



**THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH**

Technical Memorandum 5A, Evaluate Streamflow
Measurement at TxDOT Bridges

To: Shelley Pridgen
From: CTR RS/Research Team: P. Passalacqua, M. Bartos, A. Bibok
Subject: TxDOT Project 0-7095 – Tech Memo 5A
Date: 04/30/2023

Error Analysis and Data Assimilation

1. Introduction

Evaluation of measurements and flood forecasting models involves various spatial scales; from the gauge scale, at which measurements of stage and velocity are collected, to the river network scale, at which flood forecasting models run. In this report, we describe the evaluation of the National Water Model at the river network scale over the State of Texas to quantify the error associated with the model and investigate whether there are spatial and/or temporal patterns in the estimated error (Section 2).

Measurements of stage and velocity are fundamental to improve the performance of such hydrological models. The National Water Model integrates observations detected at USGS gauges through a simple nudging operation. However, more sophisticated Data Assimilation schemes are available that can improve the predictions of the model and thus flood predictability. We describe two of these approaches in this report; one, an operational Data Assimilation scheme developed by Kisters (Section 3), the other, a research investigation on Data Assimilation performed at the University of Texas at Austin (Section 4).

Finally, in Section 5 we present an analysis of the data collected at RQ-30 sites to quantify hysteresis patterns and a theoretical and modeling analysis of this behavior.

2. Error analysis at network scale

2.1.1. Overview

The National Oceanic and Atmospheric Administration (NOAA) has developed the National Water Model (NWM), a hydrologic model that can predict streamflow and other water-related variables throughout the United States. The model has four configurations (Analysis, Short range, Medium range, and Long range) for the Continental United States (CONUS) and provides analysis of streamflow, as well as other hydrologic states on the surface and near the surface. The NWM provides streamflow forecasts for short (next 18 hours), medium (next 10 days), and long-term (next 30 days) periods. The NWM's ability to provide streamflow forecasts over these three timescales for over 2.7 million river reaches in the United States has made it a useful tool for flood risk analysis and management purposes for emergency managers, water utilities, and other agencies and institutions.

In this project, the NWM forecasts provide a spatially continuous hydrological prediction over the watersheds in the State of Texas (190,000 miles of streams and rivers in Texas divided into 100,000 individual stream reaches). The availability of these predictions provides important information for issuing warnings and developing flood inundation mapping on roads. However, as with any model, the NWM is expected to have errors, particularly given that the model runs at the scale of CONUS simulating various subsurface and surface water processes, without

calibrating those processes and associated parameters locally. It is thus essential that as part of this project we assess the accuracy and performance of the NWM's streamflow forecasts to confidently use them for flood prediction purposes.

To quantify the NWM's performance in Texas, we conduct an error analysis of the hydrological forecast by comparing it with the streamflow measurements available at Texas gauges. We describe this analysis in the next sections.

2.1.2. Study area and data sources

As mentioned above, the NWM issues streamflow forecasts at three timescales: short range (up to 18 hours), medium range (up to 10 days), and long range (up to 30 days). Additionally, the NWM was run for the dataset of historical hydrological conditions spanning multiple decades to estimate the streamflow relative to those historical conditions. Version 2.1 of the NWM includes retrospective simulations for 42 years (February 1979 to December 2020). We use this product to perform an error analysis of the forecasted streamflow with respect to the observed one at USGS streamflow gauges. In particular, we use the initial prediction of each cycle of the short range 18-hour hourly streamflow forecasts (<https://docs.opendata.aws/noaa-nwm-pds/readme.html>) and the retrospective dataset (https://noaa-nwm-retrospective-2-1-pds.s3.amazonaws.com/index.html#model_output/), thus performing the error on two products of the NWM. This choice allows us to estimate the NWM performance both in terms of its short term forecast, which is the best resource for flood warnings, as well as the performance of the model when run with historical data without assimilating any information from the USGS gauges. As mentioned earlier, the NWM short term forecast does integrate streamflow information at USGS gauges. The operation performed is simply nudging (see also Section 3 for further discussion on nudging), according to which the model is adjusted to the observed streamflow at each USGS gauge and then run forward in time. Since the retrospective product is run over the historical record, no nudging is done in the model, which allows us to quantify the error in the NWM with nudging (short term forecast) as well as without nudging (retrospective product).

The United States Geological Survey (USGS) collects observations of stage and streamflow across thousands of gauging stations on rivers in the United States. Based on the available data, we have information at 553 USGS gauges distributed throughout Texas (Figure 1, <https://waterdata.usgs.gov/nwis>) at which we performed a comparison of the measured and the forecasted streamflow.

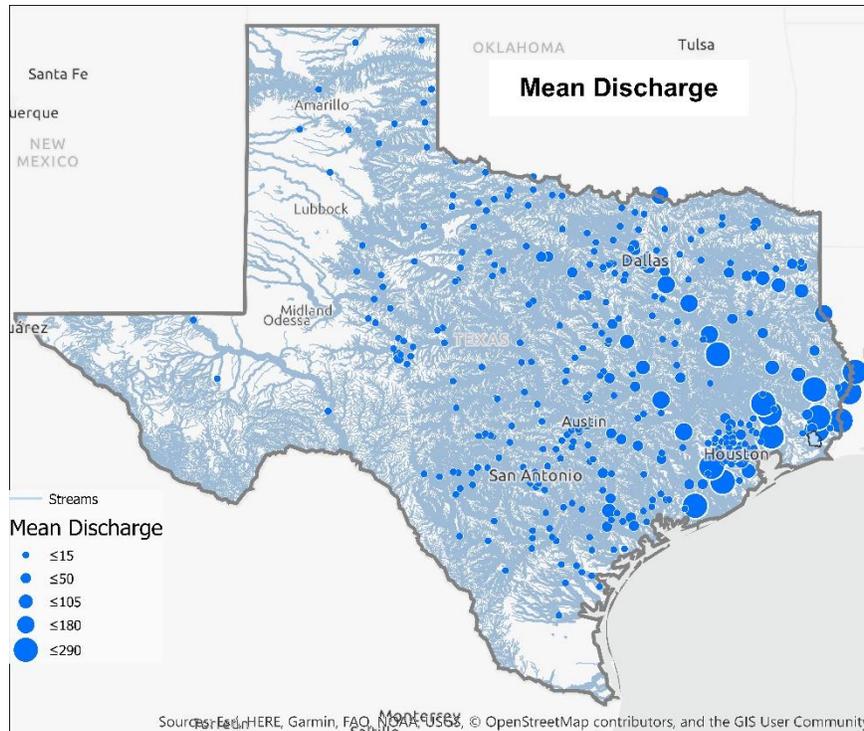


Figure 1. Location of the USGS gauges considered for the error analysis. The size of the dot indicates the magnitude of the mean annual discharge.

2.1.3. Error metrics and performance of National Water Model in Texas

To assess the NWM's performance, we compare the forecasted streamflow with the observed USGS streamflow data at the locations included in the study. While the USGS dataset includes both stage and streamflow data, the NWM provides only streamflow. Thus, we compute the error metrics only of the streamflow variable.

We find that the difference between the NWM forecasts and the observed streamflow (Figure 2) is lower during normal conditions (i.e., those without a storm) compared to storm conditions, a trend that is observed in most locations. In the example shown in Figure 2, the NWM underestimates streamflow, but this is not the case at all locations. In some locations, the NWM is able to accurately capture the time and magnitude of peak discharge and in other cases it overestimates streamflow. In some cases, the performance of the short term NWM is better at locations that are close to or at USGS gauges. This result is likely due to the nudging of the NWM with USGS observations.

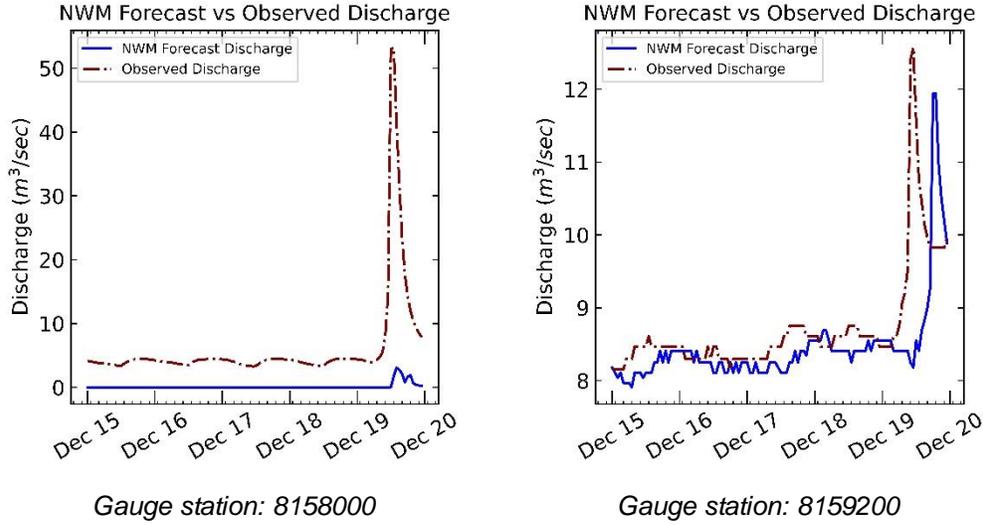


Figure 2. Example comparison of the National Water Model forecast and observed discharge at two USGS gauging stations in the Lower Colorado River for an event during December 2020.

We estimate the error between the NWM streamflow forecasts and the observed streamflow data in terms of two metrics: the root mean square error (RMSE) (in discharge units) and percent bias (dimensionless). We compute the RMSE and percent bias as:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(NWM \text{ streamflow} - \text{Observed streamflow})^2}{n}} \quad (2.1)$$

$$\text{Percent Bias} = \frac{\sum_{i=1}^n (NWM \text{ streamflow} - \text{Observed streamflow})}{\sum_{i=1}^n \text{Observed streamflow}} * 100 \quad (2.2)$$

Under low to median discharge conditions, the NWM has lower error than during storm conditions. As we are mostly concerned with flood conditions, and the ability of the model to predict them, we define a discharge threshold to select events of interest and compute the error metrics only for those events. In order to define this threshold based on the historical record, we analyze the annual maximum streamflow data from the retrospective dataset, which provide us discharge data for 42 years (1979-2020). We define the minimum annual maximum flow with a return period of 1.02 years from the 42-year as the threshold value for each location to maximize data availability and ensure an adequate amount of data for the error estimation.

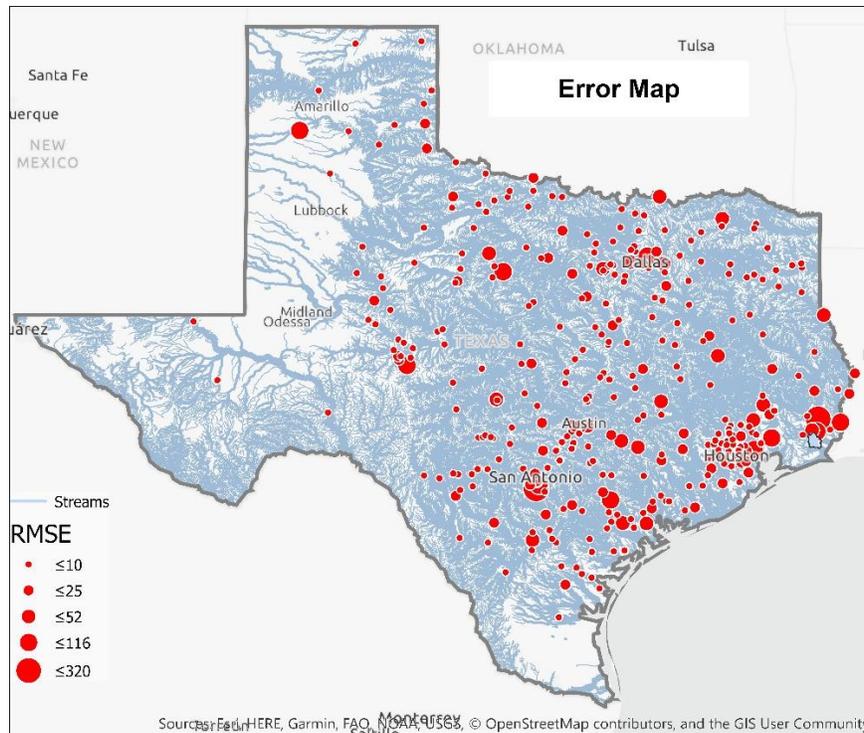


Figure 3. RMSE map for Texas for period 2020-2022. Dots represent gauge locations. The size of the dot indicates the magnitude of the RMSE (in m^3/s).

We compute the RMSE and percent bias for the short-range and retrospective streamflow for the time period between 2020 and 2022. In the future, the analysis will be expanded to the complete available record, but at this time it has been performed only for the events that happened in this time window. We compute the error metrics at all 553 USGS gauges and generate maps to analyze the spatial pattern of the error metrics throughout Texas (Figure 3 and Figure 4). The average and median RMSE are $10.19 m^3/s$ and $3.39 m^3/s$, respectively. The RMSE for the period 2020-2022 (Figure 3) shows higher errors in larger cities and the southeast region of Texas, which is more susceptible to extreme weather events.

The percent bias does not show spatial patterns (Figure 4) for the analyzed period from 2020 to 2022. *This is an important result as it shows that there is no particular spatial pattern to the errors and thus that there is no evidence that the model is less accurate in one climatic region of Texas compared to another.* The average and median percent bias are 16.57% and -0.55%, respectively. Out of the 553 gauge locations, the NWM overestimates discharge in 147 locations and underestimates it in 204 locations. At the remaining 202 gauges, the discharge did not exceed the threshold value over the 2020-2022 time period and we excluded them from the analysis. These results indicate that, on average, the NWM underestimates discharge, at least over the analyzed 2020-2022 time period.

We additionally performed a comparison of predicted and measured discharge during two historical events, Tropical Storm Imelda and Hurricane Hanna, using in the analysis both the

short-range forecast and retrospective data. As the retrospective data are not available after 2020 and short-range data are not available before 2018, we limited our analysis to these two specific storm events. For TS Imelda, the mean RMSE for short-range streamflow is 47.94 m³/s, with a median of 15.19 m³/s, a minimum RMSE of 0.031 m³/s and a maximum RMSE of 333.66 m³/s. In comparison, the mean and median RMSE for the retrospective data are 87.36 m³/s and 19.13 m³/s, respectively, with a minimum RMSE of 0.0014 m³/s and a maximum RMSE of 1169.90 m³/s. Similarly, for Hanna, the mean and median RMSE for short-range streamflow were 21.46 m³/s and 5.62 m³/s, respectively, with a minimum RMSE of 0.0035 m³/s and a maximum RMSE of 261.52 m³/s. The mean and median RMSE for retrospective streamflow were 24.57 m³/s and 5.36 m³/s, respectively, with a minimum RMSE of 0.0021 m³/s and a maximum RMSE of 240.93 m³/s.

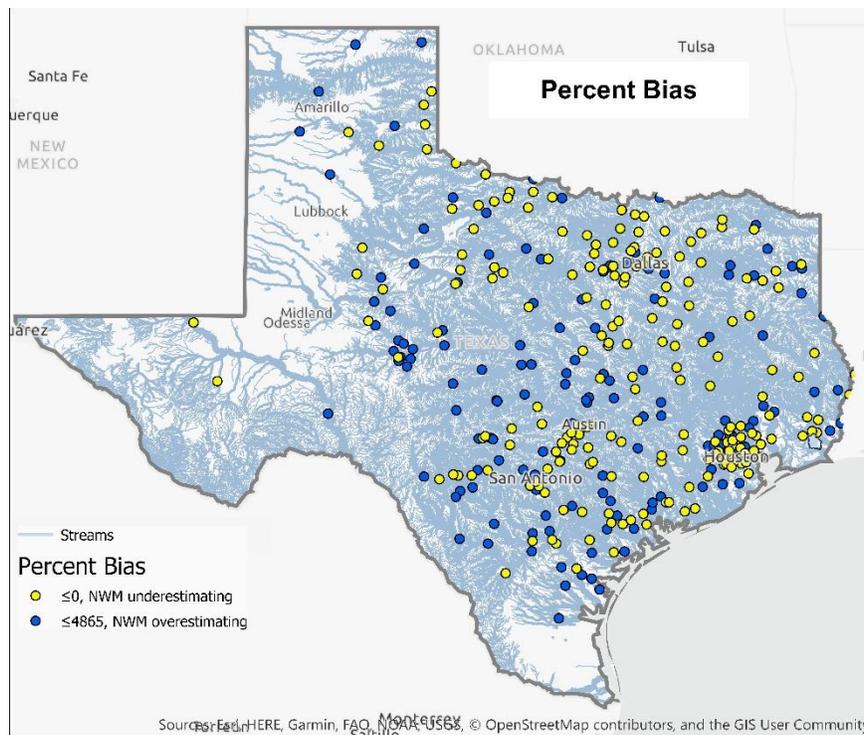


Figure 4. Percent Bias map for Texas for period 2020-2022. Dots indicate the location of the gauge. The color indicates whether the NWM is underestimating (yellow) or overestimating (blue) with respect to the observed discharge.

2.1.4. Correlation of error with hydrological and river network properties

In our analysis of the NWM error, we are also interested in quantifying whether the NWM error is correlated with precipitation and river network properties. Specifically, we analyze the NWM error and precipitation patterns of three storm events, Hurricane Harvey, Hurricane Hanna, and Tropical Storm Imelda, chosen based on the availability of retrospective precipitation data. We

obtained precipitation data from the Multi-Radar/Multi-Sensor Quantitative Precipitation Estimation (MRMS QPE) from Iowa State University's Iowa Environmental Mesonet (<https://mesonet.agron.iastate.edu/nws/>), with a horizontal resolution of 1 km (Zhang, et al., 2016).

Overlaying RMSE and percent bias with precipitation data for the three storms (Figure 5, Figure 6, Figure 7) shows that the areas with higher precipitation have greater magnitudes of RMSE. Similar observations can be made for Hurricane Hanna, where regions with higher levels of precipitation have higher values of RMSE. Thus, it can be concluded that, overall, the NWM retrospective streamflow exhibits greater errors in areas that experienced higher precipitation.

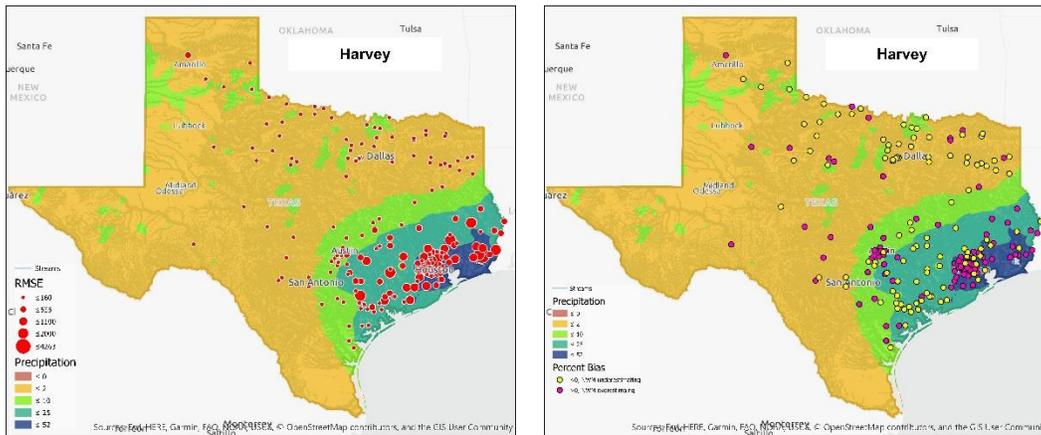


Figure 5. RMSE and percent bias overlaid on a map of precipitation (in inches) for Hurricane Harvey.

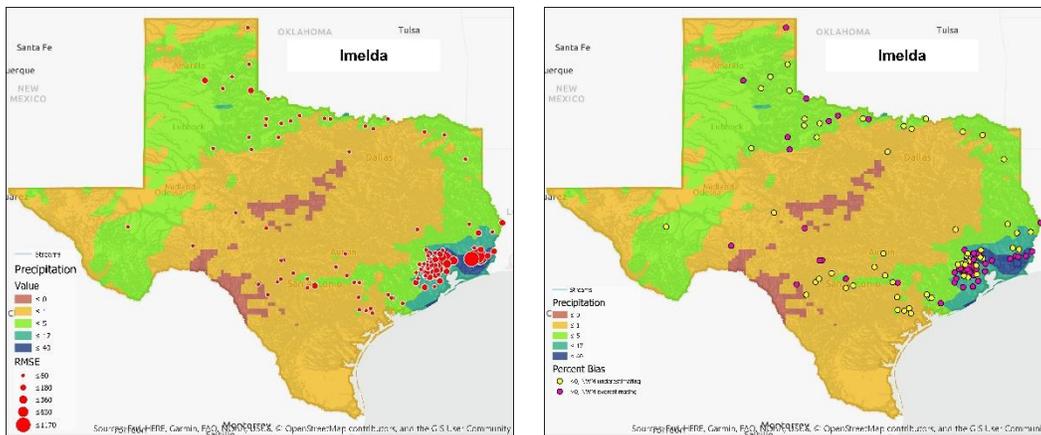


Figure 6. RMSE and percent bias with precipitation(in inches) for TS Imelda.

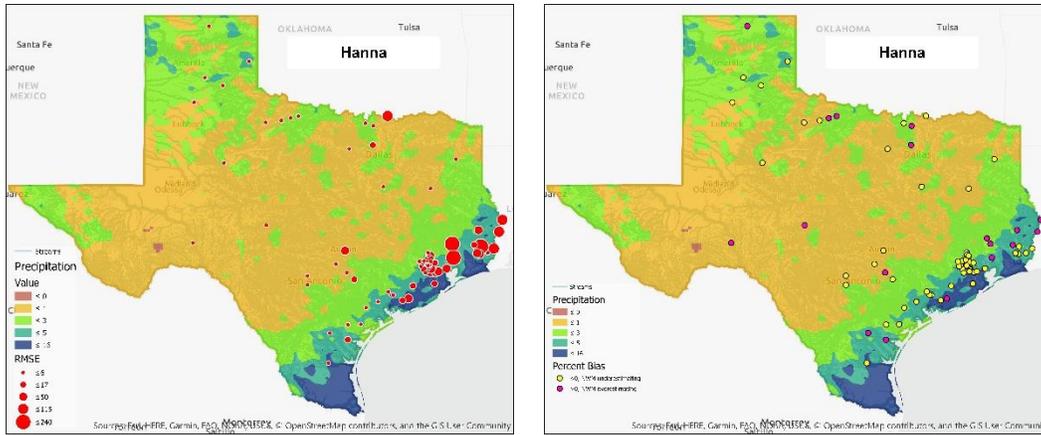


Figure 7. RMSE and percent bias with precipitation (in inches) for Hurricane Hanna.

We investigate the possible relationship between RMSE and river network properties in terms of mean annual discharge and stream order. When RMSE is plotted with respect to mean annual discharge (Figure 8), we find a positive relationship between the two variables, for the 2020-2022 time period analyzed. In particular, we find the following statistically significant relationship between RMSE and mean discharge:

$$\log(\text{rmse}) = 0.34 * \log(\text{discharge}) + 0.914$$

which allows us to predict the RMSE expected for a given discharge value.

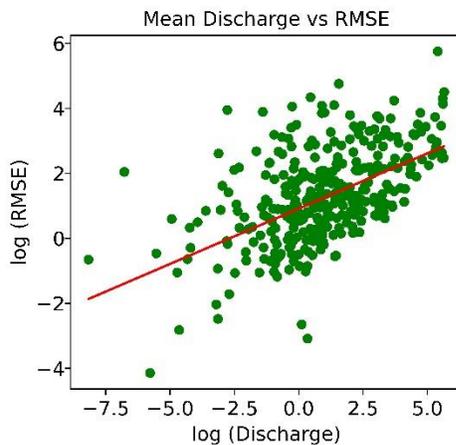


Figure 8. RMSE versus mean discharge.

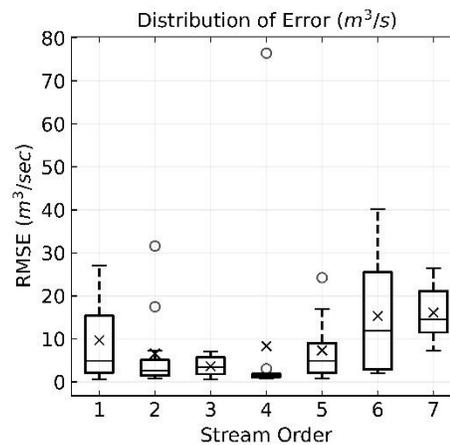


Figure 9. Box plot of RMSE versus stream order.

When we analyze RMSE with respect to stream order (Figure 9), we find that locations with stream order 4 have the lowest RMSE, whereas those with stream orders 6 and 7 have the highest RMSE values. We are currently investigating possible reasons behind this result.

2.1.5. Comparison of statistical distributions and frequency analysis

As previously mentioned, we are particularly interested in the performance of the NWM in terms of ability to accurately forecast peak flows. We compute and analyze the statistical distribution of the NWM and observed annual peak flows to assess whether or not the distributions are similar. While RMSE and percent bias provide one value to assess the similarity of the NWM and the observed flows, the analysis of the statistical distribution provides information on the full distribution of the predicted and observed flows. It is important to be able to distinguish between the cases in which the NWM's value may be off, but the overall distribution may be similar, with respect to cases in which also the distribution of the predicted flows may be off with respect to the observed one.

We analyze again the retrospective dataset from version 2.1 since it contains 42 years of retrospective data. We downloaded the observed peak flow data from 555 USGS gauges that had more than 15 years of available data to guarantee the availability of enough observations to compute reliably a statistical distribution. To quantify the similarity (or difference) of the two distributions (predicted and observed), we perform a Mann-Whitney U test, a non-parametric test commonly used to compare two datasets and determine whether they are from the same distribution. The test's null hypothesis is that both datasets come from the same distribution (thus the predicted and observed flows are similar), while the alternative hypothesis is that the two datasets do not come from the same distribution (thus the predicted and observed streamflows are different). We considered two significance levels ($\alpha = 5\%$ and $\alpha = 1\%$). The p-value represents the believability of the hypothesis; a p-value smaller than the significance level leads to a rejection of the null hypothesis. The smaller is the p-value, the stronger is the evidence against the null hypothesis.

The results show that out of the 555 gauge locations, 367 had p-values below 0.05 and 312 had p-values below 0.01 (Figure 10). This suggests that, at the 5% significance level, only 33.8% of locations have similar distributions, and at the 1% significance level, only 43.78% of locations have similar distributions. Thus, the distribution of the NWM predicted the annual peak flow is similar to the distribution of the observed streamflow in less than 44% of the locations analyzed.

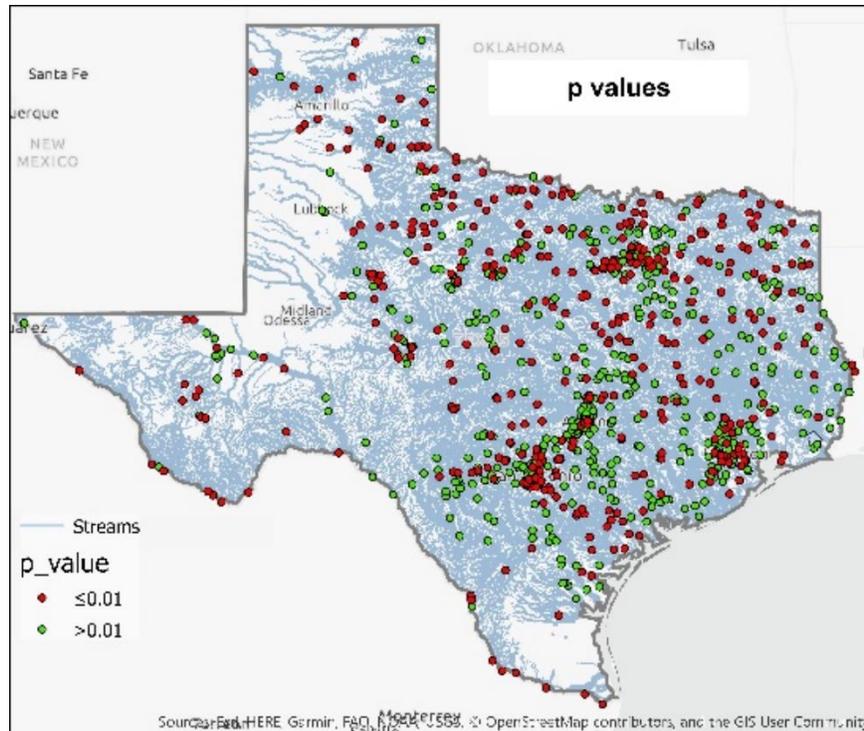


Figure 10. Map of p values for Mann-Whitney U test, with significance level $\alpha = 1\%$. The dots represent gauge locations and the color of the dot indicates where the distribution of the NWM streamflow is similar to distribution of the observed streamflow (green) or different (red).

Additionally, we performed a frequency analysis to compare the predicted and observed streamflows for a range of return periods. Figure 11 and Figure 12 show two example distributions, one with p-value of 0.99, indicating that the annual peak flow of both USGS and NWM has very similar distributions (Figure 11) and one with a p-value of 0, indicating that the USGS and NWM peak flows have different distributions (Figure 12).

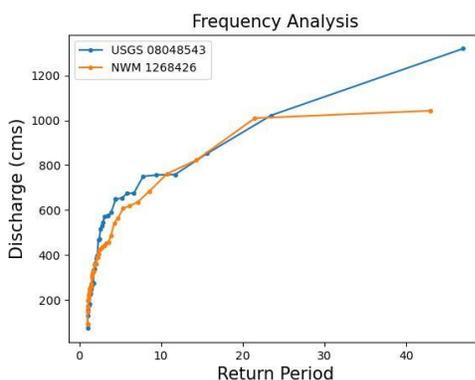


Figure 11. Frequency distribution for $p = 0.99$.

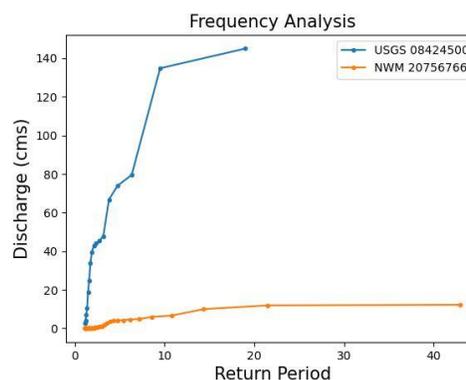


Figure 12. Frequency distribution for $p = 0$.

3. Operational Data Assimilation at network scale

In this section of the report we present the Data Assimilation we are performing to improve the performance of the NWM based on the observed data. We provide in this section a description of the KISTERS Real-time Optimization (RTO) optimization model driven by IPOPT (Interior Point Optimizer), covering the underlying mathematical formulations of the optimization problem, the governing equations of the model library, as well as its numerical schematization.

3.1. Methodology

3.1.1. Optimization problem

RTO-IPOPT (Real-time Optimization Interior Point Optimizer) implements a standard form of a continuous nonlinear optimization problem according to:

$$\min J(x, y) = J_1 + J_2 + \dots + J_n \quad (3.1)$$

subject to

$$\begin{aligned} x_{lb} &\leq x \leq x_{ub} \\ g_{lb} &\leq g(x) \leq g_{ub} \end{aligned} \quad (3.2)$$

where x is a variable of the optimization problem, $y = f(x)$ an arbitrary model, J is a sum of n objective function terms, x_{lb} , x_{ub} are the lower and upper bounds of a variable, $g()$ is an arbitrary function of x with the lower and upper bounds g_{lb} and g_{ub} , respectively.

In RTO, the solution of the nonlinear optimization problem in Equations 3.1 and 3.2 is performed by IPOPT (Interior Point OPTimizer), a software library for large-scale nonlinear optimization of continuous systems. IPOPT is part of the Computational Infrastructure for Operations Research (COIN-OR) project aiming at "Open Source for the Operations Research Community" (<https://www.coinor.org/>). IPOPT is open source and available on <https://github.com/coin-or/Ipopt>. Mathematical details of the algorithm can be found in Nocedal et al. (2008), Wächter (2002), and Wächter (2006).

The most commonly used objective function term in RTO is a quadratic or least-square penalty according to its most generic formulation:

$$J = w \begin{cases} (x_{lb} - x)^2 & x \leq x_{lb} \\ 0 & \text{if } x_{lb} < x \leq x_{ub} \\ (x - x_{ub})^2 & x_{ub} < x \end{cases} \quad (3.3)$$

where w is a weighting coefficient, x_{lb} and x_{ub} are the lower and upper bound of the penalty, and x is a single variable or a sum of n variables according to $x = \sum x_n$. Equation 3.3 can be used to penalize only a down-crossing of x_{lb} or an up-crossing of x_{ub} by setting the opposite bounds to $x_{ub} = -inf$ and $x_{lb} = -inf$, respectively.

For $x_b = x_{lb} = x_{ub}$, Equation 3.3 reduces to

$$J = w(x - x_b)^2 \quad (3.4)$$

and constitutes an important penalty function term for enforcing a variable on set points, e.g. to maintain target reservoir levels.

A corresponding, generic linear version of 3.3 is:

$$J = w \begin{cases} |x_{lb} - x| & x \leq x_{lb} \\ 0 & \text{if } x_{lb} < x \leq x_{ub} \\ |x - x_{ub}| & x_{ub} < x \end{cases} \quad (3.5)$$

However, this formulation is not favorable in a non-linear optimization due to the discontinuities in its first-order derivative. A more suitable formulation is achieved by the simplification of $x_b = x_{lb} = x_{ub}$ to obtain:

$$J = w(x - x_b) \quad (3.6)$$

as a suitable term to maximize a variable over the optimization horizon. Besides the value of a variable, we may want to penalize the variable's change over time. The available options include a penalization of the change of a variable from its previous time stamp t_{old} to its successor at t_{new} by:

$$\Delta = x_{new} - x_{old} \quad (3.7)$$

Whereas this formulation considers the absolute change, we may want to use the time derivative according to:

$$\Delta = \frac{x_{new} - x_{old}}{\Delta t} \quad (3.8)$$

where $\Delta t = t_{new} - t_{old}$. Both formulations are equivalent, but the 2nd one is more generic if the model runs with different time steps.

Another option in RTO is to refer to the relative change of the variable from one time step to another. This leads to the formulation:

$$\Delta = \frac{x_{new}}{x_{old}} \quad (3.9)$$

and may serve as a suitable approach to limit the relative increase or decrease of a variable. Table 3-1 provides an overview of the available options based on the simpler formulations in Equations 3.4 and 3.6. For completeness, this includes the "linear-absolute" option as a variation of the linear option by penalizing the set point deviation by an absolute norm. Note that this function is discontinuous in its first-order derivative and not suitable in a nonlinear optimization setup.

Table 3-1: Overview objectives.

	Derivative order = 0	Derivative order = 1 (first order derivative)
Linear	$J = w(x - x_b)$	Absolute: $J = w(x - x_{old} - x_b)$ Time derivative: $J = w\left(\frac{x - x_{old}}{\Delta t} - x_b\right)$ Relative: $J = w\left(\frac{x}{x_{old}} - x_b\right)$
Linear-absolute	$J = w x - x_b $	Absolute: $J = w x - x_{old} - x_b $ Time derivative: $J = w\left \frac{x - x_{old}}{\Delta t} - x_b\right $ Relative: $J = w\left \frac{x}{x_{old}} - x_b\right $
Quadratic	$J = w(x - x_b)^2$	Absolute: $J = w(x - x_{old} - x_b)^2$ Time derivative: $J = w\left(\frac{x - x_{old}}{\Delta t} - x_b\right)^2$ Relative:

		$J = w \left(\frac{x}{x_{old}} - x_b \right)^2$
--	--	--

A special option in the quadratic function is the optional definition of a transition threshold to a linear function. As an example, this feature is useful and advised in a data assimilation setup to not overweight outliers. Equation 3.4 extends to:

$$J = w \begin{cases} -2x_{th} * (x - x_b + x_{th}) + x_{th}^2 & x \leq x_{lb} - \Delta x_{th} \\ (x - x_b)^2 & \text{if } x_b - \Delta x_{th} < x \leq x_b + \Delta x_{th} \\ 2x_{th} * (x - x_b - x_{th}) + x_{th}^2 & x_b + \Delta x_{th} < x \end{cases} \quad (3.10)$$

Constraints are conditions of an optimization problem that a solution must satisfy. Therefore, we consider constraints only if they do not jeopardize the feasibility of the optimization problem. The implementation of constraints in RTO relies on the definition of lower and upper bounds for a variable or its first order derivative. An overview is provided in *Table 3-2*.

Table 3-2: Overview Constraints.

Derivative order = 0	Derivative order = 1 (first-order derivative)
$x_{lb} \leq x \leq x_{ub}$	<p>Absolute:</p> $g_{lb} \leq x - x_{old} \leq g_{ub}$ <p>Time derivative:</p> $g_{lb} \leq \frac{x - x_{old}}{\Delta t} \leq g_{ub}$ <p>Relative:</p> $g_{lb} \leq \frac{x}{x_{old}} \leq g_{ub}$ <p>Equivalent to</p> $x - g_{lb}x_{old} \geq 0$ $-g_{ub}x_{old} \leq 0$

Note that one can define more complex, dedicated constraint functions as part of the model library and constrain the resulting variables by the features above.

3.2. Hydrologic routing

Hydrologic routing methods are based on the principle of conservation of mass and a relation of storage and discharge by river reach. Hydrologic flow routing can impact the magnitude, velocity and shape of the flow as a function of time at one or more points along the water course (USDA-NRCS, 2014). While carrying out flow routing in a channel, the input data are inflow, lateral contributions, and channel characteristics. Required data and the computational procedure of hydrologic routing are simpler and less accurate than those of hydraulic routing, however, also more robust and computationally more efficient. Conservation of mass is expressed as:

$$\begin{aligned} \frac{dS}{dt} - I + Q &= 0 \\ S &= f(Q) \end{aligned} \quad (3.11)$$

$$\frac{S - S_{old}}{dt} - (1 - \theta)(I_{old} - Q_{old}) - \theta(I - Q) = 0 \quad (3.12)$$

where θ is a time weighting coefficient in the range [0.5,1]. The configuration of θ enables us to switch between a more accurate, second-order time weighting scheme ($\theta = 0.5$) and a fully implicit, first-order scheme ($\theta = 1$).

3.2.1. Reservoir routing

Reservoir routing is based on the mass balance equation according to Equation 3.12. The approach in RTO is referred to as “uncontrolled outflow” routing (Akram et al., 2014). It simulates the flow routing in a river reach by a cascade of reservoirs to mimic the natural flow propagation. Therefore, the reservoir storage is represented as a function of flow according to:

$$S = kQ^m \quad (3.13)$$

where Q is the flow in a river reach, k is the storage coefficient and m is the non-linearity exponential coefficient. If m equals to 1, the non-linear scheme become a simple linear reservoir routing (Akram et al., 2014).

3.2.2. Muskingum routing

The Muskingum approach has been developed in 1930s for flood protection in the Muskingum River Basin, Ohio (Ponce, 1989). It is one of the most used hydrologic routing methods. It is fully mass conservative method where storage is a function of both inflow and outflow. It is also based on the differential equation of storage (Equation 3.12). The reservoir routing method assumes a constant pool level whereas the Muskingum method assumes a sloping water surface. In the Muskingum method, storage is a linear function of inflow and outflow according to:

$$S = K[XI + (1 - X)Q] \quad (3.14)$$

where K is a storage constant $K = L/V_w =$ travel time of the flood wave through routing reach, L is the reach length, V_w is wave velocity, X is a dimensionless weighting factor ($0 \leq X \leq 0.5$).

The Muskingum method can be further improved by the Muskingum-Cunge extension deriving the parameters K and X from hydraulic characteristics of the river, making the scheme comparable to the accuracy of hydraulic models for free-flowing river reaches. However, the Muskingum approach does not have the capability of a hydraulic approach to compute backwater or flow diversions at river/channel bifurcations.

The application of the Muskingum and Muskingum-Cunge approaches depend on the fulfillment of criteria on the relation between the spatial and temporal schematization as well as flow gradients. This condition makes the schemes an optional choice in RTO for higher resolution, short-term simulation/optimization runs. Depending on the choice of the hydrological scheme, the flow propagation characteristics can be derived largely from physical characteristics, e.g. for the Muskingum-Cunge scheme, or rely more heavily on calibration.

3.3. Excess rainfall optimization (4DVAR)

The biggest shortcoming of the existing NWM 2.1 hydrologic routing model is the nudging approach used during data assimilation, as the measured streamflow only affects the discharge downstream of the gauge at which the model adjustment is made. Nudging is a technique commonly used in hydrological routing to adjust the output of a hydrological model to match observed or desired values. While nudging can be effective in certain situations, there are also several drawbacks to this approach while using it in hydrological routing:

- **Loss of information:** Nudging can result in the loss of information about the underlying physical processes that drive hydrological systems. By forcing the model to match observed or desired values, the model may not accurately represent the true behavior of the system, which can lead to erroneous conclusions.
- **Bias:** Nudging can introduce bias into the model output by favoring certain observed or desired values over others. This bias can be difficult to quantify or remove, and can lead to inaccurate conclusions.
- **Lack of transparency:** Nudging can make it difficult to understand how the model is producing its output, as the model is no longer solely based on physical principles. This lack of transparency can make it difficult to assess the model's accuracy and reliability, and can limit its usefulness for decision-making.
- **Limited applicability:** Nudging can only influence streamflow downstream of the observed stream. The streams above the last observed stream solely rely on the runoff and groundwater model.

In order to overcome these limitations, we defined data assimilation of discharge observations as an optimization problem. We estimate the excess rainfall hydrograph over the simulated watershed, in order to obtain the same discharge timeseries at the observed streamflow gauges. This optimization scheme is also referred to as 4DVAR. Four-dimensional variational data assimilation (4DVAR) is a mathematical technique used to improve the accuracy of numerical models by combining model simulations with observational data. It is commonly used in atmospheric and oceanic sciences to improve weather and climate forecasts. In 4DVAR, a numerical model of the system being studied is used to simulate the evolution of the system over a specific period of time. The model's outputs are then compared to observed data collected over the same period. The goal of 4DVAR is to adjust the initial conditions and model parameters so that the model outputs match the observed data as closely as possible. The process of adjusting the initial conditions and model parameters is an optimization problem. The objective function measures the difference between the model outputs and the observed data, and the optimization algorithm adjusts the model inputs to minimize this difference.

One of the key features of 4DVAR is that it considers the temporal evolution of the system being studied. This means that not only the initial conditions but also the model parameters are adjusted over the entire period of time considered. This makes 4DVAR different from simpler data assimilation techniques, such as 3DVAR, which only adjust the initial conditions.

After the data assimilation period is simulated, the volume stored in the streams is routed out of the system. Since every boundary condition from this point is considered a forecast, two options are available:

- route out the volume stored in the stream network
- apply the forecasts provided by the NWM at the boundary conditions.

The described approach relies solely on the discharge measurements and streamline data acquired from the NHD dataset, also used by the NWM. This data-driven approach puts emphasis on the quality and the availability of accurate discharge measurement at varying flows. Since the approach's sensitivity is high to the quality and quantity of the discharge observations available, preprocessing and validation of the input timeseries is necessary. Ensuring that the data assimilation setup is matched by the data available requires setting up a computational pipeline, as explained in the next section.

3.4. Solution Architecture

A data assimilation run can be triggered by giving the t_0 value. The time defined by t_0 is the time horizon to which the data assimilation and forecast windows are defined. This approach allows running arbitrary hindcasting runs on historic data as well, by giving a past t_0 . The data assimilation setup is run on a separate node in a Kubernetes cluster. This makes the data assimilation setup highly scalable. Multiple data assimilation setups can be run in parallel, with

different setups for different spatial and temporal scale. The data assimilation and forecast timeseries are uploaded to a temporary AWS bucket. The timeseries are associated with the streamline COMID also used by the NHD. This storage space can be accessed by Datasphere, and Datasphere’s ETL can ingest the results. The results are displayed in KISTERS’s Datasphere (Figure 13), by selecting the NHD COMID and the time horizon of the assimilation run (t_0).

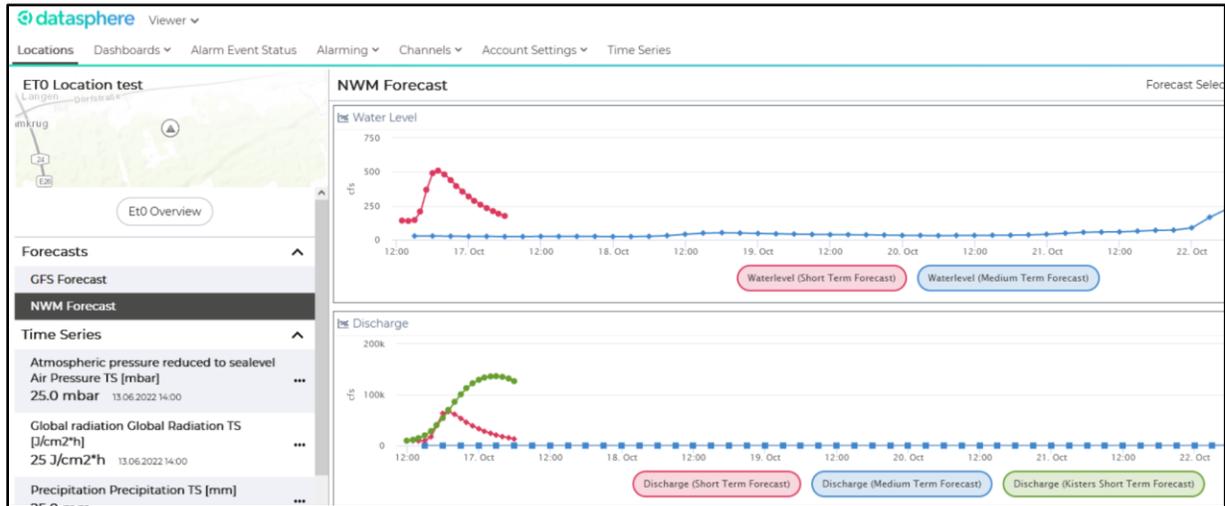


Figure 13. Adjusted forecasts displayed in Datasphere.

All of the services mentioned above are supplied as containerized REST services. Docker is a containerization technology that allows developers to create, deploy, and run applications in isolated containers. A Docker container is a lightweight, standalone, and executable package of software that contains all the necessary dependencies and libraries needed to run the application. Containers can run on any operating system and can be easily deployed and scaled. Docker containerization provides several benefits, including:

- **Portability:** Docker containers are portable, which means they can be run on any infrastructure that supports Docker. This characteristic makes it easy to deploy and move applications between different environments, such as development, testing, and production.
- **Isolation:** Each Docker container is isolated from other containers on the same system, which means that applications running in one container cannot interfere with applications running in another container.
- **Resource efficiency:** Docker containers are lightweight and consume fewer resources than virtual machines. This means that more containers can be run on the same hardware, which results in better resource utilization.
- **Consistency:** Docker containers ensure consistent deployment across different environments, which reduces the risk of configuration errors and improves application reliability.

REST (Representational State Transfer) is an architecture that is used to create web services that can be accessed over the internet. RESTful web services use HTTP methods to perform CRUD

(Create, Read, Update, Delete) operations on resources. REST APIs provide a standard interface for developers to interact with these resources. REST APIs provide several benefits, including:

- **Scalability:** RESTful web services are scalable, which means that they can handle a large number of requests without affecting performance.
- **Interoperability:** RESTful web services can be accessed using any programming language or platform that supports HTTP.
- **Flexibility:** RESTful web services are flexible, which means that they can be easily modified and extended without affecting the existing functionality.
- **Separation of concerns:** RESTful web services separate the client and server concerns, which makes it easy to modify the client and server independently.

Together, Docker containerization and REST APIs can be used to create scalable, portable, and modular software applications that can be easily deployed and managed in various computing environments. Developers can use Docker containers to package and deploy their applications, and REST APIs to provide a standardized interface for interacting with those applications. This combination allows developers to build microservices-based architectures that can be easily scaled and maintained, and enables more efficient collaboration between teams.

3.5. Automated Model generation

3.5.1. Network Tracing

An arbitrary model setup can be generated using the “`kisters.model_integration.adapters.nwm_network`” package. It is a highly customizable and flexible Python package, that generates a hydrologic and optimization model definition consumable directly by RTO. A network generation can be configured by the following configuration class:

NWMNetworkAdapterConfig contains several fields, including paths to various files and databases, definitions for hydrological features like cross-sections, catchments, and gauges, specifications for tracing upstream or downstream, settings for trimming and dropping network elements, and several other properties. The purpose of this class is to provide a configuration object for the NWM network adapter component that is used to generate a graph for hydrological modeling. It is expected that this class will be instantiated with specific values for each of its properties.

Required fields:

- `source_point_comids`: A list of NHD COMIDs which are considered as source points for tracing the network. It is traced upstream of the source streams by default.

- `nwm_streams_path`: Location of the geopackage that contains the NHD streamlines.
- `nwm_catchments_path`: Location of the geopackage that contains the NHD catchments.

Optional fields:

- `end_point_comids`: A list of NHD COMIDs, which triggers a stopping point during the network tracing. This can be used to generate a model for a mid-section of a large river basin.
- `concern_point_comids`: A list of NHD COMIDs, which should be included in the model. These streamlines cannot be trimmed.
- `trace_direction`: tracing direction from the source point(s). upstream by default.
- `trim_streamorder`: an integer of stream order under which the streamlines are excluded from the tracing.
- `gauge_comids`: a list of NHD COMIDs, which contains the available discharge measurements' location.

3.5.2. Concern Points

Concern points are designated points of interest in the network. They can be low water crossings, bridges, or any streamline of interest. The network generation considers them differently and they are protected from getting trimmed. An arbitrary list of concern points can be initially selected.

The data assimilation setup considers two types of gauges based on the network topology:

- **Boundary gauges**: Flow boundaries with the network upstream reduced into a single Flow Boundary.
- **In-line gauges**: All the gauges are considered in-line if there are streams upstream and downstream of the gauge in the generated network. Every in-line gauge has a designated subcatchment area associated with an excess rainfall hydrograph. These hydrographs are optimized during the data assimilation run.

We developed a Python package to extract timeseries from multiple sources. Archived NWM forecasts are publicly available on Amazon S3 and the GCP (Google Cloud Platform).

Discharge measurements are available from multiple sources, like USGS and river authorities, LCRA, and TRA. We obtain precipitation data from publicly available services for QPE and QPF products.

3.6. Proof of Concept

We built and ran two data assimilation models in order to prove the applicability and scalability of the excess rainfall optimization approach. We selected the Llano River and New Year Creek.

The Llano basin has multiple discharge measurements and decade-long measurement history. The New Year Creek is an area where new RQ-30 gauges have been installed and commissioned as part of this project.

3.6.1. Llano River (October 2018 flood)

The 2018 Llano River flood was a catastrophic flooding event that occurred in October 2018 in the Texas Hill Country region of the United States. The Llano River is a tributary of the Colorado River and runs through several counties, including Llano, Mason, and Kimble. The flooding event was caused by a severe weather system that brought heavy rainfall to the area. On October 15th, the National Weather Service issued a flash flood warning for the Llano River and surrounding areas. The rainfall caused the river to rise rapidly, with water levels reaching record heights. The floodwaters inundated homes, businesses, and roads, causing widespread damage and displacement. The towns of Kingsland, Llano, and Marble Falls were particularly hard hit. The flood also caused several deaths, including a family of four who were swept away by the rising waters.

We generated the network based on the available data for October 2018. The model can be loaded into any GIS application and be displayed, if geopackages are supported. Every subcatchment area (colored areas in Figure 14) has an associated excess rainfall hydrograph in the optimization setup. The data assimilation model we run (Figure 15) included the LCRA managed discharge measurements as well, resulting in a denser measurement network. The NWM only assimilated the USGS gauges.

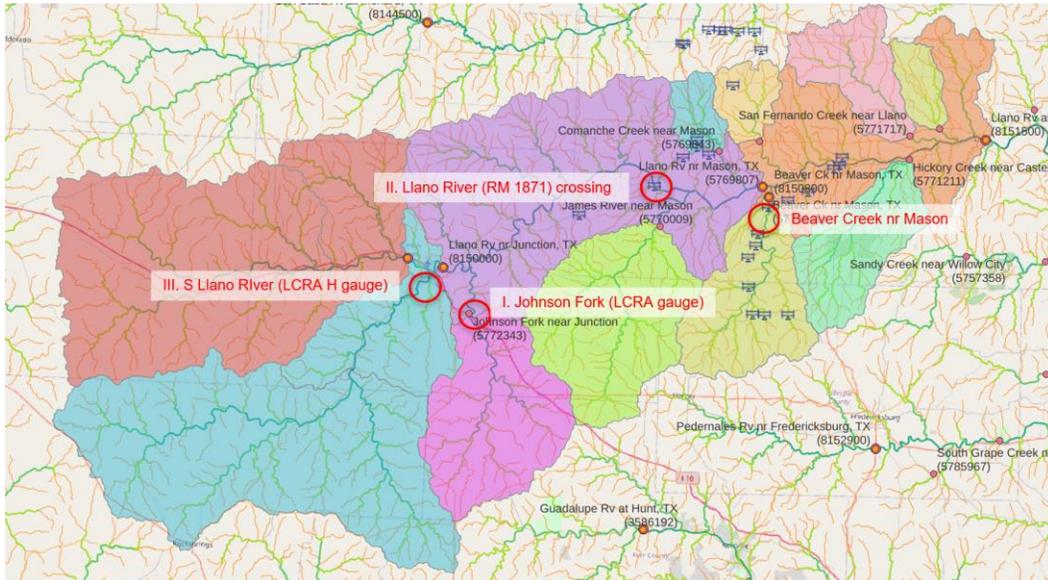


Figure 14. Llano river data assimilation network.

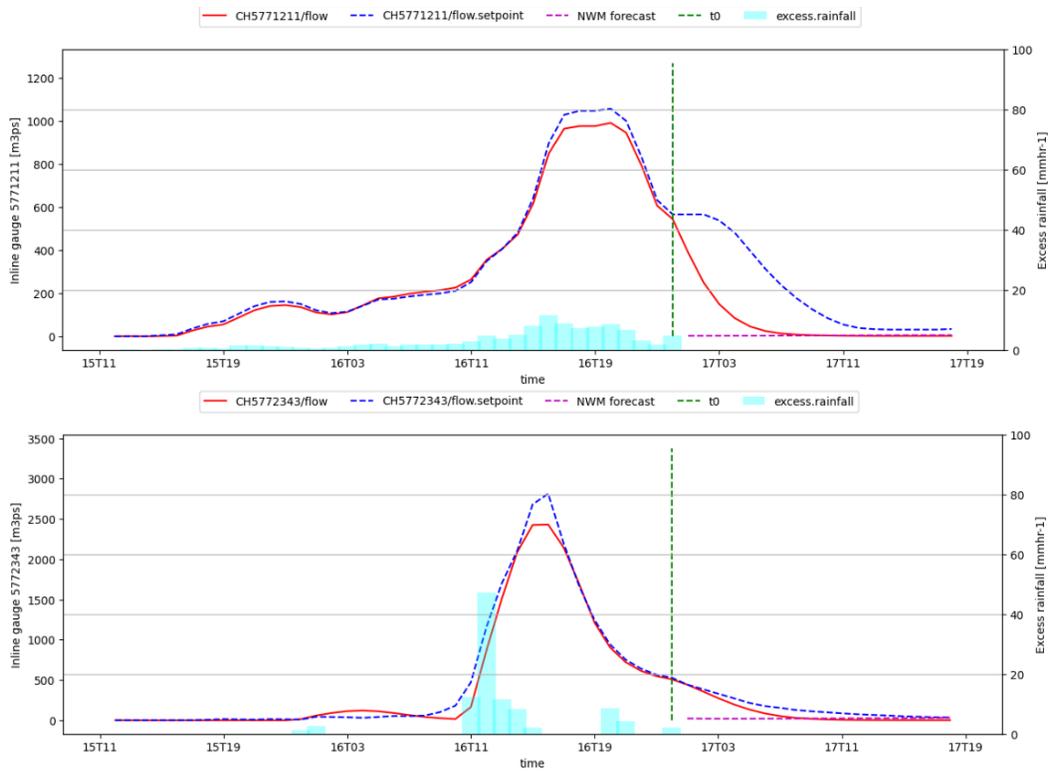


Figure 15. Data Assimilation results and adjusted forecasts at the Llano River (2018). The purple dashed line, which represents the NWM forecast ignored the rain event.

Our excess rainfall optimization approach highlights another advantage over the nudging method applied in the NWM; it makes discharge estimates more accurate even upstream of the gauges. If nudging is applied instead, and the assimilated gauges are at a large distance from each other, the discharge can be underestimated upstream of an assimilated gauge. The excess rainfall optimization approach, instead, has a domain of influence both upstream and downstream of the observed and assimilated discharge. Figure 16 shows an example of how the excess rainfall optimization provides more accurate results between two gauges, both assimilated by the NWM can, and is less susceptible to underestimating flows between gauges.

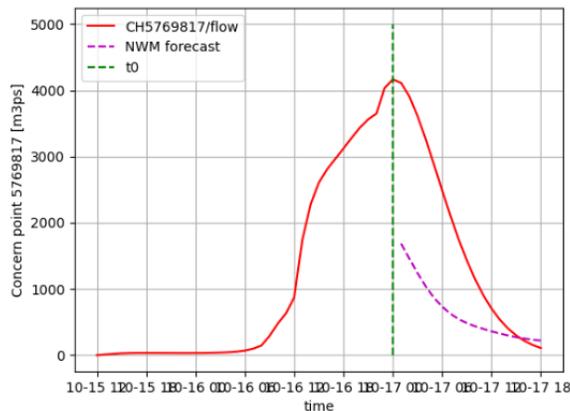


Figure 16. Llano river RM 1871 Crossing.

3.6.2. New Year Creek

The New Year Creek watershed is located in the northeastern part of Texas and covers an area of approximately 56 square miles. The watershed is situated in Collin County, and its headwaters begin near the town of Anna, Texas. The creek flows southward, passing through the city of McKinney before emptying into Lake Lavon.

The New Year Creek basin is at a much smaller spatial scale compared to the Llano basin (Figure 17). It only includes 61 streamlines compared to the Llano basin's 894 streamlines.

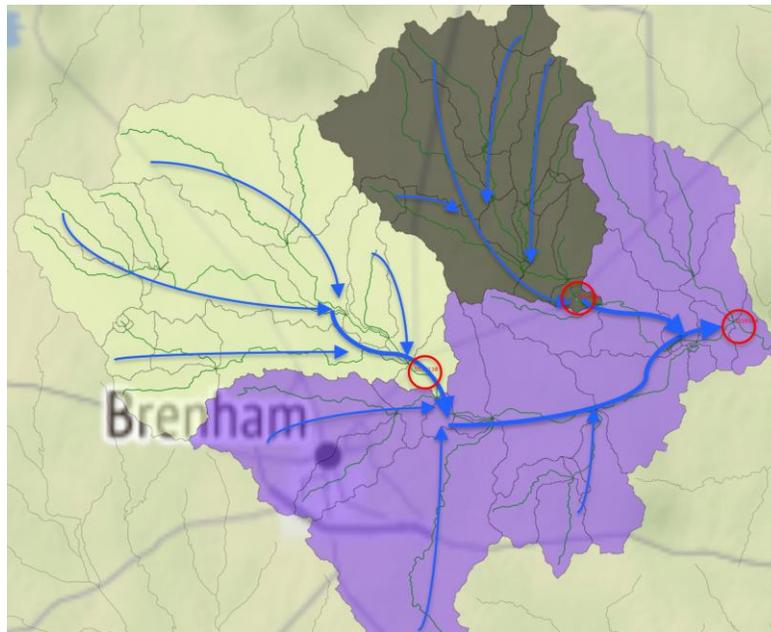


Figure 17. Data assimilation model generated for the New Year Creek.

No gauges were present in this watershed prior to the installation of sensors in this project. This explains why the NWM did not capture the rain event occurred in 2022 March (Figure 18 purple dashed line). That is a shortcoming of the nudging approach used in the NWM and that it is highly dependent on the precipitation estimate and forecast.

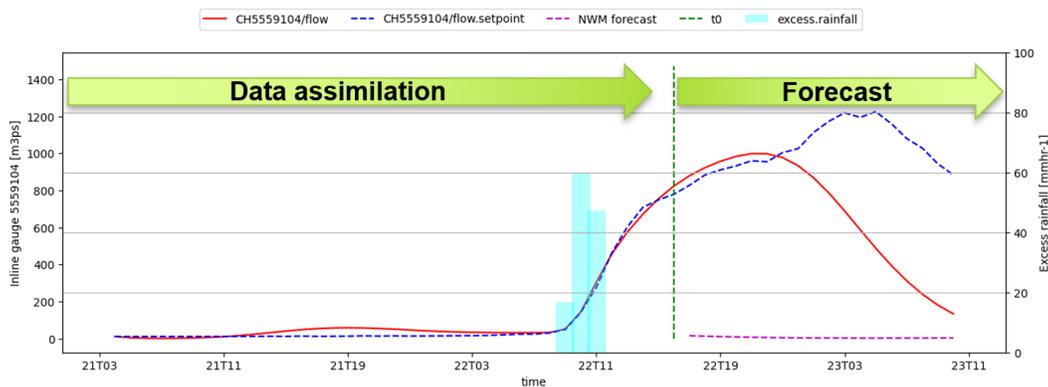


Figure 18. Data assimilation and Adjusted forecasts at New Year Creek near Chappel Hill.

4. Network-scale Data Assimilation using Saint-Venant equations with Kalman Filtering

4.1. Overview

In this section, we evaluate an alternative approach to the Kisters data assimilation approach using a physical hydraulic model based on the Saint-Venant equations for unsteady flow and using Kalman Filtering (KF) for data assimilation. Because the Saint-Venant equations are physically-based, while the Muskingum equation is conceptual, this assessment helps to validate the results of the KISTERS data assimilation procedure. This analysis also explores whether greater accuracy can be obtained using a more detailed hydraulic model, and evaluates the data requirements needed to populate such a model.

This analysis takes place in three parts. First, a physically-based model of the New Year Creek watershed is constructed using channel network topology from the NWM, and using bathymetry and roughness information from surveyed Base-Level Engineering (BLE) data available for the state of Texas, as incorporated in state-wide HEC-RAS models. Next, this channel network data is used to create a hydraulic model using a hydraulic solver based on the Saint-Venant equations. The model is then calibrated to observed stage and discharge data available at three RQ-30 gauge sites within the watershed. A holdout analysis is conducted to verify the ability for data assimilation to reconstruct river stages and discharges at ungauged locations. Specifically, the model is forced with runoff data from the NWM for a historical storm event, and data are assimilated at the two upstream sites. The assimilated model run is then compared against the observed stage and discharge data from the downstream holdout site to determine the extent to which data assimilation improves estimates of flood stage at ungauged bridges (Figure 19). From this analysis, it is concluded that data assimilation using KF improved estimates of depths and discharges downstream of sensor sites.

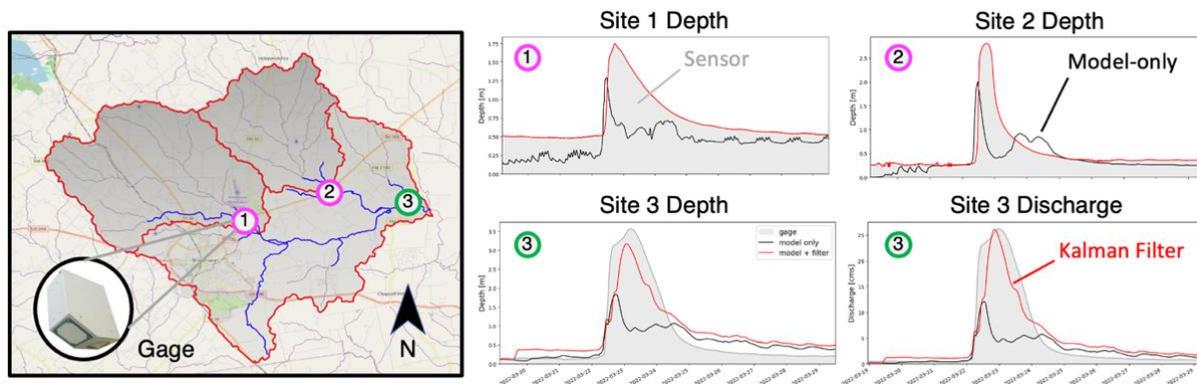


Figure 19. Overview of the data assimilation assessment with Kalman Filtering. Applying a Kalman Filter to fuse sensor data at the two upstream sites (1 & 2) improves the estimated depth and discharge at the downstream holdout site (3).

4.2. Study area

The study area chosen for this analysis is a sub-watershed of New Year Creek located near Brenham and Chappel Hill (Figure 20). The watershed is approximately 150 square miles, and hosts three RQ-30 gages located on separate tributaries. The sensor sites include New Year Creek at TX 105 near Brenham (the westernmost sensor), Rocky Creek at TX 105 near Brenham (the northernmost sensor), and New Year Creek at FM 1155 near Chappel Hill (the easternmost sensor, located near the outlet). Of these sensors, validated USGS rating curves are available for the New Year Creek at FM1155 near Chappel Hill site. This downstream site is used as the ground truth source of discharge data.

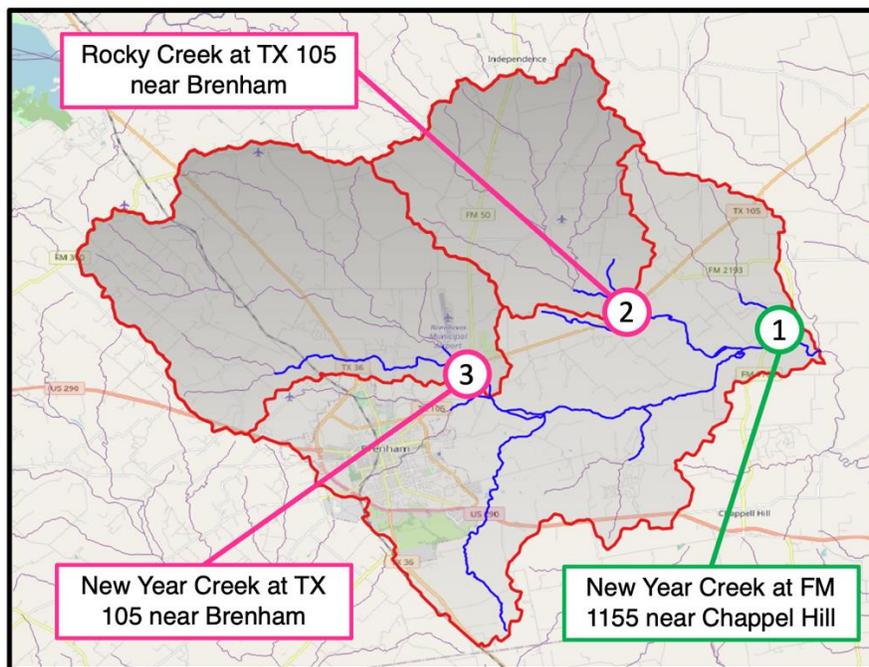


Figure 20. Study area in New Year Creek watershed near Brenham, consisting of two upstream sites (1 & 2) and one downstream site (3).

4.3. Hydraulic solver

The hydraulic solver used for this analysis is based on the Saint-Venant equations for fully unsteady flow in open channels. The Saint-Venant equations consist of two hyperbolic partial differential equations which correspond to conservation of mass and conservation of momentum.

Continuity:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_{in} \quad (4.1)$$

Momentum:

$$\frac{\partial Q}{\partial t} + \frac{\partial Qu}{\partial x} + gA \left(\frac{\partial h}{\partial x} - S_0 + S_f + S_L \right) = 0 \quad (4.2)$$

Where A is the channel cross-sectional area, Q is the discharge, q_{in} is the exogenous runoff input, u is the velocity of flow, h is the depth of flow, S_0 is the channel bottom slope, S_f is the friction slope, S_L is the local loss slope.

These coupled partial differential equations describe the full unsteady dynamics of water moving in one-dimensional open channels. The continuity equation states that the rate of change of volume within each model element must equal the flow in minus the flow out. The momentum equation is an application of Newton’s second law of motion, which states that the rate of change of momentum into the control volume plus the momentum flux through the control volume must equal the sum of forces acting on the control volume, with the forces corresponding to the hydrostatic force, the gravitational force, the friction force, and forces due to local losses. The Saint-Venant equations represent the fundamental dynamics of fluid motion in 1D channels or conduits, and are capable of capturing fluid behavior that simplified models like Muskingum routing cannot—such as backwater effects and highly inertial flows.

4.4. Model construction

The hydraulic model is constructed using the PipeDream hydraulic solver (Bartos & Kerkez, 2021), which implements an implicit method for solving the Saint-Venant equations for unsteady flow. This solver is implemented in the Python programming language, and uses the Numba just-in-time compiler to accelerate numerical code.

The hydraulic model for the New Year Creek watershed is constructed using channel network topology data from the NWM, which in turn draws its channel centerlines from the USGS National Hydrography Dataset. Channel bathymetry and roughness data are taken from surveyed Base-Level Engineering (BLE) data, which are available for rivers throughout the state of Texas. In addition to providing the hydraulic geometries, these HEC-RAS transects provide the channel bottom elevations of each reach. For each reach, an optimization approach is used to find the prismatic geometry that best matches the irregular hydraulic geometry contained within the HEC-RAS model. Using these prismatic geometries greatly facilitates model simulation times by reducing the need to simulate flows moving through complex irregular hydraulic geometries.

The hydraulic model is calibrated by modifying the hydraulic geometries and Manning’s roughness coefficients to match observed stage-discharge relationships at the RQ-30 sites. Initially, the stage-discharge relationships admitted by the HEC-RAS BLE hydraulic geometries did not accord with observed stage-discharge relationships collected by the RQ-30 gages. To correct this problem, detailed transects surveyed for the RQ-30s were input at the gage sites

within the model. The Manning's roughness factors were then modified until the observed stage-discharge relationships were reproduced.

4.5. Kalman Filtering

Kalman Filtering (KF) is an optimal state estimation technique that combines a model of a dynamical system with observed measurements to produce an updated posterior estimate of system states. For this application, the model consists of the hydraulic model based on the Saint-Venant equations, while the observed measurements consist of stage and discharge readings from RQ-30 gages. The goal of KF in this assessment is thus to improve estimates of depth and discharge at ungaged locations and to improve forecasts of future depth and discharge by optimally combining the sensor and model estimates.

From the discretization scheme introduced by the PipeDream solver, the discretized Saint-Venant equations may be expressed in state-space form (Bartos & Kerkez, 2021):

$$\mathbf{x}_{k+1} = A_k \mathbf{x}_k + B_k \mathbf{u}_k + \mathbf{v}_k \quad (4.3)$$

$$\mathbf{z}_k = H_k \mathbf{x}_k + \mathbf{w}_k \quad (4.4)$$

Where \mathbf{x}_k is the state vector of junction heads at time k ; \mathbf{z}_k is the observed heads (obtained from sensor data); \mathbf{v}_k is the driving noise (runoff input); \mathbf{w}_k is measurement noise; A_k is the state transition matrix; and H_k is the observation matrix. The Kalman recursion may then be applied to this system to produce the optimal linear estimator of system states at the next time step, $\hat{\mathbf{x}}_{k+1}$:

$$\hat{\mathbf{x}}_{k+1} = A_k \hat{\mathbf{x}}_k + B_k \mathbf{u}_k + L_{k+1} \cdot [\mathbf{z}_{k+1} - H_{k+1} \cdot (A_k \hat{\mathbf{x}}_k + B_k \mathbf{u}_k)] \quad (4.5)$$

$$L_k = P_k H_k^T (H_k P_k H_k^T + W_k)^{-1} \quad (4.6)$$

$$P_{k+1} = A_k \cdot (P_k - P_k H_k^T \cdot (H_k P_k H_k^T + W_k)^{-1} \cdot H_k P_k) \cdot A_k^T + V_k \quad (4.7)$$

Where $\hat{\mathbf{x}}_{k+1}$ is the optimal posterior state estimate; P_{k+1} is the posterior estimation error covariance, V_k is the covariance of the driving noise; and W_k is the covariance of the measurement noise. The posterior state estimate $\hat{\mathbf{x}}_{k+1}$ is known to be optimal (in the minimum mean squared error sense) if (i) the process model is perfectly accurate, (ii) the sensor and process noise are Gaussian and uncorrelated, and (iii) the covariances of the sensor and process noise are known precisely.

KF is distinct from 4DVAR in that KF produces an estimate of the current system state $\hat{\mathbf{x}}_k$ from observations occurring from the past to the present $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k$. However, 4DVAR, as implemented in the KISTERS data assimilation procedure produces estimates of the system state in both the past and present $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_k$ using observations occurring from both the past and

present $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k$. Another major difference between the two data assimilation procedures is that the KISTERS 4DVAR procedure produces a single deterministic estimate of the system state, while KF produces a probabilistic estimate with error bounds at each time step given by the estimation error covariance P_k .

Under the KF procedure described in this section, the hydraulic model is advanced forward in time, computing the hydraulic heads and discharges within each model element at each simulation time step. KF is then applied at each time step to fuse the sensor data into the model and produce an updated optimal posterior estimate of the depths and discharge throughout the entire system. This procedure continues recursively until the end of the simulation period.

4.6. Test storm event

For the data assimilation assessment, a storm event occurring on 2022-03-22 is selected. This storm event is selected because data from all three RQ-30 gages are available for this time period. A model time interval from 2022-03-19 to 2022-03-30 is selected to capture the preceding hydraulic conditions, and also to ensure that the full falling limb of the hydrograph is captured. The storm event produced a peak discharge of approximately 25 m³/s at the outlet of the watershed at FM1155 near Chappel Hill. The peak observed depth at this site was approximately 3.5 m.

Outputs from the NWM are used as forcings to simulate the hydraulic model for this storm event. Specifically, the lateral overflow into the channel ('qSfcLatRunoff') and the subsurface channel recharge ('qBucket') are used as inputs to the hydraulic model. NWM discharges are also used to establish initial conditions for the model at the start of the simulation time period.

4.7. Holdout assessment

To assess the ability for data assimilation to improve estimates of flood stage at ungaged locations, a holdout assessment is performed using the three RQ-30 gages available in the New Year Creek watershed. This assessment consists of applying data assimilation to a subset of sensor sites, and then evaluating whether the estimates of depths and discharges are improved at a set of holdout sensor sites. For this assessment, the Kalman Filter is used to fuse data only at the two upstream sites, and the estimated depths and flows produced via data assimilation at the downstream site are compared against the observed sensor measurements, taking the sensor measurements to represent ground truth. Both depth and discharge measurements are fused into the two upstream sites, and the results of the holdout analysis are assessed with respect to both variables.

4.8. Results

Data assimilation via Kalman Filtering significantly improves estimates of both depth and discharge at the downstream holdout site. Figure 21 shows the results of the KF holdout assessment for the selected storm event. RQ-30 depth and discharge data are assimilated at the two upstream sites (labeled 1 and 2 in the figure). As the model is advanced forward in time, the correction imposed by the Kalman Filter propagates downstream and improves estimates of both depth and discharge at the downstream holdout site (labeled 3 in the figure). Here, for each gage location indicated on the map, the modeled estimate using NWM inputs is shown in black, the sensor data is shown as a grey area plot, and the assimilated model run with Kalman Filtering is shown in red. The model-only approach severely underestimates both depth and discharge, due to an underprediction of precipitation from the NWM. However, the model with data assimilation shows strong agreement with observed data at the downstream holdout site. In particular, the peak discharge predicted by KF ($25 \text{ m}^3/\text{s}$) matches the true discharge at the downstream site almost exactly. Additionally, Figure 22 illustrates the comparison of root mean squared errors (RMSE) for the downstream sensor gage, both with and without the Kalman Filtering. *By incorporating data assimilation, the model achieves a substantial reduction in RMSE—approximately 46.2% for depth and 56.2% for discharge.*

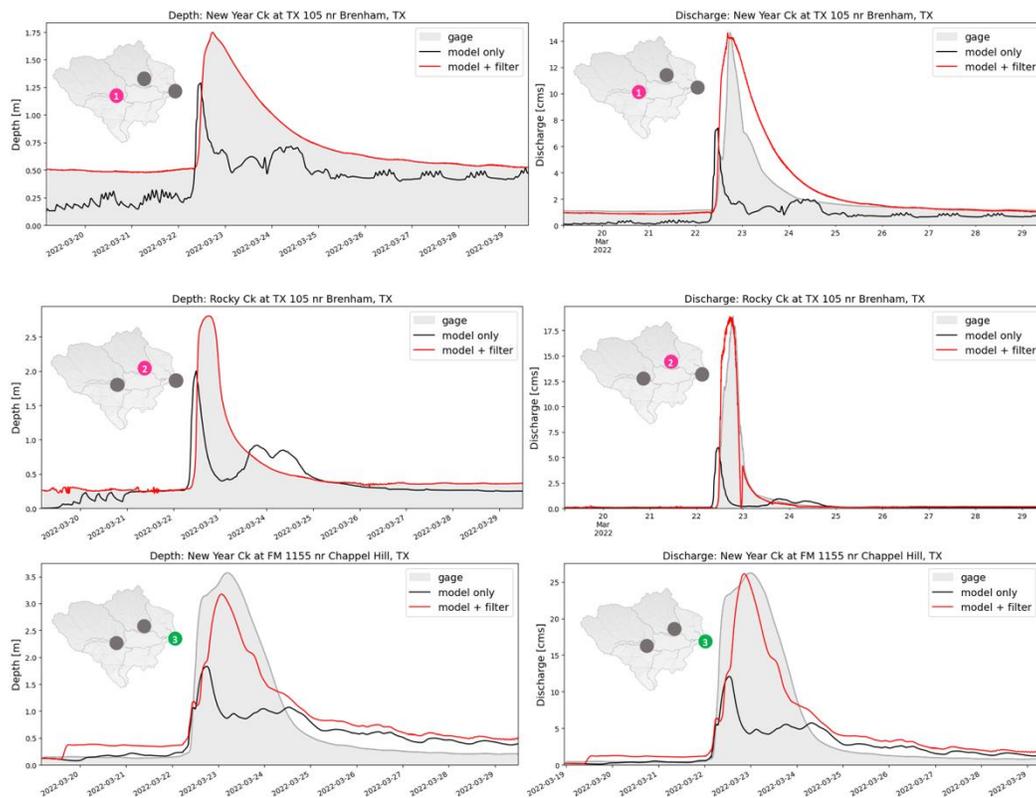


Figure 21. Results of the Kalman Filtering holdout assessment. Depths for all three sites are shown on the left while discharges are shown on the right. Model-only output is shown in black, sensor measurements are shown in grey, and the Kalman Filter output is shown in red. The mini-map indicates the location of each hydrograph in the watershed.

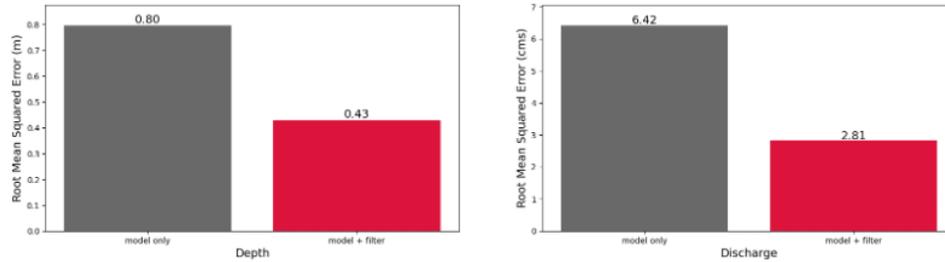


Figure 22. Evaluation of the errors with and without Kalman Filtering compared to the downstream gage: Model-only output is shown in black, while the Kalman Filter output is shown in red.

While Kalman Filtering enhances both depth and discharge estimates at the downstream holdout site, a number of potential improvements could be made that would further enhance streamflow predictions:

- First, the downstream site is fed by three major tributaries aside from the tributaries monitored by the two upstream gages. In its current implementation, the Kalman Filter does not correct the discharge from these unmonitored tributaries. As a consequence, the predicted downstream hydrograph shape is notably narrower than the ground truth downstream hydrograph shape. This problem can be corrected by specifying the process noise covariance matrix to properly capture the spatial correlation in runoff inputs across the watershed. Specifying the process noise covariance matrix in this way, runoff corrections imposed by the Kalman Filter at one location would lead to runoff corrections at spatially distant locations where the runoff input is known to be correlated. This process noise covariance matrix could be estimated empirically using time series of runoff inputs at each reach from the NWM.
- Second, it should be noted that the correction imposed by the Kalman Filter at the two upstream sites leads to a baseflow downstream that is larger than the observed baseflow. This result suggests that the stage-discharge relationships at the two upstream sites are not properly specified. Additional discharge measurements should thus be taken for low flows at the two upstream sites to ensure that the stage-discharge relationship is accurate.

4.9. Uncertainty assessment

Kalman Filtering is a stochastic data assimilation algorithm based on Bayes' theorem. As a consequence, KF produces error bounds corresponding to the posterior probability distribution of the estimated stages and discharges at each element in the modeled river network. These error bounds are defined by the estimation error covariance matrix P_k produced at each step of the Kalman Filter recursion. Figure 23 shows the 95% confidence interval of the stage (left) and discharge (right) estimates produced by the Kalman Filter at the New Year Creek at TX 105 near Brenham gage site. It should be noted that these error bounds depend on the process noise

covariance (i.e. uncertainty in runoff input) and measurement error covariance (i.e. sensor noise variance), which must be specified a priori. The error bounds here use assumed values for these parameters, and are intended for demonstration purposes only.

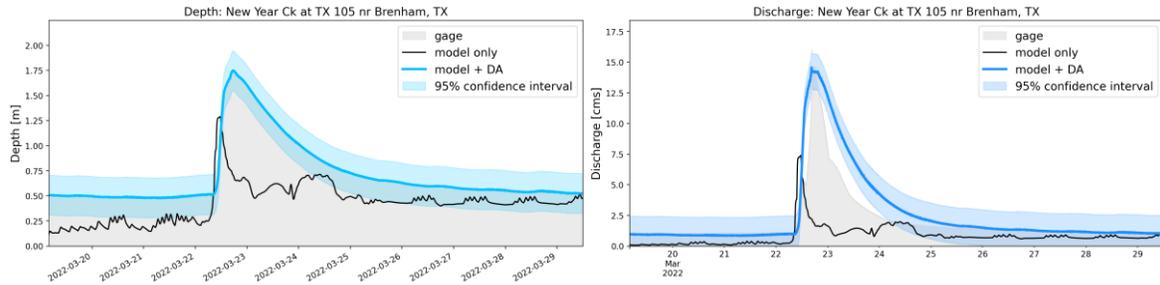


Figure 23. Kalman Filter estimate of stage (left) and discharge (right) with 95% confidence interval.

5. Improved discharge estimates at RQ-30 gages through data analysis physically-based modeling

5.1. Analysis of RQ-30 data and evidence for hysteresis behavior

Flow in open channels exhibits hysteresis behavior, such that depth and discharge differ on the rising and falling limbs of the hydrograph (Kuhle & Bowie, 1992). Typically, the depth-discharge relationship will consist of a loop, rotating either counterclockwise (meaning that the flow velocity is larger on the rising limb) or possibly clockwise (meaning that the flow velocity is larger on the falling limb). Because rating curves do not account for this looping behavior, discharge estimates produced using traditional methods may not be accurate on the rising limb, leading to inaccuracy in flood forecasts.

This flood wave hysteresis is a well-known phenomenon, but research on stream characteristics influencing its magnitude and variability is limited (Muste et al., 2020). We thus decided to take advantage of the observations collected at the RQ-30 sites to analyze patterns of hysteresis over a range of streams and to quantify hysteresis patterns, including the rapid velocity increase prior to a flood event for potential use in flood warning systems (Ho et al., 2023).

We focused this analysis in particular on a subset of streams of different characteristics (Figure 24). The availability of high temporal resolution measurements allows the detection of hysteresis patterns, as the example shown in Figure 25. As seen, velocity increases rapidly during the rising limb, and decreases slowly during the falling limb. Velocity-to-stage ratio is higher at the start of the flow event.

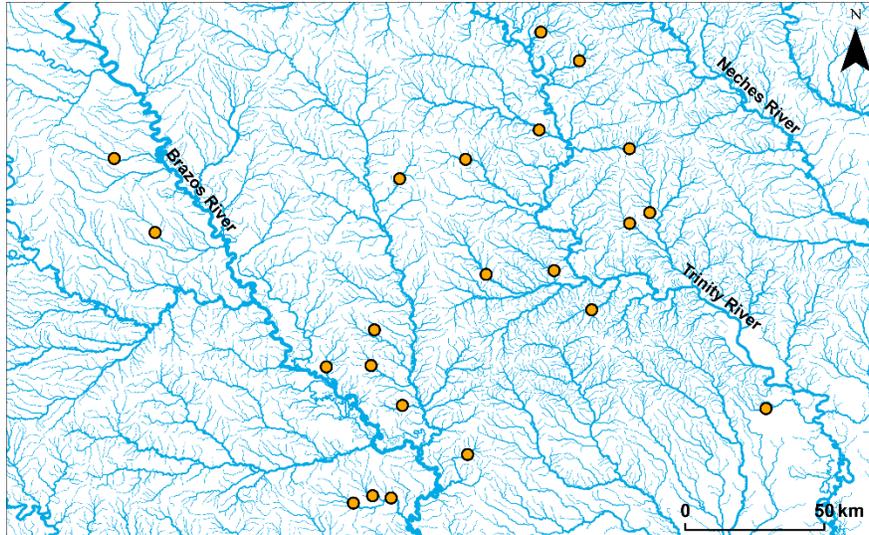


Figure 24: Location of RQ-30 sensors in Texas, USA used in this study.

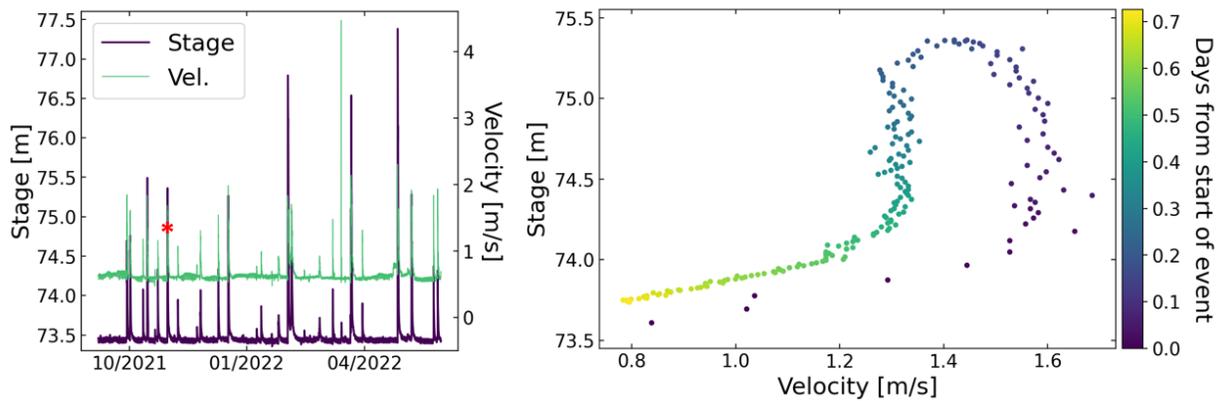


Figure 25: Stage and velocity data at a sensor in Carters Creek near College Station, TX.

To quantify the dependence on local stream characteristics, for each stream, we plotted stage-velocity curves for flow events with return periods greater than 4 weeks, resulting in a diverse set of hysteresis patterns, as it can be seen in the examples shown in Figure 26. Hysteresis varies wildly among locations, but at a given location, similar patterns emerge.

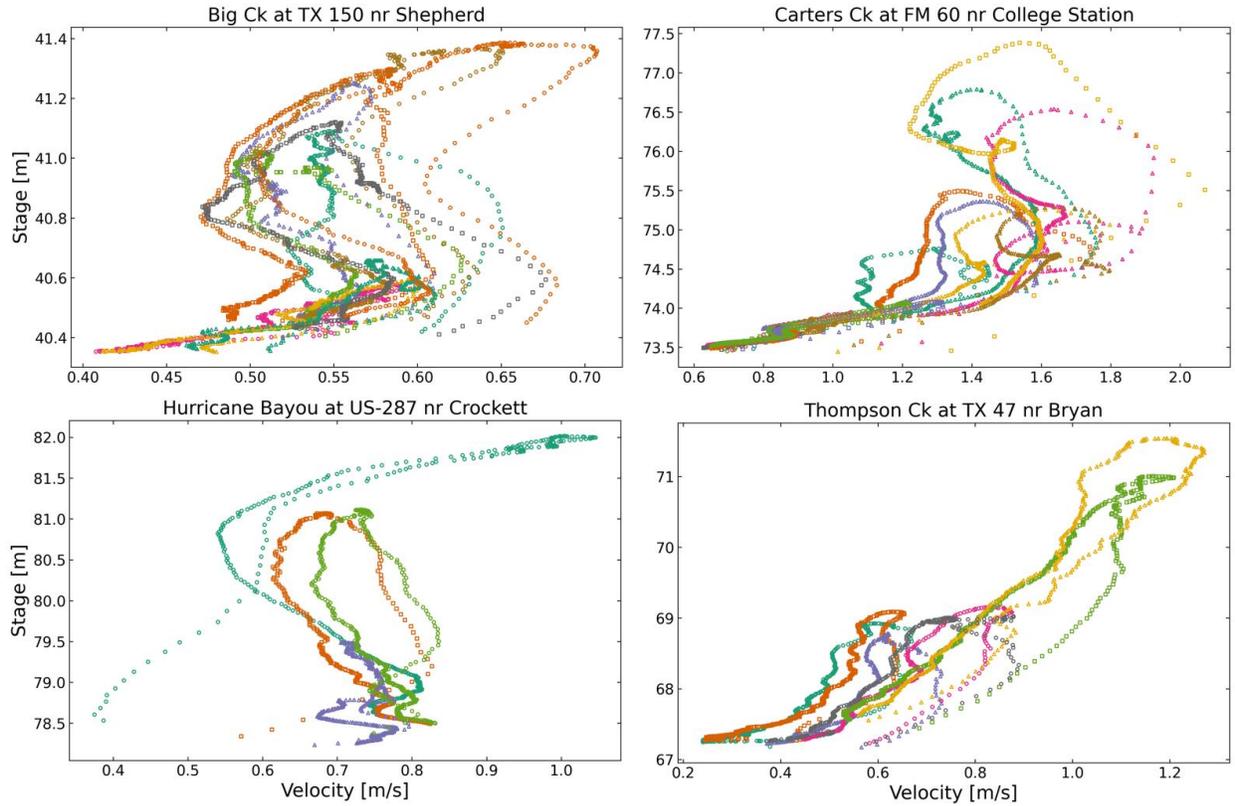


Figure 26: Plots of stage vs. velocity for four streams, where each curve is a unique flow event.

In some cases, hysteresis can potentially be explained by channel geometry (Figure 27). In order to quantify the shape of the hysteresis, we modified the h-index developed by Zuecco et al. (2016) to determine the difference in velocity for the rising and falling limbs (Figure 28). Based on the h-index, we investigated the relationship between the h-index and various stream variables (Figure 29). The analysis highlighted significant variability among streams as well as among storms for the same stream.

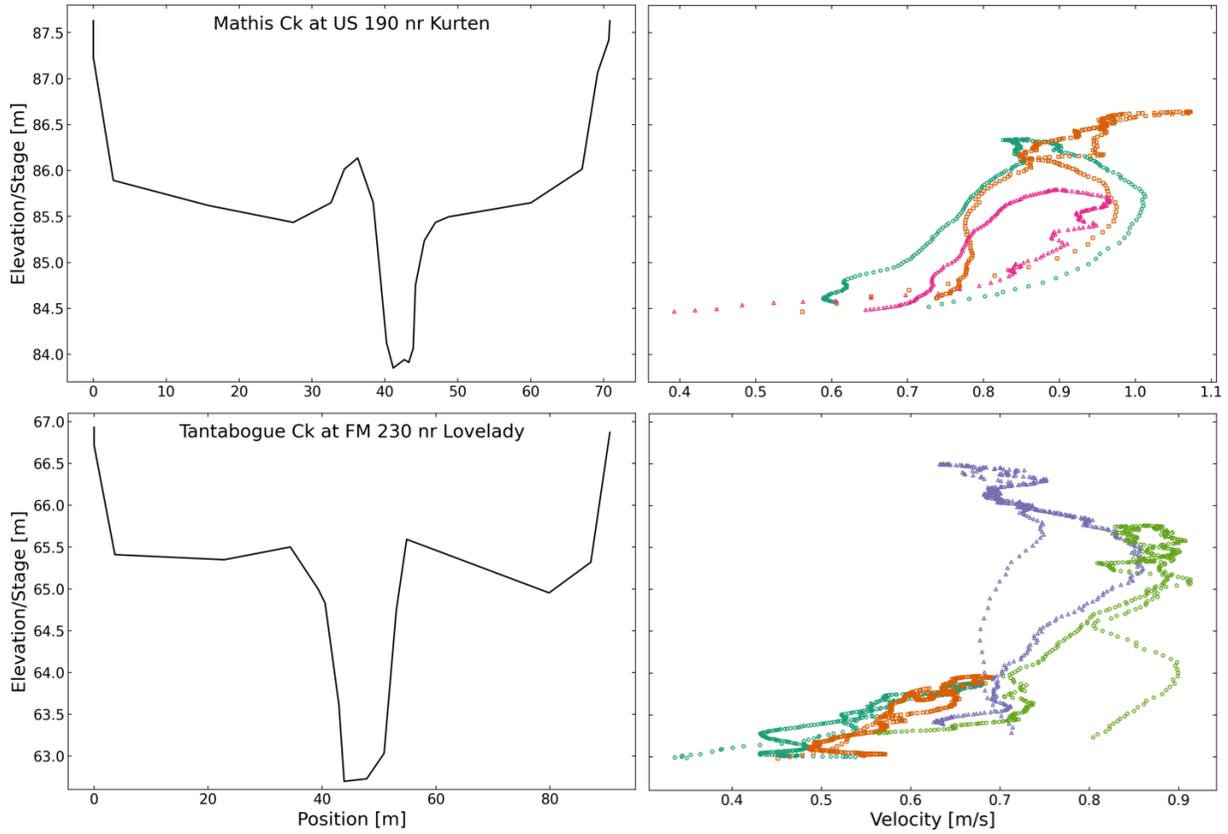


Figure 27: Comparison of stage-velocity curves with the local channel cross section.

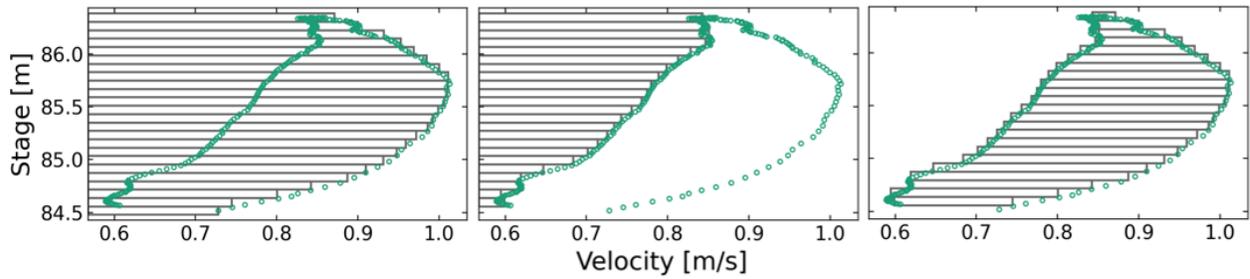


Figure 28: Conceptual integration of a hysteresis loop.

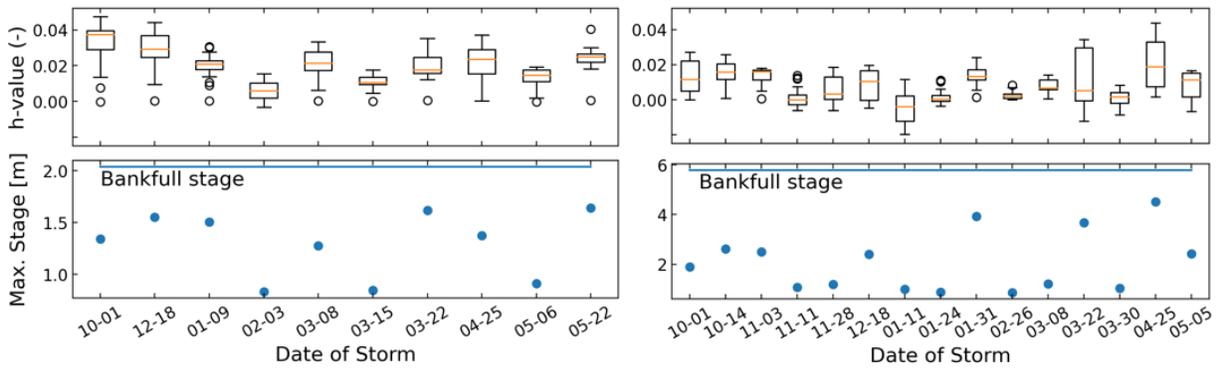


Figure 29: H-index for Big Creek (left) and Carters Creek (right).

In the context of flood monitoring, this hysteresis behavior suggests that a prominent velocity spike often occurs several hours prior to the peak of an oncoming flood wave. We thus decided to further investigate this behavior to understand the physical basis for the observed hysteresis behavior, as described in the following section.

5.2. Physical basis for hysteresis behavior

Understanding the physical basis of hysteresis will provide a pathway for potentially using the pre-emptive velocity rise information as an early flood warning indicator. Moreover, understanding hysteresis behavior will enable more accurate discharge estimates on the rising limb that can in turn be used to improve flood modeling efforts. Traditional discharge estimates based on rating curves assume a one-to-one relationship between depth in discharge, which is inconsistent with real-world open channel flow behavior, and may underestimate discharge particularly on the rising limb. Incorporating hysteresis behavior gives rise to the potential for ‘dynamic rating curves’ that better capture real-world stage-discharge behavior.

Based on the literature, clockwise hysteresis, associated with larger discharges on the falling limb, is generally associated with build-up and break-down of bedforms at the channel bottom (Kuhnle & Bowie, 1992). Bedforms consisting of sediment on the channel bottom initially impede the flow of water, causing the velocity of flow to be smaller on the rising limb, until the bedforms are broken by rapid flows during the peak of a hydrograph, whereupon the bedforms to give way and admit larger discharges on the falling limb of the hydrograph.

While this clockwise hysteresis behavior has previously been noted within the literature (Kuhnle & Bowie, 1992), it is uncommon in the TxDOT RQ-30 network. Most sites observed on the RQ-30 network exhibit counterclockwise hysteresis, where the discharges and velocities are larger on the rising limb. This behavior arises from the dynamical equations of motion describing unsteady fluid flow, and can be confirmed using simulations of open channel flow using the Saint-Venant equations. Figure 30 shows the depth time series and velocity time series of a simulated hydrograph at a single point in a prismatic open channel (top), along with the forces acting on a control volume at that point (bottom). As can be seen from this figure, the velocity peaks roughly 45 minutes before the peak of the depth hydrograph. Considering the momentum component of the Saint-Venant equations, the change of momentum in fluid flow (i.e. acceleration) is influenced by inertial, hydrostatic, gravitational, frictional, and local forces. From Figure 30 (bottom), the timing and scale of each of these forces can be observed. Here, it can be seen that the peak of the velocity hydrograph is largely driven by the hydrostatic force, while the frictional and gravitational forces lag behind the velocity peak. Specifically, the steep gradient in the water surface associated with the advancing front of a flood wave generates a hydrostatic force in the direction of flow, causing the water to accelerate. This acceleration leads to a peak in the velocity hydrograph that precedes that of the depth hydrograph.

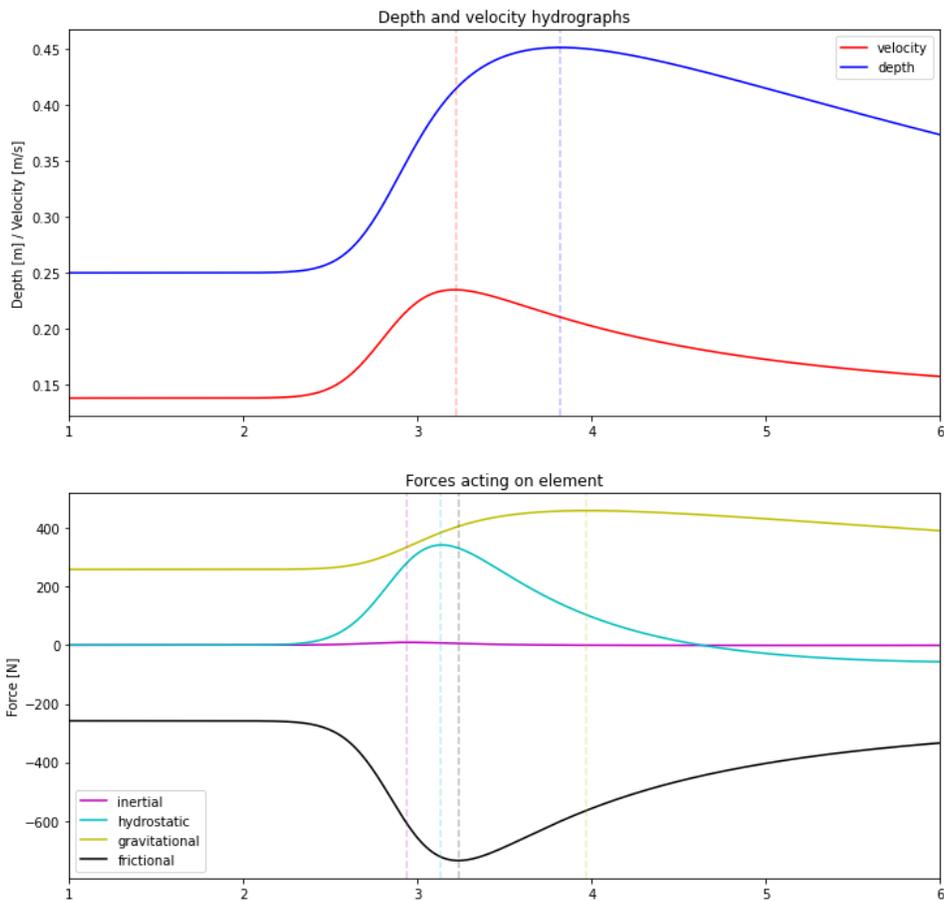


Figure 30. (Top): Depth and velocity hydrographs for a flood wave at a single point in a rectangular channel, simulated using the Saint-Venant equations. (Bottom): Forces acting on the control volume, including inertial, hydrostatic, gravitational, and frictional forces.

The observed pre-emptive velocity spike observed by the RQ-30 gauges is thus not an artefact of the sensor, but rather a real phenomenon that can be deduced from the fundamental fluid equations of motion. This finding opens the door to more accurate estimation of discharge through the development of ‘dynamic rating curves’ that account for this behavior.

5.3. Physical modeling of individual reaches

To better understand hysteresis behavior, and to assess the potential for building more accurate ‘dynamic rating curves’, physically-based models of three river reaches are constructed using a hydraulic solver based on the Saint-Venant equations for unsteady flow. The three reaches chosen include Thompson Creek, Big Creek, and Carters Creek. Each of these reaches is monitored by an RQ-30 gage. Channel geometries and bottom elevations for these reaches are taken from HEC-RAS BLE models, while the channel geometries at the RQ-30 gage site are taken from Q-Commander transects. Figure 31 shows the profile view of the reach during a flood wave (left), along with the channel cross section at the RQ-30 site (right).

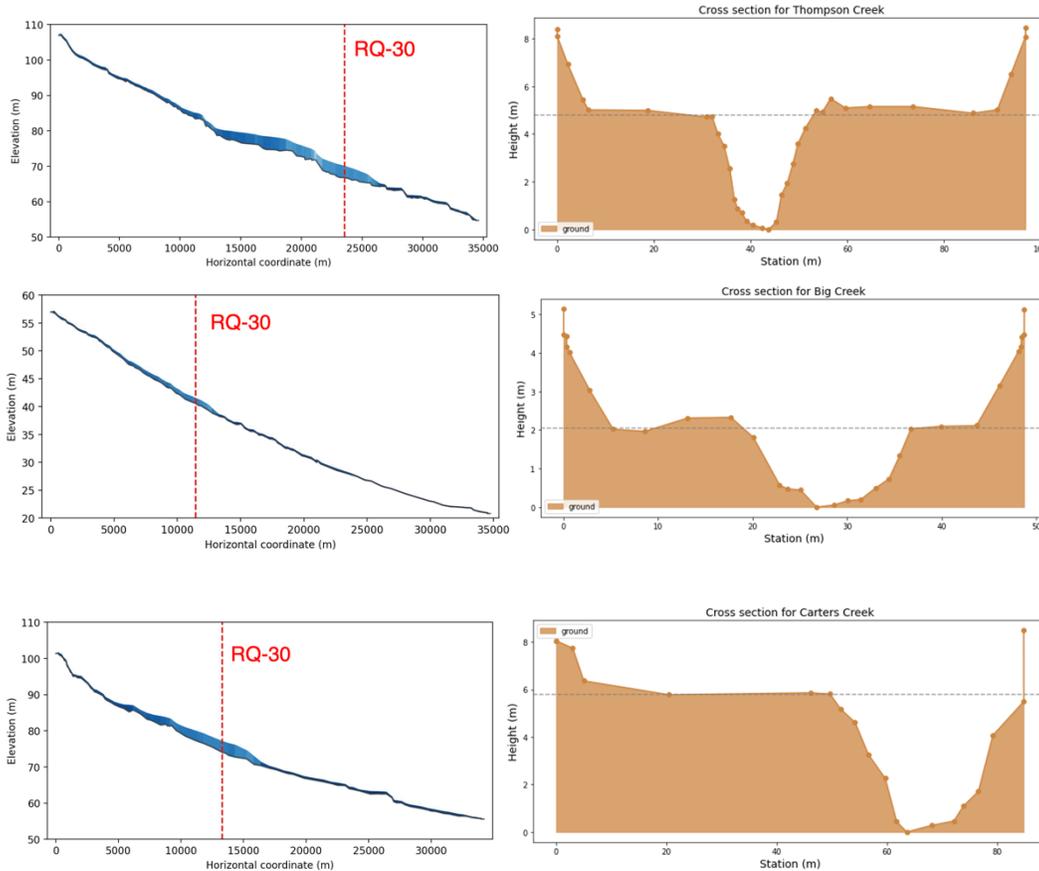


Figure 31. Profile and cross-sectional views of the hydraulic models of Thompson Creek, Big Creek, and Carters Creek. (Left): Profile view of channel during a flood wave, with the location of the RQ-30 gage indicated. (Right): Cross sectional view of channel at RQ-30 site, with floodplain height indicated.

To assess the potential for physically-based models to capture real-world hysteresis behavior, models of each reach are forced with flood waves at the upstream boundary (see Figure 31, left). The resulting modeled stage-discharge relationships at the RQ-30 location are then compared against the observed stage-discharge observations from ADCP measurements.

Figure 32 shows the modeled stage-discharge relationship (blue) along with ADCP measurements (red) for all three reaches. After adjusting the Manning's roughness factor, it is found that the upper bound of the stage-discharge hysteresis loop closely matches the observed ADCP measurements. The upper bound of the stage-discharge hysteresis loop corresponds to the falling limb of the hydrograph. Given that ADCP measurements are typically collected on the falling limb, this result indicates that the physically-based models are capable of accurately reproducing the stage-discharge relationships observed by the ground truth measurements. This result opens the door to potential physically-based 'dynamic rating curves' that can be developed from first principles for each RQ-30 location.

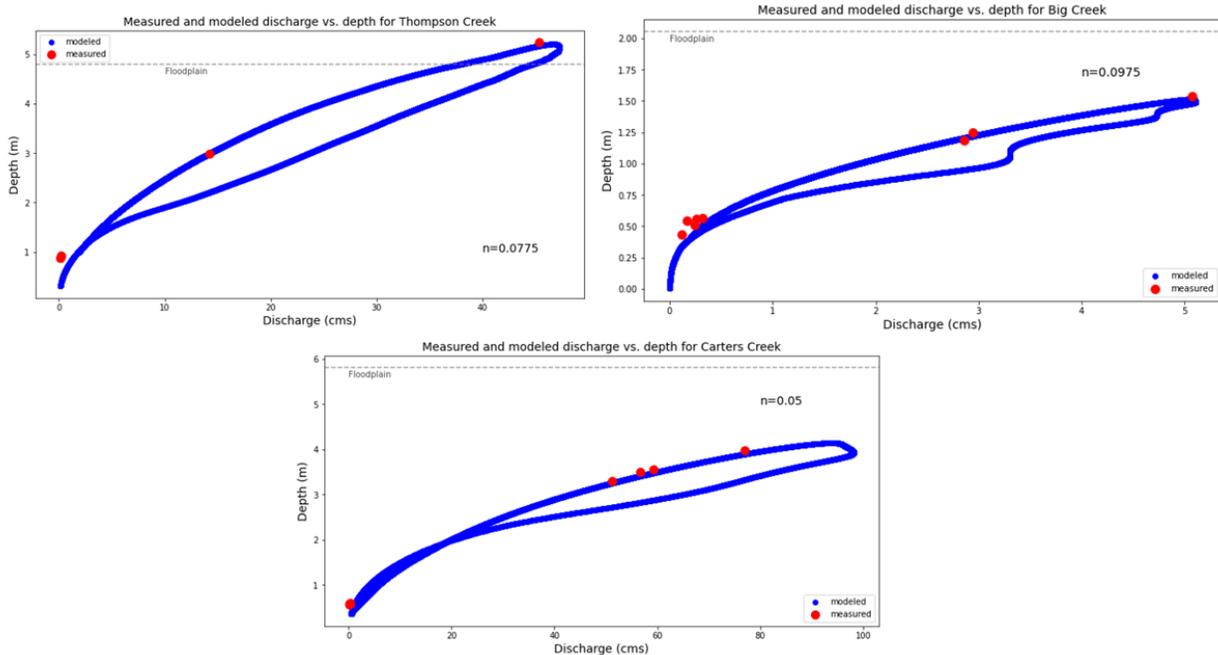


Figure 32. Stage-discharge hysteresis loops for Thompson Creek, Big Creek, and Carters Creek. Model output is shown in blue, while ADCP ground truth measurements are shown in red.

The modeled hysteresis behavior is also compared against observed stage-velocity hysteresis behavior measured by the RQ-30 gages. Figure 33 shows the observed stage-velocity hysteresis behavior measured by the RQ-30 gage (left) along with the modeled stage-velocity hysteresis behavior (right) for Carters Creek. It should be noted that the modeled and observed hysteresis loops show a similar general shape, and share inflection points corresponding to breakpoints in the channel geometry. However, the scales of the two loops differ, given that the velocity measurements from the RQ-30 correspond to the surface velocity, rather than the mean channel velocity. To better understand the hysteresis behavior observed by the RQ-30, it is necessary to better understand the relationship between the surface velocity and the mean channel velocity, as expressed by the *k-factor*.

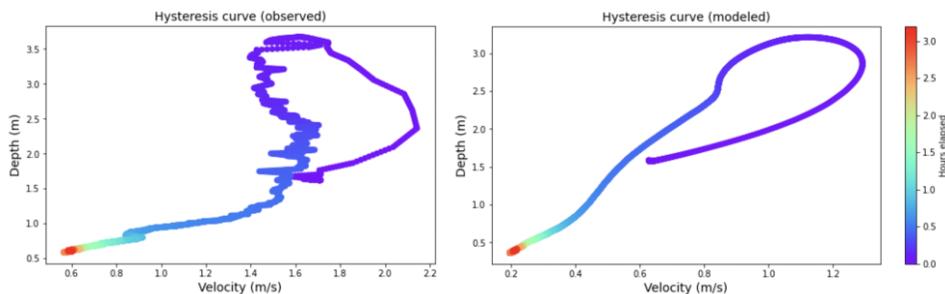


Figure 33. Stage-velocity hysteresis loops for Carters Creek. (Left): measured stage-velocity relationship, using surface velocity. (Right): modeled stage-velocity relationship, using average velocity.

5.4. Physical basis for the k-factor

The RQ-30 gages in the TxDOT streamflow monitoring network use a parameter called the k-factor to translate surface velocity measurements into discharge measurements used for flood forecasting. In this section, a physical basis for the k-factor is established in terms of the logarithmic vertical velocity profile of flow in open channels. This theoretically-derived k-factor is then applied to data collected from the RQ-30 network to improve real-time discharge estimates.

Accurate estimates of discharge are essential for real-time flood monitoring and forecasting. Because in-situ discharge measurements are expensive and difficult to collect, data providers like the USGS typically collect only continuous stage measurements. Continuous discharge measurements are estimated through the use of rating curves. In this approach, a small number of discharge measurements are sampled manually in the field during high- and low-flow events. An empirical relationship is then constructed to relate sampled discharge estimates to their corresponding stage measurements. This empirical relationship is then applied to continuous river stage measurements to produce continuous discharge estimates. As mentioned previously, these rating curves generally assume a one-to-one relationship between stage and discharge, and thus cannot represent real-world hysteresis behavior.

Unlike traditional gaging stations, the radar-based RQ-30 gages used by the TxDOT streamflow monitoring network are capable of measuring both the stage and velocity of water simultaneously. Because the discharge is equal to the average velocity times the area of flow, and because the area of flow can be deduced from the stage and the channel bathymetry, RQ-30 gauges have the potential to generate real-time estimates of discharge without the use of rating curves. Because the RQ-30 collects direct estimates of velocity on both the rising and falling limb, the derived discharge estimates account for stage-discharge hysteresis behavior that rating curves do not capture.

However, a major obstacle impeding the use of radar gages for real-time discharge measurement is the uncertain relationship between surface velocity and discharge. While the average velocity of flow is needed to compute discharge, RQ-30 gages measure only the velocity at the water surface. The relationship between the surface velocity and the average cross-sectional velocity must be estimated empirically through the k-factor:

$$K = \frac{\bar{u}}{u_s} \quad (5.1)$$

Where K is the k-factor, \bar{u} is the average cross-sectional velocity, and u_s is the velocity at the water surface. Accurately characterizing this k-factor is essential for accurately determining discharge from radar surface velocimetry measurements. The typical process for calibrating this

k-factor is to collect ADCP measurements of the cross-sectional velocity profile and relate these measurements to the RQ-30 surface velocity measurements. However, this process is as labor-intensive as the traditional rating curve approach. A method for computing the k-factor from first principles would greatly facilitate streamflow monitoring by obviating the need to collect manual discharge measurements for calibration.

Previous research has established a physical basis for the k-factor under the assumption of a logarithmic vertical velocity profile (Welber et al., 2016). In this formulation, the k-factor can be established in terms of the depth of water and the roughness length associated with the channel bottom material. In general the vertical velocity distribution for turbulent flows in wide, uniform, smooth-bed channels generally follows a logarithmic distribution, with the minimum velocity near the bottom of the channel and the maximum velocity near the water surface. The velocity as a function of the depth of water may be expressed as:

$$u(z) = \frac{u^*}{\kappa} \ln\left(\frac{z}{z_0}\right) \quad (5.2)$$

Where u is the flow velocity, z is the vertical coordinate, u^* is the friction velocity, z_0 is the roughness length, κ is the Von Karman constant (0.41). The average channel velocity may be computed by integrating this equation from the channel bottom to the free surface (Welber et al., 2016):

$$\bar{u} = \frac{1}{D - z_0} \int_{z_0}^D u(z) dz = \frac{u^*}{\kappa} \left[\ln\left(\frac{D}{z_0} - 1\right) \right] \quad (5.3)$$

Where \bar{u} is the depth-averaged velocity and D is the vertical coordinate at the water surface. The surface velocity can be recovered by setting $z = D$ in the first equation. Thus, from these two equations, a theoretical expression for the k-factor can be derived (Welber et al., 2016):

$$K = \frac{D}{D - z_0} - \ln\left(\frac{D}{z_0} - 1\right)^{-1} \quad (5.4)$$

Thus, for a wide, smooth-bottomed channel with a logarithmic velocity profile, the k-factor (K) depends only on the depth of flow and the roughness length.

Figure 34 shows the theoretical k-factor by depth of flow and bottom roughness length. In general, as the depth of flow increases, the k-factor converges to a value of approximately 0.8, which is consistent with values reported in the literature (Welber et al., 2016).

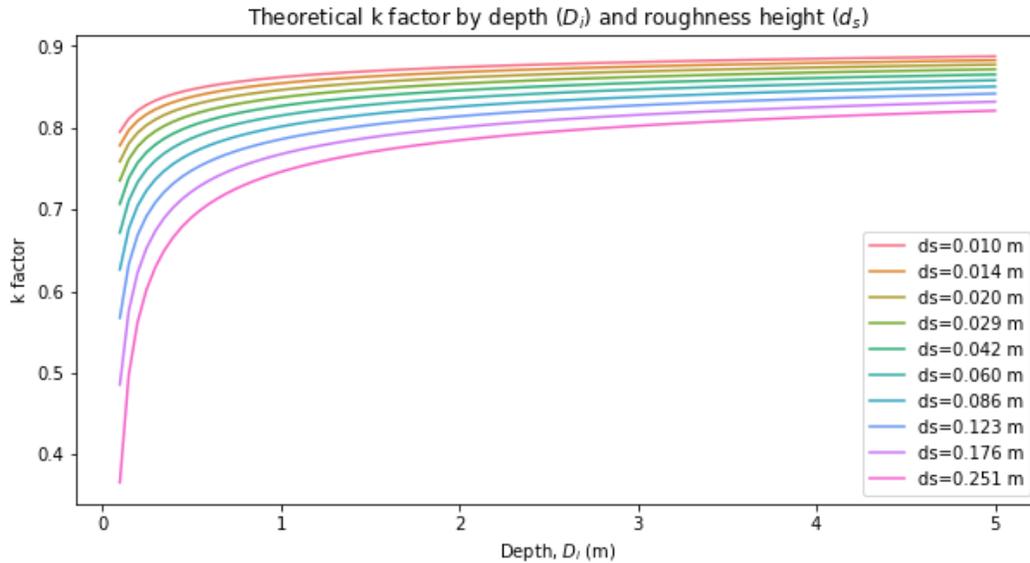


Figure 34. Theoretically-determined k-factor vs. water depth for given characteristic roughness heights, assuming a logarithmic vertical velocity profile.

These theoretical k-factors can be used to improve the accuracy of the RQ-30 depth-averaged velocity estimates. Figure 35 shows the result of applying the theoretically-derived k-factor to RQ-30 measurements in Carters Creek. Here, the discharge is shown on the horizontal axis while the stage is shown on the vertical axis. The modeled stage-discharge relationship is shown as the dark blue loop, while the ADCP ground truth measurements are shown in red, and the RQ-30 estimates of discharge are shown in cyan. The left subplot shows the RQ-30 estimates of discharge using the currently-used k-factor, while the right subplot shows the RQ-30 discharge estimates using the theoretically-derived k-factor with a calibrated roughness length value. It can be seen here that the theoretically-derived k-factor better matches the full hysteresis loop predicted by the unsteady hydraulic solver.

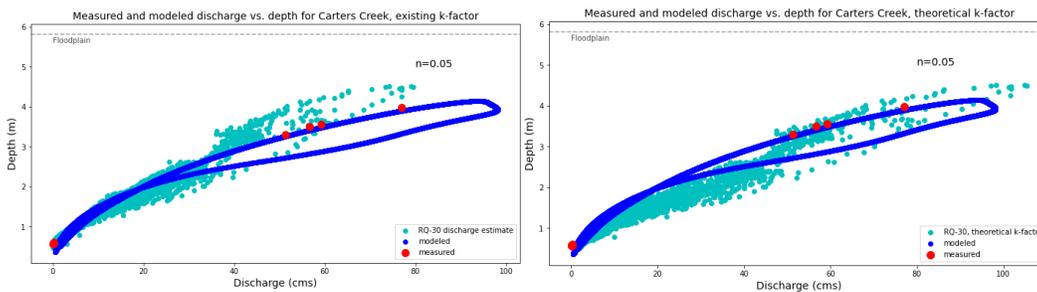


Figure 35. Stage-discharge comparison of model (blue), ADCP (red), and RQ-30 with k-factor (cyan). (Left): k-factor using existing manual calibration. (Right): theoretically-derived k-factor.

For real-world RQ-30 gage sites, the use of the theoretical k-factor is complicated by the complex bathymetry and roughness characteristics of real-world channels. Channels observed by the RQ-30 gages are not suitably approximated by a “wide” channel geometry. In most cases, these channels consist of two sections: a narrower trapezoidal channel section on the bottom, and

a wider floodplain section on top. A wide channel geometry does not accurately capture the true hydraulic behavior of these channels. Moreover, real-world channels monitored by RQ-30 gages often feature grassed or vegetated bottoms that violate the assumptions of a purely logarithmic velocity profile. Finally, the material and vegetation cover of the channel varies across its profile, meaning that the average roughness height is not constant with depth. Additional methods are needed to modify the formulation for the theoretical k-factor to account for these discrepancies.

The aforementioned issues can be corrected by allowing the roughness of the channel bottom to vary with depth—both for the hydraulic model and for the computation of the k-factor. Figure 36 shows the results of applying the theoretically-derived k-factor while allowing the roughness of the channel to vary with depth. Here, the Manning’s roughness of the channel varies from a minimum value of $n=0.03$ at a depth of 0 m to a maximum value of $n=0.0525$ at a depth of 3 m. As before, the dark blue loop shows the modeled stage-discharge relationship, the red points show the ADCP measurements, and the cyan points show the RQ-30 discharge measurements. The left subplot shows the RQ-30 discharge estimates using the existing k-factor, while the right-hand side shows the RQ-30 discharge estimates using the theoretically-derived k-factor. For this case, the stage-discharge relationship predicted by the model matches the RQ-30 measurements closely.

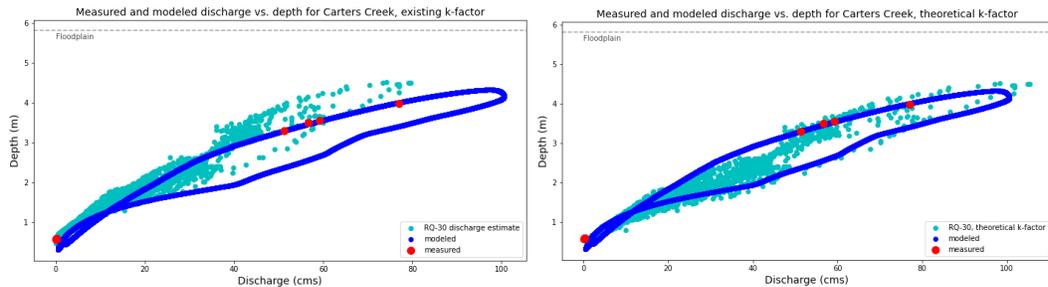


Figure 36. Stage-discharge comparison of model (blue), ADCP (red), and RQ-30 with k-factor (cyan), allowing for depth-varying channel roughness. (Left): k-factor using existing manual calibration. (Right): theoretically-derived k-factor.

6. Conclusions

Based on our analysis, we find that the performance of the National Water Model (NWM) varies spatially across the State of Texas. In particular, the NWM was found to have greater errors in areas that experienced higher precipitation, and on average, it underestimated discharge during the time period analyzed. Additionally, the analysis of mean discharge indicated that error increased with mean discharge and the analysis of stream orders indicated that stream order 4 had the lowest RMSE, while stream orders 6 and 7 had the highest RMSE values. Further, the assessment of similarity between the NWM and observed annual peak streamflow distributions showed that the NWM was able to capture the annual peak flow in less than 44% of the

locations. Additionally, no spatial bias is observed over the state of Texas, suggesting that the NWM does not preferentially overperforms or underperforms in specific areas.

We have performed a proof of concept of a data assimilation method on the Llano river basin. The assimilation method has a domain of influence upstream of the discharge measurements. This is a significant advantage over the nudging applied in the current version of the NWM, because most of the concern points and low water crossings have no discharge gauge upstream to them. We developed a scalable and solution architecture utilizing KISTERS' cloud-based services to perform data assimilation runs dynamically, based on the currently available discharge measurements. The dynamic approach to the discharge measurements makes the solution robust and less sensitive to temporary data unavailability. It automatically adapts to the currently available gauges, making fully autonomous updating the models. The data assimilation based on excess rainfall optimization can be deployed at a local, at HUC8 or larger scale if the necessary discharge observations are available.

Data assimilation via Kalman Filtering significantly improves estimates of both depth and discharge at the downstream holdout site. The approach enhances both depth and discharge estimates at the downstream holdout site. By incorporating data assimilation, the model achieves a substantial reduction in RMSE—approximately 46.2% for depth and 56.2% for discharge.

The analysis of RQ-30 stage and discharge data has revealed previously unobserved hysteresis behavior in which the discharge and velocity are larger on the rising limb of the hydrograph. An analysis of hysteresis based on the Saint-Venant equations suggests that the pattern observed at the gauges is real and this finding opens the door to more accurate estimation of discharge through the development of 'dynamic rating curves' that account for this behavior.

Finally, we theoretically-derived the k-factor used at the gauges. This shows promise in improving discharge estimates produced by RQ-30 gages. Because this k-factor can be derived from known values of channel surface roughness—a property that can be estimated directly from material properties—it can potentially be applied to new gauge sites without the need for calibration to manual discharge measurements. By reducing the need for manual calibration, and providing an alternative source of verification for discharge estimates, this methodology has the potential to help expand the number of trusted gage sites in the RQ-30 network.

7. Acknowledgments

The Streamflow research team wishes to acknowledge the financial support provided through the TxDOT RTI research program. This project has the largest grant ever made for this program. We wish in particular to acknowledge the untiring diligence of Rose Marie Klee our technical project manager, Trenton Ellis who manages the day-to-day interactions with our team, and Shelley Pridgen the RTI manager for this project.

References

- Bartos, M., and Kerkez, B. (2021). Pipedream: An interactive digital twin model for natural and urban drainage systems. *Environmental Modelling & Software*, 144, 105120.
- Ho, J., N. Tull, S. Grzyb, P. Passalacqua, D. Maidment, Understanding the factors causing flood wave hysteresis on small streams using RQ-30 measurements, American Geophysical Union AGU Fall Meeting, Chicago, IL, December 2022.
- Kuhnle, R. A., and Bowie, A. J. (1992). Loop rating curves from Goodwin Creek. *Publications from USDA-ARS/UNL Faculty*, 316.
- Muste, M., Lee, K., Kim, D., Bacotiu, C., Rojas Oliveros, M., Cheng, Z., & Quintero, F. (2020) Revisiting hysteresis of flow variables in monitoring unsteady streamflows. *Journal of Hydraulic Research*, 58:6, 867-887, DOI: 10.1080/00221686.2020.1786742.
- Muste, M., Kim, D., & Kim, K. (2022) A flood-crest forecast prototype for river floods using only in-stream measurements. *Communications Earth & Environment*, 3:1, 1-10, DOI: 10.1038/s43247-022-00402-z.
- Nocedal, J., A. Wachter, and R.A. Waltz. Adaptive barrier strategies for nonlinear interior methods. *SIAM Journal on Optimization*, 19(4):1674–1693, 2008. preprint at http://www.optimization-online.org/DB_HTML/2005/03/1089.html
- Ponce, V. M. (1989) *Engineering Hydrology, Principles and Practices*. Available at: <http://ponce.sdsu.edu/textbookhydrologyp252.html> (Accessed: 30 July 2022)
- USDA-NRCS (2014) ‘National Engineering Handbook, Part 630, Chapter 17: Flood Routing’, (AUGUST)
- Wachter, A. An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, January 2002
- Wachter, A., and L.T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006. preprint at http://www.optimizationonline.org/DB_HTML/2004/03/836.html
- Welber, M., Le Coz, J., Laronne, J. B., Zolezzi, G., Zamler, D., Dramais, G., ... & Salvaro, M. (2016). Field assessment of noncontact stream gauging using portable surface velocity radars (SVR). *Water Resources Research*, 52(2), 1108-1126.
- Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., . . . Cooper, K. B. (2016). Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities . *Bulletin of the American Meteorological Society*, 97(4), 621-638.
- Zuecco, G., Penna, D., Borga, M., & van Meerveld, H. J. (2016) A versatile index to characterize hysteresis between hydrological variables at the runoff event timescale. *Hydrol. Process.*, 30:9, 1449-1466, DOI: 10.1002/hyp.10681.