MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2024/06.24.16.52-TDI

HAILSTORM DETECTION USING GOES-16 SATELLITE DATA

Caio Átila Pereira Sena

Doctorate Thesis of the Graduate Course in Meteorology, guided by Dr. Renato Galante Negri, approved in June 14, 2024.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/4BH538S>

INPE
São José dos Campos
2024

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2024/06.24.16.52-TDI

HAILSTORM DETECTION USING GOES-16 SATELLITE DATA

Caio Átila Pereira Sena

Doctorate Thesis of the Graduate Course in Meteorology, guided by Dr. Renato Galante Negri, approved in June 14, 2024.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/4BH538S>

INPE
São José dos Campos
2024

Cataloging in Publication Data

Sena, Caio Átila Pereira.

Se55h Hailstorm detection using GOES-16 satellite data / Caio Átila Pereira Sena. – São José dos Campos : INPE, 2024.

xxvi + 142 p. ; (sid.inpe.br/mtc-m21d/2024/06.24.16.52-TDI)

Thesis (Doctorate in Meteorology) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2024.

Guiding : Dr. Renato Galante Negri.

1. Hailstorms. 2. Advanced baseline imager. 3. Machine learning. 4. Gradient boosted trees. 5. SHAP analysis. I.Título.

CDU 551.578.7

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DEFESA FINAL DE TESE DE CAIO ÁTILA PEREIRA SENA
REG.218793/2020, BANCA N°142/2024

No dia 14 de junho de 2024, em teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi **APROVADO(A)** pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutor em Meteorologia, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

Título: **" Hailstorm detection using GOES-16 satellite data ".**

Membros da banca:

Dr. Thiago Souza Biscaro – Presidente – INPE

Dr. Renato Galante Negri – Orientador – INPE

Dra. Izabelly Carvalho da Costa – Membro Interno – INPE

Dr. Fabrício Firmino de Faria – Membro Externo – UFRJ

Dr. Weber Andrade Gonçalves – Membro Externo – UFRN

Documento assinado eletronicamente por **Thiago Souza Biscaro**, **Pesquisador**, em 19/06/2024, às 19:14 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Izabelly Carvalho da Costa**, **Chefe da Divisão de Previsão de Tempo e Clima**, em 20/06/2024, às 09:10 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Renato Galante Negri**, **Chefe da Divisão de Satélites e Sensores Meteorológicos**, em 20/06/2024, às 11:55 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fabrício Firmino de Faria (E)**, **Usuário Externo**, em 28/06/2024, às 12:04 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Weber Andrade Gonçalves (E)**, **Usuário Externo**, em 04/07/2024, às 11:50 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.mcti.gov.br/verifica.html, informando o código verificador **12032800** e o código CRC **6FAED4C6**.

Referência: Processo nº 01340.004978/2024-38 SEI nº 12032800

"What I see in Nature is a magnificent structure that we can comprehend only very imperfectly, and that must fill a thinking person with a feeling of humility".

ALBERT EINSTEIN in
"Albert Einstein, the Human Side", 2014

A meus pais ***Márcia*** *e* ***Romário****, a minha irmã* ***Priscila*** *e a minha companheira* ***Lívia***

ACKNOWLEDGEMENTS

We acknowledge the contribution to the present study by Dr. Bruno Zanetti Ribeiro and the PRevots group that kindly provided the severe weather reports maintained and verified by them.

This study was financed in by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [Projeto CNPq 142291/2020-0].

ABSTRACT

Hailstorm detection is crucial for research on severe weather events and mitigating their adverse impacts. This thesis focuses on detecting convective systems with the potential to cause hailstorms in Brazil using Brightness Temperature (BT) measurements from the GOES-16 Advanced Baseline Imager (ABI). We tracked convective systems over South America from January 2018 to December 2023 using the TATHU library, creating an extensive Spatialite database of storm polygons. These polygons were matched with verified hail reports from the PRevots group to build a training set for assessing hailstorm risk. Our methodology includes the application of several machine learning techniques for classifying regions in satellite scenes likely associated with hailstorms. Our main focus was to build a lean model to preserve the interpretability of the classification problem. Therefore, our first experiment involved pixel-wise classification, where we gathered BT measurements that matched hail reports using ABI's channels in the Earth's emission region. This approach encountered challenges such as sampling bias and overfitting, but a feature importance analysis provided insights into avoiding these issues. To address these problems, we incorporated the information embedded in the TATHU dataset to detect storm polygons likely associated with hailstorms. Specifically, we used an XGBoost model in this experiment. The results indicated good performance in the southern region of Brazil but faced challenges north of approximately 20°S latitude. A SHAP (SHapley Additive exPlanations) analysis was conducted on the XGBoost model, providing valuable insights into the dynamics and spatial distribution of storm-related characteristics over the study region. The findings of this research highlight advancements in timely hailstorm detection over Brazil using measurements from geostationary satellites. Future work will focus on further enhancing the classification with more complex machine learning models, incorporating additional relevant features, and expanding the dataset to ensure robust and accurate detection across diverse regions.

Keywords: Hailstorms. Advanced Baseline Imager. Machine Learning. Gradient Boosted Trees. SHAP analysis.

DETECÇÃO DE TEMPESTADES DE GRANIZO USANDO DADOS DO SATÉLITE GOES-16

RESUMO

A detecção de tempestades de granizo é crucial para a pesquisa sobre eventos de tempo severo e para mitigar seus impactos adversos. A presente tese tem como principal objetivo a detecção de sistemas convectivos com potencial para ocasionar tempestades de granizo sobre o Brasil, utilizando medições de Temperatura de Brilho (TB) do Advanced Baseline Imager (ABI) à bordo do satélite GOES-16. Para isso, foram rastreados sistemas convectivos na América do Sul entre janeiro de 2018 e dezembro de 2023, usando a biblioteca TATHU, resultando em um banco de dados extenso em formato Spatialite com os polígonos dos sistemas. Esses polígonos foram comparados com relatos verificados de granizo do grupo PRevots com a finalidade de construir um conjunto de treinamento para avaliar o risco de granizo. Nesse sentido, a metodologia inclui a aplicação de várias técnicas de aprendizado de máquina para classificar sub-regiões em dados de satélite que provavelmente estão associadas a tempestades de granizo. O foco foi construir um modelo enxuto a fim de preservar a interpretabilidade do problema de classificação. Assim, o primeiro experimento envolveu a classificação pixel-a-pixel, em que foram consideradas medições de TB na localização correspondente aos relatos de granizo, usando para isso, os canais do ABI no espectro de emissão Terrestre. Essa abordagem encontrou desafios como viés de amostragem e "sobreajuste"do modelo, mas uma análise de importância dos atributos forneceu ideias para evitar esses problemas. De forma a resolver essas questões, incorporamos as informações embutidas no banco de dados TATHU para detectar polígonos de tempestades provavelmente associados a tempestades de granizo. Especificamente, usamos um modelo XGBoost neste experimento. Os resultados indicaram um bom desempenho na região sul do Brasil, mas enfrentaram desafios para regiões a norte de aproximadamente 20°S de latitude. Uma análise SHAP (SHapley Additive exPlanations) foi conduzida no modelo XGBoost, fornecendo insights valiosos sobre a dinâmica e a distribuição espacial das características relacionadas às tempestades na região de estudo. Os resultados desta pesquisa destacam avanços na detecção de tempestades de granizo no Brasil, utilizando dados de satélites geoestacionários. Trabalhos futuros se concentrarão em aprimorar a classificação com modelos de aprendizado de máquina mais complexos, incorporando medidas adicionais relevantes e expandindo o conjunto de dados para garantir uma detecção robusta e precisa em diversas regiões.

Palavras-chave: Tempestades de Granizo, Advanced Baseline Imager, Aprendizado de Máquina, Gradient Boosted Trees, Análise SHAP

LIST OF FIGURES

Page

1.1 Natural fiber acting as Ice Nucleus. . . . 3
1.2 Ice nucleation mechanisms. . . . 4
1.3 Re-circulation of embryo inside the storm. . . . 6
1.4 Hailstone aspects of undergoing sequences of dry and wet growth. . . . 7
1.5 Hailfall cases in Argentina and Uruguay. . . . 9
1.6 Seasonal mean of hail events in Argentina. . . . 10
1.7 Destructive hail events for the southern and southeastern Brazilian States. 12
1.8 Parallax effect on GOES-16 ABI measurements. . . . 17

2.1 Interaction of particles with electric field. . . . 19
2.2 Microwave scattering regimes. . . . 21
2.3 Impact of hydrometeors on the *Differential Reflectivity* Z_{DR}. . . . 27
2.4 Electric field interference due to polarized incident waves. . . . 28
2.5 Water cloud versus ice cloud reflectance spectra. . . . 33
2.6 Example of CTD and "CTV couplets". . . . 35

3.1 Spatial domain of study. . . . 39
3.2 Schematic view of GOES-R Series platforms with its payloads. . . . 41
3.3 General Stages of TATHU's methods for detection, characterization and tracking of targets. . . . 43
3.4 Schema of CS's possible tracked events. . . . 44
3.5 Full learning setup. . . . 48
3.6 Examples of different polynomial approximations to a function. . . . 52
3.7 The impact of increasing the size of the set of observations. . . . 53
3.8 Roadmap for choosing the Machine Learning algorithm. . . . 54
3.9 The K-Nearest Neighbour Approach. . . . 57
3.10 KNN Decision Boundary Dependency. . . . 58
3.11 Decision Tree Scheme. . . . 60
3.12 Gradient Boosted Trees Example. . . . 62
3.13 Support Vector Machine boundary. . . . 63
3.14 SVM non-linear boundary. . . . 66
3.15 K-fold cross validation. . . . 70
3.16 Example of Principal Components. . . . 75

4.1 Overview of the BT distribution in the training set. . . . 79
4.2 Comparison of Experiment 1 models. . . . 81

4.3 Models time performance. 82
4.4 Indices of refraction of water and ice. 84
4.5 Clear-sky weighting function of Channel 12 of GOES-ABI. 85
4.6 Impact of view-angle on ABI-bands. 87
4.7 Example cases of validation of the pixel-wise approach, for visual inspection. 89
4.8 Input data projected into PC space. 92
4.9 Flowchart of methods for defining positive examples of hailstorms. 96
4.10 Flowchart of methods for defining negative examples of hailstorms. . . . 98
4.11 Detection and classification examples. 104
4.12 General Feature Importance. 106
4.13 Partial Dependence Plots. 108
4.14 Dimensionality reduction of SHAP values. 110
4.15 Example of a day of strong convective activity. 112

LIST OF TABLES

Page

2.1 Typical reflectivity factors for the corresponding atmospheric phenomena. 26

3.1 ABI Spectral Attributes. 42
3.2 Parameters set used for storm tracking using TATHU. 45

4.1 Classification scores in the validation set. 83
4.2 Importance of ABI's Channels for the classification. 83
4.3 Explained Variance Ratio of Principal Components. 91
4.4 Hyperparameters Tuned for the XGBoost Model. 100
4.5 Best Hyperparameters for the XGBoost Model. 101
4.6 XGB classification scores in the validation set (2023). 102

LIST OF ABBREVIATIONS

AACP	–	Above Anvil Cirrus Plume
ABI	–	Advanced Baseline Imager
AHI	–	Advanced Himawari Imager
AMS	–	American Meteorological Society
AMSR	–	Advanced Microwave Scanning Radiometer
AMSU	–	Advanced Microwave Sounding Unit
AMV	–	Atmospheric Motion Vector
ATMS	–	Advanced Technology Microwave Sounder
BSS	–	Briar Skill Score
BT	–	Brightness Temperature
CAPE	–	Convective Available Potential Energy
CCN	–	Cloud Condensation Nuclei
CM	–	Convective Mask
CMA	–	China Meteorological Administration
CMIPF	–	Cloud and Moisture Imagery Full Disk
CNN	–	Convolutional Neural Network
CS	–	Convective System
CTD	–	Cloud Top horizontal Divergence
CTV	–	Cloud Top Vorticity
DISSM	–	Divisão de Satélites e Sensores Meteorológicos
DJF	–	December, January, February
DMSP	–	Defense Meteorological Satellite Program
EOS	–	Earth Observing System
ESWD	–	European Severe Weather Database
EUR	–	Euro
FAR	–	False Alarm Ratio
FED	–	Flash Extent Density
ForTrACC	–	Forecast and Tracking the Evolution of Cloud Clusters
FOV	–	Field Of View
GD	–	Gradient Descent
GLM	–	Geostationary Lightning Mapper
GOES	–	Geostationary Operational Environmental Satellite
GPM	–	Global Precipitation Measurement Mission
HCA	–	Hydrometheor Classification Algorithm
HDT	–	Hail Detection Tool
HM	–	Hail Mask
IBGE	–	Brazilian Institute of Geography and Statistics
IN	–	Ice Nucleus
INPE	–	Instituto Nacional de Pesquisas Espaciais
IR	–	Infrared
JJA	–	June, July, August
KDP	–	Specific Differential Propagation Phase
KNN	–	K-Nearest Neighbor

LDR	–	Linear Depolarization Ratio
MAM	–	March, April, May
MCC	–	Mesoscale Convective Complex
MDS	–	Multidimensional Scaling
MFA	–	Minimum Flash Area
ML	–	Machine Learning
MSG	–	Meteosat Third Generation
MESH	–	Maximum Expected Hail Size
MHS	–	Microwave Humidity Sounder
MWCC	–	Microwave Cloud Classification Algorithm
NCEP	–	National Center for Environmental Prediction
NEXRAD	–	Next Generation Weather Radar
NIR	–	Near Infrared
NOAA	–	National Oceanic and Atmospheric Administration
OT	–	Overshooting Top
PC	–	Principal Component
PCA	–	Principal Component Analysis
PDP	–	Parallel Distributed Processing
POD	–	Probability of Detection
POFD	–	Probability of False Detection
RAP	–	Rapid Refresh Model
ReLU	–	Rectified Linear Unit
RGB	–	Red, Green, Blue
SALLJ	–	South American Low-Level Jet
SMMR	–	Scanning Multichannel Microwave Radiometer
S-NPP	–	Suomi National Polar Partnership
SON	–	September, October, November
SRSOR	–	Super Rapid Scan Operations for GOES-R
SSMIS	–	Special Sensor Microwave Imager / Sounder
SVD	–	Singular Value Decomposition
SWIR	–	Shortwave Infrared
TATHU	–	Tracking and Analysis of Thunderstorms
TIR	–	Thermal Infrared
TMI	–	TRMM Microwave Imager
TOE	–	Total Optical Energy
TRMM	–	Tropical Rainfall Measuring Mission
US	–	United States
UTC	–	Coordinated Universal Time
VIL	–	Vertically Integrated Liquid Water Content
VIS	–	Visible (radiation spectrum)
WV	–	Water Vapor

LIST OF SYMBOLS

r	–	Particle Radius
$\vec{E}$	–	Electric field
K_{DP}	–	Specific differential phase
Z	–	Reflectivity
Z_e	–	Effective (beam-averaged) reflectivity
Z_{DR}	–	Differential reflectivity
χ	–	Particle size - wavelength relationship
λ	–	Wavelength
r_e	–	Effective radio
T_p	–	True positive
F_p	–	False positive
T_n	–	True negative
F_n	–	False negative
o_t	–	Outcome
N	–	Sample size

CONTENTS

Page

1 INTRODUCTION 1
1.1 Background and motivation 1
1.1.1 Hail formation 2
1.1.2 Hailstorms in South America 8
1.1.3 Hailstorms and lightning activity 12
1.2 Research questions and objectives 13
1.3 Significance of the study 14
1.4 Scope and limitations 15
1.5 Thesis outline 17

2 LITERATURE REVIEW 19
2.1 Key theories and models 19
2.1.1 Hail microwave scattering 19
2.1.1.1 Passive microwave remote sensing 21
2.1.1.1.1 - Algorithms 23
2.1.1.2 Active microwave remote sensing 25
2.1.1.2.1 - Dual-polarization radar 26
2.1.1.2.2 - Algorithms 30
2.1.1.3 Limitations 31
2.1.2 Hailstorms through visible and infrared wavelengths 32
2.1.2.1 Algorithms 36
2.1.2.2 Limitations 36
2.2 Gaps in existing knowledge 37

3 METHODOLOGY 39
3.1 Study area 39
3.2 Data collection 40
3.2.1 Hail reports 40
3.2.2 GOES-16 satellite 41
3.2.2.1 Advanced Baseline Imager 42
3.2.3 Database of convective storms 43
3.3 Machine learning approaches 46
3.3.1 Theoretical basis for machine learning applications 46

3.3.2 Algorithm selection and rationale . . . 50
3.3.2.1 Linear model for classification . . . 55
3.3.2.2 K-Nearest Neighbors classifier . . . 56
3.3.3 Gaussian Naive Bayes method . . . 58
3.3.3.1 Tree-based methods . . . 59
3.3.3.1.1 - Ensemble methods . . . 61
3.3.3.2 Support Vector Machines . . . 63
3.3.3.2.1 - Stochastic Gradient Descent for linear SVM . . . 64
3.3.3.2.2 - Kernel trick . . . 65
3.3.4 Optimal parameters selection . . . 66
3.3.4.1 Combinatorics . . . 67
3.3.4.2 Bayesian optimizer . . . 67
3.3.5 Model training and testing . . . 68
3.3.5.1 Resampling methods . . . 69
3.3.5.2 Evaluation metrics . . . 71
3.4 Model exploratory analysis . . . 72
3.4.1 Shapley values analysis . . . 72
3.4.2 Principal Component Analysis . . . 73

4 EXPERIMENTS AND RESULTS . . . 77
4.1 Overview of experiments . . . 77
4.2 Experiment 1: pixel-wise approach . . . 77
4.2.1 Methods . . . 77
4.2.2 Results . . . 82
4.2.2.1 The curse of dimensionality . . . 90
4.2.3 Challenges and lessons learned . . . 93
4.3 Experiment 2: classification of storm polygons . . . 94
4.3.1 Hailstorm dataset construction . . . 94
4.3.2 Class imbalance . . . 99
4.3.3 Model settings and training . . . 99
4.3.4 Model performance . . . 101
4.3.5 SHAP analysis and Dependence Plots . . . 105
4.3.6 Implications of hailstorm detection . . . 110
4.3.7 Challenges and lessons learned . . . 111

5 DISCUSSION . . . 115
5.1 Comparative analysis of experiments . . . 115
5.2 Comparison with previous studies . . . 115

5.3 Theoretical and practical implications 116
5.4 Limitations and challenges 117

6 CONCLUSION AND FUTURE WORK 119
6.1 Recommendations for future research 119

REFERENCES 123

1 INTRODUCTION

1.1 Background and motivation

Hailstones, characterized as roughly spherical ice bodies, originate from convective clouds and demonstrate considerable variability in shape, size and composition (KNIGHT et al., 2015). Typically, hail is defined as having 5 mm in diameter or in its longest dimension (AMERICAN METEOROLOGICAL SOCIETY, 2024), except in China. There, hail is defined as precipitation in the form of balls or irregular lumps of ice with a diameter larger than 2 mm (CHINA METEOROLOGICAL ADMINISTRATION, 2003 apud ALLEN et al., 2020). Then, hail is distinguished from sleet and snow by hardness, shape and fragility. Notably, hailstones can exceed 15 cm in diameter and reach terminal velocities of approximately 50 $m.s^{-1}$ (HOUZE, 2014), contributing to their potential for significant destruction upon impact.

Financially, hailstorms represent a considerable economic burden, accounting for approximately 56% of property and agricultural damage attributed to severe weather events in the United States (MURILLO, 2018), culminating in losses totaling 16 billion USD between 2013 and 2017 (GUNTURI; TIPPETT, 2017). In Europe, a notable event was the hailstorm associated with the low pressure system named Ela between June 8 and 10, 2014, which resulted in insured damages amounting to 2.3 billion EUR across France, Belgium, and Germany (PUNGE; KUNZ, 2016).

Hailfall is spatially a small-scale phenomenon, with an impact within tens of square kilometers (KNIGHT et al., 2015). Therefore, the observed data available on hailstorms and hail size usually originate from eyewitnesses reports (TUOVINEN et al., 2009) and civil protection databases (MARTINS et al., 2017). As a consequence, over regions poorly resolved by hail pads networks, the records on frequency of events are biased towards largely populated areas and the stone size data is affected by the metric adopted by the observer, that generally uses an everyday object as a reference (MURILLO; HOMEYER, 2019). Therefore, the lack of standards on observational data of hailstorms has been, for long, a worldwide issue (PROHASKA, 1902; ALLEN et al., 2020), and better spatially distributed consistent data on hail occurrence can be obtained from remote sensing instruments (PUNGE; KUNZ, 2016).

Along with its own potential damage, the hailstone precipitation is usually coupled to very energetic events. Hail occurrence is routinely related to supercells or multicell storms (BROWNING, 1977): favored by mesocyclones (FOOTE, 1984), associated with thunderstorms (KNIGHT, 1986) and storms formed under influence

of orography over long mountain ranges (KUNZ; PUSKEILER, 2010; NISI et al., 2016). For those reasons, the prediction and detection of such extreme weather is a major concern for risk management (PUSKEILER et al., 2016) and insurance pricing (BROWN et al., 2015). Hail forecast is also a matter of interest to agriculture, due to the need of protective measures prior to the hailfall (MELNYK et al., 2020; SCHMIDT et al., 2019).

Since before the 18th century there has been an effort of the scientific community to understand hail in the Earth system (PRUPPACHER; KLETT, 2012). Besides its long-established research, most of the progress was made after the World War II, once radar technology and other instruments and techniques became available for atmospheric assessments (GALVIN, 2020). Since then, new technology has been developed for this kind of investigation and algorithmic advances empower attempts to better leverage remotely sensed data. Through the next sections the current understanding of hail formation and its interaction with electromagnetic radiation is described and the state of art research in hail measurements through remote sensing will be presented.

1.1.1 Hail formation

The process of hail formation begins with an embryo, which is a small particle ($<$ 1 cm) that commonly takes the form of graupel or frozen drops (FEDERER et al., 1982; HEYMSFIELD, 1982). Less frequently natural fibers of flora may work as an embryo as well (Figure 1.1) (SAMENOW, 2015).

Figure 1.1 - Natural fiber acting as Ice Nucleus.

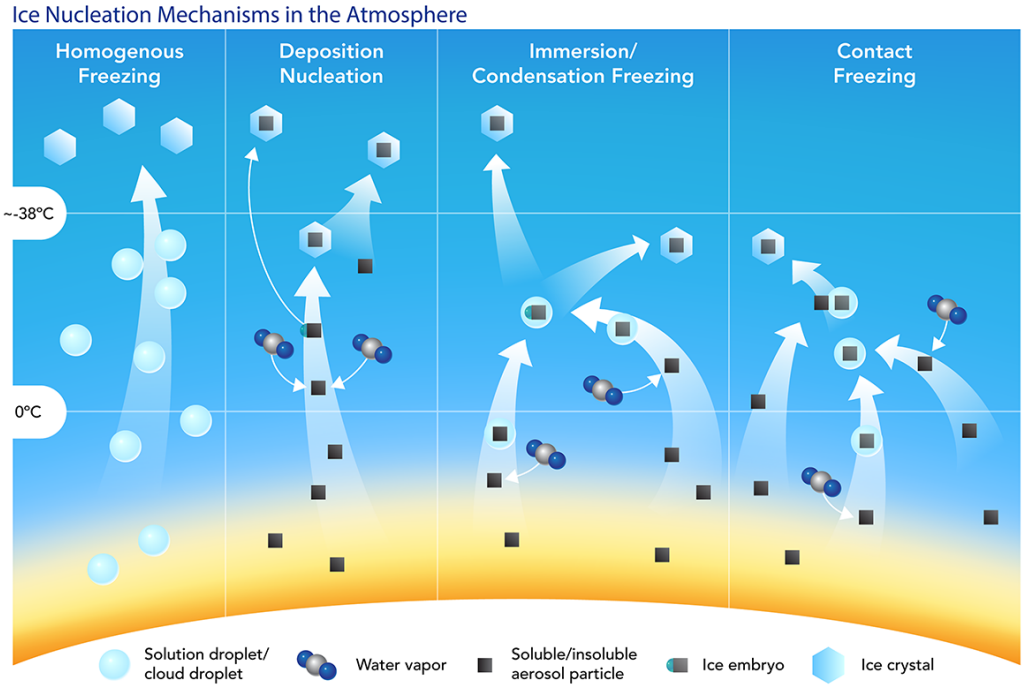
Stalk of corn encased in ice from a storm in Pampa, Texas. On November 16, 2015.
SOURCE: Samenow (2015).

The embryo, itself, is formed during ascent in an updraft current either on a nucleation active entity, i.e. an Ice Nucleus (IN) (VALI et al., 2015) or by homogeneous freezing of cloud droplets. As a general reference, the IN is the agent of heterogeneous nucleation, i.e. the process in which water molecules change to a more organized state in the presence of a foreign substance, such as aerosols and chemical compounds (e.g. silver iodide (DESSENS et al., 2016)). This type of nucleation is mainly contained between the 273.15 K and 235.15 K levels in the Atmosphere (NAGARE et al., 2016). In contrast, the homogeneous nucleation takes place when cloud droplets are lifted to a region of temperature below 235.15 K and freeze. Those nucleation mechanisms are illustrated in Figure 1.2.

The embryo formation and the hail growth from the embryo are processes involving distinct ranges of particle sizes and mass. Therefore, those events must occur in different parts of the storm, under specific updraft speeds. The embryos may be formed in "feeder cells" (DENNIS et al., 1970; HEYMSFIELD; MUSIL, 1982) or grow entirely within the main updraft. A feeder cell transfers hail embryos into the main cell for further hail growth, namely a particle injection mechanism (DENNIS; KUMJIAN, 2017). Feeder cells are typically characterized by weaker updrafts located along the side of a main cell (DENNIS et al., 1970), playing a crucial role in the transportation of pre-hail elements into the main updraft area (HEYMSFIELD

et al., 1980), by wraping around to the rear of the storm like an "embryo curtain" (BROWNING; FOOTE, 1976). Feeder cells can evolve as "daughter" cells where precipitation debris from decaying Cumulus Congestus clouds (e.g. the forward portion of the storm) enters the "feeder-cell" and can continue their growth, further contributing to the hailstorm's intensity and longevity (HEYMSFIELD et al., 1980; GUO; HUANG, 2002). When the embryo's growth takes place in the updraft, it must occur preferably in the periphery of the main current, otherwise due to its inherently small terminal velocity, if it crosses zones with large updraft speed it would be quickly ejected into the anvil and would not participate in the hail growth process.

Figure 1.2 - Ice nucleation mechanisms.



From left to right: cloud droplets lifted above the -38 °C level leading to homogeneous freezing; water vapor deposition on the surface of aerosol particles; water vapor condensation on the surface of a ice nucleating particle and further freezing; droplet freezing after colision with an ice nucleating particle.

SOURCE: Creative Commons Licensed (2014).

After the embryo formation, the growth mainly happens by the collection of supercooled liquid water at temperatures between about 263 K and 248 K (NELSON, 1983; ZIEGLER et al., 1983; KNIGHT et al., 1975) in a mechanism of *accretion*, notwithstanding that most of the supercooled liquid droplets remain present until 235.65 K (ROSENFELD; WOODLEY, 2000). The droplets may freeze immediately when in contact with the embryo, a process named "riming". Conversely, the freezing may be delayed and a liquid shell might remain around the embryo for a while after the contact. For further details on the energy balance at the hailstone's surface during growth, we refer the reader to (PRUPPACHER; KLETT, 2010).

Hailstones may also grow via homogeneous nucleation when the ambient vapor pressure in the cell's mixed-phase region falls between the saturation vapor pressure over water and the saturation vapor pressure over ice. This subsaturated environment for liquid water and supersaturated to ice lead to a ice crystal growth by water vapor deposition at the expense of the supercooled droplets (KOOP, 2004). This is called the Wegener-Bergeron-Findeisen process. It is important to notice that INs are very rare in comparison to the availability of cloud condensation nuclei (CCNs) in the Atmosphere (FINDEISEN, 1938). Hence, the presence of ice crystals in a cloud does not imply that it is in enough concentration for their growth to deplete supersaturation faster than the rate that supersaturation is generated through adiabatic cooling (STORELVMO; TAN, 2015).

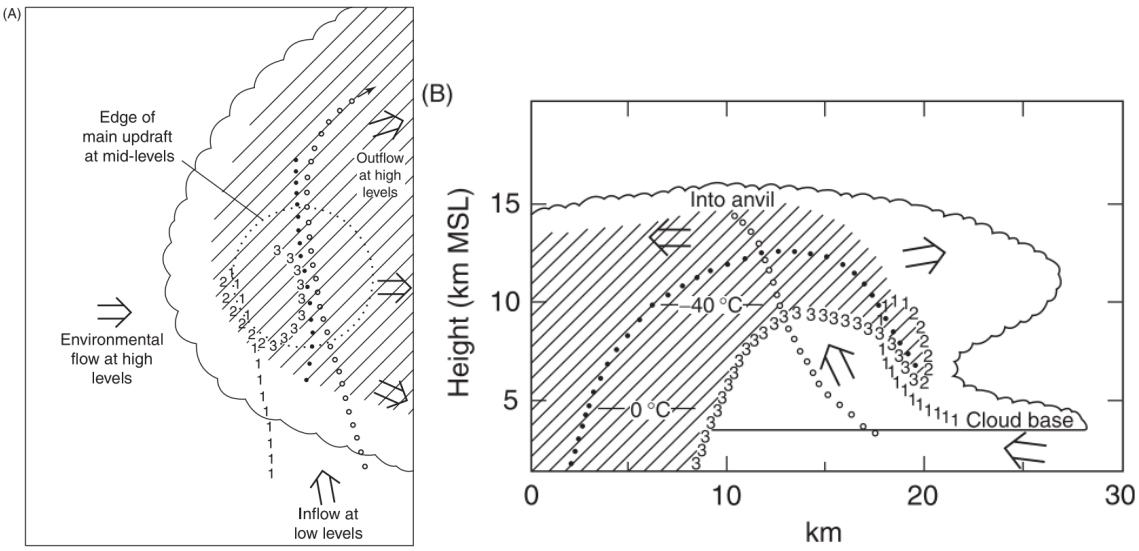
The coagulation of hailstones is not expected to occur in storms, nor is it typically observed or treated in numerical models of hailstone growth processes (ALLEN et al., 2020).

The growth of large hail is mainly supported by the time of residence of the hailstones inside the updraft current (MILLER et al., 1990) and re-circulation inside the storm (shown in Figure 1.3) by repeatedly crossing updrafts stronger than its terminal velocity, which gives it plenty of opportunities to collect super-cooled water droplets. Nelson (1983) suggests that large hailstones spend 10-15 minutes inside growth regions of storms. Since the hailstone grows continuously while collecting supercooled liquid water, it must be advected into regions of stronger updraft in order to keep suspended. Then, there must be a horizontal flux from the updraft periphery into its core.

A possible trajectory for large hail formation consists of a lifting of the embryo in the border of the core updraft (stage 1); subsequent re-insertion during it's fall trajectory, as a consequence of storm motion (stage 2); and continuous growth in

the core updraft until a final fall when the updraft current no longer overcome it's fall speed (stage 3). Embryos that are directly inserted in the main updraft will not have much time to grow as will most likely be launched quickly into the high-level outflow. An unfavourable trajectory for hail growth is the one taken by embryos that shortly enter the main updraft, which end up having a short opportunity to grow as well. This way, a broader updraft region favours the development of larger hail by providing a larger embryo injection region (NELSON, 1983) and enabling a bigger volume through which trajectories may pass (DENNIS; KUMJIAN, 2017).

Figure 1.3 - Re-circulation of embryo inside the storm.



(a) Plan view of a supercell and (B) vertical section along the storm's trajectory. The dotted circle limits the main updraft. Trajectories 1, 2 and 3 represent three stages in the growth of large hailstones. The trajectory maked by "∘" represents a complete embryo growth inside the updraught core. The dotted path illustrates the course of embryos that shortly cross the main updraft.

SOURCE: Foote and Knight (1977).

Under a completely "dry growth", there is immediate freezing happening on the outer shell of the hailstone, allowing air bubbles to be trapped inside its cavities, leading a more opaque hailstone. In contrast, under a "wet growth" when a thin layer of liquid delays to freeze outside the hailstone, the water fills the cavities before changing phase, consequently forming a more transparent hail. In a storm, its common that both processes take place leaving the hailstone with distinct layered aspects, like

the one in Figure 1.4. These alternating layers do not necessarily indicate multiple up-and-down excursions through the storm.

Additionally, wet growth generally favours the stone's growth in three ways (ALLEN et al., 2020): I) by setting a liquid layer to the hailstone which changes its thermal conductivity; II) increasing the stone's collection efficiency and III) the liquid will possibly alter the hailstone's drag characteristics. Moreover, the regime of the hailstone formation impacts on its radiation scattering properties, with main implications in the detection and sizing of hail through radar.

Figure 1.4 - Hailstone aspects of undergoing sequences of dry and wet growth.

SOURCE: Allen et al. (2020).

Overall, optimal hail growth is a function of 5 factors:

a) **appropriate updraft strength and breadth:** in order to match the terminal velocities need of different stages of hail growth (i.e. embryo and hailstones);

b) **sufficient amounts of supercooled liquid water:** so to enable a proper collection by the embryo or hailstone along its trajectory inside the updraft region;

c) **proper temperature profile:** as part of an environment conducive to hail growth;

d) **adequate number and size of embryos:** so that the particles are balanced in the updraft and the competition is not sufficient to deplete the total water content;

e) **favourable airflow patterns:** optimal trajectories inside the storm that lead the embryos to environments with proper growing conditions.

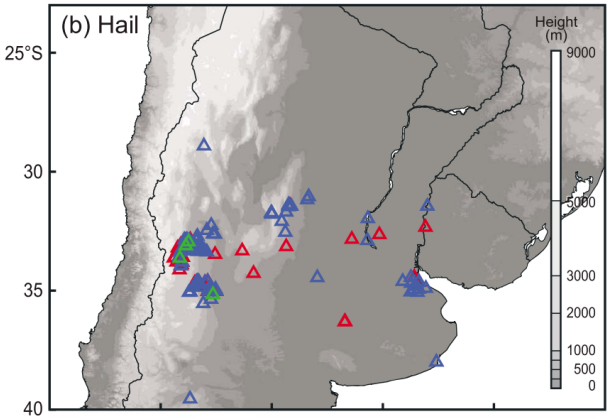
1.1.2 Hailstorms in South America

Hail is a meteorological phenomenon commonly reported across all continents except Antarctica. Hailstorms happen at most latitudes, favored by convective atmospheric conditions (FRISBY; SANSOM, 1967; ALLEN et al., 2020). In South America, hailstorms are notably intense and are influenced by distinct atmospheric dynamics (ZIPSER et al., 2006). These events are particularly associated with the frequent formation of large and persistent mesoscale convective complexes (MCCs) (LAING; FRITSCH, 1997; DURKEE; MOTE, 2010). Sub-tropical South American MCCs are known to be significantly larger and last longer than in the US and do not exhibit much poleward migration throughout the warm season (MARENGO et al., 2012).

Specifically in the southeastern region of SA, the intense convective events happen in all seasons and are favoured by the influx of moist air by the South American Low-Level Jet (SALLJ) (SILVA; BERBERY, 2006) and strong low-level wind shear. Additionally, the Andes Mountains play a crucial role in triggering convective instability by lifting low-level air masses (RASMUSSEN; HOUZE, 2011). Within the Subtropical Andes region, between 10 °S and 20 °S, local thermally driven circulations also provide dynamical forcing (boundary layer convergence) that favors convection at a time of the day that the boundary layer is most unstable (GARREAUD; WALLACE, 1997).

Overall, MCCs are more frequent east of the Andes Mountains below 20 deg S, in a region that comprehends the La Plata Basin (MARENGO et al., 2012). Simultaneously, high hailstorm activity has been identified at the extreme foothills of the Andes and Sierras de Córdoba near the cities of Mendoza and Córdoba as shown in Figure 1.5, frequently harming the crops in this area (PÉREZ; PULIAFITO, 2006). The maximum hail occurrence was found by Rasmussen et al. (2014) to coincide with the "South American Tornado Alley" in La Pampa region. Bruick et al. (2019) report that in general, hailstorms in subtropical South America occur more frequently in the overnight hours and tend to be multicellular in nature. In accordance with their methods, CSs larger than 15000 Km^2 had a higher probability of generating hailfall.

Figure 1.5 - Hailfall cases in Argentina and Uruguay.

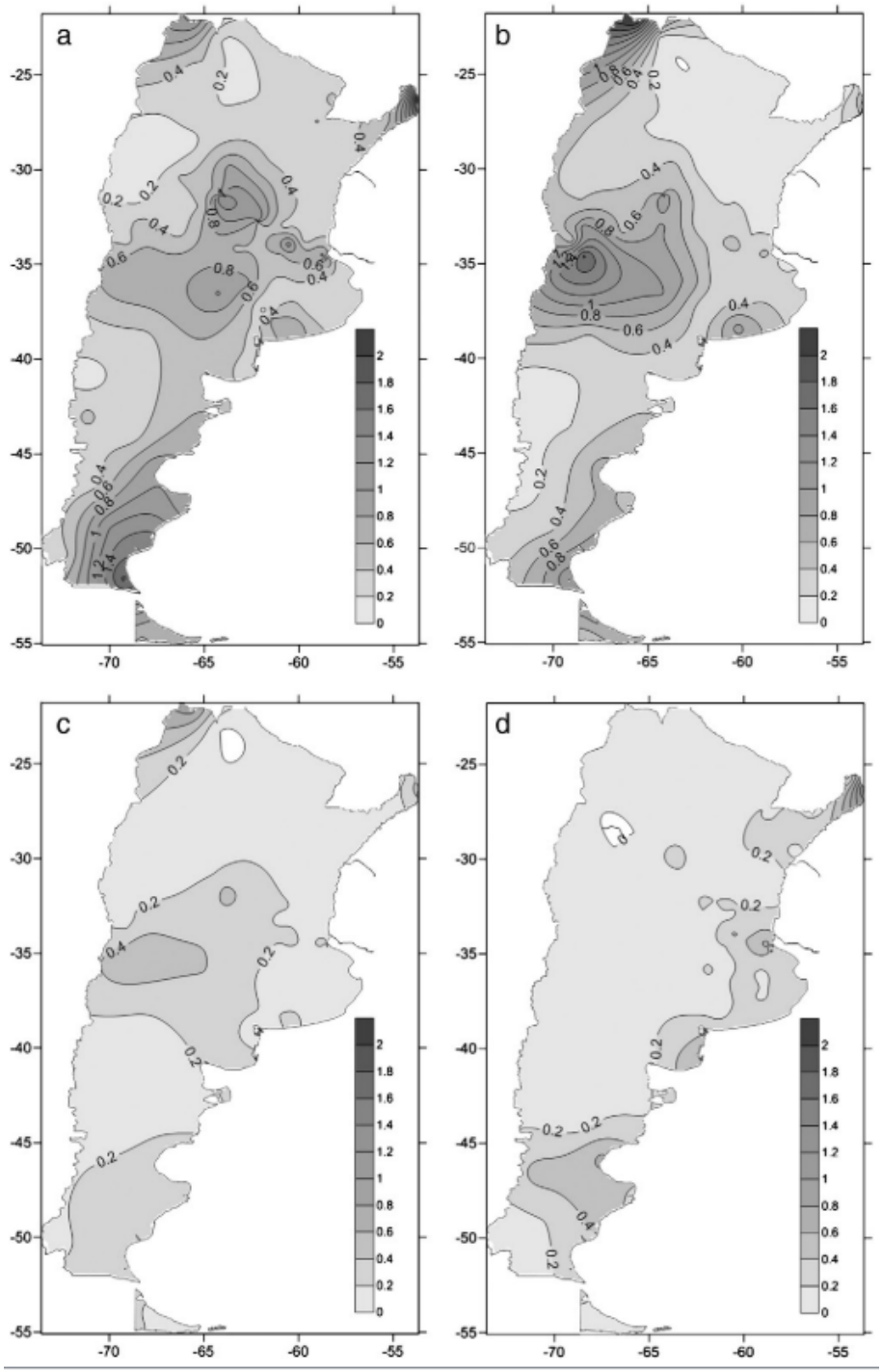


Pinned locations, gathered from local media. Triangles in red refer to Spring (SON), blue represents summer (DJF) and green is Autumn (MAM).

SOURCE: Rasmussen et al. (2014).

An in-depth climatological analysis in Argentina carried out by Mezher et al. (2012) has identified regions of high risk of hailstorms in central Argentina, outlying Buenos Aires and the coastline. As presented in Figure 1.6, hail cases are more common during Spring and Summer seasons, with highest frequency from August to January. Previous studies also linked hailstorms to the evening hours (i.e. between local 16:00 to 18:00) (SALUZZI; NUÑEZ, 1975) and revealed interdecadal oscillations in hail intensity in the Andes region with an approximate 20 year period (PRIETO et al., 2001).

Figure 1.6 - Seasonal mean of hail events in Argentina.



(a) Spring (SON), (b) Summer (DJF), (c) Autumn (MAM) and (d) Winter (JJA). SOURCE: Mezher et al. (2012).

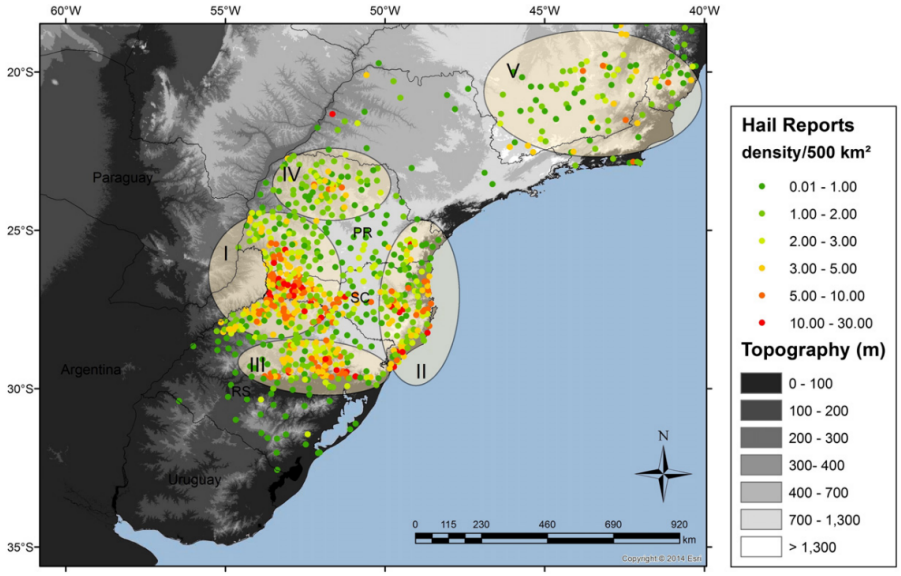
In Brazilian territory, Martins et al. (2017) gathered 1630 reports of destructive hail events (i.e. affected areas where residents are rendered homeless) from Civil Defense between 1991 and 2012. Their analysis have shown that most of the Civil Defense actuation due to hail damage were in the southern states. According to the authors, these events are overall more frequent in late afternoon and evening of the winter to spring transition. In this study, were identified four main areas in which those events tend to cluster. The highest incidence of destructive hailfall is on the western part of the three Brazilian southern states and extends to the central portion of Santa Catarina state. There is an increase in frequency of this type of report from July to December, which follow closely the temporal trends of general hailfall in the state of Santa Catarina alone (MARCELINO et al., 2004). The second main region in frequency, encircles Brazilian southern coastal strip. The third region is the central part of Rio Grande do Sul state. Other two well-defined regions were found with less frequency of destructive events, one to the North of the state of Paraná an another extending from the coast of Brazilian's Southeast region. All of those clusters and report are displayed in Figure 1.7.

According with the Civil Defense reports, the period with most frequent hail fall in Southern Brazil is from August to November, with September and October holding most of the events within this time-range. In Rio Grande do Sul state, is common to observe a spike in hail event in July, which is likely the result of large, long-lived storms affecting multiple municipalities in a single or small set of storm outbreaks. Those reports also reveal that throughout the day, it is more likely for the southern states to be hit by hailfall between late afternoon and early evening, which is closely tied to diurnal solar heating and local thermodynamics. The states of Rio Grande do Sul and Santa Catarina presented an additional late night peak in frequency of destructive hail events, likely associated to the mature phase of MCSs in the region (MACHADO et al., 1998; ZIPSER et al., 2006).

The Brazilian tropical region is very active in Convective Storms. For instance, in 47 days Machado and Laurent (2004) tracked 13409 CSs within 7.5 °N - 20 °S and 80 °W - 30 °W, during the WETAMC/LBA campaign. However, hailstorms are rare there and possibly under-reported. Until 1967 there were no official hail reports in the Amazon Valley, and amateur spotters experienced hailfall in the city of Obidos and hail of approximately 0.5 cm in Rio Cururu (FRISBY; SANSOM, 1967). Hail of up to 2 cm hailstones on the ground and mid-flight encounter with hail were reported under smoky air conditions during the LBA-SMOCC campaign (ANDREAE et al., 2004) in the amazonian region. No destructive hail event was reported in brazilian Equatorial

region in the study of Martins et al. (2017). Deep convection is known to develop in this region frequently and throughout the whole year (WILLIAMS et al., 2002; MACHADO et al., 2004), possibly boosted by an aerosol-invigoration mechanism (ROSENFELD et al., 2008) during the periods of stronger biomass burning in the Amazonian region (KOREN et al., 2004; GONÇALVES et al., 2015). However, Lebo (2018) after several simulations, suggested that overall, this mechanism is small and likely immeasurable.

Figure 1.7 - Destructive hail events for the southern and southeastern Brazilian States.



Brazilian Civil Defense reports from 1991 to 2012. Dots represent individual municipalities and the ovals encircle regions of interest.

SOURCE: Martins et al. (2017).

1.1.3 Hailstorms and lightning activity

The primary process for storm electrification is the rebounding collisions between graupel and ice crystals in the mixed-phase region of the cloud (MACGORMAN; MORGENSTERN, 1998), even though there are still uncertainties (SAUNDERS, 1993; ALLEN et al., 2020). Therefore, the local flash rate (i.e. lightning flashes over a specific geographic area within a certain time frame) is directly related to the convection's vigor and upward fluxes of water under mixed-phase conditions

(SCHULTZ et al., 2015). Accordingly, the behavior of lightning activity over time is commonly used as a proxy for the detection of intense and developing storms (MACGORMAN; BURGESS, 1994; CAREY et al., 2003), specially over land (ORVILLE; HENDERSON, 1986).

On these grounds, hailstorms have been associated to a sudden increase in the total lightning activity (both intra-cloud and cloud-to-ground lightning) during the storm developing stage, namely a "lightning jump" (WILLIAMS et al., 1999; SCHULTZ et al., 2009). Those jumps are commonly linked to a storm updraft intensification, which indicates the storm's ability to sustain convection (CHRONIS et al., 2015), indicating a favorable environmental setup for the development of large hail (SCHULTZ et al., 2011; SCHULTZ et al., 2017). In many severe hail cases, a lightning jump occurs well before the observed hail, indicating its potential to increase the lead time of warnings for severe weather, such as hailstorms (WAPLER, 2017) and windgusts associated with thunderstorms (FARNELL et al., 2017).

Reportedly, some rapidly developing hailstorms evolve from no-lightning to a lightning peak in less than 10 minutes and the subsequent hailfall occurs 4 minutes after this maximum (GOODMAN et al., 1988). In such cases, an abrupt decrease in total flash rates (i.e. storm collapse) is a precursosr to the hailfall. This fast storm development possibly curbs its lead time detection through operational 15-10 minutes full disk scans (CINTINEO et al., 2013; BEDKA et al., 2015; RIBEIRO et al., 2019). In fact, the lightning rate paired with full disk ABI window measurements has the potential to worsen automated severe severe weather detection (CINTINEO et al., 2020a).

1.2 Research questions and objectives

This research aims to advance the understanding and detection of hailstorm events in Brazil using data from the GOES-16 satellite, focusing on the contributions of the Advanced Baseline Imager (ABI) The overarching goal is to develop and validate an automated machine learning classifier for hailstorm detection, alongside a comprehensive analysis of storm characteristics and their temporal and spatial variations. The following research questions guide this investigation:

a) Development and Validation:

   - How can machine learning algorithms be effectively utilized to classify storm events as potential hailstorms using thermal infrared data from the ABI on the GOES-16 satellite?

- What are the precision and reliability of the developed classifier in detecting hailstorm events over Brazil?

b) Feature Importance Analysis:

- Which features derived from the ABI's thermal infrared bands are most important in predicting hailstorms in Brazil?

- How do these key features relate to the physical characteristics of hailstorm-producing convective storms?

c) Implications for Early Warning Systems:

- How can the findings from the classifier's performance and the analysis of storm behavior enhance the development of early warning systems for hailstorms in Brazil?

- What recommendations can be made for improving satellite-based monitoring and prediction of hailstorms, based on the identified important features and storm behaviors?

These research questions aim to contribute to a deeper understanding of hailstorm dynamics in Brazil and the development of more accurate and reliable hailstorm detection methods using satellite data. This research not only seeks to fill gaps in the current knowledge but also to provide practical insights for weather forecasting and disaster preparedness in the face of hailstorm events.

1.3 Significance of the study

This research aims to contribute to the field of meteorology by applying an innovative approach to the detection and analysis of hailstorms through satellite technology. The methodology proposed seeks to enhance the efficiency of hailstorm identification and tracking via geostationary satellites, which may improve the timeliness and accuracy of assessments conducted by meteorologists, disaster management teams, and the agricultural sector.

The study investigates various factors related to hailstorm detection, which could

offer fresh insights into the spatial and temporal characteristics of these weather phenomena. Through a systematic examination of these variables, the research aims to deepen the understanding of hailstorm behaviors, potentially aiding in the refinement of their predicted patterns, intensities, and trajectories. Such insights are expected to be beneficial to the scientific community, potentially facilitating advancements in research and refining existing weather prediction models.

Moreover, the outcomes of this research might have implications for the fields of risk management and insurance. Enhanced detection and analysis techniques could improve long-term assessments of hailstorm-related risks and damages, potentially influencing policy-making and insurance rate setting with more precise data.

In summary, this study endeavors to contribute to advancements in meteorological research through the application of novel satellite technology and analytical methods. While primarily academic in nature, the potential practical applications of this research span weather forecasting, disaster management, and economic sectors impacted by hail-related events, aiming to support better response to natural disasters.

1.4 Scope and limitations

The efficacy of our algorithm is inherently tied to the capabilities and constraints of satellite observation. We utilize the measurements provided by the GOES sensors, which means our analysis is limited to the features that can be detected through passive remote sensing across the specific wavelengths these sensors operate in. This approach provides a unique vantage point for storm observation but also introduces inherent limitations in the scope of detectable phenomena.

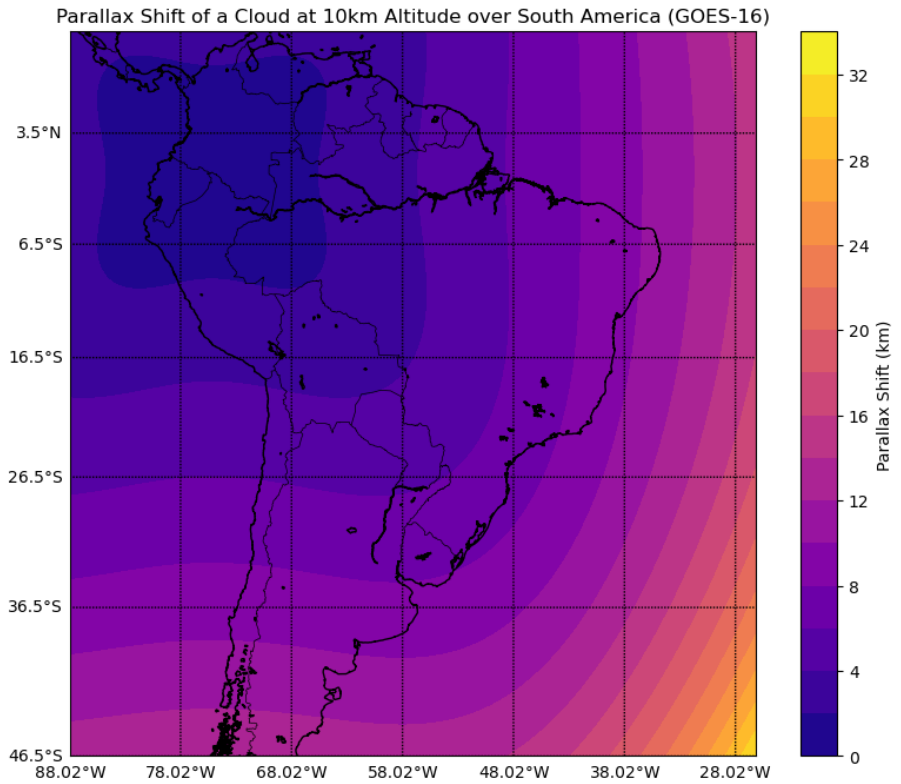
One significant limitation arises from the temporal resolution of the GOES-16 ABI Full Disk measurements. Some rapidly evolving hailstorms transition from their developing stage to collapse within the timeframe between two consecutive Full Disk observations. This rapid evolution significantly hampers our ability to detect and analyze such storms effectively, as the temporal granularity of satellite data may not be sufficient to capture these quick transitions accurately.

Regarding hail reports across Brazil, our data is primarily crowd-sourced, deriving from civil defense actions, observer reports, and newscasts. The nature of this data collection results in uneven spatial coverage across the study area. Given that hail-fall events, as verified from the ground, typically occur on a localized scale,

our records are disproportionately concentrated in more densely populated regions. This skew in data collection is critical because our algorithms depend on robust and accurate datasets to effectively learn and generalize across the study region. Consequently, areas with sparse human activity are underrepresented in our records, which could potentially compromise the model's ability to generalize effectively. This under-representation may not solely be due to limited human presence; in some cases, it reflects the actual lower frequency of hailstorms. However, the dearth of data from these regions poses challenges for both model training and validation.

Another aspect to consider is the known measurement artifact related to the positioning of hail-producing storms. Hail-fall is closely associated with large towering cumulus clouds. However, the satellite parallax effect—particularly when observing from the GOES-16 ABI's non-nadir perspective, coupled with the vertical extent of these storms—can lead to inaccuracies in pinpointing the cloud tops' actual geographic locations. This displacement issue, as outlined by Ayala et al. (2023), necessitates careful consideration of the potential uncertainties in storm positioning detected via satellite.

Figure 1.8 - Parallax effect on GOES-16 ABI measurements.



Approximate shift of a GOES-16 ABI target at 10km altitude due to Parallax effects. The color encoded distance represents a measure of target's positive displacement in the sensor's pointing direction.

SOURCE: The author.

In addressing these limitations, we aim to highlight the importance of considering both the technological constraints of our observational tools and the inherent challenges in sourcing and utilizing crowd-sourced data. Our ongoing efforts to mitigate these issues include refining data collection methods, enhancing algorithmic learning from available data, and continuously evaluating the impact of observational limitations on our findings

1.5 Thesis outline

In the subsequent chapters, this thesis explores the utilization of satellite data for hailstorm detection. Chapter 2 lays the groundwork by reviewing existing methodologies in the field, highlighting significant discoveries, and identifying existing gaps

in the assessment of hailstorms. This chapter sets the stage for understanding the current landscape of hailstorm research and the necessity for further exploration.

Chapter 3 delves into the experimental design, offering a comprehensive explanation of the experimental settings, data collection procedures, and the analytical and validation steps undertaken. This detailed overview ensures a clear understanding of the methodology and the rigorous approach to data handling.

The findings from these experiments are thoroughly examined in Chapter 4. Here, each experimental result is meticulously described, setting the foundation for a robust discussion on the implications and significance of the findings.

Chapter 5 engages in a critical discussion of the results, where the experimental outcomes are interpreted and juxtaposed with existing literature. This chapter not only elucidates the implications of the study's findings but also sheds light on its limitations and suggests paths for future research, also pointing out the study's impact.

Concluding the thesis, Chapter 6 synthesizes the discussions from previous chapters to draw final conclusions. It succinctly summarizes the key findings, underscores the study's primary contributions to the field, and offers concluding remarks that encapsulate the essence of the research. This final chapter serves as a reflection on the study's achievements and its potential influence on future hailstorm detection methodologies.

2 LITERATURE REVIEW

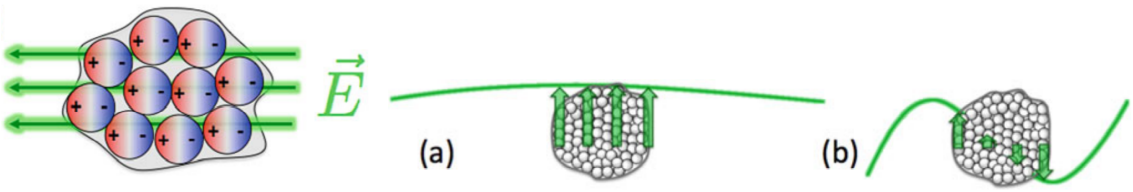
2.1 Key theories and models

In this section we introduce the theory behind the detection of potential hailstorms with use of active and passive sensors. As follows, firstly we introduce the model behind the scattering regime of hail in the Microwave (MW) region of the spectrum and following, we describe how some phenomena associated with hailstorms perceptive in the Infrared (IR) portion of the spectrum are explored as proxy for hailstorms detection.

2.1.1 Hail microwave scattering

The microwave remote sensing of large hail is based on the principles of radiation scattering by ice hydrometeors (KUMJIAN et al., 2018). Under the model of Bohren and Huffman (1983) of a particle composed of smaller oscillators, the scattered radiation pattern is highly dependent on the correspondence of the wavelength with the diameter and the level of symmetry of the particle. The resulting scattered radiation is the aftermath of the interference of the oscillation produced on the infinitesimal parts of the particle interacting with the incident wave. A scheme of this interaction between distinct wavelengths and particle sizes is presented in Figure 2.1

Figure 2.1 - Interaction of particles with electric field.



The resulting charge alignment under the model of a particle composed of smaller oscillators. In (a), the affected particle presents a negligible size compared to the wavelength; in contrast, (b) shows the interaction with an almost wavelength-sized particle.

SOURCE: Andronache (2018).

Based on this principle, the type of scattering presented by a particle is inferred

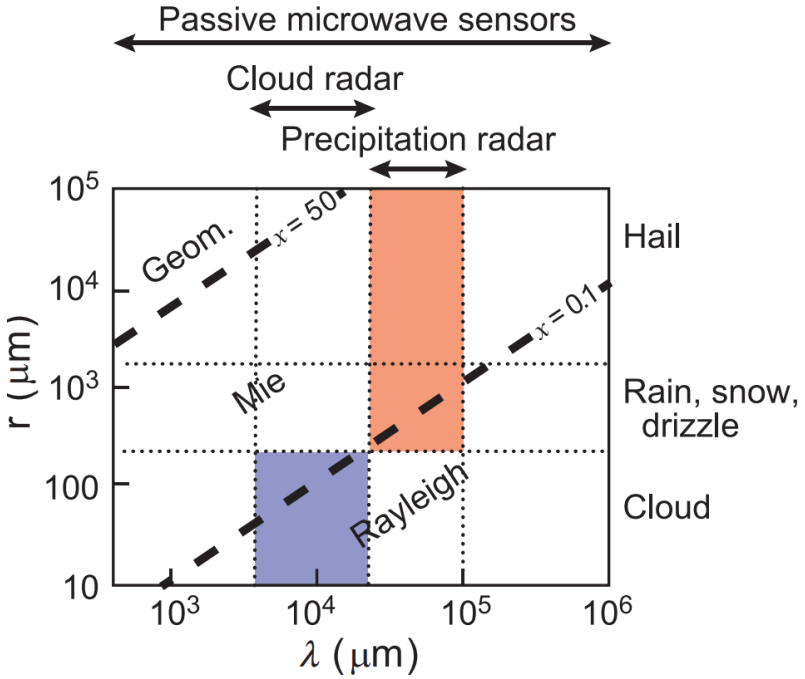
by the relationship between its size and the wavelength of interest, given by the formula:

$$\chi \doteq \frac{2\pi r}{\lambda} \tag{2.1}$$

, where r is the particle radius and λ is the wavelength. In the case of a small χ, the hydrometeor would be small compared to the wavelength of reference. In such case, the scattering by the particle is too small and the radiation extinction would be governed by the absorption (HOUZE, 2014). As a general rule, when $\chi < 0.1$, there is a Rayleigh scattering regime; for $\chi > 50$, the scattering physics is tied to the rules of geometric optics. The more extensive Mie regime, in which both scattering and absorption considerably affect the beam of radiation, applies to intermediate particle size-wavelength ratios, i.e $0.1 \leq \chi \leq 50$ (BOHREN, 1992).

Figure 2.2 summarizes the particle type that may be identified under each spectra and the respective scattering type involved. Hailstones, due to its size and external composition, tend to be Mie scattering particles, although in strict terms Mie's theory should only be applied to homogeneous dielectric spheres (MIE, 1908). Therefore, in general, hailstorms are easier detected with the aid of precipitation radars or passive microwave sensors in the wavelength range of 1 to 10 cm.

Figure 2.2 - Microwave scattering regimes.



Dashed lines refer to the relationship of particles size to wavelength χ, separating scattering patterns in the wavelength domain of atmospheric passive microwave sensors and radars.

SOURCE: Houze (2014).

2.1.1.1 Passive microwave remote sensing

Satellite-borne passive microwave radiometers offer a spatial coverage hard to achieve through ground-based sensors. However, in comparison, there is a loss on temporal resolution due to their necessary low orbit, since the most frequent site revisit by the time this work was written is 4 hours.

As previously described, hydrometeors absorb, emit and scatter radiation in the microwave spectrum (CZEKALA; SIMMER, 1998). Therefore, a good strategy for sensing ice particles in the atmosphere is to carefully choose a wavelength for which scattering by liquid precipitation particles is minimal and that has a weighting function maximum roughly corresponding to a layer of the atmosphere that is populated by hailstones.

An advantage of the use of microwave frequencies for sensing hydrometeors is that ice absortivity is very small compared to that of liquid water (HOUZE, 2014), and their scattering efficiency is very distinct as well (LIOU, 2002). One must then check the depression on different wavelengths brightness temperature (BT) to infer the amount of particles. The larger the ice size and ice volume inside the satellite field of view (FOV), the greater is the BT depression due to scattering of radiation to outside the FOV (FERRARO et al., 2020).

This impact of ice hydrometeors on the BT on the microwave spectra was first described by Spencer et al. (1983a) when comparing observed data in 37 GHz (0.8 cm) from the Nimbus 7 Scanning Multichannel Microwave Radiometer (SMMR) with the modeled transfer of 37 GHz radiation through rain clouds. By analyzing a severe rain cell over land in southeast Kansas, they were able to detect an unusual BT depression over 37 GHz that coincided with a less pronounced one in 18 GHz. More recently, Cecil et al. (2005) highlighted the impacts of hail on 85 GHz, using data from the Microwave Imager (TMI) of the Tropical Rainfall Measuring Mission (TRMM) (KUMMEROW et al., 1998) and on measurements of the Aqua Advanced Microwave Scanning Radiometer (AMSR-E) on the Earth Observing System (EOS) (CECIL; BLANKENSHIP, 2012). There is also an impact on higher frequencies, such as in the water vapor 183 ± 1 GHz, for ice hydrometeors above 7 km (HONG et al., 2005).

The more particles in the vertical column, the greater will be the attenuation of the upwelling radiation. Then, most of estimates are performed through an analysis of relative behavior in comparison to the background radiation. Therefore, different wavelengths shall better depict hail over distinct surfaces (CECIL, 2009).

In general, the extinction coefficient on channels over 85 GHz is large for millimeter sized idealized spherical ice (MROZ et al., 2017), which makes large graupel more likely to be detected than proper hailstones. In contrast, attempts to predict hailfall based on the 19 GHz channel result in less false alarm, but lack of refinement due to the low spatial resolution achieved (30 x 18 km^2 on TRMM) (SHIN; CHIU, 2005). These drawbacks, summarized in Figure 2.2, led Cecil and Blankenship (2012) to opt for 37 GHz of AMSR-E to establish hail probabilities based on BT threshold, which resulted in a global hail estimates record from 2003 to 2010.

2.1.1.1.1 Algorithms

Most of the research on improving hail detection through satellite-born passive microwave sensors focus on determining thresholds of BT depression by effect of large ice hydrometeors.

For instance, Cecil (2009) examined three frequencies from TMI passive microwave data together with radar reflectivities and lightning measurements from TRMM of severe storms to establish low BT values associated with surface hail reports. The outcome were that a BT $< 70\ K$ at 85 GHz, $< 180\ K$ at 37 GHz or $< 230\ K$ at 19 GHz represent a high likelihood of hailfall.

More recently, Ni et al. (2017) gathered surface reports of hailstones over the southeastern and south-central United States and matched with data from the TMI and Ku-band reflectivity from TRMM Precipitation Radar to achieve detection criteria for hail larger than 19 mm. The outcome was that the highest skill scores were yielded by a 230 K threshold at 37 GHz and 44 dBZ at $-22°C$ level.

A similar approach was taken by Mroz et al. (2017), but exploiting the suite of sensors from GPM Core Observatory satellite together with polarimetric measurements from the ground-based S-band Next Generation Weather Radar (NEXRAD) network to accomplish the proxies for the identification of hail. In this case, the hydrometeors were firstly classified based of the NEXRAD data, using fuzzy logic to serve as ground truth. With that, they were able to obtain the best performance by employing a threshold based hail detection algorithm combining the Ku and Ka-bands reflectivity in the mixed-phase layer. Among the GPM microwave Imager channels, the polarization-corrected temperature at 18,7 GHz showed the greatest potential for hail detection, at a threshold value of 261 K.

With respect to sounders, according to Ferraro et al. (2020), there are two primary hail detection algorithms: the Microwave Cloud Classification (MWCC) and the NOAA algorithm.

The MWCC (LAVIOLA; LEVIZZANI, 2009) was originally developed for the AMSU-B/MHS and now also adapted to the Special Sensor Microwave Imager/Sounder (SSMIS) of the Defense Meteorological Satellite Program (DMSP); the Advanced Technology Microwave Sounder (ATMS) from the Suomi National Polar Partnership (S-NPP) and for the GMI. The main task consists of leveraging high frequency channels to identify the cloud top altitude and the phase of hydrom-

eteors. The algorithm's retrieval design is to characterize the vertical development of clouds based on their impact on the water vapor channel (i.e. $\approx$ 183.31 Ghz). In order to evaluate the signal depression in this channel, an unperturbed field is dynamically estimated as the absolute maximum observed BT among the available channels.

That being the case, the hail detection is based on the reduction of the signal on high frequency channels (in the range of 90 to 190 GHz) due to the increase in ice hydrometeors on the convection development. Therefore, the BT depression is calculated in every channel by comparing to the respective clear-sky BT. The key feature of the hail detection tool (MWCC-H)(LAVIOLA et al., 2020a) is to associate probability values to perturbations from very small ice particles (low probability) to very large hailstones (i.e. diameter $>$ 10 cm), attributed to close to 1 likelihoods (LAVIOLA et al., 2020b). This transform is performed with the aid of a logistic model trained on 10 years (2000-2009) of MHS measurements co-located with hail reports over the US and MHS data extracted from selected cases of hailstorms.

The NOAA AMSU-B/MHS algorithm works on the same physical principles but the hail detection is based on permanent BT thresholds for the sensor's five channels. These values were established from a 10-year (2000-2009) training set that limited the matches to zenith angles within 25 degrees of nadir (FERRARO et al., 2015). Then, the hail presence is accused if the following conditions are met:

- $TB_{89} < 228.2K$
- $TB_{150/157} < 206.9K$
- $TB_{183+1} < 211.1K$
- $TB_{183+3} < 204.6K$
- $TB_{183+7/190.3} < 200.5K$

Therefore, there is the possibility that the thresholds will not apply for other locations because the ambient BT is not updated. False alarms for hailstorms may also be raised in locations where the freezing level is higher than the US. In such cases, the hail detected under the determined thresholds may melt completely before reaching the ground.

2.1.1.2 Active microwave remote sensing

Active remote sensing in the microwave spectrum is commonly performed through Radars. For such, measurements are performed by switching between emitting and receiving pulses of microwave radiation through an antenna (MEISCHNER, 2005). The signal received is affected by multiple targets lying along the beam's path, and a lot can be inferred after appropriate data processing. The common parameters obtained are: the reflectivity of the target, obtained through the power of the received signal; the target's velocity along the beam, estimated through the Doppler shift (HEISS et al., 1990); and information on shape and/or orientation of the target, based on the polarization of the returned signal (BRINGI; CHANDRASEKAR, 2001).

Due to its capabilities and high performance in sensing precipitation, there is a large investment worldwide on ground-base radar networks such as NEXTRAD (HEISS et al., 1990) and others (HUUSKONEN et al., 2014; BENDIX et al., 2017; SALTIKOFF et al., 2010). In addition, satellite born radars were available at TRMM and now on the GPM, operating at Ku-band (13.8 GHZ). The GMI includes a second active frequency at 36 GHz (Ka-Band). Recent technical specifications of existing and planned spaceborne radars are detailed in Battaglia et al. (2020). This represents an alternative to overcome the spatial coverage limitation of the ground networks.

In order to detect precipitating cores, radars must operate in the centimeter range, generally between $\lambda = 1 - 30cm$. The S-band, i.e. 2 - 4 GHz ($15 - 7.5cm$ wavelength) (PARKER, 2017) is the shortest at which precipitable water may be detected and estimated under no attenuation by rain (HOUZE, 2014).

A commonly exploited variable in the precipitation domain is the radar reflectivity, which is equivalent to the scattering cross section of a unit volume of air. Among other factors, this variable is a function of the power of the returned signal. Based on that, one can estimate the *Equivalent Radar Reflectivity Factor* Z_e and the dBZ_e (i.e. Z_e given in decibel units or mm^6m^{-3}).

Certain types of precipitation present distinct ranges of reflectivity (HAZENBERG et al., 2011). Typical values of dBZ_e for certain atmospheric phenomena are described in Table 2.1.

Table 2.1 - Typical reflectivity factors for the corresponding atmospheric phenomena.

dBZ_e	**Associated event**
$\sim$ -50 to 0	nonprecipitating cloud and marginally detectable precipitation
$\sim$ 0 to 10	drizzle, very light rain, light snow
$\sim$ 10 to 30	moderate rain and heavier snow
$\sim$ 30 to 45	melting snow
$\sim$ 30 to 60	moderate to heavy rain
$\sim$ 60 to 70+	**hail**

SOURCE: Adapted from Houze (2014).

2.1.1.2.1 Dual-polarization radar

With the aid of dual-polarization radars, other signal features can be exploited to analyze the hydrometeors, such as the amplitude and phase as a function of polarization (KUMJIAN et al., 2013). Under these circumstances it is possible to get a better depiction of the targets, specially in the sense of size and approximate shape (BECHINI; CHANDRASEKAR, 2015; ROBERTO et al., 2017).

With the aim of establishing more detailed aspects of the targets, two new variables may be derived from the polarized signal: (I) the *differential reflectivity* (Z_{DR}) and (II) the *linear depolarization ratio* (LDR), as describe in Equations 2.2 e 2.3.

$$Z_{DR} = 10log_{10}\left(\frac{Z_{HH}}{Z_{VV}}\right) \tag{2.2}$$

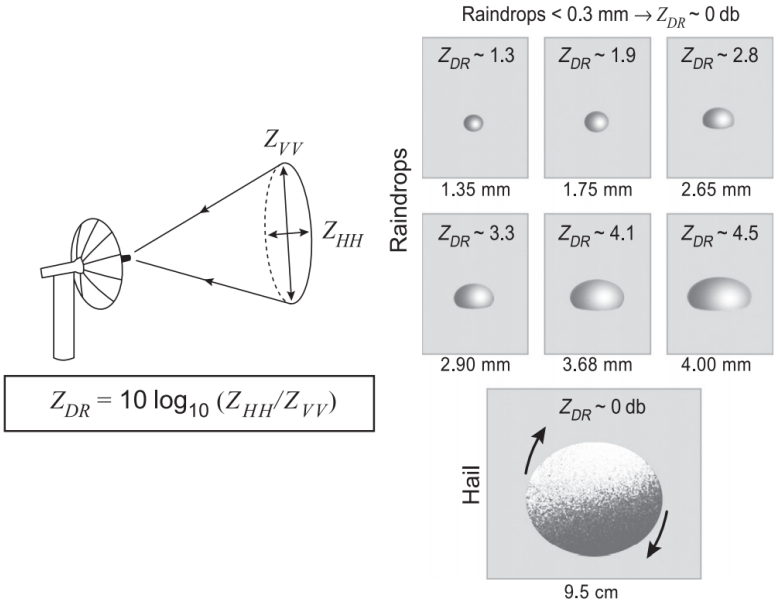
$$LDR = 10log_{10}\left(\frac{Z_{HV}}{Z_{HH}}\right) \tag{2.3}$$

Where Z_{HH}, Z_{VV} and Z_{HV} are the reflectivity values for when its horizontally transmitted and received (HH); vertically transmitted and received (VV); and horizontally transmitted and vertically received (HV), respectively.

Therefore, hydrometeors that tend to be more horizontally oriented, such as large raindrops and aggregate snowflakes, will lead to large values of Z_{DR}, as a result of its wide horizontal dimension and because melting increases the index of refraction of the particles. In contrast, small ice particles and hailstones will return near zero Z_{DR}. This is due to the fact that small ice generally presents a low index of refraction and hailstones tend to tumble on their trajectory, presenting no constant orientation.

Figure 2.3 show typical values of Z_{DR} for different hydrometeors with distinct sizes and shapes.

Figure 2.3 - Impact of hydrometeors on the *Differential Reflectivity* Z_{DR}.

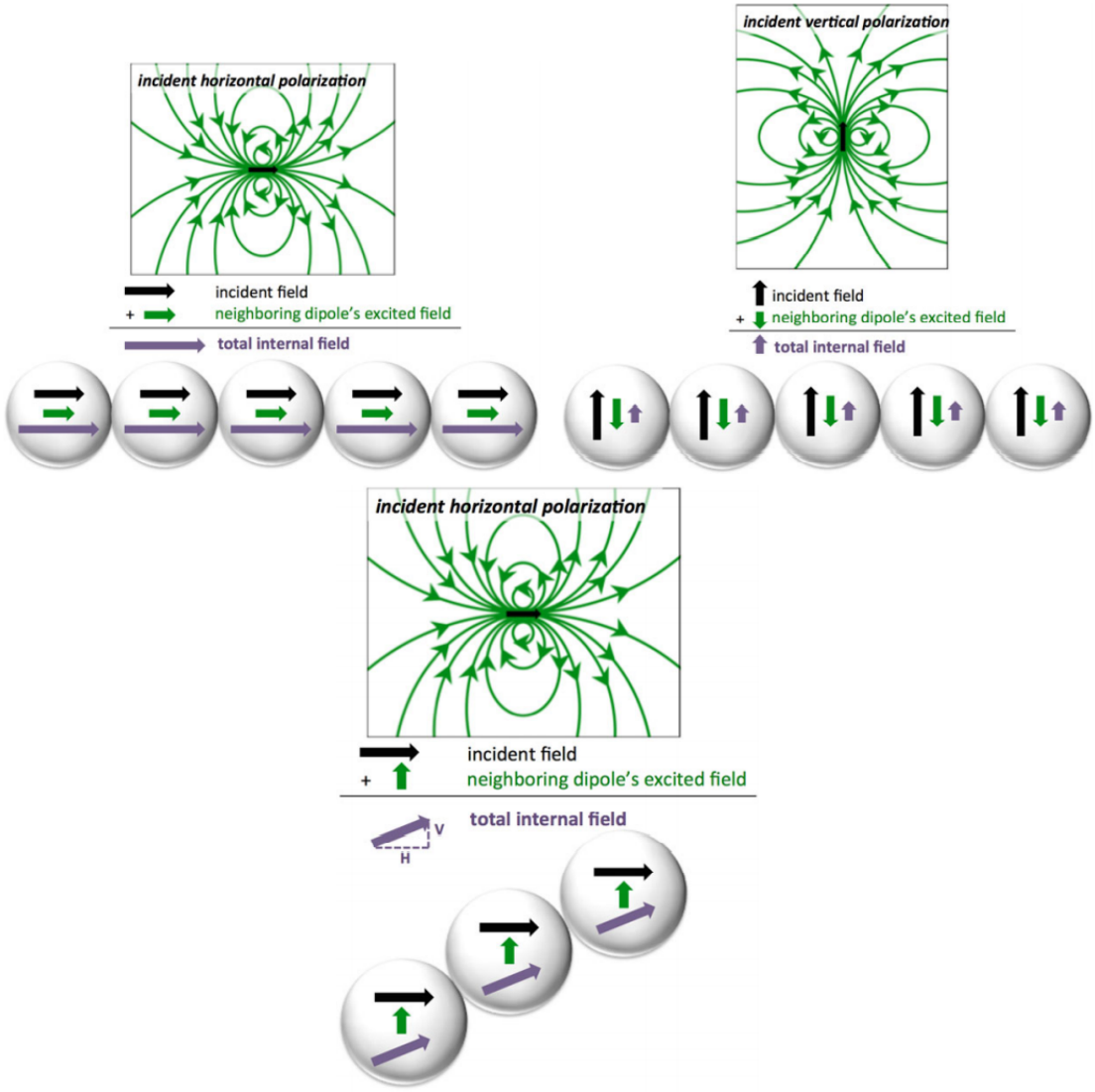


Z_{VV} refers to the vertically polarized reflectivity and Z_HH to the horizontally polarized reflectivity. The arrows around the Hailstone indicate its tumbling motion.

SOURCE: Houze (2014).

When the particles present no axis of symmetry parallel to the direction of the wave polarization, they are prone to depolarize the propagating wave (MATROSOV et al., 1996), under a scheme presented in Figure 2.4. In this case, the LDR will represent the magnitude of this effect on the horizontally transmitted signal.

Figure 2.4 - Electric field interference due to polarized incident waves.



Incident polarized radiation on a horizontally oriented planar particle composed by tiny sub-particle spheres that behave as dipole oscillators when illuminated with electromagnetic radiation. Upper left: the behavior induced by the horizontally polarized radiation. There, the induced electric field in each sphere are in the same direction to the neighbouring ones, leading to a constructive interference. Upper right: the induced response by the vertically polarized radiation. In this case, occur a destructive interference between the field components of the neighbouring cells because they are in opposite directions. Therefore, the response is reduced. In the bottom is shown the case of a slightly tipped particle in respect to the plane of incident horizontally polarized radiation. Consequently, the both aforementioned modes of interference take place, then the resulting field will be polarized in both directions.

SOURCE: Andronache (2018).

The electric field oscillations usually prevail parallel to the predominant orientation of the particle. Therefore, if the symmetry axis of a particle is not perpendicular to either the horizontal or vertical polarization directions of the incident beam, the resulting signal will present both a horizontal and a vertical component and $LDR > -\infty$ (DOVIAK; ZRNIC, 1993). This is because the horizontally oriented signal is always stronger, so the LDR tend to be finite but negative. Then the stronger the depolarization, less negative is the LDR. Also, in the backscattering by hydrometeors that are symmetric about the incident oriented electric field, the vertical components of the electric field oscillation excited within the particle by the horizontally polarized wave cancel by symmetry (OUE et al., 2015). In this case, no depolarization occurs, resulting in an $LDR = -\infty$. However, due to cross-coupling of signals (CHANDRASEKAR; KEELER, 1993), Z_{HV} is always positive, and as a consequence, LDR is finite even for symmetric particles.

The shape effects on the LDR by ice particles are often suppressed due o its low refraction, but are enhanced when the particles are wet (i.e. partially melted). Therefore, LDR can be used to estimate the melting layer and also to identify the wet growth of hail and graupel (KOWALEWSKI; PETERS, 2010).

Another consequence of the scattering by precipitation is the phase shift. When there is a forward scattering by the particle, the scattered wave becomes phase shifted as to the wave in free space. In case of non-spherical hydrometeors, this effect can be tracked by difference in phase shift of the vertically polarized wave in comparison to the horizontally polarized one, which gives the *one way differential propagation phase* (ϕ_{DP}) as shown in Equation 2.4.

$$\phi_{DP} = \phi_{HH} - \phi_{VV} \tag{2.4}$$

However, ϕ_{DP} grows as the wave crosses a region of oblate particles (KUMJIAN, 2013). Therefore, it is useful to differentiate the phase shift with respect to the path in order to remove this impact. This would give the *Specific Differential Propagation Phase* (K_{DP}). Hail and ice particles cause a K_{DP} of approximate $0°\ km^{-1}$ when dry and $1°\ km^{-1}$ under wet growth. In case of rain mixed with hail, the $K_{DP} \approx 2.5°\ km^{-1}$.

2.1.1.2.2 Algorithms

As seen above, the hail estimates and detection are not a direct output of the radar returned variables. Some post processing is needed in order to improve the inference on the hydrometeor type in the sampled volume. Common approaches involve some training based on historical data and return a probability distribution, others perform a separation of the signal belonging to classes of hydrometeors. The description of these techniques is presented below.

In the Single polarization radar domain, the vertically integrated liquid water content (VIL) is frequently used to estimate thunderstorm severity. VIL is highly dependent on the airmass characteristics and therefore, it is a hard to establish a direct relation between hail size and VIL value. Amburn and Wolf (1997) highlighted that there are high-topped thunderstorms with a large value of VIL which will not always produce large hail, as exist low-topped thunderstorms with low VIL that occasionally produce large hail. Their approach was then to normalize VIL by the echo top, based on the concept that thunderstorms of different top heights present similar maximum reflectivity when producing similar-sized hail. This new variable was called VIL density. Then, from historical surface-based reports of hail they found that the best threshold for severe hail ($\geq 19mm$) was a $3.5g.m^{-3}$ VIL density. At values $> 4.0g.m^{-3}$, every storm analyzed produced severe-criteria hail, regardless of the VIL or the thunderstorm height. A similar approach is taken by Witt et al. (1998) in the estimation of the Maximum Expected Hail Size (MESH) but using instead, a height weighting function on the vertical integration of the horizontally polarized reflectivity.

Another approach is to use fuzzy logic to identify hail (HEINSELMAN; RYZHKOV, 2006). These techniques involve taking as input the variables available and returning the probability of a certain class of hydrometeor. In order to do that, for each class considered by the algorithm, there will be a membership function which will express the probability for its class based on the set of polarimetric parameters (FERRARO et al., 2020). Environment variables gathered through sounding may also be fed to the model (CAPOZZI et al., 2018). The algorithm will then filter out the result and display the most probable class of hydrometeor in the sampled volume.

This is the case for the Hydrometeor Classification Algorithm (HCA) (PARK et al., 2009), which originally was able to discriminate between 10 classes of radar echo, based on Z_H, Z_{DR}, cross-correlation coefficient between horizontally and vertically polarized returns, K_{DP} and the magnitude of small-scale fluctuations in the

reflectivity and the differential phase fields. Further development included three hail size categories (RYZHKOV et al., 2013), and presented a better performance than operational single polarization hail detection algorithms (ORTEGA et al., 2016).

2.1.1.3 Limitations

The inference of hail-size and hailstorm intensity are still a challenge. Some of the best performing methods rely on validation data and training set that would require a dataset of trustworthy records (SHE et al., 2010).

Some limitations on the available data are related to the restrictions on the accessible network. For instance, C-Band radar data limits the hailfall prediction to a maximum warning time of about 20 min (SCHMID, 1992). This is because its operational frequency interacts with precipitation, therefore, a growing storm system will only be detected when raindrops, graupel or hailstones are present. Another example is the TRMM measurements, that only comprises observations of the precipitation within the 35°N - 35°S spatial range. Also, there is a limitation on the temporal resolution of measurements (e.g. AMSR-E present a 12 hour revisit interval) and spatial resolutions, as described before. All of these aspects, makes it harder to establish a consistent global climatology and depict a complete diurnal cycle (FERRARO et al., 2020).

There is also an obstacle when it comes precipitation estimates. It is possible directly relate the scattering and a rain rate (SPENCER et al., 1983b). However, it would be needed to properly convert the reflectivities forms Z_e to Z. Then, (I) the composition of the particles would have to be assumed, which is not always known and/or there is no clear separation (HOUZE, 2014); (II) both the curvature of the Earth and the elevation angle of the radar beam compromise the estimates of the particle-size distribution relative to a point on the Earth's surface; and (III) Z_e is calculated based on the returned power assuming that the distributed targets completely fill the resolution volume at a given range, which lead to an underestimation when it is not completely filled (MROZ et al., 2017). This effect is more pronounced the greater the range.

Some additional physical bounds apply to methods that leverage data from spaceborn sensors. Electromagnetic radiation from water surfaces is partially polarized, presenting lower emissivities on horizontally oriented radiation. The degree of polarization is a function of frequency and view-angle, affecting mostly the lower frequency channels (MROZ et al., 2017). One must therefore correct the BT to avoid

mismatches of storms with water surfaces. There are well established ways to obtain a polarization corrected temperature for some microwave frequency ranges though (SPENCER et al., 1989).

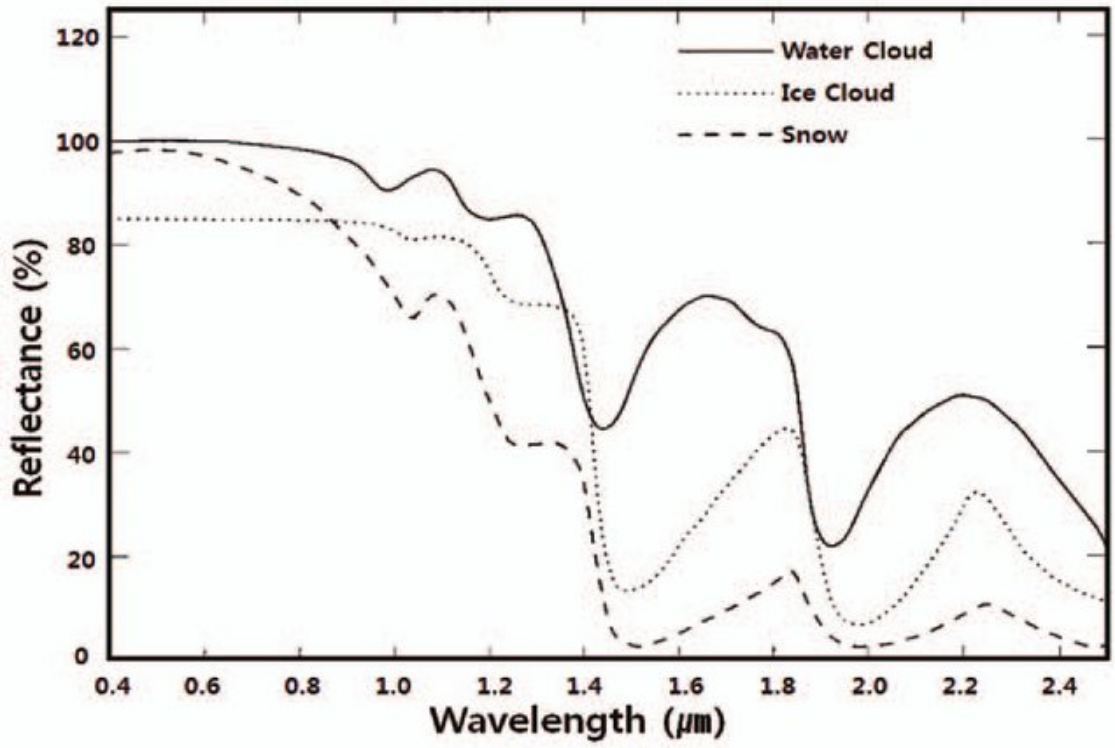
The geometry of some scans introduces a parallax-effect problem, which becomes a major issue for the analysis of tall convective systems. Mroz et al. (2017) described that for a 15 km high optically thick system viewed at a 50° angle with the GMI, there will be a $\approx 17\ km$ horizontal mismatch between the sampled volume and the point at which the bore sight intersects the Earth ellipsoid.

2.1.2 Hailstorms through visible and infrared wavelengths

One of the most intuitive factors that is critical in hailstorms is the presence of ice particles in the cloud. Then, a manner to evaluate the cloud phase is through BT in the range between $8 - 12\mu m$. The absorption coefficient by liquid water increases faster between 11 and 12 μm than between 8 and 11 μm; the opposite is true for ice particles (STRABALA et al., 1994). This feature is the base for the trispectral method.

Cloud ice crystals also present high albedo in the near infrared (NIR) and visible (VIS) channels (BERENDES et al., 2008). The reflectance at a water (or ice) absorbing channel in the NIR is mainly dependent on the cloud particle size (NAKAJIMA; KING, 1990). Therefore, cloud tops with high ice content can be distinguished of water clouds using the $1.6\mu m$ NIR channel, because the absorption of ice is approximately double that of water (ROSENFELD et al., 2004). Therefore, clouds with ice content will appear darker than water based clouds (Figure 2.5).

Figure 2.5 - Water cloud versus ice cloud reflectance spectra.



Simulation using a two-stream radiative transfer code, of directional–hemispherical reflectance spectra of a water cloud, an ice cloud, and a snow surface at an illumination angle of 60°, on the wavelength interval 0.4 - 2.5 μm. The radii of the water droplets and ice crystals are 10 and 40 μm, respectively. The snow is assumed to have grain radii of 200 mm.

SOURCE: Gao et al. (1998).

Rosenfeld et al. (2004) highlighted that this type of cloud also exhibit increased response in the $\approx 3.7\mu m$ channel due to its sensitivity to particles effective radius (r_e). Stronger radiances in NIR channels implies that particles at the cloud top are small, as a consequence of vigorous updrafts. In addition, they suggest an advantage over the NIR one because of a greater sensitivity to cloud top microstructure and smaller surface impact. These features can be exploited to estimate the intensity of storms (ROSENFELD et al., 2008), since strong updrafts are expected to produce cloud-tops full of small ice crystals.

Hail formation, as described in Chapter 1, occurs in the mixed-phase region of the

cloud (between -10 and -25 °C) (ALLEN et al., 2020) . Therefore, hailstorms are usually associated to really low BT in the IR interval.

The hailstone growth is favoured by its permanence under the effect of the storm updraft. In order for this condition to hold, the updraft speed must overcome the terminal velocity of the growing stone. This circumstance coincides with the high vertical development of the storm, causing the cloud top surpass the Tropopause. This dome-like protrusion above the cumulonimbus anvil (GLICKMAN; ZENK, 2000) is named an overshooting top (OT). Therefore, the presence of OT confirms a strong updraft within the storm (DWORAK et al., 2012).

Accordingly, convection strength may also be evaluated through the BT difference between the IR Window channel at $11\mu m$ and the water vapor absorption channel at $6.7\mu m$. When deep convective clouds penetrate the stratosphere, there is a water vapor injection into a region (i.e. the lower Stratosphere) that is warmer then the upper Troposphere. As a consequence, the BT of the $6.7\mu m$ WV channel tend to be higher than BT in the thermal IR window (SCHMETZ et al., 1997). Therefore, the $11-6.7\mu m$ BT is prone to return 0 K or less for convective clouds connected to heavy rain (KURINO, 1997). Some authors use the opposite order in the subtraction (SCHMETZ et al., 1997), in this case, positive values are associated with OTs. In Brazil, BT differences $6.19-10.35\mu m > -10K$ have been considered as an indicative of OT (RIBEIRO et al., 2019).

OTs were firstly identified in VIS imagery as a bright "cauliflower" resemblance (MIKUŠ; MAHOVIĆ, 2013), and cloud top colder than the Tropopause in the IR channels (REYNOLDS, 1980). Bedka et al. (2010) developed an OT detection through a BT spatial gradient in $11\mu m$. It was also outlined the presence of cold rings on OTs (SETVÁK et al., 2010) caused by stratification and wind shear just above the tropopause. This feature was then exploited by Žibert and Žibert (2013) on their algorithm for detecting deep convective clouds using Meteosat data.

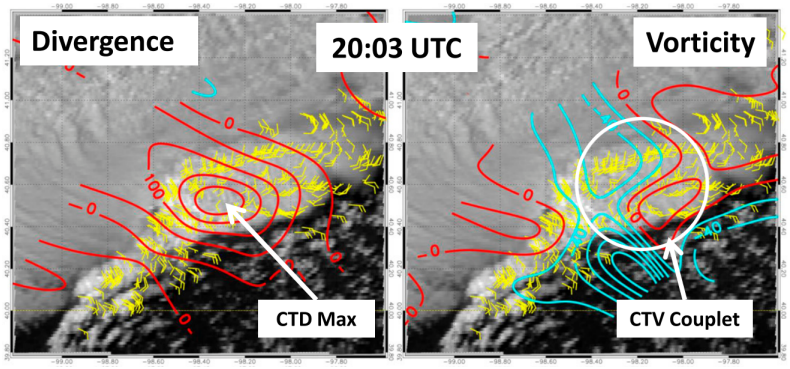
OTs often exhibits a U/V-shaped feature relatively cold BT in IR imagery (NEGRI, 1982), for long recognized as Enhanced-V. This V-shape is often connected to severe weather (ADLER et al., 1983). It was further acknowledged that this feature was owed to a Cirrus plume diverging above the anvil top (HEYMSFIELD et al., 1983) formed on the influx of ice into the lower Stratosphere under strong storm-relative wind shear. Therefore, the phenomenon is now understood as Above Anvil Cirrus Plume (AACP) (HOMEYER et al., 2017). This plume may also display a high reflectance in the NIR channels due to its ice content (LEVIZZANI; SETVÁK,

1996).

Flow field patterns at the cloud tops of deep convection also help diagnose the updraft strength (APKE et al., 2016). Fortunately, cloud motion can be derived using data from IR, WV and VIS satellite channels (HUBERT; JR, 1971; MENZEL, 2001), giving rise to atmospheric motion vectors (AMVs). Brightness temperature differences combinations can be used in the AMVs estimates, providing insights on the correspondent cloud-top microphysics structure motion as well (NEGRI et al., 2014). Thereby, it is possible to identify local wind shear and the evolution of cloud regions in distinct phases.

In addition, Bedka and Mecikalski (2005) demonstrated methods to obtain high-density AMVs, that capture synoptic-scale and mesoscale flow components. This tool, also addressed as mesoscale AMV (mAMV) (BEDKA et al., 2009), was exploited with super rapid scan data to evaluate storm-scale cloud-top horizontal divergence (CTD) and cloud top vorticity (CTV) over deep convection (APKE et al., 2016; APKE et al., 2018). With that, it was stablished that supercell storm cases tend to produce long lived maxima CTD and "couplet" signatures in CTV (Figure 2.6).

Figure 2.6 - Example of CTD and "CTV couplets".



CTD and CTV found through goes super rapid scan visible satellite data of a storm over central Nebraska on 11 May 2014. CTD contoured every 50 3 1025 s21 and (right) CTV contoured every 20 3 1025 s21 with positive (negative) values for both in red (blue).

SOURCE: Apke et al. (2016).

2.1.2.1 Algorithms

The current operational algorithms to hailstorm detection and forecast based on infrared satellite measurements, can be separated in the ones that take only satellite data as input and those that ingest different types of data from several sources (i.e. ground radar, atmospheric soundings and field variables from atmospheric models).

The currently most extensive system in terms of hailstorm detection is the ProbSevere system (CINTINEO et al., 2020b). ProbSevere is a threefold system of models to the hazards alert: (I) severe hail - ProbHail; (II) severe straight-line wind gusts - ProbWind; (III) tornadoes - ProbTor. Its complete functioning involve observation processing; followed by storm identification and tracking; and application of machine learning. The version 2.0 of ProbSevere ingested satellite data from a single source - the GOES-16, together with Multi-Radar Multi-Sensor Doppler weather radar products (MRMS) (SMITH et al., 2016), together with ground-based lightning and Rapid Refresh Model (RAP) output from each hourly forecast cycle from the National Centers for Environmental Prediction (NCEP). After the first two steps of processing, the hazards probabilities are estimated using naïve Bayesian classifiers. In total, ProbSevere works on 4 models: a naïve Bayesian classifier for each hazard, plus a combined model for any of the listed hazards, that takes the maximum probability of the three classifiers.

The Hail Detection Tool (HDT) (MERINO et al., 2014) on the other hand, is a more one-purpose algorithm. It was primarily designed to identify hail-bearing storms using multiespectral Meteosat Second Generation (MSG) data. The processing is done in two phases: the first is a Convective Mask (CM) which detects deep convection; the second is the Hail Mask (HM) which properly identifies hail-bearing clouds among the systems distinguished by the CM. Both phases of the processing are performed through a logistic regression model trained with multispectral MSG data and RGB composites from MSG channels. A physical interpretation of the model parameters after training are found in Merino et al. (2014).

2.1.2.2 Limitations

As seen in this Chapter, OTs have been the main proxy information on determining the probability of hailfall. From a microphysics point of view, strong updrafts are directly connected to hail formation. However, the hailstone starts to melt as soon as it crosses the melting level (ie. the $0°C$ environment temperature) in a downward trajectory. Therefore, the higher the melting level, more likely it is that the hail will

become liquid before touching the ground. For this reason, such methods of hailstorm detection through proxy information still lack of generality and work better over the region they were developed for.

2.2 Gaps in existing knowledge

Despite advancements in satellite meteorology and severe weather detection methods, significant gaps remain in the effective detection and analysis of hailstorms. This section outlines the critical areas where current knowledge and technological capabilities fall short:

Regional Performance Variability: Current leading methodologies for hailstorm detection exhibit inconsistent performance across different geographical areas. This inconsistency arises from varying hailstorm characteristics influenced by regional conditions. Consequently, tools and algorithms developed and validated for one geographic area may underperform or yield inaccurate results when applied to different regions. This discrepancy highlights the need for adaptable or universally applicable detection methods.

Empirical Limitations of Hail Size Diagnosis: The majority of hail detection techniques rely on indirect empirical relationships derived from storm intensity metrics to estimate hail size. These empirical models, developed based on specific regional datasets, may not be universally applicable or accurate, limiting their effectiveness globally or even regionally. This reliance on localized empirical data underscores the necessity for more robust, universally applicable models that can accurately predict hail size and impact under varied atmospheric conditions.

Operational Sampling Strategies and Radar Limitations: The effectiveness of hail detection is also constrained by current operational radar sampling strategies, which generally prioritize scanning at lower atmospheric levels. Additionally, the distance between the radar and the storm can significantly affect the resolution and accuracy of storm features captured. These limitations restrict the comprehensive detection and quantification of hail-producing storm characteristics, particularly at higher altitudes.

Inconsistencies in Methodological Approaches: The use of varied hail detection algorithms, differing radar sites, and inconsistent study periods complicates the quantitative comparison of hail detection methods. For example, Stržinar and Skok (2018) noted significant performance discrepancies among four different hail detec-

tion algorithms in Slovenia, identifying the superior performance of the classical Waldvogel et al. (1979) criteria and the SHI (WITT et al., 1998) over other commonly applied methods. This diversity in methodologies hinders the development of standardized, universally accepted hail detection practices.

Impact of Atmospheric Parameters: The role of atmospheric conditions below the freezing level, particularly in influencing the size and ground impact of hailstones, remains incompletely understood. Dessens et al. (2015) demonstrated the significant influence of melting level height on hail size distribution in France, particularly affecting smaller hailstones. Additionally, factors such as wet bulb zero height (MORGAN, 1970), low-level equivalent potential temperature (MANZATO, 2012), Convective Available Potential Energy (CAPE) (KUNZ et al., 2009), and the temperature and humidity profile (RASMUSSEN; HEYMSFIELD, 1987) are closely linked to the probability of hailstones reaching the surface but are not fully integrated into current predictive models.

Addressing these gaps requires a concerted effort towards developing more adaptable, comprehensive, and universally applicable hail detection methodologies. Enhancing our understanding of regional storm dynamics, standardizing methodological approaches, and integrating critical atmospheric parameters into predictive models will potentially address some of these gaps and advance the capability to detect, analyze, and respond to hailstorms more effectively.

3 METHODOLOGY

3.1 Study area

Our main interest is in the hail generating storms over Brazil. However, in order to better depict the behavior of those systems, the study domain boundaries are extended to track storms from origin until their dissipation. Then, the complete domain of the present study is a subset of South America, as shown in Figure 3.1. The northern latitude boundaries were chosen to intentionally introduce noise by including regions with well defined periods of intense convective activity but less frequent hailstorms (ANDREAE et al., 2004). There is no intent of covering large ocean areas.

Figure 3.1 - Spatial domain of study.

Rectangle with upper-left corner 12.54° N and 88.02° W, and bottom-right corner 46.50° S and 26.22° W, representing the bounds of the CSs tracking task.

SOURCE: The author.

3.2 Data collection

Empirical research is strongly dependent in the robustness and relevance of its data sources. In the quest to develop a predictive model for classifying hailstorm occurrences, this study leverages a combination of satellite measurements and crowd-sourced storm reports. The data collection process is bifurcated into two primary subsections, each critical to the formulation and validation of the classification algorithm.

3.2.1 Hail reports

The present study relies in crowd-sourced hail reports for ground-truth information on hailstorms over Brazil. The reports here were extracted from the Prevots dataset (RIBEIRO et al., 2022). Prevots data gathers severe weather cases over Brazil from social media posts, newscasts and volunteer storm spotters since June 2018. Those cases are verified by manually checking radar and/or satellite data of the approximate reported time of the event. In cases in which there is a lack of time of the report, it is then estimated following the approach of the European Severe Weather Database (ESWD) also by evaluating radar and satellite measurements available.

This dataset consists of three categories of severe weather reports (i.e. wind-gusts, hail and tornado) organized as tabular data. Each case holds a quality flag - related to the source of the report - date and timestamp in UTC, approximate time uncertainty in minutes - due to the gap between successive satellite and radar measurements - an approximate set of coordinates of the event and an uncertainty in kilometers. In cases in which the prospected report held no specific location but the event is confirmed, then the coordinate relative to the city center is annotated and a spatial uncertainty is added.

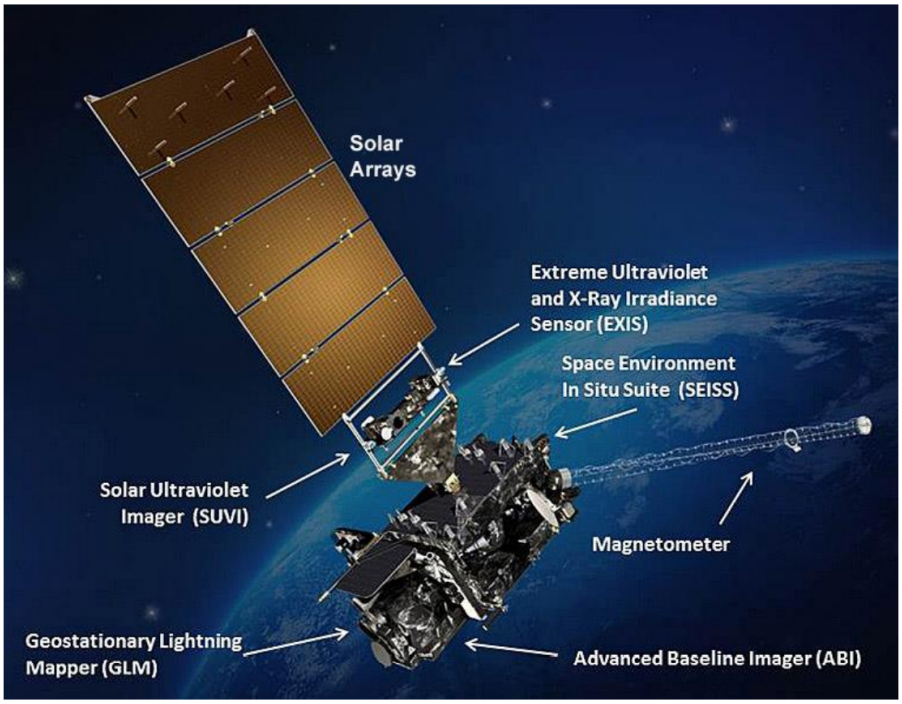
Some additional information includes the city and state of the event report, a measure of intensity of the class of severe weather and an indication on the source of the report. In the case of hail reports, the measure of severity is the approximate hail size reported. This size does not necessarily represent the maximum hail size produced by the storm, since it is known that hail reports tend to underestimate the hail size produced by a thunderstorm by nearly a factor of 2 (BLAIR et al., 2014), with differences increasing the larger the hail.

3.2.2 GOES-16 satellite

The GOES series brought in 1975 the era of of multi-channel geostationary observations (DAVIS, 2007). Since then, there have been continuous improvements in the spectral coverage and the resolution of observations (MENZEL; PURDOM, 1994). The tradition of cooperation between research centers made those platforms one of the main sources of information for meteorological services in Brazil for over 3 decades (COSTA et al., 2018).

GOES-16 is the first satellite on the GOES-R series and was positioned at the GOES-EAST location centered over 75 degrees West longitude over the Western Hemisphere in November 2017. The complete scientific payload comprises six instruments, schematically presented in Figure 3.2. For the purposes of this project, special attention is brought to the Advanced Baseline Imager and the Geostationary Lightning Mapper sensors.

Figure 3.2 - Schematic view of GOES-R Series platforms with its payloads.



SOURCE: Benz et al. (2019).

3.2.2.1 Advanced Baseline Imager

The ABI is a state-of-the-art 16-band radiometer (SCHMIT et al., 2017) on board of a geostationary platform. Its spectral bands cover the visible, near-infrared, and thermal IR portions of the electromagnetic spectrum, as presented in Table 3.1, in a similar fashion to the Advanced Himawari Imager (AHI) on board of Himawari-8/9 (DA, 2015), except for the Cirrus and veggie Bands.

Table 3.1 - ABI Spectral Attributes.

Band	**Central λ** μm	**Subpoint pixel spacing (Km)**	**Description**
1	$\sim$ 0. 47	1	Blue
2	$\sim$ 0.64	0.5	Red
3	$\sim$ 0.864	1	Veggie
4	$\sim$ 1.373	2	Cirrus
5	$\sim$ 1.61	1	Snow/Ice
6	$\sim$ 2.24	2	Cloud Particle Size
7	$\sim$ 3.9	2	SWIR Window
8	$\sim$ 6.19	2	Upper-Level WV
9	$\sim$ 6.93	2	Mid-Level WV
10	$\sim$ 7.34	2	Lower WV
11	$\sim$ 8.44	2	C_{top} Phase
12	$\sim$ 9.61	2	Ozone
13	$\sim$ 10.33	2	Clean Window
14	$\sim$ 11.21	2	Window
15	$\sim$ 12.29	2	Dirty Window
16	$\sim$ 13.28	2	CO_2

SOURCE: Adapted from Schmit et al. (2017).

Besides those improvements on the spectral aspects compared to the previous generations, another relevant advance for the study of fast developing atmospheric phenomena is the current full-disk scan every 10 minutes capability. The ABI is also able to perform mesoscale measurements with 1-km resolution through a Super Rapid Scan (SRSOR) mode, every 30 seconds to 1 minute (SCHMIT et al., 2013). Those refinements directly empower severe weather monitoring applications.

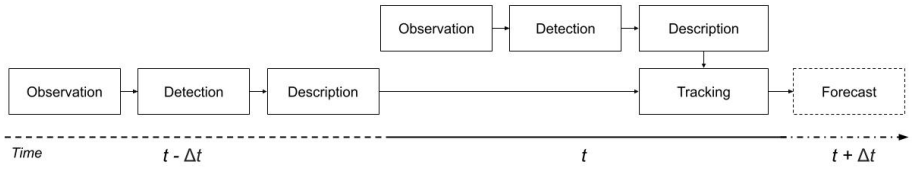
The use of ABI bands in the current study is restricted to the ABI Level 2 Cloud Moisture Imagery Product Full Disk (GROUP; PROGRAM, 2017) (ABI-L2-CMIPF, hereafter called CMIPF).

3.2.3 Database of convective storms

The data collection part of the current thesis involved the tracking of clusters of deep convection that would potentially generate hail-fall. In order to do that, we have leveraged the power of a tool for Tracking and Analysis of Thunderstorms, namely TATHU (UBA et al., 2022). TATHU was developed at INPE's Meteorological Satellites and Sensors Division (DISSM) aiming at enabling research and development in the field of Convective System (CS) tracking and the associated research. The tool handles this as a Multi Target Tracking problem (MAKRIS; PRIEUR, 2014), in which the algorithm is expected to estimate the position of a varying number of targets over time.

TATHU consists in a deterministic CS tracking method based on target overlap. That is, the CSs are detected based on a brightness temperature threshold passed by the user. Every labeled system is then characterized based on the aspects within the CS's bounds. This process is repeated in every observation, as shown in Figure 3.3, thereby, the CSs are tracked further in time, based on a selected area overlap criterion. Additionally, TATHU offers the flexibility to be set to work like the ForTrACC toolset (i.e. Forecast and Tracking the Evolution of Cloud Clusters) (VILA et al., 2008).

Figure 3.3 - General Stages of TATHU's methods for detection, characterization and tracking of targets.

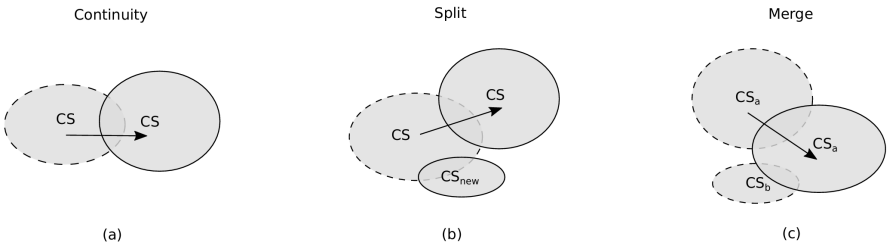


SOURCE: Uba et al. (2022).

TATHU provides a high-level compact representation of detected storms by treating them as an object (**ConvectiveSystem** class). Each of those objects is geographically bounded and holds its spatial limits as an attribute (**geom**). Besides that, the objects also retain some user-selected raster statistics in its fields and a unique universal identifier (**uuid**) that remains unchanged for the CS's lifetime.

The origin and lifelong steps of a system across several subsequent satellite measurements are annotated as well (**event** attribute). The first time a system is detected in a time-frame, that is, there is no overlap with a previous one, it is identified with a value of *spontaneous generation*. Every step further in time that the system persists, it will be tagged as *continuity*, unless the independent target evolution is violated. For this condition, an additional asset is that TATHU keeps track of dissociative and associative events between CSs, as shown in Figure 3.4. In cases where one CS overlaps with more than one in a subsequent frame, the larger area overlap determines the one that will continue the original CS life cycle, and the others intersected are then tagged as events of *split*. When the opposite occurs, more than one CS intersect with only one in a subsequent time frame, then the latter receives the *merge* tag.

Figure 3.4 - Schema of CS's possible tracked events.



The dashed hulls indicate a previous state detection and the filled shape represents the current one. The arrows point out the evolution in time for cases of a) **continuity**, b) **split** and c) **merge**.

SOURCE: Adapted from Uba et al. (2022).

Here, the storms were identified and tracked based on ABI's clean IR window brightness temperature (i.e. 10.3 μm) disposed in a regular 2 Km x 2 Km Lat-Lon WGS84 grid. This re-projection is performed by TATHU's built-in functionality *sat2grid* leveraging GDAL's Nearest Neighbour resampling. In order to detect CSs, the detector was set for systems with BT under 235 K. This threshold is likely to naturally select systems with radii larger than 72 Km (MACHADO et al., 1992). This threshold could also neglect thin Cirrus clouds associated with the CS, due to their warmer apparent BT as a function of its transparency to the IR radiation from below. Though the BT threshold for detection might be directly related to the CS size, Machado et al. (1998) used a 245 K thrshold and found no significant difference in the size-lifetime relation between predominantly continental CS that they analized

and the ones studied by Chen et al. (1996) using 208 K as BT threshold. Overall, the values of choice should be suitable for identifying clouds associated with convection in different regions of the study domain (GARREAUD; WALLACE, 1997; CARVALHO; JONES, 2001; LAURENT et al., 2002; MACHADO; LAURENT, 2004; RIBEIRO et al., 2019). In order to reduce TATHU's sensitivity to smaller scale systems, an area of at least 3000 Km^2 was defined as threshold for detection. This minimum area threshold is considered a conservative choice in terms of CSs characterization over Brazil (VILA et al., 2008), which intrinsically excludes short-lived smaller-scale storms from the analysis. Together, these parameters are likely to make the detector overlook the early (e.g. feeder cells before merging into the storm) and fading stages of a CS, when deep convection is not well developed yet or no longer present. For a deeper discussion on how these parameters are associated with the BT population inside a convective storm and their correspondence with precipitation rates, we refer the reader to the work of Machado et al. (1998), specifically the discussion therein about the top-panel of Figure 1.

A raster statistics analysis was performed in every detected CS, with parameters of BT's mean, minimum, standard deviation and count of data points inside the CS polygon (i.e. proxy value for area). In terms of tracking, the overlap area criterion was set to 10%, a default value among TATHU tutorials, which requires that the detected system has at least 0.1 rate of its area intersected by a system in a previous frame in order to be considered as part of the previous system's life cycle. These settings are summarized in Table 3.2.

Table 3.2 - Parameters set used for storm tracking using TATHU.

Parameter	**Value**
Threshold	235 K
Minimum Area	3000 km^2
Overlap Criterion	10%
Detector	*LessThan*
Resolution	2 Km
Descriptors	[*mean*, *min*, *std*, *count*]

SOURCE: The author.

The storm cells were tracked between January 1, 2018 and December 31, 2023. This 6-year dataset was exported in yearly databases with support for geospatial data, using the SpatiaLite data format (i.e. SQLite engine with spatial functions)

(FURIERI, 2023). This data format was preferred for enabling spatial data queries to the dataset which would further reduce cumbersome tasks of storm search in the registries. Besides that, there is a prompt integration for I/O in TATHU of this kind database.

3.3 Machine learning approaches

According to Abu-Mostafa (2012), the essence of Machine Learning problems lies in three conditions:

a) A pattern exists;

b) One cannot pin it down mathematically;

c) There is data on it.

In problem of detecting hailstorms from satellite data, some patterns are human detectable. For instance, a forecaster would search in a scene for the features described in Section 2.1.2. However, the procedure of solving this analytically, as shown in Section 2.1 can be cumbersome and uncertain. Fortunately, there is data on it, as the study domain has been continuously and consistently, observed for a long period (COSTA et al., 2018).

So far, there is no analytical solution for the task in hand, but there is data to construct an empirical solution. Since some cloud features associated with hailstorms are known, the problem could then be tackled with heuristics or handcrafted rules for distinguishing systems prone to be associated with hailfall. However, this practice can be cumbersome and lead to an emergence of rules and subsequent exceptions. Therefore, we aim to use ML tools to leverage the set of observations available to uncover the underlying process of detectable storms likely to cause a hailfall.

3.3.1 Theoretical basis for machine learning applications

The problem is posed as the following: every possible input to our model is held within an input space $\mathcal{X}$ in $\mathbb{R}^d$ (d here correspond to the amount of features in our dataset) and the output space $\mathcal{Y} = \{0, 1\}$, which refer to No-hail and Hail classes of storm, respectively. We assume there exists an **unknown** *target function* $f : \mathcal{X} \mapsto \mathcal{Y}$ that would map every possible input into their respective class. The model's task is then to come up with a function g so that for $x \in \mathcal{X}$, $g(x) \approx f(x)$, that is, for most of the inputs it behaves like the *target function*.

The *target function*, as stated, is not known. However, we have an idea of its behavior from previous assessments of hailstorms. Therefore, we consider this inexact foreknown information like a *target function* with noise, and change the notation from now on to an **unknown** *target distribution* $P(y|x)$. As a reference to the previous section, the underlying process is this distribution.

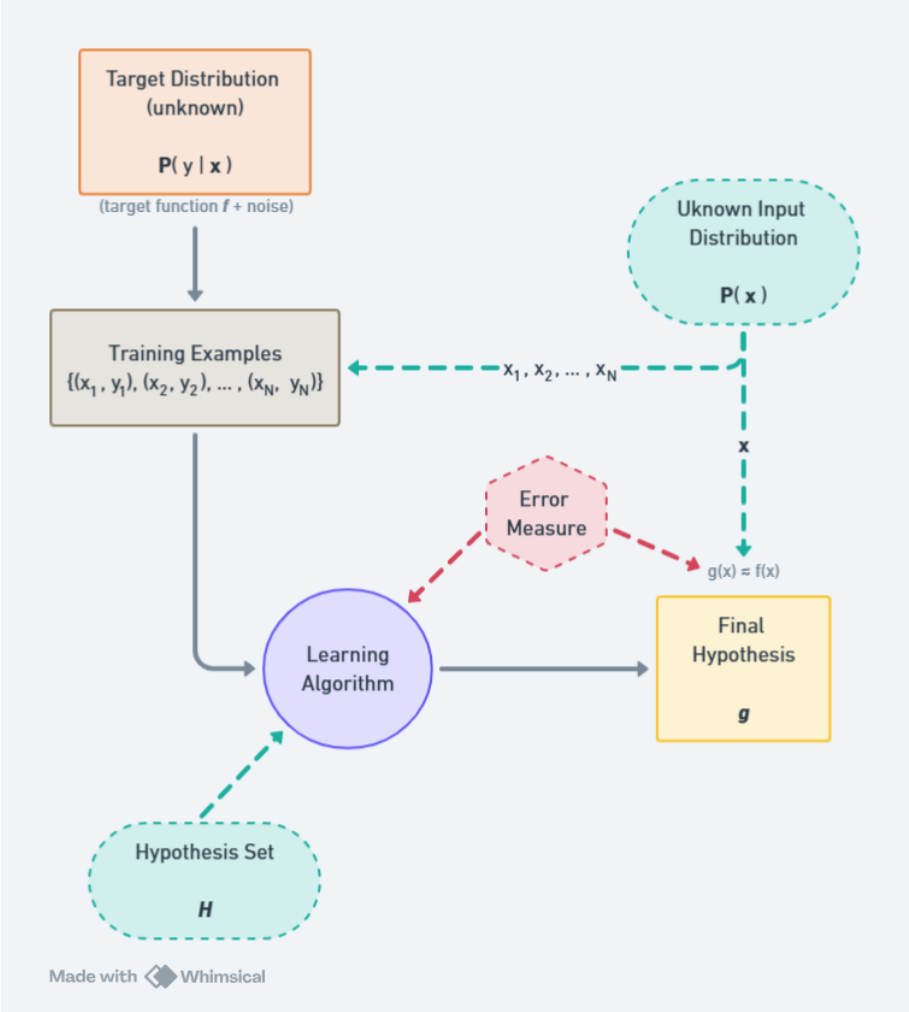
For the current study, we have set of observations with their expected output. This is basically a sample of size N drawn from the joint distribution $P(x, y) = P(x)P(y|x)$, which constitutes the *training set* $\mathcal{D} = \{(x_1, y_1), (x_2, y_2, ..., (x_N, y_N))\}$. This will be used as a reference to tune the parameters of an adaptive model. This approach is known as a *supervised learning* problem and involves rounds of model adjustments to try to minimize a cost function, a process known as *training*.

The training process can be thought of the task of reaching the function g from a set of Hypothesis $\mathcal{H}$. $\mathcal{H}$ comprehends every hypothesis the learning algorithm can generate. Therefore, its size depends on the complexity of the learning algorithm. Every $h \in \mathcal{H}$ assigns different weights to the distinct coordinates of $\mathcal{X}$. The full learning setup is displayed in Figure 3.5. In this setup, the output is categorical and encoded by discrete numbers. This is known as a *classification* task. However, in some problems, the output is better encoded by continuous variables, which would constitute a *regression problem*.

In most applications, the original input undergoes a pre-processing before being properly used. At this stage, the data can be transformed into some new space of variables in which it could easier for the pattern recognition problem to be solved. For instance, one may apply some sort of scaling to the original information (e.g. normalizing the values) or translate it into a fixed format (e.g. subsampling an original image into equally sized patches). This process is described in the literature as *feature extraction* (GÉRON, 2019; BISHOP, 2006). The aim is to find useful features that are fast to compute and simultaneously preserve useful discriminatory information.

Depending on the variability of the possible input, it can happen that the training sample only comprises a small fraction of all possible values in the population. However, after the training, it is expected that the obtained function is able to *generalize*. In other words, it should be able to handle new examples in a *test set* without much loss in performance to assign them to their respective categories (represented the same way as in the training set).

Figure 3.5 - Full learning setup.



The full process of supervised learning is illustrated from top to bottom. The underlying process we want to mimic with the model, generates the target distribution (that can be thought as the target function + noise). In order to do that, we use as a reference a sample of (x,y) drawn from the joint distribution $P(x, y) = P(x)P(y|x)$, as training examples. The learning algorithm uses this reference set to assess how far its hypothesis are from the target function. The hypothesis from $\mathcal{H}$ that emulates the target function the best will be the final one.

SOURCE: Adapted from Abu-Mostafa (2012).

Some tasks require no prior known corresponding targets, namely *unsupervised learning* problems. This normally involves the discovery of groups holding some similarity (i.e. *clustering*), or to determine the distribution of data within the input space, known as *density estimation*, or even to project the data from a high dimen-

sional space down to 2 or 3 dimensions for the sake of visualization. In the present study, unsupervised learning was also used in results interpretation and model analysis.

This task introduces three sources of uncertainty: namely, (I) an inherent stochasticity as the distribution of hail intra-cloud is itself probabilistic and the spatial distribution of hailstorms in the study domain is biased in its nature; lack of certainty also emerges due to (II) incomplete observability, because the input to the algorithm does not represent a direct observation of the variables that drive the system's behavior, as described in Section 2.1.2; and (III) the system is not completely modelled, since it is expected some level of linear dependency among the information provided to the model. Those aspects can constrain the performance of classification executed under deterministic rules.

The criteria typically considered in selecting an algorithm for a specific application hinge on several key factors, including efficiency, accuracy, interpretability, and scalability.

Efficiency refers to the computational cost required to train the model and make predictions. In real-world applications, especially those requiring real-time analysis, the efficiency of an algorithm can be a critical factor. Algorithms that require less computational power and time are preferable for such settings.

Accuracy is key in the performance of a machine learning model. It measures the model's ability to correctly predict the outcome. High accuracy is especially crucial in applications where misclassification can lead to significant consequences, such as medical diagnosis or weather prediction. The trade-off between accuracy and model complexity often guides the selection process, aiming for the highest accuracy with the least complexity.

Interpretability involves the ease with which a model's decisions can be understood by humans. For many applications, understanding why a model has made a specific prediction is as important as the prediction's accuracy. Models that offer clear insights into their decision-making processes are highly valued.

Scalability pertains to the model's ability to handle increasing amounts of data or features without a significant decrease in performance. As the volume of data continues to grow, the ability of an algorithm to scale effectively becomes a critical consideration. This includes both horizontal scaling (adding more data points) and

vertical scaling (adding more features).

In addition to these criteria, the selection of a machine learning algorithm may also consider the specific characteristics of the data, such as its dimensionality, the presence of noisy or missing data, and the type of relationship (linear or nonlinear) between features. No single algorithm excels in all scenarios, making it essential to evaluate each algorithm's strengths and weaknesses in the context of the specific problem being addressed.

Furthermore, the choice of algorithm can be influenced by the model's ability to generalize from the training data to unseen data. This generalization capability ensures that the model remains accurate and reliable when exposed to new, similar data. Techniques such as resampling methods are commonly employed to enhance a model's generalization performance.

3.3.2 Algorithm selection and rationale

When evaluating different models or algorithms, it's crucial to estimate how they will perform on unseen data. This aids in selecting models that are not just overfitting the training data but will generalize well to new data. In order to assess that, we define two types of error: $E_{in}(h)$ which indicates how much one of the model's hypothesis differ from the underlying *target function* inside the training sample; and $E_{out}(h)$ which is how different the distribution generated by the hypothesis is, compared to the target distribution (i.e. $P[h(x) \neq f(x)]$). Those measures are tied by a Hoeffding Inequality in Equation 3.1 (HOEFFDING, 1994) (for details on this proof, we refer the reader to Anguita et al. (2012)).

$$P[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon^2 N} \tag{3.1}$$

Where M is the size of $\mathcal{H}$, N is the amount of training examples and ϵ is a tolerance threshold.

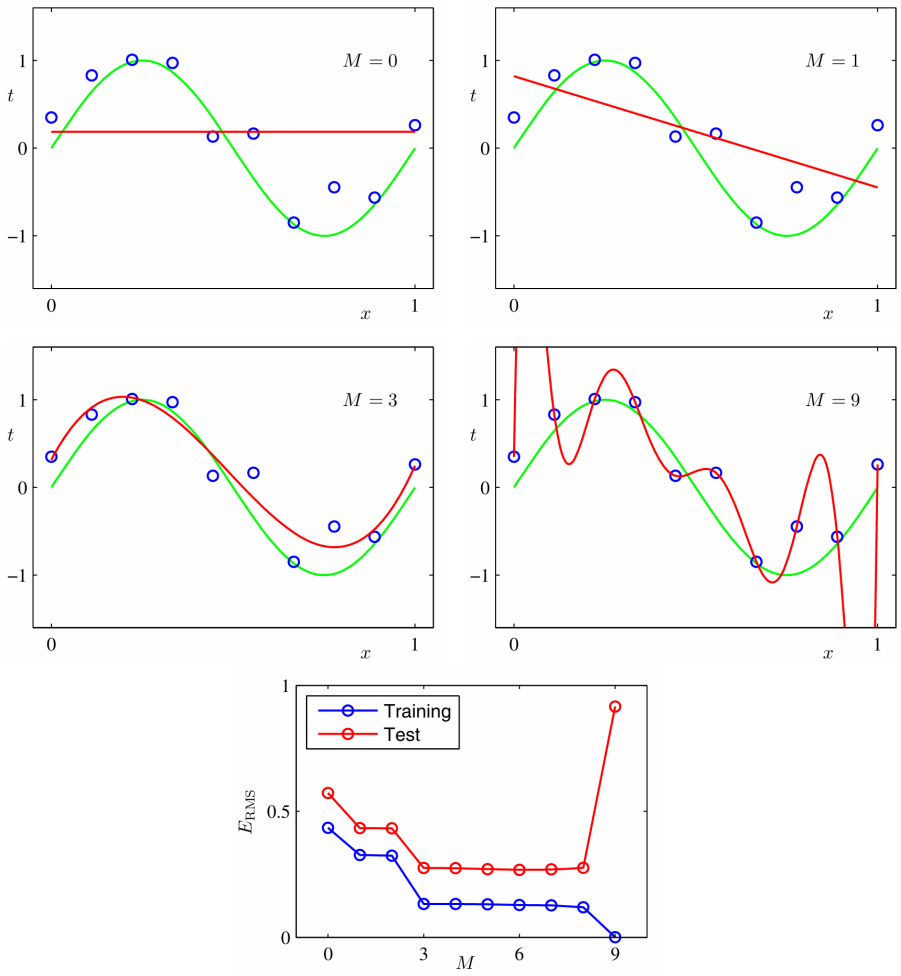
This inequality states three main things: the main one is that a larger M (i.e. model complexity) gives more flexibility for achieving a hypothesis that will efficiently minimize E_{in} but will also increase the term on the right and potentially extinguish the bound to E_{out}; secondly, a large set of training samples (large N) can tighten both errors by forcing a reduction in the exponential; and lastly, a relationship between those measures based solely on the tolerance threshold ϵ, gives that the chance of very large deviations from E_{out} (large ϵ) are small (negative exponentiation of the

tolerance squared) and small deviations from E_{out} are likely, which basically tells nothing.

This means that a complex model can generate more hypothesis, which increases the chance of it reproducing well the behavior of the training examples. However, it is likely that it will fail in handling unseen examples, which is a case of *overfitting* or lack of generality, as depicted in the lowermost panel in Figure 3.6. In small terms, it suggests that the model fitted the training sample so tightly that it will only work under the circumstances inscribed in the specific sample of the population used for training. This behavior can be observed in the red curve in Figure 3.6 using a polynomial of order 9. The desirable fit, in this example, is achieved with a leaner model polynomial curve of degree 3. This level of freedom was enough to depict well the original function behavior and is likely to return outputs for unseen observations.

On the other hand, if the chosen model is too simple for the task at hand, then it is likely to *underfit*. It means that the function needs more liberty to approximate the real behavior of the data, or the sampled features were not enough to depict the underlying patterns in the population of the data. This process is displayed in the upper 2 panels in Figure 3.6, where polynomial curves do not depict the data behavior well enough.

Figure 3.6 - Examples of different polynomial approximations to a function.

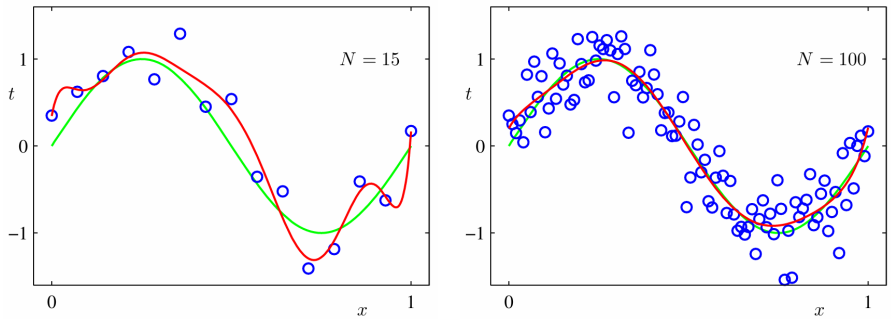


Upper 4 plots: training dataset of 10 points (blue circles), each comprising an observation of the input variable x along with the corresponding target variable t. The green curve shows the function sin(2πx) used to generate the data with a little perturbation. The red curves represent the fit of polynomials of order M to the data, by minimizing the cost function $E(W) = \frac{1}{2}\sum[y(x_n, W) - t_n]^2$, where W is the model's set of parameters used in the prediction; $y(x_n, W)$ is the model's prediction for observation x_n; t_n is the observation's known target value. The two upper plots display examples of underfitting, the lower left one shows the best fit and the lower right one is an overfit. Lowest panel: Plot of the root-mean-square error, defined by $E_{RMS} = \sqrt{\frac{2E(W*)}{N}}$, evaluated on the training set and on an independent test set.

SOURCE: Bishop (2006).

From the inequality one can also find that a larger set of observations is needed in order to afford an increase in model complexity. In Figure 3.7 this effect is demonstrated by repeating the steps in the previous example (Figure 3.6) with an expanded dataset to fit. This shows that for a given model complexity, the over-fitting problem become less severe as the size of the data set increases.

Figure 3.7 - The impact of increasing the size of the set of observations.



Plots of the solutions to the same problem in Figure 3.6, obtained by minimizing the sum-of-squares error function using the M = 9 polynomial for N = 15 data points (left plot) and N = 100 data points (right plot). Increasing the size of the data set reduces the over-fitting problem.

SOURCE: Bishop (2006).

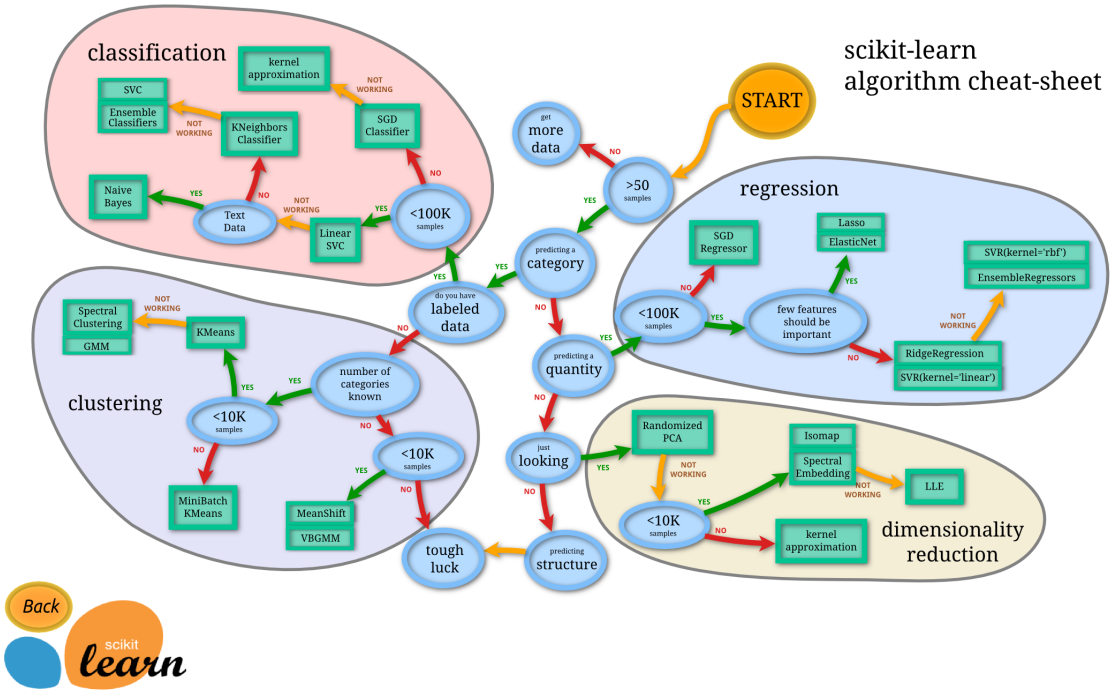
In the conduct of this study, a deliberate decision was made to prioritize interpretability over flexibility in our machine learning model selection. This choice was grounded in the necessity to address our core research questions comprehensively. Interpretability is crucial, not only for understanding the model's decision-making process but also for elucidating the environmental factors under investigation. This approach is further justified by the need to adapt to the constraints presented by our dataset's size.

Given the limited volume of data available for training, it was imperative to consider the potential risk of model overfitting. Employing a model with excessive complexity could lead to a scenario where the model performs exceptionally well on training data but fails to generalize to new, unseen data. To mitigate this risk and make an informed decision regarding the model's structure, we adopted a decision roadmap strategy. This roadmap guides the selection process by considering the training set's

size and the desired type of prediction, ensuring that the chosen machine learning model aligns with our data availability and research objectives.

Figure 3.8 illustrates this decision-making process, highlighting the balance between model complexity and data adequacy. This visual representation aids in understanding how model form is determined by the nuanced interplay between the volume of training data and the specificity of prediction tasks, ensuring that the final model selection is both data-conscious and aligned with our study's interpretative goals.

Figure 3.8 - Roadmap for choosing the Machine Learning algorithm.



SOURCE: Developers (2024).

To accurately gauge the efficacy of our model, it is essential to employ a measure of error. However, given the necessity of generalization — the model's ability to perform accurately on previously unseen data — it becomes necessary to structure each experimental iteration to include both a training set and a test set. The training set is instrumental in adjusting the model to fit the data, thereby minimizing the in-sample error (E_{in}). Conversely, the test set, which is randomly selected and consists of data not previously exposed to the model, serves as the basis for evaluating the

model's capability to generalize, hence assessing the out-of-sample error (E_{out}).

Some of the predictive algorithms utilized in this study are predicated on parametric methods, which entail presupposing the functional form of the target function. The process of training the models is thus focused on optimizing the parameters within this predefined structure. The efficacy of our choice in model form is then evaluated based on its performance against the test set, inherently introducing a bias towards minimizing E_{out} due to our commitment to a specific functional form. To counteract this and more rigorously assess the model's ability to generalize, we have instituted an additional step in our methodology. This involves the creation of a validation set — a dataset of unseen data designated exclusively for evaluating the model's generalization capabilities following the final selection in every experimental iteration.

In the following sections, every model type used in the current study is discussed in depth.

3.3.2.1 Linear model for classification

Linear regression is a method commonly applied to generate predictions as continuous variables. However, in the present study, as seen, the output is categorical.

Linear methods used for classification first predict the probability that the observation belongs to each of the categories of a qualitative variable, as the basis for making the classification. This approach only accommodates two classes, which is suitable for our case.

Two types of linear models were used in the further described experiments, Logistic Regression (Logreg) and a linear Support Vector Machine (SVM) with stochastic gradient descent training. The LogReg was one of the choices because it is one of many functions able to compress the input in values in the range $[0, 1]$. It takes the following form:

$$P(x) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 x}}$$

, where $\beta_{0,1}$ are the adjusted parameters. (3.2)

This model is fitted through a *maximum likelihood* method. That consists of seeking

estimates for β_0 and β_1 so that the predicted probability for a given input returns a value as closely as possible to its respective label. In our setting, this would mean returning a predicted probability close to 1 for observations relative to hail class, and close to zero to the rest. For a given set of parameters, the likelihood is given by:

$$L(\beta) = \prod_{i=1}^{N} P(y_i|x_i)^{y_i} \dot{(} 1 - P(y_i|x_i)^{(1-y_i)}$$

, where N is the number of observations, y_i is the observed class of the ith observation. (3.3)

Since the product of probabilities can be very small and thus numerically unstable, the model fitting process instead maximizes the log of the likelihood function, which turns the products into sums, making it easier to work with. This is equivalent to minimizing the negative log-likelihood. This negative log-likelihood is the cost function in the context of logistic regression, and it is given by:

$$J(\beta) = -log(L(\beta)) = -\sum_{i=1}^{N} [y_i log(P(y_i|x_i)) + (1 - y_i) log(1 - P(y_i|x_i))] \tag{3.4}$$

In essence, this cost function penalizes predictions that are confident and wrong more heavily than predictions that are uncertain or confident and right, encouraging the model to make correct predictions with high confidence.

3.3.2.2 K-Nearest Neighbors classifier

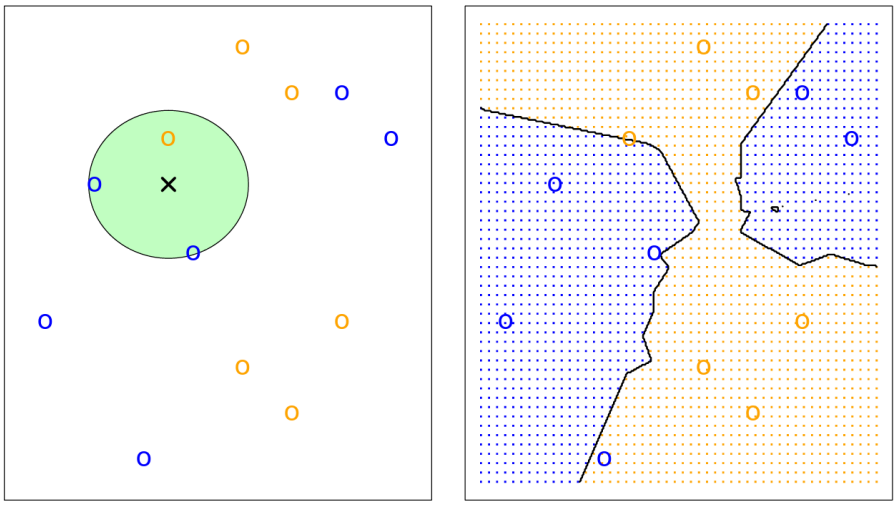
The K-nearest neighbor (KNN) classifier works by assessing K points in the training data, closest in space to the new observation, namely a set $\mathcal{N}_0$. Then, it estimates the conditional probability for a class c as a fraction of points in $\mathcal{N}_0$ whose output is c. This process is illustrated in Figure 3.9.

$$\mathrm{P}(\mathrm{Y} = \mathrm{y}_0 | \mathcal{X} = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = c)$$

$$, \text{ where } \begin{cases} I = 1, & \text{if } y_i = c \\ I = 0, & \text{otherwise} \end{cases}$$

(3.5)

Figure 3.9 - The K-Nearest Neighbour Approach.



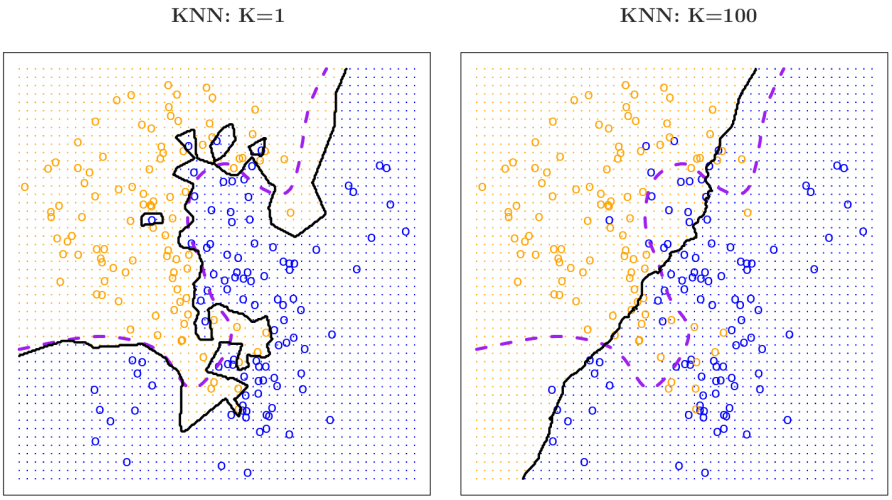
An illustration of the KNN approach, with $K = 3$, with a training set of 6 data points for each class encoded by the color of the circles. **Left**: an input indicated by the black x and the zone covered by the 3 nearest neighbors to the input (green filled circle). The input belongs to the most common class in the zone, blue. **Right**: the boundary decision produced by $K = 3$.

SOURCE: James et al. (2023).

The decision boundary for the algorithm is highly dependent on the choice of K. A smaller value, will cause a very flexible border to the classes, which leaves the model very susceptible to smaller perturbations in the training set (higher variance). A larger value will otherwise make the decision boundary very rigid, not letting the

model fit too much to the data. This difference in flexibility is displayed in Figure 3.10.

Figure 3.10 - KNN Decision Boundary Dependency.



Decision boundary flexibility as a function of the choice of K, shown by the black continuous line. The Bayes decision boundary, as a reference for a lower error rate, represented by the dashed purple line.

SOURCE: James et al. (2023).

3.3.3 Gaussian Naive Bayes method

Naive Bayes classifiers are a family of simple *probabilistic classifiers* based on applying Bayes' theorem with independence assumptions (naive) between the features.

The Gaussian Naive Bayes (GNB) algorithm extends the Naive Bayes framework to accommodate continuous data and assumes that the continuous values associated with each feature are distributed according to a Gaussian (normal) distribution.

Given a dataset with n features, $X = (x_1, x_2, ..., x_n)$, and a class variable Y, Gaussian Naive Bayes assumes that the likelihood of the features X given the outcome Y follows a Gaussian distribution:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

where μ_y is the mean of feature x_i for class y, and σ_y^2 is the variance of feature x_i for class y.

During training, the model computes the mean (μ) and variance (σ^2) of each feature x_i for each class y. This computation forms the model which can be used to make predictions.

For prediction ('predict' method), given a new set of features, X', the model uses the Bayes theorem to estimate the probability $P(y|X')$ for each class y. The model predicts the class with the highest posterior probability.

While the assumption of feature independence is rarely true in real-world data, the model can still perform well in practice. Due to its simplicity, it is very fast, making it suitable for large datasets and real-time predictions. Additionally, the model's simplicity also makes its predictions easy to interpret, which can be valuable in applications where understanding the decision-making process is important.

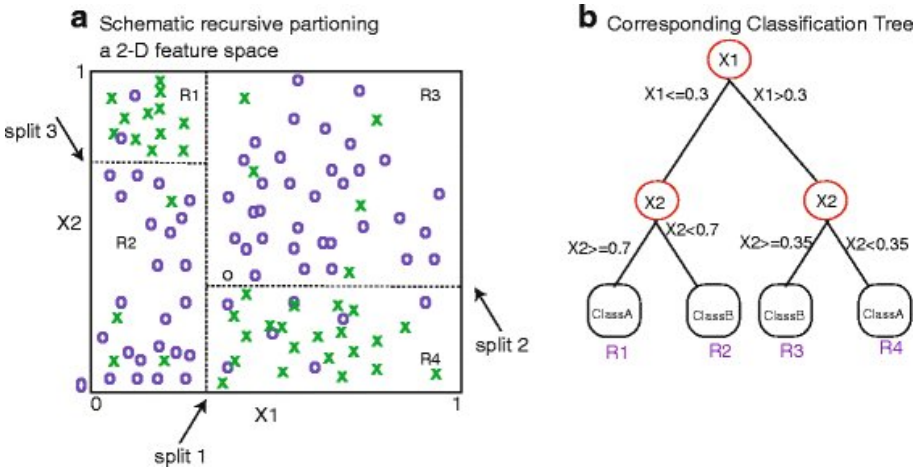
This kind of model is specifically designed for continuous data and automatically applies the Gaussian distribution to model the data. This makes it a convenient choice for datasets where features are continuous and approximately normally distributed.

3.3.3.1 Tree-based methods

This section delves into the use of tree-based methods for classification, particularly focusing on their application within geospatial and meteorological data analysis. These methods are favored for their ability to handle non-linear relationships and complex interactions between variables without necessitating data transformation.

Tree-based methods involve stratifying the feature space into simple non-overlapping regions. The prediction for a given observation is based in the region to which it belongs. The set of splitting rules used to segment the feature space are summarized in a tree structure, namely a decision Tree. Therefore, these resulting regions are known as the *leaves* of the tree. Figure 3.11 displays an schematic view a decision tree splits in the space generated by 2 features, and the equivalent tree structure.

Figure 3.11 - Decision Tree Scheme.



a) A simplified example of a 2-D model space for features X1 and X2. The recursive splitting of the feature space identifies three divisions in a sequential manner to minimize the prediction error for an outcome, Y. **b)** The corresponding decision tree displays the split nodes. The terminal nodes represent non-overlapping regions. The label given to a region is simply the majority class of the outcome, Y, in the region. Predictions for a new sample are obtained by tracing the new sample down the decision tree into a terminal region with a given label.

SOURCE: Hao and Blair (2016).

The splitting process takes a top-down *greedy* approach (JAMES et al., 2024), as it begins at the top of the *tree* and successively splits the feature space evaluating the best split at that particular step. This procedure consists in recursively selecting an observation feature X_j and a cut-point s, and binary dividing the feature space into $\{X|x_j < s\}$ and $\{X|x_j \geq s\}$. This continue until a stopping criterion is reached, for instance, the maximum number of observations per leaf. In the problem setting of the current study, the splitting is guided by the minimization of node purity. Here, this is measured through Gini index:

$$Gini = \sum_{c=1}^{C} \hat{p}_{mc}(1 - \hat{p}_{mc}) \tag{3.6}$$

, where $\hat{p}_{mc}$ is the proportion of training observations in the m-th region that are from the class c. That means that proportions close to 0 or 1 will minimize this cost function.

In the end, the prediction is based on the most commonly occurring class of training observation in the region of the split into which it falls.

Single decision trees provide a straightforward and interpretable model of decision-making processes. However, they possess a notable drawback: a lack of robustness to changes in the dataset. This characteristic introduces a high variance, where slight variations in the training data can lead to substantially different tree structures. Consequently, single decision trees can be unstable and unreliable for predictive modeling.

To counteract the high variance inherent in single decision trees, ensemble methods such as **Random Forest** and **XGBoost** have been developed. These techniques combine the predictions of numerous *weak learners*, that separately may lead to mediocre predictions, but together may enhance accuracy and stability of the outputs.

3.3.3.1.1 Ensemble methods

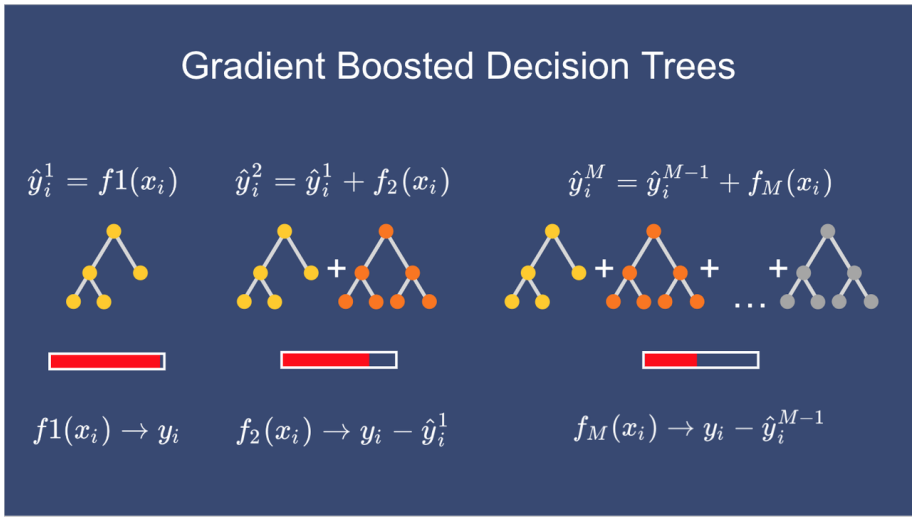
The **Random Forest** algorithm generates a forest of decision trees during the training phase and outputs the mode of the classes for classification tasks. It is particularly lauded for its I)**Capability to Process High-Dimensional Data**, as it efficiently handles hundreds to thousands of input variables without the need for variable deletion; II) **Resistance to Overfitting**, because the method of averaging multiple regression trees significantly reduces the risk of overfitting, making it well-suited for complex datasets, including those derived from geospatial and meteorological observations (SONG et al., 2023).

There is a difference in the building process of the trees that will compose the *forest*, in comparison to the single-Tree method. Firstly, during its construction, each individual tree is only able to "look" at a randomly chosen subset of the training observations, a process known in the literature as *bootstrapping* of the training set. Secondly, for every split process, a randomly chosen sample of size $m \approx \sqrt{p}$, of the features, is assessed and solely one can compose the proper split. The latter process is known for *decorrelating* the trees, because only looking at a random subset of the features for the split, make it less likely for several trees in the forest to present a strong dependence on a single important feature. In the present study, the prediction is based on the majority of votes in the forest.

In this context, there is the boosting technique. It works in a similar fashion, except

that the trees are grown sequentially as shown in Figure 3.12, using the information from previously grown trees. Boosting does not involve bootstrap sampling, instead each tree is fit on a modified version of the original data set. We then add this new decision tree into the fitted function in order to update the residuals. Each of these trees can be rather small. By fitting small trees to the residuals, the flaws from the previous trees are slowly optimized.

Figure 3.12 - Gradient Boosted Trees Example.



A schemtic view of the process of sequentially creating trees. Every tree of index t from 2 **to** M, generates a prediction $\hat{y}_i^t$ for the i-th input, which is a combination of its learned hypothesis $f_t(x_i)$ and the prediction of the previous tree $\hat{y}_i^{t-1}$. Therefore, its fitting process will regard the residual of the prediction of the previous grown tree, and its target will be $y_i - \hat{y}_i^{t-1}$.

SOURCE: Thorat (2023).

Among those methods, XGBoost (Extreme Gradient Boosting) stands out for its performance and efficiency. It was chosen for handling the data in the present study due to three key advantages:

No Requirement for Data Normalization: As it builds on decision trees that segment based on feature values rather than their magnitude, XGBoost simplifies the preprocessing step by eliminating the need for data normalization.

Effectiveness with High Dimensionality and Non-linear Relationships: It is able to handle datasets featuring a large number of variables and intricate non-linear

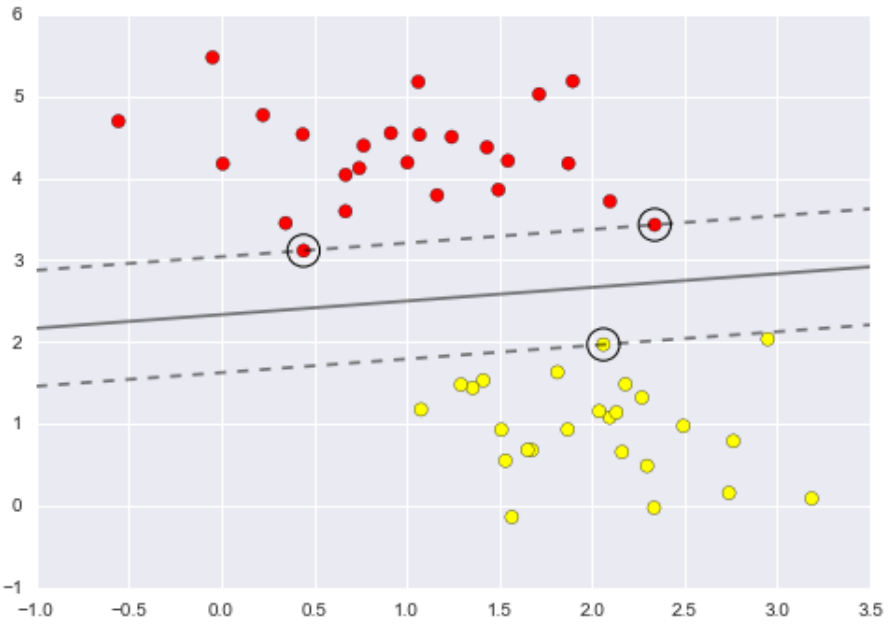
interrelations.

Outlier Management: While intrinsically resilient to outliers, in scenarios where outliers significantly impact model accuracy, applying normalization or transformation techniques can further bolster model stability and interpretability.

3.3.3.2 Support Vector Machines

The core idea of SVM is to find the hyperplane that best separates the data points into their respective classes with the maximum margin. The margin is defined as the distance between the hyperplane and the nearest points from each class, known as the support vectors, like the ones displayed in Figure 3.13. This approach aims to ensure that the decision boundary is as far away from any data point as possible, reducing the risk of misclassification. SVM works well with high-dimensional data and is versatile in handling both linear and non-linear separations.

Figure 3.13 - Support Vector Machine boundary.



An example of an SVM perfect separation of two classes of data in red and yellow. The dividing grey line maximizes the margin, indicated by the dashed line. The support vectors are the encircled data points in both classes over the dashed lines.

SOURCE: VanderPlas (2016).

3.3.3.2.1 Stochastic Gradient Descent for linear SVM

Stochastic Gradient Descent (SGD) is an optimization method used to find the values of the parameters of a function that minimize a cost function, in this case, the cost function of the linear SVM. It is called "stochastic" because, unlike traditional gradient descent that uses the entire data set to calculate the gradient of the cost function, SGD randomly selects a small batch of data points at a time to compute the gradient and update the parameters. This makes SGD particularly useful for large datasets, as it can significantly speed up the convergence process.

The process consists of an **initialization**, with an initial guess for the weights (coefficients) of the hyperplane and the bias term. These weights are usually initialized to zero or small random values. The next step is to compute the **gradient**, that is, for each training example x_i with its respective label y_i, compute the gradient of the loss function (in the present study, *hinge loss*):

$$HingeLoss = max(0, 1 - y_i(\mathbf{W}\dot{x}_i + b)) \tag{3.7}$$

, where $\mathbf{W}$ is the weight vector, b is the bias, and x_i is the input vector of the i-th observation.

Further, the weights are updated together with the bias term, based on the gradient. The update rule is:

$$\mathbf{W} = \mathbf{W} - \eta\dot{\nabla}\mathbf{W}$$

$$\text{b} = \text{b - } \eta\dot{\nabla}b \tag{3.8}$$

Here, η is the learning rate, a small positive scalar determining the step size at each iteration. $\nabla\mathbf{W}$ and ∇b are the gradients of the weight vector and bias term, respectively.

The gradients step and the weights update are repeated, iterating over the dataset multiple times (epochs) until the algorithm converges, meaning the changes in the wight vector are below a small threshold or a specified number of iterations is reached.

3.3.3.2.2 Kernel trick

SVM's capability to handle non-linear data lies in the *kernel trick*. The kernel trick involves mapping the original data into a higher-dimensional space where it is possible to find a linear hyperplane that separates the data points into their respective classes. This is done without explicitly computing the coordinates of the points in this high-dimensional space. Instead, the kernel function computes the inner products of the data points in this higher-dimensional space directly from the original input space.

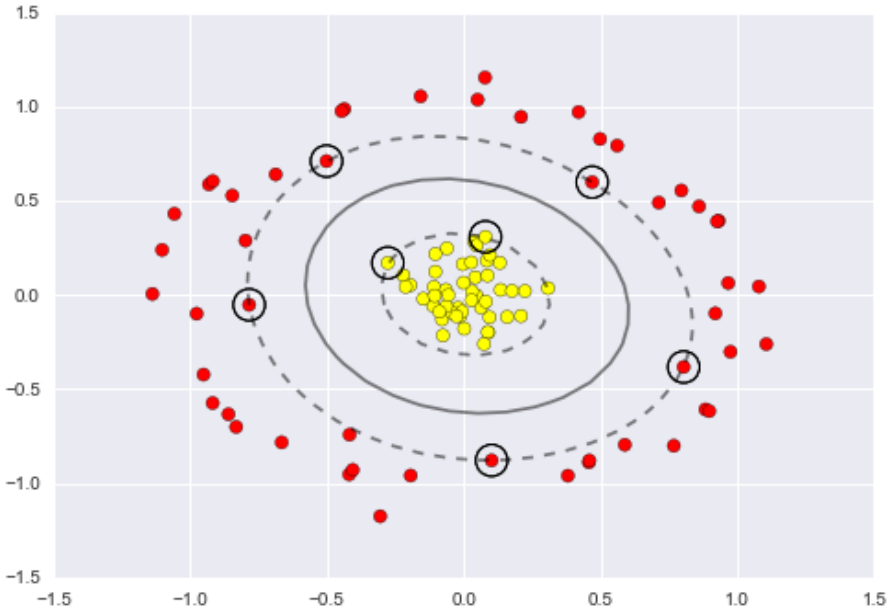
In the experimental runs of the present study, the Radial Basis Function (RBF) kernel is used. It is defined as:

$$K(x, x^{'}) = e^{-\gamma||x-x^{'}||^2} \tag{3.9}$$

Here, x and $x^{'}$ are two data points in the input space, $||x - x^{'}||^2$ denotes the squared Euclidean distance between these two points, and γ is a parameter that determines the spread of the kernel. Essentially, the RBF kernel measures the similarity between two points in the input space.

When an SVM uses an RBF kernel, the decision boundary that separates the classes in the input space is not necessarily linear. The SVM with RBF kernel effectively creates a complex, curved boundary that can follow the distribution of the points. This capability makes it especially suited for datasets where the classes are not linearly separable, this case is shown in Figure 3.14.

Figure 3.14 - SVM non-linear boundary.



A Kernelized SVM is able to fit non-linear decision boundaries using the kernel-trick. Here, the model was able to create a boundary with maximized margin between two non-linearly separable classes of data, shown in yellow and red.

SOURCE: VanderPlas (2016).

3.3.4 Optimal parameters selection

In machine learning, an estimator is a term synonymous with models or algorithms that are used for making predictions. Hyperparameters are the configuration settings used to structure the model. Unlike model parameters, which are learned from the data during training, hyperparameters are set prior to the learning process and control the model's architecture and learning dynamics. Examples of hyperparameters include the learning rate for gradient descent, the depth of a decision tree, or the number of hidden layers in a neural network.

In the present study, two types of hyperparameter searches are performed: using combinatorics and through bayesian optimizers.

3.3.4.1 Combinatorics

The *grid search* method uses combinatorics in this search. It is an exhaustive probe over a specified parameter space for an estimator. First, one must define a parameter grid containing the hyperparameters and their respective ranges of values to be tested. Then, one must specify the number of folds for cross validation (details in Section 3.3.5.1. For each possible combination of parameters in the grid, the model will be trained on the training set of each fold, the performance will be measured against the respective test set and the scores will be aggregated.

In the end of the process, the combination of parameters can be chosen based in the set that returned the best average performance among the folds. It provides the combination of parameters that yields the best model performance according to the chosen metric. The main downside is that it can be very time-consuming, especially with large datasets, complex models, or when the parameter grid is extensive. The computational cost grows exponentially with the number of parameters and their values.

3.3.4.2 Bayesian optimizer

Bayesian optimization takes a different approach to hyperparameter tuning compared to exhaustive grid searches. This method iteratively updates a probabilistic model of the objective function (i.e. the performance metric to be optimized) based on the outcomes of past evaluations. Unlike grid search, a Bayesian optimizer focuses the search on hyperparameter combinations that are more likely to yield improvements, balancing the exploration of uncharted regions with the exploitation of known promising areas.

In the present study the implementation called Optuna (AKIBA et al., 2019) was used. It begins with a phase of random searches or user-defined strategies to explore the hyperparameter space and gather initial data points. Utilizing the Tree-structured Parzen Estimator (TPE) as its default surrogate model, Optuna approximates the relationship between hyperparameters and the objective function. The details on the algorithm are out of scope of the present study, so we refer the reader to (OZAKI et al., 2020).

The TPE model helps in predicting which hyperparameter values are likely to improve model performance. Optuna employs an acquisition function to select the next set of hyperparameters for evaluation. This decision-making process is informed by

the TPE model, aiming to identify hyperparameters that balance the trade-off between exploring new areas and exploiting known good regions of the hyperparameter space.

The optimization process iterates, with each step involving the evaluation of the objective function with the chosen hyperparameters, updating the surrogate model, and selecting the next hyperparameters using the acquisition function. The cycle repeats until a stopping criterion is met, such as a specified number of iterations or a convergence threshold. Upon completion, Optuna provides the hyperparameter set that resulted in the best objective function performance, alongside insights into the optimization process and the hyperparameter importance.

Bayesian optimization approach offers several advantages over grid search, particularly in terms of efficiency and effectiveness in finding optimal hyperparameters. By intelligently narrowing down the search space based on the performance of previously evaluated configurations, Optuna can achieve comparable or superior model performance with significantly fewer evaluations. This property is especially beneficial when optimizing models like XGBoost, where the hyperparameter space is vast and model training can be computationally intensive.

3.3.5 Model training and testing

The core of this study lies in addressing a supervised learning problem, which requires the use of a reference dataset to guide the model's learning process. This is achieved through a curated set of labeled data, deemed the ground truth, which serves as the benchmark for the model's predictions. During training, the model undergoes numerous iterations where its parameters are fine-tuned to minimize the discrepancy between its outputs and the actual labels from the training data. This iterative process of error minimization is crucial for the model to learn the intricate patterns and relationships within the data, ultimately leading to convergence.

To rigorously evaluate the model's learning progression and ensure its generalizability, the data is partitioned into distinct sets: training, test, and validation. The training set is utilized to adjust the model's parameters, whereas the test set plays a critical role in tuning hyperparameters and preventing overfitting. This separation ensures that the model's performance is not inadvertently biased by the data it was trained on. The validation set, introduced only after the model has been finalized, serves as the ultimate benchmark for evaluating the model's performance. It provides an unbiased assessment of how well the model generalizes to unseen data,

reflecting its potential efficacy in real-world applications.

An essential step preceding the model training is the normalization of feature values. Normalization is a preprocessing technique that scales the feature values to a common range, enhancing the model's convergence speed and ensuring that no single feature disproportionately influences the model's performance due to its scale. This step matters for models sensitive to feature magnitude, facilitating a more balanced and efficient learning process.

After the model has been trained and hyperparameters optimized using the test set, the final assessment is carried out on the validation set. This evaluation is crucial for understanding the model's capabilities and limitations in a real-world scenario, where data may not always resemble the conditions under which the model was trained. Performance metrics, specific to the problem at hand, such as accuracy, precision, recall, and the F1 score, are employed to quantify the model's effectiveness in classifying hailstorms from satellite data.

Furthermore, model training and testing in this study adopt a rigorous approach to ensure that the findings are not only statistically significant but also practically relevant. Techniques such as k-fold cross-validation may be employed to maximize the utility of the available data, ensuring that the model's performance is robust and reliable across different subsets of data. This approach underscores the commitment to achieving a high degree of predictive accuracy while acknowledging the inherent uncertainties in modeling complex natural phenomena.

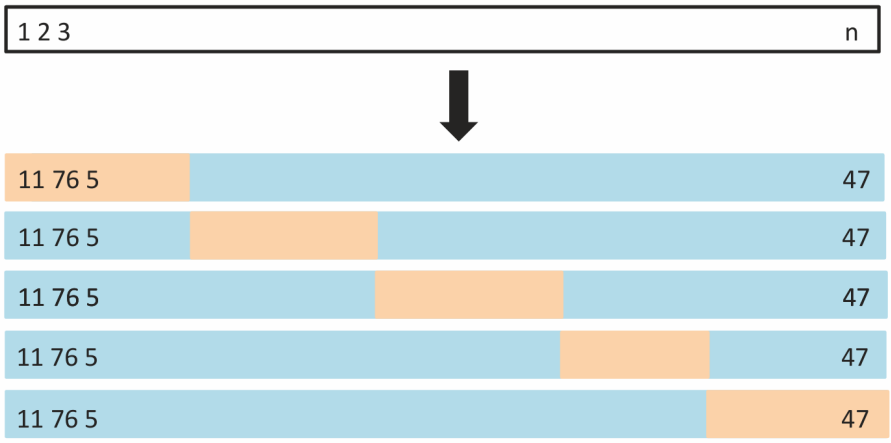
3.3.5.1 Resampling methods

In Section 3.3.2, the concept of model variance was introduced. One way of measuring the variance of the model consists of repeatedly fitting the model over different samples of the training set, and then examine the change in performance. Subtle changes in the metrics are expected, but large variations may be an indication of the model dependency on the training set arrangements. This could be an artifact of the model type, or the aftermath of a model too flexible for the task in-hand and possible overfitting. Therefore, this process is an asset in the model choice and its hyperparameters settings definition.

Two resampling methods are used in the current study: cross-validation and bootstrapping. Both methods provide insights into the generalization's ability of our models beyond the observed data.

Cross-validation is a resampling technique, commonly used to evaluate the predictive performance of a model and its generalization to an independent dataset. Here, we used the k-fold cross-validation, where the dataset is randomly divided into a number k of equal-sized folds. For each iteration, one fold is retained as the test set, and the remaining $k - 1$ folds are used as the training set. The model is trained on the $k1$ folds and tested on the retained fold. This process is repeated k times, with each fold serving as the test set exactly once, as shown in Figure 3.15.

Figure 3.15 - K-fold cross validation.



Schematic view of a 5-fold cross validation scheme. From top to bottom: a set of n observations are randomly split into five non-overlapping sub-groups. In beige, are the subsets used for testing, and in blue, the ones used for training per round.

SOURCE: James et al. (2023).

Bootstrapping is another resampling method used to estimate the distribution of a statistic (e.g., mean, variance) or model accuracy by repeatedly sampling with replacement from the observed dataset. This technique allows us to create "new" datasets by randomly selecting observations from the original dataset, with the possibility of choosing the same observation more than once. In our experiments, bootstrapping was used to assess the stability and variance of our model's performance metrics.

3.3.5.2 Evaluation metrics

Before proceeding to the evaluation, the output values must first be converted from probabilistic values to specific classes. Then, a binarization will be performed based on a probabilistic threshold p^*, so that every region with output $p \geq p^*$ will be considered "yes" for hailfall detection and everywhere else will be "no". In the present study, we have taken $p^* = 0.5$.

The model learned must be evaluated separately against the validation and the test sets. The metrics chosen for the present study are the following: Accuracy (Equation 3.10), which tests the ability to predict both classes; Recall (Equation 3.11), that depict the model's skill to correctly detect positive classes from all actual positive classes; Precision (Equation 3.12) to evaluate how correctly the model detects positive classes from all predicted positive classes and F1-test (Equation 3.13), which is the harmonic mean of precision and recall.

$$Accuracy = \frac{(T_p + T_n)}{(T_p + T_n + F_p + F_n)} \tag{3.10}$$

$$Recall = \frac{T_p}{(T_p + F_n)} \tag{3.11}$$

$$Precision = \frac{T_p}{(T_p + F_p)} \tag{3.12}$$

$$F1 = \frac{(2 \times Recall \times Precision)}{(Recall + Precision)} \tag{3.13}$$

Where, T_p are true positives and T_n true negatives, which correspond to cases that were correctly classified. F_p are false positives and F_n false negatives, meaning cases that were misclassified.

Aditionally, AUC, the Area Under the Curve of True Positive rate Vs. False Positive rate (ROC curve) at different decision thresholds is also used in some experiment settings. Here, the ROC curve is assumed to feature T_p rate on the Y axis, and F_p rate on the X axis, so that leaning to the top left corner of the plot it approaches the "ideal point" (i.e. a F_p rate of zero, and T_p rate of one). As a consequence, a larger AUC is usually better.

3.4 Model exploratory analysis

After the model is accomplished, we aim to reconstruct the weights for the features found by the model to represent the knowledge. That would enable to validate the algorithm components by checking internal states for physical consistency. It might also bring improvements in performance after tracking regions in the input space misrepresented by the training data and possibly identifying circumstances in which the model fails to generalize. Thereby, one might as well discover features in the dataset that could be relevant for novel ways of posing the problem.

3.4.1 Shapley values analysis

Shapley values are a concept from cooperative game theory that allocate payouts (or costs) to players based on their contribution to the total payout. When applied to machine learning, Shapley values explain the prediction of an instance by quantifying the contribution of each feature to the prediction.

The Shapley value is the average marginal contribution of a feature value across all possible combinations of features. In the context of machine learning, it answers how much does each feature contribute to the difference between the actual prediction and the *average prediction*. This is achieved by treating the model as black box, changing the inputs and observing the changes in the output (ŠTRUMBELJ; KONONENKO, 2014).

Calculating Shapley values directly can be extremely computationally intensive since it requires evaluating the contribution of each feature over all possible subsets of features. For a model with n features, there are $2^n - 1$ possible subsets of features to consider for each feature. Therefore, in the current study we use the SHAP library (LUNDBERG; LEE, 2017) that implements an efficient approach to compute Shapley values.

Within the SHAP library, the **KernelExplainer** is a model-agnostic tool that approximates Shapley values for any machine learning model. It works well with any model but is particularly useful when dealing with complex, black-box models.

KernelExplainer uses a weighted linear regression to estimate the Shapley values. It selects a random subset of features and observes the change in prediction when including versus excluding a specific feature. Sampling many such subsets and performing this comparison, it can approximate the Shapley values without needing to compute them over all possible subsets explicitly. The output of this process can

be interpreted directly as the contribution of each feature to the prediction for a specific instance relative to the average prediction over the dataset. The Shapley value for a certain feature i is represented as:

$$\phi_i(p) = \sum_{S \subseteq N/i} \frac{|S|!(d - |S| - 1)!}{d!}(p(S \cup i) - p(S)) \tag{3.14}$$

$|S|$ is the amount of features in a subset S truncated at the (i-1)th feature of the complete set of features N. Then, $|S|!$ is the amount of possible combinations permuting the features in S. The fraction of terms is a weight to the difference of the model prediction with the i-th feature compared to the prediction without it $(p(S \cup i) - p(S))$.

As seen, this problem is commonly tackled by assessing the difference between a prediction for an instance and the expected prediction for the same instance if the i-th feature had not been known (LEMAIRE et al., 2008; ROBNIK-ŠIKONJA; KONONENKO, 2008). This has an exponential time complexity, which makes the method infeasible for practical use. In order to compute this, the value of the i-the feature is perturbed while the values of other input features remain fixed and then the predictions are averaged. For details on the algorithm for this estimate, we refer the reader to Štrumbelj and Kononenko (2014).

One of the main motivation for the use of Shapley values is that it has a solid theoretical foundation (SHAPLEY et al., 1953; SHAPLEY; SHUBIK, 1954; SHAPLEY, 1969; ROTH, 1988), ensuring that feature contributions are fairly allocated. Additionally, a *Kernel Explainer* can be used with any machine learning model, which provides a unified measure of feature importance for studies involving several types of models.

3.4.2 Principal Component Analysis

Principal Component Analysis (PCA) serves as a fundamental technique in the realm of model exploratory analysis. Its primary objective is to project the input variables into an M-dimensional subspace, where $M < d$ (d represents the dimensions of the original input space as discussed in Section 3.3.1). This dimensionality reduction is achieved through a two-step process:

1. **Construction of a New Feature Space**: Initially, an M-dimensional feature

space is constructed. The axes of this space are linear combinations of the original input space's bases. This is designed to capture as much of the variability in the data as possible with a reduced number of dimensions.

2. **Model Fitting and Projection**: Subsequently, a model is fit using these M predictors. This model is necessary for projecting the data points from the original high-dimensional space into the newly created, reduced-dimension space.

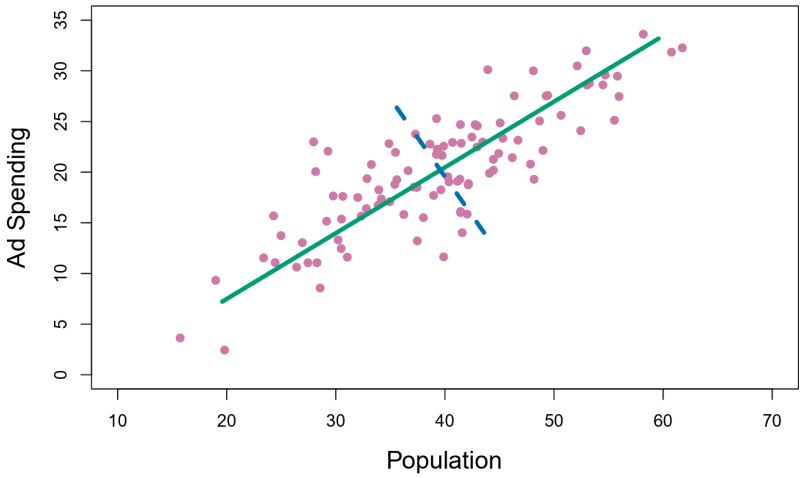
The axes, or bases, of the newly created M-dimensional space are identified as the Principal Components (PCs) of the original dataset. These components are necessarily orthogonal, ensuring that each represents a unique variance direction. Principal Components are conceptualized as the directions within the original space where the data exhibits the most significant variance. This characteristic is visually represented in Figure 3.16, illustrating how PCA captures the essence of data variability. Then, each component found will have the form:

$$Z_j = \phi_{1j}X_1 + \phi_{2j}X_2 + ... + \phi_{pj}X_p \tag{3.15}$$

That means, that each axis in the new space is a normalized linear combination of the features X_1 to X_p. Then, the factors ϕ_{1j},...,ϕpj are the *loadings* of the j-th principal component and $\sum_{j=1}^{p} \phi_{ij}^2 = 1$.

In the present study we leverage the benefits of PCA mostly for **Dimensionality Reduction**, as PCA reduces the dataset's complexity while retaining the variation present in the dataset. This reduction is important for models vulnerable to the "curse of dimensionality" (described in detail in Section 4.2.2.1 or when the goal is to simplify models to improve interpretability. Additionally, we also benefited from the assessment of **Feature Correlation**, because PCA helps in uncovering patterns in the data by highlighting the correlation between features. By analyzing principal components, one can understand the underlying structure of the data.

Figure 3.16 - Example of Principal Components.



A representation of the principal components in a 2-dimensional dataset of ad spending per Population size for 100 different cities (purple circles). The first principal component is shown the continuous line in green and the second, by the dashed line in blue.

SOURCE: James et al. (2023).

4 EXPERIMENTS AND RESULTS

4.1 Overview of experiments

The following experiments were built upon the Prevots dataset. There is a time gap between the experimental activities and, the Prevots set is under continuous update, therefore, the datasets differ in size from one experiment to another.

The classification tasks follow the same categorical sequence of of dataset construction; model training, testing, validation; and assessment of feature importance. They involve however different types of model and ways of posing the problem. Overall, 32 rounds of training and validation were performed, which were then subdivided into two main sorts in accordance with the experimental setting. Our approach to the problem was to scale in conformity with the necessity. Thereby, we expected to keep the models as lean as possible and avoid curbing the interpretability. In the following sections, these main structure of the experiments are thoroughly described.

4.2 Experiment 1: pixel-wise approach

For the present experiment, our aim was to develop a pixel-wise classification. In this type of problem, the developed model receives as input a large set of pixels and returns a value for the class each pixel belongs. Here, the input set is the BT from ABI's 10 IR bands in the terrestrial emission region, between 3.9 μm and 13.3 μm (i.e. channels 7 to 16), over the study domain. As described in Section 1.1.2, hailstorms in the area of interest are most frequent in the late afternoon and early evening hours. Then, if the algorithm was trained to receive as input measurements from channels in the solar spectrum, most of the hail events would be missed due to physical constraints. For this reason, ABI's channels 1 to 6 were left out. As output, we expect a segmented scene of regions prone to be tied to hailstorms.

4.2.1 Methods

In the beginning of this experiment we held a set of 5776 hail reports from Prevots between June 2018 and December 2020. So firstly, this dataset of records were filtered out to maintain the ones with less uncertainty. At this stage, reports that presented more than 10 minutes of uncertainty were removed from the dataset. This first filter returned 4899 listed records.

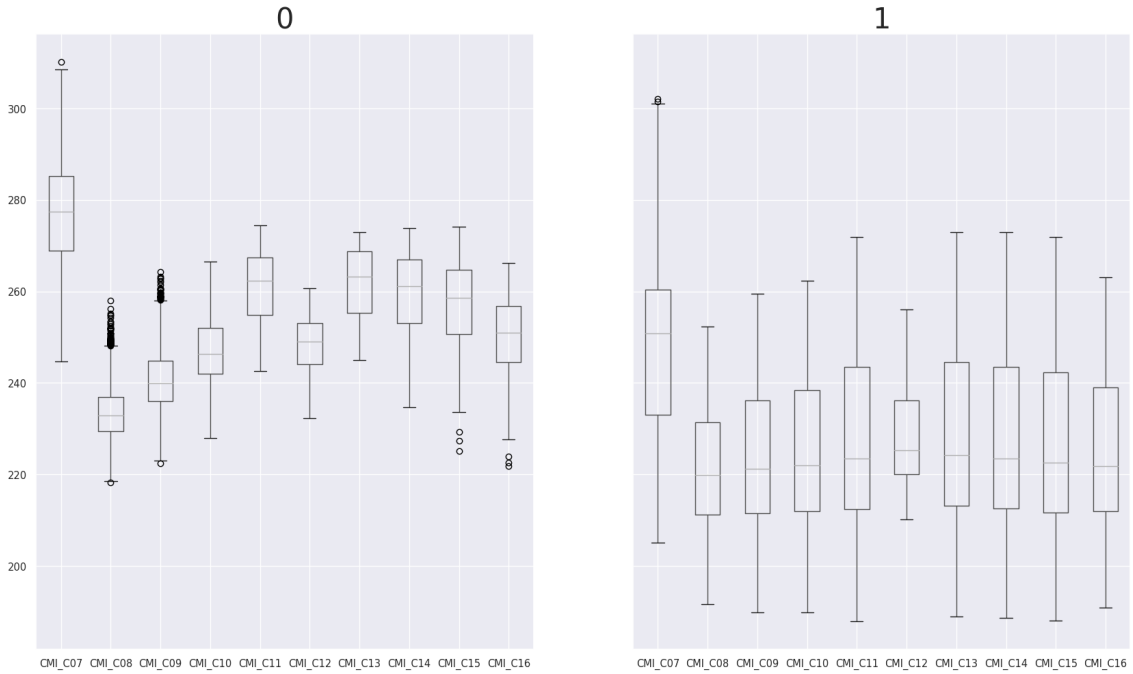
Here, the training set was built with the pointwise BT collected in the coordinate of each hail report. During this process, the filtered reports were uploaded to the

Google Earth Engine (GEE) platform to leverage its parallel computing capabilities of petabyte-scale datasets (GORELICK, 2013) to extract simultaneously the ABI measurements relative to all hail reports.

Only a portion of the reports match the exact time of ABI's records. So, the strategy to select the samples was changed to determine the best match as the value closest in time to each report within 10 minutes prior and after the Prevots record. The sampling of examples for the Class of No-Hail followed the same temporal criterion, with the addition that the BT value in 10.3 μm must be between 245K and 273K to assure a convection depiction (MACHADO et al., 2002), before being pseudo-randomly selected. This screening process assured that only samples of storms likely of generating hail were considered as there is no need for a model capable of distinguishing pixels from warm clouds.

Figure 4.1 shows the BT distribution in the training/test set for both classes of prediction. As depicted, the span of measurements of Class 1 appear much more uniform than the Class 0. This is likely an artifact of the sampling process. Most of the hail reports in South America are restricted to an area of a specific range of storm cloud top BTs. On the other hand, the pixels for Class 0 were randomly chosen over the entire domain, giving much more freedom for the sampled values. Naturally, this set required further processing stages, for instance, BTs over 273 K are not considered in neither of the classes.

Figure 4.1 - Overview of the BT distribution in the training set.



Classes of targets: 0 means No-hail and 1 for Hail.
SOURCE: The author.

The whole processing in GEE resulted in a tabular set of BTs for each report with positive and negative examples (i.e. no-hail), except in cases the CMIPF data was not available in GEE or presented artifacts, such as BT $<$ 197.31 K in in the Shortwave Window (3.9μm), which is the BT's minimum bound of the dataset. Prior to training, Hail Class examples with BT $>$ 273 K in Channel 13 were also removed, since it is physically impossible for it to be associated with hailfall. Such values were sampled likely due to the spatial uncertainty in some of the Prevots records (Section 3.2.1).

After the pre-processing, the tabular set contained 4538 instances belonging to the Hail class and 4835 to the No-Hail class. Prior to the training, every feature was normalized to speed up model convergence (Section 3.3.5). This procedure consists of applying the equation:

$$Z = \frac{x_j - \mu_x}{\sigma_x} \tag{4.1}$$

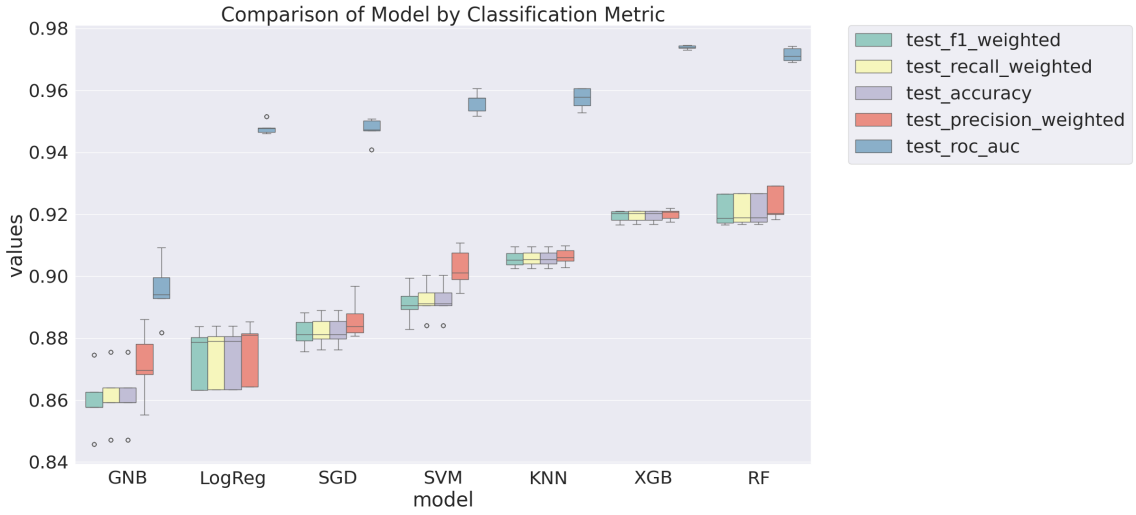
in which x_j is the value feature $\mathbf{X}$ assumes in measurement j, μ_x is the feature $\mathbf{X}$'s mean, and σ_x its standard deviation.

The measurements were randomly split with 75% of data for training and the remaining 25% for testing. These sets were given as input to 7 different types of models under SciPy library's standard configurations to get an overall assessment of their fit-time scoring and classification performance. The tested model types are a Logistic Regression Classifier (LogReg) (BISHOP, 2006) (Section 3.3.2.1); Stochastic Gradient Descent Classifier (SGD) (PEDREGOSA et al., 2024) (Section 3.3.3.2.1); Random Forest (RF) (BREIMAN, 2001) (Section 3.3.3.1.1); K-Nearest Neighbors (KNN) (COVER; HART, 1967) (Section 3.3.2.2); Support Vector Machine (SVM) (SMOLA; SCHÖLKOPF, 2004) (Section 3.3.3.2); Gaussian Naive Bayes (GNB) (CHAN et al., 1979) (Section 3.3.3) and XGBoost (XGB) (FRIEDMAN, 2001) (Section 3.3.3.1.1). The steps of data preparation and the standard model invoke were conducted through the scikit-learn (Sklearn) library (PEDREGOSA et al., 2011), with the exception of the XGB. There is a Sklearn interface to XGB, though, but we preferred to use the XGB proper library instead.

For the assessment of classification effectiveness, 5 indices were analyzed, weighted by the relative amount of instances in each class: accuracy, recall, precision, F1 test and ROC-AUC. For a deeper discussion on those metrics and the reasoning behind this choice, please refer to Section 3.3.5.2. Additionally, those metrics were evaluated over a 5-fold cross validation. Finally, empirical bootstrapping was performed at 30 samples to construct a visualization of the previously described metrics results.

From Figure 4.2, among the tested algorithms, both ensemble classifiers (RF and XGB) fitted the data very well while Gaussian Naive Bayes fitted the data rather poorly across all metrics.

Figure 4.2 - Comparison of Experiment 1 models.

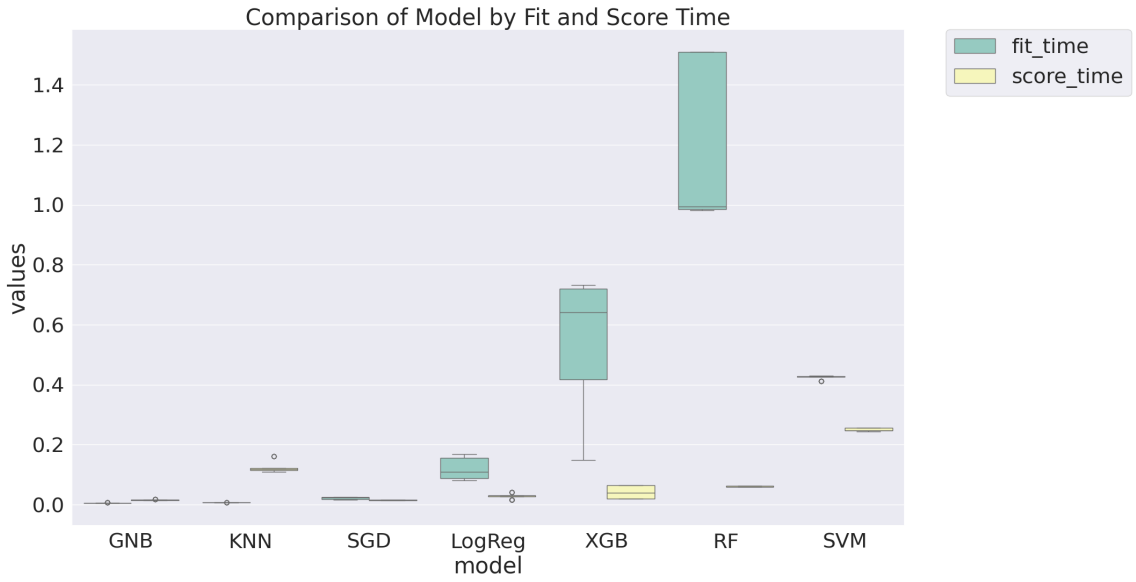


Overall classification performance of tested model types. Larger values in every metric (color encoded) indicate better performance.

SOURCE: The author.

The models were also tested regarding the time for fitting on the train set for each cross validation split (fit time) together with the time spent on scoring the estimator on the test set for each cross-validation split (score time). The results are displayed in Figure 4.3. This time, the Random Forest classifier shows a poor performance in speed of fitting scoring the data.

Figure 4.3 - Models time performance.



Comparison for fitting and classification. Smaller values are related to better performance.

SOURCE: The author.

After classification performance assessments, better results were obtained from the use of ensemble methods. Therefore, this class of models was chosen for further experiments. In this experiment we have conducted the next rounds of assessments using Random Forest models.

In order to choose the best model settings, the classification metrics of the model were again evaluated while switching possible combinations (for details, refer to Section 3.3.4.1 of a set of values for the number of estimators in the ensemble and maximum number of features per node splitting. The best performance was obtained through an ensemble of 800 estimators and a maximum of 6 features per branching and bootstrapping enabled. This model setting was trained on the previous split of 80% of the dataset and the trained model was evaluated against the remaining 20%.

4.2.2 Results

Results of the trained model predictions in the validation set are displayed in Table 2. A 94% precision was achieved in identifying hail cases, only 3% better than in no-Hail cases. Simultaneously, the model scored 95% recall in identifying no-Hail

cases, that is 5% more than for Hail cases. This is an overall good result, but also suggests that there is a higher likelihood that Hail cases are misclassified as no-Hail cases, than the opposite.

Table 4.1 - Classification scores in the validation set.

Class Type	Eval. Metric			# of Cases
	Precision	Recall	F1-score	
No-Hail	0.91	0.95	0.93	976
Hail	0.94	0.90	0.92	903

SOURCE: The author.

Subsequently, the feature importance was checked as an effort towards understanding what is "seen" by the model. This analysis takes into account how valuable the feature's information is in keeping the model's purity after branching. This generally provides a hint on how much the model's prediction would degrade if the features combinations were permuted.

Table 4.2 - Importance of ABI's Channels for the classification.

	Ch. 07 3.9 μm	**Ch. 08 6.19 μm**	**Ch. 09 6.93 μm**	**Ch. 10 7.34 μm**	**Ch. 11 8.44 μm**
Importance	0.046	0.055	0.036	0.031	0.346
	Ch. 12 9.61 μm	**Ch. 13 10.33 μm**	**Ch. 14 11.21 μm**	**Ch. 15 12.29 μm**	**Ch. 16 13.28 μm**
Importance	0.128	0.249	0.048	0.028	0.031

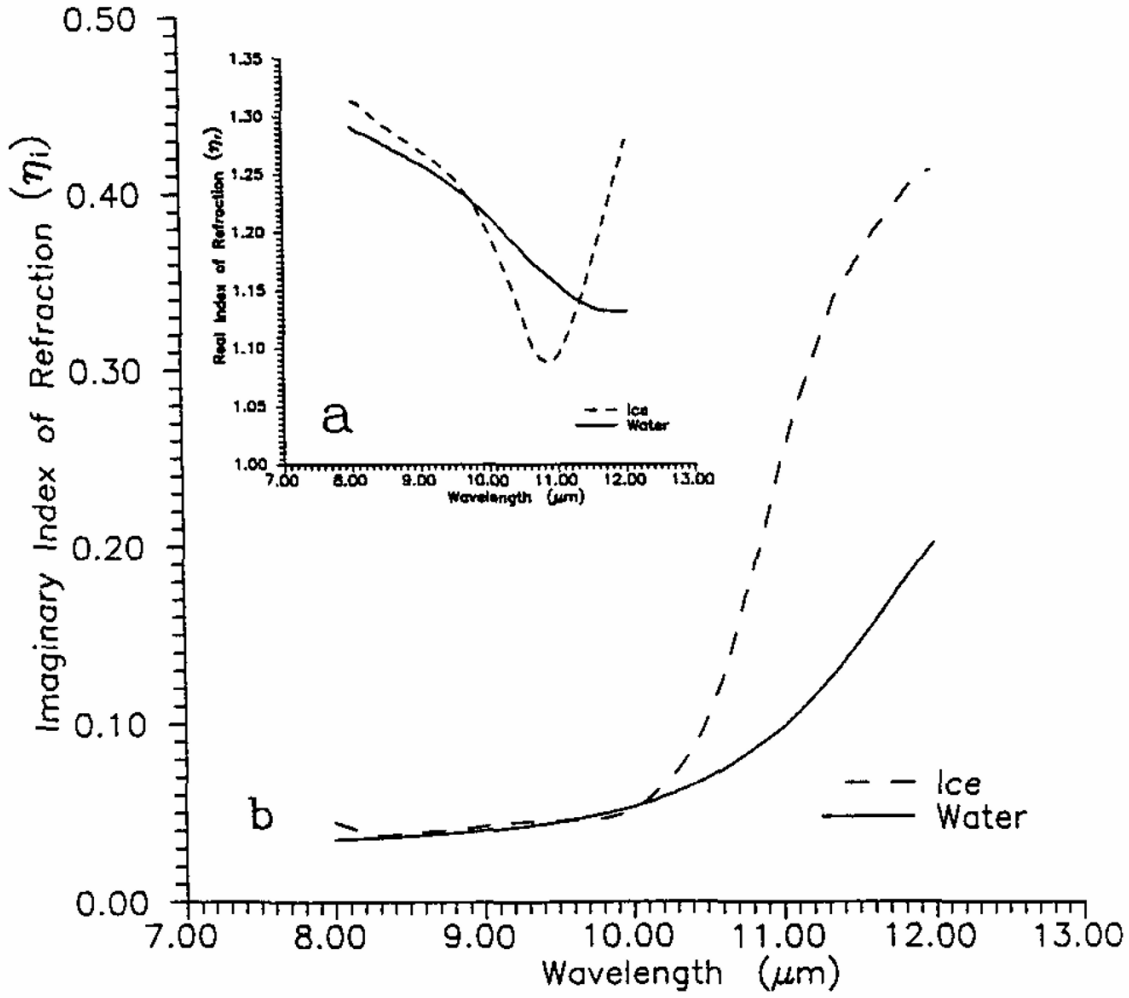
SOURCE: The author.

As depicted in Table 4.2, ABI channels 11 (cloud top phase), 12 (Ozone) and 13 (clean window) were the most important ones for the classification task. Those are bands knowingly associated with deep convection features and, together, dominate the feature importance. In comparison, the other channels presented very low influence in the final result.

Channel 11, centered in 8.5μm, is a window channel (unless in the presence of SO2). This band is a good asset in determining the phase of the cloud due to differences in absorption and scattering of water and ice particles in this wavelength (KEY;

INTRIERI, 2000). Also, more absorption takes place between 8 μm and 10.3 μm for ice particles than for water (STRABALA et al., 1994), as can be seen through the slope of the imaginary portion of the refraction index for water and ice depicted in 4.4. Therefore, brightness temperature differences between these two wavelengths tend to be larger for ice clouds.

Figure 4.4 - Indices of refraction of water and ice.

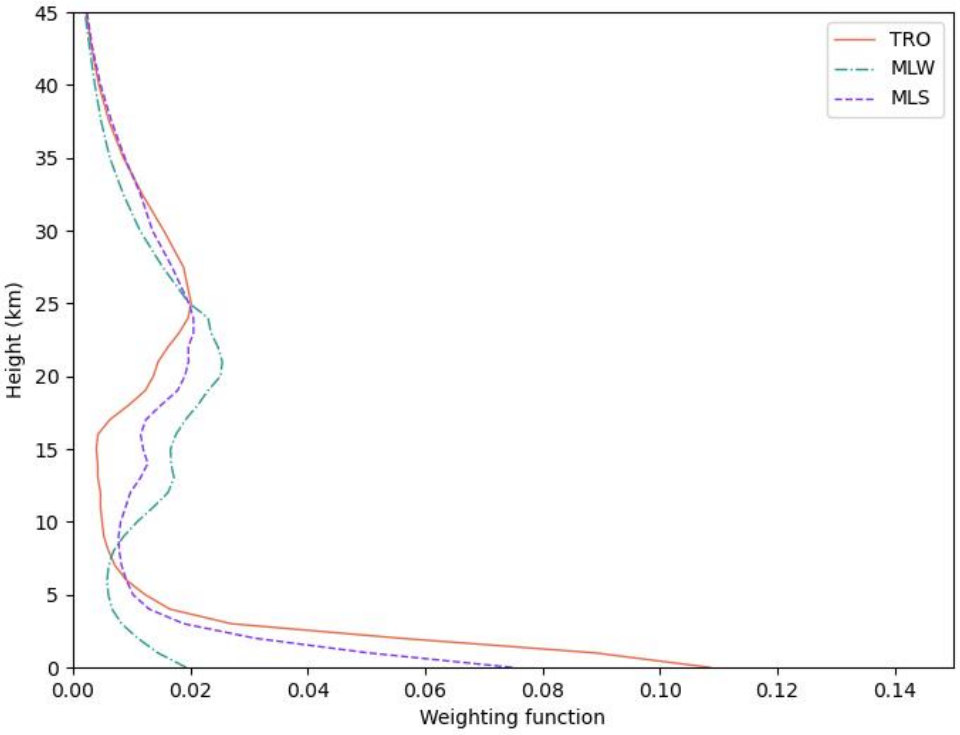


Indices of refraction of water and ice across the atmospheric window: (a) real portion, (b) imaginary portion.

SOURCE: Strabala et al. (1994).

Channel 12 is known as the ozone band. Ozone absorption at 9.6 μm makes it useful to obtain information from atmospheric aspects at high levels in the Atmosphere, as its weighting function also peaks in the Stratosphere (see Figure 4.5). Water vapor absorption also occurs in this band, complicating its use because the horizontal distribution of ozone and water vapor varies across the globe (LINDSTROM; SCHMIT; GERTH, 2017). All things considered, Overshooting Tops are noticeable in this wavelength, but due to ozone emission in the warmer stratosphere, their brightness temperature will appear warmer in this band than observed at the clean infrared window.

Figure 4.5 - Clear-sky weighting function of Channel 12 of GOES-ABI.

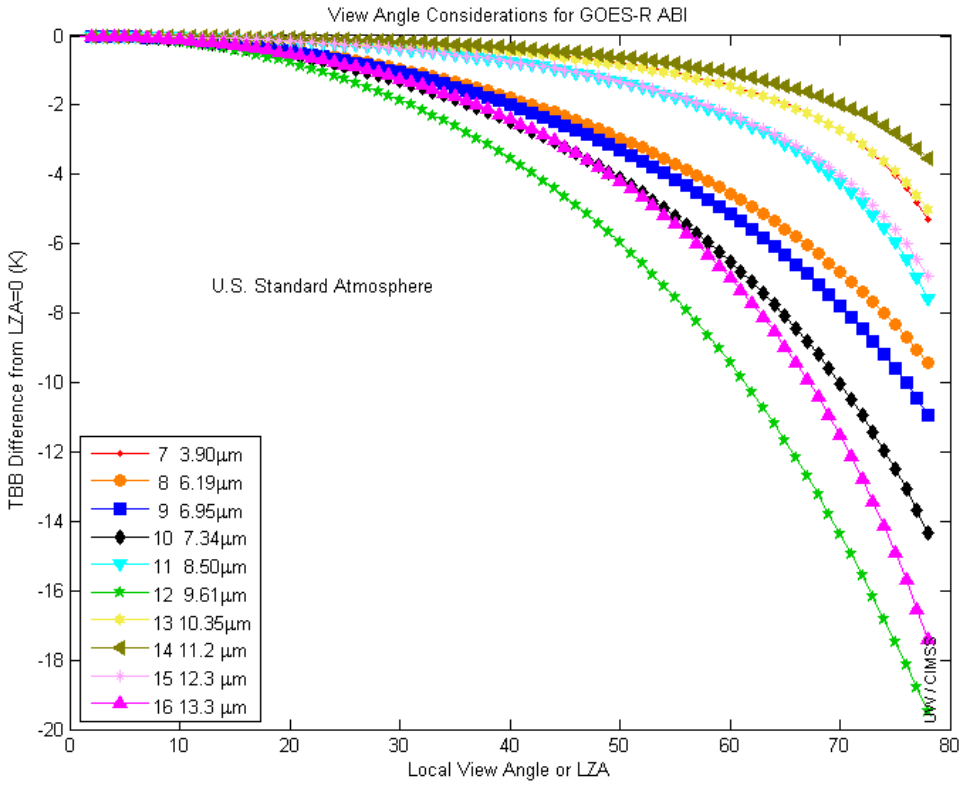


ABI's Channel 12 is centered at 9.6 μ m. Above are the weighting function over standard atmospheric profiles for the Tropical (TRO), Midlatitude Summer (MLS) and Midlatitude Winter (MLS).

SOURCE: GAVA (2024).

Additionally, Channel 12 could also be strongly impacting in the classification due to its cooling effect with increasing distance from sub-satellite viewing point. The signal measured by the ABI (or any other instrument) is affected by the view angle associated with any given field of view. The magnitude of this effect varies with wavelength and the composition of the atmosphere through which the field of view is seen. As an example, Figure 4.6 displays the cooling effect of ABI bands with the increase of ABI's view angle with the U.S. Standard Atmosphere. There, it is clear that approximately after 20^o of local zenith angle (LZA), the behavior of BT from Channel 12 completely detach from the others. At the same time, as detailed in Section 1.1.2, most of the hail generating storms in the domain of interest occur in the southernmost Brazilian states, which, in comparison to other portions of the study region, represent the furthest points to GOES-16 within the area. Therefore, it is possible that the BT behavior in Channel 12, compared to the others, is a decisive factor in selecting storms in the most prone region to hailfall.

Figure 4.6 - Impact of view-angle on ABI-bands.



Cooling effect of ABI bands with increasing distance from sub-satellite viewing point considering the U.S. Standard Atmosphere. That is, the difference between the BT measured at several Local Zenith Angles and the NADIR viewing BT. The colors encode this difference magnitudes for measurements from channels 7 to 16.

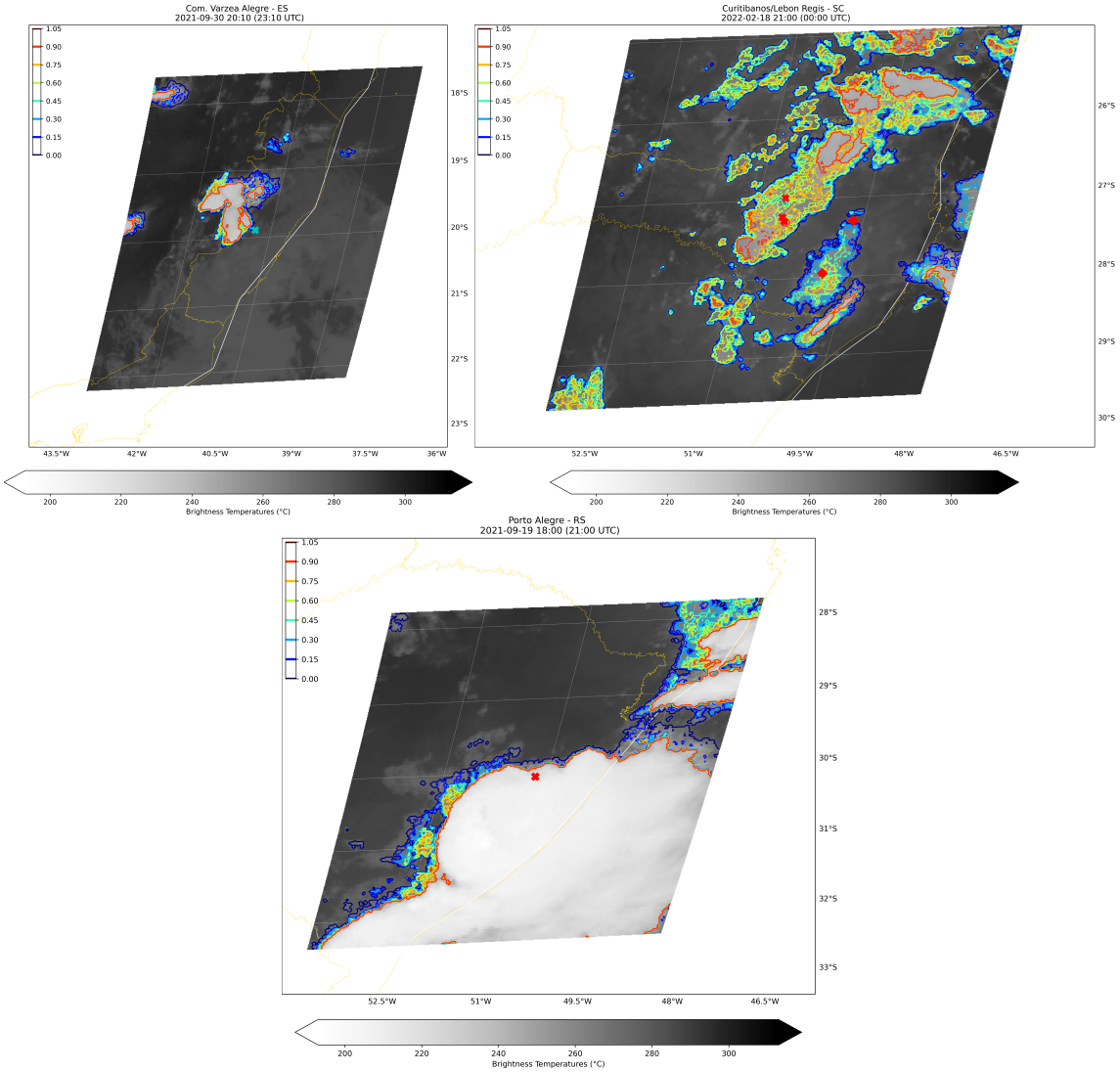
SOURCE: CIMSS (2023).

Other attempts including just the top four most important features and repeating the hyperparameter selection steps returned no improvement. In fact, the Precision in detecting pixels belonging to the Hail Class reduced 4%. Additional runs were conducted with the values for the tri-spectral method (STRABALA et al., 1994) and WV-IR window BT difference (SCHMETZ et al., 1997) which have returned little importance for the classification.

The model was then used to evaluate the risk of hail fall in 14 visually inspected distinct cases (Figure 4.7). Those cases introduce three distinct scenarios of storms

spacial placement: I) a scene with one cloud associated with the hailfall; II) several isolated clouds but only some of them precipitating hail and III) one large system with hail precipitation in parts of the system. Despite the verified approximate coordinate, the time of the news report is not exact. For this reason, some of the spots marked are not exactly under a cloud. Considering a margin of time uncertainty, in a case of an isolated cloud associated with the hailfall, the model predicted a high probability close to the reported hailstorm. Similar results are found for when there are several systems in the scene but only some generating hail. The model's prediction isolated some regions more hail prone than others. However, it failed to distinguish regions more prone to hailstorms embedded in a large system. Instead, it predicted a high probability (over 90 %) for the whole cell.

Figure 4.7 - Example cases of validation of the pixel-wise approach, for visual inspection.



Values are in predicted probabilities by the model. Top-left: an isolated hail generating storm in Várzea Alegre - ES on 2021-09-30 20:10 Local Time; Top-right: arrangement of several storms in which only some were tied to hail reports in Curitibanos/Lebon Regis - SC on 2022-02-18 21:00 Local Time; Bottom: one large system with portions of it associated with hailfall in Porto Alegre - RS on 2021-09-19 18:00 Local Time.

SOURCE: The author.

4.2.2.1 The curse of dimensionality

In the present study, the observations are comprised in high dimesionality. Every data point is embedded in the 10-dimensional space generated by the BT estimates in ABI channels 7 through 16 (input features). Considering this space, the intuition behind this classification problem is that the class to which a new observation belongs should be more strongly determined by nearby points from the training set, and less by distant points. However, as the dimensionality increases, the volume of the space increases so fast that the available data become sparse (BISHOP, 2006). Simultaneously, the difference between the nearest and farthest points tends to become negligible. Then, ML models that learn to separate those instances might face difficulties to find patterns and, therefore require exponentially more data to achieve the same level of performance (GÉRON, 2019).

At this point, some reasoning about our input features led us to believe that the measurements are confined to a region of the space having lower effective dimensionality. This is so, because the input vectors are tied to convective storms, and most of the information from the same target physically consists in small perturbations in the cloud top BT.

In order to evaluate that, we have performed a PCA on the input data through Singular Value Decomposition (SVD) (for details, we refer the reader to Sections 2.4, 6.4, 7.7 and 8.6 of Golub and Loan (2013)). Thereby, we could obtain the intrinsic components accounting for most of the useful information in the data, as well as the *explained variance ratio* (EVR). The latter represents the proportion of the dataset's total variance that is captured by each component.

EVR consists in the variance of the principal component divided by the total variance in the original dataset. Therefore, a high EVR for a principal component suggests that it captures a significant portion of the information in the dataset. Conversely, a low EVR indicates that the component accounts for a relatively small amount of the dataset's variance.

In the context of dimensionality reduction, the EVR is useful in determining how many PCs should be retained to effectively summarize the dataset. By examining the cumulative EVR of the principal components, one can decide the minimum number of components needed to capture a substantial portion of the total variance, thereby reducing the dataset's dimensionality without losing critical information. Selecting a subset of principal components based on their EVR can significantly improve the

efficiency of subsequent analyses and machine learning models by reducing computational complexity and potentially enhancing model performance due to the removal of noise and less informative variables.

The EVR calculated for the three PCs are displayed in Table 4.3. As shown, PC 1 would be enough for explaining 93.62% of the variance in the data. Together with PC 2, it would account for 98.58% of the information in the dataset. If one consider the three main PCs, then only approximately 0.3% of the information retained in the original dataset is lost. These results show that the directions over which important variations in the target variables occur, are very much confined, then, merging highly correlated features will most likely not result in too much information loss.

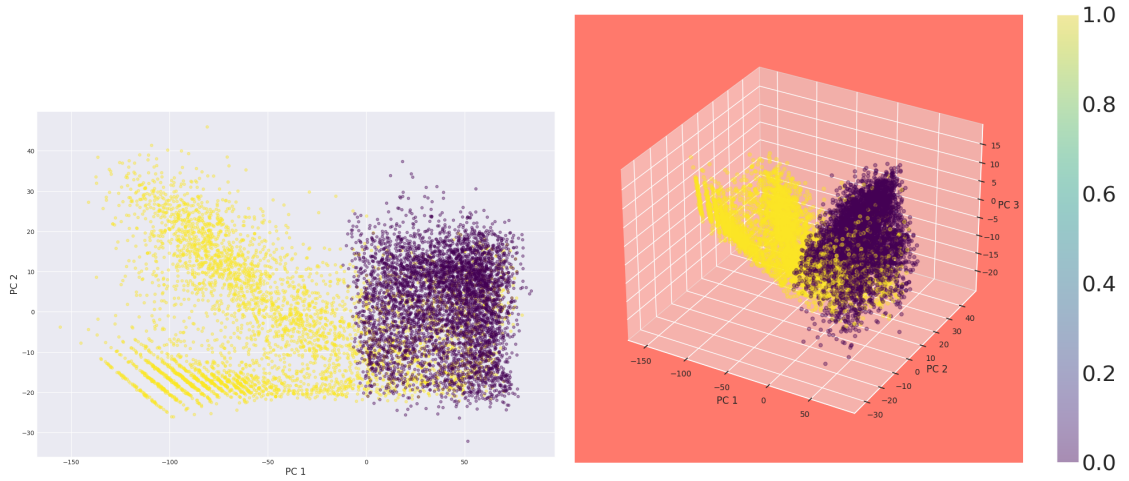
Table 4.3 - Explained Variance Ratio of Principal Components.

Cumulative	**PC 1**	**PC 2**	**PC 3**
≈ 93.62 %	0.93621037		
≈ 98.58 %		0.04959665	
≈ 99.70 %			0.01115244

SOURCE: The author.

The EVR results are an indicative that the features considered are highly correlated. This is somehow expected, since the point-wise measurements of hailstorms are necessarily over convective clouds then they should all represent small perturbations of the cloud top brightness temperature, unless in the very restrictive areas of overshooting top. There, it is expected some difference in the BT estimated from ABI water vapor channels in comparison to the IR window bands (see Section 2.1.2 for details). Then, the PCs could have been enough to depict a small difference in behavior of pixels supposedly associated with hailstorms or not. Therefore, we have decided to project the data in the space generated by the PC in order to possibly build an intuition on the separability of the observations belonging to distinct classes.

Figure 4.8 - Input data projected into PC space.



The input data projected into the space generated by 2 (left) and 3 (right) of the principal components. The classes of each input point (i.e hail or no-hail) are color encoded.

SOURCE: The author.

Figure 4.8 shows the spread of the input data in the space of the PCs. With 2 of the components the observations of both classes seem well mixed on the right portion of the plot. The third component reveal two clusters of observations of class 1 (Hailstorms) but there is still a region in which it is not possible to separate both classes. These visualizations indicated that with 99.7% of the information in the dataset preserved, there was no clear separation between classes in the data. With that, we suppose that even with the complete set, the algorithms would have trouble distinguishing the classes without contouring the data points too closely (overfitting). This result curbs the possibility of fitting a model to this data that would generalize well to new information.

Additionally, we have re-run this dimensionality reduction with a Multidimensional Scaling (MDS) algorithm (BORG; GROENEN, 2005) and PCA with the *kernel trick* (details in Section 3.14 and in Schölkopf et al. (1997)) with kernels of *Radial Basis Function* (EDWARD, 2006) and *T-distributed Stochastic Neighbour Embedding* (GISBRECHT et al., 2015). Despite generating new arrangements of the data in space, none of those approaches allowed a complete separation of the classes. The compactness of the 0-class points observed in all those attempts, is likely a conse-

quence of the threshold based approach for selecting negative examples. This choice has bounded the range of values, as previously noted in the analysis of Figure 4.1.

4.2.3 Challenges and lessons learned

The first experiment in this study brought up several challenges intrinsic to using ABI's Channels BT for hailstorm detection, alongside important lessons concerning data selection and model biases. Below are the main concerns raised from the analysis of the results that provided some insights for the subsequent activities.

Model Bias from Data Selection Criteria

The feature importance analysis revealed a significant reliance of the model on ABI's channel 13, suggesting that it may have learned the data selection criteria rather than the inherent characteristics of hailstorms. This channel was central to our sampling process; as described in Section 4.2.1, the training set's negative examples were selected based on BT thresholds in Channel 13, and similarly, the positive class observations were filtered using a BT criteria as well. This approach likely introduced a bias in the model, influencing its learning process and potentially affecting the ability to generalize the prediction.

Ineffectiveness of Existing Detection Methods

The literature suggests using the positive difference between BT in the Water Vapor strong absorption region (5.7 μm to 7.1 μm) and IR window (10.5μm to 12.5μm) to detect OTs (SCHMETZ et al., 1997) (detailed in Section 2.1.2). However, our experiments, as well as findings by Ribeiro et al. (2019), indicate that this method does not adequately capture the variability associated with updraft intensification. The tri-spectral method (STRABALA et al., 1994), while promising, appeared more effective for spatial classification rather than the localized, pixel-wise approach needed for our study. This limitation underscored the need for alternative methods that could more effectively differentiate hail-producing storm characteristics at a granular level.

Linear Dependence among Features

PCA of the data uncovered a strong linear dependence among most features. This suggests that deep convection likely homogenizes the BT values across all bands, rendering most unable to distinctly identify hail-producing storms from other severe weather types or merely high clouds. This uniformity in BT values across channels primarily reflected the cloud top temperature, subtly modified by atmospheric gas

absorption (WV and CO_2), and provided limited utility in distinguishing between the different meteorological phenomena.

Sampling Bias and Generalization Error

A possible sampling bias was identified in the selection of the no-hail class examples, which were chosen within a specific BT range. While this range aligns with physical expectations, it does not encompass all potential no-hail scenarios encountered in nature. Consequently, the model learned a criterion that, while effective within the confines of our dataset, does not adequately generalize to real-world conditions outside the specified BT range. This has highlighted the necessity for broader, more inclusive criteria in training data selection to enhance model robustness and reduce generalization errors.

4.3 Experiment 2: classification of storm polygons

On the grounds of the lessons learned from Experiment 1, for the second experiment we have changed the approach to consider spatial features instead of just pointwise assessments. Then, instead of using a pixel-wise BTs for classification, we decided to perform the classification based on aspects of the storm polygon.

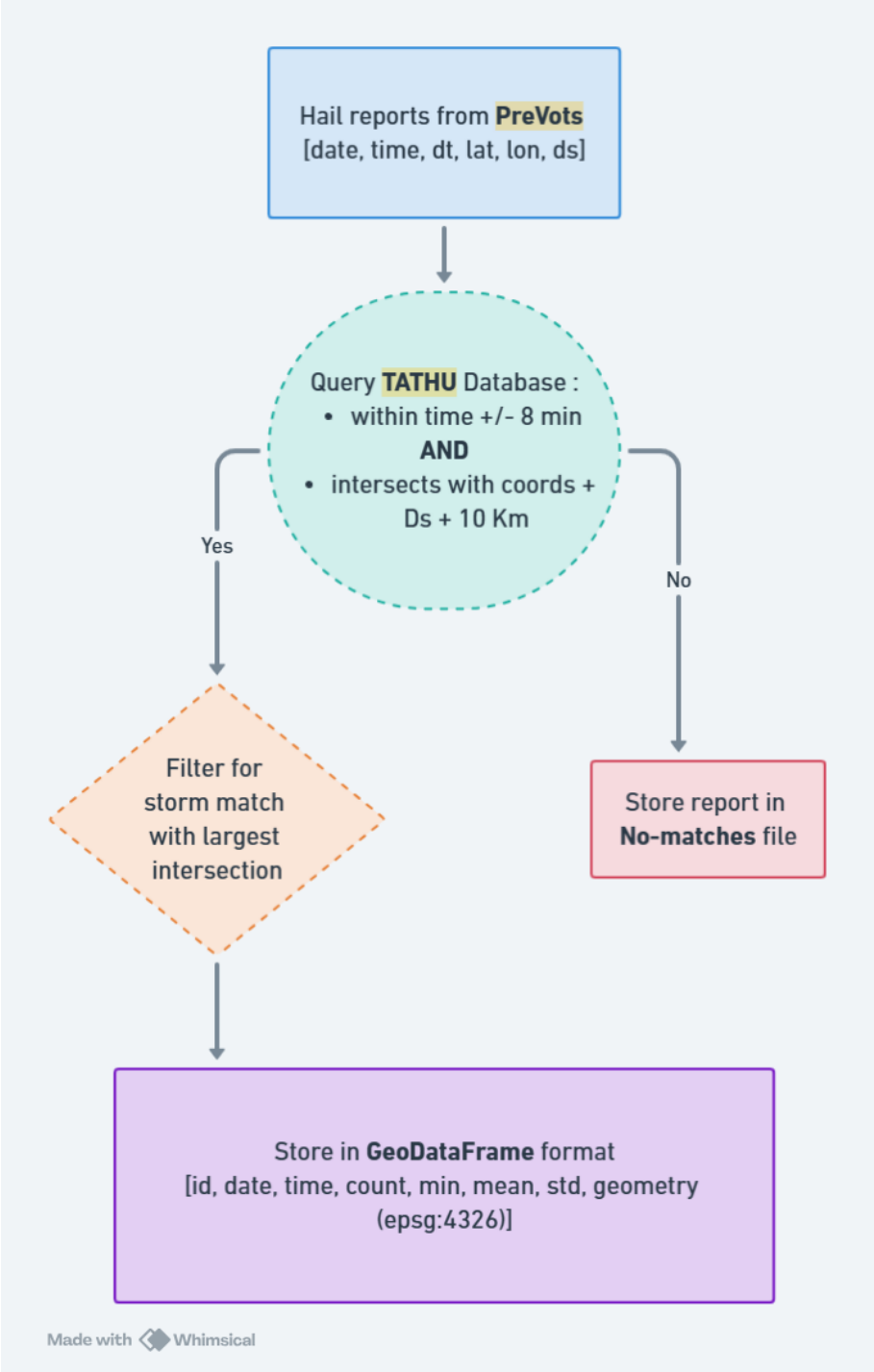
4.3.1 Hailstorm dataset construction

In order to build the dataset to be used in the algorithm training runs, the database constructed through the procedure described in Section 3.2.3 was filtered within the criteria for gathering both classes of storms. This process was generic and used soft criteria, resulting in a broad assortment of polygons.

The collection of hailstorm examples took two steps: a storm match procedure and a subsequent storm selection. Then, firstly a window of search was defined around the hail reports based on a time interval and a spatial buffer to filter the storm polygons therein. This time range, consisted of 8 minutes (i.e. the maximum time gap between two successive ABI scenes, modulo 2) before and after the report, plus the time uncertainty in the report metadata. Then, an area buffer was created centered in the report's reference coordinate by adding 10 km (converted to degrees) to the spatial uncertainty in Prevots reports. After that, TATHU's database was queried for polygons within the defined time window that intersected the spatial buffer, and sorted in decreasing order of intersection area. Then, the aspects of the storm polygon with the largest intersection was added to a table of positive examples using a GeoDataframe structure (GEOPANDAS, 2024), which is basically a set of tabular

values capable of holding geospatial information, in this case, the storm polygon of coordinates and its projection. This procedure was conducted with 5, 15, 20 and 25 km buffers before choosing the best value. The 5 km buffer was too picky for when the reports were tagged in the city center (condition described in Section 3.2.1), and the opposite happened with values over 20 km, causing an unreasonable number of matches per report. The complete flowchart of methods in this stage is shown in Figure 4.9.

Figure 4.9 - Flowchart of methods for defining positive examples of hailstorms.



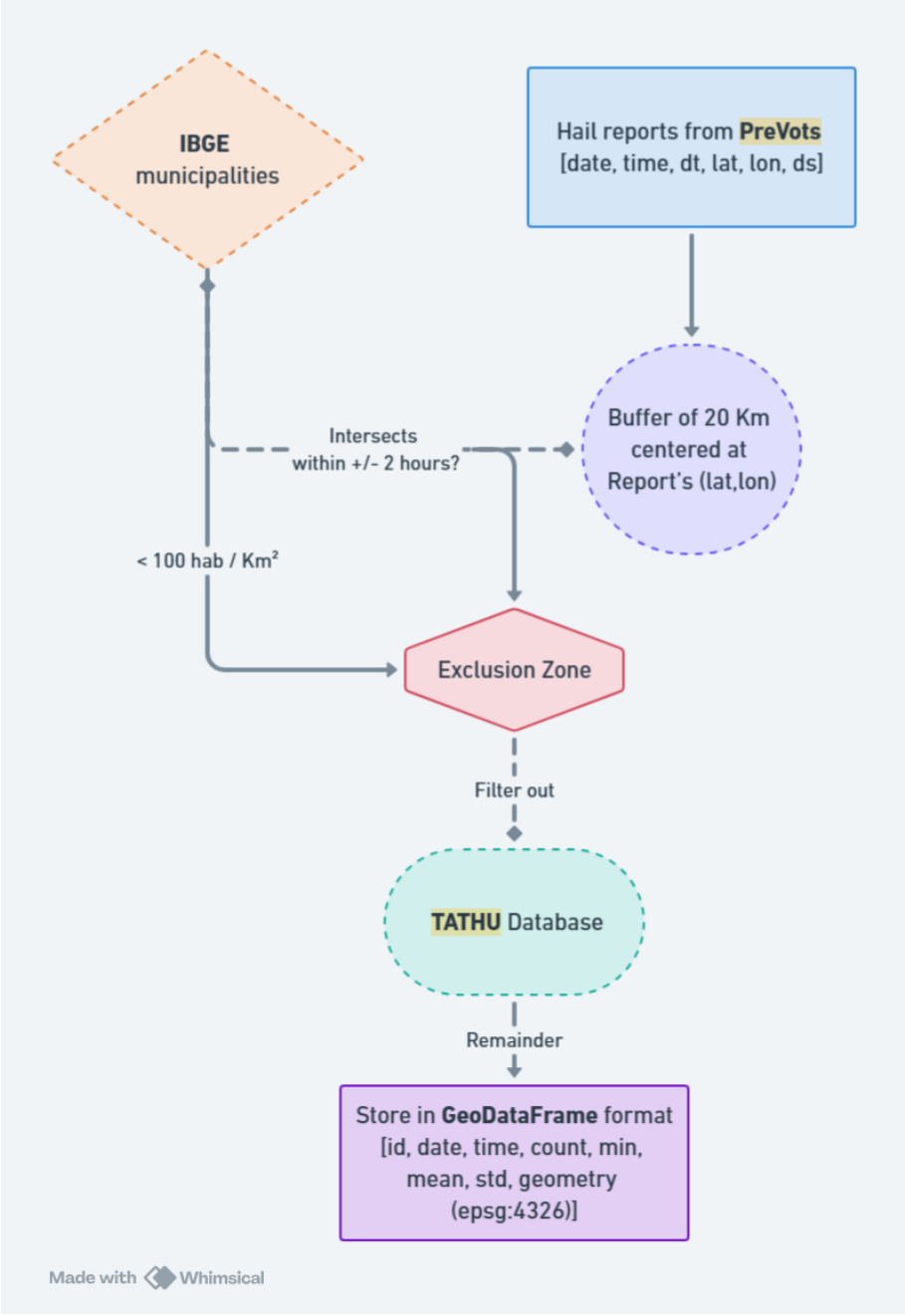
SOURCE: The author.

The selection of positive examples inserts some uncertainty in the training dataset. If there is a large uncertainty in report's time, the window of search likely comprises different stages of the same storm or even other CSs crossing the region in a day of active convection (DAWSON, 2017; BERTHET et al., 2013). In those cases, it is hard to define the proper stage of the storm that generated hail or even to determine the correct system. Another source of error arises in cases of report's large spatial uncertainty, which could match systems that are not associated with hailstorms. During the algorithm's training, some measures were taken to try to reduce those kind of uncertainties in the dataset.

The choice of negative examples was more meticulous than in Experiment 1, as well. The first step was to define some bounds to exclude possible previous or posterior stages of the reported hailstorms. Here, we have used the data from municipalities bounds from the Brazilian Institute of Geography and Statistics (IBGE) as an auxiliary asset. So, a buffer of 20 km was created around every hail report coordinate. Then, in each report, an exclusion zone was defined over every municipality that intersected with the this buffer or with the reports city. After that, we excluded every polygon that intersected with the exclusion zone at any time between 2 hours before or after the respective report.

After filtering the possible systems associated with hailstorms in different stages, we wanted to assure that the remaining were not possible cases of missed reports. For instance, if one of the CSs left was associated with a hailstorm in a low populated area, it is likely that it would not be reported (a problem described in Section 1.1). Therefore, the polygons that passed the previous scrutiny were further considered only if they intersected with municipalities with more than 100 inhabitants per squared kilometers. The complete procedure is depicted in Figure 4.10.

Figure 4.10 - Flowchart of methods for defining negative examples of hailstorms.



SOURCE: The author.

4.3.2 Class imbalance

The first step in the dataset preparation involved the removal of duplicate entries. This action was necessary because a single storm polygon could have generated multiple hail reports simultaneously. For the integrity and accuracy of our training set, it was important to account for each unique storm event only once, regardless of the number of hail reports it might have produced. This approach helped to refine the dataset by ensuring that each training example represented a distinct storm event.

Following the deduplication process, we observed a significant imbalance within the initial dataset. Specifically, it comprised approximately 10 % to 50 % more examples in the No-Hail class compared to the Hail class, depending on the time of the year. This is expected, since convective storms in Brazil more commonly do not cause hail reaching the surface. However, such an imbalance poses a challenge for machine learning algorithms, potentially leading to biased models that favor the overrepresented class. We addressed this issue through a random sampling strategy. This process involved randomly selecting a subset of examples from the overrepresented class to match the number of examples in the underrepresented one. Consequently, this measure ensured that the final training/test and validation sets contained an equal number of examples for both classes, thereby eliminating the initial bias towards the No-Hail class.

4.3.3 Model settings and training

In the end of the data preprocessing there were 13688 equally distributed examples of storms between 2018 and 2022, 20% of which (i.e. 2738) were separated for testing. For the current experiment we have selected as feature input TATHU's variables for BT (minimum, mean, standard deviation) and count, together with the storm's centroid coordinates. We have tried additional combinations using the normalized area difference using count as proxy for the area (KIDDER; HAAR, 1995) and the change in minimum BT, both compared to the previous state of the storm, but there was no increase in performance. This data is also in tabular format, pretty similar to TATHU's output.

The tabular format allowed us to choose for the current experiment another ensemble method that displayed very good performance in the trials of the prior experiment, the boosted trees from XGB library. Prior to the proper training, we have searched for the best model hyperparameters using the bayesian optimizer Optuna (see Sec-

tion 3.3.4.2). The parameters optimized together with their respective values are described in Table 4.4.

Table 4.4 - Hyperparameters Tuned for the XGBoost Model.

Parameter	Type	Range/Values	Log Scale
booster	Categorical	gbtree, gblinear, dart	No
n_estimators	Integer	200 to 1000	No
max_depth * **	Integer	3 to 25, step 2	No
subsample	Float	0.2 to 1.0	No
lambda	Float	1e-8 to 1.0	Yes
colsample_bytree	Float	0.2 to 1.0	No
min_child_weight * **	Integer	2 to 10	No
eta * **	Float	1e-8 to 1.0	Yes
gamma * **	Float	1e-8 to 1.0	Yes
grow_policy * **	Categorical	depthwise, lossguide	No
sample_type **	Categorical	uniform, weighted	No
normalize_type **	Categorical	tree, forest	No
rate_drop **	Float	1e-8 to 1.0	Yes
skip_drop **	Float	1e-8 to 1.0	Yes

* Specific for **gbtree** booster
** Specific for **dart** booster

SOURCE: The author.

The Table 4.5 displays the best hyperparameter setting from the Optuna trials, along with a short description of what each parameter does in the context of an XGBoost model. Overall, the **booster** represents the type of model to train at each iteration, where *gbtree* uses tree-based models. The **n_estimators** parameter is the number of trees to construct. More trees will usually result in better performance but can lead to overfitting. In **subsample** is defined the fraction of samples to use for each tree's growth. Smaller values make the algorithm more conservative and prevent overfitting but too small values might lead to under-fitting. The **lambda** aspect, consists of a L2 regularization term on weights (BüHLMANN; YU, 2003) - increasing this value will make model more conservative. Tuning **colsample_bytree**, determines the fraction of features to use per tree. This parameter can be used to deal with feature correlations (Section 3.3.3.1.1). The **max_depth** parameter puts a limit on the maximum depth of each tree, which controls the complexity of the model. The deeper the tree, the more specific it can get. A **min_child_weight** is the minimum

sum of instance weight needed in a child, so if the tree partition step results in a leaf node with the sum of instance weight less than the selected value, then the building process will give up further partitioning. This implies that the larger this parameter the more conservative the algorithm will be. The **eta** is analogous to a learning rate, it makes the model more robust by shrinking the weights on each step. The value of **gamma**, sets the minimum of reduction in the loss function necessary for a node to be split. Therefore, higher values lead to fewer splits. A **grow_policy** controls how trees are grown, then *depthwise* prioritizes nodes with the best split according to their depth, which balances the tree growth. The model was then constructed with this parameters and trained using the GPU for performance improvement.

Table 4.5 - Best Hyperparameters for the XGBoost Model.

Parameter	Value	Description
booster	gbtree	Type of model to run at each iteration (gbtree or gblinear)
n_estimators	894	Number of boosting rounds or trees to build
subsample	0.9995	Subsample ratio of the training instances
lambda	0.0100	L2 regularization term on weights (lambda)
colsample_bytree	0.7851	Subsample ratio of columns when constructing each tree
max_depth	15	Maximum depth of a tree
min_child_weight	2	Minimum sum of instance weight (hessian) needed in a child
eta	0.7110	Step size shrinkage used to prevent overfitting
gamma	0.0040	Minimum loss reduction required to make a further partition
grow_policy	depthwise	Strategy used to grow trees

SOURCE: The author.

4.3.4 Model performance

The model performance was assessed through the metrics described in Section 3.3.5.2. During training, the model reached 94% accuracy when compared to the testing set, suggesting that the model performs quite well. Then, the separate dataset - unseen by the model - was used for validation. The performance in classifying the storms for year 2023 are displayed in Table 4.6.

Table 4.6 - XGB classification scores in the validation set (2023).

Class Type	Eval. Metric			# of Cases
	Precision	Recall	F1-score	
No-Hail	0.92	0.93	0.92	789
Hail	0.93	0.92	0.92	789

SOURCE: The author.

From Table 4.6, one can see that the model achieved an overall accuracy of 92% on the unseen set of storm polygons. This is a strong performance, indicating that in 92% of cases, it correctly predicted whether a storm polygon was of type *no-hail* or type *hail*. Based on the precision for **Class 0** (0.92), it means that when the model predicts a polygon as type 0, it is correct 92% of the time. For polygons of class hail, the performance is slightly better (0.93). From the Recall values one can state that of all actual **Class 0** polygons, the model successfully identifies 93% of them, and for **Class 1** polygons, the model identifies 92% of them correctly.

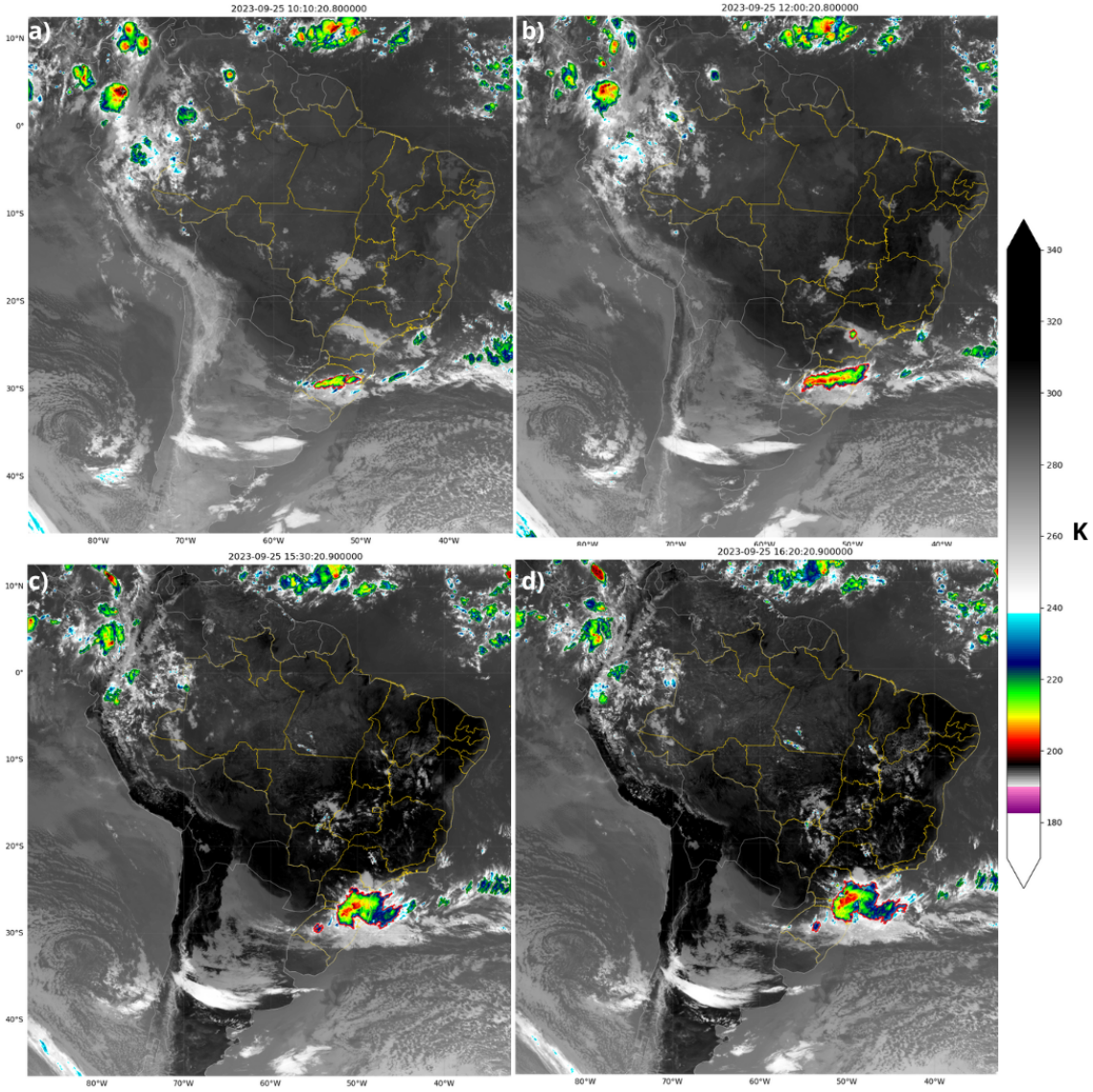
F1-Score for Class 0 and Class 1 both are 0.92, which is a harmonic mean of precision and recall. This high F1-score indicates a good balance between precision and recall for both classes. There is no indication of the model favoring one class over the other, which is particularly important in applications where both classes are equally important. The almost identical performance metrics for both classes suggest that the model is well-calibrated and does not exhibit bias toward either class, despite being equally distributed (789 instances each in the test set).

These results on an unseen dataset, provides evidence that the model generalizes well to new data. The balanced metrics across both classes indicate robustness, which is important for reliable predictions in operational settings. Although the performance is already high, one might consider experimenting with further tuning or incorporating additional features that could help improve the edge cases where the model fails.

In Figure 4.11 there is an example of using the model for classification in a case of a sequence of hail reports in the Southern region of Brazil on September 25, 2023. This is one of the cases in the validation set, which means the whole life-cycle of the storms displayed have been tracked and it is the first time the model is looking at it. On this day there were 18 hail reports over Rio Grande do Sul (RS), Santa Catarina

(SC) and Paraná (PR) states between 5:30 and 19:10 UTC. Panel **a** displays the polygon of the storm that caused at least two the reports in RS which was correctly classified by the model. This polygon remains with the same classification for the whole time period comprehended by the events. Panel **b** is two hours apart from the first one and exhibits the deep convection extending into SC. Interestingly, a small convective cluster close to the northern border of PR is also classified by the model as a region of potential hail. We have no knowledge of severe weather related to this tracked cell. Panels **c** and **d** display the moments the main tracked storm caused hail reports in different regions of SC, prior to causing severe weather in large municipalities of PR. The model also classifies a small storm polygon in RS which was not tied to any report before merging into the main cell.

Figure 4.11 - Detection and classification examples.



Different scenes of GOES-16 ABI 10.35 μm BT on September 25, 2023, close in time to several hailstorm reports over the states of Rio Grande do Sul (RS), Santa Catarina (SC) and Paraná (PR): a) 10:10 UTC, close to reports in Canudos do Vale - RS and São Francisco de Assis - RS; b) 12:00 UTC, close to reports in Vacaria - RS and Santiago - RS; c) 15:30 UTC, close to a report in Lebon Regis - SC and d) 16:20 UTC, close to report in Caçador - SC.

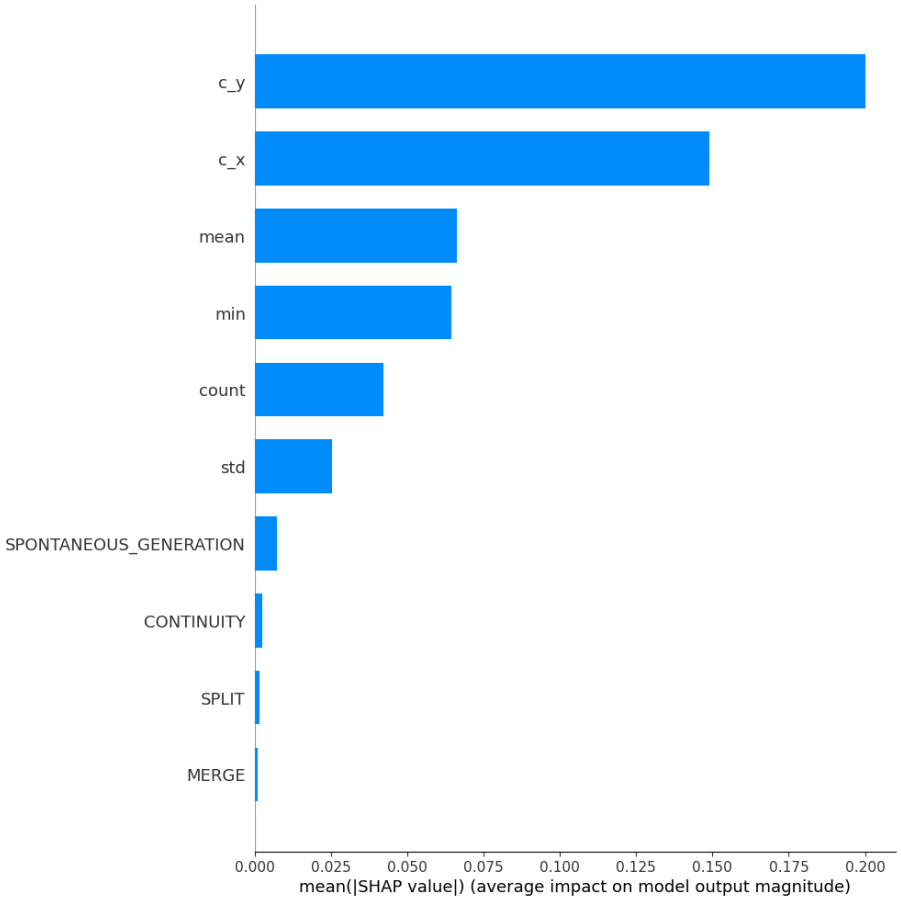
SOURCE: The author.

4.3.5 SHAP analysis and Dependence Plots

In order to gain deeper insights into the decision-making process of our XGBoost model and to understand the contribution of individual features to the prediction outcomes, we employed the SHAP library. As described in Section 3.4.1, it decomposes a prediction into contributions from each feature, thereby providing a more detailed understanding of the model's behavior.

To understand the impact of each feature on the model's predictions, we visualized the mean absolute SHAP values in a bar plot (Figure 4.12). This plot serves as a tool to quantify and compare the influence of each input variable on the predictive outcomes of our XGBoost model. In the plot, the features are arranged on the y-axis in descending order of importance, indicating their relative contribution to the model's decisions. The x-axis represents the mean absolute SHAP value for each feature, providing a measure of the average impact of a feature's change in value on the model's output. Each bar extends to the right, indicating the extent to which a particular feature drives the model's predictions. A longer bar signifies a greater impact, suggesting that variations in this feature are highly influential in determining the output of the model. This ordering and representation help identify the features that the model relies the most on for making predictions, giving insights into the data dynamics captured by the model.

Figure 4.12 - General Feature Importance.



From top to bottom, mean of absolute SHAP values in descending order. The most important features for the model classification are on top (y-axis) and the bars indicate their impact on the classification (x-axis). Here, c_x and c_y denote the storms polygons' centroid longitude and latitude, respectively.

SOURCE: The author.

From the plot analysis we observed that the central coordinates of the storms are the most impactful features on the model's output, suggesting that the intensity of other input variables depends significantly on the geographic area. This geographic dependency likely reflects the environmental conditions beneath the cloud and surrounding the storm, which are determining for the presence of hail and the chance of it reaching the surface. Among the coordinates, latitude matters the most. This aligns with literature findings that the frequency and intensity of hailstorms worldwide are indeed highly dependent on latitude, reflecting climatological and

environmental variations across different geographic regions.

Following closely in importance are the storm's mean and minimum brightness temperatures. These features, nearly equivalent in their impact, are closely tied to the storm's development, cloud phase, and updraft intensity, all key factors in the production of large hail.

The pixel count within the storm polygon, serving as a proxy for storm size, ranks next in importance. The size is indicative of the type of storm systems, with larger storms typically associated with broader weather systems such as synoptic disturbances, and smaller, more localized convective systems. Frequent CSs tied to cold fronts in southern Brazil may be comparable in size to convection associated with the South Atlantic Convergence Zone (SACZ) crossing the whole country. This is one of the scenarios in which the coordinates of the storm's centroid may help distinguish between CSs likely to generate hailfall or not.

The standard deviation of the BT also appears as a critical feature. It reflects the variable, "bubbly" appearance of storms with strong updrafts. This is useful for distinguishing between uniform high-altitude clouds with smooth appearance from the satellite view and active storm systems displaying intense convection features.

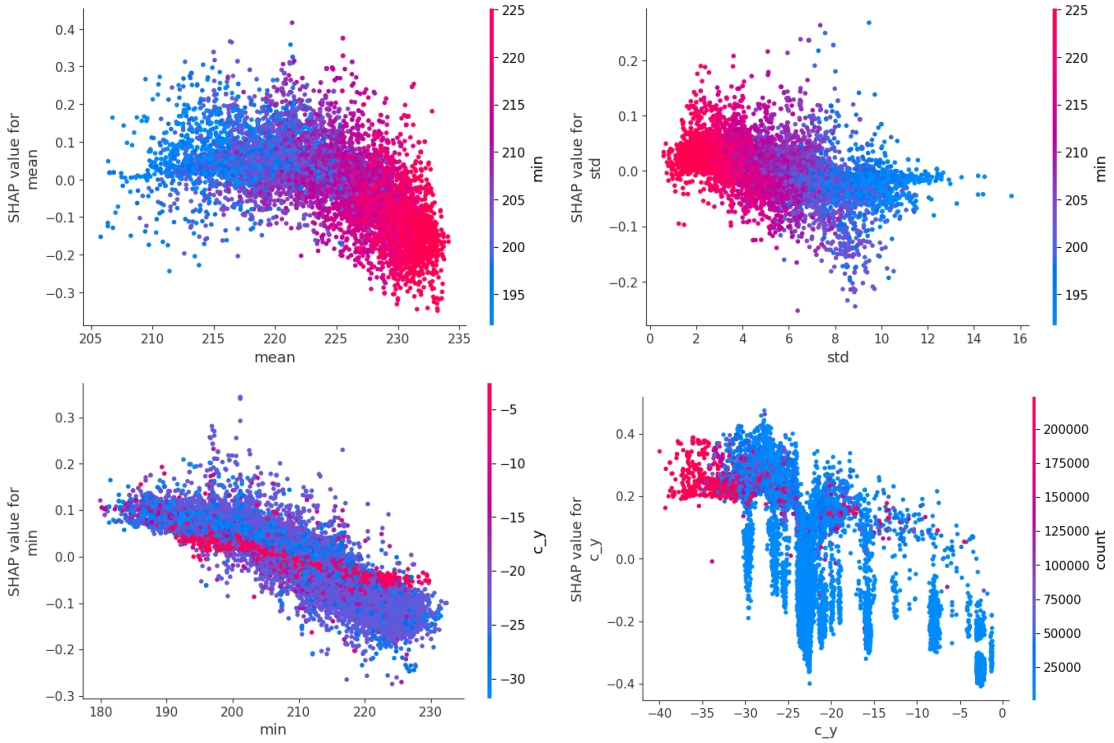
Lastly, among TATHU event types associated with storm detection, the "spontaneous generation" of storms, marking the earliest stages of convection, shows the largest importance for the classification. This type is less likely to produce hail, contrasting with the "continuity" of storms, which correlates more with sustained convection necessary for hail growth. However, these event types still demonstrate significantly lower importance compared to the geographic and meteorological features mentioned above.

Then, we utilize Dependence Plots derived from SHAP to visualize the effect of specific features on the prediction outcome. The Dependence Plots are visualizations that use SHAP values to show the effect of a single feature across the whole dataset on the predicted outcome. These plots provide insights into the relationships and interactions between features within the model, especially in cases that are otherwise difficult to interpret, such as ensemble methods.

Figure 4.13 displays a selected set of Dependence Plots among the most interesting features. Each plot focuses on a single feature and depicts how changes in the value of that feature influence the prediction output of the model. This is measured in

terms of SHAP values. Here, a second feature (colored) is used to show interactions. This means each plot shows how one feature impacts the prediction and how this impact changes depending on the value of another feature. The X-axis represents the value of the feature being analyzed, and the Y-axis shows the SHAP value associated with each feature value. This indicates how much each value of the feature pushes the model output higher or lower, that is, values above and below zero are pushing the model in different directions. With that, this visualization can reveal nonlinear relationships, such as a feature being positively correlated with the target variable up to a certain point and then negatively correlated beyond that point.

Figure 4.13 - Partial Dependence Plots.



Dependence plot to show the effect of specific features on the prediction outcome, while holding other features constant. In every plot the y-axis displays the SHAP value as a function of the feature value on the x-axis. The values for the feature to be compared are color encoded and described by the color bar on the right. The SHAP values in the y-axis above and below 0.0 indicate opposite contribution for the model prediction, as positive ones contribute for the class hail and negative for the class no-hail.

SOURCE: The author.

The top-left panel in Figure 4.13 illustrates the dependence plot of the mean brightness temperature (BT) versus the minimum BT. These features are inherently interrelated, as lower values of one typically accompany lower values of the other. Notably, the influence of mean BT changes significantly for values above 225 Kelvin. This relationship suggests that storms with mean BTs above this threshold, coupled with minimum BTs over 215 Kelvin, are substantially less likely to be associated with hailstorms.

In the top-right panel, the plot of BT's standard deviation (SD) versus minimum BT is displayed. It is observed that smaller values of minimum BT correspond with larger SDs, potentially indicative of clouds with overshooting tops or a "bubbly" appearance, as well as storms exhibiting a steep BT gradient. Interestingly, SD values above approximately 7 Kelvin are associated with negative SHAP values, indicating a higher likelihood of predictions for the no-hail class. SD values from 2 to 7 Kelvin, however, are more indicative of the hail-positive class, reflecting various storm stages. Larger SDs suggest vigorous convection, typical of storms in the development stage with growing hail. Conversely, very small SDs may denote a smoother cloud top, possibly indicative of dissipating storms or those in the early stages of development. However, given the BT threshold for storm detection, input storms are unlikely to be in the very early stages of convection development.

The bottom-left panel presents the relationship between minimum BT and the latitude of storm centroids (c_y). It is observed that for storms located at latitudes below 15°S, minimum BT has minimal impact on classification due to the rarity of hailstorm events in tropical regions. Notably, unless there is exceptionally deep convection with cloud top minimum BTs below 190 Kelvin, these polygons are unlikely to contribute to a class 1 (hail) classification. In fact, storms at latitudes below approximately 20°S with minimum BTs above 205 Kelvin are almost invariably predicted as class 0. The scattered cases near 20°S where minimum BT shows high SHAP values indicate a strong dependency on additional features, suggesting that under certain conditions, minimum BT becomes critical for prediction.

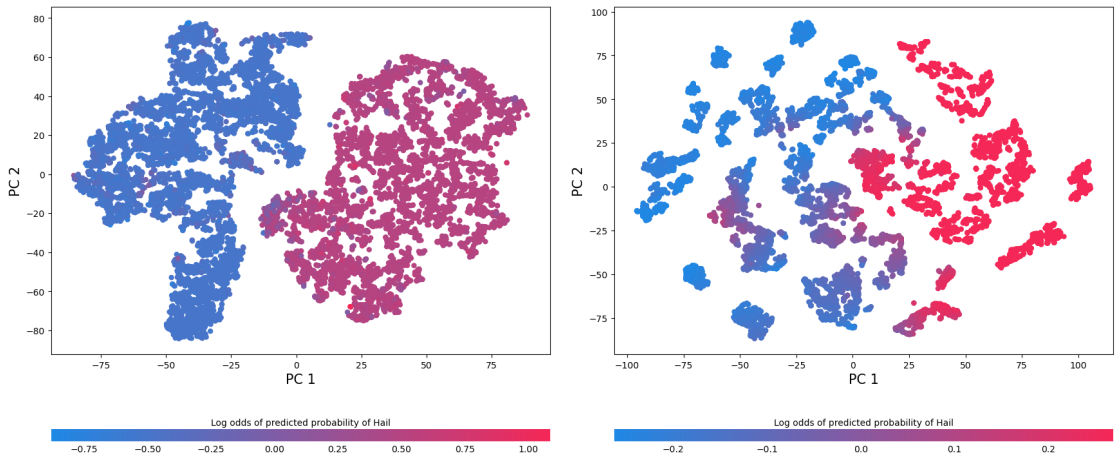
Finally, the bottom-right panel examines the relationship between the latitude of storm centroids and storm size. It is observed that larger storms predominantly occur at latitudes over 30°S. Furthermore, storms at latitudes below 20°S are almost always associated with predictions of the non-hail class, unless the storm systems are exceptionally large. In such cases, storm cell area serves as a proxy for intensity, or possibly for scenarios involving the South Atlantic Convergence Zone (SACZ),

where storm cells may extend into regions to the south 20°S.

4.3.6 Implications of hailstorm detection

The predictive capability of the XGBoost model was further investigated through a dimensionality reduction of the SHAP values matrix, which consisted of the number of observations corresponding to rows and ten input features represented by columns. Initially, Principal Component Analysis (PCA) was applied to reduce the matrix into four principal components, followed by t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce it further into two components. The transformed SHAP values were then visualized in a two-component t-SNE plots in Figure 4.14, with points colored according to the "log odds of predicted probability of hail" as determined by the model.

Figure 4.14 - Dimensionality reduction of SHAP values.



Dimensionality reduction of SHAP values for model predictions using the complete set of features (left) and a subset of the features without the latitude and longitude features. The SHAP value of each measurement is color encoded. The two dimensional plots highlight the proximity of measurements in feature space and how they contribute for model predictions.

SOURCE: The author.

This visualization (left panel in Figure 4.14) distinctly revealed two clusters of data: one predominantly characterized by negative log odds and the other by positive log

odds, indicating a clear separation in the model's predictions associated with non-hail and hail events, respectively. Such a delineation in the t-SNE plot underscores the effectiveness of the model in classifying the outcomes based on the SHAP values derived from the input features.

To assess the contribution of individual features to this classification accuracy, the same dimensionality reduction process was applied to a subset of eight features, excluding the latitude and longitude of the storm's centroid. The process to achieve that, was identical. That is, using the same training and test set but without the coordinates information, the model hyperparameters were adjusted through bayesian optimazion. The model was then trained and validated accordingly.

The resulting t-SNE plot from this reduced feature set (right panel of Figure 4.14) showed less defined clustering compared to the full ten-feature model. This observation suggests that the excluded features, particularly the geographical coordinates, significantly enhance the model's ability to discriminate between hail and non-hail events. Therefore, it is evident that the complete set of ten features, including the latitude and longitude data, markedly improves the model's classification performance.

The inclusion of the storm's centroid coordinates is particularly vital, as these serve as proxies for environmental and atmospheric conditions that are critical for hail formation yet are not directly available from the dataset. These coordinates help infer the storm's intensity and the environmental context, factors that are essential for determining whether hail produced within a storm will melt before reaching the surface.

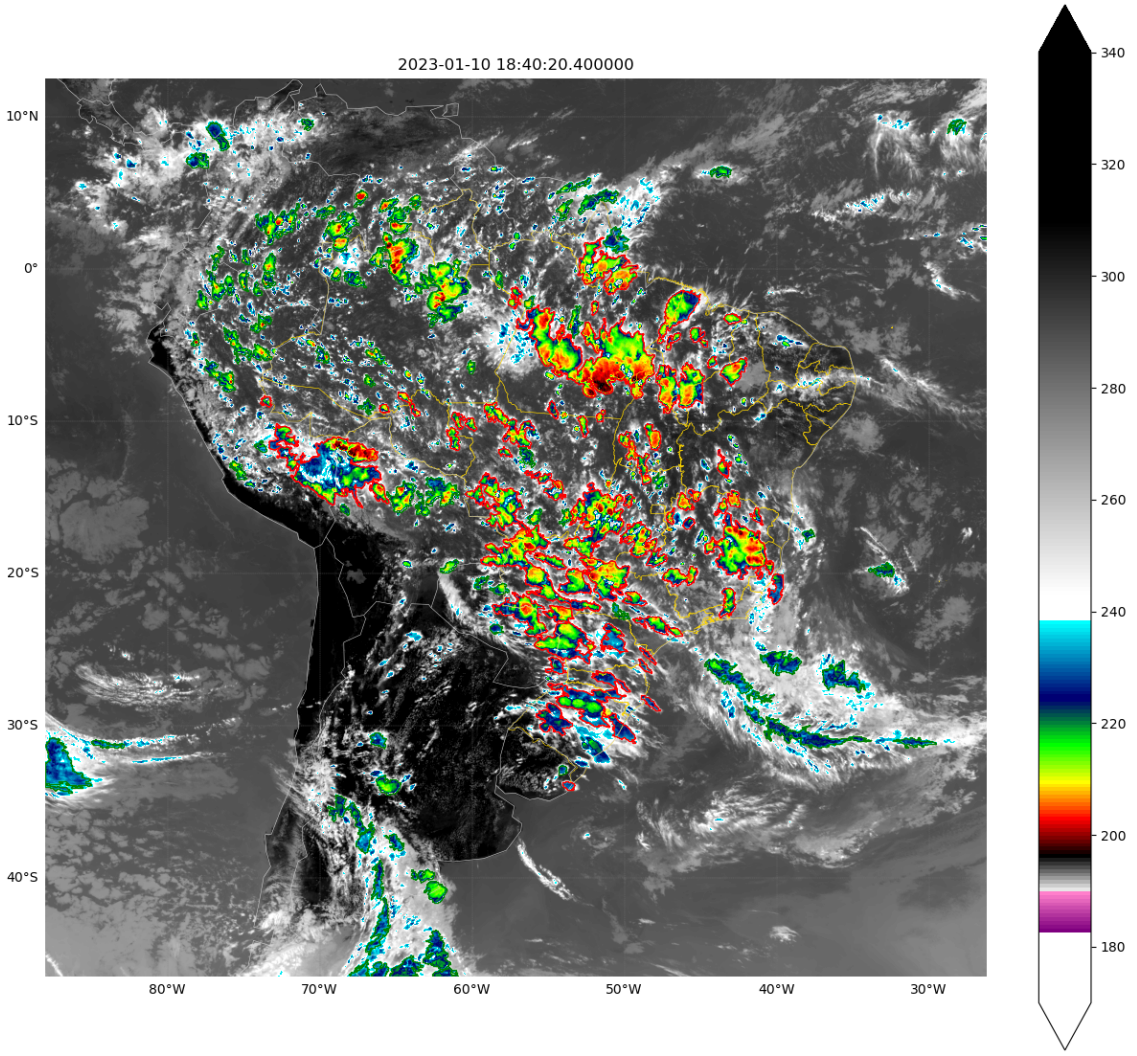
This analysis highlights the importance of including geographical information to better model hail occurrences. Thus, it is crucial for future model enhancements and data collection efforts to consider additional environmental variables surrounding and beneath the storm, which could provide further insights into the conditions favoring hail formation. The use of geographical information may have helped the detection of storm polygons likely to be associated with hailfall but it does not mean this would work for segments of the storm, for instance.

4.3.7 Challenges and lessons learned

It was observed a degradation in performance of the model for classification in Brazilian region to the north of 20 °S latitude, especially in days of strong deep

convective activity, such as the one shown in Figure 4.15. The day depicted, marked the end of a SACZ episode (CLIMATEMPO, 2023). Accordingly, there is a spread of convective activity over a large portion of the study domain.

Figure 4.15 - Example of a day of strong convective activity.



GOES-16 ABI 10.35μm BT in Kelvin from January 10, 2023 at 18:40 UTC. The model's classification is color encoded: contoured in green are polygons of No-hail class; in red are storm polygons likely to be associated with hailstorms.

SOURCE: The author.

In this particular day, there were at least 18 hailstorm reports extending over the

Southern region, in the states of Santa Catarina and Rio Grande do Sul, close in time to this satellite scene. However, there are no reports in other brazilian regions. Interestingly, both in January 9, and January 11, 2023 there were hailstorms reported in the state of Piaui, in Brazilian Northeast. Then, there could possibly exist a potential for hailstorms in the deep convection developed in Brazilian Northern and Northeastern portions in this day. From the BT features in the image, we suspect that some of the detected storms of hail class could have updraft currents strong enough to produce hail. However, since there is no report in most of the regions, we are prone to believe that the hail producing storms in there would have to sustain strong updraft currents for longer periods, in order to develop hailstones large enough not to melt before reaching the surface. This aspect requires the model to look at a time dependent feature that is not regarded in its training.

This could be a problem of unrepresentative sampling for those regions, as well. In general, there is little number of hailstorm reports in the training set for areas of latitude North of 20 °S. The low population density over certain areas together with the low frequency of such events as an aspect of local climate could impact the model training. The implications in these cases, as described in Section 3.3.1, is that the available sample might not represent the behavior of the true nature of the variable. This implies that few examples were provided during training for the model to learn well the difference between hailstorms South and North of this latitude bound.

Still on these grounds, Ni et al. (2017) combined a 16-years record from precipitation radar and microwave imager aboard TRMM with the GPM dataset and found general high hail frequency estimates in brazilian tropical region and Central Africa. Both estimates indicate the presence of hail in the clouds but could not be verified with ground hail reports. Allen et al. (2020) states that in these areas, hail is likely to melt on its way to the ground, except for high altitude locations (not the case for the Amazon rain-forest nor the Brazilian Northern region). The same limitation is found in Cecil and Blankenship (2012), and the proposed solution was to constrain the frequency estimates to regions that are known to experience hail (BANG; CECIL, 2019). Therefore, it is possible that our algorithm is correctly identifying hailstorm alike signatures in polygons over the Brazilian tropical region, but the lack of proxy information such as melting level height and sub-cloud atmospheric profile is necessary for determining if the hail would reach the ground before melting.

Another challenge that emerged from the TATHU tracking was the behavior of merging and splitting between subsequent measurements. Once the shape of the system

suffers such an abrupt change it is hard to keep track of the temporal evolution of the features of interest. In particular, the change in area over time, as described in the literature, matters for detecting severe weather. The commonly used "normalized area difference" of tracked storm polygons can reveal different stages of the storm lifecycle and depict the updraft intensity (KIDDER; HAAR, 1995). Additionally, once there is a sudden change in the bounds of the system, naturally the distribution of the brightness temperature is also affected, and split and merge events can also cause major changes in those values.

5 DISCUSSION

5.1 Comparative analysis of experiments

In Experiment 1, we employed a pixel-wise classification approach using brightness temperature (BT) values from channels 7 to 16 of the Advanced Baseline Imager (ABI) on the GOES-16 satellite. This approach relied heavily on matching pixel data to crowd-sourced hail reports, with an emphasis on temporal alignment and precise BT criteria to define hail and no-hail classes. The performance was benchmarked across multiple models, with ensemble methods like Random Forest and XGBoost showing superior classification efficacy.

Experiment 2, in contrast, utilized a more holistic approach by classifying storm polygons based on several features derived from TATHU's detection and tracking using ABI's channel 13. This method integrated spatial and temporal aspects of storms more comprehensively, leveraging both morfological and radiative properties of the storms as well as environmental conditions implied by their locations.

The comparative analysis highlights the advantages of using storm polygons over pixel-wise classification for the context of hail detection, based on the "out-of-sample error" aspect. The polygon-based approach in Experiment 2 facilitated a more nuanced understanding of storm dynamics, leading to better generalization capabilities in the model, as evidenced by the balanced accuracy and precision in classifying hail versus no-hail events. This suggests that incorporating broader contextual features, along with geographical data, provides a more robust framework for hailstorm detection.

5.2 Comparison with previous studies

Our results suggests we have achieved better skill scores than studies with similar propositions in Europe (MERINO et al., 2014; MELCÓN et al., 2016). Additionally, as far as our knowledge covers, there is only one similar approach for detecting general deep convection with use of machine learning in South America, in Cintineo et al. (2020a). Their approach uses ABI's 0.64 μm channel reflectance together with 10.35 μm channel's BT and GLM's flash extent density as input to a *Convolutional Neural Network* (CNN). As per personal communication, we have knowledge of them trying to detect deep convection cases in South America but no information on its performance. Naturally, the classification performed through this approach is restricted to sun hours due to the use of ABI's "Red" band.

It is worth mentioning thought that CNNs excel in capturing spatial patterns and could be extended to handle temporal dependencies through LSTM layers or self-attention mechanisms, but they generally lack the interpretability offered by the proposed methods here. Additionally, CNNs may not always provide the same level of performance when dealing with highly dimensional and heterogeneous data typical of geospatial and meteorological datasets.

In selecting the most appropriate methods for our analysis, the balance between algorithm complexity, interpretability, and performance was carefully considered. Tree-based methods, particularly ensemble models, emerged as the preferred approach due to their robust handling of the specific challenges posed by the available data, without compromising on interpretability and performance.

5.3 Theoretical and practical implications

The findings from these experiments have several implications for both meteorological theory and practical applications in weather forecasting.

The main results underscore the importance of considering spatial and temporal contexts in modeling storm behavior, supporting theories that emphasize the role of environmental conditions in hail formation and melting due to sub-cloud environment. By integrating geographical coordinates and storm characteristics, we enhance our understanding of regional dependence on hail formation and hailstorm development.

For operational weather forecasting, the model developed through these experiments offer a framework for a reliable automated hail detection. This can aid in disaster preparedness and response strategies, particularly in regions prone to severe hailstorms.

Additionally, one of the main benefits of using satellite data for the current study is that sensors on board of satellites are capable of sampling large regions consistently. Differences in measurements arise from the time to time replacement of sensors as systems age or retire, which could bring inconsistencies to large time-series datasets (ALLEN et al., 2020). Though there is no proper set of hail reports that would enable the validation of the algorithm on the classification of data of previous generations of GOES satellites, as shown by the study case using GOES-12 data, we are eager to believe that the algorithm achieved in Experiment 2 is capable of providing a time-series of consistent classifications that extend to previous generations of the Imager

sensor as long as CSs can be detected and tracked under TATHU's framework.

5.4 Limitations and challenges

As stated in Section 2.2 the atmospheric profile below the melting level significantly impacts the hail size distribution reaching the surface. Those parameters cannot be "seen" by the model due to physical bounds imposed by the measurements available for input. Then, in regions in which this information is potentially crucial for determining the chance of hail reaching the surface, for instance over the Brazilian tropical region and Central Africa (details in Section 4.3.7), the proposed methods are prone to underperform.

The hail-size distribution produced by a storm is determined by several microphysical and dynamical processes inside the thunderstorm (HEYMSFIELD, 1982; KNIGHT; KNIGHT, 2001; GARCÍA-ORTEGA et al., 2012; MERINO et al., 2014). Simultaneously, the sub-storm environment seems to mostly impact smaller to middle range stones (e.g. less than 11 mm in diameter) (DESSENS et al., 2015). However, the net effect of this interplay between intra-storm aspects and the atmospheric profile on the hail trajectory below the melting level, on the hail-size distribution reaching the surface, is strongly dependent on regional characteristics (DESSENS et al., 2015; DESSENS et al., 2007). Then, if these intra-cloud aspects that determine the hail-size distribution produced by the storm can be translated into something observable through ABI measurements - and we have reasons to believe it is (2.1.2) - the main challenge would be to gather a representative amount of data over the aforementioned problematic regions.

Still in this context, this **data dependency** should be considered. Both experiments heavily relied on the quality and availability of ABI data and crowd-sourced reports, which can be inconsistent and biased towards regions with higher population densities. This affects the models' ability to learn from less-reported areas, potentially skewing prediction accuracy. Although Experiment 2 showed improved generalization over Experiment 1, the models are still limited by the **representativeness** of the training data. Regions with less frequent hail events or underrepresented environmental conditions may not be well modeled, leading to potential overfitting to more commonly reported conditions.

Moving forward, addressing these limitations will require a multifaceted approach that includes expanding the dataset to cover a broader range of conditions, improving model robustness against overfitting, and experimenting with more complex models.

6 CONCLUSION AND FUTURE WORK

The present study developed a set of machine learning approaches for detecting hailstorm systems through the data of ABI GOES-16 sensor. The best performing framework leverages the processing capabilities of TATHU for standardizing the classification approach; bringing consistency for studies of long time-series of GOES payload measurements; and minimizes the necessity of on-premises data storage of large sets; since it only requires TATHU variables as input.

Simultaneously, the model's simplicity requires a small dataset for training and performs predictions in less than 20 seconds. This high performance enables standardized assessments over large databases of systems. Overall, a good performance was achieved during training/validation. A further inspection of the model's prediction to "ground truth" cases indicated no sign of a possible lack of generality by overfitting the training set.

Furthermore, the model's straightforwardness allowed a knowledge extraction by exploring the importance given to the input features. Thereby, it was feasible to perceive that it captured hailstorm characteristics available among the features and that it most likely works under physical constraints imposed by the type of measurements used as input.

All things considered, the model achieved in Experiment 2 is able to detect isolated cells that might be associated with hailstorms. However, it does not suffice to detach just small parts of large systems linked to hail reports.

Additionally, the proposed framework is easily adaptable to other regions: all of operational imagers on-board of geostationary satellites perform measurements in the clean-infrared region. A feature exploring study can be done as well following the same steps with small changes to the code. The model can be as well re-used with a fine tuning over a specific area or measurement (e.g. previous series of measurements from sensors operating in the same spectral region).

6.1 Recommendations for future research

The research conducted has laid a foundational understanding of classifying storm polygons into hailstorm occurrences or non-occurrences based on the brightness temperature readings from the Infrared Window channel of the Advanced Baseline Imager on board the GOES-16 satellite. Building upon the insights and findings from this study, several recommendations for future research endeavors are proposed to

enhance the predictive capabilities and applicability of storm classification models:

Segmentation of storm polygons: Future studies could benefit significantly from focusing on the segmentation of storm polygons to isolate regions within a polygon most likely to generate hailstorms. Storm polygons can encompass extensive areas, covering multiple regions of interest, not all of which are affected by hailfall. Precise segmentation would allow for a more targeted prediction and resource allocation, potentially increasing the accuracy and utility of storm forecasts.

Exploration of more complex models: The complexity of the models in this research was constrained by the volume of data available, to avoid overfitting. With the anticipation of more extensive datasets becoming available in the future, it becomes feasible to explore more sophisticated modeling techniques. Advanced models could offer improved accuracy and a broader generalization capability across larger areas, enhancing the predictive performance of hailstorm classification.

Focused analysis on regions southern of 20ºS Latitude: Given the higher frequency of damaging hailstorm events in the southern regions of Brazil, particularly south of 20ºS latitude, a separate analytical focus on this region is warranted. The distinct climatic and geographical conditions of the southern region compared to areas above 20ºS latitude suggest that models tailored to these specific areas could yield better performance. Such an approach would address the current limitation posed by the scarcity of classification examples in regions where hailstorms are less frequent.

Incorporation of temporal dynamics in model predictions: The incorporation of temporal behavior of storm-related variables into the classification models is recommended for future research. A model that accounts for the temporal evolution of these variables may provide a more accurate and dynamic prediction of hailstorm occurrences than models based solely on static, single-time-point data. This approach could unveil patterns and trends not observable in static models, offering deeper insights into storm development processes.

New strategies for addressing class imbalance: Experiment 2 highlighted the challenges of class imbalance in the dataset, where the natural occurrence rates of hailstorms (negative cases) far exceed positive cases. A new strategy, such as oversampling the negative examples, could better reflect the natural distribution of hailstorm occurrences. This adjustment aims to enhance the model's ability to generalize from the training data to real-world conditions, improving the reliability

of predictions.

REFERENCES

ABU-MOSTAFA, Y. S. **Learning from data: a short course**. [S.l.: s.n.], 2012. 46, 48

ADLER, R. F.; MARKUS, M. J.; FENN, D. D.; SZEJWACH, G.; SHENK, W. E. Thunderstorm top structure observed by aircraft overflights with an infrared radiometer. **Journal of Applied Meteorology and Climatology**, v. 22, n. 4, p. 579–593, 1983. 34

AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: a next-generation hyperparameter optimization framework. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 25., 2019, Anchorage-AK, USA. **Proceedings...** [S.l.]: ACM, 2019. 67

ALLEN, J. T.; GIAMMANCO, I. M.; KUMJIAN, M. R.; PUNGE, H. J.; ZHANG, Q.; GROENEMEIJER, P.; KUNZ, M.; ORTEGA, K. Understanding hail in the earth system. **Reviews of Geophysics**, v. 58, n. 1, p. e2019RG000665, 2020. 1, 5, 7, 8, 12, 34, 113, 116

AMBURN, S. A.; WOLF, P. L. Vil density as a hail indicator. **Weather and Forecasting**, v. 12, n. 3, p. 473–478, 1997. 30

AMERICAN METEOROLOGICAL SOCIETY. **Hail**. [S.l.], 2024. Available from: <<http://glossary.ametsoc.org/wiki/hail>>. 1

ANDREAE, M. O.; ROSENFELD, D.; ARTAXO, P.; COSTA, A.; FRANK, G.; LONGO, K.; SILVA-DIAS, M. d. Smoking rain clouds over the Amazon. **Science**, v. 303, n. 5662, p. 1337–1342, 2004. 11, 39

ANDRONACHE, C. **Remote sensing of clouds and precipitation**. [S.l.]: Springer, 2018. 19, 28

ANGUITA, D.; GHIO, A.; ONETO, L.; RIDELLA, S. In-sample and out-of-sample model selection and error estimation for support vector machines. **IEEE Transactions on Neural Networks and Learning Systems**, v. 23, n. 9, p. 1390–1406, 2012. 50

APKE, J. M.; MECIKALSKI, J. R.; BEDKA, K.; JR, E. W. M.; HOMEYER, C. R.; JEWETT, C. P. Relationships between deep convection updraft characteristics and satellite-based super rapid scan mesoscale atmospheric motion vector–derived flow. **Monthly Weather Review**, v. 146, n. 10, p. 3461–3480, 2018. 35

APKE, J. M.; MECIKALSKI, J. R.; JEWETT, C. P. Analysis of mesoscale atmospheric flows above mature deep convection using super rapid scan geostationary satellite data. **Journal of Applied Meteorology and Climatology**, v. 55, n. 9, p. 1859–1887, 2016. 35

AYALA, A. C. B.; GERTH, J. J.; SCHMIT, T. J.; LINDSTROM, S.; NELSON, J. P. Parallax shift in goes abi data. **Journal of Operational Meteorology**, 2023. Available from: <<https://api.semanticscholar.org/CorpusID:257012578>>. 16

BANG, S. D.; CECIL, D. J. Constructing a multifrequency passive microwave hail retrieval and climatology in the gpm domain. **Journal of Applied Meteorology and Climatology**, v. 58, n. 9, p. 1889–1904, 2019. 113

BATTAGLIA, A.; KOLLIAS, P.; DHILLON, R.; ROY, R.; TANELLI, S.; LAMER, K.; GRECU, M.; LEBSOCK, M.; WATTERS, D.; MROZ, K.; HAYMSFIELD, G.; LI, L.; FURUKAWA, K. Spaceborne cloud and precipitation radars: status, challenges, and ways forward. **Reviews of Geophysics**, v. 58, n. 3, p. e2019RG000686, 2020. 25

BECHINI, R.; CHANDRASEKAR, V. A semisupervised robust hydrometeor classification method for dual-polarization radar applications. **Journal of Atmospheric and Oceanic Technology**, v. 32, n. 1, p. 22–47, 2015. 26

BEDKA, K.; BRUNNER, J.; DWORAK, R.; FELTZ, W.; OTKIN, J.; GREENWALD, T. Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients. **Journal of Applied Meteorology and Climatology**, v. 49, n. 2, p. 181–202, 2010. 34

BEDKA, K. M.; MECIKALSKI, J. R. Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows. **Journal of Applied Meteorology and Climatology**, v. 44, n. 11, p. 1761–1772, 2005. 35

BEDKA, K. M.; VELDEN, C. S.; PETERSEN, R. A.; FELTZ, W. F.; MECIKALSKI, J. R. Comparisons of satellite-derived atmospheric motion vectors, rawinsondes, and noaa wind profiler observations. **Journal of Applied Meteorology and Climatology**, v. 48, n. 8, p. 1542–1561, 2009. 35

BEDKA, K. M.; WANG, C.; ROGERS, R.; CAREY, L. D.; FELTZ, W.; KANAK, J. Examining deep convective cloud evolution using total lightning, wsr-88d, and goes-14 super rapid scan datasets. **Weather and Forecasting**, v. 30, n. 3, p. 571–590, 2015. 13

BENDIX, J.; FRIES, A.; ZÁRATE, J.; TRACHTE, K.; ROLLENBECK, R.; PUCHA-COFREP, F.; PALADINES, R.; PALACIOS, I.; ORELLANA, J.; OÑATE-VALDIVIESO, F.; NARANJO, C.; MENDOZA, L.; MEJIA, D.; GUALLPA, M.; GORDILLO, F.; GONZALEZ-JARAMILLO, V.; DOBBERMAN, M.; CÉLLERI, R.; CARRILLO, C.; ARAQUE, A.; ARCHILLES, S. Radarnet-sur first weather radar network in tropical high mountains. **Bulletin of the American Meteorological Society**, v. 98, n. 6, p. 1235–1254, 2017. 25

BENZ, A.; CHAPEL, J.; JR., A. D.; BIRCKENSTAEDT, B.; TILLIER, C.; III, W. N.; COLON, J.; DENCE, R.; JR, D. H.; CAMPBELL, P. **NASA GOES-R Series Data Book**. [S.l.]: National Aeronautics and Space Administration (NASA), 2019. 41

BERENDES, T. A.; MECIKALSKI, J. R.; JR, W. M. M.; BEDKA, K. M.; NAIR, U. Convective cloud identification and classification in daytime satellite imagery using standard deviation limited adaptive clustering. **Journal of Geophysical Research: Atmospheres**, v. 113, n. D20, 2008. 32

BERTHET, C.; WESOLEK, E.; DESSENS, J.; SANCHEZ, J. Extreme hail day climatology in southwestern france. **Atmospheric Research**, v. 123, p. 139–150, 2013. 97

BISHOP, C. M. Linear models for classification. In: _____. [S.l.]: Springer, 2006. chapter 4, p. 179–234. ISBN 978-0387-31073-2. 47, 52, 53, 80, 90

BLAIR, S.; CAVANAUGH, D.; LAFLIN, J.; LEIGHTON, J.; SANDERS, K.; ORTEGA, K. High-resolution hail observations: implications for nws warning operations and climatological data. In: CONFERENCE ON WEATHER ANALYSIS AND FORECASTING, 26., 2014, Atlanta-GA, USA. **Proceedings...** Atlanta: American Meteorological Society, 2014. 40

BOHREN, C. F. On the absurdity and inadvisability of calling particles mie scatterers. **Optics & Photonics News**, v. 3, n. 2, p. 18–19, 1992. 20

BOHREN, C. F.; HUFFMAN, D. R. **Absorption and scattering of light by small particles**. [S.l.]: John Wiley & Sons, 1983. 19

BORG, I.; GROENEN, P. J. **Modern multidimensional scaling: theory and applications**. [S.l.]: Springer Science & Business Media, 2005. 92

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001. 80

BRINGI, V. N.; CHANDRASEKAR, V. **Polarimetric Doppler weather radar: principles and applications**. [S.l.]: Cambridge University Press, 2001. 25

BROWN, T. M.; POGORZELSKI, W. H.; GIAMMANCO, I. M. Evaluating hail damage using property insurance claims data. **Weather, Climate, and Society**, v. 7, n. 3, p. 197–210, 2015. 2

BROWNING, K. The structure and mechanisms of hailstorms. In: FOOTE, G. B.; KNIGHT, C. A. (Ed.). **Hail: a review of hail science and hail suppression**. [S.l.]: Springer, 1977. p. 1–47. 1

BROWNING, K. A.; FOOTE, G. Airflow and hail growth in supercell storms and some implications for hail suppression. **Quarterly Journal of the Royal Meteorological Society**, v. 102, n. 433, p. 499–533, 1976. 4

BRUICK, Z. S.; RASMUSSEN, K. L.; CECIL, D. J. Subtropical south american hailstorm characteristics and environments. **Monthly Weather Review**, v. 147, n. 12, p. 4289–4304, 2019. 9

BüHLMANN, P.; YU, B. Boosting with the l2 loss. **Journal of the American Statistical Association**, v. 98, n. 462, p. 324–339, 2003. Available from: <<https://doi.org/10.1198/016214503000125>>. 100

CAPOZZI, V.; PICCIOTTI, E.; MAZZARELLA, V.; MARZANO, F. S.; BUDILLON, G. Fuzzy-logic detection and probability of hail exploiting short-range x-band weather radar. **Atmospheric Research**, v. 201, p. 17–33, 2018. 30

CAREY, L. D.; RUTLEDGE, S. A.; PETERSEN, W. A. The relationship between severe storm reports and cloud-to-ground lightning polarity in the contiguous united states from 1989 to 1998. **Monthly Weather Review**, v. 131, n. 7, p. 1211–1228, 2003. 13

CARVALHO, L. M.; JONES, C. A satellite method to identify structural properties of mesoscale convective systems based on the maximum spatial correlation tracking technique (mascotte). **Journal of Applied Meteorology and Climatology**, v. 40, n. 10, p. 1683–1701, 2001. 45

CECIL, D. J. Passive microwave brightness temperatures as proxies for hailstorms. **Journal of Applied Meteorology and Climatology**, v. 48, n. 6, p. 1281–1286, 2009. 22, 23

CECIL, D. J.; BLANKENSHIP, C. B. Toward a global climatology of severe hailstorms as estimated by satellite passive microwave imagers. **Journal of Climate**, v. 25, n. 2, p. 687–703, 2012. 22, 113

CECIL, D. J.; GOODMAN, S. J.; BOCCIPPIO, D. J.; ZIPSER, E. J.; NESBITT, S. W. Three years of TRMM precipitation features. Part I: radar, radiometric, and lightning characteristics. **Monthly Weather Review**, v. 133, n. 3, p. 543–566, 2005. 22

CHAN, T. F.; GOLUB, G. H.; LEVEQUE, R. J. **Updating formulae and a pairwise algorithm for computing sample variances**. [S.l.], 1979. 80

CHANDRASEKAR, V.; KEELER, R. J. Antenna pattern analysis and measurements for multiparameter radars. **Journal of Atmospheric and Oceanic Technology**, v. 10, n. 5, p. 674–683, 1993. 29

CHEN, S.; HOUZE, R. A.; MAPES, B. E. Multiscale variability of deep convection in relation to large-scale circulation in toga coare. **Oceanographic Literature Review**, v. 43, n. 12, p. 1200, 1996. 45

CHINA METEOROLOGICAL ADMINISTRATION. **Handbook of surface meteorological observation (in Chinese)**. Beijing. China, 2003. 151p. 1

CHRONIS, T.; CAREY, L. D.; SCHULTZ, C. J.; SCHULTZ, E. V.; CALHOUN, K. M.; GOODMAN, S. J. Exploring lightning jump characteristics. **Weather and Forecasting**, v. 30, n. 1, p. 23–37, 2015. 13

CINTINEO, J. L.; PAVOLONIS, M. J.; SIEGLAFF, J. M.; HEIDINGER, A. K. Evolution of severe and nonsevere convection inferred from goes-derived cloud properties. **Journal of Applied Meteorology and Climatology**, v. 52, n. 9, p. 2009–2023, 2013. 13

CINTINEO, J. L.; PAVOLONIS, M. J.; SIEGLAFF, J. M.; WIMMERS, A.; BRUNNER, J.; BELLON, W. A deep-learning model for automated detection of intense midlatitude convection using geostationary satellite images. **Weather and Forecasting**, v. 35, n. 6, p. 2567–2588, 2020. 13, 115

CINTINEO, J. L.; PAVOLONIS, M. J.; SIEGLAFF, J. M.; CRONCE, L.; BRUNNER, J. Noaa probsevere v2. 0—probhail, probwind, and probtor. **Weather and Forecasting**, v. 35, n. 4, p. 1523–1543, 2020. 36

CLIMATEMPO, R. **ZCAS se desfaz, mas Brasil ainda tem muita chuva nesta terça**. 2023. Available from: <<https://www.climatempo.com.br/noticia/2023/01/10/zcas-se-desfaz-mas-brasil-ainda-tem-muita-chuva-nesta-terca-9020>>. Access in: 2024 May 25. 112

COOPERATIVE INSTITUTE FOR METEOROLOGICAL SATELLITE STUDIES (CIMSS). **GOES activities**: calibration. 2023. Available from: <<https://cimss.ssec.wisc.edu/goes/calibration/>>. Access in: 2024 Apr. 11. 87

COSTA, S. M.; NEGRI, R. G.; FERREIRA, N. J.; SCHMIT, T. J.; ARAI, N.; FLAUBER, W.; CEBALLOS, J.; VILA, D.; RODRIGUES, J.; MACHADO, L. A.; PEREIRA, S.; BOTTINO, M. J.; SISMANOGLU, R. A.; LANGDEN, P. A successful practical experience with dedicated geostationary operational environmental satellites goes-10 and-12 supporting Brazil. **Bulletin of the American Meteorological Society**, v. 99, n. 1, p. 33–47, 2018. 41, 46

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, v. 13, n. 1, p. 21–27, 1967. 80

CZEKALA, H.; SIMMER, C. Microwave radiative transfer with nonspherical precipitating hydrometeors. **Journal of Quantitative Spectroscopy and Radiative Transfer**, v. 60, n. 3, p. 365–374, 1998. 21

DA, C. Preliminary assessment of the advanced himawari imager (ahi) measurement onboard himawari-8 geostationary satellite. **Remote Sensing Letters**, v. 6, n. 8, p. 637–646, 2015. 42

DAVIS, G. K. History of the noaa satellite program. **Journal of Applied Remote Sensing**, v. 1, n. 1, p. 012504, 2007. 41

DAWSON, L. C. **On the predictability of multiday episodes of severe convective storms during the mesoscale predictability experiment**. Thesis (PhD) — Purdue University, 2017. 97

DENNIS, A.; SCHOCK, C. A.; KOSCIELSKI, A. Characteristics of hailstorms of western south dakota. **Journal of Applied Meteorology and Climatology**, v. 9, n. 1, p. 127–135, 1970. 3

DENNIS, E. J.; KUMJIAN, M. R. The impact of vertical wind shear on hail growth in simulated supercells. **Journal of the Atmospheric Sciences**, v. 74, n. 3, p. 641–663, 2017. 3, 6

DESSENS, J.; BERTHET, C.; SANCHEZ, J. A point hailfall classification based on hailpad measurements: the ANELFA scale. **Atmospheric Research**, v. 83, n. 2-4, p. 132–139, 2007. 117

_____. Change in hailstone size distributions with an increase in the melting level height. **Atmospheric Research**, v. 158, p. 245–253, 2015. 38, 117

DESSENS, J.; SÁNCHEZ, J.; BERTHET, C.; HERMIDA, L.; MERINO, A. Hail prevention by ground-based silver iodide generators: results of historical and modern field projects. **Atmospheric Research**, v. 170, p. 98–111, 2016. 3

DEVELOPERS, S. learn. **Choosing the right estimator**. 2024. Available from: scikit-learn documentation. Available from: <<https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html>>. Access in: 17 Mar. 2024. 54

DOVIAK, R. J.; ZRNIC, D. **Doppler radar and weather observations**. [S.l.]: Academic Press, 1993. 29

DURKEE, J. D.; MOTE, T. L. A climatology of warm-season mesoscale convective complexes in subtropical south america. **International Journal of Climatology: A Journal of the Royal Meteorological Society**, v. 30, n. 3, p. 418–431, 2010. 8

DWORAK, R.; BEDKA, K.; BRUNNER, J.; FELTZ, W. Comparison between goes-12 overshooting-top detections, wsr-88d radar reflectivity, and severe storm reports. **Weather and Forecasting**, v. 27, n. 3, p. 684–699, 2012. 34

EDWARD, C. Rasmussen and christopher ki williams. gaussian processes for machine learning. **MIT Press**, v. 211, p. 212, 2006. 92

FARNELL, C.; RIGO, T.; PINEDA, N. Lightning jump as a nowcast predictor: application to severe weather events in catalonia. **Atmospheric Research**, v. 183, p. 130–141, 2017. 13

FEDERER, B.; THALMANN, B.; JOUZEL, J. Stable isotopes in hailstones. Part II: embryo and hailstone growth in different storms. **Journal of Atmospheric Sciences**, v. 39, n. 6, p. 1336–1355, 1982. 2

FERRARO, R.; BEAUCHAMP, J.; CECIL, D.; HEYMSFIELD, G. A prototype hail detection algorithm and hail climatology developed with the advanced microwave sounding unit (amsu). **Atmospheric Research**, v. 163, p. 24–35, 2015. 24

FERRARO, R. R.; CECIL, D.; LAVIOLA, S. Hailfall detection. **Satellite Precipitation Measurement**, p. 861–879, 2020. 22, 23, 30, 31

FINDEISEN, W. Die kolloidmeteorologischen vorgänge bei der niederschlagsbildung (colloidal meteorological processes in the formation of precipitation). **Meteorologische Zeitschrift**, p. 121–133, 1938. 5

FOOTE, G. B. A study of hail growth utilizing observed storm conditions. **Journal of Applied Meteorology and Climatology**, v. 23, n. 1, p. 84–101, 1984. 1

FOOTE, G. B.; KNIGHT, C. **Hail: a review of hail science and hail suppression**. [S.l.]: Springer, 1977. 6

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of Statistics**, p. 1189–1232, 2001. 80

FRISBY, E. M.; SANSOM, H. Hail incidence in the tropics. **Journal of Applied Meteorology and Climatology**, v. 6, n. 2, p. 339–354, 1967. 8, 11

FURIERI, A. **SpatiaLite**. 2023. Version 5.1.0. Available from: <<https://www.gaia-gis.it/fossil/libspatialite/index>>. Access in: 22 Feb. 2024. 46

GALVIN, J. Meteorology and the second world war. **Weather**, v. 75, n. 10, p. 325–328, 2020. 2

GAO, B.-C.; HAN, W.; TSAY, S. C.; LARSEN, N. F. Cloud detection over the arctic region using airborne imaging spectrometer data during the daytime. **Journal of Applied Meteorology**, v. 37, n. 11, p. 1421–1429, 1998. 33

GARCÍA-ORTEGA, E.; MERINO, A.; LÓPEZ, L.; SÁNCHEZ, J. Role of mesoscale factors at the onset of deep convection on hailstorm days and their relation to the synoptic patterns. **Atmospheric Research**, v. 114, p. 91–106, 2012. 117

GARREAUD, R.; WALLACE, J. M. The diurnal march of convective cloudiness over the americas. **Monthly Weather Review**, v. 125, n. 12, p. 3157–3171, 1997. 8, 45

GEOPANDAS. **GeoPandas documentation**. 2024. Available from: <<http://geopandas.org>>. 94

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems**. [S.l.]: " O'Reilly Media", 2019. 47, 90

GISBRECHT, A.; SCHULZ, A.; HAMMER, B. Parametric nonlinear dimensionality reduction using kernel t-sne. **Neurocomputing**, v. 147, p. 71–82, 2015. 92

GLICKMAN, T. S.; ZENK, W. **Glossary of meteorology**. [S.l.]: AMS (American Meteorological Society), 2000. 34

GOLUB, G. H.; LOAN, C. F. V. **Matrix computations**. [S.l.]: JHU press, 2013. 90

GONÇALVES, W.; MACHADO, L.; KIRSTETTER, P.-E. Influence of biomass aerosol on precipitation over the central amazon: an observational study. **Atmospheric Chemistry and Physics**, v. 15, n. 12, p. 6789–6800, 2015. 12

GOODMAN, S. J.; BUECHLER, D. E.; WRIGHT, P. D.; RUST, W. D. Lightning and precipitation history of a microburst-producing storm. **Geophysical Research Letters**, v. 15, n. 11, p. 1185–1188, 1988. 13

GORELICK, N. Learning fuzzy systems with similarity relations. In: EGU GENERAL ASSEMBLY, 15., 2013, Vienna, Austria. **Proceedings...** Vienna: European Geophysical Union, 2013. 78

GROUP, G.-R. A. W.; PROGRAM, G.-R. S. **NOAA GOES-R Series Advanced Baseline Imager (ABI) level 2 Cloud and Moisture Imagery Products (CMIP) full disk**. NOAA National Centers for Environmental Information, 2017. Doi:10.7289/V5736P36. Available from: <<https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C01502>>. Access in: 22 Feb. 2024. 42

GUNTURI, P.; TIPPETT, M. **Managing severe thunderstorm risk: impact of ENSO on US tornado and hail frequencies**. [S.l.], 2017. 1

GUO, X.; HUANG, M. Hail formation and growth in a 3d cloud model with hail-bin microphysics. **Atmospheric Research**, v. 63, n. 1-2, p. 59–99, 2002. 4

HAO, F.; BLAIR, R. H. A comparative study: classification vs. user-based collaborative filtering for clinical prediction. **BMC Medical Research Methodology**, v. 16, p. 1–14, 2016. 60

HAZENBERG, P.; YU, N.; BOUDEVILLAIN, B.; DELRIEU, G.; UIJLENHOET, R. Scaling of raindrop size distributions and classification of radar reflectivity–rain rate relations in intense mediterranean precipitation. **Journal of Hydrology**, v. 402, n. 3-4, p. 179–192, 2011. 25

HEINSELMAN, P. L.; RYZHKOV, A. V. Validation of polarimetric hail detection. **Weather and Forecasting**, v. 21, n. 5, p. 839–850, 2006. 30

HEISS, W. H.; MCGREW, D. L.; SIRMANS, D. Nexrad: next generation weather radar (wsr-88d). **Microwave Journal**, v. 33, n. 1, p. 79–89, 1990. 25

HEYMSFIELD, A. J. A comparative study of the rates of development of potential graupel and hail embryos in high plains storms. **Journal of Atmospheric Sciences**, v. 39, n. 12, p. 2867–2897, 1982. 2, 117

HEYMSFIELD, A. J.; JAMESON, A. R.; FRANK, H. W. Hail growth mechanisms in a colorado storm: Part II: hail formation processes. **Journal of Atmospheric Sciences**, v. 37, n. 8, p. 1779–1807, 1980. 4

HEYMSFIELD, A. J.; MUSIL, D. J. Case study of a hailstorm in colorado. Part II: particle growth processes at mid-levels deduced from in-situ measurements. **Journal of the Atmospheric Sciences**, v. 39, n. 12, p. 2847–2866, 1982. 3

HEYMSFIELD, G. M.; SZEJWACH, G.; SCHOTZ, S.; JR, R. H. B. Upper-level structure of oklahoma tornadic storms on 2 may 1979. II: proposed explanation of "v" pattern and internal warm region in infrared observations. **Journal of Atmospheric Sciences**, v. 40, n. 7, p. 1756–1767, 1983. 34

HOEFFDING, W. Probability inequalities for sums of bounded random variables. In: FISHER, N. I.; SEN, P. K. (Ed.). **The collected works of Wassily Hoeffding**. New York: Springer, 1994. p. 409–426. 50

HOMEYER, C. R.; MCAULIFFE, J. D.; BEDKA, K. M. On the development of above-anvil cirrus plumes in extratropical convection. **Journal of the Atmospheric Sciences**, v. 74, n. 5, p. 1617–1633, 2017. 34

HONG, G.; HEYGSTER, G.; MIAO, J.; KUNZI, K. Sensitivity of microwave brightness temperatures to hydrometeors in a tropical deep convective cloud system at 89–190 ghz. **Radio Science**, v. 40, n. 04, p. 1–13, 2005. 22

HOUZE, R. A. **Cloud dynamics**. [S.l.]: Academic Press, 2014. 1, 20, 21, 22, 25, 26, 27, 31

HUBERT, L. F.; JR, L. F. W. Wind estimation from geostationary-satellite pictures. **Monthly Weather Review**, v. 99, n. 9, p. 665–672, 1971. 35

HUUSKONEN, A.; SALTIKOFF, E.; HOLLEMAN, I. The operational weather radar network in europe. **Bulletin of the American Meteorological Society**, v. 95, n. 6, p. 897–907, 2014. 25

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R.; TAYLOR, J. **An introduction to statistical learning: with applications in python**. [S.l.]: Springer Nature, 2023. 57, 58, 70, 75

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. Tree-based methods. In: _____. **Introduction to statistical learning using R book club**. Bookdown, 2024. interactive book 8. Available from: <<https://r4ds.github.io/bookclub-islr/recursive-binary-splitting.html>>. 60

KEY, J. R.; INTRIERI, J. M. Cloud particle phase determination with the avhrr. **Journal of Applied Meteorology and Climatology**, v. 39, n. 10, p. 1797–1804, 2000. 84

KIDDER, S. Q.; HAAR, T. H. V. **Satellite meteorology: an introduction**. [S.l.]: Gulf Professional Publishing, 1995. 99, 114

KNIGHT, C.; KNIGHT, N.; BROOKS, H.; SKRIPNIKOVA, K. Hail and hailstorms. In: NORTH, G. R.; PYLE, J.; ZHANG, F. (Ed.). **Encyclopedia of Atmospheric Sciences**. Radarweg: Elsevier, 2015. p. 334–338. 1

KNIGHT, C. A.; EHHALT, D. H.; ROPER, N.; KNIGHT, N. C. Radial and tangential variation of deuterium in hailstones. **Journal of Atmospheric Sciences**, v. 32, n. 10, p. 1990–2000, 1975. 5

KNIGHT, C. A.; KNIGHT, N. C. Hailstorms. In: DOSWELL, C. A. (Ed.). **Severe convective storms**. [S.l.]: Springer, 2001. p. 223–254. 117

KNIGHT, N. C. Hailstone shape factor and its relation to radar interpretation of hail. **Journal of Climate and Applied Meteorology**, v. 25, n. 12, p. 1956–1958, 1986. 1

KOOP, T. Homogeneous ice nucleation in water and aqueous solutions. **Zeitschrift für physikalische Chemie**, v. 218, n. 11, p. 1231–1258, 2004. 5

KOREN, I.; KAUFMAN, Y. J.; REMER, L. A.; MARTINS, J. V. Measurement of the effect of Amazon smoke on inhibition of cloud formation. **Science**, v. 303, n. 5662, p. 1342–1345, 2004. 12

KOWALEWSKI, S.; PETERS, G. Analysis of z–r relations based on ldr signatures within the melting layer. **Journal of Atmospheric and Oceanic Technology**, v. 27, n. 9, p. 1555–1561, 2010. 29

KUMJIAN, M. R. Principles and applications of dual-polarization weather radar. Part I: description of the polarimetric radar variables. **Journal of Operational Meteorology**, v. 1, 2013. 29

KUMJIAN, M. R.; RICHARDSON, Y. P.; MEYER, T.; KOSIBA, K. A.; WURMAN, J. Resonance scattering effects in wet hail observed with a dual-x-band-frequency, dual-polarization doppler on wheels radar. **Journal of Applied Meteorology and Climatology**, v. 57, n. 12, p. 2713–2731, 2018. 19

KUMJIAN, M. R.; RYZHKOV, A. V.; REEVES, H. D.; SCHUUR, T. J. A dual-polarization radar signature of hydrometeor refreezing in winter storms. **Journal of Applied Meteorology and Climatology**, v. 52, n. 11, p. 2549–2566, 2013. 26

KUMMEROW, C.; BARNES, W.; KOZU, T.; SHIUE, J.; SIMPSON, J. The tropical rainfall measuring mission (trmm) sensor package. **Journal of Atmospheric and Oceanic Technology**, v. 15, n. 3, p. 809–817, 1998. 22

KUNZ, M.; PUSKEILER, M. High-resolution assessment of the hail hazard over complex terrain from radar and insurance data. **Meteorologische Zeitschrift**, v. 19, n. 5, p. 427, 2010. 2

KUNZ, M.; SANDER, J.; KOTTMEIER, C. Recent trends of thunderstorm and hailstorm frequency and their relation to atmospheric characteristics in southwest germany. **International Journal of Climatology: A Journal of the Royal Meteorological Society**, v. 29, n. 15, p. 2283–2297, 2009. 38

KURINO, T. A satellite infrared technique for estimating “deep/shallow” precipitation. **Advances in Space Research**, v. 19, n. 3, p. 511–514, 1997. 34

LAING, A. G.; FRITSCH, J. M. The global population of mesoscale convective complexes. **Quarterly Journal of the Royal Meteorological Society**, v. 123, n. 538, p. 389–405, 1997. 8

LAURENT, H.; MACHADO, L. A.; MORALES, C. A.; DURIEUX, L. Characteristics of the amazonian mesoscale convective systems observed from satellite and radar during the wetamc/lba experiment. **Journal of Geophysical Research: Atmospheres**, v. 107, n. D20, p. LBA–21, 2002. 45

LAVIOLA, S.; LEVIZZANI, V. Observing precipitation by means of water vapor absor tion lines: a first check of the retrieval capabilities of p the 183-wsl rain retrieval method. **Rivista Italiana di Telerilevamento**, v. 41, n. 3, p. 39–49, 2009. 23

LAVIOLA, S.; LEVIZZANI, V.; FERRARO, R. R.; BEAUCHAMP, J. Hailstorm detection by satellite microwave radiometers. **Remote Sensing**, v. 12, n. 4, p. 621, 2020. 24

LAVIOLA, S.; MONTE, G.; LEVIZZANI, V.; FERRARO, R. R.; BEAUCHAMP, J. A new method for hail detection from the gpm constellation: A prospect for a global hailstorm climatology. **Remote Sensing**, v. 12, n. 21, p. 3553, 2020. 24

LEBO, Z. A numerical investigation of the potential effects of aerosol-induced warming and updraft width and slope on updraft intensity in deep convective clouds. **Journal of the Atmospheric Sciences**, v. 75, n. 2, p. 535–554, 2018. 12

LEMAIRE, V.; FÉRAUD, R.; VOISINE, N. Contact personalization using a score understanding method. In: IEEE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE), 18., 2008, Hong Kong, China. **Proceedings...** Hong Kong: IEEE/IJCNN, 2008. p. 649–654. 73

LEVIZZANI, V.; SETVÁK, M. Multispectral, high-resolution satellite observations of plumes on top of convective storms. **Journal of Atmospheric Sciences**, v. 53, n. 3, p. 361–369, 1996. 35

LIOU, K.-N. **An introduction to atmospheric radiation**. [S.l.]: Elsevier, 2002. 22

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). **Advances in neural information processing systems 30**. Curran Associates, 2017. p. 4765–4774. Available from: <<http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>>. 72

MACGORMAN, D. R.; BURGESS, D. W. Positive cloud-to-ground lightning in tornadic storms and hailstorms. **Monthly Weather Review**, v. 122, n. 8, p. 1671–1697, 1994. 13

MACGORMAN, D. R.; MORGENSTERN, C. D. Some characteristics of cloud-to-ground lightning in mesoscale convective systems. **Journal of Geophysical Research: Atmospheres**, v. 103, n. D12, p. 14011–14023, 1998. 12

MACHADO, L.; LAURENT, H.; DESSAY, N.; MIRANDA, I. Seasonal and diurnal variability of convection over the Amazonia: A comparison of different vegetation types and large scale forcing. **Theoretical and Applied Climatology**, v. 78, p. 61–77, 2004. 12

MACHADO, L.; ROSSOW, W.; GUEDES, R.; WALKER, A. Life cycle variations of mesoscale convective systems over the americas. **Monthly Weather Review**, v. 126, n. 6, p. 1630–1654, 1998. 11, 44, 45

MACHADO, L. A.; LAURENT, H.; LIMA, A. A. Diurnal march of the convection observed during trmm-wetamc/lba. **Journal of Geophysical Research: Atmospheres**, v. 107, n. D20, p. LBA–31, 2002. 78

MACHADO, L. A. T.; LAURENT, H. The convective system area expansion over amazonia and its relationships with convective system life duration and high-level wind divergence. **Monthly Weather Review**, v. 132, n. 3, p. 714–725, 2004. 11, 45

MACHADO, L. T.; DESBOIS, M.; DUVEL, J.-P. Structural characteristics of deep convective systems over tropical africa and the atlantic ocean. **Monthly Weather Review**, v. 120, n. 3, p. 392–406, 1992. 44

MAKRIS, A.; PRIEUR, C. Bayesian multiple-hypothesis tracking of merging and splitting targets. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 12, p. 7684–7694, 2014. 43

MANZATO, A. Hail in northeast Italy: climatology and bivariate analysis with the sounding-derived indices. **Journal of Applied Meteorology and Climatology**, v. 51, n. 3, p. 449–467, 2012. 38

MARCELINO, I.; MENDONÇA, M.; RUDORFF, F. Ocorrências de granizo no estado de santa catarina. In: SIMPOSIO BRASILEIRO DE DESASTRES NATURAIS, I., 2004, Florianópolis, Brazil. **Proceedings...** Florianópolis: GEDN/UFSC, 2004. p. 795–805. 11

MARENGO, J.; LIEBMANN, B.; GRIMM, A.; MISRA, V.; DIAS, P. d. S.; CAVALCANTI, I.; CARVALHO, L.; BERBERY, E.; AMBRIZZI, T.; VERA, C. S.; SAULO, A. C.; NOGUES-PAEGLE, J.; ZIPSER, E.; SETH, A.; ALVES, L. M. Recent developments on the south american monsoon system. **International Journal of Climatology**, v. 32, n. 1, p. 1–21, 2012. 8, 9

MARTINS, J. A.; BRAND, V. S.; CAPUCIM, M. N.; FELIX, R. R.; MARTINS, L. D.; FREITAS, E. D.; GONÇALVES, F. L.; HALLAK, R.; DIAS, M. A. S.; CECIL, D. J. Climatology of destructive hailstorms in Brazil. **Atmospheric Research**, v. 184, p. 126–138, 2017. 1, 11, 12

MATROSOV, S. Y.; REINKING, R. F.; KROPFLI, R. A.; BARTRAM, B. W. Estimation of ice hydrometeor types and shapes from radar polarization measurements. **Journal of Atmospheric and Oceanic Technology**, v. 13, n. 1, p. 85–96, 1996. 27

MEISCHNER, P. **Weather radar: principles and advanced applications**. [S.l.]: Springer Science & Business Media, 2005. 25

MELCÓN, P.; MERINO, A.; SÁNCHEZ, J. L.; LÓPEZ, L.; HERMIDA, L. Satellite remote sensing of hailstorms in france. **Atmospheric Research**, v. 182, p. 221–231, 2016. 115

MELNYK, A.; TERESHCHENKO, M.; SHARAPANIUK, O. Apple trees productivity under the hail-protective net in different soil management systems. **Horizons**, v. 5, n. 90, p. 41–49, 2020. 2

MENZEL, W. P. Cloud tracking with satellite imagery: from the pioneering work of ted fujita to the present. **Bulletin of the American Meteorological Society**, v. 82, n. 1, p. 33–48, 2001. 35

MENZEL, W. P.; PURDOM, J. F. Introducing GOES-I: the first of a new generation of geostationary operational environmental satellites. **Bulletin of the American Meteorological Society**, v. 75, n. 5, p. 757–782, 1994. 41

MERINO, A.; LÓPEZ, L.; SÁNCHEZ, J.; GARCÍA-ORTEGA, E.; CATTANI, E.; LEVIZZANI, V. Daytime identification of summer hailstorm cells from msg data. **Natural Hazards and Earth System Sciences**, v. 14, n. 4, p. 1017–1033, 2014. 36, 115, 117

MEZHER, R. N.; DOYLE, M.; BARROS, V. Climatology of hail in argentina. **Atmospheric Research**, v. 114, p. 70–82, 2012. 9, 10

MIE, G. Beiträge zur optik trüber medien, speziell kolloidaler metallösungen. **Annalen der Physik**, v. 330, n. 3, p. 377–445, 1908. 20

MIKUŠ, P.; MAHOVIĆ, N. S. Satellite-based overshooting top detection methods and an analysis of correlated weather conditions. **Atmospheric Research**, v. 123, p. 268–280, 2013. 34

MILLER, L. J.; TUTTLE, J. D.; FOOTE, G. B. Precipitation production in a large montana hailstorm: airflow and particle growth trajectories. **Journal of the Atmospheric Sciences**, v. 47, n. 13, p. 1619–1646, 1990. 5

MORGAN, G. M. An examination of the wet-bulb zero as a hail forecasting parameter in the Po Valley, Italy. **Journal of Applied Meteorology**, v. 9, n. 3, p. 537–540, 1970. 38

MROZ, K.; BATTAGLIA, A.; LANG, T. J.; CECIL, D. J.; TANELLI, S.; TRIDON, F. Hail-detection algorithm for the gpm core observatory satellite sensors. **Journal of Applied Meteorology and Climatology**, v. 56, n. 7, p. 1939–1957, 2017. 22, 23, 31, 32

MURILLO, E. M. **Severe hail detection using remote sensing observations and an evaluation of the above anvil Cirrus plume for severe weather detection**. 162 p. Master Thesis — College of Atmospheric Geographic Sciences::School of Meteorology, Oklahoma, 2018. 1

MURILLO, E. M.; HOMEYER, C. R. Severe hail fall and hailstorm detection using remote sensing observations. **Journal of Applied Meteorology and Climatology**, v. 58, n. 5, p. 947–970, 2019. 1

NAGARE, B.; MARCOLLI, C.; WELTI, A.; STETZER, O.; LOHMANN, U. Comparing contact and immersion freezing from continuous flow diffusion chambers. **Atmospheric Chemistry and Physics**, v. 16, n. 14, p. 8899–8914, 2016. 3

NAKAJIMA, T.; KING, M. D. Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: theory. **Journal of Atmospheric Sciences**, v. 47, n. 15, p. 1878–1893, 1990. 32

NEGRI, A. J. Cloud-top structure of tornadic storms on 10 april 1979 from rapid scan and stereo satellite observations. **Bulletin of the American Meteorological Society**, v. 63, n. 10, p. 1151–1159, 1982. 34

NEGRI, R.; MACHADO, L.; BORDE, R. Inner convective system cloud-top wind estimation using multichannel infrared satellite images. **International Journal of Remote Sensing**, v. 35, n. 2, p. 651–670, 2014. 35

NELSON, S. P. The influence of storm flow structure on hail growth. **Journal of Atmospheric Sciences**, v. 40, n. 8, p. 1965–1983, 1983. 5, 6

NI, X.; LIU, C.; CECIL, D. J.; ZHANG, Q. On the detection of hail using satellite passive microwave radiometers and precipitation radar. **Journal of Applied Meteorology and Climatology**, v. 56, n. 10, p. 2693–2709, 2017. 23, 113

NISI, L.; MARTIUS, O.; HERING, A.; KUNZ, M.; GERMANN, U. Spatial and temporal distribution of hailstorms in the alpine region: a long-term, high resolution, radar-based analysis. **Quarterly Journal of the Royal Meteorological Society**, v. 142, n. 697, p. 1590–1604, 2016. 2

ORTEGA, K. L.; KRAUSE, J. M.; RYZHKOV, A. V. Polarimetric radar characteristics of melting hail. Part III: validation of the algorithm for hail size discrimination. **Journal of Applied Meteorology and Climatology**, v. 55, n. 4, p. 829–848, 2016. 31

ORVILLE, R. E.; HENDERSON, R. W. Global distribution of midnight lightning: September 1977 to august 1978. **Monthly Weather Review**, v. 114, n. 12, p. 2640–2653, 1986. 13

OUE, M.; KUMJIAN, M. R.; LU, Y.; VERLINDE, J.; AYDIN, K.; CLOTHIAUX, E. E. Linear depolarization ratios of columnar ice crystals in a deep precipitating system over the arctic observed by zenith-pointing ka-band doppler radar. **Journal of Applied Meteorology and Climatology**, v. 54, n. 5, p. 1060–1068, 2015. 29

OZAKI, Y.; TANIGAKI, Y.; WATANABE, S.; ONISHI, M. Multiobjective tree-structured parzen estimator for computationally expensive optimization problems. In: PROCEEDINGS OF THE 2020 GENETIC AND EVOLUTIONARY

COMPUTATION CONFERENCE, 22., 2020, Cancun, Mexico. **Proceedings...** Cancun: ACM, 2020. p. 533–541. 67

PARK, H. S.; RYZHKOV, A.; ZRNIĆ, D.; KIM, K.-E. The hydrometeor classification algorithm for the polarimetric WSR-88D: description and application to an mcs. **Weather and Forecasting**, v. 24, n. 3, p. 730–748, 2009. 30

PARKER, M. **Digital signal processing 101: everything you need to know to get started**. [S.l.]: Newnes, 2017. 25

PEDREGOSA, F. et al. **1.5.1 Classification**. 2024. Webpage. Available from: <<https://scikit-learn.org/stable/modules/sgd.html#sgd>>. Access in: 07 Feb. 2024. 80

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. 80

PÉREZ, R. C.; PULIAFITO, E. Study of hailstorms cells producing big damages in Mendoza (Argentina). In: EUROPEAN CONFERENCE ON RADAR IN METEOROLOGY AND HYDROLOGY, 4., 2006, Barcelona, Spain. **Proceedings...** Barcelona: ERAD/EPFL, 2006. p. 18–22. 9

PRIETO, R.; HERRERA, R.; DOUSSEL, P.; GIMENO, L.; RIBERA, P.; GARCÍA, R.; HERNÁNDEZ, E. Looking for periodicities in the hail intensity in the andes region. **Atmósfera**, v. 14, n. 2, p. 87–93, 2001. 9

PROHASKA, K. **Beobachtungen über Gewitter und Hagelschläge in Steiermark, Kärnten und Oberkrain.** [S.l.], 1902. 1

PRUPPACHER, H.; KLETT, J. Growth of ice particles by accretion and ice particle melting. In: _____. **Microphysics of clouds and precipitation**. Dordrecht: Springer Netherlands, 2010. p. 659–699. ISBN 978-0-306-48100-0. Available from: <<https://doi.org/10.1007/978-0-306-48100-0_16>>. 5

PRUPPACHER, H. R.; KLETT, J. D. **Microphysics of clouds and precipitation: reprinted 1980**. [S.l.]: Springer Science & Business Media, 2012. 2

PUNGE, H. J.; KUNZ, M. Hail observations and hailstorm characteristics in Europe: a review. **Atmospheric Research**, v. 176, p. 159–184, 2016. 1

PUSKEILER, M.; KUNZ, M.; SCHMIDBERGER, M. Hail statistics for germany derived from single-polarization radar data. **Atmospheric Research**, v. 178, p. 459–470, 2016. 2

RASMUSSEN, K. L.; HOUZE, R. A. Orogenic convection in subtropical South America as seen by the TRMM satellite. **Monthly Weather Review**, v. 139, n. 8, p. 2399–2420, 2011. 8

RASMUSSEN, K. L.; ZULUAGA, M. D.; JR, R. A. H. Severe convection and lightning in subtropical South America. **Geophysical Research Letters**, v. 41, n. 20, p. 7359–7366, 2014. 9

RASMUSSEN, R. M.; HEYMSFIELD, A. J. Melting and shedding of graupel and hail. Part II: sensitivity study. **Journal of the Atmospheric Sciences**, v. 44, n. 19, p. 2764–2782, 1987. 38

REYNOLDS, D. W. Observations of damaging hailstorms from geosynchronous satellite digital data. **Monthly Weather Review**, v. 108, n. 3, p. 337–348, 1980. 34

RIBEIRO, B.; SCHILD, G.; SANTOS, L.; NASCIMENTO, E. L.; LOPES, M.; GOEDE, V.; SANTOS, L. d. O. dos; COSTA, I.; FERREIRA, V.; OLIVEIRA, M.; TELLES, J.; FURLAN, L. The brazilian severe storm reports database: methodology and a preliminary climatology. In: CONFERENCE ON SEVERE LOCAL STORMS, 30., 2022, Santa Fe, USA. **Proceedings...** Santa Fe: CSLS/AMS, 2022. p. –. 40

RIBEIRO, B. Z.; MACHADO, L. A.; CH, J. H. H.; BISCARO, T. S.; FREITAS, E. D.; MOZER, K. W.; GOODMAN, S. J. An evaluation of the GOES-16 rapid scan for nowcasting in southeastern Brazil: analysis of a severe hailstorm case. **Weather and Forecasting**, v. 34, n. 6, p. 1829–1848, 2019. 13, 34, 45, 93

ROBERTO, N.; BALDINI, L.; ADIROSI, E.; FACHERIS, L.; CUCCOLI, F.; LUPIDI, A.; GARZELLI, A. A support vector machine hydrometeor classification algorithm for dual-polarization radar. **Atmosphere**, v. 8, n. 8, p. 134, 2017. 26

ROBNIK-ŠIKONJA, M.; KONONENKO, I. Explaining classifications for individual instances. **IEEE Transactions on Knowledge and Data Engineering**, v. 20, n. 5, p. 589–600, 2008. 73

ROSENFELD, D.; CATTANI, E.; MELANI, S.; LEVIZZANI, V. Considerations on daylight operation of 1.6-versus 3.7-μm channel on noaa and metop satellites. **Bulletin of the American Meteorological Society**, v. 85, n. 6, p. 873–882, 2004. 32, 33

ROSENFELD, D.; LOHMANN, U.; RAGA, G. B.; O'DOWD, C. D.; KULMALA, M.; FUZZI, S.; REISSELL, A.; ANDREAE, M. O. Flood or drought: how do aerosols affect precipitation? **Science**, v. 321, n. 5894, p. 1309–1313, 2008. 12

ROSENFELD, D.; WOODLEY, W. L. Deep convective clouds with sustained supercooled liquid water down to-37.5 c. **Nature**, v. 405, n. 6785, p. 440–442, 2000. 5

ROSENFELD, D.; WOODLEY, W. L.; LERNER, A.; KELMAN, G.; LINDSEY, D. T. Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase. **Journal of Geophysical Research: Atmospheres**, v. 113, n. D4, 2008. 33

ROTH, A. E. **The Shapley value: essays in honor of Lloyd S. Shapley**. [S.l.]: Cambridge University Press, 1988. 73

RYZHKOV, A. V.; KUMJIAN, M. R.; GANSON, S. M.; ZHANG, P. Polarimetric radar characteristics of melting hail. Part II: practical implications. **Journal of Applied Meteorology and Climatology**, v. 52, n. 12, p. 2871–2886, 2013. 31

SALTIKOFF, E.; HUUSKONEN, A.; HOHTI, H.; KOISTINEN, J.; JÄRVINEN, H. Quality assurance in the fmi doppler weather radar network. **Boreal Environment Research Publishing Board**, 2010. 25

SALUZZI, M.; NUÑEZ, J. Comportamiento de granizadas sobre diversas áreas cultivadas del país. **Geoacta**, v. 7, p. 77–90, 1975. 9

SAMENOW, J. **Corn stalk encased in ice hailed down from tornado-producing Texas thunderstorm**. 2015. Available from: <<https://www.washingtonpost.com/news/capital-weather-gang/wp/2015/11/17/corn-stalk-encased-in-ice-hailed-down-from-tornado-producing-texas-thunderstorm/>>. 2, 3

SAUNDERS, C. A review of thunderstorm electrification processes. **Journal of Applied Meteorology and Climatology**, v. 32, n. 4, p. 642–655, 1993. 12

SCHMETZ, J.; TJEMKES, S.; GUBE, M.; BERG, L. Van de. Monitoring deep convection and convective overshooting with meteosat. **Advances in Space Research**, v. 19, n. 3, p. 433–441, 1997. 34, 87, 93

SCHMID, W. The prediction of hail. Part I: radar quantities of hail intensity. **Atmospheric Research**, v. 28, n. 1, p. 49–69, 1992. 31

SCHMIDT, T.; LAYNE, D. R.; SERRA, S.; MUPAMBI, G.; KALCSITS, L. A.; MUSACCHI, S.; HANRAHAN, I. **Use of protective netting in Washington State apple production**. [S.l.]: Pullman, Washington: Washington State University Extension, 2019. 2

SCHMIT, T. J.; GRIFFITH, P.; GUNSHOR, M. M.; DANIELS, J. M.; GOODMAN, S. J.; LEBAIR, W. J. A closer look at the abi on the goes-r series. **Bulletin of the American Meteorological Society**, v. 98, n. 4, p. 681–698, 2017. 42

SCHMIT, T. J. et al. Geostationary operational environmental satellite (GOES)-14 super rapid scan operations to prepare for GOES-R. **Journal of Applied Remote Sensing**, v. 7, n. 1, p. 073462, 2013. 42

SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K.-R. Kernel principal component analysis. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS, 10., 1997, Lausanne, Switzerland. **Proceedings...** Lausanne: ICANN/LNCS, 1997. p. 583–588. 92

SCHULTZ, C. J.; CAREY, L. D.; SCHULTZ, E. V.; BLAKESLEE, R. J. Insight into the kinematic and microphysical processes that control lightning jumps. **Weather and Forecasting**, v. 30, n. 6, p. 1591–1621, 2015. 13

_____. Kinematic and microphysical significance of lightning jumps versus nonjump increases in total flash rate. **Weather and Forecasting**, v. 32, n. 1, p. 275–288, 2017. 13

SCHULTZ, C. J.; PETERSEN, W. A.; CAREY, L. D. Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. **Journal of Applied Meteorology and Climatology**, v. 48, n. 12, p. 2543–2563, 2009. 13

_____. Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends. **Weather and Forecasting**, v. 26, n. 5, p. 744–755, 2011. 13

SETVÁK, M. et al. Satellite-observed cold-ring-shaped features atop deep convective clouds. **Atmospheric Research**, v. 97, n. 1-2, p. 80–96, 2010. 34

SHAPLEY, L. S. Utility comparison and the theory of games. **The Shapley Value. Essays in Honor of Lloyd S. Shapley**, p. 307–319, 1969. 73

SHAPLEY, L. S. et al. **A value for n-person games**. [S.l.]: Princeton University Press Princeton, 1953. 73

SHAPLEY, L. S.; SHUBIK, M. A method for evaluating the distribution of power in a committee system. **American Political Science Review**, v. 48, n. 3, p. 787–792, 1954. 73

SHE, Y.; YU, L.; WEI, Y. Application research on intelligent pattern recognition methods in hail identification of weather radar. In: INTERNATIONAL CONFERENCE ON COMPUTER APPLICATION AND SYSTEM MODELING, 1., 2010, Taiyuan, China. **Proceedings...** Taiyuan: ICCASM/IEEE, 2010. p. V11 554–558. 31

SHIN, D.-B.; CHIU, L. S. The inherent error in passive microwave rainfall estimation as inferred from the trmm data. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 1., 2005, Seoul, Korea. **Proceedings...** Seoul: IGARSS/IEEE, 2005. p. V1 1–4. 22

SILVA, V. B.; BERBERY, E. H. Intense rainfall events affecting the la plata basin. **Journal of Hydrometeorology**, v. 7, n. 4, p. 769–787, 2006. 8

SMITH, T. M. et al. Multi-radar multi-sensor (mrms) severe weather and aviation products: initial operating capabilities. **Bulletin of the American Meteorological Society**, v. 97, n. 9, p. 1617–1630, 2016. 36

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, p. 199–222, 2004. 80

SONG, G.; LI, S.; XING, J. Lightning nowcasting with aerosol-informed machine learning and satellite-enriched dataset. **NPJ Climate and Atmospheric Science**, v. 6, n. 1, p. 126, 2023. 61

SPENCER, R.; MARTIN, D.; HINTON, B.; WEINMAN, J. Satellite microwave radiances correlated with radar rain rates over land. **Nature**, v. 304, n. 5922, p. 141–143, 1983b. 31

SPENCER, R.; OLSON, W.; RONGZHANG, W.; MARTIN, D.; WEINMAN, J.; SANTEK, D. Heavy thunderstorms observed over land by the nimbus 7 scanning multichannel microwave radiometer. **Journal of Applied Meteorology and Climatology**, v. 22, n. 6, p. 1041–1046, 1983a. 22

SPENCER, R. W.; GOODMAN, H. M.; HOOD, R. E. Precipitation retrieval over land and ocean with the ssm/i: Identification and characteristics of the scattering signal. **Journal of Atmospheric and Oceanic Technology**, v. 6, n. 2, p. 254–273, 1989. 32

STORELVMO, T.; TAN, I. The wegener–bergeron–findeisen process—its discovery and vital importance for weather and climate. **Meteorologische Zeitschrift**, v. 24, n. 4, p. 455–461, 2015. 5

STRABALA, K. I.; ACKERMAN, S. A.; MENZEL, W. P. Cloud properties inferred from 8 12-μm data. **Journal of Applied Meteorology and Climatology**, v. 33, n. 2, p. 212–229, 1994. 32, 84, 87, 93

ŠTRUMBELJ, E.; KONONENKO, I. Explaining prediction models and individual predictions with feature contributions. **Knowledge and Information Systems**, v. 41, p. 647–665, 2014. 72, 73

STRŽINAR, G.; SKOK, G. Comparison and optimization of radar-based hail detection algorithms in Slovenia. **Atmospheric Research**, v. 203, p. 275–285, 2018. 37

THORAT, P. **Mastering gradient boosting in machine learning: a Comprehensive Guide!** 2023. Available from: <<https://www.linkedin.com/pulse/mastering-gradient-boosting-machine-learning-guide-pratik-thorat/>>. Access in: 2024 Apr. 5. 62

TUOVINEN, J.-P.; PUNKKA, A.-J.; RAUHALA, J.; HOHTI, H.; SCHULTZ, D. M. Climatology of severe hail in finland: 1930–2006. **Monthly Weather Review**, v. 137, n. 7, p. 2238–2249, 2009. 1

UBA, D. M.; NEGRI, R. G.; ENORé, D. P.; COSTA, I. C. da; JORGE, A. A. S. **TATHU - Software para rastreio e análise do ciclo de vida de sistemas convectivos.** São Jose dos Campos, 2022. Available from: <<http://urlib.net/ibi/8JMKD3MGP3W34T/47AF772>>. Access in: 20 Feb. 2024. 43, 44

VALI, G.; DEMOTT, P.; MÖHLER, O.; WHALE, T. A proposal for ice nucleation terminology. **Atmospheric Chemistry and Physics**, v. 15, n. 18, p. 10263–10270, 2015. 3

VANDERPLAS, J. **Python data science handbook: Essential tools for working with data**. [S.l.]: O'Reilly Media, 2016. 63, 66

VILA, D. A.; MACHADO, L. A. T.; LAURENT, H.; VELASCO, I. Forecast and tracking the evolution of cloud clusters (ForTraCC) using satellite infrared imagery: methodology and validation. **Weather and Forecasting**, v. 23, n. 2, p. 233–245, 2008. 43, 45

WALDVOGEL, A.; FEDERER, B.; GRIMM, P. Criteria for the detection of hail cells. **Journal of Applied Meteorology and Climatology**, v. 18, n. 12, p. 1521–1525, 1979. 38

WAPLER, K. The life-cycle of hailstorms: lightning, radar reflectivity and rotation characteristics. **Atmospheric Research**, v. 193, p. 60–72, 2017. 13

WILLIAMS, E.; BOLDI, B.; MATLIN, A.; WEBER, M.; HODANISH, S.; SHARP, D.; GOODMAN, S.; RAGHAVAN, R.; BUECHLER, D. The behavior of total lightning activity in severe florida thunderstorms. **Atmospheric Research**, v. 51, n. 3-4, p. 245–265, 1999. 13

WILLIAMS, E. et al. Contrasting convective regimes over the Amazon: implications for cloud electrification. **Journal of Geophysical Research: Atmospheres**, v. 107, n. D20, p. LBA–50, 2002. 12

WITT, A.; EILTS, M. D.; STUMPF, G. J.; JOHNSON, J.; MITCHELL, E. D. W.; THOMAS, K. W. An enhanced hail detection algorithm for the wsr-88d. **Weather and Forecasting**, v. 13, n. 2, p. 286–303, 1998. 30, 38

ŽIBERT, M. I.; ŽIBERT, J. Monitoring and automatic detection of the cold-ring patterns atop deep convective clouds using meteosat data. **Atmospheric Research**, v. 123, p. 281–292, 2013. 34

ZIEGLER, C. L.; RAY, P. S.; KNIGHT, N. C. Hail growth in an Oklahoma multicell storm. **Journal of Atmospheric Sciences**, v. 40, n. 7, p. 1768–1791, 1983. 5

ZIPSER, E. J.; CECIL, D. J.; LIU, C.; NESBITT, S. W.; YORTY, D. P. Where are the most intense thunderstorms on earth? **Bulletin of the American Meteorological Society**, v. 87, n. 8, p. 1057–1072, 2006. 8, 11

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.