Article

Assessing the Impact of Lightning Data Assimilation in the WRF Model

Vanderlei Vargas, Jr. [1,2], Rute Costa Ferreira [3], Osmar Pinto, Jr. [4] and Dirceu Luis Herdies [3,*]

1   Cooperative Institute for Research in the Atmosphere, Colorado State University (CSU/CIRA), Fort Collins, CO 80521, USA; vanderlei.vargas@noaa.gov
2   Global Systems Laboratory, National Oceanic and Atmospheric Administration (NOAA/GSL), Boulder, CO 80305, USA
3   Center for Weather Forecast and Climate Studies, National Institute for Space Research (INPE/CPTEC), Cachoeira Paulista 12630-000, SP, Brazil
4   Atmospheric Electricity Group, Earth Science System Center, National Institute for Space Research (INPE/CCST/ELAT), São José dos Campos 12227-010, SP, Brazil
*   Correspondence: dirceu.herdies@inpe.br

Abstract: Recent advancements in computational technologies have enhanced the importance of meteorological modeling, driven by an increased reliance on weather-dependent systems. This research implemented a lightning data assimilation technique to improve short-term weather forecasts in South America, potentially refining initialization methods used in meteorological operation centers. The main goal was to implement and enhance a data assimilation algorithm integrating lightning data into the WRF model, assessing its impact on forecast accuracy. Focusing on southern Brazil, a region with extensive observational infrastructure and frequent meteorological activity, this research utilized several data sources: precipitation data from the National Institute of Meteorology (INMET), lightning data from the Brazilian Lightning Detection Network (BrasilDAT), GOES-16 satellite images, synoptic weather charts from the National Institute for Space Research (INPE), and initial conditions from the GFS model. Employing the WRF-ARW model version 3.9.1.1 and WRFDA system version 3.9.1 with 3DVAR methodology, the study conducted three experimental setups during two meteorological events to evaluate the assimilation algorithm. These included a control (CTRL) without assimilation, a lightning data assimilation (LIGHT), and an adaptive humidity threshold assimilation (ALIGHT). Results showed that the lightning data assimilation system enhanced forecasts for large-scale systems, especially with humidity threshold adjustments. While it improved squall line timing and positioning, it had mixed effects when convection was thermally driven. The lightning data assimilation methodology represents a significant contribution to the field, indicating that using such alternative data can markedly improve short-term forecasts, benefiting various societal sectors.

Keywords: nudging; BrasilDAT; 3DVAR; lightning data assimilation

1. Introduction

Over recent decades, society's reliance on weather-sensitive technologies has increased, as has the expansion into areas prone to environmental risks, amplifying the need for precise meteorological modeling. Advancements in computing technology have enabled the development of more accurate atmospheric models, which are essential for the growing reliance on weather forecasts across various economic sectors.

Weather forecasts are typically generated using a set of equations that encompass the laws of motion and conservation principles, including mass and energy. These equations lack analytical solutions, necessitating the use of numerical methods in a process commonly referred to as Numerical Weather Prediction (NWP).

Addressing several aspects of modeling can enhance weather forecasting. These include refining the model's core and settings, such as spatial resolution and discretization techniques; enhancing the physics through more accurate parameterizations; improving data assimilation methods; and utilizing ensemble forecasting approaches.

This study focuses on how data assimilation, specifically the integration of new data sources, can refine the initial conditions of weather models, thereby improving forecast accuracy. Data assimilation involves merging observational data with model-generated estimates (the background or first guess) to produce adjusted initial conditions (the analysis), which can lead to better forecasts [1,2].

NWP is highly sensitive to its initial conditions. The complexity and high resolution of these models mean that simply interpolating observational data does not sufficiently inform the initial conditions. Thus, it is crucial to use a background derived from other models or previous runs [1,2]. The integration of observational data helps correct the model's background, reducing cumulative errors in the simulation.

The absence of comprehensive observational data often leads models to start with inaccurate initial and boundary conditions, resulting in spin-up issues [3]. However, incorporating these data at the beginning of simulations can partially mitigate these issues [4].

The Weather Research and Forecasting (WRF) model and its Data Assimilation System (WRFDA) were utilized for all experiments conducted in this study. The WRF model is a system widely employed in research and operational settings, offering a broad range of physical and dynamic parameterizations suitable for various geographic settings [5,6].

Various assimilation techniques have been explored in atmospheric science, including the successive corrections method [7], nudging [8,9], optimal interpolation [10,11], variational methods [12], Kalman filter [13], and hybrid methods [14,15]. The WRFDA system supports variational methods (3DVAR, 4DVAR) and hybrid options.

This study delves into the unique benefits of incorporating lightning data into weather models. Lightning is an excellent indicator of severe convection and is less influenced by geographic features than radar data. Continuous lightning monitoring can provide valuable insights into the dynamics and microphysics of storm clouds [16–19].

Historically, lightning data were first used to improve precipitation forecasts and other variables indirectly. Recent advances have enabled more direct integration of lightning data into models, such as through adjustments to observation operators in variational and hybrid methods [20–22].

Various lightning detection networks operate globally, each with different detection capabilities and accuracies, ranging from low-frequency networks with global coverage to high-frequency networks like the Lightning Mapping Array and the Geostationary Lightning Mapper aboard the Geostationary Operational Environmental Satellites (GOES-16), which utilizes the optical spectrum for detection from space [23].

This research marks a pioneering effort to implement a lightning data assimilation technique to enhance short-term weather forecasting in South America. The potential benefits of using lightning data, including its high spatial and temporal resolution, are explored in depth.

The primary objective of this study was to develop and refine a data assimilation algorithm that incorporates lightning data into the WRF model. The effectiveness of lightning data assimilation was evaluated by comparing experimental results with observations, focusing on the impact on short-term forecasts (up to three hours).

2. Background

Several theories have attempted to explain the formation of lightning, with the convective theories [24], inductive theories [25], non-inductive charging processes [26], and quasi-liquid layer theories [27,28] being the most prevalent. These theories elucidate how various atmospheric conditions and particle interactions contribute to charge separation and ultimately to lightning discharge.

Thunderstorm clouds, particularly cumulonimbus, exhibit significant vertical development and intense vertical motions. These motions lift humid air to higher altitudes, cooling and condensing it into ice particles. The process of cloud electrification is complex and not entirely understood but generally involves interactions between graupel and ice crystals [29,30].

Integrating lightning data into weather models is predicated on a detailed understanding of these microphysical processes. The assimilation of lightning data, which includes both intracloud (IC) and cloud-to-ground (CG) lightning types, can enhance the initialization of models by providing critical information about the atmospheric state that is not available from conventional meteorological data sources. These data are particularly valuable due to their direct relationship with atmospheric instability and convection, offering unique insights into storm dynamics and potential weather threats [16,31,32].

The Brazilian Lightning Detection Network (BrasilDAT), a primary lightning detection system in Brazil, operates in low and very low frequencies (LF/VLF). This system uses the time-of-arrival (TOA) method to triangulate the location of lightning discharges, accurately detecting both the polarity and intensity of the lightning flash [33,34]. This system's high detection efficiency ranges from 85 to 90% for CG and from 50 to 60% for IC.

The atmospheric processes that lead to lightning formation are intricately linked to the dynamics within thunderclouds. Understanding these processes is crucial for effectively incorporating lightning data into weather prediction models. The assimilation of these data can not only enhance the initialization of these models but also contribute to more accurate and reliable weather forecasting, leveraging the unique insights provided by lightning activity to anticipate and respond to meteorological events more effectively.

A pioneering study by Alexander et al. [35] marked the onset of utilizing total lightning data—comprising both IC and CG flashes—without distinction, aiming to improve the models' initial atmospheric conditions. This study employed various data sources, including satellite, infrared, and microwave sensors alongside continuous lightning observations. The study demonstrated a robust correlation between lightning data and precipitation rates, which varies regionally. By incorporating a diverse set of data, including those from lightning, their simulations effectively captured the dynamics of an extratropical cyclone in the Gulf of Mexico, showcasing enhanced precipitation rates and more accurate cyclone modeling.

In a related effort, Chang et al. [36] delved into a cyclogenesis event in the Gulf of Mexico using a dataset that included long-range lightning detection systems and satellite-based sensors. This approach allowed for a nuanced analysis of convective activity, particularly in areas beyond the reach of traditional meteorological radars. Their integration of these varied data sources into the mesoscale model (MM5) and the application of a novel algorithm led to significant improvements in forecasting accuracy, especially in predicting precipitation patterns.

Building on these foundations, subsequent studies have explored the potential of modifying physical parameterizations within weather models to incorporate lightning data effectively. Mansell et al. [37], for instance, adapted the Kain–Fritsch convective parameterization scheme within the Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS). This modification allowed the scheme to be triggered by lightning data, which notably enhanced the model's depiction of atmospheric phenomena such as cold pools, thereby improving overall forecast accuracy during the initial hours of simulations.

Further refining these methods, research by Qie et al. [38] and Fierro et al. [39] applied nudging functions to adjust the microphysics schemes within cloud systems based on lightning data. These adjustments led to significant improvements in the simulation of convective activities and precipitation fields, particularly in short-term forecasts.

The continuous evolution of lightning data assimilation techniques, especially through nudging methodologies, has been shown to significantly enhance the accuracy of weather prediction models. These techniques, which adjust model parameters in real time based on lightning observations, have improved the representation of atmospheric processes,

addressing the challenge of balancing induced convection within models to prevent prolonged, unrealistic weather patterns.

Moreover, studies like those conducted by Dixon et al. [40] and Chen et al. [41] have furthered the application of lightning data in weather modeling. Dixon et al. used high-resolution deterministic and ensemble forecasts to demonstrate that nudging techniques, which included moisture adjustments based on lightning data, could enhance convective system predictions across the United States. Chen et al. developed a specialized lightning data assimilation scheme tailored to the dynamics of squall lines, demonstrating the potential of such advanced methodologies to significantly reduce the spin-up time of models and enhance short-term forecast accuracy.

By providing detailed insights into convective processes and enabling more precise adjustments to weather models, lightning data assimilation stands as a transformative approach in the ongoing quest to enhance the accuracy and reliability of weather forecasting.

Table 1 shows some of the main studies and their characteristics in the field of lightning data assimilation.

Table 1. Characteristics of the main studies in the lightning data assimilation research field.

Ref.	Year	Study Area	Data Source	Model	Assimilation Technique
[35]	1999	Mexican Gulf	NLDN	MM5	Nudging (precipitation rate)
[36]	2001	Mexican Gulf	STARNET-1; LIS	MM5	Nudging (precipitation rate)
[18]	2005	Europe and Africa	ZEUS	SKIRON/ETA	Nudging (humidity)
[37]	2007	Midwest USA	NLDN; LMA	COAMPS	Nudging (convective scheme)
[42]	2009	North Pacific Ocean	LIS; OTD	MM5	Nudging (convective scheme)
[16]	2012	Midwest USA	ENTLN	WRF	Nudging (microphysics scheme)
[22]	2012	Alabama	WWLLN	WRF	Observation operator (CAPE)
[43]	2013	Southern France	LINET	MM5	Nudging (convective scheme)
[21]	2013	USA	ENTLN	WRF	Observation operator (CAPE)
[38]	2014	North China	SAFIR	WRF	Nudging (microphysics scheme)
[39]	2014	Northeast USA	ENTLN	WRF	Nudging (microphysics scheme)
[20]	2014	East USA	WWLLN	WRF	Observation operator (ω_{max})
[44]	2015	USA	ENTLN	WRF	Nudging (humidity)
[45]	2015	East USA	ENTLN; USPLN	WRF	Nudging (humidity)
[40]	2016	Midwest and eastern of USA	WWLLN	WRF	Nudging (humidity)
[46]	2018	East China	SAFIR	WRF	Nudging (humidity)
[41]	2019	Northern China	BLNET	WRF	Nudging (microphysics scheme)
[47]	2022	Southern China	ENTLN	WRF	Nudging (humidity)

3. Material and Methods

3.1. Data, Model Settings, and Experiment Design

This study utilized a diverse array of data sources to effectively implement and evaluate the assimilation process. We focused on two specific days known for the occurrence of meteorological systems characterized by significant lightning activity.

The geographic focus of this research is the southern region of Brazil, as depicted in Figure 1a. This region is not only well equipped with various observational stations but also frequently experiences meteorological systems conducive to the formation of intense storms accompanied by substantial lightning activity. In subsequent references, the green square highlighted in Figure 1b will be referred to as the evaluation area. This area was selected with the goal of minimizing errors related to simulation borders and/or data extrapolations. Meanwhile, the larger region shown in Figure 1b will be referred to as the study area or the simulation area.

A study discussing the WRF model's performance using a high-resolution spatial grid for a similar region can be found in [48].

To assess the simulations, precipitation data were sourced from approximately 600 meteorological surface stations provided by the National Institute of Meteorology

(INMET). The number of stations analyzed varied depending on the time of observation (Figure 1b).
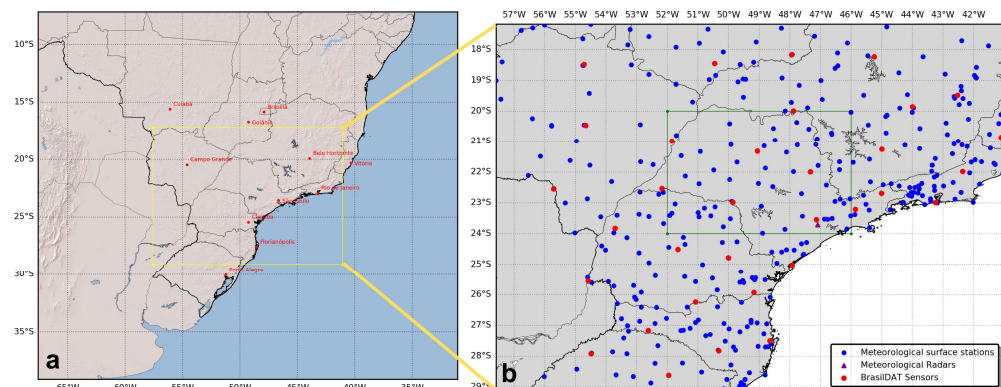

Figure 1. (**a**) The study area, i.e., the simulation domain, is highlighted by a yellow square centered over Brazil and covers an extensive part of the southeast, south, and center–west regions (major cities marked in red). (**b**) The spatial distribution of observational data within the WRF simulation domain is shown, with the green square over most of São Paulo state indicating the evaluation area for assessing the impact of the assimilation process.

Lightning data, a critical component of this study, were provided by BrasilDAT through the Atmospheric Electricity Group of the National Institute for Space Research (INPE/ELAT). These data were essential for calculating the flash rate needed to adjust the water vapor mixing ratio variable in the assimilation algorithm (see Section 3.2), thus enhancing the initial conditions. Additionally, it was instrumental in analyzing and assessing the behavior of various variables simulated by the WRF model.

Satellite imagery from GOES-16 was also integral to tracking the meteorological systems impacting the study area, allowing for the identification of the events and the regions most affected.

Synoptic weather charts from the Center for Weather Forecasting and Climate Studies (CPTEC) at INPE were utilized to accurately identify the meteorological conditions associated with each system studied.

Data from the Global Forecast System (GFS), provided by the Computational and Information Systems Laboratory of the University Corporation for Atmospheric Research, were used for model initialization [49].

This study utilized version 3.9.1.1 of the WRF-ARW model, which was officially released on 28 August 2017 [50]. The simulations incorporated a comprehensive set of parameterizations to accurately model various atmospheric processes: the Yonsei University Scheme for the Planetary Boundary Layer [51], the Thompson Scheme for Microphysics [52], the Unified Noah Land Surface Model for land surface processes [53], and the RRTMG for both shortwave and longwave radiation schemes [54].

Furthermore, cumulus parameterization was deactivated to allow the model to explicitly resolve convection. It is important to point out that to explicitly resolve convection it is necessary to run the model with a resolution of at least 3 km, which was not the case in this study. Rather, this was an attempt to partially resolve convection in the region and avoid the use of parameterizations that are not optimized for the study area. The implications of this choice are discussed Section 4.

The configuration of the simulations involved a single domain encompassing a 149×189 grid point, latitudes ranging from $-29.1184°$ S to $-17.1805°$ S, and longitudes ranging from $-57.5814°$ W to $-41.0482°$ W. The grid spacing varied longitudinally between $0.087°$ and $0.088°$ and latitudinally between $0.077°$ and $0.084°$, achieving a horizontal resolution of approximately 9 km. The vertical configuration included 50 levels, extending up to 50 hPa, and the simulations employed an adaptive timestep to optimize computational efficiency (Figure 2 provides a detailed view of the domain).
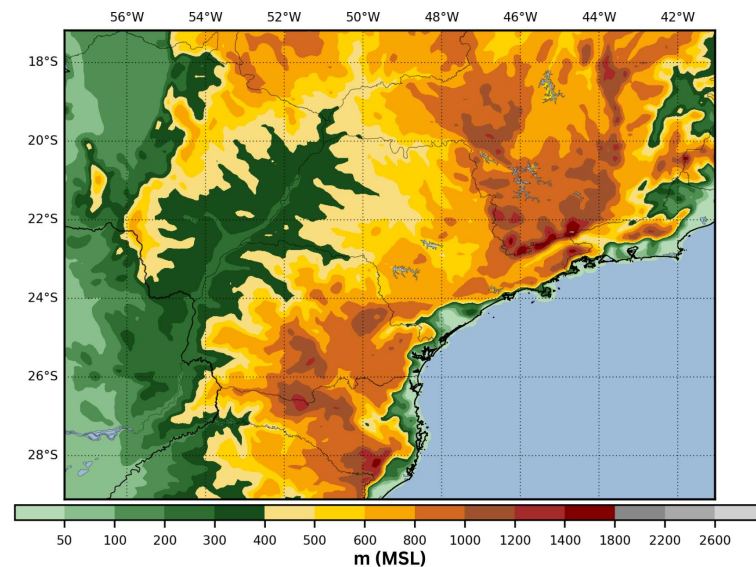
Figure 2. Terrain elevation in meters (MSL) within the simulation domain outlined by a yellow square in Figure 1a.

This specific domain was chosen to align with the spatial distribution of BrasilDAT sensors and the meteorological observation stations illustrated in Figure 1b. The selected region is known for its dynamic meteorological activity throughout the year, influenced by a diverse array of weather systems such as fronts, mesoscale convective systems (MCS), low-level jets, convergence moisture zones, and cyclones. This diversity allows for an in-depth analysis of different atmospheric configurations under well-observed conditions.

For data assimilation, the 3DVAR method was applied using the WRFDA system version 3.9.1, released on 17 August 2017. The assimilation window spanned 30 min, covering 15 min before and after the analysis time.

The assimilation process also utilized a generic background error covariance matrix (BE), designated as CV3, provided by the WRFDA for regional applications. This matrix, crucial in influencing the outcome of the assimilation, was originally generated by the National Centers for Environmental Prediction (NCEP) using the National Meteorological Center (NMC) method. This method calculates the difference between 48 and 24 h forecasts from the GFS model to model vertical covariance, as detailed in Wang et al. [55].

This research was structured around three distinct experiments, each designed to evaluate the implemented assimilation algorithm during two unique meteorological events. These experiments were organized as follows: a control group (CTRL) where no assimilation was performed, a group where lightning data was assimilated (LIGHT), and a group where lightning data assimilation was combined with an adaptive relative humidity (RH) threshold (ALIGHT).

The meteorological events chosen for detailed analysis started at 00:00 UTC on 23 January 2018 and at 00:00 UTC on 19 May 2018. These dates were selected based on the specific meteorological conditions and the distinct atmospheric synoptic scenarios associated with each event. This selection aimed to assess the assimilation process's effectiveness across varying atmospheric conditions. An in-depth discussion of the meteorological scenarios pertaining to each event will be presented in the subsequent section.

Each of the experiments consisted of eight short-term forecasts. Specifically, each simulation generated a forecast spanning three hours, with successive forecasts initiated every three hours, culminating in a total of 24 h of simulation for each experimental setup.

To mitigate the risk of numerical instabilities often associated with the insertion of mass during the assimilation process, all simulations were initiated using a cold start approach. Figure 3 illustrates the schematic structure of an experiment, depicting how initial conditions are employed directly by the WRF-ARW model to produce a three-hour

forecast for the CTRL experiment or used within the assimilation process for the LIGHT and ALIGHT experiments.
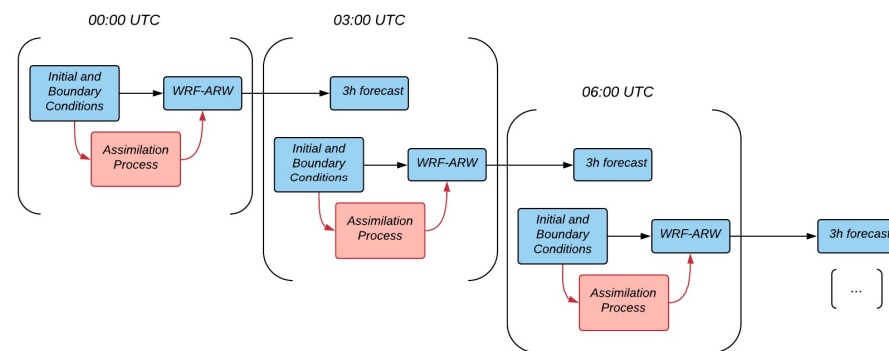

Figure 3. Diagram illustrating the schematic structure of an experiment. The assimilation process was performed over a 30 min window around the analysis time (observations from 15 min before and after the analysis time were considered). All cycles were cold started.

The experiments were scheduled to start at 00:00 UTC on the designated day and conclude at 00:00 UTC the following day. This timing was crucial, as it allowed each simulation's initialization to align with varying quantities of lightning data available throughout the day, enhancing the assessment of the assimilation methodology's performance under different conditions.

Moreover, for the experiment yielding the most favorable results, an extended 24 h forecast was conducted to analyze the impact of the assimilation process more comprehensively over a prolonged period. This extended analysis aimed to provide deeper insights into the temporal effects of data assimilation on forecast accuracy and stability.

3.2. Assimilation Algorithm

In this research, the algorithm developed to integrate lightning data into the correction of the initial model conditions was based on the framework established by Fierro et al. [16], as outlined in Equation (1). This equation modifies the water vapor mixing ratio variable Q_v in the model by considering the flash rate density X derived from lightning observations. It also involves the saturation mixing ratio Q_{sat} and the graupel mixing ratio Q_g, which are calculated from the model's initial conditions. The constants A, B, C, D, and α in the equation are assigned values of 0.81, 0.2, 0.01, 0.25, and 0.22, respectively.

$$Q_v = AQ_{sat} + BQ_{sat}\tanh(CX)\left[1 - \tanh\left(DQ_g^\alpha\right)\right] \tag{1}$$

A 30 min window is established for calculating the flash rate, which involves accumulating all lightning detections 15 min before and after the analysis time (the same 30 min window was used to perform the 3DVAR assimilation process). Subsequently, Q_{sat} is derived from the Q_v field present in the initial conditions, and Q_g is also extracted from these conditions. Utilizing Equation (1), a revised Q_v is then computed. This recalculated Q_v is specifically determined in the mixed-phase region—defined by MacGorman and Rust [31] as the atmospheric layer between the 0 °C and −20 °C isotherms where convection and electrification intensify—and when the RH falls below 81% or 60% (further details in the following section).

While Fierro et al. [16] incorporated corrections directly into the model by adjusting a microphysics parameterization, the current study adopts a method akin to that of Wang et al. [55]. Here, the assimilation is conducted using the WRFDA with the 3DVAR technique, employing a proxy field (RH) generated from Q_v as calculated by Equation (1). This process allows the flash rate to adjust the Q_v field, which is subsequently reinserted into the model through the traditional assimilation process as if it were an observational data point.

The concept of "simulated observation stations" is introduced, where, under the specified conditions, RH is assimilated. These stations are not physical but rather represent grid points within the domain where lightning data have been interpolated and the RH assimilation conditions met. Figure 4 displays the increase in Q_v (the difference between analysis and background values) resulting from the assimilation at these stations.
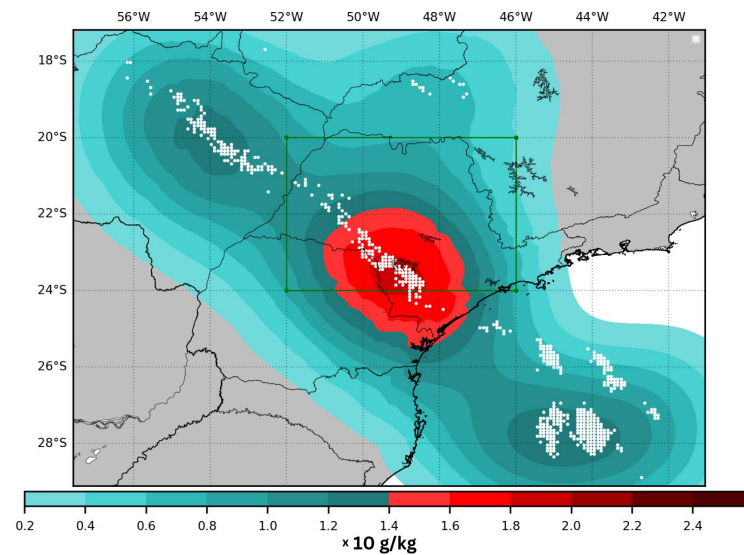

Figure 4. Illustration of the increase in Q_v (difference between analysis and background values) resulting from the application of the assimilation algorithm developed in this study. The white points indicate the locations of the simulated observation stations (604 total), where lightning data were interpolated (as depicted in Figure 5).
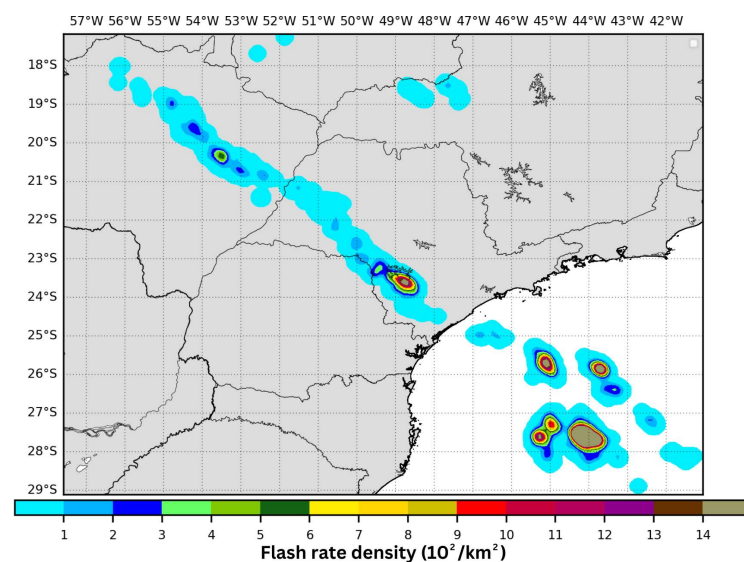

Figure 5. Flash rate density ($10^2/km^2$) derived from lightning detection data and interpolated onto the model's grid, which has a 9 km horizontal resolution. This corresponds to the same period shown in Figure 4.

It is observable that the assimilation algorithm invariably adds water vapor to the atmosphere's initial conditions. However, the number of these simulated observation stations does not directly correlate with the amount of water vapor added, as illustrated by comparing Figures 4 and 5 In regions with numerous stations, there is not necessarily a greater increase in Q_v. Notably, if a particular area was previously dry and then experiences

a lightning event, the assimilation algorithm will induce a more significant increase in Q_v than in an area that was already saturated, as shown in Figures 4 and 5.

The lightning assimilation system (LAS) is structured around its main code, from which all necessary modules and functions are executed in a specified order. Key operations involve modules dedicated to reading and writing data. One module uses the NetCDF library to read variables and grid dimensions from WRF NetCDF files, while another handles the initialization and creation of files documenting flash rate density, water vapor mixing ratios, and details about simulated observation stations.

The reading of lightning data and their interpolation to match the model's resolution are managed by specific functions within the system. Additionally, subroutines calculate the water vapor mixing ratio, temperature, and relative humidity. Other functions calculate spherical distances to map lightning detections to grid points, adjust the water vapor mixing ratio variable using the established equation, and perform linear regression to determine the relative humidity threshold. These processes are essential for refining the model's inputs and enhancing its accuracy.

Observational data from lightning detections are merged with the initial conditions produced by the "REAL" component within the LAS code (Figure 6). Following this, Q_v is updated, leading to the creation of simulated observation stations, which are then converted into the right format for the assimilation in WRFDA.
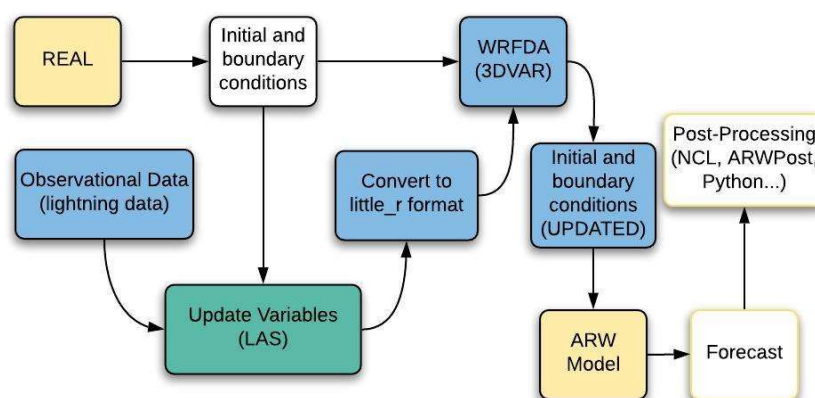

Figure 6. Diagram illustrating the integration of the assimilation algorithm with the WRF model's structure [5,6]. It shows how the normal workflow of the model was modified to incorporate the data assimilation process. Highlighted in blue are structures necessary for the assimilation, while the assimilation algorithm developed in this study is highlighted in green.

3.3. Adaptative Threshold

This research implemented an adaptive threshold for RH rather than adhering to a fixed threshold as previously established by Fierro et al. [16] and Wang et al. [55] at 81%. It was observed that an adaptive approach enhances performance, particularly during the dissipation phase of meteorological systems. When convection activity is waning yet the system still generates lightning, inserting data at this stage can inappropriately amplify convection in the model. To address this while also minimizing the model's spin-up time to accurately represent convection early in simulations, a balanced approach to the introduction of water vapor is crucial.

The adaptive RH threshold is derived from the concept of the lightning jump, which tracks the change in flash rate over time and is often utilized to signal severe weather events [56,57]. In this study, the flash rate was monitored over 30 min intervals in the two hours preceding the simulation's start. By applying linear regression to these flash rates, the optimal angle α that minimizes prediction errors was determined (Figure 7). An RH

threshold of 81% is maintained if α exceeds $30°$; otherwise, the threshold is adjusted to 60% (Equation (2)).

$$f(\alpha) = \begin{cases} 0.81, \; if \; \alpha > 30° \\ 0.60, \; otherwise \end{cases} \tag{2}$$

These thresholds and angles were empirically set during the experimental testing phase. The implementation of this adaptive threshold is anticipated to improve the model's performance even during the dissipation phases of meteorological systems.
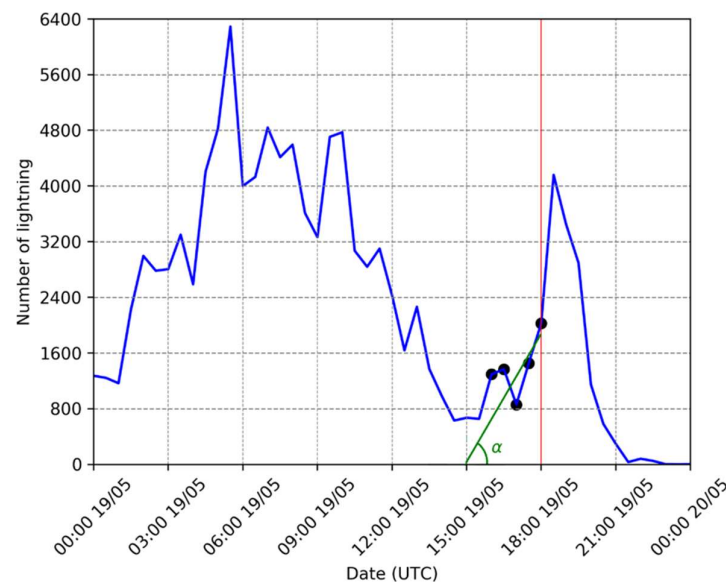

Figure 7. This graph displays the temporal variation in the number of lightning detections every 30 min for a specific case, represented by a blue line. The red line indicates the analysis time (18:00 on 19 May), while the black dots denote the data points used to compute the linear regression model. The green line represents the best fit to these points, and α is the angle used to determine the RH correction, as in Equation (2).

3.4. Performance Evaluation

The evaluation was focused on the precipitation variable. Initially, the Barnes algorithm [58] was employed to interpolate observational data onto a grid corresponding to the model's configuration (refer to Section 3.1). Equation (3) outlines the interpolation approach, where x_i^n represents the precipitation at the i-th grid point during the n-th iteration:

$$x_i^{n+1} = x_i^n + \frac{\sum_{k=1}^{K} w_{i,k} \left(y_k - x_i^n \right)}{\sum_{k=1}^{K} w_{i,k}} \tag{3}$$

where $\forall i : x_i^0 = 0$, y_k denotes the k-th observation, and $w_{i,k}$ is the weighting function for the k-th observation at the i-th grid point, defined in Equation (4).

$$w_{i,k} = exp \left[- \left(\frac{d_{i,k}}{g^n} \right)^\alpha \right] \tag{4}$$

where $d_{i,k}$ measures the distance in kilometers between the i-th grid point and the k-th observation, α is a constant greater than 1, and g^n is the decay parameter in the n-th iteration, evolving as shown in Equation (5).

$$g^{n+1} = cg^n \tag{5}$$

with $0 < c < 1$.

The parameters were fine-tuned empirically by comparing them against the observational data, resulting in two iterations with $\alpha = 1.5$, $g^o = 80$, and $c = 0.5$. The distances were calculated using spherical trigonometry.

An illustration of this method is depicted in Figure 8, showing the observed data and the interpolated precipitation field. After aligning the precipitation data with the model grid, it became feasible to employ various techniques to analyze the observational data against the simulation outcomes.
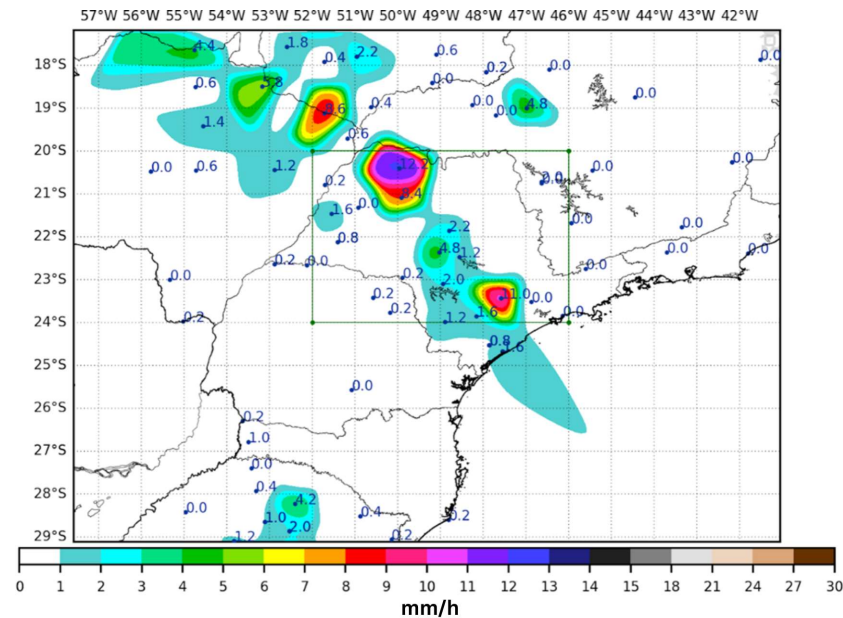

Figure 8. Illustration of the Barnes interpolation algorithm applied to precipitation data (measured in mm/h) using datasets from INPE/CPTEC.

The analysis tools included the threat scores (TS), false alarm ratio (FAR), and probability of detection (POD), defined by Equations (6)–(8), respectively:

$$TS = \frac{hits}{hits + misses + false\ alarms} \tag{6}$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms} \tag{7}$$

$$POD = \frac{hits}{hits + misses} \tag{8}$$

These metrics, defined in Table 2, outline performance with ideal scores of 1.0 for TS and POD, and 0.0 for FAR. The contingency table also elucidates terms like "hits," "false alarms," and "misses" based on model predictions and observed meteorological data. A "hit" occurs when both the model forecasts and the observations confirm precipitation in a specific area. A "false alarm" refers to the model predicting precipitation where none was observed. Conversely, a "miss" is noted when precipitation occurs but was not forecasted by the model.

Table 2. Contingency table.

Simulated/Observed	Yes	No
Yes	Hits	False alarms
No	Misses	Correct negatives

Occasionally, the precipitation forecasted by the model may not align perfectly with observations because the simulated fields are temporally delayed relative to the actual events. Nonetheless, thresholds were defined to minimize this issue. The following thresholds were established for the calculations in Equations (6)–(8):

- 30 km and 1 mm;
- 20 km and 5 mm;
- 20 km and 10 mm.

For example, the first criterion is met if a recorded or simulated precipitation amount exceeds 1 mm within a 30 km radius of the designated grid point. If this condition is fulfilled by both the model and the observations, it is classified as a hit. This method is consistently applied across all grid points in the domain, aggregating all hits, false alarms, and misses to compute the equations above.

Furthermore, a BIAS estimator was used to quantitatively assess the discrepancy between the experimental results and the observed precipitation, as specified in Equation (9):

$$BIAS = \frac{1}{N}\sum_{i}^{N}(f_i - o_i) \tag{9}$$

Here, f_i and o_i represent the forecast and observation at the i-th grid point, respectively. The methodologies outlined here, along with other standard techniques in meteorology for comparing forecasts to observational data, are more thoroughly detailed in [59].

4. Results and Discussion

This section presents and discusses the outcomes of this study. It highlights the effects of the lightning data assimilation system on WRF simulations across two separate case studies, each characterized by unique meteorological conditions. Although the analysis covers just two days, it is important to note that eight short-term simulations were conducted for each day, providing a robust sample for evaluating the methodology employed in this research.

4.1. Case Study I

The initial case study, which took place on 19 May 2018, featured the highest lightning activity at 05:00 UTC (for the whole study area), with the most intense activity within the inner domain, i.e., the evaluation area, occurring between 09:00 and 12:00 UTC. This distribution of lightning strikes throughout the day is depicted in Figure 9. As outlined in Section 2, lightning activity is closely linked to storm formation, suggesting that an uptick in lightning is a marker of increased atmospheric instability. This instability often ties back to various atmospheric dynamics, including thermodynamic processes.

Thermodynamically driven instability typically manifests towards the end of the day, stemming from the intense solar heating of the Earth's surface [60–62].

Additionally, mechanical processes can also instigate this upward movement of air layers. For example, the passage of a cold front can force atmospheric layers upwards [63,64]. During the early hours, when temperatures are at their lowest and atmospheric friction decreases, cold fronts can move more rapidly. This increased movement pushes more air layers and induces greater atmospheric instability, further promoting storm development [65].

Figure 9 illustrates a surge in lightning activity in the early morning hours, suggesting the onset of storm formation linked to a cold front. A subsequent increase in lightning around 19:00 UTC corresponds to convection triggered by the daytime solar heating in an already unstable atmosphere due to the cold front.

Figure 10 features synoptic charts from 00:00 UTC on 19 May 2018, verifying the presence of a cold front in the study area. The synoptic conditions on this day led to a significant increase in lightning activity, making it an exemplary day to employ the assimilation technique. This approach helps with assessing the impact of integrating lightning data on the accuracy of the simulations.
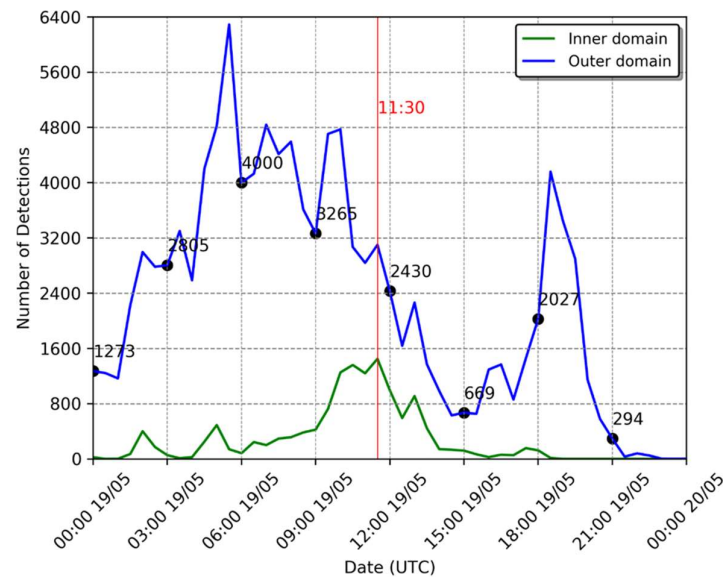
Figure 9. Lightning detections every 30 min within the inner (i.e., the evaluation area) and outer (i.e., the simulation area) domains on 19 May 2018. Outer domain refers to the simulation domain while the inner domain refers to the evaluation area.
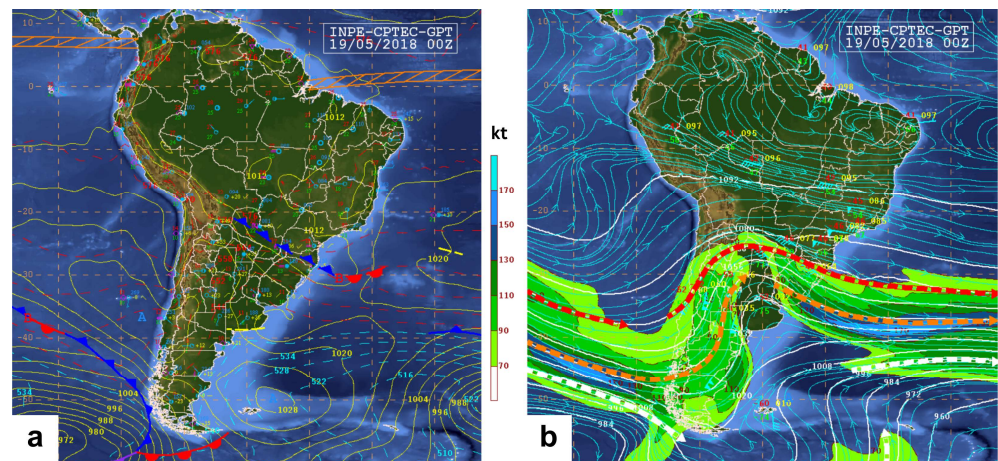

Figure 10. (**a**) Synoptic chart at 00:00 UTC on 19 May 2018, depicting the surface synoptic environment in South America. (**b**) Synoptic chart at 00:00 UTC showing the environment at 250 hPa [66].

Figure 10b displays an eastward-propagating wave at 250 hPa, inducing cyclonic curvature in the subtropical jet over southern Brazil. The misalignment of the low-pressure system at the surface (Figure 10a) indicates a baroclinic atmosphere, typical of a developing system's early stages [67].

This atmospheric configuration created a cold front that swept through the entire simulation area, reaching its peak in lightning activity in the inner domain (the São Paulo state region) at 11:30 UTC, as illustrated by the spatial distribution of lightning flashes in Figure 11.

The satellite images (Figure 12) captured the intensification of thunderstorms early in the day, aligning with the observed lightning activity in Figure 9. Although the system began to dissipate overall after 05:00 UTC, there was visible intensification of smaller storms along the southern border of São Paulo state, particularly near the coastline. These smaller storms represent the initial stages of a squall line, which later generated significant lightning activity and heavy rainfall throughout the state [68].
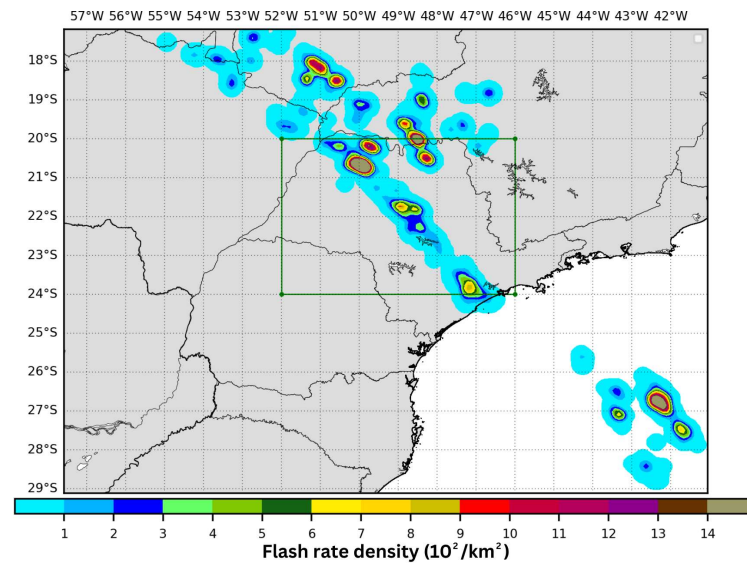
Figure 11. Flash density during the peak lightning detection period in the inner domain for a 30 min interval from 11:15 to 11:45 UTC on 19 May 2018. The green square represents the same area highlighted in Figure 1b.
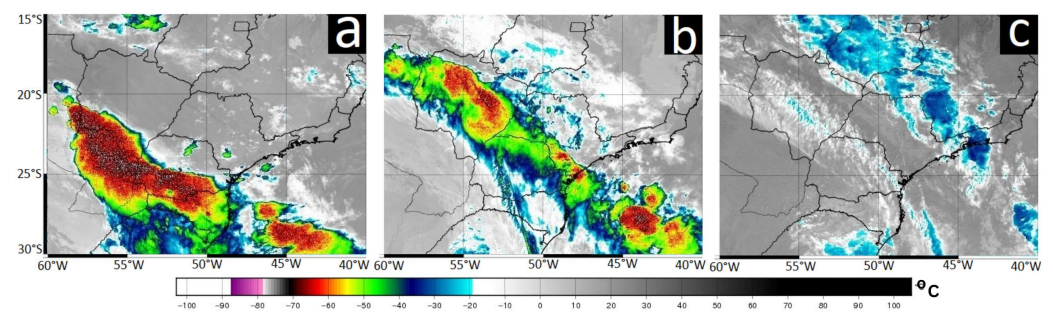
Figure 12. GOES-16 satellite images from channel 7 (3.90 μ) displaying cloud-top temperatures (°C) at 00:00 (**a**), 09:00 (**b**), and 18:00 (**c**) UTC on 19 May 2018 [69].

The lightning data assimilation technique developed by Fierro et al. [39] and later adapted by Wang et al. [55], utilized in this study alongside a variant developed for this research, primarily targeted moisture correction in the mid-levels of the troposphere. The effects and distinctions between these two approaches are depicted in Figure 13.
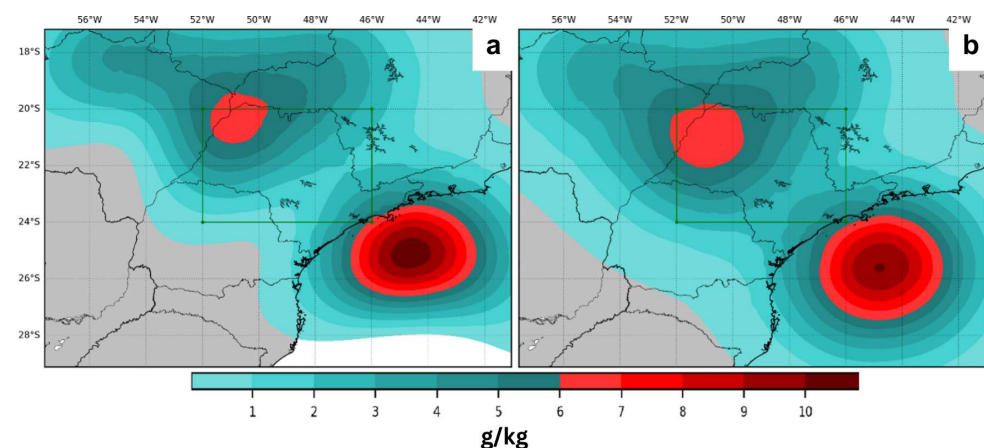
Figure 13. Comparison of Q_v increments (analysis minus CTRL) vertically integrated from the ALIGHT (**a**) and LIGHT (**b**) experiments at 09:00 UTC on 19 May 2018. The green square represents the same area highlighted in Figure 1b.

The assimilation algorithm specifically augments water vapor content, enhancing atmospheric instability. It was noted that the algorithm increased moisture, particularly in the oceanic regions near the coastline and in the northwest of São Paulo state, with slight variations observed between the two experiments. This increase in water vapor, denoted as a high Q_v increment, is typically associated with either numerous lightning detections or initially low relative humidity in areas exhibiting lightning activity.

This process is instrumental in refining short-term weather forecasts by initiating deep convection and generating corresponding cold pools, as detailed in previous studies by Fierro et al. [16] and Mansell et al. [37]. The smoothness observed in the Q_v increment field seen in Figure 13 results from the data assimilation system assimilating "simulated observations." This pattern likely emerges from recursive filters applied to two-dimensional fields of increment control variables during the background error covariance matrix computation process [70].

To assess the differences between the two experiments and quantify the added moisture in various parts of the simulation domain, the total amount of water vapor added to the atmosphere for each simulation cycle was calculated and is shown in Figure 14.
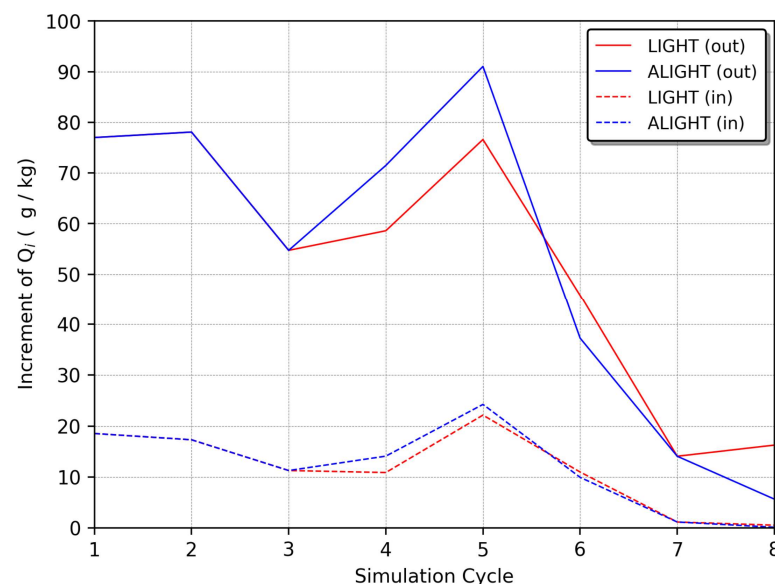

Figure 14. Total water vapor Q_v added by the assimilation algorithm in the inner and outer domains, detailed for each simulation cycle on 19 May 2018.

As mentioned in the methodology section, adjustments to the relative humidity threshold occur during the dissipating phase of the meteorological systems, leading to variations not being consistently apparent across all simulation cycles. It was observed that in certain cycles, the ALIGHT algorithm introduced more water vapor. Despite correcting fewer grid points, the ALIGHT experiment injected a greater amount of moisture due to unmodified coefficients in Equation (1).

The experiments conducted on 19 May 2018, which utilized lightning data to enhance the initial conditions of the model, demonstrated notable improvements in the early hours of the simulation compared to the control (CTRL) experiments. Specifically, the LIGHT/ALIGHT experiments showed enhanced rainfall representation and more accurate system positioning. Significant enhancements were particularly evident when observation stations recorded intense rainfall. It is known that the WRF model tends to underestimate rainfall volumes in scenarios with extreme precipitation, especially during the initial hours [71]. Thus, incorporating moisture early in the simulations through Equation (1) helped foster atmospheric instabilities, reducing the model's spin-up time and yielding quicker simulation responses.

Efforts to shorten the WRF model's typical spin-up time of 6 to 12 h have been explored in other studies, aiming for faster responses and improved initial hours of simulation [72–74]. Reducing spin-up time also significantly cuts down on computational resource usage.

Similar findings have been reported in other studies that employed lightning data assimilation techniques for initializing meteorological models, where enhancements in the representation of meteorological fields in the early simulation hours were observed [16,41,55].

When comparing the LIGHT and ALIGHT experiments, a slight improvement was noted in the latter, especially in representing the precipitation field, although these differences were only noticeable during periods with reduced lightning detections.

Figure 15 displays the cumulative precipitation field for the CTRL, LIGHT, and ALIGHT experiments, along with the observed precipitation for each simulation cycle. The experiments incorporating lightning data generally performed better, particularly in cycles with higher precipitation volumes (cycles iv and v), with the most notable performance in the ALIGHT experiment during cycle iv.
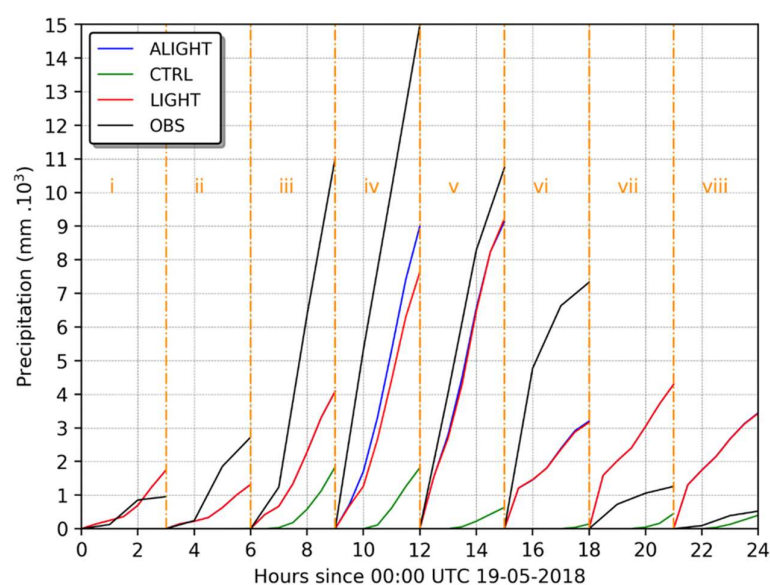

Figure 15. Accumulated precipitation totals in the evaluation area (i.e., inner domain) for each simulation cycle on 19 May 2018, broken down by experiment.

Even in the initial cycles (i and ii), where precipitation was minimal, the lightning data assimilation experiments outperformed CTRL. However, in the final cycles (vii and viii), both the LIGHT and ALIGHT experiments overestimated the observed precipitation in the domain.

This overestimation likely resulted from the simulated meteorological environment rather than the number of lightning detections by itself. During cycles i and ii, which featured 1273 and 2805 lightning detections, respectively, precipitation volumes increased sharply due to the approaching cold front, and the assimilation experiments responded well to the introduction of lightning data. Conversely, cycles vii and viii were conducted in an environment where convection was primarily thermally induced, leading to lesser rainfall despite a high incidence of lightning, a phenomenon further explored in the subsequent case study.

Large-scale systems like cold fronts typically yield higher precipitation volumes [75]. However, analyses of lightning databases indicate that significant lightning activity can occur in both large-scale and local convective systems during the southeast Brazilian summer.

The Fierro et al. [16] lightning data assimilation algorithm, developed through the analysis of numerous meteorological events across the United States, is optimized for mid-latitude environments where large-scale systems prevail. Consequently, Equation (1) is

expected to yield better results when applied to large-scale systems rather than tropical regions, where precipitation is often driven by local convection due to daytime heating.

Figure 16 shows the BIAS for the 3 h accumulated precipitation. The assimilation algorithm generally exhibited better performance during periods with higher precipitation (cycles iii, iv, v, and vi), whereas in the last cycles (vii and viii), the CTRL experiment performed better. Notably, in cycle iv, the correction implemented in the ALIGHT experiment, initially developed in this research, showed superior performance compared to both the LIGHT and CTRL experiments. Overall, the ALIGHT experiments consistently provided the best representation of precipitation volume in the domain, followed by the LIGHT experiments, as illustrated in Figure 16.
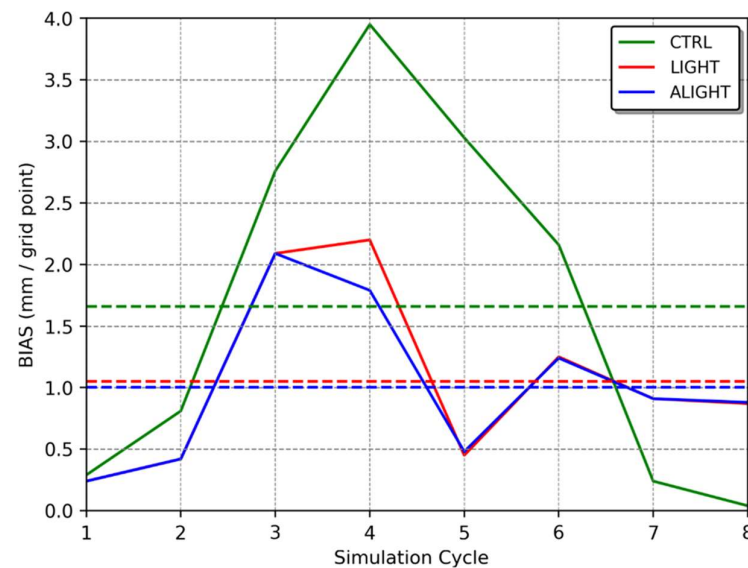

Figure 16. BIAS for three-hour accumulated precipitation corresponding to the first case for each simulation cycle, with a dashed line indicating the corresponding average.

In essence, the assimilation of lightning data effectively enhanced the precipitation field representation across almost all simulation cycles, particularly during the passage of the cold front.

In the ALIGHT experiment, more water vapor was added during cycles iv and v (as shown in Figure 14) compared to the LIGHT experiment, thereby accelerating the model's response and fostering more convection and precipitation in the initial hours.

Figure 17 spatially illustrates the impact of the data assimilation procedures by displaying the accumulated precipitation for every hour of simulation in cycle iv. The CTRL experiment struggled to accurately reproduce the precipitation field, only generating some precipitation after one hour of simulation. The precipitation field was misaligned with the meteorological system, which moved faster than the model could simulate. Fast-moving systems, such as squall lines, were particularly challenging to simulate accurately in the initial hours.

Both the LIGHT and ALIGHT experiments began producing precipitation from the outset, evolving into intense precipitation in subsequent hours. Although these experiments depicted the squall line as moving slower than observed, the quantity of rain they simulated was more closely aligned with actual observations than that of the CTRL experiment. Notably, by 12:00 UTC (three hours into the simulation), the ALIGHT experiment was the only one capable of reproducing precipitation cores with more intense rainfall (above 11 mm/h), which was closer to the observed values (Figure 17). In all experiments during cycle iv, the precipitation band associated with the squall line consistently preceded what the model could produce.
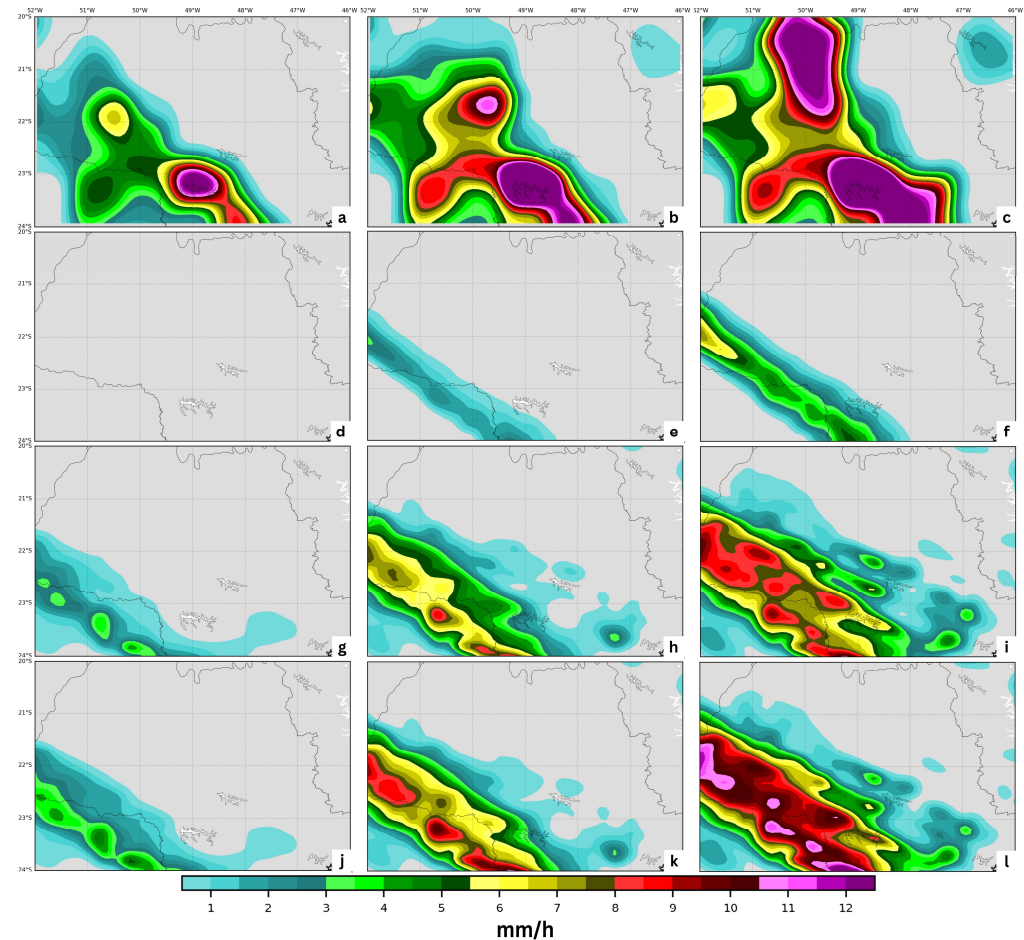
Figure 17. Precipitation accumulation in the inner domain for cycle iv (09:00 UTC on 19 May 2018), broken down by time and experiment: (**a–c**) observed precipitation; (**d–f**) CTRL experiment; (**g–i**) LIGHT experiment; (**j–l**) ALIGHT experiment. Shown at 10:00 (1 h), 11:00 (2 h), and 12:00 (3 h).

Ahasan and Debsarma [76], analyzing a squall line that affected a region in India, also employed assimilation techniques in the WRF model to simulate this extreme event more accurately. They observed that, despite some spatial and temporal biases in representing the squall line, the WRF model could appropriately simulate the synoptic environment with improvements noted in experiments utilizing data assimilation.

Similar timing and positioning biases were noted by Su et al. [77] while simulating a squall line event in eastern China using the WRF model. They highlighted the significant role of cold pools in the development and evolution of such systems, emphasizing the importance of accurately representing these features to properly simulate the squall line. Correct depiction of these atmospheric elements can enhance short-term forecasting by inducing deep convection [37].

Fierro et al. [16] also observed that accurately representing cold pools at the analysis time could enhance forecasts, particularly when the mesoscale environment was not well simulated by the model.

Despite the same timing and positioning errors observed in other studies, the incorporation of lightning data in the assimilation process yielded positive outcomes. The introduction of water vapor at the analysis time spurred deep convection in the initial hours of simulation, a phenomenon rarely achieved by the model without any data assimilation.

To quantify the improvements in system positioning due to the lightning data assimilation methodology, POD, FAR, and TS were calculated for various precipitation thresholds (refer to Section 3.4). These calculations were applied across all cycles, as shown in Table 3.

Table 3. The probability of detection (POD), false alarm ratio (FAR), and threat score (TS) calculated from the three-hour accumulated precipitation in each cycle for all three experiments (CTRL, LIGHT, and ALIGHT) across different thresholds (30 km and 1 mm, 20 km and 5 mm, 20 km and 10 mm). Null values indicate an inability to calculate due to lack of precipitation. Null means that the the specific metric was not calculated due to insufficient information.

Experiment	Simulation Cycle	30 km and 1 mm			20 km and 5 mm			20 km and 10 mm		
		POD	FAR	TS	POD	FAR	TS	POD	FAR	TS
CTRL	i	0.00	Null	0.00	0.00	Null	0.00	0.00	Null	0.00
	ii	0.00	Null	0.00	0.00	Null	0.00	0.00	Null	0.00
	iii	0.26	0.00	0.26	0.20	0.00	0.20	0.17	0.16	0.16
	iv	0.30	0.00	0.30	0.06	0.42	0.06	0.00	Null	0.00
	v	0.06	0.68	0.06	0.00	Null	0.00	0.00	Null	0.00
	vi	0.10	0.00	0.10	0.08	0.00	0.08	0.03	0.68	0.03
	vii	0.00	1.00	0.00	0.00	1.00	0.00	0.00	Null	0.00
	viii	0.17	0.94	0.04	Null	Null	Null	Null	Null	Null
	Mean	0.11	0.44	0.09	0.05	0.36	0.05	0.02	0.42	0.02
LIGHT	i	0.49	0.89	0.10	0.22	0.91	0.07	0.00	1.00	0.00
	ii	0.49	0.60	0.28	0.14	0.70	0.11	0.00	1.00	0.00
	iii	0.64	0.38	0.46	0.37	0.10	0.36	0.18	0.22	0.17
	iv	0.57	0.22	0.49	0.45	0.26	0.39	0.06	0.79	0.05
	v	0.55	0.45	0.38	0.21	0.76	0.13	0.07	0.94	0.04
	vi	0.44	0.74	0.20	0.22	0.17	0.21	0.03	0.44	0.03
	vii	0.46	0.93	0.06	0.00	1.00	0.00	0.00	1.00	0.00
	viii	0.65	0.91	0.09	Null	1.00	0.00	Null	1.00	0.00
	Mean	0.54	0.64	0.26	0.23	0.61	0.16	0.05	0.80	0.04
ALIGHT	i	0.49	0.89	0.10	0.22	0.91	0.07	0.00	1.00	0.00
	ii	0.49	0.60	0.28	0.14	0.70	0.11	0.00	1.00	0.00
	iii	0.64	0.38	0.46	0.37	0.10	0.36	0.18	0.22	0.17
	iv	0.57	0.22	0.50	0.48	0.27	0.41	0.13	0.86	0.07
	v	0.56	0.45	0.38	0.20	0.78	0.12	0.01	0.99	0.01
	vi	0.44	0.75	0.19	0.22	0.16	0.21	0.03	0.72	0.03
	vii	0.46	0.93	0.06	0.00	1.00	0.00	0.00	1.00	0.00
	viii	0.65	0.91	0.09	Null	1.00	0.00	Null	1.00	0.00
	Mean	0.54	0.64	0.26	0.23	0.62	0.15	0.05	0.85	0.04

The initial threshold (30 km and 1 mm) evaluates the model's ability to reproduce precipitation. Notably, in both experiments utilizing lightning data assimilation (LIGHT and ALIGHT), POD improved significantly, while the increase in FAR was comparatively smaller, enhancing the POD/FAR ratio. However, to accurately determine the extent of improvement in precipitation positioning due to the use of lightning data, it is essential to analyze more events, as the FAR could not always be calculated in the CTRL experiments.

This pattern persisted across other thresholds, with the lightning data assimilation experiments enhancing both POD and TS metrics, whereas FAR increased only marginally. No significant differences were observed between the LIGHT and ALIGHT experiments.

To assess how lightning data assimilation impacts medium-range forecasts, a 24 h forecast was initiated based on the promising results from cycle iv. This simulation was initiated at 09:00 UTC on 19 May 2018 and concluded at 09:00 UTC on 20 May 2018. The precipitation rate during this period is depicted in Figure 18.

The precipitation rates in the LIGHT and ALIGHT experiments displayed superior outcomes in the initial hours, and these experiments continued to more accurately represent the meteorological environment even after this period. Conversely, the CTRL experiment struggled to match the observed precipitation rate, exhibiting a delayed response, particularly in the first half of the simulation. The corresponding BIAS for this 24 h period is shown in Figure 18b.
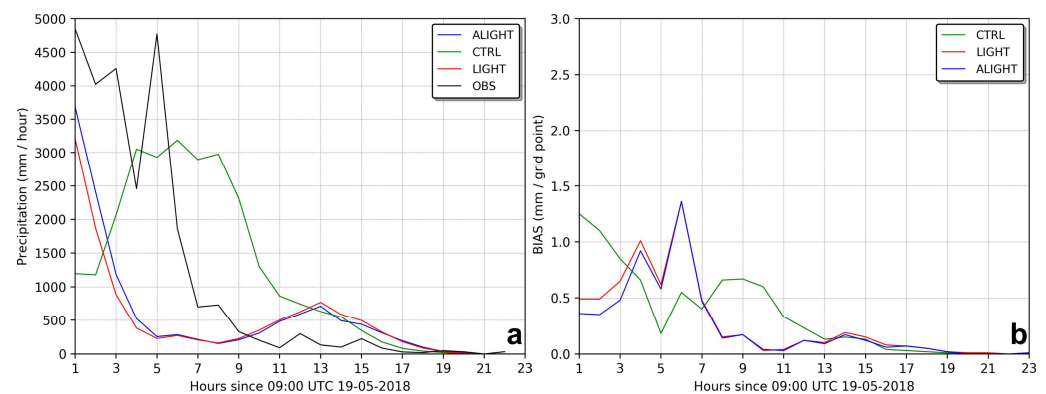
Figure 18. Simulated and observed precipitation rates from 09:00 UTC on 19 May 2018 to 08:00 UTC on 20 May 2018 (**a**), and the calculated BIAS for the precipitation rate (**b**).

An analysis of the precipitation rate BIAS indicates that the lightning data assimilation experiments more accurately reproduced rainfall amounts in the initial hours compared to the CTRL experiment. However, the apparent improvement in BIAS observed between 4 and 7 h of simulation in the CTRL experiment may not signify a genuine enhancement. As discussed previously, this could simply be a delayed response from the initial conditions, which later aligned with the observed data, contributing to the minimal BIAS noted (Figure 18b).

Fierro et al. [44] also reported enhanced performance in the accumulated precipitation forecast during the initial hours of simulation for high-impact weather events after analyzing 67 cases.

Wang et al. [55] described the algorithm developed by Fierro et al. [16] as an effective assimilation method for initiating convection where lightning data were recorded. However, it has limitations in suppressing spurious convection or modulating convection levels.

Indeed, the lightning data assimilation algorithm contributed to generating spurious convection in the simulations, which was identified as one of the sources of error in these experiments, restricting the forecast range where lightning data were beneficial during the analysis time. This spurious convection may have resulted from instabilities caused by the insertion of mass at the analysis time. Adding mass can lead to initial mass/wind imbalances and the subsequent generation of inertia–gravity waves during the model integration process [78]. Employing digital filters is one method proposed to mitigate these imbalances [79,80].

Although spurious convection occurred, the experiments incorporating lightning data demonstrated enhanced performance in short-term forecasts, with the ALIGHT experiment yielding the most favorable outcomes.

Moreover, these data assimilation experiments more accurately captured the location of precipitation compared to the CTRL experiment. As depicted in Figure 19, both POD and TS showed improvements during the first 4 h of the simulation, while the FAR was lower than that observed in the CTRL experiment at the detection threshold of 30 km and 1 mm. No substantial differences were noted between the LIGHT and ALIGHT experiments.

Wang et al. [55] also noted improved outcomes for accumulated precipitation within the initial 6 h of simulation when using the detection threshold for the frequency skill score in their experiment with lightning data assimilation. This suggests that the experiment utilizing lightning data was more effective in identifying regions with precipitation.

It is important to note that across all 24 h experiments (CTRL, LIGHT, and ALIGHT), the results tended to converge after approximately 17 h of simulation, indicating that the impact of lightning data assimilation may diminish in medium- and long-range forecasts.

Implementing data assimilation cycles at the start of the simulation, known as dynamic initialization or warm start, could potentially enhance medium-range forecasting. However, caution is advised since dynamic initialization through nudging can produce unbalanced

fields in the initial conditions, leading to both errors in the simulations and potential numerical instabilities [8].
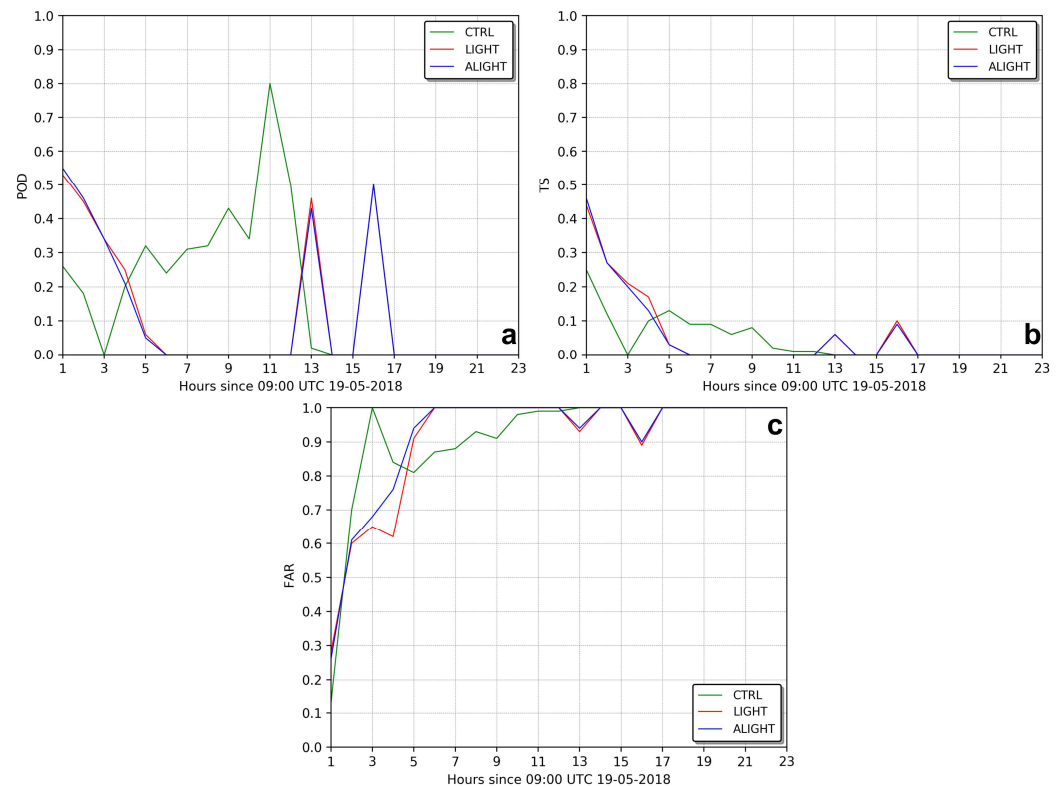

Figure 19. POD (**a**), TS (**b**), and FAR (**c**) for the precipitation rate shown in Figure 18 using the first threshold (30 km and 1 mm).

The analysis of the case from 19 May 2018 indicates that employing the lightning data assimilation system enhanced the depiction of precipitation fields and the accuracy of the meteorological system's location, particularly in the initial hours of the simulation (short-term forecast). Furthermore, it can be concluded that the algorithm adapted in this study, based on the original framework by Fierro et al. [16], demonstrated superior performance overall in comparison to other experiments conducted.

4.2. Case Study II

The second scenario examined occurred on 24 January 2018, featuring a markedly different atmospheric environment. This instance saw convective activity driven largely by thermal forcing, leading to numerous local convective systems unrelated to broader scale atmospheric dynamics. These localized systems, although driven by regional physical processes, resulted in a significant increase in lightning activity within the study area compared to the earlier case, as illustrated in Figure 20.

The peak in lightning detections occurred at 21:00 UTC, i.e., in the study area (19:00 UTC in the inner domain, i.e., in the evaluation area), a timing typical for thermally induced storms. Observations indicated that all storms within the region were triggered by similar mechanisms, as the green and blue lines in Figure 20 exhibit corresponding patterns.
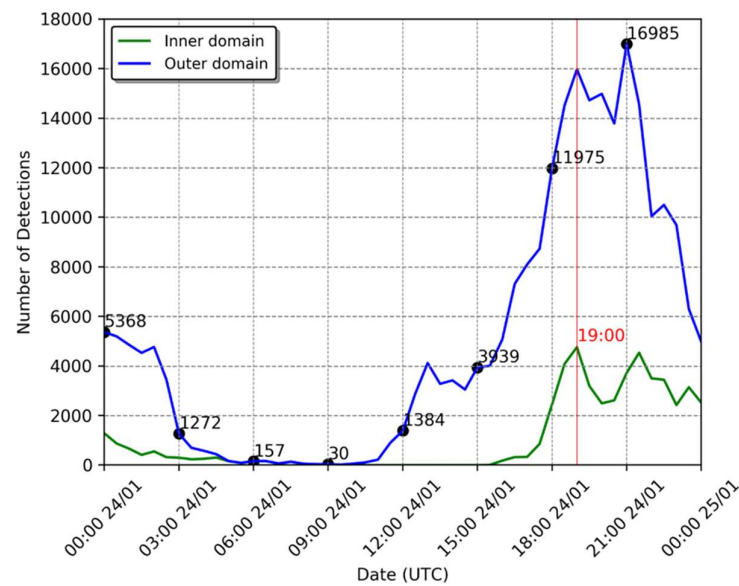
Figure 20. Lightning detections every 30 min within the inner and outer domains on 24 January 2018. Outer domain refers to the simulation domain while the inner domain refers to the evaluation area.

Further examination of the synoptic charts reveals a distinct atmospheric setup compared to the previous case (Figure 21). At the 250 hPa level, the Bolivian High was notably displaced from its climatological position, impacting the central and northern parts of the continent (Figure 21b). The Bolivian High, a high-altitude high-pressure atmospheric system, is influenced by several factors, including the Andes mountains and the Bolivian Altiplano, which force easterly winds to increase, creating divergent airflow and, consequently, a high-pressure area at higher altitudes. Additionally, intense convection in the Amazon rainforest releases substantial heat at lower levels, further bolstering the high-pressure system aloft. This displacement of the Bolivian High can occasionally occur due to atmospheric flows.
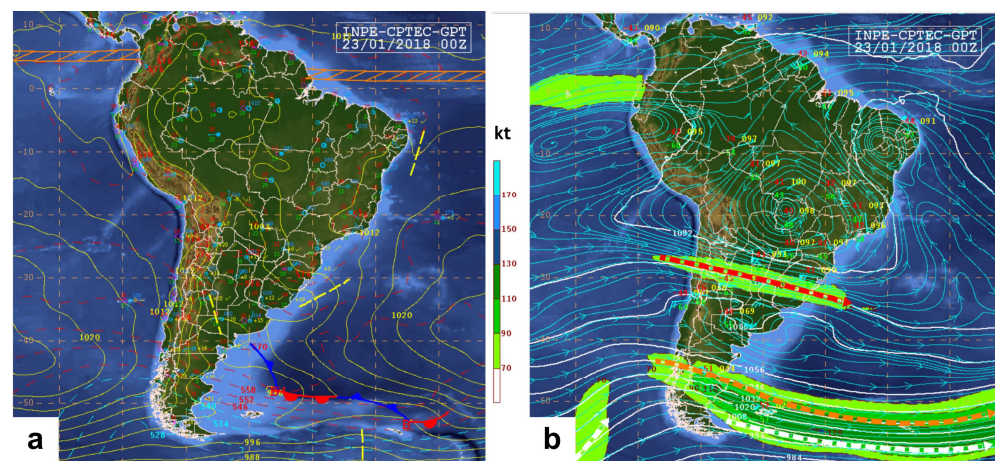

Figure 21. (**a**) Synoptic chart at 00:00 UTC on 24 January 2018, depicting the surface synoptic environment in South America. (**b**) Synoptic chart at 00:00 UTC showing the environment at 250 hPa [66].

Another significant large-scale system, the Upper Tropospheric Cyclonic Vortex (UTCV) over Northeast Brazil, is also visible [81]. The Bolivian High influences the formation of the UTCV. The UTCV, a low-pressure system, typically produces rainfall southwest of its center due to convergent airflow drawing moisture from the ocean.

This atmospheric configuration acts as an atmospheric blocking, preventing cold fronts from the south from reaching the Brazilian southeast and promoting convection driven by thermodynamics in the region [64,82].

Figure 22 depicts the maximum activity of flash density in the inner domain, showing numerous individual storms producing lightning mostly uniformly across the entire domain, a common occurrence when daytime heating is the primary driver of atmospheric instability and storm formation.
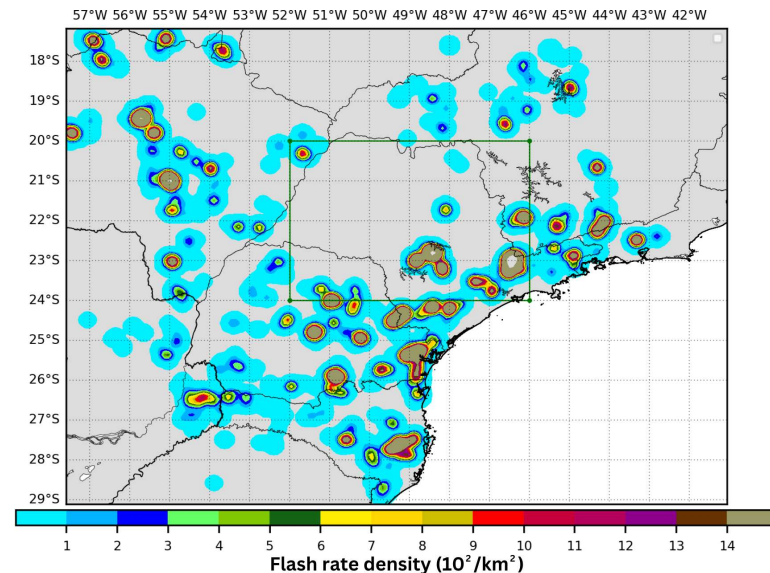

Figure 22. Flash density during the peak lightning detection period in the inner domain for a 30 min interval from 11:15 to 11:45 UTC on 24 January 2018. The green square represents the same area highlighted in Figure 1b.

The satellite images in Figure 23 illustrate the convection affecting the entire domain but with greater intensity in the states of Rio Grande do Sul and Santa Catarina. Figure 23a shows the meteorological system in its dissipation phase, having been initiated by the previous daytime heating, while Figure 23c shows that the energy from the current daytime heating began generating clouds and storms.
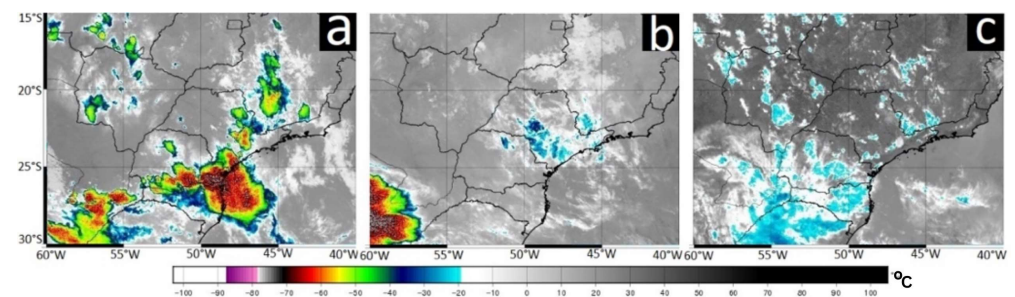

Figure 23. GOES-16 satellite images from channel 7 (3.90 μ) displaying cloud-top temperatures (°C) at 00:00 (**a**), 09:00 (**b**), and 18:00 (**c**) UTC on 24 January 2018 [69].

In contrast to the earlier case, the differences here are pronounced. In this instance, convection was primarily driven by local thermodynamics resulting from intense daytime heating, which led to a higher frequency of lightning detections, predominantly observed towards the end of the day. Consequently, the system exhibited relatively stationary behavior, characterized by localized and concentrated rainfall.

Figure 24 illustrates the impact of the lightning data assimilation algorithm on the experiment conducted on 24 January 2018, specifically showing the increase in water vapor

content (Q_v increment) at 19:00 UTC. During the periods of lightning activity (as shown in Figure 20), the algorithm, as anticipated, introduced more water vapor into the initial conditions compared to the earlier case.
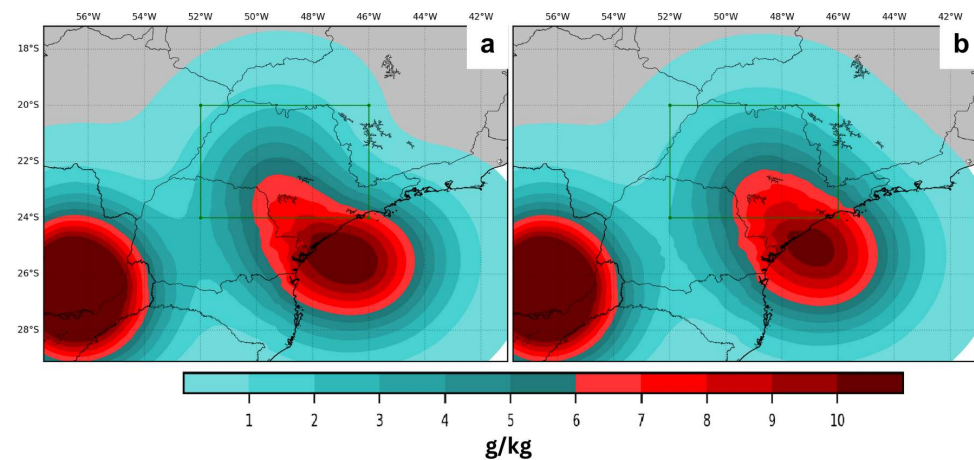

Figure 24. Comparison of Q_v increments in g/kg (analysis minus CTRL), vertically integrated, from the ALIGHT (**a**) and LIGHT (**b**) experiments at 09:00 UTC on 24 January 2018. The green square represents the same area highlighted in Figure 1b.

In the initial case during cycle iv (09:00 UTC on 19 May 2018), the algorithm accounted for 3265 lightning detections (Figure 9), leading to an addition of approximately 70 g of water vapor within the domain (Figure 14). Conversely, in cycle vi of the second case (18:00 UTC on 24 January 2018), there were 11975 lightning detections (Figure 20), resulting in the addition of about 100 g of water vapor (Figure 24).

This augmented water vapor content heightened atmospheric instability, triggering convective motions and, consequently, increased precipitation during the early hours of the simulation, as depicted in Figure 25.
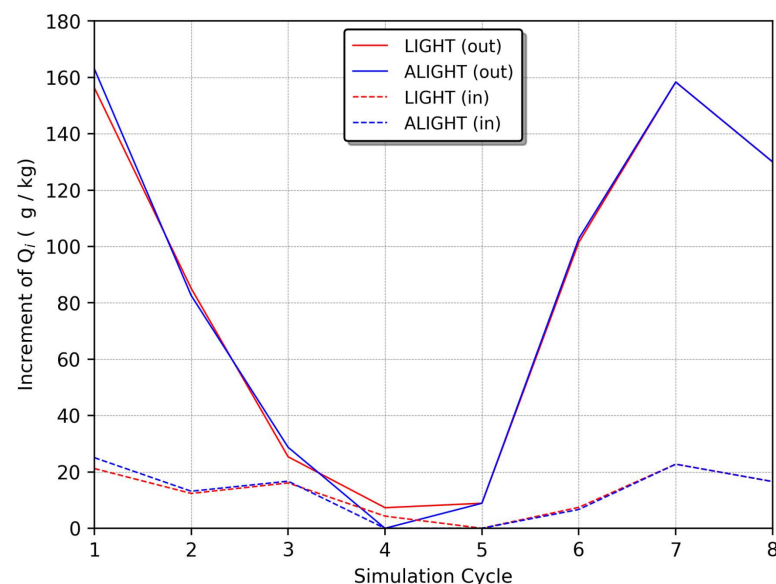

Figure 25. Total water vapor Q_v added by the assimilation algorithm in the inner and outer domains, detailed for each simulation cycle on 24 January 2018.

The experiments conducted on 24 January 2018, which utilized lightning data to enhance model initialization, did not yield the same level of improvement as those observed in the previous case. Generally, both the LIGHT and ALIGHT experiments tended

to overestimate the amount of precipitation in the study area, though they did achieve better accuracy in representing the position of the meteorological system compared to the CTRL experiment.

Contrasting with the first event studied, the lightning activity on 24 January 2018 predominantly resulted from daytime heating, peaking at 21:00 UTC, as shown in Figure 20. This pattern indicates that the lightning was primarily triggered by local convective systems, with minimal influence from larger-scale atmospheric conditions. Similar to the first case, the introduction of additional water vapor helped to foster atmospheric instabilities and subsequent convection early in the simulation. However, the experiments did not accurately capture the volume of rainfall produced by the multitude of local convection systems.

These local systems exhibited more intense lightning activity compared to the previous event. While the 19 May 2018 case recorded a higher total rainfall with a peak of around 6000 lightning detections, the 24 January 2018 event, despite having less rainfall, logged over 16,000 detections (Figures 9 and 20). These discrepancies between the number of lightning detections and the volume of rainfall affected the performance of the lightning data assimilation algorithm used in this study.

This analysis revealed that the thermal forcing on this particular day led to minimal precipitation alongside high lightning activity, which may have prompted the assimilation algorithm to add excessive water vapor, causing the LIGHT and ALIGHT experiments to overestimate accumulated precipitation.

Figure 26 illustrates the total precipitation field accumulated across the entire simulation domain for the CTRL, LIGHT, and ALIGHT experiments, alongside the observed precipitation for each of the 3 h simulation cycles.
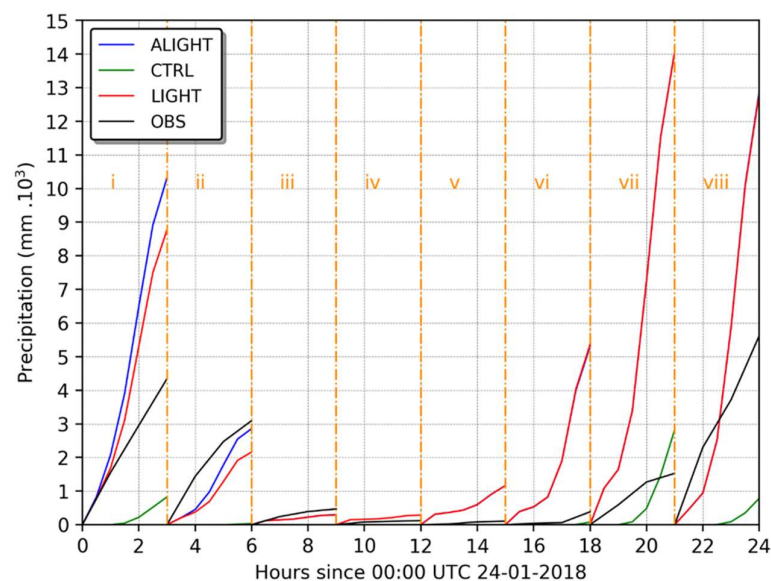

Figure 26. Accumulated precipitation totals in the evaluation area (i.e., inner domain) for each simulation cycle on 24 January 2018, broken down by experiment.

The overestimated precipitation volumes in the experiments could stem from an inadequate representation of graupel at the onset. Despite the capability to resolve convection explicitly, the experiments failed to accurately simulate graupel, possibly due to the 9 km resolution used in the simulations. Graupel plays a crucial role in thunderstorm development, particularly in local convective systems typical of the area during summer, which are characterized by deep convection producing strong updrafts and high concentrations of graupel, a key factor in lightning formation [83–85].

Further, Tao et al. [86] noted that large ice particles can suppress storm updrafts, influencing the dissipation rate of the system and consequently the precipitation output. Adams-Selin et al. [87] found that simulations allowing for large graupel formation resulted

in a minimal stratiform precipitation area with reduced convective intensity, whereas smaller graupel facilitated deeper and stronger cold pools, leading to extensive stratiform precipitation and persistent convection.

These findings align with observations from this study. The typical graupel formation within local convective systems provided the necessary conditions for frequent lightning and rapid system dissipation, consistent with the high number of lightning detections and minimal rainfall observed (Figure 26). This also elucidates why the precipitation field was overestimated by the data assimilation experiments; the inability of the model to simulate appropriate graupel concentrations led to an excessive addition of water vapor. Higher concentrations of graupel typically lead to less water vapor being added to the atmosphere.

Figure 26 demonstrates that the experiments employing data assimilation considerably overestimated the three-hour accumulated precipitation in most cycles. However, it is important to note that the first hour of accumulated precipitation simulated by the LIGHT and ALIGHT experiments showed improved performance.

Figure 27 displays the BIAS for the three-hour accumulated precipitation, where the best performance of the assimilation algorithm was seen in cycles ii, iii, and iv, with the CTRL experiment generally outperforming the others.
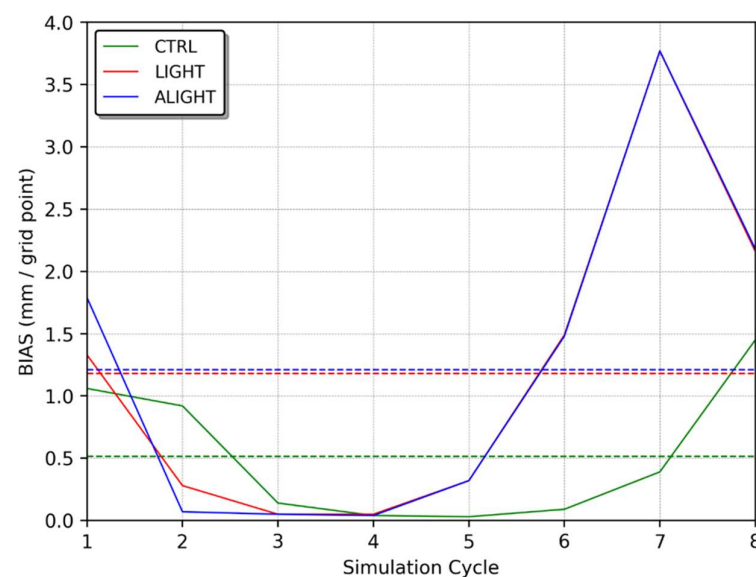

Figure 27. BIAS for three-hour accumulated precipitation corresponding to the second case for each simulation cycle, with a dashed line indicating the corresponding average.

Once again, the algorithm struggled to accurately simulate the precipitation amounts associated with thermal forcing, as evidenced by the poorer performance in the later simulation cycles of both analyzed cases (19 May and 24 January), likely due to inadequate representation of graupel mixing ratios.

In this scenario, a slight difference was observed between the ALIGHT and LIGHT experiments, attributable to the lower precipitation volumes. Despite poorer representation of the three-hour accumulated precipitation, the rapid response in the first hour of simulation suggested a beneficial impact of the assimilation.

To spatially analyze the impact of the data assimilation procedures, Figure 28 presents the accumulated precipitation for every hour of the cycle ii simulation. Although the CTRL experiment demonstrated a smaller error than other experiments, it failed to accurately reproduce the precipitation field, producing some precipitation only after the first hour of simulation. Additionally, the primary precipitation core centered around 22.5° S and 50.7° W was not accurately simulated by any of the experiments.
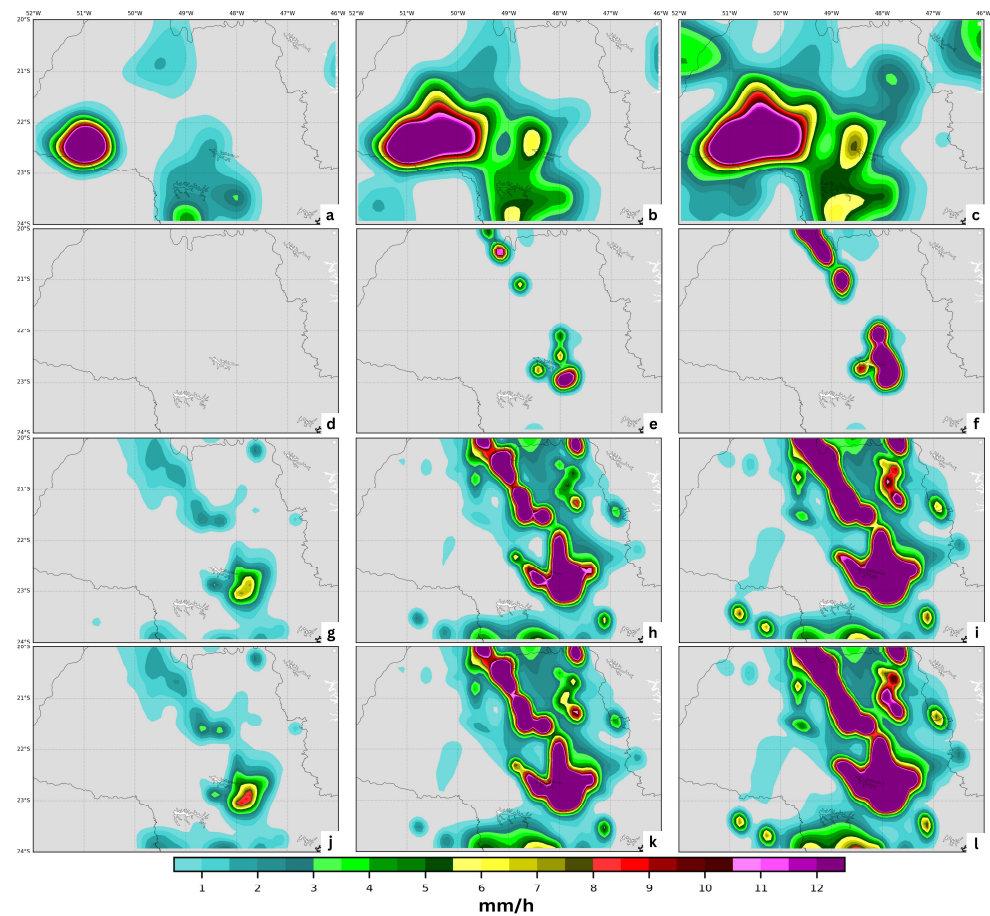
Figure 28. Precipitation accumulation in the inner domain for cycle iv (09:00 UTC on 24 January 2018), broken down by time and experiment: (**a**–**c**) observed precipitation; (**d**–**f**) CTRL experiment; (**g**–**i**) LIGHT experiment; (**j**–**l**) ALIGHT experiment. Shown at 10:00 (1 h), 11:00 (2 h), and 12:00 (3 h).

Generally, the assimilation algorithm overestimated the precipitation field, primarily associated with daytime heating. The inherent tendencies of the WRF model might also contribute to these errors, as suggested by Davis et al. [88], who, using an object-based verification methodology, found that the WRF model tends to overestimate rain area sizes during the daytime, which can lead to an overall overestimation of the precipitation field.

Although the CTRL experiment did not produce significant precipitation in the initial hours of simulation, the use of assimilation techniques reduced the spin-up period, preempting the model's natural tendency to overestimate precipitation.

Like the previous case, the variables POD, FAR, and TS were calculated for various precipitation thresholds (refer to Section 3.4) to quantitatively assess the positioning error of the simulated precipitation field. These calculations were applied across all cycles, as shown in Table 4.

Across all thresholds, the probability of detection and threat score improved slightly, with a minor increase in false alarms for the data assimilation experiments. However, null values were also noted, affecting the precise quantification of these metrics but not detracting from the overall conclusions drawn.

The ALIGHT experiment outperformed the LIGHT experiment in terms of precipitation detection, showing higher POD and TS values and fewer false alarms. When compared with the CTRL experiment, the ALIGHT experiment significantly improved the probability of detection with only a slight increase in the false alarm rate.

Across other thresholds, the experiments incorporating lightning data assimilation consistently outperformed the CTRL experiments, enhancing the accuracy of precipitation field positioning within the domain.

Table 4. The probability of detection (POD), false alarm ratio (FAR), and threat score (TS) calculated from the three-hour accumulated precipitation in each cycle for all three experiments (CTRL, LIGHT, and ALIGHT) across different thresholds (30 km and 1 mm, 20 km and 5 mm, 20 km and 10 mm). Null values indicate an inability to calculate due to lack of precipitation. Null means that the specific metric was not calculated due to insufficient information.

Experiment	Simulation Cycle	30 km and 1 mm			20 km and 5 mm			20 km and 10 mm		
		POD	FAR	TS	POD	FAR	TS	POD	FAR	TS
CTRL	i	0.24	0.67	0.16	0.05	0.82	0.04	0.03	0.97	0.02
	ii	0.00	Null	0.00	Null	1.00	Null	Null	Null	Null
	iii	0.00	Null	0.00	Null	Null	Null	Null	Null	Null
	iv	Null	Null	Null	Null	Null	Null	Null	Null	Null
	v	Null	1.00	0.00	Null	Null	Null	Null	Null	Null
	vi	0.34	0.81	0.14	Null	1.00	0.00	Null	1.00	0.00
	vii	0.40	0.82	0.14	0.19	0.98	0.02	Null	1.00	0.00
	viii	0.16	0.53	0.14	0.09	0.66	0.08	0.07	0.90	0.04
	Mean	0.19	0.77	0.08	0.11	0.86	0.03	0.05	0.97	0.01
LIGHT	i	0.44	0.77	0.18	0.29	0.88	0.09	0.27	0.96	0.04
	ii	0.28	0.82	0.12	0.10	0.92	0.05	Null	Null	Null
	iii	0.00	1.00	0.00	Null	Null	Null	Null	Null	Null
	iv	Null	1.00	0.00	Null	Null	Null	Null	Null	Null
	v	Null	1.00	0.00	Null	Null	Null	Null	Null	Null
	vi	0.47	0.96	0.04	Null	Null	Null	Null	1.00	0.00
	vii	0.47	0.87	0.11	0.36	0.98	0.02	Null	1.00	0.00
	viii	0.49	0.71	0.22	0.30	0.88	0.10	0.35	0.94	0.06
	Mean	0.36	0.89	0.08	0.26	0.91	0.06	0.31	0.97	0.02
ALIGHT	i	0.51	0.77	0.19	0.33	0.87	0.10	0.27	0.96	0.04
	ii	0.37	0.79	0.15	0.11	0.92	0.05	Null	Null	Null
	iii	0.00	1.00	0.00	Null	Null	Null	Null	Null	Null
	iv	Null	Null	Null	Null	Null	Null	Null	Null	Null
	v	Null	1.00	0.00	Null	Null	Null	Null	Null	Null
	vi	0.47	0.96	0.04	Null	Null	Null	Null	1.00	0.00
	vii	0.47	0.87	0.11	0.36	0.98	0.02	Null	1.00	0.00
	viii	0.49	0.71	0.22	0.30	0.88	0.10	0.35	0.94	0.06
	Mean	0.39	0.87	0.10	0.27	0.91	0.06	0.31	0.97	0.02

Based on the results thus far, a 24 h simulation starting at 00:00 UTC on 25 January 2018 was conducted. This cycle was selected to assess differences between the LIGHT and ALIGHT experiments during both the dissipation phase of the system and a period marked by significant precipitation and lightning activity.

During the first three hours of simulation, the precipitation rates generated by the LIGHT and ALIGHT experiments showed improvements, while the CTRL experiment exhibited a slower response. However, all experiments converged to a similar outcome after approximately 8 h. In the latter part of the simulation (beyond 16 h), the CTRL experiment responded better, producing more precipitation and aligning more closely with the observed data compared to the LIGHT and ALIGHT experiments.

The BIAS for the precipitation rate indicated that the precipitation fields in the data assimilation experiments did not closely match the observations, as shown in Figure 29.

Despite the local miscalculations in BIAS by the LIGHT and ALIGHT experiments, the metrics of POD, TS, and FAR reflected an overall improvement in the representation of the precipitation field throughout the simulation period, particularly in the initial hours, as illustrated in Figure 30.

In the first four hours of simulation, both POD and TS saw significant enhancements without an increase in false alarms compared to the CTRL experiment. Notably, the ALIGHT experiment demonstrated the best performance during this period, with increases in POD and TS and a reduction in FAR.
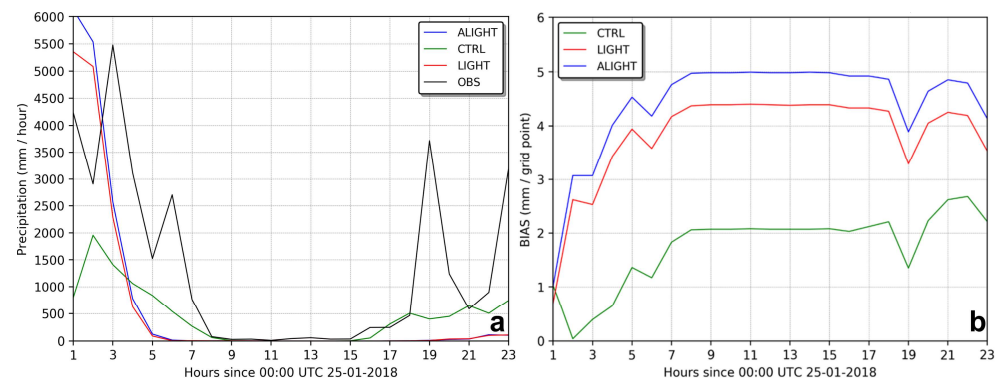
Figure 29. Simulated and observed precipitation rates from 09:00 UTC on 24 January 2018 to 08:00 UTC on 25 January 2018 (**a**), and the calculated BIAS for the precipitation rate (**b**).
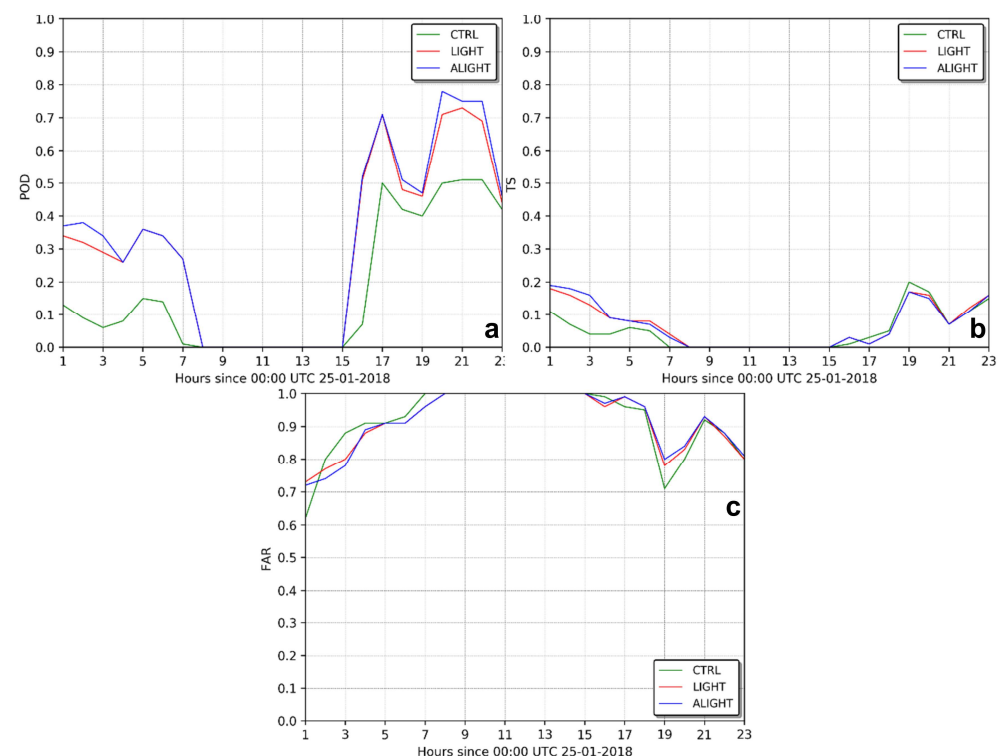

Figure 30. Details for POD (**a**), TS (**b**), and FAR (**c**) for the precipitation rate are shown in Figure 29 using the first threshold (30 km and 1 mm).

The findings from the case study on 24 January 2018 revealed that the integration of the lightning data assimilation system enhanced the timing and positioning of convection activity and the precipitation field during the initial hours of the simulation (3 h). Nonetheless, the rainfall amounts were overestimated, partially due to the surplus water vapor introduced during the analysis phase. Consequently, adjustments are still necessary for the lightning data assimilation algorithm applied during the warm season in southeastern Brazil, as demonstrated in this study.

5. Conclusions

This study evaluated the integration of lightning data into the WRF model using the WRFDA system with the 3DVAR method. It marked the first application of a lightning data assimilation technique using data from BrasilDAT to enhance short-term weather forecasting in South America.

A key contribution of this research was the development and adjustment of an assimilation algorithm for application in Brazil, including a modification to the relative humidity threshold within the original algorithm. This adjustment played a crucial role in refining the lightning data assimilation process for this region, particularly during the dissipation phase of weather systems.

The results of the experiments indicate that the lightning data assimilation system generally enhanced the accuracy of short-term precipitation forecasts for large-scale systems, particularly with the applied humidity threshold correction. Moreover, the algorithm successfully refined the timing and placement of a squall line, likely due to an accurate depiction of cold pools during the assimilation phase.

However, the study also noted limitations. For example, in scenarios where convection was driven by thermal forcing, the assimilation of lightning data had an adverse effect on the forecast. These convective systems, characterized by deep convection with high lightning activity but less precipitation, presented challenges. The model struggled to accurately simulate ice concentrations, leading the assimilation algorithm to introduce excessive moisture and overestimate precipitation. The horizontal resolution used in this study may also have influenced these outcomes, suggesting an area for future investigation.

The type of convective system significantly impacted the performance of the experiments. While the study improved the accuracy in positioning some meteorological systems, it struggled with precipitation fields generated by thermal forcing.

The introduction of lightning data as an alternative source for data assimilation represents a significant advancement, potentially enhancing short-term forecasts across various sectors. Lightning data offer high temporal and spatial resolution, providing detailed insights into storm characteristics. Furthermore, the infrastructure for lightning detection is relatively simple to deploy and maintain compared to other observational technologies like satellites and radars.

As lightning data assimilation in meteorological models is a relatively new research area, this study opens numerous avenues for further investigation. It specifically addressed the impact of such systems in South America and proposed a necessary algorithmic correction. The findings revealed that the performance of the assimilation algorithm could be influenced by the model's inability to replicate graupel concentration, highlighting a critical area for improvement.

Future research directions could include:

1. Evaluating the assimilation algorithm's performance in different seasonal and geographical contexts, particularly in mid-latitude regions during winter, where large-scale systems are predominant.
2. Adjusting assimilation algorithm coefficients, potentially using machine learning to optimize settings and enhance performance across various atmospheric conditions.
3. Exploring alternative parameterizations and/or spatial resolutions to improve model representation of graupel, which could in turn refine the assimilation process.
4. Implementing multiple assimilation cycles to enhance initial model conditions while also incorporating filters to mitigate numerical instabilities from added mass.
5. Comparing different assimilation approaches, such as 4DVAR, Kalman filtering, and hybrid methods, to further optimize the data assimilation framework.

This study's findings and recommendations pave the way for enhancing the accuracy and reliability of weather forecasting models through the use of lightning data, underscoring the need for continued research and development in this field.

Author Contributions: Conceptualization, V.V.J.; methodology, V.V.J. and R.C.F.; software, V.V.J.; validation, V.V.J. and D.L.H.; formal analysis, V.V.J.; investigation, V.V.J.; resources, V.V.J.; data curation, V.V.J.; writing—original draft preparation, V.V.J.; writing—review and editing, V.V.J. and D.L.H.; visualization, V.V.J.; supervision, O.P.J. and D.L.H.; project administration, O.P.J. and D.L.H. All authors have read and agreed to the published version of the manuscript.

References

1. Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*; Cambridge University Press: Cambridge, UK, 2003.
2. Lorenz, E.N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **1963**, *20*, 130–141. [CrossRef]
3. Davidson, N.E.; Puri, K. Tropical prediction using dynamical nudging, satellite-defined convective heat sources, and a cyclone bogus. *Mon. Weather Rev.* **1992**, *120*, 2501–2522. [CrossRef]
4. Manobianco, J.; Koch, S.; Karyampudi, V.M.; Negri, A.J. The impact of assimilating satellite-derived precipitation rates on numerical simulations of the ERICA IOP 4 cyclone. *Mon. Weather Rev.* **1994**, *122*, 341–365. [CrossRef]
5. Skamarock, W.C. *A Description of the Advanced Research WRF Version 2*; No. NCAR/TN-468+STR; University Corporation for Atmospheric Research: Boulder, CO, USA, 2005. [CrossRef]
6. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.Y.; Wang, W.; Powers, J.G. A description of the Advanced Research WRF Version 3. *NCAR Tech. Note* **2008**, *475*, 10–5065.
7. Bergthörsson, P.; Döös, B.R. Numerical weather map analysis. *Tellus* **1955**, *7*, 329–340. [CrossRef]
8. Hoke, J.E.; Anthes, R.A. The initialization of numerical models by a dynamic-initialization technique. *Mon. Weather Rev.* **1976**, *104*, 1551–1556. [CrossRef]
9. Kistler, R.E. A Study of Data Assimilation Techniques in an Autobarotropic, Primitive Equation, Channel Model. Ph.D. Thesis, Pennsylvania State University, University Park, PA, USA, 1974.
10. Lorenc, A.C. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **1986**, *112*, 1177–1194. [CrossRef]
11. Parrish, D.F.; Derber, J.C. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **1992**, *120*, 1747–1763. [CrossRef]
12. Barker, D.M.; Huang, W.; Guo, Y.-R.; Bourgeois, A.J.; Xiao, Q.N. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Weather Rev.* **2004**, *132*, 897–914. [CrossRef]
13. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]
14. Hamill, T.M.; Snyder, C. A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Weather Rev.* **2000**, *128*, 2905–2919. [CrossRef]
15. Wang, X.; Barker, D.M.; Snyder, C.; Hamill, T.M. A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Weather Rev.* **2008**, *136*, 5116–5131. [CrossRef]
16. Fierro, A.O.; Mansell, E.R.; Ziegler, C.L.; MacGorman, D.R. Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. *Mon. Weather Rev.* **2012**, *140*, 2609–2627. [CrossRef]
17. MacGorman, D.R.; Burgess, D.W.; Mazur, V.; Rust, W.D.; Taylor, W.L.; Johnson, B.C. Lightning rates relative to tornadic storm evolution on 22 May 1981. *J. Atmos. Sci.* **1989**, *46*, 221–251. [CrossRef]
18. Papadopoulos, A.; Chronis, T.G.; Anagnostou, E.N. Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale model. *Mon. Weather Rev.* **2005**, *133*, 1961–1977. [CrossRef]
19. Yang, Y.; Wang, Y.; Zhu, K. Assimilation of Chinese Doppler radar and lightning data using WRF-GSI: A case study of mesoscale convective system. *Adv. Meteorol.* **2015**, *2015*, 1–17. [CrossRef]
20. Apodaca, K.; Zupanski, M.; DeMaria, M.; Knaff, J.A.; Grasso, L.D. Development of a hybrid variational-ensemble data assimilation technique for observed lightning tested in a mesoscale model. *Nonlinear Process. Geophys.* **2014**, *21*, 1027–1041. [CrossRef]
21. Stefanescu, R.; Navon, I.M.; Fuelberg, H.; Marchand, M. 1D+ 4D-VAR data assimilation of lightning with WRFDA system using nonlinear observation operators. *arXiv* **2013**, arXiv:1306.1884.

22. Stefanescu, R.; Navon, M.I.; Marchand, M.; Fuelberg, H. Data Assimilation of Lightning in WRF 3/4-D VAR Using Observation Operators. *Mon. Weather Rev.* **2012**, *140*, 2723–2741.

23. Goodman, S.J.; Blakeslee, R.J.; Koshak, W.J.; Mach, D.; Bailey, J.; Buechler, D.; Carey, L.; Schultz, C.; Bateman, M.; McCaul, E.; et al. The GOES-R geostationary lightning mapper (GLM). *Atmos. Res.* **2013**, *125*, 34–49. [CrossRef]

24. Grenet, G. Essai d'explication de la charge electrique des nuages d'orages. *Ann. Geophys.* **1947**, *3*, 306–307.

25. Elster, J.; Geitel, H. Zur Influenztheorie der Niederschlagselektizität. *Physik Z* **1913**, *14*, 1287.

26. Reynolds, S.E.; Brook, M.; Gourley, M.F. Thunderstorm charge separation. *J. Atmos. Sci.* **1957**, *14*, 426–436. [CrossRef]

27. Baker, M.; Dash, J. Charge transfer in thunderstorms and the surface melting of ice. *J. Cryst. Growth* **1989**, *97*, 770–776. [CrossRef]

28. Fletcher, N.H. Surface structure of water and ice. *Philos. Mag.* **1962**, *7*, 255–269. [CrossRef]

29. Williams, E.R. The electrification of thunderstorms. *Sci. Am.* **1988**, *259*, 88–99. [CrossRef]

30. Cooray, G.V. *The Lightning Flash*; The Institution of Electrical Engineers: London, UK, 2003.

31. MacGorman, D.R.; Rust, W.D. *The Electrical Nature of Storms*; Oxford University Press: Oxford, UK, 1998.

32. Schultz, C.J.; Petersen, W.A.; Carey, L.D. Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends. *Weather. Forecast.* **2011**, *26*, 744–755. [CrossRef]

33. Cummins, K.L.; Murphy, M.J.; Tuel, J.V. Lightning detection methods and meteorological applications. In Proceedings of the IV International Symposium on Military Meteorology, Malbork, Poland, 26–28 September 2000; pp. 26–28.

34. Naccarato, K.P.; Bourscheidt, V.; Pinto, O. The fourth generation of the Brazilian detection efficiency model for BrasilDAT network (RDEM4). In Proceedings of the 14th International Conference on Atmospheric Electricity, Rio de Janeiro, Brazil, 8–12 August 2011; pp. 1–3.

35. Alexander, G.D.; Weinman, J.A.; Karyampudi, V.M.; Olson, W.S.; Lee, A.C.L. The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 superstorm. *Mon. Weather Rev.* **1999**, *127*, 1433–1457. [CrossRef]

36. Chang, D.-E.; Weinman, J.A.; Morales, C.A.; Olson, W.S. The effect of spaceborne microwave and ground-based continuous lightning measurements on forecasts of the 1998 Groundhog Day storm. *Mon. Weather Rev.* **2001**, *129*, 1809–1833. [CrossRef]

37. Mansell, E.R.; Ziegler, C.L.; MacGorman, D.R. A lightning data assimilation technique for mesoscale forecast models. *Mon. Weather Rev.* **2007**, *135*, 1732–1748. [CrossRef]

38. Qie, X.; Zhu, R.; Yuan, T.; Wu, X.; Li, W.; Liu, D. Application of total-lightning data assimilation in a mesoscale convective system based on the WRF model. *Atmos. Res.* **2014**, *145*, 255–266. [CrossRef]

39. Fierro, A.O.; Gao, J.; Ziegler, C.L.; Mansell, E.R.; MacGorman, D.R.; Dembek, S.R. Evaluation of a cloud-scale lightning data assimilation technique and a 3DVAR method for the analysis and short-term forecast of the 29 June 2012 derecho event. *Mon. Weather Rev.* **2014**, *142*, 183–202. [CrossRef]

40. Dixon, K.; Mass, C.F.; Hakim, G.J.; Holzworth, R.H. The impact of lightning data assimilation on deterministic and ensemble forecasts of convective events. *J. Atmos. Ocean. Technol.* **2016**, *33*, 1801–1823. [CrossRef]

41. Chen, Z.; Qie, X.; Liu, D.; Xiong, Y. Lightning data assimilation with comprehensively nudging water contents at cloud-resolving scale using WRF model. *Atmos. Res.* **2019**, *221*, 72–87. [CrossRef]

42. Pessi, A.T.; Businger, S. The impact of lightning data assimilation on a winter storm simulation over the North Pacific Ocean. *Mon. Weather Rev.* **2009**, *137*, 3177–3195. [CrossRef]

43. Lagouvardos, K.; Kotroni, V.; Defer, E.; Bousquet, O. Study of a heavy precipitation event over southern France, in the frame of HYMEX project: Observational analysis and model results using assimilation of lightning. *Atmos. Res.* **2013**, *134*, 45–55. [CrossRef]

44. Fierro, A.O.; Clark, A.J.; Mansell, E.R.; MacGorman, D.R.; Dembek, S.R.; Ziegler, C.L. Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. *Mon. Weather Rev.* **2015**, *143*, 757–777. [CrossRef]

45. Lynn, B.H.; Kelman, G.; Ellrod, G. An evaluation of the efficacy of using observed lightning to improve convective lightning forecasts. *Weather Forecast* **2015**, *30*, 405–423. [CrossRef]

46. Wang, H.; Liu, Y.; Zhao, T.; Liu, Y.; Xu, M.; Shen, S.; Jiang, Y.; Yang, H.; Feng, S. Continuous assimilation of lightning data using time-lagged ensembles for a convection-allowing numerical weather prediction model. *J. Geophys. Res. Atmos.* **2018**, *123*, 9652–9673. [CrossRef]

47. Wang, H.; Yuan, S.; Liu, Y.; Li, Y. Comparison of the WRF-FDDA-Based Radar Reflectivity and Lightning Data Assimilation for Short-Term Precipitation and Lightning Forecasts of Severe Convection. *Remote. Sens.* **2022**, *14*, 5980. [CrossRef]

48. Gonçalves, L.d.J.M.; Kaiser, J.; Palmeira, R.M.d.J.; Gallo, M.N.; Parente, C.E. Evaluation of a High Resolution WRF Model for Southeast Brazilian Coast: The Importance of Physical Parameterization to Wind Representation. *Atmosphere* **2024**, *15*, 533. [CrossRef]

49. Computational and Information Systems Laboratory (CISL). *Cheyenne: HPE/SGI ICE XA System (Climate Simulation Laboratory)*; Computational and Information Systems Laboratory (CISL): Boulder, CO, USA, 2019. [CrossRef]

50. MMM/UCAR. WRF User Page. CO, USA. 2018. Available online: https://www2.mmm.ucar.edu/wrf/users/ (accessed on 29 June 2024).

51. Hong, S.-Y.; Noh, Y.; Dudhia, J. A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Weather Rev.* **2006**, *134*, 2318–2341. [CrossRef]

52. Thompson, G.; Field, P.R.; Rasmussen, R.M.; Hall, W.D. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon Weather Rev* **2008**, *136*, 5095–5115. [CrossRef]
53. Tewari, M. Implementation and Verification of the Unified NOAH Land Surface Model in the WRF Model. In Proceedings of the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, WA, USA, 12–16 January 2004.
54. Iacono, M.J.; Delamere, J.S.; Mlawer, E.J.; Shephard, M.W.; Clough, S.A.; Collins, W.D. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res. Atmos.* **2008**, *113*, D13103. [CrossRef]
55. Wang, Y.; Yang, Y.; Liu, D.; Zhang, D.; Yao, W.; Wang, C. A case study of assimilating lightning-proxy relative humidity with WRF-3DVAR. *Atmosphere* **2017**, *8*, 55. [CrossRef]
56. Schultz, C.J.; Petersen, W.A.; Carey, L.D. Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. *J. Appl. Meteorol. Clim.* **2009**, *48*, 2543–2563. [CrossRef]
57. Williams, E.; Boldi, B.; Matlin, A.; Weber, M.; Hodanish, S.; Sharp, D.; Goodman, S.; Raghavan, R.; Buechler, D. The behavior of total lightning activity in severe Florida thunderstorms. *Atmos. Res.* **1999**, *51*, 245–265. [CrossRef]
58. Barnes, S.L. A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteorol.* **1964**, *3*, 396–409. [CrossRef]
59. Stanski, H.R.; Laurence, J.W.; Burrows, W.R. Survey of Common Verification Methods in Meteorology. 1989. Available online: https://www.cawcr.gov.au/projects/verification/Stanski_et_al/Stanski_et_al.html (accessed on 1 March 1999).
60. Browning, K.A.; Ludlam, F.H. Airflow in convective storms. *Q. J. R. Meteorol. Soc.* **1962**, *88*, 117–135. [CrossRef]
61. Petty, G.W. *A First Course in Atmospheric Thermodynamics*; Sundog Publishing LLC: Danvers, MA, USA, 2023.
62. Weisman, M.L.; Klemp, J.B. Characteristics of isolated convective storms. In *Mesoscale Meteorology and Forecasting*; American Meteorological Society: Boston, MA, USA, 1986; pp. 331–358.
63. Bergeron, T. On the physics of fronts. *Bull. Am. Meteorol. Soc.* **1937**, *18*, 265–275. [CrossRef]
64. Reboita, M.S.; Gan, M.A.; Rocha, R.P.D.; Ambrizzi, T. Regimes de precipitação na América do Sul: Uma revisão bibliográfica. *Rev. Bras. Meteorol.* **2010**, *25*, 185–204. [CrossRef]
65. Brundidge, K.C. The wind and temperature structure of nocturnal cold fronts in the first 1,420 feet. *Mon. Weather Rev.* **1965**, *93*, 587–603. [CrossRef]
66. CPTEC/INPE. Technical Report. 2019. Available online: http://tempo.cptec.inpe.br/boletimtecnico/pt (accessed on 4 April 2019).
67. Bjerknes, V. *On the Dynamics of the Circular Vortex: With Applications to the Atmosphere and Atmospheric Vortex and Wave Motions*; Cammermeyers Bokhandel: Oslo, Norway; Volume 2, No. 4.
68. Houze, R.A. Structure and dynamics of a tropical squall–line system. *Mon. Weather Rev.* **1977**, *105*, 1540–1567. [CrossRef]
69. DSA/CPTEC. *Technical Report*; DSA/CPTEC: Cachoeira Paulista, SP, Brazil, 2019.
70. Barker, D.M.; Huang, W.; Guo, Y.R.; Bourgeois, A. *A Three-Dimensional Variational (3DVAR) Data Assimilation System for Use with MM5*; No. NCAR/TN-453+STR; University Corporation for Atmospheric Research: Boulder, CO, USA, 2003. [CrossRef]
71. Moya-Álvarez, A.S.; Gálvez, J.; Holguín, A.; Estevan, R.; Kumar, S.; Villalobos, E.; Martínez-Castro, D.; Silva, Y. Extreme rainfall forecast with the WRF-ARW model in the Central Andes of Peru. *Atmosphere* **2018**, *9*, 362. [CrossRef]
72. Chen, F.; Manning, K.W.; LeMone, M.A.; Trier, S.B.; Alfieri, J.G.; Roberts, R.; Tewari, M.; Niyogi, D.; Horst, T.W.; Oncley, S.P.; et al. Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. *J. Appl. Meteorol. Clim.* **2007**, *46*, 694–713. [CrossRef]
73. Cosgrove, B.A.; Lohmann, D.; Mitchell, K.E.; Houser, P.R.; Wood, E.F.; Schaake, J.C.; Robock, A.; Sheffield, J.; Duan, Q.; Luo, L.; et al. Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). *J. Geophys. Res. Atmos.* **2003**, *108*. [CrossRef]
74. Sokol, Z. Effects of an assimilation of radar and satellite data on a very-short range forecast of heavy convective rainfalls. *Atmos. Res.* **2009**, *93*, 188–206. [CrossRef]
75. Oliveira Júnior, J.F.; Delgado, R.C.; Gois, G.; Lannes, A.; Dias, F.O.; Souza, J.C.; Souza, M. Análise da precipitação e sua relação com sistemas meteorológicos em Seropédica, Rio de Janeiro. *Floresta Ambiente* **2014**, *21*, 140–149. [CrossRef]
76. Ahasan, M.N.; Debsarma, S.K. Impact of data assimilation in simulation of thunderstorm (squall line) event over Bangladesh using WRF model, during SAARC–STORM Pilot Field Experiment 2011. *Nat. Hazards* **2015**, *75*, 1009–1022. [CrossRef]
77. Su, Y.; Ping, F.; Shen, X.-Y.; Miao, C.-S. Real-time simulation and mechanistic analysis of a squall line case in East China. *Atmos. Ocean. Sci. Lett.* **2016**, *9*, 394–400. [CrossRef]
78. Peckham, S.E.; Smirnova, T.G.; Benjamin, S.G.; Brown, J.M.; Kenyon, J.S. Implementation of a digital filter initialization in the WRF Model and its application in the Rapid Refresh. *Mon. Weather Rev.* **2016**, *144*, 99–106. [CrossRef]
79. Lynch, P.; Huang, X.-Y. Initialization of the HIRLAM model using a digital filter. *Mon. Weather Rev.* **1992**, *120*, 1019–1034. [CrossRef]
80. Harter, F.P. *Uso de Filtro Digital para Iniciar um Modelo de Área Limitada*; Instituto Nacional de Pesquisas Espaciais: São José dos Campos, Brazil, 1999.
81. Kousky, V.E.; Gan, M.A. Upper tropospheric cyclonic vortices in the tropical South Atlantic. *Tellus* **1981**, *33*, 538. [CrossRef]
82. Nascimento, E.L. Influência dos bloqueios atmosféricos na propagação de Ondas de Rossby em escoamentos de inverno no Hemisfério Sul. Ph.D. Dissertation, University of Sao Paulo, São Paulo, SP, Brazil, 1998.

83. Bringi, V.N.; Knupp, K.; Detwiler, A.; Liu, L.; Caylor, I.J.; Black, R.A. Evolution of a Florida thunderstorm during the Convection and Precipitation/Electrification Experiment: The case of 9 August 1991. *Mon. Weather Rev.* **1997**, *125*, 2131–2160. [CrossRef]
84. Dye, J.E.; Jones, J.J.; Weinheimer, A.J.; Winn, W.P. Observations within two regions of charge during initial thunderstorm electrification. *Q. J. R. Meteorol. Soc.* **1988**, *114*, 1271–1290. [CrossRef]
85. Ziegler, C.L.; MacGorman, D.R.; Dye, J.E.; Ray, P.S. A model evaluation of noninductive graupel-ice charging in the early electrification of a mountain thunderstorm. *J. Geophys. Res. Atmos.* **1991**, *96*, 12833–12855. [CrossRef]
86. Tao, Y.; Li, H.Y.; Hong, Y.C. Numerical studies on cloud physics characteristic and influence of the graupel/hail category on cloud and precipitation during a heavy rainstorm over North China. *Plateau Meteorol.* **2013**, *32*, 166–178.
87. Adams-Selin, R.D.; Heever, S.C.v.D.; Johnson, R.H. Impact of graupel parameterization schemes on idealized bow echo simulations. *Mon. Weather Rev.* **2013**, *141*, 1241–1262. [CrossRef]
88. Davis, C.; Brown, B.; Bullock, R. Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Weather Rev.* **2006**, *134*, 1772–1784. [CrossRef]