2.12 3.11 4.5

Subtask 3.11: Recommendations for the Utilization of Artificial Intelligence and Machine Learning to inform Blueprint

North Carolina Flood Resiliency Blueprint

Prepared for the North Carolina Department of Environmental Quality by AECOM

November 2023

Table of Contents

Figures

Definitions

A comprehensive list of definitions applicable to multiple North Carolina Flood Resiliency Blueprint documents is provided in a separate document.

Acronyms

AI – Artificial Intelligence

API – Application Programming Interface

AR – Augmented Reality

AEP – Annual Exceedance Probability

BCA – Benefit Cost Analysis

EP – Exceedance Probability

FEMA – Federal Emergency Management Agency

FIS – Flood Insurance Study

GPS – Global Positioning System

HEC-RAS – Hydrologic Engineering Center's River Analysis System

IIJA – Infrastructure Investment and Jobs Act

IP – Internet Protocol

LiDAR – Light Detection and Ranging

LiMWA – Limit of Moderate Wave Action

IoT – Internet of Things

ML – Machine Learning

NCDOT – North Carolina Department of Transportation

RARR – Risk Assessment and Risk Reduction

ROI – Return on Investment

SFHA – Special Flood Hazard Area

SIFT – Sediment Fingerprinting Tool

USGS – United States Geological Survey

VR – Virtual Reality

WSEL – Water Surface Elevation

2D – Two-Dimensional

1 Introduction

1.1 Purpose

Purpose: Subtask 3.11- Recommendations on the utilization of Artificial Intelligence (AI)/Machine learning (ML) to inform the development and maintenance of the Blueprint.

Industry 4.0 is the 4th industrial revolution which involves the integration of new digital technologies such as Internet of Things (IoT), cloud computing, and AI and ML into various sectors. Industry 4.0 began to revolutionize the manufacturing sector in the mid-2010s[1] and many industries are still experiencing this transformation. The Industry 5.0 era began in 2021[2] and is broadly refers to people, robots, and smart machines working together. The Industry 5.0 Report published by European Union in January 2021 pointed out that digital technologies such as AI need to be widely adopted across industries and optimizing human-machine interactions towards a human-centric, sustainable and resilient society[3]. According to the European Union Industry 5.0 "provides a vision of industry that aims beyond efficiency and productivity as the sole goals and reinforces the role and the contribution of industry to society." It is about a shift of focus from economic value to a societal value. As a matter of fact, AI technologies are surrounding everyone's daily life, from entertainment, merchandise, and transportation to medicine, security, education and more.

Since 1993, flooding has caused billions of dollars of damage across the US each year[4]. Proper hazard mitigation strategies and plans can not only save economy lost, but also lives. New technology like AI can improve the current workflow towards a more proactive natural disaster response solution. With a mature data foundation, AI solutions can support more rational, data-driven decision-making processes by providing near real-time flood hazard and risk data.

This document is recommendations for AI use in the NC Flood Resiliency Blueprint. Subtask 2.12 Artificial intelligence Machine Learning Tools gives a broader understanding of the current AI/ML environment.

[1] What is industry 4.0 and the Fourth Industrial Revolution? | McKinsey

[2] https://www.forbes.com/sites/jeroenkraaijenbrink/2022/05/24/what-is-industry-50-and-how-it-will-radically-change-your-business-strategy/

[3] European Commission, Directorate-General for Research and Innovation, Breque, M., De Nul, L., Petridis, A., *Industry 5.0 – Towards a sustainable, human-centric and resilient European industry*, Publications Office of the European Union, 2021, **https://data.europa.eu/doi/10.2777/308407**

[4] Climate change caused one-third of historical flood damages | Stanford News

1.2 Connections to Other NC Flood Resiliency Blueprint Tasks

There are various interconnections between this document *Recommendations for the Utilization of Artificial Intelligence and Machine Learning to inform Blueprint* (Task 3.11) and other tasks in the NC Flood Resiliency Blueprint, listed below.

Task 3.11 - AI/ML Utilization Recommendations

1.3 North Carolina Flood Resiliency Blueprint

The purpose of the North Carolina Flood Resiliency Blueprint ("the Blueprint") is to be a standardized, basin-wide flood resiliency approach applicable to all 17 NC river basins along with an actionable, decision-support tool that will form the backbone of the State's flood planning process. Such a process will inform how and where resources should be allocated and identify flood resiliency projects and strategies that should be implemented to increase community resiliency to flooding. The Blueprint will be applied within all NC river basins affected by past flood events and those anticipated to be impacted by future flood-related losses. It is not intended to be a static report or set of reports, but rather a dynamic tool that can be iterative and flexible to incorporate new data and information as they become available based on identified gaps and advances in science and technology. The Blueprint's outcomes will meet the requirements of the legislation based on the best available science, stakeholder engagement, and sound decision making. A key objective is to produce products that will be implementable, understandable, sustainable, and close the technical resource gaps among local planners, project developers, and communities.

End users include decision makers within state, regional, and local government, as well as businesses and members of communities affected by recurring and excessive flooding. The Blueprint will provide multi-scale modeling, scenario exploration systems, guidance documents, and iterative and interactive online planning tools which will increase decision makers' abilities to prioritize and implement projects and establish effective stormwater maintenance programs in response to current and future conditions. It will build on existing data, strategies, projects, plans, and efforts underway by academia, businesses, and nonprofits, and incorporate local knowledge. The vision for this project was developed through the legislation and through extensive communication with interagency staff, other states involved in flood mitigation and resiliency, municipalities and communities on the front line of flood events, climate and flood resiliency experts, legislators, and other stakeholders.

The Blueprint will accomplish several key goals, including:

- Develop community and river basin-specific risk management processes to identify and address flooding for NC communities.

- Develop an online decision support tool which seamlessly guides state, county, municipal, and other jurisdictions to identify and select flood mitigation strategies responsibly, systematically, equitably, and transparently.
- Establish a repeatable, statewide methodology for prioritizing and selecting flood mitigation strategies for future implementation.

2 AI Technologies Recommendations

Under the AI umbrella, there are several branches such as Machine Learning, Computer Vision, Language Processing and Robotics. AI options considered are Data enhancement & Cost Savings, improved data sharing and Interagency collaboration, improved engagement and decision making processes and digital twins. These AI/ML components are discussed in detail in Subtask 2.12. Combined with the Blueprint program, and based on the types of AI reviewed there are four major recommendations illustrated in this report.

2.1 Recommendation I- Data Governance Strategy

To ensure the performance of AI, Data Governance is the first step to ensure the data foundation is mature to feed into the AI process. Data Governance can ensure accessibility, usability, relevance, accuracy, timeliness, and integration of data to support decision making with a reliable data pipeline.

Figure 2-1 shows the overall design of the data governance operation model. The data governance plan can span over multiple years, which is considered as a long-term goal. Along the long-term plan implementation, with pain points identified, short-term AI pilot initiatives can be implemented.AI solutions can also be applied to areas where the data foundation achieves maturity. Capability Maturity Assessments (CMA) can assess data maturity. CMA is an approach to evaluate data governance knowledge areas or business processes from level 0 to level 5. The identification and description of capability maturity levels are as follows: Level 0 is the absence of capability; level 1 is initial or ad-hoc (success depends on the competence of individuals); level 2 is repeatable (minimum process discipline is in place); level 3 is defined (standards are set and used); level 4 is managed (processes are quantified and controlled); and level 5 is optimized (process improvement goals are quantified). The Data Management Body of Knowledge (DAMA) identifies ten knowledge areas: data architecture, data modeling & design, data storage & operations, data security, data integration & interoperability, document & content management, reference & master data, data warehousing & business intelligence, metadata and data quality[5]. Subtask 2.12 Artificial intelligence Machine Learning Tools gives a broader understanding of this topic in association with the Blueprint.

[5] Data Management Body of Knowledge, second edition. Technics Publications, Basking Ridge, New Jersey, United States. 2019.

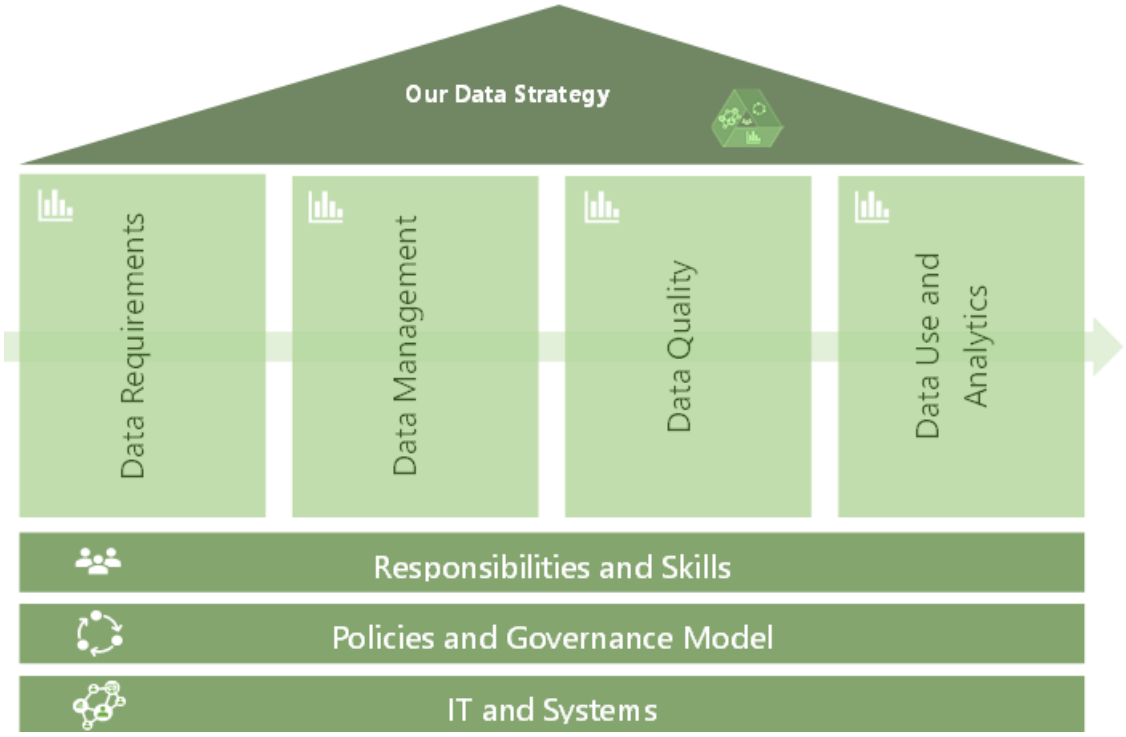

Figure 2-1 Data Strategy Operation Model

With a solid and integrated data foundation, data silos within the system can all be connected. Flood resiliency management is a complex and multi-faceted operation that requires coordinated efforts across various government agencies, as well as related Non-Governmental Organizations (NGOs) to prevent/recover from their impact and improve preparedness to withstand flooding events. A well-established data governance framework can improve data transparency and fill knowledge gaps within multiple entities.

2.2 Recommendation II- Improved Data Sharing & Interagency Collaboration via an AI-powered Chatbot

Multiple agencies are working on different data/information that could be used in the Blueprint. The desire is for everyone to continue working on their missions but coordinate their efforts. In order to streamline the multi-agency decision making process and avoid communication gaps, natural language processing can be applied to automate the interagency communication process and ultimately provide an AI-powered chatbot to serve as the single source of truth. Single source of truth

is defined as the information or data stored at one centralized location that serves as the source and basis of all business decisions. The chatbot serves as a chat agent, or the only channel to provide answers to questions to users, this would ensure users would get the same answer from multiple channels if asked the same question, it would also ensure that all agencies would be using the same data to answer these questions. The chatbot could automatically facilitate valuable, unified insights in short-term and long-term timeframe providing 24/7 support. OpenAI and others are actively using a currently available service called Kapa.ai to create project tailored "ChatGPTs" to help reduce internal support and labor costs. Additionally, Kapa.ai identifies gaps in the existing documentation and products by analyzing the generated answers. Similarly, the proposed chatbot could automate data sharing and data cleaning of related inputs from various stakeholders which may be extracted from multiple databases. The value of chatbot automation of data cleaning and sharing processes is that it would be conducted in a consistent manner to improve decision making process for better accountability and fair distribution of available resources. Figure 2-2 illustrates the concept of the single source of truth chatbot. The chatbot has the access to an integrated data source and detailed decision-making process, thus, it can guide users to locate the correct information and the correct procedures. The communication will be much more transparent and smoother than the current processes which staff may need reach out to multiple groups to gather information. There would be several steps to answering questions and chatbot would be the initial step. If chatbot has difficulty answering questions it would be routed to a human for answers.

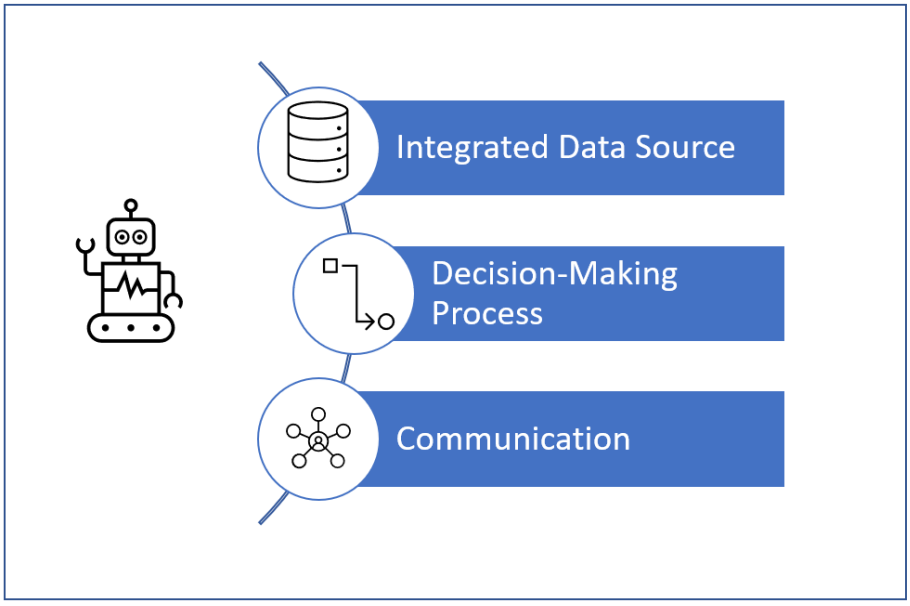

Figure 2-2 Concept of Chatbot to Provide the Single Source of Truth

A key advantage of committing to well-sought interagency collaboration for improved data sharing is comprehensive data collection for more holistic risk assessments that possibly go beyond flood management for better prevention and recovery. Furthermore, there is the ability to strategize a timely, real-time emergency response, as well as long-term resiliency planning via effective communication and quality control of resource allocations among different programs. By leveraging the power of language processing, flood resiliency planning can integrate resources such as text analysis for large complex datasets, ensuring scalability which can save significant labor cost. It can also offer sentiment analysis (e.g., citizen science on Twitter) to tailor specific community needs locally as a swift “bottom-up” response approach, as opposed to the “top-down” method of tediously running through the entire chain of commands of distinct entities to enact. Figure 2-3 shows an example of a tweet showing flooding conditions in Haywood County, NC. AI can scan through social media and gather needed information for public awareness uses or an agency decision making process. AI would not monitor social media.

Figure 2-3 Tweet about Flooding in Haywood County, NC

2.3 Recommendation III- Funding Solutions

AI, specifically machine learning and language processing technology, can efficiently identify, continuously learn about, and update project funding opportunities for flood resiliency projects. By populating project or program profiles specific to North Carolina, these solutions can provide matching and scoring for alignment of existing funding availability, perform a gap analysis on funding aggregating funding resources and needs in one place where a funding gap analysis can be conducted, and identify where more funding is needed. Additionally, these solutions provide users with the expert guidance and program insights required to navigate the rigorous requirements of the Infrastructure Investment and Jobs Act and the federal commitments of the Justice40 Initiative. Information on these initiatives is discussed in document 2.6: Flood Risk Reduction Project Funding Analysis. Most projects need to be justified, often accomplished by seeing if the benefits outweigh the costs. Having the ability to provide a rough estimate for a return on investment (ROI2.12 Artificial Intelligence Machine Learning Tools.

AI could then identify and prioritize potential mitigation sites, having calculated the cost to replace a road and the associated loss of service (flooding often causes a loss of service to transportation, utilities, employment, schooling, emergency services, etc.). The expense of the loss of services is usually not considered for large-scale planning purposes. Even without mitigation, identification of the risk alone would prepare damage cost at the time of a major event and assist stakeholders in making decisions with an optimal ROI.

2.4 Recommendation IV- Data Enhancement & Cost Savings

Data enhancement is a facet of the long-term data governance strategy. It will improve the data collection, data acquisition, data quality control and data analytics process as the data foundation. Several areas have been identified below to improve:

2.4.1 Infiltration Layer

The infiltration layer calculates the hydrology and hydraulics of a model. The advantage of using an infiltration layer is that each cell has its own calculation instead of a single averaged calculation for the entire watershed. For example, with an infiltration layer the rain that falls on a street mostly becomes runoff, while rain that falls on sand is absorbed into the ground. Unfortunately, almost all the models for North Carolina currently average the street and the sand to be considered the same. Even more challenging is that each model uses different values for the same material, often leading to results that do not match reality (e.g., sand specified as street). A standard layer allows all users to use the same input parameters and closes loopholes allowing communities to use extreme parameters to achieve goals. This would provide cost savings as each study would require the creation of a new layer while also correcting flood damage estimates from previous flawed studies.

AECOM has already created an AI-image processing tool that uses satellite imagery to determine land use at a much greater accuracy than typical data layers. This tool could be further expanded with

LiDAR (Light Detection and Ranging) point cloud data to consider vegetation density within the land cover and would use ML to produce a scientific process to calculate a curve number, manning's n value roughness coefficient, etc. The tool could also be fully automated to continuously scan new satellite imagery to detect recent changes and adjust with the seasons. ML could be further used to calibrate the standard infiltration layer based on data such as radar rainfall, antecedent moisture conditions, gage flows and stages, etc.

In 2021, the Moreton Bay Regional Council in partnership with AECOM developed an innovative approach that uses x-ray vision and point cloud data science to calculate vegetation density below the canopy. This innovative approach was successfully ground truth and significantly improves the understanding of hydraulic roughness.

Figure 2-4 shows the integration of Orthoimagery and LiDAR to generate land use data.

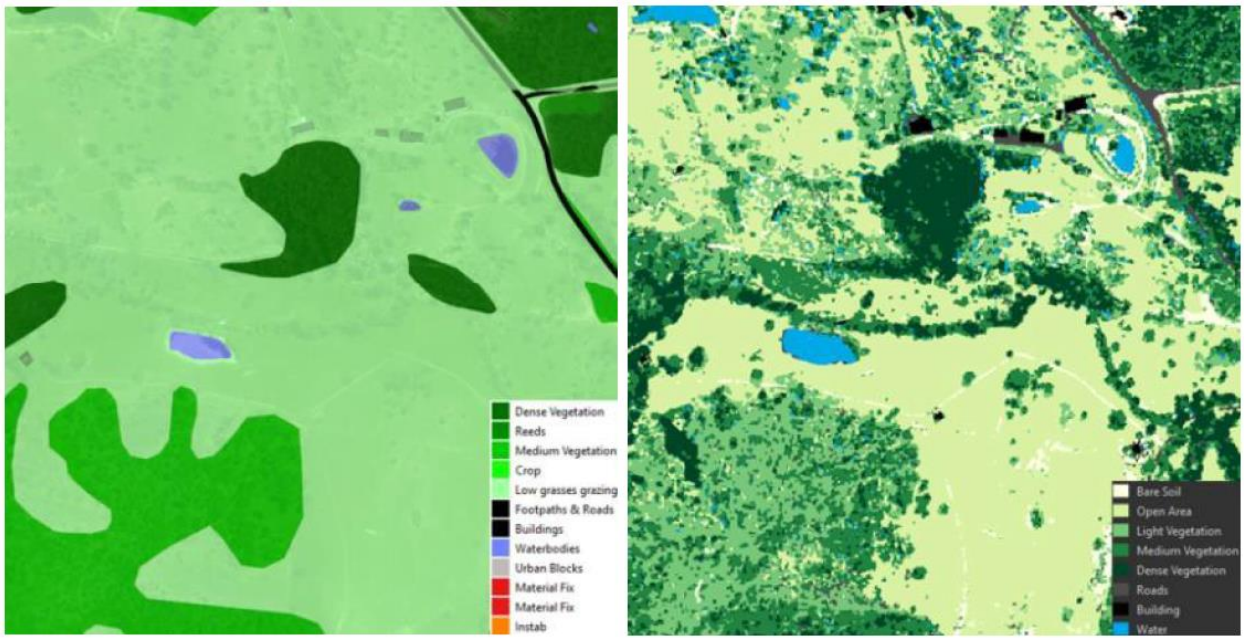

Figure 2-4 Integration of Orthoimagery and LiDAR to generate Land Use Data

- **Source**: https://www.surfacewater.biz/wp-content/uploads/2021/11/20210427_FMA21_ArtofRoughness_FINALPaper.pdf

2.4.2 Building Footprints

Multiple building footprint resources exist in several organizations. Knowing which source to use and maintaining the accuracy of these sources is challenging due to constant new construction and demolition activities. AI-image processing can scan new satellite imagery and LiDAR to detect building footprint changes and update the layer with the previously discussed infiltration layer where appropriate. The reason to standardize buildings as a statewide layer is that most models should account for these changes and currently most do not. This is a process that is ongoing with the Center for Geographic Information and Analysis (CGIA) and North Carolina Emergency Management (NCEM).

These two agencies are currently working to add new buildings from the most current state imagery and to update attributes to the structures. This is a publicly available set of data. The Geographic Information Coordinating Council (GICC) has a working group for standards on buildings.

2.4.3 Stormwater Networks

Large-scale modeling typically ignores that flooding, especially in urban areas, is often diverted and conveyed through stormwater drainage networks. It is important to capture these systems as they significantly affect model results as most mitigation projects are in areas with stormwater networks.

The initial task would be to compile a GIS (geographic information system) database of all municipal and state operated stormwater systems as a unified, statewide layer. However, there are many private systems and many municipalities that may not have a readily available data. These missing segments could be generated with AI tools. Where there is real data to validate the AI predictions, that validation process will be conducted. Culverts, outfalls, storm drains, and other structures could be identified and estimated with aerial imagery through image processing. Many of these would be simple culverts under driveways, roads, railroads, etc. These strategies would save costs as they would reduce the need to manually modify and rework layers to account for missing segments.

2.4.4 Automatic Mesh Creation Tool

Refining a 2D model's mesh is costly, and few, if any, of North Carolina's models are at the needed stage of advancement to begin the refining process. There are no practical layers to assist this process. For example, the North Carolina Department of Transportation's transportation lines layer often follows aerial imagery and not the terrain. Models will need to be refined and would benefit from an AI tool assisting with this task instead of manually creating them for the entire state. This AI tool is aiming for automating/optimizing the mesh creation processing time with the goal to reduce labor hours, it would have to be tested and compared to engineer judgement to clarify specific needs of processing and ensure the correct answers were being provided. The accuracy will be counted at the model simulation level. A refined mesh is particularly important, especially in local applications such as a mitigation project's ROI. Refined meshes include aligning cell faces with the direction of flow, aligning cell faces to extraneous features in the terrain (often via breaklines), variation in cell size (often via regions) and shape, and cell placement within the appropriate averaged calculation.

Am image processing approach could be used to determine the breakline locations instead of drawing them manually and can save many labor hours. After leveraging the breakline auto creation, 2D mesh can be created or enhanced. The task of 2D modeling is costly and presents a significant hurdle for future flood management. A high quality, refined mesh is one solution that would increase cost savings, efficiency, and overall data accuracy.

2.4.5 Streamflow Forecasting

Although USGS has more than 8,000 gages across the nation, not every single place of interest has valid record to use for flood monitoring, flood warning and Hydrology & Hydraulics modeling

verification. Leveraging nearby gage data and using the physical theory guided machine learning approach can be a good solution to predict flow condition at locations without a gauge and conduct peak flow estimation.

HydroForecast is a company using a science-guided AI solution to do streamflow prediction. It learns to represent hydrologic processes by identifying relationships between satellite observations, basin characteristics, meteorological forecasts, and streamflow measurements. The product can be used for Hydropower operations, Energy trading, Dam safety, Emergency preparedness, Water supply management and Restoration and conservation. Figure 2-5 displays an example HydroForecast prediction for forecasted flows and precipitation.

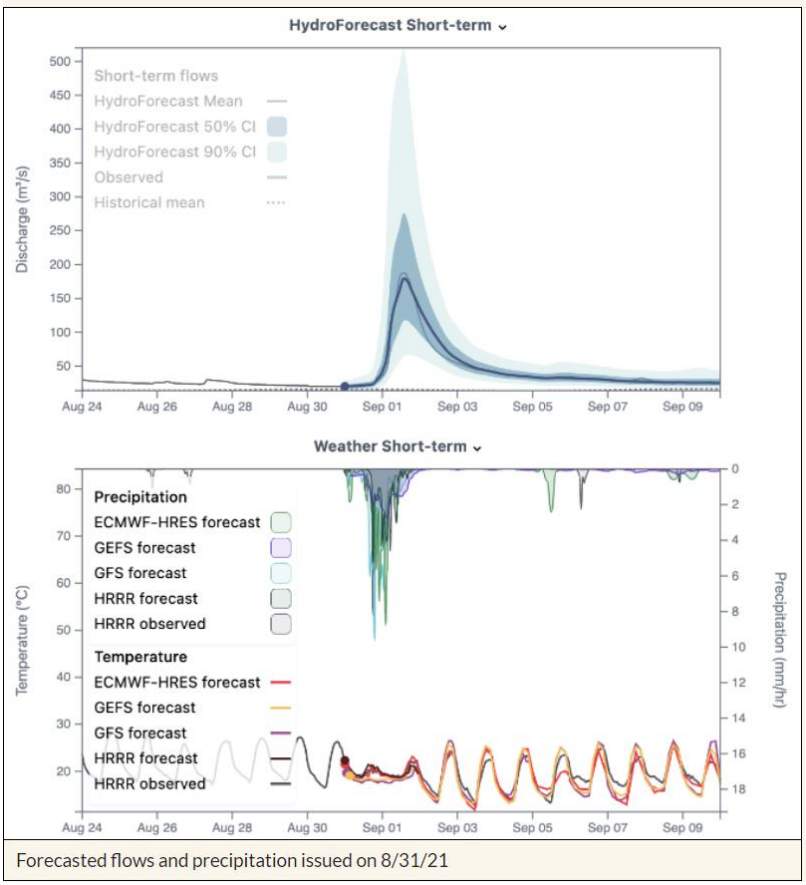

Figure 2-5 Prediction of Hurricane Ida at Little Tennessee at Needmore

3 Summary

AI is the most advanced technology in the digital era. However, it is also a technology which may change existing technology workflow, management approach and business model of traditional industry. To ensure the adoption of AI smoothly into the traditional industry needs, existing domain knowledge and physical theory will be needed to guide the right path. Meanwhile, data governance needs to be the founding strategy to ensure the data being fed into the AI process is accurate, timely, and unbiased. This report provided four recommendations as step-by-step processes for potential introductory use of AI technology to improve the Blueprint. These recommendations would move towards a better, faster and smarter platform to help flood risk mitigation and planning in North Carolina.