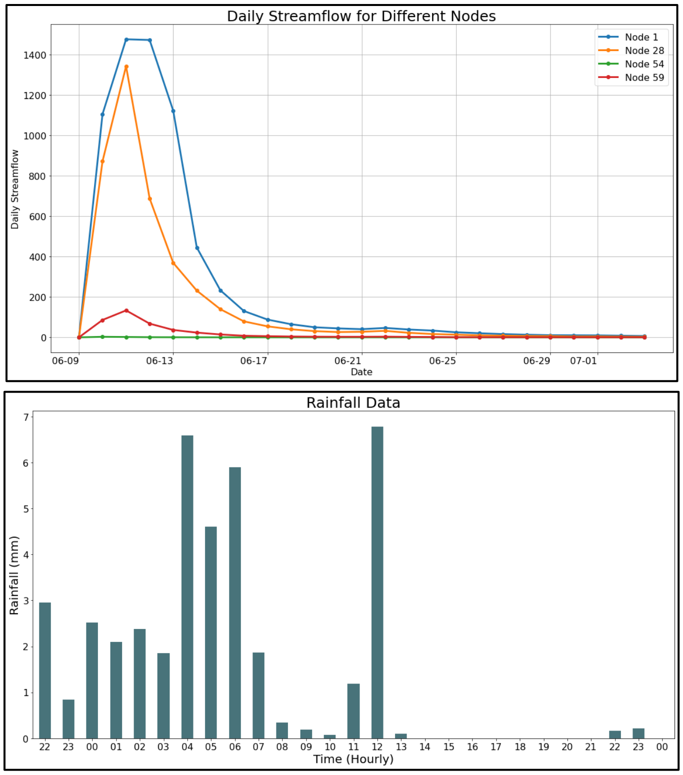

Figure 3(a): 21-day Streamflow of The Upper Wabash - Busseron Watershed for all nodes after the specific rainfall event of 1981. Very small streamflow of Node 54- small tributary, Figure 3(b): 27-hour Rainfall from 06/09/1981_22:00 to 06/11/1981_00:00. Rainfall is in mm. It can be seen that the maximum rainfall in the event occurs in the first 16 hours.

Next, 1000 random samples of Tc and S were generated in a uniform distribution, as shown in Fig 4. The HEC-HMS simulations predicted streamflow for each pair of samples. The predicted streamflow is compared with the observed streamflow of Node 1 through NSE. The NSE variation is then analyzed.

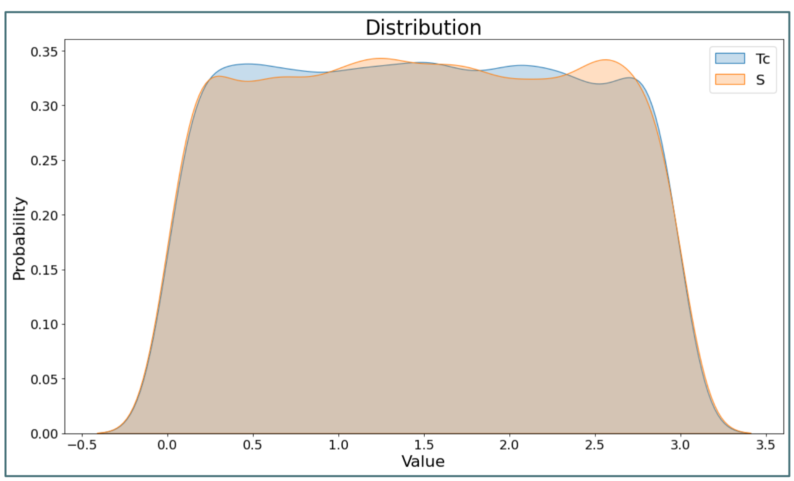

Figure 4: The image illustrates two kernel density estimate (KDE) plots, which are used to visualize the probability density function of the random variables 'Tc' and 'S'. The plots suggest an estimation of the underlying probability distribution based on a finite sample set. Given that there are only 1,000 samples, it is important to note that the KDE might not accurately represent the true underlying distribution. This limitation can be attributed to the KDE's reliance on smoothing parameters that may introduce bias when sample sizes are small. For more precise estimation, a larger sample size would be necessary to reduce the influence of the KDE's bandwidth and to provide a more accurate reflection of the true distribution.

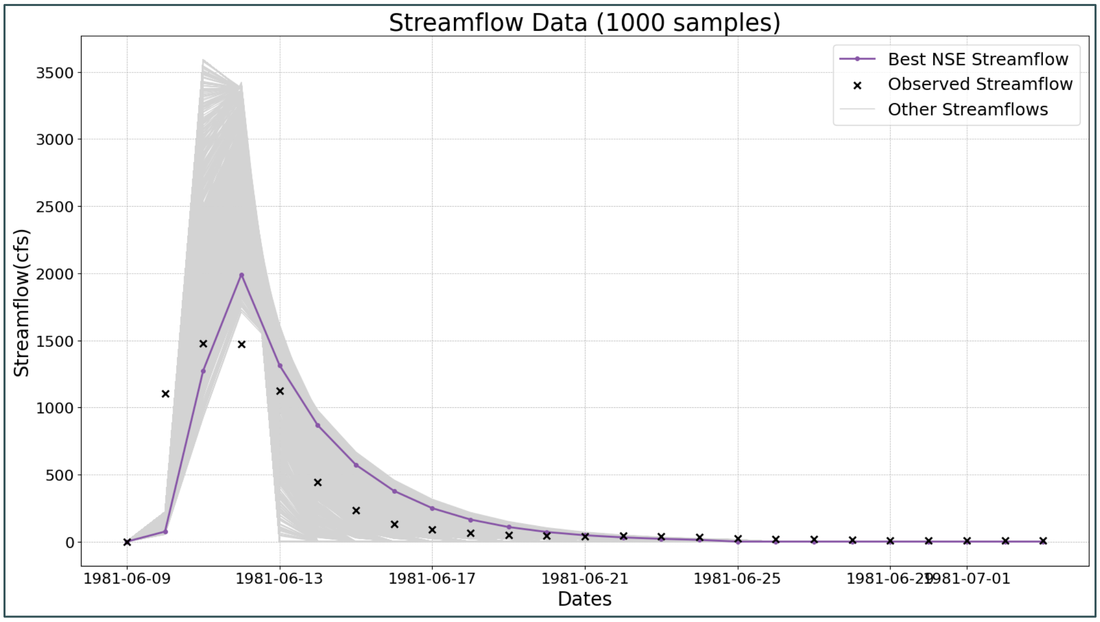

Figure 5: The predicted streamflow of NODE 1 from the simulations. The purple-colored line corresponds to the streamflow of the highest NSE of 0.667. The grey lines are the other 999 streamflow simulations. The black x's correspond to the observed streamflow values. It can be observed that the 2nd predicted value is never predicted well.

Data Quality Checking:

Overview

Data quality checking involved validating the accuracy of input datasets (rainfall, streamflow) through comparison with known benchmarks and identifying any gaps or inconsistencies.

1) **Streamflow Data:**
   a. Null Values Check
      Check if there are any null values and replace them with NaN.
   b. Gross Error Check
      Streamflow should lie between 0 to 5000cfs. Any values out of this range should be removed.
   c. Maximum Value Check:
      The streamflow of each node should be greater than 100 cfs. Any values out of this range should be removed
2) **Rainfall Data:**
   a. Null Values Check
      Check if there are any null values and replace them with NaN.
   b. Gross Error Check
      Precipitation should lie between 0 to 100 mm. Any values out of this range should be removed.
3) **Predicted Data:**
   a. Null Values Check
      Check if there are any null values and replace them with NaN.
   b. Gross error check for parameters:
      Tc and S should lie between 0 and 3. Any values out of this range should be removed
   c. Predicted Streamflow should be positive

Check	Precipitation	Observed Streamflow	Predicted Streamflow	Parameters (Tc and S)
Null Values	0	0	0	N/A
Gross Error	0	0	N/A	0
Maximum Value Check	0	1 node has a maximum streamflow of 6 cfs; hence node was removed for future study	N/A	N/A
Positive Streamflow Check	N/A	N/A	0	N/A

Table 1: Number of Corrections made in data quality checking

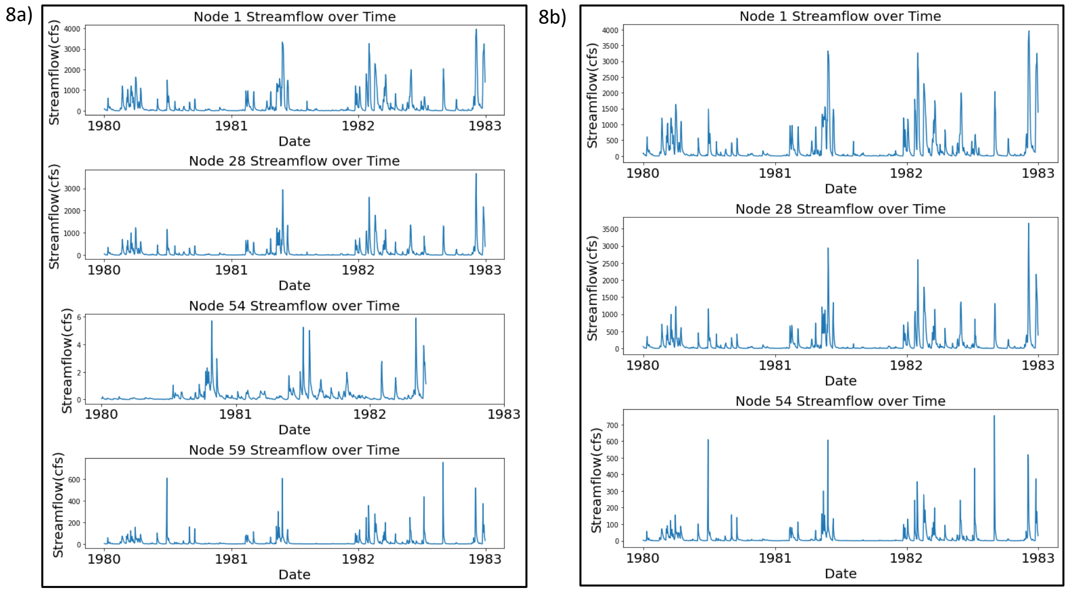

Figure 6: Streamflow of the Nodes, before and after data quality checking. It is evident from image 8b that one of the node data was removed.

Statistics And Metrics:

Overview

Statistics and Metrics calculations are performed to understand the data to a better extent. We aim to quantify and understand various aspects of streamflow patterns, evaluate model performance, identify sensitive parameters, and ultimately improve the prediction and understanding of streamflow dynamics in the study area.

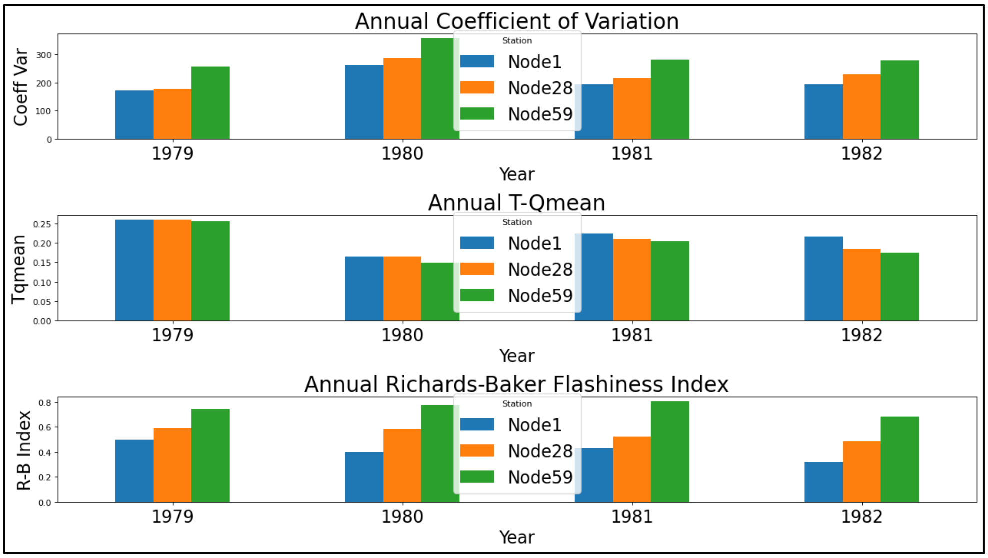

*Figure 7:: **Annual Coefficient of Variation:** A statistical measure of the dispersion of data points in a data series around the mean. The coefficient of variation appears to fluctuate from year to year for each location, with no clear trend of increase or decrease over the four years**. Annual T-Q Mean:** The fraction of time (days) that streamflow exceeds the mean annual streamflow (Qmean). Higher for 1979 because the water year consists of data from 1980 and no data from 1979. **Annual Richards-Baker Flashiness Index:** Measures the rapid changes in streamflow. Node 59 has a higher index, which indicates more rapid changes, which can be indicative of flash floods or other rapid runoff events.*

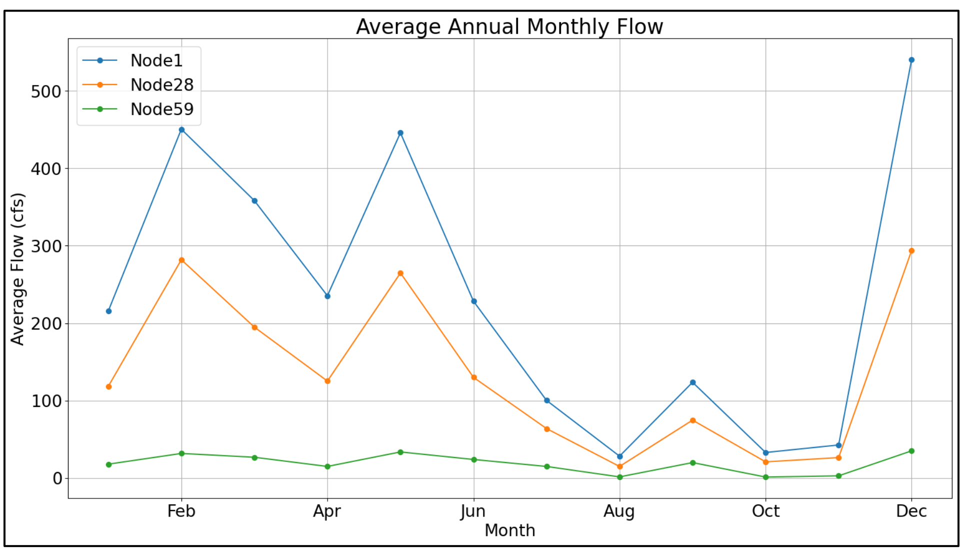

Figure 8: The average annual monthly flow can be seen in the image above. It can be seen that Node 1 and Node 28 have similar patterns, with the average flow at Node 28 being smaller in value. This implies that Node 1 is highly influenced by Node 28. Node 59 has a relatively lower average annual monthly flow every month as compared to both Node1 and Node 28.

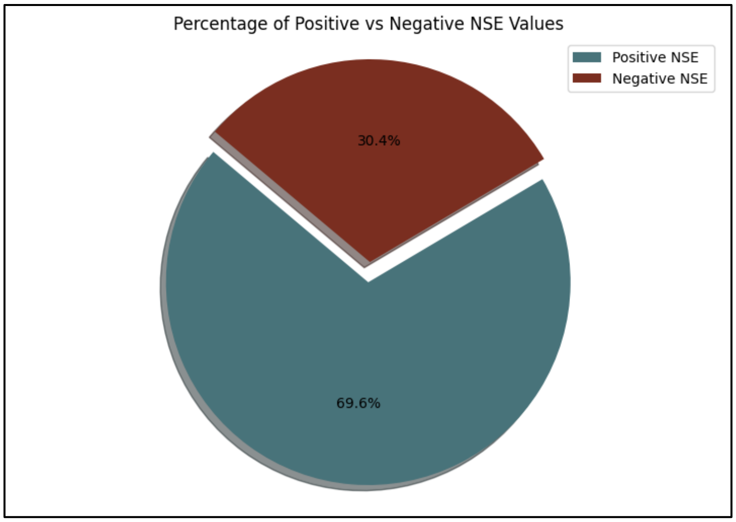

Figure 9: From the 1000 samples, 696 sets of Tc and S resulted in a positive NSE value, whereas 304 of them resulted in a negative NSE. It means that most of the sample pairs correspond to positive NSE. And can imply that our model is working fine.

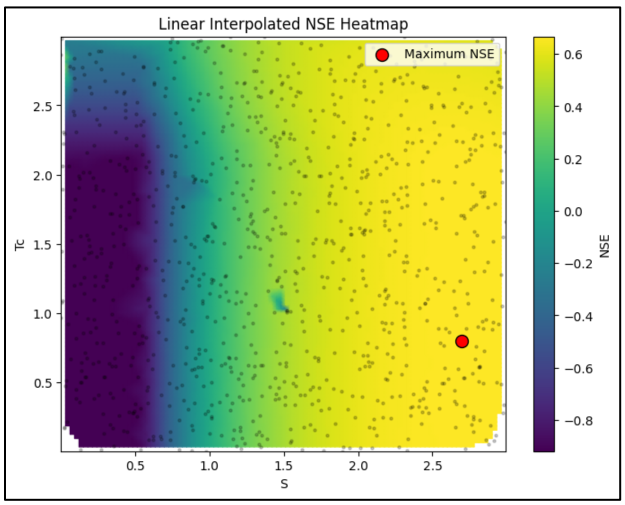

Figure 10: The red dot shows a point of Maximum NSE: 0.667 at Tc = 0.79, S = 2.69. At the blue gradient regions (Regions where S<0.75), the NSE is negative. As the gradient slowly changes towards blue, the NSE values increase. The black dots are the sample points. The gradient variation is almost the same for a fixed Tc value; hence, it is less sensitive. Some blue-colored blobs can be seen in the figure, which might be due to the problems in interpolation which can be due to the small sample size.

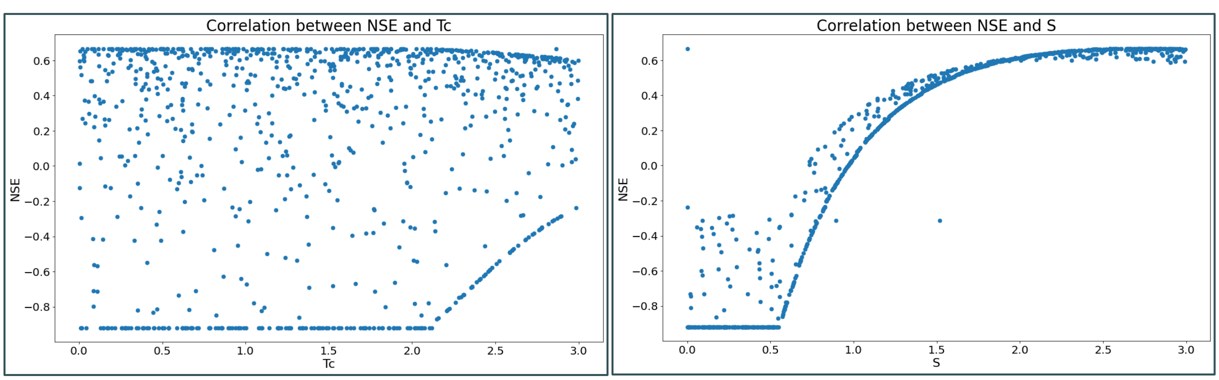

Figure 11 (a,b): The figures show the correlation between NSE and Tc and NSE and S. It can be seen in fig 12a that the variation in Tc doesn't affect the pattern in NSE. Hence it is not very sensitive. It can be seen in Fig 12b that as S increases, NSE also increases almost exponentially. Which implies that S is very sensitive.

Conclusion:

- The primary objective of the research was to leverage the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) for the prediction of streamflow dynamics and subsequently compare the model's predictions with the observed streamflow.

- Through the implementation of graphical data analysis techniques, it was ascertained that among the various streamflow monitoring nodes within the Upper Wabash-Busseron watershed, nodes 1, 28, and 59 exhibited significant streamflow volumes. Conversely, node 54, which represents a relatively minor tributary, experienced notably diminished streamflow levels.

- In the data quality checking process, checks for null values, gross errors, and range validations were performed on the streamflow, rainfall, and predicted data to ensure the integrity of the analysis.

- The maximum value attained for the Nash-Sutcliffe efficiency (NSE) metric was 0.667, which serves as an indication of the model's proficiency in accurately replicating observed streamflow patterns, thereby providing a quantitative assessment of its predictive performance.

- Sensitivity analyses conducted on the model's input parameters revealed a contrasting degree of influence exerted by the Storage Coefficient (S) and the Time of Concentration (Tc). Specifically, the Storage Coefficient exhibited a high degree of sensitivity, whereas the Time of Concentration demonstrated a comparatively lower level of sensitivity in influencing the model's output.

- Developing a comprehensive understanding of streamflow parameters is of paramount importance, as it serves as a crucial enabler for accurately predicting and mitigating the impacts of flood events, facilitating effective water resource planning and management, and addressing a multitude of other water-related challenges.

Additional Information:

This work is still under process for future steps. Data and codes are available in the GitHub repository. Python codes on automation of HECHMS are not available in the GitHub repository. They can be shared upon request.

Figure	CODE FILENAME	DATA FILENAME
FIG2	CODES/streamflow	DATA/Observed Streamflow
FIG3	CODES/event_rainfallandstreamflow	DATA/Event_rainfall DATA/Event_streamflow
FIG4	CODES/Uniform_Distribution	DATA/Predicted/Parameters_samples_1000
FIG7, FIG8	CODES/ Observed_data_StatisticsAndMetrics	DATA/Observed Streamflow
FIG9, 10, 11	CODES/ Predicted_Streamflow_Statistics	DATA/Predicted