

Real-Time Prediction of Hurricane and Tidal Flood Risk in the Gulf Coast Using Big Data

Autumn S. Wilcox

2025-10-10

Table of contents

Abstract	2
1 Introduction	2
2 Background Review	3
2.1 Prior Work in Flood Prediction	3
2.2 Big Data in Climate and Disaster Analytics	3
2.3 Identified Gaps	4
3 Method	5
3.1 Data Sources	5
3.2 Data Storage and Processing	6
3.3 Predictive Modeling	7
3.4 Visualization and Decision Support	8
3.5 Scalability and Performance Considerations	9
4 Expected Results	10
5 Limitations and Future Work	11
References	12

Abstract

The Gulf Coast of the United States experiences frequent hurricanes and tidal flooding, resulting in severe damage to infrastructure, ecosystems, and communities. Existing flood prediction systems often rely on delayed or static data sources, limiting their effectiveness for real-time decision-making. This project proposes a scalable big data framework that integrates live environmental data from the National Oceanic and Atmospheric Administration (NOAA), the U.S. Geological Survey (USGS), and NASA to support real-time flood and storm surge prediction. The framework combines distributed processing technologies such as Apache Kafka and Apache Spark Streaming with machine learning models, including Gradient Boosted Trees, Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs). These tools will be used to model complex spatial and temporal patterns in environmental data, producing near real-time risk assessments and visualizations. The proposed system aims to enhance early warning capabilities, improve emergency management response, and contribute to the broader understanding of how big data analytics can support climate resilience across the Gulf Coast region.

1 Introduction

The U.S. Gulf Coast faces recurring hurricane and tidal flood events that cause severe economic, environmental, and social disruption (Federal Emergency Management Agency, 2022; National Oceanic and Atmospheric Administration, 2023). The increasing frequency and intensity of these storms, driven by climate change and coastal development, highlight the urgent need for accurate and timely flood prediction systems (Intergovernmental Panel on Climate Change, 2021; U.S. Global Change Research Program, 2023). However, many existing forecasting models depend on static or delayed data sources and lack the ability to integrate diverse, real-time information streams. This limitation reduces their usefulness for emergency response, evacuation planning, and infrastructure protection (U.S. Geological Survey, 2022).

This project proposes a real-time big data framework for predicting hurricane and tidal flood risk in the Gulf Coast region. The system will integrate live environmental data from sources such as the National Oceanic and Atmospheric Administration (NOAA), U.S. Geological Survey (USGS) tide gauges, and satellite imagery to generate continuously updated risk forecasts. To handle the scale and speed of these data, the framework will utilize distributed streaming technologies such as Apache Kafka for ingesting real-time data and Apache Spark Streaming for large-scale analytics, which will be described in greater detail in the Methods section. Recent advances in artificial intelligence have further emphasized the global importance of data-driven flood prediction, demonstrating the potential of deep learning for extreme event forecasting (**nearing2024?**).

The primary objective of this research is to design a framework capable of identifying flood-prone zones and forecasting storm surge severity in near real time. By leveraging big data

analytics, the project aims to enhance early warning systems, improve situational awareness, and ultimately reduce disaster impacts on Gulf Coast communities.

2 Background Review

2.1 Prior Work in Flood Prediction

Traditional flood forecasting models have relied primarily on physics-based hydrologic and hydrodynamic simulations. Systems such as the Sea, Lake, and Overland Surges from Hurricanes (SLOSH) model and the Advanced Circulation (ADCIRC) model are widely used to estimate storm surge and coastal inundation levels (Jelesnianski et al., 1992; Luettich et al., 1992). While these models can accurately reproduce storm surge under ideal conditions, they require significant computational resources and are often limited by the availability of real-time input data (Resio & Westerink, 2007). This restricts their ability to update forecasts dynamically as new observations become available.

In recent years, machine learning and data-driven methods have been introduced to improve flood forecasting. Techniques such as Random Forests, Gradient Boosting, and deep learning architectures like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) have been used to model nonlinear relationships between rainfall, tide, and water-level data (Kratzert et al., 2019; Mosavi et al., 2018). These models can capture complex spatial and temporal dependencies, enabling faster and sometimes more accurate predictions than traditional approaches. However, many of these studies operate on historical datasets rather than live data streams, and few address the computational scaling required for real-time applications (Xu et al., 2021).

While progress has been made toward predictive accuracy, the real challenge lies in combining **real-time data ingestion** with scalable, distributed processing. A true big data approach must integrate diverse data sources—such as weather radar, satellite imagery, and IoT-based sensors—into continuous analytic pipelines capable of updating predictions on demand. This gap between high-resolution forecasting and real-time operational capability motivates the framework proposed in this study.

2.2 Big Data in Climate and Disaster Analytics

The increasing availability of environmental and sensor data has made big data analytics an essential tool for understanding and predicting natural hazards. With the expansion of satellite constellations, Internet of Things (IoT) networks, and open government data initiatives, researchers now have access to vast, high-velocity datasets that capture weather patterns, ocean conditions, and hydrological responses in near real time (Jin et al., 2015; Mergili et al., 2019). Big data technologies allow for the integration and analysis of these heterogeneous data streams, providing opportunities for earlier warnings and more accurate impact assessments.

Distributed computing platforms such as Apache Hadoop and Apache Spark have become the backbone of large-scale environmental analytics (Hashem et al., 2015). These frameworks enable parallel data processing across clusters of computers, making it feasible to analyze terabytes of streaming or historical data in seconds. For example, Spark Streaming has been applied to real-time weather monitoring systems to process continuous radar feeds, while cloud-based analytics pipelines have been developed to combine satellite imagery with ground sensors for flood detection (Shen et al., 2020; Wang et al., 2021).

In the disaster management domain, the integration of social media, sensor networks, and remote-sensing data has further enhanced situational awareness and decision support. Studies have demonstrated that big data analytics can significantly improve response coordination by providing real-time visualizations of affected areas and estimating damage patterns as events unfold (Imran et al., 2020). Despite these advancements, challenges remain in data quality, latency, and interoperability, especially when fusing data from multiple agencies and platforms. These challenges underscore the need for flexible, scalable architectures that can efficiently handle continuous data ingestion and real-time model updating—an approach central to this project’s proposed framework.

2.3 Identified Gaps

Although considerable progress has been made in flood modeling and disaster analytics, several key gaps remain that limit the operational usefulness of existing systems. Most current flood prediction models focus on improving accuracy using historical data but lack the capacity for continuous, real-time updating as new information becomes available (Shen et al., 2020; Xu et al., 2021). Traditional hydrodynamic models such as SLOSH and ADCIRC are computationally intensive and often constrained by the timeliness of input data, making them unsuitable for real-time deployment during rapidly evolving hurricane events (Jelesnianski et al., 1992; Luetlich et al., 1992).

Furthermore, many machine learning-based studies demonstrate high accuracy under controlled conditions but struggle to generalize across diverse geographic regions and data sources. This is due to the limited integration of heterogeneous data types—such as satellite imagery, radar observations, and IoT sensor readings—into a unified analytical pipeline (Hashem et al., 2015; Mergili et al., 2019). Even when big data frameworks are used, challenges persist in managing latency, ensuring interoperability among distributed systems, and scaling models to accommodate continuous high-volume data streams (Imran et al., 2020).

Recent reviews confirm that while machine learning approaches continue to improve predictive accuracy, most studies still rely on historical or offline data rather than real-time integration of environmental streams (**kuhaneswaran2025?**). This persistent gap highlights the lack of streaming data architectures that can fuse live observations with dynamic model updates during extreme weather events.

These challenges underscore the need for a **real-time, scalable framework** capable of merging multiple environmental data streams into a single, adaptive analytical process. Such a system would bridge the gap between traditional forecasting models and modern big data

infrastructure, enabling faster and more reliable predictions of hurricane and tidal flood risk in the Gulf Coast region.

3 Method

3.1 Data Sources

This project will draw from a variety of publicly available and authoritative datasets to support the development of a real-time flood prediction framework. Each source provides a unique perspective on hurricane behavior, ocean conditions, and flood risk in the Gulf Coast region.

1. National Oceanic and Atmospheric Administration (NOAA)

NOAA provides real-time and historical hurricane data, including storm track coordinates, wind speed, central pressure, and surge estimates through the National Hurricane Center and the Hurricane Research Division (National Oceanic and Atmospheric Administration, 2023). The agency’s Integrated Ocean Observing System (IOOS) also maintains ocean temperature, wave height, and barometric pressure data critical for surge and tide modeling.

2. U.S. Geological Survey (USGS)

The USGS National Water Information System (NWIS) supplies near real-time tide and river gauge readings (U.S. Geological Survey, 2022). These data provide ground-truth observations that can be integrated into predictive models to evaluate the timing and magnitude of storm surges, coastal flooding, and inland water-level changes during extreme events.

3. National Aeronautics and Space Administration (NASA)

NASA’s Earth Observing System Data and Information System (EOSDIS) provides access to satellite imagery capturing precipitation, cloud cover, and land surface conditions. Datasets such as the Global Precipitation Measurement (GPM) mission and the Moderate Resolution Imaging Spectroradiometer (MODIS) can be used to monitor rainfall intensity and spatial distribution during hurricanes (Mergili et al., 2019).

4. Federal Emergency Management Agency (FEMA)

FEMA’s National Risk Index and disaster declarations database supply historical flood hazard maps, event severity indicators, and community resilience metrics (Federal Emergency Management Agency, 2022). These data can assist in evaluating the socioeconomic impacts of floods and validating model predictions.

5. Local and Regional GIS Data

County and municipal governments along the Gulf Coast often maintain open-access GIS datasets on elevation, drainage infrastructure, and floodplain boundaries. Incorporating these spatial layers improves the precision of risk mapping and allows the framework to produce community-specific flood forecasts.

Each of these data sources will be integrated through a distributed data architecture, allowing real-time ingestion, transformation, and analysis. The combination of satellite, sensor, and

ground-based datasets will enable a comprehensive, multiscale understanding of hurricane and tidal flood dynamics.

3.2 Data Storage and Processing

To manage the scale, speed, and heterogeneity of environmental data used in this study, the proposed framework will employ a distributed big data architecture. This architecture is designed to handle both batch and streaming data efficiently, enabling near real-time prediction and continuous model updates.

1. Data Ingestion (Apache Kafka)

Apache Kafka will serve as the real-time data ingestion layer, collecting continuous data streams from NOAA hurricane feeds, USGS tide gauges, and satellite sensors. Kafka’s publish–subscribe model allows data from multiple producers to be transmitted simultaneously to different consumers, ensuring that all relevant components of the system receive updates instantly (Hashem et al., 2015). This streaming capability is crucial for low-latency flood prediction during rapidly changing storm conditions.

2. Distributed Storage (HDFS and Cloud Data Lake)

The Hadoop Distributed File System (HDFS) will be used for large-scale, fault-tolerant storage of both raw and processed data. HDFS partitions datasets across multiple nodes, enabling parallel read and write operations. For long-term retention and scalability, data will also be mirrored to a cloud-based data lake such as Amazon S3 or Google Cloud Storage. This hybrid approach ensures persistence, accessibility, and redundancy.

3. Real-Time Processing (Apache Spark Streaming)

Apache Spark Streaming will handle data transformation, aggregation, and analytics in real time. Spark’s in-memory computation model significantly reduces latency compared to disk-based frameworks like MapReduce, allowing for the rapid execution of machine learning tasks on streaming data (Jin et al., 2015). The framework will process incoming sensor and satellite data to compute dynamic flood-risk indicators such as storm surge height, precipitation intensity, and predicted inundation zones.

4. Data Integration and Workflow Coordination

Structured and unstructured data from different agencies will be standardized through schema mapping and metadata tagging. Workflow orchestration will be managed using open-source tools such as Apache Airflow or Oozie to automate extraction, transformation, and loading (ETL) pipelines. This ensures that models are retrained automatically as new data arrive, maintaining system adaptability during active hurricane events.

5. Security, Quality Control, and Logging

Data integrity will be maintained using validation checks for missing or anomalous values, and system logs will track data lineage from ingestion to analysis. Role-based access control (RBAC) and encrypted data transfer protocols will ensure compliance with open-data security standards.

Together, these components create a scalable, fault-tolerant environment capable of supporting continuous flood-risk prediction for the Gulf Coast region. The combination of Kafka, HDFS, and Spark Streaming allows the system to ingest, store, and process large volumes of environmental data with minimal delay—transforming raw observations into actionable insights.

3.3 Predictive Modeling

The predictive component of this project will apply machine learning techniques to estimate hurricane-driven flood and storm surge risk in near real time. The goal is to develop a model that can continuously update as new environmental data are ingested, producing dynamic predictions of water-level changes, surge intensity, and inundation probability across the Gulf Coast.

1. Model Design and Objectives

The predictive modeling process will focus on two complementary objectives:

- (1) estimating the probability of coastal flooding given meteorological and oceanic conditions, and
- (2) predicting the expected magnitude of flood depth or surge level at specific locations.

To achieve these goals, supervised learning techniques will be employed to capture the nonlinear and multivariate relationships among variables such as wind speed, pressure, rainfall rate, tide gauge readings, and coastal topography (Kratzert et al., 2019; Mosavi et al., 2018).

2. Model Selection

A hybrid modeling approach will be adopted to leverage the strengths of both ensemble learning and deep learning methods:

- **Gradient Boosted Decision Trees (GBDT):** Suitable for structured tabular data, GBDT will be used for classification and regression of flood occurrence and intensity.
- **Long Short-Term Memory (LSTM) Networks:** LSTMs will model sequential dependencies in time-series data, such as changing tide levels and rainfall accumulation, allowing the framework to capture temporal patterns during hurricane progression (Xu et al., 2021). Recent hybrid architectures that combine LSTM with Bayesian optimization have shown improved temporal forecasting performance, suggesting the potential benefits of adaptive parameter tuning within this framework (zhou2023?).
- **Convolutional Neural Networks (CNN):** For spatially distributed inputs such as satellite precipitation imagery, CNNs will extract spatial features that inform the extent and location of potential flooding (Mergili et al., 2019).

3. Model Training and Validation

Historical hurricane and flood datasets will be used to train the models under simulated streaming conditions. Cross-validation and holdout testing will ensure that model performance

generalizes to unseen events. Evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the F1-score will be used to assess predictive accuracy for both continuous and categorical outcomes. Feature importance analyses will also be performed to identify the most influential variables in flood prediction.

4. Real-Time Model Updating

As new data arrive through Kafka streams, the model will retrain incrementally using Spark’s MLlib or TensorFlow-on-Spark integration. This online learning capability will enable the system to adapt continuously, refining its predictions throughout the lifecycle of a hurricane event. Model checkpoints and version control will ensure reproducibility and rollback in case of drift or data anomalies.

5. Integration with Decision Support Systems

The final model outputs—flood probability maps and surge-level forecasts—will be passed to visualization and alerting components described in Section 3.4. These outputs can feed directly into emergency management dashboards or geographic information systems (GIS) to inform evacuation planning and public safety decisions.

By combining ensemble and deep learning models within a scalable big data pipeline, this framework will support rapid, adaptive, and accurate flood-risk prediction for the Gulf Coast region.

3.4 Visualization and Decision Support

The final stage of the proposed framework focuses on translating predictive outputs into actionable insights for emergency managers, policy makers, and coastal communities. Visualization and decision-support tools are essential for interpreting large volumes of complex data quickly, particularly during rapidly evolving storm events when timely information can save lives.

1. Real-Time Dashboards

A dynamic web-based dashboard will serve as the central interface for visualizing model predictions. Using interactive visualization libraries such as Plotly Dash, D3.js, or Leaflet, the dashboard will display continuously updating maps of predicted storm surge height, rainfall intensity, and inundation probability. Each layer will be interactive, allowing users to zoom into specific coastal regions and overlay additional data such as evacuation routes, population density, or infrastructure locations (Wang et al., 2021).

2. Geographic Information System (GIS) Integration

Model outputs will be exported as geospatial layers compatible with GIS platforms such as QGIS or ArcGIS. This will enable emergency planners to combine predicted flood zones with official FEMA flood maps and local government GIS data for operational decision-making (Federal Emergency Management Agency, 2022). Automated updates from the big data pipeline will ensure that these maps reflect the most recent conditions without requiring manual intervention.

3. Alerting and Early Warning Systems

Threshold-based alerts will be implemented using a publish–subscribe architecture that sends notifications to local agencies when flood-risk levels exceed predefined safety limits. Integration with public notification systems (e.g., SMS, email, or push alerts) will allow near-instant communication of high-risk conditions to stakeholders and the general public (Imran et al., 2020).

4. Post-Event Analysis

Following major hurricanes or flood events, archived predictions will be compared against observed outcomes from NOAA and USGS to evaluate model performance and improve system calibration. Visualization of these results will help identify spatial or temporal biases in the predictions, guiding further optimization of the model and data-processing pipeline.

Through these visualization and alerting mechanisms, the proposed framework will bridge the gap between complex data analytics and practical disaster management. By providing real-time, interpretable outputs, the system will enhance situational awareness and enable faster, evidence-based decision-making for communities along the Gulf Coast.

3.5 Scalability and Performance Considerations

Scalability and performance are critical components of any big data system designed for real-time prediction. The proposed framework incorporates several architectural and algorithmic strategies to ensure low latency, fault tolerance, and consistent performance as data volume and velocity increase during hurricane events.

1. Horizontal Scalability

The system architecture will be designed for horizontal scaling, allowing additional computing nodes to be added dynamically as data loads increase. Both Apache Kafka and Apache Spark are cluster-based frameworks that support distributed processing across multiple nodes (Hashem et al., 2015). This configuration will allow the system to handle spikes in data ingestion from NOAA and USGS sensors without performance degradation.

2. Low-Latency Data Flow

To meet real-time requirements, Spark Streaming will process micro-batches of data with sub-second latency, and Kafka will use partitioned topics to balance the load across consumers. The system will target an end-to-end processing delay of less than one minute from data ingestion to updated model output. Performance metrics such as throughput, latency, and resource utilization will be monitored continuously to maintain service reliability.

3. Fault Tolerance and Reliability

The distributed file system (HDFS) and cloud storage layers provide data replication to prevent loss in case of hardware or network failures. Kafka’s offset tracking and Spark’s checkpointing mechanisms will allow the system to resume from the last consistent state after interruptions. These features are essential during storm events, when network disruptions or sudden surges in data volume are likely.

4. Containerization and Deployment

The framework will be containerized using Docker and orchestrated with Kubernetes to simplify deployment, scaling, and version control. This container-based design will ensure that data ingestion, processing, and visualization components can run reliably across heterogeneous cloud or on-premise environments (Wang et al., 2021).

5. Performance Optimization and Resource Management

Performance will be optimized by allocating dedicated resources to high-priority tasks such as feature extraction and prediction. Data caching and memory tuning within Spark will reduce disk I/O operations, while adaptive query execution will improve efficiency under varying workloads. These optimizations are crucial for maintaining near real-time responsiveness during extreme weather events.

Through distributed design, resource elasticity, and automated fault recovery, the proposed system will achieve the scalability required for continuous, high-volume data analysis. These capabilities will enable reliable, real-time flood prediction even under the most demanding operational conditions along the Gulf Coast.

4 Expected Results

The proposed framework is expected to demonstrate the feasibility and effectiveness of real-time hurricane and tidal flood prediction using big data analytics. By integrating distributed processing, machine learning, and real-time data ingestion, the system will produce several tangible outcomes that advance both research and practical disaster management capabilities.

1. Real-Time Flood Prediction Prototype

A functional prototype will be developed to simulate real-time hurricane and flood prediction for the Gulf Coast. This prototype will integrate data streams from NOAA, USGS, and NASA sources and process them using the Kafka–Spark pipeline. The model is expected to generate updated flood-risk predictions at sub-hourly intervals, demonstrating significant reductions in processing time compared to conventional hydrologic models such as SLOSH and ADCIRC (Jelesnianski et al., 1992; Luettich et al., 1992).

2. Improved Prediction Accuracy and Timeliness

By combining machine learning and deep learning models trained on both historical and live data, the framework is expected to outperform traditional methods in forecasting storm surge height, rainfall accumulation, and flood extent. The integration of LSTM and CNN models will allow for accurate spatiotemporal pattern recognition, enhancing predictive precision for both coastal and inland flooding (Kratzert et al., 2019; Xu et al., 2021).

3. Scalable and Reproducible Architecture

The project will result in a scalable architecture that can be deployed on cloud platforms such as AWS or Google Cloud. Containerization through Docker and orchestration via Kubernetes

will ensure reproducibility and portability across different environments. This design will allow researchers and agencies to replicate the framework for other high-risk coastal regions.

4. Decision-Support Tools for Emergency Management

The visualization and alerting components are expected to demonstrate how big data systems can be operationalized for real-world decision-making. The resulting dashboard and GIS-integrated maps will provide emergency managers with continuously updated situational awareness, enabling more efficient evacuation planning and resource allocation (Federal Emergency Management Agency, 2022; Wang et al., 2021).

5. Research Contribution

The project will contribute to the growing field of environmental informatics by providing a case study in applying big data frameworks to climate-related disaster prediction. It will also highlight the challenges of integrating streaming data, distributed systems, and AI models in real-time scientific workflows—offering insights relevant to future research in both academia and government applications.

Overall, the expected outcome of this project is a validated, scalable framework that demonstrates how big data technologies can transform flood-risk forecasting from static modeling to dynamic, real-time decision support for the Gulf Coast region.

5 Limitations and Future Work

While the proposed framework provides a robust foundation for real-time hurricane and tidal flood prediction, several limitations must be acknowledged. Recognizing these constraints is essential for improving the framework’s scalability, reliability, and operational readiness in future research.

1. Data Availability and Quality

The effectiveness of the model depends heavily on the availability and consistency of high-quality data streams. Gaps in sensor coverage, communication failures during storms, and inconsistent temporal resolutions across datasets can introduce uncertainty into predictions (U.S. Geological Survey, 2022). Additionally, satellite imagery is subject to cloud cover and latency issues that may reduce the timeliness of updates.

2. Computational and Infrastructure Costs

Although the system is designed for distributed scalability, deploying a real-time big data pipeline across multiple nodes or cloud services can be resource-intensive. Smaller research teams or public agencies may face challenges in maintaining the required computing infrastructure, especially during prolonged hurricane seasons (Hashem et al., 2015).

3. Model Interpretability and Drift

Machine learning models such as LSTM and CNN are often considered “black-box” systems, making it difficult to interpret decision logic or explain predictions to non-technical users (Kratzert et al., 2019). Moreover, as environmental conditions change due to climate

variability, model drift may occur—requiring continuous retraining and validation to maintain accuracy.

4. Interagency Data Integration

Integrating data across multiple agencies (e.g., NOAA, USGS, FEMA, and local governments) introduces challenges in standardizing formats, access permissions, and update frequencies. Variations in data schemas and collection methodologies can create bottlenecks for real-time interoperability (Imran et al., 2020).

5. Future Work

Future research should focus on expanding the framework to include additional data sources such as crowd-sourced flood reports, social media activity, and high-resolution drone imagery. Incorporating ensemble learning and explainable AI (XAI) techniques could improve both accuracy and interpretability of the predictions. Additionally, integrating the system with edge-computing devices deployed at coastal monitoring stations could reduce latency and improve resilience during network outages.

By addressing these limitations, future iterations of the framework can evolve into a fully operational decision-support platform for Gulf Coast emergency management, combining scientific rigor with real-time responsiveness.

References

- Federal Emergency Management Agency. (2022). *FEMA national risk index: Coastal and flood hazards*. <https://hazards.fema.gov/nri/>.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. <https://doi.org/10.1016/j.is.2014.07.006>
- Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2020). A survey of social media data analysis for disaster response and crisis management. *Computers in Human Behavior*, 110, 106548. <https://doi.org/10.1016/j.chb.2020.106548>
- Intergovernmental Panel on Climate Change. (2021). *Climate change 2021: The physical science basis*. <https://www.ipcc.ch/report/ar6/wg1/>.
- Jelesnianski, C. P., Chen, J., & Shaffer, W. A. (1992). *SLOSH: Sea, lake, and over-land surges from hurricanes*. NOAA Technical Report NWS 48; <https://repository.library.noaa.gov/view/noaa/7131>.
- Jin, J., Gubbi, J., Marusic, I., & Palaniswami, M. (2015). A review on big data analytics in the climate and environmental sciences. *IEEE Transactions on Big Data*, 1(1), 57–66. <https://doi.org/10.1109/TBDDATA.2015.2465959>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2019). Toward improved predictions in hydrologic modeling: Combining long short-term memory networks and process-based models. *Water Resources Research*, 55(11), 7772–7794. <https://doi.org/10.1029/2019WR025528>

- Luetlich, R. A., Westerink, J. J., & Scheffner, N. W. (1992). ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. *U.S. Army Corps of Engineers, Technical Report DRP-92-6*.
- Mergili, M., Jaboyedoff, M., Pullanikkatil, D., & Pudasaini, S. (2019). Integrating remote sensing and big data for natural hazard risk analysis: A review. *Progress in Earth and Planetary Science*, 6, 1–22. <https://doi.org/10.1186/s40645-019-0288-1>
- Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>
- National Oceanic and Atmospheric Administration. (2023). *Hurricanes and tropical storms climatology*. <https://www.nhc.noaa.gov/climo/>.
- Resio, D. T., & Westerink, J. J. (2007). White paper on estimating hurricane inundation probabilities. *NOAA Technical Report*.
- Shen, L., Liu, Y., & Wu, Z. (2020). Real-time flood monitoring using apache spark and satellite data. *Environmental Modelling & Software*, 132, 104795. <https://doi.org/10.1016/j.envsoft.2020.104795>
- U.S. Geological Survey. (2022). *USGS water data for the nation: Real-time data*. <https://waterdata.usgs.gov/nwis>.
- U.S. Global Change Research Program. (2023). *Fifth national climate assessment: Southeast chapter*. <https://nca2023.globalchange.gov/>.
- Wang, X., Huang, C., Liu, M., & Chen, J. (2021). Cloud-based big data analytics for real-time flood prediction and response. *Journal of Hydrology*, 598, 126423. <https://doi.org/10.1016/j.jhydrol.2021.126423>
- Xu, Z., Liang, K., & Li, L. (2021). Deep learning for hydrologic and hydraulic prediction: A review. *Hydrology and Earth System Sciences*, 25, 3899–3923. <https://doi.org/10.5194/hess-25-3899-2021>