A Machine Learning Approach for Water Level Residual Correction Using Geospatial Terrain Features

Yan (Ethan) Xu

Abstract–**Accurate water level forecasting remains a critical challenge in hydrology, where errors in model predictions may compromise safety and flood defense decisions. This study proposes a machine learning approach that uses geospatial terrain data to correct these residuals. The study presents a LightGBM baseline model using only temporal features and a XGBoost model enhanced with topographic features, including slope, aspect, and elevation statistics derived from 30m DEMs within a 3km radius of 15 diverse National Oceanic and Atmospheric Administration (NOAA) water level stations[1]. The key finding is that the geospatially informed XGBoost model demonstrated a significant 84.38% improvement in MSE for multi-step sequential forecasting compared to the temporal LightGBM baseline, which shows that it was able to overcome the typical challenge of error accumulation in autoregressive tasks. While the simpler LightGBM architecture maintained an advantage for one-step-ahead prediction, the results demonstrate that integrating geospatial data with an appropriate model architecture is very effective for autonomous, long-term residual correction. This study aims to present a framework for incorporating static terrain data into dynamic hydrological forecasting models.**

I. Introduction

Flooding endangers human life and property, with the Federal Emergency Management Agency (FEMA) estimating that floods cause about $8.2 billion in damages every year in the United States alone[4]. Despite modern advancements in hydrodynamic modeling and machine learning techniques, persistent residuals (errors) between predicted and observed water levels remain a challenge, particularly in complex tidal environments where multiple physical processes may interact nonlinearly[5].

Current approaches to correcting residuals primarily focus on temporal features, which are variables derived from historical data to capture time-dependent patterns. Examples of these features include autoregressive components and harmonic analysis of tidal constituents such as K1, P1, K2, and S2 [6][7]. While machine learning models such as Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) have shown to be very capable in capturing temporal patterns[8][9], they often neglect the potential influence of surrounding terrain characteristics, such as local elevation gradients, slopes, and low-lying regions, on water level errors. Rather than acting as an alternative to machine learning, digital elevation models (DEMs) serve as a highly complementary data source that can provide this missing spatial context. DEMs are gridded raster datasets that express surface elevation at each pixel, similar to a height map, that permits quantitative extraction of attributes such as slope, aspect, and elevation statistics (min, max, std). By extracting these topographical features, DEMs can supply critical static inputs to machine learning algorithms, which bridges the gap between dynamic temporal forecasting and static terrain analysis. DEMs have been used extensively in other forms of hydrological modeling, such as flood simulation[10], which shows the importance of topographic features in water behaviour prediction and its potential in residual correction.

This paper investigates a novel approach to water level residual correction by incorporating geospatial elevation data features, including slope, aspect, and other statistical values derived from high-resolution DEMs within a 3-km radius of water stations. In detail, the study aims to

A. Develop a baseline LightGBM machine learning model for water level residual correction that only includes temporal features from 15 diverse NOAA stations.
B. Develop an XGBoost machine learning model for water level residual correction that integrates both temporal and geospatial features from the 15 NOAA stations and DEM data.
C. Evaluate the individual models' performances by analyzing root mean square error and R^2.
D. Compare the models' performances on one-step-ahead prediction.
E. Compare the models' performances in multi-step autoregressive forecasting scenarios where error accumulation typically degrades performance.

II. Literature Review / Background

Traditional approaches to water level residual correction have primarily relied on temporal and statistical techniques. Harmonic analysis, implemented in tools like T_TIDE and NS_TIDE, decomposes tidal signals into constituent components to predict and correct residuals[11]. Statistical time-series models such as ARIMA (AutoRegressive Integrated Moving Average) use autoregressive components and differencing to forecast residuals based on past values[12]. However, these methods predominantly focus on temporal patterns and often overlook spatial contextual factors, such as local topography, which may influence residual behaviour.

Machine learning (ML) models have shown to be powerful tools for hydrological modeling and forecasting, where they are able to effectively grasp nonlinear relationships. Techniques like Long Short-Term Memory (LSTM) networks excel in sequence prediction tasks (e.g., streamflow forecasting) by using temporal dependencies[19]. Gradient boosting algorithms (e.g., XGBoost, LightGBM) are increasingly used for regression tasks due to their strength in handling feature interactions[20].

DEMs have become fundamental in many hydrological studies, where they enable watershed delineation, flood modeling, and bathymetric mapping[13][14][15]. Global DEMs like SRTM, NASADEM, and Copernicus DEM (source of DEM data for this study)[16] provide crucial topographical data for extracting terrain features such as slope, aspect, and elevation ranges. However, these advantages do not fully translate to a large-scale coastal context, where tidal harmonics and broad spatial extents may introduce noise and dilute vital local patterns. This is the main motivation behind using hyper-local terrain features (e.g., within a 3 km radius of a station) in residual correction for water level predictions, as DEMs have extensive utility in basin-scale applications[17]. By restricting the context, the model is able to extract finer details from DEM-derived features (e.g., slope, curvature) and how they can influence water level residuals in coastal environments[18].

III. Methodology

The study utilized data from 15 diverse NOAA water level stations (ID: 9414290, 9447130, 9419750, 9432780, 8768094, 8761927, 8724580, 8658120, 8665530, 8575512, 8518750, 8443970, 9410170, 8723214, and 1612340) selected to represent a wide range of environments across the United States. When multiple stations were available within a similar environment, selections were prioritized based on the availability of continuous, verified historical data without significant recording gaps during the four-year study period. These stations provided a robust dataset for creating the proposed residual correction methodology under varying and diverse conditions. For each station, four years (1460 days) of verified historical water level data, collected at 6-minute intervals from 2021-08-20 to 2025-08-19 were obtained from the NOAA Tides and Currents database. The residual time series, which serves as the target variable for residual correction modeling, was calculated as the difference between observed and predicted water levels (Residual = Observed - Predicted), where predicted levels were extracted from NOAA's operational forecast models[21].

Geospatial data were sourced from the Copernicus Digital Elevation Model (GLO-30)[22], which provides global coverage at 30-meter resolution. For each station, a 3km by 3km raster clip centered on the station coordinates was extracted. This spatial extent was chosen to capture local topographic features without introducing unnecessary computational complexity. The Copernicus DEM was selected for its high vertical accuracy (<4m absolute error) and global consistency[23], making it suitable for detailed terrain analysis.

Terrain features were derived from each DEM clip using Rasterio for reading raster data and NumPy for geospatial calculations. These included elevation-based statistics (elev_mean, elev_std, elev_min, elev_max, elev_range, station_elev, low_elev_pct, flood_prone_area), slope metrics (slope_mean, slope_std), and aspect metrics (aspect, aspect_mean). While these features do not capture the full complexity of hydrodynamic and coastal processes, they act as significant indicators of local terrain structure. For instance, the elevation distributions and percentage of low-lying areas can give insight into the terrain's local water storage potential, thus providing a stable spatial context that allows the model to differentiate distinct coastal environments and potentially introduce a grounding context that the model can rely on for long-term forecasting.

Temporal features were constructed from the residual time series to capture autoregressive patterns. These included residual_lag1 and residual_lag2 (lagged values from 1 and 2 time steps before), hour_of_day, predicted_m (water level predicted by NOAA's model), rate_of_change_m (change of residual_m), month, hour, dayofweek, tide_phase (1 for rising, 0 for stagnant, -1 for falling), and season (0 for December, January, February, 1 for March, April, May, 2 for June, July, August, 3 for September, October, November). This simplified seasonal grouping was used to capture broad temporal variability while maintaining consistency across stations. The combination of spatial and temporal features intends to model both the historical context and static geographic data.

The XGBoost algorithm was selected for this study due to its proven effectiveness in handling tabular data with mixed feature types, its robustness to multicollinearity, and its ability to calculate feature importance, which measures how frequently and effectively each input feature contributes to reducing prediction error during model training, thereby quantifying their significance[24]. These characteristics are crucial for interpreting which terrain and temporal factors most influence residuals. The dataset was split into a 70% training set and a 30% test set. Hyperparameter tuning was performed via GridSearchCV with 5-fold temporal cross-validation, optimizing for Root Mean Squared Error (RMSE). Key hyperparameters included learning rate, maximum tree depth, number of estimators, subsample ratio, column sample ratio, and the gamma regularization term. The model was implemented using the XGBoost library in Python.

Two different validation criteria were used to assess model performance:

A. One-Step-Ahead Prediction: The model was trained and tested using true lagged residuals (residual_lag1, residual_lag2) as inputs. This evaluates the model's ability to predict the next residual when accurate recent history is available.

B. B.Multi-Step Autoregressive Forecasting: The model was trained on true data but tested autonomously over a two-day timeframe. In this criterion, the model's own predictions were fed back as inputs for subsequent time steps (e.g., predicted residual became residual_lag1 and used for other feature calculations for next prediction). This tests the model's ability to forecast without relying on continuous true data input.

Model performance was evaluated using Root Mean Squared Error (RMSE) and the coefficient of determination (R2). RMSE represents the average prediction error magnitude (lower is better) and especially punishes larger errors, as large residual misses can significantly destabilize multi-step autoregressive forecasting due to error accumulation, which makes it more critical to minimize large errors than small deviations. R2 measures the proportion of variance in the observed residuals explained by the model, where values closer to 1 indicate a better fit..

IV. Results

The initial evaluation of the models focused on their core predictive accuracy during training and validation using the conventional one-step-ahead criterion across the entire dataset of 15 stations. The dataset, as mentioned previously, was separated into a 70/30 split where 70% of the dataset is used for training and 30% for validation. The LightGBM (LGBM) model, which serves as the non-geospatial baseline, achieved a strong validation performance, with a root mean squared error (RMSE) of 0.0102 meters and a coefficient of determination (R2) of 0.9920. This indicates that the model was able to capture the dominant temporal patterns in the residual time series. The XGBoost (XGB) model, which incorporates both temporal and geospatial features, demonstrated a similar ability to learn from the training data, where it attained a RMSE of 0.0097 meters and an R2 of 0.9928. These results confirm that both architectures are capable of the one-step-ahead prediction task for which they were trained, with the XGBoost model showing a marginal improvement in RMSE in this specific context. However, it is important to note that this evaluation was performed on the validation split of the original 15-station dataset used for training, where the performance of both models stated above are when they are predicting familiar data. For reproducibility, the final XGBoost model used eta = 0.30, max_depth = 6, n_estimators = 200, subsample = 1.0, colsample_bytree = 1.0, and gamma = 0.0.

Feature importance was evaluated using the built-in importance plotting functions of both XGBoost and LightGBM, which rank features by their contribution to reducing the training objective across decision-tree splits, where high importance indicates a feature is repeatedly selected in splits that produce larger decreases in prediction error. However, it is important to note that the feature importance values are not directly comparable across different model architectures, as they have their own internal tree structure, splitting criteria, and generally different numerical scales. Therefore, the magnitude of these feature importance values should only be comparable in relation to other values in the same model. For the LGBM model, residual_lag1 was the most influential feature (feature importance score of 441352.492), followed by rate_of_change_m (feature importance score of 4495.545) and residual_lag2 (2935.392) (Figure 1). In the XGB model, residual_lag1 also ranked highest (feature importance score of approximately 1044.15), but geospatial features such as elev_mean (3rd place, score of approximately 40.29) and elev_max (4th place, score of approximately 38.25) appeared among the top five predictors, indicating their albeit subtle but consistent role in model predictions (Figure 2).

An interesting observation is that rate_of_change_m was substantially more influential in the LightGBM model as opposed to the XGBoost model, while the relative importance of residual_lag2 remained consistent. This difference likely stems from the role of temporal and spatial context in each model. In the LightGBM model, the rate of change holds value as it acts as it informs the model of short-term water level dynamics that allows it to detect rapid transitions in water level residuals that may be driven by external factors, such as tidal phase shifts. However, the XGBoost model is less dependent on the rate of change to provide external context as the geospatial features can also be used to understand environmental factors, making it less reliant on short-term temporal derivatives and instead use both temporal and spatial information to generate predictions. The consistent reliance on lagged residual values on both models suggests that autocorrelation remains a fundamental source of pattern for residual prediction, even with the presence of geospatial features.

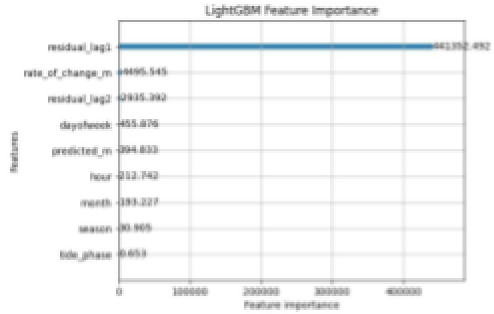

Figure 1. LightGBM Feature Importance. Ranks the relative influence of temporal variables on residual prediction accuracy across 15 training stations. Lagged residuals (residual_lag1, residual_lag2) and the rate of change are the dominant predictors.

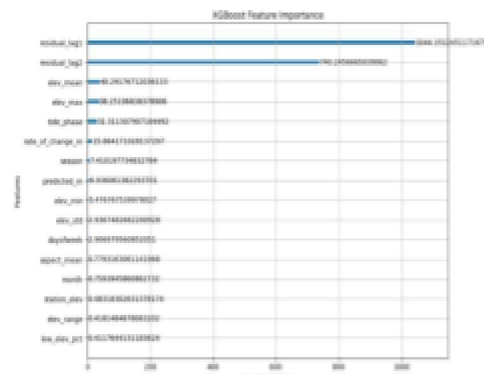

Figure 2. XGBoost Feature Importance. Ranks the relative influence of combined temporal and geospatial variables on residual prediction accuracy across 15 training stations. Alongside temporal lags, static terrain metrics like mean and maximum elevation (elev_mean, elev_max) emerge as top predictors.

To specifically test the model's ability to generalize to unseen conditions, a separate evaluation was conducted on completely new data from NOAA station 9411340, which was not included in the original training set, using water level data from 2025-08-22 to 2025-08-23. The one-step-ahead prediction results demonstrated strong performance for both models on this unseen station, though with the performance relationship reversed compared to the validation results, where the LGBM model outperforms the XGB model. The LGBM model achieved a root mean squared error (RMSE) of 0.0085 meters (Figure ?) while the XGB model achieved a RMSE of 0.0118 meters (Figure ?). This represents a 28.26% improvement in RMSE for the LGBM model compared to the XGB model on this specific unseen station. The high accuracy of both models confirms that they are able to effectively use temporal dependencies (e.g., residual_lag1) for immediate residual correction when true historical data is available, though the LGBM model demonstrated better generalization for one-step prediction on this particular unseen station. For validation visualization, three curves will be drawn. The blue line labeled "observed_m" for LightGBM and "Observed" for XGBoost represents the true measured water level at that time and serves as the goal for the model. The orange line labeled "predicted_m" for LightGBM and "Predicted (NOAA)" for XGBoost represents the water level predicted by NOAA's model before the true value was measured and serves as the data the LightGBM and XGBoost models should correct. The green line labeled "corrected_predicted_m" for LightGBM and "Corrected Predicted (XGBoost)" for XGBoost represents the sum of NOAA's predicted water level and the model's residual output and serves as the corrected version. The success of the residual models can be approximated by how much the green line overlaps with the blue line.

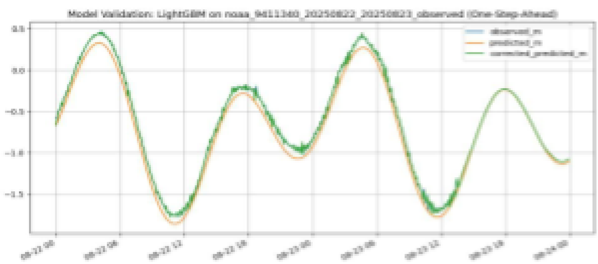

Figure 3. Model Validation: LightGBM (One-Step-Ahead). Compares water level predictions at unseen NOAA station 9411340 over a two-day period (August 22-23, 2025) using true prior residuals. The graph displays the true measured level (observed_m, blue), NOAA's baseline prediction (predicted_m, orange), and the LightGBM-corrected prediction (corrected_predicted_m, green).

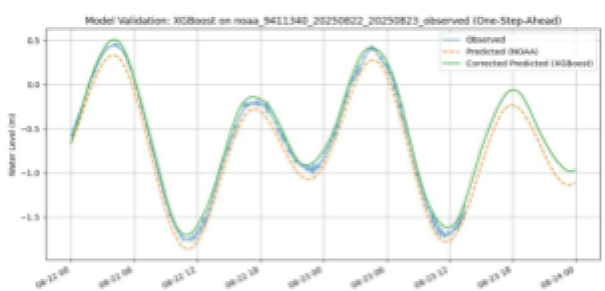

Figure 4. Model Validation: XGBoost (One-Step-Ahead). Compares water level predictions at unseen NOAA station 9411340 over a two-day period (August 22-23, 2025) using true prior residuals. The graph displays the true measured level (Observed, blue), NOAA's baseline prediction (Predicted (NOAA), orange), and the geospatially-informed XGBoost correction (Corrected Predicted (XGBoost), green).

In contrast to the one-step-ahead results, the multi-step autoregressive forecasting criterion showed a significant advantage for the XGB model incorporated with geospatial data. The LGBM baseline displayed rapid error accumulation, with its RMSE of 0.4887 meters over a two-day forecast timeframe. In this context, error accumulation refers to the recursive propagation of prediction errors, where an initial error causes later predictions to drift due as the erroneous prediction is fed back into the model for subsequent time steps. As the subsequent time steps start to produce errors as well, these errors sum up and compound, causing the model to become more inaccurate over time. In this case, the erroneous predictions occurred almost immediately at the start, and as the errors compounded, the model's performance deteriorated even more as it continued. The XGB model, however, maintained stability, achieving an RMSE of 0.0763 meters, a 84.38% improvement over the LGBM baseline. This suggests that geospatial features (e.g., terrain elevation and slope statistics) provided a net positive effect that lessens the impact of error accumulation by allowing the model to have topographic context to generate a more well-informed prediction.

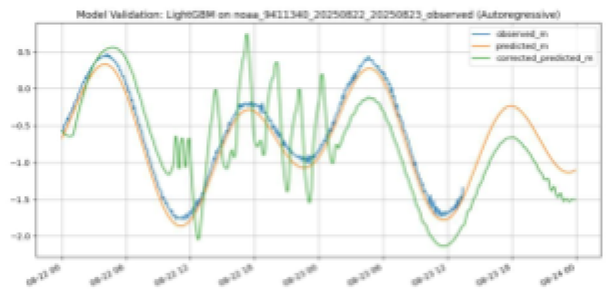

Figure 5. Model Validation: LightGBM (Autoregressive). Evaluates autonomous multi-step forecasting at unseen NOAA station 9411340 over a two-day period. The LightGBM-corrected prediction (green) deviates significantly from the true observed level (blue) and NOAA baseline (orange), demonstrating rapid error accumulation.

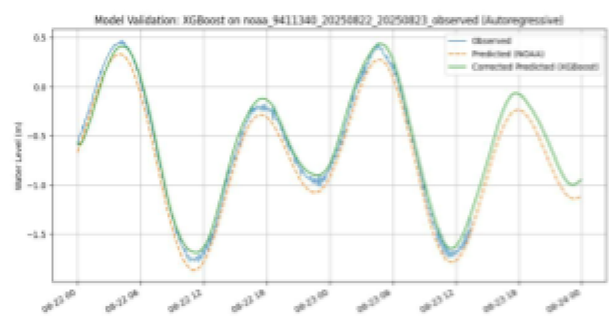

Figure 6. Model Validation: XGBoost (Autoregressive). Evaluates autonomous multi-step forecasting at unseen NOAA station 9411340 over a two-day period. The XGBoost-corrected prediction (green) closely tracks the true observed level (blue) against the NOAA baseline (orange), demonstrating model stability.

V. Discussion

The results present a nuanced role of geospatial data in machine learning-based water level residual correction, with the most significant finding being the dependence of model performance on the forecasting criterion. The superior performance of the LightGBM (LGBM) model in the one-step ahead task on unseen data suggests that for immediate correction where the true prior residual is known, a simpler model focused on temporal features is highly effective and also computationally efficient due to its lightweight architecture. However, the superior performance of the geospatially-informed XGBoost (XGB) model in the multi-step autoregressive task is the key finding of this study. Its ability to maintain an RMSE of 0.0763 meters against the LGBM's 0.4887 meters demonstrates that static terrain features provide a stabilizing context that can help mitigate the problem of error accumulation in recursive prediction. This suggests that geospatial data acts as an "anchor" that may prevent the model's predictions from drifting uncontrollably by allowing the model to have a sense of the physical reality of the station's location, which is a factor that remains constant and reliable over time.

An important consideration is whether the performance differences stem from the geospatial features or the choice of model architecture itself. This study intentionally paired LightGBM with temporal features and XGBoost with the combined feature set based on their inherent architectural complexity and strengths. LightGBM's simpler model architecture makes it very efficient for processing simpler feature sets, making it ideal for capturing patterns in temporal data without overfitting. By contrast, XGBoost's more complex algorithms and robust regularization techniques are better suited to handle the higher-dimensional, more complex feature space created by adding geospatial variables. If the architectures were swapped, using XGBoost on only temporal features might lead to increased overfitting, while applying LightGBM to the full feature set might result in underfitting and failure to capture the subtle but important geospatial patterns. Thus, the current pairing represents a match of model complexity to feature space complexity.

Still, several limitations of this study must be acknowledged. First, the geospatial data was static, but incorporating more dynamic spatial features like real-time precipitation could help the model capture more variability and patterns than terrain statistics alone. Second, the selection of a 3km radius, while reasoned to balance the amount of spatial data collected, was ultimately arbitrary. A deeper analysis on the optimal spatial context window around a station would be a valuable extension of this work. Third, the seasonal encoding uses uniform grouping across all stations. However, seasonal activity across stations varies due to differences in climate and geographical region, where some may exhibit stronger seasonal variability due to, for instance, snowmelt and regional storm cycles. Future work should improve this representation either by simply using month-level encoding, or more advanced temporal embeddings that allow the model to capture seasonal variability without the constraints of monthly categorization. Finally, the model's performance was evaluated on a limited number of stations, and while the results are promising, validation across a more extensive and global network of stations would strengthen the generalizability of the model.

Despite these limitations, this research aims to open several promising avenues for future work. The most direct extension is the integration of dynamic geospatial data into the feature set. Furthermore, exploring more complex model architectures like Transformer-based models or Graph Neural Networks (GNNs) could better capture the complex spatiotemporal relationships between residuals and the surrounding terrain. Finally, the principle of using static geospatial features to stabilize temporal forecasts could be applied to other hydrological forecasting applications, such as groundwater level modeling.

VI. Conclusion

The study demonstrated that the integration of geospatial terrain features with machine learning models presents a promising solution for improving water level residual correction, particularly in multi-step forecasting scenarios. While a simpler LightGBM model excelled at one-step-ahead predictions using temporal features alone, the XGBoost model, enriched with topographic data, achieved a significant 84% improvement in accuracy compared to the LightGBM model during autonomous multi-step forecasting. This key finding highlights the role of static terrain features in providing a stabilizing effect that helps mitigate error accumulation, which is the main challenge in recursive forecasting.

Looking forward, future work should explore the combination of real-time spatial data, such as real-time precipitation and more sophisticated architectures like graph neural networks to capture deeper underlying patterns. Ultimately, this approach intends to take a step toward a more resilient, robust, and accurate forecasting tool that can adapt to the growing challenges of coastal risk management in a world of climate change and global warming.

VII. References

[1] Center for Operational Oceanographic Products and Services, National Oceanic and Atmospheric Administration, "NOAA Tides and Currents," https://tidesandcurrents.noaa.gov/.

[2] R. Guerrero, "Fundamentals of Water Level Measurement," AEM Blog, https://blog.aem.eco/fundamentals_of_water_level_measurement.

[3] National Oceanic and Atmospheric Administration, "The Importance of Monitoring the Tides and Their Currents," NOAA Ocean Service Education, https://oceanservice.noaa.gov/education/tutorial_tides/tides09_monitor.html

[4] R. Guerrero, "Fundamentals of Water Level Measurement," AEM Blog, https://blog.aem.eco/fundamentals_of_water_level_measurement.

[5] K. Fei, H. Du, and L. Gao, "Accurate water level predictions in a tidal reach: Integration of Physics-based and Machine learning approaches," J. Hydrol., vol. 622, pt. B, p. 129705, Jul. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0022169423006479.

[6] Y. Chen, M. Gan, S. Pan, H. Pan, X. Zhu, and Z. Tao, "Application of auto-regressive (AR) analysis to improve short-term prediction of water levels in the Yangtze estuary," J. Hydrol., vol. 590, p. 125386, Nov. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0022169420308465.

[7] S. Li, L. Liu, S. Cai, and G. Wang, "Tidal harmonic analysis and prediction with least-squares estimation and inaction method," Estuarine, Coastal and Shelf Sci., vol. 220, pp. 196–208, May 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0272771418307327.

[8] M. N. A. Zakaria, A. N. Ahmed, M. A. Malek, A. H. Birima, M. M. H. Khan, M. Sherif, and A. Elshafie, "Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia," Heliyon, vol. 9, no. 7, p. e17689, Jun. 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10344711/.

[9] J. Zhou, L. Chen, T. Hu, H. Lu, Y. Shi, and L. Chen, "The comparative study of machine learning agent models in flood forecasting for tidal river reaches," Sci. Rep., vol. 15, Art. no. 19130, May 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-04633-w.

[10] K. Xu, J. Fang, Y. Fang, Q. Sun, C. Wu, and M. Liu, "The Importance of Digital Elevation Model Selection in Flood Simulation and a Proposed Method to Reduce DEM Errors: A Case Study in Shanghai," Int. J. Disaster Risk Sci., vol. 12, pp. 890–902, Nov. 2021. [Online]. Available: https://link.springer.com/article/10.1007/s13753-021-00377-z.

[11] Y. Zhang, S. Jiao, W. Zhou, Y. Wang, and X. Lv, "Exploration the effect of nonstationary signals on the tidal phenomenon using F_TIDE part I," Sci. Rep., vol. 15, Art. no. 6303, Feb. 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-90490-6.

[12] Zilliz, "What are the limitations of ARIMA models?," Milvus AI Quick Reference, https://milvus.io/ai-quick-reference/what-are-the-limitations-of-arima-models.

[13] S. Dávila-Hernández, J. González-Trinidad, H. E. Júnez-Ferreira, C. F. Bautista-Capetillo, H. M. de Ávila, J. C. Escareño, J. Ortiz-Letechipia, C. O. Robles Rovelo, and E. A. López-Baltazar, "Effects of the Digital Elevation Model and Hydrological Processing Algorithms on the Geomorphological Parameterization," Water, vol. 14, no. 15, Art. no. 2363, Jul. 2022. [Online]. Available: https://www.mdpi.com/2073-4441/14/15/2363.

[14] J. Li and D. W. S. Wong, "Effects of DEM sources on hydrologic applications," Comput. Environ. Urban Syst., vol. 34, no. 3, pp. 251–261, May 2010. [Online]. Available:

https://www.sciencedirect.com/science/article/abs/pii/S0198971509000933.

[15] K. Xu, J. Fang, Y. Fang, Q. Sun, C. Wu, and M. Liu, "The Importance of Digital Elevation Model Selection in Flood Simulation and a Proposed Method to Reduce DEM Errors: A Case Study in Shanghai," Int. J. Disaster Risk Sci., vol. 12, pp. 890–902, Nov. 2021. [Online]. Available: https://link.springer.com/article/10.1007/s13753-021-00377-z.

[16] P. L. Guth, S. Trevisani, C. H. Grohmann, J. Lindsay, D. Gesch, L. Hawker, and C. Bielski, "Ranking of 10 Global One-Arc-Second DEMs Reveals Limitations in Terrain Morphology Representation," Remote Sens., vol. 16, no. 17, Art. no. 3273, Sep. 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/17/3273.

[17] Z. Wang, J. Liu, J. Li, Y. Meng, Y. Pokhrel, and H. Zhang, "Basin-scale high-resolution extraction of drainage networks using 10-m Sentinel-2 imagery," Remote Sens. Environ., vol. 255, Art. no. 112281, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720306544.

[18] N. A. Muhadi, A. F. Abdullah, S. K. Bejo, M. R. Mahadi, and A. Mijic, "The Use of LiDAR-Derived DEM in Flood Applications: A Review," Remote Sens., vol. 12, no. 14, Art. no. 2308, Jul. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/14/2308.

[19] Y. Jia, X. Su, T. Zhang, H. Wu, Y. Jia, and Q. Wang, "Enhancing streamflow forecasting using an LSTM hybrid model with lightweight frequency-domain feature learning," Expert Syst. Appl., vol. 297, pt. B, Art. no. 129418, Feb. 2026. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0957417425030349.

[20] DataMapu, "Gradient Boosting Variants – Sklearn vs. XGBoost vs. LightGBM vs. CatBoost," DataMapu, May 8, 2024. [Online]. Available: https://datamapu.com/posts/classical_ml/gradient_boosting_variants/.

[21] National Weather Service, "About Models," Weather.gov, https://www.weather.gov/about/models.

[22] European Space Agency, "Copernicus DEM – Global and European Digital Elevation Model," Copernicus Data Space Ecosystem, https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM.

[23] European Space Agency, "Copernicus DEM – Global and European Digital Elevation Model," Copernicus Data Space Ecosystem, https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM.

[24] M. Meadows, K. Reinke, and S. Jones, "Explaining machine learning models trained to predict Copernicus DEM errors in different land cover environments," Artif. Intell. Geosci., vol. 6, no. 2, Art. no. 100141, Dec. 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666544125000371.