A Nowcasting System for Hydrometeorological Hazard Assessment of Landslides and Flooding – Part 2: On Verification and Validation Process

Um Sistema de Nowcasting para Avaliação de Riscos Hidrometeorológicos de Deslizamentos e Inundações – Parte 2: Processo de Verificação e Validação

Hugo Abi Karam 🔟

Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Meteorologia, Rio de Janeiro, RJ, Brasil
E-mail: hugo@igeo.ufrj.br

Abstract

This work presents a comprehensive analysis of the verification and evaluation of a system designed for forecasting hydrometeorological risks, with a specific focus on landslides and floods in a defined region. The proposed system continuously integrates real-time meteorological and hydrological data to provide precise and timely information on potential risk events. Through a meticulously conducted case study, the practical application of the system is highlighted, demonstrating its effectiveness in monitoring and forecasting risk events in real-world scenarios. The work addresses fundamental challenges associated with the validation of complex systems, emphasizing the imperative need for robust verification and validation methods. Furthermore, the unique characteristics of complex systems are discussed, and their implications for effective modeling and validation processes are elucidated. A detailed presentation of the benchmark case study results includes analyses of rainfall intensity, dynamic mapping of landslide susceptibility, river height monitoring, and forecast comparisons. These findings are complemented by visual aids in a web interface that facilitate a comprehensive understanding of the system's performance under real conditions. Key insights are emphasized, highlighting the crucial role of the proposed system in advancing knowledge in the field of hydrometeorological risk assessment and forecasting. The conclusions succinctly summarize the main results and underscore the critical importance of systems like the proposed one in mitigating these risks.

Keywords: Data integration; Dynamical risk mapping; Validation methods

Resumo

Este trabalho apresenta um estudo abrangente sobre a verificação e avaliação de um sistema para a previsão de riscos hidrometeorológicos, com foco específico em deslizamentos de terra e inundações em uma região delimitada. O sistema proposto integra continuamente dados meteorológicos e hidrológicos em tempo real, proporcionando informações precisas e oportunas sobre possíveis eventos de risco. Através de um estudo de caso meticulosamente conduzido, destaca-se a aplicação prática do sistema e evidenciam-se sua eficácia no monitoramento e previsão de eventos de risco em cenários reais. O trabalho aborda os desafios fundamentais associados à validação de sistemas complexos, enfatizando a necessidade imperativa de métodos robustos de verificação e validação. Além disso, são discutidas as características únicas dos sistemas complexos e elucidadas as implicações para processos de modelagem e validação. A apresentação detalhada dos resultados do estudo de caso de referência inclui análises da intensidade da chuva, mapeamento dinâmico da suscetibilidade a deslizamentos de terra, monitoramento da altura dos rios e comparações de previsões. Esses resultados são complementados por recursos visuais em uma interface web, que facilitam a compreensão do desempenho do sistema em condições reais. Destacam-se os insights obtidos, que enfatizam o papel crucial da proposta no avanço do conhecimento na avaliação e previsão de riscos hidrometeorológicos. As conclusões sintetizam os principais resultados e ressaltam a importância crítica de sistemas como o proposto na mitigação desses riscos.

Palavras-chave: Integração de dados; Mapeamento dinâmico de risco; Métodos de validação

1 Introduction

The research problem addresses the need to develop an effective system for assessing and predicting hydrometeorological risks, such as landslides and floods, in a specific region. This system should integrate real-time meteorological and hydrological data to provide accurate and timely information about potential risk events.

The validation step, particularly the assessment of complex systems, is arguably the most challenging task, requiring a strong correlation between the model system's development and data acquisition (Gentil & Blake 1981). Operational behaviors inevitably prompt consideration of methodological requirements for system validation, which encompass efficiency and adaptability to changes in response to operator behavior (Stager 1993).

The methodological structure for validating environmental indicators can be categorized into three types: "project validation," "output validation," and "final use validation." The latter, initially applied in engineering, is not inherently sufficient for verifying and validating complex systems, such as intricate agricultural systems (Bockstaller & Girardin 2003), and meteorological modeling.

Tests are vital components of the Verification and Validation (V&V) process but not the sole components. The V&V process can be "informal" (e.g., audits, process console research), static (e.g., cause-effect analysis and interface analysis), dynamic (e.g., acceptance, purpose, comparison, compliance, and failure analysis), and formal tests (i.e., inductive and deductive analysis and using evidence of predication). This has led to the technical development of the V&V process and the establishment of "Recommended Practice Guides" (RPG), contributing to an evolutionary dynamic between conceptual and procedural models (Cook & Skinner 2005).

Therefore, certification depends not only on calibration but also on the statistical trust established through initial and continuous tests during operation (e.g., using Receiver Operator Curves ROC, and Receiver Operator Area under Curve evolution analysis).

Complex Systems (CS) possess distinctive defining characteristics, including sensitivity to initial conditions, emergent behavior, and composition of components. Their increasing prevalence in modeling efforts implies additional challenges in both effective modeling and validation. CS exhibit also other characteristics such as uncertain boundaries, nesting (since components of CS can themselves be CS), state memory, non-linear relationships, and feedback loops (Petty 2018). Some workshops, such as WWRP/WGNE (2017), have provided a comprehensible review of verification methods for numerical models.

Type I error, which is less severe, occurs when inadequate Verification and Validation (V&V) processes result in the model builder not utilizing a 'valid model'. This situation exposes the final user to risks, particularly if they are overly critical, averse to false alarms, or encounter complexities in achieving the open-box approach goal. Conversely, Type II error, the most serious, arises when incorrect V&V procedures lead to the model builder employing an 'invalid model'. This exposes the final user to significant risks, especially if they assume a gray box approach, accept greater risk, experience a large number of false alarms, or encounter very small gradients on the Receiver Operator Curve (Petty 2018).

Sensitivity to initial conditions can present validation challenges, such as the analysis of the distribution of results, sensitivity analysis, and input imprecision analysis. Typically, three methods are employed to mitigate validation challenges: more trials, sensitivity analysis (e.g., by error propagation or a gradient descent method for greater efficiency, simplex calibration, etc.), and awareness of the required accuracy (Petty 2018).

Resolution of emerging behaviors (extreme or very dangerous) in risk events can be addressed with additional observations (e.g., by Kalman filter assimilation), validation assessments organized in the form of previously planned scenarios under consensus (e.g., using a complexity hierarchy of validation tests), establishment of a test bench or reference scenario space (i.e., set benchmarks), and semi-automatic model adaptation (i.e., the ability to support automated multi-scale analysis and long-term variability, such as that obtained with spectral models or inter-scale interactions and second-order parametrization) (Petty 2018; Kalnay 2003; Li et al. 2009; Groesser & Schwaninger 2012; Ota et al. 2013).

V&V of complex systems based on Human Factors (HF) criteria are gaining widespread recognition, as HF practices aim to improve system performance when applied systematically and systemically (i.e., in a participative and solution-based approach, beyond regulatory compliance) (Teperi et al. 2023). Failures and uncertainties in input data and events outside the validity conditions of equations can lead to interruptions in the continuous solution of the CS process. Considering that the system runs uninterruptedly, either a division by zero or an instability of solutions can lead to overflow (Thompson 1961; Mesinger & Arakawa 1976; Marchuk 2012; Durran 2013).

When coding CS equations, it is crucial to implement specific localized continuous tests ('weaving test') for the continuity of the model solution. The implementation of try-and-catch codes can be anticipated by modular tests, avoiding processor interruption by proposing an alternative

resolution. In general, an 'if-then-else' or 'while' block is used in the 'try-and-catch' procedure (e.g., Tatar & Mauss 2014).

The choice of a benchmark case study is highly relevant to the research problem as it permits the demonstration of the practical application of a complex system for evaluating hydrometeorological risks. By utilizing regularly updated contingency tables and long-term analyses, the study directly addresses the need for a comprehensive, real-time system to monitor and forecast risk events. Additionally, by examining the system's effectiveness in a real-world scenario, the case study provides valuable insights into its ability to address real challenges related to hydrometeorological risks.

2 Methodology

The methodology employed for assessing forecast quality incorporates nine fundamental attributes, as delineated by WWRP/WGNE (2017): Bias, Association, Accuracy, Skill, Reliability, Resolution, Sharpness, Discrimination, and Uncertainty. Here is a brief explanation of each index:

- **Bias:** This metric quantifies the congruence between the mean forecast and the observed mean.
- **Association:** It gauges the potency of the linear relationship between forecasts and observations, often expressed through the correlation coefficient.
- **Accuracy:** This parameter denotes the degree of concordance between the forecasted values and the actual observations, with the disparity representing the error.
- **Skill:** It assesses the relative accuracy of the forecast in comparison to a baseline forecast, considering the cognitive capabilities of the forecast system.
- **Reliability:** This attribute measures the average agreement between forecasted and observed values, which may vary depending on forecast stratification.
- **Resolution:** It evaluates the forecast's capacity to categorize events into subsets characterized by distinct frequency distributions.
- **Sharpness:** This characteristic describes the propensity of the forecast to predict extreme values.
- **Discrimination:** It quantifies the forecast's efficacy in differentiating among various observations.
- **Uncertainty:** This index reflects the variability inherent in observations, which can influence the complexity of forecasting endeavors.

These indices guide the selection of charts and analysis methods in our case study. Our methodology involves updating contingency tables every 15 minutes to cover a 24-hour analysis period, with cumulative tables compiled daily at midnight for long-term evaluation. We utilize Relative Operating Characteristic (ROC) curves to assess the system's performance against observational data proxies for landslide occurrence, derived from generalized linear regression models based on accumulated precipitation. The area under ROC curves serves as a robust metric for evaluating both the operational status and potential development enhancements of our nowcasting system. Furthermore, confidence scores derived from contingency table elements undergo temporal smoothing using digital filters to accurately capture temporal variations.

Numerous researchers have conducted extensive investigations into environmental concerns and risks in Rio de Janeiro and its Metropolitan Region, yielding valuable insights into the socio-environmental dynamics of the area. For example, Malta and Costa (2021) undertook an assessment of vulnerability to socio-environmental risks in Rio de Janeiro, likely utilizing an index that integrates social and environmental indicators.

Their study aims to understand vulnerability by analyzing factors such as socioeconomic status, infrastructure, and exposure to natural hazards, with potential implications for disaster preparedness and urban planning. Similarly, Sandholz et al. (2018) explored ecosystem-based strategies for mitigating landslide risk in Rio de Janeiro, emphasizing the importance of nature-based solutions.

The Metropolitan Region of Rio de Janeiro faces numerous severe environmental, social, and economic challenges, including thousands of hydro-meteorological risk areas, and water assessment issues (e.g., Bourguignon et al. 2023). These studies, among others, contribute to our understanding of environmental challenges in the region, including the impacts of environmental degradation on water, air, and soil quality. They underscore the importance of further research into local communities' perceptions of environmental risks and the effectiveness of mitigation policies.

2.1 Case Study Selection

In this study, a case study was chosen to demonstrate the system's capabilities during the transition from normal operating conditions (low risk) to a high-risk scenario involving landslides and river overflow. Specifically, it enables the assessment of temporal and spatial variations in risk indicators, emphasizing their susceptibility to spatial structure and temporal changes in precipitation

patterns. This research investigates the efficacy of a weather forecasting system during severe storms and rainfall accumulation in the Baixada Fluminense, located in the Metropolitan Region of Rio de Janeiro (MRRJ), on January 11th and 14th, 2024. The primary focus of this analysis is on the system's ability to predict and respond to hazardous weather events, particularly in storm-prone regions. By evaluating the system's performance under varying conditions, this study aims to provide insights into its reliability and effectiveness in enhancing strategies for disaster preparedness and response.

2.2 Contingent Table and Confidence Scores

In this section, we present a detailed description of the evaluation indices employed to verify and validate the proposed nowcasting system. While the methodology includes a broad range of indices to offer a comprehensive overview of the available evaluation tools, the results analysis focuses on a selected subset of these indices.

The contingency table is a fundamental tool for identifying forecast errors. Ideally, a perfect forecasting system would yield only hits and correct negatives, with no misses or false alarms. Derived from the contingency table, a range of categorical statistics offers insights into various dimensions of forecast performance. The contingency table used in this study is illustrated in Table 1. These indices collectively provide a comprehensive assessment of forecast accuracy and skill, considering various aspects such as detection, bias, reliability, and overall skillfulness.

Table 1 Contingent table.

Observed	Yes	No	Total
Forecast Yes	a (hits)	b (false alarms)	a + b
Forecast No	c (misses)	d (correct negatives)	c + d
Total	a + c	b + d	n = a + b + c + d

Commonly used evaluation indices based on contingency tables include:

Probability of Detection (POD or p_o) (range: 0 to 1, perfect score: 1) measures the fraction of actual events that were correctly predicted by the forecast, as given by Equation 1.

$$POD = p_o = \frac{a}{(a + c)} \tag{1}$$

False Alarm Ratio (FAR) (range: 0 to 1, perfect score: 0) indicates the rate of false alarms relative to the total number of forecasted events. This is computed using Equation 2.

$$FAR = \frac{b}{(a + b) + 0.1} \tag{2}$$

Threat Score (TS) or **Critical Success Index (CSI)** (range: 0 to 1, with 1 indicating a perfect score) assesses the accuracy of the forecast in predicting observed events, considering both hits and false alarms, as demonstrated in Equation 3.

$$TS = CSI = \frac{a}{(a + b + c + 0.1)} \tag{3}$$

Accuracy (AC) is a metric used to evaluate the correctness or precision of a forecast model, typically expressed as a value between 0 and 1. Also referred to as "fraction correct", it quantifies the proportion of accurate forecasts relative to the total number of forecasts made. A higher accuracy value, closer to 1, indicates greater precision and reliability in the forecast model's predictions. The accuracy is expressed by Equation 4.

$$AC = \frac{a + d}{(a + b + c + d) + 0.1} \tag{4}$$

Bias Score (BS) or **Frequency Bias or Bias Ratio (BR)** [0-1] assesses the correspondence between the mean forecast and observation mean, indicating the presence of systematic errors, as shown by Equation 5.

$$BS = BR = \frac{(a + b)}{(a + c) + 2} \tag{5}$$

Probability of False Detection (*F*) [0-1] measures the probability of incorrectly predicting an event when it does not occur, calculated using Equation 6.

$$F = \frac{b}{(b + d) + 0.1} \tag{6}$$

Success Ratio (SR) represents the ratio of successful forecasts to total forecasts, providing an overall measure of forecast accuracy (Equation 7).

$$SR = \frac{a}{(a + b)} \tag{7}$$

Heidke Skill Score (HSS) or **Cohen's *k* (k_{Cohen})** measure the improvement in forecast skill relative to random chance, as shown in Equations 8 and 9.

$$HSS = k_{Cohen} = \frac{(a + d) - (expected\ correct\ random)}{n - (expected\ correct\ random)} \tag{8}$$

where

$$(expected\ correct\ random) = \frac{\sqrt{(a + b)(a + c) + (d + b)(d + c)}}{n + 0.1} \tag{9}$$

Associated Skill Score measures the relative accuracy of the forecast compared to a reference forecast, accounting for the forecast system's skill. For this calculation, the reference model is a *uniform distribution model*, as defined in Equation 10.

$$MSE_{ref} = e^{-1} \approx 0,367879441 \approx 36.79\% \tag{10}$$

Cumulated Mean Square Error measures the accumulated error between forecasted and observed values over time, assuming a *Probabilistic Persistence Binomial Model*, as described in Equation 11.

$$Associated\ Skill\ Score = HSS \tag{11}$$

Cumulated Mean Square Error (*MSE*) represents the accumulated error between observed and modeled values over a given period. It is calculated using indicator functions (and), which assign values of 0 or 1 based on the presence or absence of observed and modeled landslides, respectively. The *MSE* provides insights into the overall accuracy of the model, considering the cumulative discrepancies between observed and modeled outcomes (Equation 12).

$$MSE = (I_{obs} - I_{mod})^2 / n \tag{12}$$

Skill score (Skill) [0-1]: Represents the overall skillfulness of the forecast compared to a reference forecast (Equation 13).

$$Skill = 1 - \left(MSE / MSE_{ref} \right) \tag{13}$$

Threat Score (TR) or **Critical Success Index (CSI)** measures the effectiveness of a forecast model in predicting events of interest, taking into account both successful forecasts and missed events. It is calculated as the ratio of the number of correct forecasts to the total number of observed events, including both hits and misses. The TR or CSI ranges from 0 to 1, where a higher score indicates better predictive accuracy and skillfulness of the model (Equation 14).

$$TR = CSI = a/(a + b + c + 0.1) \tag{14}$$

Randomic Success (a_{rand} or p_e) [0-1] metric evaluates the success of the forecast in predicting events relative to random chance. It represents an index indicating the level of success achieved randomly in a given forecast scenario (Equation 15).

$$a_{rand} = p_e = \sqrt{(a + b)(a + c)}/(n + 0.1) \tag{15}$$

Equitable Threat Score (ETS) or **Gilbert Skill Score** [0-1] is a statistic on hit prevalence function of three parameters (prevalence, sensitivity, specificity). Measures the improvement in forecast skill relative to random chance while accounting for the frequency of events (Equation 16).

$$ETS = \frac{(a - a_{rand})}{(a - a_{rand} + b + c) + 0.1} \tag{16}$$

Desproportion Ratio (DD) [0-1] indicates the degree of disproportionality between observed and forecasted events. A DD value close to 0 suggests a balanced proportionality between observed and forecasted events, while a higher DD value indicates a greater degree of disproportionality (Equation 17).

$$DD = \left| \frac{a}{(a + b + 0.1)} - \frac{c}{(c + d + 0.1)} \right| \tag{17}$$

Probability of False Detection (POFD) [0-1] measures the likelihood of incorrectly detecting an event that did not actually occur (Equation 18). A lower POFD value indicates a lower probability of false alarms, while a higher value suggests a higher likelihood of incorrectly detecting events.

$$POFD = \frac{b}{(b + d + 0.1)} \tag{18}$$

True Skill Statistic (TSS), Hanssen & Kuipers' Discriminant (HKD) or Peirce's Skill Score (PSS) [-1 to 1] provides a measure of overall forecast skill, accounting for hits, misses, false alarms, and correct rejections (Equation 19).

$$TSS = POD - POFD \tag{19}$$

Bias-Adjusted Threat Score (TSA) [-1 to 1] evaluates the skill of the forecast while accounting for systematic bias (Equations 20 and 21).

$$TSA = \frac{nume}{max(deno, 0.1)} \tag{20}$$

where

$$nume = (a + c)^{aux} - c\,t^{aux}$$
$$deno = (a + c)^{aux} + c^{aux}$$
$$aux = \frac{1}{max(0.1, BR)} \tag{21}$$

Odds Ratio (*OR*) is a statistical measure used to quantify the association between two events or conditions. It compares the likelihood of an event occurring in one group to its likelihood in another group. The formula for OR (as denoted by Equation 22) calculates this ratio, with values greater than 1 indicating a positive association, values less than 1 indicating a negative association, and a value of 1 indicating no association.

$$\text{OR} = \frac{POD}{PODF} \frac{(1 - POD)}{(1 - PODF)} \tag{22}$$

Odds ratio skill score (*ODDS*) (*Yule's q*) [-1 to 1] measures the improvement in forecast skill relative to random chance while considering the odds of events occurring (Equation 23).

$$ODDS = \frac{(ad - bc)}{max[(ad + bc), 0.1]} \tag{23}$$

Kappa statistic (*κ*) [0-1] assesses the agreement between observed and forecasted events while considering chance agreement. The equation of κ for a contingency table is given by Equation 24:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{24}$$

where represents the observed agreement between forecasted and observed events and represents the expected agreement due to chance.

These metrics offer a comprehensive evaluation of forecast performance by addressing various factors such as detection accuracy, bias, false alarms, and overall skill. Each metric provides unique insights into the quality and reliability of the forecast, allowing for a thorough assessment of model performance.

The indices are categorized into two main groups: **Accuracy Metrics** and **Skill Metrics**. **Accuracy Metrics** include the Probability of Detection (POD), False Alarm Ratio (FAR), Bias Score (BS), F Score (F), Probability of False Detection (POFD), and Sensitivity Ratio (SR). These metrics focus on aspects such as the system's ability to correctly identify events, the proportion of incorrect predictions, and the reliability of the forecast. On the other hand, **Skill Metrics** encompass Threat Score (TS), Critical Success Index (CSI), Heidke Skill Score (HSS), Mean Squared Error (MSE), Reference Mean Squared Error (MSEref), Skill Score, Random Accuracy (arand), Equitable Threat Score (ETS), Kappa Statistic (κ), Discrimination Distance (DD), True Skill Statistic (TSS), Threat Score Adjusted (TSA), and Odds Ratio (ODDS). These metrics provide insights into the overall effectiveness and skill of the prediction system, considering both correct detections and false alarms.

The indices TSS, TSA, and ODDS are commonly recommended to be presented together due to their complementary nature in providing a comprehensive evaluation of forecast performance. Conversely, POD and FAR are critical in the analysis of ROC curves, serving as coordinated axes for evaluating forecast accuracy.

- **For the analysis presented in this work, we focused on the following indices:**
- **Probability of Detection (POD):** Used to assess the system's ability to correctly identify the events of interest.
- **False Alarm Ratio (FAR):** Selected to measure the proportion of incorrect event predictions.
- **Threat Score (TS):** Used to determine the overall skill of the prediction system, considering both correct detections and false alarms.

The other indices mentioned in the methodology were included to provide a didactic context and review, allowing for a broader understanding of the various evaluation metrics available. However, the results discussed in this study are based exclusively on the indices **POD**, **FAR**, and **TS**.

2.3 Relative Operating Characteristic (ROC)

The evaluation and verification of weather forecasts play a fundamental role in continuously improving models and ensuring the accuracy of predictions. One essential tool in this process is the ROC Curve (Relative Operating Characteristic), which provides valuable insights into a

model's ability to discriminate between events and non-events. Some researchers have explored in detail the meaning and application of the ROC Curve in weather forecast verification, addressing both theoretical and practical aspects (Jolliffe & Stephenson 2012, WWRP/WGNE 2017, Brown et al. 2021).

The ROC Curve is a graphical representation that relates the hit rate (POD - Probability of Detection) to the false alarm rate (POFD - Probability of False Detection) at different probability thresholds. Through this curve, it is possible to assess the model's ability to distinguish between events and non-events. The ideal trajectory of the ROC curve goes from the bottom-left corner to the top-left corner and then up to the top-right corner, indicating perfect performance. A diagonal line represents a lack of discriminatory skill.

The Area Under the ROC Curve (AUROC) is a crucial measure in evaluating model performance. It provides a quantitative measure of the model's discriminatory ability, with values ranging from 0 to 1. An AUROC of 0.5 indicates random performance, while a score of 1.0 represents perfect performance. In addition to providing a general measure of the model's discriminatory ability, AUROC is also useful for comparing different forecast models.

The ROC Curve and the area under the curve (AUROC) are powerful tools in weather forecast verification. They provide valuable insights into a forecast model's ability to discriminate between events and non-events. By understanding and effectively applying these metrics, meteorologists can continuously improve their models and ensure more accurate and reliable predictions.

3 Results

Operational standards during high-risk days were assessed through case studies, focusing specifically on the disaster event that occurred between January 11th and 12th, 2024. The analysis provides a detailed overview of the unfolding event, highlighting key observations and preliminary conclusions. The chosen case scrutinized the specific meteorological scenario of sequential rainfall in the area to evaluate the responsiveness of the meteorological system. Operational challenges encountered during the transition from low-risk to high-risk days were documented, providing comprehensive narratives of the incidents, significant findings, and interim conclusions.

Below is a succinct overview of the case study findings, accompanied by relevant visual aids. These visual aids enhance the clarity of the case study results, providing a comprehensive understanding of the system's performance and the impacts of thunderstorms.

3.1 Case Study: Distributed Rainfall in Inhabited Complex Terrain

A case study was selected focusing on cumulative rainfall in the river headwaters of the densely populated alluvial plains within the metropolitan area of Rio de Janeiro:

- **Date:** Thunderstorms on January 11th and 14th, 2024
- **High-risk occurrence date:** Early hours of January 14th, 2024
- **Location:** Rio de Janeiro City and Baixada Fluminense Municipalities
- **Alert status:** Very high-risk situation in effect
- **Casualties:** Numerous fatalities, injuries, illnesses, evacuees, and displaced individuals

Between January 11th and 15th, 2024, Hydroestimator data indicated precipitation accumulations of approximately 40 to 60 mm, followed by a significant increase to 100 mm in the early hours of January 14th, 2024, as illustrated in Figure 1. The accumulations derived from Hydroestimator, based on area averages of 4 km by 4 km, were lower than those recorded by rain gauges during the event. For instance, the São João do Meriti pluviometric station, maintained by CEMADEN/MCTI, recorded 325 mm over the same 4-day period (Brazil, 2024).

On January 11th, a severe storm brought heavy precipitation to the Baixada Fluminense (Figure 1A), which later extended to the northern quadrant. This high-intensity rainfall caused soil saturation in the exposed urban areas, setting the stage for the flooding disaster that occurred on January 13th and 14th. On January 12th, 2024, rainfall was concentrated in the river headwaters of the Baixada Fluminense, located in the mountainous regions to the north and northwest of the alluvial sedimentary plain where the densely urbanized area of Baixada Fluminense is situated (i.e., north of the city of Rio de Janeiro) (see Figure 1B).

In the afternoon of January 13, 2024, the Baixada Fluminense (BF) area experienced another round of precipitation, though with reduced intensity until 5:00 PM local time. Concurrently, rainfall was recorded in the river headwaters to the northwest, with water being transported down slopes and flowing through the entire urban floodplain of the BF. These rivers, which have been extensively modified through channelization and rectification, ultimately discharge into Guanabara Bay, across from Ilha do Governador. Additionally, intense storms affected other regions of Rio de Janeiro (RJ), including the Lakes Region and Arraial do Cabo to the east of the Rio de Janeiro Metropolitan Region (RMRJ)
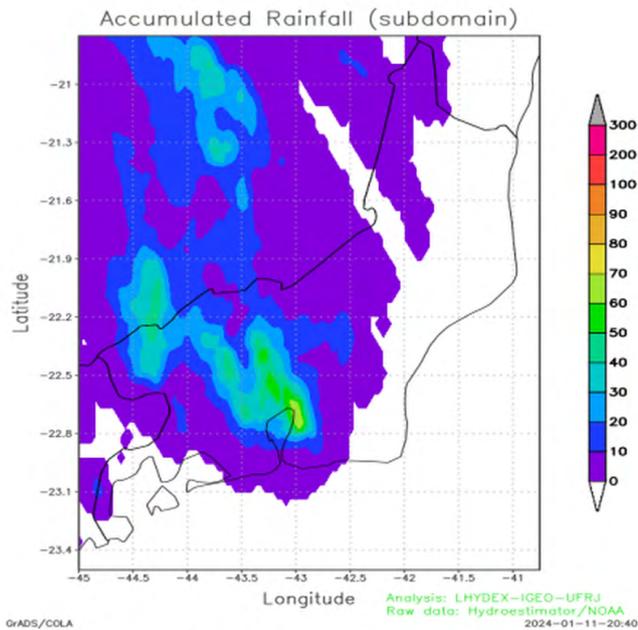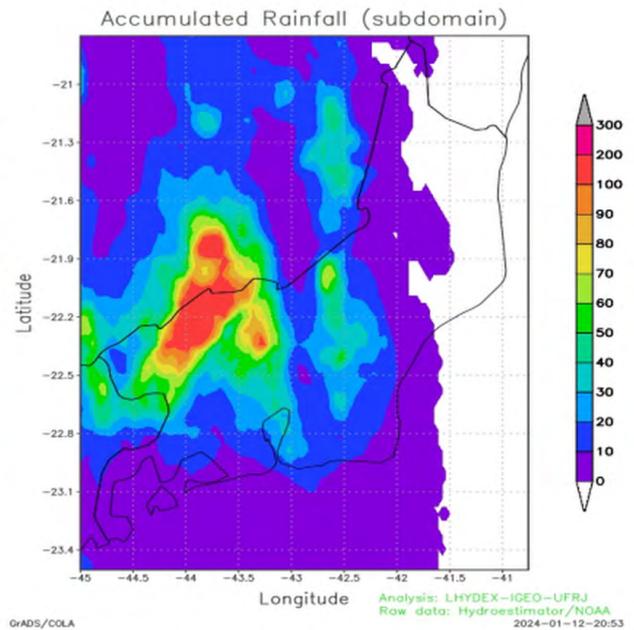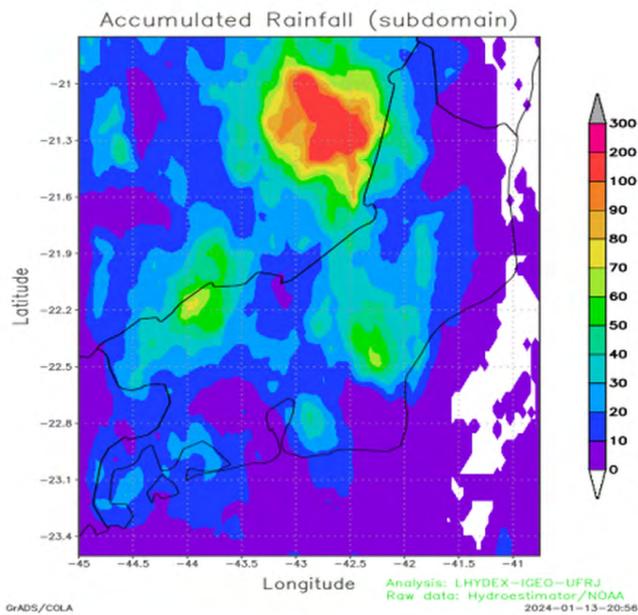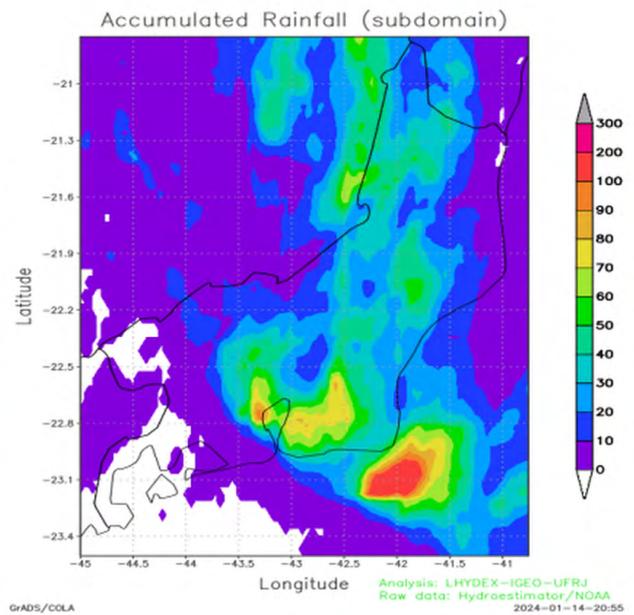
Figure 1 Spatial distribution of previous 24-hour Cumulated Rainfall (in mm) from Hydroestimator/NOAA, depicted in the mesoscale domain of the RJ State and neighboring states, during the four-day disaster event: A. January 11th at 20:40 (LT); B. January 12th at 20:53 (LT); C. January 13th at 20:56 (LT); D. January 14th at 20:55 (LT), 2024.

(Figure 1C). By late afternoon and early evening on January 13, 2024, the storms had spread from the southwest to the northeast across the state of RJ.

In the early hours of January 14th, 2024, at 01:00 AM local time, a storm developed over the Baixada Fluminense (BF) area, impacting São João do Meriti, Belford Roxo, Duque de Caxias, and other neighboring densely populated neighborhoods (Figure 1D). The storm intensified significantly, elevating the hydrometeorological risk of landslides, floods, and inundations to over 90% in the contingent model. This high risk persisted throughout the night and morning and continued to spread during the day along the entire coastline of the State of Rio de Janeiro, propagating along a convergence line of the flow at the edge of the state (not shown here).

3.2 River Height Monitoring

The hydrology of the Baixada Fluminense, located in the state of Rio de Janeiro, Brazil, is characterized by a complex network of rivers, streams, and canals of significant hydrological importance. Key waterways include:

- **Sarapuí River:** This is one of the primary rivers in the region, noted for its considerable length and crucial role in local hydrological dynamics. Together with the Iguaçu River, it forms the Iguaçu/Sarapuí Hydrographic Basin. Flooding often occurs due to soil impermeability and reduced space for water flow, which increases river volume and leads to overflow. The Sarapuí River is currently polluted, with a high contamination index.
- **Iguaçu River:** Distinguished by its extensive reach and abundant flow, the Iguaçu River is among the most significant rivers in the Baixada Fluminense, serving essential functions in water supply and drainage. Initially vital for the development of the Baixada Fluminense region and the city of Nova Iguaçu, its importance has waned due to deforestation and urbanization, leading to river silting and decreased flow.
- **Botas River:** This river meanders through various municipalities, making substantial contributions to the region's hydrographic network. However, it suffers from severe pollution caused by waste such as rubble, branches, and household garbage, which needs constant removal to prevent contamination of Guanabara Bay.
- **Meriti River:** Serving as a vital waterway for the Baixada Fluminense, the Meriti River provides essential resources and plays a key role in flood control and drainage. It has two tributaries: the Acari

River and the Pavuna River. The confluence of these rivers marks the border of Duque de Caxias and São João de Meriti. The Meriti and Pavuna Rivers also define the natural border between the Baixada Fluminense and the North Zone of Rio de Janeiro.
- **Pavuna River:** While primarily associated with Rio de Janeiro, the Pavuna River also intersects the Baixada Fluminense, enhancing its hydrological network. Flowing through the neighborhoods of Pavuna, Nilópolis, and Duque de Caxias, it reaches Guanabara Bay. Originating in the Swamp of Sítio do Retiro in the Bangu Mountain Range, it is 14 kilometers long.
- **Saracuruna River:** This river flows through Duque de Caxias to its mouth in Guanabara Bay, also passing through Nilópolis, Mesquita, Belford Roxo, Nova Iguaçu, and Duque de Caxias.

In addition to these major rivers, the Baixada Fluminense features an intricate network of smaller streams and channels that further complicate its hydrology. The hydrographic basin encompasses several municipalities, including Duque de Caxias, Nova Iguaçu, São João de Meriti, Belford Roxo, Mesquita, Nilópolis, Queimados, Japeri, Seropédica, and Magé. These areas are deeply influenced by the local hydrology, which plays a critical role in water management, flood mitigation, and overall environmental sustainability.

Efficient management and preservation of this hydrographic basin are essential for ensuring sustainable water use, mitigating flood risks, and maintaining ecological integrity. Britto, Quintslr and Pereira (2019) provide a detailed historical analysis of the Baixada Fluminense's hydrology, tracing the region's development from the 19th century to the 1930s. Their study highlights the pivotal role of the rivers in early prosperity, the subsequent economic stagnation due to negative portrayals of the area, and the transformative engineering interventions of the 1930s that spurred economic development. Currently, the rivers face severe contamination, requiring complex processes to achieve potability (Bockstaller & Girardin 2003; De Freitas et al. 2023).

3.3 Modeled Landslide Susceptibility Mapping

During the initial 10 hours of integration, from 8:32 PM local time on January 13th, 2024, to 6:32 AM local time on January 14th, 2024 (covering the night of January 13th to January 14th), the hydrometeorological risk of landslides consistently exceeded 50%, peaking at 8:32 PM on January 13th (see Figure 2). By 4:00 PM on January 14th, the risk decreased to 50%, but then rose to 70% by 7:00 PM

A. January 11th, 2024 at 23h12 (LT)   B. January 12th, 2024 at 17h37 (LT)
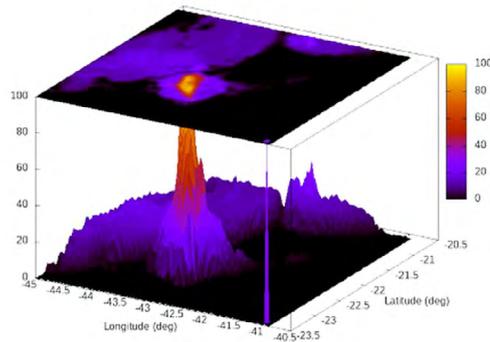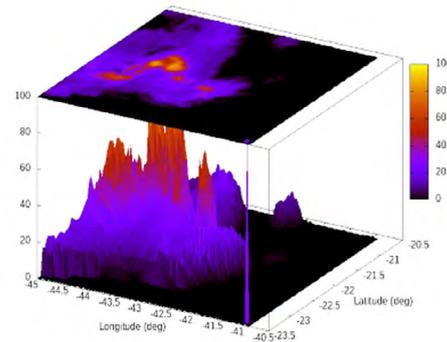
C. January 13th, 2024 at 16h38 (LT)   D. January 14th, 2024 at 04h32 (LT)
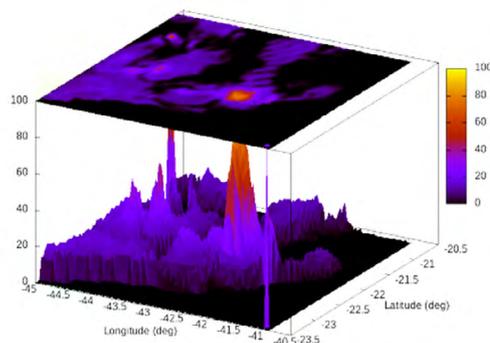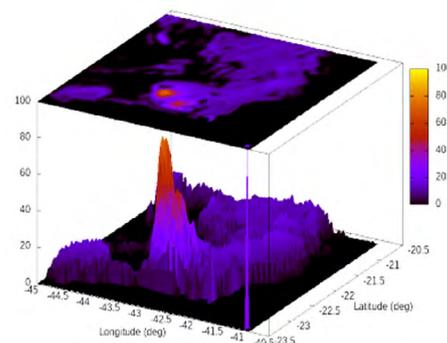
Figure 2 The three-dimensional spatial distribution of conditional landslide probability (ranging from 0 to 100%) is featured. The fields were computed by the modeling system through the assimilation of heterogeneous forcing by the variational contingent hydrological distributor. The analysis encompasses the mesoscale domain of the RJ State and neighboring states, capturing the dynamics during the four-day disaster event. Specifically, the conditions are depicted at key time points: A. January 11th, 2024 at 00h00; B. January 13th, 2024 at 00h00; C. January 14th, 2024 at 00h00; D. January 15th, 2024 at 00h00. All at local time (LT).

on the same day. In the hours leading up to these peaks, the dynamic risk model facilitated the assessment of risk escalation trends. For effective preparedness, rescue, and relief operations, it is crucial for environmental emergency managers to have sufficient lead time for issuing alerts. Even the evaluation of decreasing risk trends is vital for coordinating population rescue and relief efforts. The implementation of a dynamic risk model represents a significant advancement in developing a comprehensive regional and local hydrometeorological system.

In Figure 3, variations in risk are observed over the 24-hour duration of the event, particularly on January 14th, 2024. For disasters of this nature, there is a notable recurrence of risk at intervals of 6 to 8 hours between primary and secondary peaks. This pattern is similar to dynamic risk observed in earthquakes, where aftershocks follow the arrival of different seismic waves.

Figure 4 presents the graph of precipitation and its accumulated value as measured at the São João do Meriti pluviometric station in the Baixada Fluminense, maintained by CEMADEN/MCTI, for the period from January 11 to 14, 2024, covering the 4-day period (mm). Critical values of accumulated precipitation are observed, approximately double the expected value for January. The covariance between the observed values of precipitation rate and the accumulated value over 24 hours, considering they are spatial averages of 4 by 4 km, indicates an extremely high level of peril.

The theory of linear reservoirs provides the physical base for many bulk rainfall-runoff models (e.g., Beven 2012). Understanding the dynamics of water height excess in river systems with multiple linear reservoirs is essential for effective hydrometeorological risk management. During the studied flood event, data on river water height and flow rates were problematic. Consequently, estimates of river
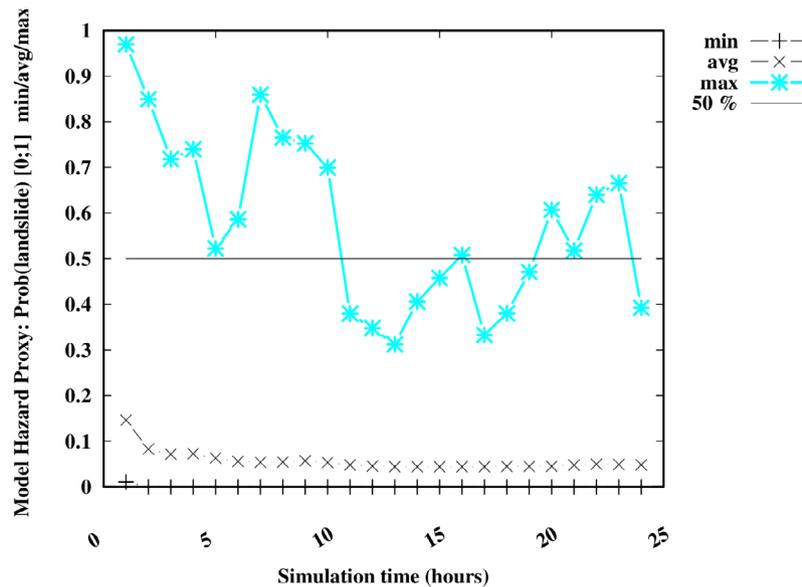
Figure 3 The Conditional Probability of Landslides (ranging from 0 to 100%) is depicted over the integration time, commencing on January 13th, 2024, at 8:32 PM local time, displaying minimum, average, and maximum values within the mesoscale domain of the State of Rio de Janeiro and neighboring states. The 50% probability threshold is denoted by the horizontal line.
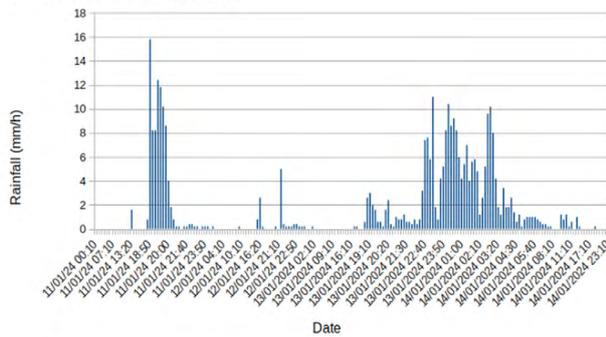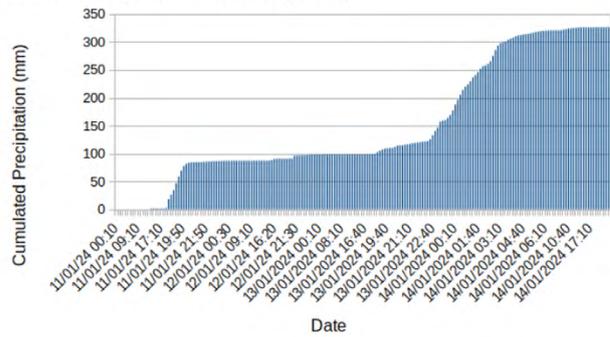
Figure 4 Precipitation measured at the São João do Meriti pluviometric station in the Baixada Fluminense maintained by CEMADEN/ MCTI for the period from January 11 to 14, 2024: A. Rainfall (mm/h); B. Accumulated precipitation over the 4-day period (mm).

height associated with water concentration in an idealized watershed are presented. This idealized watershed reflects the heterogeneous conditions expected in the Baixada Fluminense area. These conditions were approximately modeled using different time scales as parameters in a linear reservoir model.

This study investigates how varying the number of linear reservoirs in series and their associated time scales impacts water height excess. These insights are valuable for optimizing flood control strategies and enhancing water

resource management. However, it is important to note that the study area did not provide precise technical information on the maximum water heights observed during the flood event, which constrained the validation of our model.

3.4 Linear Reservoir Model for Risk Assessment

We implemented a computational model in Fortran-90 to simulate the flow and storage dynamics of linear reservoirs in series. The model uses variable time

steps, reflecting the time intervals between successive rainfall data entries. The precipitation data are from automatic weather stations (CEMADEM 2024).

The study examines four configurations of reservoirs: 1. Single reservoir with a 1-hour time scale; 2. Two reservoirs coupled in series with time scales of 1 and 3 hours; 3. Three reservoirs coupled with time scales of 1, 3, and 6 hours; and 4. Four reservoirs coupled with time scales of 1, 3, 6, and 9 hours (Figure 5). For each configuration, we calculated the reservoir storage and resulting water height excess over time, considering a hypothetical basin area of 25 square kilometers. The inflow to each reservoir is updated based on the outflow from the preceding reservoir and the time difference between observations.

The simulations reveal significant variations in water height excess among the different reservoir configurations. The single reservoir system showed the most rapid response to rainfall events, resulting in higher peak water heights over short durations. In contrast, the four-reservoir system demonstrated a more attenuated response, with lower peak heights spread over a longer period. The intermediate configurations exhibited behaviors between these extremes, with each additional reservoir contributing to a more gradual decrease in water height.

The results highlight the importance of reservoir configuration in flood risk management. Systems with multiple reservoirs in series provide enhanced attenuation of peak flows, reducing the immediate impact of heavy rainfall events. This underscores the need for tailored reservoir management strategies that consider the specific hydrodynamic characteristics and time scales of each reservoir in the system.

This analysis of water height excess over time for different arrangements of linear reservoirs in series demonstrates the critical role of reservoir configuration in modulating hydrodynamic responses to rainfall. These findings provide valuable insights for designing more effective flood control and water management strategies, ultimately contributing to improved resilience against hydrometeorological hazards. The absence of detailed technical data on the maximum water heights reached during the flooding event in the study area remains a notable limitation and highlights the need for better data collection and reporting in future studies.

The linear reservoir model is considered the simplest model to represent the hydrological response of rivers to precipitation, in unmonitored hydrological basins or even for imputation when there are flaws in the data, as occurred
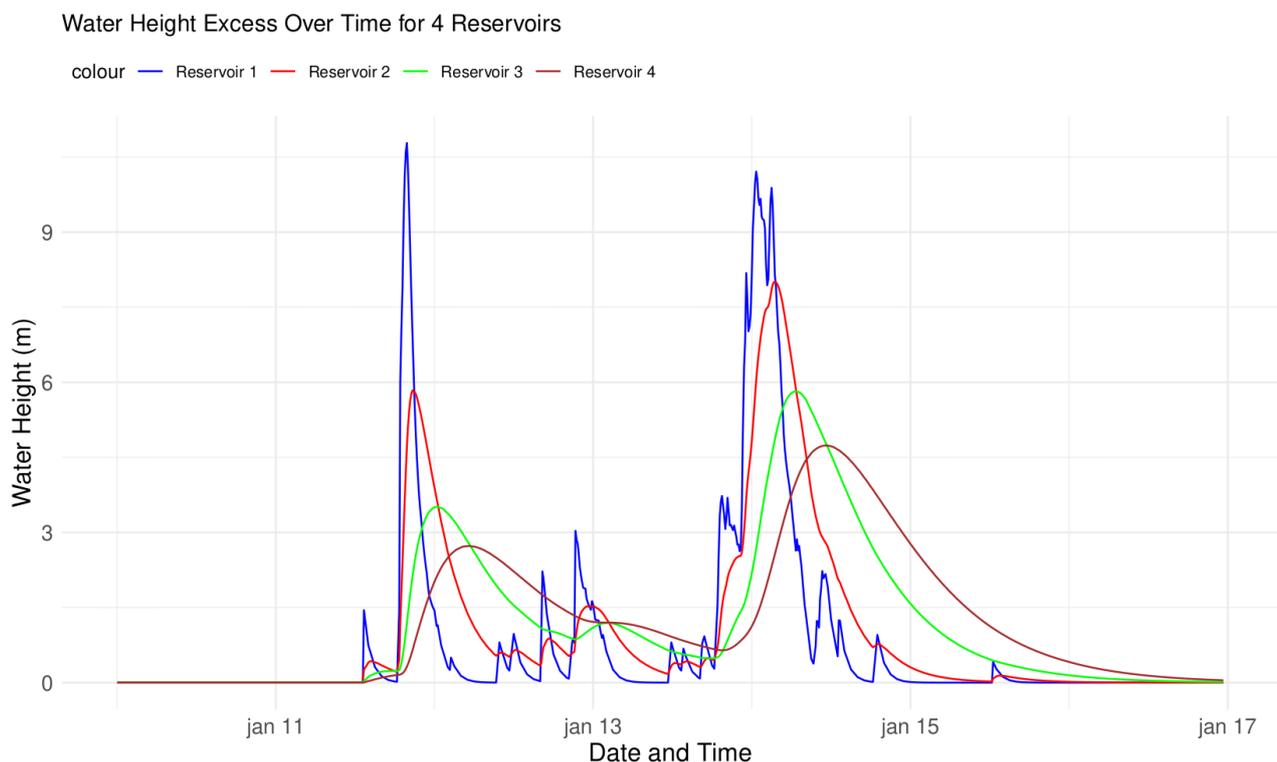
Figure 5 Water Height Excess over time for various configurations of Linear Reservoirs in series: comparative analysis of 1, 2, 3, and 4 reservoirs with coupled time scales of 1, 3, 6, and 9 hours.

in the case presented here. Time scales of 6 hours seem adequate for rivers in urbanized areas, semi-permeable, in a valley limited to the north and south by hill escarpments, as found in Baixada Fluminense. For operational purposes of evaluating the flood risk model, the linear model can serve as a proxy for predicting flood events and their impacts.

3.5 Combined Risks of Landslides and Flooding

The modeled risk, considered as a systemic association of landslides and flooding risks, was extremely high in the early hours of January 14, 2024, with the Probability of Detection (POD) of landslide exceeding 98% (Figure 6).

After 6:00 AM local time on January 14, 2024, the Probability of Detection (POD) decreased but remained stable due to continued soil saturation and ongoing rainfall of varying intensity. The POD stabilized slightly above 60% during the night of January 14, 2024. By January 15, 2024, the risk had decreased well below 50%, marking the end of the disaster period. Although graphs for January 15, 2024, are not presented, they indicated consistently low probability values throughout the day.

In summary, the POD demonstrated high values, indicating that the nowcasting system was highly effective in identifying the events of interest. The False Alarm Ratio (FAR) showed an acceptable level of false alarms, suggesting that the system effectively balanced detection with accuracy. Additionally, the Threat Score (TS) confirmed the overall skill of the system in predicting observed events, reflecting its ability to account for both correct detections and false alarms.

Failures in local measurement data and the lack of high-resolution satellite images before and after the flooding and landslides underscore the critical role of journalistic and official reports in revealing the disaster's extent, severity, and immediate impacts. In these circumstances, journalistic information becomes essential for achieving a comprehensive understanding of the environmental event. By integrating detailed descriptions, eyewitness accounts, and real-time data from news outlets, we were able to contextualize the event, assess its effects on affected communities, and guide mitigation and response efforts effectively. Although journalistic information may have limitations compared to scientific data, it significantly enhanced the analysis, contributing to a more complete understanding of the phenomenon.

On January 14th, 2024, heavy rainfall in Rio de Janeiro led to 12 deaths, primarily among elderly individuals with mobility issues, and displaced around 600 people. The President provided emergency support and resources,
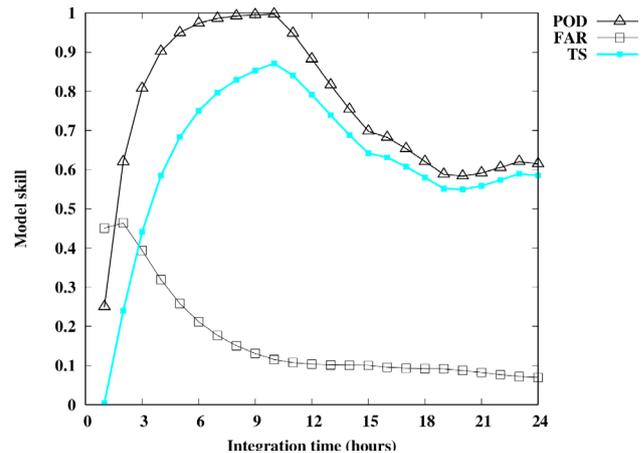
Figure 6 The temporal evolution (integration time) of the Probability of Detection (POD), False Alarm Ratio (FAR), and Threat Score (TS) from January 13, 20:30 to January 14, 20:30, 2024, as diagnosed by the conditional landslide model.

while a federal task force and various ministries assessed the damage and implemented measures such as adjusting the Bolsa Família Program payment schedule and advancing benefits. The Governor of Rio de Janeiro highlighted the urgency of street cleaning and support for the homeless, emphasizing the need for enhanced disaster prevention and response. Additionally, the Ministry of Cities proposed a R$ 780 million drainage project under the Growth Acceleration Program (PAC), including the Iguaçu Basin project, to mitigate future risks. The federal government also committed to accelerating support for affected regions and strengthening urban resilience against natural disasters (Brazil 247 News 2024; Brazil 247 News 2024a; Brazil 247 News 2024b).

The disaster on February 12th and 13th, 2024, was preceded by rainfall on January 11th, 2024, which prompted CEMADEN to issue high and very high-risk alerts for several municipalities, including Rio de Janeiro, São João do Meriti, Mesquita, and Nilópolis. This alert, based on forecast models predicting heavy rainfall for the Serrana Region, lasted over 24 hours. By February 12th, the alert level was increased for the coming days, resulting in significant casualties and displacement in the RMRJ and Baixada Fluminense (BF), including 12 deaths, 85 injuries (50% in Nilópolis), 926 displaced individuals (50% in Duque de Caxias), and 31 evacuated persons (20% in São João do Meriti).

The material damages in January 2024 were extensive. CEMADEN (2024) and Civil Defense reported 79 million Brazilian Reais in damages to public infrastructure, mainly in São João do Meriti. Additionally, damages

included 85.5 million Reais for public health facilities and 353 million Reais for housing units, with Japeri facing 296 million Reais in damages. Economic losses amounted to 9 million Reais in private losses and 9.5 million Reais in public losses, primarily for urban cleaning and waste management, with 6 million Reais allocated to São João do Meriti and Duque de Caxias.

CEMADEN (2024) also reported that the rainfall on February 13th and 14th, 2024, was twice the expected monthly average. This exceptional rainfall was attributed to several factors, including the summer rainfall period, temperatures above 30°C, high humidity, slightly elevated sea surface temperatures, the urban heat island effect, a strong El Niño, and global warming (Brazil 2024).

On January 12th, 2024, a cold front passed through Rio de Janeiro, causing rainfall between 13:00 and 14:00 hours (LT). Additional convective storms affected the city and Baixada Fluminense from 22:00 hours on January 13th until 00:00 hours on January 14th, leading to intense rainfall in densely populated alluvial plains. This resulted in 13 deaths and severe flooding. Rainfall ranged from 40 to 50 mm in some neighborhoods, exceeding 300 mm in areas like Pavuna and Anchieta. The highest accumulations were recorded in Pavuna, Anchieta, Mesquita, and Vigário Geral. Overflow from the modified rivers of the BF flooded low-lying areas, leading to a state of emergency in the eight most affected districts (Brazil 2024).

4 Discussions

The case study presented highlights the complex interactions between meteorological phenomena, hydrological dynamics, and human vulnerability during extreme weather events. This section provides a detailed discussion of the results and their implications.

The rainfall analysis reveals significant precipitation accumulations leading up to the disaster, with particularly notable totals of 100 mm recorded on January 14, 2024. This heavy precipitation, especially in densely populated areas such as Rio de Janeiro City and Baixada Fluminense, created conditions ripe for flooding and landslides. The discrepancy between the Hydroestimator-derived values and ground-based rain gauge measurements underscores the importance of accurate local data for effective risk assessment and disaster preparedness.

Landslide susceptibility mapping during the event indicated a heightened risk, peaking at over 50% probability. This underscores the critical need for timely risk assessment and intervention. River height monitoring, especially in areas such as Baixada Fluminense and Petrópolis, demonstrated the swift response of rivers to intense rainfall, reaching critical levels and leading to significant flooding in urban areas. This rapid hydrological response highlights the necessity for vigilant monitoring and effective early warning systems to mitigate flood risks.

Comparing forecast models with actual event data reveals the challenges in predicting localized extreme weather events. The models demonstrated varying degrees of accuracy, indicating the complexity of forecasting and the need for continual improvement in predictive tools. The impact assessment highlights the severe consequences of the rainfall events, including casualties, injuries, displacement, and infrastructure damage. Integrating scientific data with journalistic reports enhances the overall understanding of the event's impacts and supports more effective response efforts.

The discussion also addresses the importance of policy responses and mitigation measures designed to improve resilience and reduce vulnerability to future events. Initiatives such as drainage projects, urban planning improvements, and enhanced early warning systems are crucial for mitigating the impacts of extreme weather events. These measures aim to protect vulnerable communities and minimize the adverse effects of similar events in the future.

Comparing forecasting systems is crucial for advancing our understanding and improving the accuracy of severe weather predictions. The TITAN (Thunderstorm Identification, Tracking, Analysis, and Nowcasting) system, as detailed by Dixon and Wiener (1993), and employed in Brazil initially by Held *et al.* (2006), is specifically designed to track and predict thunderstorms' trajectories and intensities over short time-frames, typically ranging from minutes to hours. TITAN utilizes radar data to identify storm features, track their movement, and provide short-term forecasts of storm intensity. In São Paulo, TITAN has enhanced nowcasting capabilities for severe storms, though it lacks integration with vulnerability assessment models and does not include components for predicting landslides and flooding based on its forecast fields, at present.

In contrast, a variational assimilation-based expert system takes a different approach by integrating observational data with numerical models for forecasts. This system uses variational data assimilation techniques to improve forecast accuracy by dynamically updating model states with real-time data. It can incorporate detailed risk assessments related to population vulnerability, through risk polygons, and includes models for predicting landslides and flooding, offering a more comprehensive risk analysis expressed by Conditional Probabilities. However, this approach is more computationally intensive and relies on robust data assimilation processes. Future studies should explore these methodologies in greater depth to

fully understand their respective strengths and limitations. Evaluating nowcasting capabilities alongside variational assimilation-based systems could reveal how best to combine short-term forecasting with comprehensive risk assessment models.

5 Conclusion

In conclusion, the benchmark case study provides a vital perspective on disaster risk management by highlighting its complex and multifaceted nature. Through real-world examples, it elucidates the intricate interplay between meteorological, hydrological, and socioeconomic factors that influence both the occurrence and impact of disasters. By examining how these elements interact and contribute to community vulnerabilities, we gain a deeper understanding of the issues at hand.

The case study underscores the necessity of a holistic approach to disaster preparedness and response. It demonstrates that effective risk management extends beyond mere disaster reaction; it requires proactive measures that integrate scientific data, advanced forecasting tools, targeted policy interventions, and meaningful community engagement. By combining these elements, stakeholders can better mitigate risks, enhance resilience, and safeguard lives and livelihoods against increasingly frequent and severe extreme weather events.

Ultimately, the comprehensive analysis offered by the benchmark case study not only enhances our understanding of disaster risk management but also emphasizes the importance of adopting a multifaceted approach. Recognizing the interconnectedness of various factors and the need for collaborative efforts is essential for building more resilient and sustainable communities, better prepared to face future challenges.

6 References

Beven, K.J. 2012, *Rainfall-Runoff Modelling: the Primer*, 2nd edn, John Wiley & Sons, Hoboken.

Britto, A.L., Quintslr, S. & Pereira, M.S. 2019, 'Dossier: Rivers and Societies', *Revista Brasileira de História*, vol. 39, no. 81.

Bockstaller, C. & Girardin, P. 2003, 'How to validate environmental indicators', *Agricultural Systems*, vol. 76, no. 2, pp. 639-53.

Bourguignon, D.A.S., Fraga, M.S., Lyra, G.B., Cecílio, R.A., Pereira, C.R. & Abreu, M.C. 2023, 'Assessment of Water Quality and Identification of Priority Areas for Intervention in Guanabara Basin, Rio De Janeiro, Brazil, Using Nonparametric and Multivariate Statistical Methods', *Earth Science Research Network EarthSciRN*, vol. 196, no. 10, e-920.

Brazil 247 News 2024, *Ministry of Cities to include drainage project in PAC for region affected by rains in Rio de Janeiro*, viewed 16 February 2024, <https://www.brasil247.com/

regionais/sudeste/ministerio-das-cidades-deve-incluir-no-pac-projeto-de-drenagem-em-regiao-afetada-pelas-chuvas-no-rio-de-janeiro>.

Brazil 247 News 2024a, *Lula guarantees emergency support and resources to cities affected by heavy rains in RJ*, viewed 16 February 2024, <https://www.brasil247.com/regionais/sudeste/lula-garante-apoio-emergencial-e-recursos-as-cidades-atingidas-por-fortes-chuvas-no-rj>.

Brazil 247 News 2024b, *Number of deaths from Rio rains reaches 12*, viewed 16 February 2024, <https://www.brasil247.com/regionais/sudeste/chega-a-12-o-numero-de-mortos-pelas-chuvas-no-rio>.

Brazil 2024, *63rd CEMADEN/MCIT Impact Meeting*, viewed 16 February 2024, <https://www.youtube.com/watch?v=8xm1LTvEUYE>.

Brown, B., Jensen, T., Gotway, J.H., Bullock, R., Gilleland, E., Fowler, T., Newman, K., Adriaansen, D., Blank, L., Burek, T. & Harrold, M., 2021, 'The Model Evaluation Tools (MET): More than a decade of community-supported forecast verification', *Bulletin of the American Meteorological Society*, vol. 102, no. 4, pp. E782-E807.

CEMADEN 2024, *Estação Pluviométrica: Jardim Metrópole - Cidade: São João de Meriti/RJ*, Brasil, <https://resources.cemaden.gov.br/graficos/interativo/>.

Cook, D.A. & Skinner, J.M. 2005, 'How to Perform Credible Verification, Validation, and Accreditation for Modeling and Simulation', *CROSSTALK The Journal of Defense Software Engineering*, pp. 20-4.

De Freitas, A.S., Oliveira Santos, A.D., Santos, R.F., Nascimento, M.T.L., Fonseca, E.M., Félix, L.C., Bila, D.M., Aguiar, V.M.C. & Baptista Neto, J.A. 2023, 'Chemical and biological indicators of environmental pollution in the Canal do Cunha (Guanabara Bay, Rio de Janeiro, Brazil): Analysis and determination of toxins', *Journal of Coastal Research*, vol. 39, no. 6, pp. 1146-57.

Dixon, M. & Wiener G. 1993, TITAN: Thunderstorm Identification, Tracking, Analysis and Nowcasting - A Radar-Based Methodology', *Journal of Atmospheric and Oceanic Technology*, vol. 10, no. 6, 785-97.

Gentil, S. & Blake, G. 1981, 'Validation of complex ecosystems models', *Ecological Modelling*, vol. 14, no. 1-2, pp. 21-38.

Groesser, S.N. & Schwaninger, M. 2012, 'Contributions to model validation: Hierarchy, process, and cessation', *System Dynamics Review*, vol. 28, no. 2, pp. 157-81.

Held, G. Gomes, A.M., Naccarato, K.P., Pinto Jr., O. & Nascimento, E.L. 2006, 'The Structure of Three Tornado-Generating Storms Based in Doppler Radar and Lightning Observations in the State of São Paulo, Brazil', *Proceedings of 8 ICSHMO*, Foz do Iguaçu, Brazil, April 24-28, INPE, p. 1787-97.

Li, H., Kalnay, E., Miyoshi, T. & Danforth, C.M. 2009, 'Accounting for model errors in ensemble data assimilation', *Monthly Weather Review*, vol. 137, no. 10, pp. 3407-19.

Jolliffe, I.T. & Stephenson, D.B. 2012, *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, John Wiley & Sons, New York.

Kalnay, E. 2003, *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, Cambridge.

Malta, F.S. & Costa, E.D. 2021, 'Socio-environmental vulnerability index: An application to Rio de Janeiro-Brazil', *International journal of public health*, vol. 66, 584308.

Ota, Y., Derber, J.C., Kalnay, E. & Miyoshi, T. 2013, 'Ensemble-based observation impact estimates using the NCEP GFS', *Tellus A: Dynamic Meteorology and Oceanography*, vol. 65, no. 1, 20038.

Petty, M.D. 2018, 'Modeling and validation challenges for Complex Systems', in *Engineering Emergence*, CRC Press, Florida, pp. 199-216.

Sandholz, S., Lange, W. & Nehren, U. 2018, 'Governing green change: Ecosystem-based measures for reducing landslide risk in Rio de Janeiro', *International journal of disaster risk reduction*, vol. 32, pp. 75-86.

Stager, P. 1993, 'Validation in Complex Systems: Behavioral Issues', *Verification and Validation of Complex Systems: Human Factors Issues/NATO ASI Series*, Berlin, <https://doi.org/10.1007/978-3-662-02933-6_5>.

Tatar, M. & Mauss, J. 2014, 'Systematic Test and Validation of Complex Embedded Systems', *ERTS 2014 - Embedded Real Time Software and Systems*, Toulouse, France.

Teperi, A.M., Paajanen, T., Asikainen, I. & Lantto, E. 2023, 'From must to mindset: outcomes of human factor practices in aviation and railway companies', *Safety science*, vol. 158, 105968.

WWRP/WGNE Joint Working Group on Forecast Verification Research New 2017, *7th International Verification Methods Workshop: Forecast Verification Methods Across Time and Space Scales*, Berlin, Germany, pp. 1-89.

Author contributions
Hugo Abi Karam: conceptualization; formal analysis; methodology; coding; validation; writing – original draft; writing – review and editing; visualization.

Conflict of interest
The author declares no conflict of interest.

Data availability statement
Model data and reference datasets are available upon request from the author. Scripts and code can also be provided for scientific collaboration and development.

Editor-in-chief
Dr. Claudine Dereczynski

Associate Editor
Dr. Fernanda Cerqueira Vasconcellos

How to cite:
Karam, H.A. 2024, A Nowcasting System for Hydrometeorological Hazard Assessment of Landslides and Flooding – Part 2: On Verification and Validation Process, *Anuário do Instituto de Geociências,* 47:62979. https://doi.org/10.11137/1982-3908_2024_47_62979