https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Open Access Earth System Science Data Discussions

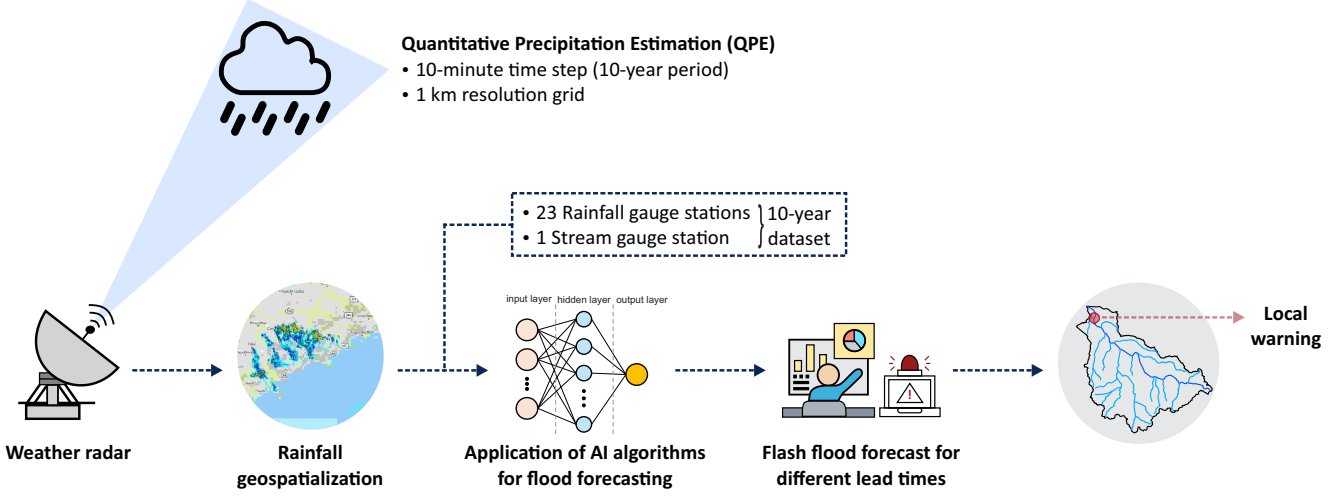
Ten years of hydrometeorological observations at 10-minute resolution and its application in machine learning hydrological models

Kleber L. Rocha-Filho[1], Lidiane S. Lima[1], Elton V. Escobar-Silva[1], Rafael M. P. Teixeira[1], Andrea S. Viteri López[1], Glauston R. T. Lima[1], Jaqueline A. J. P. Soares[1], Cristiano W. Eichholz[1], Flavio Conde[2], Carlos A. M. Rodriguez[3], Joaquin I. B. Garcia[4], and Leonardo B. L. Santos[1]

[1]National Center for Monitoring and Early Warning of Natural Disasters (Cemaden), São José dos Campos, SP, 12247-016, Brazil
[2]Hydraulics Technology Center Foundation (FTCH), São Paulo, SP, 05458-000, Brazil
[3]Institute of Astronomy, Geophysics, and Atmospheric Sciences, University of São Paulo (USP), São Paulo, SP, 05508-090, Brazil
[4]Department of Civil and Environmental Engineering, University of São Paulo (USP), São Paulo, SP, 05508-010, Brazil

Correspondence: Leonardo B. L. Santos (leonardo.santos@cemaden.gov.br)

GRAPHICAL ABSTRACT



Abstract. Accurate urban flash flood forecasting relies on well-spatialized rainfall data distribution. This study introduces and utilizes the TTI-HydroMet dataset, a publicly available and unique collection for the Tamanduateí River Watershed, in Sao Paulo (Brazil). The dataset includes rainfall measurements from 23 rain gauge stations, stage observations from a hydrological gauge near the outlet, and quantitative precipitation estimates at 1-km radar resolution, accumulated in 10-minute precipitation

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Open Access Earth System Science Data Discussions

fields over 10 years. The weather radar data presents missing values for only 0.3% of timestamps during rainfall events observed by rain gauges. The Spearman correlation coefficient between weather radar and rain gauges varies from 0.675 (full period) to 0.949 (a specific event). It was used to assess the predictive capacity of Machine Learning (ML) hydrological models trained on accumulated rainfall data from rain gauges and estimated by a weather radar. Using an advanced cross-validation framework, two representative algorithms (LinearSVR and XGBRegressor) were tested across different rainfall source configurations and showed strong performance at lead times up to 120 minutes. The Nash–Sutcliffe Efficiency index ranges from 0.781 to 0.996. The statistically comparable performance of ML models driven by radar and rain gauge rainfall indicates that radar-based ML approaches can represent a viable alternative for short-term stage forecasting in regions lacking rain gauge networks.

1 Introduction

Floods threaten lives, cause property damage, harm the environment, and disrupt economic and social activities (Lee et al., 2020). In this context, precipitation is a crucial variable and a major source of uncertainty in hydrological studies. When examining the relationship between precipitation and its impacts on urban areas, the complexity increases. This is mainly because urbanization, driven by human activities, increases soil impermeability, alters surface roughness, and alters precipitation patterns and intensity (Yang et al., 2024).

One of the main methods for recording precipitation data is using rain gauges, which provide point measurements, are cost-effective, and easy to install. However, their measurements are not precise enough to accurately represent precipitation across an entire watershed. They can have errors of 30% or more, depending on the type of instrument or local conditions. This challenge mainly arises from the intermittent nature of rain, its spatial and temporal variability, and its sensitivity to environmental factors (Van de Ven, 1990; Sokol et al., 2021). In general, some sources of error in rain gauge precipitation measurements include equipment malfunctions, systematic and random errors (instrumental errors), and the limited spatial coverage of point measurements (spatial sampling errors) (Ochoa-Rodriguez et al., 2019).

Weather radars serve as a potential alternative for quantitative precipitation estimation (QPE), offering high spatiotemporal resolution across large areas by emitting microwave pulses and analyzing the reflected and backscattered signals from raindrops, snow, and hail (Sokol et al., 2021). Early radar systems relied solely on reflectivity (Z) to measure precipitation; however, this method was affected by multiple sources of uncertainty (Doviak and Zrnic, 2014). More recently, the advent of dual-polarization radars and the use of variables such as differential reflectivity (Z_{DR}) and specific differential phase (K_{DP}) have reduced several QPE uncertainties, including those associated with drop size distribution (DSD), attenuation, and bright band contamination (Ryzhkov et al., 2022).

Weather radar data has been widely collected and utilized by national meteorological services worldwide for climatological and hydrological studies. However, most of these centers maintain archives dating only from the 2000s onward. Among the databases of 45 national centers analyzed by (Saltikoff et al., 2019), only 15 make their data available for research outside their own institutions. In Brazil, in addition to the scarcity of long-term radar data records, there are also various challenges related to data accessibility and heterogeneous formats.

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Beyond these challenges, weather radar records meet the key criteria for classifying data as big data. The concept of big data refers to datasets characterized by high volume, velocity, and variety, which require specialized computational tools for storage, processing, and analysis (Laney, 2001; Tang et al., 2022). Radar observations inherently satisfy these conditions: each scan produces large multidimensional fields with millions of data points, collected at short temporal intervals, resulting in rapidly expanding datasets over multi-year periods (Sokol et al., 2021). This complexity reinforces the need for advanced computational methods, such as machine learning (ML) algorithms, to extract hydrometeorological information and support operational applications.

Hydrodynamic models can be used to simulate realistic water flow by solving physics-based governing equations (Beven, 2012). They are often combined with hydrological models that focus on rainfall and watershed processes to provide a comprehensive understanding of floods and water systems. These models range from one-dimensional (1D) to three-dimensional (3D), each suitable for different levels of complexity and precision (Teng et al., 2017). However, their main disadvantages include high computational costs and significant data requirements for accurate setup and calibration (Kant et al., 2025; Zhu et al., 2025). On the other hand, ML techniques are increasingly being integrated into hydrological modeling to enhance predictions and address the limitations of traditional methods (Hasan et al., 2024). These approaches are typically categorized as either purely data-driven or hybrid. Data-driven models learn complex relationships between hydrological variables (e.g., rainfall, temperature, discharge) directly from data without explicitly modeling physical processes. In contrast, hybrid hydrological models, which combine ML with physics-based models, aim to utilize the strengths of both physics-based (conceptual or physically-based) and data-driven ML approaches (Chadalawada et al., 2020; Santos et al., 2025).

In Brazil, various efforts have been made in hydrological research. For example, (Amorim et al., 2022) evaluated the performance of a hydrological model in estimating runoff using distributed rainfall data applied to an urban watershed with macro drainage structures. (Hossoda et al., 2025) created an innovative flood warning system for an urban watershed, utilizing parametric and ML models. (Escobar-Silva et al., 2023) assessed the performance of HEC-RAS, a hydrodynamic model, in identifying flood-prone areas using two digital terrain models (DTMs) with different spatial resolutions. (Viteri López and Morales Rodriguez, 2020) presented a flash flood forecasting model that uses binary logistic regression to predict flash flood events in various urban watersheds within São Paulo.

This paper presents the TTI-HydroMet dataset (Escobar-Silva et al., 2025). This unique collection comprises 10 years of rainfall data from 23 rain gauges, river level data from a hydrological station, and weather radar QPE data at 10-minute temporal and 1-km spatial resolutions. To the best of our knowledge, no comparable large-scale dataset with such high temporal and spatial resolution is publicly available in the literature in such a context. Using this dataset, we compare hydrological models trained with different ML algorithms, all of which use rain gauge or radar data. We provide both a curated open dataset and a benchmark set of modeling experiments that can support the development, comparison, and operational products of ML-based flash flood forecasting in densely urbanized watersheds.

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Open Access Earth System Science Data Discussions

2 Materials and Methods

The study area is the Tamanduateí River basin (TRW), located within the São Paulo Metropolitan Region, Brazil. This region is the country's top economic zone, with the highest population concentration in Brazil, surpassing 20.6 million people (IBGE, 2023). TRW covers a drainage area of 330.41 km^2, with its main river, the Tamanduateí River, flowing for 36.5 km (Figure 1). It is a heavily urbanized sub-basin of the Tietê Upstream River (Alto Tietê) Watershed. Over 80% of TRW is impermeable, leading to frequent flooding along the riverbanks (Pereira Filho and dos Santos, 2006). The watershed also often faces extreme rainfall events (Escobar-Silva et al., 2023). In addition, the study area has a humid subtropical climate (Cwa), characterized by mild, rainy summers and moderate, dry winters, according to the Köppen climate classification (Beck et al., 2018). Lastly, the average annual temperature is 19.5°C, with July being the coldest month and February the warmest. The mean annual rainfall is about 1,500 mm.

The rainfall data used in this study were collected from 23 tipping-bucket automatic rain gauges with a sampling resolution of 0.2 mm. Conversely, the stage data were obtained from an automatic station. It is important to note that all data – rainfall and stage – were recorded at 10-minute intervals, and Station 413 collects both rainfall and stage measurements (Figure 1 (C)). The stations are part of the Alto Tietê telemetric network and are managed by the São Paulo Flood Warning System (Barros et al., 2016). Figure 1 (C) shows the spatial distribution of the stations.

Precipitation estimates were obtained from a Dual Polarization S-band Doppler weather radar (SPOL). The SPOL, located in the eastern region of the state of São Paulo, covers an area of approximately 181,000 km^2 (Figure 1 (B)) and performs volumetric scans in two operational modes: surveillance mode with two elevation angles every 15 minutes, and an operational mode with seven elevation angles every 5 minutes. The precipitation fields at a 1 × 1 km spatial resolution were generated using the DPSRI (Dual Polarization Surface Rainfall Intensity) algorithm, where the variables Z, Z_{DR}, and K_{DP} were used as described in (Ryzhkov et al., 2005). Lastly, 381003 estimated rainfall fields were summed over 10-minute intervals, assuming steady precipitation between scans.

To compare precipitation estimates from the weather radar with in situ measurements from rain gauge stations, both datasets' temporal resolutions were first standardized to regular 10-minute intervals, ensuring consistent temporal alignment. The metric adopted to represent radar-derived rainfall was based on the global mean of all radar cells within the area of interest, totaling 434 spatial units. Subsequently, the two time series, global radar and rain gauges, were merged into a single time-indexed dataset, enabling direct comparison of average rainfall values. The agreement between the sources was evaluated using the Spearman correlation coefficient, with the corresponding p-value, applied to the complete series and to the time steps with rainfall greater than 0.2 mm. For illustrative purposes, an intense rainfall event is also shown (February 14, 2024). This analysis primarily aimed to validate the radar's accuracy as a supplementary data source for hydrological applications and monitoring extreme precipitation events (Wang et al., 2015).

In preparing the hydrological dataset, the stage time series was preprocessed to remove outliers and smooth non-physical fluctuations caused by sensor malfunction and/or environmental disturbances. Outlier detection was performed using a Hampel identifier (window size set to 10 and $n_\sigma = 2.5$), which flags measurements exceeding a median-based deviation threshold within

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

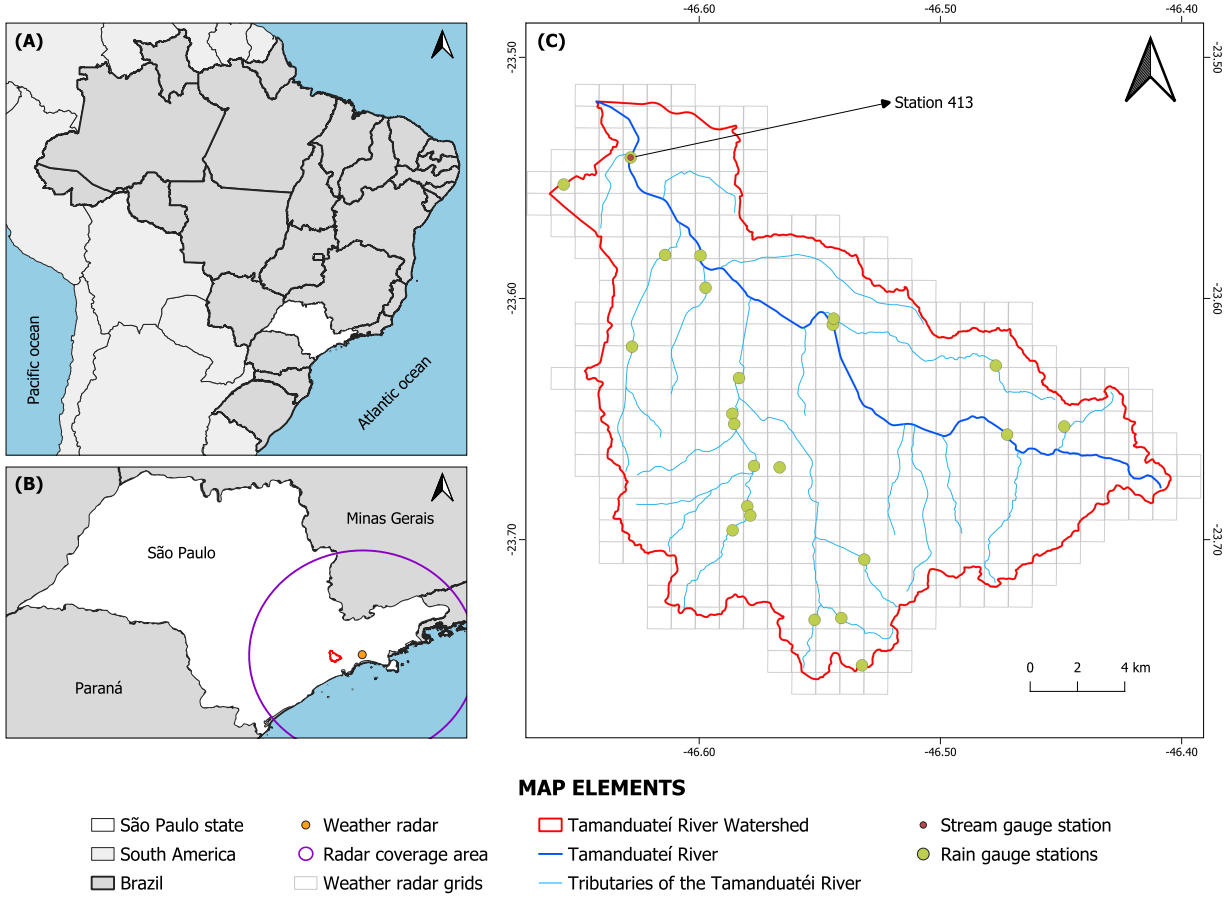


Figure 1. Location of the study area, the Tamanduateí River Watershed (TRW), within the São Paulo Metropolitan Region, Brazil. (B) shows the locations of TRW and the weather radar, along with the radar's coverage area. (C) illustrates the distribution of the 23 rain gauge stations and one stream gauge station within TRW, as well as the 1 km weather radar grids covering TRW. Station 413, the closest station to the watershed outlet, collects both rainfall and stage data.

a sliding window and replaces them with the local median (Hampel, 1974; Liu et al., 2004). Subsequently, a custom filtering procedure was applied to suppress high-frequency noise. A low-pass finite impulse response filter was used, with the cutoff frequency set to 20% of the Nyquist frequency, employing a Hann temporal window and achieving zero-phase distortion by means of a forward-reverse scheme (Gustafsson, 1996), which cancels the delays typically introduced by causal filters and preserves the timing of hydrograph peaks. To mitigate boundary transients and ringing artifacts, the signal was processed with a sliding window of size 500 with 10% reflection padding. The filtering procedure was applied twice: first to identify remaining outliers relative to the locally smoothed signal, and then again – after removing these outliers – to obtain the final time series.

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

To conduct the ML experiments, this study employed a dedicated framework that manages the entire modeling pipeline for flash flood forecasting. This framework, called ML4FF (Soares et al., 2025), provides an integrated workflow that includes data partitioning, model training, hyperparameter tuning, and performance assessment across multiple forecast lead times. In this study, two representative methods from the available framework were selected for testing: Linear Support Vector Regression (LinearSVR) and the eXtreme Gradient Boosting Regressor (XGBRegressor). In this paper, ML experiments are presented for functional validation of the dataset, not for exhaustive algorithmic comparison. For each method and lead time combination, the framework follows a workflow divided into two phases: a training-validation-test phase and a holdout assessment phase. Initially, the complete time series is divided into two mutually exclusive subsets, one for each phase. The first phase employs a nested cross-validation scheme consisting of 12 outer folds and 6 inner folds, and uses 87.5% of the data. The inner loops use Bayesian optimization to explore different hyperparameter settings, while the outer loops estimate model performance and computational cost across varying data partitions. The second phase is dedicated to final evaluation using the unseen 12.5% of the data, which helps ensure accurate estimates of how the models will perform in real-world forecasting scenarios.

In the two ML approaches based on gauge stations, accumulated rainfall data were collected from all 23 automatic rain gauges across the TRW. Combined with stage measurements from Station 413 (the gauge closest to the watershed outlet), this yielded 24 synchronized time series. To enable direct comparison with the approaches based on accumulated rainfall derived from radar fields, all 24 series were aligned to a common timeline defined by the overlapping intervals between the stage and radar data. These time series were then used as input feature vectors to form datasets for training, validation, and test the LinearSVR and XGBRegressor models, with the target variable being the stage at a specified forecast lead time. Hence, the model output corresponded to the predicted stage at each lead time (10, 60, and 120 min).

Similarly, in the radar-based setup, radar-derived accumulated rainfall at each SPOL radar pixel (434 pixels covering the watershed), combined with stage measurements, was used to build input datasets for training, validation, and test the two ML models considered in this study (LinearSVR and XGBRegressor). The resulting 435 time series were aligned using the same timeline as in the rain-gauge-based setup. These series were used as the input feature vectors for the radar-based models, which generate stage predictions for the same set of forecast lead times.

Model performance was evaluated using three standard hydrological metrics: the Nash–Sutcliffe Efficiency (NSE), the Kling–Gupta Efficiency (KGE), and the Root Mean Square Error (RMSE), following the comparison framework described in (Soares et al., 2025). NSE quantifies how well predictions reproduce observed variability; values close to 1 indicate higher skill. KGE incorporates correlation, bias, and variability components into a single metric, providing a balanced assessment of predictive performance. RMSE measures the average magnitude of the prediction error, has the same units as the target variable, and is often interpreted in relation to the standard deviation of the observations to contextualize its magnitude.

The TTI-HydroMet dataset (Escobar-Silva et al., 2025) is an original open dataset containing 10 years of hydrometeorological data from the Tamanduateí River Watershed (TRW) in São Paulo, Brazil. Rainfall data from 23 gauge stations and stage measurements from one station were collected every 10 minutes. QPE is provided at a 1 km spatial resolution and updated every 10 minutes. Additionally, the codes used and generated in this work are also available in the dataset. Lastly, it is important to note that computational procedures were performed on a central processing unit (CPU) with an i9-12900KF (Intel, Santa

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Clara, CA, USA), a PRO Z790-P WIFI (MSI, Zhonghe, Taiwan), and a GeForce RTX 4090 24GB (NVIDIA, Santa Clara, CA, USA).

3 Results and Discussion

The analysis of the radar database, covering the period from April 1, 2015, to March 29, 2025, showed an overall data gap rate of 69.3%. At first glance, this number might suggest operational failures; however, it results directly from the sensor's acquisition strategy. Radar data are collected every 5 minutes during significant precipitation events within the coverage area and every 15 minutes during surveillance mode (when no rain is observed). This characteristic matches the rainfall pattern of the study area, as shown by the clear seasonality observed in the time gaps in the data (Marcuzzo, 2020). Data absence is more noticeable during the winter months, such as July (87.3%) and August (88.0%), which are usually dry periods. In contrast, during the summer months of January (43.4%) and February (47.3%), the missing data rate is significantly lower, reflecting increased convective activity and increased rainfall frequency (Minuzzi et al., 2007). Additionally, analysis of data gap durations reveals that most (73.5%) are brief interruptions lasting 10 minutes, while 15.2% correspond to extended periods exceeding one continuous hour.

To assess the integrity of the radar database under hydrologically relevant conditions, a conditional analysis was performed, focusing only on time intervals with precipitation (> 0.2 mm). These events represent 8.6% of the total time series (45,093 out of 525,600 records), and within this subset, the missing-data rate was only 0.3%. This result demonstrates the radar's high reliability during precipitation events, highlighting its importance for monitoring weather conditions with potential impacts (Yoon and Lim, 2022). The high data availability during these crucial moments demonstrates the robustness of the sensor for applications in natural disaster studies, especially in detecting and spatially-temporally characterizing moderate to heavy rainfall events (IPCC, 2023; Haddad and Teixeira, 2015). This methodological robustness is the key to creating integrated databases that support predictive models and hydrological risk assessments.

The consistency between radar-derived accumulated rainfall and in situ rain gauge records was statistically verified using the Spearman correlation coefficient, which yielded $\rho = 0.604$ ($p < 0.001$) for the complete time series and $\rho = 0.675$ ($p < 0.001$) when considering only time steps with rainfall > 0.2 mm). These values indicate a moderate positive monotonic relationship, confirming the radar as a highly relevant complementary data source for the composition of geospatial natural hazard databases. Figure 2 illustrates this relationship: panel (a) displays the complete time series (01/04/2015–29/03/2025) of the average accumulated rainfall from radar and recorded by rain gauges, while panel (b) highlights the intense event of February 14, 2024, used as a case study, which showed a correlation of $\rho = 0.949$ ($p < 0.001$) between radar-based rainfall accumulations (QPE) and gauge stations. During this event, radar performance remained closely aligned with ground-based observations, indicating that the sensor maintained data integrity even under high-severity conditions. These findings strengthen the case for the integrated use of remote sensing technologies to support early warning systems and mitigation strategies in regions vulnerable to extreme events (Doswell III et al., 1996).

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

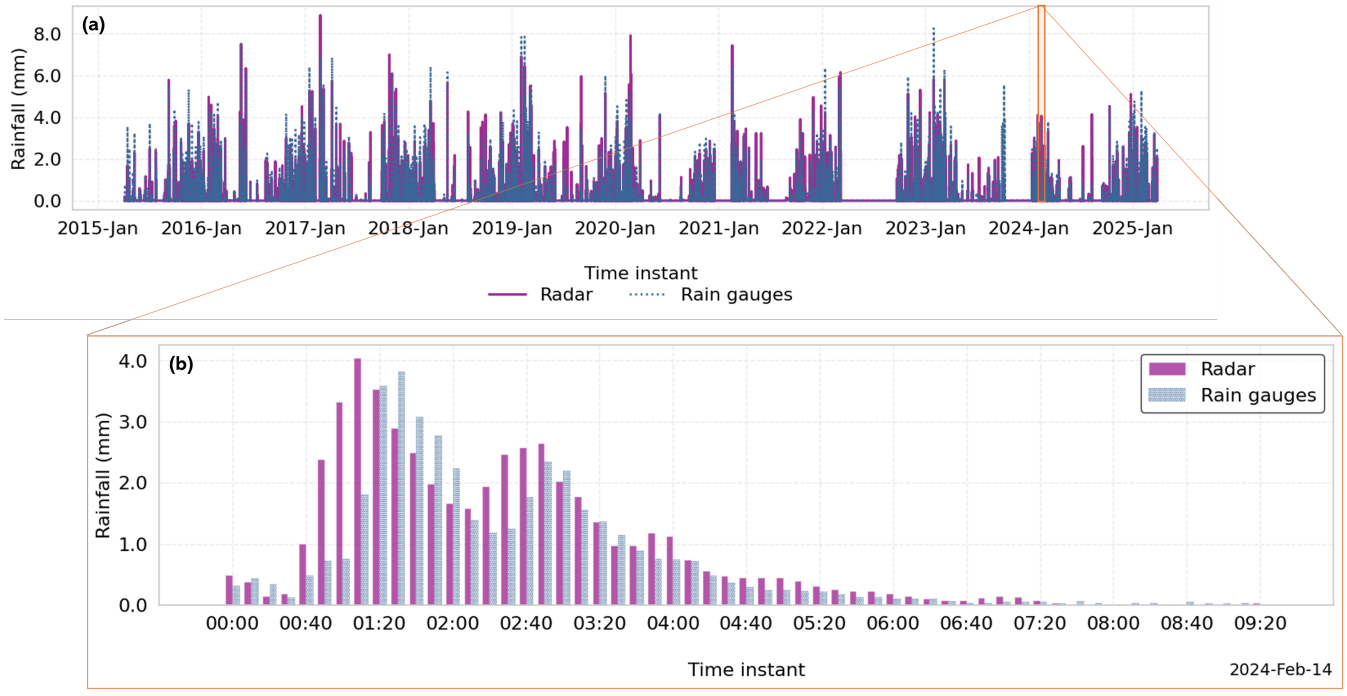


Figure 2. Comparison between average quantitative precipitation estimation (QPE) derived from radar (Radar – average of all cells, in purple) and the average records from rain gauge stations (Rain gauges, in blue dashed lines), with a temporal resolution of 10 minutes, over the period from 01/04/2015 to 29/03/2025. Panel (a) shows the complete time series (2015–2025), highlighting the seasonal correspondence and overall coherence between the two data sources. Panel (b) provides a more detailed view of the rainfall event on February 14, 2024, characterized by moderate to heavy precipitation. Statistical analysis revealed a Spearman coefficient of $\rho = 0.604$ ($p < 0.001$) for the complete time series, $\rho = 0.675$ ($p < 0.001$) when considering only time steps with rainfall above 0.2 mm, and $\rho = 0.949$ ($p < 0.001$) for the highlighted event.

Figure 3 shows the distribution of the logarithm of the Mean Field Bias (log(MFB)) as a function of daily rainfall intensity (mm/day), considering only events with accumulated rainfall greater than 0.2 mm). The calculation of log(MFB) followed a methodology similar to that proposed by (He et al., 2013), serving as a metric to quantify the systematic deviation between radar-derived accumulated rainfall and rain gauge records. The values are mostly concentrated around zero, indicating generally good agreement between the two data sources and an approximately symmetrical distribution. However, dispersion increases with rainfall intensity, especially for events exceeding 30 mm)/day, reflecting greater uncertainty in radar-based estimates under extreme conditions, a behavior widely recognized in the literature. Compared to the results reported by (He et al., 2013), the distribution in the present study is slightly wider, with log(MFB) values ranging up to approximately ± 1.5. In contrast, the visual bounds in the reference study were close to ± 1.0. This difference may be related to greater spatial heterogeneity within the basin under analysis, a higher frequency of extreme events, or the influence of local factors such as orographic blocking. The pronounced variability at the distribution extremes also suggests that, despite good average performance, radar can be significantly underestimated or overestimated in certain rainfall scenarios, reinforcing the importance of integrated validation

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Open Access Earth System Science Data Discussions

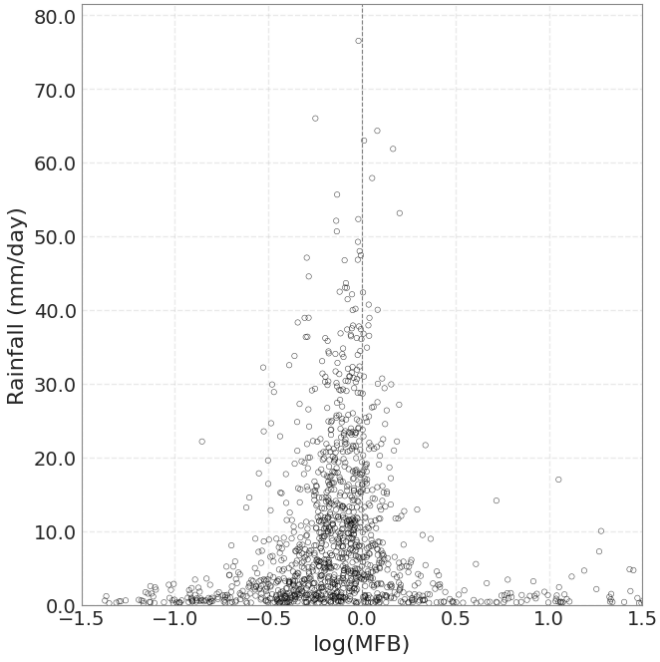


Figure 3. Logarithmic distribution of the Mean Field Bias (log(MFB)) as a function of daily precipitation intensity. Data were filtered to include only days with rainfall greater than 0.2 mm/day. The dashed vertical line indicates log(MFB) = 0, corresponding to the absence of systematic bias between radar and rain gauge data.

approaches. As highlighted by (He et al., 2013), understanding how bias evolves at different rainfall intensities is essential to improve the assimilation of radar data in operational hydrological models.

The ML4FF framework was used to obtain optimized LinearSVR and XGBRegressor models for lead times of 10, 60, and 120 minutes. To assess their performance under unseen conditions, the complete predicted stage series was analyzed using the holdout dataset. In comparison with the dataset used in the training-validation-test phase, with $N = 138128$ stage measurements totaling 959 days within a 3107-day calendar period, the holdout dataset has $N = 19733$ stage measurements totaling 137 days (546-day calendar period). The stage distributions for both datasets show very similar statistical properties, as confirmed by the Wasserstein distance (W), which measures the minimal transport cost needed to convert one distribution into the other. This metric offers a robust and understandable way to compare empirical distributions and avoids the sensitivity problems that affect hypothesis-testing methods when N is large. Using the standard dimensionless form with a significance level of $\alpha = 0.05$ and applying common scaling factors to normalize both datasets, the resulting value $W = 0.007 < \alpha$ indicates that the two-stage distributions are very similar in overall shape. This supports the conclusion that the holdout dataset provided a statistically representative basis for evaluating the trained ML models under unseen operating conditions.

Motivated by previous findings, particular attention was given to the same intense event on February 14, 2024. Along with the confirmed radar-rain gauge agreement during this episode, it is one of the few major occurrences in the holdout period that

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

is representative of a typical flash flood, exhibiting a clear two-phase stage response: a rapid rise – surpassing the 722.779 m threshold (set by the local authorities) that signals flood-warning conditions – driven by intense rainfall, followed by a slower recession phase. Figure 4 shows the predicted stage time series along with the observed values, allowing a direct comparison of how rain gauge- and radar-based models perform during intense rainfall conditions (time interval marked by vertical dashed lines). The two models successfully replicated the rapid increase of nearly 7 m during the stage rise, and – except for the radar-based XGBRegressor in panel (d) – all predictions exceed the warning threshold (horizontal dashed line), indicating that an operational alert would be triggered. Since the models were trained to forecast using only information available at the current time, the predicted series naturally exhibit some lag and occasional sharp rises, as they associate future stage increases with the rainfall signals available at the moment of prediction. For LinearSVR models, panels (a) and (c), this behavior appears as spike-like predictions whose amplitude and timing vary with lead times, with each peak at time t indicating the model's dependence on rainfall at $t - t_{(\text{lead time})}$. Rain gauge-based XGBRegressor models, panel (b), replicate the overall rising pattern with fewer abrupt changes and provide a closer match near the stage maximum. Radar-based XGBRegressor models, panel (d), also follow the general trend. However, their predictions tend to smooth out the maximum region, resulting in plateau-like values that remain slightly below the warning threshold.

An analysis of a one-day time series (Figure 4) shows consistent differences between the predicted and observed stage time series, particularly in the timing and magnitude of the peak stage. These discrepancies depend on the precipitation input type (rain gauges or radar) and the ML algorithm used (LinearSVR or XGBRegressor). For the shortest lead time (10 min), all models accurately detect the start of the hydrograph's rise, except for the radar-based XGBRegressor, which shows a slight positive lag. For longer lead times (60 and 120 min), a consistent positive lag emerges across all configurations, regardless of precipitation source or ML method. This pattern arises from the models' dependence on past rainfall to predict future stages; with radar inputs, it is further intensified by the high spatial persistence of radar rainfall fields, which tends to smooth out sudden temporal changes. The combined effects of this smoothing and reliance on past rainfall result in a delayed response. This behavior is well documented in the literature, as radar-derived rainfall estimates are affected by structural uncertainties – such as volumetric sampling, bright-band contamination, noise, and range-dependent corrections – that can alter the temporal pattern of precipitation and, consequently, the timing of runoff (He et al., 2013). In this study, these radar-related effects are especially apparent in the outputs of the XGBRegressor, which show flattened (plateau-like) peaks and, in some instances (e.g., Fig. 4d), slight underestimations of the maximum stage. Conversely, models based on rain gauge measurements tend to produce sharper peaks (reflecting the higher temporal variability of point-scale observations) and may occasionally overestimate the peak stage, with this effect being most evident in the LinearSVR.

During the recession phase of the hydrograph, all models consistently reproduced the stage decline, reinforcing their applicability despite the structural peculiarities of the precipitation inputs. LinearSVR models (for both radar and rain gauge data) generated smooth recession curves. In contrast, the XGBRegressor produced a more stepwise decay pattern, which is consistent with the piecewise-constant behavior inherent to decision-tree-based algorithms. This effect is particularly evident in radar-driven simulations because radar rainfall fields exhibit strong spatial and temporal persistence; when consecutive time steps contain very similar rainfall patterns, the model tends to follow the same decision pathways, resulting in nearly identical

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

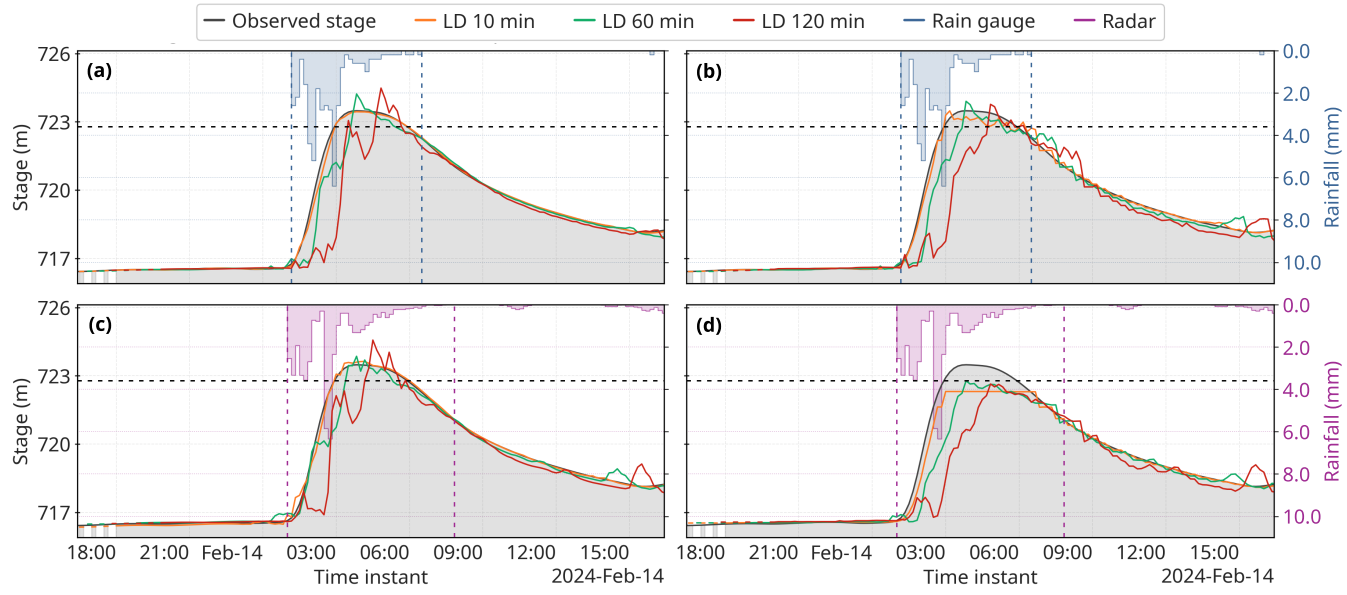


Figure 4. Stage time series comparisons over a 1-day period centered on the warning event that occurred on February 14, 2024. This example, during the holdout period, shows observed values alongside predictions from rain-gauge- and radar-based ML models trained to forecast lead times (LDs) of 10, 60, and 120 minutes. The plots provide a detailed view of each model's predictive ability, evaluated only during periods when both stage and radar data are available (gray-shaded areas). Panels (a) and (b) present results for the rain gauge-based LinearSVR and XGBRegressor models, respectively. In contrast, panels (c) and (d) show the corresponding results for the radar-based LinearSVR and XGBRegressor models. The black dashed line indicates the warning threshold of 722.779 m, and the vertical dashed lines in panels (a)-(b)/(c)-(d) mark the beginning and end of the event as determined from the gauge/radar rainfall data.

stage predictions. In contrast, rain gauge inputs, which exhibit greater temporal variability at the point scale, lead to greater fluctuations in the predicted hydrograph. These differences are most apparent during peak attenuation, yet the consistent recession behavior across all models and data sources highlights their reliability in capturing the falling limb of flood events.

To further evaluate the agreement between predicted and observed stage values, scatter plots are shown for both rain gauge- and radar-based models across all lead times (Figure 5). Overall, the results show distributions within 1 meter of the 1:1 line (shaded area), indicating that the predictions are closely aligned with the observed series. A closer examination of these distributions reveals key aspects of how both types of ML models perform: (i) for stages ≤ 720 m, the predictive skill of the models gradually declines as the lead time increases, as indicated by the increase in the dispersion of points in the scatter plots, primarily reflecting an underestimate; (ii) for stages > 720 m, this behavior is less apparent, and the predictions stay more consistent. In this range, LinearSVR models, panels (a) and (c), perform better at shorter lead times (10 and 60 minutes) but display larger deviations for longer horizons (120 minutes), where the XGBRegressor models, panels (b) and (d), show comparatively superior performance; (iii) LinearSVR models are more likely to overestimate the stage when compared with XGBRegressor models, as previously indicated by the warning event time series analysis; and (iv) the scatter plot (d) also

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

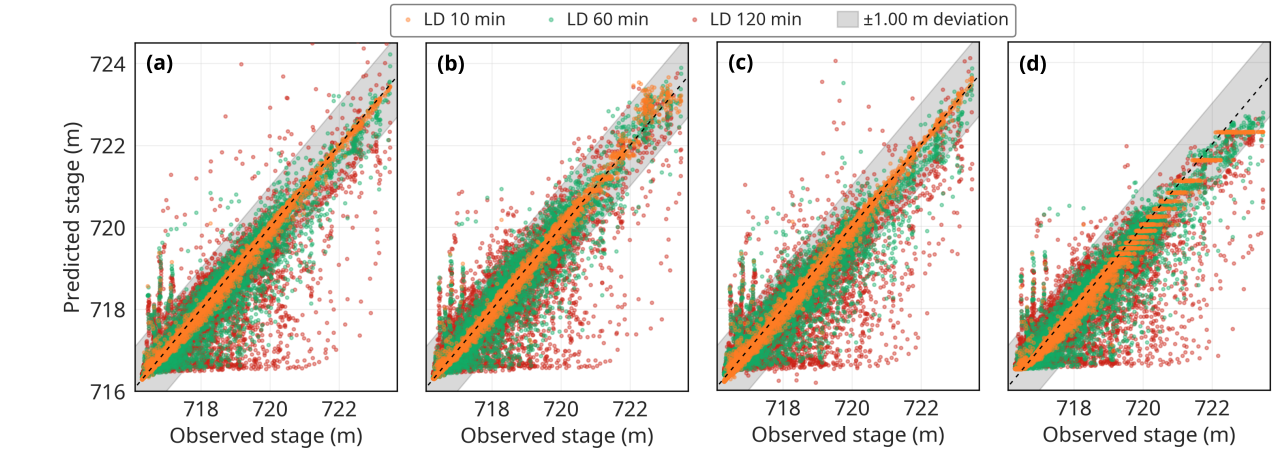


Figure 5. Scatter plots of predicted versus observed stage values within the holdout period. Panels (a) and (b) show results for the rain gauge-based LinearSVR and XGBRegressor ML models, respectively, while panels (c) and (d) display the corresponding results for the radar-based LinearSVR and XGBRegressor models. Results are displayed for models trained to forecast lead times (LDs) of 10, 60, and 120 minutes. The gray shaded band represents a ± 1.0 m deviation from the black dashed 1:1 reference line.

exhibits plateau-like structures, similar to those in Figure 4(d). Still, these do not correspond to large deviations, as they fall within the 1-meter error band in most instances.

Previous observations are supported quantitatively by the coefficient of determination (R^2), calculated for each model based on the observed stages (Table 1). The consistently high R^2 values demonstrate strong predictive ability for both the LinearSVR and XGBRegressor models. The hypothesis tests produce very small p-values ($< 10^{-6}$), even when repeated on much smaller random subsets of the data, simply showing that the correlation between predicted and observed stages is significantly different from zero. However, such statistical significance provides little insight in large datasets. In this case, the effect size (R^2) is more meaningful, as it directly indicates the strength of the linear relationship and better characterizes model performance.

Table 1. Coefficient of determination (R^2) for rain gauge- and radar-based ML models (LinearSVR and XGBRegressor) assessed from predicted versus observed values during the holdout period. Results are presented for models trained to forecast lead times of 10, 60, and 120 minutes.

Score type	LinearSVR				XGBRegressor			
	10 min	60 min	120 min	p-value	10 min	60 min	120 min	p-value
R^2 – Rain gauge	0.996	0.933	0.783	$< 10^{-6}$	0.995	0.939	0.809	$< 10^{-6}$
R^2 – Radar	0.992	0.926	0.788	$< 10^{-6}$	0.990	0.921	0.780	$< 10^{-6}$

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

The stage-duration curves for observed and predicted stages at lead times of 10, 60, and 120 min (Figure 6) show the proportion of time each stage level is reached or exceeded, providing an integrated view of how well the models capture the entire stage distribution. Predictions were generated using two ML models (LinearSVR and XGBRegressor) along with rainfall data from both rain gauges and radar-derived accumulations. The 5% exceedance stage (≈ 719.6 m) corresponds to relatively high and infrequent conditions, whereas the 95% exceedance stage (≈ 716.5 m) represents typical levels for normal low-stage periods. For lead times of 10 and 60 minutes, both ML models closely follow the observed curve across nearly all exceedance probabilities, with median deviations typically below 6 cm and maximum differences between 0.15 and 0.80 m, indicating that up to 60-minute forecasts preserve the overall statistical structure of the stage distribution. At a lead time of 120 min, deviations occur mainly at low exceedance probabilities (<10%), suggesting a slight underestimation of medium- to high-stage conditions. Median deviations increase from approximately 2.5-7 cm, with a maximum deviation of about 1.3 m (LinearSVR – rain gauge), but the models remain consistent for the more frequently occurring medium and low stages. These findings are consistent with the patterns observed in Figure 5, which show that, as the lead time increases, errors become more frequent in the moderate stage range – where greater dispersion and underestimation were also observed in the scatter plots – while predictions remain comparatively stable under the highest stage conditions. Across all lead times, the curves based on radar and gauge rainfall data remain largely consistent. Overall, the results indicate that both rainfall sources provide reliable predictions, with only a slight decrease in performance over longer lead times.

The quantitative evaluation of model performance was performed using the Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), and root mean square error (RMSE) metrics (Table 2). Rain gauge- and radar-based models show similar performance across all lead times, although rain gauge-based models consistently achieve slightly higher scores. The table also shows the CPU time required during training, validation, and test, indicating that XGBRegressor models are much more computationally intensive than LinearSVR models. Still, both models are less expensive than traditional physical models. Additionally, radar-based models incur costs nearly 10 times higher than their rain gauge-based counterparts – an expected result given the large difference in input dimensionality (23 rain gauges versus 434 radar cells). Notably, despite differences in training costs, the prediction time across the entire holdout dataset was about 0.1 s for all models. Therefore, from an operational perspective, the most time-consuming part of the workflow is obtaining stage and rainfall data from rain gauge and/or radar sources, as well as training the algorithm, since predicting stage requires minimal computational effort.

The Mann–Whitney U tests on the NSE, KGE, and RMSE distributions from the cross-validation partitions confirmed that forecast lead time was the most influential factor in model performance. The Mann–Whitney U test is a non-parametric test that evaluates whether two independent samples differ statistically by comparing the rankings of their sample values. All pairwise comparisons between 10, 60 and 120 minutes showed very low p-values for RMSE, NSE, and KGE (mean p-value ≈ 0.0007). The median RMSE gradually increased with longer lead times, while NSE and KGE values decreased (e.g., the NSE for the XGBRegressor-Radar dropped from 0.98 to 0.39 and for the LinearSVR-Rain Gauge from 0.99 to 0.35). This pattern, consistent across all scenarios, indicates a steady decline in predictive accuracy as the lead time extends.

Both rainfall data sources (rain gauges and radar) produced satisfactory results, and it can be concluded that the choice of input source does not significantly affect predictive model performance. However, minor differences were observed between

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

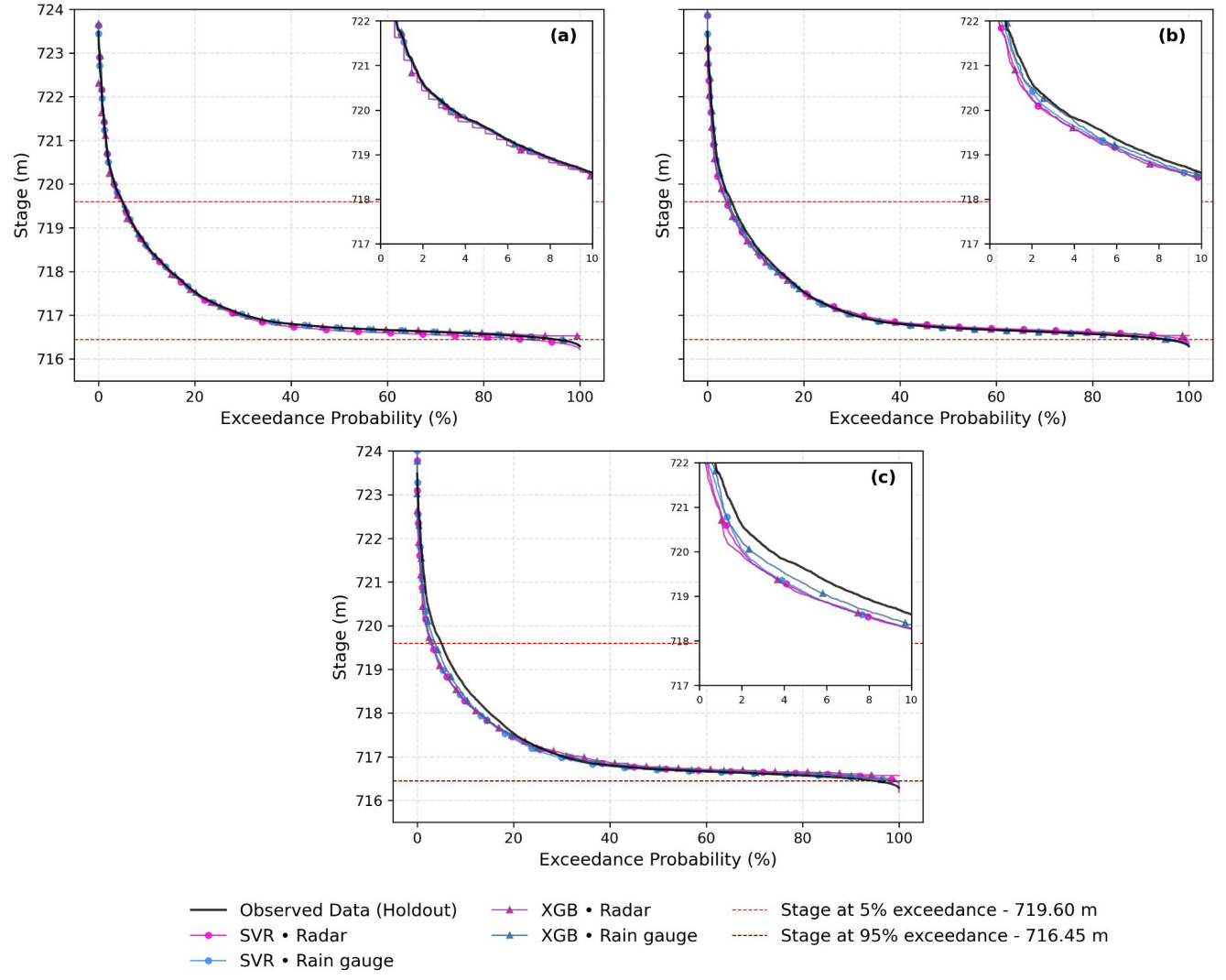


Figure 6. Comparison of stage duration curves derived from holdout data (black) and predicted stages at lead times of (a) 10, (b) 60, and (c) 120 minutes. Predictions were generated using the LinearSVR and XGBRegressor models with rainfall inputs from rain gauges (blue) and radar (purple). The red and brown dashed horizontal lines indicate the 5th and 95th percentiles of the observed stage distribution, respectively. Insets show a zoomed-in view of the low-exceedance region (0–10%).

the rainfall data sources. Rain gauge-based models generally yielded slightly higher NSE and KGE scores and lower RMSE compared to radar-based models, except for the RMSE of LinearSVR at a lead time of 120 minutes. It is important to note that KGE is a more balanced metric that considers correlation, bias, and variability (or the ratio of coefficients of variation). In contrast, NSE depends on the expected squared error.

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Open Access Earth System Science Data Discussions

Table 2. Overview of the metrics Nash–Sutcliffe Efficiency (NSE), Kling–Gupta Efficiency (KGE) and Root Mean Square Error (RMSE) for the holdout assessment phase. The CPU-time during the training and cross-validation phase is also presented. The results are shown for rain gauge- and radar-based ML LinearSVR and XGBRegressor models trained for forecasting lead times of 10, 60 and 120 minutes.

Score type	**LinearSVR**			**XGBRegressor**		
	10 min	60 min	120 min	10 min	60 min	120 min
NSE – Rain Gauge	0.996	0.933	0.783	0.995	0.939	0.809
NSE – Radar	0.992	0.926	0.788	0.990	0.921	0.781
KGE – Rain Gauge	0.990	0.934	0.830	0.990	0.961	0.857
KGE – Radar	0.989	0.903	0.806	0.951	0.893	0.776
RMSE (m) – Rain Gauge	0.070	0.279	0.500	0.080	0.266	0.469
RMSE (m) – Radar	0.094	0.291	0.494	0.105	0.301	0.503
CPU time (s) – Rain Gauge	242	108	81	2166	3128	3133
CPU time (s) – Radar	3895	3599	3469	32515	32416	32373

The comparison of the ML algorithms also showed only minor differences, suggesting that the choice of algorithm does not significantly affect the forecasts. Regarding RMSE, rain gauge-based LinearSVR forecasts had lower values for a 10-minute lead time and higher values for 60- and 120-minute lead times compared to XGBRegressor forecasts, while radar-based LinearSVR forecasts had lower values for all lead times. Concerning NSE, rain gauge-based LinearSVR forecasts had higher NSE values for a 10-minute lead time and lower values for 60- and 120-minute lead times compared to XGBRegressor forecasts, whereas radar-based LinearSVR forecasts had higher NSE values for all lead times. Finally, for KGE, rain gauge-based LinearSVR forecasts perform similarly at 10-minute lead times as XGBRegressor forecasts but produced lower KGE for 60- and 120-minute lead times; however, radar-based LinearSVR forecasts consistently outperformed XGBRegressor.

This behavior contrasts with most traditional rainfall–runoff modeling studies, in which rain gauge data generally outperform raw radar estimates unless the latter are bias-corrected or merged with gauge observations. This is illustrated by two types of rainfall-runoff simulations conducted with the HEC-HMS model using rain gauge and NEXRAD data, in which rain gauge-based precipitation produced runoff simulations with R^2 = 0.88 and 0.87, MFB-corrected radar rainfall achieved R^2 = 0.78 and 0.68, and radar-only inputs resulted in R^2 = 0.75 and 0.66 (Ahmed et al., 2022). Other works demonstrate that radar can add hydrological value when properly processed, such as the MIKE SHE groundwater simulations by (He et al., 2013), which reported approximately 5% reductions in RMSE, the debris-flow early-warning study by (Bernard and Gregoretti, 2021), and the physics-based flash flood modeling by (Looper and Vieux, 2012). Neural network models using radar have also achieved high performance; for example, the ANN model of (Santos et al., 2023) achieved NSE > 0.85 with 12-h accumulated radar

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

input, confirming the utility of radar in ML hydrology when high-resolution observations are available. Although these studies highlight the benefits of radar, only a few explicitly compare rain gauge and radar rainfall as alternative inputs to the same hydrological model, notably (He et al., 2013; Bernard and Gregoretti, 2021; Ahmed et al., 2022). None of these studies evaluates this comparison within a short-term ML forecasting framework with multiple lead times.

4 Conclusions

This paper presents the TTI-HydroMet dataset, an original dataset with 10-minute temporal resolution, 1 km radar spatial resolution, and observations of 23 rain gauges over 10 years (Escobar-Silva et al., 2025). Furthermore, this study used the TTI-HydroMet dataset with two ML algorithms (LinearSVR and XGBRegressor) for hydrological modeling.

Evidence that radar data can serve as a reliable supplementary input for hydrological modeling emerged from complementary quantitative analyses. Specifically, radar rainfall estimates exhibited high temporal continuity, with missing records accounting for 0.3% of timestamps during rainfall events observed by rain gauges (> 0.2 mm per 10 min). Additionally, the consistency between radar and gauge rainfall magnitudes was assessed quantitatively, with agreement reflected in moderate correlations computed over all rainfall periods meeting this threshold, and especially in the strong correlations observed during high-intensity events, such as the episode of 14 February 2024.

Scatter-plot analyses showed that predictions from both ML models mostly stayed within ± 1 m of observations, with a slight decline in accuracy as lead time increased (60 and 120 m). This decline was more apparent at lower stages (< 720 m), where underestimation became more common. For higher, hydrologically critical stages (> 720 m), predictions showed a small reduction in dispersion and mostly remained within the ± 1 m range across all lead times. These patterns were consistent with the high R^2 values obtained (0.78–0.996).

Stage-duration analyses revealed that both ML models, driven by radar or rain gauge rainfall, accurately reproduced the overall stage distribution for lead times up to 60 minutes. Deviations increased only at 120 minutes, mainly at low exceedance probabilities, indicating a slight underestimation of medium-to-high stages. Despite this decline, predictions remained stable under the highest stage conditions, and the two rainfall inputs produced highly consistent results. Overall, the models maintained the statistical structure of observed stages across lead times, with only a modest decrease in performance at longer horizons.

When tested on the extreme event of February 14, 2024, both ML models, driven either by rain gauge or radar inputs, effectively captured the main hydrodynamic features observed, especially the rapid rise and fall of the river stage across the three simulated lead times. This indicates their potential for short-term forecasting in fast-responding urban basins. Minor deviations in peak timing and shape – such as spikes (stage overestimation), plateaus (stage underestimation), or small phase lags – highlight the inherent temporal and spatial characteristics of rainfall inputs.

The comparative analysis of rain gauge- and radar-based models showed that both data sources produced similar predictive performance across all forecast horizons. Rain gauge-based models achieved slightly higher NSE, KGE, and RMSE scores. LinearSVR consistently used less computational power than XGBRegressor during training. In contrast, radar-based models

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

required nearly 10 times as much CPU time as their rain gauge counterparts due to the much larger input size. Despite these differences in training effort, all models took only about 0.1 s to generate predictions for the entire holdout dataset. From an operational standpoint, this indicates that the main time-consuming steps are data acquisition and model training, whereas real-time forecasting needs minimal computational resources. Overall, the performance metrics confirm that the ML models trained with both rain gauge and radar inputs provide reliable short-term stage predictions.

The statistically comparable performance of ML models driven by radar and rain-gauge rainfall indicates that radar-based ML approaches can represent a viable alternative for short-term stage forecasting in regions lacking rain-gauge networks.

A major challenge in applying ML models to hydrological modeling, as noted in existing studies, is the quality and size of available datasets, as well as the time needed to develop comprehensive datasets (Sit et al., 2020; Schmidt et al., 2020; Fang et al., 2022). The freely available TTI-HydroMet dataset (Escobar-Silva et al., 2025), which provides a 10-year record collected every 10 minutes, represents a significant step forward in addressing these issues. In addition, for operational use, improvements in radar data assimilation are crucial. The results of the log(MFB) distribution of daily radar precipitation indicate notable pointwise biases that can lead to overestimations or underestimations of larger rainfall totals.

Lastly, for future studies, it is recommended to evaluate additional ML algorithms and improve feature engineering and selection by including time-series lags and relevant watershed physical features. It is also suggested to compare results across different urban areas susceptible to flash floods, including regions without rain gauges but with radar coverage. Ultimately, analyzing short-lead-time radar forecasts will be essential for advancing hydrological predictions.

5 Code and data availability

The TTI-HydroMet dataset, which includes 10 years of hydrometeorological data for the study area along with the codes used and generated in this work, is publicly available in https://zenodo.org/records/17654660 (Escobar-Silva et al. (2025)).

Author contributions. KLRF, LSL, CWE, EVES, ASVL, RMPT, GRTL, and LBLS designed the study. KLRF, LSL, EVES, ASVL, and RMPT curated the dataset. LSL, ASVL, and RMPT performed validation and analysis. All authors contributed to manuscript writing.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Earth System Science Data.

Acknowledgements. This study was financed by the Brazilian National Council for Scientific and Technological Development (CNPq) projects 446053/2023-6 and 305205/2025-0 (LBLS), and by the São Paulo Research Foundation (FAPESP) grant 2024/02748-7 (EVES).

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

References

Ahmed, S. I., Rudra, R., Goel, P., Khan, A., Gharabaghi, B., and Sharma, R.: A Comparative Evaluation of Using Rain Gauge and NEXRAD Radar-Estimated Rainfall Data for Simulating Streamflow, Hydrology, 9, 133, https://doi.org/10.3390/hydrology9080133, 2022.

Amorim, L. F., Magalhães, A. A. B., Scarati Martins, J. R., Duarte, B. P. d. S., and Nogueira, F. F.: Hydrological modeling using distributed rainfall data to represent the flow in urban watersheds, RBRH, 27, e30, 2022.

Barros, M. T. L., Conde, F., Andrioli, C. P., and Zambon, R. C.: Flood Forecasting System in a Mega City: Challenges and Results for the São Paulo Metropolitan Region, pp. 10–19, ASCE Library, https://doi.org/10.1061/9780784479889.002, 2016.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific data, 5, 1–12, 2018.

Bernard, M. and Gregoretti, C.: The Use of Rain Gauge Measurements and Radar Data for the Model-Based Prediction of Runoff-Generated Debris-Flow Occurrence in Early Warning Systems, Water Resources Research, 57, e2020WR027 893, https://doi.org/10.1029/2020WR027893, 2021.

Beven, K. J.: Rainfall-runoff modelling: the primer, John Wiley & Sons, 2012.

Chadalawada, J., Herath, H., and Babovic, V.: Hydrologically informed machine learning for rainfall-runoff modeling: A genetic programming-based toolkit for automatic model induction, Water Resources Research, 56, e2019WR026 933, 2020.

Doswell III, C. A., Brooks, H. E., and Maddox, R. A.: Flash flood forecasting: An ingredients-based methodology, Weather and Forecasting, 11, 560–581, 1996.

Doviak, R. J. and Zrnic, D. S.: Doppler radar & weather observations, Academic press, 2014.

Escobar-Silva, E. V., Almeida, C. M. d., Silva, G. B. L. d., Bursteinas, I., Rocha Filho, K. L. d., de Oliveira, C. G., Fagundes, M. R., and Paiva, R. C. D. d.: Assessing the Extent of Flood-Prone Areas in a South-American Megacity Using Different High Resolution DTMs, Water, 15, 1127, 2023.

Escobar-Silva, E. V., Lima, L. S., Teixeira, R. M. P., Viteri, A., Rocha Filho, K. L., Eichholz, C., Soares, J. A. J. P., Conde, F., Lima, G., and Santos, L. B. L.: TTI-HydroMet: A Decade of High-Resolution Rainfall and Streamflow for the Tamanduateí River Watershed, Brazil, Zenodo Data, Version 2, https://doi.org/10.5281/zenodo.17654660, accessed: 20 November 2025, 2025.

Fang, K., Kifer, D., Lawson, K., Feng, D., and Shen, C.: The data synergy effects of time-series deep learning models in hydrology, Water Resources Research, 58, e2021WR029 583, 2022.

Gustafsson, F.: Determining the initial states in forward–backward filtering, IEEE Transactions on Signal Processing, 44, 988–992, 1996.

Haddad, E. A. and Teixeira, E.: Economic impacts of natural disasters in megacities: The case of floods in São Paulo, Brazil, Habitat International, 45, 106–113, special Issue: Exploratory Spatial Analysis of Urban Habitats, 2015.

Hampel, F. R.: The Influence Curve and Its Role in Robust Estimation, Journal of the American Statistical Association, 69, 383–393, 1974.

Hasan, F., Medley, P., Drake, J., and Chen, G.: Advancing hydrology through machine learning: insights, challenges, and future directions using the CAMELS, caravan, GRDC, CHIRPS, PERSIANN, NLDAS, GLDAS, and GRACE datasets, Water, 16, 1904, 2024.

He, X., Sonnenborg, T. O., Refsgaard, J. C., Vejen, F., and Jensen, K. H.: Evaluation of the value of radar QPE data and rain gauge data for hydrological modeling, Water Resources Research, 49, 5989–6005, https://doi.org/10.1002/wrcr.20471, 2013.

Hossoda, D. H., Perez, R. F., Tercini, J. R. B., and Bonnecarrère, J. I. G.: Data-Driven Modeling for Urban Flood Warning Systems: A Case Study in the Guarará Basin, Brazil, Journal of Flood Risk Management, 18, e70 110, 2025.

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

IBGE: Demographic Census 2022. Population and households: first results (Censo Demográfico 2022. População e domicílios: primeiros resultados), Tech. Rep. 101637, Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatistica), Rio de Janeiro, Brazil, https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102011, 2023.

IPCC: Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, https://doi.org/10.59327/IPCC/AR6-9789291691647.001, 2023.

Kant, C., Meena, R. S., and Singh, S. K.: A critical appraisal on various hydrological and hydrodynamic models, Water Conservation Science and Engineering, 10, 24, 2025.

Laney, D.: 3D Data Management: Controlling Data Volume, Velocity, and Variety, Application Delivery Strategies 949, META Group Inc., Stamford, CT, 2001.

Lee, J., Perera, D., Glickman, T., and Taing, L.: Water-related disasters and their health impacts: A global review, Progress in Disaster Science, 8, 100 123, 2020.

Liu, H., Shah, S. L., and Jiang, W.: On-line outlier detection and data cleaning, Computers and Chemical Engineering, 28, 1635–1647, 2004.

Looper, J. P. and Vieux, B. E.: An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model, Journal of Hydrology, 412, 114–132, https://doi.org/10.1016/j.jhydrol.2011.05.046, 2012.

Marcuzzo, F. F. N.: Bacia Hidrográfica do Rio Tietê: Precipitação Pluviométrica Espacializada / Tietê River Water Basin: Spacialized Rainfall, Geographia Meridionalis, 5, 243–266, https://doi.org/10.15210/gm.v5i3.16926, 2020.

Minuzzi, R. B., Sediyama, G. C., Barbosa, E. d. M., and Melo Júnior, J. C. F. d.: Climatologia do comportamento do período chuvoso da região sudeste do Brasil, Revista Brasileira de Meteorologia, 22, 338–344, https://doi.org/10.1590/S0102-77862007000300007, 2007.

Ochoa-Rodriguez, S., Wang, L.-P., Willems, P., and Onof, C.: A Review of Radar-Rain Gauge Data Merging Methods and Their Potential for Urban Hydrological Applications, Water Resources Research, 55, 6356–6391, https://doi.org/https://doi.org/10.1029/2018WR023332, 2019.

Pereira Filho, A. J. and dos Santos, C. C.: Modeling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data, Journal of Hydrology, 317, 31–48, 2006.

Ryzhkov, A., Zhang, P., Bukovčić, P., Zhang, J., and Cocks, S.: Polarimetric Radar Quantitative Precipitation Estimation, Remote Sensing, 14, https://doi.org/10.3390/rs14071695, 2022.

Ryzhkov, A. V., Giangrande, S. E., and Schuur, T. J.: Rainfall Estimation with a Polarimetric Prototype of WSR-88D, Journal of Applied Meteorology, 44, 502 – 515, https://doi.org/10.1175/JAM2213.1, 2005.

Saltikoff, E., Friedrich, K., Soderholm, J., Lengfeld, K., Nelson, B., Becker, A., Hollmann, R., Urban, B., Heistermann, M., and Tassone, C.: An Overview of Using Weather Radar for Climatological Studies: Successes, Challenges, and Potential, Bulletin of the American Meteorological Society, 100, 1739 – 1752, https://doi.org/10.1175/BAMS-D-18-0166.1, 2019.

Santos, L. B., Freitas, C. P., Bacelar, L., Soares, J. A., Diniz, M. M., Lima, G. R., and Stephany, S.: A Neural Network-Based Hydrological Model for Very High-Resolution Forecasting Using Weather Radar Data, Eng, 4, 1787–1796, https://doi.org/10.3390/eng4030101, 2023.

Santos, L. B., Escobar-Silva, E. V., Satolo, L. F., Oyarzabal, R. S., Diniz, M. M., Negri, R. G., Lima, G. R., Stephany, S., Soares, J. A., Duque, J. S., et al.: Machine learning-based hydrological models for flash floods: a systematic literature review, Smart Construction and Sustainable Cities, 3, 21, 2025.

Schmidt, L., Heße, F., Attinger, S., and Kumar, R.: Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany, Water resources research, 56, e2019WR025 924, 2020.

https://doi.org/10.5194/essd-2025-824
Preprint. Discussion started: 27 January 2026

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., and Demir, I.: A comprehensive review of deep learning applications in hydrology and water resources, Water Science and Technology, 82, 2635–2670, 2020.

Soares, J. A. J. P., Ozelim, L. C. S. M., Bacelar, L., Ribeiro, D. B., Stephany, S., and Santos, L. B. L.: ML4FF: A machine-learning framework for flash flood forecasting applied to a Brazilian watershed, Journal of Hydrology, 652, 132 674, https://doi.org/https://doi.org/10.1016/j.jhydrol.2025.132674, 2025.

Sokol, Z., Szturc, J., Orellana-Alvear, J., Popová, J., Jurczyk, A., and Célleri, R.: The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling—A Review, Remote Sensing, 13, https://doi.org/10.3390/rs13030351, 2021.

Tang, L., Li, J., Du, H., Li, L., Wu, J., and Wang, S.: Big Data in Forecasting Research: A Literature Review, Big Data Research, 27, https://doi.org/10.1016/j.bdr.2021.100289, 2022.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental modelling & software, 90, 201–216, 2017.

Van de Ven, F.: Water balances of urban areas, International Association of Hydrological Sciences Publication, 198, 21–33, 1990.

Viteri López, A. S. and Morales Rodriguez, C. A.: Flash flood forecasting in São Paulo using a binary logistic regression model, Atmosphere, 11, 473, 2020.

Wang, W., Chen, Y., Becker, S., and Liu, B.: Linear Trend Detection in Serially Dependent Hydrometeorological Data Based on a Variance Correction Spearman Rho Method, Water, 7, 7045–7065, https://doi.org/10.3390/w7126673, 2015.

Yang, L., Yang, Y., Shen, Y., Yang, J., Zheng, G., Smith, J., and Niyogi, D.: Urban development pattern's influence on extreme rainfall occurrences, Nature communications, 15, 3997, 2024.

Yoon, S.-S. and Lim, S.-H.: Analyzing the Application of X-Band Radar for Improving Rainfall Observation and Flood Forecasting in Yeongdong, South Korea, Remote Sensing, 14, 43, https://doi.org/10.3390/rs14010043, 2022.

Zhu, Y., Gao, Y., Wang, B., Nguyen, B. T., Zhang, Y., and Xue, B.: Distributed simulation of fully coupled hydrological-hydrodynamic model for predicting rainfall-induced runoff/flood in small watersheds, Journal of Hydrology: Regional Studies, 59, 102 450, 2025.