The Thesis Committee for Andrew Austin-Petersen Certifies that this is the approved version of the following Thesis:

Error Assessment of National Water Model Analysis & Assimilation and Short-range Forecasts

APPROVED BY
SUPERVISING COMMITTEE:

Paola Passalacqua, Supervisor

David Maidment, Co-Supervisor

Error Assessment of National Water Model Analysis & Assimilation and Short-range Forecasts

by

Andrew Austin-Petersen

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

December 2018

Dedication

For my mother.

In loving memory of my father.

Acknowledgements

I would like to thank my advisor Paola Passalacqua for her guidance, fairness and dedication. Thank you for your support, both academically and personally, and for the occasional kick in the pants. It has been a joy working with you.

Thank you to David Maidment. Your vision and ability to home in on important ideas is truly remarkable. Thank you for expanding the scale of my thinking.

Thank you to my EWRE besties Isha Deo and Jay Hariharan.

I would also like to thank all the EWRE students and professors I've had the pleasure of learning from, and learning with, during my time at UT. Who knew grad school was going to be so much fun?

And thank you to my friends and family for supporting me no matter what – even when I quit my job to go on tour with my band 10 years ago.

Abstract

Error Assessment of National Water Model Analysis & Assimilation and Short-range Forecasts

Andrew Austin-Petersen, M.S.E.

The University of Texas at Austin, 2018

Supervisors: Paola Passalacqua and David Maidment

Flooding is the costliest natural disaster in the United States and tragically often leads to loss of life. Flood prediction, response and mitigation are therefore critical areas of research and have been for many decades. Hydrologic and hydraulic models are key components of flood prediction methods and highly detailed models have been implemented in many areas of high risk which often correspond to areas with high population. However, the high cost and complexity of highly detailed models means that many areas of the US are not covered by flood prediction early warning systems. Recent increases in computational power and increased resolution and coverage of remotely sensed data have allowed for the development of a continental scale streamflow prediction system known as the National Water Model which is currently forecasting streamflow values for over 2.7 million stream reaches across the US.

Flood inundation predictions can be derived from the National Water Model using digital elevation data to extract reach-scale rating curves and therefore river stage height.

Using the height above nearest drainage method, flood inundation maps can be created from the stage height at relatively low computational cost at continental scale.

The National Water Model is currently operating as a deterministic model for short-term predictions and does not currently include an estimate of the uncertainty in these predictions. The final streamflow values are at the end of a chain of models which originate from precipitation forecasts and go through rainfall-runoff and finally routing modules. The total uncertainty in the streamflow predictions is therefore a function of the uncertainty in each step.

Uncertainty analysis commonly relies on an assessment of uncertainty in model parameters and boundary conditions, the use of perturbed inputs or through comparison of several different models of the same systems. Estimated uncertainty from the first model in a chain can then be propagated to the next model and so on until a final estimate is achieved. Unfortunately, the National Water Model is operated on a super computer and the details of the model are not available for perturbation analysis.

One step in the National Water Model hourly cycle is the assimilation of USGS gage data which allows for corrections to the model state before the forecast simulation is made. This excludes USGS gage data from being used as a verification dataset. Even so, it is still an informative exercise to compare NWM predictions at these sites. There are numerous local and regional gaging stations which are not assimilated into the National Water Model and can be used as an independent check on the model output. Recent flooding in the Llano River basin in central Texas provides an opportunity to compare National Water Model predictions to both USGS and non-USGS gage readings. This thesis presents an assessment of the error in National Water Model predictions in the Llano River basin.

Table of Contents

List of Tables x

List of Figures xi

Chapter 1: Introduction 1

MOTIVATION 1

BACKGROUND 2

RESEARCH QUESTIONS 4

Chapter 2: Literature Review 6

Chapter 3: Study Area and Data Sources 11

NATIONAL WATER MODEL FORECASTS 11

UNITED STATES GEOLOGICAL SURVEY STREAMFLOW 12

LOWER COLORADO RIVER AUTHORITY STREAMFLOW 12

NATIONAL HYDROGRAPHY DATASET 13

DATA PROCESSING 13

STUDY AREA 15

FLOODING EVENTS 16

Chapter 4: National Water Model Analysis & Assimilation and Short-Term Forecast Performance 17

PERFORMANCE METRICS 17

GRAPHICAL ANALYSIS 19

ANALYSIS AND ASSIMILATION PERFORMANCE 30

Graphical Analysis 30

Statistical Metrics 30

SHORT-TERM FORECAST PERFORMANCE 37

PSEUDO-ENSEMBLE ANALYSIS 39

Chapter 5: Discussion and Conclusions 51

RESEARCH QUESTIONS 51

DISCUSSION 53

FUTURE WORK 55

Appendix A 57

Appendix B 61

Appendix C 63

Works Cited 69

List of Tables

Table 1: Llano River gages and corresponding USGS, LCRA and COMID identifiers where applicable. 14

Table 2: Performance metrics for NWM analysis and assimilation forecast and observed gage readings. 32

List of Figures

Figure 1: Llano River watershed study area with locations and names of USGS and LCRA gaging sites. 15

Figure 2: Hydrograph for the USGS Llano River near Mason, TX gaging station showing two distinct flood events 16

Figure 3: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines, see text) for flood 1 at the North Llano River near Junction, TX USGS gage site. 20

Figure 4: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the North Llano River near Junction, TX USGS gage site. 20

Figure 5: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the South Llano River near Junction, TX USGS gage site. 21

Figure 6: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the South Llano River near Junction, TX USGS gage site. 21

Figure 7: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Llano River near Junction, TX USGS gage site. 22

Figure 8: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Llano River near Junction, TX USGS gage site. 22

Figure 9: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Johnson Fork near Junction, TX LCRA gage site. 23
Figure 10: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the James River near Mason, TX LCRA gage site. 23
Figure 11: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Comanche Creek near Mason, TX LCRA gage site. 24
Figure 12: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Comanche Creek near Mason, TX LCRA gage site. 24
Figure 13: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Llano River near Mason, TX USGS gage site. 25
Figure 14: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Llano River near Mason, TX USGS gage site. 25
Figure 15: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Beaver Creek near Mason, TX USGS gage site. 26
Figure 16: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Willow Creek near Mason, TX LCRA gage site. 26

Figure 17: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Hickory Creek near Castell, TX LCRA gage site. 27
Figure 18: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the San Fernando Creek near Llano, TX LCRA gage site. 27
Figure 19: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Jonson Creek near Llano, TX LCRA gage site. 28
Figure 20: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Llano River near Llano, TX USGS gage site. 28
Figure 21: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Llano River near Llano, TX USGS gage site. 29
Figure 22: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Honey Creek near Kingsland, TX LCRA gage site. 29
Figure 23: NSE in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the first flood event (closer to one is better). 33
Figure 24: NSE in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the second flood event (closer to one is better). 33
34

Figure 25: PBIAS in the Llano River basin between observed gage flow and NWM analysis assimilation forecast, first flood event (closer to zero is better). 34
Figure 26: PBIAS in the Llano River basin between observed gage flow and NWM analysis assimilation forecast, second flood event (closer to zero is better). 34
Figure 27: RSR in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the first flood event (closer to zero is better). 35
Figure 28: RSR in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the second flood event (closer to zero is better). 35
Figure 29: PEP in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the first flood event (closer to zero is better). 36
Figure 30: RSR in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the second flood event (closer to zero is better). 36
Figure 31: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the N. Llano River near Junction, TX USGS gage. 40
Figure 32: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the N. Llano River near Junction, TX USGS gage. 41

Figure 33: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Llano River near Junction, TX USGS gage. 41
Figure 34: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Llano River near Junction, TX USGS gage. 42
Figure 35: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Johnson Fork near Junction, TX LCRA gage. 42
Figure 36: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the James River near Junction, TX LCRA gage.....43
Figure 37: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Comanche Creek near Mason, TX LCRA gage. 43
Figure 38: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Comanche Creek near Mason, TX LCRA gage. 44

Figure 39: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Llano River near Mason, TX USGS gage. 44
Figure 40: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Llano River near Mason, TX LCRA gage. 45
Figure 41: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Beaver Creek near Mason, TX USGS gage. 45
Figure 42: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Willow Creek near Mason, TX LCRA gage. 46
Figure 43: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Hickory Creek near Castell, TX LCRA gage. 46
Figure 44: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the San Fernando Creek near Llano, TX LCRA gage. 47

Figure 45: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Johnson Creek near Llano, TX LCRA gage. 47
Figure 46: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Llano River at Llano, TX USGS gage. 48
Figure 47: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Llano River at Llano, TX USGS gage. 48

Chapter 1: Introduction

MOTIVATION

Flooding is the most costly natural disaster in the United States, causing an average $7.96 billion in damages annually over the last 30 years (National Weather Service 2018) with global warming expected to further exacerbate these effect (Dottori et al., 2018). Tragically, destructive flooding is also often associated with loss of life and flood prediction has, therefore, been a significant component of hydrological research for several decades. Recent Texas floods serve to underscore the continuing importance of timely and accurate flood prediction and warning systems. Hurricane Harvey in 2017 caused an estimated $125 billion in damage (NOAA Office for Coastal Management 2018). Even more recently, in October of 2018, portions of the Texas Hill Country experienced back to back flood events over the span of two weeks. The immediate effects of this "catastrophic flooding" included the inundation of homes and the destruction of the FM 2900 bridge near Kingsland, TX (KUTV Staff 2018). The downstream effects of the deluge resulted in sediment loads so high that the City of Austin was forced to issue a boil water notice to over one million customers (Anchondo 2018; AustinTexas.Gov 2018). Given the widespread impacts of flooding, modeling and prediction of floods is a major area of study in hydrology and hydraulics.

An ideal flood early warning system is able to predict both the magnitude and timing of flood events accurately as well as early enough to allow for appropriate responses from local authorities. However, due to the complexity of the hydrologic and hydraulic phenomena involved, there is often considerable uncertainty in some or all parts of the flood prediction process. It is therefore important to quantify, as accurately as possible, the uncertainty associated with flood predictions so that end users (including

government officials and citizens) can incorporate this information into their decision-making processes.

BACKGROUND

Freeze and Harlan (1969) laid out a roadmap for what they called a "digitally-simulated hydrological response model" nearly forty years ago. Since then, increases in computational power have allowed for ever more detailed simulation and forecasting models while the increases in quality and quantity of remotely sensed hydrological data have allowed for hydrological modeling over vast spatial extents. However, larger model domains are often limited by the resolution, availability and resolution of the required physical and parameterized inputs which can lead to increased uncertainty in model output.

Models can be divided into two categories: deterministic and stochastic. Deterministic models give a single output, i.e. "this is the predicted streamflow". Conversely, stochastic models provide a range of potential values, where some sense of uncertainty is built into the model. Another way to think about deterministic models is that, for a single input, the output/prediction will always be the same. In a stochastic model, a given input may give rise to a range of possible outputs.

There are tradeoffs to using the two main types of models, especially when it comes to disseminating model predictions to the public. For example, it has been argued that a single value prediction is more easily understood by the general populace. However, it has also been shown that, with proper explanation, the information content of stochastic models can be well understood.

There have been many studies that attempt to incorporate uncertainty estimates into deterministic hydraulic models. One approach is to assess the uncertainty in the

model inputs and/or the model parameters. This operation is often accomplished by using a likely range of values for the inputs and parameters in question. By observing the effects on the predicted value (often streamflow or stage height), the relative effect (sensitivity) of the model to each input or parameter can be assessed.

Recently, the National Oceanic and Atmospheric Administration (NOAA) established the National Water Model (NWM) which predicts streamflow in over 2.7 million reaches across the continental United States (CONUS). At this time, there are four distinct NWM forecast products, of which the analysis/assimilation and short-range forecasts are the focus of this work. The hourly analysis/assimilation product is the model output of current streamflow condition and is used as the basis for the model restart file from which the three forecasts products are derived. The short-range forecast is an hourly, deterministic forecast that predicts streamflow from one to eighteen hours in the future. The medium-range forecast, also deterministic, extends ten days and is run four times a day. The long-range forecast is an ensemble product which is produced daily and extends out 30 days.

Recent efforts have focused on using NWM streamflow values to predict flood inundation extent by using DEM derived synthetic rating curves and the HAND technique. This method has shown to be quite accurate at the county scale, but less accurate at the reach scale and shows promise for predicting inundation in un-gaged watersheds as well as in areas that do not have detailed hydrologic models. Efforts to improve NWM-based flood predictions have focused on using high resolution digital elevation Lidar data to improve both synthetic rating curves and HAND values as well as improving channel extraction and inter-catchment water level propagation. Fundamental to the system described above are accurate streamflow values from the NWM, without which there cannot be accurate and timely flood predictions.

Because the NWM is operated by NOAA, extends across the CONUS, and is run on a supercomputer, typical desktop approaches to sensitivity and error estimation that rely on perturbations to model parameters are not possible. Additionally, as one goal of flood-predictions based on remotely sensed data is to provide predictive capabilities in areas that are un-gaged or do not have detailed hydrologic models, inter-model comparisons are also not possible. Therefore, this study attempts to glean as much information from the NWM short-range forecasts by treating grouping the data in two distinct fashions.

RESEARCH QUESTIONS

This thesis examines the performance of the NWM analysis/assimilation and short-range streamflow predictions over the Llano River basin in central Texas during the October 2018 flood events. Four main research questions are addressed.

1. *How do the NWM analysis/assimilation and short-range streamflow predictions compare to observed streamflow at USGS gages sites?*

   Because the NWM incorporates a data assimilation step every hour, we expect those stream reaches associated with USGS gages used in the assimilation to be very well modeled by the analysis and assimilation product. Additionally, the short-range is anticipated to behave well at short lead-times because the analysis and assimilation result is used as the short-range forecast model restart file. Longer lead times streamflow predictions are likely more heavily influenced by forecasted precipitation and the rainfall-runoff portion of the NWM and are therefore more likely to show some divergence from the observed stream flow values.

2. *How do the NWM analysis/assimilation and short-range streamflow predictions compare to the observed streamflow at non-USGS gage sites?*

   This study uses gages operated by the Lower Colorado River Authority (LCRA) that are not used in the NWM data assimilation step and thus serve as an independent verification dataset. The LCRA gage sites are located on tributaries of the Llano River and thus are also representative of a typical small un-gaged watershed with no detailed hydraulic model. Because the LCRA gages are located on tributaries, they also do not benefit from being downstream of a USGS gage site which would be expected to improve model accuracy. Instead, the LCRA gage sites are wholly predicted by the precipitation, rainfall-runoff and routing components of the NWM.

3. *Is there a significant difference in the quality of the prediction made between USGS and non-USGS gage sites?*

   Because access to the NWM structure is not available, a comprehensive discussion of errors is not possible. However, an analysis of available precipitation forecasts is presented.

4. *How does the magnitude of the error in NWM streamflow predictions change through time and space?*

   The USGS gage sites are located upstream, mid-river and downstream on the Llano River. Thus, we will be able to compare the rainfall-runoff dominated gages (upstream) with the routing-dominated (downstream) USGS gages.

Chapter 2: Literature Review

Flood Early Warning Systems (FEWS) can be used as an alternative, or in addition, to engineered flood mitigation structures to reduce the negative impacts of flooding on a community (Krzysztofowicz et al., 1994). These systems are built by combining real-time hydrological and meteorological monitoring stations, weather forecasts, and hydrologic models to make predictions of flow and water level which can be used to issue warnings (Haggett, 1998; Werner et al., 2005) early enough to allow for appropriate civil and public responses (Penning-Rowsell et al., 2000). In addition to riverine flooding – the focus of this study – FEWS are important for navigation and bridge clearance, fishing, recreation and industry (Parker and Fordham, 1996). While the implementation of FEWSs is complex and they can be costly to operate, the return on investment (ROI) for such systems has been estimated at greater than 100:1; with a potential ROI of 400:1 with improved weather forecasts and model performance (Pappenberger et al., 2015). The effectiveness of any FEWS is dependent on accurate, timely and actionable information, which is in turn dependent on appropriate assessment, analysis and dissemination of input and forecast uncertainty to the end users (Todini, 2004). The uncertainty in streamflow forecasts is often low for large rivers with detailed hydraulic models; however, uncertainty is significantly larger for small streams prone to flash flooding or when using precipitation forecasts and rainfall-runoff models (Todini, 2004).

Modeling of the hydrologic cycle to predict flooding has been a subject of interest for decades. To do so requires a mathematical description of the relevant subsystems including precipitation and evapotranspiration, as well as groundwater, unsaturated, overland and open channel flow (Freeze and Harlan, 1969). A detailed description of the

physical characteristics of the study area is also needed including elevation, soil type and ground cover characteristics. For these reasons, hydrologic models have typically been based on detailed engineering surveys at the local watershed scale; however, these types of models are labor intensive and prohibitively costly at continental scale.

With increased computational power, analysis of large-scale, remotely sensed hydrologic data sets has become possible. Early work focused on the extraction of drainage networks from digital elevation datasets (O'Callaghan and Mark, 1984; Tarboton et al., 1991). Increases in spatial resolution of digital elevation datasets has improved the accuracy of such channel extractions dramatically (Passalacqua et al., 2010b, 2010a). These flow networks, derived from remotely-sense data, are then used as the hydraulic framework around which hydraulic models are constructed. Recent work has further improved the accuracy of channel extraction using lidar derived digital elevation models (Zheng et al., 2018).

The National Water Model (NWM), operated by the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS), is a near-real-time flood prediction model operating across 2.7 million reaches in the continental US (Maidment, 2017). This work is focused on the analysis and assimilation, and the short-range NWM products, both of which are run hourly. Importantly, the analysis and assimilation product, which is the NWM best estimate of current conditions, is used as an initialization file for the short-range forecasts. Analysis and assimilation meteorological forcing is provided by the Multi-Radar/Multi-Sensor System (NOAA National Severe Storms Laboratory, 2015) while the short-range forecast uses High Resolution Rapid Refresh (HRRR) and Rapid Refresh (RR) forecasts (Benjamin et al., 2016). Routing in the NWM is along the National Hydrography Dataset plus version two (NHDPlusV2, McKay et al., 2018).

The NWM streamflow forecasts are at the end of a chain of models, with each model having an associated uncertainty in both the input forcing data and the output from the previous model in the chain.

One approach to improving the integration of uncertainty into both the modeling and dissemination of FEWS products is to use ensemble forecasting techniques (Cloke and Pappenberger, 2009; Renner et al., 2009; J. C. Schaake et al., 2007). Zhu et al., (2002) used a cost-loss framework to show that, in a majority of cases, ensemble weather forecasts provide greater "economic value" than single-valued (control) forecasts alone, with the increased benefit becoming near-universal at lead times greater than 72 hours. The location of the ensemble within a typical FEWSs model chain can vary from, for example, the initial meteorological inputs (Schaake et al., 2007) or by conducting a multi-model hydraulic analysis in parallel (Zarzar et al., 2018). Further, Schaake et al., (2007) have presented a method for deriving ensemble precipitation and temperature predictions from a single model output.

Uncertainty in distributed models can also be assessed using the generalized likelihood uncertainty estimation (GLUE) framework which asserts that errors and uncertainty in model structure, parameters and boundary conditions can lead to many equally likely descriptions of the system in question (Beven and Binley, 1992). The GLUE methodology has been used to show the uncertainty in flood inundation extent in an area with a detailed engineering model (Pappenberger et al., 2006). At its core, the GLUE method is based on testing many possible combinations of parameter sets that are varied within a set range to find a set of possible model setups which can then be used to assess the range of likely model outcomes and thus is a recognition that there is likely no single "best" set of calibration parameters.

Inundation mapping is often an important component of FEWSs and hydrological modeling. Potential inundation maps are used to determine areas at risk for flooding and determine requirements for flood insurance, determine risk and plan response activities (National Academy of Sciences, 2009). Retrospective flood inundation maps are an important component in recovery and disaster aid applications.

A key component in flood risk assessment are flood elevation profiles which are generally costly to generate and require detailed study of an area. This leads to poor or non-existent maps in many parts of the United States (National Academy of Sciences, 2009). Recent work has focused on generation of flood elevation profiles from remotely sensed data. One method of particular interest is the Height Above Nearest Drainage (HAND), which is used to determine the relative height of any point on the land surface relative to the nearest streambed using digital elevation raster data (Nobre et al., 2016; Rennó et al., 2008). Zheng et al. (2018) present a method for extraction of river geometry and rating curves using remotely sensed data. When combined with the HAND method this allows for near real-time generation of flood inundation maps at continental scale based on NWM predictions (Zheng et al., 2018). Uncertainty in flood inundation prediction is caused by uncertainties in flow rates, topography, and uncertainty in the underlying hydraulic model (Merwade et al., 2008). Because these inundation methods rely on rating curves – the relationship between discharge and stage height – any error or uncertainty in streamflow measurements or predictions will result in a corresponding error or uncertainty in the state height and therefore inundation extent. Uncertainty in the rating curve itself is reported to play a small role in the overall uncertainty in flood predictions (Ocio et al., 2017), while uncertainty in model boundary conditions can play a significant role in the uncertainty in inundation extent (Pappenberger et al., 2006). Within a chain of models such as the NWM, boundary conditions are at least partially defined by

the output of the previous step and in this way uncertainty is propagated through the model.

Chapter 3: Study Area and Data Sources

The acquisition, processing and organization of data are critical to this investigation, which is the focus of this chapter. First, the four main data sources are explored, followed by an overview of the data storage and pre-processing methods and finally the study area is described.

NATIONAL WATER MODEL FORECASTS

NOAA does not archive National Water Model forecasts. This presents a problem for retrospective analyses unless data are continually retrieved and stored locally. The Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI), via the HydroShare web platform (https://hs-apps.hydroshare.org/apps/nwm-forecasts/api-page/) provides a 40-day rolling window of the full suite of NWM products that can be accessed through their National Water Model Viewer web app or accessed via the application programmers interface (API). Importantly, the API service allows users to subset the full CONUS NWM output either with a list of COMIDs of interest or by spatial extent. This allows for a greatly reduced file size and lowers processing time for extended time-series analysis. For this work, subsets of NWM short-term streamflow forecasts were collected and archived for the state of Texas starting on April 05, 2018. Though the current analysis is focused on only a fraction of the total stream reaches in Texas, the archive will allow for future work in other areas of Texas than are addressed in this work. The Python script used to subset NWM forecasts is shown in Appendix One. The bulk of this script was provided by Dr. Tim Whiteaker and was modified by the author to allow for repeated calls to the API service based on a range of dates as well as to retrieve each of the twenty-four hourly forecasts released each calendar day.

NWM short range streamflow forecasts are released hourly and contain hourly streamflow predictions from one to eighteen hours from the prediction time. Each forecast is obtained as an individual netcdf4 formatted file. Because the short-term forecast product contains eighteen predictions, there is substantial overlap between sequential forecasts.

UNITED STATES GEOLOGICAL SURVEY STREAMFLOW

The United States Geological Survey (USGS) maintains or co-maintains numerous weather and streamflow gages throughout the United States. River monitoring stations record, at a minimum, streamflow discharges and many stations also include stage height measurements. These data are freely available and can be automatically retrieved by USGS gage number. The Python script used to retrieve these data is shown in Appendix A.

LOWER COLORADO RIVER AUTHORITY STREAMFLOW

The Lower Colorado River Authority (LCRA) maintains or co-maintains numerous weather and streamflow gages throughout the lower Colorado River basin in central Texas. Some of the gages are co-operated with the USGS and thus do not provide additional information, however there are several gages on tributaries that are solely operated by the LCRA and thus can serve as an independent check of NWM streamflow predictions. These data were retrieved by manual download for each gage from the LCRA Hydromet website (https://hydromet.lcra.org/). Gages that are co-operated by LCRA and USGS are referenced by USGS gage number.

NATIONAL HYDROGRAPHY DATASET

The National Hydrography Dataset *Plus* Version 2 (NHDPlusV2) serves as the hydrologic framework for NWM streamflow predictions. The fundamental unit of the NHDPlusV2 is the reach, each of which is assigned a unique COMID. To compare gage data to NWM forecasts, it is necessary to associate each gage with the appropriate COMID. For this study, this was accomplished visually using a Geographic Information System (GIS) interface to display the NHDPlusV2 reaches and USGS gages based on geographic coordinate and LCRA gages based on geography, road crossings and stream names within the NHDPlusV2. A summary of streamflow gages in the Llano River basin with USGS or LCRA gage identifiers and corresponding NHDPlusV2 COMIDs is shown in Table 1.

DATA PROCESSING

Data were downloaded and archived in the default format. NWM files are provided as netcdf4, USGS gage data are provided in tab-delimited text format, LCRA gage data are provided as comma-delimited text files, and the NHDPlusV2 is provided as a GIS shapefile. USGS, LCRA and the NHDPlusV2 are all available within a single file, so require no extra processing. In contrast, NWM forecasts are provided as individual files for each hourly forecast. Thus, an additional step of extracting the predictions for the COMIDs of interest and combing them into a single file is necessary. As described above, the NWM forecast archive was subset for the entire state of Texas and contains streamflow forecasts for over 100,000 reaches.

Gage Site	USGS Gage Number	LCRA Site ID	NHDPlusV2 COMID
N. Llano River near Junction, TX	08148500	--	5761759
S. Llano River at Flat Rock Ln at Junction, TX	08149900	--	5765175
Llano River near Junction, TX	08150000	--	5770577
Johnson Fork near Junction, TX	--	2313	5772343
James River near Mason, TX	--	2399	5770009
Comanche Creek near Mason, TX	--	2424	5769643
Llano River near Mason, TX	08150700	--	5769807
Willow Creek near Mason, TX	--	2443	5770503
Hickory Creek near Castell, TX	--	2498	5771211
San Fernando Creek near Llano, TX	--	2616	5771717
Johnson Creek near Llano, TX	--	2625	5771703
Llano River at Llano, TX	08151500	--	5771725
Honey Creek near Kingsland, TX	--	2694	5771417

Table 1: Llano River gages and corresponding USGS, LCRA and COMID identifiers where applicable.

STUDY AREA

The area under investigation in this study is the Llano River basin in central Texas. The Llano River starts as two main tributaries that merge at Junction, TX and flows east-north-east across the Edwards Plateau – a region known as "flash-flood alley" – to drain into Lake Travis, one of two main reservoirs for the Austin, TX area. The Llano River basin is relatively well gaged by both the USGS and the LCRA, with the USGS gages primarily on the main stem of the river and the LCRA gages on some of the major tributaries, as shown in Figure 1.

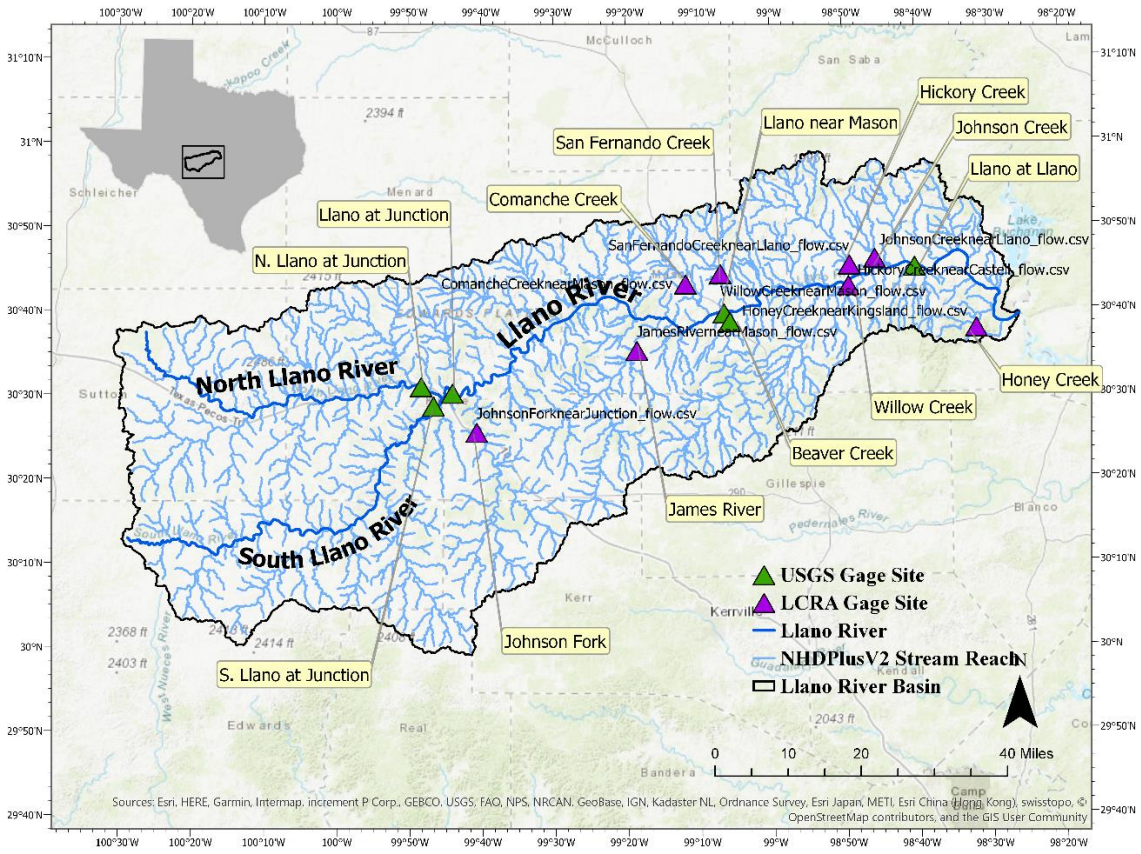

Figure 1: Llano River watershed study area with locations and names of USGS and LCRA gaging sites.

FLOODING EVENTS

The Llano River experienced historic flooding during the first and second weeks of October 2018. During this time, two distinct high flow events are observed in many of the gage sites in the basin, with an intervening period of significantly lower flow rates. The hydrograph for the USGS Llano River near Mason gage is shown below in Figure 2.

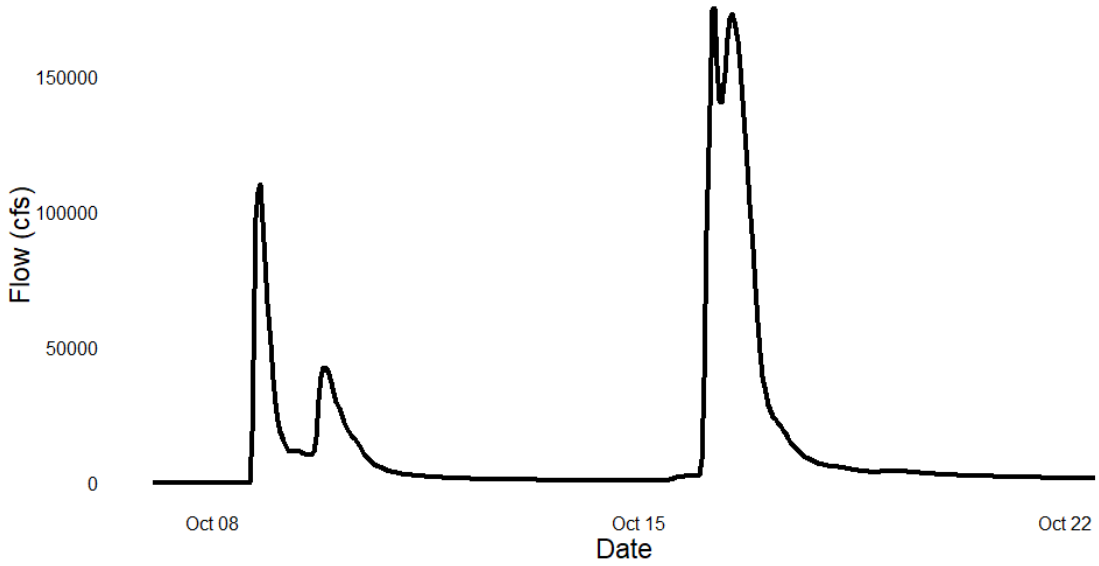

Figure 2: Hydrograph for the USGS Llano River near Mason, TX gaging station showing two distinct flood events.

At this site, there are clearly two extremely high flow events, where the first flooding event is somewhat split into two flow peaks and the second showing a small decrease in flow at the apex. Due to variations in the spatial distribution of precipitation during the flooding events, not every gage site showed extreme flow values for both events. In these cases, analysis is limited to the single observed flooding event.

Chapter 4: National Water Model Analysis & Assimilation and Short-Term Forecast Performance

This section will examine the performance of each individual short-term NWM forecast as compared to the observed streamflow at every USGS and LCRA gage site in the Llano River basin. As a convention, each short-term forecast will be referenced to the initialization time of the forecast which, for example, corresponds to one hour earlier than the one-hour lead time forecast.

PERFORMANCE METRICS

Moriasi et al. (2007) present a review of statistical methods used in the comparison between model output and observations in watershed simulations. Several methods are then recommended based on robustness, common use and strength in model evaluation, five of which are used in this study and are presented below. For all equations, Y_i^{obs} is the ith observed streamflow gage reading (indexed by time), Y_i^{pred} is the ith NWM predicted streamflow value, and $\bar{Y}^{obs}$ is the average observed streamflow over the interval $i = 1{:}n$, where n = 18 for all NWM short-range forecasts.

1. Nash-Sutcliffe efficiency (NSE), which compares the residual variance to the variance of the observations, is calculated using equation 1 (Nash and Sutcliffe, 1970):

$$\text{NES} = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - \bar{Y}^{obs}\right)^2}\right] \quad (1)$$

   where the range of values is $(-\infty, 1]$, with values ≥ 0 indicating model performance better than using the mean of the observations as a predictor.

2. Percent bias (PBIAS), which is an assessment of the tendency of predicted values to be above or below the observed values, is calculated using equation 2 (Gupta et al., 1999):

$$\text{PBIAS} = \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{pred}\right) * 100}{\sum_{i=1}^{n}\left(Y_i^{obs}\right)}\right] \quad (2)$$

where values closer to zero indicate closer agreement between predictions and observations.

3. RMSE-observation standard deviation ratio (RSR), which is RMSE standardized to the standard deviation of the observed variable, is calculated using equation 3:

$$\text{RSR} = \left[\frac{\sqrt{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{pred}\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(Y_i^{obs} - \bar{Y}^{obs}\right)^2}}\right] \quad (3)$$

where values closer to zero indicate a smaller RMSE and therefore better agreement between model and observation.

4. Percent error in peak flowrates (PEP), which is the percent error between model and observation for any given storm event, is calculated using equation 4:

$$\text{PEP} = \frac{Y_{max}^{pred} - Y_{max}^{obs}}{Y_{max}^{obs}} * 100 \quad (4)$$

where values close to zero indicate close agreement in magnitude of the peak flow event, though do not address differences in the timing of the peaks.

5. Slope and Pearson's coefficient of determination (R^2) are calculated using normal least squares regression between the simulated and observed flowrates during the storm period. T- and p-values ($\alpha = 0.05$) are provided for the slopes.

GRAPHICAL ANALYSIS

The large amount of data makes presentation in tabular format inefficient and difficult to interpret. Therefore, as a supplement to the above analysis, the performance of the NWM short-range forecasts at each gage site is presented below in a graphical format. Because the spatial distribution of precipitation over the Llano River watershed was not uniform between the two storms in October of 2018, some gage sites only show extreme flow events for one of the two storms. For gage sites that did experience high streamflow for both storms, the figures are notated as "Flood Event 1" and "Flood Event 2" as appropriate. Each figure shows the observed streamflow (blue filled area), the NWM analysis and assimilation forecast (black line) and the overlapping NWM short-range forecasts (colored lines). The line color of the short-term forecasts is based on the forecast effective time. Colors range from orange to yellow to green to blue to purple to red for earlier to later forecast, respectively. In some cases, the observed streamflow is so much larger than the NWM predictions that any variation in the forecasts is obfuscated by the scale of the ordinate. In these cases, a supplemental figure of just the NWM analysis and assimilation and short-range forecasts is provided. The gage sites are presented in approximately west to east order.

The domain for each flood event was selected automatically based on the maximum observed streamflow. For each flooding event, all times with observed streamflow values greater than or equal to 10% of the maximum observed streamflow were selected and the corresponding NWM forecasts were selected using the same timestamp range.

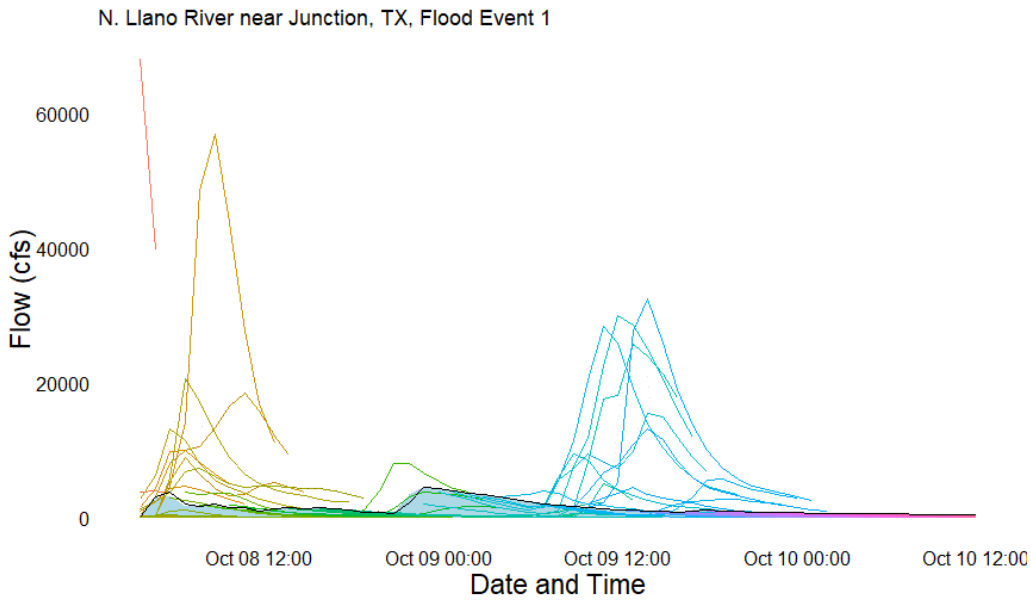

Figure 3: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines, see text) for flood 1 at the North Llano River near Junction, TX USGS gage site.

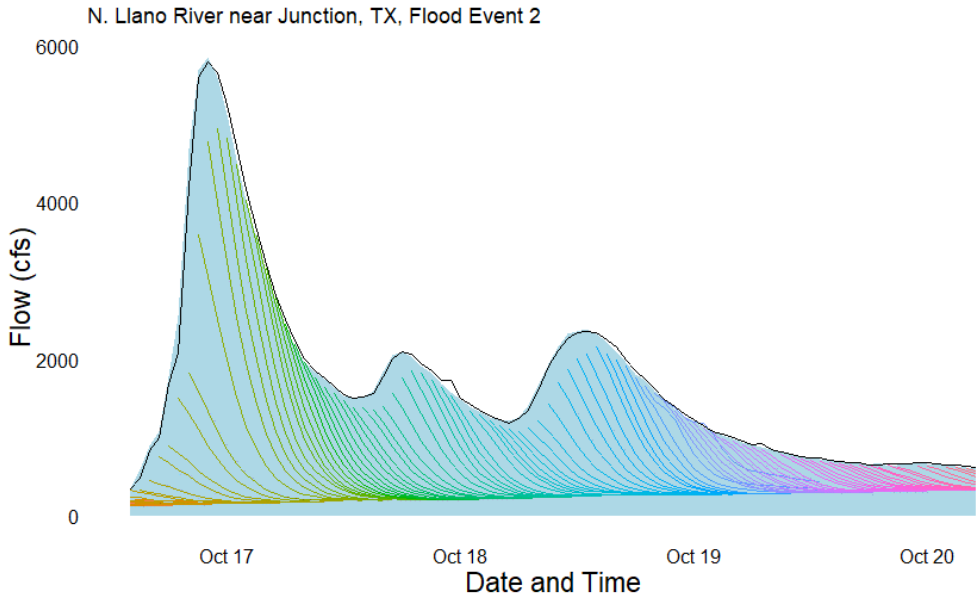

Figure 4: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the North Llano River near Junction, TX USGS gage site.

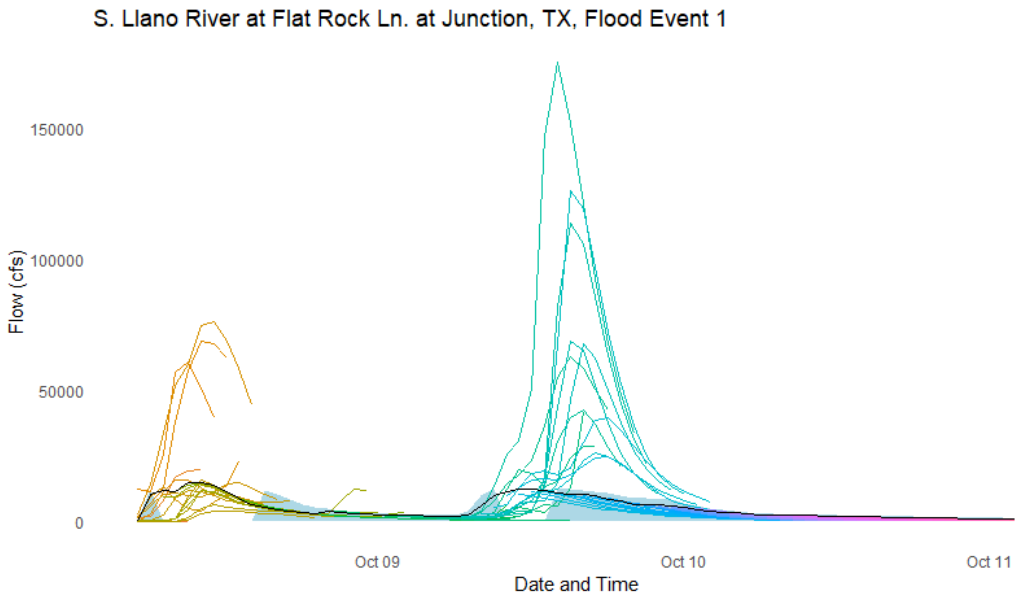

Figure 5: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the South Llano River near Junction, TX USGS gage site.

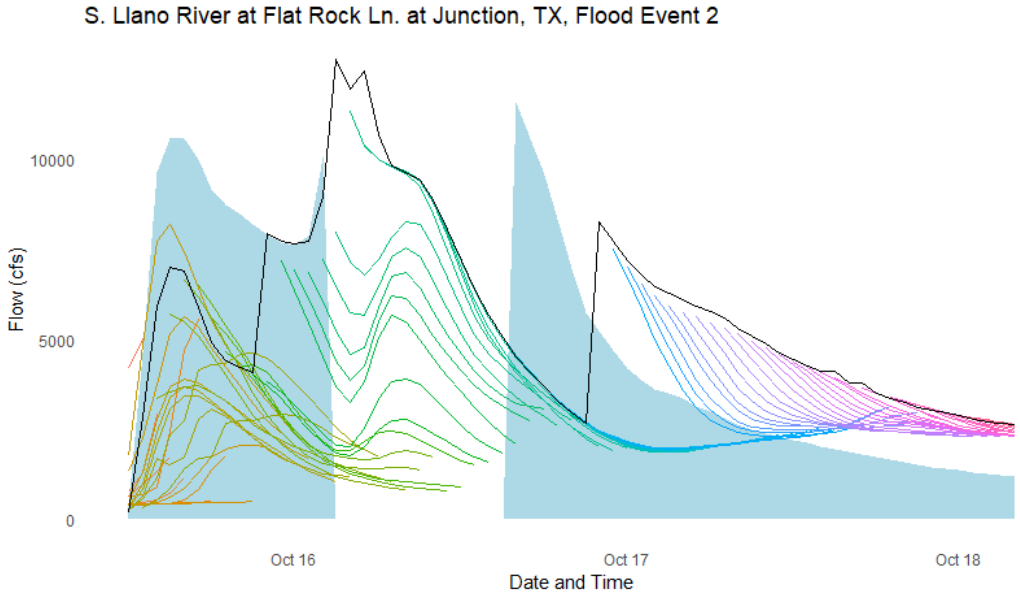

Figure 6: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the South Llano River near Junction, TX USGS gage site.

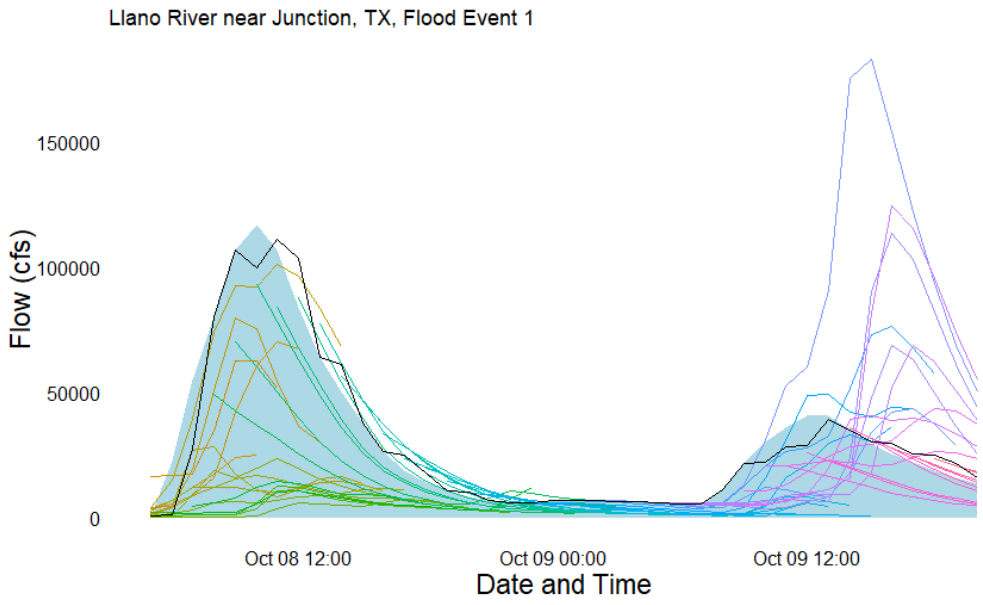

Figure 7: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Llano River near Junction, TX USGS gage site.

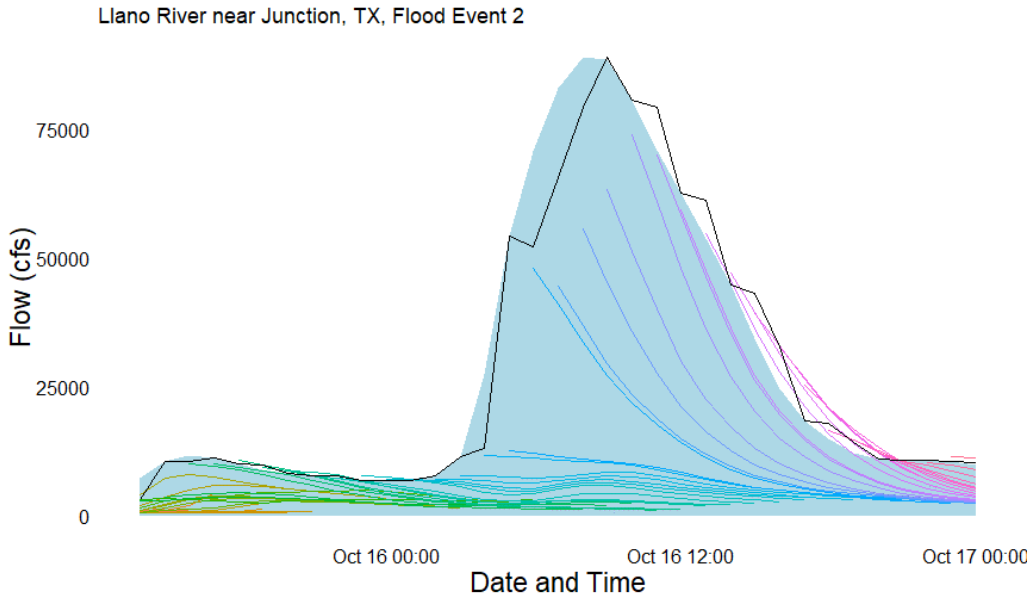

Figure 8: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Llano River near Junction, TX USGS gage site.

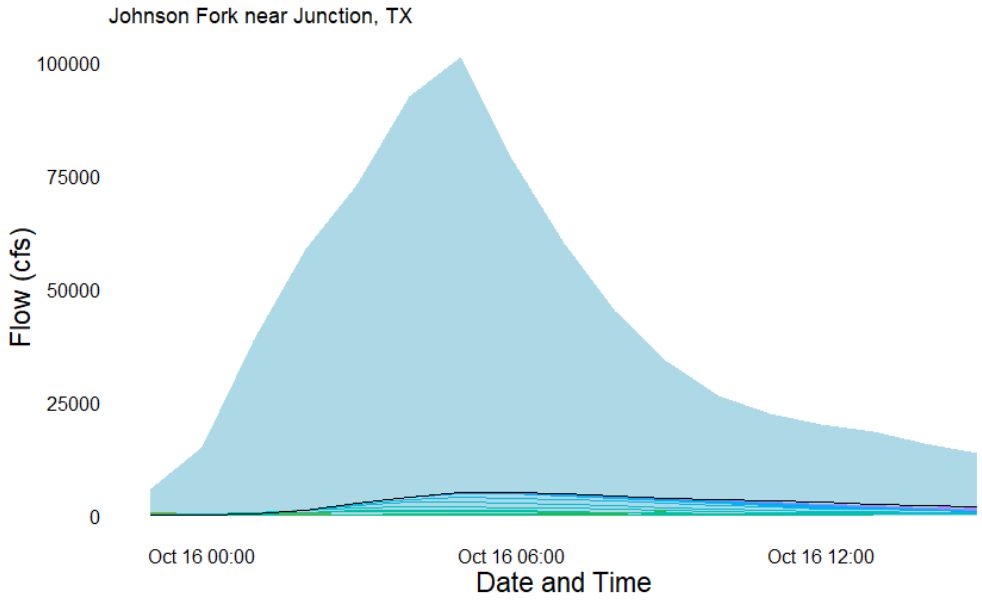

Figure 9: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Johnson Fork near Junction, TX LCRA gage site.

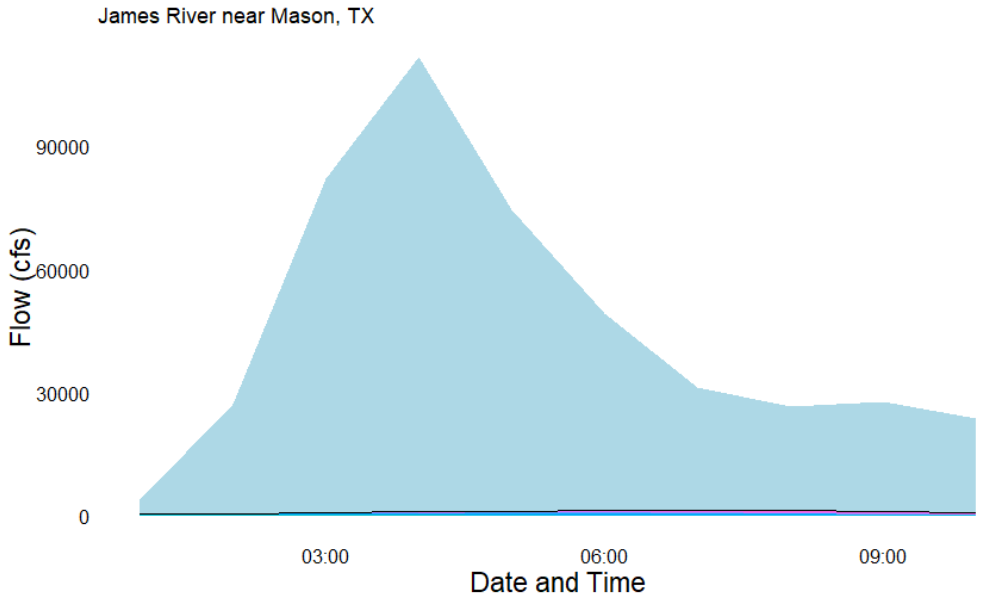

Figure 10: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the James River near Mason, TX LCRA gage site.

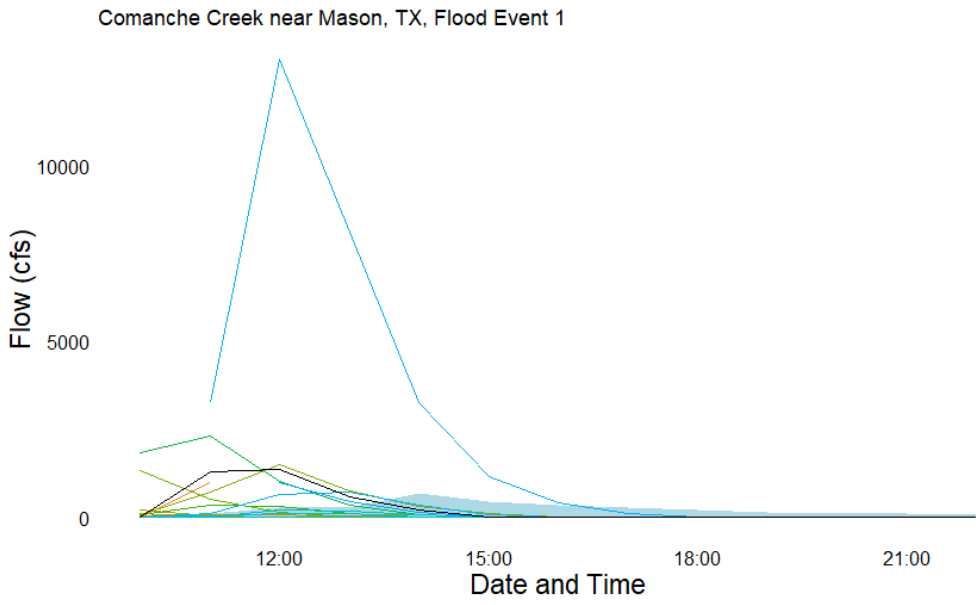

Figure 11: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Comanche Creek near Mason, TX LCRA gage site.

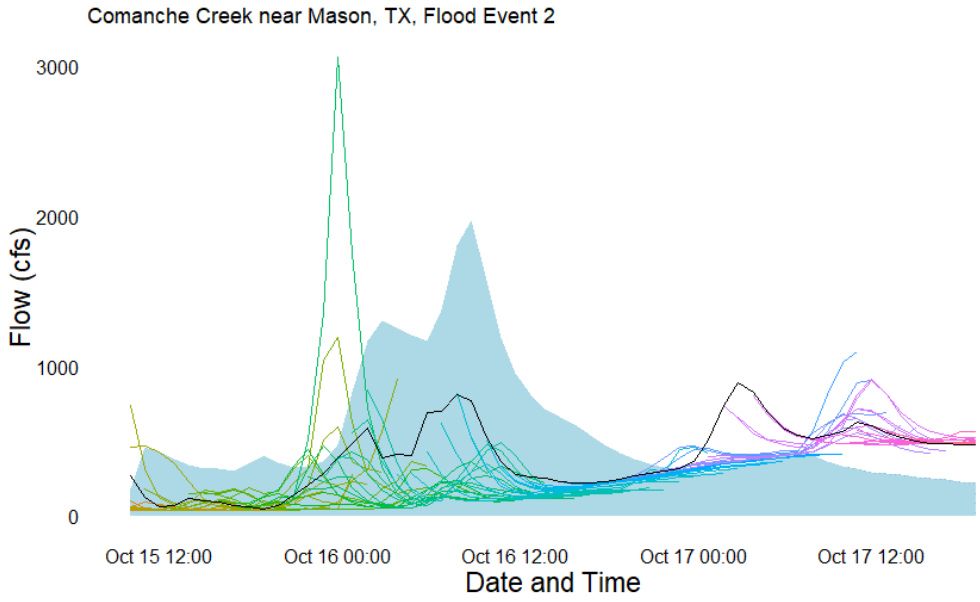

Figure 12: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Comanche Creek near Mason, TX LCRA gage site.

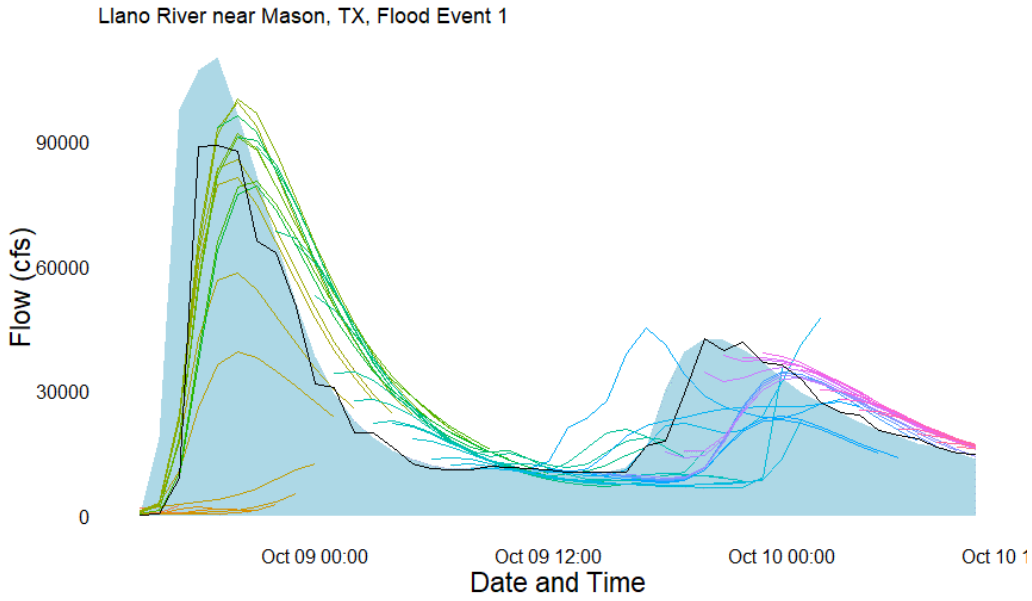

Figure 13: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Llano River near Mason, TX USGS gage site.

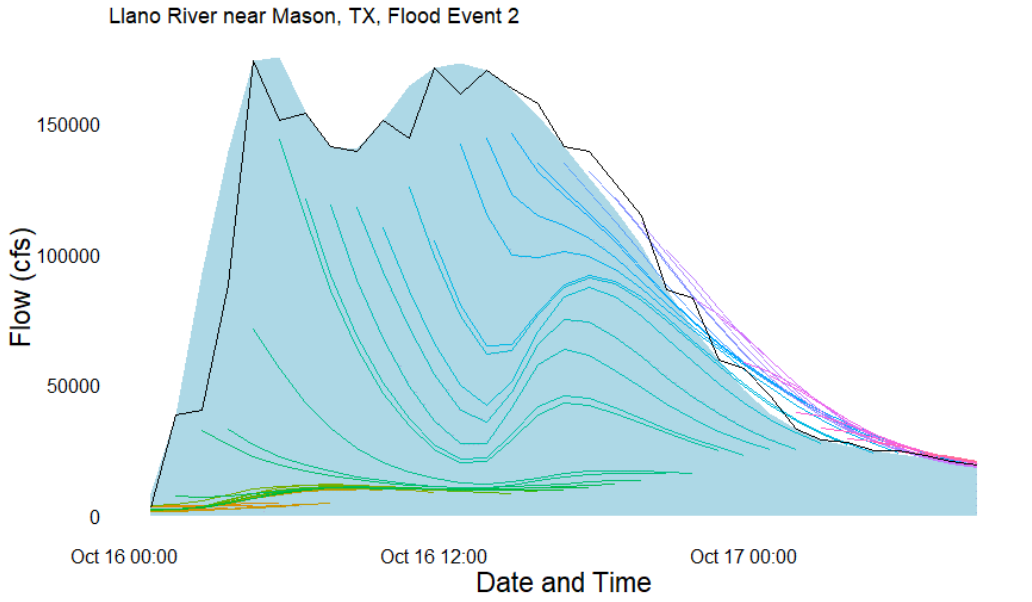

Figure 14: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Llano River near Mason, TX USGS gage site.

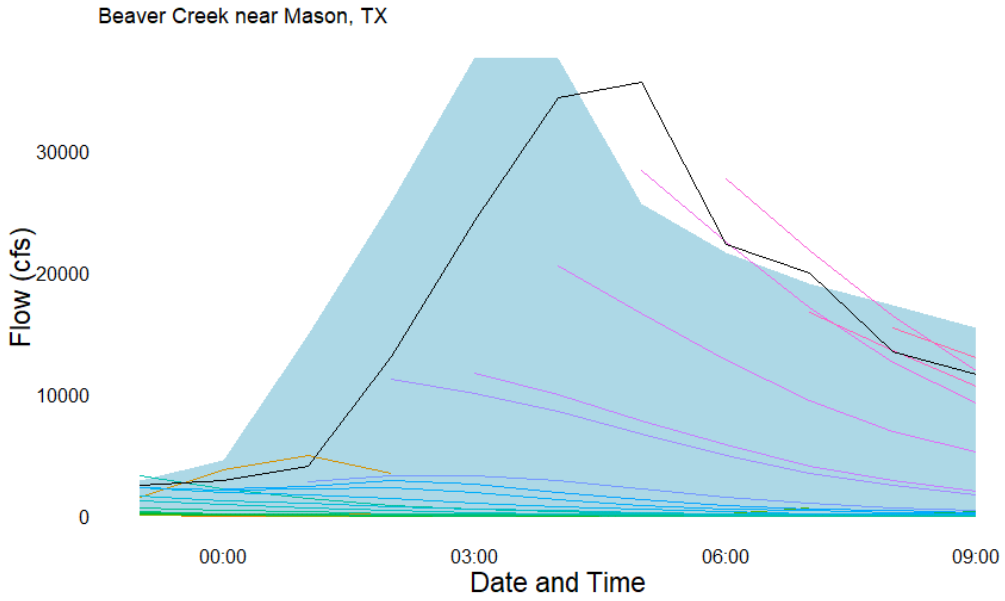

Figure 15: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Beaver Creek near Mason, TX USGS gage site.

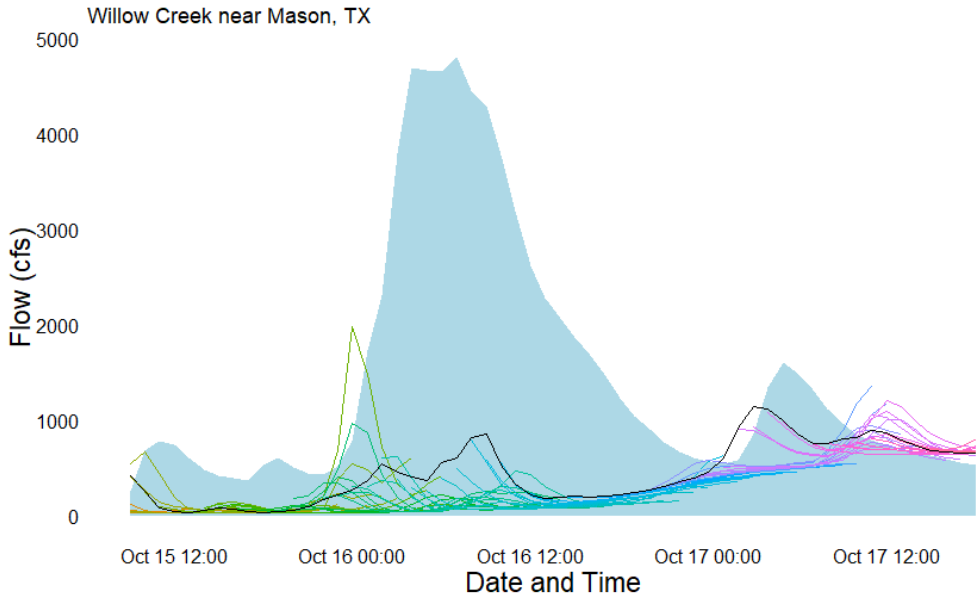

Figure 16: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Willow Creek near Mason, TX LCRA gage site.

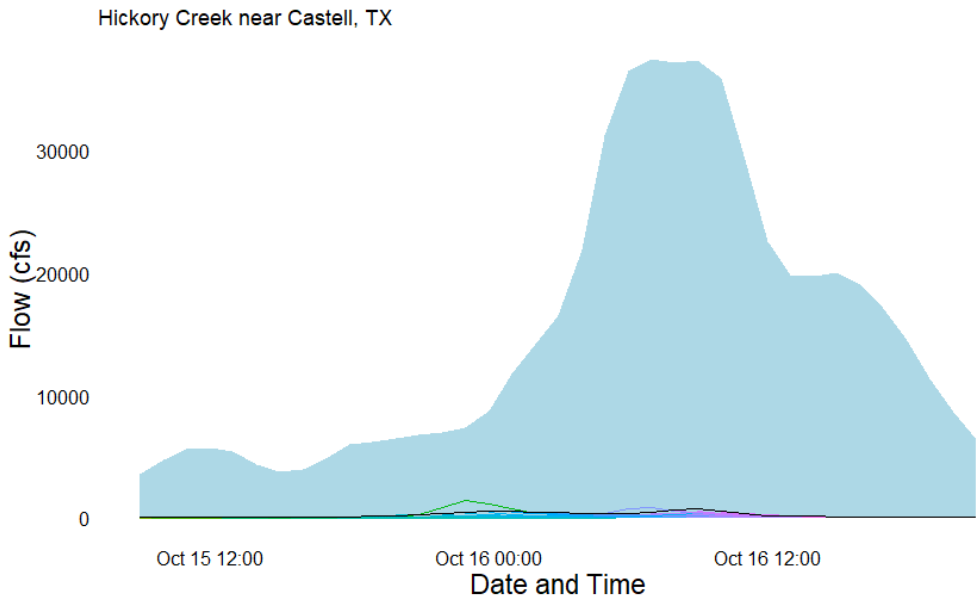

Figure 17: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Hickory Creek near Castell, TX LCRA gage site.

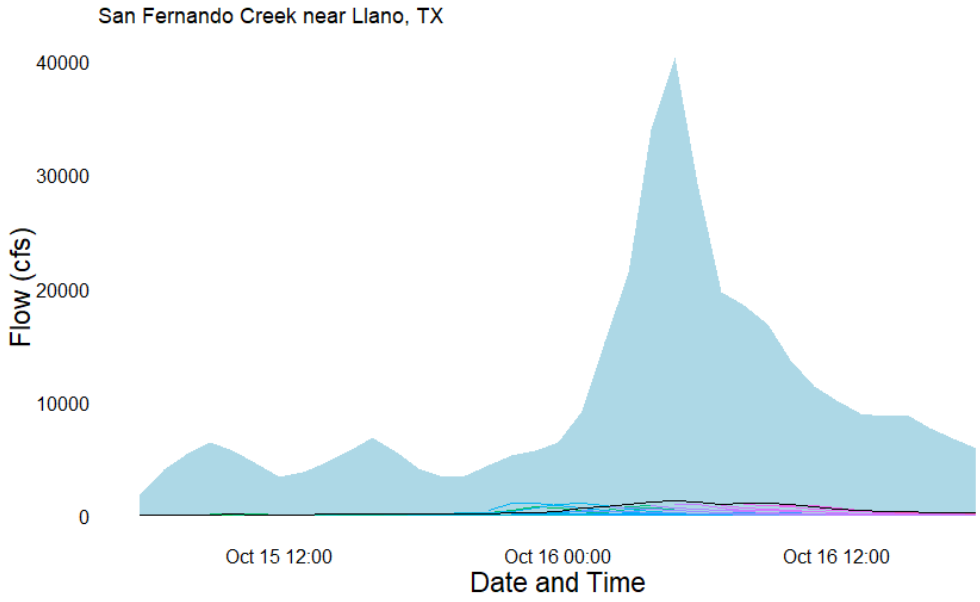

Figure 18: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the San Fernando Creek near Llano, TX LCRA gage site.

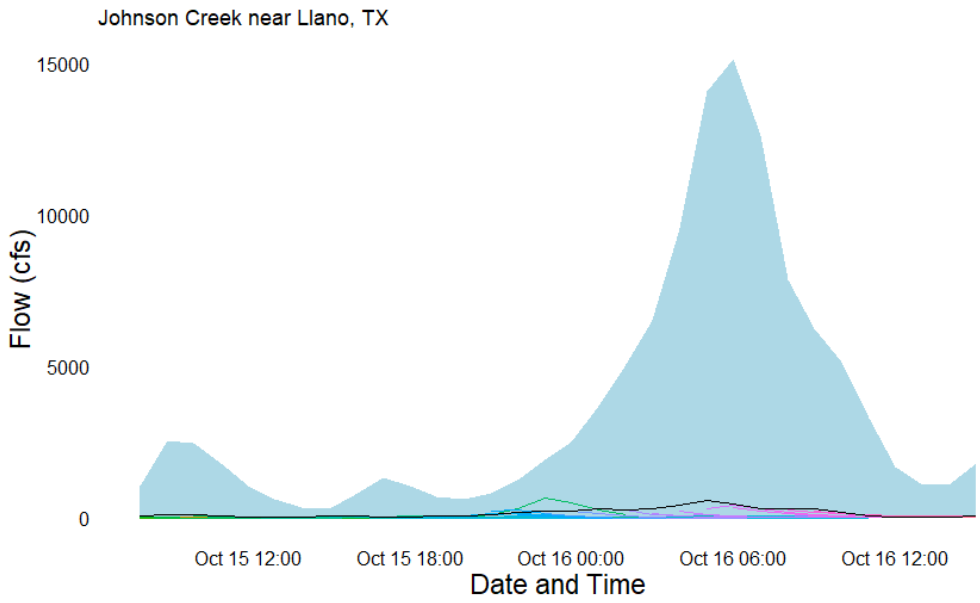

Figure 19: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Jonson Creek near Llano, TX LCRA gage site.

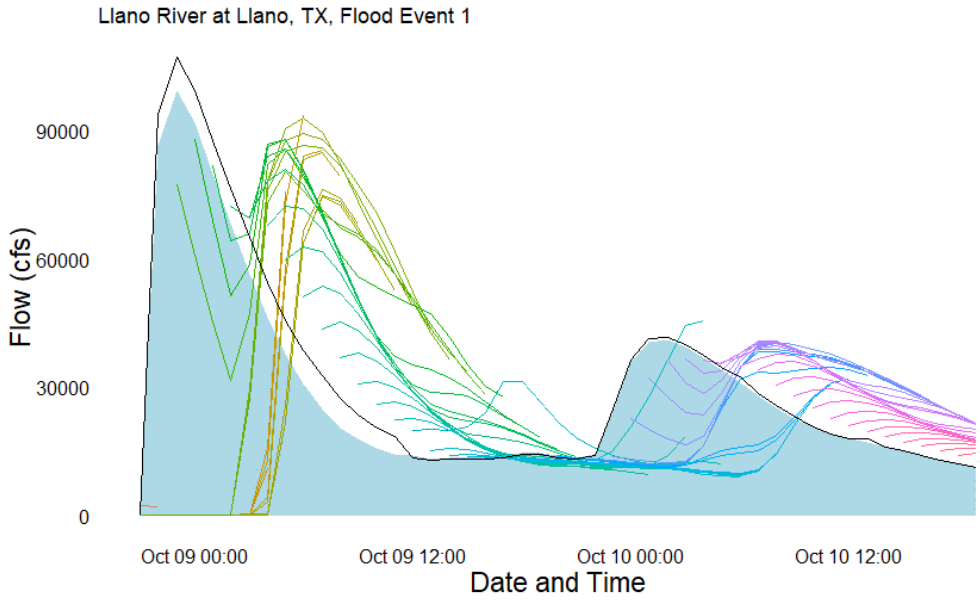

Figure 20: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 1 at the Llano River near Llano, TX USGS gage site.

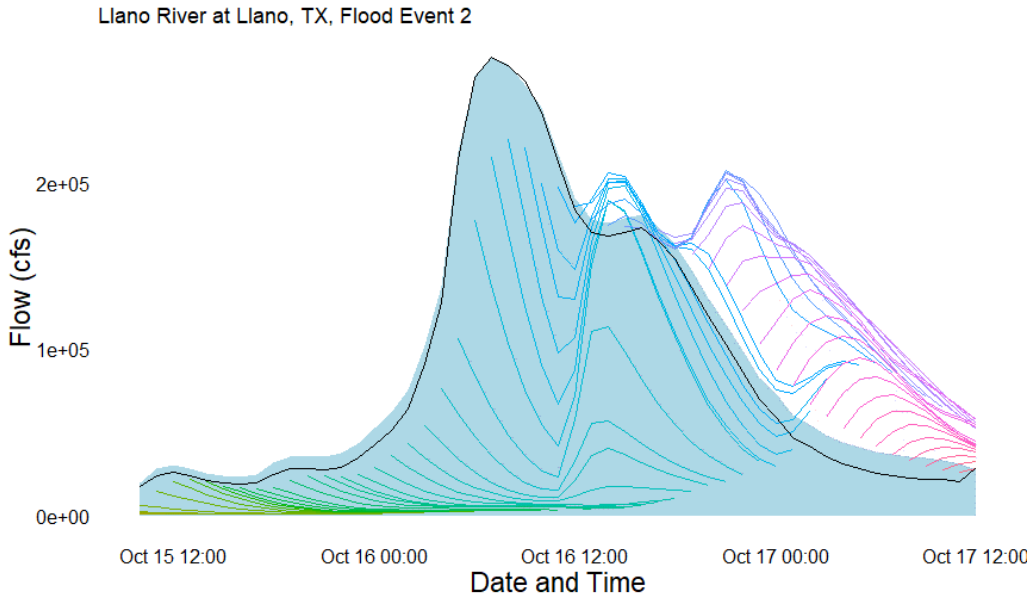

Figure 21: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) for flood 2 at the Llano River near Llano, TX USGS gage site.

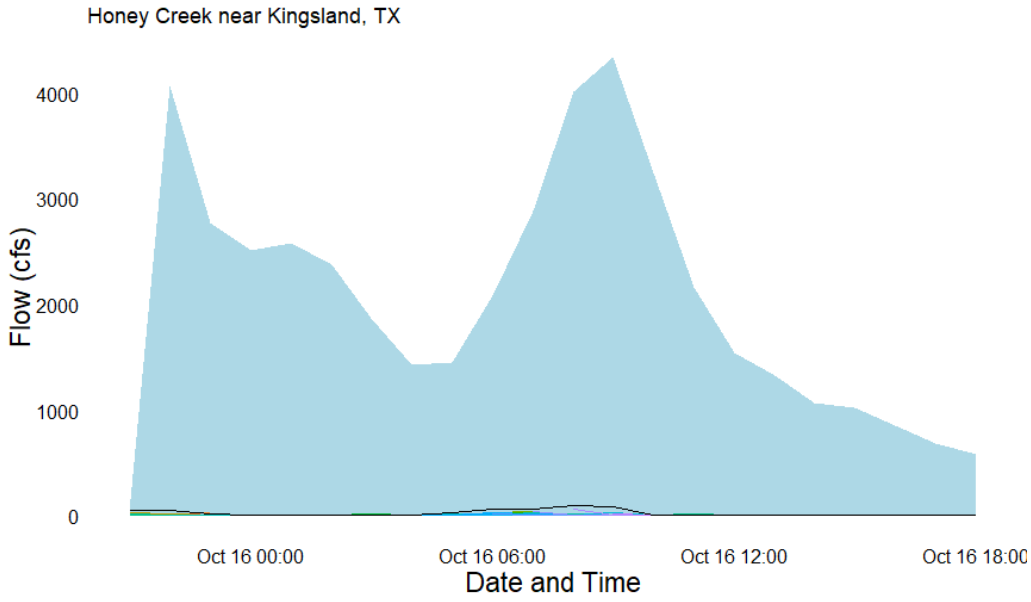

Figure 22: Observed flows (light blue area), NWM analysis and assimilation prediction (black line) and NWM short-range forecasts (colored lines) at the Honey Creek near Kingsland, TX LCRA gage site.

ANALYSIS AND ASSIMILATION PERFORMANCE

Graphical Analysis

The most obvious trend seen in the comparison between observed flows and the NWM analysis and assimilation prediction is the dichotomy between USGS and LCRA gaged sites. Said another way, when gage observations are incorporated into the NWM the agreement is quite good. This is expected given that the analysis and assimilation product is a best estimate of current conditions and therefore should be nearly identical to the actual gage readings at USGS gage sites because these are the sites being assimilated. Conversely, it is clear the NWM analysis and assimilation product does not accurately capture the observed streamflow at the LCRA gage sites, which are not part of the assimilation process. Potential causes of this split behavior will be discussed in a later section.

Statistical Metrics

For both flooding events NSE, PBIAS, RSR and PEP were calculated using the observed gage flows and the NWM analysis and assimilation forecast, the results of which are shown in Table 2 below. Note that only five of the gage sites have significant flows for the second flooding event. These sites have two entries for each column where the left entry is the first flooding event and the right entry is the second flooding event. Each metric is also shown graphically as seen in Figure 23 – Figure 30. As was the case for the graphical analysis, it is clear from the performance metrics that the NWM analysis and assimilation forecast displays a high degree of skill at USGS sites and very little skill at LCRA gage sites. NSE values are all above 0.93 (with a max of one with perfect agreement) except for the first flood event at the Llano River near Mason, TX gage, where the analysis and assimilation forecast is notably below the observed flow and an

NSE value of 0.710 is calculated; however, at the same gage during the second flood event the NSE is 0.939, so there is no evidence for a systematic error. The NSE for the LCRA gage sites, conversely, is negative for each gage site, indicating no model skill. Both the S. Llano River and Beaver Creek USGS gages went offline during the flooding, and therefore the skill has been reduced giving a hybrid type behavior and NSE values between the other USGS and LCRA gages.

A similar trend to NSE is seen in the PBIAS values where USGS gage sites are generally within 10%, again with the exception of the first flood event at Llano River near Mason and the mixed performance of S. Llano at Mason and Beaver Creek. The majority of LCRA gages show underprediction on the order of 90% or more. The largest exception to this trend is at Comanche Creek where, during the first flooding event the NWM analysis and assimilation forecast tends to overpredict by around 11% and tends to underpredict by around 30% for the second flooding event. Willow Creek shows a tendency to underpredict of about 60%. Both Comanche Creek and Willow Creek experienced the lowest flow magnitudes of all the gages during either of the two flooding events, so the magnitude of the NWM analysis and assimilation forecast is similar in magnitude to the observed flows in these cases. This result is likely due to the precipitation forecast aligning with the actual rainfall in the drainage areas of these two gages, which is discussed further in a following section.

RSR, a ratio of the RMSE to the standard deviation of the observed flows shows the same pattern as above for NSE and PBIAS. Given the expected large standard deviations in the observed stream flows, where the flow rate ranges over an order of magnitude, any value of RSR over one is definitely cause for concern and indicates that the RMSE is larger than the standard deviation of the observations.

PEP again shows similar trends to the above three metrics, with the exception of Comanche Creek during the first flood event where the NWM peak flow is nearly double the observed maximum flow, though again these are the lowest magnitude flows seen at any of the gage sites.

Gage Site	NSE		PBIAS		RSR		PEP	
N. Llano River near Junction, TX	0.965	0.994	-1.679	0.003	0.188	0.075	-0.654	-1.022
S. Llano River at Flat Rock Ln. at Junction, TX								
Llano River near Junction, TX	0.934	0.954	1.992	2.574	0.257	0.215	-4.656	0.000
Johnson Fork near Junction, TX	-1.805		93.040		1.675		-94.773	
James River near Mason, TX	-2.011		97.816		1.735		-98.785	
Comanche Creek near Mason, TX	-7.341	-0.119	-11.304	30.825	2.888	1.058	94.432	-54.647
Llano River near Mason, TX	0.710	0.939	14.306	3.073	0.539	0.246	-19.091	-0.571
Beaver Creek near Mason, TX	0.527		17.225		0.688		-5.384	
Willow Creek near Mason, TX	-0.562		69.175		1.250		-76.209	
Hickory Creek near Castell, TX	-1.745		98.395		1.657		-98.082	
San Fernando Creek near Llano, TX	-1.166		95.410		1.472		-96.579	
Johnson Creek near Llano, TX	-0.675		95.544		1.294		-96.237	
Llano River at Llano, TX	0.969	0.987	-7.672	8.281	0.177	0.114	7.864	0.001
Honey Creek near Kingsland, TX	-3.076		98.770		2.019		-97.737	

Table 2: Performance metrics for NWM analysis and assimilation forecast and observed gage readings.

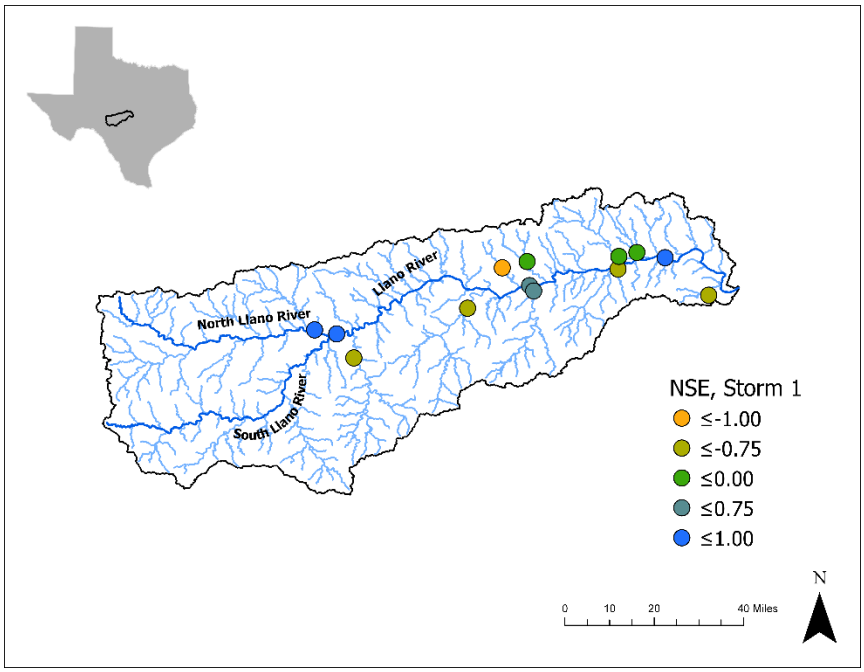

Figure 23: NSE in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the first flood event (closer to one is better).

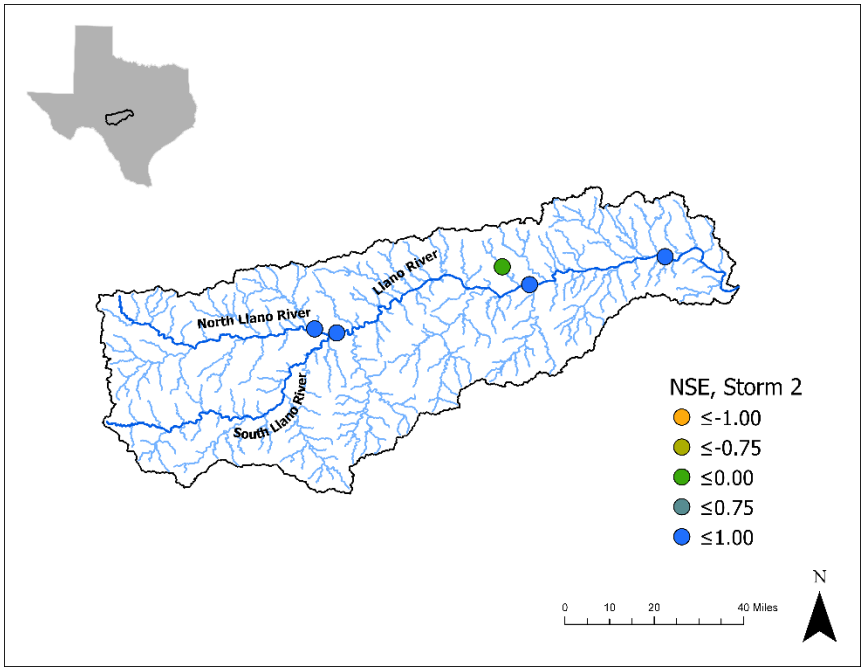

Figure 24: NSE in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the second flood event (closer to one is better).

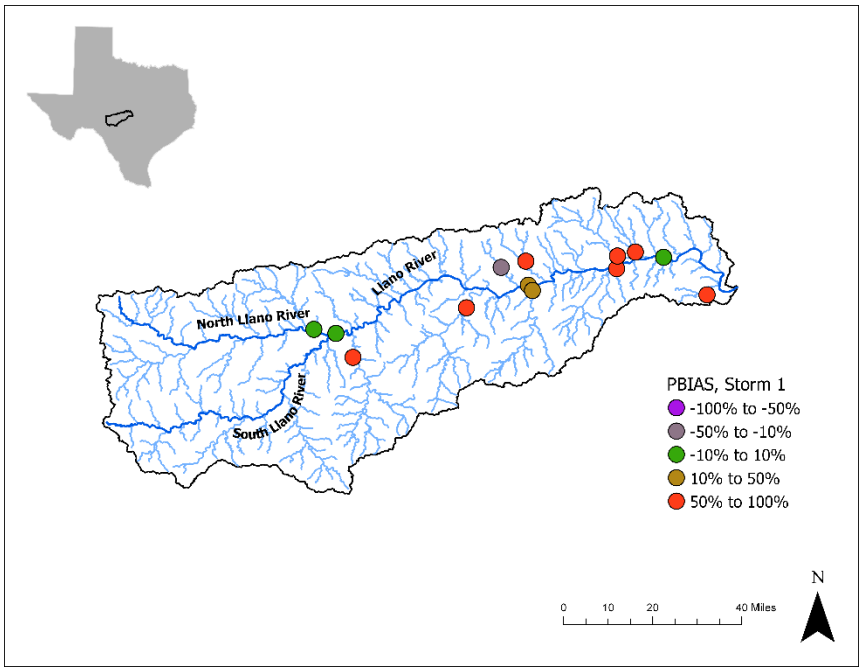

Figure 25: PBIAS in the Llano River basin between observed gage flow and NWM analysis assimilation forecast, first flood event (closer to zero is better).

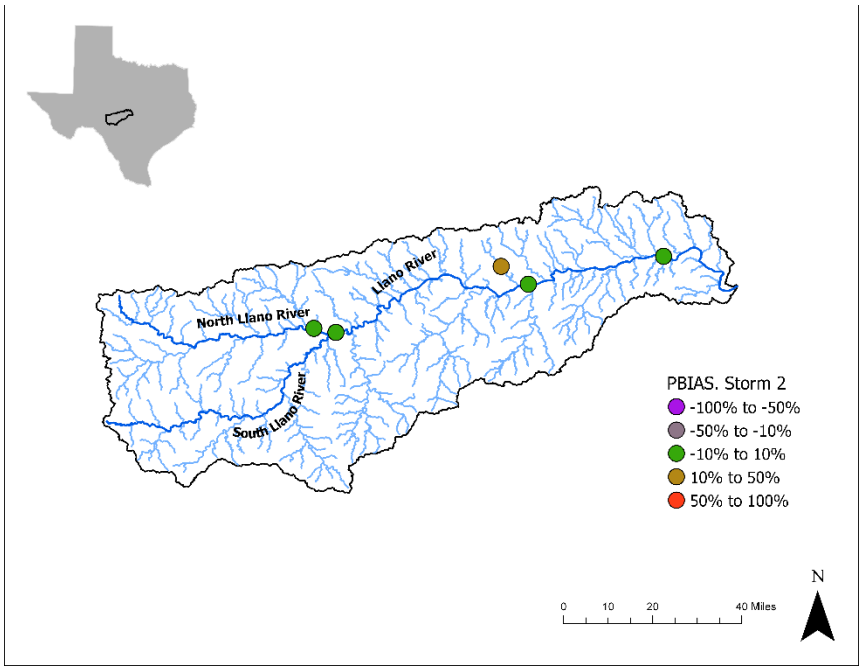

Figure 26: PBIAS in the Llano River basin between observed gage flow and NWM analysis assimilation forecast, second flood event (closer to zero is better).

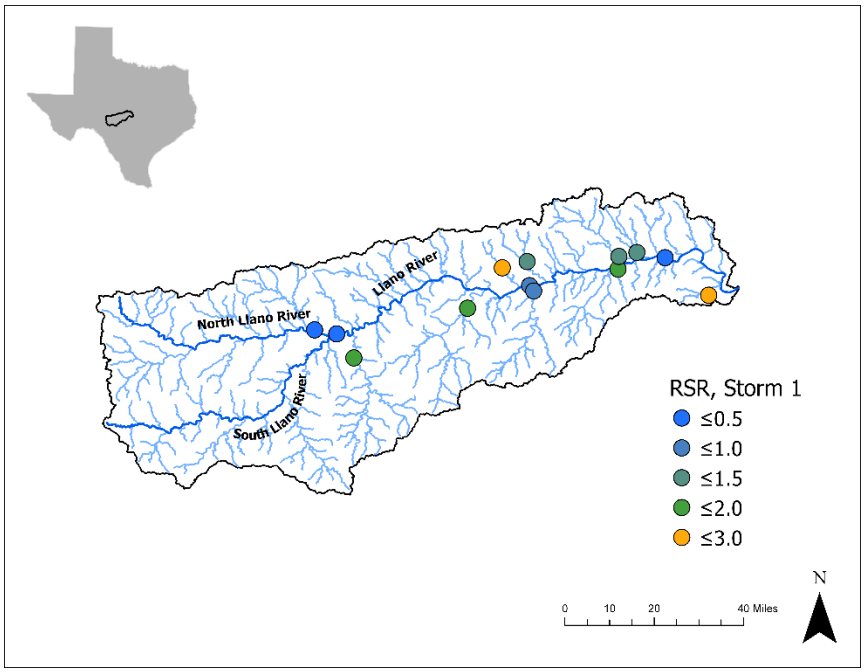

Figure 27: RSR in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the first flood event (closer to zero is better).

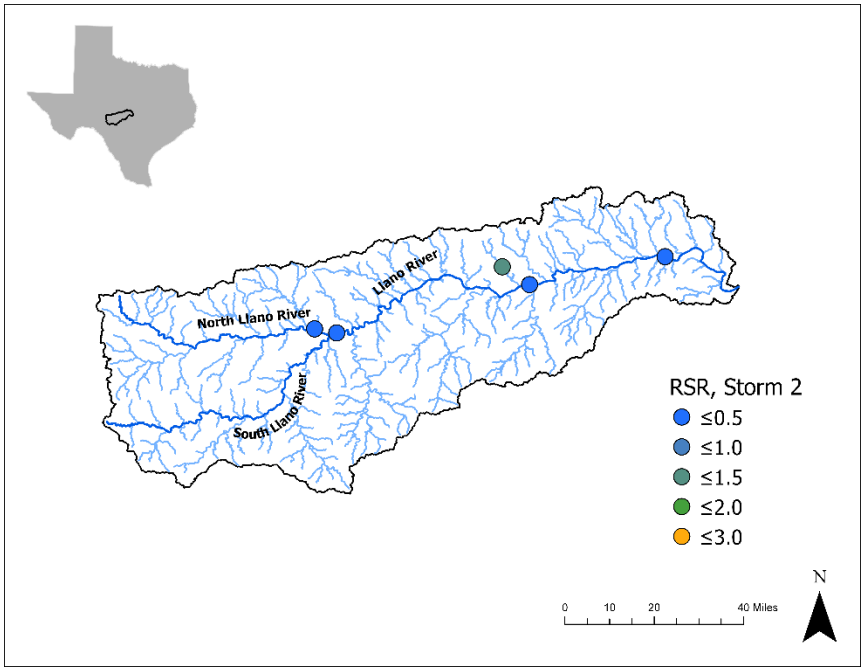

Figure 28: RSR in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the second flood event (closer to zero is better).

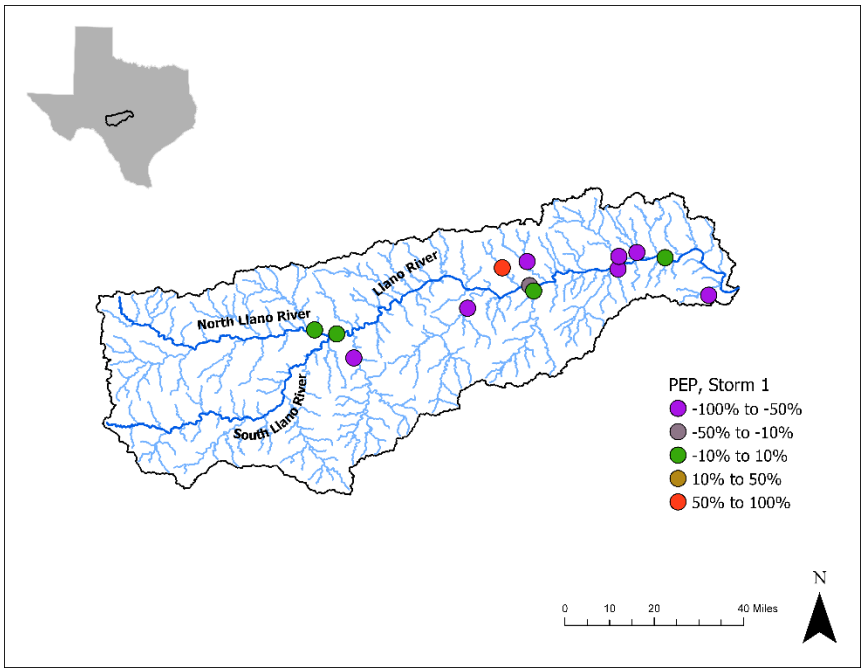

Figure 29: PEP in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the first flood event (closer to zero is better).

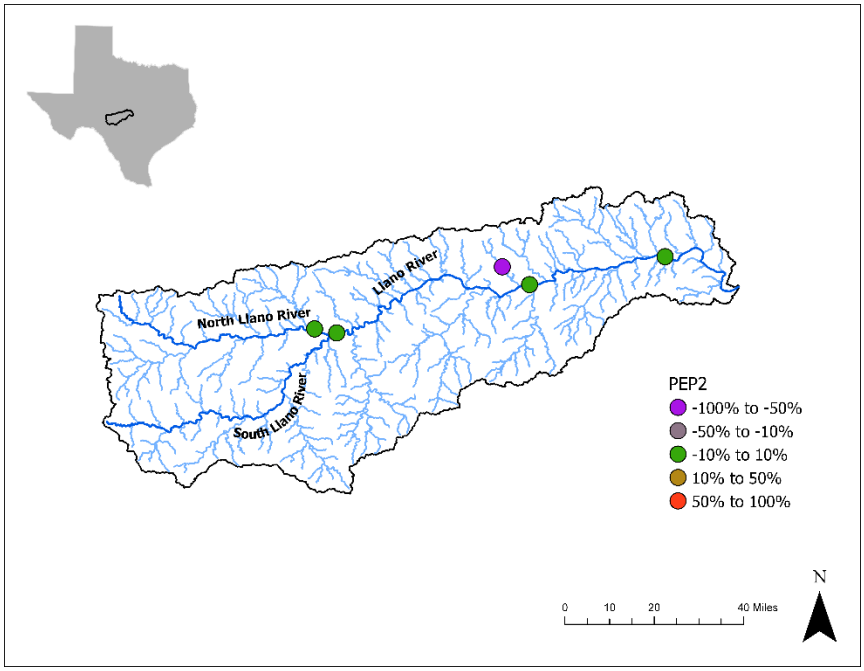

Figure 30: RSR in the Llano River basin between observed gage flow and NWM analysis assimilation forecast for the second flood event (closer to zero is better).

SHORT-TERM FORECAST PERFORMANCE

At the majority of the LCRA gages, the short-term forecasts mirror the poor performance of the analysis and assimilation forecast and do not warrant further discussion beyond what is presented above. However, the behavior of the short-range forecast at the USGS gage sites bears further investigation as several distinct patterns arise. Note that the short-term forecasts can be tracked through time based on the color of the lines, which transitions from orange to yellow to green to blue to purple corresponding to earlier and later forecasts, respectively. Additionally, the analysis will proceed from upstream to downstream gages, starting with the three gages near Junction, TX, followed by the two gages near Mason, TX and finally with the gage at Llano, TX. Each gage saw elevated flow rates for both flooding events except for the Beaver Creek gage.

The first flooding event was relatively minor at the North Llano River gage even though many of the short-term forecasts predicted much higher flow rates. This difference is likely attributed to changes in the rainfall forecast where the actual rainfall occurring primarily over the South Llano River basin, as seen in the extremely high flows at Llano River at Junction gage, which is just downstream of the confluence of the North and South forks. Unfortunately, the South Llano River USGS gage failed during both flooding events (though we can infer that the flows were significant based on the difference between the Llano River at Junction and the North Llano River observed streamflow values). However, the predicted streamflow at the South Llano River gage is significantly lower, again indicating that the actual rainfall did not coincide with the predictions. The first flooding event really was two overlapping events and the second half shows consistent overprediction at both the North and South Llano gages, which is then carried over to the Llano at Junction gage.

The second flooding event shows very different behavior in the NWM short-range forecasts. At the North Llano River gage there is an unusual striped appearance to the short-range forecasts indicating that the model did not predict any increases in flow over the time period in question. This result is likely due to the opposite causes as in the first flood, where rain fell in areas where it was not predicted, and the model could not compensate even given the updates from the analysis assimilation step. Again we are faced with missing data from the south fork, but can still infer, based on measured downstream flows, that the streamflow was underpredicted significantly. This assessment is borne out by noting that the short-range predictions at the Junction gage are well below observed flow rates and that only after assimilation of the rising edge of the flood wave do the short-term forecasts begin to predict the falling edge well.

The NWM short-term predictions show good agreement with observed flows at the Mason, TX USGS gage during the first flooding event. The predicted peak of the flood wave is later in time than the observed peak, indicating that the flood wave velocity was modeled to be slower than observed, but the overall intensity was preserved from the upstream assimilated streamflow values. The second flood wave is predicted much more poorly. Notably, the shape of the flood differs significantly from the second flood peak at the Junction, TX gage, with the initial peak at Mason occurring earlier in time than the peak upstream, with the secondary peak occurring several hours after the observed upstream peak, with this peak being somewhat better predicted by the short-term forecasts. Precipitation likely plays a role in this observed behavior as well. Because we see the initial peak flow before the flood wave could have propagated from upstream at Junction, we can infer that portion of the flow is from excess runoff between the two gages and, since the peak is not predicted, that rain fell in the areas between the gages where it was not predicted.

Finally, the gage at Llano, TX shows the performance of the NWM short-range at the downstream end of the Llano River basin. In both flood events, we see a delayed forecast of the peak floods relative to the observed streamflow values; however, the shape of the predicted flood wave matches well in both cases. Here again there is likely a routing velocity underestimation in the model. Again, we see that the propagation of the flood wave based on assimilated data upstream is relatively successful.

PSEUDO-ENSEMBLE ANALYSIS

To further examine the changes in the short-term forecast uncertainty over time, we can slice the data in a different fashion than previously described by taking advantage of the overlap between subsequent NWM short-range forecasts. If we consider two short-range forecasts initialized one hour apart, we note that 17 of the 18 streamflow predictions are made at overlapping times. In other words, there are two predictions for the streamflow at these 17 overlapping times. By extending this to a long series of short-range forecast, we end up with 18 predictions for each hour thereby generating a pseudo-ensemble forecast. From there, at each time step, the average and standard deviation are calculated, the results of which are shown in Figure 31 through Figure 47. In each figure, the average and standard deviation of the pseudo-ensemble at each time step are shown as orange and green dots, respectively, while the observed streamflow is shown as a black line in the left panel. The right panel shows the standard deviation versus the average streamflow color-coded to indicate the order of the observation with low time step values corresponding to earlier observations and high time step values corresponding to later observations. The combination of forecasts across lead-times is referred to as pseudo-ensemble because the starting point of each point in the analysis is different, with the

predictions associated with the longest lead times expected to be less accurate due to changes in both the quantity and accuracy of the forecasted precipitation over time.

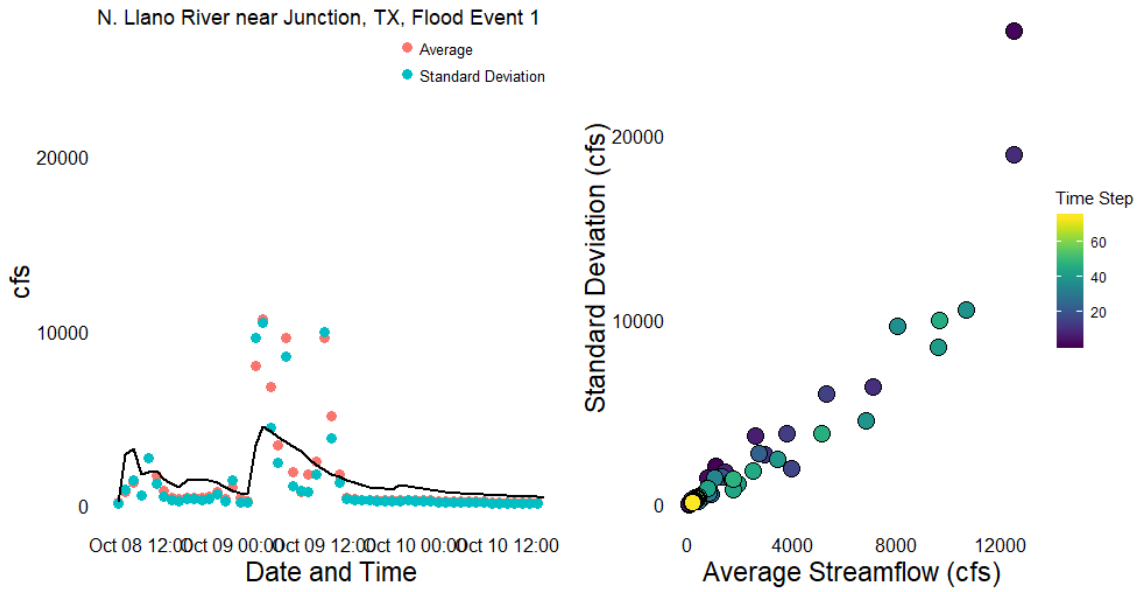

Figure 31: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the N. Llano River near Junction, TX USGS gage.

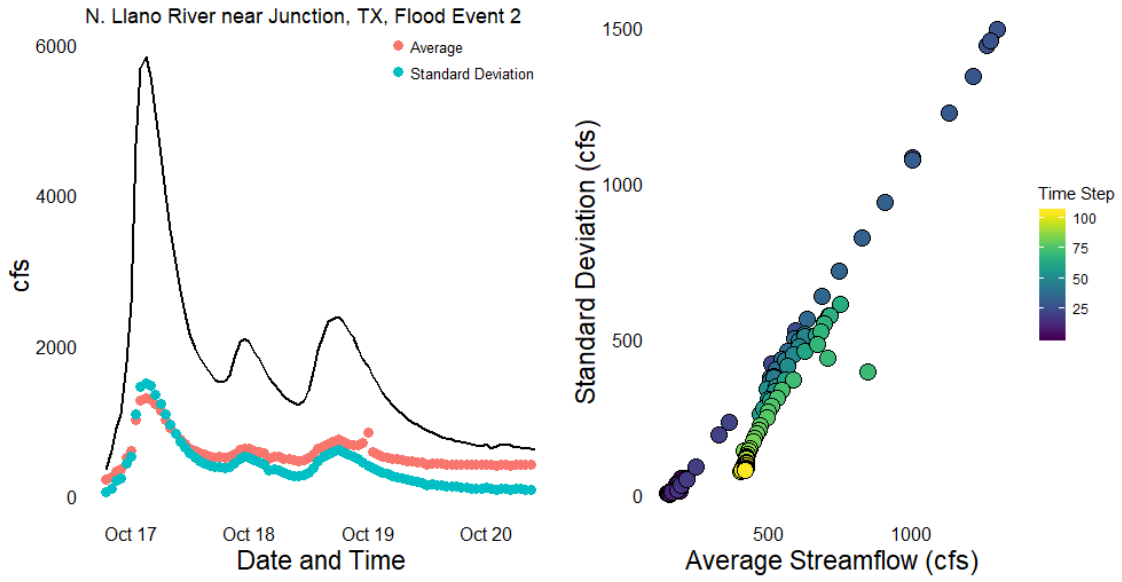

Figure 32: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the N. Llano River near Junction, TX USGS gage.

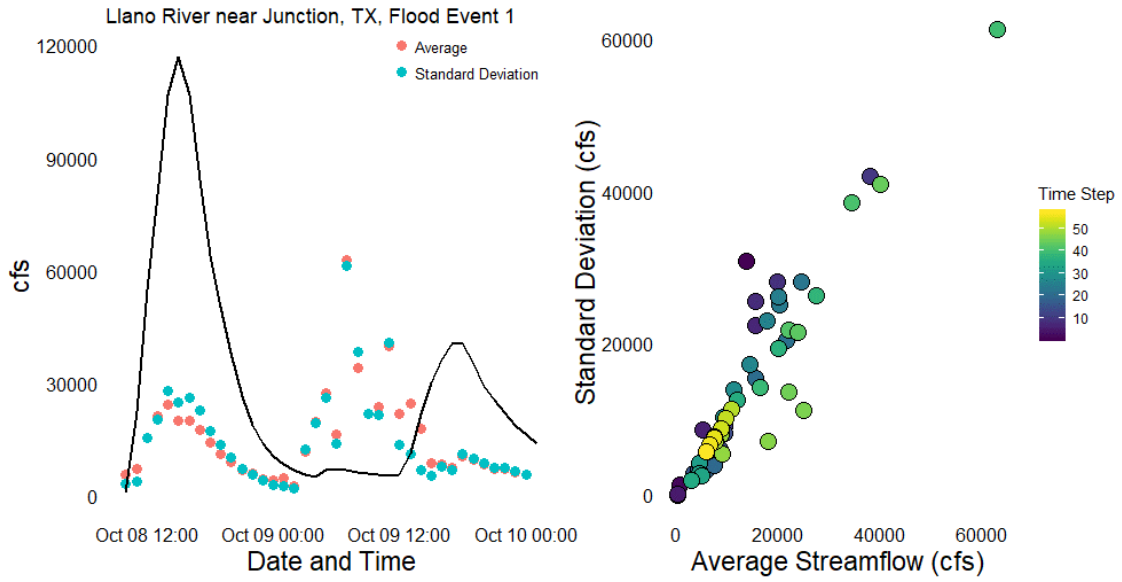

Figure 33: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Llano River near Junction, TX USGS gage.

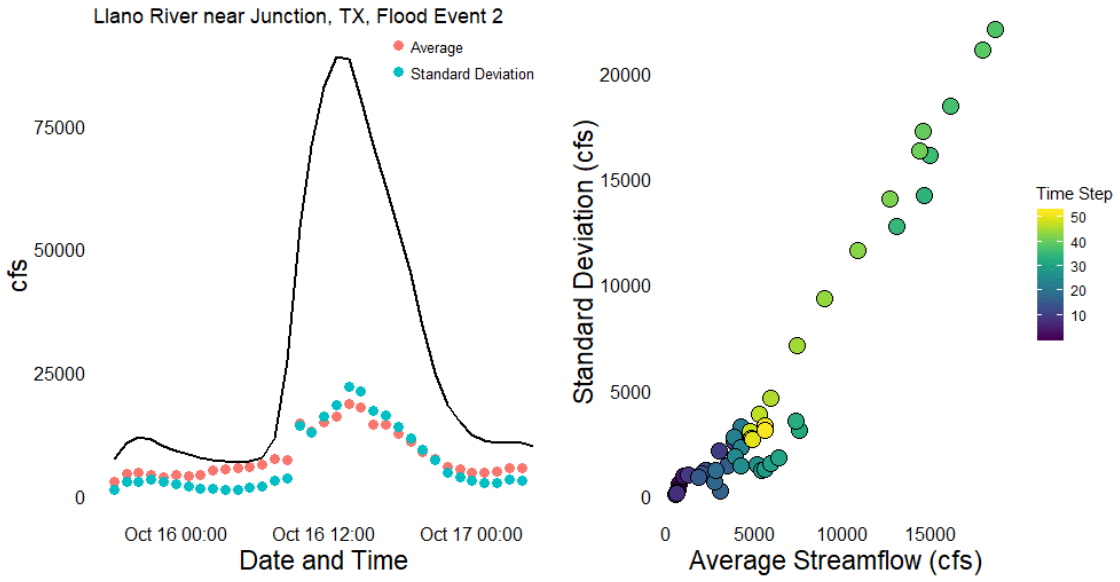

Figure 34: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Llano River near Junction, TX USGS gage.

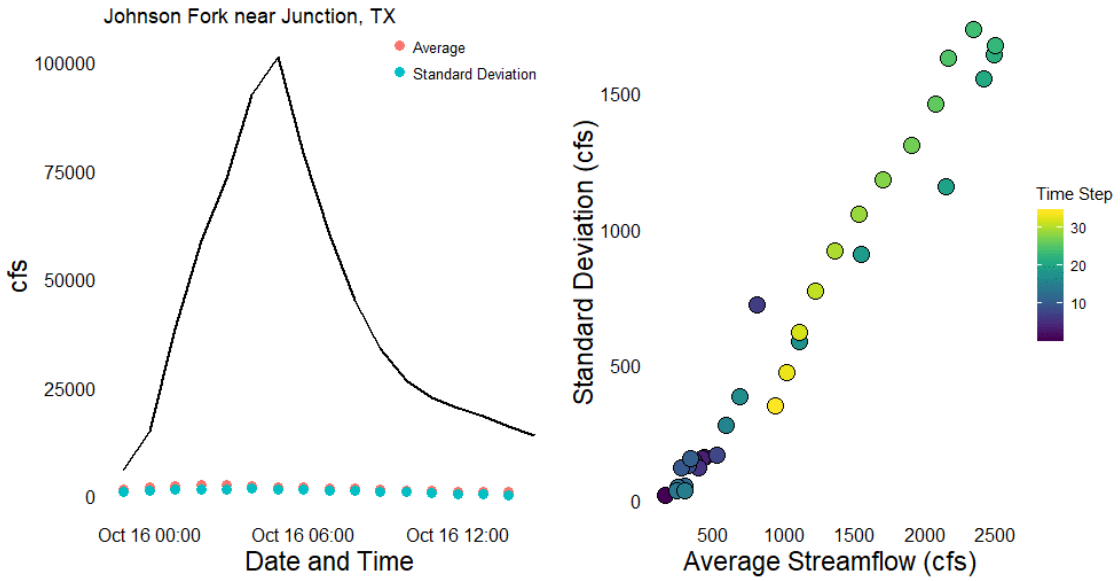

Figure 35: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Johnson Fork near Junction, TX LCRA gage.

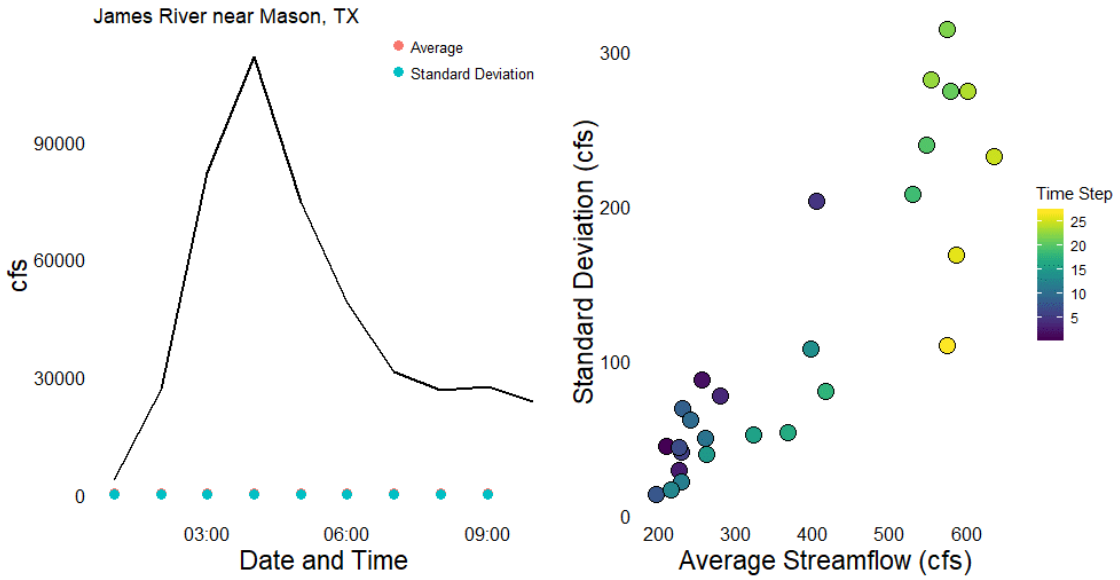

Figure 36: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the James River near Junction, TX LCRA gage.

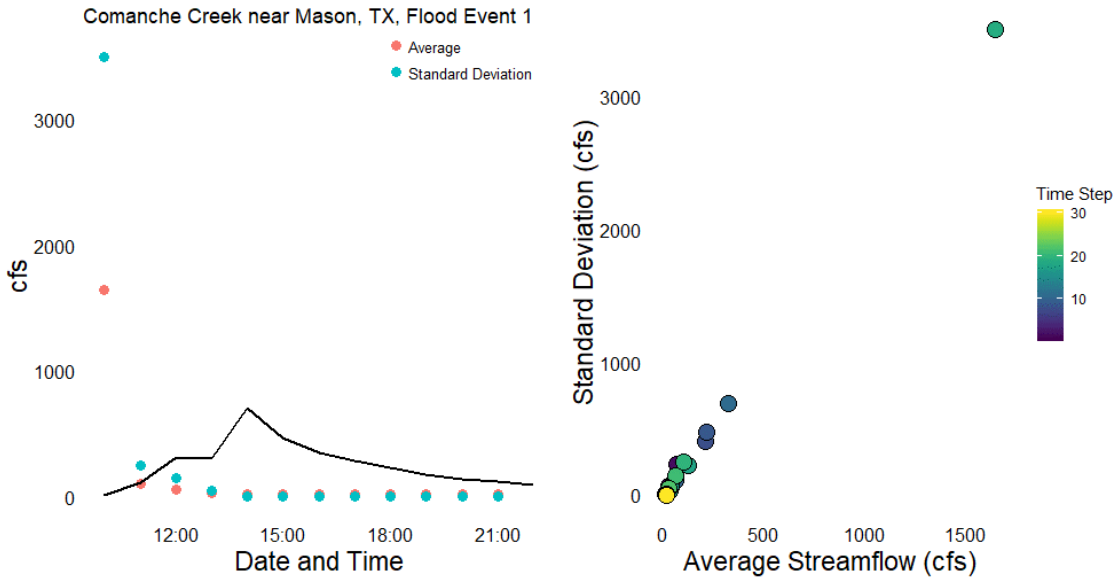

Figure 37: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Comanche Creek near Mason, TX LCRA gage.

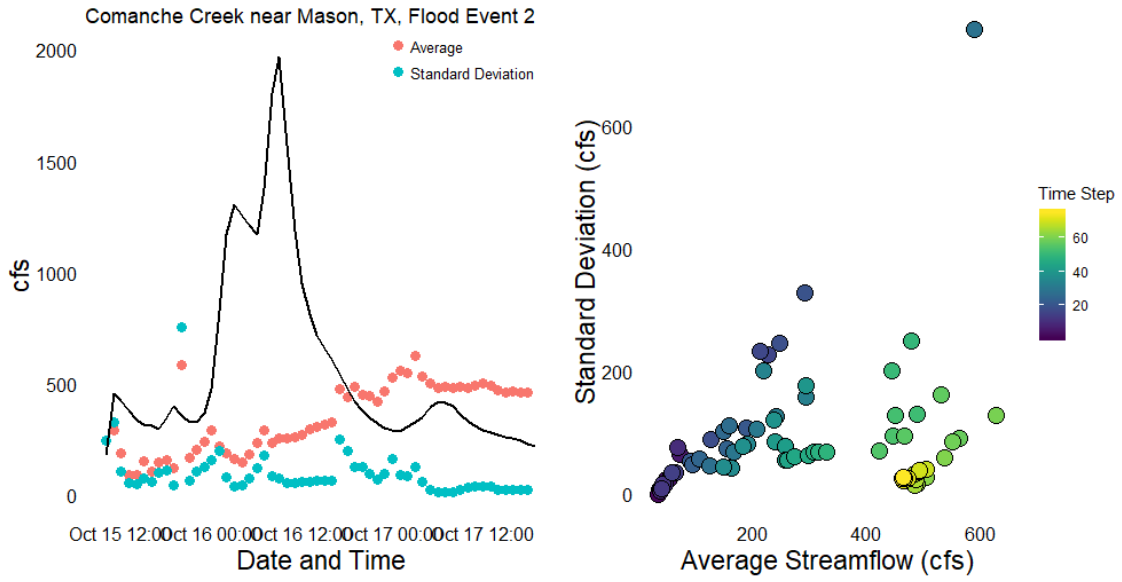

Figure 38: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Comanche Creek near Mason, TX LCRA gage.

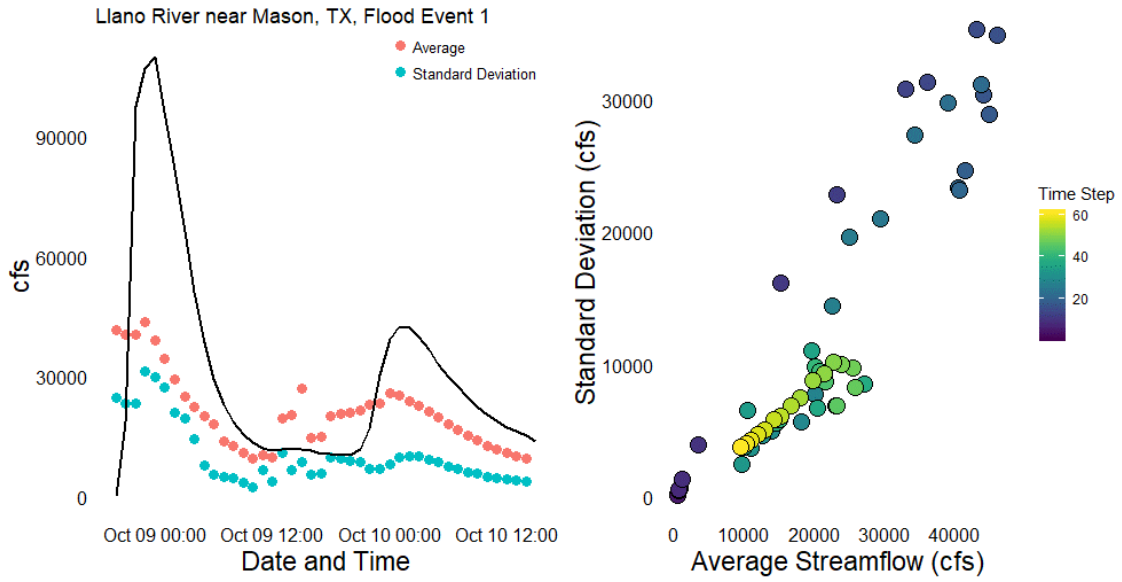

Figure 39: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Llano River near Mason, TX USGS gage.

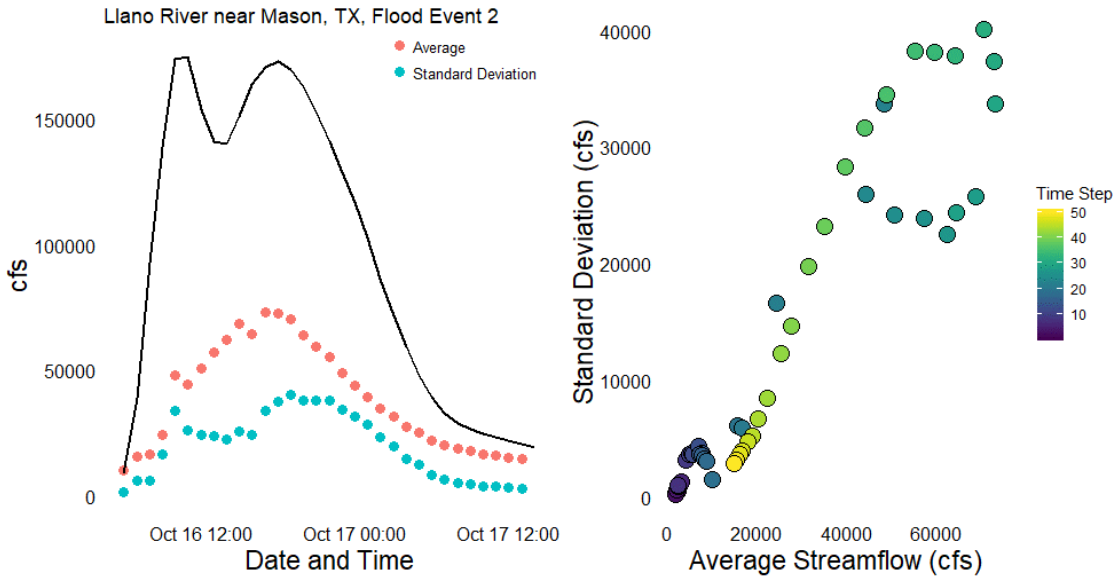

Figure 40: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Llano River near Mason, TX LCRA gage.

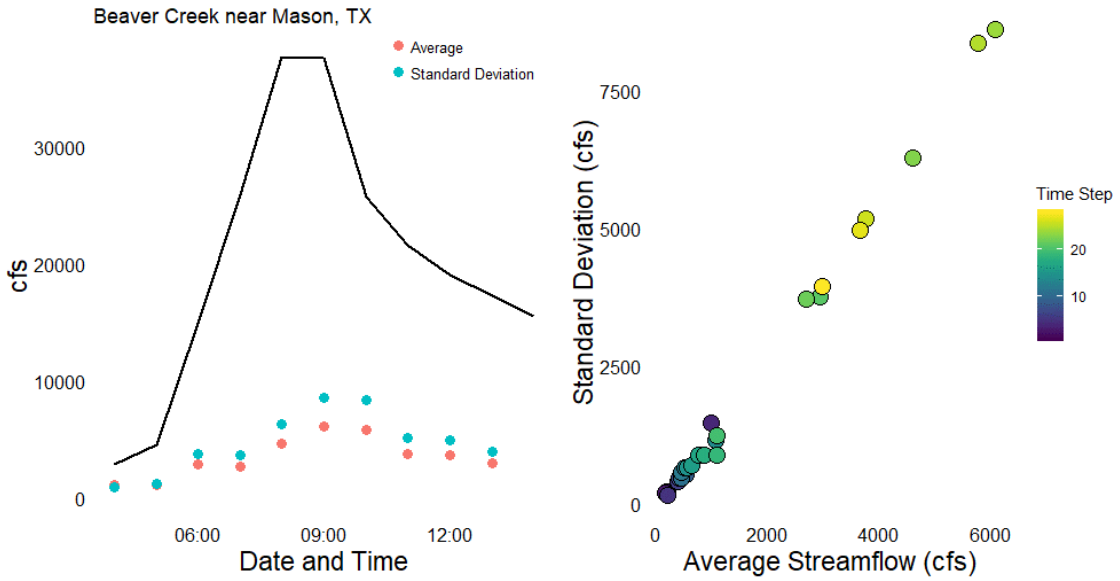

Figure 41: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Beaver Creek near Mason, TX USGS gage.

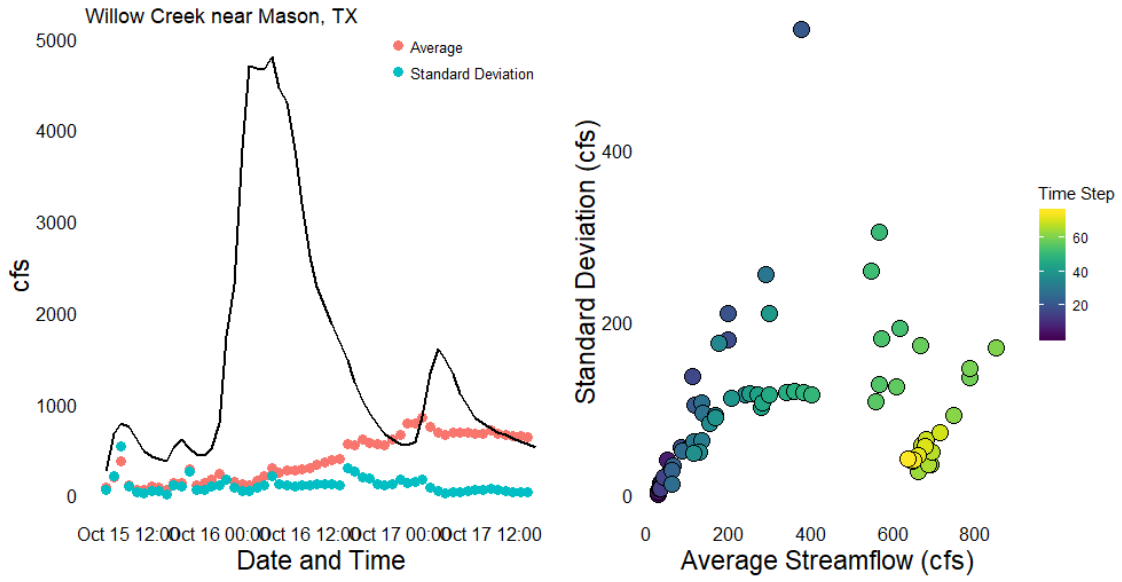

Figure 42: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Willow Creek near Mason, TX LCRA gage.

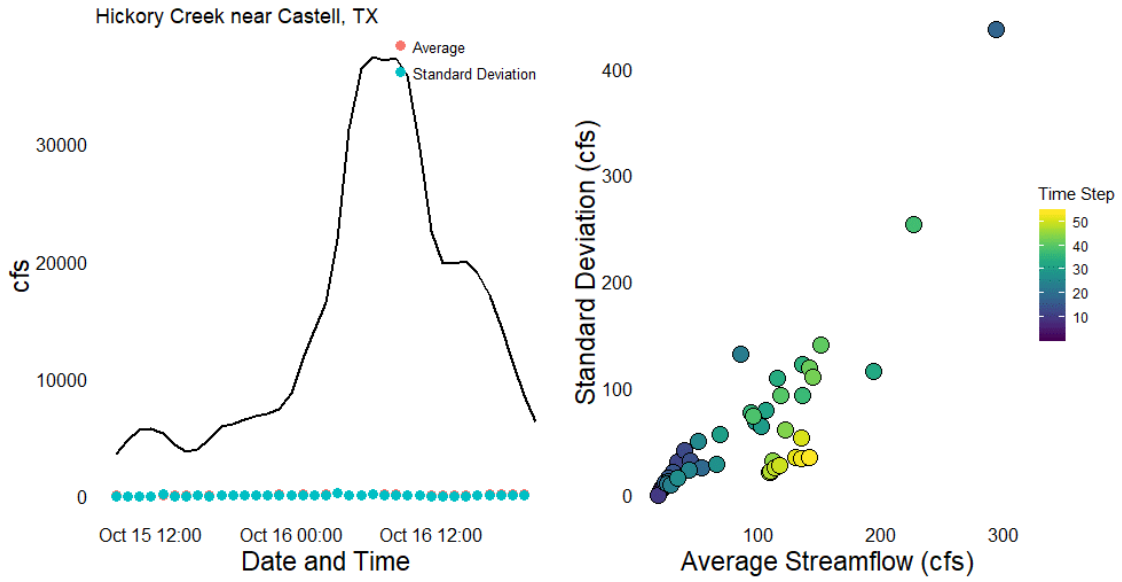

Figure 43: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Hickory Creek near Castell, TX LCRA gage.

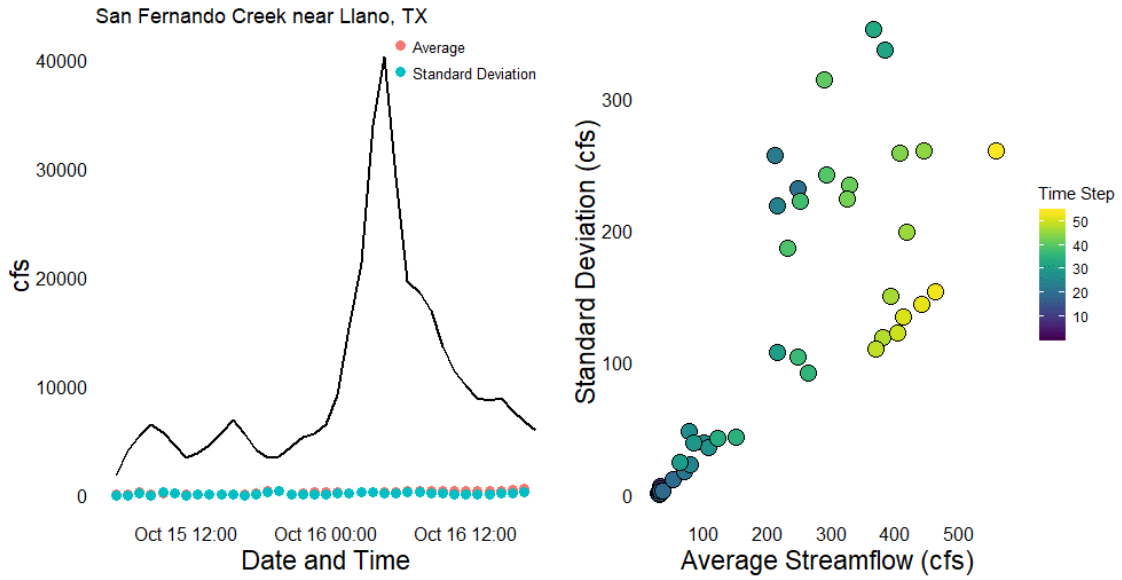

Figure 44: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the San Fernando Creek near Llano, TX LCRA gage.

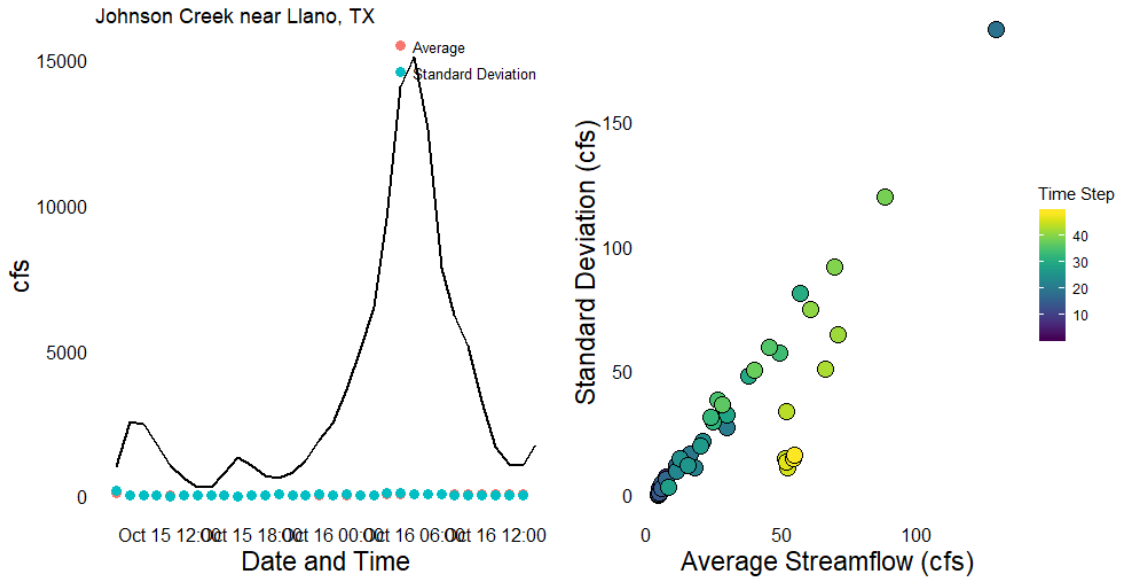

Figure 45: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) at the Johnson Creek near Llano, TX LCRA gage.

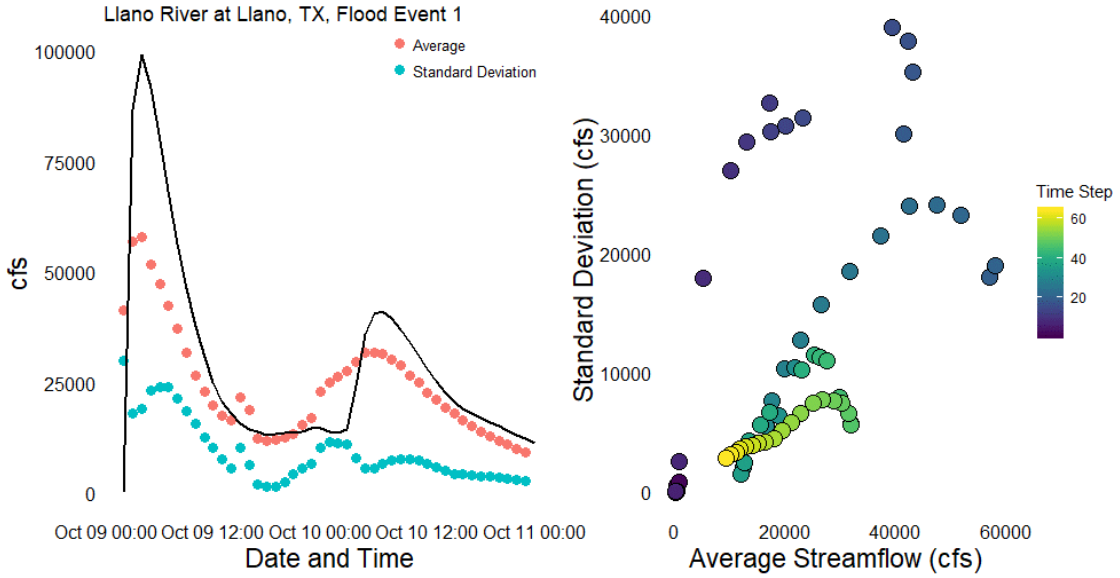

Figure 46: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event one at the Llano River at Llano, TX USGS gage.

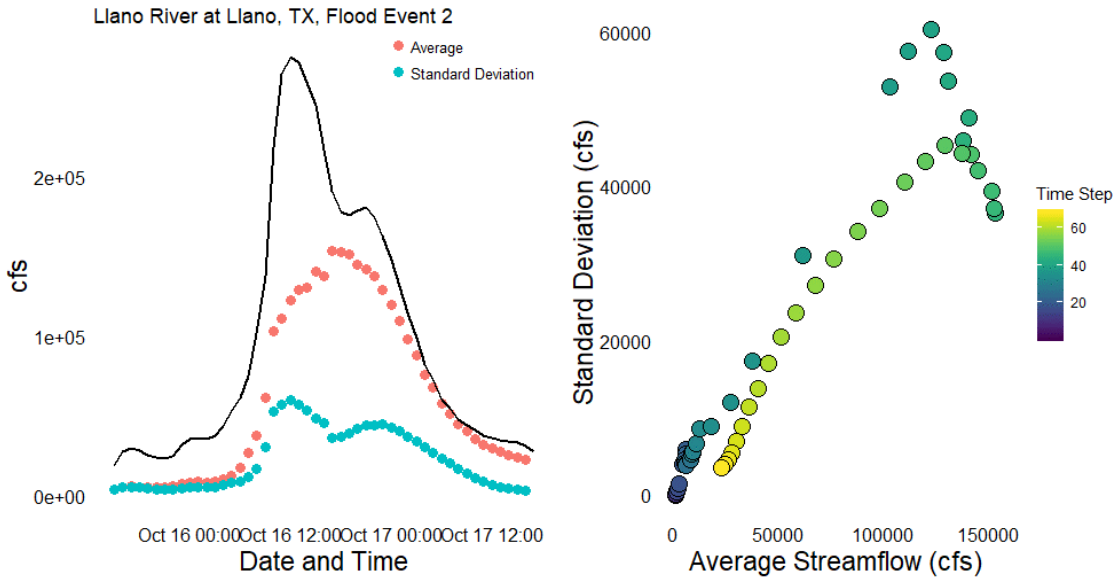

Figure 47: Average and standard deviation of the pseudo-ensemble NWM short-range forecasts with the observed gage flow (black line) through time (left) and standard deviation vs. average streamflow (right) for flood event two at the Llano River at Llano, TX USGS gage.

Again, and in similar fashion to the previous analyses, the LCRA gage sites are poorly predicted by the NWM short-range streamflow forecasts. For the USGS gages, this view neatly describes the average performance of the NWM short-range predictions through time. Looking at the N. Llano River near Junction, TX gage, we see that the average prediction is much lower than the observed flow, while the standard deviation is similar in magnitude to the average, indicating a high degree of uncertainty. This corresponds to the ever-decreasing pattern observed from the individual short-range forecasts and succinctly captures the poor overlap between subsequent forecasts. A similar pattern is observed at the Llano River at Junction, TX gage that is just downstream. Additionally, both of these gages show areas of high average and standard deviation, which correspond to long-lead time variation and likely to changes in the precipitation forecast.

At the Llano River at Llano, TX gage, we see the average predicted streamflow values aligned with the observed flows much more closely in magnitude and the standard deviation relative to the average values is noticeably reduced. Here we see increases in the standard deviation that correspond to the rising edge of the flood wave and decreasing standard deviation on the descending side. This result indicates that there is more uncertainty in the NWM short-range forecast for the rising edge of flood waves, which makes sense given changes in the timing and exact location of precipitation at different lead times. Again, the Llano River near Mason, TX gage site shows a mix of behavior relative to the upstream and downstream gages, with poorer performance during the leading section of the flood wave and better agreement in the latter section.

The right panel of the previous figures demonstrates a hysteresis effect, with the majority of flood peaks showing higher standard deviations for early observations (rising edge) than later observations (falling edge). This observation leads us to the conclusion

that the uncertainty is higher for the rising edge of flood waves, which is similar to what was observed in the left panel of the above figures. The increased uncertainty in the rising edge of the flood wave again is due to uncertainty in the precipitation and rainfall-runoff components of the model chain as well as the timing of the routing for the downstream gages. The falling edge is generally well modelled by the NWM, as see in Figures 3-22.

Chapter 5: Discussion and Conclusions

The behavior of the NWM analysis/assimilation and short-range NWM forecasts is analyzed in this work. This chapter will revisit the research questions posed in Chapter One, followed by final conclusions and recommendations for future work.

RESEARCH QUESTIONS

1. *How do the NWM analysis/assimilation and short-range streamflow predictions compare to observed streamflow at USGS gages sites?*

   Because the NWM uses the USGS gage site in the data assimilation step, we expected that the analysis/assimilation forecast would closely match the observed streamflow reading and this is indeed the case. In all cases but one, the analysis/assimilation prediction was very closely matched with the gage readings. During the first flood event at the Llano River near Mason, TX gage, the NWM underpredicted by about 14%. Notably, this is not the case during the second flooding event, where streamflow values were underpredicted by only 3.1% which is counter to the presence of a systematic error.

   The performance of the short-term forecasts was mixed. In some cases, the model predicted decreasing streamflow at every time step (though with starting points that mirrored the observed flows). This result is problematic in that every prediction indicates that the highest flowrates have passed. In the context of flood warnings and response, this situation could result in no warning being issued and potentially the re-allocation of limited response resources to other areas. This trend is observed more in the upstream gages, which indicated error in the rainfall predictions (which are observed) and/or issues with the rainfall-runoff portion of the NWM.

In other cases, the model predicted the shape of the flood wave well, but the timing was delayed relative to observed streamflow values. This trend is most obvious at the Llano River at Llano, TX gage where peak flows are predicted five to six hours later than occurred. In this context of flood response, this is again problematic as warnings and response could be delayed relative to the actual arrival of the flood wave. This behavior is seen most clearly at the USGS gage farthest downstream on the Llano River which indicates that the NWM is successfully assimilating and routing water from the upstream USGS gages.

Additionally, in some cases (e.g., the N. Llano River near Junction, TX flood event 1) there are large over-predictions at longer lead times that are generally reduced at shorter lead times. This result is clearly attributed to changes in the precipitation forecast and indicates the model is responding well to those changes.

2. *How do the NWM analysis/assimilation and short-range streamflow predictions compare to the observed streamflow at non-USGS gage sites?*

The performance of the NWM when compared to the LCRA (non-USGS) gage sites in the Llano River basin for the two flooding events of October 2018 is very poor. This observation holds true for both the analysis/assimilation and short-range forecasts. Because the LCRA gages are not part of the data assimilation step, there is no way for the model to make corrections from previous states relative to the actual streamflow values. Additionally, each of the LCRA gages is positioned on a tributary of the Llano River and thus none have an upstream USGS gage. Given that the NWM has been shown to successfully route flood waves in the Llano River basin, it is possible that an upstream USGS would improve the performance at the LCRA sites significantly.

3. *Is there a significant difference in the quality of the prediction made between USGS and non-USGS gage sites?*

There is a significant difference in the quality of the analysis/assimilation NWM forecasts between USGS and LCRA gage sites. Though the performance of the short-range NWM forecast varies between USGS gages, the overall magnitude was comparable to the actual gage readings while the same cannot be said at the LCRA gage sites.

4. *How does the magnitude of the error in NWM streamflow predictions change through time and space?*

The USGS sites exhibit different behavior which depends on the relative position of the gage in the watershed. The NWM predictions at the upstream gages show the highest variability, with several large spikes at long lead-times, and the worst match in the shape of the flood wave. The NWM predictions at the downstream gage show the best agreement in flood wave shape but with a loss of temporal accuracy. The NWM predictions at the mid-river USGS gage shows a mixed behavior; flood event number one more closely resembles the behavior of the upstream gages, while flood event number two resembles the behavior of the downstream gage. Hysteresis is observed at the majority of gage sites, with higher uncertainty associated with the rising edge of the flood wave.

DISCUSSION

At USGS gages, the performance of the NWM short-range forecasts varies with the location of the gage relative to the outlet of the watershed in the Llano River Basin, with the upstream gages showing poorer performance and the downstream gage showing good performance when considering the shape of the flood wave and moderate

performance when considering the timing of peak flows. The NWM did not model the streamflow at the LCRA gage sites well where, in most cases, there was little indication of the flood in the NWM output. It is important to note that these results only consider two extreme rainfall events and future performance in not guaranteed to replicate this result. In fact, we expect the predictive quality of the NWM to improve with continued refinement of the model itself and improvements to the meteorological forcing.

The performance of the model at UGSG gaging sites implies differing performance in the various modules of the NWM. The upstream gage sites did not indicate the presence of a flood wave, instead predicting decreasing flow at most time steps. Though the starting point of each short-range forecast was highly accurate because of the assimilation step, the resulting forecasts were not. The flow in this portion is likely to be more heavily influenced by the rainfall-runoff portion of the model and, importantly, there are no upstream USGS gages. This means there is no possibility of correction to inflow errors in the model upstream and therefore, though the routing appears to be working well based on the downstream gages, the amount of water in the channel is underestimated.

The mid-river USGS gage showed mixed success of the NWM short-range forecasts. There is clear evidence of peak propagation from the upstream USGS gage, but, in the case of the first flood event, there is also peak in the observed streamflow that is prior to the peak at the upstream gage and that is not well modelled by the NWM short-range forecasts.

The Llano River basin is within an area of central Texas known as “flash flood alley” which is characterized by short reaction times and rapid flood wave propagation. These characteristics may explain the inferred low performance of the rainfall-runoff portion of the NWM given that prediction of flash floods requires high spatial and

temporal resolution of input data (J. C. Schaake et al., 2007). The one-hour timestep of the NWM may be too long to adequate capture the behavior of the catchments in the Llano River basin and the 1 km^2 resolution of the precipitation forecasting may be too granular.

FUTURE WORK

A logical extension of this study is to incorporate either individual short-range forecasts, or the maximum standard deviation at each site, into the HAND method to generate a range of flood inundation maps. Applying the maximum standard deviation (or some multiple based on a desired confidence level) to predicted flow maxima would provide high and low estimates to predicted flow peaks and potential upper and lower bounds to predicted inundation extent. Floods exist in the upper regions of the rating curve, where the slope stage height to streamflow is quite flat which will reduce the relative uncertainty in stage height as compared to the streamflow.

Second, if possible, the NWM should be obtained for a portion of the US that is small enough to be run on a local machine. With access to the model the full suite of uncertainty analysis tools become available and the relative contribution of each step in the chain of models can be assessed. As a corollary, since it is likely the precipitation forecast contributes a large portion of the overall uncertainty, running the model on an ensemble of precipitation forecasts is also recommended.

If running the NWM water model locally proves unfeasible, additional data assimilation can be considered. Because the NWM only incorporates USGS gage data in the assimilation step, post processing of the forecasts with non-USGS gage data (such as the LCRA gages in this case) may improve the forecast in the area of the alternate gages.

This could be especially appealing to local entities that wish to incorporate NWM predictions in to their decision making.

Appendix A

Python script to subset NWM forecasts from Hydroshare

```
# coding: utf-8
"""Downloads subsetted NWM files from HydroShare."""

from datetime import date, datetime
from datetime import timedelta
import json
import os
import requests
from shutil import move
import stat
import tempfile
import time
import warnings
from zipfile import ZipFile

from dateutil import parser as p


def download_zip(download_url):
    """Downloads a zip file to a temporary directory.

    Args:
        download_url: URL of zip file.

    Returns:
        Path to downloaded zip file.
    """
    resp = requests.get(download_url, verify=False)
    if resp.status_code == 200:
        zip_filename = resp.headers['Content-Disposition'].split(';')[1].split('=')[1].replace('"', '')
        zip_path = os.path.join(tempfile.gettempdir(), zip_filename)
        with open(zip_path, 'wb') as f:
            for chunk in resp.iter_content(chunk_size=1024):
                f.write(chunk)
    else:
        zip_path = None
    return zip_path


def get_subset(geojson_file, nwm_config, file_type, start_date,
               forecast_hour,end_date=None,  output_file=None):
    """Downloads subsetted netCDF file and returns output filename.

    Requests HydroShare to subset National Water Model (NWM) results for a
    given area, and downloads the result. Only one study area polygon with a
    single part is processed by HydroShare. For example, if you had a multipart
    polygon like Hawaii, only the first island in the GeoJSON file would be
    processed.

    Args:
        geojson_file: GeoJSON file with your study area polygon.
        nwm_config: 'analysis_assim', 'short_range', or 'medium_range'.
        file_type: 'channel_rt', 'land', or 'forcing'.
        start_date: Forecast date, or start date of analysis_assim files to
            subset, in yyyy-mm-dd format.
        end_date: End date (inclusive) of data to retrieve in yyyy-mm-dd
            format. Valid for analysis_assim configuration only.
        forecast_hour: Optional forecast hour of model simulation. If not
            provided, '06' is used since that is common to all NWM
            configurations.
        output_file: Optional output filename. If not provided, filename as
            downloaded from HydroShare is used.
```

```
Returns:
    Path to downloaded file.
"""

# Validation
with open(geojson_file) as f:
    geojson = json.load(f)
geojson_polygon_str_4326 = json.dumps(geojson['features'][0]['geometry'])

allowable_configs = ['analysis_assim', 'short_range', 'medium_range']
if nwm_config not in allowable_configs:
    configs = ', '.join(allowable_configs)
    raise ValueError('NWM config must be one of ' + configs)

allowable_types = ['channel_rt', 'land', 'forcing']
if file_type not in allowable_types:
    types = ', '.join(allowable_types)
    raise ValueError('NWM file type must be one of ' + types)

if start_date != 'latest':
    start_date = p.parse(str(start_date)).strftime('%Y-%m-%d')

if end_date is None:
    end_date = date.today().strftime('%Y-%m-%d')

try:
    hour = int(forecast_hour)
    if hour < 0 or hour > 23:
        raise ValueError('Forecast hour must be 0 to 23')
    hour = str(hour).zfill(2)
except:
    raise ValueError('Forecast hour must be 0 to 23')

# Submit the job and get job_id

server = 'https://hs-apps.hydroshare.org/apps/nwm-forecasts/api/'
data = {
        'subset_parameter': {
        'config': nwm_config,
        'startDate': start_date,
        'endDate': end_date,
        'geom': file_type,
        'time': hour,
        'merge': True
    },
    'watershed_epsg': 4326,
    'watershed_geometry': geojson_polygon_str_4326
}

print('Submitting job')
resp = requests.post(server + 'submit-subsetting-job/',
                     headers = {'Authorization': 'Token [redacted]'},
                     data=json.dumps(data),
                     verify=False)
print resp
resp_json_obj = json.loads(resp.content.decode('latin1'))
print resp_json_obj
job_id = resp_json_obj['job_id']
print('Job id: ' + job_id)
```

```
    # Check job status
    retry_interval_seconds = 3
    max_tries = 200
    check_url = server + 'check-subsetting-job-status/?job_id=' + job_id
    print('URL to manually check status:\n' + check_url)

    retry_counter = 0
    job_done = False
    while not job_done:
        retry_counter += 1
        resp_check_status = requests.get(check_url, verify=False)
        print resp_check_status
        job_status = json.loads(resp_check_status.content.decode('latin1'))['status']
        print('Job status {0}: {1}'.format(retry_counter, job_status))
        if job_status.lower() == 'success':
            job_done = True
            break
        elif job_status.lower() == 'failure':
            raise Exception('Job failed to complete')  # todo: traceback?
        elif retry_counter > max_tries:
            msg = 'Max retry reached. Please check status manually.'
            raise Exception(msg)
        time.sleep(retry_interval_seconds)

    # Download result
    if job_done:
        download_url = server + 'download-subsetting-results/?job_id=' + job_id
        print('URL to manually download result:\n' + download_url)
        print('Downloading result')
        zip_file = download_zip(download_url)
        if zip_file is None:
            raise Exception('Failed to subset watershed')
        print('Zip downloaded to ' + zip_file)

        nc = extract_nc(zip_file, output_file)
        if nc is not None:
            print('Result saved to ' + nc)
            os.chmod(zip_file, stat.S_IWRITE)  # HydroShare gives us read-only file
            os.remove(zip_file)
        else:
            raise Exception('NetCDF file not found in downloaded zip file')
    return output_file


def extract_nc(zip_file, output_file=None):
    tempdir = tempfile.gettempdir()
    with ZipFile(zip_file) as z:
        names = z.namelist()
        for filename in names:
            if filename.endswith('.nc'):
                if output_file is None:
                    output_file = os.path.basename(filename)
                extracted = z.extract(filename, tempdir)
                move(extracted, output_file)
                return output_file
    return None
```

```
if __name__ == '__main__':


    # SET THESE VARIABLES TO MATCH YOUR DATA AND DESIRED DOWNLOAD PARAMETERS.
  geojson_file = 'test.json'  # GeoJSON file with study area polygon
  nwm_config = 'short_range'  # analysis_assim, short_range, or medium_range
#  nwm_config = 'analysis_assim'
  datestoget = [date.today()-timedelta(days=34)+n*timedelta(days=1) for n in range(34)]
#  datestoget = 'a'
  for i, Date in enumerate(datestoget):
    start_date = str(Date)  # yyyy-mm-dd format. Can use 'latest' for short or medium range.
#    start_date = '2018-03-07'
    end_date = '2018-05-06'  # yyyy-mm-dd format. Only used for analysis_assim. Can use 'latest'

    # To see what dates are available in HydroShare for subsetting, see
    # "Available Data for Subsetting" near the end of this page:
    # https://appsdev.hydroshare.org/apps/nwm-forecasts/api-page/
    for hour in range(24):
#    for hour in range(6,7):
       file_types = 'channel_rt'
       output_files = 'nwmfiles/'+datetime.strftime(Date,'%y_%m_%d')+'_'+str(hour).zfill(2)+'.nc'
#        output_files = 'nwmfiles/analysis_assim/Onion171226_180506.nc'
       forecast_hour = str(hour)
#    for i, filename in enumerate(output_files):
       print('Getting {0} file'.format(file_types))
       with warnings.catch_warnings():
           warnings.simplefilter('ignore')  # ignore HydroShare https warning
           get_subset(geojson_file, nwm_config, file_types, start_date,
                  forecast_hour,end_date, output_file = output_files)
       print('All files retrieved.')
```

Appendix B

R script for organizing individual NWM short-range forecasts based on a COMID list.

```
library(ncdf4)
library(lubridate)
library(zoo)
library(tidyverse)


filedirstr <- 'C:/Users/andre/Box Sync/NWM/ErrorEstimation/Texas/nwmfil
es/'
raw_files <- list.files(path = filedirstr)
file_ext <- raw_files[1:3]


llano_COMIDS <- c(5761759, 5765175, 5772343, 5770577, 5770009, 5769643,
5769807, 5770503, 5771211, 5771717, 5771703, 5771725, 5771417,5770545)

out <- tibble()

start <- Sys.time()
for (file_name in raw_files){
  file <- nc_open(paste0(filedirstr, file_name))

  featurids <- ncvar_get(file, varid = 'feature_id')
  comid_index <- sapply(llano_COMIDS, function(x) which(featurids == x)
)

  streamflow <-  ncvar_get(file, varid = 'streamflow')[comid_index,]

  times <- ncvar_get(file, varid = 'time')
  pred_times <- as.POSIXct(times*60, origin = '1970-01-01', tz = 'UTC')

  flow_tib <- as.tibble(streamflow)
  colnames(flow_tib) <- pred_times
  flow_tib %>%
    mutate(COMID = llano_COMIDS) %>%
    gather(key = pred_time, value = flow, 1:18) %>%
    mutate(lead_time = rep(1:18, each = length(llano_COMIDS)))-> flow_t
ib
```

```
  out <- rbind(out, flow_tib)
  nc_close(file)
}
Sys.time() - start

write_csv(x = out, path = 'data/llanoFlood.csv')

#assim

nc <- nc_open("data/assim/test.nc")
featurids <- ncvar_get(nc, varid = 'feature_id')
comid_index <- sapply(llano_COMIDS, function(x) which(featurids == x))
streamflow <-  ncvar_get(nc, varid = 'streamflow')[comid_index,]
times <- ncvar_get(nc, varid = 'time')
pred_times <- as.POSIXct(times*60, origin = '1970-01-01', tz = 'UTC')
flow_tib <- as.tibble(t(streamflow))
colnames(flow_tib) <- llano_COMIDS
flow_tib %>% gather(key = COMID, value = Flow) %>%
  mutate(DateTime = rep(pred_times,length(llano_COMIDS))) -> flow_tib

write_csv(flow_tib, path = "data/assim_flow.csv")


nc_close(nc)
```

Appendix C

R script for data analysis and plotting.

```
library(ggplot2)
library(plotly)
library(reshape2)
library(lubridate)
library(magick)
library(xts)
library(viridis)
library(cowplot)
library(tidyverse)

flood_cut <- function(gage_data, start, end){

  flood_cut <- gage_data %>% filter(DateTime >= start & DateTime <= end
)
  index <- which(flood_cut$Flow>= max(flood_cut$Flow, na.rm = T)/10)
  only_flood <- flood_cut[(min(index)-1):(max(index)-1), ]

}

plot_flood <- function(gage_flood, assim_flood, short_range_flood, n =
NA){

  gage_flood$Flow[is.na(gage_flood$Flow)] <- 0

  if (is.numeric(n)){

    title <- paste0(gage_name,', Flood Event ',n)

  } else {

    title <- gage_name
  }

  r <- ggplot() + geom_area(data = gage_flood, aes(x = DateTime, y = Fl
ow), fill = 'lightblue') + geom_line(data = short_range_flood, aes(x =
pred_time, y = Flow, col = as.factor(DateTime)), show.legend = F) + ggt
itle(title) +
  geom_line(data = assim_flood, aes(x = DateTime, y = Flow)) +
   theme_minimal() + theme( panel.grid.major = element_blank(), panel.g
rid.minor = element_blank(), axis.text=element_text(size=12, color = 'b
```

```
lack'),axis.title = element_text(size = 17)) + xlab("Date and Time") +
ylab("Flow (cfs)")
  #print(r)

  for_avg_sd <- unique(gather(short_range_flood[,c(5,7,8)],key = 'stat'
, value = 'value', -DateTime ))
  t <- ggplot() + geom_point(data = for_avg_sd, aes(x = DateTime, y = v
alue, col = stat), size = 3) + geom_line(data = gage_flood, aes(x = Dat
eTime, y = Flow), size =1) + theme_minimal() + theme(legend.title=eleme
nt_blank(), panel.grid.major = element_blank(), panel.grid.minor = elem
ent_blank(), legend.position=c(0.8,.95), axis.text=element_text(size=12
, color = 'black'),axis.title = element_text(size = 17),legend.text = e
lement_text(size = 10)) + xlab("Date and Time") + ylab("cfs") + scale_c
olor_manual(labels = c("Average", "Standard Deviation"), values = c('#F
8766D','#00BFC4')) +ggtitle(title)+xlim(c(min(gage_flood$DateTime),max(
gage_flood$DateTime)))
  #print(t)

  for_sdvsavg <- unique(short_range_flood[,c(5,7,8)])
  for_sdvsavg$n <- 1:length(for_sdvsavg$avg)
  u <- ggplot() + geom_point(data = for_sdvsavg, aes(x = avg, y = sd, f
ill = n), colour="black",pch=21, size=5)  + scale_fill_viridis(name = "
Time Step")  +  theme_minimal() + theme(axis.text=element_text(size=10,
color = 'black'),axis.title = element_text(size = 17), panel.grid.major
= element_blank(), panel.grid.minor = element_blank()) + xlab('Average
Streamflow (cfs)') + ylab('Standard Deviation (cfs)')
  #print(u)
  v <- plot_grid(t,u)
  print(v)
}

flood_cor <- function(gage_data, short_range){

  dateTimes <- unique(gage_data$DateTime)

  cor <- NULL

  for (i in 1:(length(dateTimes)-18)){
    date <- dateTimes[i]
    forecast <- short_range %>% filter(DateTime == date)
    gage <- gage_data %>% filter(DateTime >= min(forecast$pred_time) &
DateTime <= max(forecast$pred_time))
    cor <- append(cor, cor(gage$Flow,forecast$Flow))

  }
```

```
  ret <- tibble(DateTime = head(dateTimes, -18), cor = cor)
  ret
}

NSE <- function(obs,pred){
  1-sum((obs$Flow-pred$Flow)^2)/sum((obs$Flow-mean(obs$Flow))^2)
}

PBIAS <- function(obs,pred){
  sum(obs$Flow-pred$Flow)*100/sum(obs$Flow)
}

RSR <- function(obs, pred){
  sqrt(sum((obs$Flow-pred$Flow)^2))/sqrt(sum((obs$Flow-mean(obs$Flow))^
2))
}

PEP <- function(obs,pred){
  (max(pred$Flow)-max(obs$Flow))/max(obs$Flow)*100

}

avg_sd <- function(pred) {

  res <- pred %>% group_by(DateTime) %>% mutate(avg = base::mean(Flow),
sd = stats::sd(Flow)) %>% ungroup()
  res <- res[,c(5,6,7)]

}

gage_info <- read_csv('data/SiteInfo.csv')[-2,]
short_range_all <- read_csv('data/llanoFlood.csv') %>%
  mutate(DateTime = pred_time - hours(lead_time), flow = 35.3146667*flo
w) %>%
  rename(Flow = flow) %>%
  mutate(DateTime = with_tz(DateTime, tzone = "US/Central"), pred_time
= with_tz(pred_time, tzone = "US/Central"))
assim_all <- read_csv("data/assim_flow.csv") %>% mutate(DateTime = with
_tz(DateTime, tzone = "US/Central"), Flow = Flow*35.3146667)
performance1 <- tibble(ShortName = gage_info$ShortName, NSE = NA, PBIAS
= NA, RSR = NA, PEP = NA)
performance2 <- tibble(ShortName = gage_info$ShortName, NSE = NA, PBIAS
= NA, RSR = NA, PEP = NA)

DoPlots <- TRUE
```

```r
for (i in 1:nrow(gage_info)) {

  comid <- gage_info[[i,1]]
  gage_file <- paste0('data/',gage_info[[i,2]])
  LCRA <- gage_info[[i,3]]
  gage_name <- gage_info[[i,4]]
  twofloods <- gage_info[[i,5]]
  assim <- assim_all %>% filter(COMID == comid)
  short_range <- short_range_all %>% filter(COMID == comid)

  if (LCRA) {

    gage_data <- read_csv(gage_file) %>%
      mutate(`Date - Time` = mdy_hm(`Date - Time`, tz = "US/Central"))
%>%
      rename(Flow = `Flow (cfs)`, DateTime = `Date - Time`) %>%
      mutate(DateTime = DateTime + 5*60) %>%
      filter(minute(DateTime) == 00) %>%
      arrange(., DateTime)

  } else {

    gage_data <- read_tsv(gage_file) %>%
      mutate(DateTime = mdy_hm(datetime,tz=Sys.timezone())) %>%
      select(., c(5,9))
    colnames(gage_data) <- c('Flow','DateTime')
    gage_data <- gage_data %>% filter(minute(DateTime) == 00)
    gage_data$Flow[gage_data$Flow == -9999] <- NaN

  }

  cor <- flood_cor(gage_data, short_range)
  avg_Sd <- avg_sd(short_range)
  short_range <- left_join(short_range, cor, by = "DateTime")
  short_range <- left_join(short_range, avg_Sd, by = "DateTime")


  if (twofloods){

    gage_flood1 <- flood_cut(gage_data, '2018-10-07', '2018-10-15')
    assim_flood1 <- assim %>%
      filter(DateTime >= min(gage_flood1$DateTime) & DateTime <= max(ga
ge_flood1$DateTime))
    short_range_flood1 <- short_range %>%
      filter(pred_time >= min(gage_flood1$DateTime) & pred_time <= max(
gage_flood1$DateTime))
```

```
    if(DoPlots){
      plot_flood(gage_flood1, assim_flood1, short_range_flood1, n = 1)
    }

    performance1[i,2:5] <- c(NSE(gage_flood1, assim_flood1), PBIAS(gage
_flood1, assim_flood1), RSR(gage_flood1, assim_flood1), PEP(gage_flood1
, assim_flood1))


    gage_flood2 <- flood_cut(gage_data, '2018-10-14', '2018-10-30')
    assim_flood2 <- assim %>%
      filter(DateTime >= min(gage_flood2$DateTime) & DateTime <= max(ga
ge_flood2$DateTime))
    short_range_flood2 <- short_range %>%
      filter(pred_time >= min(gage_flood2$DateTime) & pred_time <= max(
gage_flood2$DateTime))

    if(DoPlots){
      plot_flood(gage_flood2, assim_flood2, short_range_flood2, n = 2)
    }

    performance2[i,2:5] <- c(NSE(gage_flood2, assim_flood2), PBIAS(gage
_flood2, assim_flood2), RSR(gage_flood2, assim_flood2), PEP(gage_flood2
, assim_flood2))

  } else {

    gage_flood <- flood_cut(gage_data, '2018-10-15', '2018-10-30')
    assim_flood <- assim %>%
      filter(DateTime >= min(gage_flood$DateTime) & DateTime <= max(gag
e_flood$DateTime))
    short_range_flood <- short_range %>%
      filter(pred_time >= min(gage_flood$DateTime) & pred_time <= max(g
age_flood$DateTime))
    if (DoPlots){
      plot_flood(gage_flood, assim_flood, short_range_flood)
    }


    performance1[i,2:5] <- c(NSE(gage_flood, assim_flood), PBIAS(gage_f
lood, assim_flood), RSR(gage_flood, assim_flood), PEP(gage_flood, assim
_flood))

  }
```

```
}

to_plot <- gage_data %>% filter(DateTime >'2018-10-07')
plot <- ggplot(data = to_plot) + geom_line(aes(DateTime, Flow), size =
1.5) + theme_minimal() + theme(panel.grid.major = element_blank(), pane
l.grid.minor = element_blank(), axis.text=element_text(size=12, color =
'black'),axis.title = element_text(size = 17),legend.text = element_tex
t(size = 10)) + ylab('Flow (cfs)') + xlab('Date')
plot
```

Works Cited

Austin Water Statistics | AustinTexas.gov - The Official Website of the City of Austin [WWW Document], n.d. URL https://www.austintexas.gov/department/austin-water-utility-statistics (accessed 12.2.18).

Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G., Olson, J.B., James, E.P., Dowell, D.C., Grell, G.A., Lin, H., Peckham, S.E., Smith, T.L., Moninger, W.R., Kenyon, J.S., Manikin, G.S., 2016. A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. Monthly Weather Review 144, 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6, 279–298. https://doi.org/10.1002/hyp.3360060305

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. Journal of Hydrology 375, 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R.A., Feyen, L., 2018. Increased human and economic losses from river flooding with anthropogenic warming. Nature Climate Change 8, 781. https://doi.org/10.1038/s41558-018-0257-z

Freeze, R.A., Harlan, R.L., 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. Journal of Hydrology 9, 237–258. https://doi.org/10.1016/0022-1694(69)90020-1

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. Journal of Hydrologic Engineering 4, 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)

Haggett, C., 1998. An Integrated Approach to Flood Forecasting and Warning in England and Wales. Water and Environment Journal 12, 425–432. https://doi.org/10.1111/j.1747-6593.1998.tb00211.x

HydroShare Apps Portal - National Water Model Forecast Viewer [WWW Document], n.d. URL https://hs-apps.hydroshare.org/apps/nwm-forecasts/api-page/ (accessed 12.7.18).

Krzysztofowicz, R., Kelly, K.S., Long, D., 1994. Reliability of Flood Warning Systems. Journal of Water Resources Planning and Management 120, 906–926. https://doi.org/10.1061/(ASCE)0733-9496(1994)120:6(906)

Maidment, D.R., 2017. Conceptual Framework for the National Flood Interoperability Experiment. JAWRA Journal of the American Water Resources Association 53, 245–257. https://doi.org/10.1111/1752-1688.12474

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., Rea, A., 2018. NHDPlus Version2: User Guide.

Merwade, V., Olivera Francisco, Arabi Mazdak, Edleman Scott, 2008. Uncertainty in Flood Inundation Mapping: Current Issues and Future Directions. Journal of Hydrologic Engineering 13, 608–620. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:7(608)

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

National Academy of Sciences, 2009. Mapping the Zone: Improving Flood Map Accuracy. National Academies Press, Washington, D.C. https://doi.org/10.17226/12573

NOAA National Severe Storms Laboratory, 2015. NOAA National Severe Storms Laboratory MRMS Multiple Radar/Multiple Sensor.

NOAA Office for Coastal Management, n.d. Hurricane Costs [WWW Document]. URL https://coast.noaa.gov/states/fast-facts/hurricane-costs.html (accessed 12.2.18).

Nobre, A.D., Cuartas, L.A., Momo, M.R., Severo, D.L., Pinheiro, A., Nobre, C.A., 2016. HAND contour: a new proxy predictor of inundation extent. Hydrological Processes 30, 320–333. https://doi.org/10.1002/hyp.10581

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing 28, 323–344. https://doi.org/10.1016/S0734-189X(84)80011-0

Ocio, D., Le Vine, N., Westerberg, I., Pappenberger, F., Buytaert, W., 2017. The role of rating curve uncertainty in real-time flood forecasting: THE ROLE OF RCU IN FLOOD FORECASTING. Water Resources Research 53, 4197–4213. https://doi.org/10.1002/2016WR020225

Pappenberger, F., Cloke, H.L., Parker, D.J., Wetterhall, F., Richardson, D.S., Thielen, J., 2015. The monetary benefit of early flood warnings in Europe. Environmental Science & Policy 51, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016

Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.-B., Pfister, L., Fraipont, P., 2006. Influence of uncertain boundary conditions and model structure on flood inundation predictions. Advances in Water Resources 29, 1430–1449. https://doi.org/10.1016/j.advwatres.2005.11.012

Parker, D., Fordham, M., 1996. An evaluation of flood forecasting, warning and response systems in the European Union. Water Resources Management 10, 279–302. https://doi.org/10.1007/BF00508897

Passalacqua, P., Tarolli, P., Foufoula-Georgiou, E., 2010a. Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. Water Resources Research 46. https://doi.org/10.1029/2009WR008812

Passalacqua, P., Trung, T.D., Foufoula-Georgiou, E., Sapiro, G., Dietrich, W.E., 2010b. A geometric framework for channel network extraction from lidar: Nonlinear

diffusion and geodesic paths. Journal of Geophysical Research: Earth Surface 115. https://doi.org/10.1029/2009JF001254
Penning-Rowsell, E.C., Tunstall, S.M., Tapsell, S.M., Parker, D.J., 2000. The Benefits of Flood Warnings: Real But Elusive, and Politically Significant. Water and Environment Journal 14, 7–14. https://doi.org/10.1111/j.1747-6593.2000.tb00219.x
Renner, M., Werner, M.G.F., Rademacher, S., Sprokkereef, E., 2009. Verification of ensemble flow forecasts for the River Rhine. Journal of Hydrology 376, 463–475. https://doi.org/10.1016/j.jhydrol.2009.07.059
Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M.J., 2008. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment 112, 3469–3481. https://doi.org/10.1016/j.rse.2008.03.018
Schaake, J, Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., Seo, D.J., 2007. Ensemble forecasts 63.
Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., Seo, D.J., 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth System Sciences Discussions 4, 655–717. https://doi.org/10.5194/hessd-4-655-2007
Schaake, J.C., Hamill, T.M., Buizza, R., Clark, M., 2007. HEPEX: The Hydrological Ensemble Prediction Experiment. Bulletin of the American Meteorological Society 88, 1541–1548. https://doi.org/10.1175/BAMS-88-10-1541
Service, N.N.W., n.d. NOAA's National Weather Service, Hydrologic Information Center [WWW Document]. URL http://www.nws.noaa.gov/hic/ (accessed 12.2.18).
Staff, K., 2018. VIDEO: Bridge collapses from "catastrophic flooding" on Llano River in Texas [WWW Document]. KUTV. URL http://kutv.com/news/nation-world/video-catastrophic-flooding-on-llano-river-in-texas (accessed 12.2.18).
Tarboton, D.G., Bras, R.L., Rodriguez-Iturbe, I., 1991. On the extraction of channel networks from digital elevation data. Hydrological Processes 5, 81–100. https://doi.org/10.1002/hyp.3360050107
Todini, E., 2004. Role and treatment of uncertainty in real-time flood forecasting. Hydrological Processes 18, 2743–2746. https://doi.org/10.1002/hyp.5687
Tribune, T.T., Anchondo, C., 2018. Austin issues city-wide boil water notice; calls for action "to avoid running out of water" [WWW Document]. The Texas Tribune. URL https://www.texastribune.org/2018/10/22/austin-water-boil-water-notice-after-historic-flooding/ (accessed 12.2.18).
Werner, M.G.F., Schellekens, J., Kwadijk, J.C.J., 2005. Flood Early Warning Systems for Hydrological (Sub) Catchments, in: Anderson, M.G., McDonnell, J.J. (Eds.), Encyclopedia of Hydrological Sciences. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/0470848944.hsa022
Zarzar, C.M., Hosseiny, H., Siddique, R., Gomez, M., Smith, V., Mejia, A., Dyer, J., 2018. A Hydraulic MultiModel Ensemble Framework for Visualizing Flood

Inundation Uncertainty. JAWRA Journal of the American Water Resources Association 54, 807–819. https://doi.org/10.1111/1752-1688.12656
Zheng, X., Maidment, D., Liu, Y., Tarboton, D.G., Lin, P., 2016. From Forecast Hydrology to Real-Time Inundation Mapping at Continental Scale. AGU Fall Meeting Abstracts 42.
Zheng, X., Tarboton, D.G., Maidment, D.R., Liu, Y.Y., Passalacqua, P., 2018. River Channel Geometry and Rating Curve Estimation Using Height above the Nearest Drainage. JAWRA Journal of the American Water Resources Association 54, 785–806. https://doi.org/10.1111/1752-1688.12661
Zhu, Y., Toth, Z., Wobus, R., Richardson, D., Mylne, K., 2002. The Economic Value Of Ensemble-Based Weather Forecasts. Bulletin of the American Meteorological Society 83, 73–83. https://doi.org/10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2