Integrated Methodologies for Flood Susceptibility Mapping and Urban Flood Prediction

Maria Clara Fava[1], Marcos Roberto Benso[2], Maurício Jonas Ferreira[3], Marina Batalini de Macedo[4], Anaí Floriano Vasconcelos[5], Beliana Cavalcante Sawada de Carvalho[6], Alan Vaz Lopes[7], Maria Elisa Leite Costa[8], Thiago Biscaro[9], Javier Tomasella[10]

Federal University of São Carlos, São Carlos, São Paulo, Brazil[1]
E-mail: mcfava@ufscar.br

University of São Paulo, São Carlos, Piracicaba, Brazil[2]
E-mail: marcosbenso@usp.br

University of São Paulo, São Carlos, São Paulo, Brazil[3]
E-mail: mjferradura@hotmail.com

Federal University of Itajubá, Itajubá, Minas Gerais, Brazil[4]
E-mail: marinamacedo@unifei.edu.br

Federal University of São Carlos, São Carlos, São Paulo, Brazil[5]
E-mail: anai.vasconcelos@ufscar.br

National Institute for Space Research, São José dos Campos, São Paulo, Brazil[6]
E-mail: beliana.sawada@unesp.br

National Water and Basic Sanitation Agency, Brasília, Distrito Federal, Brazil[7]
E-mail: alanvazlopes@gmail.com

Ministry of Cities, Brasília, Distrito Federal, Brazil[8]
E-mail: elisa.costa@cidades.gov.br

National Institute for Space Research, Cachoeira Paulista, São Paulo, Brazil[9]
E-mail: thiago.biscaro@inpe.br

National Institute for Space Research, Cachoeira Paulista, São Paulo, Brazil[10]
E-mail: javier.tomasella@inpe.br

ABSTRACT

Extreme precipitation events, such as urban flooding and flash floods, represent major challenges for risk management, particularly in rapidly urbanized regions. Climate change has increased the frequency and intensity of these events, resulting in significant social and economic impacts. In Brazil, many cities are marked by dense urbanization, unplanned expansion, and extensive impervious surfaces, which intensify flood risks. These characteristics underscore the need for reliable, robust, and easily replicable methodologies to support flood prediction and mitigation, especially in data-scarce contexts. This study focuses on the Metropolitan Area of São Paulo, a megacity highly vulnerable to urban flooding, aiming to develop integrated methodologies that combine flood prediction with susceptibility mapping. The hydrological component employs historical flood records and modeling tools to simulate surface runoff processes and estimate flood extents under different scenarios. In parallel, data-driven approaches are applied to produce detailed flood susceptibility maps. Using Geographic Information Systems (GIS) and statistical analyses, the study evaluates key physical and socioeconomic drivers of flood risk, including land use, slope, proximity to water bodies, population density, and socioeconomic vulnerability indicators. These variables are correlated with historical flood occurrences to develop predictive susceptibility models

that can support urban planning and risk reduction strategies. The outcomes include high-resolution flood susceptibility maps and predictive tools suitable for integration into operational early warning systems. These products are intended to support city planners and disaster management agencies by providing actionable information to improve preparedness, response, and long-term resilience. Moreover, the methodological framework developed in this research can be transferred to other Brazilian municipalities, particularly those with limited hydrometeorological data availability.

KEYWORDS: Urban flooding; Hydrological modeling; Machine learning; GIS-based analysis; Data-scarce environments.

1 INTRODUCTION

The consequences of high-intensity precipitation events, including urban flooding and flash floods, pose significant challenges for risk management. The impact of these events is even more concerning in vulnerable communities (Deria et al., 2020). The increasing frequency and intensity of these events, driven by climate change (Simonovic and Peck, 2022), have led to severe social and economic impacts (Chaudhary and Piracha, 2021), particularly in densely populated areas with limited infrastructure (IPCC, 2022). Many Brazilian cities and countries in the Global South are characterized by high levels of urbanization, unplanned expansion, and significant impervious surface coverage, which exacerbate flood risks (Adelekan, 2016; Tadesse et al., 2024; Shatanawia et al., 2024; Zhang et al., 2025). These conditions highlight the urgent need for reliable and easily replicable methodologies to predict and mitigate the impacts of urban flooding.

This study focuses on the Metropolitan Area of São Paulo, a megacity highly vulnerable to urban flooding, to develop integrated methodologies that combine flood prediction and susceptibility mapping. The city of São Paulo is the most populous municipality in Brazil, with a population of approximately 11.4 million inhabitants in 2022 (IBGE, 2022). Every summer, the megacity experiences flooding in various locations throughout its urban area (Haddad, 2015). Situated among three major rivers, Tietê, Pinheiros, and Tamanduateí, São Paulo's municipal territory encompasses around 1,521 km² and contains 166 sub-basins within its urban landscape (Gouveia, 2016). Considering the city's flood history, the heterogeneity of areas exposed to flood risk, and São Paulo's significant social and economic importance, this study adopts an integrated methodological approach. For the pilot case study, the assessments were conducted in the Aricanduva catchment, a highly urbanized catchment with frequent flood events.

In this context, this study aims to develop and evaluate an integrated and replicable framework for urban flood assessment in data-scarce environments. Two hydrological models with different levels of complexity, HYMOD and HEC-HMS, are applied and compared to assess the trade-off between model performance, computational cost, and data requirements. In parallel, flood susceptibility mapping is conducted using machine-learning classification algorithms, including Random Forest, Regularized Random Forest, and XGBoost, trained with historical flood occurrence data. By integrating hydrological modeling and data-driven susceptibility analysis, this study seeks to provide a cost-effective and robust methodology to support flood risk management in urban basins, particularly in large metropolitan areas of the Global South.

2 METHODOLOGY

The methodological framework of this study integrates hydrological modeling and data-driven susceptibility mapping to assess and predict urban flooding in a highly urbanized catchment. The approach combines physically based rainfall–runoff simulations with machine learning classification techniques, allowing both the dynamic representation of flood-generating processes and the spatial identification of areas prone to flooding. The following sections describe the study area, the hydrological modeling procedures, and the flood susceptibility assessment.

2.1 Case Study

The study area is the Aricanduva River basin, located in the eastern zone of the municipality of São Paulo, Brazil (Figure 1). The catchment covers an area of approximately 102.5 km² and is fully urbanized, with a high population density and a long history of recurrent flood events, including flash floods.

The region is characterized by a humid subtropical climate, with average monthly temperatures ranging from 17.2 °C in July to 23.5 °C in February. Mean monthly precipitation varies markedly throughout the year, from 32.3 mm in August to 292.1 mm in January, reflecting strong seasonal rainfall patterns (INMET, 2020).

According to the most recent demographic data, the basin has a population density of approximately 7,400 inhabitants per square kilometer (IBGE, 2022), which significantly increases exposure to flood-related hazards, including population displacement, injuries, and economic losses (Simas, 2017). In addition, the Aricanduva basin is well instrumented, with continuous monitoring of rainfall, streamflow (Gauge control in Fig. 1), and water level (Gauges at Fig. 1) at 10-minute intervals over a period of approximately ten years, providing a robust dataset for modeling and flood analysis.

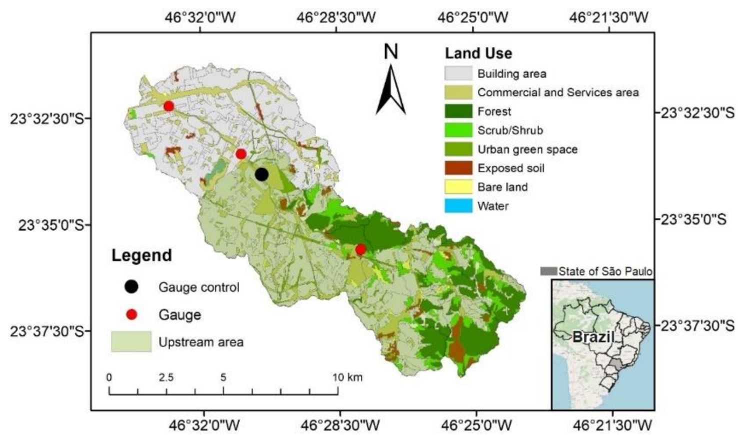


Figure 1: Land use and land cover were mapped using high-resolution orthorectified imagery

2.2 Hydrological Modeling

Hydrological modeling was carried out to simulate rainfall–runoff processes and support flood forecasting in the Aricanduva River catchment. Two widely used rainfall–runoff models with different levels of complexity were applied and adjusted: the conceptual HYMOD model and the semi-distributed HEC-HMS model. The joint use of these models allows the comparison of hydrological responses derived from simplified and spatially distributed representations of urban catchment processes.

Both models were calibrated and validated using the same observational dataset, consisting of precipitation and streamflow records collected within the catchment. Precipitation data were obtained from four monitoring stations, while discharge data were measured at a control section of the Aricanduva stream. All time series were processed at a 10-minute time step, allowing the representation of short-duration, high-intensity rainfall events typical of urban flooding.

Model calibration was performed utilizing automatic optimization techniques complemented by manual adjustments as needed. Additional details regarding model structure, parameterization, calibration strategies, and performance assessment are presented separately for each model in Sections 2.1.1 and 2.1.2, which describe the configuration of the HYMOD and HEC-HMS models, respectively.

2.2.1 HYMOD Model Configuration and Calibration

The HYMOD model was applied as a lumped conceptual rainfall–runoff model to simulate streamflow response in the catchment. The model was forced with time series of precipitation (P), potential evapotranspiration (PET), and calibrated using observed discharge (Q_{obs}).

The canonical five-parameter structure of HYMOD was adopted, including the maximum soil moisture storage capacity (C_{max}), the spatial variability of soil moisture storage (B_{exp}), the fraction of excess rainfall routed to the fast flow component (α), and the recession coefficients of the slow (Rs) and fast (Rq) flow reservoirs. Excess rainfall was computed using a probabilistic soil moisture capacity distribution and partitioned between fast and slow flow paths, represented by a cascade of linear reservoirs.

Parameter calibration was performed using the SciPy Optimize Minimize function with the L-BFGS-B algorithm, constrained by physically plausible parameter bounds reported in previous studies. Automatic calibration was complemented by manual adjustments when necessary to improve model performance. The optimized parameter set was subsequently applied to the full dataset for validation, and model performance was assessed by comparing simulated and observed discharges using standard statistical performance metrics.

2.2.2 HEC-HMS Model Configuration and Calibration

The HEC-HMS model was implemented as a semi-distributed hydrological model to represent spatial variability in land use and hydrological response across the catchment. Initial estimates of the Initial Loss (IL) and Constant Rate (CR) parameters were derived from HEC-HMS application guidelines and hydrological handbooks for urban catchments. The impervious area fraction (IM) was defined based on land use and land cover classification, while basin lag time (LT) was estimated during model pre-processing according to the spatial distribution of urban occupation. The basin was subdivided into 13 sub-catchments, allowing localized calibration of hydrological parameters.

Automatic calibration was conducted using the Univariate Gradient optimization method, with a maximum of 100 iterations and a convergence tolerance of 0.01. In cases where automatic calibration did not yield satisfactory results, manual calibration was applied. Final parameter values for IL, CR, IM, and LT were optimized for each sub-catchment based on the Runoff Curve Number (RCN) methodology and selected to maximize agreement between simulated and observed streamflow, following established statistical performance criteria.

2.3 Susceptibility Assessment

The study employs classification-based machine learning models to map urban flood susceptibility. The Random Forest, Regularized Random Forest, and XGBoost algorithms were employed and trained using flood occurrence data recorded between 2013 and 2024, classified into three categories: non-susceptible (1,000 data points), waterlogging of drainage system (196 data points), and fluvial flooding (207 data points), totaling 1,403 data points (Figure 2). Explanatory variables were derived from LiDAR data originally acquired at 0.5 m spatial resolution and resampled to 5 m for computational efficiency, using a Triangulated Irregular Network (TIN). From this dataset, the following topographic and hydrological factors were calculated: slope, slope length, slope steepness, Topographic Position Index (TPI), Topographic Index (TI), Height Above the Nearest Drainage (HAND), Profile Convergence Index (TCI), Stream Power Index (SPI), aspect, and drainage direction. The dataset was split into training (80%) and testing (20%) samples, and model performance was assessed using repeated 10-fold cross-validation. Model accuracy and Cohen's Kappa coefficient were used as evaluation metrics, allowing for both performance evaluation and identification of the most relevant factors influencing flood susceptibility.

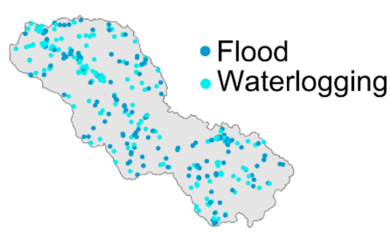


Figure 2: Flood occurrence data used in the susceptibility analysis (2013–2024; n = 1,403)

3 RESULTS AND DISCUSSION

3.1 Flood Modeling

The calibration and validation results of the HYMOD and HEC-HMS models showed a high level of agreement between simulated and observed hydrographs, particularly with respect to peak discharge and runoff volume. During the validation stage, both models were executed using the same calibrated parameters, demonstrating a consistent ability to reproduce the main hydrological characteristics of the analyzed events (Figure 3).

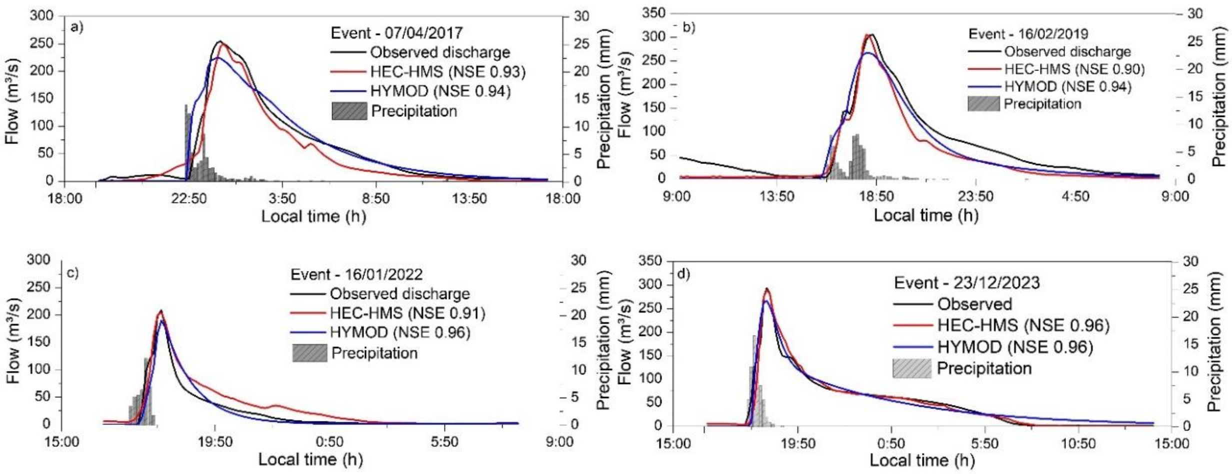


Figure 3: Measured and simulated discharge (m³ s⁻¹) for calibration (a–c) and validation (d).

Table 1: Statistic index of HEC-HMS and HYMOD calibration (dark grey) and validation (light grey)

Statistic index	Calibration/Validation									
	Date/time (duration of rain)									
	07Apr2017/ (22:50h to 04:00h)		16Feb2019/ (16:00 to 20:50h)		16Jan2022/ (16:00h to 17:10h)		23Dez2023/ (17:10h to 19:00h)		Performance Rating Moriasi et al. (2007)	
	HEC-HMS	HYMOD	HEC-HMS	HYMOD	HEC-HMS	HYMOD	HEC-HMS	HYMOD		
PEV (%)	**17.48**	**1.18**	**19.89**	**21.46**	**7.80**	**4.90**	**0.64**	**4.98**	Very good	<±10
									Good	±10–±15
									Satisfactory	±15–±25
									Unsatisfactory	>±25
PEPF (%)	**1.91**	**11.54**	**0.19**	**12.53**	**1.98**	**9.23**	**1.33**	**9.22**	Very good	<15
									Good	15 to 30
									Satisfactory	30 to 40
									Unsatisfactory	>40
R^2	**0.97**	**0.94**	**0.96**	**0.95**	**0.95**	**0.92**	**0.97**	**0.96**	Very good	0.75 to 1
									Good	0.65 to 0.75
									Satisfactory	0.50 to 0.65
									Unsatisfactory	<0.50
RMSE	**0.22**	**0.24**	**0.32**	**0.29**	**0.25**	**0.28**	**0.19**	**0.20**	Very good	0 to 0.5
									Good	0.5 to 0.6
									Satisfactory	0.6 to 0.7
									Unsatisfactory	> 0.7
NSE	**0.93**	**0.94**	**0.90**	**0.94**	**0.91**	**0.96**	**0.96**	**0.96**	Very good	0.75 to Unity
									Good	0.65 to 0.75
									Satisfactory	0.50 to 0.65
									Unsatisfactory	<0.50
PBIAS (%)	**17.48**	**-1.18**	**19.89**	**21.46**	**7.80**	**4.90**	**-0.64**	**-4.98**	Very good	<±10
									Good	±10 to ±15
									Satisfactory	±15 to ±25
									Unsatisfactory	>±25

Model performance was evaluated using six statistical indices (Table 1), allowing a comprehensive assessment of simulation accuracy. Across all simulations, R^2 values exceeded 0.90, and NSE ranged from 0.90 to 0.96, classifying the performance of both models as very good according to the criteria proposed by Moriasi et al. (2007). Low RMSE values, RSR values below 0.33, and reduced PBIAS further confirm the high reliability of the simulations.

In terms of runoff volume, HYMOD exhibited better overall performance, with percentage errors (PEV) below ±5% in three of the four analyzed events and satisfactory performance in only one event. HEC-HMS presented higher volumetric errors in earlier events, but achieved very good performance in the most recent events.

Peak discharge reproduction was satisfactory for both models, with percentage errors below 15% in all cases. HEC-HMS showed superior performance in this aspect, with consistently low errors (below 2%), while HYMOD exhibited slightly higher errors, although still classified as very good.

Overall, the results demonstrate that the HYMOD model, despite its simpler conceptual structure, was able to simulate urban flood events satisfactorily, accurately reproducing both peak flows and runoff volumes, and confirming its suitability for flood modeling in highly urbanized catchments.

3.2 Susceptibility Mapping

The susceptibility models achieved high predictive performance, with accuracy values above 0.80 and Kappa coefficients indicating good to substantial agreement across calibration and validation stages (Figure 4a). Random Forest-based models showed slightly better generalization capability compared to XGBoost. Variable importance analysis revealed that HAND and elevation were the dominant predictors, highlighting the critical role of relative topographic position with respect to the drainage network in controlling urban flood susceptibility (Figure 4b). Secondary variables related to flow convergence and hydraulic energy contributed to model performance, while slope length and flow direction showed limited influence, reflecting the strong impact of urban infrastructure on runoff dynamics.

Figure 4: a) Evaluation indices for calibration and validation of ML models, b) Variable importance

a) b)

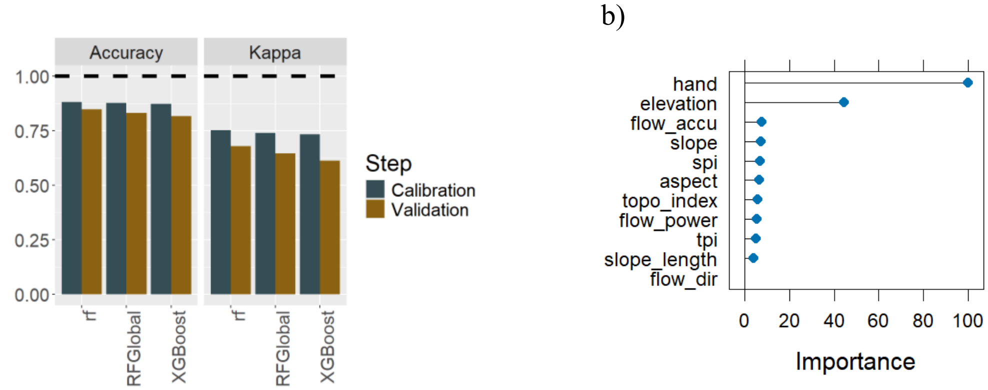
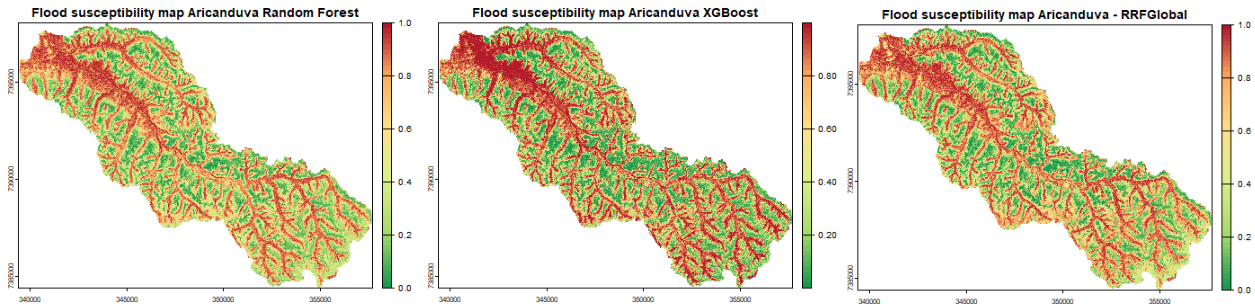


Figure 5 shows the flood susceptibility maps generated by the Random Forest, XGBoost, and Regularized Random Forest models. Despite methodological differences, all models consistently identify high-susceptibility areas along the main drainage network and low-lying regions of the basin, indicating robust spatial patterns. Minor differences are observed in the spatial continuity of susceptible areas, with XGBoost producing more fragmented patterns and Random Forest-based models yielding smoother susceptibility gradients. These patterns are consistent with the dominance of HAND and elevation identified in the variable importance analysis, reinforcing the physical interpretability of the results. Areas classified as highly susceptible by all models represent priority zones for flood risk management and urban drainage interventions.

Figure 5: Flood susceptibility maps of the Aricanduva catchment – red areas are the most susceptible



4 CONCLUSION

This study investigated hydrological modeling and flood susceptibility mapping to support flood risk assessment in a highly urbanized catchment. The results demonstrated that both HYMOD and HEC-HMS models were able to accurately reproduce urban flood events, with very good performance in terms of peak discharge and runoff volume. Despite its simpler conceptual structure, HYMOD showed reliable results, highlighting its potential for operational flood forecasting in urban basins.

The susceptibility mapping complemented the hydrological analysis by providing a spatially explicit identification of flood-prone areas. Machine learning models consistently highlighted low-lying areas near the drainage network as highly susceptible, with physically interpretable patterns strongly controlled by HAND and elevation. The convergence of results across different algorithms reinforces the robustness of the susceptibility maps and their applicability for urban flood management.

The combined use of hydrological modeling and susceptibility mapping offers a practical framework for urban flood risk mitigation. While high-resolution hydraulic models are computationally demanding for real-time applications in large urban areas, susceptibility maps can be used to pre-identify priority zones. In this context, when alert discharges are issued by the hydrological model, hydraulic simulations can be selectively applied to the most susceptible areas, enabling faster and more targeted flood response and decision-making.

Some limitations should be acknowledged. The hydrological models rely on lumped or semi-distributed representations, which may not fully capture the localized effects of urban drainage infrastructure. The susceptibility mapping is based on historical flood occurrences and static topographic variables, not explicitly accounting for temporal changes in land use, drainage capacity, or rainfall characteristics. Additionally, uncertainties associated with input data resolution and event reporting may affect model performance. Future research should focus on integrating dynamic rainfall forecasts, improving the representation of urban drainage systems, and coupling the proposed framework with real-time hydraulic simulations for operational flood warning systems.

5 ACKNOWLEDGEMENTS

The authors acknowledge the support of the National Council for Scientific and Technological Development (CNPq) through research grants (processes 381989/2024-0 and 174232/2023-3) and funding provided by the CNPq/MCTI Call No. 15/2023 – *Extreme Meteorological Events: Natural Disaster Prevention and Damage Minimization* (process 446029/2023-8).

REFERENCES

Adelekan I. O. (2010). Vulnerability of poor urban coastal communities to flooding in Lagos, Nigeria. Environment and urbanization, 22(2), 433-450. https://doi.org/10.1177/09562478103801

Chaudhary M. T. and Piracha A. Natural Disasters—Origins, Impacts, Management. Encyclopedia 1, no. 4 (2021): 1101–1131. https://doi.org/10.3390/encyclopedia1040084

Gouveia I. C. M. C. (2016). A cidade de São Paulo e seus rios: uma história repleta de paradoxos. Confins. Revue franco-brésilienne de géographie/Revista franco-brasilera de geografia, (27).

Haddad E. A. and Teixeira E. (2015). Economic impacts of natural disasters in megacities: The case of floods in São Paulo, Brazil. Habitat International, 45, 106-113. https://doi.org/10.1016/j.habitatint.2014.06.023

Intergovernmental Panel on Climate Change. (2022). Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg2/

INMET - Instituto Nacional de Meteorologia [National Meteorology Institute]. (2020). Normais climatológicas do Brasil [Brazilian climate normal]. Available at: https://portal.inmet.gov.br/normais#

IBGE - Instituto Brasileiro de Geografia e Estatística [Brazilian Institute of Geography and Statistics], https://www.ibge.gov.br/cidades-e-estados/sp/sao-paulo.html, consulted 19 December 2025.

Moriasi D.N. et al. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900. https://doi.org/ 10.13031/2013.23153.

Shatanawia K., Al-Weshaha R., Ta'anyc R., Mohammadd A. H., Kassaba G. and Halalshehd M. Assessment and mapping of flash flood hazard and risk at wadi yuyum basin in jordan: integrating hydrological modeling techniques. Water Conservation & Management (WCM) 8(3) (2024) 315-325. http://doi.org/10.26480/wcm.03.2024.315.325

Simas I. T. H. (2017). Análise retrospectiva de episódios de inundações na Bacia Hidrográfica do Rio Aricanduva – São Paulo [Retrospective analysis of flooding episodes in the Aricanduva River catchment – São Paulo]. Thesis, Universidade de São Paulo, São Paulo. https://doi.org/10.11606/D.8.2017.tde-21072017-162915

Simonovic S. P. and Peck, A. (2022). Resilience to climate change-caused flooding—Metro Vancouver case study. River, 1(1), 47-59. https://doi.org/10.1002/rvr2.6

Tadesse D., Suryabhagavan K. V., Nedaw D. and Hailu B. T. (2024). A model-based flood hazard mapping in Itang District of the Gambella region, Ethiopia. Geology, Ecology, and Landscapes, 8(1), 8-25. https://doi.org/10.1080/24749508.2021.2022833

Zhang Q., Li C., Wen D., Kang J., Chen T., Zhang B. and Slater L. (2025). Global South shows higher urban flood exposures than the Global North under current and future scenarios. Communications Earth & Environment, 6(1), 594. https://doi.org/10.1038/s43247-025-02585-7