

Integrating Hydraulic Modeling and Citizen Science for Past Flood Reconstruction in the Upper Paranapanema River Basin

Jaqueline Carolino Santos ¹; Javier Tomasella ²; Maria Clara Fava ³; Anai Floriano Vasconcelos ⁴; Hélio Rodrigues Bassanelli ⁵; Carlos Eugênio Pereira ⁶; Pedro Ramos Galvão ⁷; Gabriela Gonçalves Cildo ⁸.

Impacts, Adaptation, and Vulnerabilities Division, National Institute for Space Research, São José dos Campos, SP, Brazil ¹

Email: jaqueline.santos@inpe.br

Impacts, Adaptation, and Vulnerabilities Division, National Institute for Space Research, São José dos Campos, SP, Brazil ²

Email: javier.tomasella@inpe.br

Civil Engineering Department, Federal University of São Carlos, São Carlos, SP, Brazil ³

Email: mcfava@ufscar.br

Environmental Sciences Department, Federal University of São Carlos, São Carlos, SP, Brazil ⁴

Email: anai.vasconcelos@ufscar.br

Natural Disasters, São Paulo State University “Júlio de Mesquita Filho”, São José dos Campos, SP, Brazil ⁵

Email: helio.bassanelli@unesp.br

Faculty of Civil Engineering, Federal University of Uberlândia, Uberlândia, MG, Brazil ⁶

Email: cepereira@ufu.br

Environmental Sciences Department, Federal University of São Carlos, São Carlos, SP, Brazil ⁷

Email: gabrielacildo@estudante.ufscar.br

Environmental Sciences Department, Federal University of São Carlos, São Carlos, SP, Brazil ⁸

Email: pedroorg@estudante.ufscar.br

ABSTRACT

The quantification of flood hazards is frequently hindered by the lack of systematized hydrological records, particularly in developing nations. To address this challenge, this study evaluates a methodology that integrates citizen science with hydrodynamic modeling. Focusing on the Upper Paranapanema River Basin in southeastern Brazil, the HEC-RAS model was employed to reconstruct historical flood events. The approach was validated using resident-supplied flood marks and satellite imagery, yielding a Kappa statistic of 0.7 and an NSE of up to 0.79. Additionally, collective memory was instrumental in supplementing peak flow statistics to generate scenarios across multiple return periods. The study further investigated the hydraulic effects of rural earth dikes, revealing that these structures exert negligible influence on flood extent within the urban area. These results highlight the efficacy of combining local knowledge with numerical modeling, validating citizen science as a practical and effective alternative for flood management in data-scarce contexts.

KEYWORDS: flood risk; data scarcity; citizen science; hydrodynamic model; flood scenarios; return period analysis.

1 INTRODUCTION

Floods are among the leading causes of socioeconomic and environmental. Their increasing frequency, linked to global warming (IPCC, 2021), challenges the traditional stationarity assumptions underpinning water resources management. Understanding these events is complicated due to natural and anthropogenic changes in river basins (Berghuijs et al., 2016). Despite advances in modeling, data scarcity remains a critical limitation in developing countries (Krabbenhoft et al., 2022).

In this context of scarcity, evidence from local observations and indirect methods is valuable (Sy et al., 2020). Citizen science emerges as a promising approach to fill data gaps and enrich knowledge about floods, offering a flexible framework for public participation. Beyond data collection, it helps to understand cultural interactions with hydrological systems, relying on collective memory to validate scientific initiatives (Sarmiento Buarque et al., 2020).

In Brazil, studies report changes in the magnitude and frequency of floods, with positive trends in the South and Southeast regions (Bartiko et al., 2019), often attributed to extreme precipitation (Chagas et al., 2021), despite recent drought trends in some periods. Flood risk mapping in the country typically relies on limited field observations and coarse extrapolations (Marchezini et al., 2022), highlighting the need for alternative approaches to facilitate knowledge exchange between scientists and communities.

This study focuses on the Upper Paranapanema River Basin in southeastern Brazil, an area frequently affected by major floods and with fragmented hydrological data. The region experienced significant events in 2004, 2010, and 2016, causing substantial losses to agriculture and residents (SISSER, 2016; SEBRAE, 2025). The historical presence of earthen dikes, built in the 1970s and destroyed in 2004, raises questions about flow alteration and the reduction of the natural attenuation capacity of floodplains, which may exacerbate damage in downstream urban areas.

To address these challenges, this study examines whether citizen science can enhance the reliability of hydrodynamic modeling and flood frequency analysis in data-poor regions. The objective is to integrate resident accounts with detailed mapping and hydrological data to calibrate a model that reconstructs the 2004, 2010, and 2016 flood events, as well as to estimate historical peak discharges beyond the instrumental period using citizen-provided information. By combining historical and systematic records, the study assesses flood impacts associated with return periods of up to 500 years. Under these scenarios, the reconstruction of earthen dikes for rural protection is considered, and their effects on the urban area are evaluated, accounting for the reduction in upstream flood attenuation capacity.

2 METHODOLOGY

2.1 Study area

The study site (Figure 1) covers 128.85 km² in the Upper Paranapanema basin, encompassing the confluence of the Paranapanema and Itapetininga rivers. Upstream, the area is dominated by floodplains and a meandering morphology, where the Paranapanema flows north with a low average slope (0.00016 m.m⁻¹), while the Itapetininga flows west (0.00020 m.m⁻¹). Downstream of the confluence, the river straightens due to topographic constraints; although remnants of old dikes exist along the banks, these anthropogenic structures have been partially destroyed by recurrent flooding.

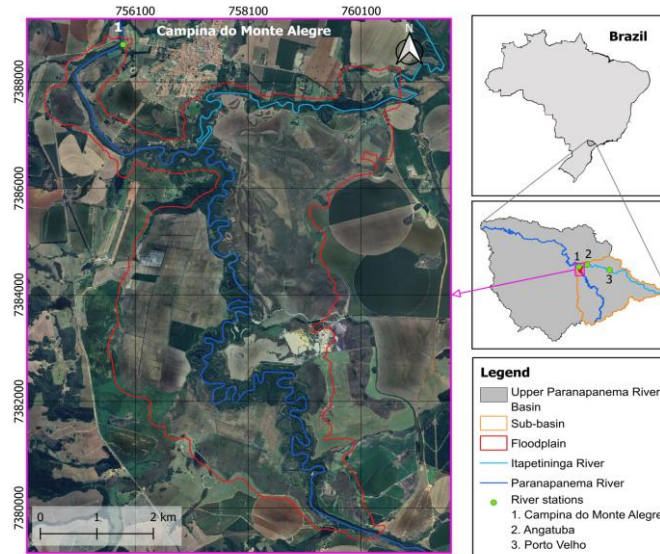


Figure 1: Location of the study site in the Upper Paranapanema River Basin. Projection: SIRGAS 2000, UTM zone 22S. Data sources: Google Satellite (2024), IBGE (2022), ANA (2024).

The main urban center, Campina do Monte Alegre, is located downstream and houses 5,954 inhabitants (IBGE, 2022), with an economy primarily based on agriculture and livestock (SEBRAE, 2025). The region features a Cfa climate with an annual average temperature of 20.2°C and precipitation of 1,324 mm, situated within a geomorphological context dominated by the sedimentary Paraná basin and the Atlantic orogenic belt. Despite economic activities, the area retains significant native vegetation cover from the Atlantic Forest and Cerrado biomes (CBH-ALPA, 2023).

2.2 Citizen science data

To address the absence of systematic records prior to the 2000s, a citizen science approach was applied in Campina do Monte Alegre, combining interviews and physical flood mark documentation. Conducted between January and April 2023 in accordance with ethical standards (CAAE 61253522.3.0000.5504). This method enabled the identification of eight major flood events between 1983 and 2016, with the 2004 flood being the most significant. Local flood-level markings on utility poles supported data validation with rainfall series and satellite imagery, enhancing the reliability of the historical reconstruction in a data-scarce environment, as further detailed in Vasconcelos et al. (2025).

2.3 Topography, river cross-section survey, satellite image, and hydrological data

Topographic data consisted of the ANADEM digital terrain model (ANA, 2025), a 30-m resolution product. This was complemented by a bathymetric survey conducted in May 2023 using a RiverSurveyor M9 ADCP to map cross-sections and longitudinal profiles along the Itapetininga and Paranapanema rivers. Additionally, the precise elevations of historical flood marks on utility poles were acquired using Geomax GNSS RTK receivers, with data post-processed via the IBGE PPP service to ensure high positional accuracy.

For model validation and boundary conditions, a Landsat 8 satellite image (ID: LC08_L2SP_220076_20160120_20200907_02_T1_SR) capturing the 2016 flood was processed in QGIS using spectral indices to isolate open water areas for comparison with simulations. Hydrological data were obtained from the National Water Agency (ANA) platform, Hidroweb (Figure 1). The Angatuba and Campina do Monte Alegre stations defined the boundary conditions, while the Porto Velho station was used to estimate stages at Angatuba for the 2016 event, as direct records at this station were no longer available following its deactivation in 2010.

2.4 Flood Frequency Analysis

Flood frequency analysis was performed using peak flow records from the Angatuba and Campina do Monte Alegre stations. L-Moments were derived through Probability Weighted Moments (Greenwood et al., 1979; Hosking and Wallis, 1997), selected for their reliability with limited datasets (Fischer, 2024). The Angatuba station provided a continuous 76-year series, while the Campina do Monte Alegre dataset required the integration of historical flood events (1983 to 2016) documented by residents. These flood marks were georeferenced and correlated with observed water levels to estimate peak discharges based on the official rating curve.

To address discontinuities in the Campina do Monte Alegre record, partial Probability Weighted Moments were applied (Wang, 1990), splitting the dataset according to a predefined threshold (Fischer, 2024). Twelve probability distributions were evaluated using the methodology developed by Hosking (1985), and the best fit was selected based on the lowest root mean square error and visual inspection. Plotting positions followed the Weibull method for Angatuba and the Hirsch approach (1987) for the historical dataset. Despite potential long-term trends, a stationary assumption was maintained to reduce the high uncertainty typically associated with non-stationary models (Anzolin et al., 2024).

2.5 Hydrodynamic Modeling Setup

Hydrodynamic simulations were performed using the HEC-RAS 6.6 two-dimensional model (USACE, 2024), incorporating geospatial inputs via RAS Mapper and applying a uniform Manning roughness coefficient of 0.07. The model was calibrated under transient conditions with a five-day warm-up period to reconstruct the 2005, 2010, and 2016 flood events. Validation employed NSE, KGE, and RMSE metrics based on observed stages at Campina do Monte Alegre. Boundary conditions included daily discharges from Angatuba, a downstream normal depth slope of 0.00012 metres per metre, and upstream inflow from the Paranapanema River estimated by mass balance, justified by its dominant contribution over the Itapetininga tributary.

In accordance with Brazilian Law 12.608 (Brasil, 2012) and CEMS-Flood guidelines (2025), steady-state scenarios were simulated for return periods of 50, 100, and 500 years. Two configurations were evaluated: one reflecting current condition without flood embankments, and another assuming the reconstruction of dikes lost in 2004 to safeguard agricultural areas. Impacts were assessed by comparing water levels and flow velocities along longitudinal profiles covering three kilometres upstream and six kilometres downstream of the confluence, including urban peripheries.

3 RESULTS

3.1 Reconstruction of Floodplain inundation during the floods of 2004, 2010, and 2016

The hydrodynamic reconstruction of the flood events that occurred in 2004, 2010, and 2016 demonstrated strong model performance in both spatial and temporal dimensions. Spatial validation using Landsat imagery from the 2016 flood resulted in a Kappa coefficient of 0.7 (Cohen, 1960), indicating satisfactory spatial agreement despite minor errors of commission and omission near the riparian boundaries. From a hydrological perspective, the model accurately reproduced discharge magnitude and timing at the gauge located in Campina do Monte Alegre, with Nash-Sutcliffe efficiency values ranging from 0.79 to 0.91, Kling-Gupta efficiency values between 0.74 and 0.84, and root mean square errors below 0.40. In addition, comparison with flood marks identified through citizen science and georeferenced using RTK measurements confirmed the model performance in terms of vertical accuracy, with absolute elevation differences limited to 0.36 meters for the 2004 flood and -0.41 meters for the 2016 flood.

Analysis of inundation patterns reveals that the expansion of the flooded area is topographically constrained by surrounding higher elevations, particularly to the west and near the urban center. Consequently, variations in flood magnitude are primarily reflected in water surface elevations and depths rather than in the lateral extent of the flooded area.

3.2 Statistical Analysis of Peak Flows

The integration of historical flood marks obtained through citizen science with systematic records emphasized the relevance of pre-2000 events, notably the 1983 flood, ranked as the second largest in the 1983–2020 series. By applying a validated elevation offset between utility pole marks (measured with RTK-GPS) and gauge data from the 2004 and 2016 events, peak discharges were estimated for historical floods, four of which ranked among the top seven. For frequency analysis, the Generalized Pareto distribution provided the best fit at both Angatuba and Campina do Monte Alegre. While Angatuba used a continuous series (1935–2010), Campina do Monte Alegre required a partial Probability Weighted Moments approach with a 300 cubic metres per second threshold to incorporate nine events into a censored sample, combining systematic and historical peaks.

The reconstructed flood series revealed significant hydrological differences between the contributing basins. Inclusion of citizen-derived historical events notably altered the frequency curve at Campina do Monte Alegre, underscoring the methodological value of integrating local knowledge to refine flood hazard assessments.

3.3 Scenarios analysis for different return periods and potential impacts of the reconstruction of the earth dike in the floodplain

The comparative analysis of "no dike" and "dike" scenarios for 50, 100, and 500-year return periods reveals that the reconstruction of flood defenses would have negligible impacts on the spatial extent of flooding in the downstream urban area (Figure 2). While the total flooded area and depth increase with return periods, the presence of dikes does not alter the flood limits surrounding the town, as confirmed by the longitudinal profiles downstream of the confluence, where water surface elevations remain virtually identical across scenarios. Instead, the hydraulic impact of the dikes is concentrated upstream of the confluence. On the Paranapanema River, the dikes cause a rise in water levels that intensifies upstream and correlates positively with flood magnitude, whereas on the Itapetininga River, hydraulic differences are more pronounced for lower return periods.

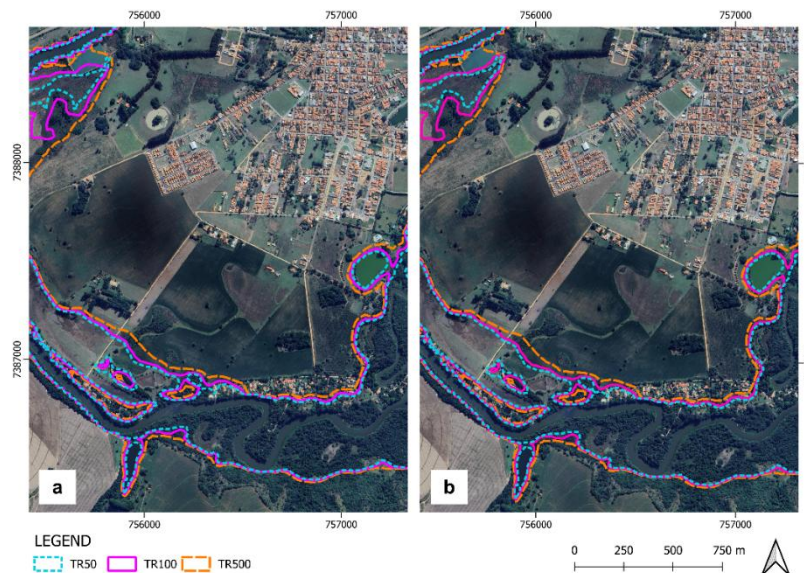


Figure 2: Limits of the flooded area for the 50, 100 e 500-year floods in the vicinity of Campina do Monte Alegre for the “no dike” (a) scenario and for the “dike” scenario (b).

In relation to stream dynamics, velocity profiles confirm that the dikes exert a significant influence on flow energy in the upstream sector, while downstream conditions remain largely unaffected. The topography of the valley limits lateral floodplain expansion, causing flood magnitude to be expressed

primarily through increased depth. Under these constraints, the presence of dikes enables the recovery of agricultural areas in upstream regions without intensifying flood risk in the downstream urban zone. However, this benefit comes with the trade-off of elevated hydraulic energy in rural upstream sections.

4 DISCUSSION

Citizen science played a key role in enhancing flood reconstruction in a data-scarce region, supporting global findings on the reliability of non-expert observations. HEC-RAS simulations compared with citizen-recorded water marks showed vertical errors of 0.3–0.4 m, attributed to DEM and resolution limitations (Patel et al., 2017; Mohanty et al., 2020; Artiglieri et al., 2025). Despite such discrepancies, citizen data extended the flood record to 1983, preventing underestimation of critical flows and improving hydrological accuracy in unmonitored areas (Scheller et al., 2024).

The model demonstrated good performance, with high NSE values, especially for the 2004, 2010, and 2016 events. However, the 2004 peak was slightly underestimated due to uncertainties in friction slopes, mass balance, and rating curve extrapolations. Spatial validation using Landsat imagery confirmed that ANADEM reliably represents floodplain morphology, with differences limited to riparian vegetation zones.

Simulations indicated a limited influence of old earth dikes on downstream flooding because of channelization effects, whereas upstream impacts were more pronounced. Along the Paranapanema River, the presence of dikes reduced the effective flow section, leading to localized increases in flow velocity. In the Itapetininga River, backwater effects became more evident in the vicinity of the confluence. Despite remaining limitations, including uncertainties related to the digital elevation model and Manning roughness assumptions, the results emphasize the spatial heterogeneity of dike effects and underscore the importance of flood management strategies informed by geomorphological controls.

5 CONCLUSION

This study demonstrated the efficacy of integrating remote sensing, hydraulic modeling, and citizen science to reconstruct floodplain dynamics in data-scarce regions. While HEC-RAS simulations showed good agreement with observations, the results highlighted that citizen-derived data is crucial not only for model validation and prioritizing high-resolution topographic surveys in critical areas but, most importantly, for recovering historical peak flows beyond the instrumental period. This methodological integration significantly improves extreme-value statistics and the reliability of flood hazard scenarios, which are imperative given the increasing frequency of extreme events. Ultimately, this approach offers a cost-effective, empowering framework for incorporating disaster resilience into community planning and early warning systems in municipalities across Brazil that lack conventional monitoring networks.

6 ACKNOWLEDGEMENTS

The authors would like to thank the National Council for Scientific and Technological Development (CNPq) for funding the project (Grants No. 409527/2021-1 and 440016/2024-0). We also thank the Postgraduate Program in Natural Disasters at the São Paulo State University "Júlio de Mesquita Filho" (UNESP), as well as the Postgraduate Program in Earth System Science at the National Institute for Space Research (INPE).

REFERENCES

Anzolin G, de Oliveira DY, Vrugt JA, AghaKouchak A, Chaffe PLB (2024) Nonstationary frequency analysis of extreme precipitation: Embracing trends in observations. *J Hydrol* 637:131300. <https://doi.org/10.1016/j.jhydrol.2024.131300>

- Artiglieri, Pasqualino, et al. Performance of HEC-RAS v6. 5 at basin scale for calculating the flow hydrograph: comparison with TUFLOW. *Natural Hazards*, 121.8 (2025): 9477-9490. <https://doi.org/10.1007/s11069-025-07170-x>
- Bartiko D, Oliveira D. Y., Bonumá N. B., and Chaffe P. L. B. (2019) Spatial and seasonal patterns of flood change across Brazil, *Hydrological Sciences Journal*, 64:9, 1071-1079. <https://doi.org/10.1080/02626667.2019.1619081>
- Berghuijs, W.R., et al., 2016. Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, 43 (9), 4382–4390. <https://doi.org/10.1002/2016GL068070>
- Chagas, V.B.P., Chaffe, P.L.B. & Blöschl, G. Climate and land management accelerate the Brazilian water cycle. *Nat Commun* 13, 5136 (2022). <https://doi.org/10.1038/s41467-022-32580-x>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Fischer, S (2024) Incorporating historical flood events in type-based statistics. *Journal of Hydrology*, 636. 131289. <https://doi.org/10.1016/j.jhydrol.2024.131289>
- Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.* 15 (5), 1049–1054. <https://doi.org/10.1029/WR015i005p01049>
- Hirsch, R.M., 1987. Probability plotting position formulas for flood records with historical information. *J. Hydrol.* 96 (1), 185–199. [https://doi.org/10.1016/0022-1694\(87\)90152-1](https://doi.org/10.1016/0022-1694(87)90152-1)
- Hosking JRM. 1985. Research Report: Fortran Routines for use with the Method of L-Moments. Version 3.04. IBM Research Division, T. J. Watson Research Center, Yorktown Heights, NY 10598. RC 20525 (90933) 8/5/96, revised 7/25/05. Available from: <http://lib.stat.cmu.edu/general/lmoments>
- Hosking JRM, Wallis JR (1997) *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, New York. <https://doi.org/10.1017/CBO9780511529443>
- IPCC. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva: IPCC; 2021. <https://doi.org/10.1017/9781009157896.001>
- Krabbenhof, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., ... & Olden, J. D. (2022). Assessing placement bias of the global river gauge network. *Nature Sustainability*, 5(7), 586-592. <https://doi.org/10.1038/s41893-022-00873-0>
- Marchezini V, Porto de Albuquerque J, Pitidis V, Rudorff CM, Lima-Silva F, Klonner, C and Martins, MHM (2022) Flood risk governance in Brazil and the UK : facilitating knowledge exchange through research gaps and the potential of citizen-generated data. *Disaster Prevention and Management: An International Journal*, 31 (6). pp. 30-44. <https://doi.org/10.1108/dpm-01-2022-001>.
- Mohanty, M.P., Indu, J., Ghosh, S., Rao, G., Nithya, S., Karmakar, S., 2020b. Sensitivity of various topographic data in flood management: Implications on inundation mapping over large data-scarce regions. *J. Hydrology* 590, 125523. <https://doi.org/10.1016/j.jhydrol.2020.125523>.
- Patel, D.P., Ramirez, J.A., Srivastava, P.K. Bray M, Han D. Assessment of flood inundation mapping of Surat city by coupled 1D/2D hydrodynamic modeling: a case application of the new HEC-RAS 5. *Nat Hazards* 89, 93–130 (2017). <https://doi.org/10.1007/s11069-017-2956-6>
- Sarmiento Buarque, A. C., Bhattacharya-Mis, N., Fava, M. C., Souza, F. A. A. D., & MENDIONDO, E. M. (2020). Using historical source data to understand urban flood risk: a socio-hydrological modelling application at Gregório

Creek, Brazil. Hydrological Sciences Journal, 65(7), 1075-1083.
<https://doi.org/10.1080/02626667.2020.1740705>

Scheller, Mirjam, Ilja van Meerveld, and Jan Seibert. "How well can people observe the flow state of temporary streams?." *Frontiers in Environmental Science* 12 (2024): 1352697.

Sy, B., Frischknecht, C., Dao, H., Consuegra, D. and Giuliani, G. (2020) Reconstituting Past Flood Events: The Contribution of Citizen Science. *Hydrology and Earth System*, 24, 61-74. <https://doi.org/10.5194/hess-24-61-2020>

Vasconcelos, A. F., Santos, J. C., Cinto, G. G., Galvão, P. R., Fava, M. C., Menezes Filho, F. C. M., Pereira, C. E., Bassanelli, H. R. e Tomasella. Contribuições da ciência cidadã para o gerenciamento de riscos de inundações: o caso de Campina do Monte Alegre – SP, Brasil. In: Souto, Raquel Dezidério (Org.). *Estudos de caso em mapeamento colaborativo*. Rio de Janeiro: IVIDES.org. p. 303-334. ISBN: 978-65-985676-2-0. doi: 10.5281/zenodo.16808938

Wang, Q.J., 1990. Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution. *J. Hydrol.* 120 (1–4), 115–124. [https://doi.org/10.1016/0022-1694\(90\)90145-N](https://doi.org/10.1016/0022-1694(90)90145-N)

Web sites:

ANA, (2024): <https://www.snirh.gov.br/hidroweb/serieshistoricas>, consulted 13 May 2024.

ANA, (2025): [https://metadados.snirh.gov.br/geonetwork/srv/api/records/93664c15-1ff8-4e87-bbed-2bb69d321309#:~:text=](https://metadados.snirh.gov.br/geonetwork/srv/api/records/93664c15-1ff8-4e87-bbed-2bb69d321309#:~:text=,), consulted 13 Jan 2025.

Brasil. (2012): http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112608.htm, consulted 07 Ago 2025.

CBH-ALPA. (2023): https://sigrh.sp.gov.br/public/uploads/documents/8124/rs_2014_cbhalpa_vfinal_of.pdf, consulted 5 Jun 2024.

Copernicus Emergency Management Service (2025): <https://confluence.ecmwf.int/display/CEMS/CEMS-Flood+flood+inundation+maps>, consulted 28 Jul 2025.

IBGE (2022): <https://cidades.ibge.gov.br/brasil/sp/campina-do-monte-alegre/panorama>, consulted 16 Jun 2024.

SEBRAE (2025): <https://datampe.sebrae.com.br/profile/geo/campina-do-monte-alegre>, consulted 1 Mar 2025.

SISSER (2022): <https://dados.agricultura.gov.br/dataset/sisser3/resource/54e04a6b-15b3-4bda-a330-b8e805deabe4>, consulted 11 Aug 2024.

USACE (2024): <https://www.hec.usace.army.mil/software/hec-ras/>, consulted 17 Jan 2025.