MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21c/2019/10.29.16.19-TDI

ASSESSING THE IMPACT OF LIGHTNING DATA ASSIMILATION IN THE WRF MODEL

Vanderlei Rocha de Vargas Junior

Doctorate Thesis of the Graduate Course in Space Geophysics/Atmospheric Sciences, guided by Drs. Osmar Pinto Junior, and Dirceu Luis Herdies, approved in October 09, 2019.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/3UAML3H>

INPE
São José dos Campos
2019

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GBDIR)
Serviço de Informação e Documentação (SESID)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21c/2019/10.29.16.19-TDI

ASSESSING THE IMPACT OF LIGHTNING DATA ASSIMILATION IN THE WRF MODEL

Vanderlei Rocha de Vargas Junior

Doctorate Thesis of the Graduate Course in Space Geophysics/Atmospheric Sciences, guided by Drs. Osmar Pinto Junior, and Dirceu Luis Herdies, approved in October 09, 2019.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/3UAML3H>

INPE
São José dos Campos
2019

Cataloging in Publication Data

Vargas Junior, Vanderlei Rocha de.

V426a Assessing the impact of lightning data assimilation in the WRF model / Vanderlei Rocha de Vargas Junior. – São José dos Campos : INPE, 2019.

xxiv + 117 p. ; (sid.inpe.br/mtc-m21c/2019/10.29.16.19-TDI)

Thesis (Doctorate in Space Geophysics/Atmospheric Sciences) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2019.

Guiding : Drs. Osmar Pinto Junior, and Dirceu Luis Herdies.

1. Assimilation. 2. Lightning. 3. BrasilDAT. 4. WRF. I.Title.

CDU 551.594.21

Aluno (a): *Vanderlei Rocha de Vargas Junior*

Titulo: 'ASSESSING THE IMPACT OF LIGHTNING DATA ASSIMILATION IN THE WRF MODEL"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Geofísica Espacial/Ciências Atmosféricas

Dr. Marcelo Magalhães Fares Saba

Presidente / INPE / SJCampos - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Dr. Osmar Pinto Junior

Orientador(a) / INPE / SJCampos - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Dr. Dirceu Luis Herdies

Orientador(a) / INPE / Cachoeira Paulista - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Dr. Kleber Pinheiro Naccarato

Membro da Banca / INPE / São José dos Campos - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 09 de outubro de 2019

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Geofísica Espacial/Ciências Atmosféricas

Dr. Fabricio Pereira Harter

Convidado(a) / UFPEL / Pelotas - RS

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Dr. Mario Francisco Leal de Quadro

Convidado(a) / IFSC / Florianópolis - SC

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 09 de outubro de 2019

ACKNOWLEDGEMENTS

Agradeço primeiramente aos meus pais, Zoraia Magnoria Abud e Vanderlei Rocha de Vargas, e à minha irmã, Camila Abud de Vargas, por sempre estarem ao meu lado, sendo sempre o meu maior apoio em todas as minhas conquistas.

À Isabelly M. S. A. Benites por todo apoio e companheirismo durante a maior parte do meu doutorado.

Aos meus orientadores, Dr. Osmar Pinto Junior e Dr. Dirceu Luis Herdies, os quais desempenharam um papel de fundamental importância tanto na minha vida pessoal quanto profissional ao longo desses quatro anos de doutorado.

À Dra. Karina Martinez Apodaca pela sua orientação técnica e companheirismo durante o meu período de intercâmbio.

Aos Drs. Kleber Pinheiro Naccarato, Marcelo Magalhães Fares Saba, Fabrício Pereira Harter e Mario Francisco Leal de Quadro por terem aceitado fazer parte da minha banca examinadora de doutorado e contribuído com sugestões técnicas para que este trabalho ficasse ainda mais elaborado.

A todos os membros do Grupo de Eletricidade Atmosférica (ELAT) e do Centro de Previsão e Estudos Climáticos (CPTEC) do Instituto Nacional de Pesquisas Espaciais (INPE) que direta ou indiretamente contribuíram para a conclusão deste trabalho.

Aos docentes, discentes e funcionários do curso de pós-graduação em Geofísica Espacial do INPE, pelo apoio técnico e compartilhamento de seus conhecimentos e experiências que foram fundamentais ao longo do meu doutorado.

Ao INPE, ao National Oceanic and Atmospheric Administration (NOAA), à Colorado State University (CSU), e ao Cooperative Institute for Research in the Atmosphere (CIRA) por disponibilizarem a infraestrutura necessário para a realização deste trabalho.

Finalmente, agradeço às agências de fomento à pesquisa no Brasil, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), por viabilizarem financeiramente o desenvolvimento desta pesquisa.

ABSTRACT

The increasing dependence of society on weather-sensitive technologies as well as the expansion of urban centers to risk areas are making meteorological modeling even more important in the last decades. Moreover, the development of powerful computational systems has made the implementation of new physical models capable of representing more precisely the atmosphere inducing several sectors of the economy to become even more dependent on weather forecasting. This present work is the first one to apply a lightning data assimilation technique in order to improve the short-term weather forecasting in South America. The use of this new data source in the assimilation procedures has the potential to increase the efficiency of the initialization methods currently used in meteorological operation centers, especially in South America. The main goal of this research was to implement and improve a data assimilation algorithm responsible for inserting lightning data into the WRF model. Specifically, it was intended to evaluate the performance of the experiments with lightning data assimilation comparing them with the experiments with no assimilation procedures applied, focusing on the impact in short-term forecasts. The area selected for this work was set in South America specifically over the southern portion of Brazil. This area is well covered by many types of observation stations and at the same time, it has favorable conditions for the occurrence of several meteorological systems which implies in the occurrence of many storms with a high incidence of lightning. In order to perform the simulations, evaluate the experiments and track the meteorological system it was used data from different sources such as: Precipitation data from the National Institute of Meteorology; Lightning data from BrasilDAT provided by the Atmospheric Electricity Group of the National Institute for Space Research (INPE); Satellite images from GOES-16 and synoptic weather charts from the Center for Weather Forecasting and Climate Studies of INPE; and initial and boundary conditions from the GFS model provided by the Computational and Information Systems Laboratory from University Corporation for Atmospheric Research. This study used the WRF-ARW model version 3.9.1.1 and the WRFDA system version 3.9.1 with the 3DVAR methodology. The assimilation algorithm developed in this study to assimilate lightning data and correct the initial conditions of the model was based on the equation developed by Fierro et al. (2012). This study proceeded with three different experiments during the occurrence of two distinct meteorological events aiming to assess the assimilation algorithm implemented here. The experiments were basically divided in: control (CTRL), where no assimilation procedures were used; lightning data assimilation (LIGHT), where lightning data was assimilated using the equation developed by Fierro et al. (2012); and ALIGHT, where lightning data was assimilated using the equation with an adaptative relative humidity threshold developed in this study. Based on the experiments performed in this study, it was possible to conclude that in general, the use of the Lightning Data Assimilation System improved the short-term weather forecast for the precipitation field induced by large-scale systems, especially when the correction in the relative humidity threshold was applied. Additionally, the assimilation algorithm also improved the timing and positioning of a squall line

that affected the study area possibly due to the correct representation of cold pools during the assimilation process. In the second case analyzed, the assimilation algorithm improved the representation of the precipitation field in a few simulation cycles but it was noticed that when the convection is associated with thermal forcing the assimilation of lightning data using the algorithm presented in this study had a negative impact in the experiments. The assimilation methodology for lightning data presented in this study represents a significative contribution to the data assimilation field. The operational use of an alternative data source such as lightning has the potential to improve the short-term forecasts impacting positively several sectors of society.

Keywords: Assimilation. Lightning. BrasilDAT. WRF.

AVALIAÇÃO DO IMPACTO DA ASSIMILAÇÃO DE RELÂMPAGOS NO MODELO WRF

RESUMO

A crescente dependência da sociedade em tecnologias sensíveis ao tempo bem como a expansão de centros urbanos para áreas de risco estão tornando a modelagem meteorológica ainda mais importante nas últimas décadas. Além disso, o desenvolvimento de sistemas computacionais mais eficientes tornou a implementação de novos modelos físicos não apenas capazes de representar com mais precisão a atmosfera, mas também fez com que vários setores da economia se tornassem ainda mais dependentes da previsão do tempo. Este trabalho é o primeiro a aplicar uma técnica de assimilação de dados de relâmpagos a fim de melhorar as previsões meteorológicas de curto prazo na América do Sul. O uso dessa nova fonte de dados nos procedimentos de assimilação tem o potencial de aumentar a eficiência dos métodos de inicialização atualmente utilizados em centros de operações meteorológicas, especialmente na América do Sul. O principal objetivo desta pesquisa foi implementar e aperfeiçoar um algoritmo de assimilação de dados responsável pela inserção de dados de relâmpagos no modelo WRF. Especificamente, pretendeu-se avaliar o desempenho dos experimentos com assimilação de dados de relâmpagos, comparando-os com experimentos sem procedimentos de assimilação de dados, com foco no impacto dos algoritmos de assimilação nas previsões de curto prazo. A área selecionada para este trabalho foi definida na América do Sul, especificamente na parte sul do Brasil. Esta área apresenta uma boa cobertura de estações de observação e, ao mesmo tempo, possui condições favoráveis para a ocorrência de vários sistemas meteorológicos, o que implica na ocorrência de muitas tempestades com alta incidência de relâmpagos. Para realizar as simulações, avaliar os experimentos e acompanhar os sistemas meteorológicos, foram utilizados dados de diferentes fontes, tais como: Dados de precipitação do Instituto Nacional de Meteorologia; Dados de relâmpagos da BrasilDAT fornecidos pelo Grupo de Eletricidade Atmosférica do Instituto Nacional de Pesquisas Espaciais (INPE); Imagens de satélite do GOES-16 e cartas sinóticas do Centro de Previsão Meteorológica e Estudos Climáticos do INPE; e condições iniciais e de contorno do modelo GFS fornecido pelo Computational and Information Systems Laboratory from University Corporation for Atmospheric Research. Este estudo utilizou o modelo WRF-ARW versão 3.9.1.1 e o sistema WRFDA versão 3.9.1 com a metodologia 3DVAR. O algoritmo de assimilação desenvolvido neste estudo para assimilar dados de relâmpagos e corrigir as condições iniciais do modelo foi baseado na equação desenvolvida por Fierro et al. (2012). Este estudo prosseguiu com três experimentos diferentes durante a ocorrência de dois eventos meteorológicos distintos, com o objetivo de avaliar o algoritmo de assimilação implementado. Os experimentos foram basicamente divididos em: controle (CTRL), onde não foram utilizados procedimentos de assimilação, em assimilação de dados de relâmpagos

(LIGHT), onde os dados de relâmpagos foram assimilados, e em assimilação de dados de relâmpagos com um limiar de umidade relativa adaptativo (ALIGHT). Com base nos experimentos realizados neste estudo, foi possível concluir que, em geral, o uso do Sistema de Assimilação de Dados de Relâmpagos melhorou a previsão de curto prazo para o campo de precipitação induzido por sistemas de grande escala, especialmente quando a correção do limiar de umidade relativa do ar foi aplicada. Além disso, o algoritmo de assimilação também melhorou o timing e o posicionamento de uma linha de tempestade que afetou a área de estudo, possivelmente devido à melhor representação das piscinas frias durante o processo de assimilação. No segundo caso analisado, o algoritmo de assimilação melhorou a representação do campo de precipitação em alguns ciclos de simulação, mas notou-se que, quando a convecção está associada à forçantes térmicas, a assimilação de dados de relâmpagos usando o algoritmo apresentado neste estudo teve um impacto negativo nos experimentos. A metodologia de assimilação de dados de relâmpagos apresentada neste estudo representa uma contribuição significativa para o campo de assimilação de dados. O uso operacional de uma fonte de dados alternativa como os relâmpagos tem o potencial de melhorar as previsões de curto prazo, impactando positivamente vários setores da sociedade.

Palavras-chave: Assimilação. Relâmpagos. BrasilDAT. WRF.

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS AND ABBREVIATIONS

3DVAR	Tridimensional Variational
4DVAR	Four-dimensional Variational
ARW	Advanced Research WRF
BE	Background Error Covariance Matrix
BLNET	Beijing Lightning Network
BrasilDAT	Brazilian Lightning Detection Network
BRN	Bulk Richardson Number
CAPE	Convective Available Potential Energy
CFL	Courant–Friedrichs–Lewy
CG	Cloud-to-Ground
CISL	Computational and Information Systems Laboratory
COAMPS	Couple Ocean-Atmosphere Mesoscale Prediction System
CPTEC	Center for Weather Forecasting and Climate Studies
ELAT	Atmospheric Electricity Group
ENTLN	Earth Networks Total Lightning Network
FAR	False Alarm Ratio
GFS	Global Forecast System
GLM	Geostationary Lightning Mapper
GOES-16	Geostationary Operational Environmental Satellite 16
GSI	Gridpoint Statistical Interpolation
IC	Intracloud
INMET	National Institute of Meteorology
INPE	National Institute for Space Research
LAS	Lightning Assimilation System
LF	Low Frequency
LIS	Lightning Imaging Sensor
LMA	Lightning Mapping Array
MCAPE	Maximum Convective Available Potential Energy
MCIN	Maximum Convective Inhibition Energy
MM5	Mesoscale Model
NMC	National Meteorological Center
NMM	Nonhydrostatic Mesoscale Model

NWP	Numerical Weather Prediction
PDF	Probability Density Function
POD	Probability of Detection
RH	Relative Humidity
RK3	third order Runge-Kutta
STARNET	Spheris Timing and Ranging Network
TOA	Time-of-Arrival
TS	Threat Scores
UCAR	University Corporation for Atmospheric Research
UTCV	Upper Tropospheric Cyclonic Vortex
VLF	Very Low Frequency
WPS	WRF Preprocessing System
WRF	Weather Research and Forecasting
WRFDA	WRF Data Assimilation
WWLLN	World Wide Lightning Location Network

TABLE OF CONTENTS

Page

1 INTRODUCTION

The increasing dependence of society on weather-sensitive technologies as well as the expansion of urban centers to risk areas are making meteorological modeling even more important in the last decades. Moreover, the development of powerful computational systems has made the implementation of new physical models capable of representing more precisely the atmosphere inducing several sectors of the economy to become even more dependent on weather forecasting.

The prediction of the atmospheric conditions is usually made by using a set of equations that comprehend the laws of motion and principles like the conservation of mass and energy. Those equations do not have analytical solutions and it is necessary to solve them by applying numerical methods, that is why the weather forecasting procedures it is often referred to as Numerical Weather Prediction (NWP).

In order to improve the weather forecasts, it is possible to approach different modeling aspects. We can improve the forecasts (1) improving the model core and configurations like the spatial resolution, discretization methods, etc.; (2) Improving the physics, using different parameterizations or equations which describes the reality more precisely; (3) improving the data assimilation procedures; Or (4) applying ensemble techniques.

This work approaches the weather forecasts from the point of view of the data assimilation explaining how the insertion of a new data source can improve the initial conditions and subsequently the result of the simulation. Basically, the assimilation is the process of the combination of observational data from different sources with the simulated data, known as background or first guess, in order to obtain adjusted initial conditions, known as analysis, resulting in better weather forecasts (KALNAY, 2003; LORENZ, 1963).

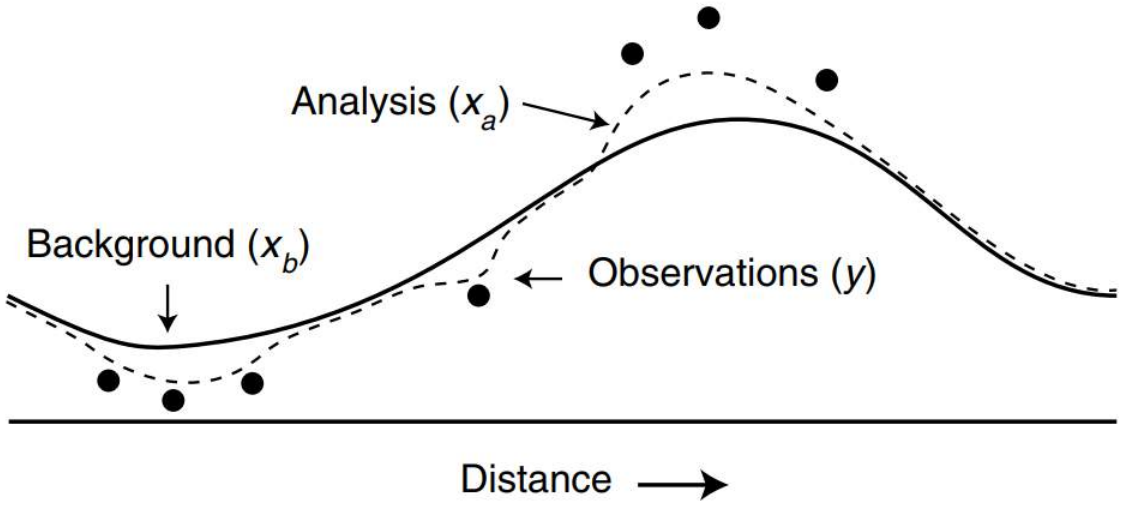
The NWP is an initial value problem, i.e., the physical equations used in atmospheric modeling not only need initial conditions but also are highly sensitive to them. The sensibility of the models associated with the initial conditions as well as the high resolution of them make the simple interpolation

of the observational data insufficient to provide proper information for the initial conditions, so the use of a background which can be obtained from another model or from a previous run with the same model is indispensable (KALNAY, 2003; LORENZ, 1963). The insertion of the observational data can correct the model background and, at the same time, decrease the cumulative errors from the simulation.

The lack of observational data in the atmosphere causes the models to initialize with imprecise initial and boundaries conditions which results in the known spin-up problem (DAVIDSON; PURI, 1992). Nevertheless, the assimilation of those observational data in the initial stages of the simulations can still smooth this problem (MANOBIANCO et al., 1994). More details are discussed in the next chapter.

Figure 1.1 shows a schematic illustration of how the observations are combined with the background in order to generate the analysis. It is possible to notice that the assimilation methodology gives more weight to the observations when they are available and when they are not, the background has more weight over the analysis.

Figure 1.1 – One-dimension schematic illustration showing how the background is combined with the observations in order to produce an analysis.



Source: Warner (2011).

In this study, the Weather Research and Forecasting (WRF) and the WRF Data Assimilation (WRFDA) were used to perform all the experiments. The WRF model is an NWP that has been widely used for both applications research and operation (SKAMAROCK et al., 2005, 2008). It is also a community system where developers can contribute to improvements in the code. This model, typically used in regional applications, has a large spectrum of physical and dynamical parameterizations which allows it to adapt to almost any place (see section 2.3 for more details).

There are many assimilation methodologies available in the atmospheric sciences which aim to combine the observation with the background such as: Successive Corrections Method (BERGTHÓRSSON; DÖÖS, 1955), Nudging (HOKE; ANTHES, 1976; KISTLER, 1974), Optimal Interpolation (LORENC et al., 2000; PARRISH; DERBER, 1992), Variational Methods (BARKER et al., 2004), Kalman Filter (KALMAN, 1960), Hybrid Methods (HAMILL; SNYDER, 2000; X. WANG et al., 2008), and others. The WRFDA system supports variational methods such as tri and four-dimensional (3DVAR, 4DVAR) and hybrid capabilities (see section 2.2 for more details about 3DVAR).

The assimilation methodologies have been widely used in the meteorological models and its impact is well studied by the scientific community. However, additional studies are necessary in order to understand which methodology is the best for operational purposes and how the integration of different types of data sources can affect the forecasts. Each data source has its own limitation, the satellite, and upper air soundings data, for example, do not provide sufficient information to elaborate appropriate initial conditions in micro and mesoscales, while the meteorological radars are limited to their coverage radius, it has just a few observations over the ocean and mountain regions can easily affect its coverage (YANG; WANG, 2015).

The use of lightning data provides information about the atmosphere that other types of observations cannot. Lightning is a good indicator of deep convection and unlike the radar observations, they are less affected by geography features (FIERRO et al., 2012; MACGORMAN et al., 1989; YANG; WANG, 2015). In

addition, the continuous observation of lightning can provide information about the development, life cycle and microphysics of a storm cloud (PAPADOPOULOS et al., 2005).

The first lightning assimilation studies used lightning data as an alternative source to generate more accurately precipitation fields and other variables (proxy fields), i.e., from the information of lightning occurrence, it was possible to infer meteorological variables indirectly for the assimilation process (ALEXANDER et al., 1999; CHANG et al., 2001).

Posteriorly, with the development of more efficient computers and assimilation methods, other forms to incorporate lightning data in the models emerged. For instance, nudging assimilation methods started to be developed for the purpose of lightning data assimilation. This technique consists in the modification of one or more equations in order to force the model to develop atmospheric features related to lightning occurrences (FIERRO et al., 2012, 2015, 2014; MANSELL et al., 2007).

Recently, other studies have shown techniques of directly lightning data incorporation through the modification of the observation operators in variational and hybrid methods (APODACA et al., 2014; STEFANESCU et al., 2012, 2013).

Currently, there are many lightning detection networks working in different ways around the world. The networks operating in the range of low frequencies are able to detect lightning in the whole world, but their detection efficiency and precision are about 50% and 10 km, respectively (RODGER et al., 2006). On the other hand, there are networks which operate with medium frequencies, although their coverage is more limited when compared to the low frequency's networks, the efficiency and precision are about 90% and 500 m, respectively (BIAGI et al., 2007; CUMMINS; KRIDER; MALONE, 1998; CUMMINS; MURPHY, 2009). Also, there are systems known as Lightning Mapping Array (LMA) which operate in the range of high frequencies (KOSHAK et al., 2004) and the Geostationary Lightning Mapper (GLM) aboard of Geostationary Operational Environmental Satellite 16 (GOES-16) which uses the optical spectrum to detect lightning from the space (GOODMAN et al., 2013).

In this study, it was implemented a methodology to calculate proxy fields which were assimilated by the WRFDA system. These proxy fields were calculated based on lightning data obtained from the Brazilian Lightning Detection Network (BrasilDAT), a medium frequency lightning detection network capable of distinguishing between Cloud-to-Ground (CG) and Intracloud (IC) flash. More details about the network will be discussed in section 2.4.

1.1 Motivation and objective

This present work is the first to apply a lightning data assimilation technique in order to improve the short-term weather forecasting in South America. The use of this new data source in the assimilation procedures has the potential to increase the efficiency of the initialization methods currently used in meteorological operation centers, especially in South America.

The use of lightning data has innumerous advantages that will be discussed in the next sections. For instance, one of the main advantages is the fact that lightning data has a high spatial and temporal resolution and it can provide information about the thunderstorms that other type of convectional observation are not able to do. At the same time, it has the potential to reduce the computational cost of the meteorological operations.

The main goal of this research was to implement and improve a data assimilation algorithm responsible for inserting lightning data into the WRF model. Specifically, it was intended to evaluate the performance of the experiments with lightning data assimilation comparing them with experiments with no assimilation procedures applied, focusing on the impact of the assimilation algorithms in short-term forecasts (≤3 hours).

1.2 Organizational structure

This work is structured in five chapters: (1) This is the present chapter which includes the introduction of the subject, basic concepts, the motivation and the goals of this research; (2) The focus of this chapter is to provide a complete review of the necessary concepts for a better understanding of the assimilation process; (3) In the chapter 3 are presented a description of the data and study area, model configurations, assimilation algorithm, experiments performed and

methods used to analyze the results; (4) This chapter shows the results obtained from the experiments performed with the model analyzing the impact of lightning data assimilation procedures and the differences between the different cases analyzed here; (5) Finally, in the last chapter are presented the conclusions and some suggestions of future works based on the results obtained in this research.

2 BACKGROUND

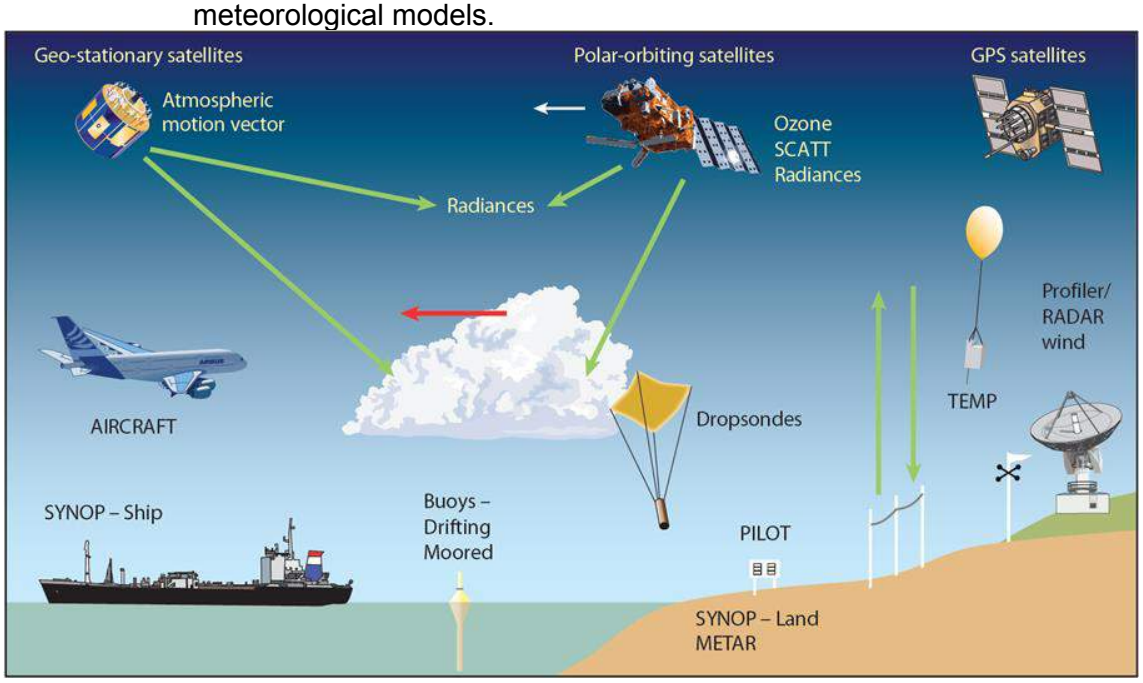
This chapter introduces all the necessary concepts for a proper understanding of this work including a background on data assimilation focusing on the algorithm used in this research. It is also discussed in more details the assimilation system (WRFDA), the meteorological model (WRF), how they were employed and how the use of lightning data can help to improve the initial conditions.

2.1 Data assimilation

The assimilation process is commonly used in the meteorological operation of the main weather prediction centers in the world. This process is applied to improve the initial conditions of the models.

Currently, many different data sources are available and all these data can be combined and used in the assimilation methodologies (Figure 2.1).

Figure 2.1 – Illustration showing different data sources commonly used in meteorological models.



Source: ECMWF (2019)

The combination of all those observational data with data from the meteorological models is used in order to minimize the errors and find a better solution for the initial conditions.

Even with all data available today, we still do not have enough observations for the proper elaboration of initial conditions in most of the meteorological models running globally. Also, the observational data has intrinsic errors associated with many factors such as calibration and other limitations associated with each data source. Even though the data from the model (background) contains enough information to elaborate the initial conditions for the next cycle, usually it has errors associated with itself (cumulative numerical errors from the integration of the equations or from not properly representing the atmospheric physics due to the approximations in the set of physical equations used in the model, for example).

The assimilation process aims to process the observational data, filter the imprecise information, and combine it with the background from the model in a statistical way decreasing the errors associated with each of it and adjusting the initial conditions for the best possible solution. The background can be provided from the simulation of another model or from a previous simulation of the same model.

The goal of the assimilation process is to minimize the errors through the combination of both data, observational and simulated (RABIER; LIU, 2003). Each data has its own error (ϵ). Taking a variable X scalar, its true value is given by X^* and the error associated with the background and the observation is given by the Equations 2.1 and 2.2, respectively.

$$\epsilon_b = X_b - X^* \tag{2.1}$$

$$\epsilon_o = X_o - X^* \tag{2.2}$$

The true value X^* is a theoretical value which is not possible to measure, and it is just used for mathematical purposes.

It is also assumed that the errors are unbiased which means that the error average is zero (Equation 2.3).

$$\overline{\epsilon_b} = \overline{\epsilon_o} = 0 \tag{2.3}$$

Moreover, the analysis X_a is given by the linear combination of the background and the observation (Equation 2.4).

$$X_a = \alpha X_o + \beta X_b + \gamma \tag{2.4}$$

Where α, β and γ are the coefficients of the equation.

The same assumption made in the Equations 2.1 and 2.2 is valid for the analysis (Equations 2.5 and 2.6).

$$\epsilon_a = X_a - X^* \tag{2.5}$$

$$\overline{\epsilon_a} = 0 \tag{2.6}$$

With the equations above, it is possible to derive the following equations (Equations 2.7 and 2.8):

$$X_a = \alpha X_o + (1 - \alpha) X_b \tag{2.7}$$

$$\overline{\epsilon_a{}^2} = \alpha \overline{\epsilon_o{}^2} + (1 - \alpha)^2 \overline{\epsilon_b{}^2} \tag{2.8}$$

Where $\overline{\epsilon_a{}^2}$, $\overline{\epsilon_b{}^2}$ and $\overline{\epsilon_o{}^2}$ are the analysis error variances ${\sigma_a}^2$, background ${\sigma_b}^2$, and observation ${\sigma_o}^2$. It was also considered that the errors are not correlated.

In order to obtain the minimum analysis error, the Equation 2.8 is minimized and the value for α is given by the Equation 2.9 ($\frac{d{\sigma_a}^2}{d\alpha} = 0$).

$$\alpha = \frac{{\sigma_b}^2}{{\sigma_b}^2 + {\sigma_o}^2} \tag{2.9}$$

The complete set of equations and more details about general data assimilation methods and how to derive the equations can be found in Bouttier and Courtier (2002).

The corrections in the initial conditions using assimilation procedures reduce the spin-up time which means that the period where the model balances the mass/wind fields and develops realistic three-dimensional features into the simulation domain is reduced. The typically spin-up time for regional models like WRF varies between 6 to 12 hours of simulation. The corrections made by the observational data in the initial conditions of the simulation provide an atmospheric state partially balanced which allows the reduction in the spin-up period (KALNAY, 2003).

The model can also be initialized using a previous run from another model or with the same model initialized before of the analysis time which allows the model to initialize with atmospheric conditions partially spun-up, i.e., balanced, this procedure is known as Warm Start or Dynamic Initialization. The opposite process is called Cold Start or Static Initialization, where the model is initialized in the analysis time and no other run is used (WARNER, 2011). The assimilation process can be used in both types of initializations, reducing in both cases the spin-up time.

Basically, the data assimilation is a branch of the Estimation Theory, i.e., with redundant measures, each of them with different range errors, it is possible to

estimate a better solution with a smaller error combining both measures statically (ATLAS; TODLING, 1999).

2.2 Variational methodology (3DVAR)

The variational methodology is one of the most used data assimilation techniques in meteorological models. This technique is based on Bayes Theorem. This theorem says that the likelihood of an event A occurs given that event B is true is given by Equation 2.10:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{2.10}$$

Where $P(A)$ and $P(B)$ are the probabilities of the events A and B occur independently of each other. While $P(B|A)$ is the probability of an event B occurs given that the event A is true.

This theorem is very useful when we already know that a certain event already happened. Considering the event $A = T$ where T is an estimative of the temperature, and $B = T_o$ where T_o is the observed temperature, the Equation 2.10 can be written as it follows (Equation 2.11).

$$p(T|T_o) = \frac{p(T)p(T_o|T)}{p(T_o)} \tag{2.11}$$

This means to say that the probability of an event T occurs given the observations T_o is equal to the probability of T multiplied by the probability of an event T_o given T divided by the probability of T_o.

Assuming that the Probability Density Function (PDF) is given by the Gaussian Distribution, the probabilities can be written as it is shown by Equations 2.12 and 2.13.

$$p(T) = \frac{1}{\sqrt{2\pi{\sigma_b}^2}} \exp\left(-\frac{1}{2}\frac{(T-T_b)^2}{{\sigma_b}^2}\right) \tag{2.12}$$

$$p(T_o) = \frac{1}{\sqrt{2\pi{\sigma_o}^2}} \exp\left(-\frac{1}{2}\frac{(T_o-T)^2}{{\sigma_o}^2}\right) \tag{2.13}$$

Where ${\sigma_b}^2$ and ${\sigma_o}^2$ are the background and the observation error variances, respectively, and T_b is the background or first guess of the model.

Using Equations 2.12 and 2.13 and replacing them into the Equation 2.11 we can obtain the proportionality relationship shown in Equation 2.14.

$$p(T|T_o) \propto \frac{1}{2\pi\sqrt{\sigma_b\sigma_o}} \exp\left(-\frac{1}{2}\left\{\frac{(T-T_b)^2}{{\sigma_b}^2} + \frac{(T_o-T)^2}{{\sigma_o}^2}\right\}\right) \tag{2.14}$$

It is possible to write another proportionality relationship based on the Equation 2.14 (Equation 2.15):

$$\ln\big(p(T|T_o)\big) \propto -\frac{1}{2}\left\{\frac{(T-T_b)^2}{{\sigma_b}^2} + \frac{(T_o-T)^2}{{\sigma_o}^2}\right\} \tag{2.15}$$

Since $p(T|T_o) < 1$ and $\ln(1) = 0$, the goal is to find the temperature T that minimizes Equation 2.16. This equation is known as the cost function.

$$J(T) = \frac{1}{2}\left\{\frac{(T-T_b)^2}{{\sigma_b}^2} + \frac{(T_o-T)^2}{{\sigma_o}^2}\right\} \tag{2.16}$$

In order to find the optimal temperature that minimizes Equation 2.16 (T_a) it is necessary to calculate $\frac{dJ(T)}{dT} = 0$ which leads to the Equation 2.17.

$$T \equiv T_a = \frac{{\sigma_b}^2 T_o}{{\sigma_o}^2 + {\sigma_b}^2} + \frac{{\sigma_o}^2 T_b}{{\sigma_o}^2 + {\sigma_b}^2} \tag{2.17}$$

Which is the same equation obtained in the previous section.

Figure 2.2 shows how the combination of J_o which is the first term on the right side of the Equation 2.16, is combined with J_b which is the second term to produce the complete cost function where the minimum of the function is what we call as analysis T_a.

The equations showed until now demonstrate the application of the assimilation process for a scalar case. In the real case, the variables are matrices containing not only other variables but also the space information. For the tridimensional variational (3DVAR) case the probability function is given by Equation 2.18.

Now, considering the event $A = \boldsymbol{x}$ where $\boldsymbol{x}$ is the estimative, a matrix resulting from the assimilation process, and $B = \boldsymbol{y_o}$ where $\boldsymbol{y_o}$ is the matrix containing the observational data, the Equation 2.10 can be rewrite as it is shown by Equation 2.18.

$$p(\boldsymbol{x}|\boldsymbol{y_o}) = \frac{p(\boldsymbol{x})p(\boldsymbol{y_o}|\boldsymbol{x})}{p(\boldsymbol{y_o})} \tag{2.18}$$

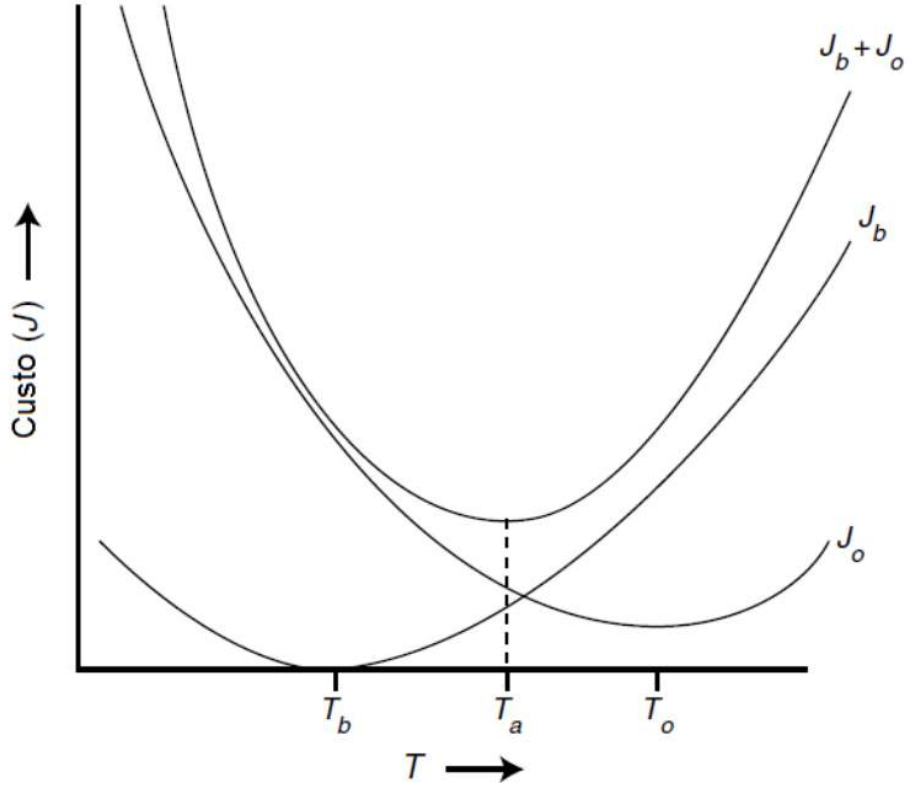
Assuming that the PDF's are given again by the Gaussian Distribution, but this time for a multidimensional case, the probabilities can be written as it is shown by Equations 2.19 and 2.20.

$$p(\boldsymbol{x}) = \frac{1}{|\boldsymbol{B}|^{1/2}\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x_b})^T \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x_b})\right) \tag{2.19}$$

$$p(\boldsymbol{y_o}|\boldsymbol{x}) = \frac{1}{|\boldsymbol{R}|^{1/2}\sqrt{(2\pi)^p}} \exp\left(-\frac{1}{2}(\boldsymbol{y_o} - \boldsymbol{h}(\boldsymbol{x}))^T \boldsymbol{R}^{-1}(\boldsymbol{y_o} - \boldsymbol{h}(\boldsymbol{x}))\right) \quad (2.20)$$

Where $\boldsymbol{B}$ and $\boldsymbol{R}$ are the background and the observation covariance error matrices, respectively. $\boldsymbol{h}(\,)$ is the observation operator responsible for bringing $\boldsymbol{x}$ from the model space to the observation space which allows it to be compared with the observations $\boldsymbol{y_o}$, and $\boldsymbol{x_b}$ is the background or first guess of the model.

Figure 2.2 – Schematic representation of the cost function for a scalar case (Equation 2.16).



Source: Adapted from Warner (2011).

Using the Equations 2.19 and 2.20 into Equation 2.18 we can obtain the proportionality relationship shown in Equation 2.21.

$$p(\boldsymbol{x}|\boldsymbol{y_o}) \propto \frac{1}{\sqrt{|\boldsymbol{R}||\boldsymbol{B}|(2\pi)^{np}}} \exp\left(-\frac{1}{2}\{(\boldsymbol{x} - \boldsymbol{x_b})^T \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x_b}) + [\boldsymbol{y_o} - \boldsymbol{h}(\boldsymbol{x})]^T \boldsymbol{R}^{-1}[\boldsymbol{y_o} - \boldsymbol{h}(\boldsymbol{x})]\}\right) \quad (2.21)$$

Equivalently in Equation 2.22:

$$\ln\big(p(\boldsymbol{x}|\boldsymbol{y_o})\big) \propto -\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{x_b})^T\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x_b}) + [\boldsymbol{y_o}-\boldsymbol{h}(\boldsymbol{x})]^T\boldsymbol{R}^{-1}[\boldsymbol{y_o}-\boldsymbol{h}(\boldsymbol{x})]\} \quad (2.22)$$

Since $p(\boldsymbol{x}|\boldsymbol{y_o}) < 1$ and $\ln(1) = 0$, the cost function for the 3DVAR case is given by Equation 2.23 (more details are shown in Lorenc (1986)).

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x_b})^T\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x_b}) + \frac{1}{2}[\boldsymbol{y_o}-\boldsymbol{h}(\boldsymbol{x})]^T\boldsymbol{R}^{-1}[\boldsymbol{y_o}-\boldsymbol{h}(\boldsymbol{x})] \quad (2.23)$$

In order to find the optimal solution that minimizes Equation 2.23 ($\boldsymbol{x_a}$) it is necessary to calculate $\nabla J(\boldsymbol{x}) = 0$ which gives us Equation 2.24.

$$\boldsymbol{x_a} = \boldsymbol{x_b} + (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})[\boldsymbol{y_o}-\boldsymbol{H}(\boldsymbol{x_b})] \quad (2.24)$$

However, the exact solution is not commonly used in the operational environment due to its high computational cost to solve it. Instead, iterative methods are used aiming to obtain an approximated solution as it is shown by Ming (2006).

2.3 Weather forecasting and data assimilation systems

The Weather Prediction System, used to simulate the meteorological conditions of the atmosphere, was the WRF model while the WRFDA was the Data Assimilation System responsible for inserting the observational information into the model and correcting the initial and boundary conditions for the simulation. Those systems were introduced briefly in the first chapter, in this section it is shown in more details the structure and the main components of them.

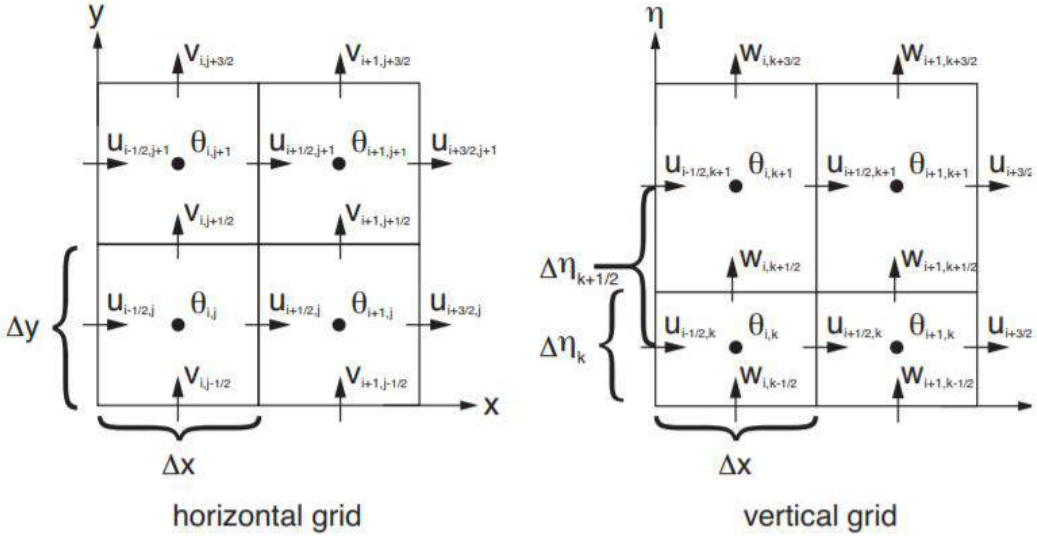
Inside the WRF infrastructure, there are what we call as dynamic solvers, these are the main components responsible for solving the physical equations of the model. The WRF model has two different dynamic solvers, the Nonhydrostatic

Mesoscale Model (NMM) or the Advanced Research WRF (ARW), the last one used in this study.

The complete set of physical equations that represents the atmospheric physics does not have an exact solution, this way, it is necessary to apply numerical methods in order to solve these equations which implies in the discretization of space and time. Some of these characteristics will be discussed below.

The discretization of the space requires a grid, in the case of the ARW the grid used was Arakawa-C (ARAKAWA; LAMB, 1977). In this type of grid, the wind components (u, v, w) and the thermodynamic variables (θ) are calculated as it is shown by Figure 2.3.

Figure 2.3 – Horizontal and vertical grids (Arakawa-C) used by ARW solver.



Source: Skamarock et al. (2008)

The Δy and Δx are constant and the correction for different projections is made by a map factor m.

Note that the zonal and meridional wind components u and v are calculated between the grid points, while the potential temperature θ is calculated exactly on the grid points. The same applies to the vertical grid where the vertical wind component w is also calculated between grid points.

Different types of grids affect directly the format of the equations as well as the propagation of perturbations in the model, these features and more details can be seen in Arakawa and Lamb (1977).

With the grid defined above, the ARW solver uses 2nd and 6th order advection options in horizontal and vertical spaces.

Another important aspect is the horizontal resolution, i.e., the space between each grid point. The resolution must be defined based on the characteristics of the atmosphere that we want to study. A high-resolution grid to study a global scale atmospheric system might spend unnecessary computational resources while a low-resolution grid to study a smaller scale system might not represent properly the physics involved. An example of the impact of the horizontal resolution can be seen in Figure 2.4.

The Figure 2.4 illustrates an ideal case where the observations (pluviometers) are spaced in different configurations aiming to represent the precipitation caused by the systems with length L. The expected spatial distribution of the precipitation is shown in “b”. Note that in “c” and “d” the observation instruments are spaced with $s = L$ and because of that the precipitation is not properly represented, the same happens when $s = \frac{3}{4}L$. When the pluviometers are spaced with $s = \frac{1}{2}L$ the spatial distribution of the precipitation is more realistic.

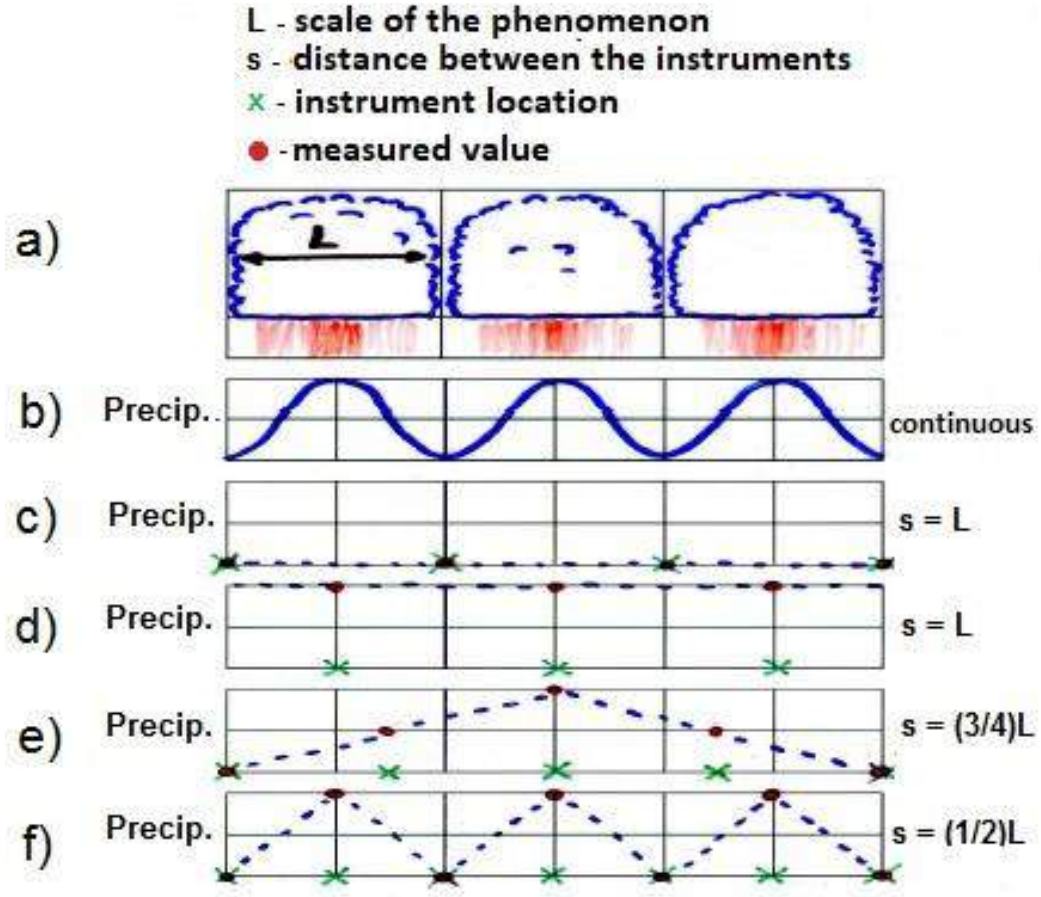
The same logic can be applied for the horizontal resolution of the model where it is necessary that the space between the grid points are at maximum half of the length of the system that we want to simulate, i.e., $L \geq 2\Delta s$. However, due to the relative differences between the meteorological systems and the grid points, the ideal for operational purposes is to use $L \geq 5\Delta s$.

Moreover, the ARW dynamic core equations are based on a vertical coordinate normalized by the hydrostatic-pressure in the surface defined by Equation 2.25 (LAPRISE, 1992).

$$\eta = \frac{p_h - p_{ht}}{p_{hs} - p_{ht}} \tag{2.25}$$

Where p_h is the hydrostatic pressure, p_{ht} is the hydrostatic pressure at the top and p_{hs} is the hydrostatic pressure at the surface.

Figure 2.4 – Schematic illustration showing the representation of the precipitation field for different distances between the observation stations.

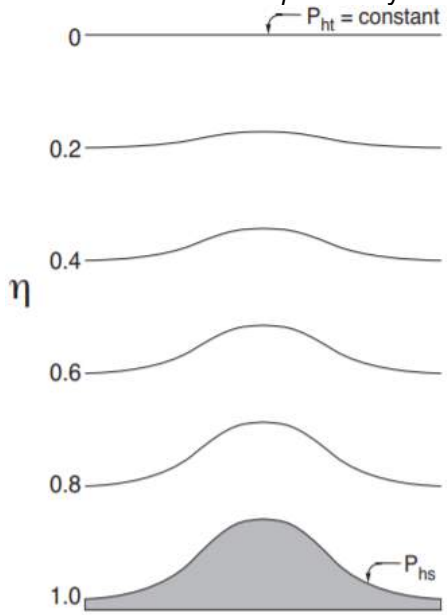


Source: Adapted from Gandu (2005).

This coordinate is usually referred to σ which is used in many hydrostatic atmospheric models. The coordinate varies between 1 at the surface and 0 at the top layer of the model (Figure 2.5).

The advantage of the use of η is due to the fact that it is a surface coordinate. However, errors in the pressure gradient are common in regions near to mountains with a high altitude gradient (PHILLIPS, 1957). More details about each type of vertical coordinates used in atmospheric models and the differences between each other can be found in AMS (2019).

Figure 2.5 – Vertical coordinate η used by ARW solver.



Source: Skamarock et al. (2008).

The differential equations used in the atmospheric sciences generally are composed by terms that describe a large range of oscillation modes, some of these modes are generated by the integration of these equations and does not have any meteorological meaning but the low-frequency modes are associated with atmospheric conditions (SKAMAROCK et al., 2008). Hence, the time integration in ARW is calculated differently for different modes. For high-frequency acoustic modes, the integration is made over smaller steps in relation to the low-frequency modes, where the third-order Runge-Kutta (RK3) is used allowing it to maintain the numerical stability of the system (SKAMAROCK et al., 2008).

The numerical stability is an important factor that needs to be considered in any simulation. The numerical schemes of the model must take into account criteria like consistency, convergence, and stability in order to solve the equations (CUNNINGHAM et al., 2001).

In nonlinear equations, which is the case of the models that use primitive equations (differential equations used to describe the atmospheric flow) like WRF, the occurrence of nonlinear instabilities is commonly observed. This type

of instability happens when different oscillation modes generated by the integration of the model equations interact with each other generating errors in the redistribution of energy in the simulation domain, it is usually observed in wavelengths with about 2 to $4\Delta x$ (KALNAY, 2003; WARNER, 2011).

The primitive equations also need to obey the criteria for convergence known as Courant–Friedrichs–Lewy or CFL (COURANT et al., 1967). The Equation 2.26 shows that the horizontal speed of the fastest wave in the model domain V_H (usually considered $300m/s$) multiplied by the time step Δt and divided by the horizontal resolution Δx must be less or equal to 1.

$$V_H \frac{\Delta t}{\Delta x} \leq 1 \tag{2.26}$$

Moreover, the physical equations of the model have a wide range of parameterizations which allows it to adapt to a specific region or period of the year. Basically, the parametrizations are simplified equations responsible for describing a specific physical process. These simplifications are used for two main reasons: (1) for when we do not have enough knowledge about some specific physical process which makes it necessary to develop equations according to known variables, or (2) for when we want to reduce the computational cost (WARNER, 2011).

Besides that, most of the parameterizations are developed empirically, i.e., they are developed based on the observational data and usually, their use is limited to a specific period of the year or to a specific region. Stensrud (2007) presents a good review of the parametrizations used in meteorological models in general.

In the ARW-WRF model, there are simplified parameterization schemes for idealized studies and also complex schemes associated with Microphysics, Cumulus parameterizations, Surface physics, Planetary Boundary Layer Physics and Atmospheric radiation physics. More details in Powers et al. (2017).

Figure 2.6 shows the main WRF model components, including the assimilation system WRFDA, and how they are connected between each other in order to produce a weather forecast.

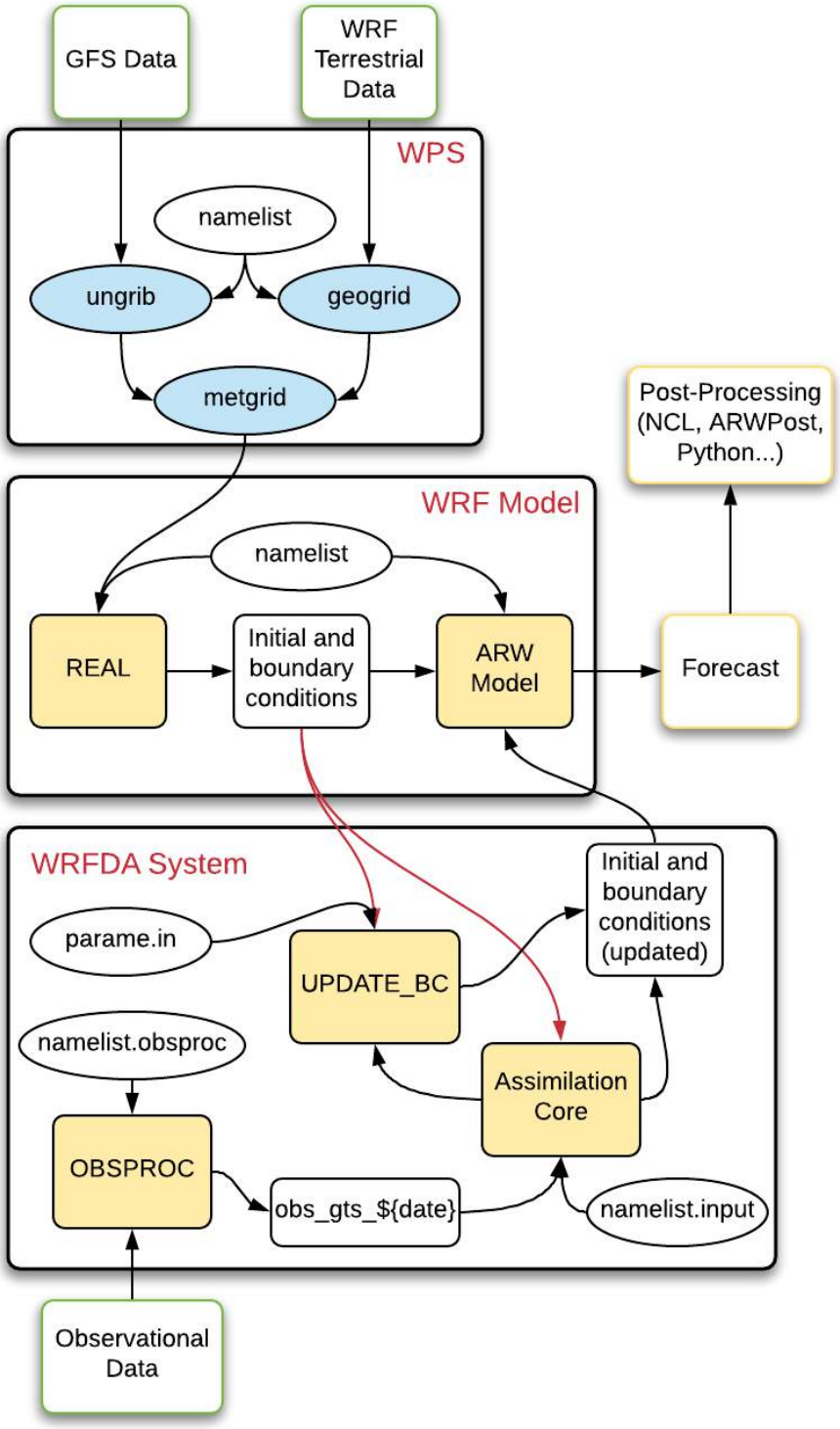
The first component of the system is the WRF Preprocessing System (WPS). This component is responsible for processing the necessary external data for the model. The geogrid program defines some of the domain configurations including the area and the interpolation of the terrestrial data to the study area, while the ungrib program read, unpack and transform the GRIB data (initial and boundary conditions from another model, in this case, the Global Forecast System (GFS)) into a simpler format. The metgrid program is responsible for interpolating horizontally the data from ungrib to the domain defined by geogrid. The configurations of this procedure (input data, domain, resolution, etc.) are defined by a namelist file.

The second component of the weather forecasting system includes the real program where the initial and boundary conditions are prepared for the ARW, which is the core of the whole system and where the set of equations are integrated in time generating a forecast. Posteriorly, the output is processed for operational or research purposes. Similar to the previous component, a namelist file defines some of the configurations in this component like the physical parameterizations used in the equations, for example.

The WRFDA system receives the initial and boundary conditions processed by the real program in the second component and the assimilation core uses an assimilation methodology, like 3DVAR, to correct the initial conditions based on the observations (previously processed by the obsproc code). The update_bc code corrects the boundary conditions based on the corrections made by the assimilation core and then the corrected initial and boundary conditions are used by the ARW which is expected to generate a better forecast. The files parame.in, namelist.obsproc and namelist.input are also configuration files.

Another component (optional) is the assimilation system. As it was discussed in the previous sections, the assimilation system is not strictly necessary, however, when used it can improve significantly the weather forecast.

Figure 2.6 – Diagram showing how all the components, including the WRF model and WRFDA system, used in this study interact with each other.



The diagram in Figure 2.6 presents the main components of the weather forecasting system used in this study, the complete description of all

components, including the mathematical formulation and the code of the whole weather forecasting system can be found in Skamarock et al. (2008).

2.4 Lightning and its application in data assimilation

In this section, we discuss the atmospheric processes responsible for the formation of lightning, its relationship with the meteorological variables and why the use of this information in assimilation methodologies can improve the model initialization.

There are many theories that explain the electrical structure of the clouds and most of them are based on how the particles (ice crystal, graupel, water droplets…) interact with the surrounding environment to form the electric charge centers. The convective (GRENET, 1947; VONNEGUT, 1963), the inductive (ELSTER; GEITEL, 1913), the non-inductive (REYNOLDS, 1957) and the quasi-liquid (BAKER; DASH, 1989; DASH, 1989; FLETCHER, 1968) are the most common ones.

The thunderstorm clouds, also known as Cumulonimbus, usually have great vertical development associated with intense vertical motion. The vertical motions push the humid air from the surface to higher altitudes condensing it as the air parcel reaches the Lift Condensation Level. If this process continues, the water vapor will eventually freeze forming ice particles.

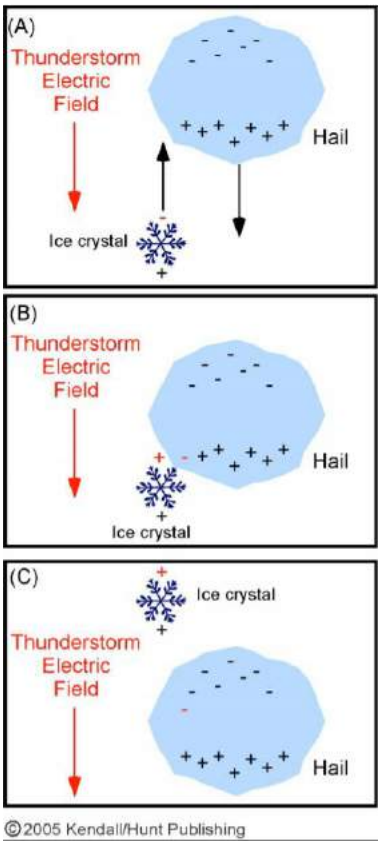
In the inductive process, it is considered that an external electric field is capable of to polarize the particles inside the cloud. The particles with larger size (graupel) fall while the small particles (ice crystals) are pushed to the top of the cloud by the vertical wind, during this process the top of the small particles (negatively charged due to the polarization by the electrical field) collides with the bottom of the falling particles (positively charged) as it is shown by Figure 2.7. The collision process between the particles of different sizes creates a positive charge center in the top of the cloud and a negative charge center in the bottom (Figure 2.8).

When the electric potential difference is high enough, an arc discharge may occur between clouds or inside of the same cloud (IC) or even between a cloud and the ground (CG).

However, the non-inductive process seems to be the most viable explanation for the observed electric structure of a storm cloud and it is generally thought to be the best way to explain the rapid electrification of thunderstorms. Unlike the inductive process, the non-inductive does not consider the existing of an electrical field (COORAY, 2003; WILLIAMS, 1988).

Basically, the graupel collides with ice crystals, in the presence of supercooled water droplets, and if the environment temperature is above $-15℃$, the small particles (ice crystals) become negatively charged while the graupel particles become positively charged. In colder temperatures, the electrification process is the opposite (Figure 2.9).

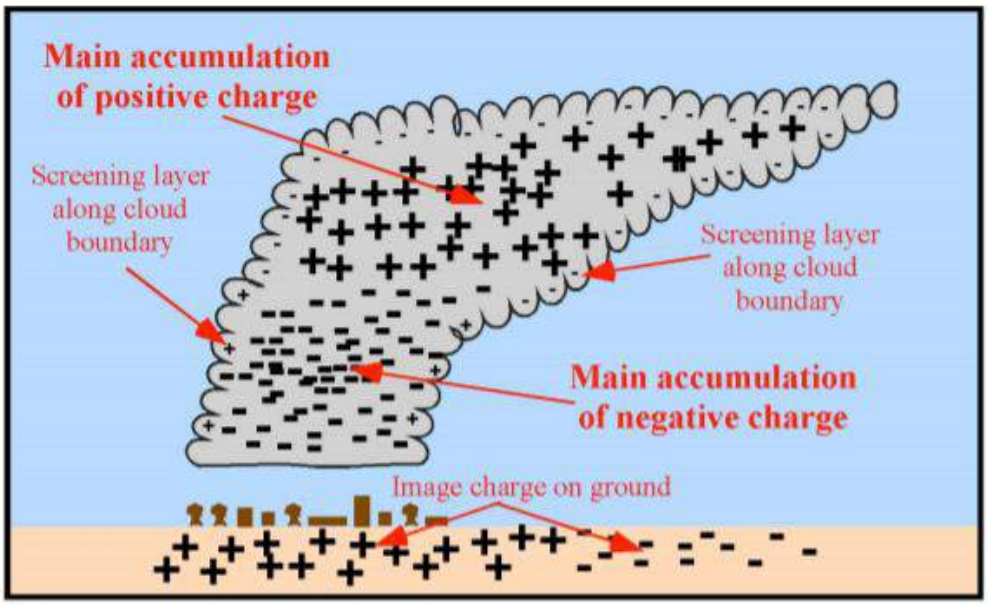
Figure 2.7 – Illustration showing the interaction between ice crystal and graupel (hail), and how these particles interact with each other under the influence of an external electric field.



This way, above the isotherm of $-15℃$ the heavier particles tend to accumulate in the bottom while the small particles are carried out to the top of the cloud. Below the isotherm of $-15℃$, the process is reversed creating a tripolar structure inside the cloud, with the top and the bottom positively charged and

the middle negatively charged. Figure 2.10 shows this tripolar structure for different regions and in different periods of the year.

Figure 2.8 – Illustration of the electric charge centers according to the inductive process.

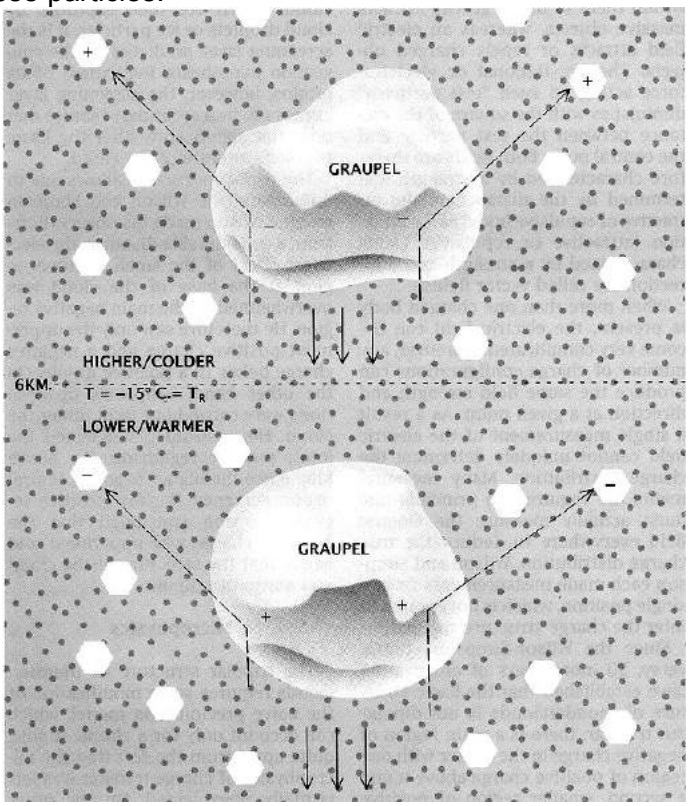


The lightning is composed of multiple electrical discharges typically separated by tens of milliseconds and each of these individual discharges is called return stroke. Commonly, the return strokes are grouped into what we call as a flash. The flashes can be IC or CG, as already mentioned before.

Moreover, the CG flash also can be classified as negative, when it is observed the transference of negative charges, or positive when the net electric charge transferred is positive. Also, they can be either descending, when the discharge leaves a cloud towards the ground or ascending when the discharge leaves the ground towards a cloud (Figure 2.11).

The whole process of initiation, formation, and dissipation of the charge centers in thunderclouds resulting in the occurrence of lightning is a complex process and we will not discuss in this study. However, more details about this process can be found in Campos (2016), Cooray (2003), Naccarato (2006), Uman (2001) and Williams (1988).

Figure 2.9 – Illustration showing the interaction between ice crystals and graupel in different temperatures, and how the charges are transferred between these particles.

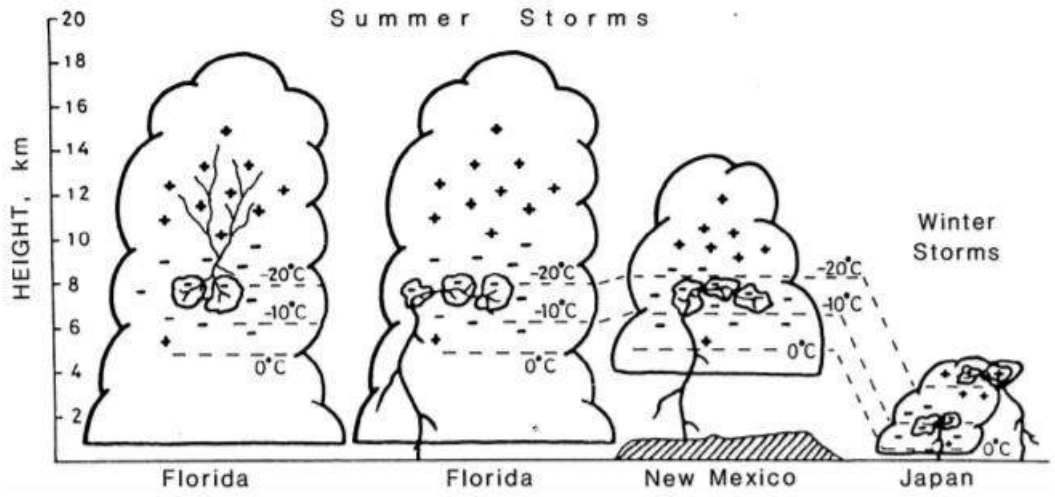


Source: Williams (1988).

Based on what we discussed until now, it is clear now that lightning is related to the microphysics processes inside the thunderclouds and to atmospheric instability and thermodynamics. All these physical processes corroborate the fact that lightning data can provide different information about the atmosphere when compared to conventional data sources commonly used in meteorology. Moreover, the detection systems responsible for detecting lightning flashes also present several advantages.

The BrasilDAT is the main lightning detection system in Brazil. This system operates in the range of Low and Very Low Frequencies (LF/VLF) using the Time-of-Arrival (TOA) algorithm to triangulate the position where the electromagnetic pulse was detected by the sensors (CUMMINS; MURPHY; TUEL, 2000).

Figure 2.10 – Illustration showing the charge centers and how the temperature affects their distribution inside the cloud.



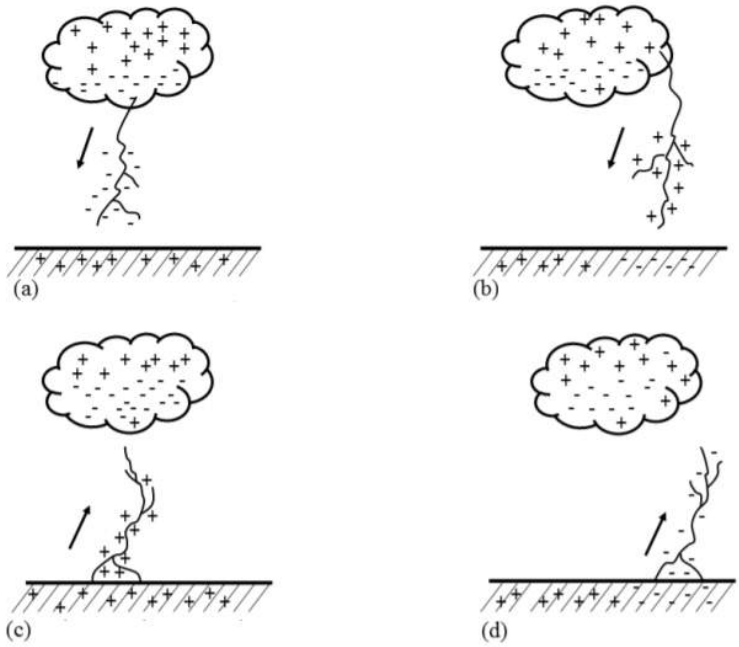
Source: Krehbiel (1981).

The location of the BrasilDAT sensors currently in operation can be seen in Figure 2.12. The study area, as well as the sensors inside it, can be seen in the next chapter in more details (Figure 3.2).

This detection system is able to determine both the polarity (positive or negative) of the flash detected and if it is an IC or a CG. This system is also able to measure the electric current intensity produced by the flashes, this is made analyzing the signal and the amplitude of the first wave peak. More details about the operation of this system can be found in Naccarato and Pinto Junior (2012).

The BrasilDAT network has an average detection efficiency of 85 to 90% for CG, while for IC this detection efficiency is between 50 to 60% (NACCARATO; PINTO JUNIOR, 2012). However, this can varies depending on the region and on the number of sensors (Figure 2.12 and 2.13).

Figure 2.11 – Illustration showing the different types of CG flashes. (a) Negative downward flash; (b) Positive downward flash; (c) Negative upward flash; and (d) Positive upward flash.



Source: Rakov and Uman (2003).

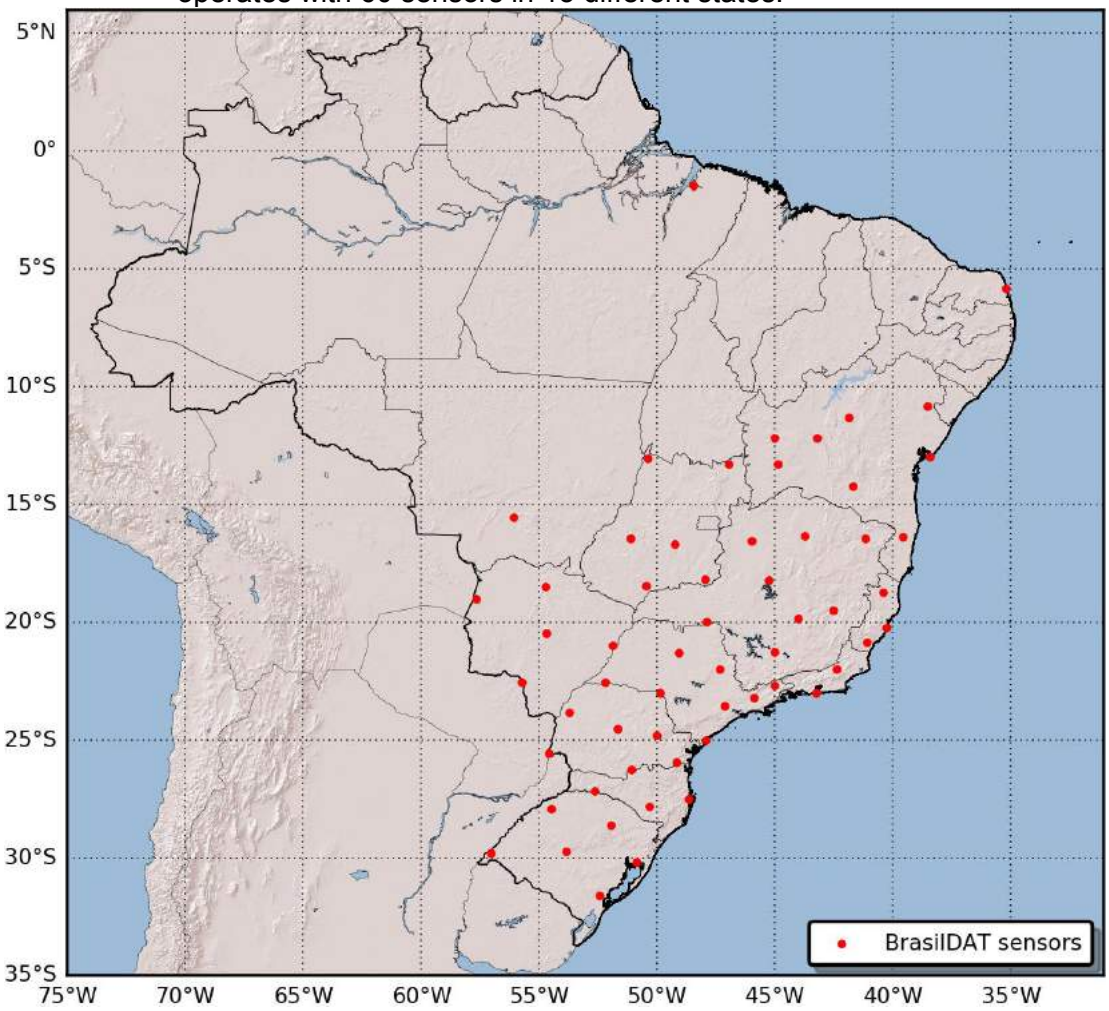
The differences in the detection efficiency of IC and CG are mainly due to the electric current and the direction of propagation associated with each them. IC flashes tend to have a smaller electric current when compared to CG flashes since the resistance necessary for the occurrence of one discharge inside the cloud is smaller than between a cloud and the ground. Also, part of the IC flashes occurs horizontally while the sensors are capable of detecting only the vertical component of the electric field.

The IC and CG flashes are associated with different processes inside the clouds, for example, IC flashes are correlated to updraft strength within deep continental storms (FIERRO et al., 2012; MACGORMAN; RUST, 1998; SCHULTZ; PETERSEN; CAREY, 2011) while the CG flashes are more correlated to downdraft strength and rain (CAREY; RUTLEDGE, 1996).

The spatial distribution of flash, as well as the area selected for this study, located in southeastern Brazil (South America), is presented in Figure 2.14. It is possible to notice that the study region has a relatively high incidence of

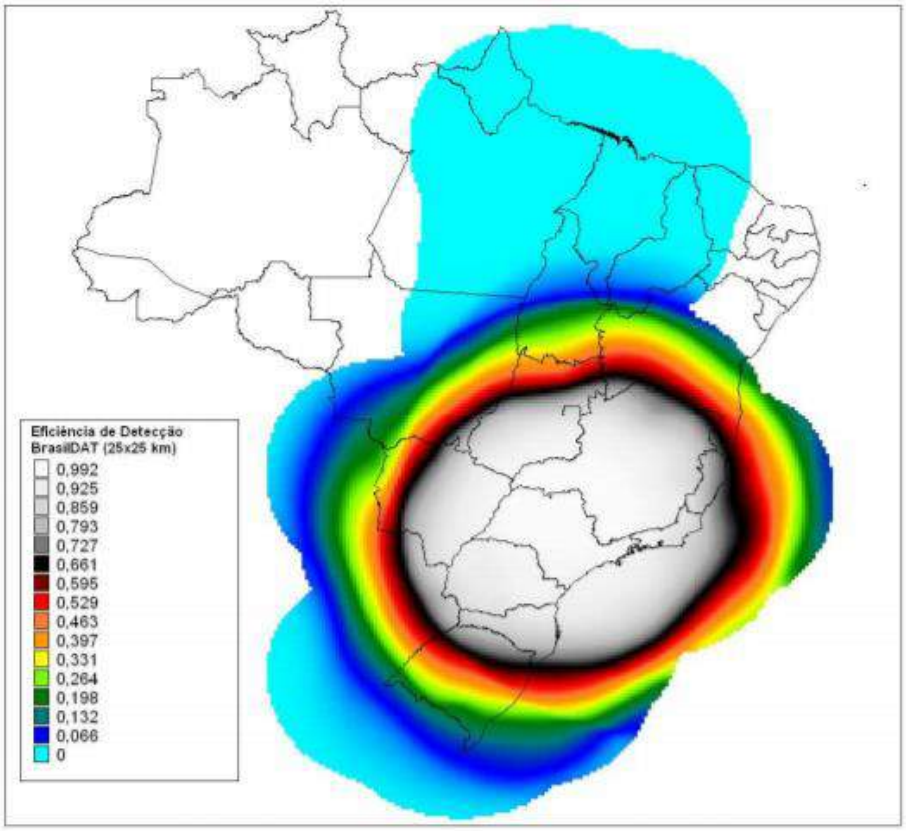
lightning which makes it a good region to evaluate the impact of lightning data in the WRF model.

Figure 2.12 – Spatial distribution of the BrasilDAT sensors. Currently, the network operates with 60 sensors in 13 different states.



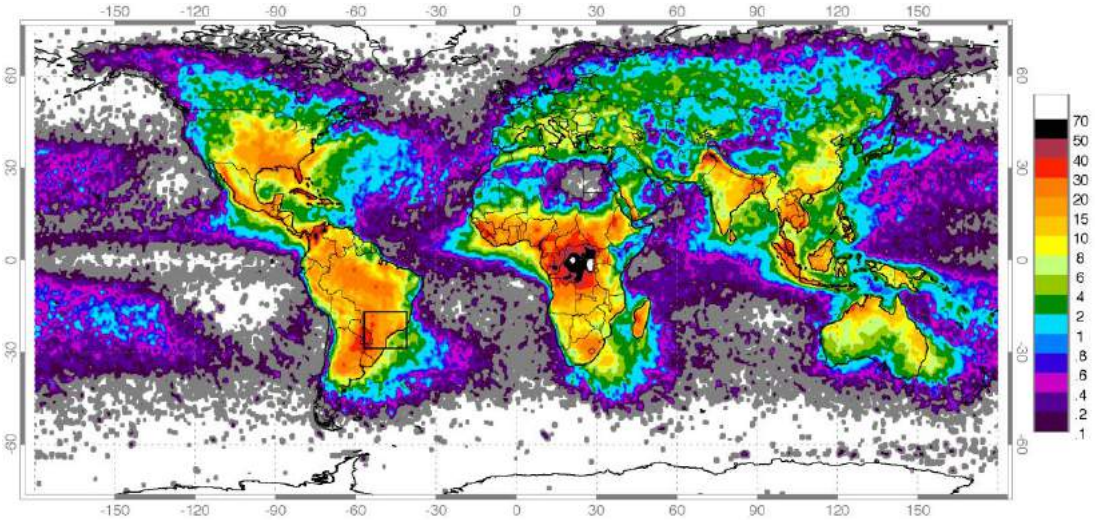
In more detail, Figure 2.15 shows the spatial distribution of CG flashes in Brazil. Note that the incidence of flashes is higher in the west portion of the study area, especially in the southwest. These characteristics are due to a wide range of meteorological systems that affect the region interacting with different terrain features.

Figure 2.13 – Lightning detection efficiency map for BrasilDAT in 2005.



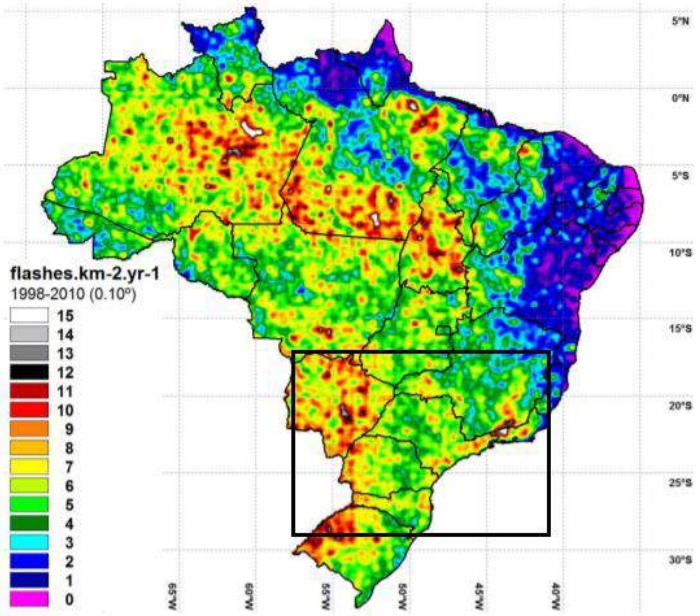
Source: Naccarato and Bourscheidt (2011).

Figure 2.14 – Global flash density ($flash/km^2/yr$) detected by the Lightning Imaging Sensor (LIS) from space-based optical sensors between 1995 and 2003. The black square over South America indicates the study area used in this work.



Source: Adapted from NASA (2019).

Figure 2.15 – Spatial distribution of CG flashes during the period of 1998 to 2010. The black square indicates the study area.



Source: Adapted from Naccarato et al. (2011).

2.5 Lightning research review

This section presents some of the main studies developed in the lightning data assimilation field showing the impact that this procedure has in the weather forecasts and how the different techniques use this information in order to obtain better results. The following works support the research developed in this present study showing that, in fact, the use of lightning data can improve the model simulations, especially when nudging techniques are used.

Alexander et al. (1999) developed one of the first studies aiming to assimilate total lightning data, i.e., without separating IC and CG flashes, into a meteorological model in order to improve the initial conditions.

In this study was used different data sources (satellite and lightning) in order to elaborate more precisely the initial conditions for a proper application of the assimilation procedure. Data from infrared and microwaves sensors with 3 and 12 hours of time resolution, respectively, and from continuous observation of lightning were used.

As mentioned before, each type of data source has its own limitations. For instance, the infrared sensors present a higher temporal resolution than

microwave sensors but they are affected by cirrus clouds. In the case of lightning data, there is a high temporal resolution and there is a good correlation with precipitation rates, although this relationship depends on the region (MACGORMAN; RUST, 1998).

This way, Alexander et al. (1999) proceeded with simulations using the Mesoscale Model (MM5) over North America and assimilating latent heat derived from precipitation rate estimates obtained from a variety of combinations between different data sources (cited above) in order to analyze the occurrence of an extratropical cyclone in the Mexican Gulf region. It was observed that the combination of all data sources (infrared, microwave and lightning) was able to reproduce better the precipitation rates and, consequently, the experiment that assimilated these data was able to represent more precisely the extratropical cyclone in the study region.

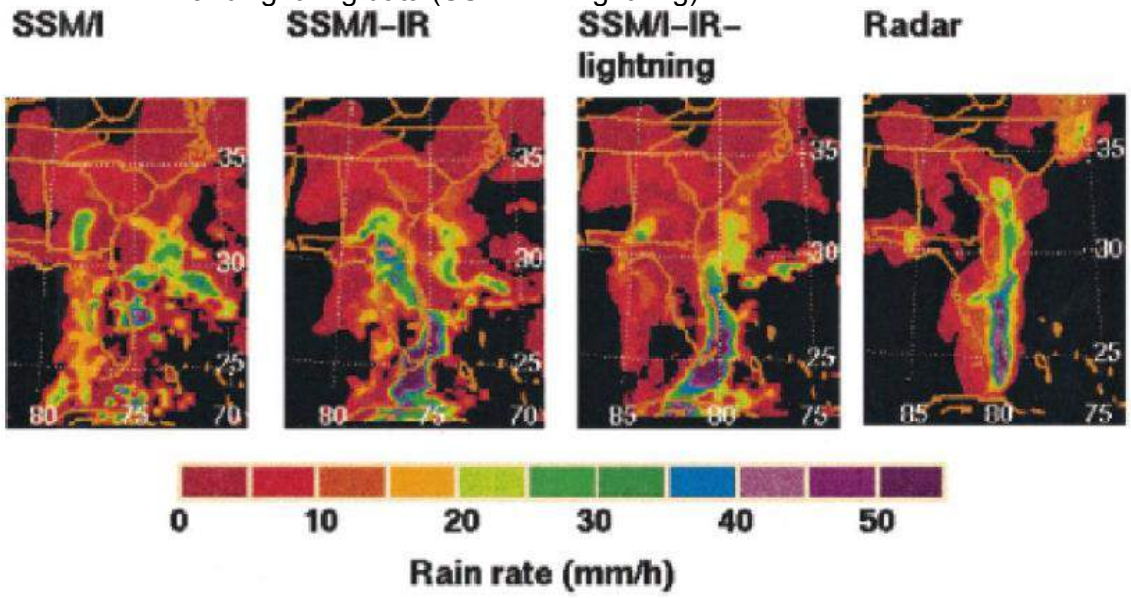
The impact of the different experiments with data assimilation proceeded in that study can be observed in Figure 2.16 where the precipitation rates observed and simulated are shown.

In a similar case, Chang et al. (2001) also analyzed a cyclogenesis that occurred in the Mexican Gulf, far away from the meteorological radars coverage, from the lightning detection network and from the conventional observation systems. Instead Alexander et al. (1999), this study had data from a long-range lightning detection system (Spheris Timing and Ranging Network 1 – STARNET-1) and from LIS aboard of a satellite, both capable of to monitor lightning activity and, consequently, the convective activity in the study region. Besides these data, this study used information about the sea surface temperature, additional microwave radiometer data, and data from the first meteorological radar aboard a satellite. Thus, the work developed by Chang et al. (2001) had a larger amount of data than the work developed by Alexander et al. (1999).

Chang et al. (2001) proceeded with several experiments with the MM5 model applying an algorithm developed in that work and assimilating data from

different data sources. Figure 2.17 shows the results from some of the experiments performed in that study.

Figure 2.16 – Precipitation rate at 0900 UTC on March 13th, 1993 obtained from the meteorological radar WSR-57 and from the simulations with the assimilation of microwave from the Special Sensor Microwave/Imager (SSM/I), microwave and infrared (SSM/I-IR) and microwave, infrared and lightning data (SSM/I-IR-Lightning).



Source: Alexander et al. (1999).

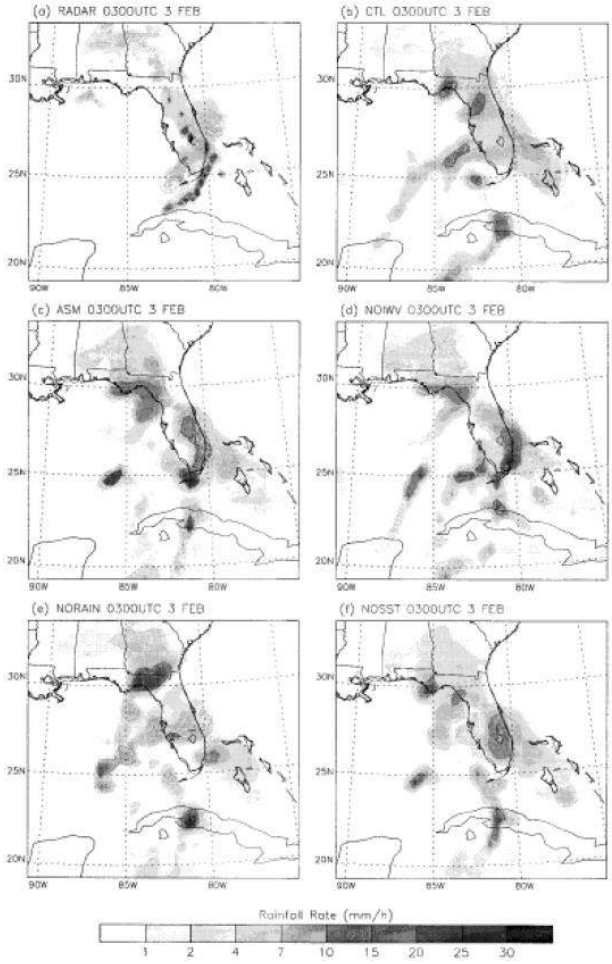
In a subjective analyze, it is possible to notice that the ASM experiment in which includes the assimilation of precipitation (variable derived from a combination of satellite and lightning data) obtained the most accurate results when compared with the observational data. Thus, it was shown that the continuous observation of lightning, through the use of satellite and long-range systems can be applied in the assimilation procedures to improve weather forecasting.

The following studies have in common the modification of specific physical parameterizations where the information of lightning detection was incorporated in the model through a nudging function (HOKE; ANTHES, 1976; KISTLER, 1974).

Using an assimilation technique similar to what was developed by Rogers et al. (2000) to assimilate radar data, Mansell et al. (2007) modified the Kain-Fritsch convective parameterization scheme (KAIN; FRITSCH, 1993) to allow lightning

data to control the activation of this scheme inside the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS). This study verified that lightning assimilation was well succeeded in the representation of cold pools (isolated cores with lower temperature when compared to the surrounding environment). These cold pools were presented on the surface during the model initialization. Besides that, it was concluded that the forecast was significantly improved when compared with the control experiment, especially in the first hours of simulation.

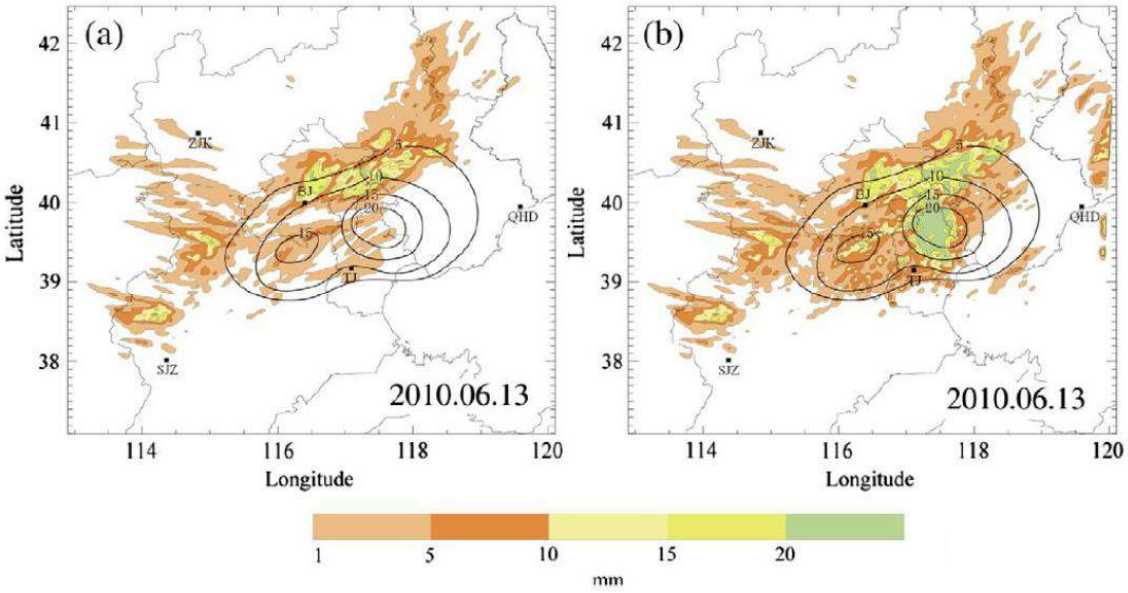
Figure 2.17 – Spatial distribution of precipitation rate at 0300 UTC on February 3rd, 1998 for (a) NWS radar observation and forecasts for the following experiments: control (CTL – b); with precipitation assimilation, integrated water vapor in the atmosphere and sea surface temperature (ASM – c); the same of (c) but without the assimilation of water vapor (NOIWV – d); the same of (c) but without the assimilation of precipitation (NORAIN – e); and the same of (c) but without the assimilation of sea surface temperature (NOSST – f).



Source: Chang et al. (2001).

Qie et al. (2014) performed simulations with the WRF model with the domain located in north China in order to simulate a Mesoscale Convective System. The study applied lightning data assimilation through a method that uses a nudging function similar to what was developed by Fierro et al. (2012), in which the values of the microphysics scheme are adjusted in the mixed-phase region of the cloud. Figure 2.18 shows the differences between the simulations and what was observed.

Figure 2.18 – Observed precipitation accumulated in six hours and predicted starting at 2000 UTC on June 13th, 2010: Experiment without lightning data assimilation (a) and with assimilation (b). The black contours represent the data from the surface stations and the shaded area indicates the simulated precipitation.



Source: Qie et al. (2014).

Note that there is a clear improvement in the experiment with the assimilation of lightning data when compared to the experiment without assimilation in both location and module of the accumulated precipitation in that period. Also, the study concluded that the results have shown improvements in the representation of the regions with convection, mainly in the areas with a high incidence of lightning.

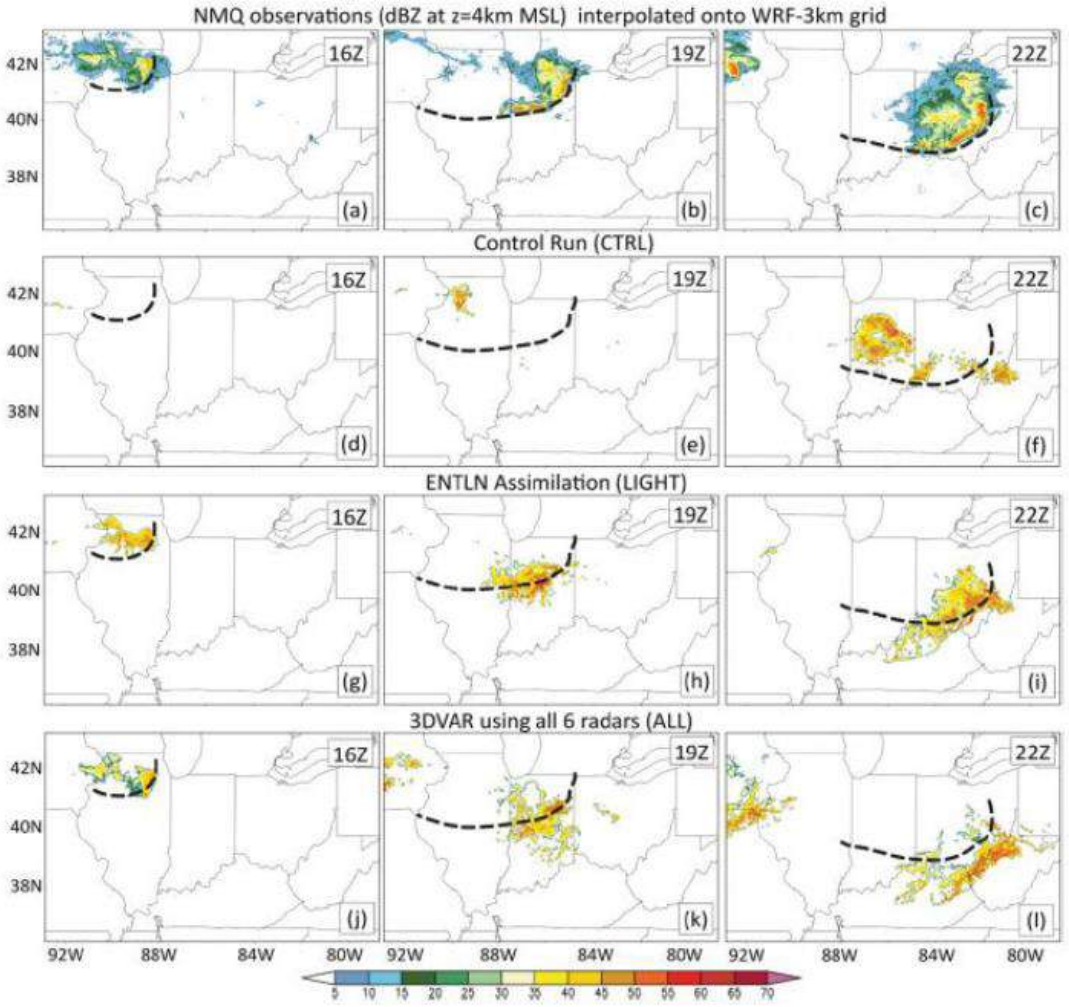
In the case of Fierro et al. (2014), the focus of the work was to compare two different assimilation methodologies that aims to improve the representation of

the convection during the analysis time and in the short-term forecast (6h) for a Mesoscale Convective System and for a derecho event (which is equivalent to a squall line). The first assimilation methodology was elaborated by Fierro et al. (2012), while the second is the same used in the 3DVAR in the Advanced Regional Prediction System (GAO et al., 2013). Then, a nudging assimilation technique (lightning assimilation) and a 3DVAR (with radar assimilation) were applied. As expected, it was observed that the 3DVAR assimilation technique had a better performance in the representation during the analysis time. However, using the nudging technique, it was observed better results in the short-term forecast. These features can be observed analyzing Figure 2.19, which shows the observed and simulated reflectivity field.

Meanwhile, Dixon et al. (2016), using lightning data from the World Wide Lightning Location Network (WWLLN) and a lightning data assimilation technique based on moisture nudging, simulated two different cases in the Midwest and eastern of the United States using high resolution (3km) deterministic and ensemble forecasts. The work concluded that, for the deterministic simulations, the nudging technique increased the integrated water vapor and Convective Available Potential Energy (CAPE) improving the forecast while the control experiments failed to reproduce the initiation and organization of the convective system. In the ensemble experiments, the use of lightning also improved the representation of the convection improving the forecasts in general, although, the impact was smaller when compared to the deterministic experiments.

More recently, combining the equations developed by Qie et al. (2014) and Fierro et al. (2012), Chen et al. (2019) developed a lightning data assimilation scheme (C18) where the parameters of the set of equations were defined based on the Bulk Richardson Number (BRN). The BRN is related to the occurrence and evolution of the convective systems (WEISMAN; KLEMP, 1982) and using that it is possible to correct the coefficients of the equations according to the level of instability of the atmosphere.

Figure 2.19 – Reflectivity field (dBZ) in 4km predicted and observed on June 29th, 2012. (a) reflectivity observed from National Mosaic and Multi-sensor QPE interpolated to the WRF model grid; (d) – (f) results from the CTRL experiment; (g) – (i) results from the LIGHT experiment; and (j) – (l) results from the ALL experiment.



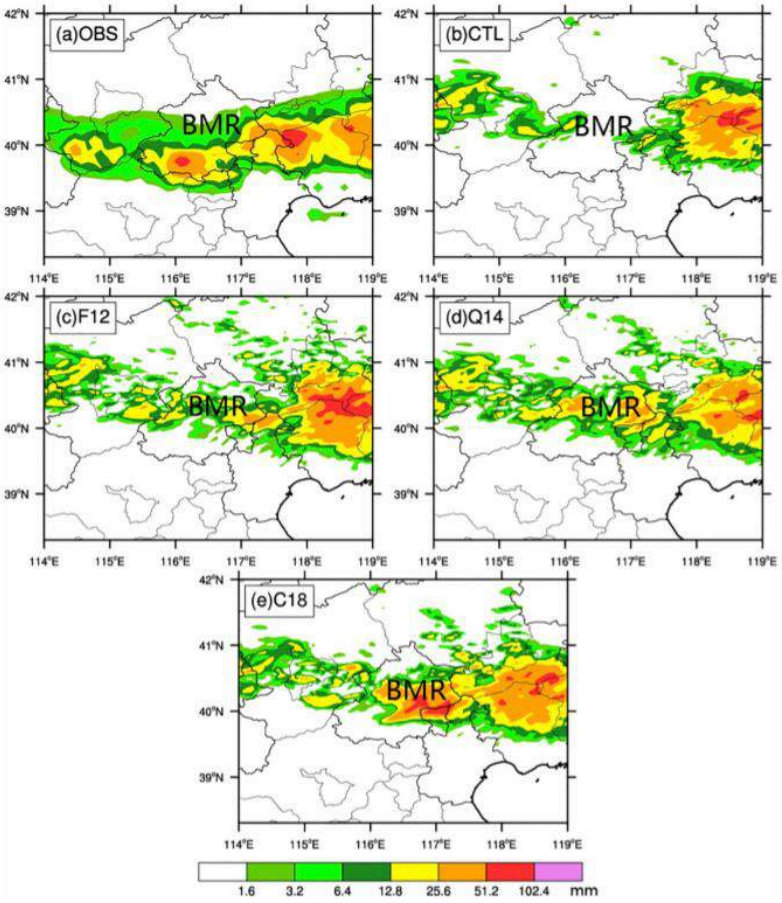
Source: Fierro et al. (2014).

The assimilation scheme was implemented using the WRF model configured with two nested domains with 6 and 2km of horizontal resolution located in northeast China, in the region of Beijing, with the lightning data from Beijing Lightning Network (BLNET). The study applied a lightning data assimilation scheme aiming to improve the simulation of two squall line cases that produced heavy precipitation in the study area. Figure 2.20 shows the comparison between different experiments and the observation for 3 hours accumulated precipitation field for one of the squall line cases.

It is visible the difference between the control experiment and the experiment with lightning data assimilation. The performance of the assimilation experiments was far more superior than the control reducing the spin-up period, inducing convection at the beginning of the simulations and improving significantly the short-term weather forecasts. Similar results were observed in both cases analyzed.

The algorithm developed by Chen et al. (2019) to assimilate lightning data proved to be better in representing the precipitation field than the other assimilation algorithms in the first hours of simulation (4 hours) and the study concluded that C18 seems to be more appropriate to simulate squall lines in the study area.

Figure 2.20 – 3 hours accumulated precipitation field from 2100 to 2400 LST on July 27th, 2015 (mm). (a) Observed; (b) Control experiment; (c) Fierro algorithm; (d) Qie algorithm; and (e) New algorithm implemented.



Source: Chen et al. (2019).

Other studies that applied different variations of nudging techniques to assimilate lightning data in NWP models also presented promising results (FIERRO et al., 2015; LAGOUVARDOS et al., 2013; LYNN; KELMAN; ELLROD, 2015; PAPADOPOULOS; CHRONIS; ANAGNOSTOU, 2005; PESSI; BUSINGER, 2009; WANG et al., 2018).

Besides the studies cited above, there are studies like the one developed by Stefanescu et al. (2012) where a different assimilation methodology was implemented to incorporate lightning data into the WRF model.

In that study, lightning data was used from the Earth Networks Total Lightning Network (ENTLN) during the simulation of two severe weather events with the occurrence of tornadoes in Alabama state in the United States. The method developed was implemented using a non-linear observation operator in the WRF – 3DVAR based on the CAPE variable as a proxy field for lightning. One of the main conclusions of this study showed that this assimilation scheme using lightning data improved the short-term forecast of the temperature profile.

Meanwhile, Stefanescu et al. (2013) proceeded with similar experiments and included one more that modified the WRF – 4DVAR observation operator. They also imposed certain conditions for the model to develop physical characteristics according with the observed flash rate. It was noticed that the application of that data assimilation method refined the precipitation statistics inside the assimilation window and for the period of 3 to 7 hours immediately after.

Apodaca et al. (2014) implemented an observation operator for lightning flash rate into the Gridpoint Statistical Interpolation (GSI) system by following a hybrid variational-ensemble approach. This assimilation method uses an observation operator based on the vertical updraft regression. The operator uses the maximum vertical velocity variable, which is related to the standard model variables through the continuity equation and, at the same time, has a relationship with lightning occurrence (PRICE; RIND, 1992). Therefore, it is possible to use lightning flash rate to update the initial model state modifying directly the standard model variables through the observation operator.

Even though it was not observed significative differences in the CAPE values (possibly because there were no observations of lightning in the region with high atmospheric instability), it was observed that the assimilation improved the analysis and the short-term forecasting.

Despite the development of many different techniques to assimilate lightning, especially in the last years, it seems that most of the advances in this research field are coming from the nudging methodologies. Even though nudging equations are not exactly based on physics, since additional terms are inserted in the equations in order to correct the imprecision of the physical equations used in the models, due to the results obtained in many studies most of the scientific community is focusing on the development of better nudging methodologies or on the improvement of the representation of lightning using proxy fields. The relationships between lightning and proxy fields are a challenge especially when those proxies are based on graupel and water vapor (FIERRO; MANSELL, 2018).

Although most of the studies applying nudging techniques have shown good results in the improvement of the convection and other aspects, especially in the first hours of simulation reducing the spin-up period, there are still many issues that need to be solved in order to improve this assimilation methodology. For instance, forcing the model to develop convection due to the insertion of moisture or through the modification of another proxy field may help the model to develop convection, however, in some cases, the convection persists way too long unrealistically delaying the dissipation of the system or inducing the formation of spurious convection.

So, a proper assimilation technique has to induce the formation of convection just enough in the first hours in order to reduce the spin-up period and avoid the excess of atmospheric instability that could lead the convection systems to persist in the late hours of simulation.

The main studies in the Lightning Data Assimilation research field are summarized in Table 2.1.

Table 2.1 – Characteristics of the main studies in the Lightning Data Assimilation research field.

ID	Year	Study Area	Data Source	Model	Assimilation Technique
1	1999	Mexican Gulf	NLDN	MM5	Nudging (precipitation rate)
2	2001	Mexican Gulf	STARNET-1; LIS	MM5	Nudging (precipitation rate)
3	2005	European continent and north Africa	ZEUS	SKIRON/Eta	Nudging (humidity)
4	2007	Midwest USA	NLDN; LMA	COAMPS	Nudging (convective scheme)
5	2009	North Pacific Ocean	LIS; OTD	MM5	Nudging (convective scheme)
6	2012	Midwest USA	ENTLN	WRF	Nudging (microphysics scheme)
7	2012	Alabama	WWLLN	WRF	Observation operator (CAPE)
8	2013	Southern France	LINET	MM5	Nudging (convective scheme)
9	2013	USA	ENTLN	WRF	Observation operator (CAPE)
10	2014	North China	SAFIR	WRF	Nudging (microphysics scheme)
11	2014	Northeast USA	ENTLN	WRF	Nudging (microphysics scheme)
12	2014	East USA	WWLLN	WRF	Observation operator (ω_{max})
13	2015	USA	ENTLN	WRF	Nudging (humidity)
14	2015	East USA	ENTLN; USPLN	WRF	Nudging (humidity)
15	2016	Midwest and eastern of USA	WWLLN	WRF	Nudging (humidity)
16	2018	East China	SAFIR	WRF	Nudging (humidity)
17	2019	Northern China	BLNET	WRF	Nudging (microphysics scheme)

The correspondent study associated with the ID of Tab. 2.1 can be seen below:

[1] Alexander et al. (1999); [2] Chang et al. (2001); [3] Papadopoulos, Chronis and Anagnostou (2005); [4] Mansell et al. (2007); [5] Pessi and Businger (2009); [6] Fierro et al. (2012); [7] Stefanescu et al. (2012); [8] Lagouvardos et al. (2013); [9] Stefanescu et al. (2013); [10] Qie et al. (2014); [11] Fierro et al. (2014); [12] Apodaca et al. (2014); [13] Fierro et al. (2015); [14] Lynn, Kelman and Ellrod (2015); [15] Dixon et al. (2016); [16] Wang et al. (2018); [17] Chen et al. (2019).

3 METHODOLOGY

In this present section the structure of the whole assimilation process is discussed including the algorithm developed for this study, the data used and their respective sources, the configuration of the model, the experiments and the evaluation methods, aiming to proceed with an objective analysis in order to define the advantages and disadvantages of the implementation of this technique operationally.

3.1 Data and study area

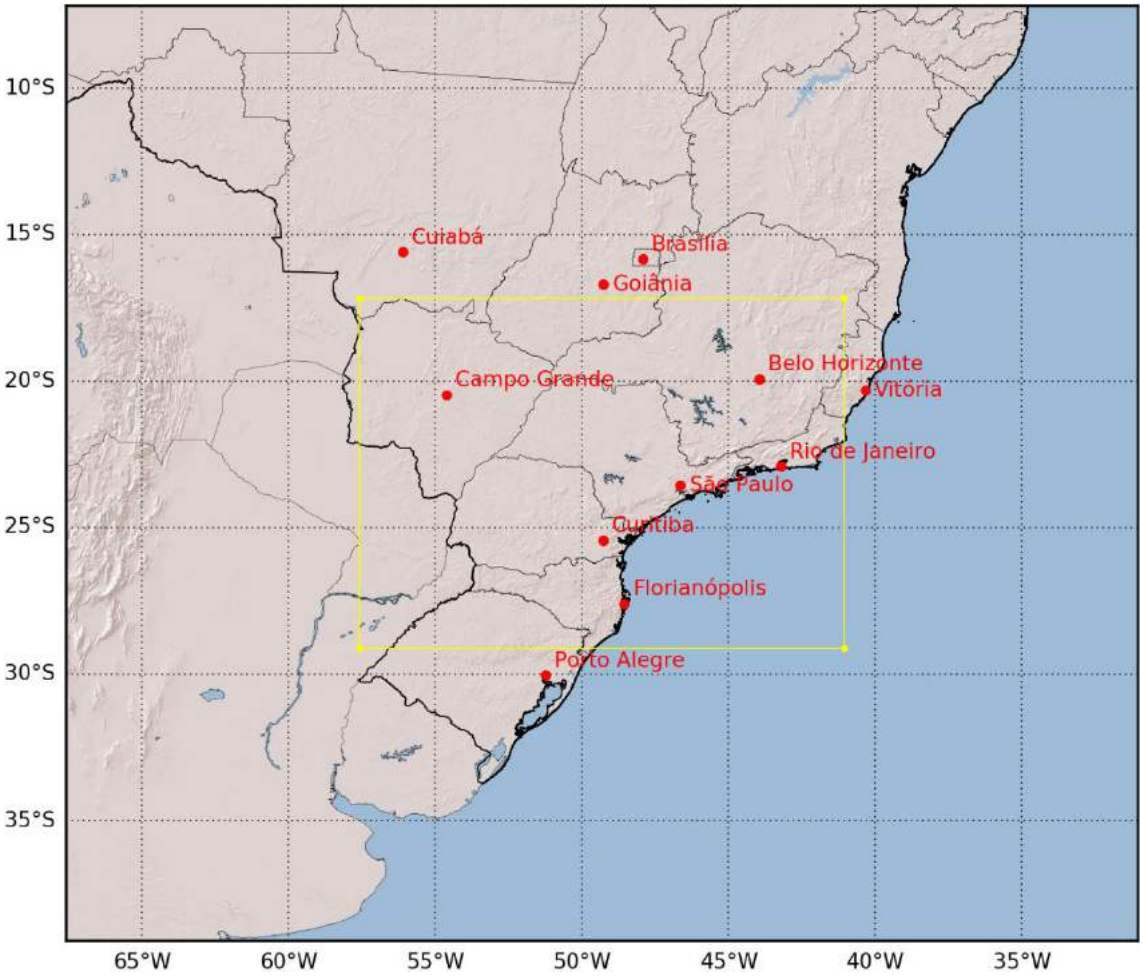
For a better evaluation and implementation of the assimilation process, many different data sources were used in this work. It was analyzed two different days where the occurrence of meteorological systems with a high incidence of lightning was verified. More details of the system and the experiments developed are shown in section 3.5.

The area selected for this work was set in South America specifically over the southern portion of Brazil as it is shown by Figure 3.1. This area is well covered by many types of observation stations and at the same time it has favorable conditions for the occurrence of several meteorological systems which implies in the occurrence of many storms with a high incidence of lightning.

The precipitation data used in this study to evaluate the simulations were obtained from about 600 meteorological surface stations (the number of stations depends on the hour analyzed) from the National Institute of Meteorology (INMET). The data has hourly resolution and the spatial distribution of those stations inside the simulation domain can be seen in Figure 3.2.

The green square in Figure 3.2 shows the area considered in the evaluation methods (read more about it in section 3.6). This area was selected with the intention to avoid errors associated with the borders either because of instabilities in the simulation or because of the interpolation of observational data in regions with no data.

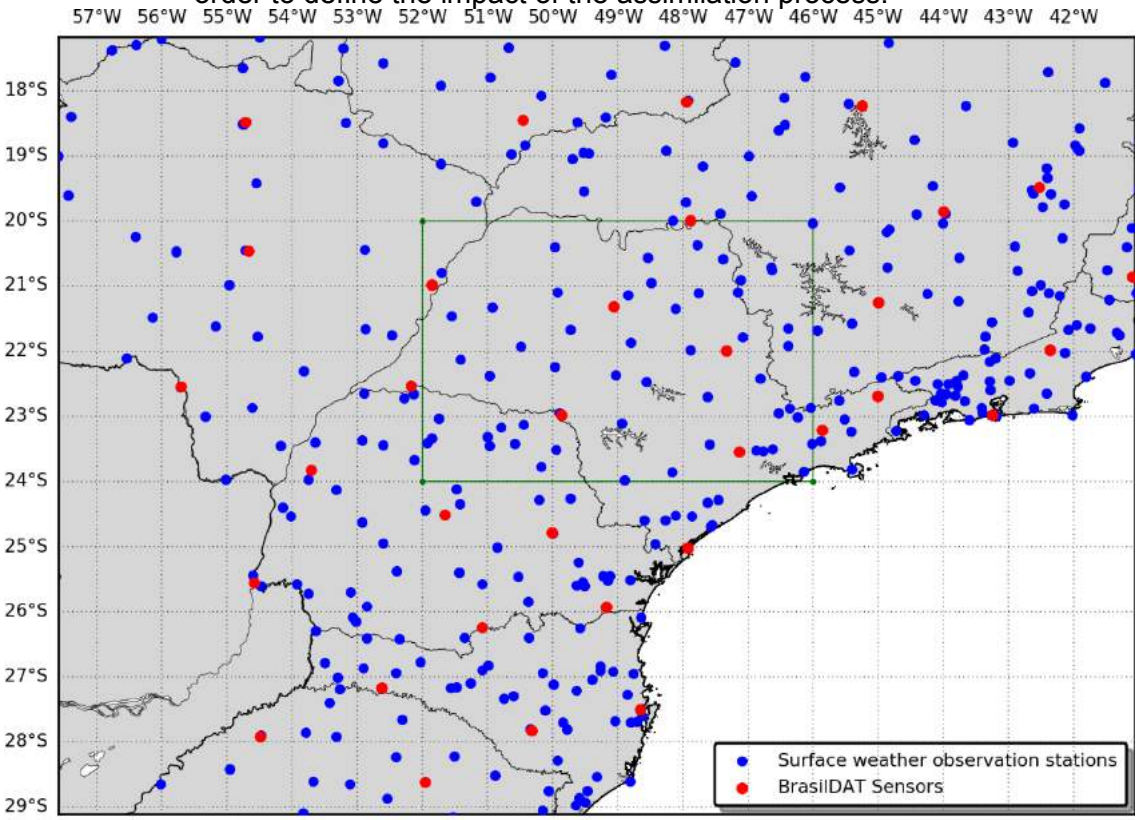
Figure 3.1 - Study area delimited by the yellow square used in the simulations with the WRF model. The domain has its center over Brazil covering a big portion of the southeast, south, and center-west of this country (the main cities are shown in red).



Lightning data was the main data source used in this study. It was obtained from BrasilDAT provided by the Atmospheric Electricity Group of the National Institute for Space Research (ELAT/INPE). Lightning data was necessary to calculate the flash rate density that provided the necessary information to insert inside the assimilation algorithm responsible for correction the water vapor mixing ratio variable in order to obtain a better initial condition. It was also important to analyze and evaluate the behavior of some variables simulated by the WRF model.

Satellite images from GOES-16 were considered in order to track the meteorological systems affecting the study area. It was possible to identify the type of meteorological system as well as the most affected regions.

Figure 3.2 - Spatial distribution of the observational data inside the domain. The green square shows the region where the evaluation methods were applied in order to define the impact of the assimilation process.



Moreover, synoptic weather charts from the Center for Weather Forecasting and Climate Studies (CPTEC) of INPE were used for proper identification of the meteorological conditions associated with each meteorological system analyzed in this study.

Finally, for the purpose of initialization of the regional model, it was also used data from the GFS (with 0.25° of spatial resolution) provided by the Computational and Information Systems Laboratory from University Corporation for Atmospheric Research (CISL/UCAR, 2017).

3.2 Model configuration

This study used the WRF-ARW model version 3.9.1.1 which was released on August 28th, 2017 (MMM/UCAR, 2018). The main parameterizations set for the simulations were: Yonsei University Scheme for Planetary Boundary Layer (HONG; NOH; DUDHIA, 2006), Thompson Scheme for Microphysics (THOMPSON et al., 2008), Unified Noah Land Surface Model for Land Surface

(TEWARI et al., 2004) and RRTMG for Shortwave and Longwave schemes (IACONO et al., 2008). Also, the Cumulus parameterization was deactivated which allowed the model to solve the convection explicitly.

The simulations were configured with only one domain with a total of 149x189 grid points located between the latitudes -29.1184°S and -17.1805°S and between the longitudes -57.5814°W and -41.0482°W. The longitudinal distance between the grid points varies between -0.087° and -0.088° and between -0.077° and -0.084° latitudinally which gives a horizontal resolution of approximately 9km (Figure 3.3 shows the domain in more details). Vertically, 50 vertical levels with the top in 50hPa and an adaptive timestep were used.

This domain was selected considering the spatial distribution of the BrasilDAT sensors and meteorological observation stations (Figure 3.2) as well as the meteorological systems that commonly affect this region. The region selected is dynamically active the whole year showing a wide variety of meteorological systems like Fronts, Mesoscale Convective Systems, Low-Level Jets, Convergence Moisture Zones, Cyclones, and other systems. All that gives us the opportunity to analyze different physical configurations of the atmosphere with good coverage of observational data.

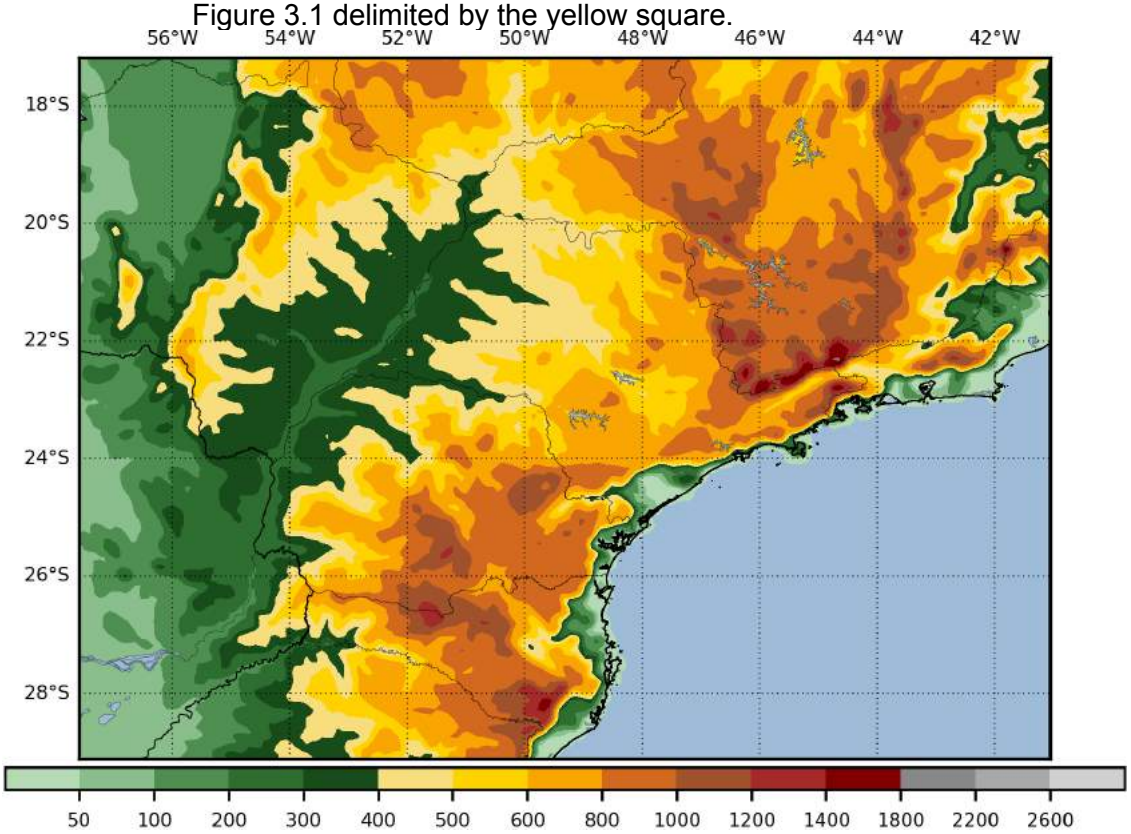
The 3DVAR methodology in the WRFDA system version 3.9.1 (released on August 17th, 2017) with a time window of 30 minutes (covering 15 minutes before and 15 minutes after the time analysis) was used to assimilate the relative humidity corrected by the flash rate density obtained through the interpolation of lightning observations.

The Background Error Covariance Matrix (BE) employed during the assimilation process was a generic matrix (CV3) provided by the WRFDA which can be used for any regional application. This matrix is an important factor that can affect the assimilation process significantly, however, the use of different types of BE was not explored in this study.

The CV3 was generated by NCEP using the National Meteorological Center (NMC) method (PARRISH; DERBER, 1992) as a difference of 48 and 24-hours

forecasts using the GFS model. This option uses the vertical recursive filter to model the vertical covariance. More details can be found in Wang et al. (2017).

Figure 3.3 -The terrain elevation (m) in the simulation domain which was showed in Figure 3.1 delimited by the yellow square.



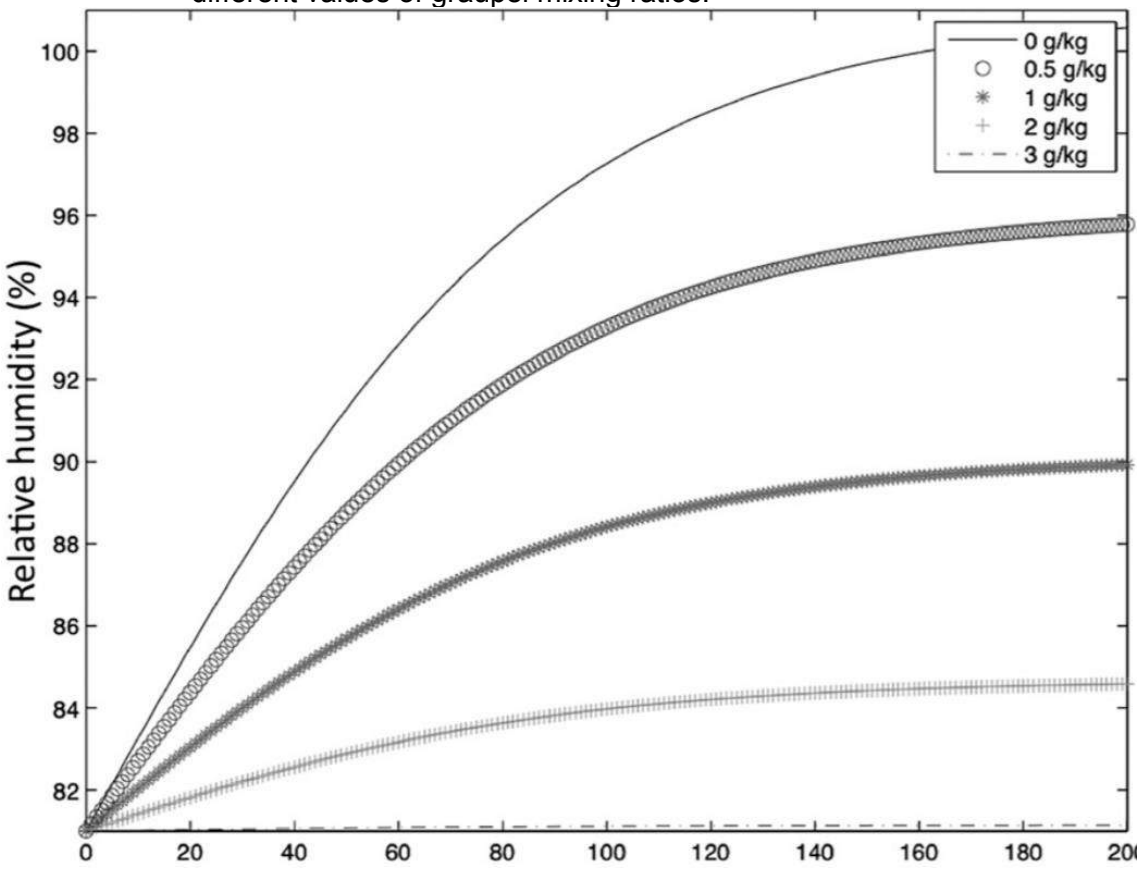
3.3 Assimilation Algorithm

The assimilation algorithm developed in this study to assimilate lightning data and correct the initial conditions of the model was based on the equation developed by Fierro et al. (2012) (Equation 3.1). The Equation 3.1 corrects the water vapor mixing ratio variable Q_v of the model based on the flash rate density X calculated using lightning detections. The saturation mixing ratio Q_{sat} and the graupel mixing ratio Q_g are obtained from the initial conditions of the model. The coefficients A, B, C, D, and α are constants and their values are defined as 0.81, 0.2, 0.01, 0.25 and 0.22, respectively.

$$Q_v = AQ_{sat} + BQ_{sat}\tanh(CX)\left[1 - \tanh(DQ_g^{\alpha})\right] \quad (3.1)$$

The behavior of the function defined by Equation 3.1 for a given Q_g can be seen in Figure 3.4.

Figure 3.4 – Plot showing the behavior of the function given by Equation 3.1 for different values of graupel mixing ratios.



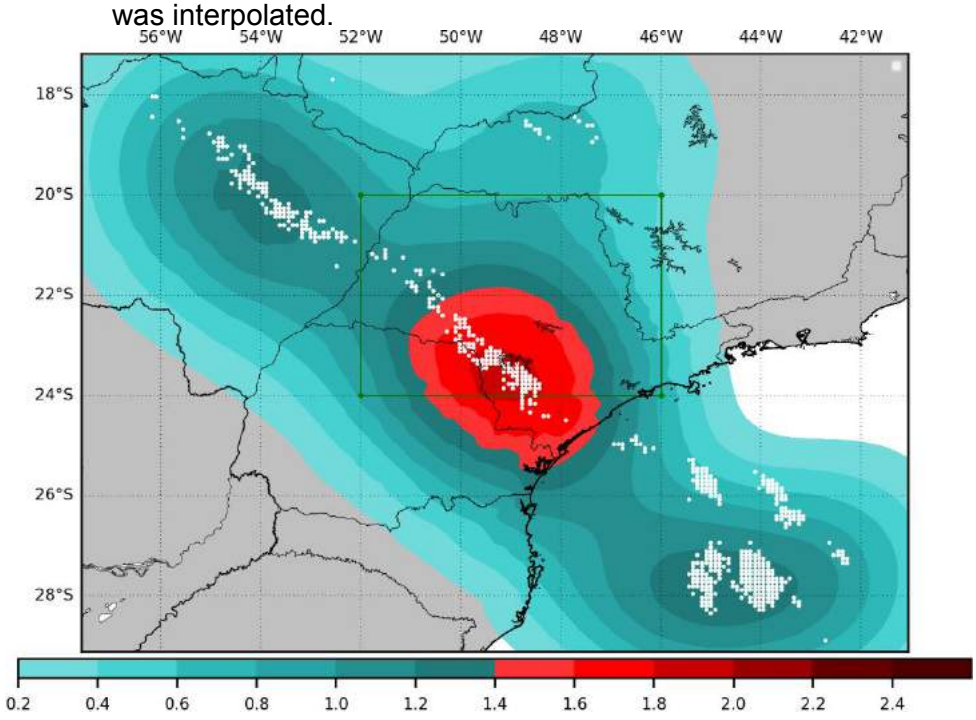
Source: Fierro et al. (2012)

It was defined a 30 minutes time window to calculate the flash rate which means that all lightning detections were accumulated for 30 minutes around the time analysis (15 minutes before and after a central time). After that, Q_{sat} is calculated using Q_v field from the initial conditions, Q_g is extracted also from the initial conditions and then using Equation 3.1 a new Q_v is calculated. Q_v is only calculated in the mixed-phase region (a region in the atmosphere between the 0°C and -20°C isotherms where the convection and the electrification are more intense – MacGorman; Rust (1998)) and when the Relative Humidity (RH) is below 81% or 60% (see next section).

However, in the case of the study developed by Fierro et al. (2012), the Q_v field was corrected inside the model through the modification of a microphysics parameterization. In this present study, the methodology applied is similar to that applied by Wang et al. (2017) where the assimilation process was made using the WRFDA with the 3DVAR technique to assimilate a proxy field (RH) obtained from Q_v calculated with the Equation 3.1. Therefore, the flash rate density is used to correct the Q_v field which is reinserted in the model as an observation. A diagram of the code developed for this purpose can be seen in Figure 3.7.

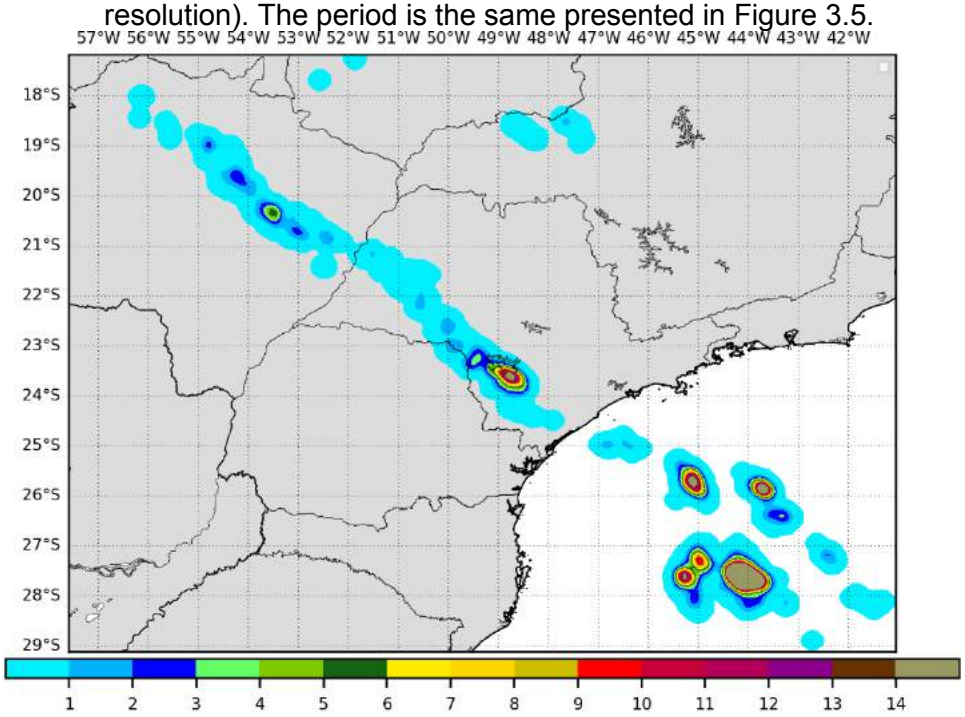
The intention was to generate "simulated observation stations" where, following the conditions described above, the variable RH is assimilated. These "simulated observation stations" are actually grid points of the domain to where lightning data was interpolated and the conditions to assimilate the RH were satisfied. Figure 3.5 shows the increment of Q_v (the difference between analysis and background) due to the assimilation of these simulated stations.

Figure 3.5 – Example of the Q_v increment in g/kg (the difference between analysis and background) induced by the application of the assimilation algorithm implemented in this present work. The white points represent the simulated observation stations (604) to where lightning data (Figure 3.6) was interpolated.



It is possible to notice that the assimilation algorithm always adds water vapor to the initial conditions of the atmosphere and the number of simulated observation stations is not proportional to the amount of water added to the atmosphere which is visible comparing Figures 3.5 and 3.6. The regions with many stations not necessarily are the regions with higher Q_v increment. In fact, if a specific region of the domain was previously dry and flashes are registered, the assimilation algorithm will induce a higher Q_v increment than the region that was previously saturated (Figure 3.5 and 3.6).

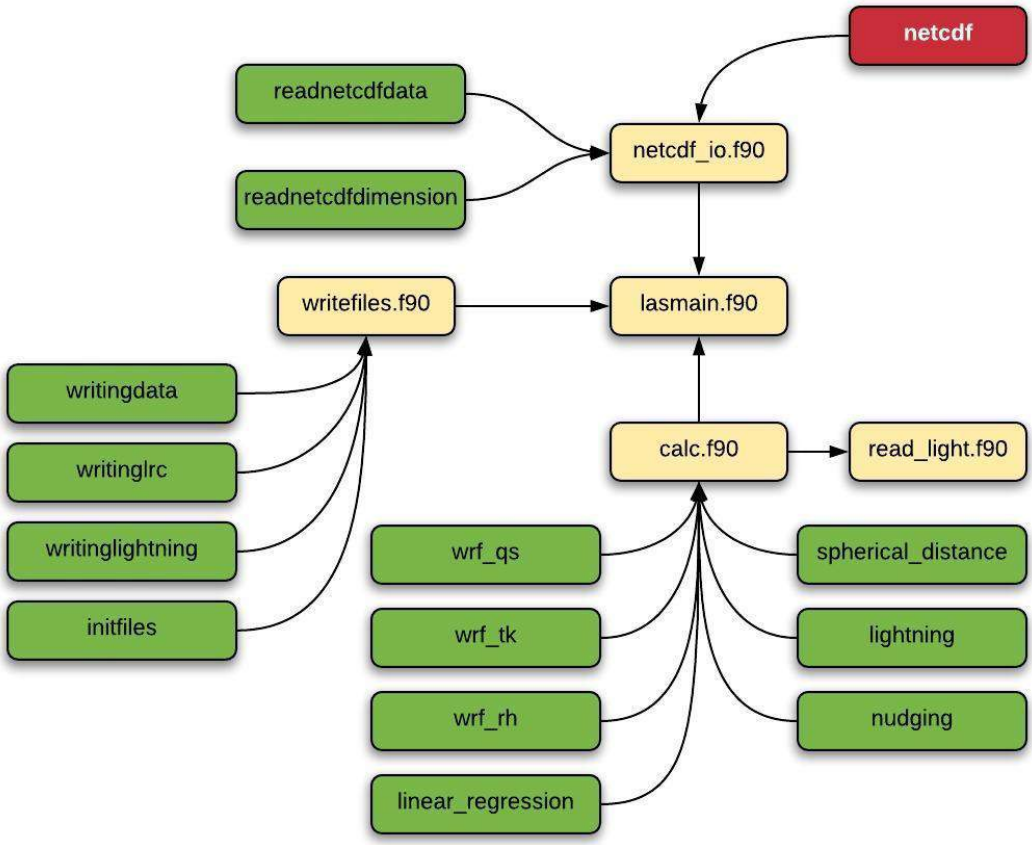
Figure 3.6 – Flash rate density ($\cdot 10^2/km^2$) calculated from lightning data detections and interpolated to the same grid used in the model (9km horizontal resolution). The period is the same presented in Figure 3.5.



The Lightning Assimilation System (LAS) has as main code what was named as “lasmain.f90”. It is from there where all the modules and functions are called following a specific flow. The module “netcdf_io.f90” uses the NetCDF library (external code) and contains the subroutines “readnetcdfdata” and “readnetcdfdimension” which are responsible for reading the variables from the WRF NetCDF files and the dimensions of the grids, respectively. The module “writefiles.f90” contains subroutines responsible for initializing (“initfiles”) and writing files containing the flash rate density (“writinglightning”),

The “writefiles.f90” code is also responsible for writing the file containing the water vapor mixing ratio information, the details about the simulated observation stations (“writingdata”) and the final file containing the simulated observations in little_r format (“writinglrc”).

Figure 3.7 - Structure of the code implemented in Fortran to convert lightning data into a little_r file which is assimilated by the WRFDA system. The functions and subroutines are shown in green while the core of the code is shown in yellow. An external library was also used (red).



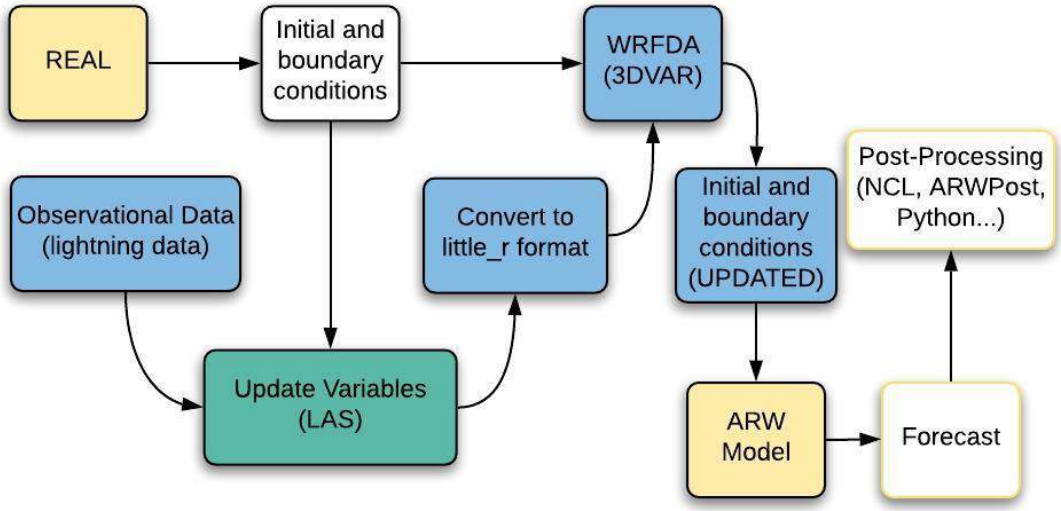
The reading of lightning data is made by the module “read_light.f90” while the interpolation of it, in order to produce the flash rate density in the same resolution of the model, is made by the subroutine “lightning” inside the module “calc.f90”. The subroutines “wrf_qs”, “wrf_tk” and “wrf_rh” are used to calculate Q_v, the temperature, and RH, respectively.

Besides that, the module “calc.f90” has three more subroutines: (1) the “spherical_distance” subroutine calculates the distance between two points on the surface of the Earth, this subroutine is used to calculate the distance of

each lightning detection to the grid points of the domain; (2) the “nudging” subroutine is where the correction in the Q_v variable is made using the Equation 3.1 with the conditions previously discussed; and (3) the “linear_regression” subroutine is used to find the RH threshold (see section 3.4).

The diagram in Figure 3.8 shows how the assimilation algorithm interacts with the structure of the WRF model (the WRF model diagram was previously shown in Figure 2.6).

Figure 3.8 - Diagram showing how the normal flow of the model showed in Figure 2.6 was adapted in order to proceed with the data assimilation process. In blue, it is shown what was added to the original structure and in green, the assimilation algorithm developed in this study and showed in Figure 3.7.



The observational data (lightning detections) are combined with the initial conditions generated by “REAL” inside the LAS code, Q_v is updated and then the simulated observation stations are generated and converted to the little_r file. Thus, the little_r file (which contains the simulated observations) is used in the WRFDA system in order to generate new initial and boundary conditions.

3.4 Adaptative threshold

One of the main contributions of this present research is associated with the ideal definition of the threshold for RH. While in the original algorithm developed by Fierro et al. (2012) and adapted by Wang et al. (2017) the RH threshold is

fixed in 81%, this research found out that it is more appropriate to use an adaptative threshold.

During the development of the experiments it was noticed that the algorithm does not have a good performance during the dissipation period of the meteorological system, i.e., when the simulation is started during the period of dissipation of the convection activity and the system is still producing flashes, the insertion of lightning data at this point can force the model to intensify the convection when it actually should be getting weaker. However, we still need to force a rapid response from the model in order to decrease the spin-up period and represent properly the convection in the first hours of simulation. This way, there is a delicate balance between the amount of water vapor that needs to be inserted in the model and the life cycle of the system.

The solution found for this problem was to use an adaptative RH threshold based on the concept of Lightning Jump. This variable gives the variation of flashes in a given time interval and it is commonly used to identify the occurrence of severe weather events (SCHULTZ et al., 2009; WILLIAMS et al., 1999).

In the case of this study, the number of flashes was calculated using 30 minutes intervals during 2 hours before the initial time of the simulation. Applying a linear regression model to these calculated values it is possible to find the angle formed by the line that minimizes the errors (Figure 3.9).

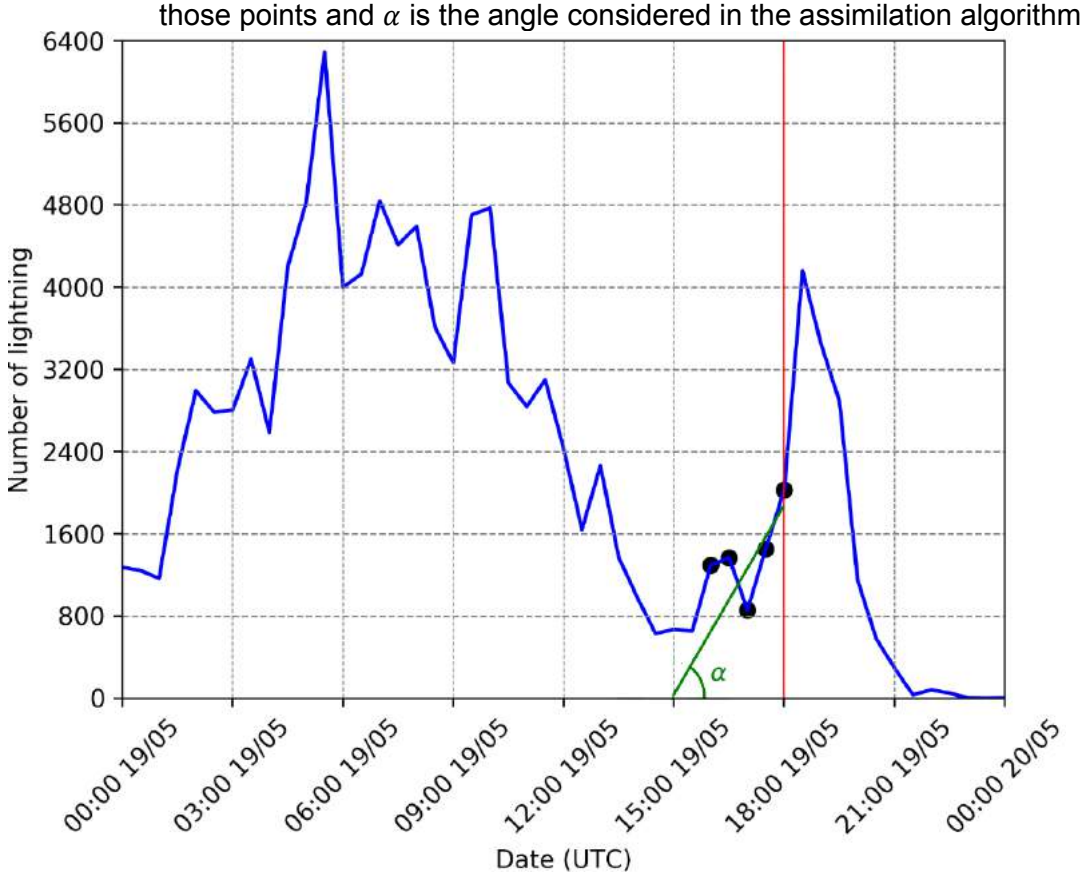
When the angle α is above $30°$, the RH threshold is the same used in other studies, 81%. However, when the angle α is below to that value the RH threshold is considered 60% (Equation 3.2).

$$f(\alpha) = \begin{cases} 0.81, & if\ \alpha > 30° \\ 0.60, & otherwise \end{cases} \tag{3.2}$$

Where $RH = f(\alpha)$. The thresholds and angles were defined empirically during the testing phase of the experiments.

With this new implementation, it is expected a better response from the model in the dissipating period of the system.

Figure 3.9 – Graphic showing the variation in time of the number of lightning detections per 30 minutes for a real case (blue line). The red line shows analysis time (18:00 19/05), the black dots represent the points considered to calculate the linear regression model, in green the line that fits best to those points and α is the angle considered in the assimilation algorithm.



3.5 Experiment design

The study proceeded with three different experiments during the occurrence of two distinct meteorological events aiming to assess the assimilation algorithm implemented here. The experiments were basically divided in control (CTRL) where no assimilation procedures were used, in lightning data assimilation (LIGHT) where lightning data was assimilated and in lightning data assimilation with an adaptative RH threshold (ALIGHT).

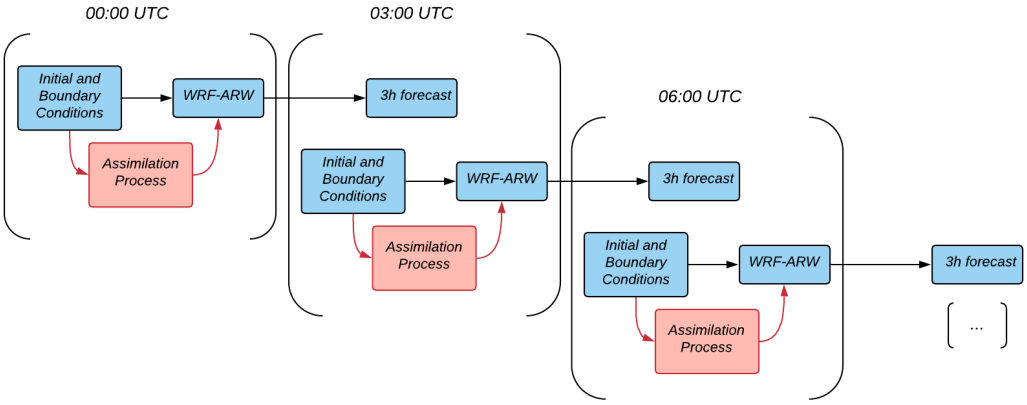
Moreover, the meteorological events selected for this study started at 00:00 UTC on 23/01/2018 and at 00:00 UTC on 19/05/2018. These specific days were selected due to the meteorological conditions associated with them. The systems analyzed had important morphological differences with each of them associated with a different synoptic scenario in the atmosphere. This way, the

assimilation process was evaluated in different situations. The discussion about the meteorological scenario in each event is presented in the next chapter.

All the experiments were composed by 8 short-term forecasts, i.e., for each simulation were performed 3 hours forecast, with an interval of 3 hours amongst each initialization totalizing 24 hours of simulation.

Furthermore, the simulations were all Cold Started which is expected to avoid the emergence of numerical instabilities due to the insertion of mass in the assimilation process. Figure 3.10 shows schematically how it is structured one experiment.

Figure 3.10 - Diagram showing the schematic structure of an experiment.



The initial conditions can be directly used by WRF-ARW in order to produce a 3 hours forecast (experiment CTRL) or it can be used in the assimilation process instead (Figure 3.8) which also produce a 3 hours forecast and define the experiment LIGHT or ALIGHT.

The experiments started at 00:00 UTC of the current day analyzed and ended at 00:00 UTC of the next day which means that the last simulation started at 21:00 UTC. As the number of lightning detections varies throughout the day, the initialization of the model in different times allows the assimilation process works with different quantities of lightning flashes which is useful when we want to assess the performance of this methodology.

Besides that, for the experiment that showed the best results, a 24 hours forecast was performed in order to analyze more deeply the impact of the assimilation process in a larger range of time.

3.6 Evaluation methods

This section shows some of the methods applied to evaluate the experiments performed in this study. The goal was to verify objectively/quantitively the impact of the lightning data assimilation process.

3.6.1 Precipitation analysis

One of the methods used to assess the experiments was the precipitation analysis. Firstly, it was necessary to use Barnes algorithm (BARNES, 1964) to interpolate the observational data in a grid with the same configuration as the one used by the model (see section 3.2).

The Equation 3.3 defines the interpolation method where x_i^n is the variable (precipitation in this case) in the i-th grid point in the n-th iteration:

$$x_i^{n+1} = x_i^n + \frac{\sum_{k=1}^{K} w_{i,k}\,(y_k - x_i^n)}{\sum_{k=1}^{K} w_{i,k}} \tag{3.3}$$

Where $\forall\, i: x_i^0 = 0$, y_k is the k-th observation and $w_{i,k}$ is the weight function for the k-th observation in the i-th grid point given by Equation 3.4:

$$w_{i,k} = exp\left[-\left(\frac{d_{i,k}}{g^n}\right)^{\alpha}\right] \tag{3.4}$$

$d_{i,k}$ is the distance in km between the k-th observation and the i-th grid point, α is a constant greater than 1 and g^n is the falloff parameter in the n-th iteration which is given by Equation 3.5:

$$g^{n+1} = cg^n \tag{3.5}$$

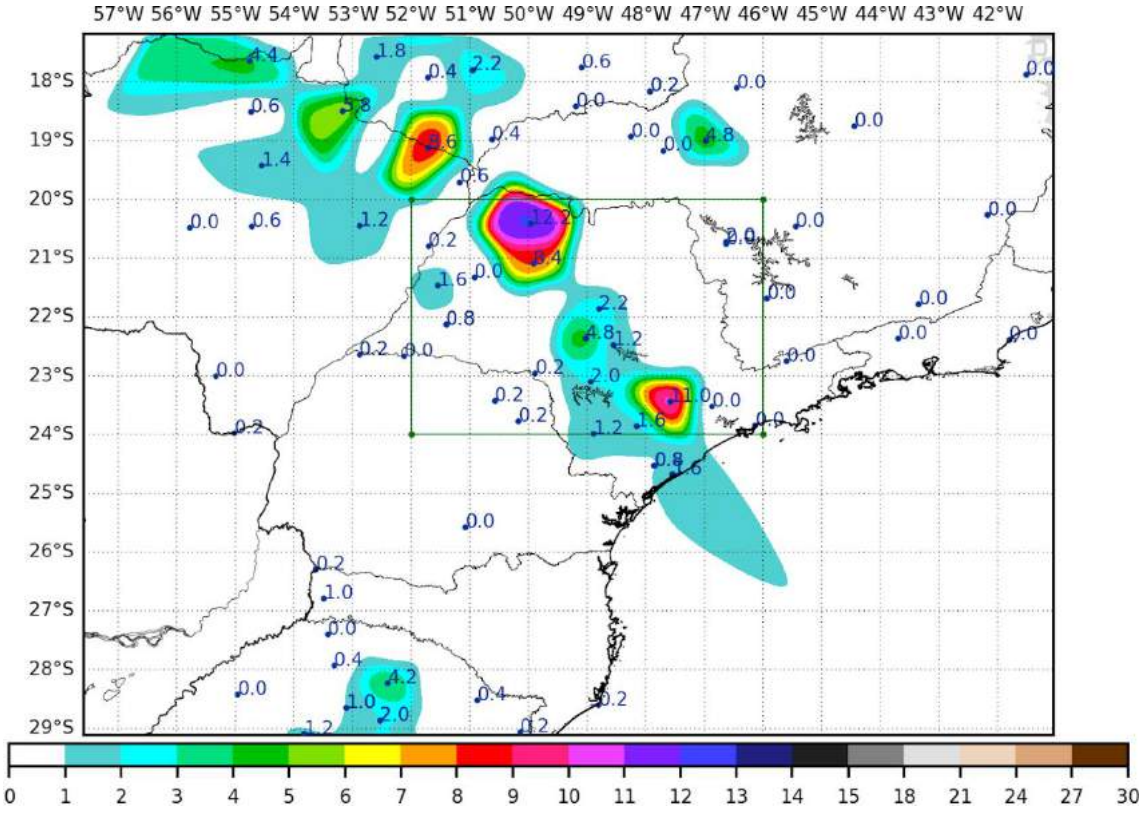
Where $0 < c < 1$.

The parameters were adjusted empirically through the comparison with observational data. Based on that, it was defined as a total of two iterations with $\alpha = 1.5$, $g^0 = 80$ and $c = 0.5$. The distance between the observations and the grid points was calculated using spherical trigonometry.

One example of this interpolation method can be seen in Figure 3.11 where it shows the data registered by the observation stations and the resulting precipitation field.

Once the precipitation data is interpolated to the same grid used in the model it is possible to apply techniques to compare the observational data with the simulation.

Figure 3.11 - Example of the application of Barnes Interpolation Algorithm for the precipitation filed (mm/h) using data provided by CPTEC/INPE.



The methods used to evaluate the positioning of the precipitation and compare the different experiments with the observational data were Threat Scores (TS), False Alarm Ratio (FAR) and Probability of Detection (POD). The Equations 3.6, 3.7 and 3.8 describe each one of these methods.

$$TS = \frac{hits}{hits + misses + false\ alarms} \tag{3.6}$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms} \tag{3.7}$$

$$POD = \frac{hits}{hits + misses} \tag{3.8}$$

Where $hits$, $misses$, and $false\ alarms$ are defined by Tab. 3.1. The perfect score for TS and POD is 1.0, while for FAR is 0.0.

Table 3.1 – Contingency table.

$Simulated/Observed$	yes	no
yes	$hits$	$false\ alarms$
no	$misses$	$correct\ negatives$

The Tab. 3.1 basically defines some variables used to calculate the equations above. For example, when the model predicts precipitation in some region and it is also observed by the meteorological observation station this is called hit; When the model predicts precipitation in a region where the observations have shown no precipitation this is called $false\ alarm$. A $miss$ happens when the model was not able to predict the precipitation where it was observed.

Sometimes it happens that the precipitation predicted by the model is not synchronized with the observation because the fields simulated by the model are delayed in relation to what was observed. However, it is considered right (hit) when the model predicts a certain amount of precipitation accumulated just a few km from where it was in fact observed. Based on that, it was defined some thresholds to calculate the set of equations defined above (Equations 3.6, 3.7 and 3.8):

1. 30 km and 1mm;
2. 20 km and 5mm;
3. 20km and 10mm.

For example, the first condition is satisfied if it is registered/simulated a volume of precipitation greater than 1mm over a radius of 30km around the grid point considered. So, if the model and the observation satisfy this condition, we have a hit. This procedure is applied for all grid points of the domain and all the $hits$, $false\ alarms$ and $misses$ are counted and the equations mentioned above are calculated.

Moreover, it was also applied a BIAS estimator aiming to determine the difference quantitively between the experiments and the observed precipitation. The BIAS is given by Equation 3.9.

$$BIAS = \frac{1}{N}\sum_{i}^{N}(f_i - o_i) \tag{3.9}$$

Where f_i and o_i are the forecast and the observation, respectively, in the i-th grid point. The techniques here described and others commonly used in meteorology to compare the forecasts with the observational data are presented in more details in Stanski et al. (1989).

This present section showed the methods applied to all the short-term forecasts experiments. Based on the results, the evaluation methods were applied again with an additional evaluation method (described in the next section) on the experiment with the best performance in order to analyze other characteristics of the assimilation process.

3.6.2 Reflectivity analysis

Another way to analyze and evaluate the experiments performed in this study was through the comparison of the reflectivity fields simulated by the model with lightning data using specific algorithms developed for this purpose.

This advanced analysis was made only for a specific experiment with a larger range of simulation (24 hours) where a good performance of the model during the short-term experiments was observed considering previous evaluation methods discussed in section 3.6.1.

One way to identify if the reflectivity field was well simulated is comparing what was called here as “center of mass”. Basically, it is expected that regions with higher values of reflectivity (3km of height) will contain clouds with a high concentration of ice and consequently a high amount of lightning flashes will be detected.

Therefore, the center of mass of the reflectivity (which is the weighted average) will have similar behavior to the arithmetic center of lightning detections.

The definition of center of mass is given by the Equations 3.10 and 3.11:

$$x_c = \frac{\sum_i^N w_i x_i}{\sum_i^N w_i} \tag{3.10}$$

$$y_c = \frac{\sum_i^N w_i y_i}{\sum_i^N w_i} \tag{3.11}$$

Where x_c and y_c are the coordinates of the center of mass (longitude and latitude, respectively), w_i is the weight given by the reflectivity in the i-th grid point and x_i and y_i are the coordinates in that grid point.

4 RESULTS AND DISCUSSION

In this section, the results obtained in this research are presented and discussed. The impact of the Lightning Data Assimilation System on the WRF simulations is evaluated for two distinct days, each one with different meteorological characteristics. Even though only two days were analyzed, it is emphasized that eight short-term simulations for each day were performed, giving us a good sample of simulations for evaluating the methodology applied on this work.

4.1 Case study I

4.1.1 Observational analysis

The first case analyzed occurred on May 19th, 2018 and presented the peak of lightning activity at 05:00 UTC. The inner domain registered the maximum activity between 09:00 and 12:00 UTC. The number of lightning detections and the distribution along the day inside the simulation domain can be seen in Figure 4.1.

As discussed in chapter 2, lightning activity is intrinsically associated with the formation of storms which means that an increase in the lightning activity also indicates an increase in the atmospheric instability. This instability can be associated with different atmospheric mechanisms such as thermodynamics.

The instability associated with thermodynamics usually is observed at the end of the day and it is due to the intense heating of the Earth's surface by the Sun which in turn warms the lower layers of the atmosphere. The warm layers tend to ascend in the atmosphere dragging moisture to higher altitudes which condense creating clouds and eventually thunderstorms (BROWNING; LUDLAM, 1962; PETTY, 2008; WEISMAN; KLEMP, 1986).

Moreover, this process can also occur mechanically, i.e., the layers of the atmosphere are forced to ascend due to a mechanical force like during the passage of a Cold Front. The colder and denser air slides under a warmer air mass ahead pushing the layers to the top of the troposphere inducing the formation of clouds and storms (BERGERON, 1937; REBOITA et al., 2010). At

night, especially in the beginning of the day when the temperature is minimum, the friction of the atmosphere is smaller which allows the Cold Front to move faster pushing more layers ahead and inducing more atmospheric instability which contributes even more for the formation of storms (BRUNDIDGE, 1965).

Figure 4.1 – Number of Lightning Detections per 30 minutes in the inner and outer domains on 19/05/2018. The black dots over the blue line and the numbers associated with them indicate the number of lightning detections considered for the assimilation at that time analysis.

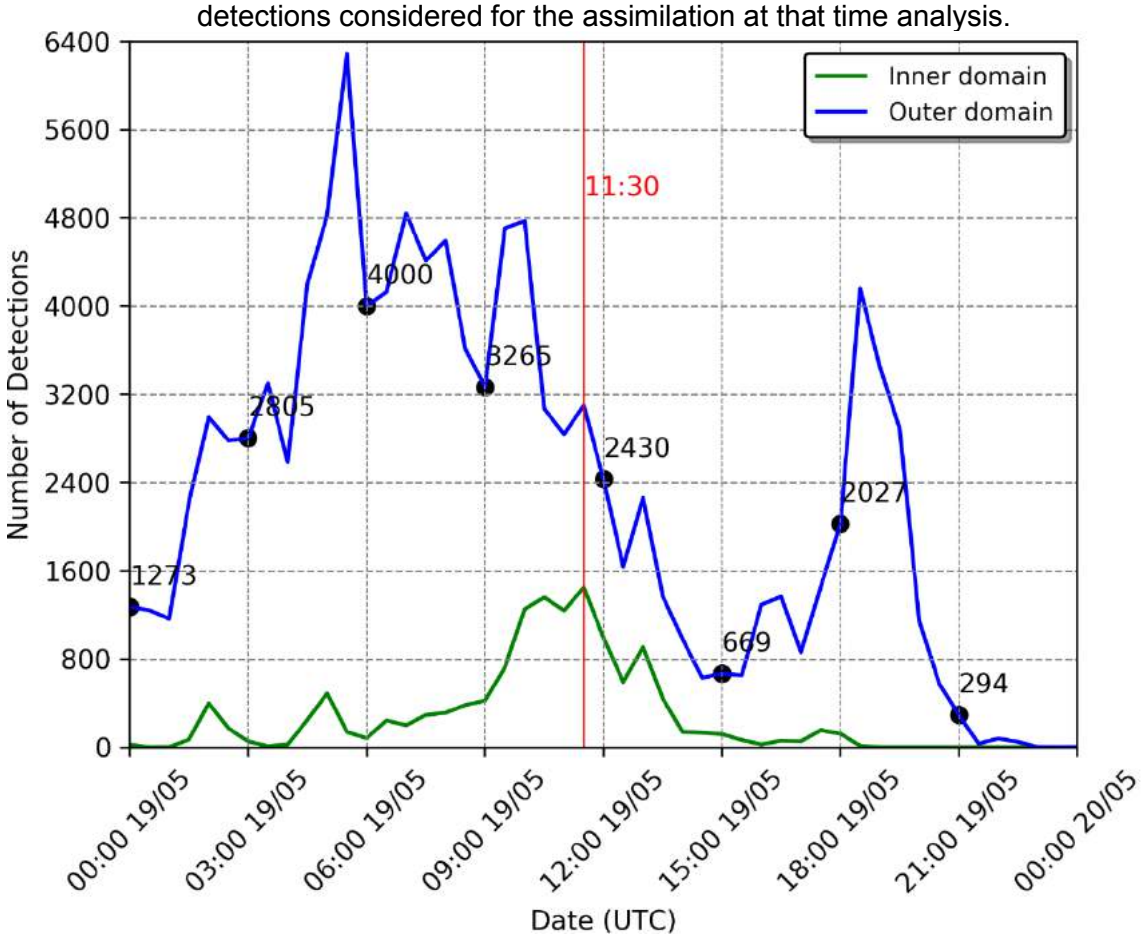
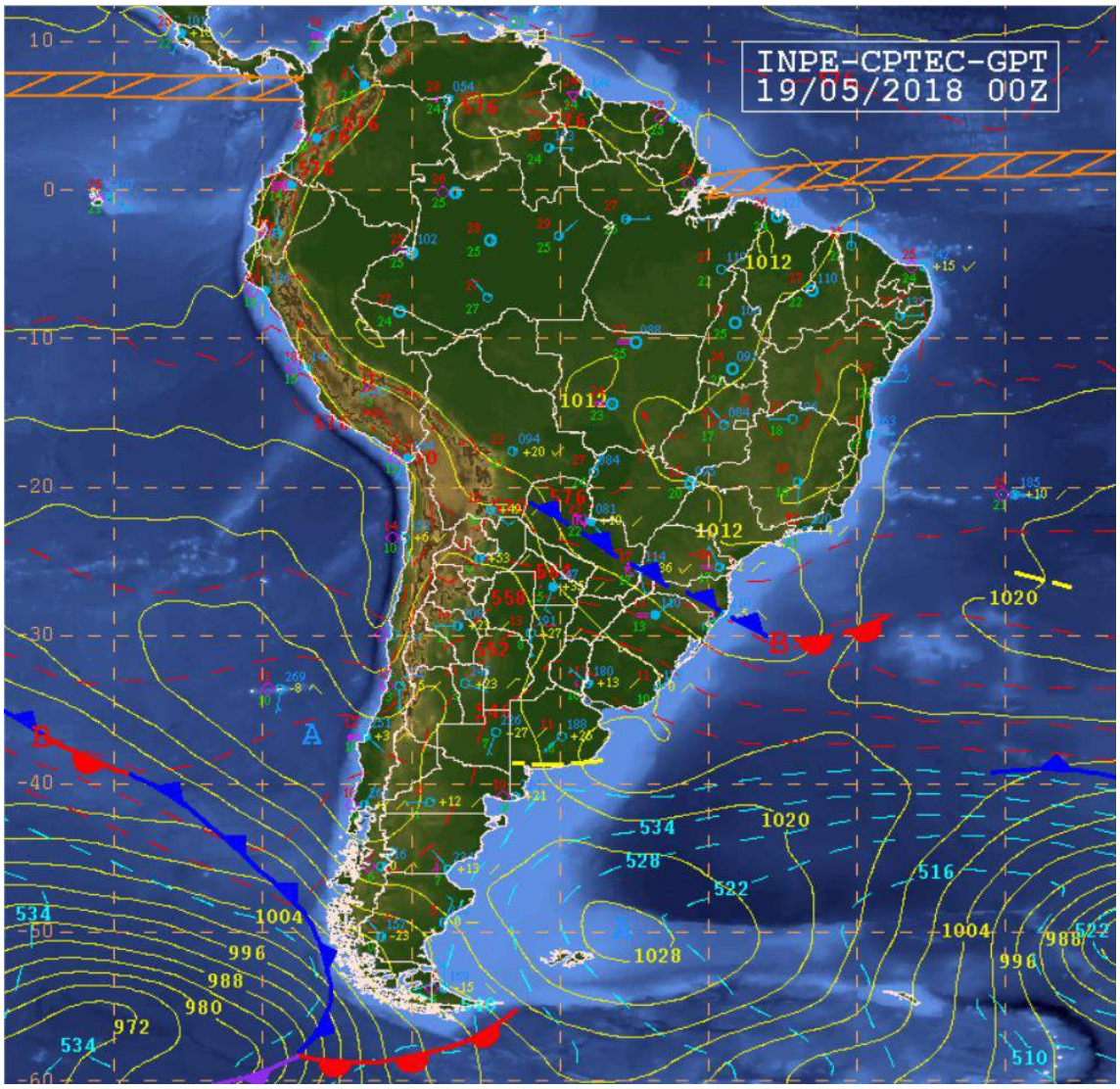


Figure 4.1 shows a peak of detections in the first hours of the day, which indicates the formation of storms associated with the occurrence of a Cold Front. The second peak around 19:00 UTC is associated with the convection induced by the daytime heating in an atmosphere already unstable due to the Cold Front in the region.

The Figures 4.2 and 4.3 show synoptic charts at 00:00 UTC on 19/05/2018 confirming the occurrence of a Cold Front in the study area. The synoptic environment on this day contributed to a high incidence of lightning in the region

throughout all day making this specific day ideal to apply the assimilation technique in order to understand how the use of lightning data can affect the simulation.

Figure 4.2 – Synoptic chart at 00:00 UTC on 19/05/2018 showing the synoptic environment on the surface in South America.

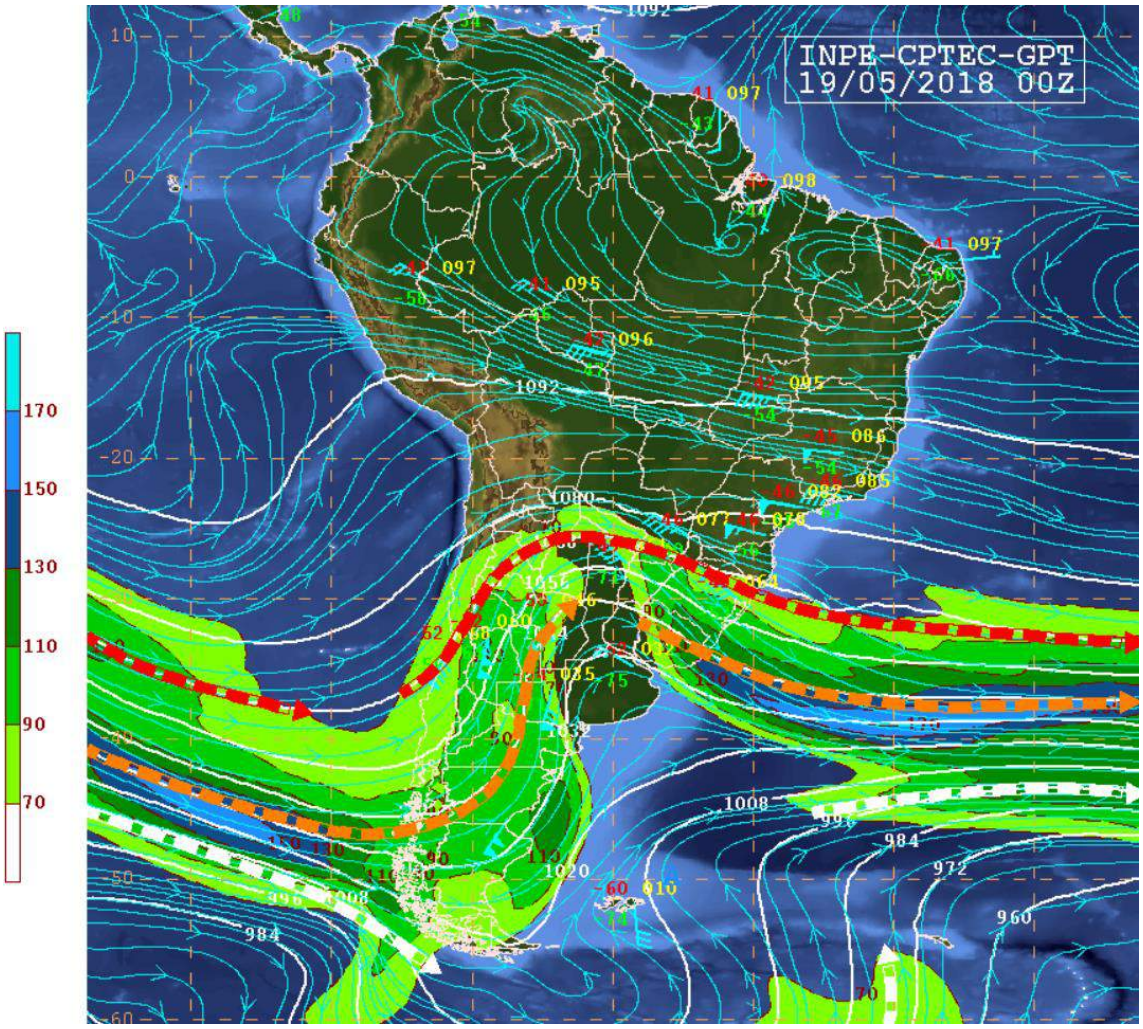


Source: CPTEC (2019).

The airflow in 250hPa shows a wave propagating eastward inducing cyclonic curvature in the subtropical jet over southern Brazil (Figure 4.3). Also known as Trough, this atmospheric system induces negative vorticity advection on its right side which means that there is a higher-pressure atmospheric system ahead with divergent atmospheric airflow. By the principle of mass conservation,

surface convergent airflow is required which generates a center of low atmospheric pressure (PETTY, 2008), as it can be seen near the southern coast of Brazil in Figure 4.2.

Figure 4.3 – Synoptic chart at 00:00 UTC on 19/05/2018 showing the synoptic environment in 250hPa in South America.



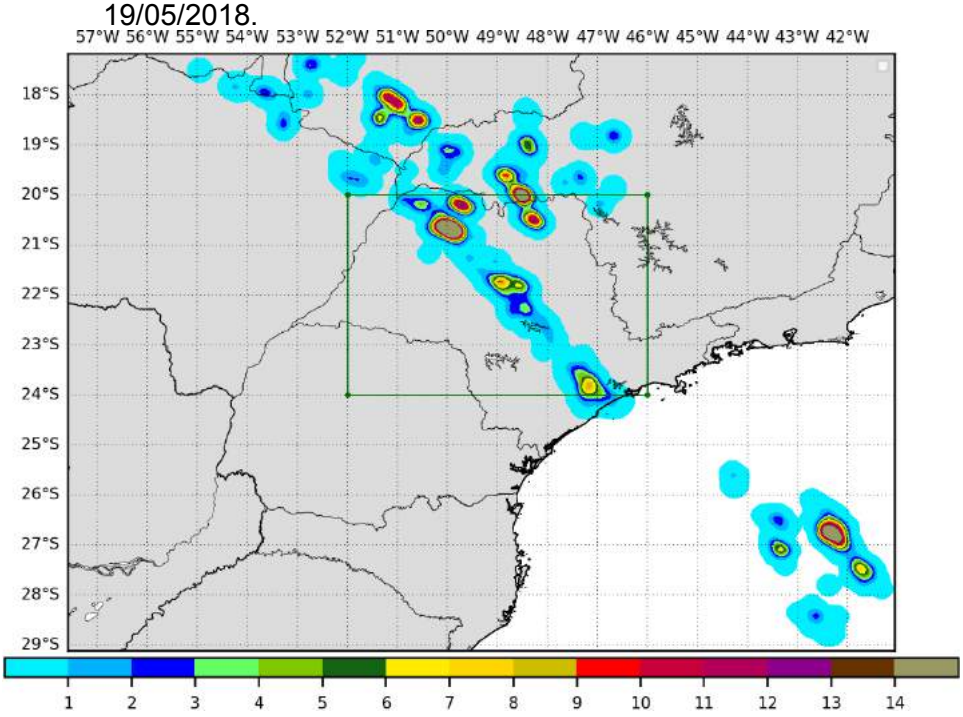
Source: CPTEC (2019).

Due to the Coriolis force, a convergent airflow creates an enclosure low-pressure system which rotates cyclonically (clockwise direction) moving air masses throughout the continent. On its left side, the low-pressure system pushes air masses from the south towards the north creating a Cold Front (transition zone where a cold air mass moves towards a warmer one) and, a Warm Front on the opposite side (BJERKNES, 1921).

The Figures 4.2 and 4.3 show a low-pressure system on the surface right below the right side of the Trough on high levels of the atmosphere which characterizes a barotropic atmosphere, the early stages of a system under development (BJERKNES, 1921).

The Cold Front affected the entire simulation area reaching the peak of lightning activity in the inner domain (the region of São Paulo state) at 11:30 UTC as it can be seen in the spatial distribution of flashes in Figure 4.4.

Figure 4.4 – Flash density for the peak of Lightning Detections inside the inner domain (Figure 4.1) for a time window of 30 minutes, from 11:15 to 11:45 UTC 19/05/2018.



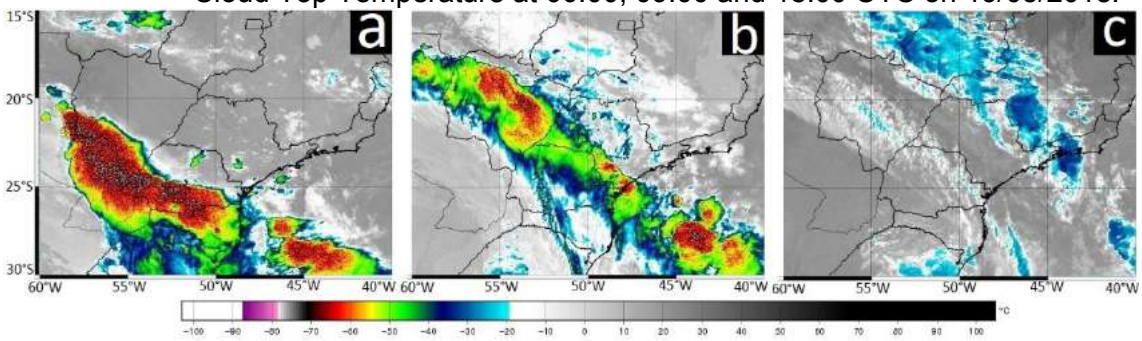
The cooling of the continent after the summer provides the necessary environment for the advance of the South Atlantic Anticyclone, a high atmospheric pressure system typically observed on the surface of the South Atlantic during the whole year, towards to South America continent.

This atmospheric configuration induces the formation of an alley with winds bringing large amounts of moisture from the Amazon rainforest to Paraguay region, northeast of Argentina and South of Brazil. These winds, also known as Low-Level Jets, play an important role in the formation and development of Convective Mesoscale Systems and Cold Fronts (BROWNING; PARDOE, 1973; PAEGLE, 1998; SALIO; NICOLINI; ZIPSER, 2007). At the same time, a

colder continent facilitates the advance of the Cold Fronts further north until the point where it can reach the northeast of Brazil (CARDOZO et al., 2015).

The displacement of the Cold Front, as well as the thunderstorms, can be seen in Figure 4.5.

Figure 4.5 – Images from GOES-16 satellite of the channel 7 (3.90μ) showing the Cloud Top Temperature at 00:00, 09:00 and 18:00 UTC on 19/05/2018.



Source: DSA/CPTEC (2019).

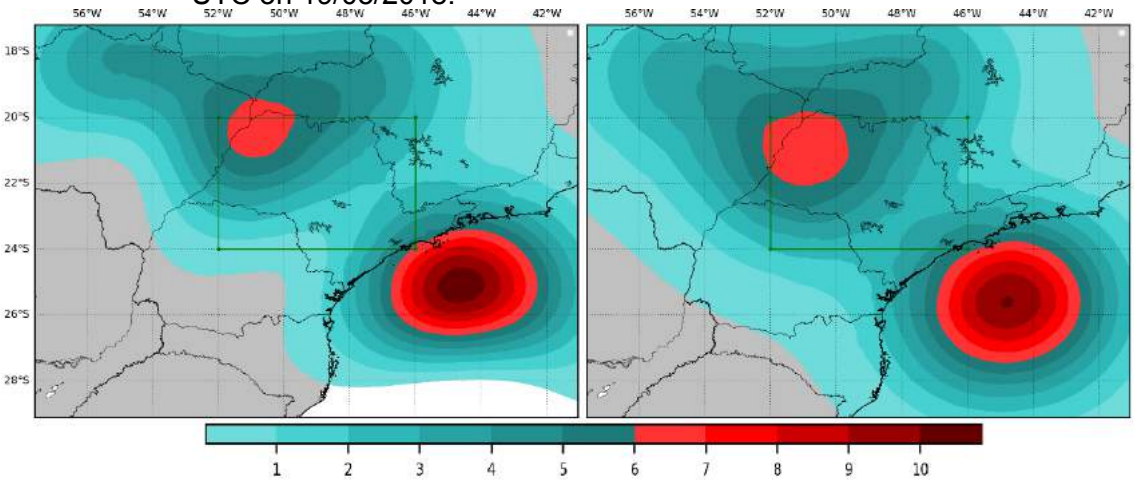
The satellite images show the intensification of the thunderstorms in the first hours of the day which is compatible with the observations of lightning (Figure 4.1). Even though the system starts to dissipate as a whole after 05:00 UTC, it is possible to see the intensification of small storms in the south border of São Paulo state, especially near the coastline. These small storms were the early stages of a squall line that produced a high incidence of lightning and a large volume of rainfall in São Paulo state (HOUZE JUNIOR, 1977).

Based on this synoptic analysis, the assimilation technique was applied during the occurrence of a large-scale system (Cold Front) that also induced a squall line in the region of São Paulo state. This way, it is expected a more uniform displacement and spatial distribution of lightning detections.

4.1.2 Impact of the assimilation algorithm

The impact of the lightning data assimilation technique developed by Fierro et al. (2014) and adapted by Wang et al. (2017) used in this study as well as a version of it developed here are focused on the correction of the moisture content in the middle levels of the troposphere. Its impact and the differences between these two experiments can be seen in Figure 4.6.

Figure 4.6 – ALIGHT (left) and LIGHT (right) experiments showing Q_v increment in g/kg (Analysis minus CTRL) integrated vertically in the atmosphere at 09:00 UTC on 19/05/2018.



Note that the algorithm added more water vapor on the ocean near the coastline and on the northwest of São Paulo state with a small difference between both experiments. A high Q_v increment means that either it was observed a high number of lightning detections or the relative humidity in the region with lightning activity was previously low.

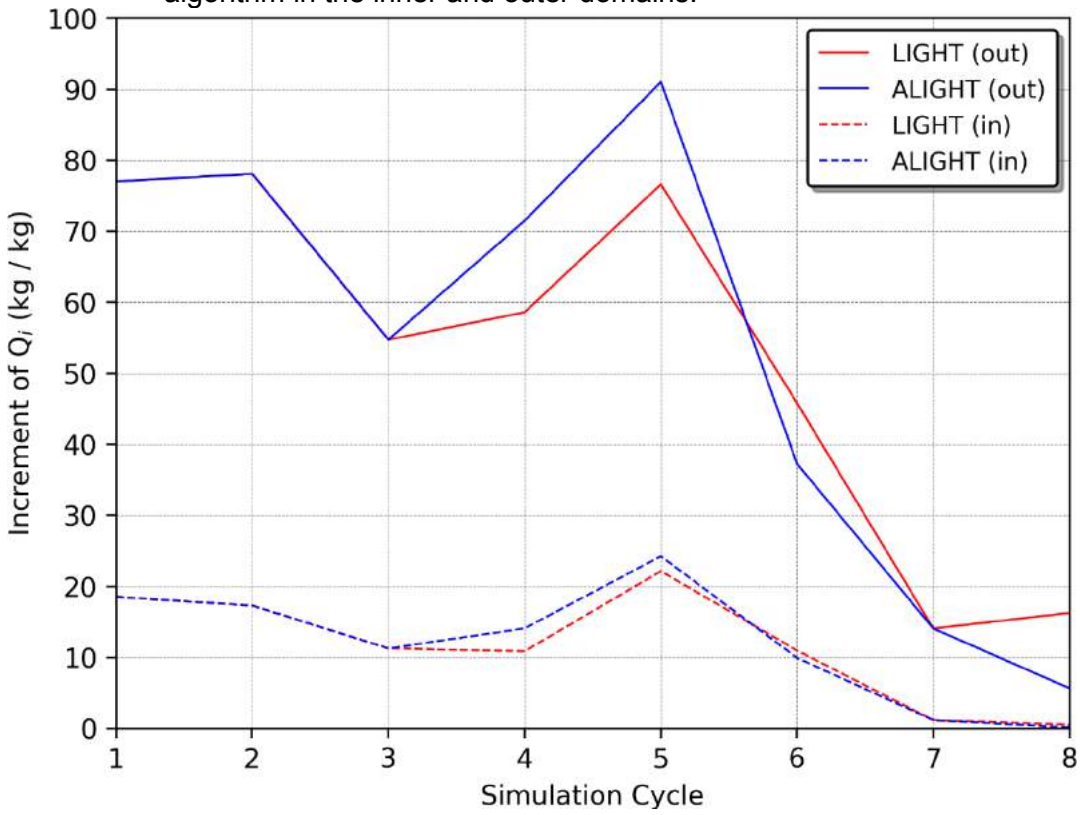
The assimilation algorithm works adding water vapor to the atmosphere which increases the atmospheric instability. Lightning activity induces the algorithm to add water vapor in the locations where thunderstorms are observed. This procedure helps to improve the short-term weather forecasting by starting deep convection and forming the correspondent cold pools (FIERRO et al., 2012; MANSELL et al., 2007).

Moreover, this smooth pattern in the Q_v increment field observed in Figure 4.6 is the result of the data assimilation system used to assimilate the simulated observation stations. Possibly, this pattern is observed due to recursive filters applied to the two-dimensional fields of control variables increments during the process to obtain the background error covariance matrix (BARKER; HUANG; GUO, 2003).

In order to analyze the difference between both experiments as well as the amount of water vapor added in different areas of the simulation domain, the

total amount of water vapor added to the atmosphere in the whole domain for each simulation cycle was calculated (Figure 4.7).

Figure 4.7 – Q_v increment for each simulation cycle performed on 19/05/2018. The chart shows the total amount of water vapor added by the assimilation algorithm in the inner and outer domains.



As it was discussed in the methodology section, the correction in the relative humidity threshold is applied in the dissipating phase of the meteorological system. Therefore, differences between the experiments are not observed in all simulation cycles.

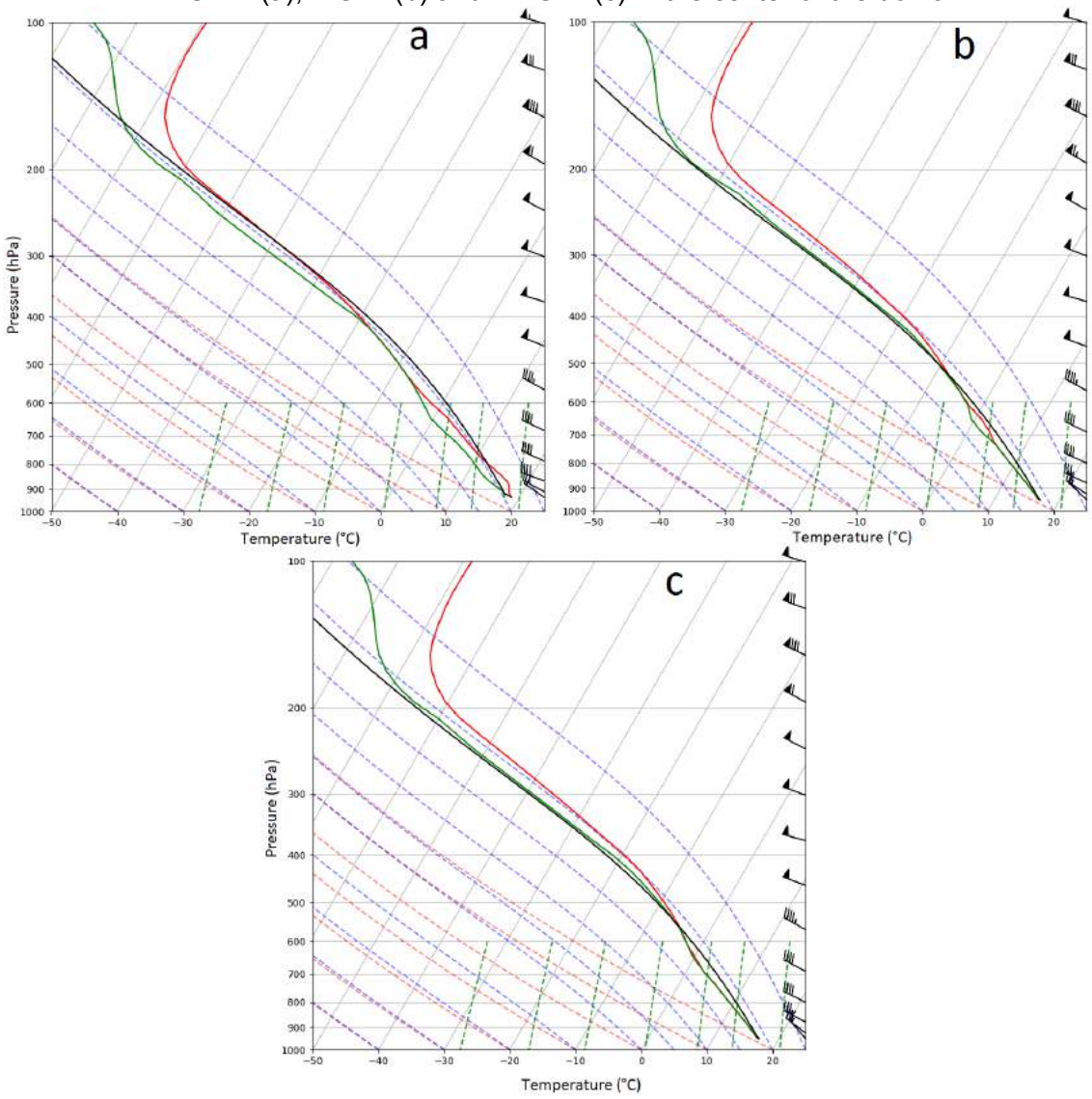
It is possible to see that in some cycles the ALIGHT algorithm added more water vapor. Even though the ALIGHT experiment corrected fewer grid points than the LIGHT experiment, the amount of water vapor inserted is greater since the coefficients of the Equation 3.1 were not modified.

Note that when we analyze the vertical profile of the simulations (Figure 4.8), the experiments with data assimilation presented a more unstable atmosphere after 1 hour of simulation. The amount of moisture in the lower levels of the

atmosphere (from the surface up to 600hPa) was increased while the wind speed and direction were affected on a smaller scale.

The ALIGHT and LIGHT experiments also presented differences between each other with a significative increase in the atmospheric instability in the ALIGHT experiment. Especially, due to the higher values of relative humidity observed between 700 and 600 hPa.

Figure 4.8 – Skew-T chart showing vertical profile of air temperature, dewpoint and wind at 10:00 UTC on 19/05/2018 (1-hour simulation) for the experiment CTRL(a), LIGHT(b) and ALIGHT(c) in the center of the domain.



4.1.3 Model performance

The analyses of the experiments performed on 19/05/2018 using lightning data to improve the initial conditions of the model showed significant improvements

in the first hours of simulation when compared to the CTRL experiments. In general, the LIGHT/ALIGHT experiments presented improvements in the representation of the rainfall as well as in the location of the system.

The main improvements were observed when the observation stations registered a high intensity in the rainfall. It is known that the WRF model underestimates the volume of rain in most of the cases with extreme rainfall observed, especially in the first hours (MOYA-ÁLVAREZ et al., 2018). Therefore, the addition of moisture in the initial conditions using the Equation 3.1 supported the development of instabilities as soon as the model started which decreased the spin-up time and returned a rapid response from the simulation.

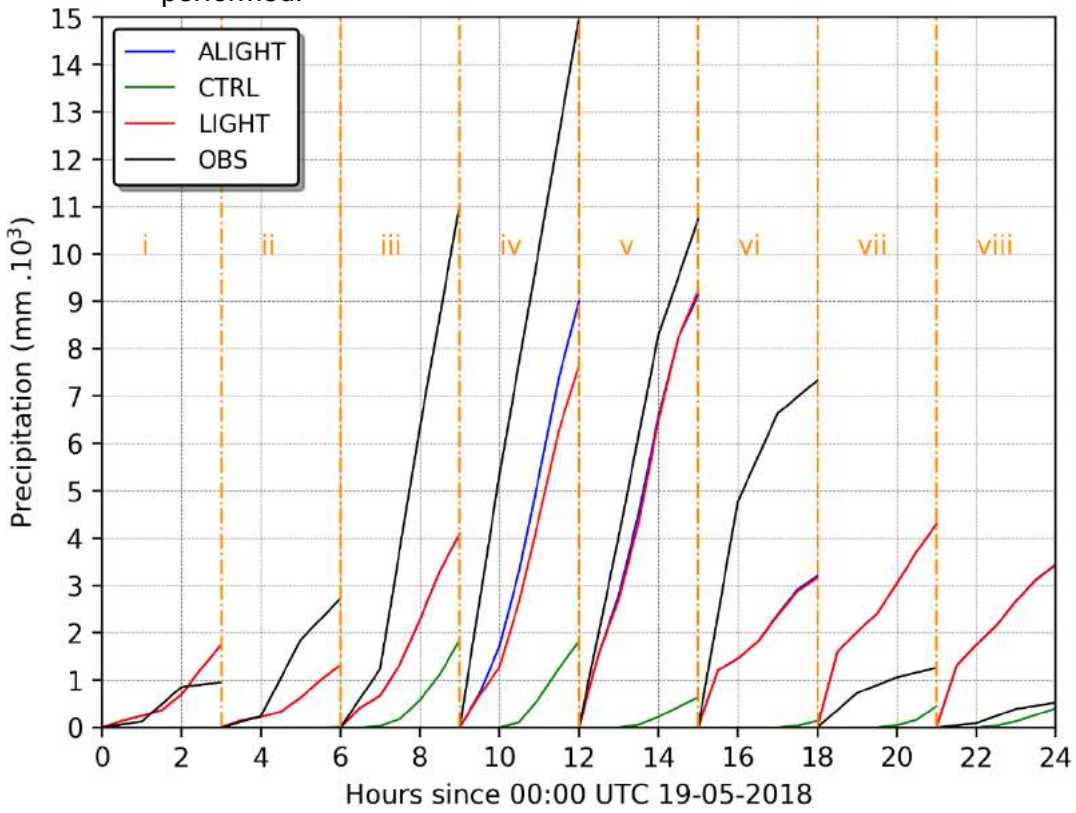
Since the spin-up time of the WRF model is about 6 to 12 hours many other studies have tried to decrease this period using different approaches in order to obtain a faster response and improve the first hours of simulation (CHEN, Fei et al., 2007; COSGROVE et al., 2003; SOKOL, 2009). The reduction in the spin-up time means a significant reduction in the use of computation resources.

Similar results were observed in other studies that used lightning data assimilation techniques to initialize the meteorological model. It was also observed a better representation of the meteorological fields in the first hours of simulation (CHEN, Zhixiong et al., 2019; FIERRO et al., 2012; WANG, Y. et al., 2017).

Analyzing the differences between the LIGHT and ALIGHT experiments it was possible to observe that the second had a small improvement over the first one especially in representing the precipitation field, even though the differences were only observed in periods where a decrease in the number of lightning detections was also observed.

The Figure 4.9 presents the sum of the precipitation field accumulated in the whole simulation domain for the CTRL, LIGHT and ALIGHT experiments and for the precipitation field observed for each one of the simulation cycles.

Figure 4.9 – The sum of the accumulated precipitation in the inner domain on 19/05/2018 for each simulation cycle and for each experiment performed.



Note that the experiments with lightning data assimilation had better performances in the cycles where the amount of precipitation was higher (cycles iv and v), with the best performance observed in the ALIGHT experiment especially in cycle-iv.

Even in the first cycles (i and ii) where the amount of precipitation was small the experiments with lightning data assimilation still had a better performance. However, in the last cycles (vii and viii) the LIGHT and ALIGHT experiments overestimated the amount of precipitation observed in the domain.

This overestimation is possibly explained by the meteorological environment simulated and not specifically because of the number of lightning detections used to initialize the model. According to Figure 4.1, it was used in the assimilation process 1273 lightning detections in cycle-i and 2805 in cycle-ii. During these two cycles, the precipitation volume increased rapidly due to the approximation of the Cold Front and the experiments with assimilation had a good response to the insertion of lightning data. While the cycles vii and viii

(with the assimilation of 2027 and 294 lightning detections, respectively) were performed in an environment with the convection predominantly induced by thermal forcing which generated a small amount of rain even though it was still observed a high incidence of lightning (this phenomenon is explained in more details in the next case).

The highest volumes of precipitation are usually observed in large scale systems such as Cold Fronts (OLIVEIRA JÚNIOR et al., 2014). Meanwhile, the analyses of lightning databases suggest that a large number of lightning can be detected in both, large scale and local convective systems in the summer of the Brazilian southeast since the last one tends to occur all over the study area in this period of the year.

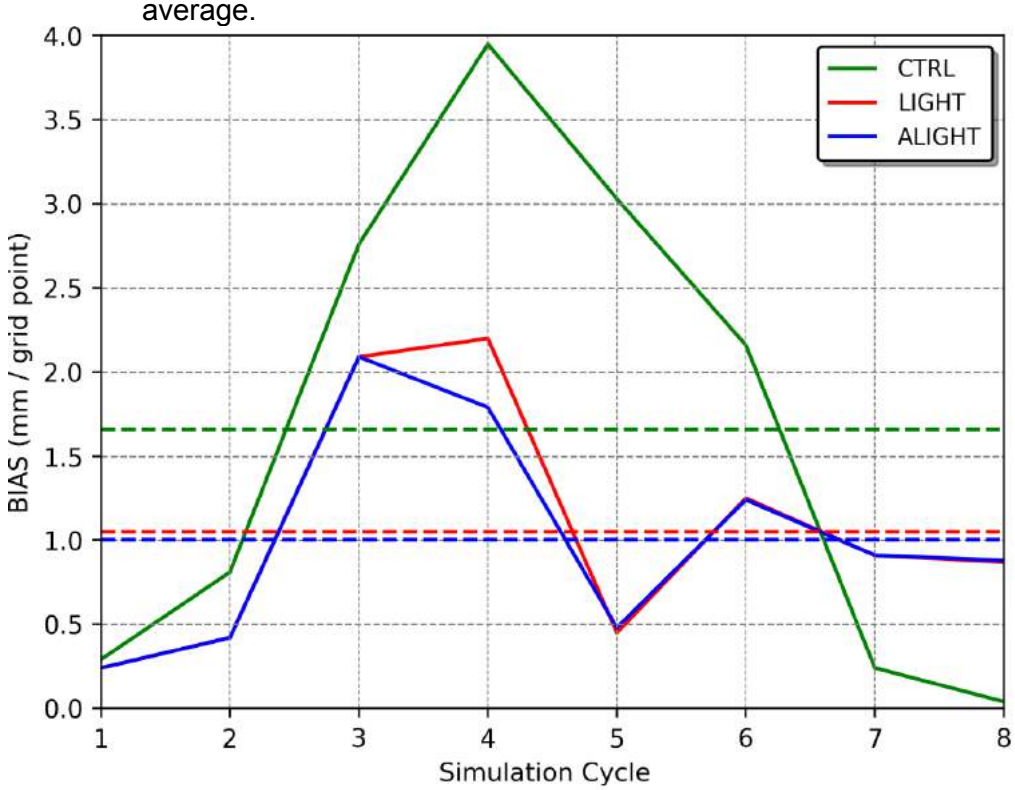
Fierro (2012) developed a lightning data assimilation algorithm analyzing dozens of meteorological events over the United States, i.e., the algorithm was developed for midlatitudes where large-scale systems are predominant. This way, it is expected that the Equation 3.1 presents better results when associated with large-scale systems instead of tropical regions where the synoptic environment is completely different, and the precipitation is frequently associated with local convection induced by the daytime heating.

The BIAS for the 3 hours accumulated precipitation is presented in Figure 4.10. As mentioned above, the assimilation algorithm presents a better performance during the periods with higher amount of precipitation (cycles iii, iv, v and vi) while in the cycles last cycles (vii and viii) the CTRL experiment had a better performance.

Note that in cycle-iv, the correction made in the data assimilation algorithm originally in this research (ALIGHT experiment) had the best performance when compared to the LIGHT and CTRL experiments. In general, the ALIGHT experiments obtained the best results in representing the volume of precipitation in the domain, followed by the LIGHT experiments (Figure 4.10).

In fact, it is possible to see that the assimilation of lightning data improved the representation of the precipitation field in almost all the simulation cycles during the passage of the Cold Front.

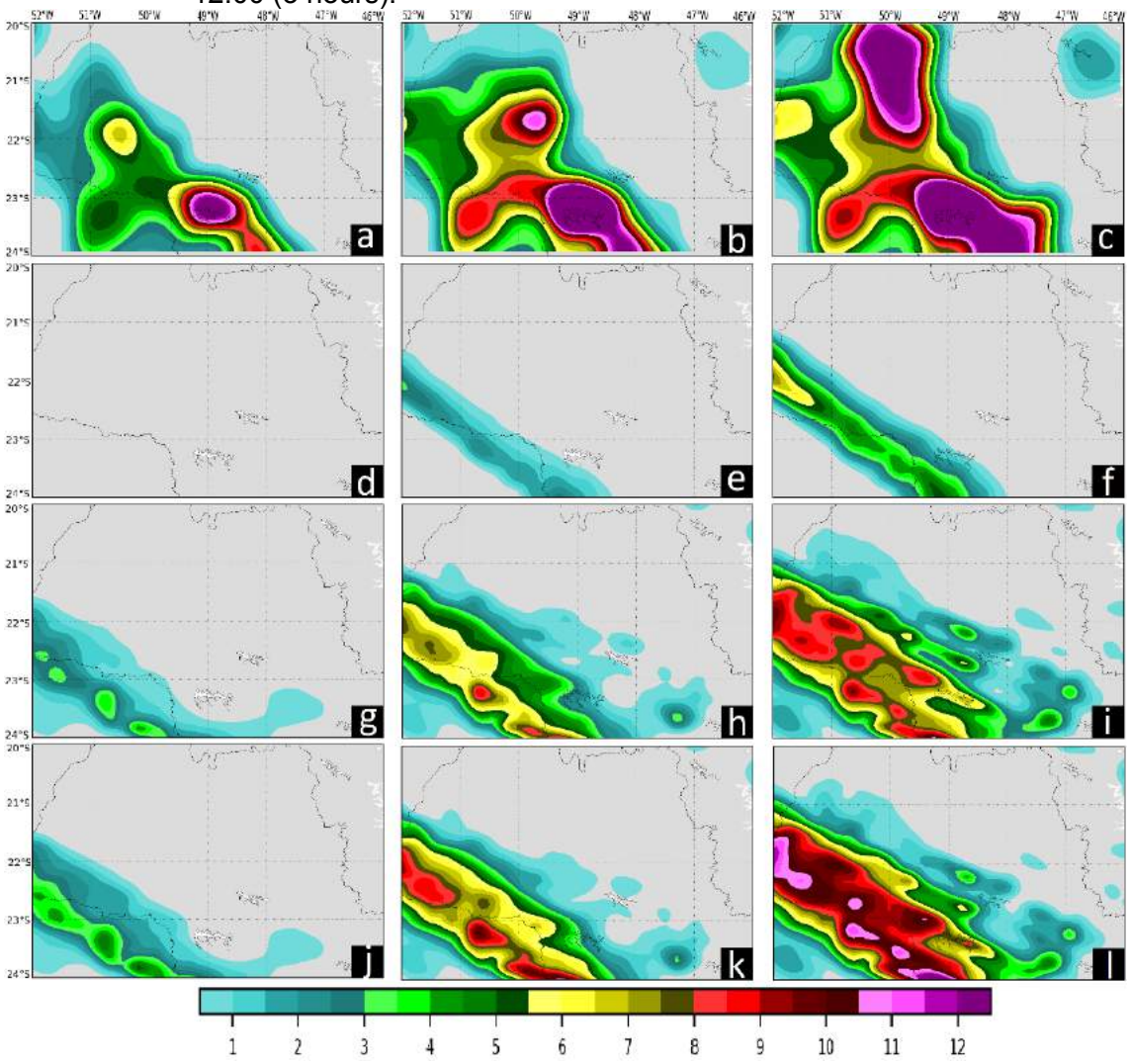
Figure 4.10 – BIAS for the 3 hours accumulated precipitation calculated for each simulation cycle with the dashed line showing the correspondent average.



The algorithm applied in the ALIGHT experiment added more water vapor in the cycles iv and v (Figure 4.7) than the LIGHT experiment accelerating, even more, the model response and inducing the development of more convection and precipitation in the first hours.

In order to analyze spatially the impact of the data assimilation procedures, the accumulated precipitation is presented in Figure 4.11 for every hour of simulation in cycle-iv. The CTRL experiment was not able to reproduce the precipitation field properly, producing some precipitation only after 1 hour of simulation. The precipitation field is not aligned with the meteorological system that was moving faster than the model could reproduce. A fast-moving system like squall lines is hard to simulate, especially in the first hours.

Figure 4.11 – Accumulated precipitation in mm in the inner domain for the cycle-iv (09:00 UTC 19/05/2018). (a), (b) and (c) present the observed precipitation; (c), (d) and (f) present the CTRL experiment; (g), (h) and (i) the LIGHT experiment; (j), (k) and (l) the ALIGHT experiment. (a), (d), (g) and (j) present the accumulated precipitation at 10:00 (1 hour of simulation); (b), (e), (h) and (k) at 11:00 (2 hours); (c), (f), (i) and (l) at 12:00 (3 hours).

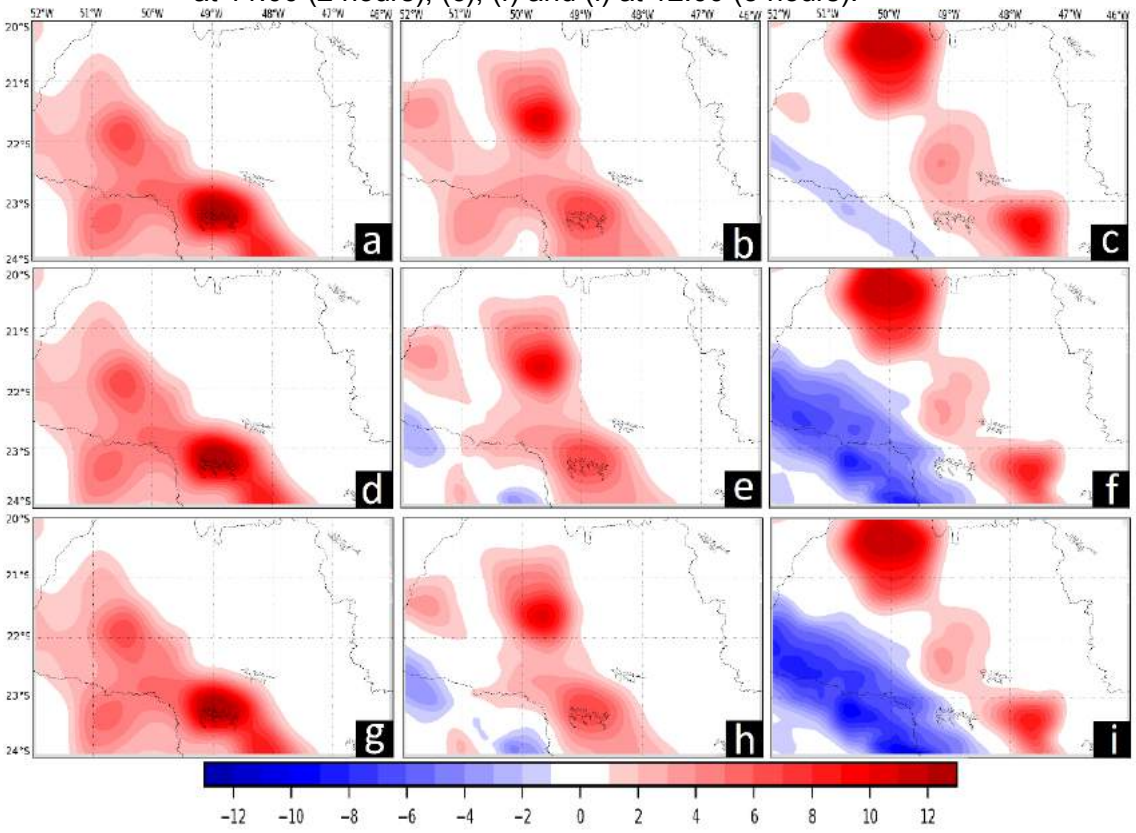


The LIGHT and ALIGHT experiments started to produce precipitation from the very first hour evolving to intense precipitation in the following hours. Even though these experiments showed the squall line moving slower than the observations indicated, the amount of rain simulated was more compatible with the observations than the CTRL experiment.

Note that at 12:00 UTC (3 hours of simulation) the ALIGHT experiment was the only experiment capable of to reproduce cores with more intense rainfall (above 11mm) closer to what was observed (Figure 4.11).

The difference between the precipitation field observed and simulated is shown in Figure 4.12.

Figure 4.12 – Difference between the accumulated precipitation observed and simulated in mm in the inner domain for the cycle-iv (09:00 UTC 19/05/2018). (a), (b) and (c) present the observation minus CTRL experiment; (c), (d) and (f) present the observation minus LIGHT; (g), (h) and (i); the observation minus ALIGHT. (a), (d), and (g) present the accumulated precipitation at 10:00 (1 hour of simulation); (b), (e) and (h) at 11:00 (2 hours); (c), (f) and (i) at 12:00 (3 hours).



The desynchronization between the simulated fields and the observations is clear when Figure 4.12 is analyzed. The red color indicates that the observation registered more precipitation than what it was reproduced by the simulations while the blue color indicates that the precipitation was overestimated by the model. In all experiments in cycle-iv, the band of precipitation associated with the squall line was always ahead than what the model was capable of to produce.

Ahasan and Debsarma (2015) analyzing a squall line that affected a region of India also applied assimilation techniques in the WRF model in order to reproduce more precisely this extreme event. They noticed that even for this

unusual meteorological system the WRF model was able to simulate the synoptic environment properly with an improvement in the experiments that used data assimilation procedures. However, they also noticed some spatial and temporal biases in the representation of the squall line.

The same timing and positioning biases were also observed by SU et al. (2016) when they were simulating a squall line event that occurred in the east of China with the WRF model. According to them, the cold pools play an important role in the development and evolution of this type of system which makes the representation of them very important to simulate the squall line properly. The correct representation of these atmospheric features can improve the short-term forecasting by inducing deep convection (MANSELL; ZIEGLER; MACGORMAN, 2007).

Fierro et al. (2012) also notice that the representation of these cold pools in the analysis time can improve the forecast, especially when the mesoscale environment is not well simulated by the model.

Although, the results also presented the same timing and positioning errors observed in other studies, the use of lightning data in the assimilation process still brought positive results. The addition of water vapor in the time analysis induced deep convection in the first hours of simulation which is rarely reproduced by the model without any type of data assimilation process.

In order to quantify the positioning improvement of the system caused by the use of the lightning data assimilation methodology, POD, FAR and TS were calculated for different precipitation thresholds (see section 3.6.1). This calculus was applied for all cycles as it is shown by Tab. 4.1.

The first threshold (30km and 1mm) shows the capability of the model in reproducing precipitation. Note that in both experiments with lightning data assimilation (LIGHT and ALIGHT) the probability of detection was improved significantly while the false alarms increased by a smaller rate, improving the ratio POD/FAR. However, for the precise quantification of how much the use of lightning data improved the positioning of the precipitation, it is necessary the

analyses of more events since in multiple cases in the CTRL experiment was not possible to calculate FAR.

Table 4.1 – Table showing POD, FAR and TS variables calculated based on the 3 hours accumulated precipitation simulated in each cycle for all three experiments (CTRL, LIGHT and ALIGHT) with different thresholds (30km and 1mm, 20km and 5mm, and 20km and 10 mm). Null values mean that it was not possible to calculate since there was no precipitation.

Experiment	*Simulation Cycle*	*30km and 1mm*			*20km and 5mm*			*20km and 10mm*		
		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 pattern remained the same for other thresholds with the lightning data assimilation experiments improving the variables POD and TS while the variable FAR also increased by a smaller rate. It was not observed significant differences between experiments LIGHT and ALIGHT.

In order to understand how the assimilation of lightning data affects a medium-range forecast, a 24 hours forecast was performed. Based on the results obtained above, the cycle-iv was selected to proceed with this simulation, i.e., it was performed a 24 hours simulation starting at 09:00 UTC on 19/05/2018 and ending at 09:00 UTC on 20/05/2018. The precipitation rate for the simulation period can be seen in Figure 4.13.

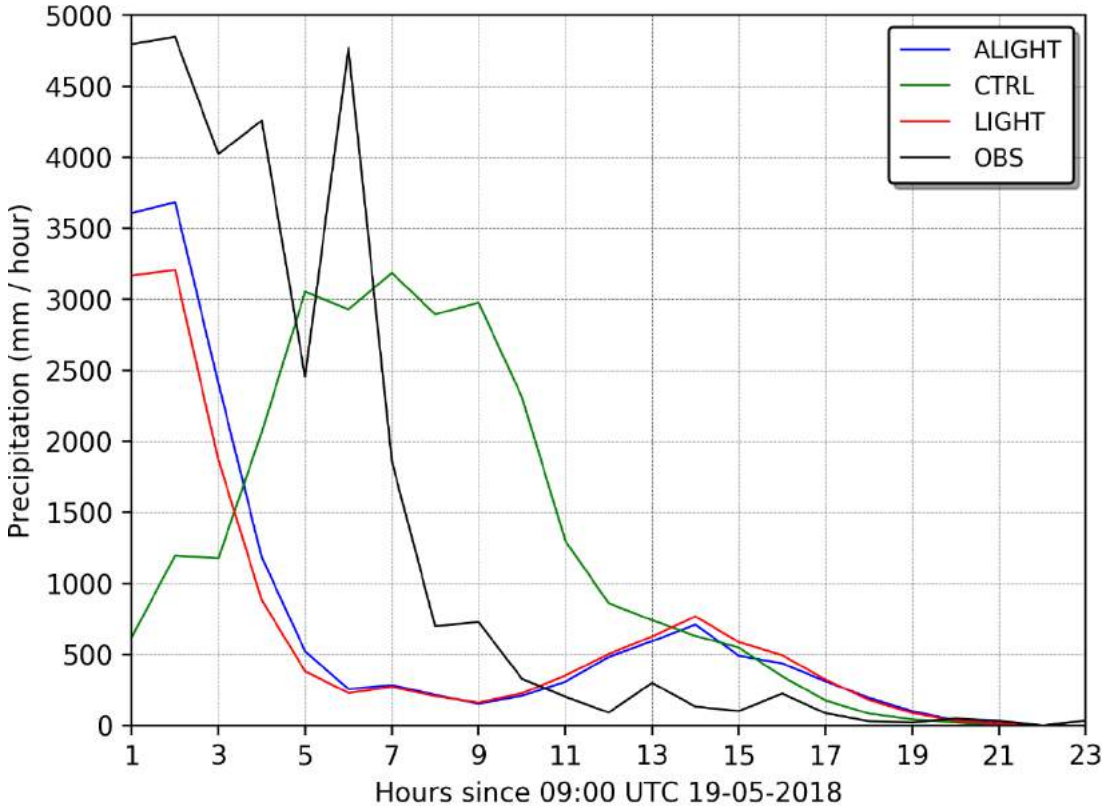
The precipitation rate simulated by the experiments LIGHT and ALIGHT showed better results in the first hours and even after this period the pattern observed in these experiments suggested that they kept reproducing the meteorological environment more accurately. The CTRL experiment seemed to have difficulties in following the precipitation rate observed with a delayed response, especially in the first half of the simulation. The correspondent BIAS of this 24 hours experiment can be seen in Figure 4.14.

In fact, when the precipitation rate BIAS is analyzed, it is clear that the experiments with lightning data assimilation reproduced the amount of rain better than the CTRL experiment in the first hours. However, the better BIAS observed between 4 and 7 hours of simulation in the experiment CTRL might not represent a real improvement. As it was discussed above, this improvement might be only a delayed response from the initial conditions that crossed the observation line which caused the small BIAS observed (Figure 4.13 and 4.14).

Fierro et al. (2015) also observed an improvement in the performance of the accumulated precipitation forecast in the first hours of simulation during high-impact weather events after analyzing 67 cases.

According to Wang et al. (2017), the algorithm developed by Fierro et al. (2012) is an efficient assimilation method for initializing convection where lightning data was registered. Nevertheless, it has limited to suppress spurious convection or modulate the convection.

Figure 4.13 – Precipitation rate simulated and observed from 09:00 UTC 19/05/2018 to 08:00 UTC 20/05/2018.



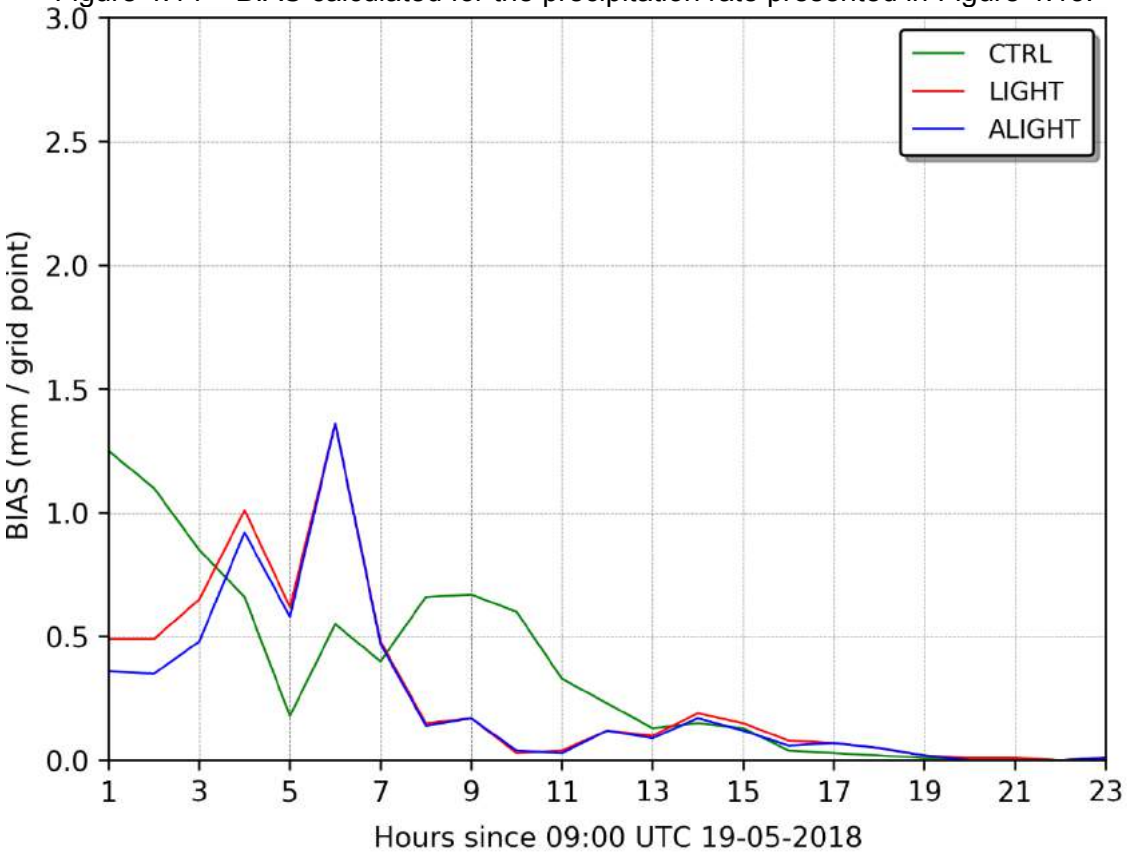
In fact, the lightning data assimilation algorithm was responsible for generating spurious convection in the simulation which was one of the causes of errors in these experiments limiting the forecast range where the use of lightning data in the analyses time is beneficial. The spurious convection observed in the experiments could have been caused by instabilities created by the insertion of mass in the analyses time. The addition of mass in the simulations can contribute to the initial mass/wind imbalances and subsequent generation of inertia-gravity waves during the integration of the model equations (PECKHAM et al., 2016). One way to reduce these imbalances is by using digital filters (HARTER, 1999; LYNCH; XIANG-YU, 1992).

Despite the generation of spurious convection, the experiments with lightning data assimilation had a better performance when we consider a short-term forecast with the ALIGHT experiment having the best results.

The experiments with data assimilation were also able to reproduce the precipitation location more accurately than the experiment CTRL. Figure 4.14

shows that POD and TS were improved in the first 4 hours of simulation while FAR remained smaller than CTRL experiment for the detection threshold (30km and 1mm). It was not observed significant differences between the experiments LIGHT and ALIGHT.

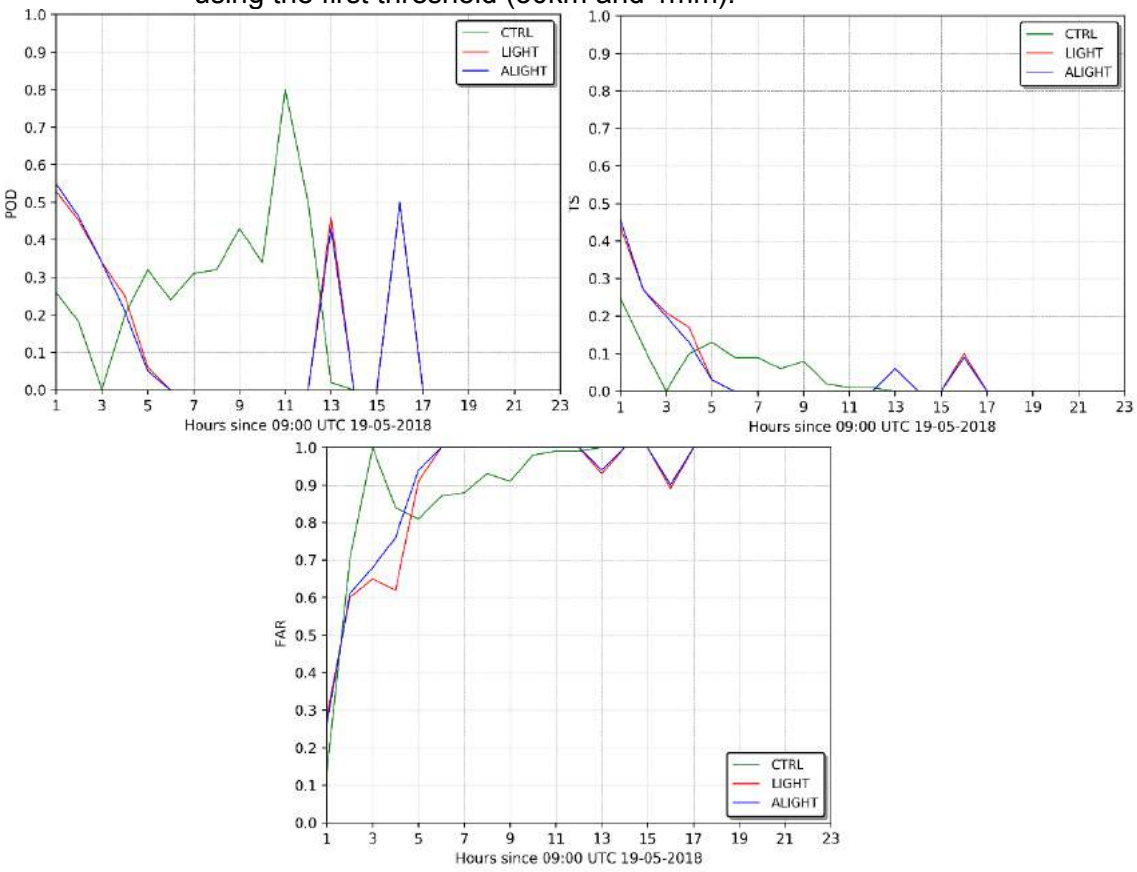
Figure 4.14 – BIAS calculated for the precipitation rate presented in Figure 4.13.



Wang et al. (2017) also observed better results for the accumulated precipitation in the first 6 hours of simulation for the experiment with lightning data assimilation using the detection threshold for the Frequency Skill Score. This means that the experiment using lightning data was capable to detect better regions with precipitation.

Note that all the 24 hours experiments (CTRL, LIGHT, and ALIGHT) tend to converge to the same result after about 17 hours of simulation which can indicate that the lightning data assimilation process has a small impact in medium and long-range forecasts.

Figure 4.15 – POD, TS and FAR for the precipitation rate presented in Figure 4.13 using the first threshold (30km and 1mm).



The use of data assimilation cycles at the beginning of the simulation (Dynamic Initialization/Warm Start) to initialize the model could be an option to improve the medium-range forecasting. However, precautions are needed since the Dynamic Initialization by nudging is capable of to create unbalanced fields in the initial conditions which can cause not only errors in the simulations but also numerical instabilities (HOKE; ANTHES, 1976).

Comparing the geometric center of lightning detections with the center of mass of the Reflectivity field in 3km of height simulated by the experiments with the WRF model, it is possible to track the displacement of the simulated storms and analyze if the observations were in agreement with what was simulated (Figure 4.16).

Figure 4.16 – Representation of the location of the center of mass calculated based on the Reflectivity field in 3 km simulated by the experiments and of the geometric center based on lightning detections for the period between 09:00 UTC 19/05/2018 and 09:00 UTC 20/05/2018. The direction of movement is from light to dark colors.

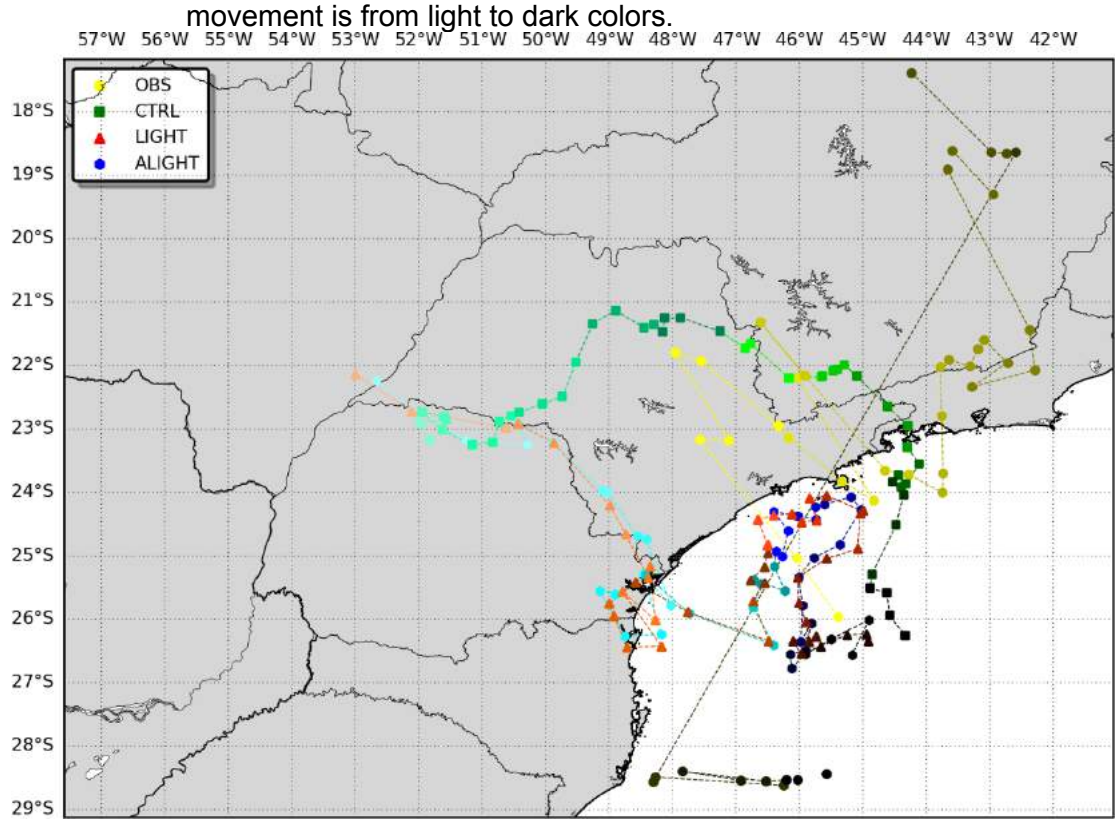
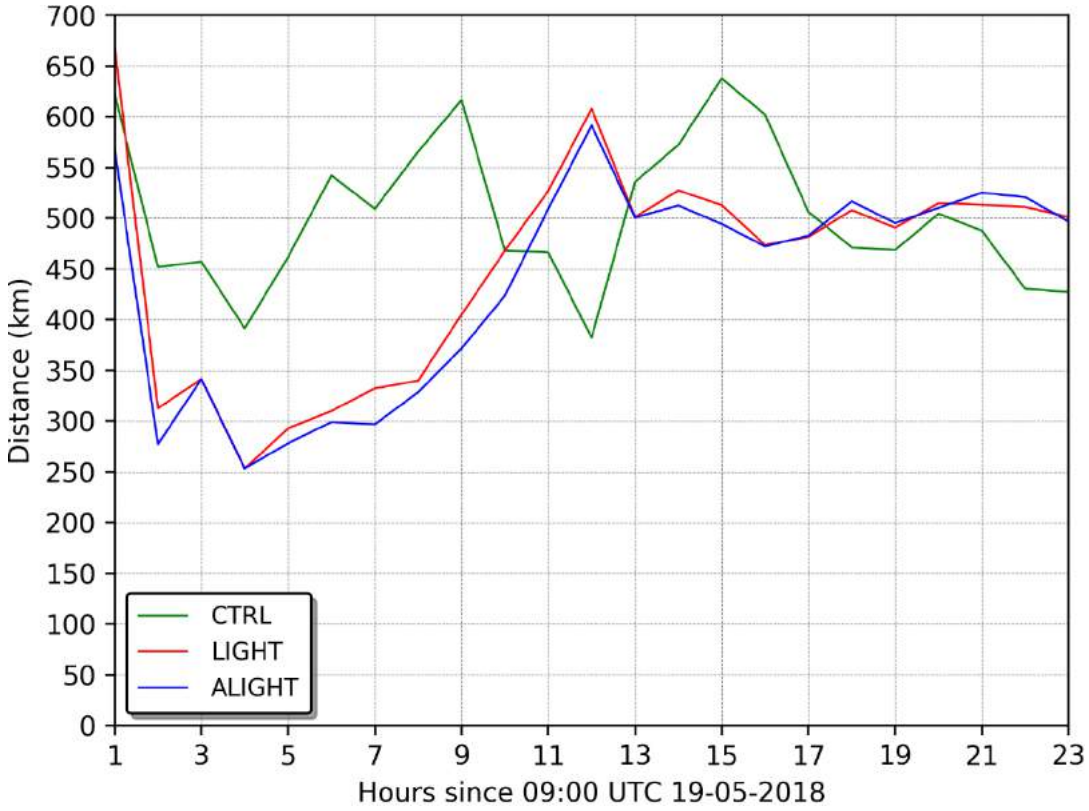


Figure 4.16 shows the center of mass of the Reflectivity field simulated by the experiments as well as the geometric center of lightning detections moving northeastward. The analyses of the location error associated with the difference between the simulations and the observations show that the experiments with data assimilation had better performances.

Even though the precipitation rate was improved only in the first 4 hours of simulation, the capability of the experiments with lightning data assimilation in representing the location and displacement of the meteorological system that affected the study area remained better than the CTRL experiment up to 10 hours of simulation with no significative difference between the experiments after that (Figure 4.17).

Figure 4.17 – Location error of the center of mass in relation to the geometric center based on lightning observations for each experiment between 09:00 UTC 19/05/2018 and 09:00 UTC 20/05/2018.



Based on the analyses of the case occurred on 19/05/2018 presented in this section, the implementation of the lightning data assimilation system improved the representation of the precipitation fields as well as the location of the meteorological system, especially in the first hours of simulation (short-term forecast). Additionally, it is also possible to conclude that the algorithm developed in this research (an adaptation of the original algorithm developed by Fierro et al. (2012)) applied in this case showed the best results in general when compared to other experiments.

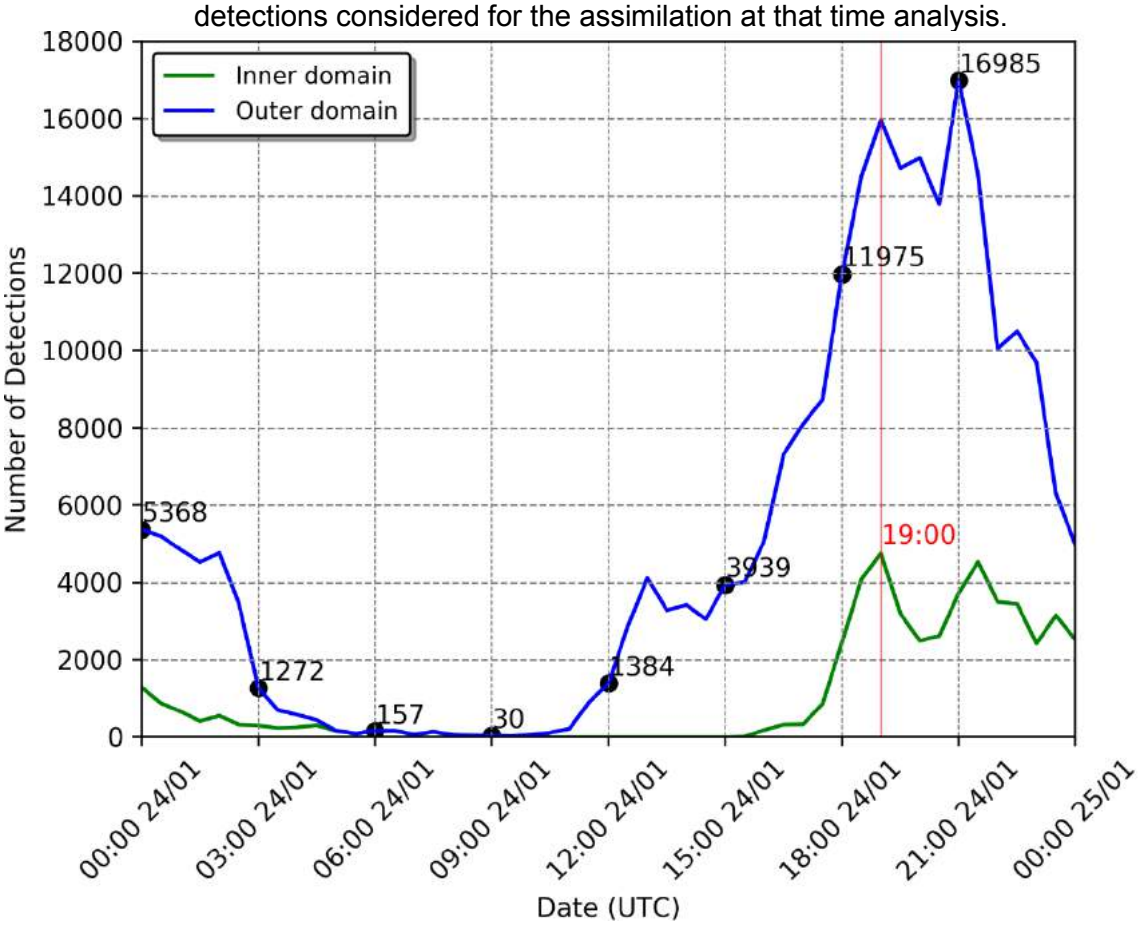
4.2 Case study II

4.2.1 Observational analysis

The second case analyzed occurred on January 24th, 2018 and it was associated with a complete different atmospheric environment. In this case, the convective activity was predominantly produced by thermal forcing which generated many local convective systems not directly related to the large-scale

environment. Even though the meteorological systems were caused by the local physics the quantity and the intensity of these individuals systems produced a higher incidence of lightning in the study area when compared to the previous case (Figure 4.18).

Figure 4.18 – Number of Lightning Detections per 30 minutes in the inner and outer domains on 24/01/2018. The black dots over the blue line and the numbers associated with them indicate the number of lightning detections considered for the assimilation at that time analysis.



Since the thunderstorms were triggered by the intense daytime heating the peak of lightning detections was observed at 21:00 UTC (at 19:00 UTC in the inner domain) which is expected for storms modulated by this type of physics. Note that the green and blue lines in Figure 4.18 follow the same behavior meaning that all the storms in the region were triggered by the same mechanism. In the previous case, the Cold Front, moving from South to North, started to produce lightning in the outer domain first until it reaches the inner domain (Figure 4.1).

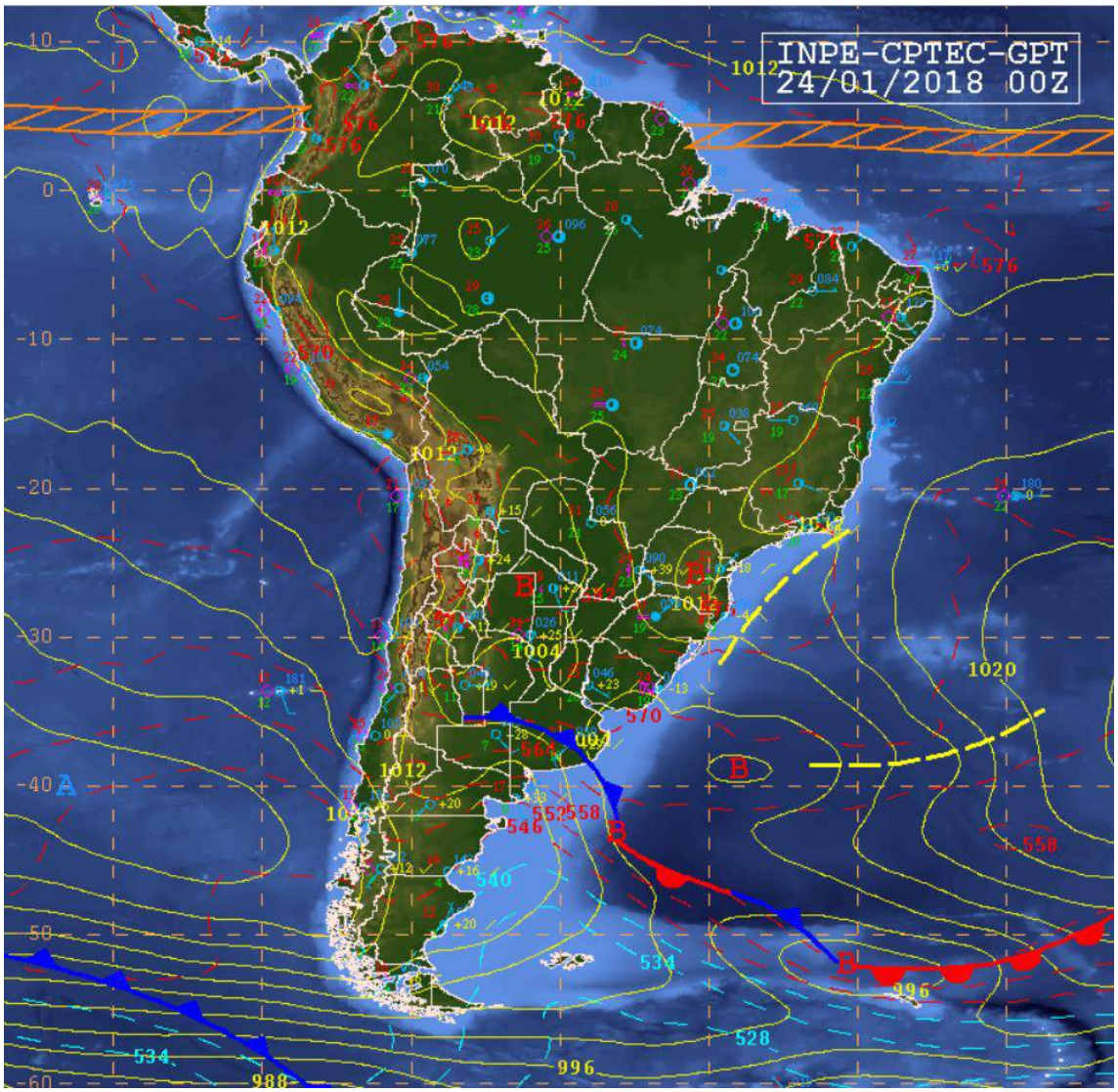
Analyzing the synoptic charts, it is also possible to notice that the configuration of the atmosphere is different from what was observed before (Figure 4.19 and 4.20).

At the 250hPa level, it is possible to see the Bolivian High displaced from its climatological position affecting part of the central and northern of the continent (Figure 4.20). The Bolivian High is a high-pressure atmospheric system that occurs at high levels of the atmosphere and its origin can be explained by different factors (DIAS; SCHUBERT; DEMARIA, 1983). The Andes mountains and the altitude of the Bolivian Altiplano has an important role in this process forcing the air coming from the east to ascend creating a divergent airflow and, consequently, a high-pressure system at higher levels of the atmosphere. Also, the intense convection in the Amazon rainforest releases large amounts of heat (latent heat) at lower levels of the atmosphere contributing to the high-pressure system at higher levels. Due to the atmospheric flow, eventually, the Bolivian High can be displaced from its typical position, as it is observed on this specific day.

Another import large-scale system to be noticed is the Upper Tropospheric Cyclonic Vortex (UTCV) over the Northeast of Brazil (FERREIRA; MELLO, 2005; KOUSKY; GAN, 1981). The Bolivian High is one of the factors that contribute to the formation of UTCV. This high-pressure system over Bolivia curves the atmospheric flow counterclockwise and by the conservation of absolute vorticity it generates a clockwise atmospheric flow ahead (SILBERMAN, 1954). The UTCV is a low-pressure system and it usually generates rainfall at the southwest of its center (due to the convergent airflow bringing moisture from the ocean).

This configuration works as an atmospheric blocking preventing the Cold Fronts coming from the south to reach the Brazilian southeast and facilitating the convection associated with atmospheric thermodynamics in the region (NASCIMENTO, 1998; REBOITA et al., 2010).

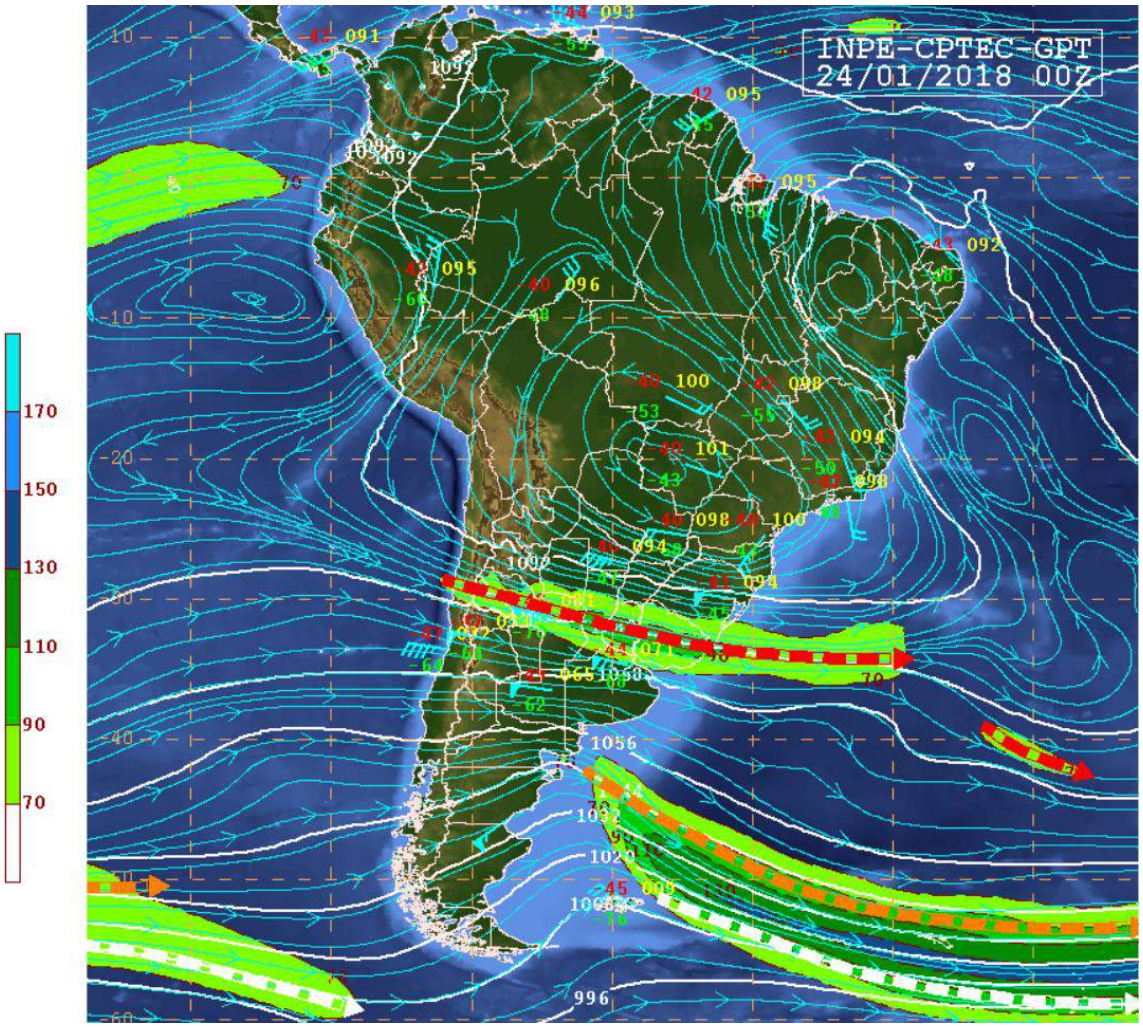
Figure 4.19 – Synoptic chart at 00:00 UTC on 24/01/2018 showing the synoptic environment on the surface in South America.



Source: CPTEC (2019).

The spatial distribution of flash density at its maximum activity in the inner domain can be seen in Figure 4.22. Note that there many individual storms generating lightning uniformly in the whole domain which is commonly observed when the daytime heating is the main factor generating atmospheric instability and storms.

Figure 4.20 – Synoptic chart at 00:00 UTC on 24/01/2018 showing the synoptic environment in 250hPa in South America.



Source: CPTEC (2019).

The satellite images in Figure 4.22 show the convection affecting the entire domain but with more intensity on the Rio Grande do Sul and Santa Catarina states. In Figure 4.22a the meteorological system is in its dissipation phase since it was formed by the daytime heating of the previous day. In Figure 4.22c the energy of the daytime heating started to produce clouds and storms once again.

Comparing with the previous case analyzed, it is obvious the difference between both. The convection activity, in this case, was predominantly induced by the local thermodynamics due to the intense daytime heating of the surface generating a higher number of lightning detections where most of it was

observed at the end of the day. Therefore, it is expected that the system presents a relative stationary behavior with more local and concentrated rainfall.

Figure 4.21 - Flash density for the peak of Lightning Detections inside the inner domain (Figure 4.18) for a time window of 30 minutes, from 18:45 to 19:15 UTC 24/01/2018.

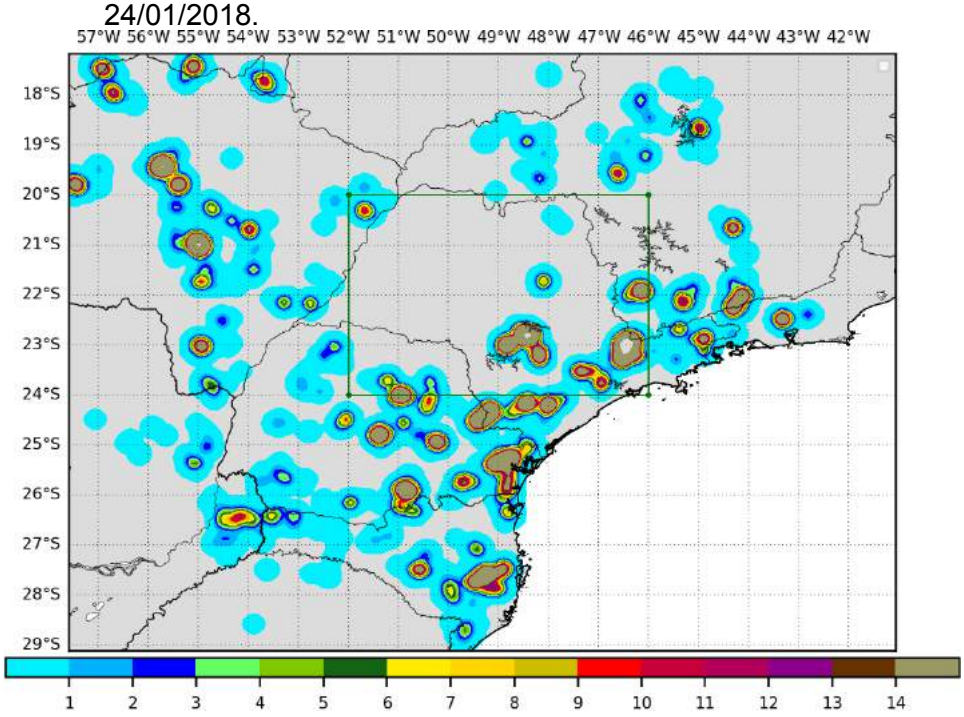
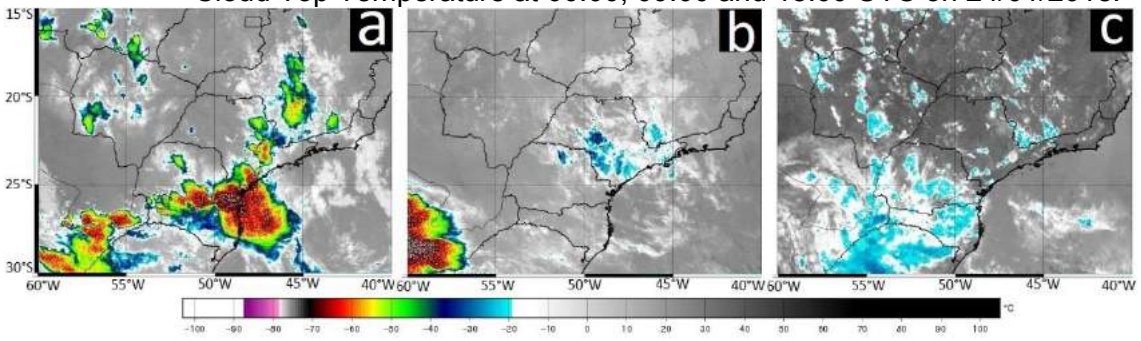


Figure 4.22 – Images from GOES-16 satellite of the channel 7 (3.90μ) showing the Cloud Top Temperature at 00:00, 09:00 and 18:00 UTC on 24/01/2018.



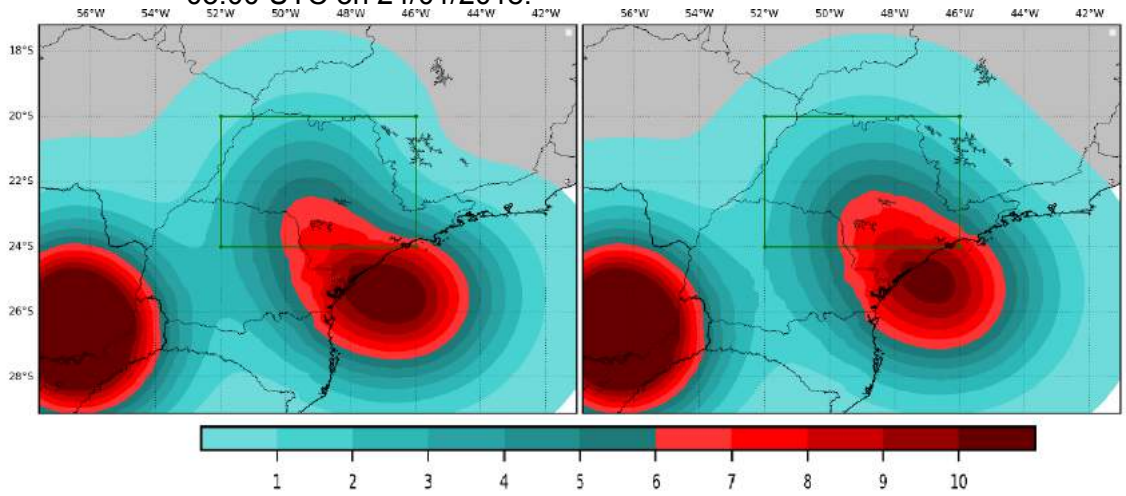
Source: DSA/CPTEC (2019).

By analyzing many short-term forecasts throughout the day in two different cases with synoptic environments completely distinct from each other it was expected to obtain a definitive conclusion about the use of lightning data in the WRF model.

4.2.2 Assimilation algorithm impact

The impact of the lightning data assimilation algorithm on the experiment performed for 24/01/2018 can be seen in Figure 4.23 which shows the Q_v increment at 19:00 UTC.

Figure 4.23 – LIGHT (left) and ALIGHT (right) experiments showing Q_v increment in g/kg (Analysis minus CTRL) integrated vertically in the atmosphere at 03:00 UTC on 24/01/2018.



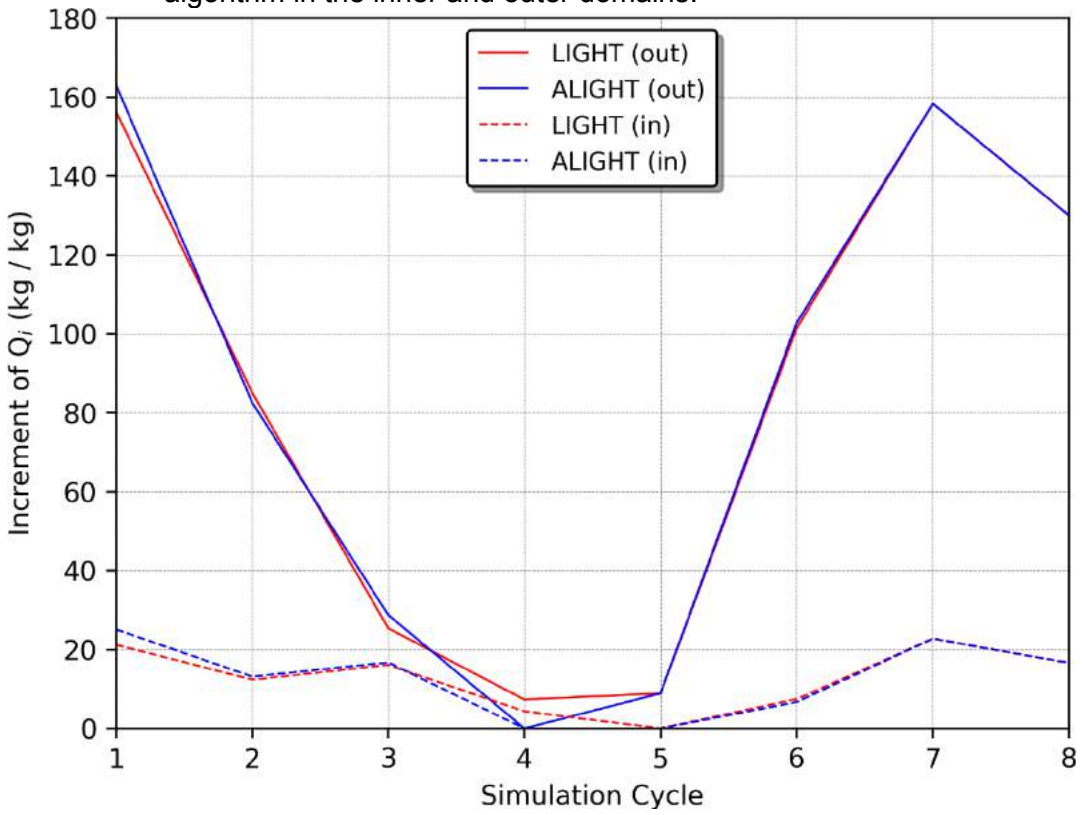
During the lightning activity periods (Figure 4.17), the assimilation algorithm, as expected, added more water vapor to the initial conditions when compared with the previous case analyzed.

In the first case, for the cycle-iv (09:00 UTC 19/05/2018), the number of lightning detections considered by the assimilation algorithm was 3265 (Figure 4.1) which caused the algorithm to insert an additional of about 70g of water vapor in the domain (Figure 4.7). Meanwhile, in the cycle-vi of the second case (18:00 UTC 24/01/2018), it was considered 11975 lightning detections (Figure 4.18) which added about 100g of water vapor (Figure 4.24).

Note that the amount of water vapor added also depends on factors other than the number of lightning detections. The previous state of the atmosphere as well as how lightning detections are grouped (which change the number of simulated observation stations and the values of flash density in each grid point) can also affect the water vapor increment. This is visible in the first cycle of this case where 5368 lightning detections caused a Q_v increment of about 160g,

possibly because the initial conditions before the use of the lightning data assimilation system was previously dry.

Figure 4.24 – Q_v increment for each simulation cycle performed on 24/01/2018. The chart shows the total amount of water vapor added by the assimilation algorithm in the inner and outer domains.

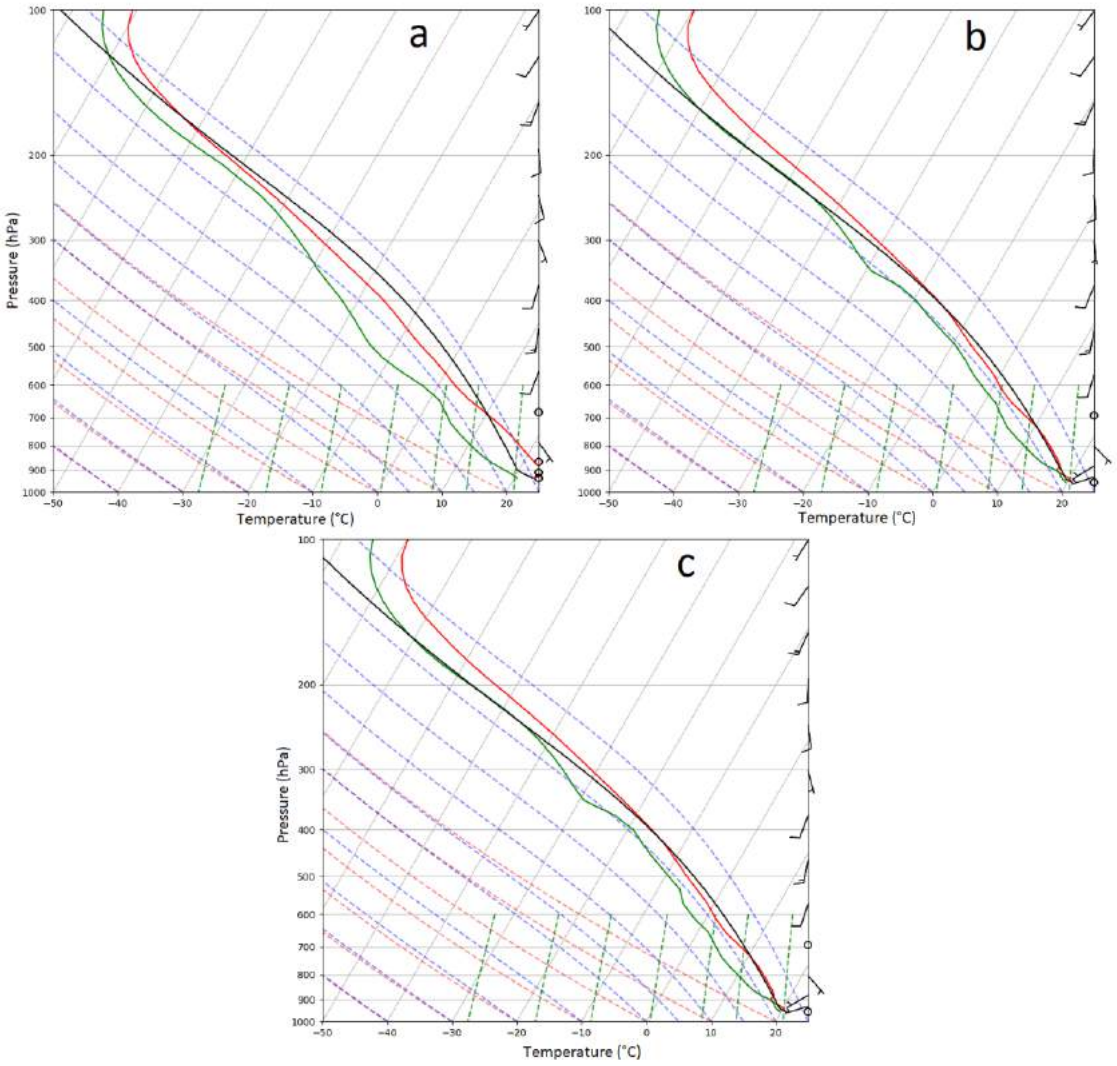


This additional water vapor added increased the atmospheric instability inducing convection motions and consequently more precipitation in the first hours of simulation (Figure 4.26).

In this case the vertical profile simulated by the model presented different physical characteristics (Figure 4.25). Even though the atmospheric instability was again increased, CAPE was drastically reduced after the assimilation process inducing the formation of isolated thunderstorm clouds.

It was not observed any significative difference between the LIGHT and ALIGHT experiment.

Figure 4.25 – Skew-T chart showing vertical profile of air temperature, dewpoint and wind at 04:00 UTC on 24/01/2018 (1-hour simulation) for the experiment CTRL(a), LIGHT(b) and ALIGHT(c) in the center of the domain.



4.2.3 Model performance

The analyses of the experiments performed on 24/01/2018 using lightning data to improve the initial conditions of the model did not have the same improvement observed in the analyses of the previous case. In general, the experiments LIGHT and ALIGHT overestimated the amount of precipitation in the study area. However, the positioning of the meteorological system was better represented than the CTRL experiment.

Unlike the first event analyzed, the lightning activity occurred on 24/01/2018 was mainly induced by daytime heating showing a peak of lightning detections at 21:00 UTC (Figure 4.18). This means that lightning activity was induced by

local convective systems with a smaller impact of the large-scale environment. As presented in the first case, the addition of water vapor also induced the development of instabilities and subsequent convection in the first hours of simulation, but the amount of rainfall generated by the set of many local convection systems was not properly represented by the experiments with lightning data assimilation.

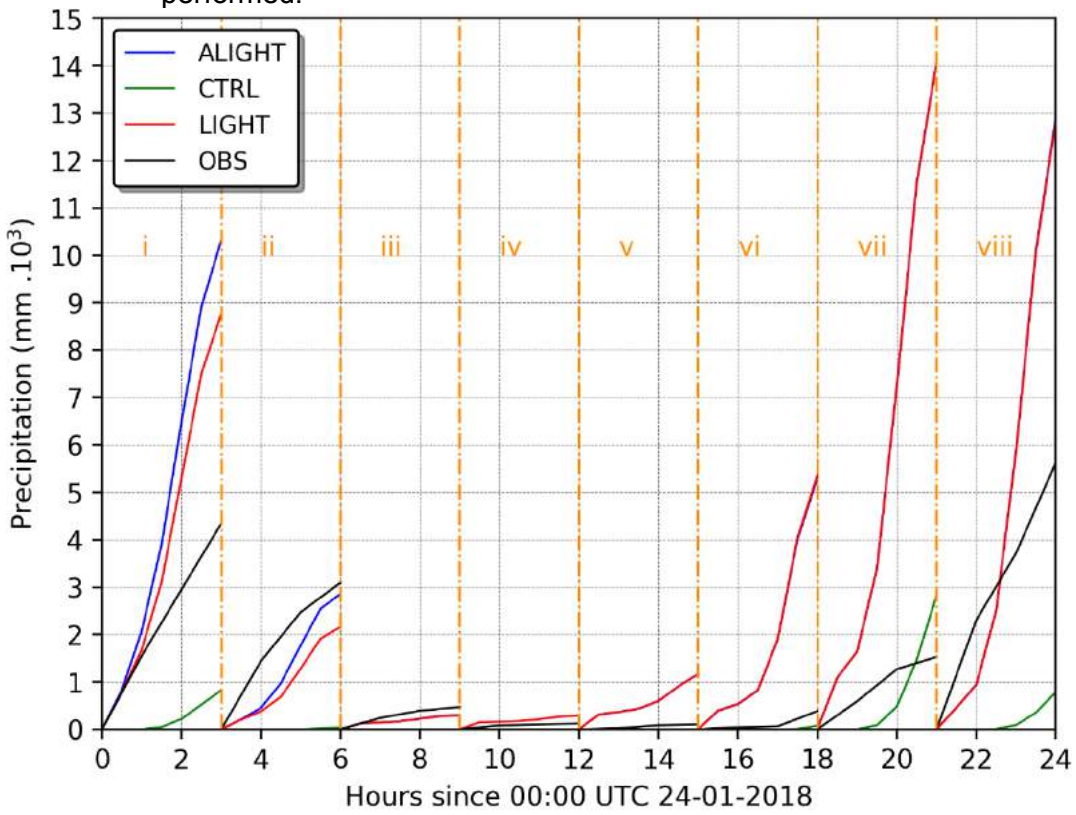
It was observed that these local convective systems that occurred in the simulation domain had a higher lightning activity when compared to the previous case analyzed. Even though the case occurred on 19/05/2018 registered a higher amount of rainfall it had a peak of about 6,000 lightning detections while the case occurred on 24/01/2018 with a smaller amount of rainfall reached more than 16,000 lightning detections (Figures 4.1 and 4.18). These differences between lightning detections and volume of rainfall ratios changed the performance of the lightning data assimilation algorithm implemented in this study.

Therefore, this case analyzed in this section was induced by thermal forcing causing a small amount of precipitation with a high lightning activity which could have caused the assimilation algorithm to add more water vapor than necessary making the experiments LIGHT and ALIGHT overestimate the accumulated precipitation.

The Figure 4.26 presents the sum of the precipitation field accumulated in the whole simulation domain for the CTRL, LIGHT and ALIGHT experiments and for the precipitation field observed for each one of the simulation cycles in the first 3 hours.

The overestimation of the volume of precipitation simulated by the experiments with lightning data assimilation could have been caused due to the poor representation of graupel concentrations in the initial time. Even though the experiments could solve the convection explicitly, they were not able to reproduce graupel in the atmosphere maybe because of the 9km resolution used in the simulations.

Figure 4.26 – The sum of the accumulated precipitation in the inner domain on 24/01/2018 for each simulation cycle and for each experiment performed.



Graupel has an important role in the development of a thunderstorm. The local convection systems observed commonly during the summer in the study area are usually associated with the daytime heating and present deep convection. This type of convection generates intense updrafts and consequently, a high concentration of graupel which is one the main factor responsible for the formation of lightning (BRINGI et al., 1997; DYE et al., 1988; ZIEGLER et al., 1991).

Additionally, Tao et al. (2013) observed that the storm's updraft is suppressed in the presence of large ice particles which can affect the dissipation time of the system and by consequence affecting the amount of precipitation generated.

Similar results were also observed by Adams-Selin et al. (2013). They concluded that simulations with parametrizations that allowed the formation of large graupel had a minimal stratiform precipitation region with reduced convective intensity. Meanwhile, the simulations with smaller graupel allowed

the formation of deeper and strong cold pools which caused a wide region with stratiform precipitation as well as persistent convection.

Those studies agree with what was observed in this section. The typical formation of graupel observed in local convective systems provided the necessary environment for the occurrence of lightning and induced a rapid dissipation of the systems agreeing with the high number of lightning detections (Figure 4.18) and with a small amount of observed rain (Figure 4.26).

Furthermore, it also explains the overestimation of the precipitation field simulated by the experiments with data assimilation. Since the model was not able to reproduce the graupel concentration in the atmosphere, the experiments with lightning data assimilation which depend on it added more water vapor than necessary. As it was showed by Figure 3.4, higher concentrations of graupel contributes to the addition of less water vapor in the atmosphere.

The Thompson Scheme used for Microphysics in the simulations (THOMPSON et al., 2008) was not enough to simulate graupel concentrations, even this parametrization contains prognostic equations for cloud water, rain water, ice, snow, and graupel mixing ratios. An investigation needs to be conducted in order to find the optimal parameterization to represent graupel concentrations, especially coupled with the lightning data assimilation algorithm developed in this study.

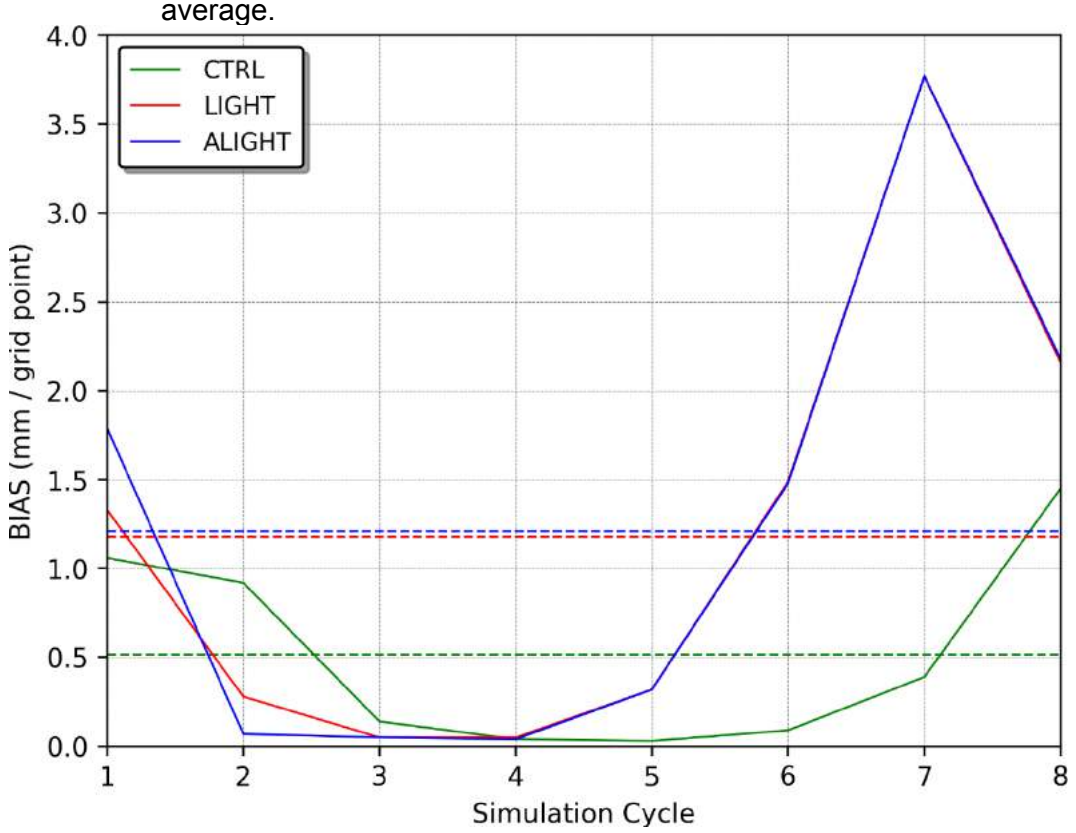
It is possible to notice in Figure 4.26 that the experiments with data assimilation overestimated by far the 3 hours accumulated precipitation in most of the cycles. However, note that the first 1 hour accumulated precipitation simulated by the experiments LIGHT and ALIGHT had a better performance.

The BIAS for the 3 hours accumulated precipitation is presented in Figure 4.27. The best performance of the assimilation algorithm was observed in the cycles ii, iii and iv. In general, the CTRL experiment had better performance.

Once more, the assimilation algorithm was not able to reproduce the amount of precipitation associated with thermal forcing and it becomes obvious when we notice that the worse performance in both cases analyzed (19/05 and 24/01)

occurred in the last simulations cycles possibly because of the poor representation of graupel mixing ratios, as discussed above.

Figure 4.27 – BIAS for the 3 hours accumulated precipitation calculated for each simulation cycle with the dashed line showing the correspondent average.

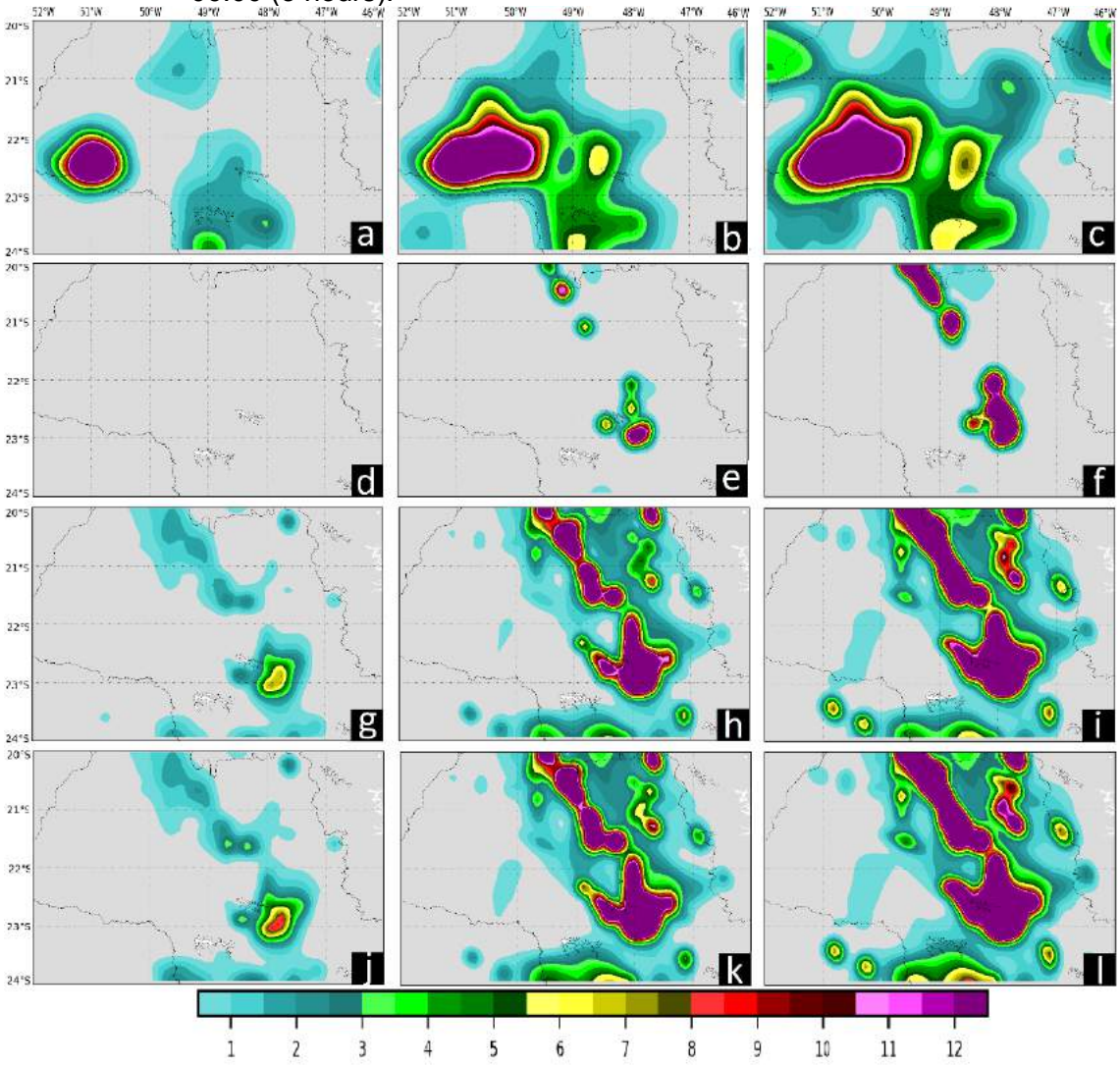


In this case, there was a small difference observed between the experiments ALIGHT and LIGHT due to the smaller volume of precipitation. The experiments had a worse representation of the 3 hours accumulated precipitation but the rapid response in the first hour of simulation indicated a positive impact.

In order to analyze spatially the impact of the data assimilation procedures, the accumulated precipitation for every hour of simulation in cycle-ii is presented (Figure 4.28). Even though the error was smaller than other experiments, the CTRL experiment could not reproduce the precipitation field properly producing some precipitation only after 1 hour of simulation. Additionally, the main precipitation core with a center located around 22.5°S and 50.7°W was not reproduced by any of the experiments.

This inability in reproducing the main precipitation core is more visible when Figure 4.29 is analyzed. In fact, there was a timing and a positioning error simulating the convective activity in this case.

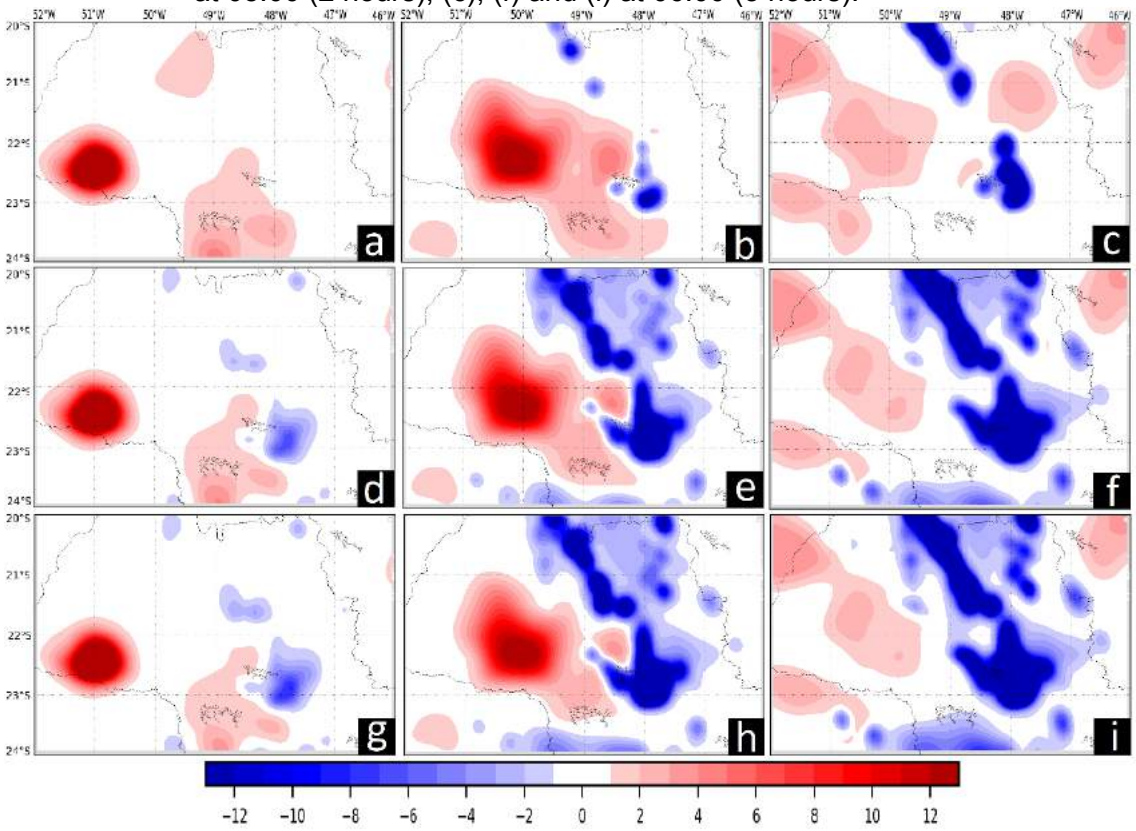
Figure 4.28 – Accumulated precipitation in mm in the inner domain for a cycle starting at 03:00 UTC on 24/01/2018). (a), (b) and (c) present the observed precipitation; (c), (d) and (f) present the CTRL experiment; (g), (h) and (i) the LIGHT experiment; (j), (k) and (l) the ALIGHT experiment. (a), (d), (g) and (j) present the accumulated precipitation at 04:00 (1 hour of simulation); (b), (e), (h) and (k) at 05:00 (2 hours); (c), (f), (i) and (l) at 06:00 (3 hours).



In general, the assimilation algorithm overestimated the precipitation field once it was mainly associated with daytime heating. However, the WRF model by itself might be also inducing the error associated with this type of atmospheric environment.

Davis et al. (2006) applying an object-based verification methodology in order to evaluate the forecast of rainfall during the warm season over the United States showed that the WRF model generally tends to overestimate the size of the rain areas during the day which can induce the model to overestimate the precipitation field.

Figure 4.29 – Difference between the accumulated precipitation observed and simulated in mm in the inner domain for a cycle starting at 03:00 UTC on 24/01/2018. (a), (b) and (c) present the observation minus CTRL experiment; (c), (d) and (f) present the observation minus LIGHT; (g), (h) and (i); the observation minus ALIGHT. (a), (d), and (g) present the accumulated precipitation at 04:00 (1 hour of simulation); (b), (e) and (h) at 05:00 (2 hours); (c), (f) and (i) at 06:00 (3 hours).



Even though the CTRL experiment was not able to generate a significative amount of precipitation in the first hours of simulation, the use of assimilation techniques reduced the spin-up period anticipating a natural overestimation response of the precipitation field from the model.

Similarly, to the previous case, the variables POD, FAR and TS were calculated for different precipitation thresholds (see section 3.6.1) in order to define

quantitively the position error of the precipitation field simulated by the experiments. This calculus was applied for all cycles as it is shown by Tab. 4.2.

Table 4.2 – Contingency table showing the variables POD, FAR and TS calculated based on the accumulated precipitation in 3 hours predicted for each simulation cycle (i, ii, iii, iv, v, vi, vii, viii) for all the experiments (CTRL, LIGHT and ALIGHT) with different thresholds (30km and 1mm, 20km and 5mm, and 20km and 10 mm) starting on 24/01/2018. For example, the cycle "iv" corresponds to the precipitation accumulated between 09:00 and 12:00 of 24/01/2018 with the difference of 3 hours between each cycle.

Experiment	*Simulation Cycle*	*30km and 1mm*			*20km and 5mm*			*20km and 10mm*		
		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

In all thresholds analyzed the probability of detection and the threat score variables were improved with a small increment in the false alarms for the experiments with data assimilation. However, in this case, null values were also observed affecting the precise quantification of these variables but not impacting in the final conclusion.

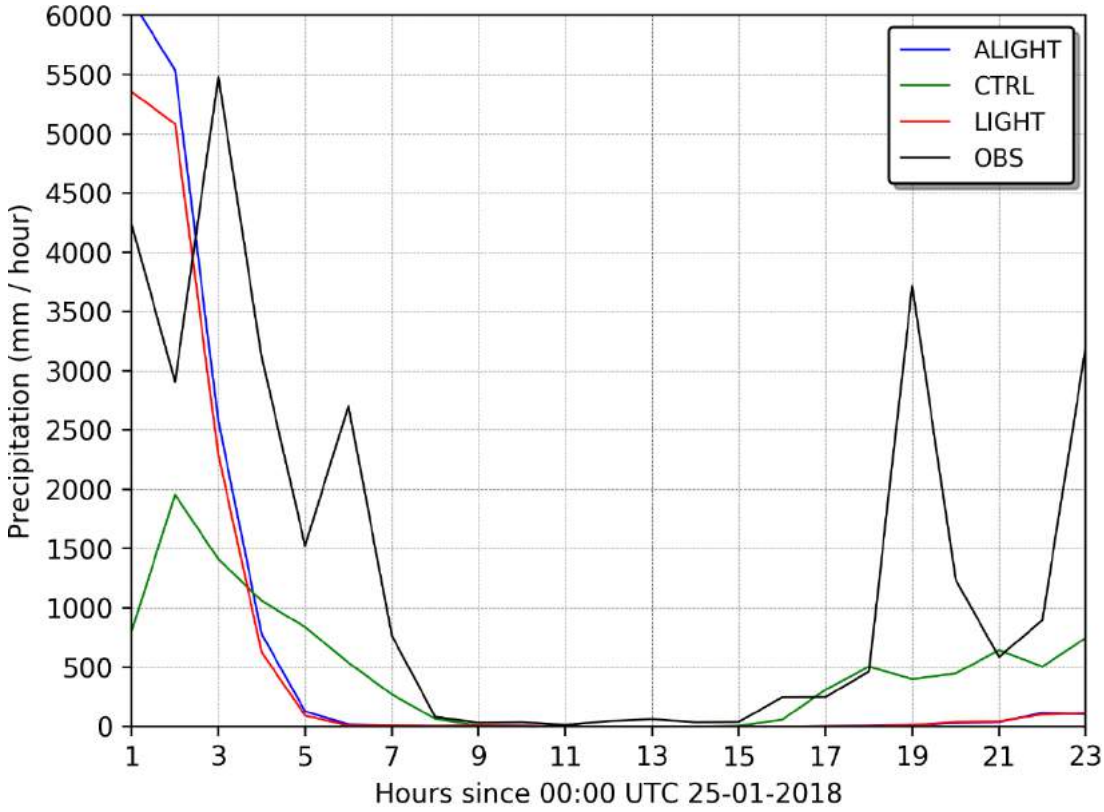
The ALIGHT experiment had the best performance in detecting precipitation increasing both POD and TS values and decreasing the false alarms when compared to the LIGHT experiment. When ALIGHT with CTRL are compared, it is visible that the probability of detection was drastically increased with a small increment in the false alarm variable.

In other thresholds, the experiments with lightning data assimilation proved to be better than the CTRL experiments improving the positioning of the precipitation field in the domain.

Analyzing the results obtained until now, a 24 hours run starting at 00:00 UTC on 25/01/2018 was performed. This simulation cycle was chosen in order to evaluate the differences between the experiments LIGHT and ALIGHT during the dissipation phase of the system and at the same time during a period with a significative amount of precipitation and lightning detections registered. The precipitation rate for this simulation period can be seen in Figure 4.30.

The precipitation rate simulated (the sum for the whole domain) by the experiments with data assimilation (LIGHT and ALIGHT) was improved in the first 3 hours of simulation while the CTRL experiment once again had a delayed response with all of them converging to the same solution after about 8 hours of simulation. In the last part of the simulation (after 16 hours of simulation), the CTRL experiment had a better response generating more precipitation and following more precisely the observations when compared to the LIGHT and ALIGHT experiments.

Figure 4.30 – Precipitation rate simulated and observed from 00:00 UTC 25/01/2018 to 23:00 UTC 25/01/2018.



The correspondent BIAS for the precipitation rate seems to indicate that the precipitation field in the data assimilation experiments was not aligned with the observations (Figure 4.31).

Although the BIAS variable was not well represented by the experiments LIGHT and ALIGHT (this variable is calculated locally), the variables POD, TS and FAR show an improvement in the representation of the precipitation field during all period of the simulation but especially in the first hours (Figure 4.32).

The variables POD and TS had a great improvement in the first hours with no additional false alarms when the experiments with data assimilation are compared with the CTRL experiment. Note that the ALIGHT experiment had the best performance in the first 4 hours of simulation increasing POD and TS, and decreasing FAR.

Figure 4.31 – BIAS calculated for the precipitation rate presented in Figure 4.30.

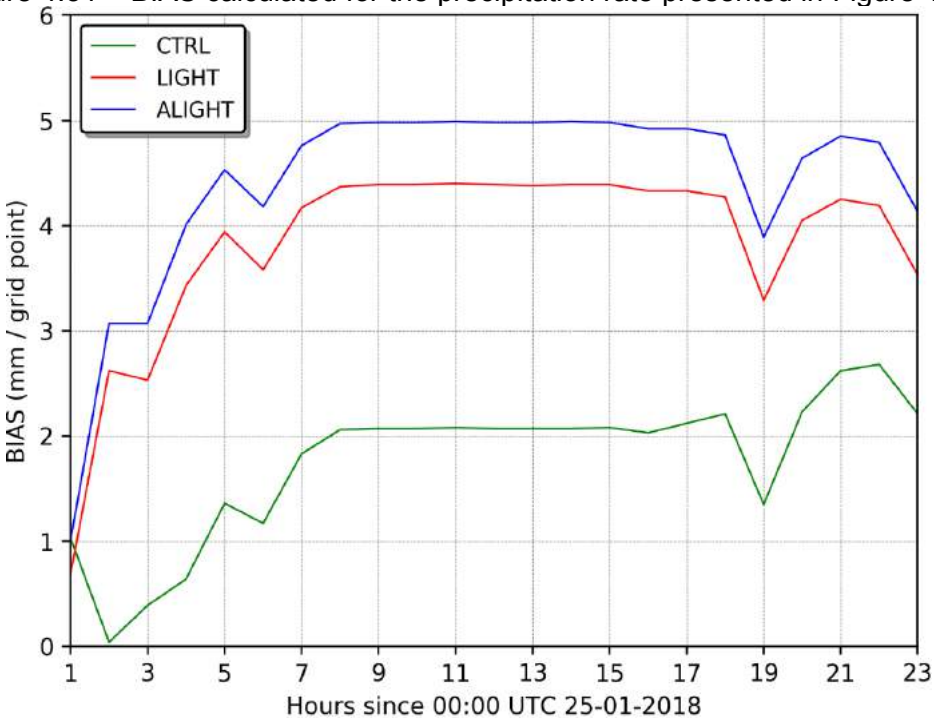
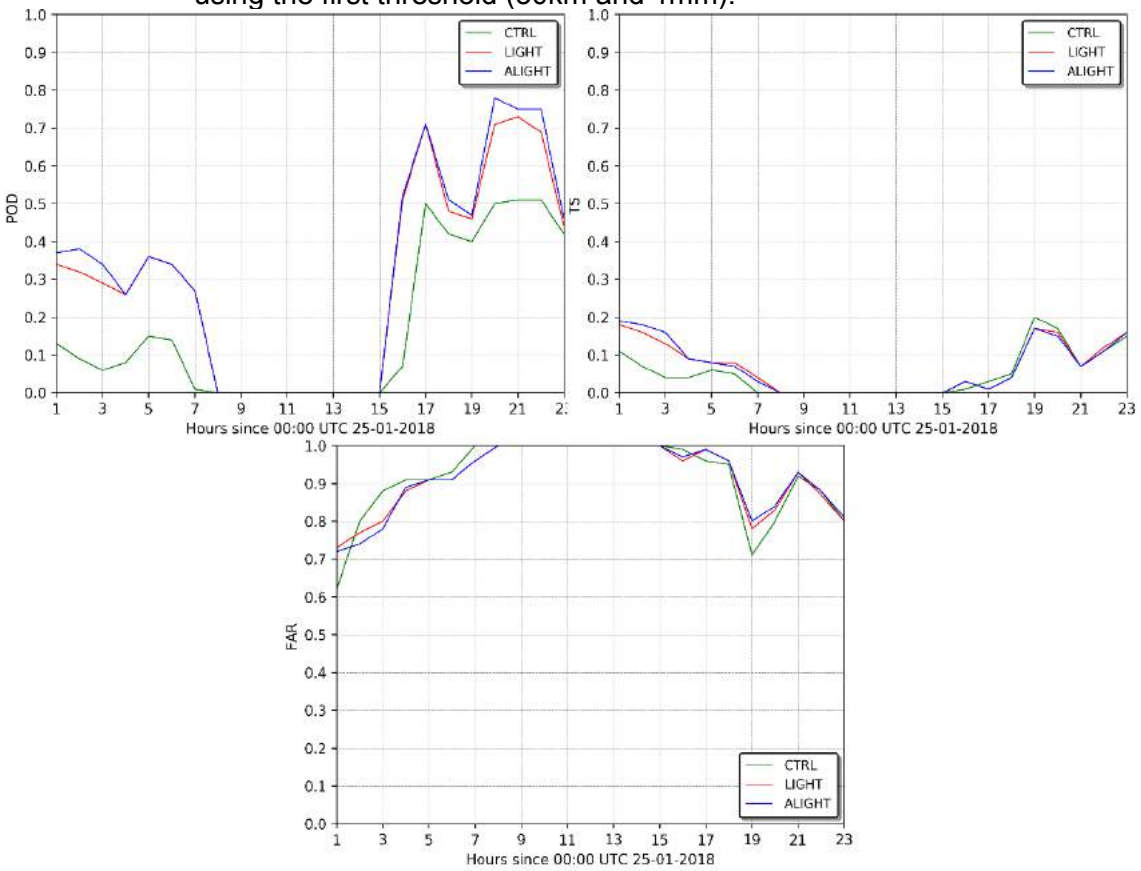


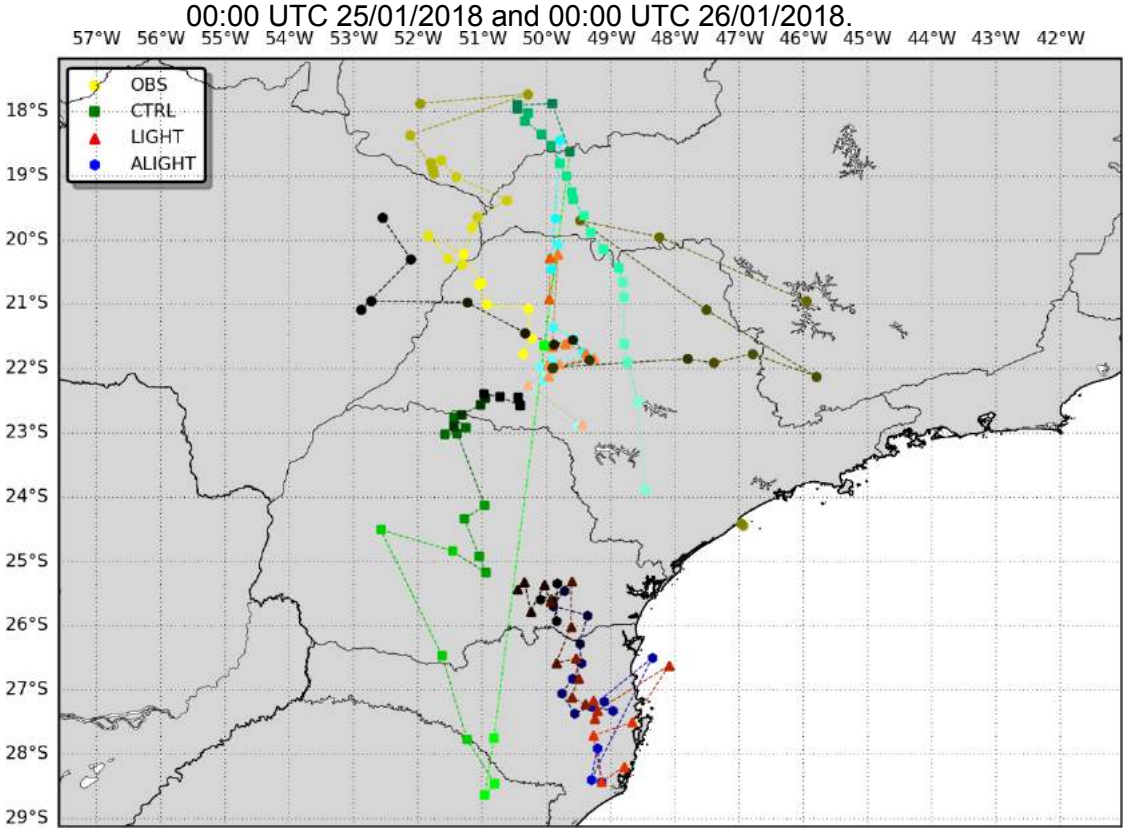
Figure 4.32 – POD, TS and FAR for the precipitation rate presented in Figure 4.30 using the first threshold (30km and 1mm).



In Figure 4.33 is possible to see the center of mass of the convection activity simulated by the model based on the Reflectivity field as well as the convection activity observed based on lightning detections. The simulation error correspondent to each experiment performed for this case can be seen in Figure 4.34.

Note that differently from the previous case, the experiments with lightning data assimilation were capable to improve the convection positioning only for the first 3 hours of simulation with no significative difference between each other.

Figure 4.33 – Representation of the location of the center of mass calculated based on the Reflectivity field in 3 km simulated by the experiments and of the geometric center based on lightning detections for the period between 00:00 UTC 25/01/2018 and 00:00 UTC 26/01/2018.

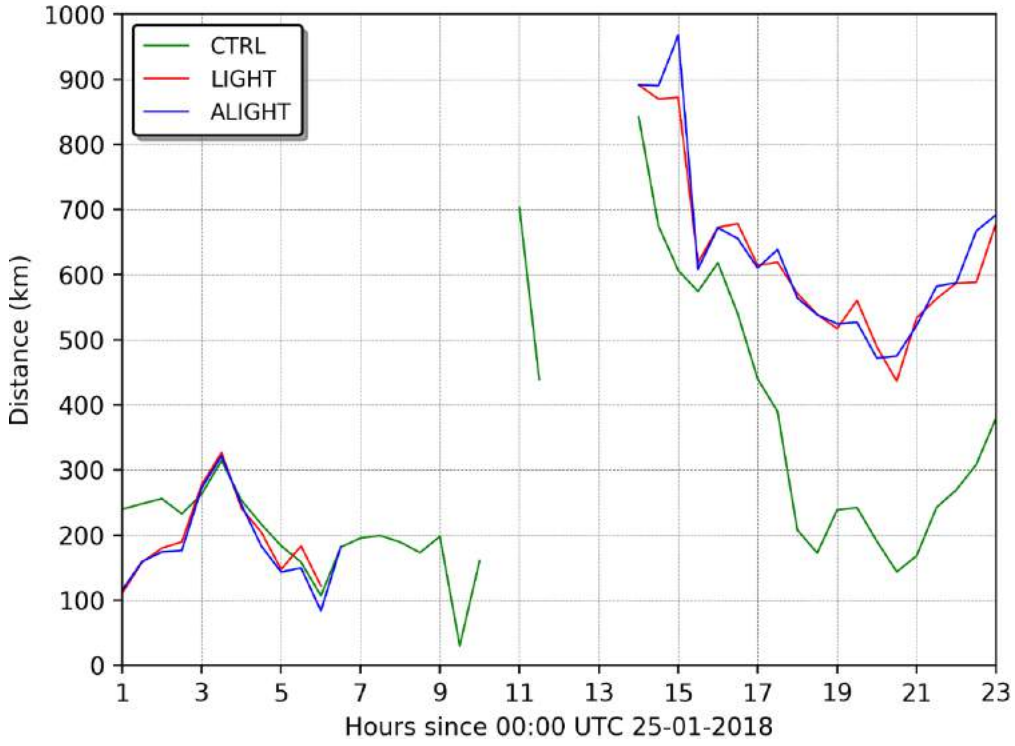


The deactivation of the cumulus parameterization allowed the model to solve the convection explicitly, but the 9km resolution used to perform the experiments might not have been ideal to solve the convection induced by the daytime heating (GILLILAND; ROWE, 2007). Although it was not observed significative difference simulating a large-scale system (as it was shown in the previous case), the convection associated with the daytime heating happens

locally and, in many cases, the thunderstorm core is smaller than 9km which could have affected the simulation of small systems.

The results presented in the analysis of the case occurred on 24/01/2018 showed that the application of the lightning data assimilation system improved the timing and positioning of the convection activity and the precipitation field in the first hours of simulation (3 hours). However, the amount of rainfall was overestimated in part because of the excess of water vapor added in the time analysis. Therefore, the use of the lightning data assimilation algorithm presented in this study during the warm season in the southeastern Brazil still need to be adjusted.

Figure 4.34 – Location error of the center of mass in relation to the geometric center based on lightning observations for each experiment between 00:00 UTC 25/01/2018 and 00:00 UTC 26/01/2018.



5 FINAL CONSIDERATIONS

This research assessed the impact of a Data Assimilation System responsible for inserting lightning data into the WRF model using the WRFDA-3DVAR system. Also, it was implemented and adapted an assimilation algorithm in order applying it to Brazil. For the first time, a lightning data assimilation technique using BrasilDAT data and aiming to improve the short-term weather forecasting was applied in South America.

This study presented a technique to correct the relative humidity threshold used in the original assimilation algorithm which contributed to improving the lightning data assimilation system during the dissipation phase of the meteorological system.

Based on the experiments performed on this study, it was possible to conclude that in general, the Lightning Data Assimilation System improved the short-term weather forecast of precipitation field for large-scale systems, especially when the correction in the relative humidity threshold was applied. Additionally, the assimilation algorithm improved the timing and positioning of a squall line that affected the study area possibly due to the correct representation of cold pools during the assimilation process.

Even though the assimilation algorithm improved the representation of the precipitation field in a few simulation cycles in the second case analyzed, it was noticed that when the convection is associated with thermal forcing the assimilation of lightning data using the algorithm presented in this study had a negative impact on the experiments. These convective systems are usually associated with deep convection generating a high incidence of lightning with a smaller amount of precipitation and a high concentration of ice on the top of the atmosphere. The inability of the model in reproducing ice concentrations induced the assimilation algorithm to add more moisture than necessary and subsequently to overestimate the precipitation. However, the horizontal resolution used in this study could have played an important role in simulating this type of meteorological system making it an important feature to be evaluated in future works.

The performance of the experiments was drastically affected by the type of convective system. The incapacity in reproducing the precipitation field associated with thermal forcing is visible in both cases analyzed. However, the capability in representing the correct positioning of the meteorological systems was improved in both cases independently on what mechanism induced the convection.

The assimilation methodology for lightning data presented in this study represents a significative contribution to the data assimilation field. The operational use of an alternative data source such as lightning has the potential to improve the short-term forecasts impacting positively several sectors of society. Lightning data has a high temporal and spatial resolution which can provide detailed information about different characteristics of thunderstorms. Moreover, the necessary instrumentation to detect lightning flashes is easier to deploy and maintain when compared with other data sources such as satellites and meteorological radars.

Since the assimilation of lightning data in meteorological models is a new field with most of the scientific papers published in the last ten years, there are many study possibilities to investigate. This present research investigated the impact of a lightning data assimilation system applied in South America and proposed a correction/adaptation in the algorithm. However, as it was showed previously the assimilation algorithm depends on the graupel concentration and since the model was not able to reproduce this variable the assimilation algorithm was affected negatively while trying to represent the precipitation field of convective systems generated by the daytime heating.

Based on the advantages and disadvantages observed in the results presented in this research, some suggestions for future works are proposed:

1. Due to the observed performance of the experiments for different types of meteorological systems, improving the representation of the precipitation field during the passage a Cold Front and impacting negatively in local convective systems, it is possible that the best performance of the assimilation algorithm would be observed during the

winter or in an area located in midlatitudes (since most of the precipitation is generated by large-scale systems). This way, it is needed a performance evaluation of the algorithm under those conditions;

2. A major adjustment in the coefficients of the assimilation equation has the potential to improve drastically the performance of lightning data assimilation procedures here presented. The use of machine learning algorithms for this purpose would help to define the ideal coefficients as well as the ideal conditions for when and where the data assimilation algorithm should be applied defining the layers of the atmosphere and the relative humidity threshold for when the performance is best;
3. Since the misrepresentation of the precipitation field for local convective systems was associated with the difficulties of the model in reproducing the graupel variable, the use of different parameterizations and/or spatial resolution in the experiments could improve the assimilation technique through the improvement of the graupel variable;
4. Another suggestion to improve lightning data assimilation procedures is using multiples assimilation cycles to improve the initial conditions of the model. However, it would necessary the use of filters during the initialization processes in order to reduce numerical instabilities due to the addition of mass.
5. Finally, it is also important to evaluate the performance of the lightning data assimilation process using different approaches other than 3DVAR, such as 4DVAR, Kalman Filter, Hybrid Methods…

REFERENCES

ADAMS-SELIN, R. D.; HEEVER, S. C. van den; JOHNSON, R. H. Impact of graupel parameterization schemes on idealized bow echo simulations. **Monthly Weather Review**, v. 141, n. 4, p. 1241–1262, 2013.

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. **Natural Hazards**, v. 75, n. 2, p. 1009–1022, 2015.

ALEXANDER, G. D. et al. The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 superstorm. **Monthly Weather Review**, v. 127, n. 7, p. 1433–1457, 1999. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1999)127%3C1433%3ATEOARR%3E2.0.CO%3B2>.

AMERICAN METEOROLOGICAL SOCIETY - AMS. **Meteorology glossary**. Available from: <http://glossary.ametsoc.org>. Access in: 5 feb. 2019.

APODACA, K. et al. Development of a hybrid variational-ensemble data assimilation technique for observed lightning tested in a mesoscale model. **Nonlinear Processes in Geophysics**, v. 21, n. 5, p. 1027–1041, 2014.

ARAKAWA, A.; LAMB, V. R. Computational design of the basic dynamical processes of the UCLA general circulation model. In: CHANG, J. (Ed.). **General Circulation Models of the Atmosphere.** New York: Academic Press, 1977. p. 174–265.

ATLAS, R.; TODLING, R. **Estimation theory and foundations of atmospheric data assimilation**. Greenbelt: NASA, 1999. (DAO Office Note 1999-01).

BAKER, M.B.; DASH, J.G. Charge transfer in thunderstorms and the surface melting of ice. **Journal of Crystal Growth**, v. 97, n. 3/4, p. 770–776, 1989.

BARKER, D. M. et al. A three-dimensional variational data assimilation system for MM5: implementation and initial results. **Monthly Weather Review**, v. 132, n. 4, p. 897–914, 2004.

BARKER, D. M.; HUANG, Wei; GUO, Yong-Run. **A three-dimensional variational (3DVAR) data assimilation system for use with MM5**. Boulder, CO: MMM Division, 2003. (NCAR Tech Note, 68).

BARNES, S. L. A technique for maximizing details in numerical weather map analysis. **Journal of Applied Meteorology**, v.3, p.396-409, 1964

BERGERON, T. On the physics of fronts. **Bulletin of the American Meteorological Society**, v. 18, n. 9, p. 265–275, 1937.

BERGTHÓRSSON, P.; DÖÖS, B. R. Numerical weather map analysis. **Tellus**, v. 7, n. 3, p. 329–340, 1955.

BIAGI, C. J. et al. National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003-2004. **Journal of Geophysical Research Atmospheres**, v. 112, n. 5, p. 1–17, 2007.

BJERKNES, V. On the dyanamics of the circular vortex with applications to the atmosphere and atmospheric vortex and wave motions. **Geophysiske Publkationer**, p. 1–81, 1921.

BOUTTIER, F; COURTIER, P. Data assimilation concepts and methods. **Meteorological training course lecture series. ECMWF**, p. 1–58, 2002. Available from: <http://www.msi.ttu.ee/~elken/Assim_concepts.pdf>.

BRINGI, V. N. et al. Evolution of a Florida thunderstorm during the Convection and Precipitation/Electrification Experiment: the case of 9 August 1991. **Monthly Weather Review**, v. 125, n. 9, p. 2131–2160, 1997.

BROWNING, K. A.; LUDLAM, F. H. Airflow in convective storms. **Quarterly Journal of the Royal Meteorological Society**, v. 88, n. 376, p. 117–135, 1962.

BROWNING, K. A.; PARDOE, C. W. Structure of low-level jet streams ahead of mid-latitude cold fronts. **Quarterly Journal of the Royal Meteorological Society**, v. 99, n. 422, p. 619–638, 1973.

BRUNDIDGE, K. C. The wind and temperature structure of nocturnal cold fronts in the first 1,420 feet. **Monthly Weather Review**, v. 93, n. 10, p. 587–603, 1965.

CAMPOS, L. Z. S. **On the mechanisms that lead to multiple ground contacts in lightning**. 2016. 280p. Tese (Doutorado em Geofísica Espacial) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.

CARDOZO, A. B. et al. Climatologia de frentes frias na América do Sul e sua relação com o modo anular sul. **Revista Brasileira de Climatologia**, v. 17, p.9-26, 2015.

CAREY, L. D.; RUTLEDGE, S. A multiparameter radar case study of the microphysical and kinematic evolution of a lightning producing storm. **Meteorology and Atmospheric Physics**, v. 59, n. 1/2, p. 33–64, 1996.

CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS. DIVISÃO DE SATÉLITES E ESTUDOS AMBIENTAIS - DSA/CPTEC. **CPTEC**. Available from: <http://satelite.cptec.inpe.br/>. Access in: 2 july 2019.

CHANG, D.-E. et al. The effect of spaceborne microwave and ground-based continuous lightning measurements on forecasts of the 1998 groundhog day storm. **Monthly Weather Review**, v. 129, n. 8, p. 1809–1833, 2001. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282001%29129%3C1809%3ATEOSMA%3E2.0.CO%3B2>.

CHEN, F. et al. Description and evaluation of the characteristics of the NCAR high-resolution land. **Journal of Applied Meteorology and Climatology**, v. 46, n. 6, p. 694–713, 2007.

CHEN, Z. et al. Lightning data assimilation with comprehensively nudging water contents at cloud-resolving scale using WRF model. **Atmospheric Research**, v. 221, p. 72–87, 2019. Available from: <https://doi.org/10.1016/j.atmosres.2019.02.001>.

COORAY, G. V. **The lightning flash**. London: The Institution of Electrical Engineers, 2003.

COSGROVE, B. A. et al. Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). **Journal of Geophysical Research**, v. 108, n. D22, 2003.

COURANT, R.; FRIEDRICHS, K.; LEWY, H. On the partial difference equations of mathematical physics. **IBM journal of Research and Development**, v. 11, n. 2, p. 215–234, 1967.

CUMMINS, K. L.; KRIDER, E. P.; MALONE, M. D. The U.S. national lightning detection network and applications of cloud-to-ground lightning data by electric power utilities. **IEEE Transactions on Electromagnetic Compatibility**, v. 40, n. 4, p. 465–480, 1998.

CUMMINS, K. L.; MURPHY, M. J. An overview of lightning locating systems: history, techniques, and data uses, with an. **IEEE Transactions on Electromagnetic Compatibility**, v. 51, n. 3, p. 499–518, 2009.

CUMMINS, K. L; MURPHY, M. J.; TUEL, J. V. Lightning detection methods and meteorological applications. In: INTERNATIONAL SYMPOSIUM ON MILITARY METEOROLOGY, 4., 2000, Malbork, Poland. **Proceedings…** 2000.

CUNNINGHAM, W. A.; PREACHER, K. J.; BANAJI, M. R. Research article consistency, stability, and convergent validity. **Psychological Science**, v. 12, n. 2, p. 163–170, 2001. Available from: <http://pss.sagepub.com/content/12/2/163.abstract>.

DASH, J. G. Surface melting. **Contemporary Physics**, v. 30, n. 2, p. 89–100, 1989.

DAVIDSON, N. E.; PURI, K. Tropical prediction using dynamical nudging,

satellite-defined convective heat sources, and a cyclone bogus. **Monthly Weather Review**, v. 120, p. 2501–2522, 1992.

DAVIS, C.; BROWN, B,; BULLOCK, R. Object-based verification of precipitation forecasts - Part I: methodology and application to mesoscale rain areas. **Monthly Weather Review**, v. 134, n. 7, p. 1772–1784, 2006.

DIAS, P. L. S.; SCHUBERT, W. H.; DEMARIA, M. Large-scale response of the tropical atmosphere to transient convection. **Journal of the Atmospheric Sciences**, v. 40, n. 11, p. 2689–2707, 1983.

DIXON, K. et al. The impact of lightning data assimilation on deterministic and ensemble forecasts of convective events. **Journal of Atmospheric and Ocean Technology**, p. 1801–1823, 2016.

DYE, J. E. et al. Observations within two regions of charge during initial thunderstorm electrification. **Quarterly Journal of the Royal Meteorological Society1**, v. 114, n. 483, p. 1271–1290, 1988.

ELSTER, Julius; GEITEL, Hans. Zur Influenztheorie der Niederschlagselektizität. **Physik Z**, v. 14, p. 1287, 1913.

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS. **ECMWF**. Available from:<https://www.ecmwf.int/>. Access in: 1 feb. 2019.

FERREIRA, A, G.; MELLO, N. G. S. Principais sistemas atmosféricos atuantes sobre a região nordeste do Brasil e a influência dos oceanos. **Revista Brasileira de Climatologia**, v. 1, n. 1, p. 15–28, 2005.

FIERRO, A. O. et al. Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. **Monthly Weather Review**, v. 140, n. 8, p. 2609–2627, 2012.

_____. Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. **Monthly Weather Review**, v. 143, p. 757–777, 2015.

FIERRO, A. O.; MANSELL, E. R. Relationships between electrification and storm-scale properties based on idealized simulations of an intensifying hurricane-like vortex. **Journal of the Atmospheric Sciences**, v. 75, n. 2, p. 657–674, 2018.

FIERRO, A. O et al. 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. **Monthly Weather Review**, v. 142, n. 1, p. 183–202, 2014. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-13-00142.1>.

FLETCHER, N. H. Surface structure of water and ice: II: a revised model. **Philosophical Magazine**, v. 18, n. 156, p. 1287–1300, 1968.

GANDU, A. W. **Notas de aula**. Available from: <http://www.dca.iag.usp.br/www/material/adwgandu/DESATIVADOS/PRE-2008/Gandu_UNALM_Peru_2007_old/MATERIALES/APOSTILAS_DIDATICAS/Gandu_Apostila_Teorica/ModelMeso-Gandu-Versao_1.pdf>. Access om: 1 sept. 2017.

GAO, J. et al. A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. **Weather and Forecasting**, v. 28, n. 3, p. 727–745, 2013.

GILLILAND, E. K.; ROWE, C. M. A comparison of cumulus parameterization schemes in the WRF model. [S.l: s.n.], 2007. p. P2.

GRENET, G. Essai d'explication de la charge électrique des nuages d'orages. **Annales Geophysicae**, v. 3, p. 306–307, 1947.

HAMILL, T. M.; SNYDER, C. A hybrid ensemble Kalman filter–3D variational analysis scheme. **Monthly Weather Review**, v. 128, p. 2905–2919, 2000.

HARTER, F. P. **Uso de filtro digital para iniciar um modelo de área limitada**. 1999. 111 f. Dissertação (Mestrado em Meteorologia) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1999.

HOKE, J. E.; ANTHES, R.. The initialization of numerical models by a dynamic-initialization technique. **Monthly Weather Review**, v.104, p.1551-1556, 1976.

HONG, S.-Y.; NOH, Y.; DUDHIA, J. A new vertical diffusion package with an explicit treatment of entrainment processes. **Monthly Weather Review**, v. 134, n. 9, p. 2318–2341, 2006.

HOUZE JUNIOR, R. A. Structure and dynamics of a tropical squall–line system. **Monthly Weather Review**, v. 105, n. 12, p. 1540–1567, 1977.

IACONO, M. J. et al. Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. **Journal of Geophysical Research Atmospheres**, v. 113, n. 13, p. 2–9, 2008.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS - CPTEC/INPE. **Tech. Rep.** Available from: <http://tempo.cptec.inpe.br/boletimtecnico/pt>. Acsess in: 5 apr. 2019.

KAIN, J. S.; FRITSCH, J. M. Convective parameterization for mesoscale models: the Kain-Fritsch scheme. **Meteorological Monography**, v. 46, p. 165–170, 1993.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, v. 82, n. 1, p. 35, 1960.

KALNAY, E. **Atmospheric modeling, data assimilation and predictability**.

[S.l.]: Cambridge University Press, 2003. Available from: <http://books.google.com/books?id=Uqc7zC7NULMC&pgis=1>.

KISTLER, R. E. **A study of data assimilation techniques in an autobarotropic, primitive equation, channel model**. 1974. Dissertation (PhD) - Pennsylvania State University, State College, PA, 1974.

KOSHAK, W. J. et al. North Alabama Lightning Mapping Array (LMA): VHF source retrieval algorithm and error analyses. **Journal of Atmospheric and Oceanic Technology**, v. 21, n. 4, p. 543–558, 2004.

KOUSKY, V. E.; GAN, M. A. Upper tropospheric cyclonic vortices in the tropical South Atlantic. **Tellus**, v. 33, n. 6, p. 538–551, 1981.

KREHBIEL, P. R. **An analysis of the electric field change produced by lightning**. [S.l.]: Geophysical Research Center, 1981.

LAGOUVARDOS, K. et al. Study of a heavy precipitation event over southern france, in the frame of HYMEX project: observational analysis and model results using assimilation of lightning. **Atmospheric Research**, v. 134, p. 45–55, 2013. Available from: <http://dx.doi.org/10.1016/j.atmosres.2013.07.003>.

LAPRISE, R. The Euler equations of motion with hydrostatic pressure as an independent variable. **Monthly Weather Review,** v.120, p.197-207, 1992. Available from: <http://dx.doi.org/10.1175/1520-0493(1992)120%3C0197:TEEOMW%3E2.0.CO;2>.

LORENC, A. C. Analysis methods for numerical weather prediction. **Quarterly Journal of the Royal Meteorological Society**, v. 112, n. 474, p. 1177–1194, 1986. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/qj.49711247414/abstract>.

_____. The Met Office global three-dimensional variational data assimilation scheme. **Quarterly Journal of the Royal Meteorological Society**, v. 126, n. 570, p. 2991–3012, 2000.

LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the Atmospheric Sciences**, v.20, p.130-141, 1963.

LYNCH, P.; XIANG-YU, H. Initialization of the HIRLAM model using a digital filter. **Monthly Weather Review1**, v. 120, n. 6, p. 1019–1034, 1992.

LYNN, B. H.; KELMAN, G.; ELLROD, G. An evaluation of the efficacy of using observed lightning to improve convective lightning forecasts. **Weather and Forecasting**, v. 30, p. 405–423, 2015.

MACGORMAN, D. R. et al. Lightning rates relative to tornadic storm evolution on 22 May 1981. **Journal of the Atmospheric Sciences**, v.46, n.2, p.221-250,1989

MACGORMAN, D. R.; RUST, W. D. **The electrical nature of storms**. [S.l.]: Oxford University Press, 1998.

MANOBIANCO, J. et al. The impact of assimilating satellite-derived precipitation rates on numerical simulations of the ERICA IOP 4 cyclone. **Monthly Weather Review**, v. 122, p. 341–365, 1994.

MANSELL, E. R.; ZIEGLER, C. L.; MACGORMAN, D. R. A Lightning data assimilation technique for mesoscale forecast models. **Monthly Weather Review**, v. 135, n. 5, p. 1732–1748, 2007.

MING, X. **Notas de aula**. Available from: <http://twister.caps.ou.edu/OBAN2014/3DVAR.pdf>. Access in: 29 mar. 2015.

MOYA-ÁLVAREZ, A. et al. Extreme rainfall forecast with the WRF-ARW model in the Central Andes of Peru. **Atmosphere**, v. 9, n. 9, p. 362, 2018.

NACCARATO, K. P.; ALBRECHT, R. I.; PINTO JUNIOR, O. Cloud-to-ground lightning density over Brazil based on high-resolution lightning imaging sensor (LIS) data. In: INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY, 14., 2011, Rio de Janeiro, Brazil. **Proceedings…** 2011. p. 12–15.

NACCARATO, K. P.; BOURSCHEIDT, V. The fourth generation of the Brazilian detection efficiency model for BrasilDAT network (RDEM4). In: INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY, 14., 2011, Rio de Janeiro, Brazil. **Proceedings…** 2011. NACCARATO, K. P.; PINTO JUNIOR, O. Lightning detection in Southeastern Brazil from the new Brazilian Total Lightning Network (BrasilDAT). In: INTERNATIONAL CONFERENCE ON LIGHTNING PROTECTION (ICLP), 2012, Vienna, Austria. **Proceedings...** 2012. p. 8.

NACCARATO, K. P. **Análise das características dos relâmpagos na Região Sudeste do Brasil**. 2006. 364 f. Tese (Doutorado em Geofísica Espacial) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006.

NASCIMENTO, E. L. **Influência dos bloqueios atmosféricos na propagação de ondas de Rossby em escoamentos de Inverno no Hemisfério Sul**. 1998. Dissertação (Mestrado em Ciências Atmosféricas) - Universidade de São Paulo, São Paulo, 1998.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION - NASA. **GODDARD**. Available from: <https://www.nasa.gov/centers/goddard/news/topstory/2004/0621lightning.html>. Access in: 19 feb. 2019.

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH. MESOSCALE AND MICROSCALE METEOROLOGY LABORATORY - MMM/UCAR. **WRF user page**. Available from: <http://www2.mmm.ucar.edu>. Access in: 20 jan. 2018.

OLIVEIRA JÚNIOR, J. F. et al. Análise da precipitação e sua relação com sistemas meteorológicos em Seropédica, Rio de Janeiro. **Floresta e Ambiente**, v. 21, n. 2, p. 140–149, 2014.

PAEGLE, J. A comparative review of South American low level jets. **Meterologica**, v. 23, p. 73–81, 1998.

PAPADOPOULOS, A.; CHRONIS, T. G.; ANAGNOSTOU, E. N. Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale nodel. **Monthly Weather Review**, v. 133, n. 7, p. 1961–1977, 2005.

PARRISH, D. F.; DERBER, J. C. The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. **Monthly Weather Review**, v.120, p.1747-1763, 1992.

PECKHAM, S. E. et al. Implementation of a digital filter initialization in the WRF model and its application in the rapid refresh. **Monthly Weather Review**, v. 144, n. 1, p. 99–106, 2016.

PESSI, A. T.; BUSINGER, S. The impact of lightning data assimilation on a winter storm simulation over the North Pacific Ocean. **Monthly Weather Review**, v. 137, n. 10, p. 3177–3195, 2009. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/2009MWR2765.1>.

PETTY, G. W. **A first course in atmospheric thermodynamics**. [S.l.]: Sundog Publishing, 2008.

PHILLIPS, N. A coordinate system having some special advantages for numerical forecasting. **Journal of Meteorology**, v.14, p.184-185, 1957

PINTO JUNIOR, O. **Lightning in the tropics**: from a source of fire to a monitoring system of climatic changes. [S.l.]: Nova Science, 2009.

POWERS, J. G. et al. The weather research and forecasting model: Overview, system efforts, and future directions. **Bulletin of the American Meteorological Society**, v. 98, n. 8, p. 1717–1737, 2017.

PRICE, C.; RIND, D. A simple lightning parameterization for calculating global lightning distributions. **Journal of Geophysical Research**, v. 97, p. 9919–9933, 1992. Available from: <http://www.agu.org/pubs/crossref/1992/92JD00719.shtml>.

QIE, X. et al. Application of total-lightning data assimilation in a mesoscale convective system based on the WRF model. **Atmospheric Research**, v. 145–146, p. 255–266, 2014. Available from: <http://dx.doi.org/10.1016/j.atmosres.2014.04.012>.

RABIER, F.; LIU, Z. **Variational data assimilation:** theory and overview. 2003. Available from: <https://www.ecmwf.int/sites/default/files/elibrary/2003/11805-variational-data-assimiltion-theory-and-overview.pdf>.

RAKOV, V. A.; UMAN, M. A. **Lightning:** physics and effects. [S.l.]: Cambridge University Press, 2003.

REBOITA, M. S. et al. Regimes de precipitação na América do Sul: uma revisão bibliográfica. **Revista Brasileira de Meteorologia**, v. 25, n. 2, p. 185–204, 2010.

REYNOLDS, S. E. Thunderstorm charge separation. **Journal of Meteorology**, v.14, p.426-436, 1957.

RODGER, C. J. et al. Detection efficiency of the VLF World-Wide Lightning Location Network (WWLLN): initial case study. **Annales Geophysicae**, v. 24, p. 3197–3214, 2006.

ROGERS, R. F.; FRITSCH, J. M.; LAMBERT, W. C. A simple technique for using radar data in the dynamic initialization of a mesoscale model. **Monthly weather review**, v. 128, p. 2560–2574, 2000. Available from: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Simple+Technique+for+Using+Radar+Data+in+the+Dynamic+Initialization+of+a+Mesoscale+Model#0>.

SALIO, P.; NICOLINI, M.; ZIPSER, E. J. Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet. **Monthly Weather Review**, v. 135, n. 4, p. 1290–1309, 2007.

SCHULTZ, C. J.; PETERSEN, W.; CAREY, L. D. Lightning and severe weather: a comparison between total and cloud-to-ground lightning trends. **Weather and Forecasting**, v. 26, n. 5, p. 744–755, 2011.

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. **Journal of Applied Meteorology and Climatology**, v. 48, n. 12, p. 2543–2563, 2009.

SILBERMAN, I. Planetary waves in the atmosphere. **Journal of Meteorology**, v. 11, n. 1, p. 27–34, 1954.

SKAMAROCK, W.C. et al. **A description of the advanced research WRF version 2**. Boulder, Colorado: NCAR, 2005 (Technical Report, 113).

_____. **A description of the advanced research WRF version 3**. Boulder, Colorado: NCAR, 2009. (Technical Report).

SOKOL, Z. Effects of an assimilation of radar and satellite data on a very short

range forecast of heavy convective rainfalls. **Atmospheric Research**, v. 93, n. 1–3, p. 188–206, 2009.

STANSKI, H. R.; LAURENCE, J.W.; BURROWS, W. R. **Survey of common verification methods in meteorology**. [S.l.]: WMO, 1989. (Technical Resport, 358). Available from: <https://www.cawcr.gov.au/projects/verification/Stanski_et_al/Stanski_et_al.html>.

STEFANESCU, R. et al. **1D+4D-VAR data assimilation of lightning with WRFDA system using nonlinear observation operators**. Available from: <https://arxiv.org/ftp/arxiv/papers/1306/1306.1884.pdf>. Access in: 20 out. 2016.

_____. **Data assimilation of lightning in WRF 3/4-D VAR using observation operators**. 2012. Available from: <http://myweb.fsu.edu/rstefanescu/Papers/Stefanescu_Navon.pdf>.

STENSRUD, D. J. **Parameterization schemes:** keys to understanding numerical weather prediction models. [S.l.]: Cambridge University Press, 2007.

SU, Y. et al. Real-time simulation and mechanistic analysis of a squall line case in East China. **Atmospheric and Oceanic Science Letters**, v. 9, n. 5, p. 394–400, 2016. Available from: <http://dx.doi.org/10.1080/16742834.2016.1219975>.

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 Meteorology**, v. 1, 2013.

TEWARI, M. et al. **Implementation and verification of the unified noah land surface model in the WRF model**. Seattle, Washington: [s.n.], 2004. Available from: <http://www2.mmm.ucar.edu/wrf/users/phys_refs/LAND_SURFACE/noah.pdf>.

THOMPSON, G. et al. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. part II: implementation of a new snow parameterization. **Monthly Weather Review**, v. 136, p. 5095–5115, 2008.

UMAN, M. **The lightning discharge**. [S.l.]: Courier Co, 2001.

VONNEGUT, B.. Some facts and speculations concerning the origin and role of thunderstorm electricity. **Meteorological Monographs**, v. 5, n. 27, p. 224–241, 1963.

WANG, H. et al. Continuous assimilation of lightning data using time-lagged ensembles for a convection-allowing numerical weather prediction model. **Journal of Geophysical Research: Atmospheres**, v. 123, n. 17, p. 9652–9673, 2018.

WANG, Y. et al. A case study of assimilating lightning-proxy relative humidity with WRF-3DVAR. **Atmosphere**, v. 8, n. 3, 2017.

WANG, X. et al. A hybrid ETKF–3DVAR data assimilation scheme for the WRF model - part II: real observation experiments. **Monthly Weather Review**, v. 136, n. 12, p. 5132–5147, 2008. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/2008MWR2444.1>.

WARNER, T. T. **Numerical weather and climate prediction**. Boulder, Colorado: Cambridge University Press, 2011.

WEISMAN, M. L.; KLEMP, J. B. **Characteristics of isolated convective storms**. Boston, MA: American Meteorological Society, 1986.

_____. The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. **Monthly Weather Review**, v. 110, n. 6, p. 17, 1982.

WILLIAMS, E. et al. The behavior of total lightning activity in severe Florida thunderstorms. **Atmospheric Research**, v. 51, n. 3, p. 245–265, 1999.

WILLIAMS, E. R. The electrification of thunderstorms. **Scientific America**, v. 107, n. 452, p. 277–298, 1988.

YANG, Y.; WANG, S. Assimilation of chinese doppler radar and lightning data with WRF-GSI: a case study of mesoescale convective system. **Advances in Meteorology**, v. 2015, p. 4517, 2015.

ZIEGLER, C. L. et al. A model evaluation of noninductive graupel-ice charging in the early electrification of a mountain thunderstorm. **Journal of Geophysical Research: Atmospheres**, v. 96, n. D7, p. 12833–12855, 1991.