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

sid.inpe.br/mtc-m21c/2019/05.14.19.52-TDI

DEVELOPMENT OF A PASSIVE MICROWAVE-BASED SATELLITE PRECIPITATION ESTIMATION ALGORITHM FOR BRAZIL

Lia Martins Costa do Amaral

Doctorate Thesis of the Graduate Course in Meteorology, guided by Dr. Daniel Alejandro Vila, approved in May 20, 2019.

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

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/2019/05.14.19.52-TDI

DEVELOPMENT OF A PASSIVE MICROWAVE-BASED SATELLITE PRECIPITATION ESTIMATION ALGORITHM FOR BRAZIL

Lia Martins Costa do Amaral

Doctorate Thesis of the Graduate Course in Meteorology, guided by Dr. Daniel Alejandro Vila, approved in May 20, 2019.

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

INPE
São José dos Campos
2019

Cataloging in Publication Data

Amaral, Lia Martins Costa do.

Am13d Development of a passive microwave-based satellite precipitation estimation algorithm for Brazil / Lia Martins Costa do Amaral. – São José dos Campos : INPE, 2019.

xxii + 115 p. ; (sid.inpe.br/mtc-m21c/2019/05.14.19.52-TDI)

Thesis (Doctorate in Meteorology) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2019.

Guiding : Dr. Daniel Alejandro Vila.

1. Satellite. 2. Precipitation. 3. Artificial neural networks. 4. Passive microwave. 5. GPM Microwave Imager (GMI). I.Title.

CDU 551.577:004.021

Aluno (a): *Lia Martins Costa do Amaral*

Título: "DEVELOPMENT OF A PASSIVE MICROWAVE-BASED SATELLITE PRECIPITATION ESTIMATION ALGORITHM FOR BRAZIL"

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

Dra. Simone Marilene Sievert da Costa Coelho

Presidente / INPE / Cachoeira Paulista - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Dr. Daniel Alejandro Vila

Orientador(a) / INPE / Cachoeira Paulista - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Dr. Luiz Augusto Toledo Machado

Membro da Banca / INPE / SJCampos - SP

( ) Participação por Video - Conferência

(X) Aprovado *( ) Reprovado*

Dr. Giulia Panegrossi

Convidado(a) / ISAC/CNR / Roma - IT

(X) Participação por Video - Conferência

( ) Aprovado *( ) Reprovado*

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 20 de maio de 2019

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

Meteorologia

Dr. Enrique Vieira Mattos

Enrique Vieira Mattos

Convidado(a) / UNIFEI / Itajubá - MG

( ) Participação por Video - Conferência

(X) Aprovado ( ) Reprovado

Este trabalho foi aprovado por:

( ) maioria simples

( ) unanimidade

São José dos Campos, 20 de maio de 2019

"Life isn't about waiting for the storm to pass... It's about learning how to dance in the rain."

Vivian Greene

ACKNOWLEDGEMENTS

Firstly, the author would like to thank the essential the financial support and fellowship of the National Council for Scientific and Technological Development (CNPq)-Brazil (Reference number-140715/2017-7 and Universal Call number 403987/2016-4) and the Coordination for the Improvement of Higher Education Personnel (CAPES) Brazil during fellowship PhD studies and also the Science Without Borders (CSF) Program by the Sandwich Doctorate (SWE)(201237/2015-6) for the internship opportunity. The author would like to thank the CHUVA Project (FAPESP Grant 2009/15235-8) and H-SAF Consortium for the data provided for this study. Additionally, mention also the acknowledge for the infrastructure provided by the National Institute for Space Research (INPE), Institute of Atmospheric Sciences and Climate (CNR-ISAC) and the Italian Civil Protection Department (DPC) and for the EUMETSAT for the financial support with the Visiting Scientist Activities funded by the H-SAF Consortium.

The author would also like to effusively thank and acknowledge the MCTIC/FINEP (CT-INFRA grant 0112052700) and the Embrace Space Weather Program for the computing facilities at INPE, specially the Manager for Operations Dr. Marcelo Banik, without whom the execution of the NNIMBUS algorithm would never reach the maturity level presented in here. Huge thanks Banik! For all the nights and weekends that you have left the clusters online for us. :)

I would thank my advisor Dr. Daniel Vila for believing in me in the realization of this work. To my external advisor, Dra. Giulia Panegrossi, for the amazing opportunity to work with you and your group at CNR. My deep grateful to Ana Cinzia Marra and Paolo Sanò for the patience in teach me several technical, linguistic and cultural teachings and for all my italian friends and colleagues that receive me very all. To my committee members, Dra. Giulia, Dra. Simone, Dr. Enrique, Dr. Luiz and Dr. Daniel for contribute to my study. To Dr. Luiz Augusto Machado for the opportunity to participate in the CHUVA campaign in Manaus and work with the CHUVA dataset and Thiago Biscaro for the radar data support. To Dr. Enrique Mattos by the nice experience to teach his students at UNIFEI. My eternal gratitude to my parents and family members that always gave me all support to carry out this work. My special thanks to my husband David França for the love and huge patience with me. I would not have arrived here without you! My special thanks also to Sofia Scomazzon that was my angel when I needed her most and who became one of my best friends. And my thanks to all my friends who directly and indirectly collaborated in this work.

ABSTRACT

In order to develop a passive microwave-based satellite precipitation estimation algorithm optimized for Brazil, this work was divided in two parts. The first part consisted in extending the cloud-radiation database used as *a priori* information for the Cloud Dynamics and Radiation Database (CDRD) Bayesian algorithm in order to include the cloud resolving model simulations representative of brazilian rainfall regimes. Simulations of microphysical, dynamical and meteorological profiles were then generated using the University of Wisconsin - Nonhydrostatic Modeling System and the brightness temperature (TB) simulations were generated using the Radiative Transfer Equation Modeling System for the CHUVA (Amazon and Vale) golden cases and compared with observed TB. The results demonstrated that the simulations detected perturbations in the TB fields (in space and time) however in terms of the range of temperature values, the model did not reproduce the lowest values of TB that were present in the observations. The model also seemed to struggle with the riming process on graupel formation, providing small amounts of graupel content. These results demonstrated that the models needed adjustments to be able to describe the regional features of TB across a wide range of meteorological systems in Brazil. For these reasons, the second part of the work was developed by making use of an observational database from the sensors GPM Microwave Imager and Dual-frequency Precipitation Radar (GMI/DPR-CMB) in order to develop a screening of precipitation and rainfall retrieval algorithm over Brazil, based on artificial neural networks (ANN) and called **N**eural **N**etwork **IM**plementation of the **B**razilian M**U**ltilayer Perceptron for **S**creening and precipitation retrieval (NNIMBUS). The precipitation screening proved to be very effective in both detecting larger systems and smaller or isolated systems. Regarding the GMI/DPR-CMB validation dataset, the screening performed well, with an accuracy of 0.95, POD of 0.80, FAR of 0.39 and bias of 1.34. When compared to the Goddard profiling algorithm (GPROF) the screening still had good performance, however with slightly smaller scores. It was observed that through the comparison maps with GPROF the NNIMBUS can detect agglomerates very similarly, however it does not detect the borders of the systems very well. This behavior might be associated with the precipitation thresholds that were configured with the training dataset (0.2 a 60 mm/h), which might be leading more stratiform regions of the systems to go undetected. The rainfall retrieval model also performed well when compared to the GMI/DPR-CMB observations, with an MAE of 4.19, standard deviation of 3.23 and RMSE of 5.59 for the validation dataset. Analyzing the rain rate classes, the retrieval tends to underestimate classes between 0.2 and 1 mm/h, overestimate classes between 1 and 10 mm/h and underestimate classes greater than 10 mm/h. These features can be associated with the input dataset distribution, as well as with the criteria applied in data cleaning process.

Keywords: Satellite. Precipitation. Artificial neural networks. Passive microwave. GPM Microwave Imager (GMI). Observational database.

DESENVOLVIMENTO DE UM ALGORITMO DE ESTIMATIVA DE PRECIPITAÇÃO BASEADA EM MICROONDAS PASSIVO PARA O BRASIL

RESUMO

Com objetivo de desenvolver um algoritmo de estimativa de precipitação por satélite baseado em microondas passivo otimizado para o Brasil, este trabalho foi dividido em duas partes. A primeira parte consistiu em estender o banco de dados de radiação de nuvens usado como informação *a priori* para o algoritmo Bayesiano *Cloud Dynamics and Radiation Database* (CDRD), a fim de incluir as simulações representativas dos regimes de precipitação do Brazil. Simulações dos perfis microfísicos, dinâmicos e meteorológicos foram geradas usando o *University of Wisconsin – Nonhydrostatic Modeling System* e as simulações de temperatura de brilho (TB) foram geradas usando o *Radiative Transfer Equation Modeling System* para os sitemas precipitantes observados durante o projeto CHUVA (campanhas do Vale do Paraíba e Manaus). Os resultados demonstraram que as simulações detectaram as perturbações nos campos de TB (no espaço e no tempo), porém em termos do intervalos de TB, o modelo não reproduziu os menores valores de TB presentes nas observações. O modelo aparentou ter dificuldade em gerar o processo de formação de graupel, gerando pequenas valores de conteúdo de graupel. Esses resultados demonstraram que os modelos precisavam de ajustes para poder descrever as características regionais da TB para ampla gama de sistemas meteorológicos no Brasil. Por estas razões, a segunda parte do trabalho consistiu no desenvolvimento de um algoritmo de redes neurais artificais (denominado *Neural Network Implementation of the Brazilian Multilayer Perceptron for Screening and precipitation retrieval* (NNIMBUS), tanto para detecção da área precipitante *screening* como para recuperação da intensidade da precipitação, utilizando um banco de dados observacionais provindos dos sensores *GPM Microwave Imager* e *Dual-frequency Precipitation Radar* (GMI/DPR-CMB). A detecção de precipitação (*screening*) provou ser muito eficaz na detecção de sistemas maiores e sistemas menores ou isolados. Em relação ao conjunto de dados de validação do GMI/DPR-CMB, o algoritmo apresentou bom desempenho, com acurácia de 0,95, POD de 0,80, FAR de 0,39 e viés de 1,34. Quando comparado ao algoritmo *Goddard profiling algorithm* (GPROF), a detecção de precipitação ainda apresentava bom desempenho, porém com estatísticas ligeiramente menores. Através dos mapas de comparação com o GPROF, foi possível perceber que o NNIMBUS consegue detectar os aglomerados de forma muito semelhante, porém não detecta muito bem as bordas dos sistemas. Esse comportamento pode estar associado aos limiares de precipitação que foram configurados com o conjunto de dados de treinamento (0,2 a 60 mm/h), o que pode estar levando a que regiões mais estratiformes dos sistemas não sejam detectadas. O modelo de recuperação da precipitação também teve um bom desempenho quando comparado com as observações GMI/DPR-CMB, com um MAE de 4,19, desvio padrão de 3,23 e RMSE de 5,59 para o conjunto de dados de validação. Analisando as classes de taxa de chuva, a recuperação tende a subestimar as classes entre 0,2 e 1 mm/h, superestimar as classes entre 1 e 10 mm/h e subestimar as classes acima de 10 mm/h. Essas características podem estar

associadass à distribuição do conjunto de dados de entrada, bem como aos critérios aplicados no processo de limpeza de dados.

Palavras-chave: Precipitação por satélite. Redes Neurais Artificais. Microondas passivo. GMI. DPR. CMB. Banco de dados observacional.

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

AI	–	Artificial Intelligence
AMSR-E	–	Advanced Microwave Scanning Radiometer-Earth Observing System
AMSU	–	Advanced microwave Sound Unit
ANN	–	Artificial neural network
API	–	Application programming interface
CDRD	–	Cloud Dynamics and Radiation Database
CHUVA	–	Cloud processes of the main precipitation systems in Brazil: A contribution to cloud resolving modeling and to the Global Precipitation Measurement
CLASS	–	Comprehensive Large Array-data Stewardship System
CMB	–	Combined Radar-Radiometer Algorithm
CRM	–	Cloud resolving model
CSU	–	Colorado State University
DPR	–	Dual-frequency Precipitation Radar
DSD	–	Drop size distribution
GCOMW1	–	Global Change Observation Mission-Water 1
GFS	–	Global Forecast System
GMI	–	GPM Microwave Imager
GPM	–	Global Precipitation Measurement
GPM-CO	–	GPM Core Observatory
GPROF	–	Goddard Profiling Algorithm
H	–	Horizontal polarization
H-SAF	–	Satellite Application Facility on Support to Operational Hydrology and Water Management
ILW	–	Integrated liquid water
IMERG	–	Integrated Multi-satellite Retrievals for GPM
IR	–	Infrared
ITCZ	–	Intertropical Convergence Zone
IWP	–	Ice Water Path
JAXA	–	Japan Aerospace Exploration Agency
LWC	–	liquid water content
LWP	–	Liquid Water Path
MCC	–	Mesoscale Convective Complexes
METOP(A/B)	–	European Meteorological Operational A and B
MHS	–	Microwave Humidity Sounder
ML	–	Machine Learning
MRR	–	Micro Rain Radar
MT1	–	Megha-Tropiques 1
MW	–	Microwave
NN	–	Neural network

NESDIS	–	National Environmental Satellite, Data, and Information Service
NOAA	–	National Oceanic and Atmospheric Administration
PCT	–	Polarization-corrected brightness temperature
PERSIANN	–	Precipitation Estimation from Remotely Sensed Information using Neural Network
PMW	–	Passive microwave
PNPR	–	Neural Network Precipitation Retrieval
PR	–	Precipitation Radar
PUSH	–	Precipitation Uncertainties for Satellite Hydrology
RR	–	Rain Rate
RTE	–	Radiative transfer equation
RTE/RMS	–	Radiative Transfer Equation Modeling System
SACZ	–	South Atlantic Convergence Zone
SAPHIR	–	Sondeur Atmosphérique du Profil d'Humidité Intertropicale par Radiométrie
SDR	–	Sensor data record
SIPAM	–	Sistema de Proteção da Amazônia
SL	–	Squall lines
S-Pol	–	S-band dual-polarization Doppler radar
SSM/I	–	Special Sensor Microwave/Imager
SVM	–	Support Vector Machine
TB	–	Brightness temperature
TMI	–	TRMM's Microwave Imager
TRMM	–	Tropical Rainfall Measuring Mission
UW-NMS	–	University of Wisconsin - Nonhydrostatic Modeling System
V	–	Vertical polarization
VIS	–	Visible
VPR	–	Vertical profiles of reflectivity
XPOL	–	X-band dual polarization radar

CONTENTS

1 INTRODUCTION

1.1 Introduction and context

The knowledge about the distribution of water around the globe is an aspect of extreme importance for the management of natural resources. Within the hydrological cycle, the precipitation acts as a central component regulating the energy balance through the interactions of water vapor (a primary greenhouse gas) and clouds, which redistribute latent heat in the atmosphere. On the surface, the precipitation, besides affecting from the domestic to the industrial sectors, is a primary source of freshwater in a world that has been facing emerging crises in terms of water availability.

The quantification and estimation of precipitation around the globe, in terms of frequency and intensity, is made by pluviometric networks and meteorological radars [single and double polarization] on ground. But, in turn, they have a sparse and unequal distribution, especially in mountain regions, forest and ocean. Thus, satellite precipitation estimates can fill these gaps.

As exposed by Kummerow et al. (1996) the microwave remote sensing of clouds and precipitation has shown great promise by virtue of its direct interaction with hydrometeors and differently from the infrared measurements which interact only to the highest layer of clouds, the microwave radiation has the ability to penetrate through the clouds and offer insight about the rainfall structure. In this context, several algorithms have been developed within the purpose to improve rainfall estimation based on passive microwave, some main algorithms can be cited: the Goddard Profiling Algorithm (GPROF) (KUMMEROW et al., 1996), which uses an Bayesian approximation to retrieve the instantaneous rainfall globally; the Bayesian approach Cloud Dynamics and Radiation Database (CDRD) (CASELLA et al., 2013) and the Passive Microwave Neural Network Precipitation Retrieval (PNPR) (SANÒ et al., 2015).

In the last decades, many studies were able to provide improvements in rainfall retrievals and one method for this was the development of precipitating cloud–radiation databases that can be derived from cloud resolving model (CRM) simulations or from radar measurements, or from a combination of them (Evans et al. (1995), Kummerow et al. (2001), Mugnai et al. (1993). Normally, these databases are composed by thousands of microphysical-meteorological reproduced by numerical and cloud resolving models concerning the occurrence of precipitating events

and they are coupled to radiative transfer models that reproduce the simulated brightness temperatures (TB) associated with their respective profiles (CASELLA et al., 2012). This type of database is affected by the classical non-uniqueness (ambiguity) problem inherent to multi-value mathematical functions, which it implies that multiple solutions are possible by reason of different vertical profiles of microphysical hydrometeors can lead to exactly or nearly exactly the same TB vector. And in order to minimize the ambiguity Smith et al. (2013) identified and applied optimal geophysical and *meteorological tags* that constrain and refine the selection of candidate microphysical profiles used for the Bayesian retrieval (CDRD).

In order to improve the understanding of the precipitation physics and its space–time variability, besides providing global rainfall monitoring, the Global Precipitation Measurement (GPM) mission launched the GPM Core Observatory (GPM-CO) spacecraft which carries the most advanced precipitation sensors currently in space. These sensors include the Ku- and Ka-band Dual-frequency Precipitation Radar (DPR) provided by Japan Aerospace Exploration Agency (JAXA) that measures three-dimensional (3D) structures of precipitation, and the GPM Microwave Imager (GMI), a well-calibrated multi-frequency radiometer capable of providing wide-swath precipitation data. Therefore the GPM-CO, in addition to the Tropical Rainfall Measuring Mission (TRMM) (previous mission), offered the opportunity to use a large observational/empirical database of coincident TRMM/GPM space-borne radar (precipitation radar (PR)/DPR) and radiometer observations (TMI/GMI) instead of a database built from several simulations (Cloud Resolving Models (CRM) coupled to Radiative Transfer Equation (RTE) models). As mentioned by Skofronick-Jackson et al. (2018) the first three years of GPM operations showed some noteworthy achievements, including: algorithms updated with improved calibration of the DPR, an observational database for the GMI retrieval algorithm, and updates on the representation of light rain, falling snow, and non-spherical particles.

In addition to these advances, from the computational point of view, Machine Learning (ML) techniques have proven to be excellent tools to handle with difficult problems from a variety of applications, among them Atmospheric Science, Remote Sensing and Weather Prediction. These are typically problems related to classification, prediction, optimization, among others. Increasingly facilitated access to these techniques (mainly the open source languages and tools) provided a larger understanding of how they work and allowed their broader applicability.

Ground-based measurements of the cloud processes of the main precipitating sys-

tems over Brazil (e.g., squall lines (SL); Mesoscale Convective Complexes (MCC), cold fronts, South Atlantic Convergence Zone (SACZ), Intertropical Convergence Zone (ITCZ), warm clouds, local convection) were collected during field campaigns by the CHUVA project (MACHADO et al., 2014). These measurements provided a representative set of case studies to evaluate the performance of the algorithms for different types of events and brought knowledge about cloud microphysical processes. It is worth to remember that the country has an area of 8.5 million square kilometers and lies primarily south of the equator (and within the tropics), being ideally situated for studies of tropical continental convection over a broad range of precipitation regimes within a single country. The territorial extension also implies a surface type and land use heterogeneity, large coastline length and orographic diversity, which contributed to increase the complexity of estimate rainfall over the country.

Taking into account the absence of a optimized microwave precipitation retrieval to reproduce the brazilian rainfall variability, and even more so for operational purposes, this study proposes the development of a passive microwave-based satellite precipitation estimation algorithm for Brazil. To achieve this goal, this study seeks to answer the following specific questions:

a) What is the Cloud Dynamics and Radiation Database (CDRD) algorithm performance (optimized for Europe and Africa) to retrieve the rainfall rates over the central Amazon and Southeast region of Brazil? (Chapter 3)

b) Is the cloud resolving model from University of Wisconsin - Nonhydrostatisc Modeling System (UW-NMS) used in CDRD able to reproduce the typical precipitating systems in the Amazon and Southeast region of Brazil? (Chapter 4)

c) Assuming the GMI/DPR-CMB as the ground truth, is it possible to achieve a representative rainfall retrieval over Brazil through the use of Artificial Neural Network (ANN) methods? (Chapter 5)

1.2 Objectives

The main objective of this study is to develop a passive microwave-based satellite precipitation estimation algorithm for Brazil. As an initial proposal of the PhD, the goal was to extend the cloud-radiation database used as *a priori* information for the CDRD Bayesian algorithm (SSMIS - denominated H01) developed at the

EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF), by means of cloud resolving model simulations for Brazil. Thus, the representativeness of the algorithm for this region could be improved (the algorithm is currently optimized for Europe and Mediterranean area, Africa and Southern Atlantic). Once the simulations were done, it was intended to identify the *meteorological tags* characteristic of the brazilian rainfall regimes.

After the first verification studies of the simulations, using the current CDRD algorithm, some constraints and difficulties associated with the cloud resolving model and radiative transfer equation system were found. As such limitations could not be overcome, for technical reasons, a new line was explored. Considering that during last years the scientific community was already migrating from cloud resolving model simulations to observational/empirical database as *a priori* information, it was opted out for a new approach which consists in the development of a land rainfall retrieval algorithm based on artificial neural Network (ANN) for the Brazilian territory. The ANN will be trained with observational measurements from GMI/DPR matchups (DPR/GMI Combined (CMB) - Version 5) from the GPM satellite core provided by the Colorado State University (CSU). In this way, the new specific objectives are presented:

- Verify the performance (limitations and potentials) of the current *a priori* database CDRD algorithm (not optimized for Brazil, i.e. composed by simulations only for Europe and Africa) in comparison with the ground based information collected during CHUVA (Amazon and Vale) campaigns.

- Generate simulations of microphysical, dynamical and meteorological profiles using the University of Wisconsin – Nonhydrostatic Modeling System (UW-NMS) and generate brightness temperature (TB) simulations using the RTE/RMS (Radiative Transfer Equation Modeling System) for the CHUVA (Amazon and Vale) representative case studies (named as golden cases) and compare with observed TB.

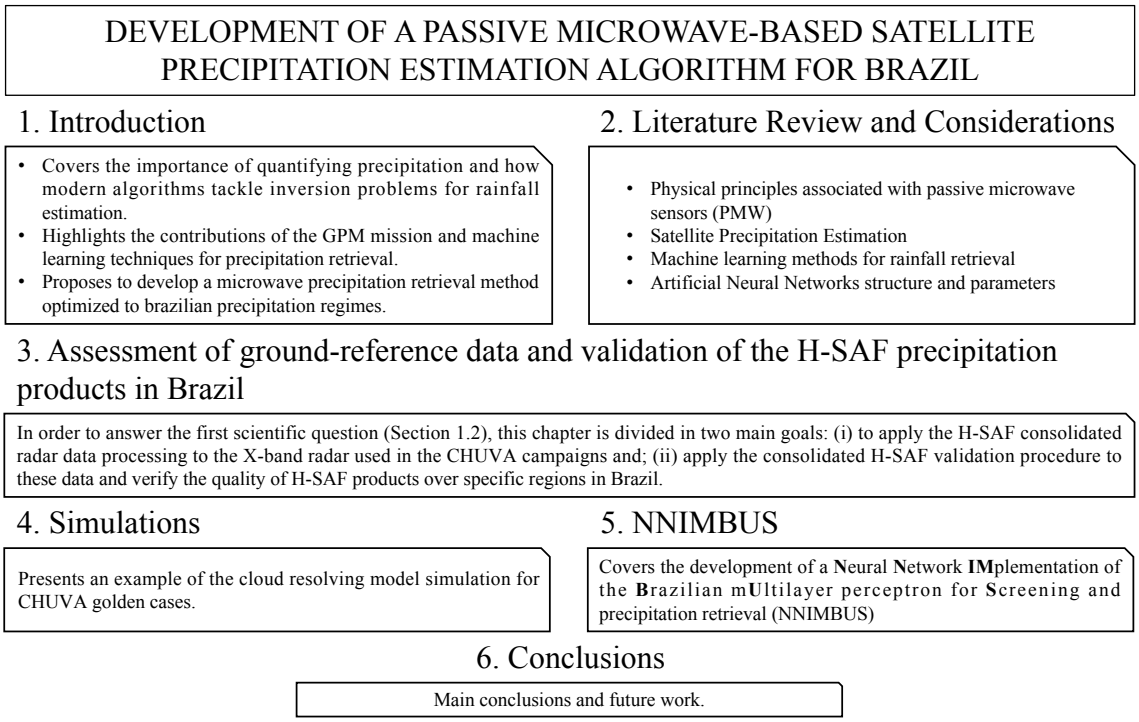
- Develop an optimal ANN architecture for precipitation (rain/no-rain area delineation) and retrieval, through statistical and empirical analysis over the training database.

- Verify the performance of the optimal ANN using independent GMI observations (not used in the training process).

1.3 Document structure

This thesis is structured in five chapters that are presented on Figure 1.1. A brief introduction to the thesis theme with the motivation, scientific questions and objectives is provided in **Chapter 1**. The **Chapter 2** presents a literature overview of the physical principles associated with passive microwave sensors; satellite precipitation estimation; machine learning methods for rainfall retrieval and structure, and main parameters related to artificial neural networks. The **Chapter 3** shows the validation of the H-SAF precipitation products in Brazil during two CHUVA field campaigns. The **Chapter 4** discusses the results concerning the CHUVA golden cases simulated with the UW-NMS and RTE/RMS models. The **Chapter 5** brings up all the steps of the ANN development, from the *a priori* dataset treatment up to the ANN training process and performance verification. Finally, the main conclusions and suggestions for future works are addressed in **Chapter 6**.

Figure 1.1 - Overview of the thesis and its main addressed points.



SOURCE: Author's production

The studies presented from chapters 3 to 5 resulted in the following publications:

Chapter 3: Martins Costa do Amaral, L.; Barbieri, S.; Vila, D.; Puca, S.; Vulpiani, G.; Panegrossi, G.; Biscaro, T.; Sanò, P.; Petracca, M.; Marra, A.C.; Gosset, M.; Dietrich, S. Assessment of Ground-Reference Data and Validation of the H-SAF Precipitation Products in Brazil. Remote Sens. 2018, 10, 1743.

Chapter 4: Poster "Investigation on CDRD and GPROF performance over central Amazon region during GoAmazon/CHUVA campaigns" presented at 8° International Precipitation Working Group (IPWG) and 5° International Workshop on Space-based Snowfall Measurement (IWSSM) Joint Workshop, 2016, Bologna, Italy.

Chapter 5: "Regional land rainfall retrieval based on artificial neural network for the Brazilian territory." *Abstract submitted for the 2019 Joint Satellite Conference - American Meteorological Society.*

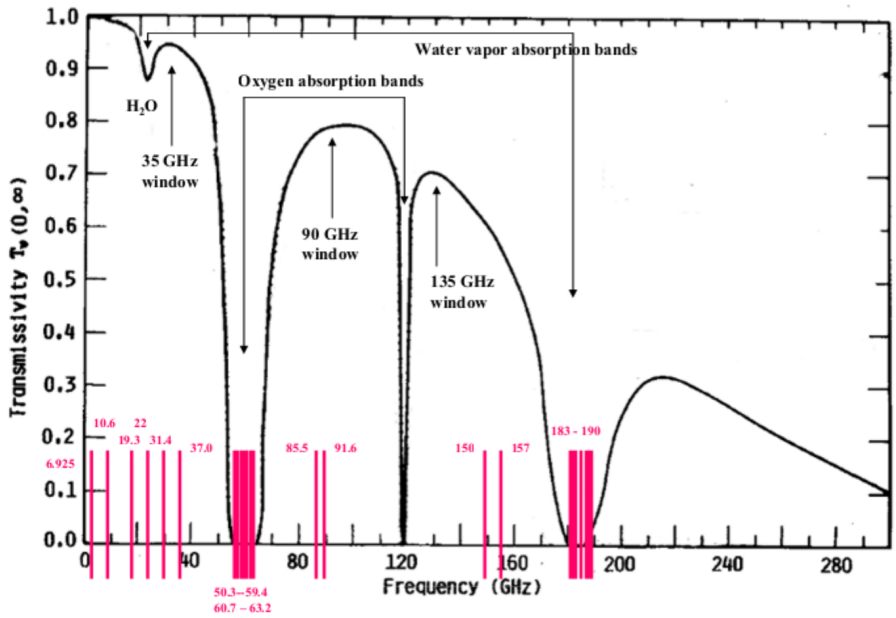
2 LITERATURE REVIEW AND CONSIDERATIONS

This work is based on the passive microwave observations in order to propose an algorithm optimized to retrieve precipitation over the Brazilian territory. Therefore, the following subsections present a review about the main topics related to the this subject: physical principles associated with passive microwave sensors; satellite precipitation estimation; screening of precipitation; recent studies over Brazilian territory; machine learning methods for rainfall retrieval and structure, and main parameters related to artificial neural networks (ANN or NN).

2.1 Physical principles associated with passive microwave sensors (PMW)

Differently from the visible (VIS) and infrared (IR) electromagnetic spectra, the remote sensing techniques using microwave frequencies(MW), millimeter waves and sub-millimeter waves are able to penetrate through clouds and precipitation with sensitivity to hydrometeors (STAELIN, 1981). In the frequencies of MW, hydrometeors are the main source of radiation attenuation. Therefore, the rainfall in MW are physically more direct than VIS/IR. In the absence of scatter targets the predominant process in MW is the absorption and emission with stronger effect at certain intervals due to presence of certain atmospheric constituents, such as O_2 and H_2O. The Figure 2.1 shows the atmospheric transmittance in the MW spectrum, where some absorption bands and atmospheric windows can be observed.

Figure 2.1 - Spectrum of atmospheric transmittance in MW, for a standard atmosphere. Exhibition of the absorption bands and atmospheric windows and channels used by different radiometers in passive MW.



SOURCE: Casella (2010).

Passive microwave sensors (PMW) can be classified as: imager and sounder sensors. The imagers operate, essentially, in the window regions of the MW spectrum, far from absorption lines of water vapor and oxygen. The atmosphere tends to be transparent in these regions, thereby a strong signal can be obtained as a function of the total water in the atmospheric column. Examples of imager sensors that use atmospheric windows are: Special Sensor Microwave/Imager(SSM/I) (HOLLINGER et al., 1990), TRMM's Microwave Imager (TMI) (KUMMEROW et al., 1998), Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (KAWANISHI et al., 2003), AMSR-2 (SHIMODA, 2005) and (IMAOKA et al., 2010) and GPM Microwave Imager (GMI) (DRAPER et al., 2015). These sensors typically operate on conical scan with a fixed tilt angle of 53° along the scan where the signals of the vertical (V) and horizontal (H) polarizations are not mixed as in the cross-track sensors.

The sounder sensors operate in the oxygen absorption (60 and 118GHz) and water vapor (183GHz) lines. These frequencies are a proxy for the temperature and water vapor atmospheric profiles. Example of sounder sensors are: Advanced microwave Sound Unit (AMSU-A and AMSU-B) (ATKINSON, 2001), Microwave Hu-

midity Sounder (MHS) (COSTES et al., 1999) and the Advanced Technology Microwave Sounder (ATMS) (BOUKABARA S. A. AND GARRET K. AND BLACKWELL B., 2011).

The hydrometeors (liquid or fronzen water particles suspended or falling in the atmosphere) can cause measurable changes in the microwave brightness temperature by the processes of absorption, emission and scattering. The liquid hydrometeors (radius less than 50 μm) absorb substantially MW radiation, but scatter barely. Larger liquid hydrometeors (e.g., rain) or the majority of the frozen hydrometeors (e.g., snow, hail and cirrus ice) MW scattering can be significant, mainly in high frequencies. Additionally,these processes of absorption and scattering are dependent of some features, as: hydrometeor phase, shape parameters, particle-size distribution and total water density (GASIEWSKI, 1993).

In term of MW frequency channels, the channels bellow 22GHz are affected by the absorption of the liquid hydrometeors and the ice particles above the rain layer are practically transparent. Meanwhile the channels above 60GHz are affected by the presence of ice scattering. For the channels in the this range (between 22GHz and 60GHz) the radiation interacts with the variety of hydrometeors, being water particles or droplets (liquid or frozen) (LEVIZZANI V., 2002).

The surface type has a important influence in rainfall retrieval. The oceans present relatively constant and low emissivity (approximately 0.4) so that the radiation emitted from it is small being "colder" and precipitation (emissivity around 0.8) appears "warmer" due the increased radiation detected by the sensor through emission. Additionally, the high sea surface polarization contrasts a lot with the low polarization of precipitation. On the other hand, the land surfaces have high and variable emissivities (between 0.7 and 0.9), low polarization and the precipitation increase the upwelling radiation at the same time that absorbs radiation, making it difficult to distinguish rainy areas (Levizzani V. (2002), Biscaro e Morales (2008), Michaelides et al. (2009)). For this reason, the use of high frequencies channels (associated to ice scattering signal) are suitable to retrieve precipitation over continent.

Based on the characteristic that liquid surfaces have different emissivities at different polarizations, Spencer et al. (1989) developed a linear combination of horizontally polarized and vertically polarized brightness temperatures to correct this difference (BISCARO, 2006). As exposed by Cecil e Chronis (2018): "the PCT is then useful for identifying scenes with precipitation, with less ambiguity related to the underlying surface type or surface conditions". Recently, these authors developed a new

formulation for the PCT coefficients for the 10-, 19-, 37- and 89-GHz (and similar) frequencies to minimize differences between land and water surfaces, and also in order to discern deep convection from water surfaces. The authors also highlight that for 10- and 19-GHz frequencies have been non-existent or seldom used in the past and the new formulations for these frequencies hold promise for identification and investigation of intense convection.

2.2 Satellite Precipitation Estimation

The satellite precipitation estimation based in PMW is based on the detection of the emerging electromagnetic radiation from hydrometeors and its conversion in brightness temperature (TB). To obtain the estimate of some variable (precipitation rate, water vapor content, etc.) from the brightness temperatures it is necessary to know the relation between the observed brightness temperatures and the variable in question.

The main algorithms to retrieve precipitation in PMW can be classified as physical, statistical or the combination of both approaches. The physical algorithms use the radiative transfer equation to explain the interaction of the emerging electromagnetic energy and the atmospheric constituents. Meanwhile, the statistical algorithms seek to adjust the relationship between observed precipitation data with the measured TB through empirical and statistical relations (BISCARO; MORALES, 2008). The physical algorithms are composed by two parts: an *a priori* database of atmospheric information and a radiative transfer code. This *a priori* database is usually derived from cloud resolving and numerical weather models, such as Mugnai et al. (1993), Evans et al. (1995), Kummerow et al. (1996), while the radiative transfer code is used to identify the brightness temperature profile associated with hydrometeor profiles and rain rate.

A well-known example of this type of algorithm is the Goddard Profiling Algorithm (GPROF), initially described by (KUMMEROW et al., 1996) for retrieving the instantaneous rainfall rate globally by using a Bayesian approach to match the brightness temperatures to hydrometeor profiles derived from cloud resolving models (CRM). Originally, the algorithm was developed for the SSM/I (KUMMEROW et al., 1996), applied to TMI, through the algorithm 2A12 (KUMMEROW et al., 2001) was also later adapted to the SSMI/S, from the F-16, F-17 and F-18 satellites (VILA et al., 2013). It has been adapted over the years and also improved for GPM, being applicable for different PMW sensors (imagers and sounders) (OLIVEIRA, 2017).

Still in this context, there are also the PMW precipitation products within the EUMETSAT H-SAF (MUGNAI et al., 2013) which are in constant development and refinement to exploit all available radiometers in the GPM constellation and validation activity. The physically based Bayesian Cloud Dynamics and Radiation Database (CDRD) algorithm (CASELLA et al., 2013), (SMITH et al., 2013), (CASELLA et al., 2015) and (CASELLA et al., 2017) is applied to conically scanning radiometers and relies on the incorporation of meteorological and geographical parameters (meteorological tags) to reduce the ambiguities. There is also the Passive Microwave Neural Network Precipitation Retrieval (PNPR) which has been developed and applied primarily to cross-track sensors (Sanò et al. (2015), Sanò et al. (2016)) and recently adapted to the connical GMI radiometer (SANÒ et al., 2018).

Several factors make the quantitative estimation of precipitation from space particularly challenging: the intrinsic large spatial and temporal variability of precipitation; the conversion of radiometric measurements into quantitative precipitation estimates due to the non uniqueness relationship (ambiguity) between surface rain rate (RR) and observed TB (PANEGROSSI et al., 1998); the impact of the background surface on the upwelling radiation, that under some conditions masks the signal originating from the precipitating cloud; and uncertainties associated with rain gauges and radar measurements (i.e., attenuation, beam-blocking) used for verification of consistency and accuracy of precipitation retrieval (TANG et al., 2014).

Notably, for the physical algorithms the quality of the final estimates relies critically on the ability of the coupled "cloud resolving model-weather numerical model-radiative transfer code" in representing the atmospheric variables, microphysical properties and ultimately the multi-frequency PMW TB observation (TB vector) (CASELLA et al., 2013). For example, to improve the quality of the database (VILTARD et al., 2006) created a hybrid database using PR (KUMMEROW et al., 1998) reflectivity profiles and numerical simulations. The PR profiles add a realistic character to the database, but it represents ice poorly. This deficiency is just partly supplied by the cloud models. However, CRM also has difficulties when representing the different types of ice (hail, snow, etc.)

The advent of satellites/missions such as the Tropical Rainfall Measuring Mission (TRMM) (KUMMEROW et al., 1998) and the Global Precipitation Measurement (GPM) (HOU et al., 2014), offered the opportunity to use of a large observational/empirical database of coincident TRMM/GPM space-borne radar (PR/DPR) and radiometer observations (TMI/GMI) instead of a database built from several

simulations (of a CRM coupled to an RTE model).

As exposed by Sanò et al. (2018), an empirical database overcomes the limitations of cloud-radiation model generated databases, such as uncertainties in surface properties characterization (e.g., surface emissivity), single scattering properties of ice or mixed phase hydrometeors and cloud microphysics parameterizations. On the other hand, the use of empirical databases is dependent of the accuracy and limitations of the precipitation products used as reference (GMI/DPR CMB). And the use of the database is limited to radiometers with similar characteristics (i.e., frequencies, polarizations, scanning geometry and spatial resolution). Description of detailed algorithms built with empirical databases can be found in Kummerow et al. (2011), Casella et al. (2017) and Sanò et al. (2018).

For deeper knowledge about the vast majority of satellite retrievals (IR, MW, blended techniques) and its main features (input data, spatial and temporal scales, areal coverage, developers, etc) the reader can access a detailed survey made by (TAPIADOR et al., 2012).

2.3 Rain/no-rain classification (RNC)/screening of precipitation

A main part of the PMW satellite rainfall retrievals is the module that precedes rainfall retrieval denominated rain/no-rain classification (RNC) or screening of precipitation, which is the delineation of rainfall signatures from microwave footprints and is an essential pre-processing step which assists in the estimation of the rainfall intensity (the rainfall retrieval itself). Due the influence of the surface emissivity in the MW frequencies, typically over ocean screening techniques are essentially emission based. Unlike the oceans, over land the algorithms rely solely on the ice scattering phenomenon using the PMW high-frequencies.

The Grody–Ferraro screening methodology Ferraro et al. (1986) and Grody (1991)) has been the most applied methodology for microwave land precipitation since its introduction. As the precipitation is radiometrically characterized at higher frequencies by the scattering by ice, it leads to a depression of TB in 89GHz in relation to TB in channel 24 (both in vertical polarization). The presence of rain was identified when TB in 89GHz (V) was less than 270K and the scattering index (SI) [TB24V – TB89V] was greater than 8K. Subsequent calculations were necessary to determine if the instantaneous fields of view (IFOV) was contaminated by desert, snow, or ice surfaces and are presented by Meyers et al. (2015) as updates for GPROF2010 (KUMMEROW et al., 2001) rainfall screening over land, focusing on recognizing inac-

curacies in separating precipitation from frozen surfaces by applying a climatological snow screen.

Sanò et al. (2018) presented a new rain/no-rain classification scheme (RNC) based on the ANN approach, which provides different rainfall masks for different RR classes. The scheme is based on two years of GPM observational database and due the extremely large size of the database, they used two neural networks, which was more suitable than use multiple parameters in one single complex network. The first ANN input selection was based on the PCA with the addition of two TB-based indexes, SI and PCT. The second ANN is based on the use the polarization signal difference (TB (V) - TB(H)) from the 5 window channels at 10.65, 18.70, 36.5, 89.0 and 166.0 GHz, and the ancillary variables: surface type and the ECMWF T2m. According to the categorization of the training database (associated to three RNC flag values), the two networks return the outputs in terms of the same flag values (0, 1 and 2). At this way, a rain/no-rain classification index (RNCI) is therefore built by combining the outputs from the two ANN.

2.4 Rainfall retrievals studies over Brazil

On the subject of satellite-based precipitation studies over Brazil using microwave channels, it is possible to notice few studies where each one of them was focused in different topics to contribute for the improvement of rainfall satellite estimation over Brazil. They cover topics such as error modeling, microphysical characteristics associated with different regimes of precipitation and also associated with larger error in the estimates, development of methods and rainfall retrieval, etc. It is possible start by mentioning the development of a statistical algorithm to estimate precipitation for the Amazon region made by Biscaro (2006). The algorithm was built based on coincident observations of TMI and PR sensors during the period from January to April 1999. Besides the retrieval of precipitation, the algorithm also includes the development of a screening of precipitation. The screening of precipitation was created based on the identification of the relation of the parameters: a) scatter index (SI), b) a maximum threshold of the PCT and c) the standard deviation of TB in 85GHz (V) in a neighborhood of 5×5 pixels. For the rainfall retrieval, the algorithm considered the Probability Matching Method (PMM) which relates the calculation the cumulative distribution of rain rates and PCT, which give us information about the probabilities of both variables in order to establish a relation between them. For validation procedure, the author compared the performance of the its algorithm (named USProb) with another four widely used algorithms: GPROF, GSCAT and two ver-

sions of the method from Ferraro e Marks (1995) FM_1 and FM_2) used operationally at National Environmental Satellite, Data, and Information Service (NESDIS). In terms of screening of precipitation the GPROF showed the best performance with probability of detection (POD) of 0.763 and false alarm ratio (FAR) of 0.298 in comparison with POD of 0.710 and FAR of 0.398 for the algorithm. In terms of precipitation distribution the USProb has a good performance presenting rainfall average value of 7% higher than the PR, meanwhile FM_1, GSCAT and GPROF presented 250%, 11% and 43% higher and FM_2 presented 16.4% less than PR mean values.

In order to contribute to the estimation of satellite precipitation from warm clouds, Calheiros (2013) analyzed the microphysical and radiative properties in four field campaigns of the CHUVA project (North, Northeast and Southeast regions of Brazil). For this, the author considered measurements from ground-based instruments as MW radiometer, disdrometers and the rainfall events were identified by the vertical profile of reflectivity (VPR) and polarimetric variables derived from a X-Band radar. Non-precipitating cloud characteristics were also analyzed. For precipitating clouds liquid water content (LWC_R) and integrated liquid water content (ILW_R) were analyzed. In relation to non-precipitating clouds the integrated liquid water (ILW_C) and integrated water vapor content (IWV_C) were considered and they presented the highest values in locations close to the coast in the Northeast region. The convective clouds presented higher ILW_R values, followed by stratiform systems and warm clouds. The Vale do Paraíba region and Belém presented the highest reflectivities at higher levels, while in Alcântara and Fortaleza the highest reflectivity values were located in the warm cloud layer. The authors also performed analyzes to quantify the impact of the microphysical and thermodynamic parameters on the emerging radiation signal, but this one was identified as small. However, they observed that the integration of the channel difference in MW can serve as an indicative of the presence of warm clouds and may be useful in the estimation of precipitation by satellite.

As mention by Oliveira (2017) the comprehension and quantification of errors in satellite rainfall estimate products are extremely important for many applications (hydrological modeling, data assimilation systems, etc) and for this reason the author verified the uncertainty of Integrated Multi-satellitE Retrievals for GPM (IMERG) (level-3) and the Goddard Profiling Algorithm (GPROF) (level-2) algorithms in representing the main precipitation systems from Vale do Paraíba and Amazonian regions. The algorithms were compared in relation to ground-based radar obser-

vations, from the Amazon Protection National System (SIPAM) and the X-band dual polarization weather radar (X-band CHUVA radar). In their results, for the Vale do Paraíba campaign the GPROF presented relatively good agreement (spatial distribution and accumulated rainfall), notably for convective events, were there were significant presence of ice scattering. However for the intensity and volume of light/moderate rain rates the algorithm presented overestimation, meanwhile showed underestimation for the light/heavy rain rate classes. For the Amazon region, the IMERG product (using GPROF2014 for GMI sensor) showed significantly overestimation of the frequency of heavy rainfall volumes at around 00:00–04:00 UTC and 15:00–18:00 UTC during the wet season. It was noticed by the authors an evident overestimation over the Negro, Solimões and Amazon rivers due the poorly representation of water bodies surfaces in the algorithm. Diversely, in the dry season the IMERG underestimate the mean precipitation in comparison to the S-band SIPAM radar, since it is derived from isolated convective cells not detected by the algorithm. The authors also carried out a study of error modeling through the Precipitation Uncertainties for Satellite Hydrology (PUSH) framework which demonstrated that the PUSH model was able to predict the error distribution in terms of spatial and intensity distributions.

Recently, Costa (2018) studied the radiative and microphysical properties of clouds that show errors in the rainfall estimation by satellite from passive MW sensors, focusing mainly on clouds where the ice content is not directly associated with precipitation, which are usually the cases of warm clouds. It was analyzed ten years of TRMM data (2002-2011) on the central region of the Amazon. They identified that the frequency distribution of the error resembles Gaussian distribution and they divided it in three classes: consistent (from 20 to 80 percentiles) and the distribution tails representing the underestimation and overestimation. For these last two classes they evaluated the vertical structure of the clouds and it was verified that the underestimation error is related to all cloud properties (rain rate, cloud high, liquid water path (LWP), ice content (IWP), polarization and polarized corrected temperature at 85GHz (PCT85)) meanwhile the overestimation error is only function of IWP. It was also studied the relation between ice content and precipitation, as well the error in the precipitation estimate and the ice content estimators considering the GPM (DPR/GMI Combined) observations from September 2014 until August 2015. From this analysis, they pointed out that there is a problem in determining IWP by GPROF caused by the Bayesian method, where the approach tends to adjust the higher values, which occur less frequently, to lower values that occur more frequently.

2.5 Machine learning methods for rainfall retrieval

The Machine Learning (ML) term can be described as a science that studies the prediction of patterns in data through the use of computational algorithms and statistical models without using explicit instructions to perform this task (KOZA et al., 1996). ML is often considered a subset of the Artificial Intelligence (AI) field and, currently, it is widely used in several daily tasks such as anti-spam filters in emails (GUZELLA; CAMINHAS, 2009), classification of skin cancer (ESTEVA et al., 2017), etc.

There is a plethora of material available on ML topics and methods in the literature, among these Haykin (2009) and Brownlee (2016) makes a great work in comprehensively describing a theoretical foundation of ML algorithms. Broadly speaking, they describe ML in two main topics: supervised and unsupervised learning. The supervised learning is where given a set of input variables X and an output variable Y and an algorithm is used to learn the relation between (by means of a function) the input and output information (like trying to predict for example $Y = rainrate$ given the available $X = TB$). The goal is to approximate this function so well that when there is new input data (TB), the algorithm can predict the output variables (rain rate) effectively from that learned process (generalization capability). In this kind of learning, the network is able to compare its own outcomes with the observed outputs provided, adjusting the weight given to the synapses in order to reduce the difference between the observed outputs and the network estimation (TAPIADOR et al., 2004). Example of this type of algorithms are:

- Classification: is when the output variable is a category.
- Regression: is when the output problem variable is a real value.

Some popular examples of supervised ML algorithms are: Linear regression (for regression problems), random forest (for classification and regression problems) and support vector machine (SVM) for classification problems.

On the other hand the unsupervised learning consists of input data (TB) but no corresponding output target variables (rain rate) and the aim is to model the fundamental structure or distribution in the data in order to learn more about that data and there is not a reference to compared check if the model is performing well during the training phase. Examples of unsupervised algorithms are: K-means (for clustering problems) and Apriori algorithm for association rule problems.

In the ML context, the Artificial Neural Networks (ANN) are computational models inspired by the neural structure of intelligent organisms and the imitation of the elements and organisms from nature itself with the purpose of enhancing the human experience within its environment is called Biomimetics or Biomimicry (VINCENT et al.,). Much like the human brain, the ANN is also composed of interconnected artificial neurons that transfer signals to each other through edges (similar to the axons in the organic neuron), further propagating these signals through weights and activation functions (illustrated in Figure 2.2 that apply mathematical transformations on those signals in one or several hidden layers of the network. The specific parameters that compose a ANN will be further detailed in the next Subsection 2.6.

In the atmospheric sciences field, we can cite the use of multilayer perceptron Neural Networks (NN) for many applications as Marzban e Stumpf (1996) to predict the existence of tornadoes, McCann (1992) to forecast the presence or absence of significant thunderstorms and Navone e Ceccatto (1994) for Indian monsoon rainfall forecasting.

Compared to pure physically-based methods for rainfall retrieval, some potentialities can be considered in the use of ANN: the ability to extract nonlinear relationships is a very valuable feature, it can also deal with large amounts of data in a automated context, presenting adaptability, learning from examples, also offering parallel and distributed processing capabilities, generalization capacity, fault tolerance and low computational costs (after training) (HAYKIN, 1998) and (ANOCHI, 2015). Moreover, the literature suggests that ANN models outperform other statistical modeling approaches when it comes to data that exhibits non-parametric behaviour (GARDNER; DORLING, 1998), not necessarily fitting a normal distribution. Which is often the case of the nature of the precipitation data. Gardner e Dorling (1998) also suggests that the dynamics of severe weather phenomena are not easily included in current numerical weather prediction models due to their small scale, and often cases, the currently limited understanding of them.

It is worth mentioning as an example of ANN applied to rainfall estimation, the H-SAF Passive microwave Neural network Precipitation Retrieval algorithm (PNPR) adapted for cross-track scanning radiometers AMSU/MHS (SANÒ et al., 2015), ATMS (SANÒ et al., 2016) and for the conical radiometer GMI (SANÒ et al., 2018). Another algorithm based in ANN neural networks as an adaptive system is the PERSIANN (Precipitation Estimation from Remotely Sensed Information using Neural Network), described in detail by (HSU et al., 1997). The fundamental algorithm is

based on a ANN and can, therefore, be easily adapted to incorporate relevant information as it becomes available (IR and MW imagery, rain gauge and radar data).

Meyer et al. (2015) compared four ML algorithms — random forests (RF), neural networks (NN), averaged neural networks (AVNN) and support vector machines (SVM) — for rainfall area detection and rainfall rate assignment using MSG SEVIRI data over Germany. Satellite-based proxies for cloud top height, cloud top temperature, cloud phase and cloud water path were considered as predictor variables. The results indicated an overestimation of rainfall area delineation regardless of the ML algorithm (averaged bias = 1.8) but a high probability of detection ranging from 81% (SVM) to 85% (NN). Though the differences in the algorithms performance were rather small, NN and AVNN were identified as the most suitable algorithms. On average, they demonstrated the best performance in rainfall area delineation as well as in rainfall rate assignment. The authors also concluded that NN computational speed is an additional advantage when working with large datasets such as in remote sensing based rainfall retrievals.

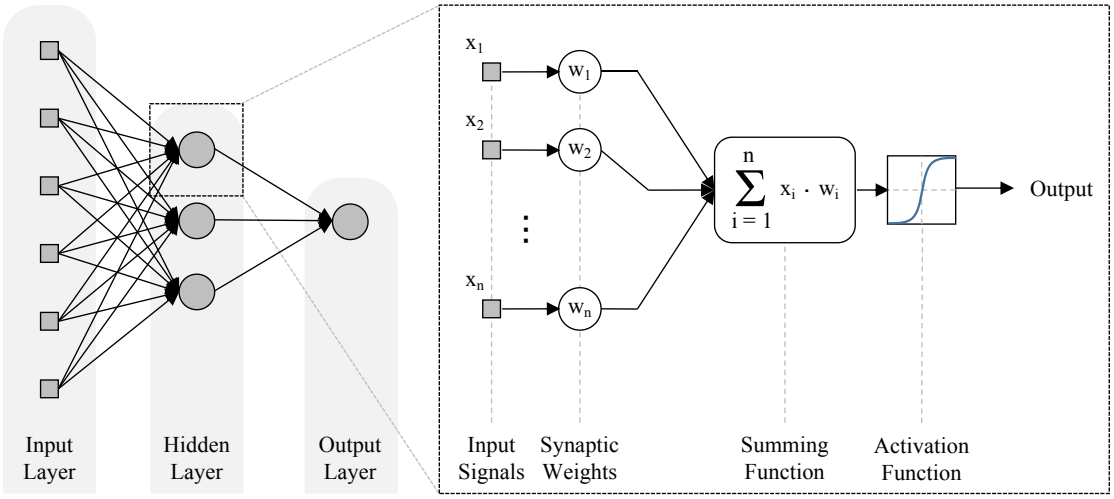
Beusch et al. (2018) developed and compared satellite rainfall retrievals based on generalized linear models and ANN. They used as input predictors the SEVIRI IR channels, their differences, the Satellite Application Facility nowcasting (NWC-SAF) products and geographical auxiliary variables. Their rainfall retrieval was built by two modules: the screening of precipitation, treated as a classification problem and the rainfall intensity retrieval, treated as a regression problem. They demonstrated that use of ANN improved the prediction skill and reduced false alarms.

2.6 Artificial Neural Networks structure and parameters

The schemes in Figure 2.2 show the main segments that compose a ANN architecture. The initial information of the ANN are the input variables (or **predictor attributes**) as independent variables and the output variable (**predicted attributes**) as the dependent variables. The **neuron** is the information processing unit, in which the refinement of the function will be done iteratively by adjusting **weights** for each neuron. Broadly speaking, it is possible to say that the learning of the network is to identify the best set of weights for a database, therefore the knowledge of the network lies in the weights found for each neuron. The calculation of the best set of weights occurs in the **synapses** where the input data is multiplied by the synaptic weight w_i and this process is represented by the **summing function**.

Ultimately the **activation function** defines the neuron output by limiting the neu-

Figure 2.2 - Generic architecture of an ANN and a neuron model description.



SOURCE: Adapted from Haykin (2009).

ron amplitude. There is a vast set of functions and activation functions, and they can be chosen depending on the nature of the data and problem (classification and regression). The following list describes some of the most popular activation functions and their respective ranges: identity $(-\infty, \infty)$, sigmoid (0,1), tanh (-1,1), relu $[0,\infty)$ and softplus $(0,\infty)$.

As explained by Brownlee (2016) s the error related to each iteration and the aim is to minimize the error. Important parameters arise from the iterative process. The **cost function** represents the error calculated from the observed output and the outcome estimated by the ANN and the aim is to minimize the error. For example, the **batch size** is how many records will be considered for each weight update and the **number of epochs** is how many times the weight update will be executed. The **learning rate** refers to the amount that the weights are updated in the end of each batch size, meanwhile the **momentum** controls how much the previous update should influence the update of the actual weight. There is also the **weight initialization** that refers to a suite of different techniques to choose the initial small random values.

Another crucial component is the **optimizer** which represents the mathematical method to make the weight update in order to minimize the loss function and make the predictions as correct as possible. The loss function by pointing out whether the

error is decreasing or increasing serves as a guide to the optimizer appointing if it is moving in the right or wrong direction.

The most popular optimizer is called **Gradient Descent** being fast, robust, and flexible. Basically, it calculates partial derivatives and then it determines which specific operation should be performed on the weights and adjusts each individual weight based on its gradient. As most popular gradient-based optimizer methods, can be found: Stochastic Gradiente Descent (SGD), Batch Gradient Descent, Adam, Adagrad, Adadelta, Nesterov accelerated gradient, etc.

Although there are currently a wide range of ANN types, for the sake of brevity this work will only concern with the multilayer perceptron, which is the method used in this thesis. For further understanding of the several types of ANN, referring to Haykin (2009) is suggested.

3 ASSESSMENT OF GROUND-REFERENCE DATA AND VALIDATION OF THE H-SAF PRECIPITATION PRODCUTS IN BRAZIL

In order to verify the performance (limitations and potentials) of the current *a priori* database CDRD algorithm (not optimized for Brazil, i.e. composed by simulations only for Europe and Africa) in comparison with the ground based information collected during CHUVA (Amazon and Vale) campaign, a validation of CDRD and PNPR algorithms in Brazil are carried out.

First, an assessment of the reference data (ground-based radar) was done with the collaboration of researcher Stefano Barbieri and a brief summary of this part will be presented here. Full access to the details of data processing can be done through Barbieri S. (2017) and Amaral et al. (2018).

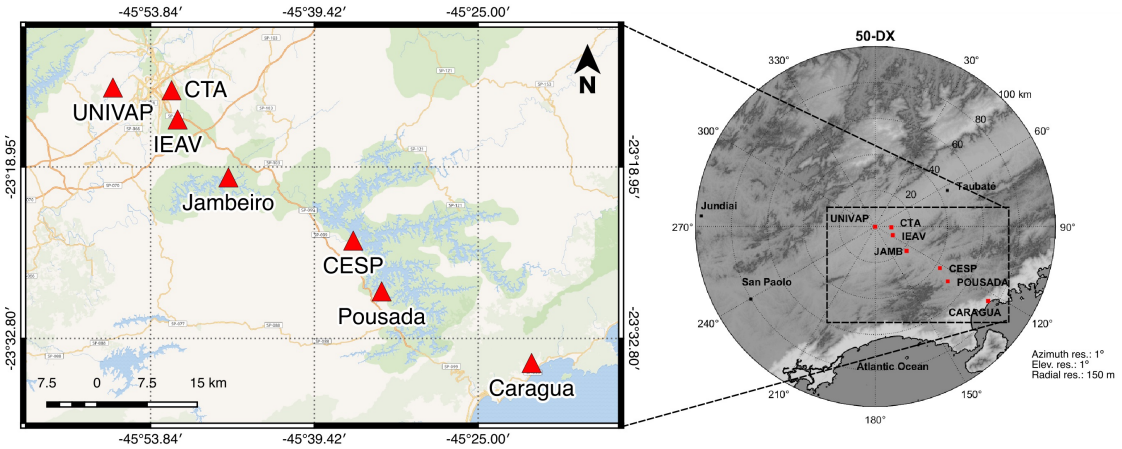
In Section 3.1, the study areas, radar characteristic and rain gauge distribution are described. In the same section, the methodology for radar data treatment (quality index) and radar-based rainfall algorithms are briefly described. A brief description of satellite algorithms is also provided. In Section 3.3 the results of the evaluation of rainfall radar estimates with respect to the rain gauges are presented, as well the results from the satellite verification process, where statistical and pixel by pixel evaluations are performed. Finally, a summary of the main results is presented in Section 3.5.

3.1 Study Area and Data Sources (Radar and Rain Gauges)

This study was developed based on two CHUVA field campaigns. The first experiment occurred in the Vale do Paraíba region located in the southeastern part of Brazil and the second campaign took place in Manaus city in the northern region of Brazil (centrally located in the Amazon basin). The satellite-based evaluation process considers the precipitating events with largest rain rates for both campaigns. Specifically, the case studies analyzed in Manaus took place on 15, 21, 23, 24, 25, 26 February and 2 and 8 March 2014 (8 days) while in the Vale do Paraíba campaign 6 days were considered: 11, 13 November and 1, 8, 14 and 20 December 2011 for a total of 14 precipitating events. For both campaigns the X-band polarimetric radar, manufactured by Gematronik (Germany) has been employed with the main characteristics: Magnetron with 35 Kw per channel, simultaneous horizontal and vertical polarization, pulse width of 0.5μs, operative pulse repetition frequency (PRF) of 1500 Hz, 1.8 m antenna diameter, 1.3° beam width, operation frequency of 9.375 GHz, 150 m of range resolution and maximum distance of 100 km.

The Vale do Paraíba campaign occurred in São Paulo State in an elevated valley between the Serra da Mantiqueira and Serra do Mar mountain ranges. This field campaign had the longest duration for the CHUVA experiment, with an Intensive Observation Period (IOP) starting from 1 November to 22 December 2011, followed by a second period with less intensive measurements through 31 March 2012. The site strategy is indicated in Figure 3.1, where the X-band polarimetric radar was installed near São José dos Campos (Lat. 23°12′31.33″S, Lon. 45°57′7.87″W, 650 m ASL) above the roof of the UNIVAP building (Vale do Paraíba University) being approximately 82 km inland from the ocean. The radar scanning strategy produced a volume scan with 13 elevations (varying from 1 to 25 degrees) with repetition time each 6 minutes. Seven measurement sites (called UNIVAP, CTA, IEAV, Jambeiro, CESP, Pousada and Caragua) were established and equipped with rain gauges located at 9, 11, 22, 43, 51 and 75 km from the radar site, respectively, along a perpendicular line towards the ocean (Figure 3.1).

Figure 3.1 - Position of the X-band radar and of the rain gauges indicated with a thumbtack in each site, during the Vale do Paraíba campaign.



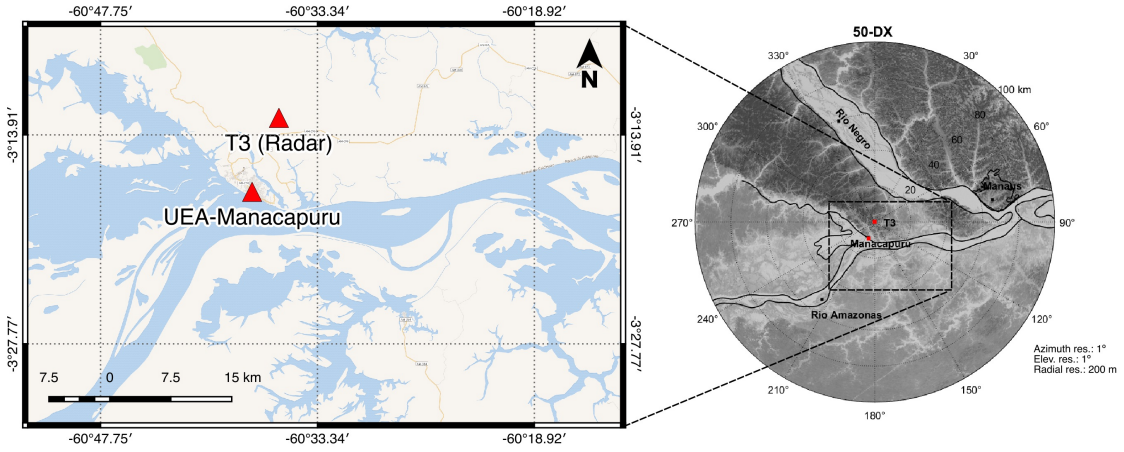
SOURCE: Adapted from Barbieri S. (2017).

The main rainfall systems that were observed during the campaign were caused by the penetration of cold fronts, local convection and organized mesoscale systems. The presence of a 500 hPa trough to the east (first half of the period) and west (second part) was responsible for the atmospheric moisture flow over the region and,

consequently, the precipitation regime. Several thunderstorm events, some associated with hail, were reported during the campaign (CALHEIROS; MACHADO, 2014).

During the Manaus campaign, the X-band polarimetric radar was installed in the Amazon rainforest about 60 km from Manaus, between the Negro and Amazon rivers (Lat. 3°12′46.86″S, Lon. 60°35′53.92″W, 69 m ASL). The X-band radar scanning strategy produced one volume scan with 15 elevations (varying from 0.5 to 30 degrees) every 10 min. Two measurement sites (called T3 and Manacapuru) were established and equipped with rain gauges. T3 is located in the same position of the radar site while Manacapuru is approximately 10 km from the radar site (Figure 3.2). The field campaign occurred in two IOPs, the first happened from 13 February to 31 March 2014 during the wet season, and the second one between 1 and 30 September 2014 at the end of the dry season.

Figure 3.2 - Position of the X-band radar and of the rain gauges indicated with a thumbtack in each site, during Manaus campaign.



SOURCE: Adapted from Barbieri S. (2017).

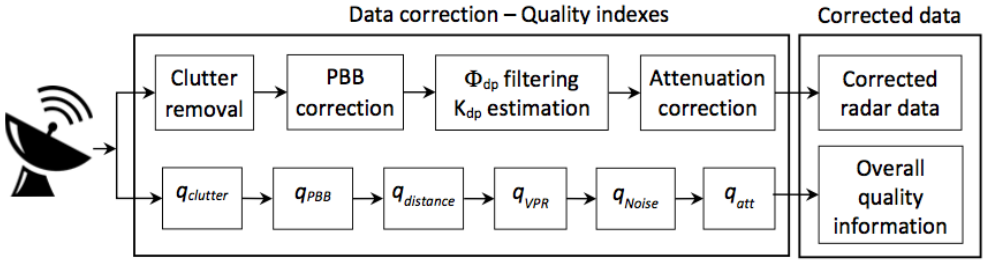
The austral winter corresponds to the dry season in most of the Amazon region, although it represents a rainy maximum for the far northwest of the basin (NUNES et al., 2016). Well-defined wet and dry seasons are associated with the so-called South American Monsoon System (SAMS) (CARVALHO et al., 2011).

3.2 Data and methodology

3.2.1 Radar Data Quality Index

As considered by Amaral et al. (2018) to perform a reliable validation procedure, the common validation methodology developed by the H-SAF Precipitation Product Validation Group (PPVG) was applied to the CHUVA radar data. The methodology aims to compensate, minimize or eliminate the uncertainties by applying a quality index where each source of error is considered. The main error sources considered were: ground clutter, partial beam blocking (PBB), range distance, non-uniform vertical profiles of reflectivity (VPR), differential phase processing and rain induced attenuation (BRINGI; CHANDRASEKAR, 2001) and are shown in Figure 3.3.

Figure 3.3 - Weather radar data processing chain.



SOURCE: Adapted from Barbieri S. (2017).

As exposed by Barbieri S. (2017) the overall quality index is composed by multiplying the partial indices as shown in the following equations:

$$Q = q_{blank} \cdot q_{range} \cdot max\left\{q_{loss} \cdot q_{noise}\right\} \tag{3.1}$$

where q_{range} consider the geometric characteristics (as distance, elevation) defined as:

$$q_{range} = q_{distance} \cdot q_{VPR} \tag{3.2}$$

And the q_{loss} group the quality index related to the signal (Z_h) correction and q_{noise} related to the signal K_{DP}:

$$q_{loss} = q_{PBB} \cdot q_{att} \tag{3.3}$$

Once the quality control was applied to the radar data, tests were also carried out with different precipitation estimation algorithms for the radar rainfall retrieval in order to identify which would be more suitable for the areas under analysis. In the tests, several combinations were analyzed including simple Z-R relations and polarimetric variables as K_{DP}-R relations. As mentioned on the beginning of this chapter the detailed methodology applied for each error source and radar rainfall estimation is presented in detail in Barbieri S. (2017) and Amaral et al. (2018).

3.2.2 Satellite Products and Dataset Generation

The passive microwave (PMW) precipitation products within the EUMETSAT H-SAF are based on the development and refinement of retrieval techniques exploiting all available radiometers in the GPM constellation (MUGNAI et al., 2013). In this context, operational PMW precipitation products for the different radiometers are being released within H-SAF and they are based on two approaches (MUGNAI et al., 2013): the physically based Bayesian Cloud Dynamics and Radiation Database (CDRD) algorithm (CASELLA et al., 2013; SMITH et al., 2013) for conically scanning radiometers and the Passive microwave Neural network Precipitation Retrieval algorithm (PNPR) for cross-track scanning radiometers (SANÒ et al., 2015; SANÒ et al., 2016). Three PMW H-SAF products were considered in this study: H01 (CDRD approach applied to SSMIS), H02 (PNPR developed for AMSU/MHS), and H18 (PNPR adapted to ATMS).

The algorithms are based on the use of a cloud-radiation database made up of thousands of microphysical-meteorological profiles derived from cloud-resolving model simulations of different precipitation events including 60 simulations over the European/Mediterranean area (CASELLA et al., 2013) and 34 simulations over Africa and Southern Atlantic. The main features of each product can be accessed in detail in the respective references cited above.

To perform the validation over Brazil, we had to acquire all the input data (brightness temperatures in TB) for the the CHUVA experiments timeframe, and then, to process the H-SAF products to extend the coverage of the retrievals in order to include the whole country (extended to 75°N–60°S and 80°W–80°E). The inputs for the H01 were the SSMI/S orbits of the DMSP F16, F17 and F18 satellites. For the H02 algorithm, the input files were the AMSU-A, AMSU-B and MHS orbits

from NOAA-18, NOAA-19, MetOp-A and MetOp-B. In addition, lastly, we used the ATMS orbits from the Suomi-NPP satellite as input for the H18 algorithm. Though we have analyzed the three products: H01, H02 and H18 for the Manaus campaign, we only analyzed the H01 and H02 products for the Vale do Paraíba campaign because the ATMS data was not available in 2011.

3.2.3 Application of the Common Validation Code (CVC)

The common validation code (CVC) developed by the Precipitation Product Validation Group (PPVG) enables implementation of a common validation procedure to make the validation results comparable (PUCA et al., 2014). The products to be validated differ in terms of retrieval technique, spatial and temporal resolutions. Therefore, each product requires a specific validation procedure. The methodology can be divided on the following general steps:

- Ground data error analysis;
- Upscaling of radar data to match the satellite product nominal resolution;
- Temporal matching of precipitation products (satellite and ground);
- Application of evaluation statistical methods (continuous and multi-categorical) to all available overpasses, for each pixel pair (satellite-ground).

In order to use the CVC on the Brazilian radar data, the code had to be adapted by including the new radar coordinates (Vale do Paraíba and Manaus) and changing the reading routine to the new radar data. As each campaign had a slightly different radar scan time (6 and 10 min, Vale do Paraíba and Manaus) the CVC was configured to match a maximum temporal difference between satellite and radar of 16 min. Although this may impact a difference of a few minutes in the comparison between the radar and the satellite, at least it ensures that there are at least 3 or 2 radar images to be comparable with one satellite overpass. The radar data was upscaled to the satellite product nominal resolution, considering the antenna pattern (Gaussian function), viewing geometry, and scanning strategy (conical and cross-track) of the MW radiometers. As the radar data was filtered beforehand (quality control), the pixels with low quality were eliminated in the upscale processing. To investigate the performance of the precipitation products, the statistical scores commonly used in the pixel-based validation by the H-SAF PPVG were considered.

3.3 Results

In this section, we summarize the results of the evaluation of rain gauge measurements and the performance radar-based rainfall algorithms. Sequentially, we performed the validation of the H01, H02 and H18 algorithms, breaking it down into two phases, the statistical evaluation and pixel-by-pixel analysis. In order to summarize the results for all case studies, we chose to exhibit one representative event for each campaign; however, the results of the remaining cases will also be discussed further ahead.

A performance verification of the radar-based rainfall retrievals was made regarding the rain gauges (figures and tables were not shown here). The comparison considering the nearest value shows better results. In general, the radar-based rainfall R_Z (just based on reflectivity) presented underestimation, the K_{DP}-based algorithms had good performance when compared with rain gauges measurements. The best estimator was the R_{q2Vu15}, because it presented the perfect correlation coefficient score, which is one. And bias, mean absolute error and root mean square not higher than 0.11, 0.13 and 0.41, respectively.

3.3.1 Validation of H-SAF Precipitation Products

The second part, which is the main focus of this study, is to identify the performance of the H-SAF products in order to provide the algorithm developers with information on the limitations and issues of the retrievals over specific regions in Brazil. The case studies under analysis are the same 14 cases (for both campaigns) that were considered previously on the radar data quality analysis. As mentioned before, three satellite rainfall products were analyzed: H01 (CDRD approach applied to SSMIS), H02b (PNPR v1 developed for AMSU/MHS), and H18 (PNPR v2 adapted to ATMS). We present a statistical analysis based on continuous and dichotomous statistical scores computed within the CVC, along with an in-depth analysis of the selected cases, followed by a pixel by pixel analysis. The considerations concerning the acquisition of the matching pairs (radar $\times$ satellite) are exposed in Section 3.2.3.

3.3.1.1 Statistical Evaluation

As exposed in Barbieri S. (2017) the R_{q2Vu15} algorithm for radar rainfall retrieval presented the best estimate performance and for this reason it was chosen for the statistical evaluation. In order to investigate the impact of different quality indexes in the radar retrievals, we performed a sensitivity analysis (not shown here). According

to this analysis, we found that a quality index value equal or greater than 0.7 could be considered a good compromise between the desired performance of the statistical scores and the size of the dataset (that guarantees the reliability of the results), i.e., the number of pixels in the sample. The statistical scores used in the validation procedure are presented in Tables A.3 and A.2. The continuous statistical scores were computed for the pixels in which both radar and satellite give rainfall estimates larger than 0.25 mm/h (hits only). On the other hand, all pixels are considered for the multicategory scores. The number of pixels for each algorithm is different because it depends on the number of available satellite overpasses.

For Manaus, the events under analysis counted a total of 33 overpass matchings (which means correspondence in time and space for both the satellite and radar measurements) for H01, 49 matchings for H02 and 13 matchings for H18. Since ATMS is aboard a single satellite, the number of overpasses over the region of interest is less than the other sensors (consequently, lower number of matched pixel pairs).

Table 3.1 presents continuous scores for the algorithms in analysis, the number inside the brackets refers to the number of matched pixel pairs for each algorithm. All algorithms tend to overestimate the X-band radar estimates. The H01 presented the largest values for ME, RMSE and FSE. In general H02 had slightly better scores than H01, such as lower mean rainfall rate values, ME, RMSE, SD and FSE. Finally, the H18 presents intermediate scores in relation to the other algorithms, having just the smallest FSE. It is worth to mention also the possibility that radar mean value may be lower due some radar attenuation process, which should be investigated in greater detail in the future. Regarding multicategory scores (Table 3.2), it is worth noticing that both algorithms based on the neural network approach presented better detection skills than H01, the highest POD (0.96 and 0.81), the lowest FAR (0.47 and 0.39) and highest critical CSI (0.51 and 0.53), respectively.

Table 3.1 - Statistical continuous scores for Manaus for H01, H02 and H18. The number inside parenthesis represents the number of matched pixel pairs for each algorithm.

Algorithm	**Sat. Mean**	**Rad. Mean**	**ME**	**RMSE**	**SD**	**FSE**	**CORR**
H01 (406)	4.60	0.93	3.67	5.20	3.69	5.59	0.40
H02 (631)	3.78	1.18	2.59	5.11	4.40	4.32	0.32
H18 (140)	4.53	1.46	3.07	5.10	4.07	3.48	0.34

Table 3.2 - Multicategory scores for Manaus for H01, H02 and H18. The number inside parenthesis concerns the number of matched pixel pairs for each algorithm.

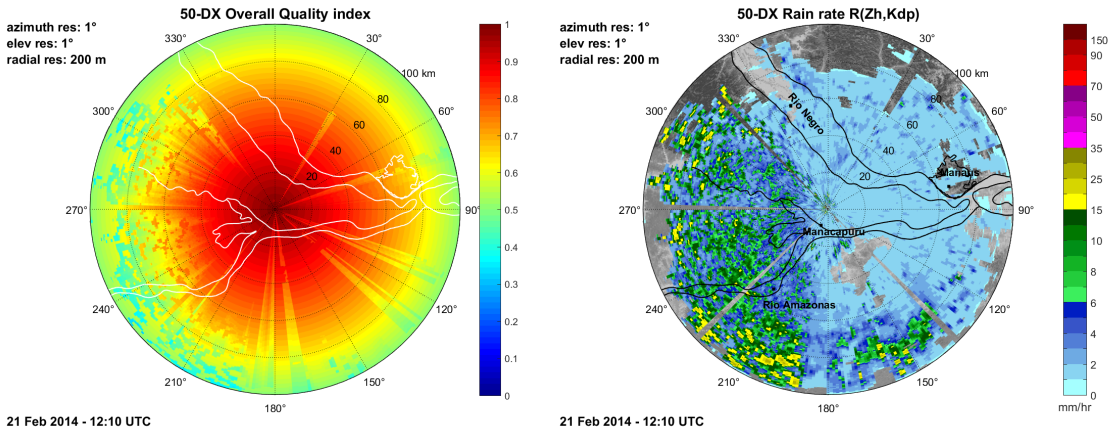
Algorithm	POD	FAR	CSI
H01 (3149)	0.75	0.54	0.39
H02 (1889)	0.96	0.47	0.51
H18 (916)	0.81	0.39	0.53

3.3.1.2 Pixel by Pixel Analysis

In order to further analyze the algorithms performance, a case study analysis for each product and each campaign, with pixel by pixel comparison, was carried out.

A case study for Manaus, which occurred on 21 December 2014 was presented, with the overall quality index shown in Figure 3.3.1.2a and the radar rainfall field in its original resolution in Figure3.3.1.2b. On this day, a well defined squall line approached the X-band radar region. The systems persisted on the region from the morning around 08:20 UTC to the afternoon 15:40 UTC. It is possible to see the presence of convective cores with rainfall rate upwards to 35 mm/h in the regions between the north-west and the south-west quadrants in Figure 3.3.1.2b. The remaining regions are dominated by light and stratiform precipitation with rainfall rates going up to 6 mm/h. In Figure 3.5a it is possible to see the filtered radar data by applying the quality index threshold at 0.7 and upscaled to the satellite native grid, and in Figure 3.5b the respective H01 rainfall retrieval is shown. The light rain rate provided by the radar (light blue), is associated with moderate precipitation values by H01 (shades of green). Additionally, the moderate rain rates from radar (green and dark blue) is strongly overestimated by the H01 (orange to red) algorithm. We can say that H01 has a general tendency to overestimate all rainfall classes.

Figure 3.4 - (**a**) Overall radar quality index and (**b**) Rain rate from radar on 21 February 2014 at 12:10 UTC.



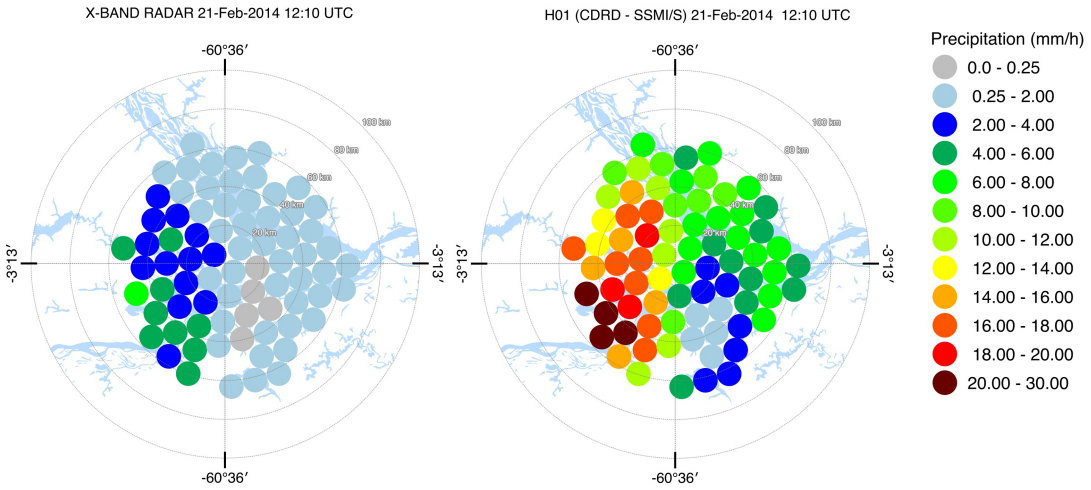
(**a**) (**b**)

production.

SOURCE: Adapted from Barbieri S. (2017).

For H02, one overpass for the case which occurred on 8 March 2014 (at 05:30 UTC) is shown. Convective cells are visible along the border of the Rio Negro River, followed by smaller convective cells over the Southeast quadrant (Figure 3.6). Figure 3.7a,b shows that the H02 algorithm provides precipitations from light to moderate values (shades of green) in pixels where the radar detects absence of precipitation (gray pixels). This tendency to produce a large area of precipitating pixels is related to the precipitation screening, which is the algorithm module where potential precipitating pixels are selected. This feature is related to the relatively high FAR scores (shown in Table A.2).

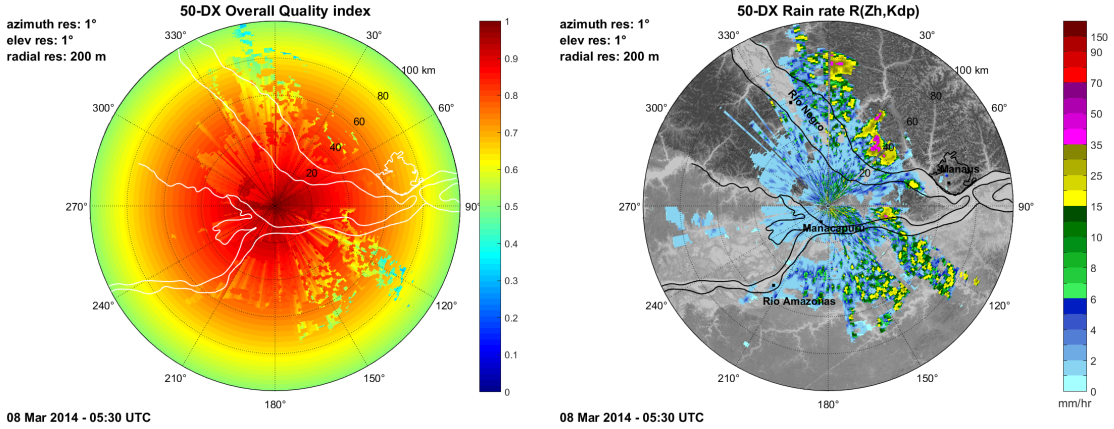
Figure 3.5 - (**a**) Rain rate from radar upscaled to the satellite grid and (**b**) H01 rain rate retrieval in Manaus on 21 February 2014 at 12:10 UTC.



(**a**) (**b**)

SOURCE: Author's production.

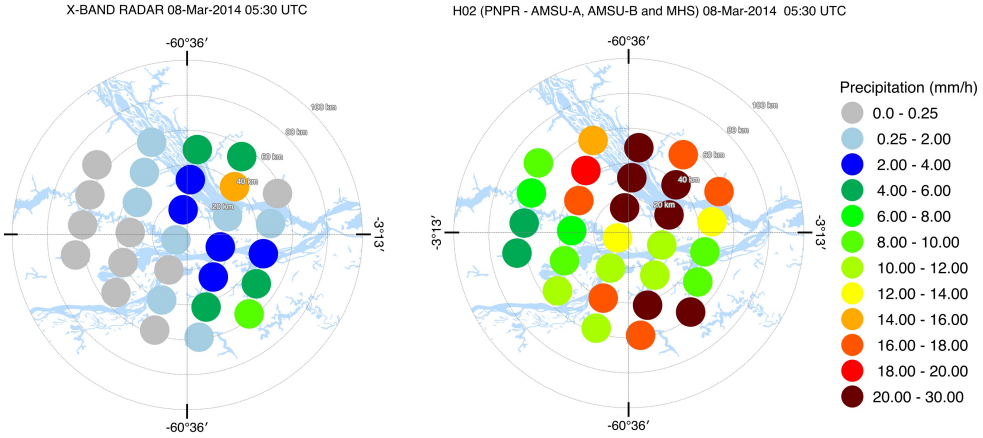
Figure 3.6 - (**a**) Overall radar quality index and (**b**) Rain rate from radar on 8 March 2014 at 05:30 UTC.



(**a**) (**b**)

SOURCE: Adapted from Barbieri S. (2017).

Figure 3.7 - (**a**) Rain rate from radar upscaled to the satellite grid and (**b**) H02 rain rate map in Manaus on 8 March 2014 at 05:30 UTC.

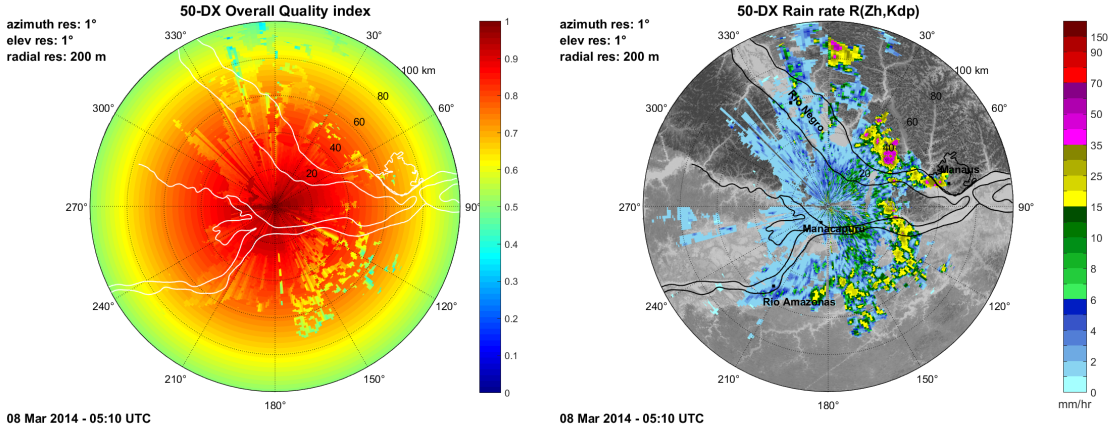


(**a**) (**b**)

SOURCE: Author's production.

The example for H18 shown in Figures 3.8 and 3.9 is for the same event presented in H02, on 8 March 2014 but at 05:10 UTC. Despite the fact that the satellite overpasses (H02 and H18) were very close in time (20 min of difference) it was enough time to see differences in the precipitation field. The H18 precipitation pattern is similar to H02 because they are based on the same precipitation screening method. It is worth pointing out that the screening of precipitation is equal for all products in analysis in this study and it is applied over all background surfaces, except over desert. It is very likely that the screening procedure is not well tuned for the atmospheric conditions of the Amazon region, characterized by the high water vapor content. On the other hand, in the area with the most intense convective cores, the H18 tends to produce a lower overestimation rate when compared to H02. This aspect is related to the difference in the two retrieval algorithms (PNPR for AMSU/MHS and PNPR for ATMS) (as pointed out by (SANÒ et al., 2016), where H18 is based on just one neural network for all surface types, trained with a unique database with the additional channels in the water vapor absorption band at $TB_{183\pm3}$ (SANÒ et al., 2016).

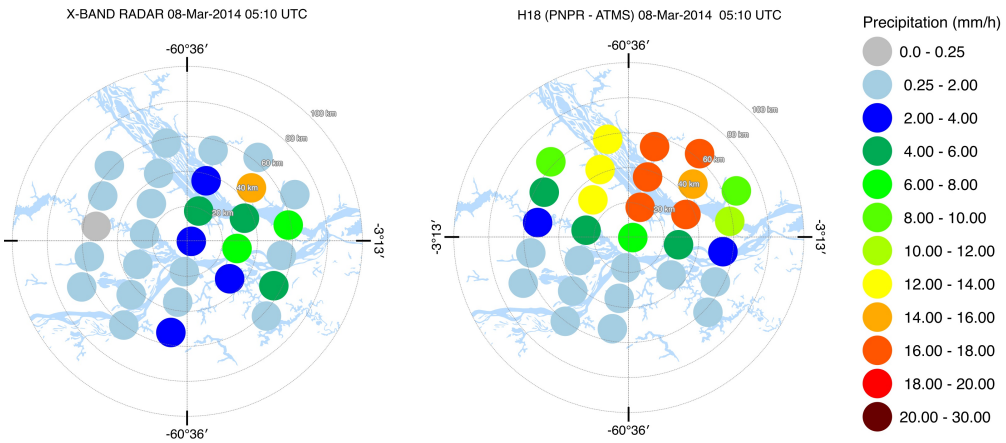
Figure 3.8 - (**a**) Overall radar quality index and (**b**) Rain rate from radar on 8 March 2014 at 05:10 UTC.



(**a**) (**b**)

SOURCE: Adapted from Barbieri S. (2017).

Figure 3.9 - (**a**) Rain rate from radar upscaled to the satellite grid and (**b**) H18 rain rate map in Manaus on 8 March 2014 at 05:10 UTC.



(**a**) (**b**)

SOURCE: Author's production.

It is noticeable that the three algorithms have a tendency to overestimate the larger rain rates. The study in reference (OLIVEIRA et al., 2016) evaluated the ability of two GPM rainfall algorithms (GPROF2014 and IMERG) in reproducing the main characteristics and the diurnal cycle of precipitation as observed by the S-band SIPAM radar, in the Manaus region during the CHUVA campaign. The authors

have found similar results where GPROF2014 (GMI) presents large overestimation of the rain rate volume and occurrence greater than 10 mm h^{-1} during IOP1.

The first example for the Vale do Paraíba campaign refers to the presence of convective clusters associated with the SACZ (South Atlantic Convergence Zone). The coincident overpass between the satellite and the radar occurred at 21:18 UTC and shows the presence of a few intense convective cells in the northwest, north and northeast quadrants of the radar (Figure 3.10b). The overall quality index (Figure 3.10a) in this region is most affected by the blocking effects caused by two mountain ranges, the Serra da Mantiqueira (western) and Serra do Mar (eastern). Analyzing the upscaled maps (Figure 3.11) it is clear that, differently from Manaus, the precipitation pattern is well detected by the H01. Concerning the rainfall estimation, in this case, the tendency of the algorithm is to underestimate the highest precipitation rates seen by the radar. Despite this feature, the estimates are in better agreement with the radar than the estimates in the Manaus campaign (as shown by the lower ME for this region in Table 3.3).

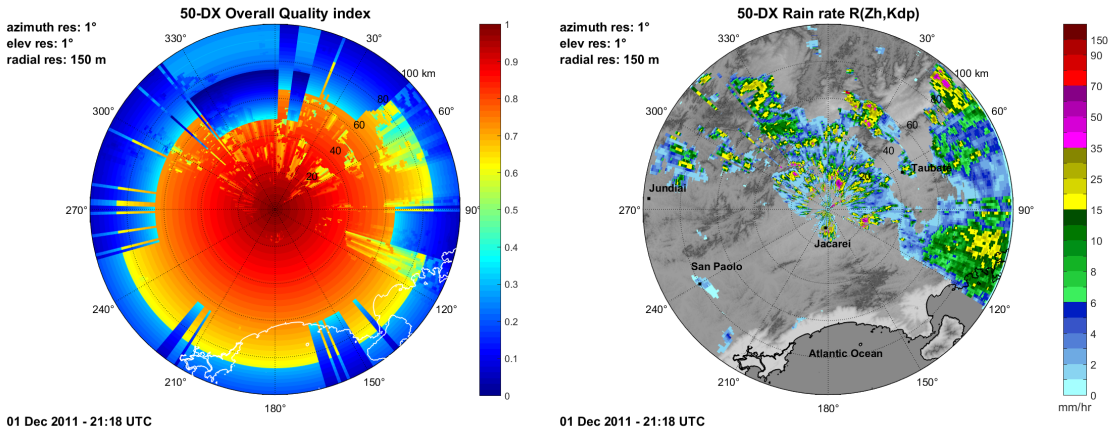
Table 3.3 - Continuous scores for Vale do Paraíba for H01 and H02. The number inside parenthesis concerns the number of matched pixel pairs for each algorithm.

Algorithm	**Sat. Mean**	**Rad. Mean**	**ME**	**RMSE**	**SD**	**FSE**	**CORR**
H01 (181)	2.53	2.89	−0.35	4.00	3.98	1.38	0.44
H02 (124)	2.45	2.76	−0.30	3.56	3.55	1.29	0.37

Table 3.4 - Multi-category scores for Vale do Paraíba for H01 and H02. The number inside parenthesis concerns the number of matched pixel pairs for each algorithm.

Algorithm	**POD**	**FAR**	**CSI**
H01 (2420)	0.49	0.49	0.33
H02 (991)	0.43	0.21	0.39

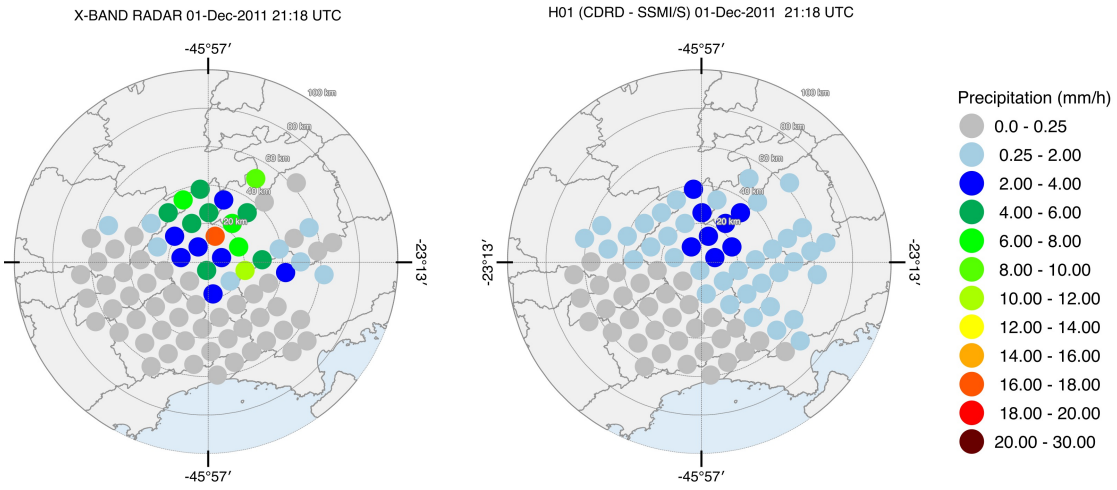
Figure 3.10 - (**a**) Overall radar quality index and (**b**) Rain rate from radar on 1 December 2011 at 21:18 UTC.



(**a**) (**b**)

SOURCE: Adapted from Barbieri S. (2017).

Figure 3.11 - (**a**) Rain rate from radar upscaled to the satellite grid and (**b**) H01 rain rate map in Vale do Paraíba on 1 December 2011 at 21:18 UTC.



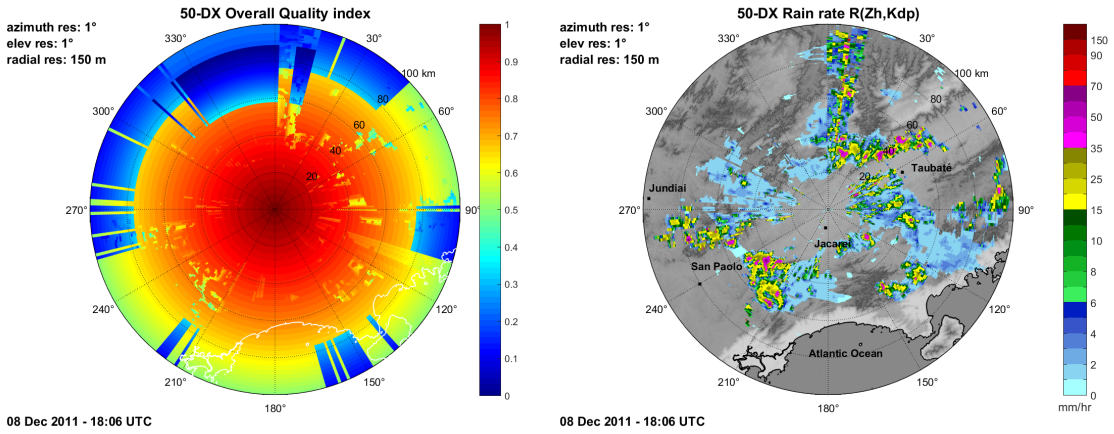
(**b**)

SOURCE: Author's production.

The second example for the Vale do Paraíba campaign (Figures 3.12 and 3.13) consists of the occurrence of local convection with sparse intense convective cells

occurring predominantly in the afternoon (Figure 3.12b). During this event, the precipitation pattern is also quite close to the ground-based reference, which is also reflected on the lowest FAR values (Table 3.4). Concerning the rainfall estimation, in this case, the algorithm tends to underestimate the highest precipitation rates observed by the radar. However, the PMW estimates are in better agreement with the radar observations than the estimates generated in the Manaus campaign (as indicated by the lower ME for this region in Table 3.3).

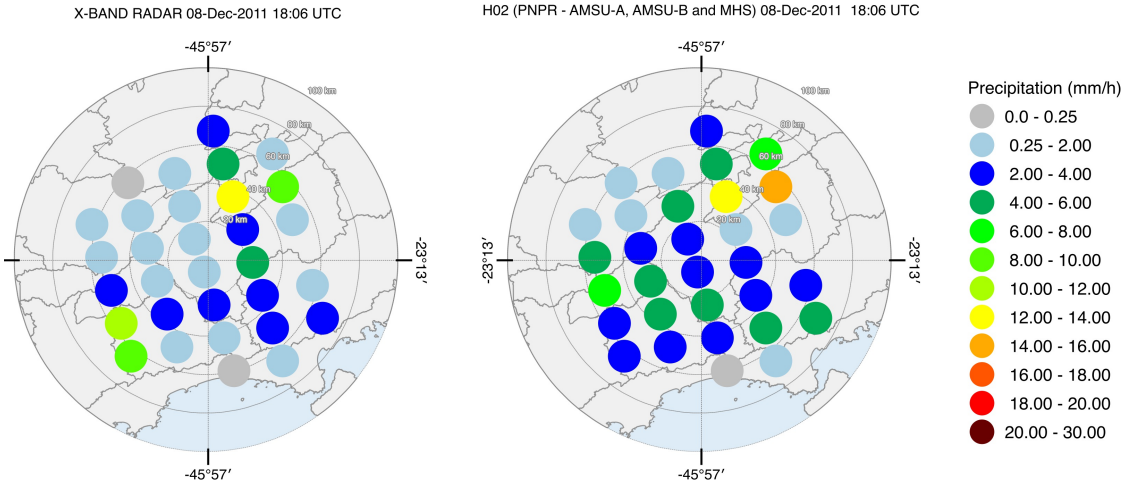
Figure 3.12 - (**a**) Overall radar quality index and (**b**) Rain rate from radar on 8 December 2011 at 18:06 UTC.



(**a**) (**b**)

SOURCE: Adapted from Barbieri S. (2017).

Figure 3.13 - (**a**) Rain rate from radar upscaled to the satellite grid and (**b**) H02 rain rate map in Vale do Paraíba on 8 December 2011 at 18:06 UTC.



(**a**) (**b**)

SOURCE: Author's production.

3.4 Discussion

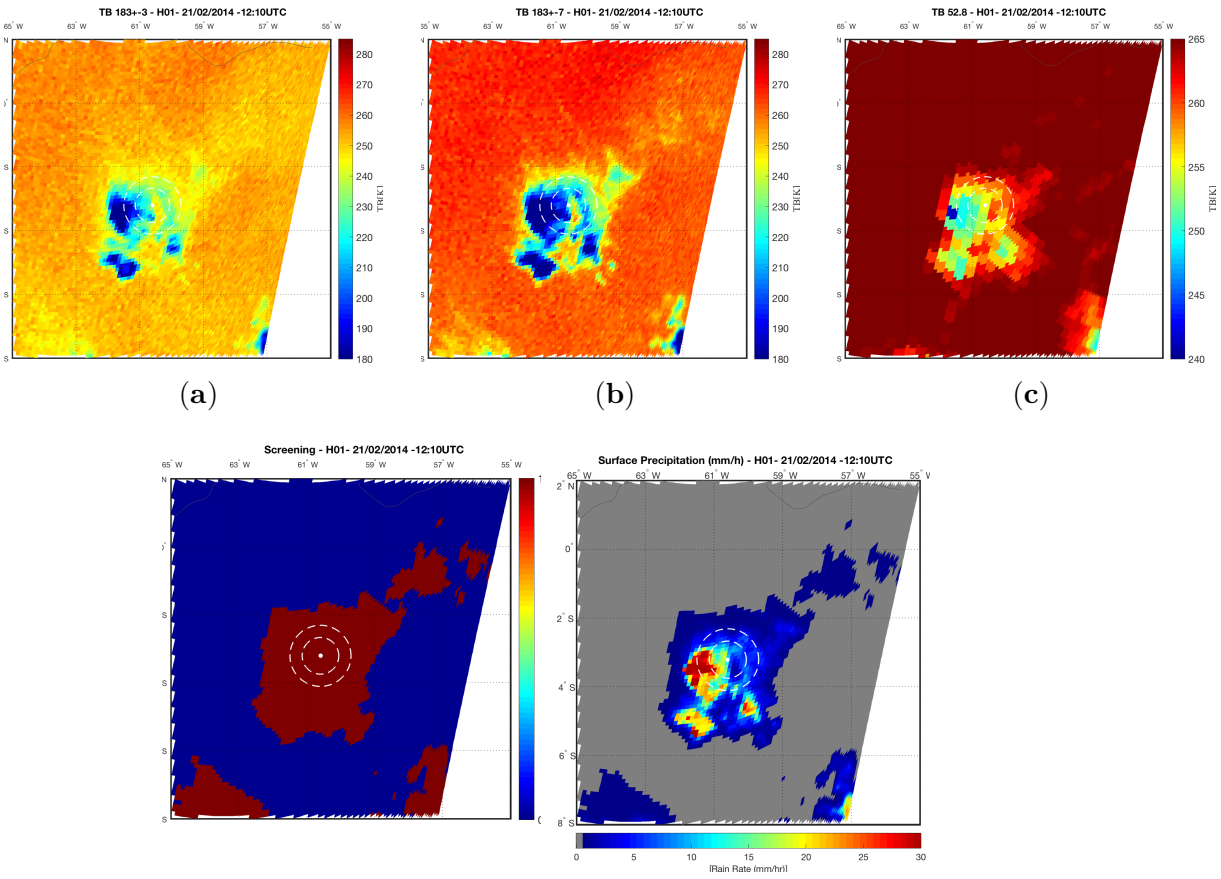
We will discuss the main features of the H01 algorithm that mostly influence the results in Brazilian areas: the precipitation screening process, the surface classification and the representativeness of the cloud model simulations in the *a priori* database. The first part of the retrieval processing, the screening of precipitation, is a important module because defines the amount of pixels selected as potentially precipitating pixels and thus affects the final precipitation field. The screening used in the algorithms is based on the method described by Chen e Staelin (2003), which uses the comparison of the TBs, which are the water vapor absorption band at $TB_{183\pm3}$ and $TB_{183\pm7}$ GHz and the TB_{53}GHz channel in the oxygen absorption band around 50 GHz. A detailed description of the implementation of the H-SAF products screening process can be accessed in Mugnai et al. (2013), Sanò et al. (2015).

Figure 3.14 present the radiative signatures from the $TB_{183\pm3}$, $TB_{183\pm7}$ and TB_{53}GHz channels for Manaus for the case study on 21 February 2014. and we can notice that in Amazon region (Figure 3.14 a,b) both water vapor absorption channels present a strong TB decay (minimal values around 160–180 K). This pattern seems to be associated with a change in the weight function peak of these channels in the presence of high water content, typical of the Amazon rainforest. As

a consequence, the efficiency of the screening can be affected generating an extensive area of precipitation can be seen in the screening map and in the final rainfall rates (Figure 3.14 d,e).

On the other hand, for the Vale do Paraíba region (Figure 3.15), the channels 183 ± 7 and 183 ± 3 GHz do not present an extensive are of strong decay in TB inside the radar area and the screening map (Figure 3.15 d) follows the same pattern that was observed in 183 ± 7 GHz channel (Figure 3.15 b). Differently from Manaus, we can see that the retrieval does not generate precipitation (gray pixels in Figure 3.15 e) in the entire area provided by the screening as potentially precipitating (indicated by red in Figure 3.15 d).

Figure 3.14 - (**a**) $TB_{183\pm3}$ GHz, (**b**) $TB_{183\pm7}$ GHz, (**c**) TB_{53} GHz, (**d**) Screening of precipitation (0—no rain, 1—rain) and (**e**) Surface Precipitation (mm/h) on 21 February 2014 at 12:10 UTC in the Manaus region.

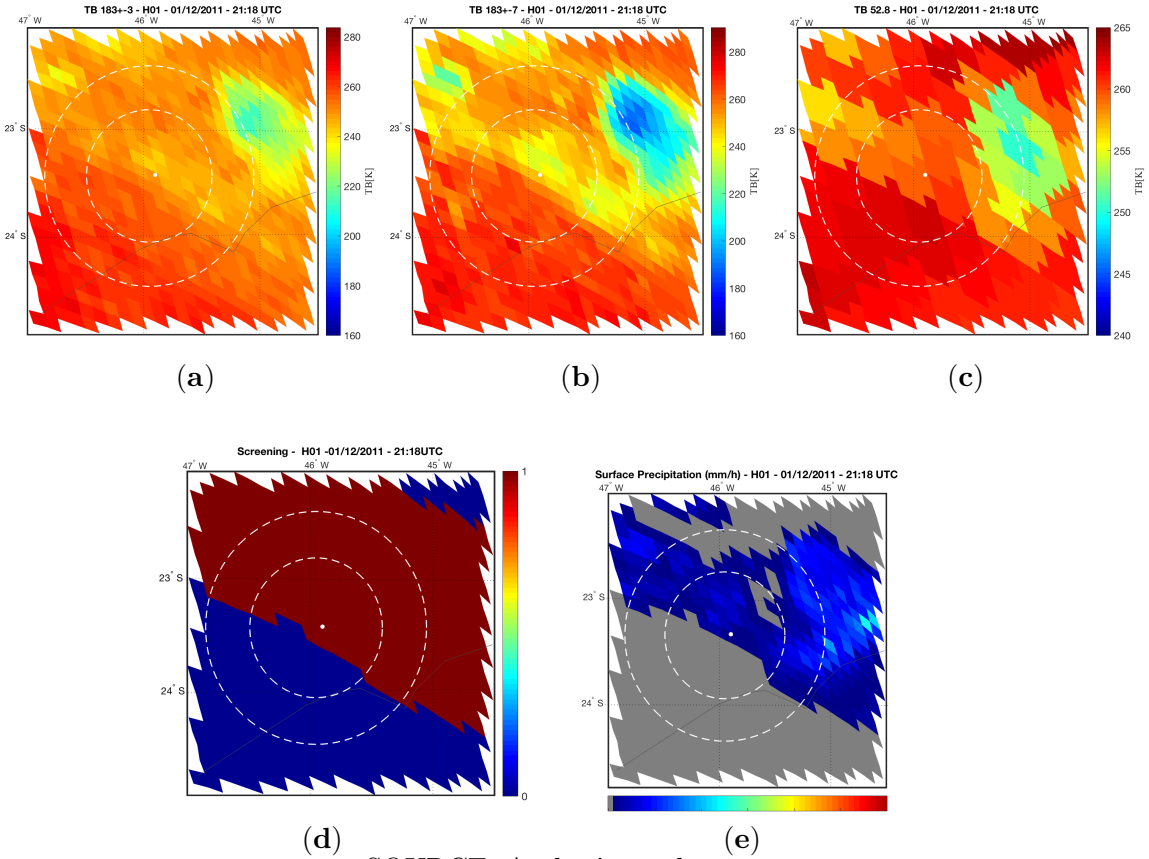


(**a**) (**b**) (**c**)

(**d**) (**e**)

SOURCE: Author's production.

Figure 3.15 - (**a**) $TB_{183\pm3}$ GHz, (**b**) $TB_{183\pm7}$ GHz, (**c**) TB_{53} GHz, (**d**) Screening of precipitation (0—no rain, 1—rain) and (**e**) Surface Precipitation (mm/h) on 1 December 2011 at 21:18 UTC in the Vale do Paraíba region.



(**a**) (**b**) (**c**)

(**d**) (**e**)

SOURCE: Author's production.

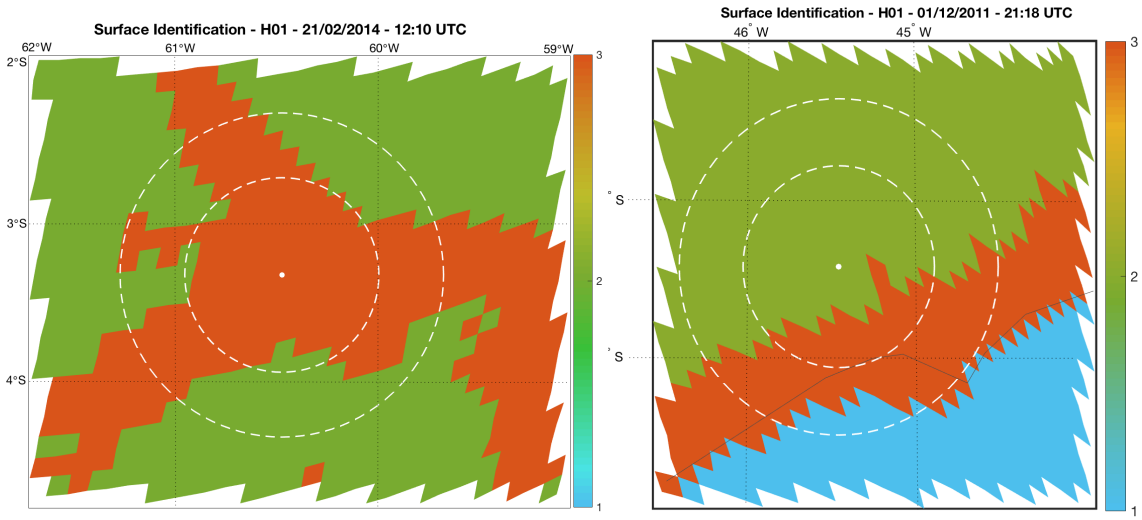
In the Amazon region we have noticed that the surface characteristics were not properly classified. This region is dominated by large inland water bodies (Amazonas and Rio Negro rivers) and surrounded by vegetated land. Figure 3.16a shows that, in the radar area, the algorithm classifies the surface as coastal pixels, which is the least populated database, therefore, the least representative and usually affected by larger uncertainties. In the Vale do Paraíba region (Figure 3.16b) the algorithm considers large part of the radar area as vegetated land and the Bayesian approach look for land profiles, which leads to the choice of more appropriate rain rate profiles.

In the Bayesian approach, the surface classification impacts the choice of channels used in retrieval processing, which influences the selection of hydrometeor profiles

in the *a priori* database and, ultimately, the rainfall rate retrieval.

Finally, the representativeness of the cloud model simulations in the *a priori* database, which is currently optimized for Europe and Africa, does not properly represent the typical microphysical and rainfall profiles of brazilian regions. Improving the database including simulations that represent the brazilian rainfall regimes could solve the issue.

Figure 3.16 - Surface Identification (1—Ocean, 2—Land, 3—Coast) for (**a**) Manaus and (**b**) Vale do Paraíba.



(**a**) (**b**)

SOURCE: Author's production.

3.5 Conclusions

In this chapter was analyzed the data collected by mobile X-band polarimetric radars during two campaigns of the CHUVA project, both in the Vale do Paraíba and Manaus regions. First, in order to ensure that the verification of the H-SAF precipitation products is based on well-treated data, the ground radar data was submitted to a quality control procedure and the rainfall estimates were tested with different algorithms, obtaining the best results the algorithm R_{q2Vu15} based on polarimetric variables (Barbieri S. (2017) and Amaral et al. (2018)).

Regarding the results from the satellite algorithms validation, for the Manaus region, the CDRD algorithm (H01) tends to overestimate all rain rates classes (light to heavy). The PNPR algorithm for AMSU-A/AMSU-B/MHS sensors (H02) presents better POD than H01, but it also presents high FAR values. The PNPR for the ATMS sensor (H18) presents lower overestimation of heavy rain rates when compared to H02, probably due to the different neural network used on H18. It is worth considering that the sample size for this verification study (14 cases) was quite small, and a more extensive validation with a larger ground-based dataset would be suggested to perform a more comprehensive quality-assessment.

All analyzed rainfall retrieval algorithms for the Amazon region showed high FAR values and larger precipitation patterns which are deeply related to the precipitation screening scheme. The screening seems to be substantially affected by the high water vapor content in this region. The H01 (CDRD for SSMIS-Bayesian approach), in specific, it was affected by highly variable surface emissivities, impacting the selection of hydrometeor profiles in the *a priori* database and, ultimately, the rainfall rate retrieval produzing the results found in this assessment. Moreover, we must take into account the uncertainties on precipitation estimates, which in the Bayesian approach are represented by the coefficients of error covariance matrices (for TBs, and ancillary and meteorological parameters) as analyzed in Smith et al. (2013). For the Vale do Paraíba region, both algorithms, H01 and H02, produced ME values that were quite close to zero (or negative) and lower FAR values (from 0.21 to 0.49) than Manaus. Unlike the Manaus case, the precipitation patterns were well detected and the estimations were in good agreement with the reference as indicated by the low ME values.

4 DEVELOPMENT OF THE CLOUD DYNAMICS AND RADIATION DATABASE (CDRD) FOR RAINFALL SATELLITE RETRIEVALS OVER BRAZIL

The initial aim of this study was to develop the cloud-radiation database representative of the physical, microphysical and radiative processes of precipitation systems that operate in different regions of Brazil. Such database to be used in the passive microwave (PMW) Cloud Dynamic Radiation Database retrieval algorithms (CDRD) in order to improve satellite rainfall estimates. The CDRD is a Bayesian algorithm performing the retrieval of precipitation from measurements of conical scanning PMW radiometers (i.e., SSMIS, AMSR-2, GMI). And the focus was to develop the database based on SSMIS overpasses.

For generation of the database a characterization of rainfall in Brazil was take into account the data collected during the campaigns of the CHUVA project (based on selection of golden cases). After identifying these precipitating events, cloud resolving-model simulations for these cases were carried out in order to simulate microphysical profiles associated with radiometric measurements of passive microwaves. The simulations were made from the combination of a model mesoscale University of Wiscosin Nonhydrostatic Modeling System (UW-NMS) and a model of Radiative Transfer Equation Model (RMS). The UW-NMS is a non-hydrostatic, three-dimensional, nested model, ranging from the microphysical scale to the synoptic and with non-Boussinessq dynamics (TRIPOLI, 1992; TRIPOLI; SMITH, 2014a; TRIPOLI; SMITH, 2014b). Once the previous steps are reached the cloud-radiation database can be generated from the simulations and can be implemented in the Bayesian inversion algorithm for generating the estimate of precipitation over Brazil. To achieve this goal some steps were delineated, namely:

For the elaboration of the database,

- Identify the representative cases (“golden cases”) of several type of precipitation systems over Brazil, using as initial basis the CHUVA project data;
- Generate simulations of microphysical, dynamical and meteorological profiles using the University of Wisconsin – Nonhydrostatic Modeling System (UW-NMS) and generate brightness temperature (TB) simulations using the RTE/RMS (Radiative Transfer Equation Modeling System) for the CHUVA (Amazon and Vale) golden cases and compare with observed TB

- Develop the cloud-radiation database simulating the brightness temperatures for different PMW radiometers using of radiative transfer model (Radiative Transfer Equation Modeling System - RMS)
- Identify which are the main meteorological variables (meteorological tags) that influence and characterize the formation of different precipitation systems.
- Evaluate the performance of the CDRD algorithm using the new database developed for the Brazilian territory.

4.1 CHUVA golden cases and model configurations

From the six CHUVA field campaigns, it was decided to initiate the generation of the simulations from the Vale do Paraíba and Manaus campaigns, because they had longer period of data observations. From these campaigns, the selected golden cases are presented in Table 4.1.

Table 4.1 - Selected golden cases of Vale do Paraíba and Manaus campaigns for initial simulations.

CHUVA Campaign	Date	Precipitation System
Vale do Paraíba	11/11/2011	Local convection
Vale do Paraíba	13/11/2011	Frontal system
Vale do Paraíba	01/12/2011	Umidity convergence zone (ZCOU)
Vale do Paraíba	08/12/2011	Local convection
Manaus	08/03/2014	Convective/stratiform system
Manaus	12/03/2014	Convective system
Manaus	19/03/2014	Convective system
Manaus	20/03/2014	Convective system

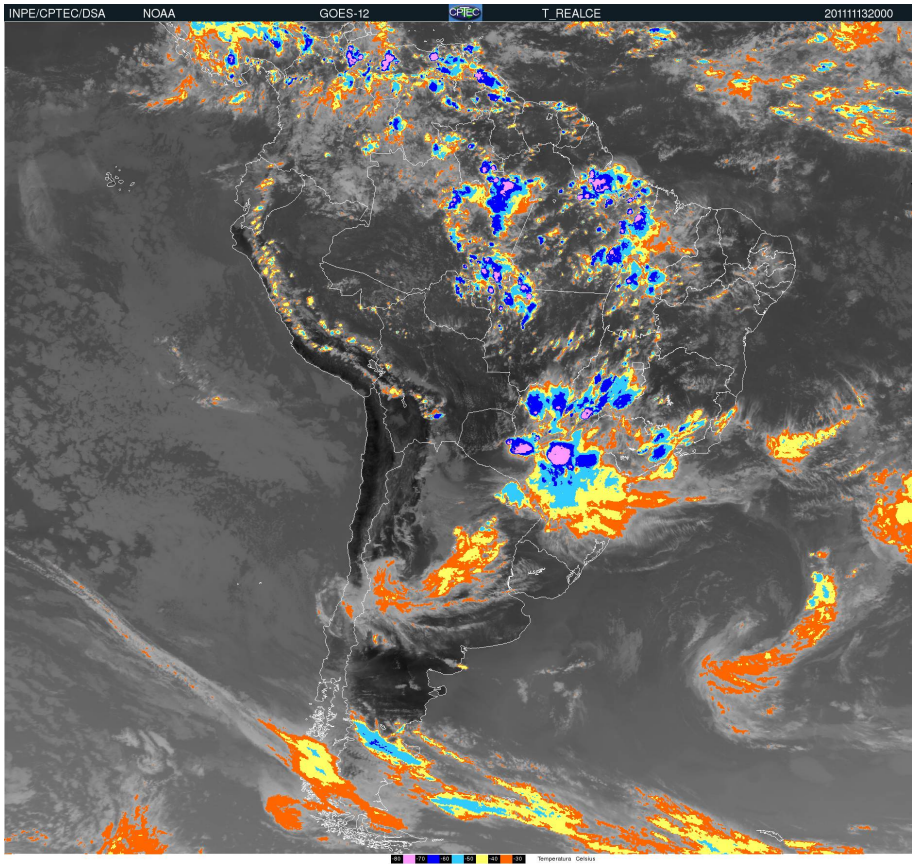
After selection of the golden cases, it is necessary to check the matching satellite PMW sensors (SSMIS F16, F17 and F18) with the region and duration of each event. Beyond to serve as input data for the Bayesian inversion algorithm, this data also allows to characterize the brightness temperatures (TB) measured by these sensors in each precipitation regime. Checking the matching passages of the SSMI/S sensor and obtaining the data (SDR - Data Data Sensor) were made through the Integrated Management System (CLASS-NOAA), which consists of an electronic library for the distribution of data from NOAA operational satellites and derived data.

The initial and boundary conditions for the NMS were from the operational analysis data (every six hours) of the GFS/NOAA model. The model was configured with three nested grids with: 92×92 points (Δ x = $50km \times 50km$), 92×92 points (Δ x = $10km \times 10km$) and 252×252 points (Δ x = $2km \times 2km$). The two most refined grids make up the CRM module and the grid vertical was configured with 36 vertical levels up to 17km with grid-dependent spacing with height. Finally, the RTE system represents the integration of a single scatter model for the multicomponent and heterogeneous hydrometeor cloud-precipitation medium and an Eddington approximation solution to the classical, inelastic, steady state RTE (CASELLA, 2010).

4.2 Simulation results for Vale do Paraíba golden case

The case study that will be presented as an example is the golden case for the Vale do Paraíba campaign that occurred on 13 November 2011 associated with the displacement of a frontal system with strong convection on the southeastern part of South America (Rio Grande do Sul, Santa Catarina and Paraná), as shown in Figure 4.1. Over the northern part of São Paulo some isolated convective activity can be seen. The observational CHUVA data campaign registered precipitation between 20:00 and 00:00UTC. From the CHUVA ground-based measurements the pluviometers and disdrometers on the IEAV site registered precipitation with maximum value around 45mm/h between 20:00 and 00:00 UTC.

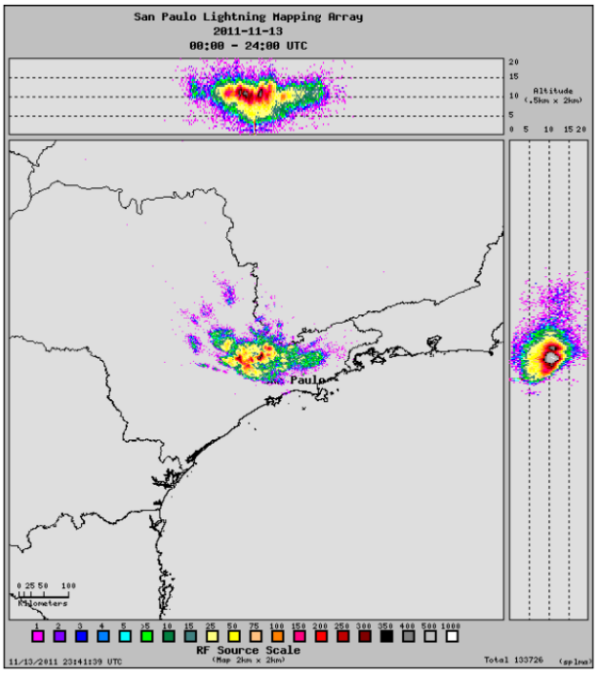
Figure 4.1 - Infrared brightness temperature on 13/11/2011 at 20:00UTC (GOES 12).



SOURCE: DSA/CPTEC/INPE (2011).

Figure 4.2 shows the density of electromagnetic pulses in daily VHF of the LMA network with a resolution of $2km \times 2km$ for the day of 13 November 2011. Although the precipitation was detected from 20:00 UTC, the electrical activity began before recording pulses in VHF as of 17:00 UTC. The VHF pulses were detected with increased density and displacement from the northeast of the state of São Paulo to the south of the state of Rio de Janeiro, with a maximum density of more than 1,000 pulses per 4 km^2, which implies that the region in analysis was affected by strong convective activity.

Figure 4.2 - Density of electromagnetic pulses in daily VHF of the LMA network with resolution of $2km \times 2km$ on 13 November 2011.

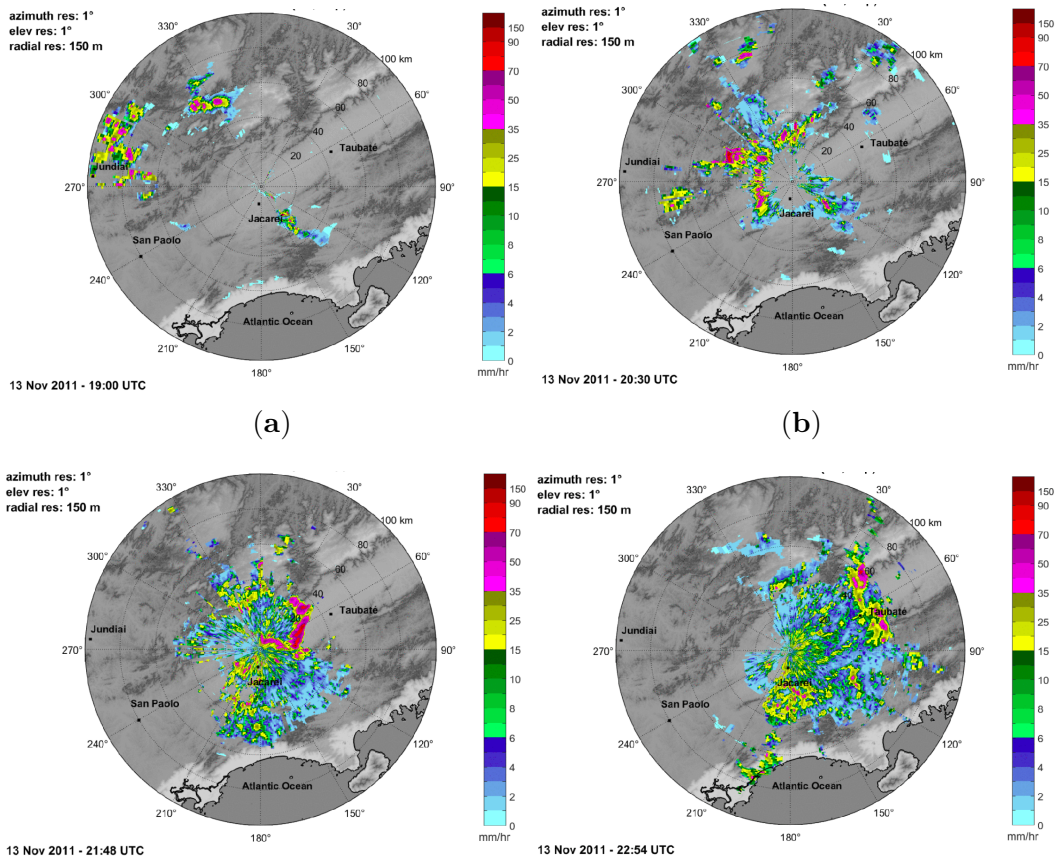


SOURCE: DSA/CPTEC/INPE (2011).

From the ground-based X Band dual-pol radar, Figure 4.3 can be observed, in which systems arise around 19:00 UTC, and convective cores with maximum values of rain rate varying from 50 – 80 mm/h are visible on the northwest sector of the radar (between 270º to 330º azimuth angle). These nuclei move on northwest – southeast direction, crossing the center of the radar around 21:40 UTC which is fairly close to the first coincident satellite overpass (F16 – 21:23 UTC). At around 20:30 UTC the occurrence of attenuation due to intense precipitation (northwest sector) is already noticeable. During the second coincident satellite overpass (F18 – 22:52 UTC) there are few intense convective cores and they are located northeast of the radar (close to the city of Taubaté).

The simulation was made for São José dos Campos from 13 November 2011 starting at 00:00 UTC with 30 hours of duration (with hourly output) where the first 12 hours corresponded to the model spin up time. The simulation ran with the "default" parameters used for the european database. During this event there were 2 SSMIS coincident overpasses (F16 at 21:23 UTC and F18 at 22:52 UTC).

Figure 4.3 - X Band radar observations on 13/11/2011 at a) 19:00 UTC, b) 20:30 UTC, c) 21:48 UTC and d) 22:54 UTC.



(a) (b)

(c) (d)

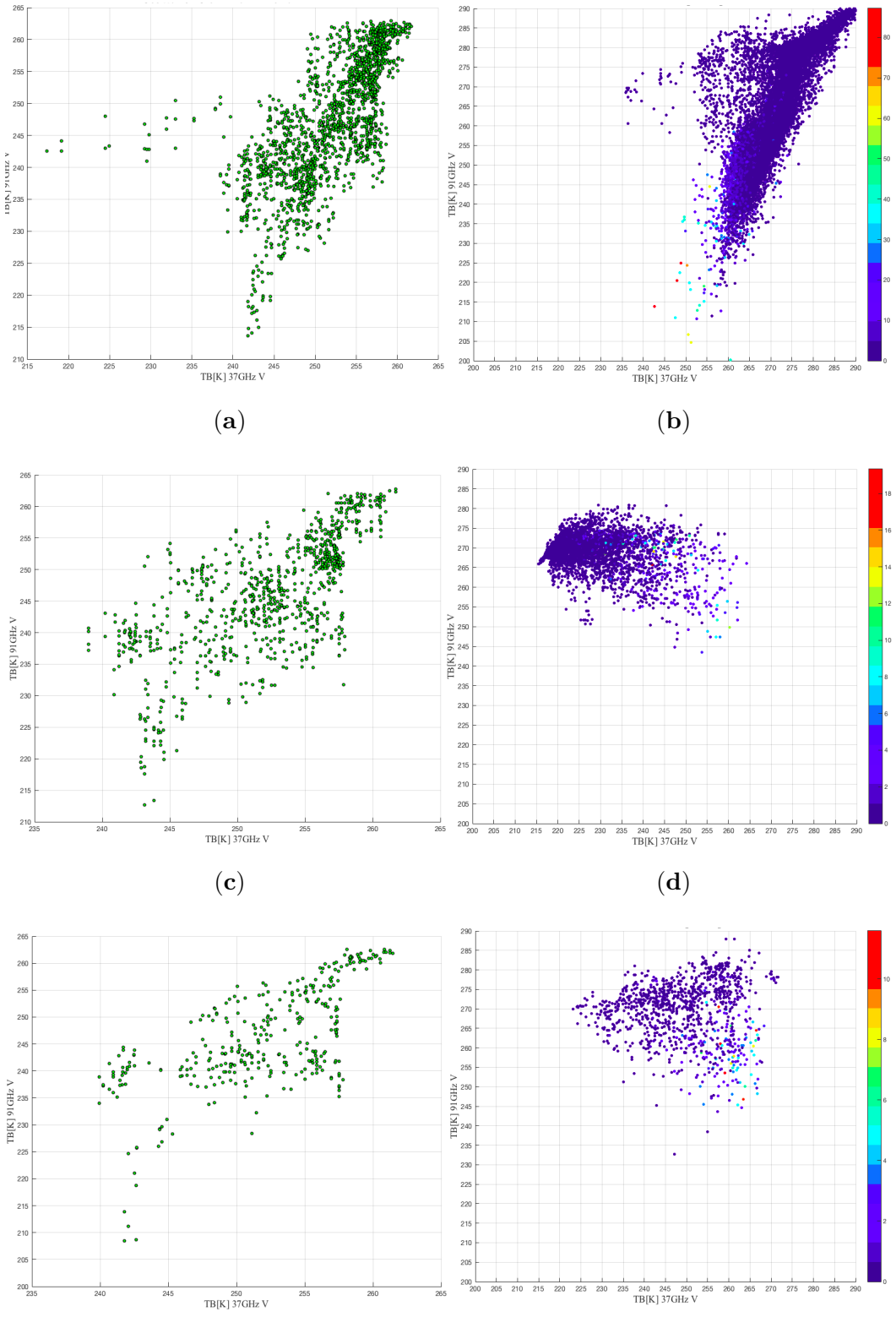
SOURCE: Adapted from Barbieri S. (2017).

To verify the general behavior of the model during the simulations the next figures present the scatter plots of TB observed by the SSMIS (for the two overpasses) and simulated scatter plot of TB x TB x RR divided by surface for the 18 hours of simulations.

The Figure 4.4 presents information from the vertically polarized channels 37GHz and 91GHz from the observations (left) and from simulations (right) for the different surface types. We can see for land pixels (Fig. 4.4a and b) that both populations present a downward linear tendency and the radiative transfer model (Fig. 4.4b) has the tendency to provide a branch of warmer pixels on this type of surface as seen between 270-300K for both temperatures. The precipitating pixels are associated with the coldest pixels for both frequencies, however the simulations gave coldest minimal values for 91GHz (180K) against 215K measured on the observations.

In relation to ocean pixels the observations (Fig. 4.4c) have a relatively high dispersion with a colder branch in 91GHz reaching minimal values around 213K, meanwhile the simulations (Fig. 4.4d) show a different distribution pattern. This population presented a higher number of pixels associated with higher 91GHz temperature values associated with lower 37GHz temperature values and the precipitating pixels does not have a specific pattern, they are highly spread. The observations associated with coast pixels show a tendency relatively linear with a small branch of colder pixels (205K) in 91GHz (Fig. 4.4e) meanwhile the simulations does not reach such cold temperatures (minimum of 230K for 91GHz in Fig. 4.4f). The precipitating pixels are associated with a tendency of colder TB in 91GHz associated with warmer temperatures in 37GHz.

Figure 4.4 - Scatter plot from TB at 37GHz × 91GHz at vertical polarization from SSMIS observations (left) and TB at 37GHz × 91GHz × Rain Rate (mm/h) from all simulation times (right) for a) and b) for land pixels; c) and d) for ocean pixels and e) and f) for coast pixels, respectively.



SOURCE: Author's production.

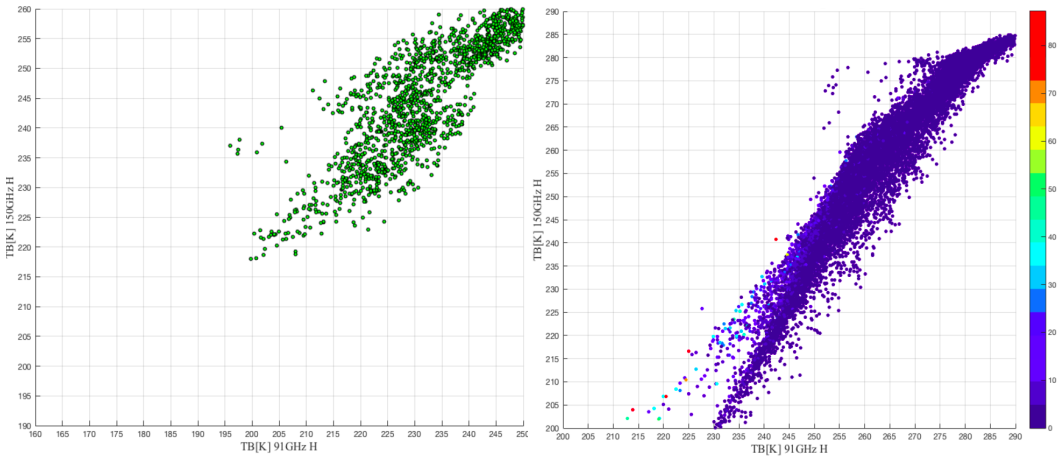
The Figure 4.5 presents information from the horizontally polarized TB from the channels 150GHz × 90GHz × observed by the SSMIS (for the two overpasses) and the simulated scatter plot of 150GHz × 90GHz × Rain Rate (mm/h) for the different surface types.

For the land pixels both datasets has a similar behavior with a downward linear tendency relative to the TB decrease by the presence of ice content. The large difference is that the radiative transfer model (Fig. 4.5b) provided a trend reaching colder temperatures around 165K in 150GHz and around 180K in 91GHz meanwhile the observations detected minimal temperatures around 215K in 150GHz and 195K in 91GHz. Although they present the same tendency they present great differences in terms of the minimum values. The simulation (Fig. 4.5b) presented the precipitating pixels associated with temperatures lower than 250K for both frequencies. The simulations presented a tendency to generated much higher TB values associated with lower rain rate values (between 0 and 5 mm/h), showing values reaching 295K meanwhile the observations were restricted to values not higher than 260K.

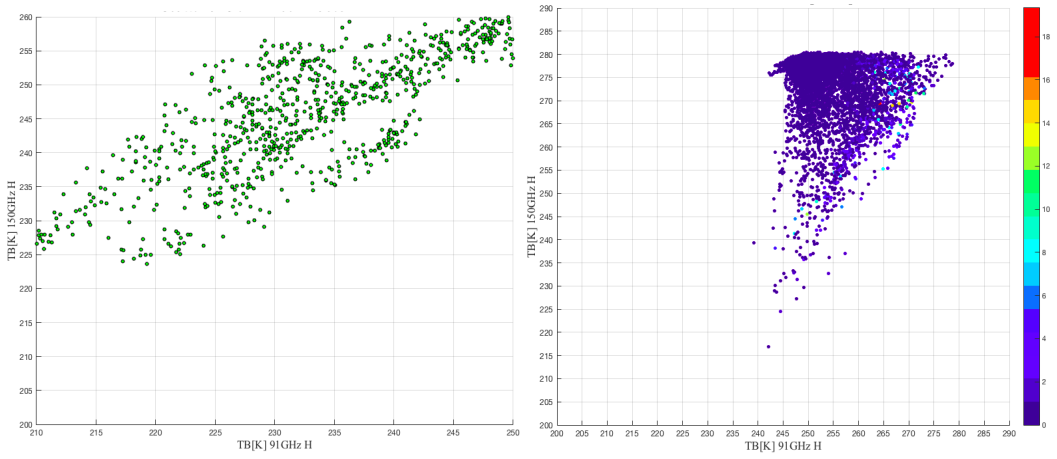
For the ocean pixels the observations (Fig. 4.5c) show a downward tendency reaching minimal values of 225K on 15GHz0H-Pol and 210K on 91GHz. It is possible to see the tendency of the model (Fig. 4.5d) in provide a high density of pixels concentrated, in this case, between 270-280K on 150GHz and 240-260K 91GHz and producing a population of pixels with warmer values (until 280K). It is possible to notice that the precipitating pixels does not show a specific trend (Fig. 4.5d).

The population of the coast pixels has the smaller number of points and presents the greater spread (Fig. 4.5e and Fig. 4.5f). On these frequencies the observations has a similar behavior of the ocean pixels, with a downward tendency reaching minimal values of 225K on 150GHz and 210K on 91GHz (Fig. 4.5e) and with the simulations presenting a high density of pixels concentrated, in this case, between 275-280K on 150GHz and 245-285K on 91GHz and the precipitating pixels show a slightly tendency with decreasing of temperature in both frequencies (Fig. 4.5f).

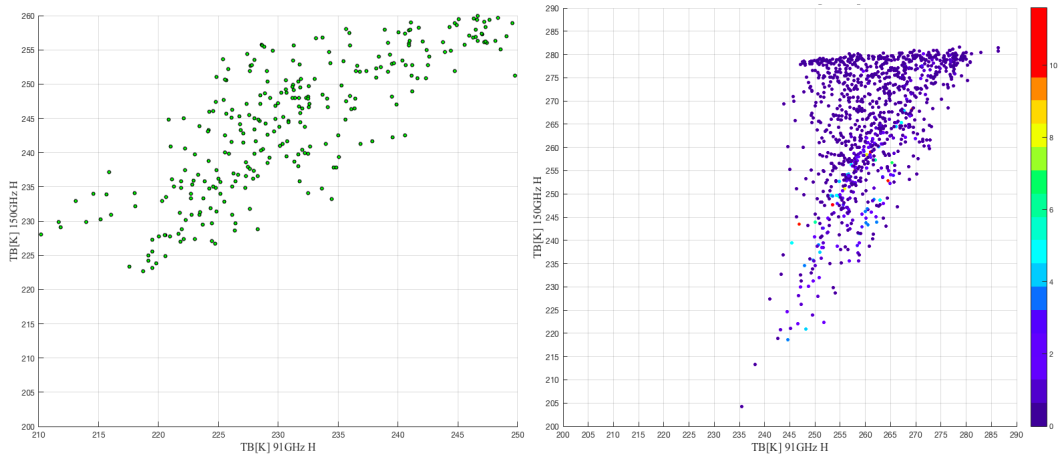
Figure 4.5 - Scatter plot from TB TB at 150GHz × 91GHz at horizontal polarization from SSMIS observations (left) and TB at 150GHz × 91GHz × Rain Rate (mm/h) from all simulation times (right) for a) and b) for land pixels; c) and d) for ocean pixels and e) and f) for coast pixels, respectively.



(**a**) (**b**)



(**c**) (**d**)



(**e**) (**f**)

SOURCE: Author's production.

This brief analysis represented the patterns seen on the simulations for the Vale do Paraíba and Manaus (not shown here) where the simulations presented discrepancies in representing the manifold of brightness temperatures observed by the SSMIS overpasses. Must be taking into account that this differences in the population of points are also related with the fact that the simulations refer to a larger number of pixels. In general, the simulations for land pixels presented a tendency to generate lower minimum temperatures (less than 200K) while the observations had minimum thresholds above 200K. This pattern was seen mainly in the high frequency (91GHz and 150GHz) channels. The simulations for the ocean tended to produce warmer temperatures on the higher frequency channels (91GHz and 150GHz) and the distributions for the coast pixels were those with the highest scatter and lower agreement in the temperature thresholds. Similar results were found for the Amazonian simulations (figures not included here) and due its greater surface variability in the central area of study, the simulated TB values are even more affected due to the presence of the Negro and Amazonas rivers.

It should be noted that a deeper analysis with an extensive number of simulations could ensure good representativeness of the simulated multichannel TBs in order to better describe the regional features of TB across a wide range of meteorological systems and microphysical conditions containing precipitation.

Deep adjustments in mesoscale models and improvements mostly in the radiative transfer (mainly the scattering by non-spherical ice particles) could contribute for gains in performance of the simulations in different regions of Brazil. Unfortunately, by the time this stage of work was completed, we did not have the technical conditions to implement such changes. For this reason, we migrated towards the use of an observational database in order to develop a rainfall retrieval algorithm over Brazil, as will be shown in the next chapter.

5 NNIMBUS

In this chapter, the development of the rainfall retrieval based on artificial neural network (ANN) optimized for continental territory of Brazil is presented. For practical purposes, the algorithm will be called : **N**eural **N**etwork **IM**plementation of the **B**razilian m**U**ltilayer perceptron for **S**creening and precipitation retrieval (**NNIMBUS**) and this name will be used hereafter. Furthermore, all Python 3 codes used in this thesis and their respective libraries will be available in the NNIMBUS GitHub Repository or available through the link: `https://github.com/acmlia/nnimbus/`.

Section 5.1 is subdivided into three topics: subsection 5.1.1 describes the main features from the DPR instrument and GPM core scan patterns and swaths, subsection 5.1.2 shows the GMI radiometer characteristics as channels, footprints and resolutions and subsection 5.1.3 describes the observational dataset used as *a priori* dataset to train the ANN.

An exploratory data analysis of the observational dataset is described in section 5.2 in order to further identify and summarize the main characteristics of the variables contained in the dataset, and if they are representative in terms of typology of meteorological events in order to allow for a reliable training procedure. The design methodology of the NNIMBUS algorithm, as well as network training settings, parameters and network architecture are described in details in subsection 5.3. The verification of NNIMBUS performance is made for one month of data extracted from the database (i.e., an independent part of the observational database, based on 2B-CMB (V05), and not used in the training and design phase of the algorithm) is also presented in subsection 5.4.

5.1 Data

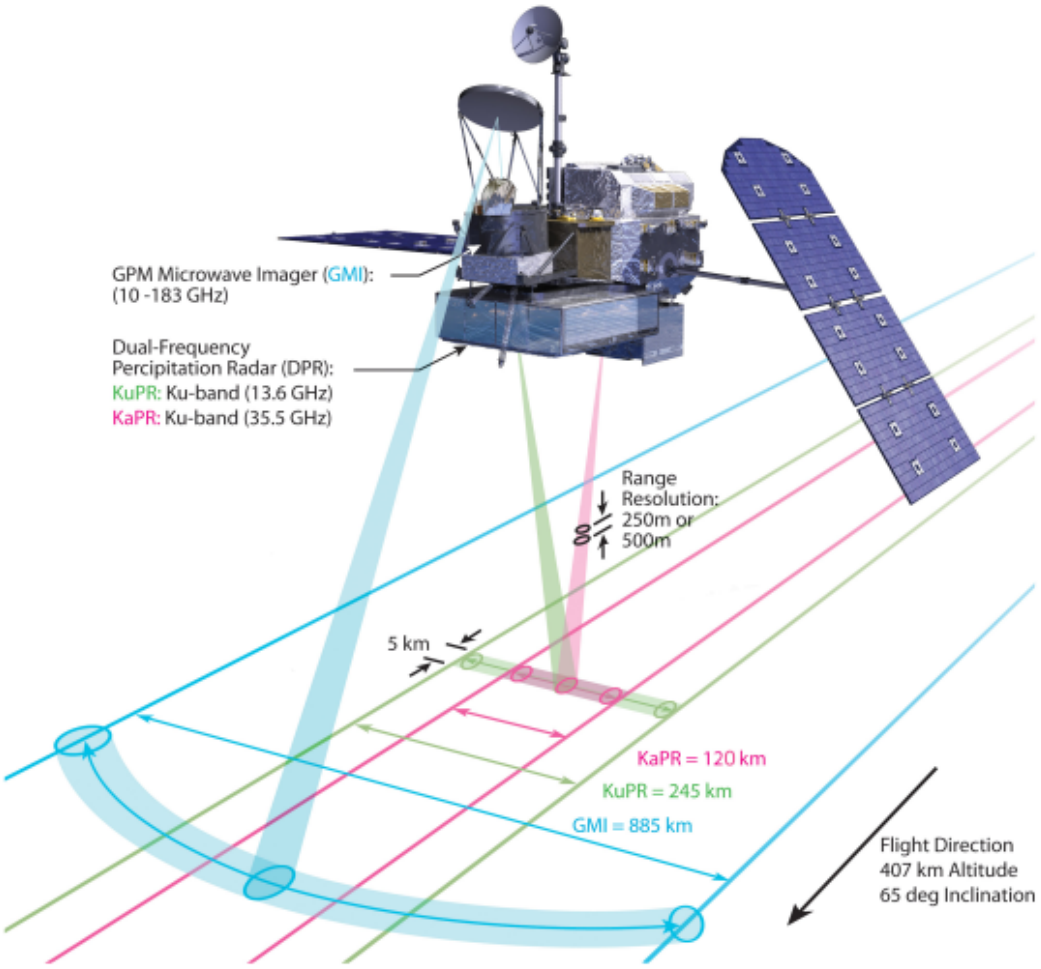
5.1.1 DPR instrument

The DPR instrument is comprised of two, essentially independent radars. One radar operates in the Ku-band at 13.6 GHz, also known as KuPR (Ku-band Precipitation Radar). The other radar operates in the Ka-band at 35.55 GHz, known as KaPR (Ka-band Precipitation Radar). By measuring the reflectivities of rain at two widely different radar frequencies, it is possible to infer information regarding rain rate, cloud type and its three-dimensional structure, and drop-size distribution. The DPR has an improvement in the detection of light precipitation, in the information on the

distribution of particle size of precipitation and snow, as well as in the identification of vertical structure consistent to the particles in the mixed, ice and liquid phases (IGUCHI et al., 2002).

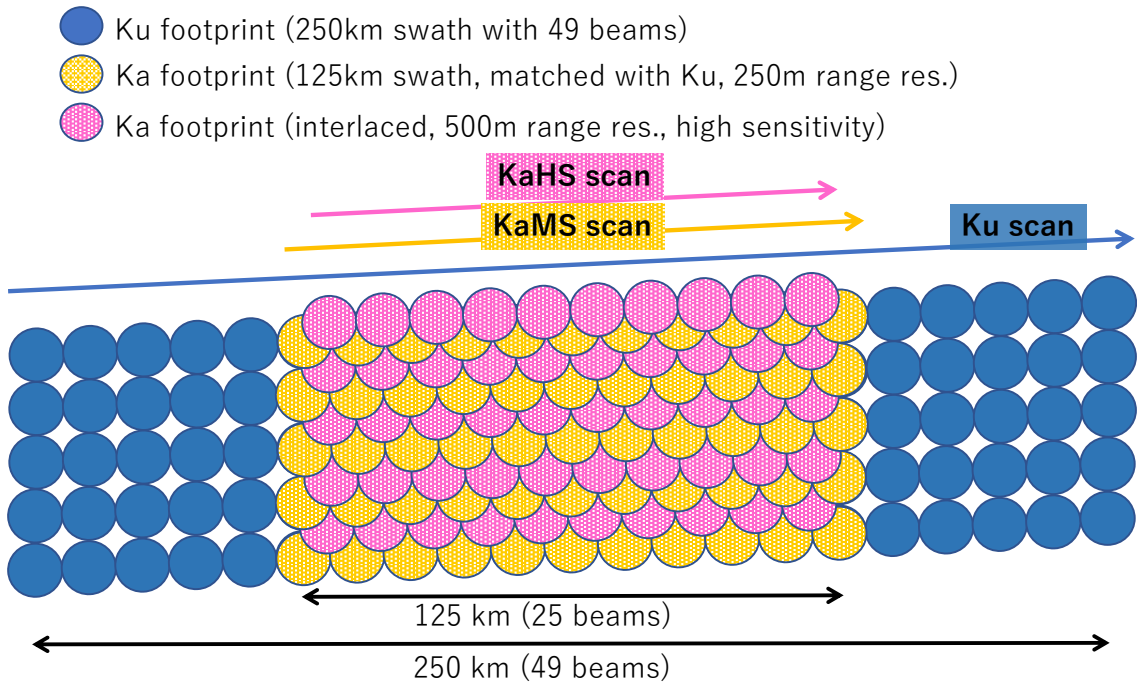
The KuPR has a 245 km measurement swath and the KaPR has a high sensitivity mode which provides an interlacing scan with a swath width of 120 km. This high sensitivity mode contributes to the measurement of light rain and snow. Figure 5.2 shows the KuPR, KaPR and KaPR (interlaced) footprints disposal. The KuPR has 49 footprints, meanwhile the KaPR has 24 footprints, and both have swath width of 5km. The footprints relating to the KaPR high sensitive mode are highlighted in pink color.

Figure 5.1 - GPM Core observation geometries.



SOURCE: NASA (2018).

Figure 5.2 - DPR antenna scanning and footprints.



SOURCE: JAXA (2018).

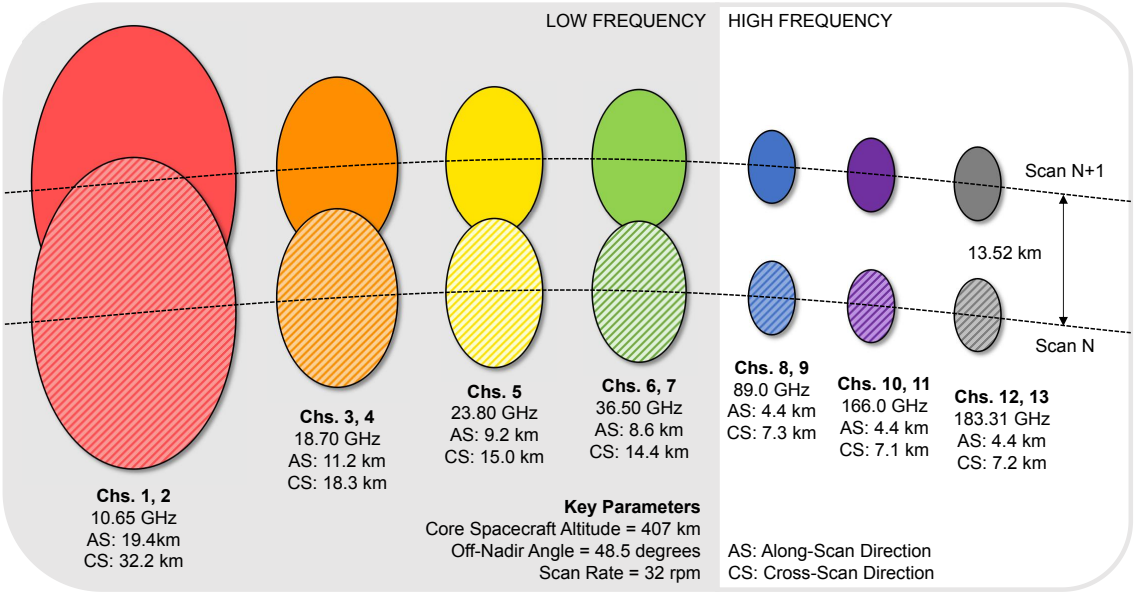
5.1.2 GMI instrument

The GMI instrument is a conical-scanning, polarization-sensitive and multi-frequency passive radiometer for rainfall measurement. The conical scan geometry of GMI is shown previously in 5.1 where the off-nadir-angle defining the cone swept out by the GMI is set at 48.5º which represents an Earth incidence angle of 52.8º (identical to that of TMI on TRMM). The swath represents 885 km on Earth's surface. The radiometer is a multichannel radiometer with frequencies ranging from 10 to 183GHz.

The GMI is a multichannel, conical-scanning, total power MW radiometer equipped with 10 dual-polarization (V and H) window channels at 5 frequencies (10.65, 18.70, 36.5, 89.0 and 166.0 GHz) and three single-polarization (V) channels, one at 23.8 GHz and two in the water vapour absorption band at 183.31 GHz (V polarization) (SANÒ et al., 2018).

All these frequencies are actually considered as the most appropriate for detecting the wide spectrum (heavy, moderate and light) of precipitation intensities (HOU et al., 2014). The four high frequency, millimetre-wave channels at 166 GHz and 183.31 GHz, can be exploited for light precipitation and snowfall retrieval at higher latitudes (e.g., Panegrossi et al. (2017)). A better spatial resolution than most of the previous radiometers up to roughly 4 km × 7 km at the high frequency channels and around 11 km × 18 km at 19 GHz, as can be seen in Figure 5.3.

Figure 5.3 - GMI instantaneous field-of-view (IFOV) projections on Earth surface



SOURCE: Adapted from NASA (2018).

5.1.3 GMI/DPR observational database

For the algorithm development, the observational measurements from GMI/DPR matchups (DPR/GMI Combined (CMB)-Version 5) from the GPM satellite core provided by the Colorado State University (CSU) were used. The GPROF for GPM V5 retains the previous version (i.e. GPM V4) of the CMB and DPR-Ku products for its databases. Because the CMB V04 algorithm appeared to significantly over-estimate precipitation over land, the *a priori* databases were constructed from the CMB (V4) over ocean and from the DPR Ku (V4) over land and coastal regions.

Some practical details about the database are: a) approximately 400 million matched

GMI/Surface precipitation profiles from all sources (globally); b) CMB database uses only the middle 21 pixels; c) TBs are in their native resolution; d) ECMWF , and GANAL databases are used for defining pixels total column water content (TCWV) , and temperature at 2m (T2m) and surface wet bulb temperature and e) the dataset surface classification is divided in 14 surface types.

5.2 Exploratory Analysis

In order to build an accurate model it is necessary to understand and extract insights from the *a priori* dataset that the model will be built upon. This section will present an overview of the behaviour of the observational database introduced in Section 5.1.3. The database is composed of 2D (two dimensions) variables and of vertical profiles (derived from DPR). For this study, only the 2D variables were considered, which are: latitude, longitude, surface code (sfccode) temperature at 2m (T2m), total water vapor content (tcwv), skin temperature (skint), surface precipitation (sfcprcp), convective precipitation (cvprcp), all 13 GMI frequencies and their respective surface emissivities.

5.2.1 Statistical analysis of observational database over Brazil

A first analysis is made using box plots to provide insights about the distribution and discriminate the range of data falling between the 25^{th} and 75^{th} percentiles (the box), the median value is represented by the horizontal line inside the box and the complete range of the data is represented by the whiskers.

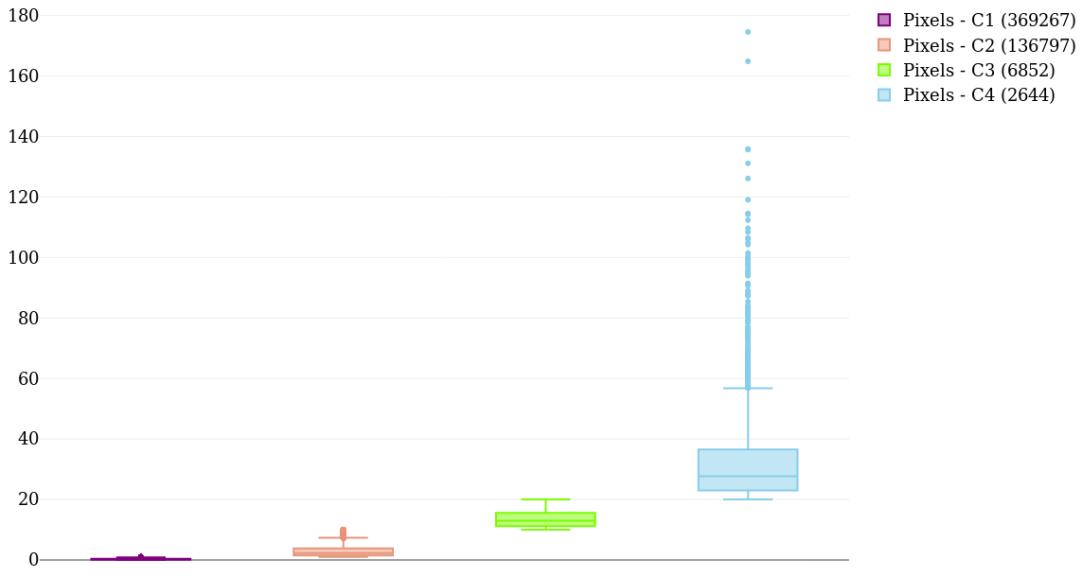
For this analysis, the variables were divided according to four precipitation classes, which are designated as: values between **0-1 mm/h** (C1), **1-10 mm/h** (C2), **10-20 mm/h** (C3) and larger than **20 mm/h** (C4). For the Brazilian territory, the dataset contains observations from 3603771 pixels.

Figure 5.4 shows the total number of pixels, where only 14,3% (515560) are precipitating pixels. From this sample, the C1 represents 71,6% (369267), C2 contains 26,5% (136797), C3 has 1,3% (6852), and C4 contains the smaller percentage of 0.5% (2644) pixels.

In terms of occurrence, C2 and C3 present median values around 2,25 and 13 mm/h, respectively. Class C4 presents the distribution with higher range, with the box percentiles between 23 and 36 mm/h and the remaining range (and outlier points) with values up to 56 mm/h.

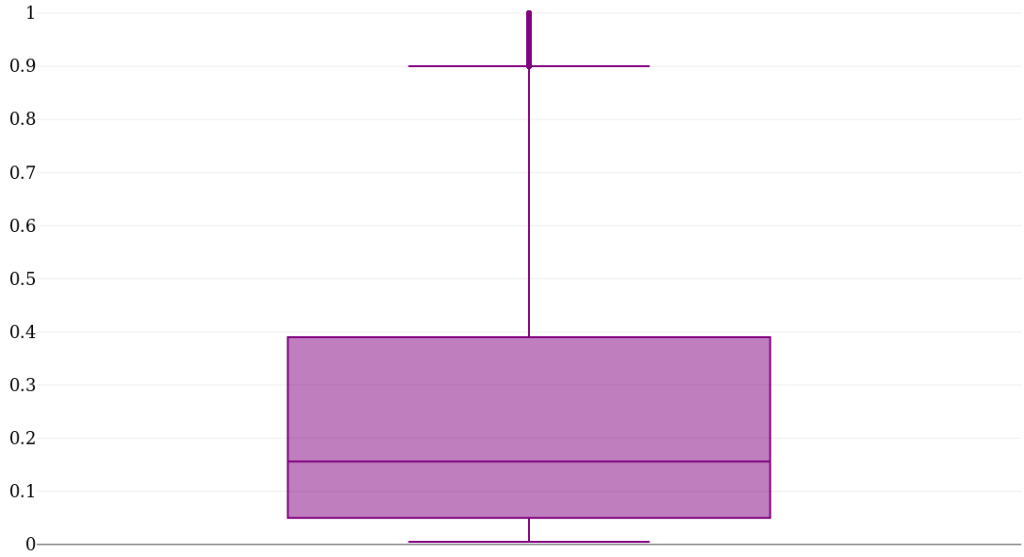
Figure 5.5, shows that the values responsible for the larger population of C1 are the values from 0,05 and 0,4 mm/h (the box percentiles).

Figure 5.4 - Box plot of surface precipitation (mm/h) divided by classes.



SOURCE: Author's production.

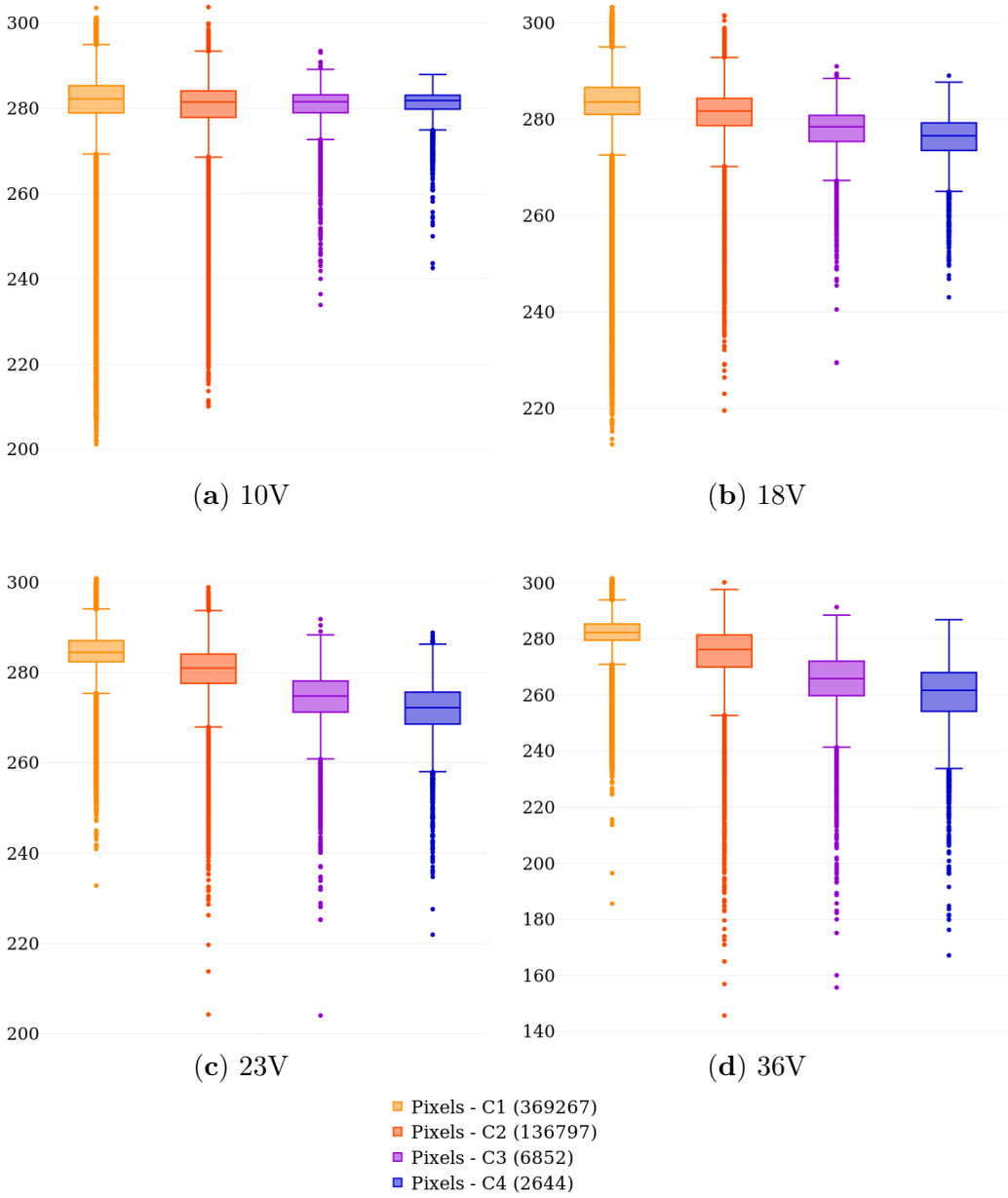
Figure 5.5 - Box plot of Surface Precipitation in class C1 (0-1 mm/h).



SOURCE: Author's production.

Regarding the TB signatures, the analyses for the lower frequency channels 10-36GHz (V) are shown in Figure 5.6. These frequencies have the box plot distributions shifted to high TB values (higher than 260K) for all the four precipitation classes (C1 to C4) because of the high emission surface signal. The 10 and 18GHz frequencies associate classes C1 and C2 with a larger range of TB, extending the outliers up to 200K. This could be due to the low emissivity of relatively clear inland water bodies. Frequencies between 23 and 36 GHz presented larger range of TB for classes C2, C3 and C4, respectively. In these channels, the TB value starts to decay with the increase of the precipitation rate, as is expected due to ice scattering the signal. The 36GHz frequency channel presents the lowest TB values for all analyzed classes, reaching 140K in some cases.

Figure 5.6 - Box plot of brightness temperature at 10, 18, 23 and 36GHz (V).

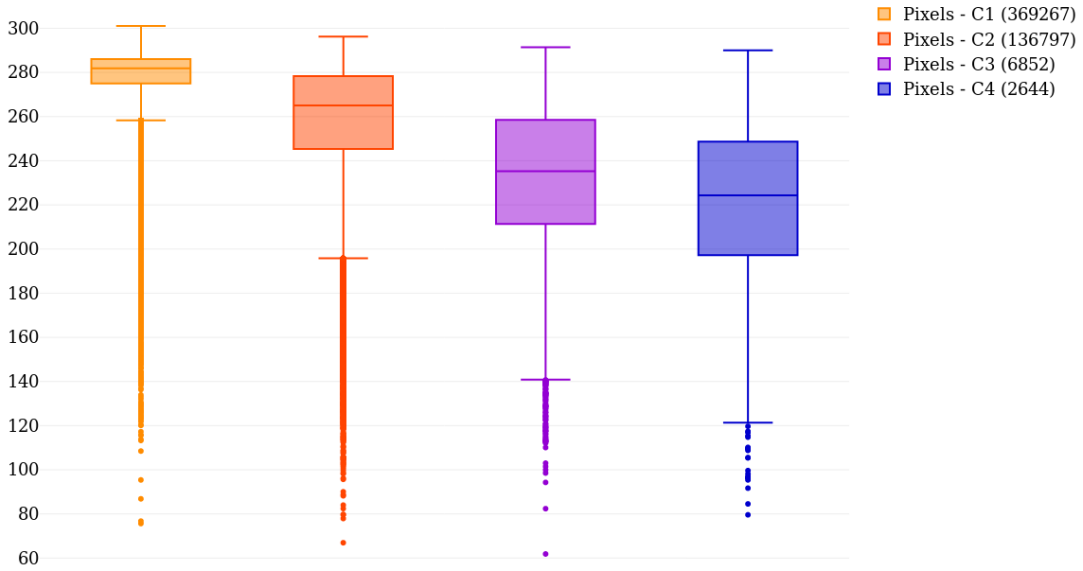


SOURCE: Author's production.

Figure 5.7 shows the signal scattering behaviour of ice, better visible at higher rain rates, at 89GHz with vertical polarization. All classes present minimum values of around 60K, and C1 presents the larger range for these outliers. Very low TB associated with light precipitation rates or no rain can be associated with high scattering processes in cold surfaces like deserts (see Figure 17 in Grody e Wen

(2008)).

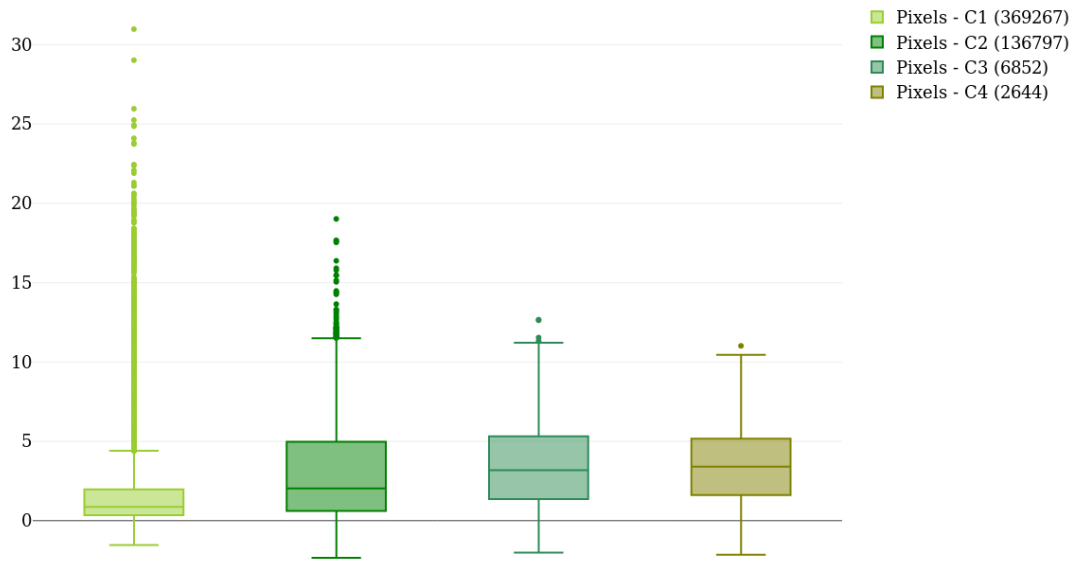
Figure 5.7 - Box plot of brightness temperature at 89GHz (V).



SOURCE: Author's production.

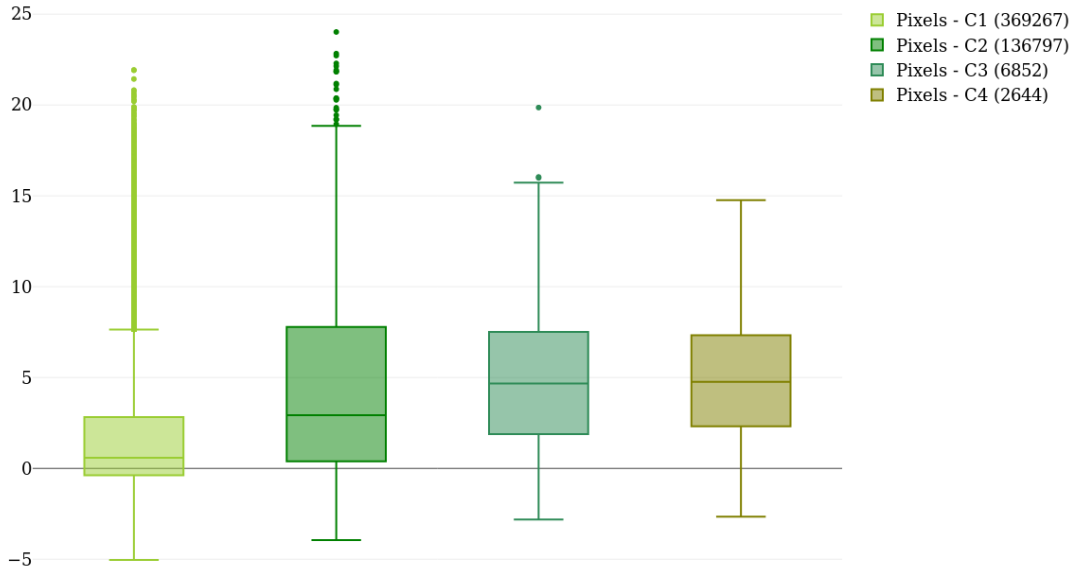
Considering the possibility of extracting additional information using the vertically (V) and horizontally (H) polarized radiances, the polarimetric differences (PD) from GMI frequencies were also computed. An example of the PD application is the use of TB at 89 and 166GHz. As found by Gong e Wu (2017) the scattering by frozen particles is highly polarized, with V–H PD being positive throughout the tropics and the winter hemisphere mid-latitude jet regions. High polarization occurs mostly near convective outflow regions (i.e., anvils or stratiform precipitation), while the polarization signal is small inside deep convective cores as well as at the remote cirrus regions. Figures 5.8 and 5.9 present PD values of 89GHz and 166GHz, respectively. It is noticeable that 166GHz PD reaches more negative values for all classes, and for classes C2, C3 and C4 it has the largest amplitude. This behaviour can be interpreted as the better ability of 166GHz PD to distinguish better the variation from convective cores to anvil outflow and further to clear sky than 89GHz PD. This occurs due to the increasing contribution of ice particle scattering at higher-frequency microwave channels. This kind of statistical analysis can bring complementary information for the screening and rainfall retrieval process.

Figure 5.8 - Box plot of polarimetric difference (PD) at 89GHz (V-H).



SOURCE: Author's production.

Figure 5.9 - Box plot of polarimetric difference (PD) at 166GHz (V-H).



SOURCE: Author's production.

Based on this analysis, the expected physical relationship between precipitation and the TB signatures is detected and it is well represented in the dataset. Furthermore, variables derived from polarimetric differences can be also be exploited to supply additional information from microphysical features of the precipitating systems.

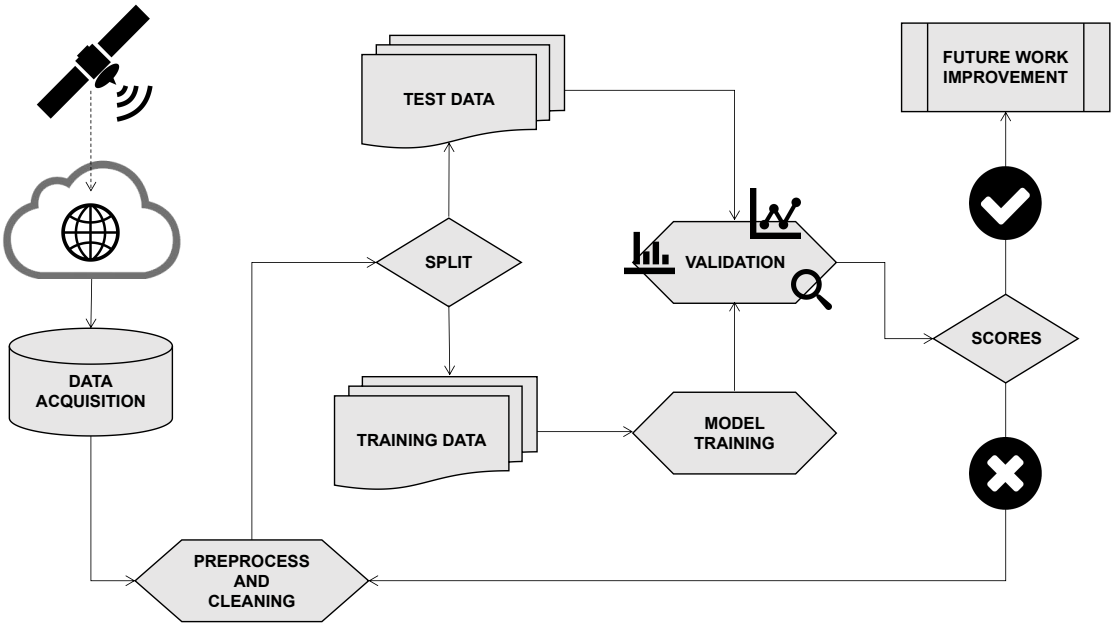
5.3 Methodology

The NNIMBUS algorithm is composed of two main modules, where the first identifies the precipitating areas (screening of precipitation) and the second retrieves the rainfall intensity. Both modules are based in the use of ANN and the detailed methodology used to build each module will be presented in details in subsections 5.3.1 and 5.3.2.

The main steps taken in the construction of a neural network are presented in Figure 5.10. The 2D variables from the GMI/DPR database (and additional variables derived from it) are used as input predictors and a thorough cleaning and preprocessing treatment is applied in order to make the data readable for the network. In this stage, some of the treatments are: data format conversion, regional extraction from the global dataset, file concatenation, creation of additional variables and exploratory analysis. Right after that, the treated data is inserted in the ANN training process, where it is split into three samples. The first sample is the training dataset, which is the data the ANN will be trained with (this is the sample from which the network will learn the knowledge), the second is the test sample used to check the validity and usefulness of different ANN models during the training phase, and the third sample is called the validation dataset.

Until an optimal ANN design is not found, the training processing is repeated as many times as necessary until an ideal weight adjustment is defined. Once an optimal ANN adjustment is found, its performance is verified by comparing its results to the validation dataset. In our case, since only one year of observations is available, one month of the observational data was extracted to be used as the validation dataset (coincident geo-colocated CMB data). We decided to extract the month of November 2014 as the validation dataset because it presented neither excess nor the lack of rainfall events (verified by the total precipitation observed by rainfall monitoring from CPTEC/INPE).

Figure 5.10 - Workflow of the general steps used to train ANNs.



SOURCE: Author's production.

The whole NNIMBUS algorithm was writen and tested for Python 3.6.8 (also available through the link: `https://www.python.org/downloads/`). It have also made extensive use of several open-source libraries that will no be cited here for the sake of brevity, but special emphasis must be given to the machine learning libraries: TensorFlow, Keras and Scikit Learn. In brief, TensorFlow is an end-to-end open source platform for machine learning and it has a comprehensive and flexible ecosystem of tools. It is used for both research and production at Google and it can be considered as the state-of-the-art in ML (ABADI et al., 2015). Keras is a high-level neural networks application programming interface (API) capable of running on top of TensorFlow and with focus on enabling fast experimentation (CHOLLET, FRANÇOIS AND OTHERS, 2015) and the ScikitLearn is also a ML library with simple and efficient tools for data mining and data analysis, also providing many useful tools for data processing (PEDREGOSA et al., 2011).

The capabilities of the hardware in charge of executing all the test rounds during the NNIMBUS ANNs development is also summarized in the List 5.1. The machines used for this work were two mirrored clusters, *Orion-d11* and *Helios*. Each cluster had 20 physical CPUs running at 2.20GHz and hyper-threading capabilities enabled,

summing up for a total of 40 virtual CPUs in each cluster. A third cluster with fewer processing cores where also used, but in which the cores operated at a higher frequency.

Besides the general thesis workflow presented in figure 5.10, the Annex A.1 of this work also present a more in-depth pipeline synthesizing all the inputs of NNIMBUS, their processing steps, functionalities and outputs.

```
[user@orion-d11 ~]$ lscpu

Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                40
On-line CPU(s) list:   0-39
Thread(s) per core:    2
Core(s) per socket:    10
Socket(s):             2
NUMA node(s):          2
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 62
Model name:            Intel(R) Xeon(R) CPU E5-2660 v2 @ 2.20GHz
Stepping:              4
CPU MHz:               1217.218
CPU max MHz:           3000,0000
CPU min MHz:           1200,0000
BogoMIPS:              4399.95
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              256K
L3 cache:              25600K
NUMA node0 CPU(s):     0-9,20-29
NUMA node1 CPU(s):     10-19,30-39
```

Listing 5.1 - CPU summary for the Orion-d11 cluster

5.3.1 Screening of precipitation

The next sections describe the main pre-processing data treatment done for the screening model, as well the hyperparameters selection process and model training architecture.

5.3.1.1 Pre-processing of the input data

The available inputs to be used as predictors for the ANN training were the following 30 variables: lat, lon, sfccode, T2m, tcwv, skint, 10V, 10H, 18V, 18H, 23V, 36V, 36H, 89V, 89H, 166V, 166H, 186V, 190V, 10VH, 18VH, 36VH, 89VH, 166VH, delta183V, SI(23V-89V), PCT10, PCT18, PCT36, and PCT89. In order to decide which input variables could contribute more during the learning process, it is possible to perform a process called feature selection. In this case, the chosen method of feature selection is known as recursive feature elimination (RFE). This approach creates a ranking with recursive feature elimination. Given an external estimator that assigns weights to features (e.g., the coefficients of a linear model), the goal of the RFE approach is to select features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and the importance of each feature is obtained either through a coefficient attribute or through a feature importance attribute. Then, the least important features are pruned out from the current set of features. That procedure is recursively repeated on the pruned set until the desired number of features to be selected is eventually reached. Among the most important variables determined by the RFE for the screening, the 10 following predictors were used: 36V, 89V, 89VH, 166V, 166VH, 186V, 190V, SI(23V-89V), PCT36 and PCT89.

Another important aspect to be taken into account is that the different input attributes comprise a wide scale range. Most of the ML methods usually expect, or are more effective, if the data attributes have a consistent scale or distribution, therefore, two popular data scaling methods are normalization and standardization. Since precipitation screening is a classification problem (binary classification) the normalization method was chosen, and it is a method in which the attribute values are rescaled within the range of 0 and 1.

As mentioned before, the dataset must be split into three smaller datasets, for training, test and validation. The complete dataset (eleven months) used in this step contains 3259296 pixels, of which 80% are used for the training process and 20% are used as test set. The splitting process selected the samples randomly. It is important to bear in mind that there are no general guidelines on how to split the complete database into smaller parts, and in this case the sizes of the datasets were defined in order to balance between obtaining sufficiently large and representative samples while at the same time attempting to reduce the computational time consumption.

5.3.1.2 Hyperparameters selection and model training

Considering the large number of input attributes and parameters that need to be set in neural networks, which make them notoriously difficult to configure, the model hyperparameter optimization techniques are an important step to assist in this configuration. The GridSearchCV class from Scikit Learn library is one of such techniques. This technique performs an exhaustive search over specified parameter values for an estimator. It implements a "fit" and a "score" method and the parameters of the estimator used to apply these methods are optimized by cross-validated grid-search over a parameter grid. By default, accuracy is the score that is optimized, but other scores can be specified. In terms of processing power by default, the grid search will only use one thread, but it can be set to use all machine cores. Therefore the GridSearchCV process will then construct and evaluate one model for each combination of parameters; cross validation is then used to evaluate each individual model.

The model to be analyzed is instantiated using the **Keras model** API. Then, the **Sequential model** that is the linear stack of layers is created. And in the layers structure is where the **Dense model** is applied, in which the weights matrix created by each layer and the element-wise activation functions are implemented. As the screening is a classification problem, the KerasClassifier function is used.

Through the use of the GridSearchCV, several hyperparameter tuning tests were processed. The optimal ANN for the screening was found using the following parameters.

- The ANN was trained with 10 predictors: 36V, 89V, 89VH, 166V, 166VH, 186V, 190V, SI, PCT36, and PCT89.
- It has **four layers**, where the first layer has **10 neurons**, the second and third layers have **32 neurons** and the last layer has **one** neuron (the output).
- The activation functions which control the non-linearity of the individual neurons and when they are fired are the **relu** for the first three layers and the **sigmoid** function.
- The **batch size** in iterative gradient descent is the number of patterns shown to the network before the weights are updated, and in this case it was defined as **10**. The number of **epochs**, that is, the number of times that the entire training dataset is shown to the network during training

phase, was defined as **100**.

- The optimization algorithm used to find the values of parameters (coefficients) of a function (f) that minimizes a cost function (cost) was the **Stochastic Gradient Descent (SGD)**. The SGD is a simple yet very efficient approach to discriminative learning of linear classifiers.
- The learning rate, which controls how much to update the weight at the end of each batch and the momentum that controls how much to let the previous update influence the current weight update, were set with default values of **lr=0.01** and **momentum=0.0**, respectively.
- The network weight initialization, which determines how the weight initialization matrix is configured, was set as **uniform** where small random values are determined following an uniform distribution.

The categorical scores for the test dataset and for the validation dataset are presented in the results subsection 5.5. Where comparisons between the NNIMBUS output and GMI/DP CMB and GPROF/GMI rain/no-rain maps will also be presented.

5.3.2 Rainfall retrieval

The next subsections described the main pre-processing data treatment done for the retrieval model, as well the cleaning criteria applied in the dataset and the model training architecture.

5.3.2.1 Pre-processing of the input data and model training

After several tests for different ANN architectures and hyperparameter analysis, the optimal ANN for the rainfall retrieval had 17 input predictors, and the variables were: sfccode, T2m, tcwv, 10V, 10H, 18V, 18H , 36V , 36H , 89V , 89H, 166V, 166H, 183VH, PCT36, PCT89 and 89VH. A principal component analysis (PCA) was performed on GMI channels with similar characteristics to explore the possibility of decrease in the number of inputs and to reduce the effect of surface emission variability on the measured TBs, which would consequently improve the network performance in retrieving surface precipitation (SURUSSAVADEE; STAELIN, 2008). Thereby, the PCA transformation was applied for the subsets: subset PCA1 for 10V, 10H, 18V and 18H channels; and and subset PCA2 for 36V, 36H, 89V, 89H, 166V and 166H channels.

As mention in subsection 5.3.1.1 is necessary to scale the input attributes, and since the rainfall is a discrete variable the standardization was applied (instead of a normalization, like it was previously done in the screening).

As the performance of the ANN is largely dependent on the consistency of the input data with the current observations, some quality standards have been applied to the training data-set before feeding the ANN. Therefore, in order to comply with this premise, the input set was treated using these three criteria: The first criterion consists in removing samples below 125K at 89GHz (V), since the distribution at this frequency for all precipitating classes (C1 to C4) presents outlier values below 125K, i.e., they are outside the 25^{th} and 75^{th} percentiles range as shown in the exploratory analysis (Figure 5.7). Furthermore, the analysis also displayed that the precipitation class between 0 and 1 mm/h (named C1) corresponds to the larger portion of the input data (71.6%) with more than 3 million elements (Figure 5.4). For this reason, the second criterion consisted in randomly under-sampling 90% of the C1 (which remained with 165313 elements), in such a way that the original data distribution was still preserved. As demonstrated by Skofronick-Jackson et al. (2018), the DPR and CORRA quantify rain rates between 0.2 and 110 mm/h, while GMI estimates rain rates up to 60 mm/h due to the averaging of strong convective cells by the coarser-resolution radiometer observations. It is worth noting that for the DPR (GMI) 5 km (15 km) footprint scales, rain rates $\leq$ 110 mm/h (60 mm/h) are very infrequent. Thus, the third criteria considered just range of observed values between 0.2–60 mm/h.

The retrieval model was also instantiated using the **Keras model** API, with the **Sequential model** and **Dense model**. However, as it is a regression problem, different from the screening, this module used the KerasRegression fuction. And after several of hyperparameters tuning tests, the optimal ANN architecture for the rainfall retrieval was configured as:

- The ANN was trained with 10 predictors: sfccode, T2m, tcwv, 10V, 10H, 18V, 18H , 36V , 36H , 89V , 89H, 166V, 166H, 183VH, PCT36, PCT89 and 89VH. And the PCA was applied in two subsets: PCA1 (10V, 10H, 18V, 18H) and PCA2(36V , 36H , 89V , 89H, 166V, 166H).
- It has **four layers**, where the first layer has **17 neurons**, the second layer has 51 neurons,the third layer has **10 neurons** and the last layer has **one** neuron (the output).

- The activation function which controls the non-linearity of the individual neurons and when they are fired was the **linear** function for all the layers.
- The **batch size** was defined as **10** and the number of **epochs** was defined as **1000**.
- The optimization algorithm used to find the values of parameters (coefficients) of a function (f) that minimizes a cost function (cost) was the gradient-based stochastic optimization algorithm **Adam**.
- After to application of the cleaning criteria, the input dataset contained 165313 pixels, where it was divided in 80% (132250 pixels) for training and 20% (33062) for the test dataset.

The continuous scores for test dataset and for the validation dataset are presented in the results subsection 5.6. A comparison between the NNIMBUS output and GMI/DP CMB and GPROF/GMI retrievals will also be presented.

5.4 Results

In this section, the performance of the screening and retrieval NNIMBUS modules, considering a set of categorical and continuous scores, are presented (the formulas are given in Annex A). Those scores are performed for two datasets described above: the test dataset (randomly selected) and the validation dataset (one month of data). Several case studies over Brazilian territory for a single time slot with each algorithm (single orbits) are performed in order to compare the algorithm's performance qualitatively.

5.5 Validation of NNIMBUS screening

For the classification problem, several categorical scores were obtained by comparing the satellite-estimated precipitating regions for NNIMBUS output and GMI/DPR-CMB observations as reference. They were calculated from confusion matrices showing agreement and disagreement between predicted and observed rainfall areas.

To identify the level of agreement between the forecast and the observation, the **accuracy** is used. The difference between the forecast and the observation is the error. The lower the error, the greater the accuracy. The bias is the correspondence between the mean forecast and mean observation, if the rain is underestimated in the model if the bias is <1; values of >1 indicate overestimation of rain in the model. This is a pure quantitative measure that does not account for an agreement between observed and predicted rainfall areas. The **probability of detection** (POD) gives

the percentage of rain pixels that the model correctly identified as rain and the **probability of false detection** (POFD) gives the proportion of non-rain pixels that the model incorrectly classified as rain. Similar to POFD, the **false alarm ratio** (FAR) gives the proportion of predicted rain where no rain is observed. The **critical success index** (CSI) is the proportion of true positives to both kinds of errors and correctly classified non-rain pixels are not considered. The **Hansen–Kuipers discriminant** (HKD) and **Heidke skill score** (HSS) account for chance agreement, where the HSS gives the proportion of correct classifications (both rain pixels and non-rain pixels) after eliminating expected chance agreement and the HKD is the difference between POD and POFD provides insight into whether predicting a pixel as rain leads to a considerable increase in false alarms.

A first assessment of the final NNIMBUS screening module has been carried out using the test dataset which contains 293894 pixels and it is presented in Table 5.1. The model presented an accuracy of 0.94 or, in other words, it could be said that from the total number of predictions, the model was in agreement with the truth for 94% of the pixels. The model presented the tendency to overestimate the rainfall areas with a mean bias value of 1.27. In terms of discrimination between precipitating and non-precipitating pixels, POD presented value 0.80 and POFD of 0.03, that we can translate that the model predict correctly 80% of the rain pixels, and with only 3% of no-rain pixels predicted as rain. The FAR corresponded to 0.36(or 36%) and the CSI to 0.55 (or 55%). The HSS presented a value of 0.68 and the HKD is equal to 0.76.

Table 5.1 - Categorical statistical scores obtained by the NNIMBUS screening for the test dataset.

Accuracy	**Bias**	**POD**	**POFD**	**FAR**	**CSI**	**HSS**	**HKD**	**N°of pixels**
0.94	1.27	0.80	0.03	0.36	0.55	0.68	0.76	293894

From those scores, an indication of the behaviour of the screening module considering the test dataset is obtained. To verify the performance of the network in relation to an a new set of unseen data, Table 5.2 presents the statistical scores for the validation dataset that contains 651860 pixels. The accuracy was of 0.95, the bias of 1.34, FAR of 0.39 and HKD of 0.77 shows a small increase, which can be associated with a larger number of pixels. The CSI and HSS are reduced when compared with

the test database to 0.52 and 0.66, respectively.

Table 5.2 - Categorical statistical scores obtained by the NNIMBUS screening for the validation dataset.

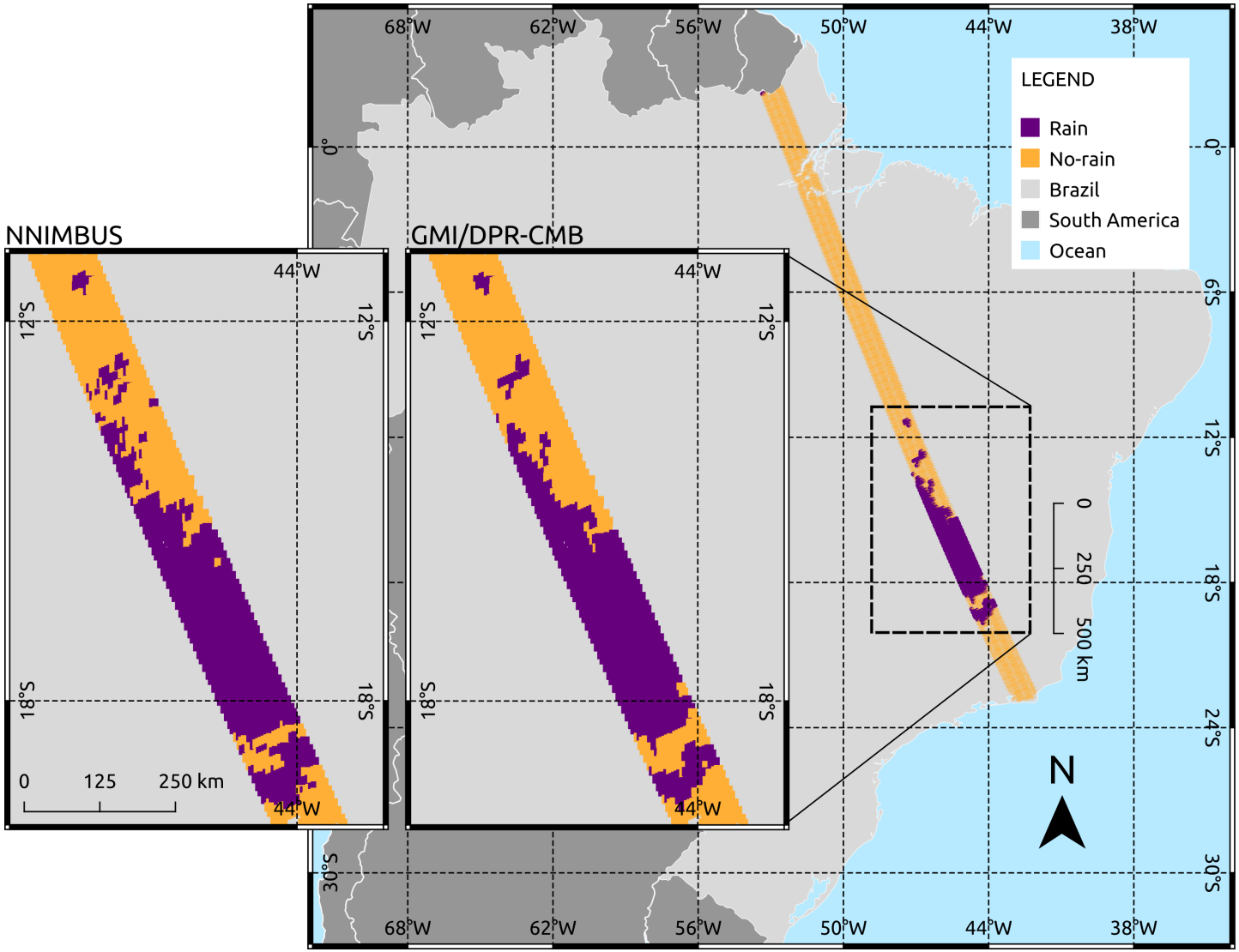
Accuracy	**Bias**	**POD**	**POFD**	**FAR**	**CSI**	**HSS**	**HKD**	**N° of pixels**
0.95	1.34	0.80	0.03	0.39	0.52	0.66	0.77	651860

In order to illustrate an example of the detection of rain/no-rain pattern, Figure 5.11 shows the GMI/DPR orbit on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4922). This satellite overpass detected the presence of raining pixels over the southeast region of Brazil. It is visible that the NNIMBUS screening (on the left) fits very well wit the GMI/DPR-CMB data (right). Small differences can be noticed in both panels, where NNIMBUS presents a noisier pattern at the edges of the systems, presenting a more pixelated coverage. This behaviour is confirmed by the scores shown in Table 5.3.

Table 5.3 - Categorical statistical scores obtained by the NNIMBUS screening for the orbit of GMI on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4922).

Accuracy	**Bias**	**POD**	**POFD**	**FAR**	**CSI**	**HSS**	**HKD**	**N° of pixels**
0.97	0.95	0.89	0.012	0.05	0.85	0.90	0.88	787

Figure 5.11 - Screening of precipitation by NNIMBUS (left), GMI/DPR-CMB (right) on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4299)



SOURCE: Author's production.

From the statistical comparison shown above, it is clear that the NNIMBUS screening procedure performed very well when compared to the DPR/GMI-CMB database. However, it should be noted that even though these examples refer to different periods, the reference is derived from the same product. This consideration may be causing the results presented above to be very similar, since the distribution of the reference data belongs to the same distribution.

In order to have a different perspective on the performance of this model, mainly related to the case-to-case variability, a comparison between the NNIMBUS screening and the GPROF-GMI (V05) is presented for new dates which are not present in the observational database. Figure 5.12 presents the GMI overpass on 12 March 2014 at 04:44-06:17 UTC (orbit n° 195) with 10980 pixels. The NNIMBUS screening (left) is able to represent both the large and smaller clusters detected by GPROF-GMI (center). The right panel of the figure shows the overlap of both screenings, where the GPROF is presented in green and the NNIMBUS is shown in red and it is visible that the missing regions in NNIMBUS are the edges of the systems, where NNIMBUS presents the noisiest pattern (pixelated in left panel).

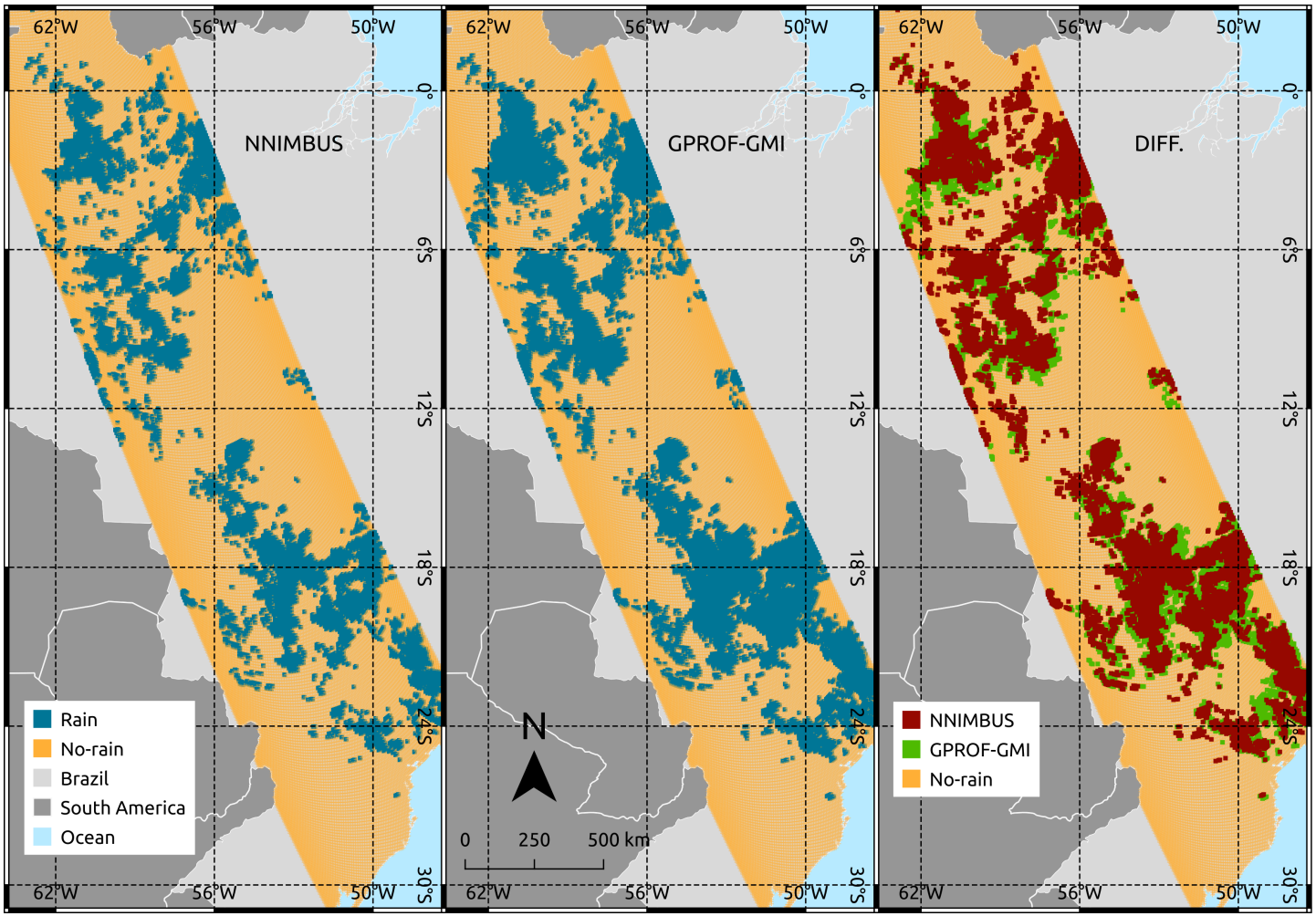
This omission in the edges of the systems is associated with the fact that the screening was trained for precipitation values with a minimum threshold of 0.2 mm/h. The design of the NNIMBUS screening was intentionally made more rigorous for the lower rain rate classes (< 0.2 mm/h) in order to block those pixels to advance to the next stage: the rainfall retrieval neural network.

Although it is known that the GPROF does not represent the complete truth, it represents the state-of-the-art of rainfall retrieval algorithms, and the statistical comparison between NNIMBUS output and GPROF is an important step in this study. Table 5.4 presents the categorical scores comparing NNIMBUS to GPROF, corresponding to those 10890 pixels. In this case the bias is smaller when compared with the DPR/GMI CMB database (0.13) and also, POFD (0.07), FAR (0.30) and CSI (0.58) but it also presents a small decay in the accuracy of (0.89), POD 0.78, HSS 0.67 and HKD 0.70.

Table 5.4 - Categorical statistical scores obtained by the NNIMBUS screening for the orbit of GMI on 12 March 2014 at 04:44-06:17 UTC (orbit n° 195).

Accuracy	**Bias**	**POD**	**POFD**	**FAR**	**CSI**	**HSS**	**HKD**	**N° of pixels**
0.89	1.13	0.78	0.07	0.30	0.58	0.67	0.70	10890

Figure 5.12 - Screening of precipitation by NNIMBUS (left), GMI/DPR-CMB (center) and the overlap between of both (right) on 12 March 2014 at 04:44-06:17 UTC (orbit n° 195).



SOURCE: Author's production.

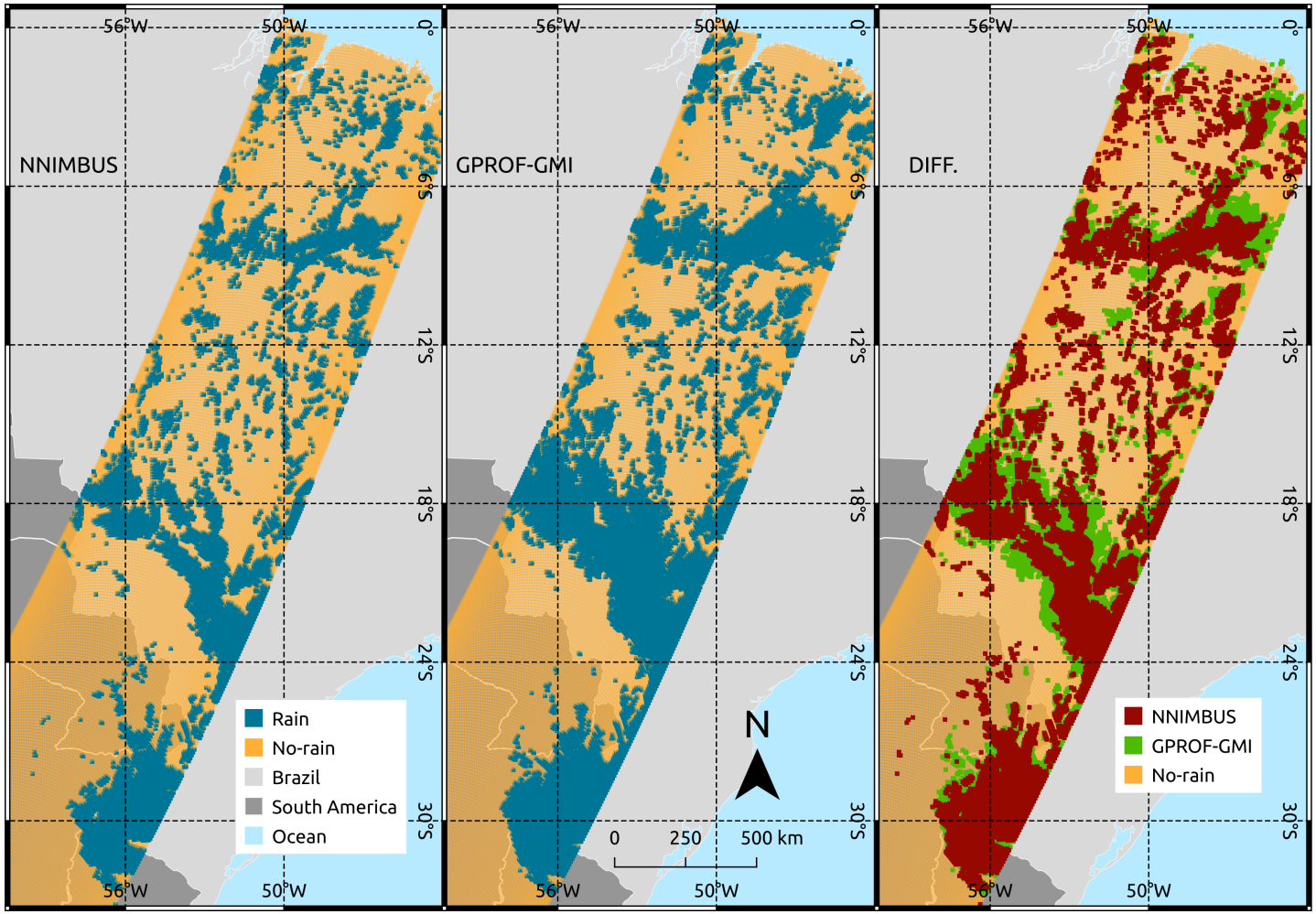
Figure 5.13 shows another case study of GPROF-GMI and NNIMBUS. This case belongs to the overpass on 23 November 2018 at 04:44-06:17 UTC (orbit n° 26914), with 14568 pixels. As in the example before, the NNIMBUS screening (left) is able to represent both the large and small clusters detected by GPROF-GMI (center), but it seems to miss more points on the edges of the systems (green pixels in the right panel) when compared with the previous case. This aspect is confirmed by the decay of the scores as shown in Table 5.5.

Table 5.5 - Categorical statistical scores obtained by the NNIMBUS screening for the orbit of GMI on 23 November 2018 at 04:44-06:17 UTC (orbit n° 26914).

Accuracy	**Bias**	**POD**	**POFD**	**FAR**	**CSI**	**HSS**	**HKD**	**N° of pixels**
0.87	1.08	0.72	0.09	0.33	0.53	0.61	0.63	14568

A very similar study involving neural networks for the GMI, trained with a larger sample of DPR/GMI-CMB observations and with a new rain/no-rain methodology, was the study developed by Sanò et al. (2018). In their verification analysis, they used 100 million pixels from the 2B-CMB rainfall used as truth, and obtained for the screeening of precipitation the best scores of: POD of 0.72, FAR of 0.15, CSI of 0.63 and HSS of 0.67. Beusch et al. (2018) also developed their screening and rainfall retrieval based in ANN, however optimized for geostationary satellite infrared (IR) brightness temperatures(SEVIRI MSG). For their screening model, they found performance with POD of 0.65, FAR of 0.4, CSI of 0.45 and HSS and HKD of 0.6. Another study involving ML approaches was made by Meyer et al. (2015) where they compared four ML algorithms — random forests (RF), neural networks (NNET), averaged neural networks (AVNNET) and support vector machines (SVM) — for rainfall area detection and rainfall rate assignment using MSG SEVIRI data. In general the NNs presented the best performance. For the screening using NNs, they obtained values for POD of 0.80, FAR of 0.55, CSI of 0.4, HSS of 0.55 and HKD of 0.6. Comparing our results with these similar studies in the literature, NNIMBUS screening methodology is capable of achieving a good performance both in relation to DPR/GMI-CMB (V05) and GPROF (V05) and comparable scores with the recent literature. Remembering that it is necessary to perform an evaluation of the model in comparison with ground-based measurements.

Figure 5.13 - Screening of precipitation by NNIMBUS (left), GMI/DPR-CMB (center) and the overlap between of both (right) on 23 November 2018 at 04:44-06:17 UTC (orbit n° 26914).



SOURCE: Author's production.

5.6 Validation of NNIMBUS retrieval

The performance of the NNIMBUS retrieval is obtained considering a set of continuous scores (the formulas are given in Apendix A. Similar to the screening analysis, those scores are performed for two datasets: the test dataset and the validation dataset. The same satellite overpass showed for screening, for 30 November 2014 (orbit n 4299), is analyzed.

The **mean absolute error** (MAE) gives the average magnitude of the prediction errors which can be translated as the strength of the prediction error. The **root mean square error** (RMSE) is more sensitive to large errors because it squares the errors and therefore penalizes large deviations from the mean. The difference between MAE and RMSE error gives valuable information about the variance of the errors. The **standard deviation** is a measure that is used to quantify the amount of variation or dispersion of a set of data values. The **fractional standard error** (FSE) gives an idea of the extent to which an estimate is likely to deviate from the true population. And finally, the **correlation coefficient** measures the strength and direction of a linear relationship between the predictions and observations.

Table 5.6 presents the scores for the test dataset, which corresponds to 29298 pixels. We can see that the observed mean value and the predicted mean value are quite close (3.44 and 3.48, respectively). The MAE has value of 1.95 and RMSE 3.68. The standard deviation was equal 2.78, the FSE was equal 1.05 and correlation of 0.56.

Table 5.6 - Continuous scores obtained by the NNIMBUS rainfall retrieval for the test dataset.

Mean(obs)	Mean(pred)	MAE	RMSE	STD	FSE	CC	N° of pixels
3.44	3.48	1.95	3.68	2.78	1.05	0.56	29298

For the validation dataset the scores are presented in Table 5.7 and these refer to 22671 pixels. It is noticed that scores are higher when compared with the test dataset. The mean precipitation values are higher, presenting mean values of 5.84 in comparison with 2.54 from the observed dataset. The MAE has value of 4.19 and RMSE of 5.59. The standard deviation was a little bit higher than in the test dataset with value of 3.23 (compared to 2.78). The FSE also increased, presenting a value of 2.19 and the correlation remained low with value of 0.53.

Table 5.7 - Continuous scores obtained by the NNIMBUS rainfall retrieval for the validation dataset.

Mean(obs)	Mean(pred)	MAE	RMSE	STD	FSE	CC	N° of pixels
5.84	2.54	4.19	5.59	3.23	2.19	0.53	22671

The evaluation for same the satellite overpass showed for the screening is presented in Table 5.8. Despite the fact that this overpass refers to the same month of the validation dataset, it was previously removed from that set in order to be independently analyzed. It can be seen that the predicted mean precipitation values are similar to the validation, 5.34 versus the mean observed value of 2.66. The MAE, RMSE and FSE were lower than the validation dataset, with values of 3.35, 4.01 and 1.50, respectively. The correlation was higher (0.64) probably due the smaller number of pixels.

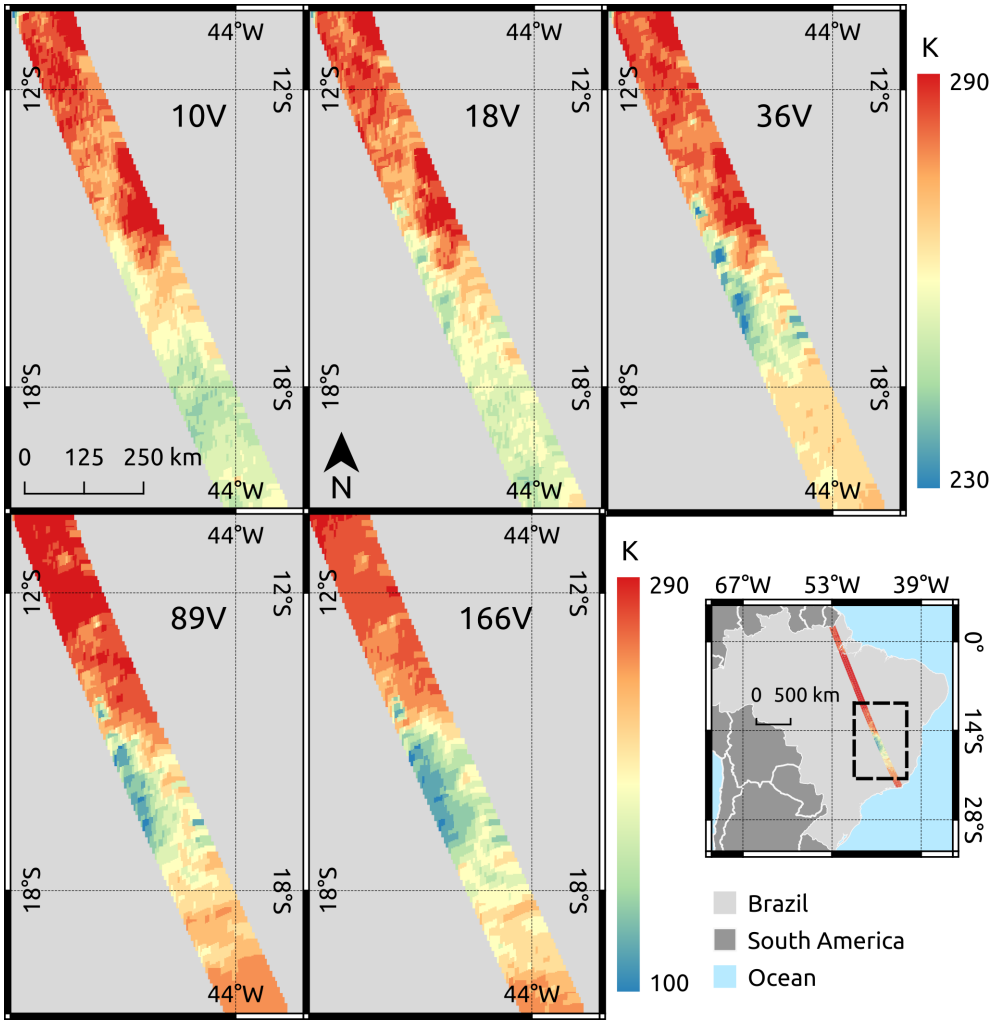
Table 5.8 - Continuous scores obtained by the NNIMBUS rainfall retrieval for the orbit of DPR/GMI-CMB on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4922).

Mean(obs)	Mean(pred)	MAE	RMSE	STD	FSE	CC	N° of pixels
5.34	2.66	3.35	4.01	3.29	1.50	0.64	784

Contrasting the three evaluations shown above, it is clear that the test dataset tends to present better scores, even if it is selected randomly from the training database, and although it is not included in the training, it seems to carry the signature from the input dataset distribution, which produces better scores when compared to the other sets that were analyzed. Unlike the analysis which was shown for the screening (in which we also compared our predictions with GPROF), here we will cover the rainfall retrieval (Figure 5.14) compared only with the GMI/DPR-CMB, the reason being that this is the only dataset in which all the input predictors are geo-colocated. Figure 5.14 also presents the GMI TBs: 10V, 18V, 36V, 89V and 166V on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4922). The analysis of the TB values is important in order to obtain a characterization of the atmosphere for the orbit to be analyzed, as well as to have an idea of how the input predictors are behaving in relation to precipitation in the rainfall maps. Thereby, the low frequency channel of 10GHz, presents the predominance of warm regions (280-290K, orange

to red colors) with a small region a little colder (250K - light green) at the left edge of the swath that can be related to presence of some clouds. In the same area at 18GHz there are more defined colder clusters (240K - green) associated with the presence of clouds. Also, the 36GHz clearly depicts a larger area with distinguishable signature from the environment, having cores with lower values around (230K). High frequencies are shown with a colder color scale, and we can see clusters with colder cores (130K) at 89GHz. Finally, we can see the larger coverage of the system in 166GHz which is probably depicting the presence of small ice in the anvil and cirrus clouds, and colder cores in the convective cores (reaching a minimum of 100K).

Figure 5.14 - GMI brightness temperatures for the channels 10V, 18V, 36V, 189V and 166V on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4922).



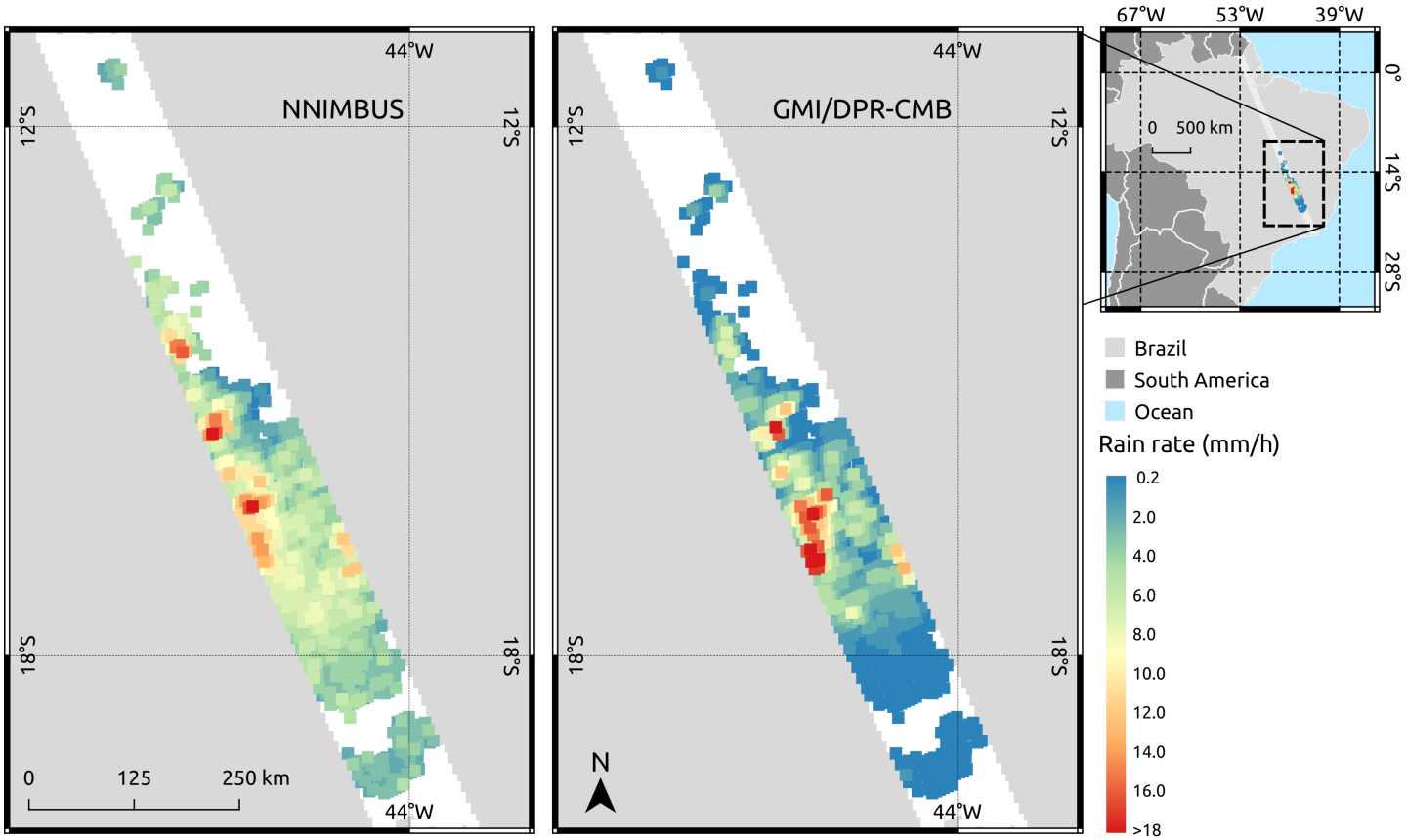
SOURCE: Author's production.

Insights about the model performance can also be drawn by verifying the predicted versus the observed rainfall maps (5.15). The left panel presents the NNIMBUS retrieval results against the GMI/DPR-CMB on the right with a minimum rainfall threshold set to 0.2 mm/h. At first glance the model captures the overall rainfall spatial distribution and detects the most intense cores (orange and reddish pixels). It is worth noticing that for the classes between 0.2 and 1 mm/h (blue colors) observed by the CMB, the NNIMBUS tends to overestimate, producing values between 4 and 6 mm/h (blue to light green). And by analyzing in detail the precipitation fields matrix for this orbit, the maximum observed precipitation rain rate for this overpass was 44 mm/h, meanwhile the NNIMBUS maximum value is in the order of 20 mm/h.

These observed patterns, even for a single orbit, combined with the fact that NNIMBUS did not generate classes between 0.2 and 1 (blue pixels) can lead us to suspect that the second cleaning criterion, the random under-sampling of C1 was very strict and, in turn, penalized the learning of the network for these classes. Thus, the network tended towards learning about the moderate classes (1 - 10 mm/h) which may have become more representative, reproducing more values in this range. Another notable aspect was the inability of NNIMBUS in reproducing the heavy classes (values $>$ 10 mm/h). This behaviour was observed by **??**), where the author analyzed the variability of the error in the precipitation estimation from the GPROF and CMB algorithms (V05), i.e., for the same observational database used in our study, but for subregions in the Amazon region. Their results saw an overestimation of the precipitation rain rate assessed by GPROF for lower values ($<$ 5 mm/h), regardless of the considered subregion, and an underestimation for higher values ($>$ 10 mm/h). Even though they have found this behaviour on the Amazon region, given the distribution seen in our exploratory analysis (section 5.2), it can be seen that this pattern is present in the dataset for the entire brazilian territory. Analyzing mean precipitation values estimated by the GPROF according to the CMB classes of estimation for the whole period, they found values no higher than 20 mm/h. In addition, for further understanding of the differences that they had found, they analyzed the relation by rain rate and ice content estimated by CMB product. And the comparison between the ice water path (IWP) estimates (CMB and GPROF) were those that presented more concrete differences, indicating that the errors in the precipitation estimation by GPROF are related to the IWP estimation. This error is probably related to the Bayesian approach, which tends to favor the median, inducing a decrease of the highest precipitation values, since these values are the most frequent in the *a priori* Bayesian database.

The results found by Costa (2018) corroborate the hypothesis that the distribution of the input data greatly impacts the final network's ability. Also, the hypothesis that the network architecture requires more adjustments should not be discarded, since the process is empirical, but from these considerations seen above, it is believed that a new treatment to make the dataset more properly balanced is the first step to see improvements in estimates. Futhermore, the construction of a longer dataset with observations from the new versions of the GPROF and CMB products added to the neural network adjustments could possibly bring improvements to the estimates generated by NNIMBUS.

Figure 5.15 - Rainfall retrieval from NNIMBUS (left) and GMI/DPR-CMB (right) on 30 November 2014 at 23:00-03:54 UTC (orbit n° 4922).



SOURCE: Author's production.

6 CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In order to develop a passive microwave-based satellite precipitation estimation algorithm for Brazil, this work was divided in two parts. As an initial proposal, the goal was to extend the cloud-radiation database used as a priori information for the CDRD Bayesian algorithm (SSMIS - H01) developed through the generation of cloud resolving model simulations for Brazil. As a first step it was necessary to evaluate the performance of CDRD in retrieving rainfall rates over the central Amazon and Southeast region of Brazil. The results obtained in the assessment of the CDRD (plus PNPR) products over Brazil concluded that CDRD tends to overestimate all the rain rate classes (light to heavy) in the Amazon region.

The performances of the PNPR (H02) for AMSU-A/AMSU-B/MHS sensors and PNPR (H18) for the AMTS sensor were also analyzed. For this region all the algorithms showed high FAR values and larger precipitation patterns which are deeply related to the precipitation screening scheme. The screening seems to be substantially affected by the high water vapor content in this region. The H01 specifically, was affected by highly variable surface emissivities, impacting the selection of hydrometeor profiles in the *a priori* database and, ultimately, the rainfall rate retrieval. On the other hand, for the Vale do Paraíba region, both algorithms, H01 and H02, produced ME values that were quite close to zero (or negative) and lower FAR values (from 0.21 to 0.49) than Manaus. Unlike the Manaus case, the precipitation patterns were well detected and the estimations were in good agreement with the reference, as indicated by the low ME values.

Regarding the generation of cloud resolving model simulations for the Vale do Paraíba and Amazon regions, it was observed that the simulations detected the occurrence of perturbations on TB fields in terms of position and time of occurrence, however in terms of the range of temperature values, the model did not reproduce the lowest values of TB seen on the observations. As for the microphysical columnar contents, the field that showed an unexpected behavior was the graupel content where the model seemed to struggle with the riming process on graupel formation. The largest amount of columnar content was associated with rain. The model seems to generate a great content of aggregate and snow. Concerning the microphysical rates the model was able to generate only the rain rate (maximum around 20 mm/h), being unable to generate the other contents. It should be noted that a deeper analysis with an extensive number of simulations could ensure good representativeness

of the simulated multichannel TBs in order to better describe the regional features of TB across a wide range of meteorological systems and microphysical conditions containing precipitation. Deep adjustments in mesoscale models and improvements mostly in the radiative transfer (mainly the scattering by non-spherical ice particles) could contribute towards gains in performance of the simulations in different regions of Brazil. Unfortunately, by the time this stage of the work was completed, we did not have the technical conditions to implement such changes.

For these reasons, the second part of the work was developed, by making use of an observational database in order to develop a screening of precipitation and rainfall retrieval algorithm over Brazil, based on ANN.

The NNIMBUS precipitation screening method proved to be very effective in both detecting larger systems and smaller or isolated systems. Regarding the GMI/DPR-CMB observations, the screening performed well, with an accuracy of 0.95, POD of 0.80, FAR of 0.39 and bias of 1.34. When compared to the GPROF the screening still had good performance, however with lower scores. It was observed that through the comparison maps with GPROF the NNIMBUS can detect agglomerates very similarly, however it does not detect the borders of the systems very well. This behavior might be associated with the precipitation thresholds that were configured with the training dataset (0.2 a 60 mm/h), which might be leading more stratiform regions of the systems to go undetected.

The NNIMBUS rainfall retrieval model also performed well when compared to the GMI/DPR-CMB observations, with an MAE of 4.19, standard deviation of 3.23 and RMSE of 5.59 for the validation dataset. When analyzing the precipitation field that was forecast for an orbit, it became clear that the retrieval tends to underestimate classes between 0.2 and 1 mm/h, overestimate classes between 1 and 10 mm/h and underestimate classes greater than 10 mm/h. The observed behavior towards lower classes might be related to the data cleaning treatment process, in which a "random reduction" was performed in the input database, for those values were exceptionally more frequent than the other classes (76% of the dataset), and that might have penalized said values during the learning process. Thus, the network leaned towards learning about the moderate classes (1 - 10 mm/h) which may have become more representative, reproducing more values in this range. The struggle in forecasting higher classes, on the other hand, is associated with their distribution in the input dataset in which these classes are the least frequent and, therefore, poorly represented. With these results in mind, it becomes clear that a new dataset

is required, one with greater timeframe and more representativity for higher classes.

Although the literature presents a wide variety of mathematical approaches to tackle inversion problems, the use of ANN proved through its statistical scores showed to be highly efficient in solving complex non-linear problems and also dealing with the high variability of the input data that is inherent to the nature of the rainfall phenomena.

The NNIMBUS time performance, once trained, was shown to be very effective in terms of prediction time, generating, in a matter of seconds, a new forecast based on data from dataset it had not worked with yet. Thus, the ANN is capable of making complex and accurate predictions (accuracy = 0.80) in less than 60 seconds in new satellite overpasses, making it extremely fast and highly adequate for operational environments.

It is also worth mentioning that the time taken to train the NNIMBUS ANN was calculated for a given hardware configuration and using only the CPU cores. The Orion and Helios clusters in which the ANN where trained were also equipped with three NVidia Tesla K20 graphic cards each. These GPU cards where not even used at the time of this work for library-compatibility reasons, given that the TensorFlow python library is highly capable of parallelization over GPU resources. The reason for not using the GPU cores is that the TF, NVIDIA, CUDA, OS and its Kernel needs to be tightly fitted together regarding the version of the libraries for each of these softwares. This means that the NNIMBUS has a lot of space for performance improvement without even changing its code, but instead by calling the tensorflow-gpu libraries in its core python imports.

6.2 Future work

This work presented important results regarding the development of an artificial neural network for microwave rainfall retrieval optimized for Brazil, based on the GMI and DPR (CMB), which are state-of-the-art precipitation sensors. Future work could build upon the knowledge expressed in this work to delve deeper into some issues, such as:

- To perform new verifications and validation in a denser and more comprehensive ground-based measurements.
- Identification of new criteria to clean and balance the distribution of the input dataset.

- To build a longer observational dataset in order to obtain more representativeness of the rain rate distribution classes.
- Exploitation of the capability of TensorFlow python library to do processing using GPUs.

REFERENCES

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANé, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIéGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. **TensorFlow: large-scale machine learning on heterogeneous systems**. 2015. Software available from tensorflow.org. Disponível em: <http://tensorflow.org/>. Acesso em: 20 nov. 2018. 66

AMARAL, L. M. C.; BARBIERI, S.; VILA, D.; PUCA, S.; VULPIANI, G.; PANEGROSSI, G.; BISCARO, T.; SANÒ, P.; PETRACCA, M.; MARRA, A. C.; GOSSET, M.; DIETRICH, S. Assessment of ground-reference data and validation of the H-SAF precipitation products in Brazil. **Remote Sensing**, v. 10, p. 1–24, 2018. Disponível em: <http://www.mdpi.com/2072-4292/10/11/1743>. Acesso em: 10 dez. 2018. 21, 24, 25, 40, 113

ANOCHI, J. A. **Previsão climática de precipitação por redes neurais autoconfiguradas**. 161 p. (09.16.22.02-TDI). Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2015. Disponível em: <http://urlib.net/8JMKD3MGP3W34P/3K98PDP>. Acesso em: 2 set. 2018. 17

ATKINSON, N. C. Calibration, monitoring and validation of AMSU-B. **Advances in Space Research**, v. 28, n. 1, p. 117–126, 2001. Disponível em: <https://doi.org/10.1016/S0273-1177(01)00312-X>. Acesso em: 10 maio 2016. 8

BARBIERI S. **Testing of dual-polarization processing algorithms for radar rainfall estimation in tropical scenarios**: H-SAF visiting scientist report (EUMETSAT). 2017. Disponível em: <http://hsaf.meteoam.it/documents/visiting-scientist/HSAF_VS17_Final_Report.pdf>. Acesso em: 20 nov. 2017. 21, 22, 23, 24, 25, 27, 30, 31, 33, 35, 36, 40, 48

BEUSCH, L.; FORESTI, L.; GABELLA, M.; HAMMAN, U. Satellite-based rainfall retrieval: from generalized linear models to artificial neural networks.

Remote Sensing, v. 10, p. 1–24, 2018. Disponível em: <https://www.mdpi.com/2072-4292/10/6/939>. Acesso em: 8 de jul. 2018. 18, 78

BISCARO, T. S. **Estimativa de precipitação via sensoriamento remoto em microondas na Bacia Amazônica**. 134 p. Dissertação (Mestrado em Meteorologia) — Universidade de São Paulo (USP), São Paulo, 2006. Disponível em: <http://www.iag.usp.br/pos/sites/default/files/m_thiago_s_biscaro.pdf>. Acesso em: 2 fev. 2016. 9, 13

BISCARO, T. S.; MORALES, C. A. Continental passive microwave-based rainfall estimation algorithm: application to the amazon basin. **Journal of Applied Meteorology and Climatology**, v. 47, p. 1962–1981, 2008. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/2007JAMC1744.1>. Acesso em: 5 fev. 2016. 9, 10

BOUKABARA S. A. AND GARRET K. AND BLACKWELL B. **ATMS description expected performances**: Darmstadt, germany. 2011. Disponível em: <htpps://www.eumetsat.int>. Acesso em: 20 abr. 2018. 9

BRINGI, V.; CHANDRASEKAR, V. **Polarimetric doppler radar: principles and applications**. Cambridge University Press, 2001. 636 p. (ISBN: 9780511541094). Disponível em: <https://www.cambridge.org/core/books/polarimetric-doppler-weather-radar/C93A3269D222708EADB81C1E4504BB4F>. Acesso em: 03 jun. 2018. 24

BROWNLEE, J. **Master machine learning algorithms**. Machine learning mastery, 2016. Disponível em: <https://machinelearningmastery.com/master-machine-learning-algorithms/>. 16, 19

CALHEIROS, A. J. P. **Propriedades radiativas e microfísicas das nuvens continentais: uma contribuição para a estimativa da precipitação de nuvens quentes por satélite**. 204 p. (10.07.19.43-TDI). Tese (Doutorado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013. Disponível em: <http://mtc-m16d.sid.inpe.br/rep/sid.inpe.br/mtc-m19/2013/10.07.19.43?metadatarepository=sid.inpe.br/mtc-m19/2013/10.07.19.43.37{&}ibiurl.backgroundlanguage=pt{&}ibiurl.requiredsite=mtc-m16d.sid.

inpe.br+806{&}requiredmirror=sid.inpe.br/mtc-m19@80/2009/08.21.17.0>. Acesso em: 2 mar.2018. 14

CALHEIROS, A. J. P.; MACHADO, L. A. T. Cloud and rain liquid water statistics in the CHUVA campaign. **Atmospheric Research**, v. 144, p. 126–140, 2014. Disponível em: <http://dx.doi.org/10.1016/j.atmosres.2014.03.006>. Acesso em: 5 ago. 2016. 23

CARVALHO, L. M. V.; JONES, C.; SILVA, A. E.; LIEBMNAN, B.; DIAS, P. L. S. The South American Monsoon System and the 1970s climate transition. **International Journal of Climatology**, v. 31, p. 1248–1256, 2011. Disponível em: <https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/joc.2147>. Acesso em: 10 jun. 2016. 23

CASELLA, D. **Analisi e verifica del metodo Cloud Dynamics and radiation database (CDRD) per la stima Bayesiana della precipitazione con radiometri alle microonde montati su satellite**. 122 p. Tese (Doutorado em Sensoriamento Remoto) — Università degli Studi di Roma La Sapienza, Roma, 2010. Disponível em: <https://www.researchgate.net/publication/236646056_Analisi_e_verifica_del_metodo_Cloud_Dinamics_and_Radiation_Database_per_la_stima_bayesiana_della_precipitazione_con_radiometri_alle_microonde_montati_su_satellite>. Acesso em: 24 set. 2015. 8, 45

CASELLA, D.; AMARAL, L. M. C.; DIETRICH, S.; MARRA, A. C.; SANÒ, P.; PANEGROSSI, G. The cloud dynamics and radiation database algorithm for AMSR2: exploitation of the GPM observational dataset for operational applications. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 10, p. 3985–4001, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7959562>. Acesso em: 2 fev. 2018. 114

CASELLA, D.; DIETRICH, S.; FORMENTON, M.; MUGNAI, A.; PANEGROSSI, G.; SANÒ, P.; SMITH, E.; TRIPOLI, G. J. Verification of Cloud Dynamics and Radiation Database (CDRD) passive microwave precipitation retrieval algorithm using TRMM satellite radar and radiometer measurements over Southern Mediterranean Basin. In: 12TH SPECIALIST MEETING ON MICROWAVE RADIOMETRY AND REMOTE SENSING OF THE ENVIRONMENT (MICRORAD), 2012, Rome, Italy. **Proceedings...** Rome, 2012. p. 1–4. 2

CASELLA, D.; PANEGROSSI, G.; P.SANÒ, L. M.; PETRACCA, M.; DIETRICH, S. A novel algorithm for detection of precipitation in tropical regions using PMW radiometers. **Atmospheric Measurement Techniques**, v. 8, p. 1217–1232, 2015. Disponível em: <http://www.atmos-meas-tech.net/8/1217/2015/>. Acesso em: 20 out. 2016. 11

CASELLA, D.; PANEGROSSI, G.; SANÒ, P.; DIETRICH, S.; MUGNAI, A.; SMITH, E.; TRIPOLI, G. J.; FORMENTON, M.; PAOLA, F.; LEUNG, W. Y. Transitioning from CRD to CDRD in Bayesian retrieval of rainfall from satellite passive microwave measurements: Part 2. overcoming database profile selection ambiguity by consideration of meteorological control on microphysics. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, p. 4650–4671, 2013. Disponível em: <https://ieeexplore.ieee.org/document/6527927>. Acesso em: 13 out. 2015. 1, 11, 25

CASELLA, D.; PANEGROSSI, G.; SANÒ, P.; MARRA, A. C.; DIETRICH, S.; JOHNSON, B. T.; KULIE, M. S. Evaluation of the GPM-DPR snowfall detection capability: Comparison with CloudSat-CPR. **Atmospheric Research**, Elsevier, v. 197, p. 64–75, 2017. ISSN 01698095. Disponível em: <https://doi.org/10.1016/j.atmosres.2017.06.018>. 11, 12

CECIL, D. J.; CHRONIS, T. Polarization-corrected temperatures for 10-, 19-, 37-, and 89-GHz passive microwave frequencies. **Journal of Applied Meteorology and Climatology**, v. 57, p. 2249–2265, 2018. Disponível em: <https://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-18-0022.1>. Acesso em: 20 dez.2018. 9

CHEN, F.; STAELIN, D. H. AIRS/AMSU/HSB precipitation estimates. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, p. 410–417, 2003. Disponível em: <http://ieeexplore.ieee.org/document/1196057/>. Acesso em: 20 abr.2016. 37

CHOLLET, FRANÇOIS AND OTHERS. **Keras: the python deep learning library**: Keras. 2015. Disponível em: <https://keras.io>. Acesso em: 20 nov. 2018. 66

COSTA, I. C. da. **Influências das propriedades microfísicas e radiativas das nuvens nas incertezas das estimativas de precipitação por sensores remotos passivos**. 131 p. (04.26.12.59-TDI). Tese (Doutorado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2018.

Disponível em: <http://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2018/04.26.12.59/doc/publicacao.pdf>. Acesso em: 2 nov. 2018. 15, 85

COSTES, L.; BUCKLEY, C. B. ad M. J.; MASON, G. Microwave Humidity Sounder (MHS) antenna. In: SENSORS, SYSTEMS, AND NEXT-GENERATION SATELLITES III, 1999, Florence, Italy. **Proceedings...** Florence, 1999. p. 412–426. 9

DRAPER, D. W.; NEWELL, D. A.; WENTZ, F. J.; KRIMCHANSKY, S.; SKOFRONICK-JACKSON, G. M. The Global Precipitation Measurement (GPM) microwave imager (GMI): instrument overview and early on-orbit performance. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, p. 3452–3462, 2015. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=7052302>. Acesso em: 21 jul. 2016. 8

ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, p. 115 EP –, 2017. Disponível em: <https://doi.org/10.1038/nature21056>. Acesso em: 19 abr. 2017. 16

EVANS, K. F.; TURK, J.; WONG, T.; STEPHENS, G. L. A Bayesian Approach to Microwave Precipitation Profile Retrieval. **Journal of Applied Meteorology**, v. 34, p. 260–279, 1995. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0450-34.1.260>. Acesso em: 8 abr. 2015. 1, 10

FERRARO, R. R.; GRODY, N. C.; KOGUT, J. A. Classification of Geophysical Parameters Using Passive Microwave Satellite Measurements. **IEEE Transactions on Geoscience and Remote Sensing**, v. 24, p. 1008–1013, 1986. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=4072573>. Acesso em: 1 abr. 2019. 12

FERRARO, R. R.; MARKS, G. F. The development of SSM/I rain-rate retrieval algorithms using ground-based radar measurements. **Journal of Atmospheric and Oceanic Technology**, v. 12, p. 755–770, 1995. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0426%281995%29012%3C0755%3ATDOSRR%3E2.0.CO%3B2>. Acesso em: 5 mar. 2017. 14

GARDNER, M.; DORLING, S. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. **Atmospheric Environment**, v. 32, p. 2627–2636, 1998. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1352231097004470>. Acesso em: 29 jan. 2018. 17

GASIEWSKI, A. J. Microwave radiative transfer in hydrometeors. In: JANSSEN, M. A. (Ed.). **Atmospheric remote sensing by microwave radiometry**. New York: John Wiley Sons, 1993. p. 91–250. 9

GONG, J.; WU, D. L. Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements. **Atmospheric Chemistry and Physics**, v. 17, p. 2741–2757, 2017. Disponível em: <https://www.atmos-chem-phys.net/17/2741/2017/>. Acesso em: 3 mar. 2018. 63

GRODY, N. C. Classification of snow cover and precipitation using the special sensor microwave imager. **Journal of Geophysical Research**, v. 96, p. 7423–7435, 1991. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/91JD00045>. Acesso em: 3 maio 2017. 12

GRODY, N. C.; WEN, F. Microwave emission and scattering from deserts: theory compared with satellite measurements. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, p. 361–375, 2008. Disponível em: <https://ieeexplore.ieee.org/document/4432709>. Acesso em: 12 nov. 2016. 63

GUZELLA, T. S.; CAMINHAS, W. M. A review of machine learning approaches to spam filtering. **Expert Systems with Applications**, v. 36, p. 10206 – 10222, 2009. Disponível em: <http://www.sciencedirect.com/science/article/pii/S095741740900181X>. Acesso em: 22 ago. 2018. 16

HAYKIN, S. **Neural networks: a comprehensive foundation**. Pearson, 1998. 823 p. 2ed. Disponível em: <https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural%20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf>. Acesso em: 20 out. 2018. 17

HAYKIN, S. **Neural networks and learning machines**. Pearson, 2009. 938 p. 3ed. Disponível em: <https://cours.etsmtl.ca/sys843/REFS/Books/ebook_Haykin09.pdf>. Acesso em: 11 out. 2018. 16, 19, 20

HOLLINGER, J. P.; PEIRCE, J. L.; POE, G. A. SSM/I instrument evaluation. **IEEE Transactions on Geoscience and Remote Sensing**, v. 28, p. 781–790, 1990. Disponível em: <https://ieeexplore.ieee.org/document/58964>. Acesso em: 29 set. 2018. 8

HOU, A. Y.; KAKAR, R. K.; NEECK, S.; AZARBARZIN, A. A.; KUMMEROW, C. D.; KOJIMA, M.; OKI, R.; NAKAMURA, K.; IGUCHI, T. The global precipitation measurement mission. **Bulletin of the American Meteorological Society**, v. 95, p. 701–722, 2014. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-13-00164.1>. Acesso em: 12 ago. 2016. 11, 58

HSU, K.-l.; GAO, X.; SOROOSHIAN, S.; GUPTA, H. V. Precipitation estimation from remotely sensed information using artificial neural networks. **Journal of Applied Meteorology**, v. 36, p. 1176–1190, 1997. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0450{%}281997{%}29036{%}3C1176{%}3APEFRSI{%}3E2.0.CO{%}3B2>. Acesso em: 17 maio 2016. 17

IGUCHI, T.; OKI, R.; SMITH, E. A.; FURUHAMA, Y. Global precipitation measurement program and the development of dual-frequency precipitation radar. **Journal of the Communications Research Laboratory**, v. 49, p. 37–45, 2002. Disponível em: <https://pdfs.semanticscholar.org/3b17/564c88e9060b8a8f811165da314c33a7b1a2.pdf?_ga=2.14844533.2125990033.1562593147-141899584.1562593147>. Acesso em: 15 fev. 2017. 56

IMAOKA, K.; KACHI, M.; KASAHARA, M.; ITO, N.; NAKAGAWA, K.; OKI, T. Instrument performance and calibration of AMSR-E and AMSR2. **ISPRS Archives**, v. 38, p. 13–16, 2010. Disponível em: <https://www.isprs.org/proceedings/XXXVIII/part8/pdf/JTS13{_}20100322190615.pdf>. Acesso em: 23 jun. 2017. 8

KAWANISHI, T.; SEZAI, T.; ITO, Y.; IMAOKA, K.; TAKESHIMA, T.; ISHIDO, Y.; SHIBATA, A.; MIURA, M.; INAHATA, H.; SPENCER, R. W. The Advanced

Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. **IEEE Transacions on Geoscience and Remote Sensing**, v. 41, p. 184–194, 2003. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=1196037>. Acesso em: 28 abr. 2016. 8

KOZA, J. R.; BENNETT, F. H.; ANDRE, D.; KEANE, M. A. An automated design of both the topology and sizing of analog electrical circuits using genetic programming. In: GERO, J. S.; SUDWEEKS, F. (Ed.). **Artificial Intelligence in Design '96**. Dordrecht: Springer Netherlands, 1996. p. 151–170. 16

KUMMEROW, C.; HONG, Y.; OLSON, W. S.; YANG, S.; ADLER, R. F.; MCCOLLUM, J.; FERRARO, R.; PETTY, G.; SHIN, D.-B.; WILHEIT, T. T. The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. **Journal of Applied Meteorology**, v. 40, p. 1801–1820, 2001. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0450{%}282001{%}29040{%}3C1801{%}3ATEOTGP{%}3E2.0.CO{%}3B2>. Acesso em: 13 jan. 2016. 1, 10, 12

KUMMEROW, C.; KOZU, W. B. ans T.; SHIUE, J.; SIMPSON, J. The Tropical Rainfall Measuring Mission (TRMM) sensor package. **Journal of Atmospheric and Oceanic Technology**, v. 15, p. 809–817, 1998. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0426{%}281998{%}29015{%}3C0809{%}3ATTRMMT{%}3E2.0.CO{%}3B2>. Acesso em: 15 jul. 2016. 8, 11

KUMMEROW, C.; OISON, W. S.; GIGLI., L. A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. **IEEE Transacions on Geoscience and Remote Sensing**, v. 34, p. 1213–1232, 1996. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=536538>. Acesso em: 28 abr. 2016. 1, 10

KUMMEROW, C. D.; RINGERUD, S.; CROOK, J.; RANDEL, D.; BERG, W. An observationally generated a priori database for microwave rainfall retrievals. **Journal of Atmospheric and Oceanic Technology**, v. 28, p. 113–130, 2011. Disponível em:

<https://journals.ametsoc.org/doi/full/10.1175/2010JTECHA1468.1>. Acesso em: 8 jan. 2016. 12

LEVIZZANI V. AND AMORATI R. AND MENEGUZZO F. **MUSIC – MUltiple-Sensor precipitation measurements, Integration, Calibration and flood forecasting**: A Review of satellite-based rainfall estimation methods. 2002. Disponível em: <http://satmet.isac.cnr.it/papers/MUSIC-Rep-Sat-Precip-6.1.pdf>. Acesso em: 16 nov. 2015. 9

MACHADO, L. A. T.; DIAS, M. A. F. S.; MORALES, C.; FISCH, C.; VILA, D.; ALBRECHT, R.; GOODMAN, S.; CALHEIROS, A. J. P.; BISCARO, T.; KUMMEROW, C.; COHEN, J.; FITZJARRALD, D.; NASCIMENTO, E. L.; SAKAMOTO, M.; CUNNINGHAM, C.; CHABOUREAU, J. P.; PETERSEN, W. A.; ADAMS, D. K.; BALDINI, L.; ANGELIS, C. F.; SAPUCCI, L. F.; SALIO, P.; BARBOSA, H. M. J.; LANDULFO, E.; SOUZA, R. A. F.; BLAKESLEE, R. J.; BAILEY, J.; FREITAS, S.; LIMA, W.; TOKAY, A. The Chuva Project: how does convection vary across Brazil? **Bulletin of the American Meteorological Society**, v. 95, p. 1365–1380, 2014. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-13-00084.1>. Acesso em: 29 jan. 2015. 3

MARZBAN, C.; STUMPF, G. J. A neural network for tornado prediction based on doppler radar-derived attributes. **Journal of Applied Meteorology**, v. 35, p. 617–626, 1996. Disponível em: <https://doi.org/10.1175/1520-0450(1996)035<0617:ANNFTP>2.0.CO;2>. Acesso em: 17 fev. 2018. 17

MCCANN, D. W. A neural network short-term forecast of significant thunderstorms. **Weather and Forecasting**, v. 7, p. 525–534, 1992. Disponível em: <https://doi.org/10.1175/1520-0434(1992)007<0525:ANNSTF>2.0.CO;2>. Acesso em: 10 set. 2018. 17

MEYER, H.; KÜHNLEIN, M.; APPELHANS, T.; NAUSS, T. A neural network short-term forecast of significant thunderstorms. **Atmospheric Research**, v. 169, p. 424–433, 2015. Disponível em: <http://dx.doi.org/10.1016/j.atmosres.2015.09.021>. Acesso em: 15 set. 2018. 18, 78

MEYERS, P.; FERRARO, R. R.; WANG, N. Y. Updated screening procedures for gprof2010 over land: utilization for AMSR-E. **Journal of Atmospheric and**

Oceanic Technology, v. 32, p. 1015–1028, 2015. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/JTECH-D-14-00149.1>. Acesso em: 21 set. 2018. 12

MICHAELIDES, S.; LEVIZZANI, V.; ANAGNOSTOU, E.; BAUER, P.; KASPARIS, T.; LANE, L. E. Precipitation: measurement, remote sensing, climatology and modeling. **Atmospheric Research**, v. 94, p. 512–533, 2009. Disponível em: <http://dx.doi.org/10.1016/j.atmosres.2009.08.017>. Acesso em: 21 out. 2017. 9

MUGNAI, A.; CASELLA, D.; CATTANI, E.; DIETRICH, S.; LAVIOLA, S.; LEVIZZANI, V.; PANEGROSSI, G.; PETRACCA, M.; SANÒ, P.; PAOLA, F. D.; BIRON, D.; LEONIBUS, L. D.; MELFI, D.; ROSCI, P.; VOCINO, A.; ZAULI, F.; PAGLIARA, P.; PUCA, S.; RINOLLO, A.; MILANI, L.; PORCÙ, F.; GATTARI, F. Precipitation products from the hydrology SAF. **Natural Hazards and Earth System Science**, v. 13, p. 1959–1981, 2013. Disponível em: <http://www.nat-hazards-earth-syst-sci.net/13/1959/2013/>. Acesso em: 4 maio 2016. 25

MUGNAI, A.; SMITH, E. A.; TRIPOLI, G. J. Foundations for statistical-physical precipitation retrieval from passive microwave satellite measurements. part ii: emission-source and generalized weighting-function properties of a time-dependent cloud-radiation model. **Journal of Applied Meteorology**, v. 32, p. 17–39, 1993. Disponível em: <https://doi.org/10.1175/1520-0450(1993)032<0017:FFSPRF>2.0.CO;2>. Acesso em: 15 out. 2016. 1, 10

MUGNAI, A.; SMITH, E. A.; TRIPOLI, G. J.; BIZZARRI, B.; CASELLA, D.; DIETRICH, S.; PAOLA, F. D.; PANEGROSSI, G.; SANÒ, P. CDRD and PNPR satellite passive microwave precipitation retrieval algorithms: EuroTRMM/EURAINSAT origins and H-SAF operations. **Natural Hazards and Earth System Science**, v. 13, p. 887–912, 2013. Disponível em: <https://www.nat-hazards-earth-syst-sci.net/13/887/2013/>. Acesso em: 2 dez. 2015. 11, 25, 37

NAVONE, H. D.; CECCATTO, H. A. Predicting indian monsoon rainfall: a neural network approach. **Climate Dynamics**, v. 10, p. 1432–0894, 1994. Disponível em: <https://doi.org/10.1007/BF00228029>. Acesso em: 1 nov. 2018. 17

NUNES, A. M. P.; DIAS, M. A. F. S.; ANSELMO, E. M.; MORALES, C. A. Severe convection features in the amazon basin: A trmm-based 15-year evaluation.

Frontiers in Earth Science, v. 4, p. 1–14, 2016. Disponível em: <http://journal.frontiersin.org/article/10.3389/feart.2016.00037>. Acesso em: 18 nov. 2016. 23

OLIVEIRA, R.; MAGGIONI, V.; VILA, D.; MORALES, C. Characteristics and diurnal cycle of GPM rainfall estimates over the central amazon region. **Remote Sensing**, v. 8, p. 20, 2016. Disponível em: <http://www.mdpi.com/2072-4292/8/7/544/htm>. Acesso em: 15 mar. 2017. 33

OLIVEIRA, R. A. J. de. **Characteristics and error modeling of GPM satellite rainfall estimates during chuva campaign in Brazil**. 163 p. (05.22.17.16-TDI). Tese (Doutorado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017. Disponível em: <http://urlib.net/sid.inpe.br/mtc-m21b/2017/05.22.17.16>. Acesso em: 13 fev. 2018. 10, 14

PANEGROSSI, G.; DIETRICH, S.; MARZANO, F. S.; MUGNAI, A.; SMITH, E. A.; TRIPOLI, G. J.; WANG, P. K.; BAPTISTA, J. P. V. P. Use of cloud model microphysics for passive microwave-based precipitation retrieval: significance of consistency between model and measurement manifolds. **Journal of the Atmospheric Sciences**, v. 55, p. 20, 1998. Disponível em: <https://doi.org/10.1175/1520-0469(1998)055<1644:UOCMMF>2.0.CO;2>. Acesso em: 19 mar. 2016. 11

PANEGROSSI, G.; RYSMAN, J. F.; CASELLA, D.; MARRA, A. C.; SANÒ, P.; KULIE, M. S. Cloudsat-based assessment of GPM microwave imager snowfall observation capabilities. **Remote Sensing**, v. 9, p. 28, 2017. Disponível em: <https://doi.org/10.1175/1520-0469(1998)055<1644:UOCMMF>2.0.CO;2>. Acesso em: 28 jan. 2018. 58

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: machine learning in python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. 66

PUCA, S.; PORCU, F.; RINOLLO, A.; VULPIANI, G.; BAGUIS, P.; BALABANOVA, S.; CAMPIONE, E.; ERTÜRK, A.; GABELLANI, S.; IWANSKI, R.; JURAsEK, M.; KAnÁK, J.; KERÉNYI, J.; KOSHINCHANOV, G.; KOZINAROVA, G.; KRAHE, P.; LAPETA, B.; LÁBÓ, E.; MILANI, L.; OKON,

L.; ÖZTOPAL, A.; PAGLIARA, P.; PIGNONE, F.; RACHIMOW, C.; REBORA, N.; ROULIN, E.; SÖNMEZ, I.; TONIAZZO, A.; BIRON, D.; CASELLA, D.; CATTANI, E.; DIETRICH, S.; PAOLA, F. D.; LAVIOLA, S.; LEVIZZANI, V.; MELFI, D.; MUGNAI, A.; PANEGROSSI, G.; PETRACCA, M.; SANÒ, P.; ZAULI, F.; ROSCI, P.; LEONIBUS, L. D.; AGOSTA, E.; GATTARI, F. The validation service of the hydrological SAF geostationary and polar satellite precipitation products. **Natural Hazards and Earth System Sciences**, v. 14, p. 871–889, 2014. Disponível em: <https://www.nat-hazards-earth-syst-sci.net/14/871/2014/>. Acesso em: 6 maio 2016. 26

SANÒ, P.; PANEGROSSI, G.; CASELLA, D.; PAOLA, F. D.; MILANI, L.; MUGNAI, A.; PETRACCA, M.; DIETRICH, S. The passive microwave neural network precipitation retrieval (PNPR) algorithm for AMSU/MHS observations: description and application to european case studies. **Atmospheric Measurement Techniques**, v. 8, p. 837–857, 2015. Disponível em: <http://www.atmos-meas-tech.net/8/837/2015/>. Acesso em: 13 jul. 2016. 1, 11, 17, 25, 37

SANÒ, P.; PANEGROSSI, G.; CASELLA, D.; MARRA, A. C.; PAOLA, F. D.; DIETRICH, S. The new passive microwave neural network precipitation retrieval (PNPR) algorithm for the cross-track scanning ATMS radiometer: description and verification study over europe and africa using GPM and TRMM spaceborne radars. **Atmospheric Measurement Techniques**, v. 9, p. 5441–5460, 2016. Disponível em: <http://www.atmos-meas-tech.net/9/5441/2016/>. Acesso em: 12 maio 2017. 11, 17, 25, 32

SANÒ, P.; PANEGROSSI, G.; CASELLA, D.; MARRA, A. C.; D'ADDERIO, L.; RYSMAN, J. F.; DIETRICH, S. The passive microwave neural network precipitation retrieval (PNPR) algorithm for the CONICAL scanning global microwave imager (GMI) radiometer. **Remote Sensing**, v. 10, p. 21, 2018. Disponível em: <https://www.mdpi.com/2072-4292/10/7/1122>. Acesso em: 24 out. 2018. 11, 12, 13, 17, 57, 78

SHIMODA, H. GCOM missions. **IEEE International Geoscience and Remote Sensing Symposium (IGARSS)**, v. 6, p. 4201–4204, 2005. Disponível em: <https://ieeexplore.ieee.org/document/1525844>. Acesso em: 27 jan. 2016. 8

SKOFRONICK-JACKSON, G.; KIRSCHBAUM, D.; PETERSEN, W.; HUFFMAN, G.; KIDD, C.; STOCKER, E.; KAKAR, R. The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: reviewing four years of advanced rain and snow observations. **Quarterly Journal of the Royal Meteorological Society**, v. 144, n. July 2017, p. 27–48, 2018. ISSN 1477870X. Disponível em: <https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/qj.3313>. 2, 71

SMITH, E. A.; LEUNG, H. W. Y.; ELSNER, J. B.; MEHTA, A. V.; TRIPOLI, G. J.; CASELLA, D.; DIETRICH, S.; MUGNAI, A.; PANEGROSSI, G.; SANÒ, P. Transitioning from CRD to CDRD in bayesian retrieval of rainfall and physics from satellite passive microwave measurements: Part 3 – atmospheric identification of optimal meteorological tags. **Natural Hazards and Earth System Sciences**, v. 51, p. 4119–4143, 2013. Disponível em: <https://www.nat-hazards-earth-syst-sci.net/13/1185/2013/nhess-13-1185-2013.pdf>. Acesso em: 15 set. 2015. 2, 11, 25, 41

SPENCER, R.; GOODMAN, M. H.; HOOD, R. Precipitation retrieval over land and ocean with the SSMI identification ans characteristics of the scattering signal. **Journal of atmospheric and oceanic technology**, v. 6, p. 254–273, 1989. Disponível em: <https://doi.org/10.1175/1520-0426(1989)006<0254:PROLAO>2.0.CO;2>. Acesso em: 23 jan. 2017. 9

STAELIN, D. H. Passive microwave techniques for geophysical sensing of the earth from satellites. **Journal of atmospheric and oceanic technology**, v. 29, p. 683–687, 1981. Disponível em: <https://ieeexplore.ieee.org/abstract/document/1142630/>. Acesso em: 14 set. 2016. 7

SURUSSAVADEE, C.; STAELIN, D. H. Global millimeter-wave precipitation retrievals trained with a cloud-resolving numerical weather prediction model, part ii: performance evaluation. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, p. 109–118, 2008. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=4407633>. Acesso em: 5 abr. 2016. 70

TANG, L.; TIAN, Y.; LIN, X. Validation of precipitation retrievals over land from satellite-based passive microwave sensors. **Journal of Geophysical Research:**

Atmospheres, v. 6, p. 4201–4204, 2014. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1002/2013JD020933>. Acesso em: 22 out. 2016. 11

TAPIADOR, F.; TURK, F. J.; PETERSEN, W.; HOU, A. Y.; GARCÍA-ORTEGA, E.; MACHADO, L. A. T.; ANGELIS, C. F.; SALIO, P.; KIDD, C.; KUMMEROW, C.; HUFFMAN, G. J.; CASTRO, M. Global precipitation measurement: methods, datasets and applications. **Atmospheric Research**, v. 104-105, p. 70–97, 2012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169809511003607>. Acesso em: 19 ago. 2015. 12

TAPIADOR, F. J.; KIDD, C.; HSU, K. L.; MARZANO, F. Neural networks in satellite rainfall estimation. **Journal of Geophysical Research: Atmospheres**, v. 11, p. 83–91, 2004. Disponível em: <https://www.cambridge.org/core/article/neural-networks-in-satellite-rainfall-estimation/ED3BF8AE3038A0D617DA589C18A148BB>. Acesso em: 7 ago. 2018. 16

TRIPOLI, G. J. A nonhydrostatic mesoscale model designed to simulate scale interaction. v. 120, p. 1342–1359, 1992. Disponível em: <https://doi.org/10.1175/1520-0493(1992)120<1342:ANMMDT>2.0.CO;2>. Acesso em: 13 dez. 2015. 43

TRIPOLI, G. J.; SMITH, E. A. Introducing variable-step topography (VST) coordinates within dynamically constrained nonhydrostatic modeling system (NMS). part 1: VST formulation within NMS host model framework. v. 66, p. 28–57, 2014. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0377026514000025>. Acesso em: 11 jul. 2016. 43

_____. Introducing variable-step topography (VST) coordinates within dynamically constrained nonhydrostatic modeling system (NMS). part 2: VST performance on orthodox obstacle flows, journal = "dynamics of atmospheres and oceans. v. 10-27, p. 10–27, 2014. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0377026514000049>. Acesso em: 11 jul. 2016. 43

VILA, D.; HERNANDEZ, C.; FERRARO, R. R.; SEMUNEGUS, H. The performance of hydrological monthly products using SSM/I–SSMI/S sensors. **Journal of Hydrometeorology**, v. 14, p. 266–274, 2013. Disponível em:

<https://journals.ametsoc.org/doi/pdf/10.1175/JHM-D-12-056.1>. Acesso em: 15 dez. 2015. 10

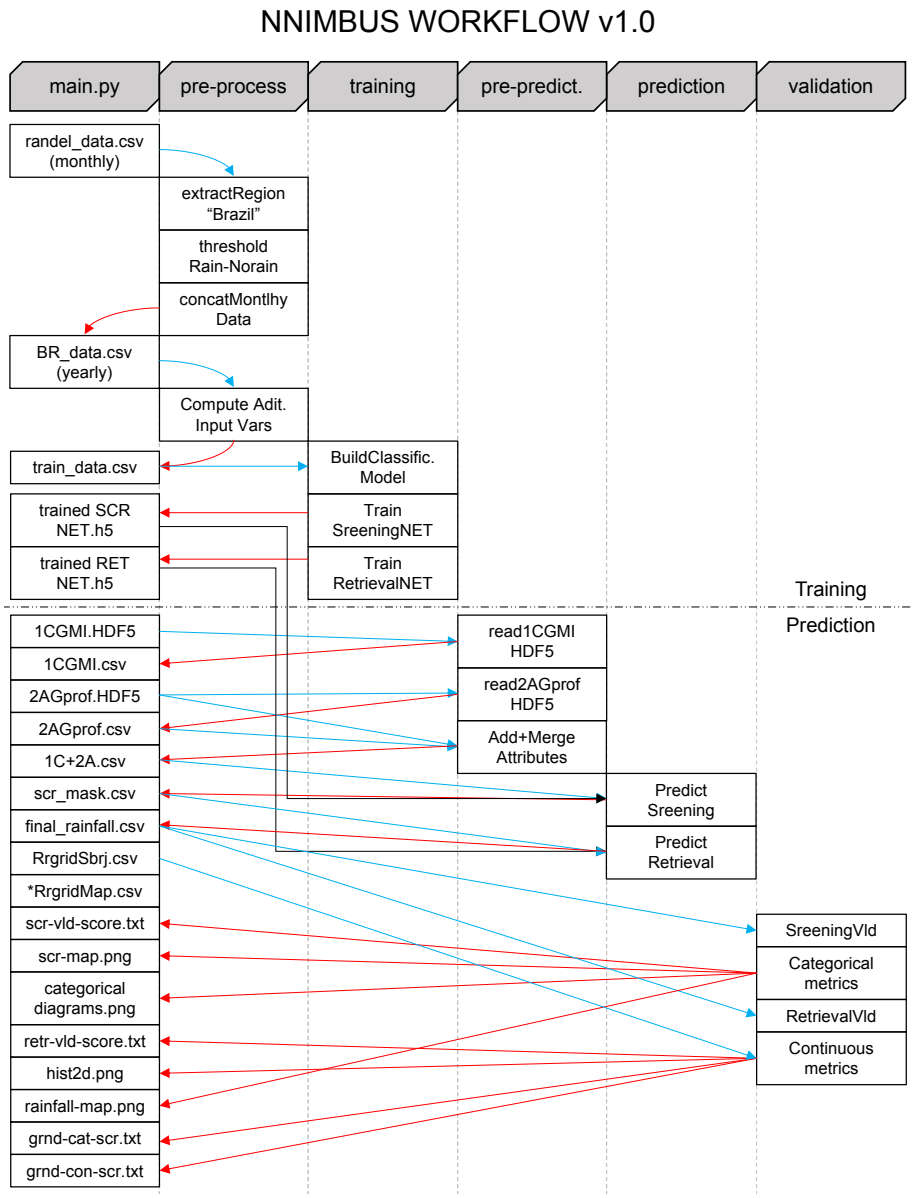
VILTARD, N.; BURLAUD, C.; KUMMEROW, C. D. Rain retrieval from TMI brightness temperature measurements using a TRMM PR-based database. **Journal of Applied Meteorology and Climatology**, v. 45, p. 455–466, 2006. Disponível em: <https://journals.ametsoc.org/doi/pdf/10.1175/JAM2346.1>. Acesso em: 27 fev. 2016. 11

VINCENT, J. F.; BOGATYREVA, O. A.; BOGATYREV, N. R.; BOWYER, A.; PAHL, A. Biomimetics: its practice and theory. **Journal of The Royal Society Interface**, v. 3. Disponível em: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2006.0127>. Acesso em: 19 fev. 2019. 17

APENDIX A - NNIMBUS ALGORITHM WORKFLOW

This sections presents a summarized execution sequence of the NNIMBUS pipeline from its beginning to end of execution, when it outputs the prediction and validation graphics. The blue lines represent input data whereas the red lines are the output.

Figure A.1 - Summarized execution sequence of the NNIMBUS pipeline.



SOURCE: Author's production

ANNEX A - MULTI-CATEGORICAL AND CONTINUOUS SCORES EQUATIONS

This section presents the equations used to compute both discrete and categorical data analysis and statistics.

Table A.1 - Confusion matrix for the base definition of the categorical score variables.

		Observation	
		Rain	No-rain
Prediction	Rain	True positives (TP)	False positives (FP)
	No-rain	False negatives (FN)	True negatives (TN)

Table A.2 - Reference table for the categorical score equations.

Name	**Equation**	**Range**	**Optimum**
Bias	**Bias** $= \frac{TP+FP}{TP+FN}$	0 to ∞	1
Probability of detection	**POD** $= \frac{TP}{TP+FN}$	0 to 1	1
Probability of false detection	**POFD** $= \frac{FP}{FP+TN}$	0 to 1	0
False alarm ratio	**FAR** $= \frac{FP}{TP+FP}$	0 to 1	0
Critical success index	**CSI** $= \frac{TP}{TP+FP+FN}$	0 to 1	1
Heidke skill score	**HSS** $= \frac{TP*TN-FP*FN}{[(TP+FN)*(FN+TN)+(TP+FP)*(FP+TN)]/2}$	-∞ to 1	1
Hansen-Kuipers discriminant	**HKD** $= \frac{TP}{(TP+FN)} - \frac{FP}{(FP+TN)}$	-1 to 1	1

Table A.3 - Reference table for the continuous statistical score equations.

Score	**Perfect Score**	**Calculation**
Mean error or bias (ME)	0	$ME = \frac{1}{N}\sum_{k=1}^{n}(sat_k - obs_k)$
Standard deviation (SD)	0	$SD = \sqrt{\frac{1}{N}\sum_{k=1}^{n}(sat_k - obs_k - ME)^2}$
Root mean square error (RMSE)	0	$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{n}(sat_k - obs_k)^2}$
Fractional standard error (FSE)	0	$FSE = \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{n}(sat_k - obs_k)^2}}{\frac{1}{N}\sum_{1}^{n} obs_k} = \frac{RMSE}{\frac{1}{N}\sum_{1}^{n} obs_k}$
Correlation coefficient (CC)	1	$CC = \frac{\sum_{k=1}^{n}(sat - \bar{sat})(obs - \bar{obs}))}{\sqrt{\sum_{k=1}^{n}(sat - \bar{sat})^2}\sqrt{\sum_{k=1}^{n}(obs - \bar{obs})^2}}$

ANNEX B - MANUSCRIPT PUBLICATIONS AND DERIVED WORK

This section describes all the material derived from the elaboration of this thesis.

Figure B.1 - First page of the manuscript published in Remote Sensing Journal, Volume 10, Issue 11 (November 2018).

remote sensing

MDPI

Article

Assessment of Ground-Reference Data and Validation of the H-SAF Precipitation Products in Brazil

Lia Martins Costa do Amaral [1,*], Stefano Barbieri [2], Daniel Vila [1], Silvia Puca [3], Gianfranco Vulpiani [3], Giulia Panegrossi [4], Thiago Biscaro [1], Paolo Sanò [4], Marco Petracca [3], Anna Cinzia Marra [4], Marielle Gosset [5] and Stefano Dietrich [4]

1 Weather Forecast Center and Climate Studies, National Institute for Space Research (CPTEC/INPE), São José dos Campos, SP 12227-010, Brazil; daniel.vila@inpe.br (D.V.); thiago.biscaro@inpe.br (T.B.)
2 CETEMPS, University of L'Aquila, 67100 L'Aquila, Italy; barbieris380@gmail.com
3 Italian Civil Protection Department, 00189 Rome, Italy; silvia.puca@protezionecivile.it (S.P.); gianfranco.vulpiani@protezionecivile.it (G.V.); Marco.Petracca@protezionecivile.it (M.P.)
4 Institute of Atmospheric Sciences and Climate (ISAC) National Research Council of Italy (CNR), 00133 Rome, Italy; giulia.panegrossi@artov.isac.cnr.it (G.P.); paolo.sano@artov.isac.cnr.it (P.S.); anna.cinzia.marra@gmail.com (A.C.M.); s.dietrich@isac.cnr.it (S.D.)
5 Institute of Research for Development (IRD), 13572 Marseille, France; marielle.gosset@ird.fr
* Correspondence: lia.meteorologia@gmail.com; Tel.: +55-12-3208-6645

Received: 13 August 2018; Accepted: 24 October 2018; Published: 5 November 2018

check for updates

Abstract: The uncertainties associated with rainfall estimates comprise various measurement scales: from rain gauges and ground-based radars to the satellite rainfall retrievals. The quality of satellite rainfall products has improved significantly in recent decades; however, such algorithms require validation studies using observational rainfall data. For this reason, this study aims to apply the H-SAF consolidated radar data processing to the X-band radar used in the CHUVA campaigns and apply the well established H-SAF validation procedure to these data and verify the quality of EUMETSAT H-SAF operational passive microwave precipitation products in two regions of Brazil (Vale do Paraíba and Manaus). These products are based on two rainfall retrieval algorithms: the physically based Bayesian Cloud Dynamics and Radiation Database (CDRD algorithm) for SSMI/S sensors and the Passive microwave Neural network Precipitation Retrieval algorithm (PNPR) for cross-track scanning radiometers (AMSU-A/AMSU-B/MHS sensors) and for the ATMS sensor. These algorithms, optimized for Europe, Africa and the Southern Atlantic region, provide estimates for the MSG full disk area. Firstly, the radar data was treated with an overall quality index which includes corrections for different error sources like ground clutter, range distance, rain-induced attenuation, among others. Different polarimetric and non-polarimetric QPE algorithms have been tested and the Vulpiani algorithm (hereafter, R_{q2Vul5}) presents the best precipitation retrievals when compared with independent rain gauges. Regarding the results from satellite-based algorithms, generally, all rainfall retrievals tend to detect a larger precipitation area than the ground-based radar and overestimate intense rain rates for the Manaus region. Such behavior is related to the fact that the environmental and meteorological conditions of the Amazon region are not well represented in the algorithms. Differently, for the Vale do Paraíba region, the precipitation patterns were well detected and the estimates are in accordance with the reference as indicated by the low mean bias values.

Keywords: rain gauges; radar; quality indexes; satellite rainfall retrievals; validation

1. Introduction

The knowledge about the distribution of water around the globe is an aspect of extreme relevance for the management of natural resources. The precipitation is, within the hydrological

SOURCE: Amaral et al. (2018)

Figure B.2 - First page of the manuscript published in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 10, Issue 9 (September 2017).

The Cloud Dynamics and Radiation Database Algorithm for AMSR2: Exploitation of the GPM Observational Dataset for Operational Applications

Daniele Casella, Lia Martins Costa do Amaral, Stefano Dietrich, Anna Cinzia Marra, Paolo Sanò, and Giulia Panegrossi

***Abstract*—A new precipitation retrieval algorithm for the AMSR2 is described. The algorithm is based on the cloud dynamics and radiation database (CDRD) Bayesian approach and represents an evolution of the previous version applied to SSMIS observations, and used operationally within the EUMETSAT H-SAF program. This new product presents as main innovation the use of a very large database entirely empirical, derived from coincident radar and radiometer observations from the NASA/JAXA GPM-CO launched on February 28, 2014. The other new aspects are: 1) a new rain-/no-rain screening approach; 2) use of EOF and CCA for dimensionality reduction; 3) use of new ancillary variables to categorize the database and mitigate the problem of non-uniqueness of the retrieval solution; and 4) development and implementations of modules for computation time minimization. A verification study for case studies over Italy and for coincident AMSR2/GPM-CO observations over the MSG full disk area has been carried out. Results show remarkable AMSR2 capabilities for RR retrieval over ocean (for RR $>$ 0.1 mm/h), good capabilities over vegetated land (for RR $>$ 1 mm/h), while for coastal areas the results are less certain. Comparisons with NASA GPM products, and with ground-based radar data, show that the new CDRD for AMSR2 is able to depict very well the areas of high precipitation over all surface types. The algorithm is also able to handle an extremely large observational database available from GPM-CO and to provide rainfall estimate with minimum latency, making it suitable for NRT hydrological and operational applications.**

***Index Terms*—Advanced Microwave Scanning Radiometer 2 (AMSR2), Bayesian retrieval algorithm, Cloud Dynamics and Radiation Database (CDRD), Global Precipitation Measurement (GPM) mission, passive microwave (PMW) radiometer, satellite precipitation.**

Manuscript received December 15, 2016; revised April 18, 2017; accepted May 24, 2017. This work was supported in part by the EUMETSAT through the project "Satellite Application Facility on Support to Operational Hydrology and Water Management" (H-SAF), and in part by the EarthH2Observe "Global Earth Observation for Integrated Water Resource Assessment" project funded under the EU DG Research 7th Framework Programme. The work of L. M. C. Amaral was supported in part by the National Council for Scientific and Technological Development, in part by the Coordination for the Improvement of Higher Education Personnel, and in part by the Ministry of Science, Technology and Innovation, Brazil, during her Ph.D. studies and in part by the Institutional Program of Overseas Sandwich Doctorate (SWE) (process number 201237/2015-6) for the internship opportunity at Institute of Atmospheric Sciences and Climate/National Research Council. *(Corresponding author: Giulia Panegrossi.)*

D. Casella was with the Institute of Atmospheric Sciences and Climate – National Research Council, Rome 00133, Italy. He is now with SERCO S.p.A., Frascati 00044, Italy (e-mail: daniele.casella@artov.isac.cnr.it).

L. M. C. Amaral is with the Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP 12227-010, Brazil, and also with the Institute of Atmospheric Sciences and Climate – National Research Council, Rome 00133, Italy (e-mail: lia.amaral@artov.isac.cnr.it).

S. Dietrich, A. C. Marra, P. Sanò, and G. Panegrossi are with the Institute of Atmospheric Sciences and Climate – National Research Council, Rome 00133, Italy (e-mail: stefano.dietrich@artov.isac.cnr.it; a.marra@isac.cnr.it; paolo.sano@artov.isac.cnr.it; g.panegrossi@isac.cnr.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2017.2713485

List of Acronyms

AMSR2	Advanced Microwave Scanning Radiometer 2.
AMSU	Advanced Microwave Sounding Unit.
CAL/VAL	CALibration/VALidation.
CC	Correlation Coefficient.
CCA	Canonical Correlation Analysis.
CDRD	Cloud Dynamics and Radiation Database.
CRM	Cloud Resolving Model.
CPR	Cloud Profiling Radar.
CV	Canonical Variable.
DMSP	Defense Meteorological Satellite Program.
DPR	Dual-frequency Precipitation Radar.
DTH	Dynamical-Thermodynamical-Hydrological.
ECMWF	European Centre for Medium Range Weather Forecasts.
EOF	Empirical Orthogonal Function.
EPS-SG	EUMETSAT Polar System-Second Generation.
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites.
FAR	False Alarm Ratio.
FSE	Fractional Standard Error.
AFSE	Adjusted Fractional Standard Error.
GANAL	JMA Global ANALysis.
GCOM-W1	Global Change Observation Mission-Water 1.
GMI	GPM Microwave Imager.
GPM	Global Precipitation Measurement.
GPM-CO	Global Precipitation Measurement-Core Observatory.
GPROF	Goddard PROFiling algorithm.
H-SAF	Satellite Application Facility on Support to Hydrology and Water Management.
HSS	Heidke Skill Score.
HTS	High-Temperature noise Source.

SOURCE: Casella et al. (2017)

Table B.1 - Posters and abstracts submitted and presented during the PhD.

- Poster “The new Cloud Dynamics and Radiation Database algorithms for AMSR2 and GMI: exploitation of the GPM observational database for operational applications.”, presented at European Geophysical Union - EGU, 2017, Vienna, Austria.
- Poster “The new Bayesian algorithms for AMSR2 and GMI for operational applications in the EUMETSAT H-SAF”, EUMETSAT - Meteorological Satellite Conference, 2017, Rome, Italy.
- Poster “The Passive Microwave Neural Network Precipitation Retrieval (PNPR) for the Conical Scanning GMI”, EUMETSAT - Meteorological Satellite Conference, 2017, Rome, Italy.
- Poster “Investigation on CDRD and GPROF performance over central Amazon region during GoAmazon/CHUVA campaigns” presented at 8o International Precipitation Working Group (IPWG) and 5o International Workshop on Space-based Snowfall Measurement (IWSSM) Joint Workshop, 2016, Bologna, Italy.

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.