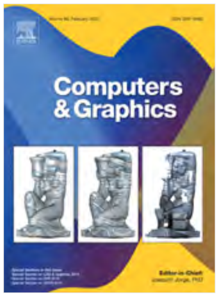
Journal Pre-proof



Visualizing simulation ensembles of extreme weather events

Carolina Veiga Ferreira de Souza, Priscila Cunha Luz Barcellos, Lhaylla Crissaff, Marcio Cataldi, Fabio Miranda, Marcos Lage

PII: S0097-8493(22)00007-3
DOI: https://doi.org/10.1016/j.cag.2022.01.007
Reference: CAG 3481

To appear in: *Computers & Graphics*

Received date : 30 July 2021
Revised date : 21 January 2022
Accepted date : 24 January 2022

Computers & Graphics (2022)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Visualizing Simulation Ensembles of Extreme Weather Events

ARTICLE INFO

Article history:
Received January 12, 2022

Visual analytics, Weather visualization, Ensemble visualization

ABSTRACT

In the last 20 years, extreme weather-related events like floods, landslides, droughts, and wildfires have caused the death of 1.23 million people and a loss of 2.97 trillion dollars. Studies show that low and lower-middle income countries are the most impacted ones given the lack of investment in disaster risk management. To reduce the impact of these events, weather researchers have been developing numerical weather models that inform public agencies about the impending extreme events in advance. Despite being powerful tools, these models can suffer from several sources of uncertainty, ranging from the approximation of micro-scale physical processes to the location-dependent calibration of parameters, which is especially critical in developing countries. To minimize uncertainty effects, researchers generate several different weather scenarios to compose an ensemble of simulations that typically are inspected using manual, laborious, and error-prone approaches. In this paper, we propose an interactive visual analytics system, called X-Weather, developed in close collaboration with weather researchers from Brazil. Our system contributes a set of statistics and probability-based visualizations that allows the assessment of extreme weather events by effortlessly navigating through and comparing ensemble members. We demonstrate the effectiveness of the system through two case studies analyzing tragic events that happened in the mountain region of Rio de Janeiro in Brazil.

1. Introduction

In recent times, the world has seen a dramatic increase in the number of climate-related disasters. Between 1980 and 1999, 4,212 reported disasters claimed the lives of 1.19 million people, with a total cost of over 1.63 trillion dollars. In the last 20 years, the number of reported disasters grew to 7,348, causing the death of 1.23 million people and more than 2.97 trillion dollars in damages [1]. This scenario can be attributed in part to the staggering rise in the number of extreme weather-related events, including floods, storms, landslides, droughts, and wildfires. By comparison, in the last 20 years, the number of flooding occurrences more than doubled: 3,254 versus 1,389 in 1980-1999. Studies show that these events disproportionately impact low and lower-middle income countries: while they experienced 43% of all major recorded disasters, they suffered 63% of the fatalities [2]. In Brazil, for example, two tragedies caused by extreme weather events caused the death of more than 1,000 people in the state of Rio de Janeiro. In April 2010, a severe storm in the metropolitan region caused landslides and floods that resulted in more than 200 deaths and displaced more than 15,000 people [3]. One year later, another storm in the mountain region caused the death of more than 900 people, with thousands displaced from their homes. This event is considered the worst climate-related disaster that happened in Brazil [4, 5].

Disaster risk management plays a key role in minimizing the catastrophic consequences of extreme weather events [6]. In the case of floods, being able to accurately forecast severe storms and downpours and adequately notify the population in a timely manner can save thousands of lives [7]. To this end, weather researchers have been developing numerical weather prediction (NWP) models that allow public agents to know beforehand about destructive events and enable the development of prevention plans to minimize environmental, material, and human disasters. Although these models are powerful tools, they can suffer from several sources of uncertainty. One of the approaches used by weather researchers is to then create an ensemble of simulations for a given region and time. For example, one can use different weather models [8], or use a single model with perturbed parameters (e.g., initial conditions, spatial and temporal resolutions, parametrizations). Building ensembles is interesting since probabilistic studies of the simulations members, which are individually deterministic, become possible and generally demonstrate better results than a single simulation [9]. We highlight the importance of weather forecast studies that consider different physical parametrizations, since the forecast results may differ from each other and, consequently, misinterpretations may occur. This situation is aggravated in the case of developing countries [10], which rely on weather models primarily developed for regions in North America and Europe.

One challenge of the ensemble approach is that each of its members is a multivariate spatiotemporal data set describing a different weather forecast. Since the combination of multiple factors can indicate a looming extreme weather event, it is paramount for weather researchers to analyze these forecasts not only across space and time but also across multiple variables. Manually inspecting these results, while necessary to make sense of forecast uncertainty, is exhausting and error-prone. For this reason, it is necessary to employ new strategies to facilitate the analysis of these ensembles.

In this paper, we propose X-Weather, a visual analytics system built in close collaboration with weather researchers in Brazil, interested in studying extreme weather events in the mountain region of Rio de Janeiro by investigating ensembles created through perturbations in physical features (i.e., physical ensembles). Therefore, the proposed tool allows the assessment of extreme weather events that can potentially lead to weather disasters by enabling effortless navigation through multiple weather ensemble members grouped by physical features and allowing their evaluation and comparison. Figure 2 presents an overview of the X-Weather interface. More precisely, our contributions are as follows:

- We introduce a set of statistics-based visualizations that allows weather researchers to easily identify the multiple weather scenarios contained in a large simulation ensemble, taking into account the inherent uncertainty of weather models.
- We introduce a set of probability-based visualizations that enables the assessment of extreme weather events by exploring the chances of observing target scenarios.
- We introduce X-Weather, a web-based system that enables the investigation of weather ensembles through the visual, interactive, and integrated evaluation/comparison of the multivariate spatiotemporal ensemble members.
- We demonstrate the effectiveness of the system through two case studies using simulations of extreme weather events that happened in the mountain region of Rio de Janeiro in Brazil.

It is important to reinforce that, although our web-based system is built on well-known visualization techniques, the proposed set of visualizations was designed to be familiar to weather specialists, while being a powerful tool that can be used to obtain nontrivial insights into ensembles of weather simulations.

2. Related Work

Visualization and visual analytics enable complex data investigation that allows identifying patterns, trends, and outliers in weather data. This is an important area that has seen numerous research papers in the past few years, including visualization of aviation weather [11], vector fields [12] and iso-contours [13], weather forecasts [14] and climate simulation [15], computational fluid dynamics [16], scientific data in general [17, 18], and similarity exploration of climate data [19]. In particular, recently, several ensemble visualization systems have been developed to help experts in different areas. These include systems for network security [20] and public health [21], and systems that leverage biomedical images [22] and time-varying data [23]. Due to the complexity of the data, ensemble visualization faces a variety of research challenges [24]. Wang et al. [25] presented a complete survey of visualization and visual analysis of ensemble data, discussing how traditional visualization techniques have been adapted to handle the specificities of ensemble data.

Rautenhaus et al. [26] presented a detailed survey with state-of-the-art techniques in meteorological data visualization. The authors draw attention to the fact that, sometimes, domain experts are not open to interactive functionalities and novel visualization metaphors, like those in 3D. They are more familiarized with line-command tools (e.g., Ferret [27], GrADS [28], GMT [29]) or general programming languages (e.g. Matlab, Python). In this regard, visualization systems' developers must be aware of the domain's demands and concerns, and concentrate efforts on attracting and encouraging data exploration. Potter et al. [8] presented Ensemble-Vis, a framework that supports visual analysis of weather ensemble data through a combination of statistical visualization techniques and user interactions. The system provides a view of the data that enables experts to perform analysis at multiple scales from high-level abstraction to the direct display of data values. The goal is to enable the user to explore the general results and the results from each member of the ensemble in spatial and temporal dimensions for different atmospheric variables. Sanyal et al. [30] created Noodles, a tool to visualize ensemble uncertainty of a weather event data set using glyphs, ribbons, and spaghetti plots. The authors demonstrated their work with an ensemble composed of only 18 members of the 1933 Superstorm simulation, representing the standard deviation, interquartile range, and the width of the 95% confidence interval of the data. In another direction, Diehl et al. [31] developed a system for the visual analysis of data from weather forecasts that allow in-depth studies of selected areas and the comparison between simulated outputs and observed data. This web-based tool provides a timeline with an integrated map view, a forecast operation tool, a curve-pattern selector, spatial filters, and a linked meteogram. In a more recent paper, Diehl et al. [32] created Albero, a system focused on probabilistic weather forecasting analysis. This tool helps to identify patterns, trends, and their associated errors in the forecast model. Besides that, the system improves decision-making and simplifies the measure of forecast uncertainty. Biswas et al. [33] and Wang et al. [34] proposed analysis tools for three ensembles, each one including 150 members built using different calibrations of the same physical parametrization scheme. Rautenhaus et al. [35] present Met.3D, a robust open-source tool developed with the initial purpose of assisting air route planning, but also allowing ensemble investigation. The tool offers statistical and probabilistic methods applied mainly to three-dimensional structures. As two-dimensional images are very common in domain-specific tasks, the authors added 2D functionalities linked to the 3D visualizations. Santos et al. [36] and Williams et al. [37] introduce UV-CDAT, a system that integrates several tools (e.g., Python, ParaView, VisTrails [38]), to allow the analysis of a large collection of climate data.

Another important aspect of our work is the consideration of the underlying data uncertainty. Previous work has tackled this challenge by proposing visualization of summary statistics [39, 40], considering geospatial data [41]. A complete review of uncertainty visualization can be found in Broadlie et al. [42] and Bounneau et al. [43], and taxonomy of uncertainty visualization can be found in Potter et al. [44]. In the weather domain, uncertainty is particularly important, and different studies have analyzed its impact when taking into account global temperature [45], climate change [46], and different climate variables [47, 48, 49].

In terms of weather forecasters, Novak et al. [50] presented a survey of US operational forecast managers regarding the communication of forecast uncertainty, highlighting the need to address uncertainty information in weather ensembles. Schumacher and Davis [51] presented an analysis of heavy rainfall events (and their uncertainties), also highlighting in their conclusion the need to better inform about forecast uncertainty. Nadav-Greenberg et al. studied different common visualizations to understand their impact on the decision-making process of weather forecasters [52], highlighting the importance of understanding user interaction and forecasting tasks. They also highlight that trust in forecasts is very important, as wrong decisions can create false alarms and safety problems due to non-compliance.

In summary, previous works greatly contributed to the understanding of weather forecast models, and also highlighted the importance of taking into account domain-specific needs in the assessment of uncertainty during the weather-forecast decision-making process. However, they focused on different goals: sensitivity of parameters [32, 34, 33]; uncertainty analysis [31, 33, 35, 30]; general and broad investigation of ensembles and their members individually [8, 36, 37, 35]; the path of vector variables over time [47]; the comparison with observed data [31]; improving weather forecasting using neural networks [53]; and developing techniques for weather modeling with ensembles for forecasting extreme events [54]. They do not target the discovery of risks of extreme rainfall events from groups of members of a physical ensemble. To the best of our knowledge, no other system tackles this problem. In other words, none of them were designed to facilitate 1) the understanding of large ensembles, with members built using different physical process parametrizations; and 2) the effects of these parametrizations in the prediction of *extreme* weather events.

To better understand the impact of parametrizations in the predicted scenarios and interpret the chances of observing heavy precipitation values, it is important to analyze groups of ensemble members that share a parametrization. For this reason, our design privileges the visualization of collections of members instead of individual simulations. We stress that this problem is extremely relevant for developing countries, especially Brazil, given its climate influenced by the Amazon region, the atmospheric characteristics of the South Hemisphere, and the occurrence of cold fronts and convection rains.

3. Background

Numerical models. Mathematical models are usually employed to represent weather phenomena. Weather and climate numerical models, for instance, use physics-based equations to represent the state of the atmosphere, following Newton's Second Law, Thermodynamics laws, and conservation of mass. Since they do not have an analytical solution, they are solved through numerical methods. Climate models are usually used for global simulations using long time ranges, such as weeks, months, or even years. Weather models, on the other hand, are specific to a region and phenomena that can occur in minutes, hours, or days. The Weather Research and Forecast (WRF) model, developed at the National Center for Atmospheric Research (NCAR) and first introduced in 2000, is a numerical weather prediction (NWP) model widely utilized by numerous universities and research centers [55]. WRF's adoption is mostly driven by a few factors: it is provided without cost, including no restrictions on modifications; it is highly portable, able to run on several platforms, from laptops to supercomputers; and it disposes of a host of tailored capabilities, from air chemistry [56] to solar and wind energy [57, 58].

In order to perform a single weather simulation using the WRF model, a user (e.g., weather researcher) must define the initial and boundary conditions that describe the atmospheric state in the time and location of interest. Although the definition of these conditions is complex, historical data describing atmosphere states all over the world are available in the Global Forecast System (GFS) [59] and can be directly used by simulations performed using the WRF model. One important source of uncertainty is that these initial conditions depended on in-situ measurements, highly susceptible to calibration errors and instrument precision. Beforehand, a $n_x \times n_y$ grid covering the region of interest, the start/end dates and number of time steps n_t of the simulation must be provided. The simulation results are given in terms of the variables that describe atmospheric conditions, such as temperature, pressure, wind, and precipitation.

Parametrizations and ensembles. The weather behavior depends on micro-scale physical processes that, due to its complexity and computational resource limitations, are approximated by parametrizations. A parametrization is basically composed of a set of algorithmic or statistical approximations of a physical process; given its complexity, the same process can be described by different parametrizations, each introducing different levels of inaccuracy to the simulated results.

Given the different sources of uncertainty in a weather model, experts need to adopt strategies to minimize the possibility of misjudging a result. One common practice is to run an ensemble of simulations for the same region and period of time, each simulation with a different characteristic (e.g., initial conditions, domain and temporal discretizations, parametrizations). Ensemble analysis supports studying the probability of observing special weather events based on the proportion of simulations that predict a target scenario.

According to Rautenhaus et al. [26], a usual practice in weather forecasting is to simulate the whole ensemble at low space/time resolutions and the most promising member at higher resolutions. Although ensembles with different physical parametrizations are more common outside the context of operational weather forecasting [26], we highlight the importance of encouraging weather forecast studies that consider this type of perturbation, since the forecast results may differ from each other and, consequently, misinterpretations may occur. We chose this type of ensemble since the success of atmospheric modeling in extreme event detection depends mainly on the relationship between the chosen physical parameterizations and the nature of the atmospheric phenomenon [60]. This has been observed in practice by two domain experts with over 20 years of experience – both of them are co-authors of this paper.

Analysis workflow. The usual weather data analysis workflow can be summarized in four main tasks. First, the weather forecaster sets up the proper parametrizations, and initial conditions for the simulation, leveraging domain expertise and especially their knowledge of the region of interest. Second, the scientist runs the ensemble of simulations. The output of the simulation is then visualized as static plots using standard tools, such as GRADS or UV-CDAT. During this exhaustive process of analyzing the simulation outputs, manually going through potentially *several hundred* different maps, the researcher is able to determine if there is a chance of a target weather event in the region of interest. Even though popular tools facilitate this workflow in some capacity (by providing mechanisms to slice and dice, or aggregate the data) it still boils down to a manual, laborious, and error-prone process of visualizing and comparing a very large set of static maps.

Challenges. Ensemble data contains multiple dimensions (e.g. variable, space, time, etc.) that must be explored by the experts to perform reliable weather predictions, which makes weather ensemble analysis a complex task. General purpose tools (e.g., GrADS, Python) do not support a broad and off-the-shelf investigation of ensembles, so answering tasks like "the identification of ensemble members that represent scenarios with a high volume of rain", would require individually browsing through a large collection of members or employing an ad-hoc strategy that may require programming skills. Also, two main challenges of analyzing ensembles are to bring to light and democratize the access to information that is hidden in the large and complex mass of data that composes an ensemble.

In order to properly investigate ensembles, domain expertise is paramount to determine if input parametrizations have generated outputs with good representations of the underlying physical processes of atmospheric events. With this in mind, the expert needs to be constantly aware of the parametrizations input throughout the analysis of the ensemble's outputs, which is not easily possible with well-known tools.

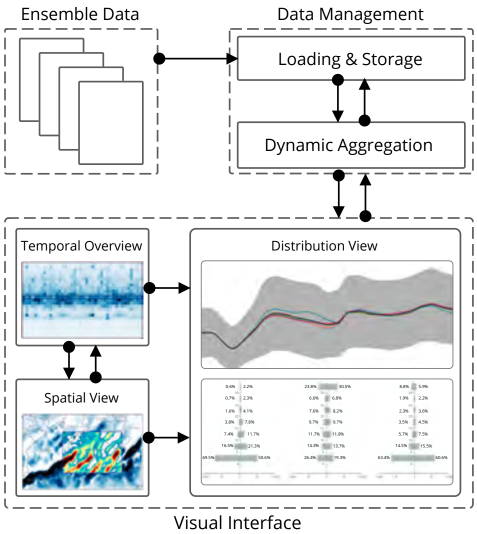


Fig. 1. X-Weather is a web application composed of a data management component (see Section 6) and a visual exploration interface (see Section 7). The data management component is responsible for loading, storing, and dynamically aggregating the ensemble data. The visual exploration interface implements several linked visualizations and interactions that facilitate the analysis of weather ensembles.

4. Requirements

In our collaboration with weather scientists and forecasters, we had several meetings and sessions where we established a set of requirements for a visual analytics system in order to facilitate their analysis workflow. During these meetings, we identified two main tasks that the experts want to perform with the tool: 1) identify the multiple weather scenarios contained in a large ensemble of simulations produced, taking into consideration the sources of uncertainty inherent to weather simulation models, especially the ones introduced by the parametrization of micro-scale physical processes; 2) assess the occurrence of extreme weather events, using the ensemble data to estimate the probability of observing target situations (e.g., the occurrence of accumulated precipitation greater than 20 mm in a period of 3 hours). In order to accomplish the listed tasks, we identified that our system should satisfy the following requirements:

[R1] Support the exploration of spatiotemporal patterns. Explore the spatial and temporal patterns of the multiple output variables of the ensemble members, so the forecaster can identify regions and time periods to which they should focus their attention.

[R2] Support the ensemble members comparison. Compare predictions of different ensemble members, so the forecaster can contrast different weather scenarios.

[R3] Support the analysis of the weather model's uncertainty. Analyze subgroups of ensemble members that share the same sources of uncertainty (e.g., group the members according to the parameterization of a given micro-scale physical process).

[R4] Support the exploration of target events probabilities. Assess the probability of observing target weather scenarios, especially extreme weather events like heavy rain and dry weather.

[R5] Support interactive response times. React to user actions in the interactive time since responses slower than 500 ms can significantly impact visual analysis, reducing the rate at which users make observations [61].

5. X-Weather System

In order to satisfy the previously detailed requirements, we propose X-Weather, a web-based visual analytics tool composed of two main modules: a data management backend, and an interactive visual interface. The data management backend is responsible for managing the weather simulation ensemble data and handling the interface queries. The visual interface implements several visualizations and user interactions that enable the visual exploration of the ensemble. Figure 1 shows an overview of the system. We briefly describe these modules next.

Data management. Our system supports the interactive exploration of a large collection of simulation outputs (**R5**). We accomplish this by 1) efficiently storing the data in order to maximize coalesced memory access; and 2) making use of precomputed schemes that allow for the interactive computation of aggregates, including order statistics (e.g., percentiles). We detail this component in Section 6.

Visual interface. The visual interface was designed to support the investigation of weather simulation ensembles constructed using different parametrizations to approximate micro-scale physical processes over a region of interest and/or a user-defined subregion. This design choice brings to light risks of extreme rainfall events regardless of a specific choice of parametrization used to reproduce each physical process. In this sense and to support the exploration of spatiotemporal patterns, we designed an interface with three main components. The first component, *Temporal Overview*, is composed of heat matrices that display summary statistics (e.g., average, percentiles) or probability distributions (e.g., output variable greater than a certain threshold) of a subset of members of the ensemble (following **R1** and **R2**). The component allows the user to apply a temporal constraint by selecting a particular time step of interest. The second component, *Spatial View*, primarily satisfies **R1** and **R2** by allowing the expert to visualize and compare the spatial distribution of multiple ensemble predictions, considering summary statistics or probability distributions. In the component, the user can apply a spatial constraint by brushing a region of interest. The third component, *Distribution View*, consists of two views: a line chart showing mean and twice the standard deviation of ensemble members aggregated over time; and three histograms with the distribution of values of the time step of interest (center), and the previous and next time steps (left and right). This component satisfies requisites **R3** and **R4**. The components are detailed in Section 7.

6. Data Management

The data management backend is responsible for loading, storing, and dynamically aggregating the ensemble data in order to handle the interface requests. In what follows, we describe the strategies used to ensure that the server can handle the queries interactively, one of the requisites that X-Weather system should satisfy, as discussed in **R5** of Section 4.

Data loading and storage. Numerical weather models generate and store simulation outputs in NetCDF files. Different outputs are stored in a single file, but only a few of them might be relevant for analysis. For this reason, in this work, the outputs of interest were extracted from NetCDF files and stored as CSV files, which are reduced, light, and easily manipulated. When the backend starts, the content of the CSV files is stored in a one-dimensional row-major vector, with a straightforward indexing mapping between multi-dimension and linear positions. As we show next, this strategy accelerates the computation of the order statistics and interface requests since it favors coalescent memory access.

Dynamic data aggregation. The X-Weather system's visual interface requires on-the-fly computation of user-defined scenario probabilities and summary statistics (e.g., average and percentiles of the ensemble members). Probabilities and averages can be efficiently computed since it only requires access to the members' data. The computation of the percentiles, on the other hand, requires an additional step of sorting the data. Using our storage approach, we can accelerate this operation by copying chunks of data that are sequentially stored in memory.

Moreover, after the system is initialized, the user can apply spatial constraints and define a region of interest. When that happens, the backend filters the grid points of each ensemble member that should be considered during the aggregations. Precomputing strategies would require the use of advanced data structures such as Nanocubes [62, 63] or its extended version that supports the computation of order statistics [64]. Using a one-dimensional storage strategy we are able to interactively compute a time series with the percentiles of the output variables considering a subgroup of ensemble members predictions over the entire grid. In fact, we can compute a time series ($n_t = 25$) with the median of the precipitation values over the entire grid ($n_x \times n_y = 5,472$) of a subgroup with 40 ensemble members in 2 seconds on average. We observe that when the user defines a region of the grid to focus the analysis, the computation times are even faster and the queries are typically returned in less than 1 second. To accelerate the response times when the entire grid is considered, we cache the statistical summaries and probabilities of the output variables. The previous acceleration strategies, although simple, sufficiently satisfied our requirements, given the data set size and the case studies designed by our collaborators. We emphasize that larger data

Fig. 2. The X-Weather interface. (a) The Temporal Overview allows users to globally inspect the output variables of each simulation in each time step. (b) The Spatial View allows users to study and compare the spatial distribution of an atmospheric variable in two subsets of ensemble members at a particular instant of time. (c) The Distribution View enables a better understanding of the ensemble distribution using line charts showing the temporal distribution or histograms showing the probability mass functions of ensemble groups. (d) The menu allows the user to change the system parameters like the active output variable and aggregation function.

sets may require the adoption of more complex solutions (e.g., Nanocubes [62]).

7. Visual Exploration Interface

We worked closely with weather forecasters in the design of X-Weather's user interface in order to support the tasks described in Section 4. The results of our interview sessions with domain experts indicate that they usually shy away from using systems and frameworks offering too many options, visualizations, or widgets. The same occurs with 3D structures, as observed by Rautenhaus et al. [26]. Furthermore, another reason why 3D does not suit our purpose is that the experts were interested in inherently 2D outputs (e.g., surface-level precipitation). Our goal is to develop a system that experts are interested in and feel comfortable using it. Therefore, we have chosen well-known techniques to bring previously mostly inaccessible information to light.

The visual interface is composed of three components highlighted in Figure 2: (a) *Temporal Overview*, (b) *Spatial View*, and (c) *Distribution View*. The interface also contains a menu that allows the user to change the system parameters. When the system starts, those parameters have been previously selected by default, and the user can change them to perform the analysis. Thus, the interface is never empty.

In each one of these views, the simulations are organized in subsets, one for each available parametrization of a given physical process, chosen by the user in the menu. Such grouping allows the exploration of the ensemble from different perspectives and increases the chances of uncovering extreme weather events. In addition, the user can use the menu to select a global atmospheric variable (e.g., rain, humidity) that will be used to populate the visualizations.

7.1. Temporal Overview

This component is composed of heat matrices each one representing a subset of simulations. Each column of a matrix corresponds to a simulation, and each cell of a column an instant in time. Considering that a simulation output is, for a given time step, a set of values in the spatial dimension, a statistical summary (e.g., mean, percentile, standard deviation) or probability distribution defined by the user (e.g., probability of accumulated precipitation greater than 10mm) of these values will be calculated and assigned to the appropriate matrix cell. In other words, the matrices show a measure of the values predicted in space by each simulation in each time step.

The main purpose of the matrices is to allow for the visualization of an atmospheric variable over time according to each ensemble member (meeting **R1** and **R3**), coupled with a probability scenario investigation (**R2**). This property provides an overview of the existence of a risk of extreme events, the moment in which it might occur, and its proportion. This helps the weather forecaster identify and, consequently, further analyze the spatial components of a subset of simulations, avoiding unnecessary access to those that do not contribute to present useful information. Furthermore, the Temporal Overview enables the user to add temporal constraints by selecting specific time steps of interest, which will update both the Spatial View as well as the Distribution View (see Figure 3). It is important to notice that this component can produce effective visualizations of ensembles with a limited, but large, number of members (in the case studies we considered 160 members). In fact, only a few previous proposals successfully handle ensembles with comparable size [33, 34]. In order to support the visualization of larger ensembles, we could adapt the proposed visualizations by adding filtering strategies or zoom and pan interactions.

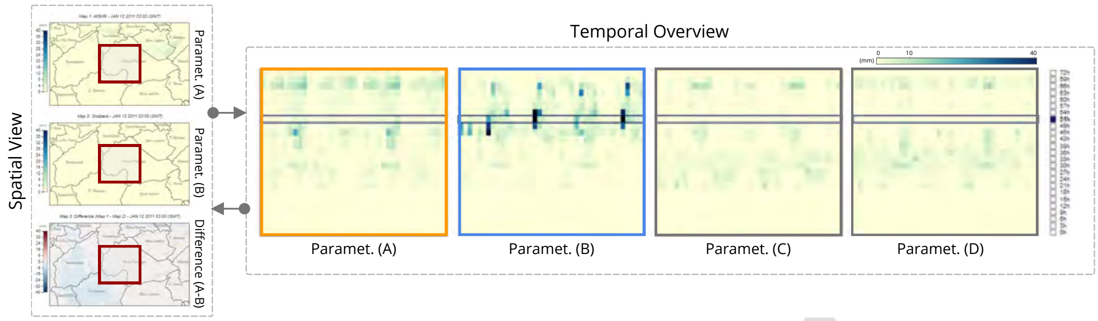


Fig. 3. Temporal Overview and Spatial View interactions. The user can define spatial constraints by brushing on the first map of the Spatial View. If a spatial constraint is active, the heat matrices of the Temporal Overview only consider the grid points inside the constraint. Similarly, the user can define temporal constraints by clicking on the labels of the Temporal Overview matrices rows. Also, by clicking on the matrices, the user selects the parametrizations that, together with the temporal constraint, are used to build the Spatial View. In this example, the Temporal Overview state reflects the visualization of the 160 ensemble members organized in groups (matrices). Each group was formed according to the parametrization used for cloud microphysics' physical process. That is, there are four matrices (members' groups), each one gathering forty columns (members that used the same parametrization) and twenty-five rows (time steps).

7.2. Spatial View

This component displays a set of heat maps showing the spatial distribution of an atmospheric variable at a particular time instant, enabling the weather forecaster to perform analyses of the ensemble data in the spatial dimension, primarily satisfying **R1**. Given an atmospheric variable, a selected time instant, and two groups of simulations in the Temporal Overview, the data from each simulation subset will be aggregated by grid point according to the active statistical summary (e.g., mean, percentile, standard deviation) or probability distribution defined by the user. Below the map of the two groups of simulations, this view will also display the difference between the two maps (see Figure 3(left)).

This view also provides a lens functionality: the user moves the lens, and the area within it shows a variable while the outside area shows another one. This is highlighted in Figure 4, with the visualization of different variables/metrics or the conditional probability of *another* scenario occurring for a second atmospheric variable, i.e., given that the scenario investigated in the maps occurred for one variable, what is the probability of a second scenario occurring simultaneously? This information is relevant mainly for the expert to relate the probabilities between two variables and, with their domain expertise about their characteristics, understand the real dimension of the risk of an extreme event. Again, it is possible to explore scenario probabilities (**R2**), comparison of ensemble member groups in the spatial dimension (**R3**), and spatiotemporal patterns (**R1**), since a spatial constrain updates the other views of the interface.

7.3. Distribution View

The Distribution View is composed of two different visualization widgets (shown in Figure 5) specifically designed to allow a better understanding of the underlying data distribution. In the first widget (Figure 5(top)), the ensemble data is aggregated in the spatial dimension, and grouped by simulations. Each group is represented by different line color and represents an active statistical measurement of the data over time (in the entire region or a region of interest if selected in the Spatial View). This visualization also allows the inspection of the active statistical measurement plus/minus twice the standard deviation associated with each distribution. This particular visualization allows the expert to identify outliers in time, which can indicate the occurrence of extreme events. It is important to note also that line charts are a visual metaphor known to the expert, which facilitates their analysis. By describing the behavior of ensemble member groups over time according to a region of interest, this visualization supports requirements **R1** and **R3**. The expert can also visualize the probability mass functions of two groups of simulations (Figure 5(bottom)). The histogram at the center of the widget represents the instant of time selected in the Temporal Overview; the histograms on the left and right correspond to the time step immediately before and after the time instant of interest, respectively. This is particularly important to the expert so that they can find time instants with the possibility of a severe event occurring based on a regional selection. This widget meets requirement **R1** for representing the entire region or a specific region in a certain time step, **R2** and **R4** for allowing the exploration of probabilities of previously established scenarios, and **R3** for allowing the comparison of groups of ensemble members.

7.4. Implementation

The X-Weather system was implemented following a web-based client-server architecture, such that the visual interface could be easily accessed by experts through a web browser, without the need of installing any additional software. We used NodeJS and Express to implement the backend and ReactJS and D3 to implement the front-end components of the system. For data preprocessing, we used Python 3, and the NumPy and netCDF4 libraries. The case studies were executed on a computer with an AMD Ryzen 7 3700X 3.6GHZ, 16GB RAM, and GeForce GT 210 1GB.

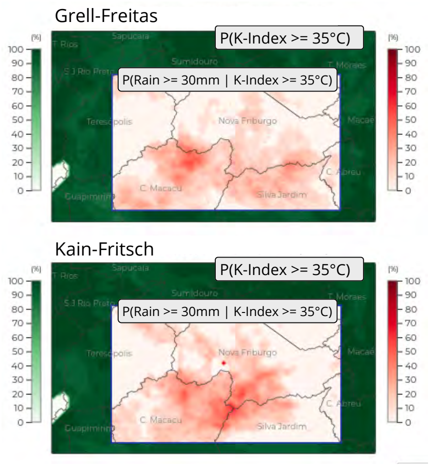


Fig. 4. Lens tool of the Spatial View. The tool allows the user to explore two output variables and/or aggregations methods simultaneously, one shown in the entire domain and the other inside the lens. Also, the tool can show the conditional probability of a target value of an output variable occurring given the probability of observing a given scenario of another variable.

8. Case Studies

To demonstrate X-Weather, our partner meteorologists used the system to study weather ensembles containing simulations of two intense precipitation events that occurred in the mountain region of Rio de Janeiro in 2011 and 2020. We emphasize that we used previously known extreme events instead of weather forecast data for a future date to highlight how the system would augment the analysis pipelines. In this way, the experts evaluated the system's effectiveness by comparing the information extracted from it with available data. In both events, severe rain caused a lot of destruction. In the 2011 episode, landslides killed several people and destroyed a number of buildings [4, 5]. The simulations were run using the 4.2.1 version of the WRF model [65]. The ensembles constructed for the case studies contain $n_m = 160$ simulations with different parametrization setups of five physical processes related to the development of storms: Cloud Microphysics, Cumulus Convection, Land Surface, Surface Layer, and Planetary Boundary Layer. The considered parametrizations for each of the previous physical processes are:

- Cloud Microphysics: WSM6, Kessler, Goddard, Eta (Ferrier);
- Cumulus Convection: Betts-Miller-Janjic, Grell-Freitas, Grell-3D, Grell-Devenyi, Kain-Fritsch;
- Surface Layer: MM5, MM5 Old;
- Land Surface: Noah MP, Dudhia 1996;
- Planetary Boundary Layer: MRF, MYNN3.

The simulations were run on a grid with $n_x = 96$ and $n_y = 57$ cells, composed of $n_t = 25$ time steps representing 3 hour intervals. For each simulation, $n_v = 7$ output atmospheric variables that can indicate the development of storms were produced: accumulated precipitation, the temperature at 2 meters from the surface, relative humidity at 850 hPa (850 hectopascal, i.e. 1.5 km above sea level), upward vertical wind at 500 hPa (5.5 km above sea level), divergence at 300 hPa (10 km above sea level), convergence at 850 hPa and the k-index (an indicator of atmospheric instability). Boundary and initial conditions were downloaded from GFS [59].

X-Weather was introduced to the experts in sessions lasting up to 15 minutes. They then spent up to 10 minutes extracting information, making decisions for each use case, and conducting the experiments without our help, which suggests that X-Weather was easy to operate.

8.1. *Extreme Rainfall Event in 2011*

On the night of January 11th, 2011 a system called South Atlantic Convergence Zone caused an intense storm, with 150 mm of accumulated precipitation in 24 hours, that devastated the mountain region of Rio de Janeiro and was considered the worst weather disaster in Brazil's history [4, 5, 60]. In the 7 days prior to the disaster, the area had already registered a persistent rain, which made the soil wet and unstable. On the event's night, satellite images showed the generation of clouds with substantial vertical development and potential for severe storms.

To explore the ensemble with X-Weather, the meteorologists first used the Spatial View component to select the region of Nova Friburgo, the region most impacted by the storm (see Figure 6(a)). Also, they configured the system to build visualizations using the 90th percentile of the rain atmospheric variable since this measure helps the investigation of extreme values. By grouping the ensemble members based on the Cloud Microphysics parametrization type, the Temporal Overview and the line chart of the Distribution View showed that the majority of the members predicted rain throughout the day (see Figure 6(b,c)). In fact, Kessler was the parametrization that better predicted the accumulated precipitation of the event.

Similarly, it was observed that the Betts-Miller-Janjic and the MYNN3 parametrizations predicted the highest amounts of accumulated precipitation among the parametrizations of Cumulus Convection and Planet Boundary Layer, respectively. The parametrizations of Land Surface and Surface Layer had a minor influence on the predicted accumulated precipitation. However, both the Temporal Overview and the line chart of the Distribution View showed that very few members predicted high accumulated precipitation *at the time of the event* (see Figure 6(b,c)). In fact, the mass probability function visualization of the Distribution View indicated that the probability of observing more than 20 mm of rain at the time of the event was close to zero (see Figure 6(d)). The rain output indicated the unlikely occurrence of an extreme event, and most likely warning systems would not be triggered if only considering this variable.

To properly study the occurrence of severe rain, the meteorologists must investigate not only the predicted accumulated precipitation values but also other atmospheric variables, such as the ascending vertical wind at 500 hPa and the convergence at 850 hPa. To do so, they observed the Spatial View component shown in Figure 6(e), built using the lens to show the 90th percentiles of the convergence at 850 hPa (entire region) and the ascending vertical wind at 500 hPa (lens region). These two variables indicate the existence of energy capable of raising the

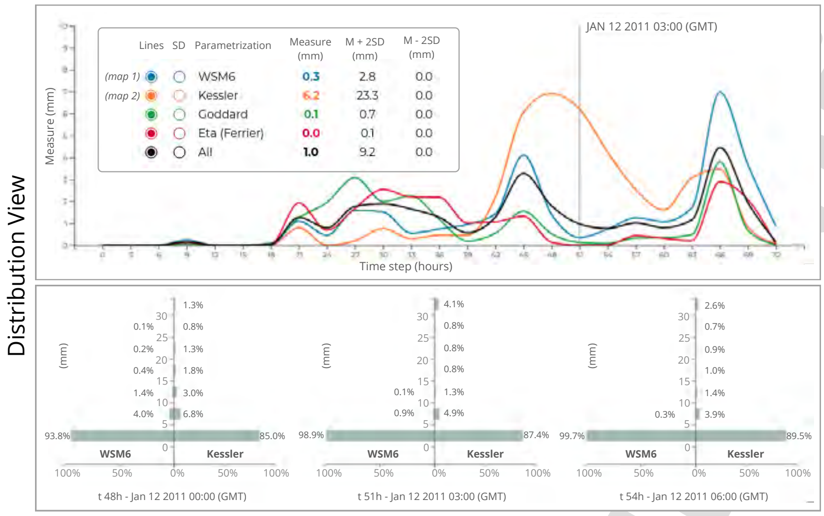


Fig. 5. The Distribution View can show two different visualizations. The line chart on the top is built by aggregating the ensemble members that use the same parametrizations in the spatial dimension over time. The color of the lines represents the different parametrizations. The standard deviation of each group of members can also be shown. The three pairs of bar charts on the bottom show the probability mass functions of two groups of simulations on the time step selected using the Temporal Overview (center) as well as on the previous (left) and on the next (right) time steps.

humidity to form rain clouds. In fact, the Temporal Overview shows that all ensemble members predicted close to 100% relative humidity at 800 hPa (see Figure 6(f)). Finally, the experts observed that the many members predicted k-index higher than 35 °C, which indicates that the high humidity led to high atmospheric instability (see Figure 6(g)).

8.2. *Heavy Rainfall Event in 2020*

On January 8th, 2020, close to 90 mm of accumulated precipitation was registered in just one hour in the mountain region of Rio de Janeiro. Unlike the 2011 event, this episode was caused by the passage of a cold front associated with the formation of a low-pressure area on the continent, due to the strong heat and the high levels of air humidity left by the summer rains that hit the region in the previous six days. This event caused the overflow of urban rivers and several landslides.

Exploring the rain atmospheric variable using the Temporal Overview, the weather experts saw that, differently from the previous case study, the ensemble members predicted with good precision both the day and the time when that the severe event occurred. In fact, many ensemble members predicted high average and 90th percentile values of precipitation in the late afternoon of January 6, 7, and 8, 2020, characterizing the typical summer rains that occur in the region. Moreover, some simulations predicted even higher accumulations in the late afternoon of the 8th, especially those that used the Grell-Freitas and Kain-Fritsch parametrizations to model the physical process of Cumulus Convection.

For this reason, these two parametrizations were selected for close inspection in the Spatial View (see Figure 7(a)). The map showing the probability of observing more than 30 mm of rain on January 8th at 9 pm (GMT) (6 pm local time) indicates that the areas with a higher probability of having large volumes of rain are located in the south region of Rio de Janeiro. By selecting this region, the experts used the probability mass function visualization of the Distribution View to confirm that 25.3% of the members that used the Grell-Freitas and 37.9% of the members with the Kain-Fritsch predicted heavy rain in the region (see Figure 7(b)).

The analysis of other variables, like the temperature at 2 m, convergence at 850 hPa, divergence at 300 hPa, humidity at 850 hPa, and k-index allowed the meteorologists to clearly identify patterns that characterized the formation of summer rain at the end of the day, demonstrating that the system, again, was able to bring to light the possibilities of a severe event occurring. For example, setting the Spatial View to show the probability of having k-index greater than 35 °C and activating the lens to show the conditional probability of observing more than 30 mm of rain given that the values of the k-index are greater than 35 °C, the meteorologists can see that both variables were likely to achieve high values simultaneously on the night of the event (see Figure 4).

Now considering the atmospheric variable k-index, using X-Weather it was possible to notice that the vast majority of the members of the ensemble indicates the occurrence of values greater than 35 °C. This represents the possibility of atmospheric instability (see Figure 7(c)). Maintaining the k-index variable active with minimum value of 35 °C and activating the lens with minimum rain value of 30 mm, the meteorologists observed that even with high probability of k-index values greater than 35 °C, only in the late afternoon of the 8th there was a higher probability of rainfall values greater than 30 mm, considering the Grell-Freitas and Kain-Fritsch parametrizations. This shows that the k-index is, individually, an incomplete indicator of storm formation. However, the conditional probability of rainfall values greater than 30 mm was also important consid-

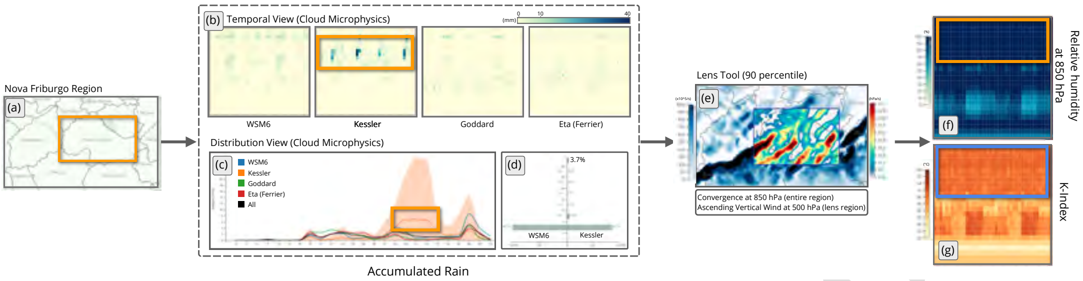


Fig. 6. Example of interactive exploration using X-Weather of a weather ensemble with simulations of a severe rain even that occurred in the mountain region of Rio de Janeiro in 2011. The region of Nova Friburgo (a), the most affected by the storm, was investigated by weather experts. Using the system they observed that only a small number of simulations using the Kessler parametrization of the Cloud Microphysics process predicted large amounts of rain (highlighted regions in (b) and (c)). More precisely, the probability of observing large amounts of rain based on the predictions that use Kessler was 3.7% (d). However, a closer look in other variables associated to the development of storms, such as the ascending vertical wind at 500 hPa and the convergence at 850 hPa, showed the existence of energy capable of raising humidity to form rain clouds (e). This fact was then confirmed observing that all members predicted close to 100% relative humidity at 850 hPa (highlighted in (f)) and values of k-index greater than 35 °C, which indicates atmospheric instability (highlighted in (g)).

ering the high relative humidity values at 850 hPa. This shows, once again, coherence concerning the physical transformations of atmospheric variables by the two parametrizations.

The ability to better predict extreme weather events in specific regions by visually inspecting a large number of ensemble members (and different atmospheric variables) with different parametrizations is something that can greatly improve alert systems and possibly minimize the human and financial costs of weather-related disasters.

9. Experts Feedback

Throughout the research and development of X-Weather, we kept close contact with the domain experts, tuning the interface and exploration aspects of the tool to better satisfy their needs. We requested their feedback regarding ease of use, utility, and feature requests.

The users agreed that the tool is very useful in its capability to augment extreme weather alert systems, since the visualizations and interactions, together with different statistical metrics bring to light often hidden information that can make a difference when it comes to alerting about the possible occurrence of natural disasters. The users also highlighted the ability to visualize the results of simulations with different parametrizations, as some sets of ensemble members can present better performance than others when considering different regions. The experts also highlighted the usefulness of the temporal heat matrices view, a visualization that is not part of a meteorologists' daily routine, unlike heatmaps and line charts. They realized that the matrices, in fact, presented general information about each member of a large ensemble in a practical and optimized way. This can be especially useful when they face situations where a model not forecasting a high volume of rain does not necessarily mean that there is no possibility of a severe event. In this sense, it was important that the matrices enabled them to visualize multiple variables over time since only the direct result of rain can mask the existence of risks.

One of the features requested by the experts was the ability to set arbitrary time intervals for aggregation. The current version of X-Weather aggregates the data with a fixed window of three hours. Important events might happen at a finer temporal resolution (e.g., rain over a short period of time), or coarser resolution (e.g., accumulated precipitation over a day), so it is important for the specialist to choose their own aggregation bins. Another request was related to the Map View lens widget; since the data visualized with the lens is linked with the current underlying map data, the expert suggested that it would be useful to select different instants of time, one for the base map itself and another one for the lens maps. This would be especially useful because some atmospheric variables are related at different times (e.g., it is common for air to rise hours before a storm, such as the 2011 event, as well as movements of convergence and divergence at different times).

10. Conclusion and Future Work

In this work, we presented X-Weather, a visual analytics tool built specifically for the analysis of a large ensemble of simulations generated by a numerical weather model, configured with different parametrizations to represent various physical processes. By using three different visualization components, weather forecasters can explore the ensemble and investigate the possibilities and probabilities of extreme weather events. We also presented a set of case studies that show the usefulness of the tool in the analysis of extreme weather events in the mountain region of Rio de Janeiro; the experts who used the tool highlighted its capability to augment extreme weather alert systems, and potentially prevent some of the consequences of heavy rainfalls that lead to landslides. One of the most important outcomes of X-Weather is to increase the forecaster's ability to interpret weather simulations, specifically when numerical models were not designed with a certain region (developing countries) in mind. In doing this, we believe that different stakeholders in the alert system infrastructure (e.g., city, state, and federal agencies, private entities) will ultimately be more open and secure to take actions that can save lives.

Furthermore, the interdisciplinary interaction between weather and visualization experts during the development and

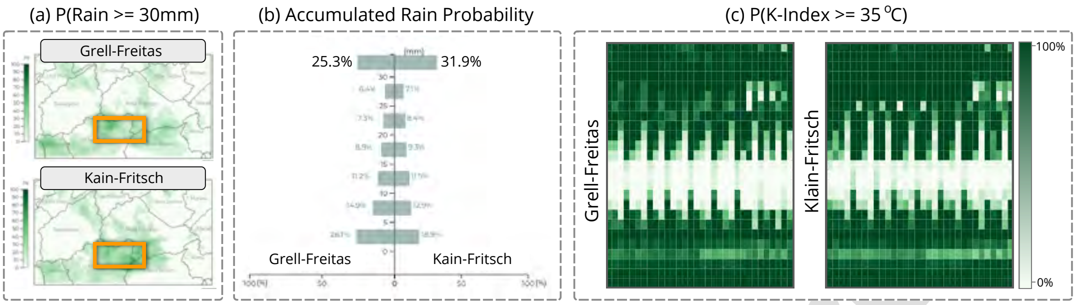


Fig. 7. Heavy rainfall event in the mountain region of Rio de Janeiro in 2020. The simulations of the constructed ensemble predicted high values of accumulated precipitation (more than 30 mm), especially in the south of the studied area (a). In fact, considering simulations using the Grell-Freitas and Kain-Fritsch parametrizations to approximate the Cumulus Convection process, the chances of raining more than 30 mm was 25.3% and 31.9%, respectively (b). Considering the same parametrizations, we see that the probability of having k-index greater than $35\,^{\circ}\mathrm{C}$ **was close to 100% during the storm period (c).**

use of the system has provided valuable lessons to guide future work: (1) The experts highlighted the importance to set arbitrary time intervals for aggregations; (2) The experts mentioned that some weather phenomena happen due to the previous occurrence of others, i.e., the relationship between them exists at different time steps. In this context, it is essential to facilitate the investigation of patterns by visualizing them not only at the same time steps, as it is done in X-Weather, but also at different steps; (3) Although the organization of the interface was sufficient for the experts to properly use X-Weather, we noticed that they needed to switch screens frequently to investigate different atmospheric variables. Presenting these variables on the same screen (not just the lens) could improve the analysis workflow.

On top of the previously mentioned directions, we also plan to incorporate terrain and building models into our system as well as landslide and flooding historical data so that the weather forecaster can have a view of the impact of rain on regions that are usually impacted by extreme rain, and make more informed decisions regarding possible emergency evacuation. We also plan to make the tool available to a wider audience, deploying it on a reliable and robust server. On top of this, we plan to investigate, in collaboration with weather experts, other regions in Brazil that also suffer from heavy rain and landslides.

References

[1] Mizutori, M, Guha-Sapir, D. Human cost of disasters 2000-2019. Tech. Rep.; United Nations Office for Disaster Risk Reduction; 2020. URL: https://www.undrr.org/publication/human-cost-disasters-2000-2019.

[2] Wallemacq, P, House, R. Economic losses, poverty & disasters: 1998-2017. Tech. Rep.; UN Office for Disaster Risk Reduction; 2018. URL: https://www.undrr.org/publication/economic-losses-poverty-disasters-1998-2017.

[3] Treinish, L, Praino, A, Cipriani, J, Mello, U, Real, LV, Sesini, P, et al. Enabling a high-resolution, coupled hydro-meteorological system for operational forecasting of severe weather and flooding events in Rio de Janeiro. Conference on Transition of Research to Operations; 2013,.

[4] Pinheiro, H, Andrade, K, Moura, C. A maior catástrofe climática do brasil sob a visão operacional do CPTEC/INPE. In: Annals... International Symposium on Climatology; 2011,.

[5] Busch, A, Amorim, S. A tragédia da região serrana do Rio de Janeiro em 2011: procurando respostas. Tech. Rep. 2; Escola Nacional de Administração Pública (ENAP); 2011. URL: http://repositorio.enap.gov.br/handle/1/328.

[6] Wannous, C, Velasquez, G. UNISDR's contribution to science and technology for disaster risk reduction and the role of the international consortium on landslides. In: World Landslide Forum. 2017, p. 109–115.

[7] UNDRR, . Sendai framework for disaster risk reduction 2015 - 2030. 2015. URL: https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030.

[8] Potter, K, Wilson, A, Bremer, PT, Williams, D, Doutriaux, C, Pascucci, V, et al. Ensemble-vis: A framework for the statistical visualization of ensemble data. In: 2009 IEEE International Conference on Data Mining Workshops. IEEE; 2009, p. 233–240.

[9] Palmer, T, Hagedorn, R. Predictability of weather and climate. Cambridge University Press; 2006.

[10] Webster, P. Improve weather forecasts for the developing world. Nature 2013;493:3.

[11] Stähli, L, Rudi, D, Raubal, M. Turbulence ahead - a 3d web-based aviation weather visualizer. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology. UIST '18; Association for Computing Machinery; 2018, p. 299–311.

[12] Ferstl, F, Kanzler, M, Rautenhaus, M, Westermann, R. Visual analysis of spatial variability and global correlations in ensembles of iso-contours. Computer Graphics Forum 2016;35(3):221–230.

[13] Ferstl, F, Kanzler, M, Rautenhaus, M, Westermann, R. Time-hierarchical clustering and visualization of weather forecast ensembles. IEEE Trans Vis Comput Graphics 2017;23(1):831–840.

[14] Ferstl, F, Bürger, K, Westermann, R. Streamline variability plots for characterizing the uncertainty in vector field ensembles. IEEE Trans Vis Comput Graphics 2016;22(1):767–776.

[15] Kehrer, J, Muigg, P, Doleisch, H, Hauser, H. Interactive visual analysis of heterogeneous scientific data across an interface. IEEE Trans Vis Comput Graphics 2011;17(7):934–946.

[16] Obermaier, H, Bensema, K, Joy, KI. Visual trends analysis in time-varying ensembles. IEEE Trans Vis Comput Graphics 2016;22(10):2331–2342.

[17] Matkovic, K, Gracanin, D, Klarin, B, Hauser, H. Interactive visual analysis of complex scientific data as families of data surfaces. IEEE Trans Vis Comput Graphics 2009;15(6):1351–1358.

[18] Hummel, M, Obermaier, H, Garth, C, Joy, KI. Comparative visual analysis of lagrangian transport in cfd ensembles. IEEE Trans Vis Comput Graphics 2013;19(12):2743–2752.

[19] Poco, J, Dasgupta, A, Wei, Y, Hargrove, W, Schwalm, C, Cook, R, et al. Similarityexplorer: A visual inter-comparison tool for multifaceted climate data. Computer Graphics Forum 2014;33(3):341–350.

[20] Hao, L, Healey, CG, Hutchinson, SE. Ensemble visualization for cyber situation awareness of network security data. In: 2015 IEEE Symposium on Visualization for Cyber Security (VizSec). 2015, p. 1–8.

[21] Srabanti, S, Marai, GE, Miranda, F. COVID-19 EnsembleVis: Visual analysis of county-level ensemble forecast models. In: 12th workshop on Visual Analytics in Healthcare (VAHC). 2021,.

[22] Hermann, M, Schunke, AC, Schultz, T, Klein, R. Accurate interactive visualization of large deformations and variability in biomedical image ensembles. IEEE Trans Vis Comput Graphics 2016;22(1):708–717.
[23] Nocke, T, Flechsig, M, Bohm, U. Visual exploration and evaluation of climate-related simulation data. In: Winter Simulation Conf. 2007,.
[24] Obermaier, H, Joy, KI. Future challenges for ensemble visualization. IEEE Computer Graphics and Applications 2014;34(3):8–11.
[25] Wang, J, Hazarika, S, Li, C, Shen, HW. Visualization and visual analysis of ensemble data: A survey. IEEE Trans Vis Comput Graphics 2019;25(9):2853–2872.
[26] Rautenhaus, M, Böttinger, M, Siemen, S, Hoffman, R, Kirby, RM, Mirzargar, M, et al. Visualization in meteorology—a survey of techniques and tools for data analysis tasks. IEEE Trans Vis Comput Graphics 2018;24(12):3268–3296.
[27] NOAA Pacific Marine Environmental Laboratory, . Ferret. 2021. URL: https://ferret.pmel.noaa.gov/Ferret/.
[28] Center for Ocean-Land-Atmosphere Studies, . Grid analysis and display system (grads). 2020.
[29] School of Ocean and Earth Science, , Technology of the University of Hawaii, . The generic mapping tools. 2021. URL: https://www.generic-mapping-tools.org/.
[30] Sanyal, J, Zhang, S, Dyer, J, Mercer, A, Amburn, P, Moorhead, R. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. IEEE Trans Vis Comput Graphics 2010;16(6):1421–1430.
[31] Diehl, A, Pelorosso, L, Delrieux, C, Saulo, C, Ruiz, J, Gröller, ME, et al. Visual analysis of spatio-temporal data: Applications in weather forecasting. Computer Graphics Forum 2015;34(3):381–390.
[32] Diehl, A, Pelorosso, L, Delrieux, C, Matković, K, Ruiz, J, Gröller, ME, et al. Albero: A visual analytics approach for probabilistic weather forecasting. Computer Graphics Forum 2017;36(7):135–144.
[33] Biswas, A, Lin, G, Liu, X, Shen, HW. Visualization of time-varying weather ensembles across multiple resolutions. IEEE Trans Vis Comput Graphics 2017;23(1):841–850.
[34] Wang, J, Liu, X, Shen, HW, Lin, G. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. IEEE Trans Vis Comput Graphics 2017;23(1):81–90.
[35] Rautenhaus, M, Kern, M, Schäfler, A, Westermann, R. Three-dimensional visualization of ensemble weather forecasts – part 1: The visualization tool met.3d (version 1.0). Geoscientific Model Development 2015;8(7):2329–2353.
[36] Santos, E, Poco, J, Wei, Y, Liu, S, Cook, B, Williams, DN, et al. Uv-cdat: Analyzing climate datasets from a user's perspective. Computing in Science & Engineering 2013;15(1):94–103.
[37] Williams, D, Bremer, PT, Doutriaux, CM, Patchett, J, Williams, S, Shipman, G, et al. Ultrascale visualization of climate data. Computer 2013;46(9):68–76.
[38] Callahan, SP, Freire, J, Santos, E, Scheidegger, CE, Silva, CT, Vo, HT. Vistrails: Visualization meets data management. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. SIGMOD '06; 2006, p. 745–747.
[39] Potter, K, Kniss, J, Riesenfeld, R, Johnson, CR. Visualizing summary statistics and uncertainty. In: Proceedings of EuroVis'10. The Eurographs Association &; John Wiley &; Sons, Ltd.; 2010, p. 823–832.
[40] Pang, A, Wittenbrink, C, Lodha, S. Approaches to uncertainty visualization. Tech. Rep.; 1996.
[41] MacEachren, AM, Robinson, A, Hopper, S, Gardner, S, Murray, R, Gahegan, M, et al. Visualizing geospatial information uncertainty: What we know and what we need to know. Cartography and Geographic Information Science 2005;32(3):139–160.
[42] Brodlie, K, Allendes Osorio, R, Lopes, A. A Review of Uncertainty in Data Visualization. Springer London; 2012, p. 81–109.
[43] Bonneau, GP, Hege, HC, Johnson, CR, Oliveira, MM, Potter, K, Rheingans, P, et al. Overview and State-of-the-Art of Uncertainty Visualization. London: Springer London; 2014, p. 3–27.
[44] Potter, K, Rosen, P, Johnson, CR. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In: Dienstfrey, AM, Boisvert, RF, editors. Uncertainty Quantification in Scientific Computing. Springer Berlin Heidelberg; 2012, p. 226–249.
[45] Retchless, DP, Brewer, CA. Guidance for representing uncertainty on global temperature change maps. International Journal of Climatology 2016;36(3):1143–1159.
[46] Collins, M, Booth, BB, Harris, GR, Murphy, JM, Sexton, DM, Webb, MJ. Towards quantifying uncertainty in transient climate change. Climate dynamics 2006;27(2-3):127–147.
[47] Potter, K, Wilson, A, Bremer, PT, Williams, D, Doutriaux, C, Pascucci, V, et al. Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. Journal of Physics: Conference Series 2009;180:012089.
[48] Kaye, NR, Hartley, A, Hemming, D. Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty. Geoscientific Model Development 2012;5(1):245–256. URL: https://gmd.copernicus.org/articles/5/245/2012/.
[49] Stephens, EM, Edwards, TL, Demeritt, D. Communicating probabilistic information from climate model ensembles—lessons from numerical weather prediction. WIREs Climate Change 2012;3(5):409–426.
[50] Novak, DR, Bright, DR, Brennan, MJ. Operational forecaster uncertainty needs and future roles. Weather and Forecasting 01 Dec. 2008;23(6):1069 – 1084.
[51] Schumacher, RS, Davis, CA. Ensemble-based forecast uncertainty analysis of diverse heavy rainfall events. Weather and Forecasting 01 Aug. 2010;25(4):1103 – 1122.
[52] Nadav-Greenberg, L, Joslyn, SL, Taing, MU. The effect of uncertainty visualizations on decision making in weather forecasting. Journal of Cognitive Engineering and Decision Making 2008;2(1):24–47.
[53] Grönquist, P, Yao, C, Ben-Nun, T, Dryden, N, Dueben, P, Li, S, et al. Deep learning for post-processing ensemble weather forecasts. Philosophical Transactions of the Royal Society A 2021;379(2194):20200092.
[54] Navascués, B, Calvo, J, Morales, G, Santos, C, Callado, A, Cansado, A, et al. Long-term verification of hirlam and ecmwf forecasts over southern europe: History and perspectives of numerical weather prediction at aemet. Atmospheric Research 2013;125-126:20–33.
[55] Powers, JG, Klemp, JB, Skamarock, WC, Davis, CA, Dudhia, J, Gill, DO, et al. The weather research and forecasting model: Overview, system efforts, and future directions. Bulletin of the American Meteorological Society 2017;98:1717–1737.
[56] Jimenez, PA, Hacker, JP, Dudhia, J, Haupt, SE, Ruiz-Arias, JA, Gueymard, CA, et al. Wrf-solar: Description and clear-sky assessment of an augmented nwp model for solar power prediction. Bulletin of the American Meteorological Society 2016;97(7):1249–1264.
[57] Grell, GA, Peckham, SE, Schmitz, R, McKeen, SA, Frost, G, Skamarock, WC, et al. Fully coupled "online" chemistry within the wrf model. Atmospheric Environment 2005;39(37):6957 – 6975.
[58] Fast, JD, Jr., WIG, Easter, RC, Zaveri, RA, Barnard, JC, Chapman, EG, et al. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of houston using a fully coupled meteorology-chemistry-aerosol model. Journal of Geophysical Research: Atmospheres 2006;111(D21).
[59] National Centers for Environmental Prediction, National Weather Service, NOAA, . Global forecast system. 2020. URL: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs.
[60] Barcellos, PCL, Cataldi, M. Flash flood and extreme rainfall forecast through one-way coupling of wrf-smap models: Natural hazards in Rio de Janeiro State. Atmosphere 2020;11.
[61] Liu, Z, Heer, J. The effects of interactive latency on exploratory visual analysis. IEEE Trans Vis Comput Graphics 2014;20(12):2122–2131.
[62] Lins, L, Klosowski, JT, Scheidegger, C. Nanocubes for real-time exploration of spatiotemporal datasets. IEEE Trans Vis Comput Graphics 2013;19(12):2456–2465.
[63] Miranda, F, Lins, L, Klosowski, JT, Silva, CT. Topkube: A rank-aware data cube for real-time exploration of spatiotemporal data. IEEE Trans Vis Comput Graphics 2018;24(3):1394–1407.
[64] Pahins, CAL, Ferreira, N, Comba, JL. Real-time exploration of large spatiotemporal datasets based on order statistics. IEEE Trans Vis Comput Graphics 2019;26(11):3314–3326.
[65] Skamarock, WC, Klemp, JB, Dudhia, J, Gill, DO, Liu, Z, Berner, J, et al. A description of the advanced research wrf model version 4. 2019. NCAR Tech. Note NCAR/TN-556+STR.

Highlights:

- Our visualizations help identify extreme weather scenarios in large simulation ensembles.
- Our web-based system enables the investigation of individual weather ensemble members.
- Two case studies highlight our system's utility by analyzing extreme weather events.

Visualizing Simulation Ensembles of Extreme Weather Events

Carolina Veiga Ferreira de Souza [(a)], Priscila Cunha Luz Barcellos [(a)], Lhaylla Crissaff [(a)], Marcio Cataldi [(a)], Fabio Miranda [(b)], Marcos Lage [(a)]

(a) Universidade Federal Fluminense

(b) University of Illinois at Chicago

Detailed information:

Carolina Veiga Ferreira de Souza (Corresponding author)

PhD student: Universidade Federal Fluminense,

Email: carolinavfs@id.uff.br

Priscila da Cunha Luz Barcellos

Postdoctoral researcher: Universidade Federal Fluminense

Email: luz.priscila@gmail.com

Lhaylla Crissaff

Professor

Universidade Federal Fluminense

Email: lhayllacrissaff@id.uff.br

Fabio Miranda

Assistant Professor

University of Illinois at Chicago

Email: fabiom@uic.edu

Marcio Cataldi

Professor

Universidade Federal Fluminense

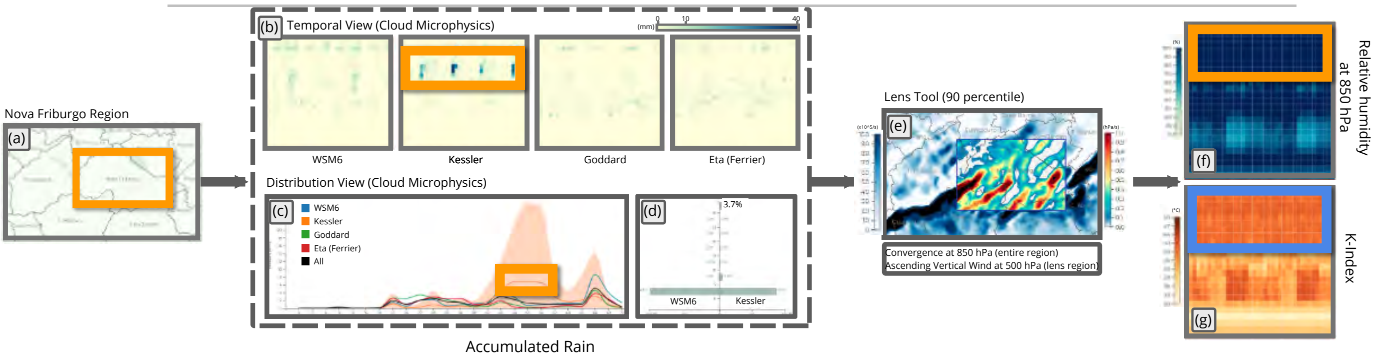
Email: marcio.cataldi@gmail.com

Marcos Lage

Professor

Universidade Federal Fluminense

Email: mlage@ic.uff.br


Nova Friburgo Region
(a)
(b)
Temporal View (Cloud Microphysics)
(mm)
0
10
40
WSM6
Kessler
Goddard
Eta (Ferrier)
Distribution View (Cloud Microphysics)
(c)
WSM6
Kessler
Goddard
Eta (Ferrier)
All
(d)
3.7%
WSM6
Kessler
Accumulated Rain
Lens Tool (90 percentile)
(e)
Convergence at 850 hPa (entire region)
Ascending Vertical Wind at 500 hPa (lens region)
(f)
Relative humidity
at 850 hPa
(g)
K-Index

CRediT author statement

Carolina Veiga: Conceptualization, Methodology, Software, Writing - Original Draft, Visualization.

Priscila Luz: Data Curation, Formal analysis, Investigation, Resources.

Lhaylla Crissaff: Validation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Marcio Cataldi: Data Curation, Formal analysis, Investigation, Resources.

Fabio Miranda: Conceptualization, Methodology, Validation, Supervision, Writing - Original Draft, Writing - Original Draft, Writing - Review & Editing, Visualization.

Marcos Lage: Conceptualization, Methodology, Validation, Software, Supervision, Writing - Original Draft, Writing - Review & Editing, Visualization.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: