



SISTEMA DE PREVISÃO DE EVENTOS METEOROLÓGICOS CONVECTIVOS PARA A ÁREA TERMINAL DO RIO DE JANEIRO

Vinícius Albuquerque de Almeida

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Civil.

Orientadores: Nelson Francisco Favilla
Ebecken
Gutemberg Borges França
Haroldo Fraga de Campos Velho

Rio de Janeiro
Setembro de 2020

SISTEMA DE PREVISÃO DE EVENTOS METEOROLÓGICOS
CONVECTIVOS PARA A ÁREA TERMINAL DO RIO DE JANEIRO

Vinícius Albuquerque de Almeida

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA CIVIL.

Orientadores: Nelson Francisco Favilla Ebecken
Gutemberg Borges França
Haroldo Fraga de Campos Velho

Aprovada por: Prof. Nelson Francisco Favilla Ebecken
Prof. Gutemberg Borges França
Prof. Haroldo Fraga de Campos Velho
Prof. Rogério Pinto Espíndola
Prof. Elton Fernandes
Prof. Lucio Pereira de Andrade
Prof. Saulo Ribeiro de Freitas

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2020

Almeida, Vinícius Albuquerque de

Sistema de previsão de eventos meteorológicos convectivos para a área terminal do Rio de Janeiro/Vinícius Albuquerque de Almeida. – Rio de Janeiro: UFRJ/COPPE, 2020.

XVII, 99 p.: il.; 29, 7cm.

Orientadores: Nelson Francisco Favilla Ebecken

Gutemberg Borges França

Haroldo Fraga de Campos Velho

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Civil, 2020.

Referências Bibliográficas: p. 89 – 97.

1. nowcasting. 2. machine learning. 3. eventos convectivos. 4. assimilação de dados. I. Ebecken, Nelson Francisco Favilla *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

*Aos meus pais e minha esposa
por sempre estarem ao meu lado
em minhas escolhas.*

Agradecimentos

Agradeço primeiramente a Deus por mais uma conquista.

Agradeço a meus pais, Antônio de Almeida Filho e Vera Regina Albuquerque Ferreira, pelo constante apoio e pela paciência e atenção que sempre tiveram, pelo entusiasmo e orgulho com que sempre reagiram aos resultados acadêmicos ao longo dos anos.

À minha esposa, pelo companheirismo apesar da falta de atenção e das ausências ao longo dos últimos anos. Este resultado não seria possível sem você ao meu lado.

Agradeço ao meu orientador Dr. Gutemberg Borges França pela amizade e por ter acreditado em meu potencial em 2018, quando resolvi deixar o mercado privado e retornar ao ambiente acadêmico.

Agradeço ao meu orientador Dr. Haroldo Fraga de Campos Velho, pela inspiração científica e por todos os ensinamentos como profissional e como ser humano.

Agradeço ao meu orientador Dr. Nelson F. Favilla Ebecken por ter acreditado no potencial do trabalho que foi proposto no ano de 2018 antes do meu ingresso no Curso de Doutorado e por todas as contribuições científicas em nossas reuniões e seminários.

Agradeço a todos os professores e pesquisadores da Pós-Graduação do Programa de Engenharia Civil da COPPE/UFRJ e do Programa de Pós-graduação em Meteorologia do IGEO/UFRJ, com os quais muito aprendi através das disciplinas do curso e seminários.

Agradeço ao Laboratório de Meteorologia Aplicada (LMA) por proporcionar a infraestrutura para desenvolvimento do trabalho de pesquisa nos últimos anos.

Agradeço a todos os meus amigos do LMA, em especial ao Dr. Francisco Leite de Albuquerque Neto pela amizade e pelas contribuições sempre muito pertinentes nas discussões do meu trabalho.

Ao CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo apoio financeiro.

Ao Departamento de Controle do Espaço Aéreo (DECEA) por meio da Organização Brasileira para o Desenvolvimento Científico e Tecnológico do Controle do Espaço Aéreo (CTCEA) pelas discussões técnicas semestrais ao longo dos ciclos de seminário sobre Meteorologia Aeronáutica.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

SISTEMA DE PREVISÃO DE EVENTOS METEOROLÓGICOS CONVECTIVOS PARA A ÁREA TERMINAL DO RIO DE JANEIRO

Vinícius Albuquerque de Almeida

Setembro/2020

Orientadores: Nelson Francisco Favilla Ebecken
Gutemberg Borges França
Haroldo Fraga de Campos Velho

Programa: Engenharia Civil

Neste trabalho são utilizadas duas abordagens de previsão de tempo usando algoritmos de aprendizado de máquina para a área terminal de voo do Aeroporto Internacional do Rio de Janeiro. O primeiro método utiliza índices termodinâmicos extraídos das sondagens do aeroporto como variável de entrada para o modelo e, como variável objetivo, a previsão de (a) ocorrência e (b) severidade de eventos convectivos – classificados a partir de descargas atmosféricas registradas pela rede RINDAT entre 2001 e 2016. Os resultados encontrados para previsão de ocorrência (severidade) apontam mais de 90% (84%) de probabilidade de detecção e cerca de 10% (20%) de falso alarme. Um segundo método, consiste no treinamento de redes neurais artificiais para assimilação de dados locais (superfície e perfil) no modelo atmosférico regional Weather Research Forecasting aplicada em campos de previsão de 6 horas. Este método se mostrou eficiente para emular a abordagem tradicional de assimilação (3d-var) sem perda significativa de informação e com duas grandes vantagens: tempo de execução cerca de 70 vezes mais rápido e poder ser executado em máquinas de pequeno porte, sem a necessidade de supercomputadores.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

FORECAST SYSTEM OF METEOROLOGICAL CONVECTIVE EVENTS AT
THE TERMINAL AREA OF RIO DE JANEIRO

Vinícius Albuquerque de Almeida

September/2020

Advisors: Nelson Francisco Favilla Ebecken
Gutemberg Borges França
Haroldo Fraga de Campos Velho

Department: Civil Engineering

In this work, two weather forecasting approaches using machine learning algorithms for the flight terminal area of the International Airport of Rio de Janeiro are explored. The first method uses thermodynamic stability indices extracted from the airport soundings as input variable for the model and, as target variable, the prediction of (a) occurrence and (b) severity of meteorological convective events - classified from atmospheric discharges recorded by the RINDAT network between 2001 and 2016. The results found for the prediction of occurrence (severity) indicate more than 90% (84%) probability of detection and about 10% (20%) false alarm. A second method consists of training artificial neural networks for local data assimilation (surface and profile) in the regional atmospheric model Weather Research Forecasting applied to 6-hour forecast fields. This method proved to be efficient to emulate the traditional data assimilation approach (3d-var) without significant loss of information and with two great advantages: execution time around 70 times faster and can be used without high-performance hardwares.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiv
Lista de Abreviaturas	xv
1 Introdução	1
1.1 Importância e planos da aviação mundial	4
1.2 Estado da arte	6
1.2.1 <i>Nowcasting</i>	6
1.2.2 Eventos convectivos	7
1.2.3 Redes Neurais Artificiais no contexto da meteorologia	10
1.3 Iniciativa pioneira no Brasil	13
1.4 Contribuição	14
1.5 Objetivos	14
1.6 Estrutura do trabalho	15
2 Fundamentação Teórica	16
2.1 Eventos meteorológicos convectivos (EMC)	16
2.1.1 Índices de Instabilidade	16
2.1.2 Descargas atmosféricas	17
2.2 Modelos físicos	19
2.2.1 Conceitos	19
2.2.2 Assimilação de dados	20
2.2.3 Limitações dos métodos tradicionais	20
2.2.4 <i>Weather Research and Forecasting</i>	21
2.2.5 Método de assimilação 3D-Var	21
2.2.6 Assimilação de dados por redes neurais artificiais	22
2.2.7 Algoritmo de colisão multipartículas	23
3 Área de Estudo e Dados	25
3.1 Área de Estudo	25

3.2	Dados	28
4	Aprendizado de máquina e Meteorologia Aeronáutica	31
4.1	Estrutura geral para <i>nowcasting</i>	31
4.2	Método I - Previsão de eventos meteorológicos convectivos para a área terminal do Rio de Janeiro	33
4.2.1	Fluxograma	33
4.2.2	Organização dos dados	34
4.2.3	Seleção e integração dos dados	37
4.2.4	Higienização	37
4.2.5	Classificação dos eventos	37
4.2.6	Modelagem	38
4.2.7	Avaliação dos resultados	42
4.3	Método II - Assimilação de dados locais usando redes neurais artificiais	43
4.3.1	Sensibilidade do modelo WRF ao método 3D-Var	43
4.3.2	Redes neurais para assimilação de dados sintéticos	43
4.3.3	Redes neurais para assimilação de dados observados da RE-DEMET	44
5	Resultados e Discussões	46
5.1	<i>Nowcasting</i> : modelo de previsão de EMC	46
5.1.1	Treinamento e validação	46
5.1.2	Teste para o período de 2011 a 2016	52
5.1.3	Caracterização de EMC mais intensos	54
5.2	Assimilação de dados usando redes neurais	56
5.2.1	Sensibilidade do modelo WRF à assimilação (3D-Var)	56
5.2.2	Assimilação de dados sintéticos com redes neurais	67
5.2.3	Assimilação de dados das estações com redes neurais	77
6	Conclusões e Trabalhos Futuros	84
6.1	Conclusões	84
6.2	Trabalhos Futuros	86
6.3	Contribuições	87
	Referências Bibliográficas	89
A	Sistema NOW-BR	98

Lista de Figuras

1.1	(a) e (b) Estatísticas para acidentes relacionados a aeronaves relacionados a parâmetros meteorológicos de 1994 a 2000: um número real de acidentes e distribuições probabilísticas em gráfico de pizza (dados do NTSB 2010; 1996). (c) e (d) Condições de tempo como causa / fator durante todos os acidentes para o período de 2000-2011. As estatísticas baseadas no NTSB (<i>National Transportation Safety Board</i>) que resultaram em 19.441 acidentes e 29% desses acidentes estavam relacionados às condições meteorológicas (c) estatísticas específicas de uma companhia aérea no período 2000-2011 (d). Adaptado de GULTEPE <i>et al.</i> (2019).	2
3.1	Área de estudo centrada no Aeroporto Internacional do Rio de Janeiro (Galeão) com 150 km de raio.	26
3.2	Domínio computacional e localização dos pontos de coleta de dados (aeroportos existentes na área de estudo). Os aeroportos são representados por indicadores de localidade definidos pela Organização da Aviação Civil Internacional (OACI), a saber: SBAF - Base Aérea dos Afonsos; SBSC - Base Aérea de Santa Cruz; SBJR - Aeroporto de Jacarepaguá; SBRJ - Aeroporto Santos Dumont; SBGL - Aeroporto Internacional do Rio de Janeiro - Galeão.	27
4.1	Esquema ilustrativo do processo de previsão de eventos meteorológicos convectivos.	32
4.2	Fluxograma com a sequência de atividades que foram realizadas no desenvolvimento do Método I.	33
4.3	Arquitetura para o modelo DL-FC para previsão de (a) ocorrência e (b) severidade.	41
5.1	Gráfico de dispersão POD x (1-FAR) para resultados de validação cruzada no conjunto de dados de treinamento (2001-2010) para previsão de ocorrência do evento meteorológico convectivo (EMC). . . .	50

5.2	Gráfico de dispersão POD x (1-FAR) para resultados de validação cruzada no conjunto de dados de treinamento (2001-2010) para previsão de severidade do evento meteorológico convectivo (EMC).	51
5.3	Análise do conjunto de teste (2011 a 2016) considerando diferentes limiares de descargas atmosféricas nos eventos meteorológicos convectivos.	53
5.4	Metodologia de teste de sistema de assimilação de dados. Adaptado de um curto do <i>Data Assimilation Research Testbed</i> disponível em https://dart.ucar.edu/DART_LAB/presentation/DART_LAB_Section01.pdf , slide 45.	57
5.5	Teste de sensibilidade de domínios de grade única e aninhada realizado com processo de assimilação de dados aplicado à previsões de 6 horas nas estações referentes aos aeroportos na Região Metropolitana do Rio de Janeiro.	58
5.6	Campo de temperatura do ar para (a) previsão de 6 horas a partir de condição inicial sem assimilação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimilação, (c) previsão de 6 horas com assimilação 3D-Var e (d) diferença (c) e (a), para 1º de fevereiro de 2014 06 UTC.	61
5.7	Campo de umidade relativa para (a) previsão de 6 horas a partir de condição inicial sem assimilação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimilação, (c) previsão de 6 horas com assimilação 3D-Var e (d) diferença (c) e (a), para 1º de fevereiro de 2014 06 UTC.	62
5.8	Campo de velocidade do vento para (a) previsão de 6 horas a partir de condição inicial sem assimilação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimilação, (c) previsão de 6 horas com assimilação 3D-Var e (d) diferença (c) e (a), para 1º de fevereiro de 2014 06 UTC.	63
5.9	Perfil de vento para (a) previsão de 6 horas a partir de condição inicial sem assimilação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimilação, (c) previsão de 6 horas com assimilação 3D-Var e (d) diferença (c) e (d), para 1º de fevereiro de 2014 06 UTC.	64
5.10	Erro no campo de temperatura do ar da observação para previsão de 6 horas sem assimilação e previsão de 6 horas com assimilação no SBGL entre 1 e 7 de fevereiro de 2014 e de 1 a 7 de fevereiro de 2015.	66

5.11	Gráfico quantil-quantil comparando a distribuição de probabilidade da temperatura do ar gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.	71
5.12	Gráfico quantil-quantil comparando a distribuição de probabilidade da umidade relativa gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.	72
5.13	Gráfico quantil-quantil comparando a distribuição de probabilidade da componente u do vento gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.	73
5.14	Gráfico quantil-quantil comparando a distribuição de probabilidade da componente v do vento gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.	74
5.15	Campo de temperatura do ar em 01/fev/2014 06 UTC para: (a) previsão de 6 horas; (b) previsão de 6 horas com assimilação 3D-var; (c) previsão de 6 horas com assimilação MPCA; e (d) diferença de (c) e (b). Os pontos vermelhos representam a localização das estações de superfície onde os dados assimilados foram medidos.	76
5.16	Mapa de temperatura do campo (a) previsão de 6 horas sem assimilação de dados (controle), (b) previsão de 6 horas com assimilação de dados 3D-Var e previsão de 6 horas com assimilação de dados por redes neurais - treinado no (c) TensorFlow e (d) Weka - em 1º de fevereiro de 2019 12 UTC.	79
5.17	Mapa de erro de temperatura dos campos de previsão de 6 horas com assimilação de dados (a) 3D-Var e redes neurais - treinados em (b) TensorFlow e (c) Weka - em relação ao mapa observado em 1º de fevereiro de 2019 12 UTC.	81

5.18	Perfil no período 1 a 7/2/2019 para os erros médios de previsão de 6 horas entre: perfil de previsão de 6 horas com assimilação de dados usando o <i>tensorflow</i> e dados observados (tf-obs); perfil de previsão de 6 horas com assimilação de dados usando o <i>tensorflow</i> e perfil de previsão de 6 horas com assimilação de dados usando o 3D-Var (tf-3dvar); perfil de previsão de 6 horas com assimilação de dados usando 3dvar e perfil médio observado (3dvar-obs); previsão de 6 horas sem assimilação de dados e perfil médio observado (back-obs); previsão de 6 horas com assimilação de dados usando o <i>weka</i> e previsão de 6 horas com assimilação de dados usando o 3dvar (weka-3dvar); previsão de 6 horas com assimilação de dados usando o <i>weka</i> e perfil médio observado (weka-obs0. O eixo y contém alturas da superfície até cerca de 20 km.	83
A.1	Esquema ilustrativo dos tipos de dado de entrada, modelos e fenômenos contemplados pelo NOW-BR.	99

Lista de Tabelas

2.1	Índices de instabilidade atmosféricos normalmente utilizados como indicativos de tempestades.	17
3.1	Dados utilizados para construção do modelo de previsão de eventos convectivos na Área Terminal do Rio de Janeiro e para o processo de assimilação de dados utilizando redes neurais artificiais.	30
4.1	Formato dos dados originais de descargas atmosféricas.	35
4.2	Formato dos dados originais de sondagens	36
4.3	Estrutura dos dados integrados	37
4.4	Variáveis de entrada, saída e suas respectivas descrições.	38
4.5	Classificadores utilizados para treinamento	40
5.1	Resultados da previsão de ocorrência de tempestade.	47
5.2	Resultados da previsão de tempestade para os 30 principais eventos do conjunto de dados de teste.	54
5.3	Estatísticas para todos os pontos de grade e variáveis de superfície no domínio de estudo.	67
5.4	Resultados do MPCA para o conjunto de treinamento e validação para variáveis meteorológicas.	69
5.5	Estatísticas do MPCA para o conjunto de generalização. Na tabela a sigla REQM representa a raiz do erro quadrático médio, DESVPAD o desvio padrão e R2 o coeficiente de determinação.	70
5.6	Topologia final e outras características de redes neurais nos experimentos do TensorFlow e Weka.	78
5.7	Tempo de treinamento para redes neurais. A hora é formatada como hh: mm: ss, em que hh representa as horas, mm representa os minutos e ss representa os segundos ou fração.	78
6.1	Resumo da contribuição científica realizada no âmbito da tese.	88

Lista de Abreviaturas

ANAC	Agência Nacional de Aviação Civil, p. 3, 25
API	<i>Application Programming Interface</i> , p. 39
ATM	<i>Air traffic management</i> , p. 5
CAN-Now	<i>Canadian Airport Nowcasting Project</i> , p. 7
CAPE	<i>Convective Available Potential Energy</i> , p. 38
CENIPA	Centro de Investigação e Prevenção de Acidentes Aeronáuticos, p. 3
CINE	<i>Convective inhibition energy</i> , p. 38
CSI	<i>Critical Success Index</i> , p. 47
DART	Data Assimilation Research Testbed, p. 56
DA	Descargas Atmosféricas, p. 9
DECEA	Departamento do Controle do Espaço Aéreo, p. 6
DL-FC	<i>Deep Learning Fully-Connected network</i> , p. 38
EMA	Estação Meteorológica de Altitude, p. 28
EMC	Eventos Meteorológicos Convectivos, p. 15
FAA	<i>Federal Aviation Administration</i> , p. 5, 21
FAR	<i>False alarm rate</i> , p. 47
GANP	<i>Global Air Navigation Plan</i> , p. 5
GFS	<i>Global Forecast System</i> , p. 29
GPU	<i>Graphics Processing Unit</i> , p. 39
IA	Inteligência Artificial, p. 9

LETKF	<i>Local Ensemble Transform Kalman Filter</i> , p. 12
MLP	<i>Multi-layer perceptron</i> , p. 12
MPCA	<i>Multi-Particle Collision Algorithm</i> , p. 13
MRJ	Município do Rio de Janeiro, p. 10
NAS	<i>National Airspace System</i> , p. 5
NCAR	<i>National Center for Atmospheric Research</i> , p. 21
NCEP	<i>National Centers for Environmental Prediction</i> , p. 21
NN	<i>Neural Networks</i> , p. 12
NOAA	<i>National Oceanic and Atmospheric Administration</i> , p. 3
NOW-BR	Sistema Brasileiro de Nowcasting, p. 13
NWS	<i>National Weather Service</i> , p. 3
OACI	Organização da Aviação Civil Internacional, p. 4
PCA	Particle Collision Algorithm, p. 23
PMC	<i>Perceptrons</i> multicamadas, p. 11
PNT	Previsão Numérica do Tempo, p. 20
POD	<i>Probability of detection</i> , p. 47
RBF	<i>Radial Basis Function</i> , p. 42
RINDAT	Rede Integrada Nacional de Detecção de Descargas Atmosféricas, p. 17
RMRJ	Região Metropolitana do Rio de Janeiro, p. 9
RNA	Redes Neurais Artificiais, p. 22
SBGL	Aeroporto Internacional do Galeão, p. 25
SCIT	<i>Storm Cell Identification and Tracking</i> , p. 6
SISCEAB	Sistema de Controle do Espaço Aéreo Brasileiro, p. 6
SPEEDY	<i>Simplified parameterizations, primitive-Equation Dynamics</i> , p. 12

TS	<i>Threat Score</i> , p. 47
UCAR	<i>University Corporation for Atmospheric Research</i> , p. 56
WRFDA	<i>Weather Research and Forecasting Data Assimilation</i> , p. 21
WRF	<i>Weather Research and Forecasting</i> , p. 8

Capítulo 1

Introdução

Numerosos estudos mostraram que o tempo afeta severamente as operações de aviação civil e de defesa. O impacto dos processos atmosféricos na aviação é reconhecido desde 1900. Por exemplo, DINES (1917) afirmou que "assim parece que a exigência do aviador ao meteorologista é que ele seja capaz de prever vento e neblina e, em menor medida, nuvens, na rota, o aviador está planejando seguir.". Atualmente, seus comentários sobre parâmetros relacionados à aviação, como velocidade do vento e visibilidade, ainda são válidos (fig. 1.1) (GULTEPE *et al.* (2019)).

Condições de tempo que causam ou contribuem para o acidentes de aviação incluem vento, visibilidade / teto, turbulência, congelamento do carburador, correntes ascendentes/descendentes, precipitação, gelo, tempestades, cisalhamento do vento, elevação térmica, temperaturas extremas e relâmpagos (NTSB (2010)). A figura 1.1a mostra um gráfico de barras das estatísticas relacionadas ao clima condições de 1994 a 2003 que afetam a superfície próximas operações de aviação; eles são principalmente vento e visibilidade (NTSB (2010)). Na figura 1.1a, condições de visibilidade, altura do teto, e precipitação ocorreram 485 vezes, vento e turbulência 1381 vezes, e gelo e congelamento do motor 150 vezes. Este trabalho sugeriu que de 2003 a 2007, houve 8.657 aviação acidentes relacionados e as condições de tempo foram um fator em 1740 de esses acidentes. A figura 1.1b mostra esses parâmetros em percentis também para o período de 1994–2003; vento e visibilidade ainda são os parâmetros mais críticos. Para aeronaves pequenas e não comerciais a causa primária de acidentes relacionados com o clima de 2000 a 2011 foram ventos adversos, seguidos de tetos baixos (fig. 1.1c). Em níveis de cruzeiro de aeronaves comerciais, esta imagem é diferente, com mais de 70% dos acidentes associados à condições de tempo de 2000 a 2011 sendo relacionado à turbulência (fig. 1.1d) (GULTEPE *et al.* (2019)).

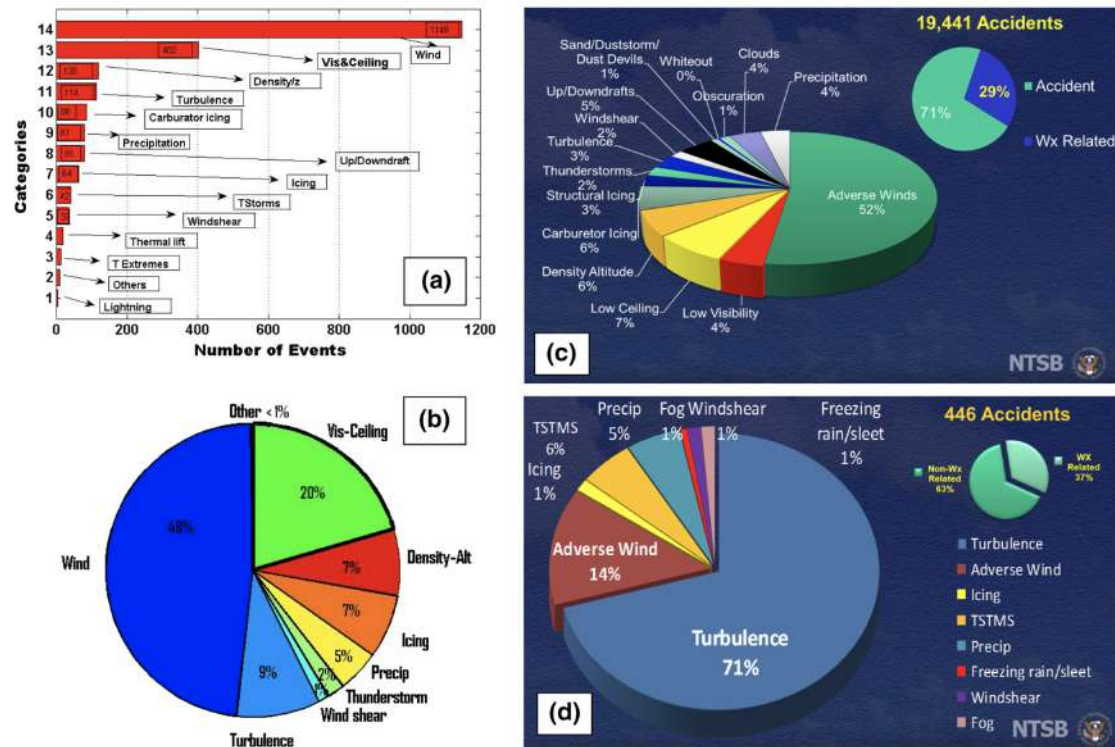


Figura 1.1: (a) e (b) Estatísticas para acidentes relacionados a aeronaves relacionados a parâmetros meteorológicos de 1994 a 2000: um número real de acidentes e distribuições probabilísticas em gráfico de pizza (dados do NTSB 2010; 1996). (c) e (d) Condições de tempo como causa / fator durante todos os acidentes para o período de 2000-2011. As estatísticas baseadas no NTSB (*National Transportation Safety Board*) que resultaram em 19.441 acidentes e 29% desses acidentes estavam relacionados às condições meteorológicas (c) estatísticas específicas de uma companhia aérea no período 2000-2011 (d). Adaptado de GULTEPE *et al.* (2019).

Atrasos e danos à aeronave durante o pouso, decolagem e voo em nível de cruzeiro devido a condições perigosas do tempo podem ocorrer frequentemente. GULTEPE *et al.* (2017a,b) afirmaram que um evento de névoa congelante de três dias no Reino Unido resultou em cerca de 50 M US\$ em perdas financeiras para as empresas. KESSINGER *et al.* (2016) também afirmaram que os riscos da aviação para voos oceânicos podem afetar a segurança, a eficiência econômica e a produtividade, o que leva a um custo total de cerca de 62,7 M US\$ por ano. COOK *et al.* (2009) quantificaram o impacto dos fatores climáticos nos atrasos de voos e descobriram que os fatores contribuintes são altamente dependentes do aeroporto. Devido à sua importância para a aviação, o *National Weather Service* (NWS) da *National Oceanic and Atmospheric Administration* (NOAA) fornece previsões de altura do topo das nuvens, turbulência, detecção de raios, precipitação, gelo, nuvens baixas e neblina e cinzas vulcânicas. Essas previsões estão disponíveis ao público em seu site (<https://www.aviationweather.gov/>) (GULTEPE *et al.* (2019)).

Em particular, a área terminal de um aeroporto é a área onde a aeronave está esperando para pousar ou decolar e, portanto, é bastante sensível às condições meteorológicas. Os controladores de tráfego aéreo e os pilotos requerem informações precisas sobre as condições meteorológicas na área terminal para tomar decisões de curto prazo que se enquadram na escala de tempo de previsão a curto prazo, que varia do intervalo de alguns minutos até 6 h (FRANÇA *et al.* (2016)).

No Brasil, um meteorologista está atualmente usando sua experiência para integrar diferentes observações meteorológicas *in situ* e/ou saídas de modelos atmosféricos usando modelos conceituais sobre como a atmosfera funciona para gerar previsões atuais nos principais aeroportos (FRANÇA *et al.* (2016)).

De acordo com a Agência Nacional de Aviação Civil (ANAC) e informações do Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA), entre 2007 e 2017, foram registradas no Brasil 121 ocorrências em que as condições meteorológicas adversas atuaram como possíveis fatores contribuintes. Entre os fatores meteorológicos de destaque podem-se citar: trovoadas, rajadas de vento e nevoeiro. Mais informações podem ser obtidas a partir dos relatórios estatísticos disponíveis em <https://www2.fab.mil.br/cenipa/index.php/estatisticas>.

A previsão local de tempestades é uma das tarefas mais desafiadoras na previsão do tempo devido à sua alta variabilidade espaço-temporal. As condições do tempo interferem no cotidiano de todos, mas poucas atividades humanas são tão dependentes das condições da atmosfera quanto a navegação aérea. Vários sistemas e fenômenos meteorológicos têm sido apontados como os responsáveis pelos acidentes e incidentes aeronáuticos ocorridos no mundo; o que significa que a compreensão da atmosfera torna possível a condução de voos mais seguros.

Os eventos convectivos estão entre as principais causas das catástrofes naturais

que impactam diretamente a sociedade e o meio ambiente. Portanto é de fundamental importância a caracterização desses fenômenos e conseqüentemente o conhecimento termodinâmico do mesmo para aumentar a assertividade de suas previsões.

O crescente aumento no número de aeronaves em operação exige um melhor aproveitamento do espaço aéreo, e assim, com a devida segurança, diminuindo o espaçamento entre elas. Nesse contexto, as informações meteorológicas passam a ser cada vez mais decisivas.

1.1 Importância e planos da aviação mundial

Segundo a Organização da Aviação Civil Internacional (OACI), o transporte aéreo desempenha papel de destaque na condução de um desenvolvimento econômico e social sustentável. A aviação direta e indiretamente emprega 58,1 milhões de pessoas, contribui com mais de US\$ 2,4 trilhões no Produto Doméstico Bruto global e carrega mais de 3,3 bilhões de passageiros e cargas com valor de cerca de US\$ 6,4 trilhões anualmente. As operações da aviação envolvem a distribuição de suprimentos, fluxos de turismo e permitem o acesso de produtores locais a mercados distantes. Para garantir o avanço das melhorias de segurança e modernização da navegação aérea, a OACI desenvolveu uma abordagem estratégica para o progresso em ambas as áreas (ICAO (2016)).

O Plano Global de Navegação Aérea (GANP) da OACI tem por objetivo guiar o progresso do sistema de transporte aéreo global no período entre 2016-2030 e é revisado trienalmente por um conselho da OACI. Na 5ª edição do GANP (ICAO (2016)) a OACI introduz uma metodologia estratégica de atualização por blocos (*Block Upgrades*) que aproveita as tecnologias existentes no mercado e busca antecipar desenvolvimentos futuros baseados nos objetivos operacionais acordados entre a indústria da aviação e os programas dos países membros de seu conselho (ex: SIRIUS no Brasil, NextGen nos Estados Unidos da América, entre outros). Estes blocos são organizados em períodos de 6 anos não sobrepostos começando em 2013 e continuando até 2031 (e adiante). Estes blocos fornecem a base das estratégias de investimento e um compromisso dos países membros, da indústria, dos operadores e provedores de serviços. A OACI desenvolveu a metodologia global de atualização por blocos primariamente para garantir que a segurança da aviação seja mantida e aprimorada, que os programas de aprimoramento do Gerenciamento do Tráfego Aéreo (ATM, em inglês) estejam em harmonia, e que as barreiras aos futuros ganhos de eficiência e ambientais da aviação possam ser removidas a um custo razoável. As principais áreas de aprimoramento de desempenho da aviação são definidas como: operações de aeroportos, interoperabilidade global de sistemas e dados, capacidade ótima, voos flexíveis e rotas aéreas eficientes. Uma sequência de 4 blocos foi definida

com base em temas: bloco 0 - tecnologias e conhecimento que já foi desenvolvido e implementado em várias partes do mundo; blocos 1 ao 3 - caracterizados por soluções existentes e projetadas com marcos de início em 2019, 2025 e 2031, respectivamente.

Segundo o GANP, a meteorologia aeronáutica encontra-se na segunda grande área de aprimoramento de desempenho da aviação, nomeada "Interoperabilidade global de sistemas e dados", fazendo parte da *thread* (linha de trabalho) AMET que é citada nos blocos 0, 1 e 3. No bloco 0 (B0-AMET), o papel da meteorologia é fornecer informação para o suporte a melhorias da eficiência e segurança operacional. No bloco 1 (B1-AMET), há a expectativa de que a meteorologia auxilie a tomada de decisão no ambiente operacional através de informações meteorológicas integradas que forneçam suporte ao ATM (Planejamento). Por último, no bloco 3 (B3-AMET), o papel da meteorologia passa a ser fundamental para o suporte (aéreo e em solo) e para a implementação de estratégias de mitigação de eventos meteorológicos em tempo real.

A *Federal Aviation Administration* (FAA), parte do Departamento de Transportes dos Estados Unidos da América, através do documento *The Future of the NAS* (FAA (2016)) descreve o estágio ,até 2016, da transformação do *National Airspace System* (NAS) no programa *Next Generation air transportation system* (Next-Gen), cujos objetivos principais são: aumento da capacidade e eficiência do sistema garantindo a sua segurança e a mitigação dos impactos ambientais. Importante notar que as condições do tempo são citadas diversas vezes neste relatório como fator importante para o alcance dos objetivos anteriormente listados, entre diversos outros aspectos que não serão aqui abordados. Segundo o relatório, a maior parte dos atrasos existentes no tráfego aéreo do sistema atual estão associados às condições do tempo. Os avanços propostos pelo NextGen fornecerão aos tomadores de decisão acesso a recursos avançados de integração das informações meteorológicas nos processos e ferramentas de tomada de decisão. A subjetividade das interpretações meteorológicas no presente será reduzida de forma significativa. A melhoria dos dados meteorológicos e das previsões permitem a automação da tradução dos dados meteorológicos disponíveis em restrições à aviação e a avaliação dos impactos no NAS, assim como a assimilação destes dados nas tomadas de decisão do ATM.

No Brasil, o Comando da Aeronáutica, através do Departamento do Controle do Espaço Aéreo (DECEA) no documento DCA 351-2/2011 "Concepção Operacional ATM Nacional" (DECEA (2011)), no contexto do programa nacional SIRIUS, destaca os principais benefícios da informação meteorológica para o Sistema ATM: (a) melhor adequação do planejamento e a previsão das trajetórias de voo, beneficiando a segurança operacional e eficiência do sistema ATM; (b) maior disponibilidade de informações meteorológicas visando a melhor performance de voo e à otimização operacional do espaço aéreo; (c) melhor formatação e apresentação das condições

meteorológicas adversas atenuando seus efeitos sobre as aeronaves; (d) melhor acesso à informação para melhorar a utilização da capacidade disponível; (e) melhor representatividade das condições meteorológicas para o emprego operacional direto dos dados obtidos por aeronaves e a melhoria das previsões meteorológicas; e (f) redução do efeito das aeronaves sobre o meio ambiente. O documento ainda destaca a necessidade da transição do atual sistema de meteorologia aeronáutica para um sistema automatizado e em tempo real que atenda às necessidades do Sistema de Controle do Espaço Aéreo Brasileiro (SISCEAB) como parte do Sistema ATM Nacional e Global.

1.2 Estado da arte

1.2.1 *Nowcasting*

MASS (2012), SUN *et al.* (2013), FRANÇA *et al.* (2016) e GULTEPE *et al.* (2019) apresentam revisões abrangentes sobre os esforços recentes que foram feitos no contexto do nowcasting desde a década de 1960. Os trabalhos mais antigos sobre *nowcasting* eram limitados principalmente à interpretação subjetiva e extrapolação temporal de radar meteorológico (WILSON (1966); BATAN (1973); WILSON e WILK (1981)) e satélite (PURDOM (1976)).

Na década de 90, o algoritmo TITAN (DIXON e WIENER (1993)) apresenta uma metodologia de acompanhamento de células convectivas incluindo a sua evolução temporal e de intensidade utilizando-se dados de radar. Neste mesmo contexto, JOHNSON *et al.* (1998) apresentou o algoritmo *Storm Cell Identification and Tracking* (SCIT) para monitoramento do centroide das células e complexos convectivos. Em seguida, trabalhos incluindo a assimilação de dados observados (WILSON *et al.* (1998); XUE *et al.* (2003); SCHROEDER *et al.* (2006); LIU *et al.* (2008)) foram desenvolvidos para o aperfeiçoamento de modelos de previsão numérica do tempo. Mais recentemente HAN *et al.* (2009) apresentou uma versão melhorada do algoritmo TITAN para identificação e acompanhamento de células convectivas.

Alguns experimentos de campo foram conduzidos para permitir o desenvolvimento dos métodos de *nowcasting*, tais como: *Chesapeake Bay Nowcasting Experiment* (SCOFIELD e WEISS (1977)), Olimpíadas de verão em Atlanta (ROTHFUSZ *et al.* (1998)), *Spring Forecast Experiment* (CONIGLIO *et al.* (2010); KAIN *et al.* (2003); KAIN *et al.* (2010)) e um experimento em Helsinque, Finlândia (KOSKINEN *et al.* (2011)).

Considerando a meteorologia aeronáutica, ISAAC *et al.* (2006), ISAAC *et al.* (2011) e ISAAC *et al.* (2014) apresentaram uma sequência de trabalhos que resultou em um sistema de previsão de tempo refinado para aviação que usa dados de

modelos numéricos, observações de superfície, radar, satélite e um radiômetro de microondas para gerar nowcasting para os principais aeroportos no Canadá, até aproximadamente 6h.

Recentemente, FRANÇA *et al.* (2018) apresentaram um modelo de *nowcasting* para perfis de vento próximo à superfície (até 500 m) no Aeroporto Internacional de Guarulhos, São Paulo. Um importante aspecto destacado neste trabalho é que no Brasil, a previsão da aviação é baseada principalmente em uma interpretação subjetiva da observação disponível e dados de saída do modelo, juntamente com modelos conceituais de atmosfera.

CAN-NOW

ISAAC *et al.* (2012, 2014) apresentam o *Canadian Airport Nowcasting Project* (CAN-Now), um protótipo avançado de um sistema de *nowcasting* que pode ser usado nos aeroportos. Este sistema utiliza dados de modelos numéricos, relatórios de piloto, observações de sensores *in situ* (precipitação, gelo, teto, visibilidade, ventos), sensoriamento remoto no local (como radar e radiômetro de micro-ondas) e sensores remotos (informações de satélite e radar) para fornecer *nowcasting* detalhados para aproximadamente 6 h. O sistema foi desenvolvido e testado no Aeroporto Internacional Pearson de Toronto (CYYZ) e no Aeroporto Internacional de Vancouver (CYVR). Uma carta de situação foi desenvolvida para permitir que os usuários tenham um produto de alto valor, que identifica problemas de tempo significativo nos aeroportos. Novos produtos que combinam observações e saída numérica do modelo em *nowcasting* foram testados. Verificações estatísticas foram feitas dos produtos de previsão, com comparações com a persistência, abrangendo um período de inverno (2009/2010) e verão (2010). Problemas com a previsão de umidade relativa e direção do vento são descritos. A capacidade de prever variáveis categóricas, como teto, visibilidade, bem como taxa de precipitação e tipo com precisão são discutidas. No geral, para a maioria das variáveis, os sistemas de *nowcasting* podem superar a persistência após as primeiras 2 h e fornecer previsões mais precisas do que os modelos individuais de Previsão Numérica do Tempo para 6 h.

1.2.2 Eventos convectivos

O conhecimento e a previsão dos sistemas convectivos têm adquirido grande relevância ao longo dos anos já que esses sistemas são responsáveis por desligamentos da rede elétrica, alagamentos, enchentes e deslizamentos de encostas, trazendo transtornos às atividades humanas e, até mesmo, levando a perdas de vidas.

NASCIMENTO (2005) faz uma revisão abrangente sobre eventos convectivos com o objetivo de alimentar uma discussão maior sobre a questão da previsão de

tempo severo no Brasil, onde conduz uma descrição de parâmetros atmosféricos úteis no auxílio à identificação de ambientes favoráveis à ocorrência de tempestades convectivas severas (tempestades capazes de produzir granizo, vendavais e tornados). O cálculo de tais parâmetros atmosféricos — que destacam condições de intensa instabilidade convectiva e cisalhamento vertical do vento — fazendo-se uso, dentre outras fontes de informação, de saídas de modelos de mesoescala representa uma estratégia operacional de possível adoção no Brasil. Neste sentido, o autor explora indícios de que a ocorrência de tempestades severas no Brasil não é tão rara como tipicamente considerada, e o fato de que alguns índices de tempo severo originalmente concebidos para as latitudes médias do hemisfério norte podem: (i) ser úteis para a previsão de sistemas convectivos no Brasil, em particular para a região sul do país, e (ii) servir de base conceitual para a elaboração de índices mais adequados para as regiões tropicais do país. Estes pontos justificam uma maior atenção ao tópico de convecção severa no Brasil, tanto do ponto de vista acadêmico quanto operacional.

POLIFKE *et al.* (2015) procuraram avaliar o desempenho de indicadores atmosféricos em eventos de chuva que ocorreram na cidade do Rio de Janeiro, no período entre 1997-2012, com o objetivo de identificar quais dentre os mesmos apresentam maior confiabilidade na identificação das condições atmosféricas favoráveis à ocorrência de chuvas severas. A partir das saídas de simulações feitas com o modelo numérico *Weather Research and Forecasting* WRF, foram estabelecidos limiares para cada um dos indicadores atmosféricos analisados, assim como o cálculo da probabilidade de detecção e razão de alarme falso destes indicadores para os dias de ocorrência de chuvas severas. Com os resultados, procurou-se determinar os indicadores que apresentaram maior confiabilidade e verificou-se que a divergência do vento em 300 hPa, umidade específica em 850 hPa, energia total das tempestades severas em 500 hPa e água precipitável foram os indicadores mais eficientes na identificação das condições atmosféricas favoráveis à ocorrência de chuvas severas que ocorreram no município do Rio de Janeiro durante o período de estudo.

FRANÇA *et al.* (2016) apresentam um modelo, baseado em técnicas de redes neurais, para produzir previsões de curto prazo de instabilidade significativa para voos na área terminal do Aeroporto do Galeão, Rio de Janeiro, Brasil. Doze anos de dados foram utilizados para treinamento/validação e teste de rede neural. Os dados são originalmente de quatro fontes: (1) observações meteorológicas por hora de estações meteorológicas de superfície em cinco aeroportos distribuídos ao redor da área de estudo; (2) perfis atmosféricos coletados duas vezes por dia na estação meteorológica do Aeroporto do Galeão; (3) dados de chuva coletados de uma rede de 29 pluviômetros na área de estudo; e (4) dados de descargas atmosféricas (DA) regularmente coletados por redes nacionais de detecção. Uma investigação foi realizada sobre a capacidade de uma rede neural produzir sinais de alerta de eventos

de instabilidade significativa na área de estudo. O modelo automatizado de now-casting foi testado com resultados de cinco estatísticas categóricas, indicados entre parênteses nas previsões da primeira, segunda e terceira horas, respectivamente, a proporção correta (0,99, 0,97 e 0,94), BIAS (1,10, 1,42 e 2,31), a probabilidade de detecção (0,79, 0,78 e 0,67), relação de alarme falso (0,28, 0,45 e 0,73) e *threat score* (0,61, 0,47 e 0,25). Possíveis fontes de erro relacionadas ao procedimento de teste são apresentadas e discutidas. O teste mostrou que o modelo proposto pode identificar as relações físicas existentes no conjunto de dados, e seu desempenho é bastante encorajador para a primeira e segunda horas para previsão de eventos de instabilidade significativa na área de estudo.

DE FREITAS *et al.* (2019) fez um estudo sobre as descargas atmosféricas e sua previsão através de Inteligência Artificial (IA), neste caso, as árvores de decisão, na Região Metropolitana do Rio de Janeiro (RMRJ), com o propósito de melhor entender este fenômeno e aumentar a previsibilidade de tempestades convectivas profundas. Os resultados indicaram árvores de decisão com taxa de acerto geral acima de 80% e taxa de erro geral abaixo de 20% nos eventos. Nos estudos de caso, foi verificado que a representatividade espacial do conjunto de índices utilizado e, conseqüentemente, a previsão da AD ótima obtida é representativa além de 100 km dos limites da RMRJ definida para teste.

NASCIMENTO (2018) estudou a climatologia de descargas atmosféricas na região Sudeste do Brasil, no qual procurou conhecer as características das suas distribuições temporais e espaciais. Foi utilizado um conjunto de aproximadamente 15 anos de dados, com 48,4 milhões de descargas atmosféricas do tipo nuvem solo (41.022.466 de polaridade negativa e 7.380.663 de polaridade positiva) registradas pela rede de detecção RINDAT, no quadrilátero delimitado pelas latitudes de 14,3°S e 25,3°S e longitudes de 39,7°W e 53,2°W, durante o período de 2001 a 2016. Avaliaram-se as variabilidades (interanual, sazonal e mensal), distribuições (horária e de intensidade dos picos de corrente) e espacialização do número de descargas atmosféricas, identificando as regiões com maior densidade de descargas positivas, negativas no sudeste do Brasil. De acordo com os resultados, 84,8% das descargas atmosféricas na região Sudeste são negativas. Há uma grande variabilidade no número de descargas por ano, com o máximo em 2001 (4.580.460) e mínimo em 2011 (1.860.488). A variabilidade mensal mostra que os meses de outubro a março apresentam as maiores médias mensais e há uma queda brusca durante os meses de junho a agosto. Portanto, a maior atividade de descargas ocorre durante o verão e a menor durante o inverno. A distribuição horária apresenta uma variação bem definida durante a maior parte do ano, com os maiores valores de acumulado na parte da tarde e da noite e os menores na parte da manhã. A distribuição do pico de corrente mostra que a partir de 30 kA de intensidade as descargas atmosféricas

tendem a ser mais raras. A distribuição espacial parece ter relação com a topografia e com a presença de centros urbanos.

PRISTO *et al.* (2018) apresenta uma climatologia de eventos de chuvas intensas (ECIs) no Município do Rio de Janeiro (MRJ) a partir dos dados observados pelo Sistema Alerta Rio no período 1997-2016. Um critério objetivo de identificação de ECIs é avaliado, confrontando-se dados de chuva com informações sobre os transtornos deflagrados por tais eventos sobre a cidade. O critério estabelece que um ECI se caracteriza por um total pluviométrico diário igual ou superior a seu percentil 95, com ocorrência no mesmo dia de pelo menos um registro de chuva em 15 minutos também igual ou superior a seu percentil 95. Ademais a climatologia da precipitação no MRJ é revisitada. Os resultados indicam que 33% do total pluviométrico anual médio no MRJ (1192 mm) ocorre durante 6,6 ECIs. Os máximos (mínimos) de ECIs são observados no Sumaré e Mendanha (na Saúde e Tijuca), numa média 30 dias por ano. Em média os ECIs ocorrem com maior frequência no verão (43,7%), seguido do outono (30,9%), na primavera (18,2%) e inverno (7,2%). Nota-se em geral uma tendência de aumento no número de ECIs durante o período analisado.

PAULUCCI *et al.* (2019) estudou a variabilidade espacial e temporal dos relâmpagos nuvem-solo na região metropolitana do Rio de Janeiro, área afetada por um alto nível de descarga atmosférica a cada ano. Os autores realizaram uma análise estatística dos dados de raios retirados de um sistema de localização de raios para o período de 16 anos de 2001-2016 e analisaram as características como polaridade, correntes de pico, distribuições geográficas e diurna, variabilidade intra e interanual. Foram observados níveis extremamente elevados de atividade a partir de 258.794 eventos de descargas nuvem-solo registrados durante o período analisado: 64,3% de eventos ocorridos no verão, 20,5% de eventos ocorridos na primavera e 12,9% e 2,3% de eventos ocorridos no outono. Os níveis máximos de atividade elétrica foram observados a partir de aproximadamente 18:00 às 19:00 hora local, quando há mais energia potencial disponível para convecção. Os resultados da análise espacial revelam que a maioria dos relâmpagos observados sobre o Rio de Janeiro deriva do efeito orográfico, que estimula a formação de tempestades convectivas ao longo da parte sul da encosta.

1.2.3 Redes Neurais Artificiais no contexto da meteorologia

NOWOSAD (2001) examina duas novas abordagens para o problema de assimilação de dados atmosféricos, o uso de um filtro de Kalman estendido adaptativo e redes neurais supervisionadas do tipo *perceptrons* multicamadas (PMC). Primeiro, o filtro de Kalman estendido adaptativo é usado para assimilação de dados em três sistemas dinâmicos não-lineares: sistema de tempo discreto caótico de Hénon, equação

diferencial caótica de Lorenz e modelo computacional DYNAMO para a atmosfera. Esta abordagem não requereu que o erro de modelagem fosse estacionário e usa um filtro de Kalman linear para estimar este erro. Este método foi comparado aos métodos usando transformada de Laplace, filtro de Kalman linear e estendido. A conclusão foi que somente o filtro de Kalman estendido adaptativo pode ser usado com confiabilidade em todos os sistemas. Segundo, redes neurais de *perceptrons* multicamadas foram usadas para assimilação de dados nos mesmos sistemas dinâmicos. Nesta abordagem, a rede é treinada para "emular" a assimilação de dados por filtro de Kalman, com o objetivo de evitar recálculo da matriz de ganho a cada instante de assimilação. Um novo procedimento para treinar as redes também foi proposto, baseado numa modificação do algoritmo de retropropagação. Resultados de testes encorajadores são mostrados. Mostrou-se também que para sistemas de alta dimensão as redes treinadas podem ser computacionalmente mais rápidas que os filtros de Kalman.

HARTER (2004) aplica diferentes tipos de redes neurais no contexto de assimilação de dados em dinâmica não-linear: perceptron de múltiplas camadas, funções de base radial, Elman e Jordan. Avalia-se a eficiência das redes neurais em emular o filtro de Kalman e a possível aplicabilidade desta técnica a problemas de dimensão maior, como por exemplo, previsão de tempo operacional. Os resultados obtidos com os modelos simples (sistema de Lorenz em regime caótico e o modelo de água rasa 1D (*shallow water* 1D) mostram que as redes foram eficientes na solução do problema, sendo que o principal resultado foi a redução do espaço de busca durante o treinamento das redes. Pela forma como os exemplos de treinamento são apresentados às redes neurais, estima-se ter reduzido a dimensão do espaço de busca da ordem de 10^9 para 10^3 em problemas práticos de previsão de tempo e clima em relação ao trabalho pioneiro de NOWOSAD (2001). Apresenta-se também, resultados da aplicação da perceptron de múltiplas camadas em assimilação de dados em Clima Espacial, o que abre uma nova linha de pesquisa em Geofísica Espacial.

ALMEIDA (2009) apresenta um sistema de previsão de curtíssimo prazo (*now-casting*) para previsão de teto de nuvens e visibilidade horizontal no Aeroporto de Guarulhos - SP, desenvolvido e testado operacionalmente. O desenvolvimento do sistema foi baseado em 56 anos de dados meteorológicos da Estação Meteorológica de Superfície de Cumbica/Guarulhos. Três abordagens metodológicas foram testadas: uma com Regressão Linear Múltipla, duas com Redes Neurais Artificiais: *Probabilistic Neural Network* e *Network General Regression Neural Network*. Os resultados são apresentados e discutidos para as três abordagens. Em particular, os resultados com *Probabilistic Neural Network* indicaram um melhor desempenho, para a previsão de curtíssimo prazo (3 horas) do teto de nuvens e da visibilidade horizontal, do que as outras duas técnicas mencionadas. Em resumo, o desempenho

do sistema pode ser considerado aceitável para ser implementado operacionalmente. As respostas do sistema de previsão para as médias de três horas dos acertos e dos alarmes falsos, respectivamente, são: a) 85% e 8% para visibilidade menor que 400 m, b) 92% e 6% para nevoeiro, c) 68% e 6% para teto de nuvem abaixo de 30 m, d) 77% e 5% para teto de nuvem abaixo de 60 m.

CINTRA (2010) e CINTRA e CAMPOS VELHO (2012) são trabalhos onde um experimento numérico é realizado utilizando a assimilação de dados a partir de redes neurais artificiais no modelo global *SPEEDY* – modelo espectral com parametrização física simplificada. O processo de assimilação dos dados utiliza uma RNA do tipo perceptron de multicamadas com treinamento supervisionado. A MLP-NN (*multi-layer perceptron neural network*) é capaz de emular o *Local Ensemble Transform Kalman Filter* (LETKF). Neste experimento, o algoritmo foi treinado com supervisão para os primeiros seis meses, considerando os anos de 1982, 1983 e 1984. Um experimento de difusão para assimilação de dados realizou um ciclo para janeiro de 1985 com o modelo MLP-NN, LETKF e *SPEEDY*. As observações sintéticas da temperatura foram utilizadas. Os resultados numéricos demonstram a eficácia desta técnica de RNA na assimilação de dados atmosféricos. Os resultados com RNA são muito próximos dos resultados da assimilação de dados LETKF. As simulações mostram que a principal vantagem do uso do MLP-NN é o melhor desempenho computacional, com qualidade de campos semelhante. Considerando o tempo de CPU no processo de assimilação, a MLP-NN é 75 vezes mais eficiente do que o LETKF com as mesmas observações. Na verdade, considerando a RNA supervisionada para assimilação de dados, a questão mais relevante é a aceleração computacional para calcular a condição inicial usada em execução operacional de um modelo de estado, que acelera todo o processo de previsão numérica do tempo.

ANOCHI (2015) faz uso de técnicas de otimização da configuração de redes neurais supervisionadas, buscando determinar a arquitetura ideal da rede de forma automática – o problema de otimização é resolvido pela meta-heurística denominada algoritmo de colisão de múltiplas partículas (MPCA: *Multi-Particle Collision Algorithm*). A técnica é usada em várias aplicações: desenvolver um modelo neural de previsão climática sazonal do campo de precipitação em mesoescala; usar rede neural artificial como uma técnica não convencional para assimilação de dados; utilizar meta-heurísticas para o treinamento das redes neurais supervisionadas; entre outros. Os resultados apontam que a arquitetura ótima determinada pelos algoritmos de otimização é encontrada em um tempo menor comparado com um especialista para encontrar uma topologia aceitável. Ainda, os resultados apontam que mesmo que os modelos tenham sido aplicados em dois problemas específicos de meteorologia (assimilação de dados e previsão climática), a metodologia proposta é uma metodologia genérica e mostra-se robusta o suficiente para ser adaptada e aplicada

em qualquer problema que possa ser resolvido usando redes neurais. Uma grande vantagem do uso de redes neurais é a sua implementação em hardware.

ALMEIDA *et al.* (2020), como parte do trabalho da presente tese, apresentam um método para previsão de eventos convectivos meteorológicos na área de controle de terminais do Aeroporto Internacional do Galeão, Rio de Janeiro, Brasil, utilizando dados de aprendizado de máquina, sondagens e descargas atmosféricas detectadas remotamente de 2001 a 2016. Seis modelos de aprendizado de máquina foram treinados e validados cruzadamente por 10 anos (2001–2010) e um teste foi produzido por 6 anos (2011–2016). Os resultados mostraram que o algoritmo de *deep learning* obteve os melhores resultados para previsão de ocorrência e intensidade de tempestades com base nas seguintes estatísticas: probabilidade de detecção (0,91 e 0,85), BIAS (1,03 e 1,07), taxa de falso alarme (0,12 e 0,20) e CSI (0,81 e 0,69), respectivamente. A análise do teste de 6 anos mostrou que o modelo tem desempenho crescente para eventos de alto impacto, e esse desempenho diminui gradualmente à medida que os eventos se tornam mais fracos e mais frequentes. Os autores sugerem que o modelo proposto pode ser uma ferramenta útil para fins de gerenciamento do tráfego aéreo.

1.3 Iniciativa pioneira no Brasil

Neste contexto, é uma relevante questão estratégica a criação de um conjunto de ações específicas direcionadas à ciência e tecnologia que promovam o desenvolvimento de ações que subsidiem o aperfeiçoamento do SISCEAB para que o mesmo esteja em harmonia com as convenções e tratados internacionais. Com objetivo de modernização da Meteorologia Aeronáutica brasileira, está inserido o NOW-BR, tendo como objetivos gerais a serem atingidos: pesquisa continuada, estudo, elaboração, desenvolvimento e/ou aprimoramento de técnicas de previsão meteorológicas que visem maior segurança e otimização da navegação aérea, em suas diversas fases, dentro do território brasileiro, através do subsídio de informações meteorológicas aos aeronavegantes, aos Centros Meteorológicos, e aos órgãos de controle dos Serviço de Tráfego Aéreo. Em relação às previsões meteorológicas para a aviação destacam-se em especial as previsões de curtíssimo prazo (*nowcasting*), de minutos até cerca de 6h, que têm sido cada vez mais estudadas devido a sua importância no ambiente operacional.

Este trabalho faz parte de uma sequência de estudos relacionados à previsão do tempo para aviação que vem sendo executada pelo grupo de pesquisadores e estudantes do Laboratório de Meteorologia Aplicada da Universidade Federal do Rio de Janeiro, seguindo ALMEIDA (2009), SILVA *et al.* (2016), FRANÇA *et al.* (2016), FRANÇA *et al.* (2018), DE FREITAS *et al.* (2019), NASCIMENTO (2018),

PAULUCCI *et al.* (2019), GULTEPE *et al.* (2019), ALMEIDA *et al.* (2020) e outros trabalhos em andamento ou em processo de publicação. Este trabalho faz parte de um conjunto de iniciativas com objetivo do desenvolvimento de um sistema nacional de previsão de *nowcasting* para a aeronáutica (ver apêndice A).

1.4 Contribuição

No contexto da atividade operacional da Meteorologia Aeronáutica, faz-se necessário ferramentas de previsão dos principais fenômenos meteorológicos que possam de alguma forma impactar as atividades na área terminal dos aeródromos e em rotas aéreas. O caráter inovador do presente trabalho é apresentar a análise de diferentes técnicas para previsão de fenômenos significativos à aviação usando modelos de *machine learning*.

Em relação a eventos convectivos, embora a previsão de núcleos com grande desenvolvimento vertical tenha evoluído desde a década de 1960 - onde apenas métodos de extrapolação eram utilizados - ainda há uma limitação muito grande na previsão objetiva deste fenômeno. De forma geral, no Brasil e no mundo a previsão de eventos convectivos ainda é muito dependente do conhecimento do previsor, ou seja, é feita com um elevado grau de subjetividade. Desta forma, o modelo objetivo de previsão de ocorrência e severidade de eventos meteorológicos convectivos mostra-se como uma iniciativa bastante inovadora em relação a meteorologia no Brasil e no mundo, com potencial para ser utilizado em ambiente operacional.

Um outro aspecto inovador do presente trabalho é a proposta de métodos otimizados para assimilação de dados locais aplicados a previsões de curto prazo em modelos de previsão do tempo em escala regional. Considerando os centros operacionais, onde existe uma pequena janela para a elaboração dos boletins de previsão, métodos otimizados para assimilação de dados são de vital importância por vários aspectos: possibilidade de assimilação de uma maior quantidade de dados, mais tempo para a análise dos resultados e menor demanda por recursos computacionais.

1.5 Objetivos

O objetivo geral deste trabalho é a aplicação de modelos de aprendizado de máquina para previsão de eventos meteorológicos de impacto à aviação para a Área Terminal do Aeroporto Internacional do Rio de Janeiro. Os objetivos específicos deste trabalho podem ser divididos em dois, a saber:

1. Utilização de algoritmos de aprendizado de máquina para previsão de Eventos Meteorológicos Convectivos (EMC); e

2. Utilização de redes neurais artificiais para assimilação de dados locais no modelo regional WRF de forma mais rápida do que os métodos tradicionais e com menor custo computacional.

1.6 Estrutura do trabalho

Este texto é estruturado em 6 capítulos.

O presente capítulo apresenta o contexto no qual se insere a meteorologia no planejamento global de navegação aérea. Em seguida, são apresentadas as iniciativas brasileiras a este respeito, a contribuição e os objetivos. Por fim, a estrutura do trabalho é apresentada de forma a facilitar a leitura e compreensão do mesmo.

No capítulo 2, é apresentada uma fundamentação teórica com todos os conceitos relevantes para uma boa compreensão dos resultados e discussões apresentados.

No capítulo 3, são apresentados a área de estudo e os dados meteorológicos utilizados.

No capítulo 4, são apresentados os métodos empregados para geração dos resultados apresentados.

No capítulo 5, são apresentados os resultados e discussões.

No capítulo 6, as conclusões e sugestões de trabalhos futuros são apresentados.

Em seguida, o referencial bibliográfico utilizado é apresentado.

Por fim, apêndices com informações complementares são apresentados.

Capítulo 2

Fundamentação Teórica

2.1 Eventos meteorológicos convectivos (EMC)

Eventos Meteorológicos Convectivos (EMC) são fenômenos compostos por um conjunto de células com grande desenvolvimento vertical. Células convectivas são formações de nuvens cujo crescimento, amadurecimento e dissipação são de grande interesse entre meteorologistas uma vez que estão associados com tempestades severas (NASCIMENTO (2005)).

Em especial, o sudeste do Brasil é uma região favorável a ocorrência de fenômenos convectivos por dois fatores principais:

1. Transporte de umidade da Região Amazônica para latitudes mais altas através de correntes (jatos) em baixos níveis;
2. Forte aquecimento da região nos meses mais quentes do ano.

A ocorrência dos EMC está associada à mudança da estrutura termodinâmica da atmosfera que é monitorada diariamente nos aeroportos a partir de sondagens atmosféricas e cálculo de índices termodinâmicos.

As tempestades severas normalmente estão associadas a diversos fenômenos impactantes a aviação, tais como: precipitação, granizo, rajadas de vento, fortes correntes descendentes, rajadas de vento e grande densidade de descargas atmosféricas.

2.1.1 Índices de Instabilidade

O estudo da instabilidade atmosférica é normalmente feito a partir do cálculo de parâmetros convectivos a partir de perfis atmosféricos termodinâmicos obtidos *in situ* – sondagem atmosférica - ou obtidos por sensoriamento remoto (sensores a bordo de satélites). Os parâmetros são concebidos para ressaltar, em essência, ambientes em escala sinótica e em mesoescala com altos valores de instabilidade convectiva e cisalhamento vertical do vento.

A tabela 2.1 apresenta alguns dos mais importantes índices atmosféricos utilizados no contexto do Brasil e do mundo como parâmetros indicativos de tempestades com potencial para se tornarem tempestades severas.

Tabela 2.1: Índices de instabilidade atmosféricos normalmente utilizados como indicativos de tempestades.

Nome	Equação	Referência
Showalter	$S = T_{500} - T_{(p)500(de850hPa)}$	SHOWALTER (1953)
Índice Levantamento	$LI = T_{500} - T_{(p)500}$	GALWAY (1956)
Total Totals	$TT = T_{850} - T_{d850} - 2 * T_{500}$	MILLER (1972)
K	$K = (T_{850} + T_{d850}) - (T_{700} - T_{d700}) - T_{500}$	GEORGE (1961)
CAPE	$CAPE = g \int_{NCE}^{NE} \frac{T_{v(p)} - T_{v(a)}}{T_{v(a)}} dz$	HOUZE (2014)
CINE	$CINE = g \int_{SUP}^{NCE} \frac{T_{v(p)} - T_{v(a)}}{T_{v(a)}} dz$	HOUZE (2014)

Obs: T_i representa a temperatura no nível i , $T_v(a)$ representa a temperatura virtual do ambiente, $T_v(p)$ representa a temperatura virtual da parcela, T_{di} representa a temperatura do ponto de orvalho no nível i , SUP indica o nível da superfície, NCE indica o nível de convecção espontânea, NE indica o nível de equilíbrio (ou nível de perda de empuxo).

2.1.2 Descargas atmosféricas

A seguir são apresentados conceitos importantes sobre as descargas atmosféricas extraídos da página oficial da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT) que é uma rede de sensores especializados e centrais de processamento que permitem detectar, em tempo real, as descargas atmosféricas nuvem-solo, isto é, a maior parte das descargas que atingem o solo, em parte do território brasileiro, principalmente na região centro-sul.

Descargas atmosféricas são descargas elétricas de grande extensão (alguns quilômetros) e de grande intensidade (picos de intensidade de corrente acima de um quiloampere), que ocorrem devido ao acúmulo de cargas elétricas em regiões localizadas da atmosfera, em geral dentro de tempestades. A descarga inicia quando o campo elétrico produzido por estas cargas excede a capacidade isolante, também conhecida como rigidez dielétrica, do ar em um dado local na atmosfera, que pode ser dentro da nuvem ou próximo ao solo. Quebrada a rigidez, tem início um rápido movimento de elétrons de uma região de cargas negativas para uma região de cargas positivas. Existem diversos tipos de descargas, classificadas em função do local onde se originam e do local onde terminam.

Descargas atmosféricas podem ocorrer da nuvem para o solo, do solo para a

nuvem, dentro da nuvem, da nuvem para um ponto qualquer na atmosfera, denominados descargas no ar, ou ainda entre nuvens.

De todos os tipos de descargas, as descargas atmosféricas intra-nuvem são as mais frequentes, em parte devido ao fato de a capacidade isolante do ar diminuir com a altura em função da diminuição da densidade do ar, em parte devido às regiões de cargas opostas dentro da nuvem estarem mais próximas. Globalmente, as descargas atmosféricas representam cerca de 70% do número total de descargas. Este percentual varia com a latitude geográfica, sendo em torno de 80-90% em regiões próximas ao equador geográfico e em torno de 50-60% em regiões de médias latitudes.

As descargas nuvem-solo, também denominados raios, são as mais estudadas devido ao seu caráter destrutivo. Elas podem ser divididas em dois tipos ou polaridades, definidas em função do sinal da carga efetiva transferida da nuvem ao solo: negativas e positivas.

Os raios negativos, globalmente cerca de 90% dos raios, transferem cargas negativas (elétrons) de uma região carregada negativamente dentro da nuvem para o solo. Os raios positivos, cerca de 10%, transferem cargas positivas de uma região carregada positivamente dentro da nuvem para o solo.

Os raios duram em média em torno de um quarto de segundo, embora valores variando desde um décimo de segundo a dois segundos têm sido registrados. Durante este período, percorrem na atmosfera trajetórias com comprimentos desde alguns quilômetros até algumas dezenas de quilômetros.

A corrente elétrica, por sua vez, sofre grandes variações desde algumas centenas de ampères até centenas de quiloampères. A corrente flui em um canal com um diâmetro de uns poucos centímetros, denominado canal do relâmpago, onde a temperatura atinge valores máximos tão elevados quanto algumas dezenas de milhares de graus e a pressão valores de dezenas de atmosferas.

Embora o raio possa parecer para o olho humano uma descarga contínua, na verdade em geral ele é formado de múltiplas descargas, denominadas descargas de retorno, que se sucedem em intervalos de tempo muito curtos.

2.2 Modelos físicos

2.2.1 Conceitos

A atmosfera é um fluido cujo funcionamento (estrutura e evolução) baseia-se nos princípios de conservação de quantidade de movimento, massa e energia. Esses princípios são expressos em um sistema de equações diferenciais parciais não-lineares – as equações de Navier-Stokes (N-S), cuja solução até o momento só pode ser obtida de forma aproximada por métodos numéricos. Nesse contexto, surgiu a modelagem numérica de dinâmica da atmosfera, que leva em consideração os processos físicos (microfísica de nuvens, radiação, camada limite planetária turbulenta, modelo de superfície (interação solo-atmosfera), entre outros), dinâmicos (processos advectivos e convectivos) e químicos (deposição seca, dispersão, efeitos diretos e indiretos de aerossóis, fotólise, entre outros) da atmosfera.

A solução numérica de um sistema de equações demanda um processo de discretização das derivadas no espaço e no tempo. Os modelos podem ser globais ou cobrir apenas uma região – os chamados modelos regionais ou de área limitada.

A escolha do domínio e da resolução depende do tipo de fenômeno que se quer representar, a região que se quer abranger e da capacidade computacional para conduzir a simulação. Fenômenos cujas dimensões são menores do que a resolução do modelo são tratados como funções das variáveis resolvidas nos pontos de grade do modelo. Essa técnica é chamada de parametrização.

Os modelos numéricos atmosféricos utilizam as leis físicas básicas para prever o estado futuro desse sistema complexo que é a atmosfera. Eles devem ser capazes de reproduzir as interações que ocorrem com os outros componentes climáticos (oceano, criosfera, biosfera e litosfera) descrevendo o comportamento geral da troposfera e estratosfera a partir de condições iniciais do estado da atmosfera e de condições de contorno da superfície (terra ou oceano). São compostos por uma parte dinâmica e outra que constitui os processos físicos (as parametrizações). As equações primitivas da segunda lei de Newton, do gás ideal, da primeira lei da termodinâmica, da conservação de massa representam a dinâmica; enquanto os processos físicos como fluxos de calor e umidade da superfície para a atmosfera, radiação, turbulência, processos convectivos e de precipitação, precisam ser parametrizados a partir de variáveis calculadas pelo próprio modelo. Algumas aproximações podem ser usadas, reduzindo o número de variáveis do modelo matemático, como a aproximação hidrostática (equilíbrio entre o campo de pressão e o peso da coluna de ar) ou a aproximação geostrófica (equilíbrio entre o campo de vento horizontal e o campo de pressão).

Além da previsão de tempo e clima, os modelos meteorológicos podem ser utilizados como suporte para decisões nas mais diversas áreas, como por exemplo: ae-

ronáutica e marinha; qualidade do ar; agrometeorologia; geração de energias eólica, fotovoltaica, hidrelétrica; indústria do turismo; precificação de seguros e outros.

2.2.2 Assimilação de dados

A Previsão Numérica do Tempo (PNT) é considerado um problema de valor inicial, onde uma aproximação do estado atual da atmosfera é utilizada como entrada em um modelo numérico para simulação ou previsão da sua evolução espaço-temporal. O problema de determinação das condições iniciais de modelos de PNT é essencial e complexo e é uma área de intensa pesquisa. Vários métodos foram desenvolvidos desde a década de 1950 para solucionar este problema. DALEY (1991) e KALNAY (2003a) descrevem amplas revisões sobre estas técnicas.

Em Meteorologia, existe uma grande quantidade e variedade de fonte de dados que podem ser assimiladas para estimar de forma adequada o estado da atmosfera (exemplo: observações de superfície, sondagens atmosféricas, coleta de dados por aeronaves, coleta de dados em boias oceânicas, dados de satélite, perfiladores verticais de vento e radares meteorológicos).

De forma geral, os métodos de assimilação resultam de um processo de otimização para encontrar o campo ótimo em determinado instante, que descreve o estado atmosférico de uma forma acurada com base nas previsões do tempo anteriores e os dados coletados em um intervalo de tempo próximo ao instante da assimilação.

Por exemplo, em problemas de qualidade do ar, considerando que a dispersão de poluentes e reações químicas decorrentes das emissões ocorrem em um intervalo de poucas horas, por vezes é necessário que o modelo seja alimentado com ciclos de assimilação rápida (ou seja, atualização de dados observacionais de alta frequência).

2.2.3 Limitações dos métodos tradicionais

Segundo GULTEPE *et al.* (2019), modelos de PNT são importantes para a melhoria da previsão de curto e médio prazo para a aviação. O sucesso das previsões de curto prazo destes modelos depende em grande parte da incorporação de observações. A qualidade dos dados utilizados para identificar a condição inicial dos modelos é essencial para o sucesso das simulações numéricas. Vários modelos são iniciados a partir de dados globais com baixa resolução espacial – tipicamente entre 15 e 50 km – e temporal – normalmente com frequência de 3-6 h. Estes dados atmosféricos podem não ser precisos o suficiente para simulações de alta resolução de características locais como por exemplo o comportamento da visibilidade das superfície ou os processo turbulentos de pequena escala. A qualidade dos dados para inicialização do modelo depende principalmente nas técnicas de assimilação de dados e na qualidade das

observações. A qualidade e cobertura de observações utilizada para criação do campo inicial dos modelos também possuem influência na acurácia das condições iniciais.

2.2.4 *Weather Research and Forecasting*

O WRF é um modelo de previsão numérica de tempo de área limitada, projetado para pesquisa atmosférica e previsão operacional. As equações de Navier-Stokes são discretizadas pelo método de diferenças finitas, tem um sistema de assimilação de dados e uma arquitetura de software que suporta computação paralela. O esforço para desenvolver o WRF começou nos anos 90 e foi uma parceria do *National Center for Atmospheric Research* (NCAR), da Administração Oceânica e Atmosférica Nacional (representada pelos *National Centers for Environmental Prediction* (NCEP) e do Laboratório de Pesquisa do Sistema Terrestre), a Força Aérea dos Estados Unidos da América, o Laboratório de Pesquisa Naval, a Universidade de Oklahoma e a *Federal Aviation Administration* (FAA). A consulta ao guia do usuário do WRF e o documento da *Nota Técnica* disponível em <http://www2.mmm.ucar.edu/wrf/users/> fornece uma descrição completa do modelo WRF (SKAMAROCK *et al.* (2019)).

O modelo WRF resolve um conjunto de equações modelando o estado e a evolução da atmosfera, incluindo: (i) conservação do momento; (ii) conservação de energia termodinâmica; (iii) conservação em massa; (iv) relação geopotencial; e (v) a equação de estado. Além disso, vários processos físicos são parametrizados – por exemplo, transferência de radiação de ondas curtas e longas, modelagem de superfície, turbulência, convecção de cumulus, microfísica de nuvens e precipitação. Estes processos ocorrem em uma escala muito pequena, em um intervalo muito curto, são muito complexos, são pouco compreendidos ou muito caros em termos computacionais para serem explicitamente representados.

2.2.5 **Método de assimilação 3D-Var**

As idéias básicas da assimilação de dados pelo método variacional e, especificamente, o sistema WRF Data Assimilation (WRFDA) são discutidas em profundidade em BARKER *et al.* (2004) e BARKER *et al.* (2012).

Entre os vários métodos de assimilação de dados disponíveis, a abordagem variacional têm sido amplamente utilizada em meteorologia, especificamente o método 3D-Var, devido ao seu menor custo computacional se comparado a outros (4D-Var, 4D-Var híbrido). Na abordagem 3D-Var, é definida uma função de custo (Eq. 2.1) proporcional à diferença quadrática ponderada de dois termos, o primeiro entre a análise (x^a) e uma previsão anterior (ex: 6 horas antes) (x^b) e o segundo termo da diferença entre análise e as observações (y^o) (KALNAY (2003b)). O campo de análise é calculado pela minimização direta de tal função. É importante notar que

as matrizes de erro do modelo (B) e de observação (R) são consideradas no processo de minimização. O operador H mapeia a análise dos pontos da grade computacional para o domínio das observações, para comparação com o vetor y^o . A análise x^a é calculada minimizando a função de custo (J) expressa como:

$$J = 1/2\{[y^o - H(x)]^T R^{-1}[y^o - H(x)] + [x - x^b]^T B^{-1}[x - x^b]\} \quad (2.1)$$

onde R é a matriz de covariância dos erros dos sensores e B é a matriz de covariância dos erros de tendência. A matriz B é calculada como um produto vetorial a partir da diferença de duas execuções WRF para uma determinada condição inicial (BARKER *et al.* (2004)).

A abordagem 3D-Var consiste no processamento de informações observadas em uma janela temporal (normalmente de 1 hora antes do tempo de análise até 1 hora depois) sobre um domínio espacial.

2.2.6 Assimilação de dados por redes neurais artificiais

As redes neurais artificiais são um ramo da inteligência artificial pertencente à classe dos algoritmos de aprendizado de máquina (*machine learning*) – ver HAYKIN (1999). Uma RNA é um arranjo de várias unidades de processamento conectadas. Essas unidades são chamadas *neurônios*, nas quais as entradas ponderadas podem ou não ser combinadas com um viés para alimentar uma função de ativação não linear. A RNA pode ser classificada em dois grupos: redes neurais supervisionadas e não supervisionadas. Para o primeiro, há um conjunto de dados de referência a ser usado para identificar os pesos de conexão. Uma RNA supervisionada muito empregada é o *perceptron de múltiplas camadas* (MLP-NN).

A MLP-NN é uma rede supervisionada e geralmente contém um conjunto de camadas: a camada de entrada (uma ou mais entradas), uma ou mais camadas ocultas e a camada de saída (uma ou mais saídas). O conhecido algoritmo de retropropagação de erro é um procedimento padrão para determinar os pesos ótimos das conexões – o processo é nomeado como *a fase de treinamento ou aprendizado* (HAYKIN (1999), Seção 4.3).

Existem muitos parâmetros ou funções a serem escolhidos para configurar a MLP-NN: número de camadas ocultas, número de neurônios para cada camada oculta, o tipo de função de ativação e os parâmetros para a fase de treinamento (taxa de aprendizado e momento). Para se encontrar a melhor arquitetura para a MLP-NN, foram desenvolvidos diversos enfoques, baseados em meta-heurísticas e conhecimento. É possível abordar este problema como um problema de otimização, com a minimização do funcional (ANOCHI e CAMPOS VELHO (2014)):

$$L(\mathbf{Q}) = \textit{penalidade} \times \left[\frac{\rho_1 E_{\text{trein}}(\mathbf{Q}) + \rho_2 E_{\text{gen}}(\mathbf{Q})}{\rho_1 + \rho_2} \right] \quad (2.2)$$

$$\textit{penality} = c_1 \exp\{[\#\text{ neur\u00f4nios}]^2\} + c_2 \{\#\text{ \u00e9pocas}\} + 1 \quad (2.3)$$

onde \mathbf{Q} \u00e9 o vetor desconhecido; E_{train} e E_{gen} s\u00e3o respectivamente erros de treinamento e generaliza\u00e7\u00e3o (a diferen\u00e7a quadrada entre a sa\u00edda NN e a an\u00e1lise produzida pelo 3D-Var); finalmente *penalidade* \u00e9 uma medida da complexidade da rede neural. Portanto, a topologia ideal para o MLP-NN \u00e9 a rede neural mais simples e com melhor concord\u00e2ncia em rela\u00e7\u00e3o aos dados de refer\u00eancia (treinamento e generaliza\u00e7\u00e3o).

A solu\u00e7\u00e3o ideal \mathbf{Q}^* \u00e9 calculada minimizando o funcional 2.2 acima. O problema de otimiza\u00e7\u00e3o \u00e9 resolvido pelo m\u00e9todo MPCA descrito na pr\u00f3xima se\u00e7\u00e3o.

2.2.7 Algoritmo de colis\u00e3o multipart\u00edculas

O MPCA \u00e9 uma metaheur\u00edstica baseada no algoritmo can\u00f4nico de colis\u00e3o de part\u00edculas (PCA) desenvolvido por SACCO e DE OLIVEIRA (2005) – ver tamb\u00e9m SACCO *et al.* (2006, 2007, 2008), inspirado em um n\u00e9utron viajando dentro de um reator nuclear sob fen\u00f4menos de absor\u00e7\u00e3o e dispers\u00e3o.

H\u00e1 semelhan\u00e7as com o esquema de *Simulated Annealing* (KIRKPATRICK *et al.* (1983)). O MPCA segue a estrat\u00e9gia do PCA, mas com um novo recurso: o uso de v\u00e1rias part\u00edculas, em vez de apenas uma part\u00edcula para atuar no espa\u00e7o de pesquisa. A base te\u00f3rica do algoritmo MPCA \u00e9 detalhada por PACHECO DA LUZ *et al.* (2011). A coordena\u00e7\u00e3o entre as part\u00edculas foi poss\u00edvel atrav\u00e9s de uma estrat\u00e9gia de *black-board*, onde as melhores informa\u00e7\u00f5es de condicionamento s\u00e3o compartilhadas entre todas as part\u00edculas no processo. O MPCA \u00e9 implementado usando bibliotecas MPI em uma arquitetura de multiprocessador com mem\u00f3ria distribu\u00edda.

A codifica\u00e7\u00e3o MPCA \u00e9 pr\u00f3xima ao PCA. Supondo que o n\u00famero de chamadas para o *operador de absor\u00e7\u00e3o* seja igual ao n\u00famero de chamadas do *operador de dispers\u00e3o* e ambas iguais a N , resulta em uma complexidade $O(N^2)$, apenas opera\u00e7\u00f5es de verifica\u00e7\u00e3o nos *loops* internos. Por\u00e9m, devido ao novo loop, introduzido pela t\u00e9cnica de m\u00faltiplas part\u00edculas, o n\u00famero de opera\u00e7\u00f5es de verifica\u00e7\u00e3o pode ser aumentado para N^3 , considerando o n\u00famero de part\u00edculas igual ao n\u00famero de itera\u00e7\u00f5es. Portanto, a complexidade associada ao MPCA ser\u00e1 $O(N^3)$. A computa\u00e7\u00e3o em paralelo pode melhorar o processamento, distribuindo as tarefas entre os p -processadores. Se o n\u00famero de processadores puder ser $p = N$, sendo N o n\u00famero de part\u00edculas, o esfor\u00e7o computacional ser\u00e1 reduzido para $O(N^2)$, como o PCA padr\u00e3o.

A partir de uma solu\u00e7\u00e3o inicial candidata, o PCA modificada a solu\u00e7\u00e3o por uma

perturbação estocástica, levando à construção de uma nova solução. A nova solução é comparada à antiga (as soluções são comparadas calculando a adequação de cada uma) e a nova solução pode ou não ser aceita. Se uma nova solução for melhor que a anterior, essa nova solução será absorvida – *absorção* é um fenômeno envolvido no processo de colisão real. Se uma solução pior for encontrada, uma probabilidade é calculada para encontrar uma partícula em um local diferente – afastado da região da solução candidata – no espaço de busca, dando ao algoritmo a capacidade de escapar de um mínimo local. O último procedimento é inspirado no processo de *espalhamneto*.

PACHECO DA LUZ *et al.* (2011) apresentam uma aplicação do algoritmo MPCA para resolver dois problemas inversos - formulados como problemas de otimização. Os autores concluem que o MPCA é a alternativa para determinar soluções inversas. Hoje em dia, mesmo computadores pessoais são encontrados com arquiteturas com múltiplos núcleos de processamento, o que permite aplicar a execução de um algoritmo desenvolvido para ambientes de alto desempenho. Os resultados também demonstram a convergência do MPCA para calcular uma solução ótima dentro de uma quantidade razoável de recursos disponíveis.

ANOCHI (2015) usou o MPCA para previsão de campo de precipitação climática nas regiões Sul, Sudeste e Nordeste do Brasil. Os resultados sugerem que a arquitetura ideal determinada pelo MPCA foi encontrada em um tempo menor em comparação com o tempo que um especialista levaria para encontrar uma topologia aceitável. Outra vantagem é que a estratégia automática descarta a necessidade de um especialista em redes neurais, tornando o uso de redes neurais acessível a um público maior.

Capítulo 3

Área de Estudo e Dados

3.1 Área de Estudo

O Aeroporto Internacional do Galeão (SBGL) está localizado na cidade do Rio de Janeiro (RJ), Brasil, com latitude e a longitude de aproximadamente $22^{\circ}48'18.9''S$ e $43^{\circ}15'23,9''W$, respectivamente. A figura 3.1 apresenta a área de estudo que representa o limite aproximado de influência dos dados de sondagens do aeroporto para avaliar as condições atmosféricas em sua vizinhança. Este é o domínio utilizado para construção do modelo de previsão de eventos meteorológicos convectivos na área terminal do Rio de Janeiro. Esta área é definida pela Agência Nacional de Aviação Civil (ANAC) (https://www2.anac.gov.br/anacpedia/por_ing/tr2209.ht) como a área de controle situada geralmente na confluência de rotas e nas imediações de um ou mais aeródromos.

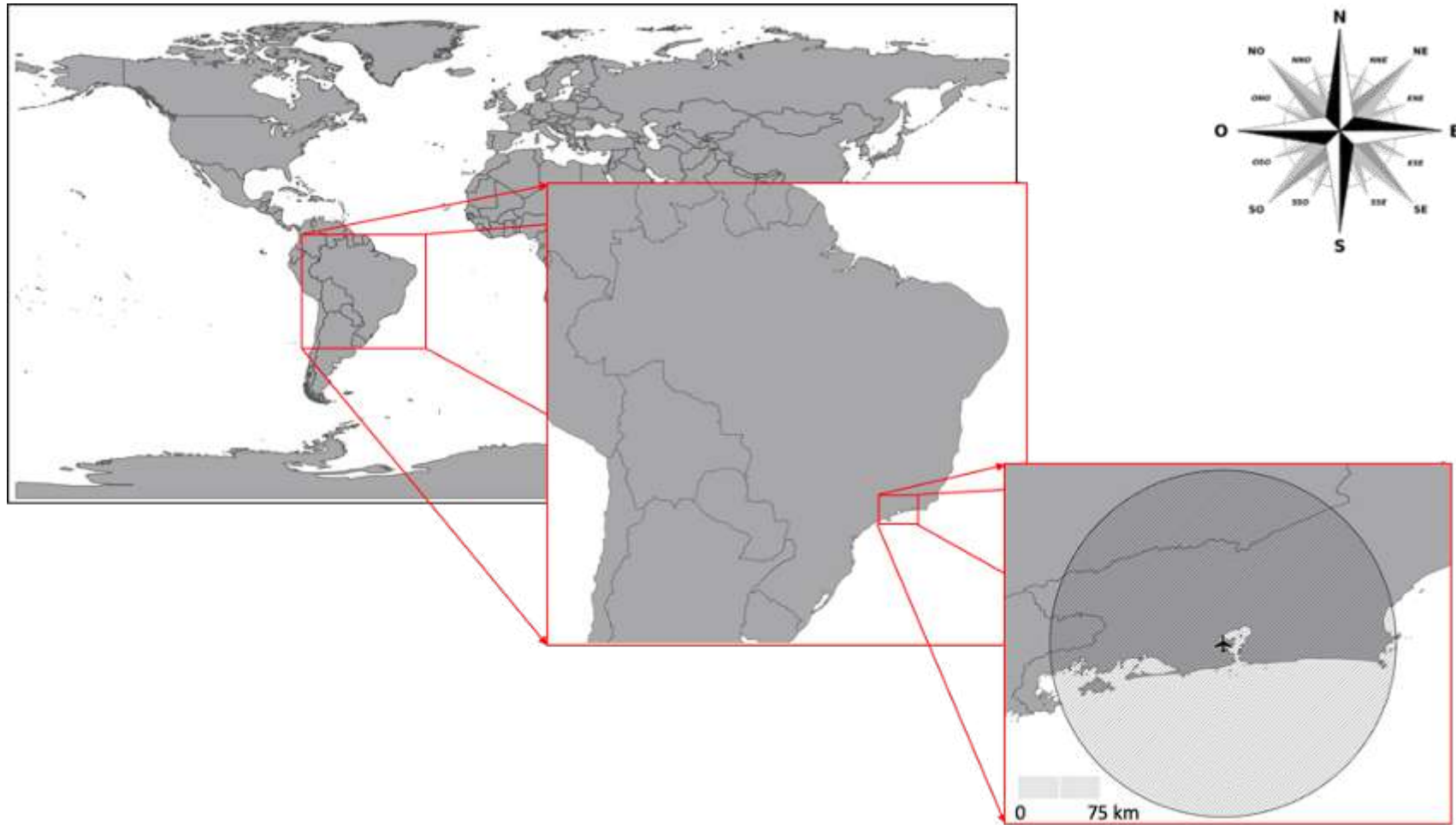


Figura 3.1: Área de estudo centrada no Aeroporto Internacional do Rio de Janeiro (Galeão) com 150 km de raio.

A figura 3.2 apresenta o domínio computacional - com duas grades aninhadas com 8 e 2,6 km de resolução horizontal, respectivamente - utilizado para a avaliação de sensibilidade do modelo regional WRF para assimilação de dados de superfície e de perfil. Ressalta-se que os dados utilizados no treinamento das redes neurais artificiais para emulação do método de assimilação de dados (3d-var) foram extraídos dos pontos de grade existentes no domínio apresentado na figura 3.2.

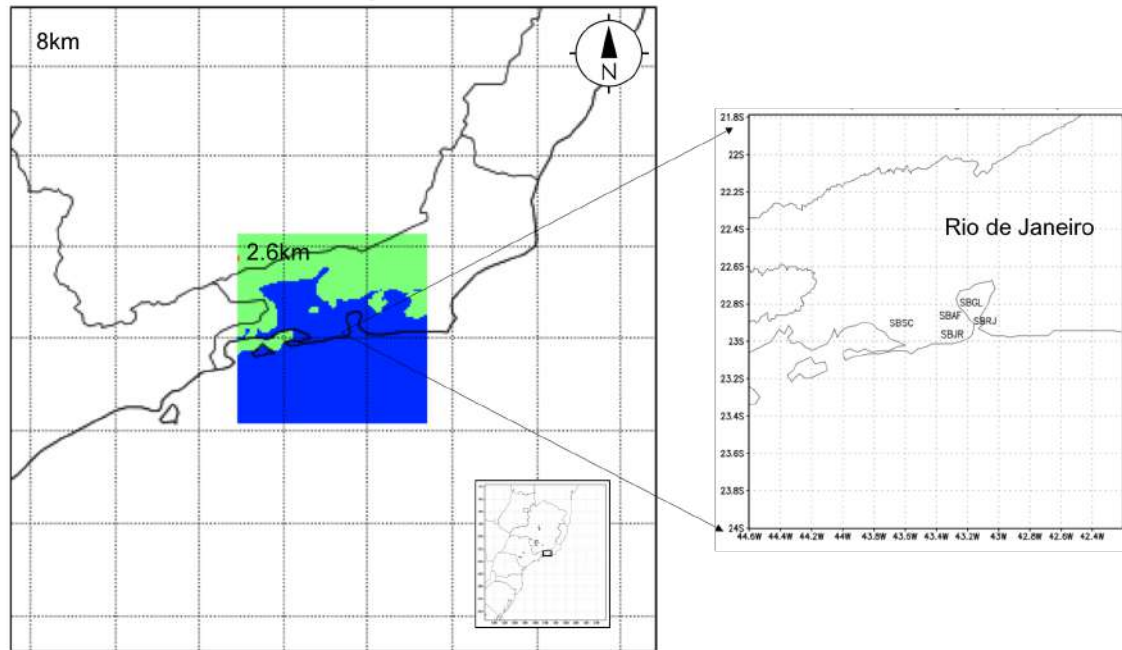


Figura 3.2: Domínio computacional e localização dos pontos de coleta de dados (aeroportos existentes na área de estudo). Os aeroportos são representados por indicadores de localidade definidos pela Organização da Aviação Civil Internacional (OACI), a saber: SBAF - Base Aérea dos Afonsos; SBSC - Base Aérea de Santa Cruz; SBJR - Aeroporto de Jacarepaguá; SBRJ - Aeroporto Santos Dumont; SBGL - Aeroporto Internacional do Rio de Janeiro - Galeão.

3.2 Dados

Para a realização do presente trabalho, dados de diversas fontes nacionais e internacionais foram utilizados. A seguir são apresentadas informações relevantes de cada uma dessas fontes e os tipos de dados disponíveis.

As variáveis meteorológicas observadas de superfície foram extraídas de diversos aeroportos na Região Metropolitana do Rio de Janeiro (ver figura ??). Nestes aeródromos são geradas mensagens horárias (denominadas METAR) a partir de dados meteorológicos coletados de uma estação meteorológica de superfície e observados por um meteorologista, tais como: vento à superfície, visibilidade, tempo presente (ex: chuva, nevoeiro, entre outros), grupo de nuvens, temperatura do ar e do ponto de orvalho, e pressão reduzida ao nível médio do mar. Quando há alguma mudança significativa fora do horário de confecção da mensagem METAR (horas inteiras), o observador meteorológico confecciona uma mensagem do tipo SPECI, que possui a mesma estrutura de dados meteorológicos da mensagem METAR. Uma descrição detalhada dos tipos de dados disponíveis em uma mensagem METAR está na página oficial da Rede Meteorológica do Comando da Aeronáutica: <https://www.redemet.aer.mil.br/?i=facilidades&p=decodificacao-metar>.

As variáveis meteorológicas observadas de altitude foram extraídas do Aeroporto Internacional do Galeão (ver figura ??), que é o único dos aeroportos na área de estudo que faz este tipo de observação regularmente em dois horários: 00 UTC (21h horário local) e 12 UTC (9h horário local). As estações meteorológicas de altitude (EMAs) têm por finalidade, através de radiossondagem, coletar dados meteorológicos nos diversos níveis da atmosfera, permitindo o conhecimento de condições atmosféricas e/ou meteorológicas dominantes em um determinado local num certo momento. As radiossondagens são realizadas por meio de um balão lançado na atmosfera, contendo gás hidrogênio (ou hélio), ao qual é presa uma sonda com sensores para medir a temperatura do ar, umidade relativa e pressão atmosférica, bem como de uma antena para registrar o deslocamento da sonda, que permite registros da direção e velocidade do vento. Os dados coletados a partir de EMAs podem ser utilizados em diversas aplicações, a saber: iniciação de modelos numéricos de previsão do tempo, previsão local/regional, previsão de curto e curtíssimo prazo (*nowcasting*), distribuição do conteúdo em vapor de água, estudos de poluição atmosférica e do clima em altitude, entre outras aplicações.

Os dados de descargas atmosféricas utilizados no presente trabalho foram extraídos a partir da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT). Esta rede é composta por uma rede de sensores especializados e centrais de processamento que permitem detectar, em tempo real, as descargas atmosféricas nuvem-solo, isto é, a maior parte das descargas que atingem o solo, em parte do

território brasileiro. Segundo informações da RINDAT, em 2012, a sua rede possuía cerca de 40 sensores instalados nas Regiões Sul, Sudeste e Centro-Oeste do Brasil. Em área de monitoramento, a RINDAT ocupa a terceira posição no mundo, sendo superada somente pelas redes existentes nos Estados Unidos e Canadá. Mais informações sobre a rede e o acompanhamento em tempo real das áreas de incidências de raios em <http://www.rindat.com.br/>.

Os dados utilizados como condições iniciais e de controle para as simulações feitas no presente trabalho foram extraídos do arquivo histórico do modelo global do *National Centers for Environmental Prediction* (NCEP) denominado *Global Forecast System* (GFS). Para simulações com dados sintéticos, foram utilizados dados disponíveis pelo NCAR na internet: <https://rda.ucar.edu/datasets/ds083.2>, com resolução horizontal de $1^\circ \times 1^\circ$ com cobertura de dados de Julho de 1999 até o presente. Para as simulações com dados observados foram utilizados dados disponíveis em: <https://rda.ucar.edu/datasets/ds084.1>, com resolução superior, de $0,25^\circ \times 0,25^\circ$, com cobertura de dados de janeiro de 2015 até o presente. Ambos os produtos incluem dados de análise e previsões, com a diferença de que o produto com maior resolução horizontal possui dados de previsão a cada 3h, enquanto que o outro possui previsões a cada 6h. Para utilização no presente trabalho, dos produtos acima descritos foram extraídos os dados de análise para todo o período de simulação contendo uma série de variáveis meteorológicas necessárias à inicialização do modelo regional WRF. Estes campos contêm tanto dados de superfície quanto dados em diversos níveis verticais da atmosfera.

A tabela 3.1 apresenta um resumo de todos os dados utilizados no presente trabalho. A coluna I (Método) informa, para cada tipo de dado, se ele foi utilizado para (I) construção do modelo de previsão de eventos meteorológicos convectivos na Área Terminal do Rio de Janeiro utilizando técnicas de *machine learning* ou (II) para testes de sensibilidade e treinamento de redes neurais artificiais para assimilação de dados.

Tabela 3.1: Dados utilizados para construção do modelo de previsão de eventos convectivos na Área Terminal do Rio de Janeiro e para o processo de assimilação de dados utilizando redes neurais artificiais.

Método	Dados	Freq.	Variáveis	Fonte	Cobertura	Período
I e II	Estação Meteorológica de Altitude (EMA)	12 hr	Perfil de temperatura, umidade, direção e intensidade do vento e índices termodinâmicos	REDEMETS ¹	Brasil	1977-2019
II	METAR	1 hr	Dados meteorológicos de superfície (temperatura, umidade relativa e vento)	REDEMETS ¹	Brasil	1977-2019
I	Descargas atmosféricas	300 ns	Localização da descarga atmosférica, Polaridade (+/-) e intensidade (V)	RINDAT ²	Centro-Sul do Brasil	2001-2019
II	<i>NCEP FNL 1 Degree Operational Model Global Tropospheric Analyses</i>	3h	Descrição completa das variáveis disponível na página oficial do produto ³	NCEP ds083.2 ³	Global	1999-presente
II	<i>NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive</i>	6h	Descrição completa das variáveis disponível na página oficial do produto ⁴	NCEP ds084.1 ⁴	Global	2015-presente

¹<https://www.redemet.aer.mil.br>

²<http://www.rindat.com.br/>

³<https://rda.ucar.edu/datasets/ds083.2/>

⁴<https://rda.ucar.edu/datasets/ds084.1/>

Capítulo 4

Aprendizado de máquina e Meteorologia Aeronáutica

4.1 Estrutura geral para *nowcasting*

Atualmente os centros de previsão do tempo são capazes de dar uma previsão com até 24h de antecedência da probabilidade de regiões de instabilidade em determinada localidade, mas não são capazes de dar alertas precisos sobre o risco de eventos significativos. A figura 4.1 apresenta um diagrama esquemático com as três fases de previsão propostas no presente trabalho: (i) previsão de mesoescala feita nos centros de previsão; (ii) a previsão de ocorrência e severidade de eventos meteorológicos convectivos a partir dos dados obtidos a partir da sondagem atmosférica; e (iii) a previsão regional com ciclos rápidos de assimilação de dados locais acelerada com a utilização de redes neurais artificiais. Nota-se no canto superior direito da figura 4.1 a referência ao trabalho de PAULUCCI *et al.* (2019) que mostra a climatologia de raios na região metropolitana do Rio de Janeiro, corroborando a relevância da previsão de eventos meteorológicos convectivos no final da tarde, período com os maiores acumulados de descargas atmosféricas.

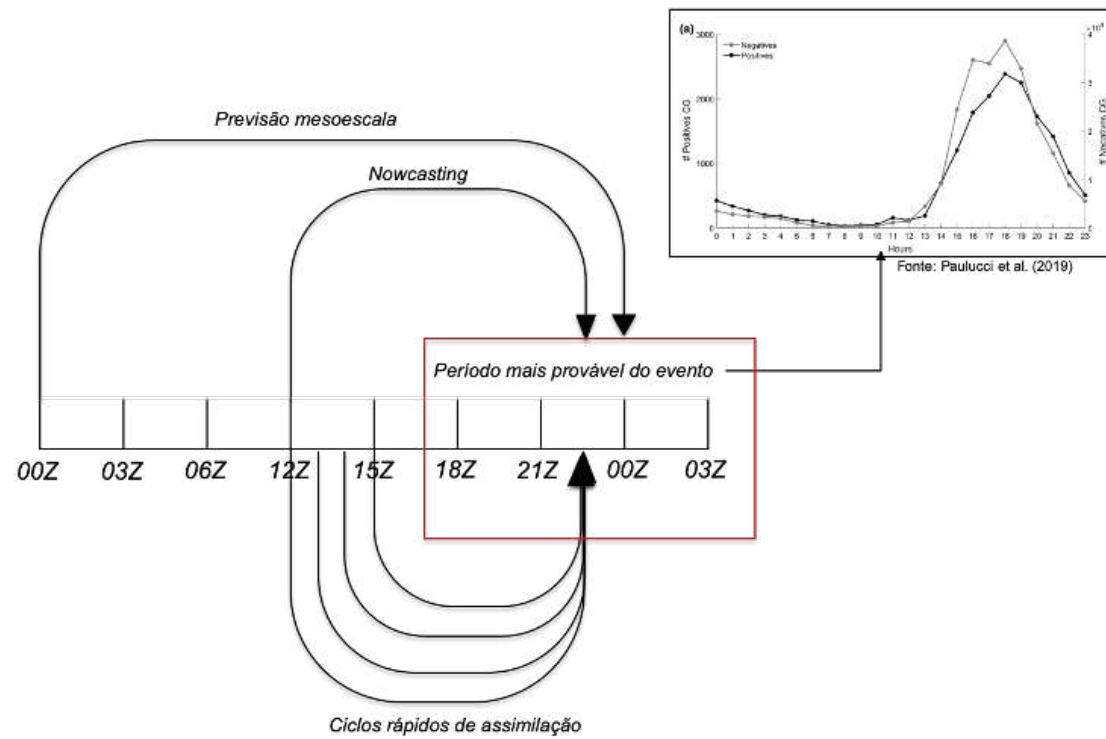


Figura 4.1: Esquema ilustrativo do processo de previsão de eventos meteorológicos convectivos.

4.2 Método I - Previsão de eventos meteorológicos convectivos para a área terminal do Rio de Janeiro

4.2.1 Fluxograma

A figura 4.2 apresenta o fluxo de atividades que foram realizadas no desenvolvimento do presente método. Os passos podem ser resumidos em: (i) coleta e organização dos dados de sondagens e descargas atmosféricas em um bando de dados espacial PostgreSQL; (ii) aplicação de um filtro espacial para seleção de um *subset* dos dados com os dados da área de estudo; (iii) um filtro temporal para agrupamento dos dados de descargas atmosféricas em acumulados por dia e redução dos dados de sondagem para apenas os registros das 12 UTC, ou seja, a sondagem da manhã no aeroporto SBGL; (iv) a definição de classe de eventos com e sem descargas atmosféricas; (v) a definição de classe de severidade para os eventos que tiveram ao menos uma descarga atmosférica; (vi) a separação dos dados disponíveis em um conjunto de treino e validação (2001 a 2010) e de teste (2011 a 2016).

Nas próximas seções é apresentado um detalhamento das atividades realizadas em cada uma dessas etapas, incluindo as dificuldades encontradas para organização e limpeza dos dados.

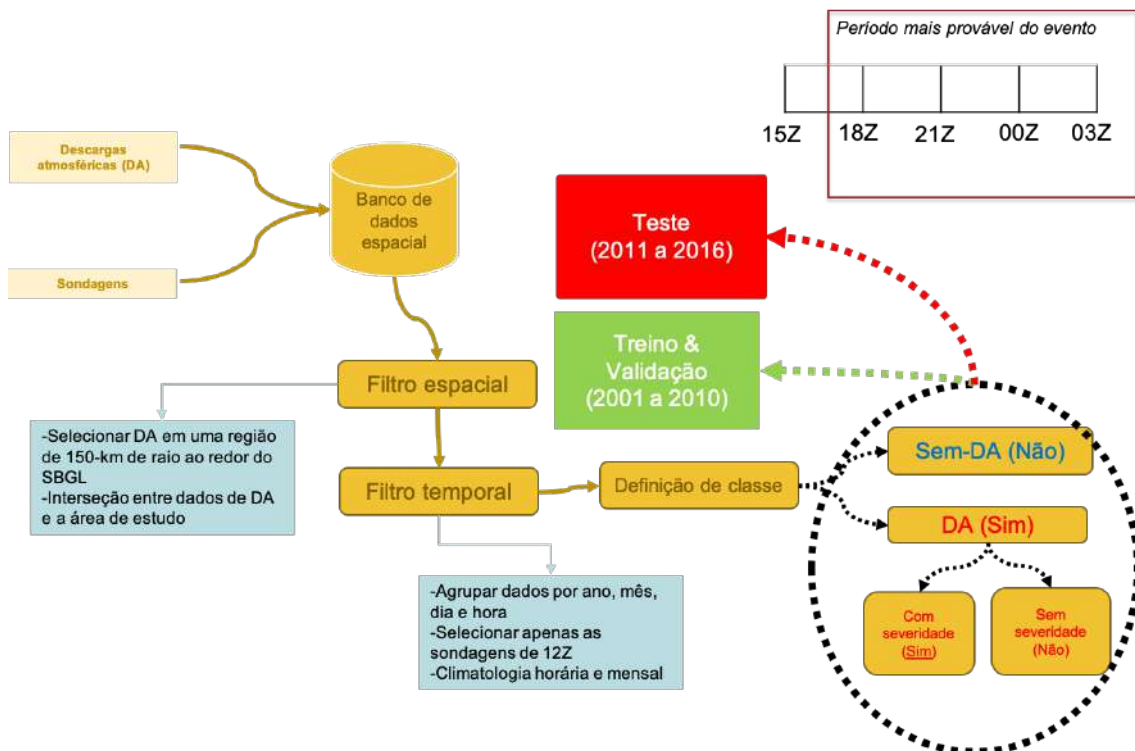


Figura 4.2: Fluxograma com a sequência de atividades que foram realizadas no desenvolvimento do Método I.

4.2.2 Organização dos dados

A rede RINDAT registra os dados de descargas atmosféricas por sua localização espacial (latitude e longitude), horário e uma série de atributos alfanuméricos, como por exemplo: sentido (ex: nuvem-solo) e intensidade. A tabela 4.1 apresenta um exemplo do formato original destes dados.

Já as sondagens do aeroporto internacional do Galeão são armazenadas em matrizes, onde as colunas são as variáveis meteorológicas observadas (ex: temperatura, direção e intensidade do vento) e as linhas representam os níveis de observação (ex: 1000 hPa, 500 hPa) - a tabela 4.2 apresenta um exemplo do formato original dos dados.

De forma a facilitar a manipulação e integração dos dados foi selecionado o banco de dados *PostgreSQL/PostGIS* como infraestrutura para organização dos dados. Importante notar que as sondagens não foram carregadas em sua forma original no banco de dados. Somente os índices atmosféricos de cada dia - derivados das variáveis extraídas nas sondagens - foram armazenados por questões de otimização de armazenamento e para reduzir a dimensão de dados de entrada do problema.

Tabela 4.1: Formato dos dados originais de descargas atmosféricas.

Date and time	Latitude	Longitude	Discrimination	Signal (kA)	Chi square	Ell. smaj	Ell. smin	Ell. an- gle	Freedom	Multi.	Sensors nb.
01/01/2014 16:41:49.484	-22.7081	-43.3476	CG-	-9	1.6	44.2	2.7	85.3	1	0	2
01/01/2014 16:49:53.648	-22.7179	-43.3394	CG-	-14	10	34.8	2.4	85.7	1	0	2
01/01/2014 16:58:53.369	-22.7212	-43.9814	CG-	-36	1.8	0.4	0.4	321.5	12	0	11
01/01/2014 17:02:58.173	-22.7234	-43.9774	CG-	-26	1.7	0.6	0.4	8.2	3	0	5
01/01/2014 17:04:19.963	-22.7024	-43.9516	CG-	-34	3	0.4	0.4	335.3	5	0	7
01/01/2014 17:07:47.805	-22.7693	-43.9333	CG-	-34	2.1	0.4	0.4	326.2	6	0	8
02/01/2014 01:32:07.021	-22.849	-43.7966	CG-	-58	1.3	1.2	0.4	302.2	6	0	9
02/01/2014 01:42:53.655	-22.9595	-43.7484	CG+	15	1.8	0.6	0.4	349.9	2	0	4
02/01/2014 01:43:15.873	-22.9958	-43.6112	CG-	-19	2.1	1.9	0.4	310.6	4	0	5
02/01/2014 01:48:21.366	-22.9548	-43.7309	CG-	-21	1.6	1.8	0.4	305.2	4	0	5
02/01/2014 01:48:21.610	-22.9566	-43.7298	CG-	-30	2.1	1.4	0.4	299.6	5	0	6

Tabela 4.2: Formato dos dados originais de sondagens

83746 SBGL Galeao Observations at 12 UTC 10 Nov 2019
 PRES HGHT TEMP DWPT RELH MIXR DRCT SKNT THTA THTE THTV
 hPa m C C % g/kg deg knot K K K
 1013.0 9 26.8 23.1 80 17.93 110 3 298.9 351.3 302.0
 1000.0 123 25.2 22.2 83 17.17 300 3 298.4 348.5 301.4
 954.0 536 21.6 21.6 100 17.36 19 4 298.7 349.5 301.8
 933.0 729 20.4 18.7 90 14.77 56 4 299.4 342.7 302.1
 925.0 804 21.6 18.7 84 14.90 70 4 301.4 345.4 304.1
 907.0 975 21.8 17.4 76 13.98 43 6 303.3 345.0 305.8
 864.0 1397 20.2 13.2 64 11.15 337 11 305.9 339.6 307.9
 850.0 1538 19.0 13.0 68 11.19 315 13 306.0 339.9 308.1

4.2.3 Seleção e integração dos dados

Após a organização dos dados no banco de dados ainda havia um volume muito grande de dados como (i) sondagens desde 1977; (ii) descargas atmosféricas de todo o Centro-Sul do Brasil; (iii) múltiplas ocorrências de descargas atmosféricas para um mesmo dia.

Inicialmente foi criada uma tabela auxiliar contendo a geometria associada a área de estudo e aplicada a função de interseção do *PostGIS* para extrair somente as descargas atmosféricas registradas na região de interesse. Em seguida, as descargas atmosféricas foram agrupadas por dia e hora, resultando na densidade horária registrada pela rede RINDAT. Em seguida, os dados diários dos índices atmosféricos e a densidade de descargas atmosféricas foram integrados, resultando em uma estrutura similar a ilustrada na tabela 4.3.

4.2.4 Higienização

A série dos dados integrados não é contínua já que existem dias sem descargas atmosféricas na região de estudo. Desta forma, nesta etapa foi feito o preenchimento da série de dados informando zero para o número de raios dos dias sem descargas atmosféricas na base de dados da RINDAT.

Além disso, existiram casos onde um ou mais índices não estavam disponíveis para determinado dia. Devido a baixa ocorrência destes casos esses dias foram simplesmente excluídos da análise para não atrapalhar a fase de treinamento.

4.2.5 Classificação dos eventos

Normalmente, o treinamento de problemas com múltiplas classes é mais complexo do que o treinamento de classes binárias. Desta forma, decidiu-se por uma estratégia de decisão hierárquica para atacar o problema de previsão de ocorrência e severidade de eventos meteorológicos na área de estudo. Primeiro a intenção é definir sobre a ocorrência ou não de um evento com raios na área de interesse. Em seguida, deseja-se saber no subconjunto de eventos com raios, se os mesmos podem ser considerados severos ou não. Não desejávamos arbitrar o limiar a partir do qual o evento seria considerado severo. Desta forma, foram testados os percentis de 50% a 90%, decidindo-se como limiar o que apresentasse a melhor relação custo-benefício entre detecção do evento e falso alarme.

Tabela 4.3: Estrutura dos dados integrados

ano	mês	dia	hora	índice 1	índice n	raios
2001	1	2	18	3.32	40	50

Após a fase de classificação, a estrutura de dados de entrada e saída (objetivo) do modelo que será desenvolvido está pronta. A tabela 4.4 apresenta as variáveis utilizadas para o treinamento dos modelos

Tabela 4.4: Variáveis de entrada, saída e suas respectivas descrições.

Variável	Tipo	Descrição
Showalter	Entrada	Valor numérico extraído da sondagem do SBGL
Índice de Levantamento	Entrada	Valor numérico extraído da sondagem do SBGL
K	Entrada	Valor numérico extraído da sondagem do SBGL
Total totals	Entrada	Valor numérico extraído da sondagem do SBGL
CAPE	Entrada	Valor numérico extraído da sondagem do SBGL
CINE	Entrada	Valor numérico extraído da sondagem do SBGL
Tempestate? (sim/não)	Saída	Valor binário representando a presença (sim) ou ausência (não) de descarga atmosférica nesse registro
SeveridadeP0.X? (sim/não)	Saída	Valor binário representando se o número de AD nesse registro é maior (sim) ou menor (não) do que o percentil 0.X do número de descargas atmosféricas em eventos de tempestade

4.2.6 Modelagem

Os modelos de *machine learning* construídos na seção 4.2 tem como objetivo criar a relação entre o estado termodinâmico da atmosfera e ocorrência e severidade dos eventos meteorológicos convectivos na área terminal do Rio de Janeiro.

Diversas plataformas estão disponíveis para implementação de modelos de *machine learning* entre os quais destacam-se o software Weka (desenvolvido pela Universidade de Waikato, EIBE FRANK e WITTEN (2016)) e a biblioteca *Tensorflow* implementada na linguagem *Python*.

No software Weka diversos algoritmos de *machine learning* estão disponíveis para serem treinados mas para o presente estudo foram selecionados somente os cinco classificadores descritos na tabela 4.5 (linha 2-6).

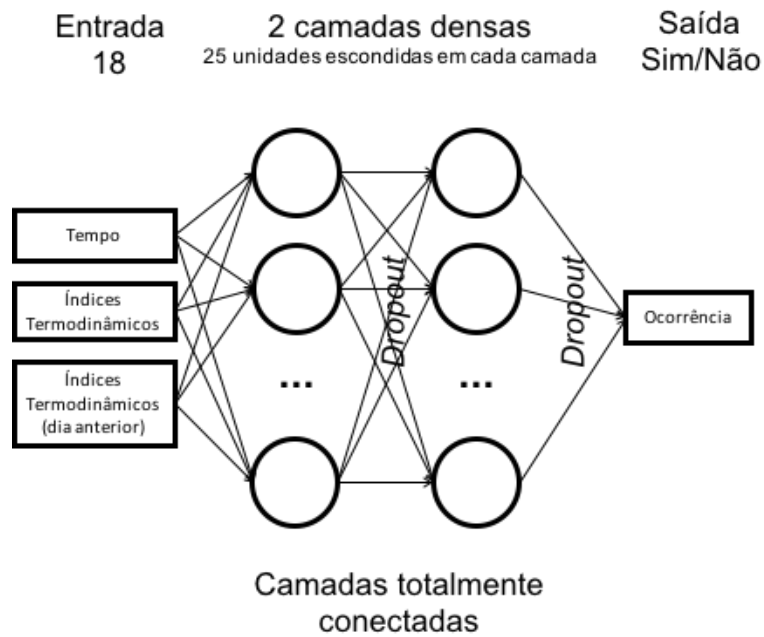
Nos últimos anos, redes neurais com múltiplos níveis de abstração (*Deep Learning*) têm sido muito utilizadas no contexto de aprendizado de máquina (LECUN *et al.* (2015); GOODFELLOW *et al.* (2016); KAMANGIR *et al.* (2020); ALMEIDA *et al.* (2020)). Desta forma, uma rotina em Python foi desenvolvida para implementação de um algoritmo utilizando a biblioteca *Tensorflow*, aqui denominado *Deep Learning Full-Connected network* (DL-FC), cuja topologia é apresentada na figura 4.3a e 4.3b, para previsão de tempestade e severidade, respectivamente.

TensorFlow é uma interface para desenvolvimento de algoritmos de aprendizado de máquina e uma implementação para execução desses algoritmos. Suas origens

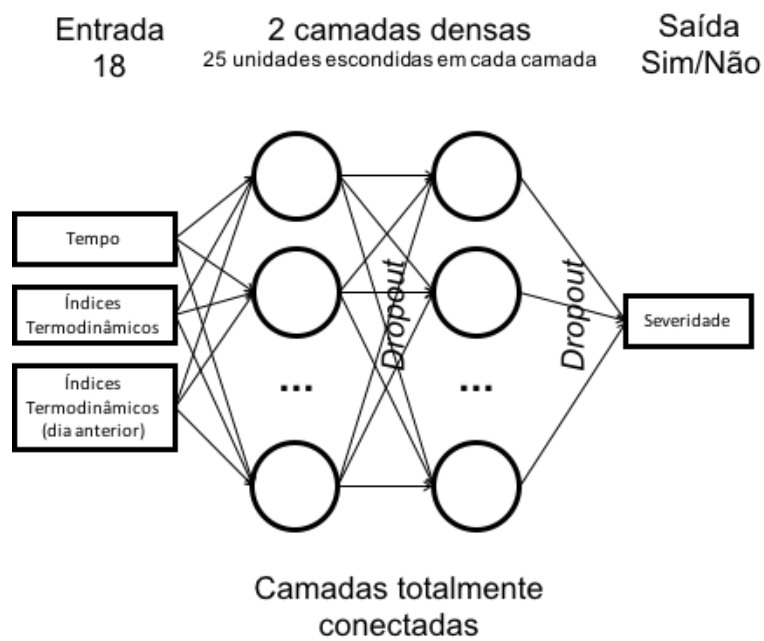
referem-se ao projeto Google Brain iniciado em 2011 para explorar o uso de redes neurais profundas em grande escala, tanto para pesquisa quanto para uso em produtos do Google. Com base na experiência de outros projetos dentro do grupo (por exemplo, DistBelief) e um entendimento mais completo das propriedades e requisitos do sistema desejáveis para treinamento e uso de redes neurais, o TensorFlow foi construído, um sistema de segunda geração para a implementação e implantação de modelos de aprendizado de máquina em grande escala. O TensorFlow usa cálculos descritos usando um modelo semelhante ao fluxo de dados e os mapeia em uma ampla variedade de plataformas de hardware diferentes, que vão desde a execução de inferência em plataformas de dispositivos móveis, como Android e iOS, até treinamento de tamanho médio usando máquinas contendo um ou mais GPUs para sistemas de treinamento em larga escala rodando em centenas de máquinas especializadas com milhares de GPUs. O sistema é flexível e pode ser usado para expressar uma ampla variedade de algoritmos, incluindo algoritmos de treinamento e inferência para modelos de rede neural profunda, e tem sido usado para conduzir pesquisas e implantar sistemas de aprendizado de máquina em produção em mais de uma dúzia de áreas de ciência da computação e outros campos, incluindo reconhecimento de fala, visão computacional, robótica, recuperação de informações, processamento de linguagem natural, extração de informações geográficas e descoberta computacional de drogas. A API TensorFlow e uma implementação de referência foram lançadas como um pacote de código aberto sob a licença Apache 2.0 em novembro de 2015 e estão disponíveis em <https://www.tensorflow.org> (ABADI *et al.* (2015)).

Tabela 4.5: Classificadores utilizados para treinamento

Classificador	Descrição	Configuração	Referência
Random forest	Cria árvores de decisão treinadas em diferentes subconjuntos de entradas.	Configuração padrão com 100 árvores usando a redução de variação padrão como critério de divisão	BREIMAN (2001)
<i>Decision tree</i> (J48)	Cria uma única árvore de decisão com base em todos as entradas disponíveis	Árvore de decisão não podada com um mínimo de dois registros por a folha.	QUINLAN (1993)
<i>Multi-layer Perceptron Classifier</i> (MLP)	Perceptron multicamada com uma camada oculta com número personalizado de unidades ocultas	Perceptron padrão com dez unidades ocultas usando a função de ativação sigmoid e otimização pela minimização da função de erro quadrático.	FRANK <i>et al.</i> (2016)
<i>Radial Basis Function Classifier</i> (RBF)	Classe que implementa uma rede com funções de base radial	Classificador de função de base radial com dez unidades ocultas treinadas minimizando o erro quadrático.	FRANK <i>et al.</i> (2016)
Voting committee	Classe para combinar classificadores	Configuração padrão usada para <i>Random Forest</i> e J48 e as versões personalizadas do <i>MLPClassifier</i> e <i>RBFClassifier</i> com dez unidades ocultas	KUNCHEVA (2002)
<i>Deep Learning fully-connected</i> (DL-FC) layers with dropout	Modelos sequenciais do Keras para a aprendizagem profunda.	Implementação em Python utilizando o framework tensorflow. Duas camadas totalmente conectadas (densas) com 25 unidades cada, regularização do tipo dropout entre as camadas, otimizador adam, função de perda entropia cruzada, ativação ReLu para camadas intermediárias e softmax para camada de saída.	ABADI <i>et al.</i> (2015)



(a)



(b)

Figura 4.3: Arquitetura para o modelo DL-FC para previsão de (a) ocorrência e (b) severidade.

4.2.7 Avaliação dos resultados

A avaliação dos resultados dos classificadores descritos na tabela 4.5 foi feita em duas etapas:

1. **Validação cruzada:** neste método, o conjunto de dados de treinamento é dividido em partes iguais (ex: 10 subconjuntos) e um processo iterativo de treinamento e validação é iniciado. O processo será executado continuamente até que todos os subconjuntos tenham sido utilizados tanto na fase de treinamento quanto na fase de validação. O resultado final a ser avaliado neste processo não é o resultado individual da validação de cada subconjunto, mas a média do resultado obtido para todos os subconjuntos. Este método é bastante utilizado porque remove um dos problemas clássicos do método de repartição arbitrária de treinamento e validação, que é normalmente dependente do processo de repartição adotado;
2. **Teste:** como o presente trabalho tem como um dos seus objetivos a operacionalização do produto, também foi realizado um processo de teste. Neste processo, um conjunto grande dos dados (2011 a 2016) foi separado do conjunto utilizado em (1) para execução de ciclo completo ininterrupto de previsões simulando o ambiente operacional. Com isso, foi possível avaliar o desempenho do modelo em casos relevantes de eventos convectivos na região de interesse.

4.3 Método II - Assimilação de dados locais usando redes neurais artificiais

4.3.1 Sensibilidade do modelo WRF ao método 3D-Var

Nesta seção, a sensibilidade do modelo WRF à assimilação de dados locais de superfície e perfil é avaliada.

Experimentos com assimilação de dados de uma semana são realizados usando o modelo WRF em 2014 e 2015. A assimilação de dados é aplicada a campos de previsão de 6 horas. Este método é realizado a cada 6 horas para variáveis de superfície (temperatura do ar, umidade relativa e direção e velocidade do vento) nos locais do aeroporto e a cada 12 horas para variáveis do ar superior (temperatura do ar, umidade relativa e direção do vento e velocidade) no local da SBGL.

O experimento foi realizado nas seguintes etapas:

1. Uma perturbação (ruído branco gaussiano) é aplicada ao campo de análise do GFS/FNL (modelo global interpolado para a grade regional) em 01/Fev às 00 UTC nos locais do aeroporto para dados de superfície e de sondagens gerando observações sintéticas;
2. Observações sintéticas são posicionadas nas coordenadas exatas onde os sensores reais estão localizados;
3. Novo campo é gerado a partir de observações sintéticas e dos dados do GFS/FNL utilizando a técnica de assimilação de dados 3D-Var existente no framework do modelo WRF;
4. O modelo WRF é integrado por 6 horas;
5. Os passos 1–4 são repetidos até 08/Fev 00 UTC com aplicação da assimilação de dados aos campos de previsão de 6 horas em superfície (a cada 6h) e sondagem (a cada 12h);
6. Os passos 1–5 são repetidos pelo mesmo período de 168h para 2014 e 2015.

4.3.2 Redes neurais para assimilação de dados sintéticos

Nesta seção, os dados sintéticos gerados na seção ?? são utilizados para treinamento de redes neurais artificiais para assimilação de dados aplicada em campos de previsão de 6 horas.

As etapas da experiência são descritas a seguir:

1. São coletados os dados gerados a partir no método da seção ??

2. O impacto das observações sintéticas na vizinhança é calculado usando o raio de 5 pontos da grade. Nos pontos de grade sob a influência de mais de uma estação, o inverso da distância é usado como fator de ponderação;
3. Para cada variável (temperatura, umidade relativa, componente u e v do vento) e ponto de grade são utilizados dados sintéticos e campo de previsão, como entrada, e dados de previsão de 6 horas com assimilação de dados usando o método 3D-var como variável objetivo, para treinamento no MPCA;
4. Um pré-processamento é executado para limpeza e normalização de dados;
5. Aleatoriamente os dados são divididos definindo 60% para treinamento, 20% de validação e 20% de generalização;
6. São treinadas uma rede para cada uma das variáveis (temperatura, umidade relativa, componente u e v do vento) no software MPCA;
7. Uma avaliação é realizada comparando os resultados do processo de assimilação de dados pelos dados 3D-Var e pela rede neural auto-configurada.

4.3.3 Redes neurais para assimilação de dados observados da REDEMET

Nesta seção, os métodos descritos nas seções ?? e 4.3.2 são mais uma vez aplicados para assimilação de dados locais usando redes neurais artificiais em campos de previsão de 6 horas. Contudo, nesta etapa são utilizados dados reais de observação de superfície e perfil extraídos dos aeroportos da área de estudo (e não sinteticamente gerados), a partir da REDEMET. Mais informações sobre os dados utilizados ver a tabela 3.1.

O experimento foi realizado nas seguintes etapas:

1. Campo é gerado a partir de observações dos aeroportos e dos dados do GFS ($0,25 \times 0,25$) utilizando a técnica de assimilação de dados 3D-Var existente no *framework* do modelo WRF para o dia 01/Fev 00 UTC;
2. O modelo WRF é integrado por 6 horas;
3. Os passos 1–2 são repetidos até 08/Fev 00 UTC com aplicação de assimilação de dados em campo de previsão de 6 horas em superfície (a cada 6h) e sondagem (a cada 12h);
4. Os passos 1–3 são repetidos pelo mesmo período de 168h para os anos de 2015 a 2019.

5. O impacto das observações sintéticas na vizinhança é calculado usando o raio de 5 pontos da grade. Nos pontos de grade sob a influência de mais de uma estação, o inverso da distância é usado como fator de ponderação. Os dados criados nesta etapa são chamados de pseudo-observações.;
6. Para cada variável (temperatura, umidade relativa, componente u e v do vento) e ponto de grade são utilizados dados observacionais e campo de previsão, como entrada, e dados de previsão de 6 horas com assimilação de dados usando o método 3D-var como variável objetivo, para treinamento das redes neurais;
7. Um pré-processamento é executado para limpeza e normalização de dados;
8. Aleatoriamente os dados entre 2015 e 2018 são divididos definindo 80% para treinamento e 20% de validação. Os dados de 2019 são guardados como conjunto de generalização;
9. São executados treinamentos utilizando redes neurais de múltiplas camadas nos softwares WEKA e também uma implementação em Python, usando a biblioteca TensorFlow. Em ambas as plataformas são utilizadas 2 camadas totalmente conectadas com 25 unidades escondidas cada. A diferença está apenas na função de ativação que foi *sigmoid* e *relu* para as implementações em pacotes de software WEKA e desenvolvidos na linguagem Python, respectivamente. Ambas as topologias possuem na camada de saída a função de ativação linear.
10. São treinadas uma rede para cada uma das variáveis (temperatura, umidade relativa, componente u e v do vento);
11. Uma avaliação é realizada comparando os resultados do processo de assimilação de dados pelos métodos 3D-Var e redes neurais.

Capítulo 5

Resultados e Discussões

Neste capítulo, são apresentados os resultados para:

- Método I: Previsão de ocorrência e severidade de EMC utilizando *machine learning*;
- Método II: sensibilidade do modelo WRF para assimilação de dados locais de superfície e perfil, assimilação de dados sintéticos utilizando redes neurais artificiais e, por fim, os resultados da assimilação de dados observados nos aeroportos obtidos a partir da REDEMET usando redes neurais artificiais.

5.1 *Nowcasting*: modelo de previsão de EMC

Esta seção apresenta os resultados do artigo de ALMEIDA *et al.* (2020) (*Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges*) publicado na revista *International Journal of Remote Sensing*, como parte do trabalho desenvolvido ao longo da Tese.

5.1.1 Treinamento e validação

Esta seção descreve os resultados para a previsão de pelo menos 9h de antecedência (ver figura 4.1) com base na sondagem feita às 12 UTC no Aeroporto Internacional do Rio de Janeiro. Eventos positivos indicam (i) a ocorrência de atividade convectiva com presença de descargas atmosféricas; e (ii) densidade de descargas atmosféricas superior a um determinado limiar, para os modelos de previsão e severidade, respectivamente. Ressalta-se que não foi arbitrado nenhum valor prévio do limiar de severidade, testando-se limiares de 50 a 90% e ficando com o melhor resultado encontrado a partir do treinamento com os algoritmos de aprendizado de máquina.

Embora amplamente utilizadas na literatura (veja o capítulo 7 de WILKS (2011) para uma revisão completa das estatísticas de verificação de modelos classificatórios), abaixo são citadas as estatísticas utilizadas para avaliação dos resultados, as informações obtidas a partir de cada uma delas e o valor ideal desejado para um modelo com bons resultados:

- POD: A *probability of detection* (POD) ou *hit rate* (H) representa o número de previsões corretas de uma determinada classe em relação ao total de vezes em que a classe foi observada. O valor ideal desta estatística é um;
- FAR: A *false alarm rate* (FAR) representa o número de previsões de uma classe que não ocorreram. O valor ideal desta estatística é zero;
- BIAS: A BIAS é uma estatística que mede a correspondência entre a previsão e a observação de determinada classe. Valores maiores (menores) que um indicam que a ocorrência da classe tende a ser superestimada (subestimada) pelo modelo. O valor ideal desta estatística é um;
- CSI: O *critical success index* (CSI) ou *threat score* (TS) representa o número de previsões corretas de determinada classe dividido pelo número total de eventos onde a classe foi prevista e/ou observada. O valor ideal desta estatística é um.

A tabela 5.1 apresenta o resultado do treinamento – para o período de 2001 a 2010 - dos algoritmos de aprendizado de máquina considerados no presente trabalho para a previsão de tempestade com aproximadamente 9h de antecedência para a área terminal do Rio de Janeiro. Destaca-se um desempenho levemente superior do algoritmo DL-FC que obteve 0,91, 0,12, 1,03 e 0,81 para as estatísticas POD, FAR, BIAS e CSI, respectivamente. De forma geral, com exceção das redes neurais MLP e RBF do Weka, todos os demais algoritmos apresentaram um POD acima de 80% e um FAR inferior a 20%, o que indica uma boa convergência dos modelos para a previsão de ocorrência de eventos com descargas atmosféricas.

As figuras 5.1 e 5.2 apresentam dois gráficos com estrutura similar: o eixo x indica o complemento da estatística FAR (1-FAR) e o eixo y indica a POD. Os

Tabela 5.1: Resultados da previsão de ocorrência de tempestade.

Algoritmo	POD	FAR	1-FAR	BIAS	CSI
J48	0.84	0.17	0.83	1.01	0.71
Random Forest	0.88	0.13	0.87	1.01	0.78
Classificador MLP	0.71	0.26	0.74	0.96	0.57
Classificador RBF	0.69	0.27	0.73	0.94	0.55
Comitê de Votação	0.84	0.15	0.85	0.99	0.73
DL-FC	0.91	0.12	0.88	1.03	0.81

resultados desejados encontram-se no canto superior direito, ou seja, na região de maximização de POD e (1-FAR) simultaneamente. Analisando-se a figura 5.1, com o resultado da validação cruzada para previsão de ocorrência de EMC, nota-se que todos os modelos possuem valores superiores a 70% para ambas as estatísticas, indicando que conseguem identificar os padrões termodinâmicos para ocorrência de descargas atmosféricas. Destaca-se o modelo DL-FC (como mencionado na análise da tabela 5.1) com resultados próximos a 90% para ambas as estatísticas. Neste ponto é importante uma discussão sobre as características dos algoritmos utilizados e as possíveis razões que justifiquem as diferenças de performance observadas. Os algoritmos utilizados são divididos em dois grandes grupos: árvores de decisão (J48 e Random Forest) e redes neurais (*Multilayer Perceptron* - MLP, *Radial Basis Function* - RBF e *Deep Learning* - DL). Em relação aos algoritmos de árvore de decisão, o aumento observado de desempenho do algoritmo *J48* para a *Random Forest* está relacionado à habilidade deste último para aprender mais padrões, uma vez que consiste em múltiplas árvores treinadas a partir de combinação de diferentes parâmetros de entrada. Em relação aos algoritmos de rede neural, o algoritmo *DL* supera os demais métodos (MLP e RBF) uma vez que, por exemplo, o POD de *DL* é maior do que o *MLP* e o *RBF* em 20% e 22%, respectivamente. A razão para as melhorias de aprendizagem no *DL* versus *MLP* e *RBF* (mostrado nas estatísticas da tabela 5.1) são principalmente o uso de (i) mais camadas e unidades ocultas; (ii) método de regularização (*dropout*); e (iii) utilização do otimizador *adam*. Além disso, os resultados da *DL* são ligeiramente melhores do que os algoritmos de árvore de decisão, como mostram os valores POD, FAR e CSI na tabela 5.1. Finalmente, o comitê de votação tem um desempenho pior do que o *Random Forest*, possivelmente penalizado pelo desempenho inferior dos métodos *MLP* e *RBF*.

A figura 5.2 apresenta os resultados de validação cruzada no conjunto de dados de treinamento (2001-2010) para severidade de tempestades usando algoritmos definidos na tabela 4.5 para os decis de 50 a 90% de todos os eventos com ao menos uma descarga atmosférica. Como pode-se notar, a maioria dos modelos estão localizados na região ideal - aqui definido como a região onde as estatísticas de POD e (1-FAR) são superiores a 70%. No entanto, há uma maior dispersão dos resultados em comparação com os resultados apresentados na figura 5.1, o que pode ser explicado pelo fato de neste modelo terem sido utilizados limiares para a severidade da tempestade dos decis de 50% a 90%, o que implica em um aumento gradual da dificuldade de modelagem dos eventos. Os modelos tiveram um desempenho inferior para classificar eventos com mais de 438,0 DA h⁻¹ (percentil de 90%) com valores CSI na maior parte abaixo de 0,6. O desempenho aumenta gradualmente para limites mais baixos, uma vez que os eventos ficam mais frequentes (maior equilíbrio de registros entre as classes). Para o limiar de 91 DA h⁻¹ (percentil 70%), o algoritmo DL-FC mostra

valores de 0,845, 0,207 e 0,69 para POD, FAR e CSI, respectivamente, tendo sido o melhor resultado observado entre todos os modelos avaliados.

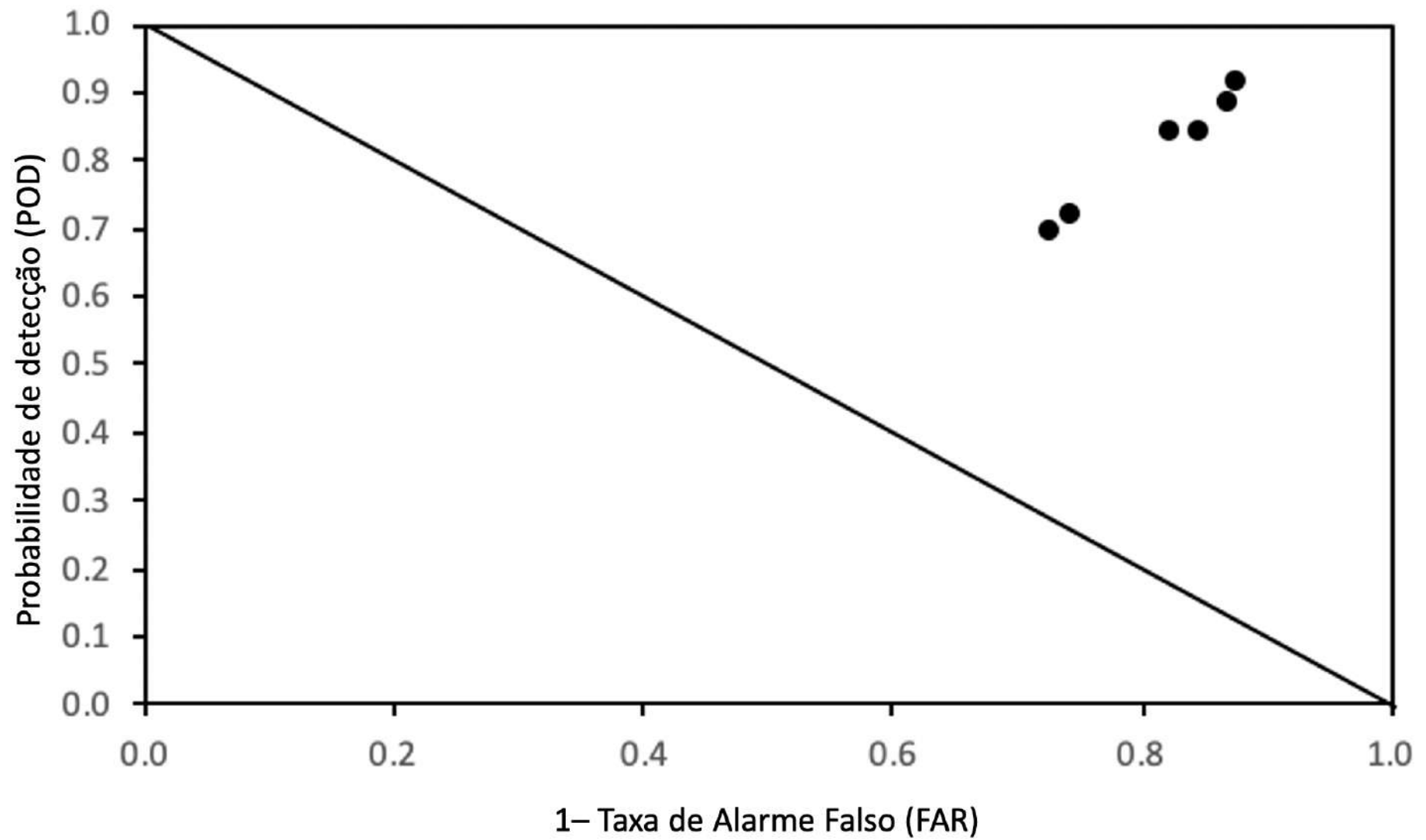


Figura 5.1: Gráfico de dispersão POD x (1-FAR) para resultados de validação cruzada no conjunto de dados de treinamento (2001-2010) para previsão de ocorrência do evento meteorológico convectivo (EMC).

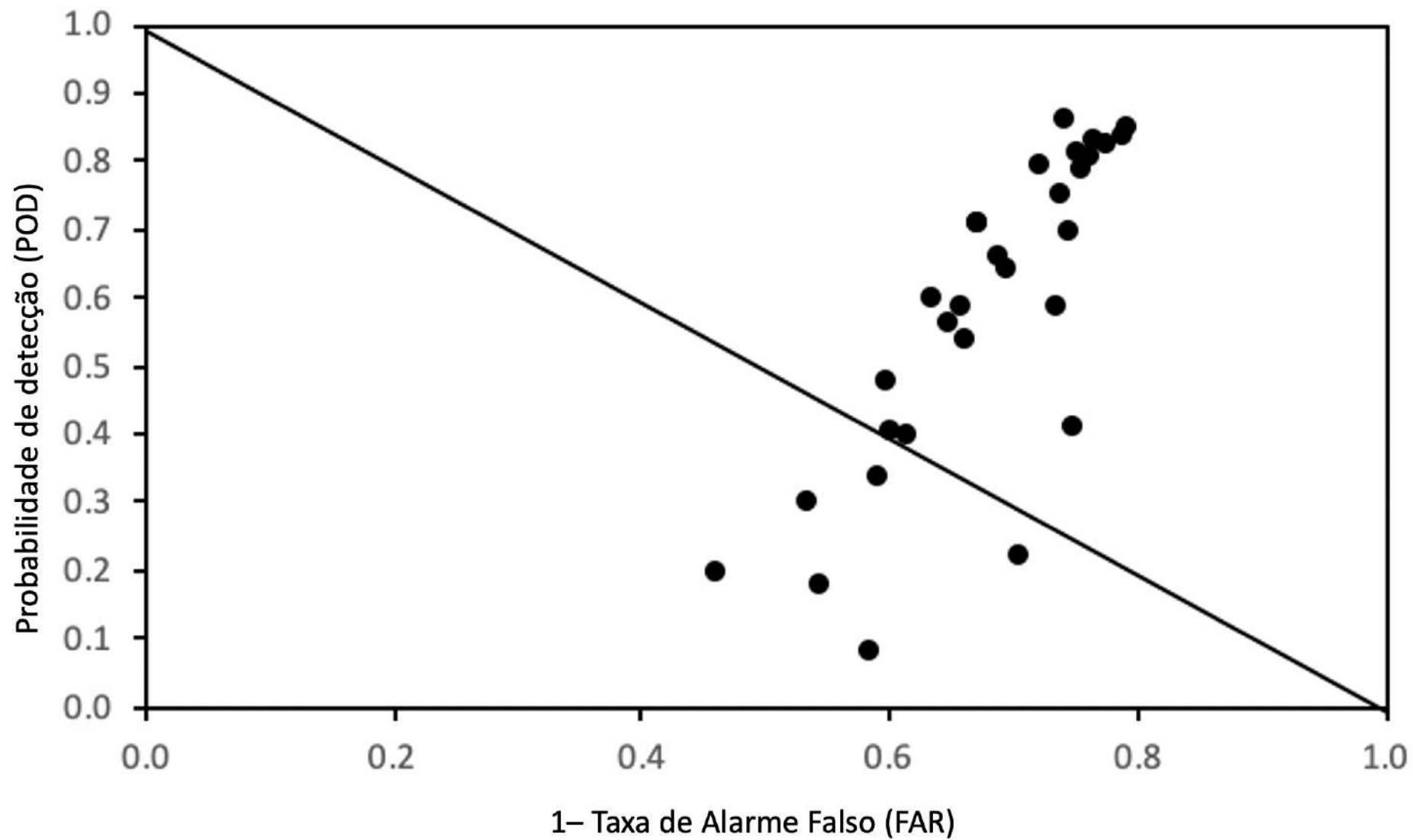


Figura 5.2: Gráfico de dispersão POD x (1-FAR) para resultados de validação cruzada no conjunto de dados de treinamento (2001-2010) para previsão de severidade do evento meteorológico convectivo (EMC).

5.1.2 Teste para o período de 2011 a 2016

Em resumo, para se fazer a previsão de um EMC são realizadas as seguintes etapas:

- Os índices termodinâmicos foram computados usando os perfis de temperatura, umidade e vento extraídos das sondagens de 12 UTC (9 h horário local), que estão regularmente disponíveis no SBGL aproximadamente 13h30 UTC (10h30 horário local);
- A ocorrência de tempestade foi definida com base na existência de pelo menos uma descarga atmosférica entre 18 h e meia-noite (horário local);
- Com base no percentil 70% da DA em eventos de tempestade (91 DA h^{-1}), o parâmetro de severidade foi definido.

A figura 5.3 apresenta o volume de acertos e erros para todos os eventos no conjunto de dados de teste (2011 a 2016), usando o algoritmo DL-FC, que foi selecionado aqui com base na análise do POD máximo e FAR mínimo para previsão de ocorrência e severidade de EMC. Vale ressaltar que, à medida que os eventos se tornam mais intensos e raros, o modelo de previsão se torna mais preciso na previsão de EMC. No entanto, eventos mais fracos tendem a ser mais frequentes, levando a maiores erros de previsão. Eventos mais fracos (menos de 1.000 DA), possuem desempenho inferior para previsão de ocorrência de tempestade. À medida que os eventos se tornam mais fortes (com maior número de descargas atmosféricas), progressivamente, houve um aumento no desempenho. Eventos com mais de 3.000 DA são quase sempre previstos (apenas 3 erros) no conjunto de testes.

Para a completude deste trabalho, a tabela 5.2 apresenta os eventos mais intensos no conjunto de dados de teste – ordenados em ordem decrescente pelo acumulado diário de descargas atmosféricas entre 18h e meia-noite (horário local). É importante notar, para os eventos selecionados, que o acumulado de descargas atmosféricas variou entre 4.966 e 15.122; ou seja, foram eventos relevantes ocorridos na área do estudo. Considerando este subconjunto de eventos, o modelo ideal (DL-FC) foi capaz de prever corretamente todos os EMC. Isto é importante porque indica que o previsor no centro meteorológico teria sido capaz de alertar na manhã de cada dia a todas as aeronaves que fossem voar para o Rio de Janeiro no final do dia que havia risco de núcleos convectivos e possíveis desvios de rota na aproximação dos aeroportos. Sem dúvida, esta informação é de vital importância para a segurança dos vôos e para o gerenciamento do tráfego aéreo.

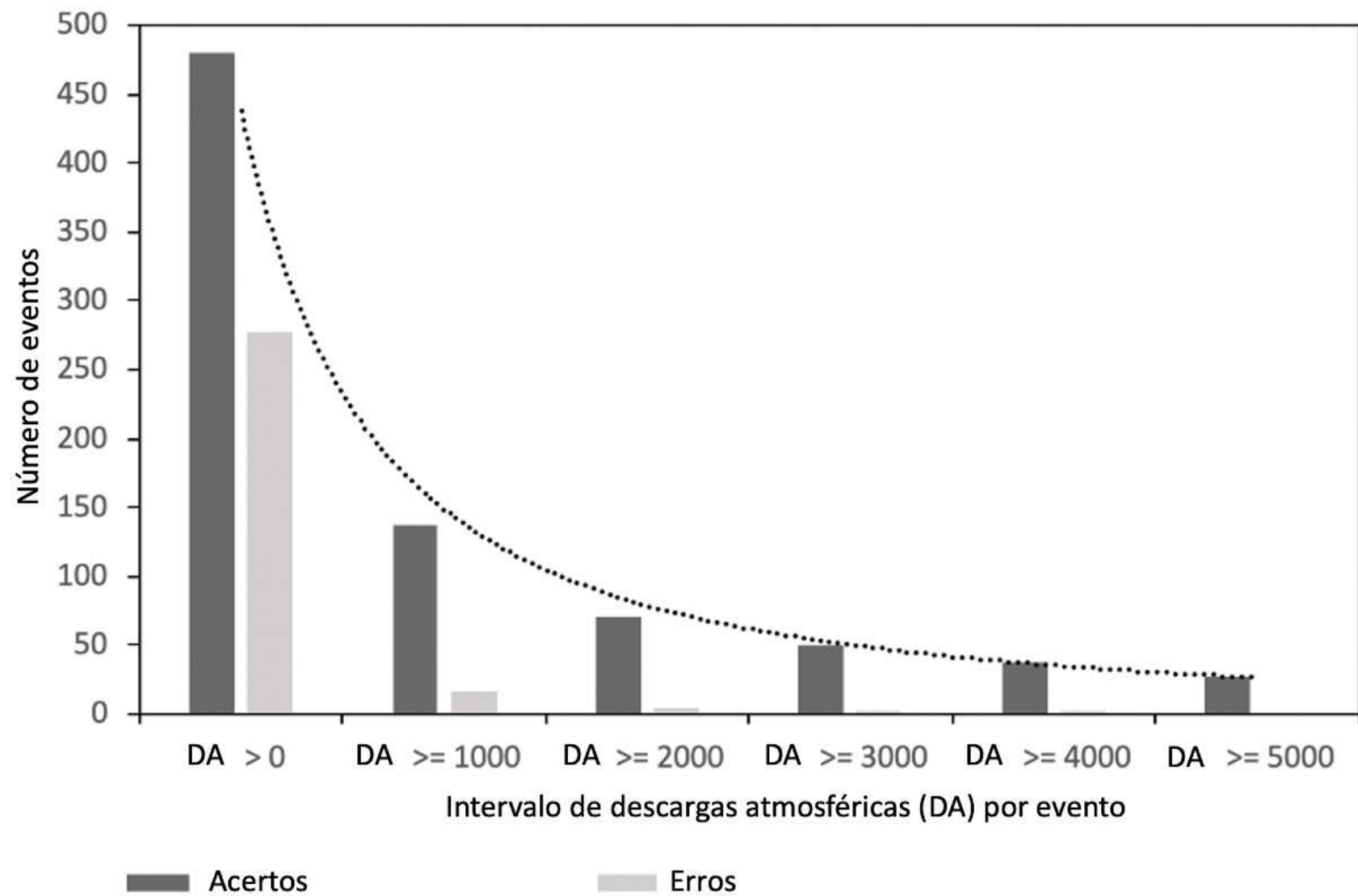


Figura 5.3: Análise do conjunto de teste (2011 a 2016) considerando diferentes limiares de descargas atmosféricas nos eventos meteorológicos convectivos.

Tabela 5.2: Resultados da previsão de tempestade para os 30 principais eventos do conjunto de dados de teste.

Ano	Mês	Dia	Previsão	DA
2016	2	20	Sim	15122
2016	2	17	Sim	12973
2014	1	16	Sim	11075
2012	12	30	Sim	10988
2013	3	5	Sim	10478
2016	3	6	Sim	10053
2016	3	23	Sim	9750
2012	12	27	Sim	9026
2016	12	12	Sim	8936
2016	12	22	Sim	8437
2016	12	8	Sim	8129
2012	12	29	Sim	7922
2015	11	8	Sim	7720
2016	2	22	Sim	7518
2016	2	15	Sim	7359
2014	2	28	Sim	6956
2016	2	28	Sim	6696
2012	12	13	Sim	6525
2016	12	31	Sim	6198
2013	3	4	Sim	5867
2013	12	10	Sim	5809
2016	2	25	Sim	5753
2013	3	9	Sim	5681
2016	2	27	Sim	5568
2013	2	26	Sim	5541
2012	12	12	Sim	5473
2012	12	17	Sim	5436
2013	1	15	Sim	5358
2013	1	14	Sim	4980
2012	12	14	Sim	4966

5.1.3 Caracterização de EMC mais intensos

Eventos convectivos estão normalmente associados a diversos fenômenos meteorológicos, tais como: rajadas de vento, granizo, chuva forte, cisalhamento de vento, turbulência e descargas atmosféricas.

Conforme descrito no capítulo de métodos, a classificação de EMC utilizada no presente trabalho foi feita apenas com base na densidade de descargas atmosféricas na região de estudo. Desta forma, uma caracterização destes eventos é necessária para identificar se os eventos convectivos que foram definidos estão de fato associados a fenômenos meteorológicos de impacto à sociedade e, principalmente, às atividades

aéreas.

Uma caracterização sinótica dos eventos definidos na tabela 5.2 foi possível a partir da análise de diversos tipos de dados, tais como: dados de superfície, imagens de satélite, cartas sinóticas e sondagens atmosféricas. Os resultados desta caracterização podem ser resumidos como:

- Os índices termodinâmicos obtidos a partir das sondagens atmosféricas apontam que os eventos, em sua grande maioria, possuíam índices termodinâmicos (K, TT, LI, CAPE e CINE) em faixas críticas definidas na literatura (ex: NASCIMENTO (2005)) que apontam riscos de eventos meteorológicos extremos. Em alguns casos foram observados perfis termodinâmicos com inibição de convecção pela manhã, contribuindo para a sua intensificação no final da tarde;
- Em termos sinóticos, os eventos estavam associados a condições típicas de verão, ou seja, condição dinâmica de convergência em baixos níveis associada a presença de cavados, sistemas frontais, zonas de convergência, ou forçantes locais;
- Os eventos em sua maioria estiveram associados a precipitação de moderada a muito forte na cidade do Rio de Janeiro e com relato de rajadas de vento nas estações localizadas nos aeroportos.

Desta forma, é possível concluir que a maior parte dos EMC estudados – definidos a partir da densidade de descargas atmosféricas – está associado com a ocorrência de precipitação (moderada ou intensa) e rajadas de vento, fenômenos meteorológicos de grande impacto a sociedade e a aviação. Vale ressaltar que a análise feita no trabalho foi restrita aos casos da tabela 5.2 e precisa ser expandida para os demais casos de forma a se criar um modelo conceitual da formação, desenvolvimento e dissipação dos EMC. Outras variáveis prognósticas também podem ser úteis neste modelo conceitual, tais como: temperatura da superfície do mar, convergência de umidade, divergência em altos níveis, além de simulações numéricas dos eventos.

5.2 Assimilação de dados usando redes neurais

5.2.1 Sensibilidade do modelo WRF à assimilação (3D-Var)

Nesta seção, a abordagem de assimilação de dados 3D-Var implementada no WRFDA é empregada para avaliar o processo de assimilação aplicado à área do aeroporto do Rio de Janeiro. Qualquer método de assimilação de dados consiste basicamente em combinar campos de observação e previsão para gerar a melhor aproximação do verdadeiro estado da condição inicial para um sistema dinâmico. As observações não representam uma fidelidade completa da “realidade”, mas a estimativa mais próxima do estado verdadeiro sobreposto com algum ruído devido aos erros de medição do sensor. Portanto, uma estratégia padrão para avaliar um algoritmo de assimilação de dados é adicionar algum ruído na localização do sensor (dinâmica verdadeira), gerando uma observação sintética. Se o algoritmo de assimilação de dados funcionar, o resultado esperado é uma redução significativa do ruído nos dados previsão de 6 horas, obtendo uma dinâmica mais próxima do campo original (antes da aplicação da perturbação), ou seja, nosso verdadeiro estado do sistema dinâmico.

Uma atividade de treinamento em assimilação de dados foi desenvolvida pela *University Corporation for Atmospheric Research* (UCAR) e pelo *National Center for Atmospheric Research* (NCAR)¹. A estratégia citada para avaliar esquemas de assimilação de dados é realizada na primeira seção do *Data Assimilation Research Testbed* (DART)², onde um experimento numérico de simulação semelhante é descrito para sistemas 1D. A figura 5.4 é uma adaptação de uma figura disponível no slide 45, onde a estratégia mencionada para o teste de assimilação de dados é representada graficamente. Em suma, o processo de assimilação dos dados reduz o ruído adicionado na previsão “anterior” (*prior*) e torna o campo “posterior” mais próximo da “dinâmica real”, ou seja, mais próximo do estado real do sistema.

¹https://dart.ucar.edu/pages/dart_lab.html

²https://dart.ucar.edu/DART_LAB/presentation/DART_LAB_Section01.pdf

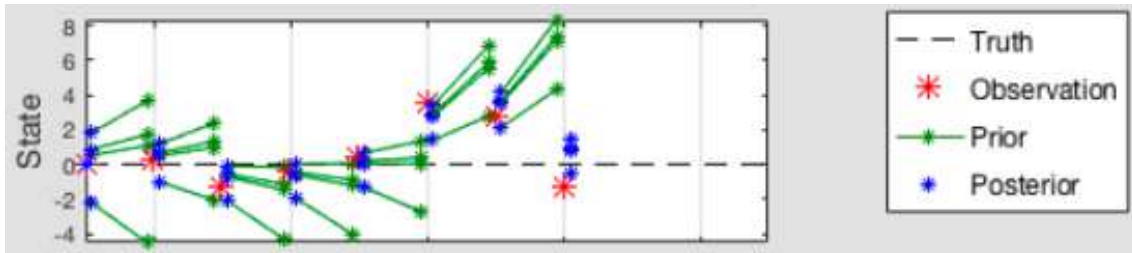


Figura 5.4: Metodologia de teste de sistema de assimilação de dados. Adaptado de um curto do *Data Assimilation Research Testbed* disponível em https://dart.ucar.edu/DART_LAB/presentation/DART_LAB_Section01.pdf, slide 45.

Um outro ponto importante a ser discutido é se o uso de domínios únicos ou aninhados possui algum impacto no resultado no experimento realizado para avaliação da sensibilidade de assimilação de dados sintéticos usando o método 3D-Var no modelo WRF para a área terminal de voo do Rio de Janeiro. Desta forma, um teste de sensibilidade foi realizado para avaliar o impacto das condições iniciais e de contorno no domínio de estudo. Portanto, um experimento foi realizado usando três grades aninhadas para avaliar se o processo de assimilação seria diferente dos resultados do experimento de grade única. A figura 5.5 mostra o resultado do ciclo de assimilação de 6 h para os experimentos de domínio único (fig. 5.5a) e de domínio aninhado (fig. 5.5b), em cada aeroporto dentro da área de estudo - a propósito, estamos adotando o mesmo padrão de plotagem mostrado na figura 5.4. A comparação foi realizada no domínio de maior resolução horizontal (2 km). A figura 5.5 mostra que o efeito do processo de assimilação é muito semelhante nos dois experimentos, ou seja, tem o efeito de remover parcialmente o ruído branco da previsão de 6 horas, tornando os resultados próximos do experimento controle, o estado “verdadeiro” do sistema. Os resultados do experimento indicam um uso seguro de um único domínio de nossos experimentos numéricos.

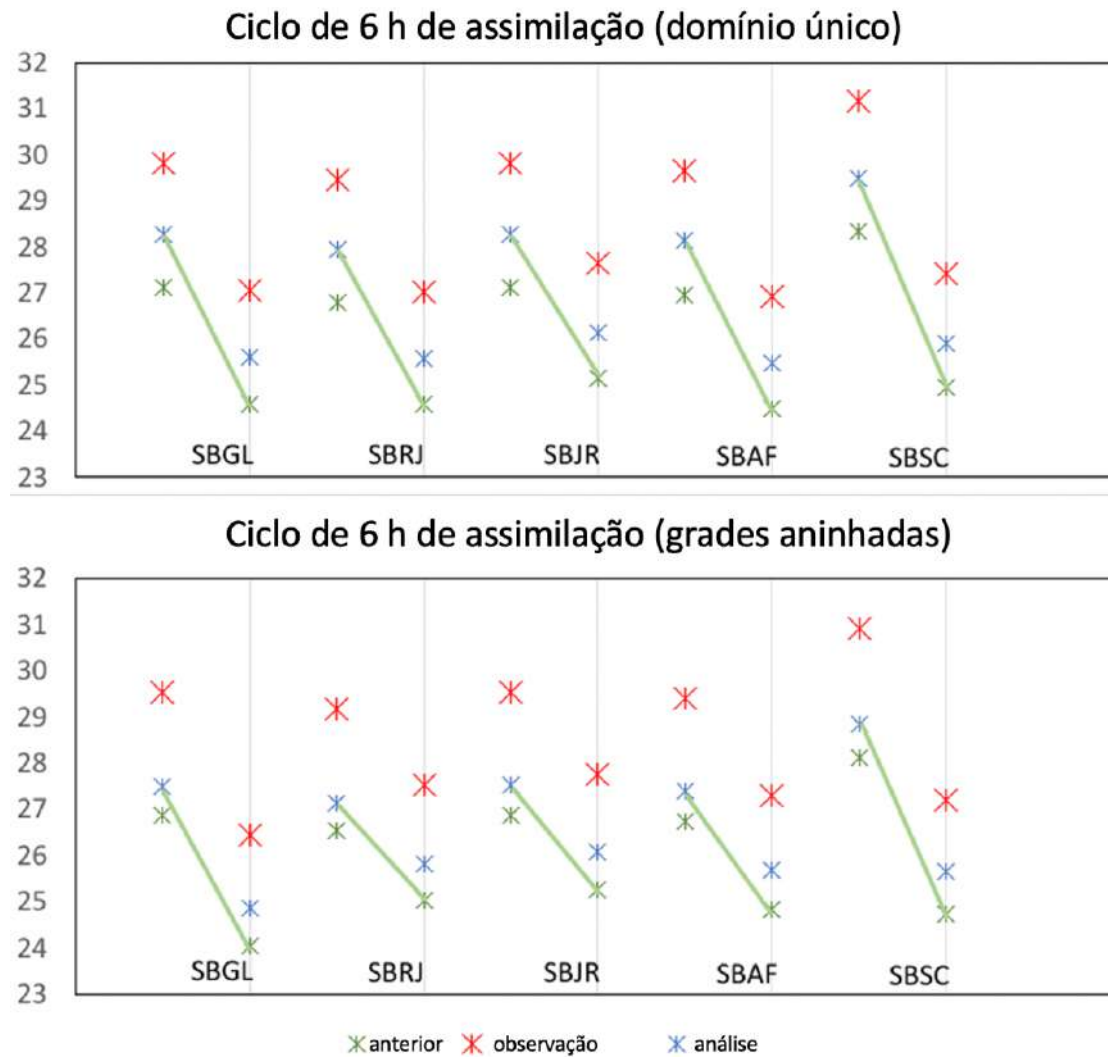


Figura 5.5: Teste de sensibilidade de domínios de grade única e aninhada realizado com processo de assimilação de dados aplicado à previsões de 6 horas nas estações referentes aos aeroportos na Região Metropolitana do Rio de Janeiro.

As figuras 5.6–5.9 apresentam os resultados do processo de assimilação de dados de superfície e perfil no domínio de estudo aplicado a campos de previsão de 6 horas. Todas as figuras têm a seguinte estrutura: (a) previsão de 6 horas do campo inicial sem assimilação; (b) previsão de 6 horas a partir de um campo inicial com assimilação; (c) resultado da assimilação de dados da superfície e do ar superior usando o método 3D-Var aplicada a campos de previsão de 6 horas; e (d) diferença do campo de previsão de 6 horas com assimilação usando método 3D-Var e o campo de previsão 6 horas a partir de uma condição inicial sem assimilação de dados (controle). Todas as figuras mostram um comportamento semelhante, com pequenos erros entre o campo de previsão de 6 horas com assimilação de dados 3D-Var e o campo controle, como esperado, mostrando que o processo de assimilação removeu a maior parte da perturbação do ruído branco existente nos dados da observação.

A figura 5.6a mostra que no dia 1º de fevereiro de 2014 06 UTC o processo de assimilação dos dados gerou um campo (fig. 5.6c) com maiores valores de temperatura no entorno das localidades das estações, principalmente próximo ao SBAF. Considerando que a magnitude da perturbação do ruído branco era um número real entre 0 e 3 K, as diferenças entre o campo de previsão de 6 horas com assimilação de dados 3D-Var e a previsão de 6 horas a partir de condição sem assimilação de dados (fig. 5.6d) mostram que o processo de assimilação de dados removeu 80-90 % do ruído, e a magnitude resultante da inovação não era maior que $|0,6|$.

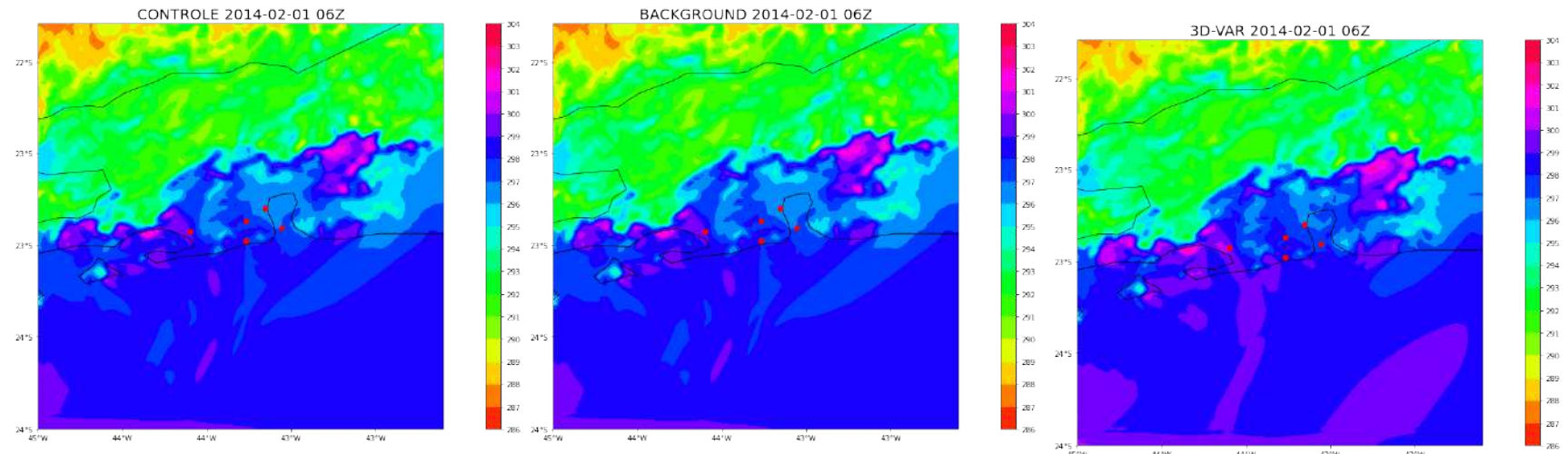
Em uma análise semelhante, a figura 5.7 mostra o impacto do processo de assimilação de dados para o campo de umidade relativa de 2 m. Diferentemente do que foi observado para o campo de temperatura na figura 5.6, a matriz de inovação resultante gerada pelo 3D-Var (fig. 5.7d) atingiu valores maiores, com regiões com inovações de até 50 % da perturbação inicial.

A velocidade do vento (fig. 5.8) mostra que o processo de assimilação de dados gerou uma inovação (fig. 5.8d) de até 60 % da perturbação de ruído branco aplicada ao campo de controle - que teve uma magnitude de até 1 m s^{-1} . As dificuldades envolvidas na análise de um campo vetorial (a variável do vento) em comparação com as variáveis escalares (por exemplo, temperatura do ar e umidade relativa) são notáveis. Conforme mostrado na figura 5d, pequenas perturbações em campos vetoriais parecem causar perturbação em quase todo o domínio, enquanto a inovação em campos escalares (fig. 5.6d e 5.7d) são mais restritas aos arredores das estações - onde a perturbação foi aplicada. A análise do perfil de velocidade do vento (fig. 5.9) mostra que a diferença entre a previsão de 6 horas com assimilação de dados e a previsão de 6 horas a partir de uma condição inicial sem assimilação de dados (fig. 5.9d) foi próxima de zero a partir de 850 hPa para cima, com pequeno aumento na camada entre 800 e 650 hPa. Maior impacto positivo é observado próximo à superfície, possivelmente relacionado à contribuição da assimilação dos dados de

superfície. Diferentemente do campo de superfície, o impacto no perfil vertical não excede 20 % da perturbação do ruído branco. O pequeno efeito do processo de assimilação pode estar relacionado à forma como 3D-Var calcula o impacto dos dados da estação em seus arredores, ou seja, o processo de interpolação da grade de observação para a grade do modelo pelo operador H (ver eq. (2.1)) após o cálculo da inovação.

Em resumo, todas as figuras apresentam comportamento semelhante, com pequenos erros (em relação à magnitude da perturbação) entre o campo com assimilação de dados aplicado à previsão de 6 horas e a rodada controle, como esperado, mostrando que o processo de assimilação removeu a maior parte da perturbação do ruído branco existente nos dados de observação.

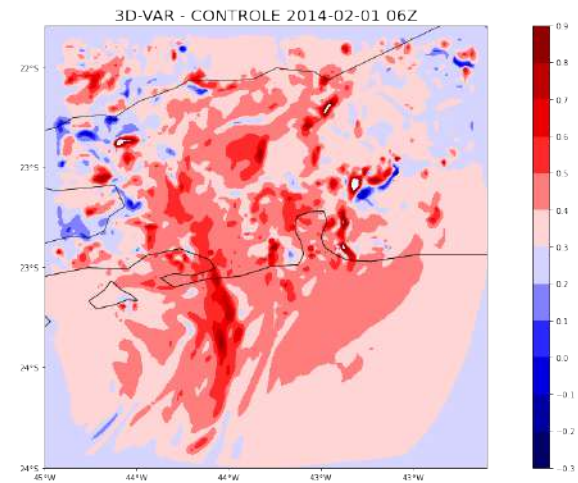
Portanto, os resultados do processo de assimilação de dados utilizando dados sintéticos (temperatura do ar, umidade relativa e velocidade do vento), mostram que o método 3D-Var no sistema WRFDA é capaz de realizar uma boa estimativa do campo de controle, aqui representando o estado "verdadeiro" do sistema dinâmico. As implicações de tais resultados são importantes, uma vez que afirma implicitamente que apenas usando dados locais na inicialização do modelo atmosférico regional as previsões meteorológicas tendem a ser melhoradas. Atualmente, muitos dados observacionais não são usados porque não são considerados pelo sistema de assimilação do modelo global ou porque não fazem parte do sistema de observação global.



(a) Previsão de 6h a partir de condição inicial sem assimiação e pontos de observação

(b) Previsão de 6h a partir de condição inicial com assimiação

(c) Previsão de 6h com assimiação 3D-Var



(d) Previsão de 6h com assimiação 3D-Var - Previsão de 6h a partir de condição inicial sem assimiação

Figura 5.6: Campo de temperatura do ar para (a) previsão de 6 horas a partir de condição inicial sem assimiação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimiação, (c) previsão de 6 horas com assimiação 3D-Var e (d) diferença (c) e (a), para 1° de fevereiro de 2014 06 UTC.

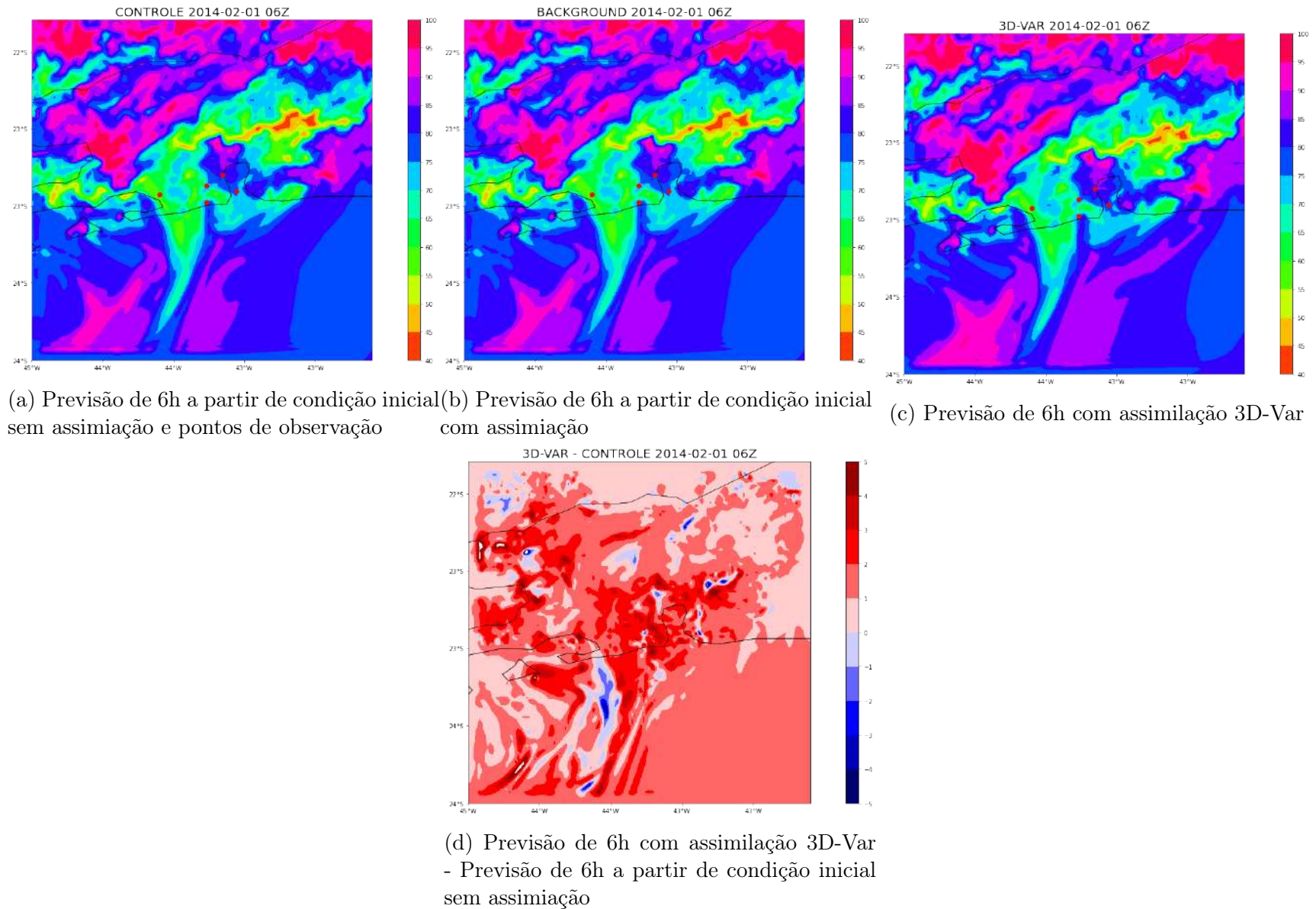


Figura 5.7: Campo de umidade relativa para (a) previsão de 6 horas a partir de condição inicial sem assimilação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimilação, (c) previsão de 6 horas com assimilação 3D-Var e (d) diferença (c) e (a), para 1º de fevereiro de 2014 06 UTC.

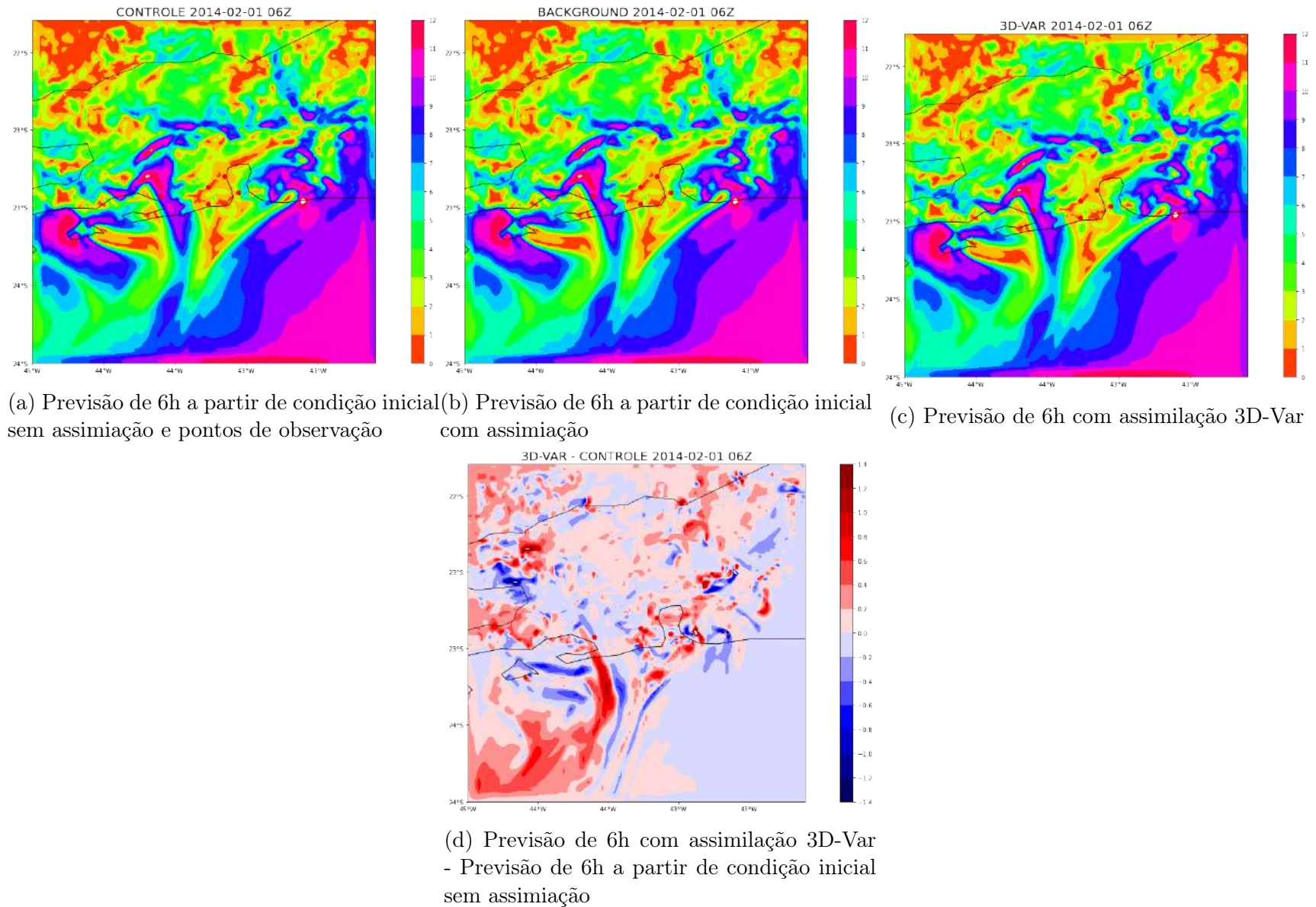
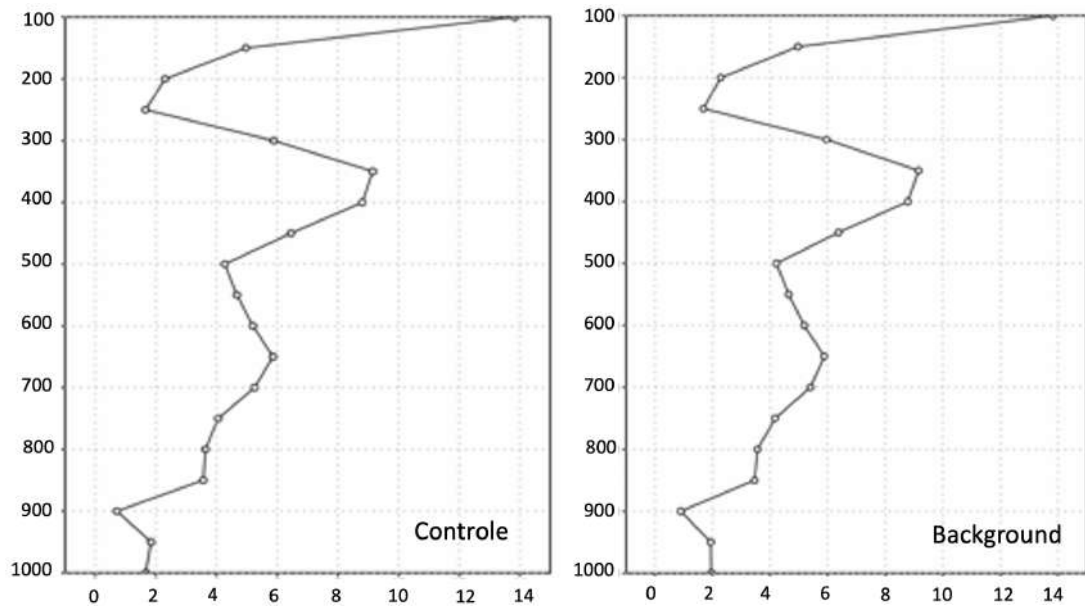
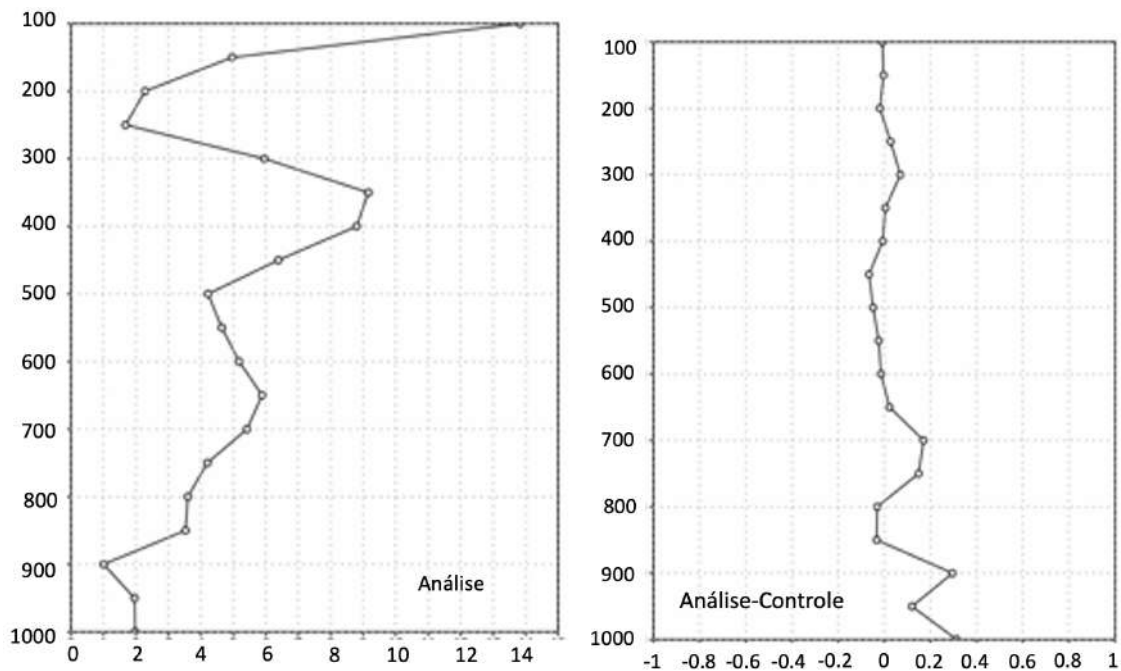


Figura 5.8: Campo de velocidade do vento para (a) previsão de 6 horas a partir de condição inicial sem assimiação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimiação, (c) previsão de 6 horas com assimiação 3D-Var e (d) diferença (c) e (a), para 1º de fevereiro de 2014 06 UTC.



(a) Previsão de 6 horas a partir de condição inicial sem assimilação (b) Previsão de 6 horas a partir de condição inicial com assimilação



(c) Previsão de 6 horas com assimilação 3D-Var

(d) Diferença entre (c) e (a)

Figura 5.9: Perfil de vento para (a) previsão de 6 horas a partir de condição inicial sem assimilação e pontos de observação, (b) previsão de 6 horas a partir de condição inicial com assimilação, (c) previsão de 6 horas com assimilação 3D-Var e (d) diferença (c) e (a), para 1º de fevereiro de 2014 06 UTC.

A figura 5.10 apresenta o resultado do processo de assimilação na localidade do aeroporto internacional do Rio de Janeiro (SBGL) entre 1 e 7 de fevereiro nos anos de 2014 e 2015. Como esperado, o processo de assimilação remove a maior parte da perturbação de ruído branco existente na os dados de observação, representados por erros mais baixos nos dados (linha sólida) em comparação com os valores de previsão (linha tracejada).

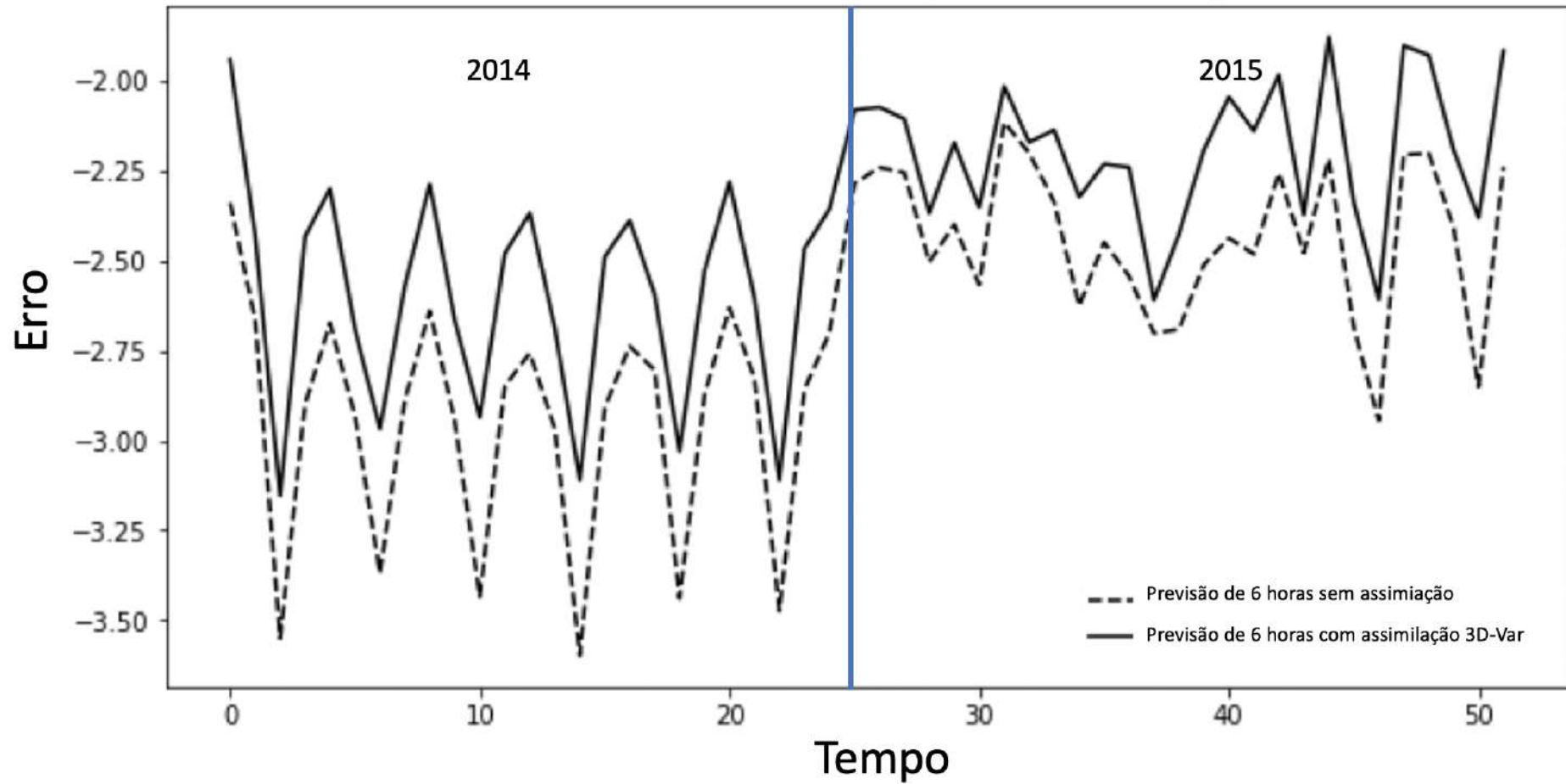


Figura 5.10: Erro no campo de temperatura do ar da observação para previsão de 6 horas sem assimilação e previsão de 6 horas com assimilação no SBGL entre 1 e 7 de fevereiro de 2014 e de 1 a 7 de fevereiro de 2015.

A tabela 5.3 exibe as estatísticas para todos os pontos da grade e variáveis de superfície usadas no processo de assimilação. O impacto geral da assimilação é positivo para variáveis escalares (temperatura do ar e umidade relativa), conforme mostrado pelos valores menores do erro de previsão de 6 horas com assimilação em comparação com o erro de previsão de 6 horas sem assimilação. Quase não houve impacto no processo de assimilação da velocidade do vento (campo vetorial) e o efeito foi apenas uma redução da média geral.

Este comportamento indica que o processo de atenuação no campo de vento é maior do que nos demais campos ou a perturbação aplicada aos dados não foi suficiente para modificar significativamente o campo inicial, o que precisará ser investigado em trabalhos futuros.

De forma geral, pode-se destacar que os erros foram sempre abaixo do valor da perturbação de ruído branco aplicado aos dados de *background*. Este resultado é importante porque mostra mais uma vez que o processo de assimilação possui uma tendência de redução do ruído dos dados de forma a aproximar o campo de previsão de 6 horas com assimilação da dinâmica verdadeira do sistema. Outra característica relevante a ser ressaltada é que o a dispersão próxima ao valor da perturbação mostra que o processo é relativamente estável para todo o domínio, ou seja, não insere grandes perturbações que possam evoluir negativamente durante a integração do modelo em modo previsão.

Tabela 5.3: Estatísticas para todos os pontos de grade e variáveis de superfície no domínio de estudo.

		Temperatura	Umidade Relativa	Veloc. vento
Observação	média (μ_o)	30,13	69,81	4,80
	std (σ_o)	3,84	18,71	3,37
Prev. 6h	média (μ_p)	27,48	63,63	4,81
	std (σ_p)	3,45	17,56	3,40
	erro (e_p)	9,00%	9,00%	0,21%
Prev. 6h com assim. 3D-Var	média (μ_a)	27,74	65,22	4,79
	std (σ_a)	3,49	17,46	3,38
	erro (e_a)	8,00%	7,00%	0,21%

5.2.2 Assimilação de dados sintéticos com redes neurais

A configuração da rede neural artificial de múltiplas camadas (MLP-NN) é uma tarefa demorada, exigindo a avaliação de diferentes topologias para obter uma boa configuração. A tabela 5.4 apresenta os resultados de uma MLP-NN auto-configurada para 20 experimentos realizados para 4 variáveis meteorológicas usando o algoritmo

MPCA. O software MPCA foi aplicado para determinar diferentes parâmetros de uma MLP-NN, como: número de camadas ocultas, número de neurônios em cada camada oculta, tipo de função de ativação e parâmetros do processo de aprendizagem - momento (α) e taxa de aprendizado (η). Na tabela 5.4, os valores 1, 2 e 3 usado para as funções de ativação representam as funções logística, tangente hiperbólica e gaussiana, respectivamente. Os resultados para os experimentos número 3, 1, 1 e 1 (tabela 5.4, colunas 5, 3, 3 e 3, respectivamente) mostram as melhores topologias para as variáveis temperatura do ar, umidade relativa, componentes u e v do vento, respectivamente.

A tabela 5.5 apresenta os valores estatísticos do erro médio (Erro), desvio padrão (DESVPAD), raiz quadrada do erro quadrático médio (REQM) e coeficiente de determinação (R2) calculados para cada variável meteorológica do conjunto de dados de teste. As previsões foram obtidas utilizando-se as melhores topologias de rede neural obtidas no software MPCA - consulte a tabela 5.4. As estatísticas apresentadas na tabela 5.5 mostram que os erros de assimilação a partir do modelo treinado com o MPCA foram inferiores aos valores de perturbação aplicados (10% do valor extraído das variáveis na localização dos sensores das estações dos aeroportos). Por exemplo, a perturbação no vento foi de até $1,0 \text{ ms}^{-1}$ e os erros absolutos foram, no pior resultado, $|0,83| \text{ ms}^{-1}$ para a componente v do vento. O desvio padrão indica que em média há uma baixa dispersão dos valores obtidos pelo MPCA em relação aos obtidos pelo método tradicional de assimilação de dados (3D-Var). Outro aspecto interessante a ser analisado na tabela 5.5 é o coeficiente de determinação (R2) que nos fornece uma medida objetiva do ajuste dos resultados do MPCA em relação aos resultados do 3D-Var. Nota-se que as variáveis temperatura, umidade relativa e a componente u do vento possuem mais de 90% de seus resultados do 3D-Var explicados pelos resultados dos modelos treinados com o MPCA. A componente v do vento possui um valor R2 de 0,83 relativamente mais baixo do que o das demais variáveis. De forma geral, é possível concluir que o modelo MPCA é capaz de aprender o método 3D-Var com baixa dispersão se considerada a parturbação inicial que foi aplicada aos dados, conforme discutido anteriormente.

Tabela 5.4: Resultados do MPCA para o conjunto de treinamento e validação para variáveis meteorológicas.

Variável	Parâmetros	Experimentos				
Temperatura	Valor da função custo	0.0924	0.1233	0.0899	0.0924	0.0770
	Número de camadas	1	1	1	1	2
	Número de neurônios na camada 1	5	16	15	15	11
	Número de neurônios na camada 2	0	0	0	0	24
	Função de ativação	1	3	2	2	1
	Alpha	0.2545	0.6352	0.3492	0.669	0.5306
	Eta	0.0428	0.5617	0.5573	0.2724	0.8033
Umidade Relativa	Valor da função custo value	0.0827	0.0908	0.0891	0.0912	0.0909
	Número de camadas	1	2	1	2	1
	Número de neurônios na camada 1	9	6	20	11	9
	Número de neurônios na camada 2	0	14	0	23	0
	Função de ativação	2	1	1	1	2
	Alpha	0.6701	0.3604	0.2851	0.1981	0.0139
	Eta	0.8110	0.2713	0.5577	0.0318	0.3069
Componente u DO VENTO	Valor da função custo value	0.0737	0.0739	0.0751	0.0754	0.0577
	Número de camadas	1	1	1	1	1
	Número de neurônios na camada 1	5	5	12	12	16
	Número de neurônios na camada 2	0	0	0	0	0
	Função de ativação	1	1	2	2	2
	Alpha	0.1021	0.2733	0.2722	0.1559	0.5584
	Eta	0.4563	0.3299	0.0282	0.0520	0.5277
Componente v do vento	Valor da função custo value	0.0492	0.0841	0.0424	0.0698	0.0772
	Número de camadas	2	2	2	1	1
	Número de neurônios na camada 1	21	14	19	6	10
	Número de neurônios na camada 2	11	7	12	0	0
	Função de ativação	1	2	1	2	1
	Alpha	0.0946	0.7286	0.2388	0.1467	0.3568
	Eta	0.8424	0.1335	0.7601	0.7047	0.5013

Tabela 5.5: Estatísticas do MPCA para o conjunto de generalização. Na tabela a sigla REQM representa a raiz do erro quadrático médio, DESVPAD o desvio padrão e R2 o coeficiente de determinação.

Variáveis	Erro	REQM	DESVPAD	R2
Temperatura (K)	-0,12	0,39	0,37	0,99
Umidade Relativa (%)	1,02	4,17	4,04	0,94
Componente u do vento (ms^{-1})	-0,19	1,07	1,05	0,94
Componente v do vento (ms^{-1})	-0,83	1,77	1,56	0,83

As figuras 5.11 a 5.14 apresentam o gráfico *quantile - quantile* (gráfico para comparar duas distribuições de probabilidade) para temperatura do ar, umidade relativa, as componentes u e v do vento, respectivamente, para o conjunto de dados de teste. Esse tipo de gráfico é muito útil para encontrar viés no modelo de previsão para regiões específicas da distribuição da variável. Observando as figuras 5.11 e 5.12, existe uma tendência de subestimação para temperatura do ar em valores superiores a 35 °C) e umidade relativa no intervalo [80-100%]. Além disso, existe uma leve tendência de superestimação da umidade relativa para percentis mais baixos (abaixo de 30%). Conforme mostrado na tabela 5.11, maiores diferenças são encontradas para as previsões de vento. A figura 5.13 mostra o componente u , onde há uma tendência de subestimação para valores maiores que 5 ms^{-1} . Para a componente v do vento (figura 5.14), há uma tendência de subestimação para valores maiores que 5 ms^{-1} e um viés positivo alto para todos os valores negativos.

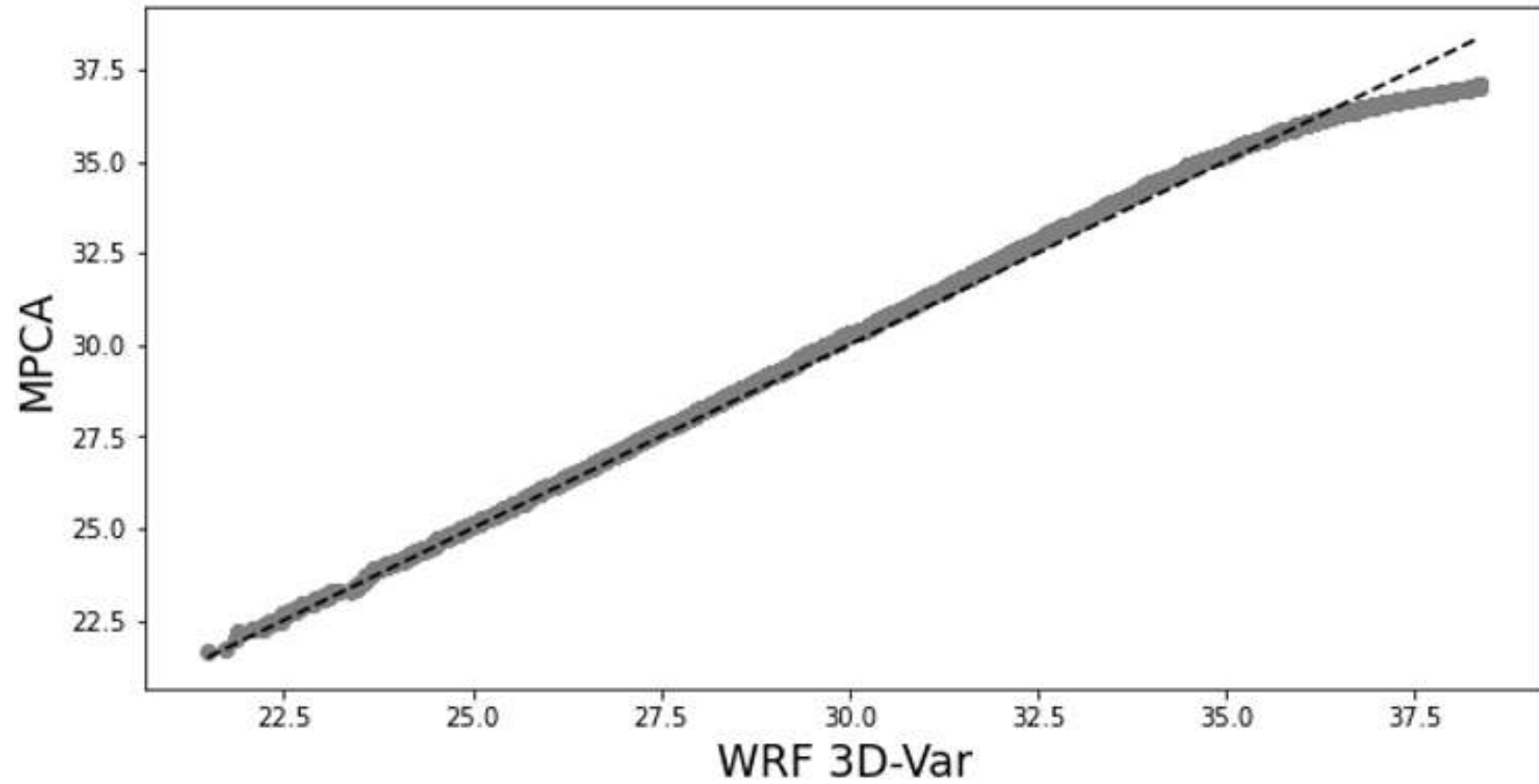


Figura 5.11: Gráfico quantil-quantil comparando a distribuição de probabilidade da temperatura do ar gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.

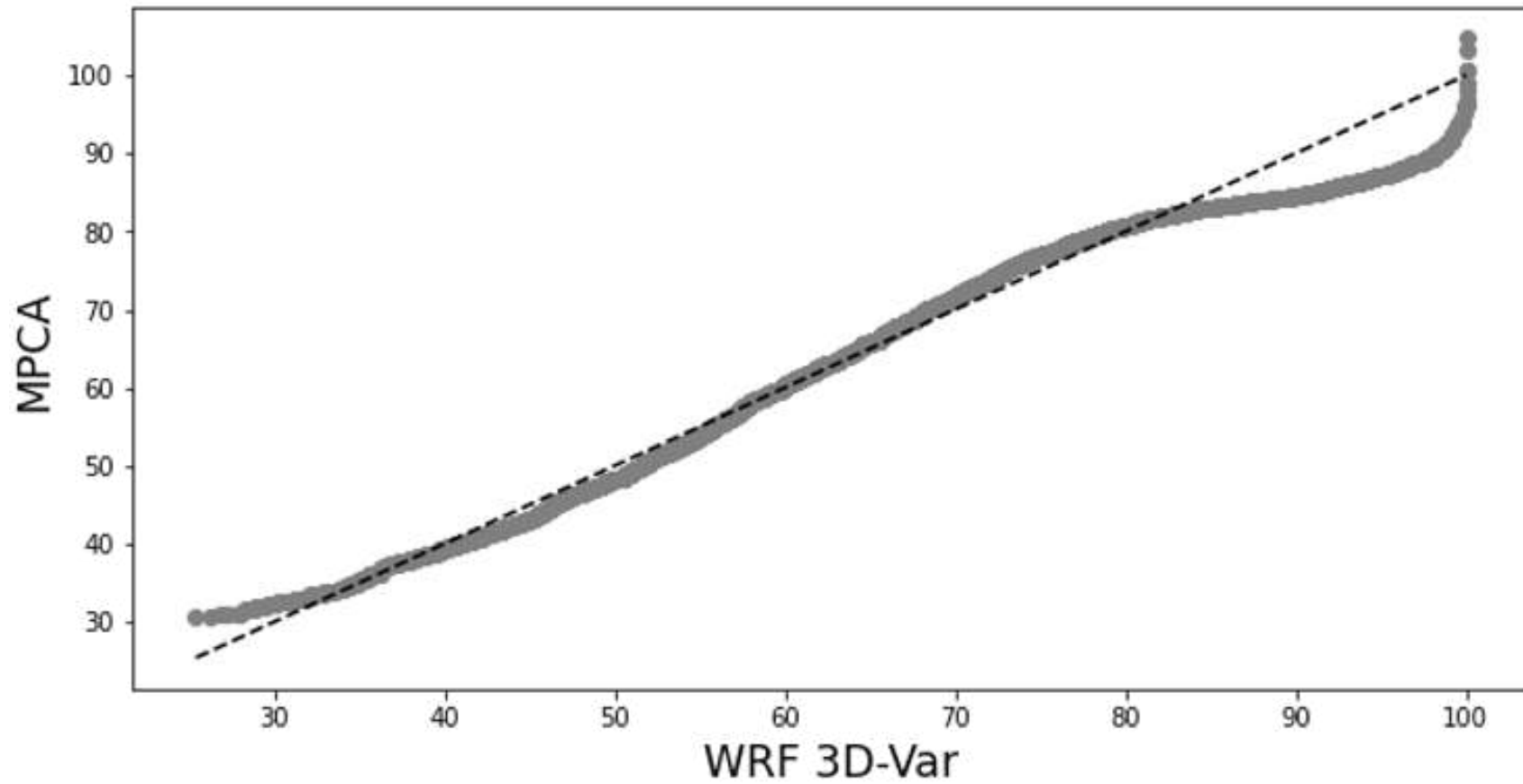


Figura 5.12: Gráfico quantil-quantil comparando a distribuição de probabilidade da umidade relativa gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.

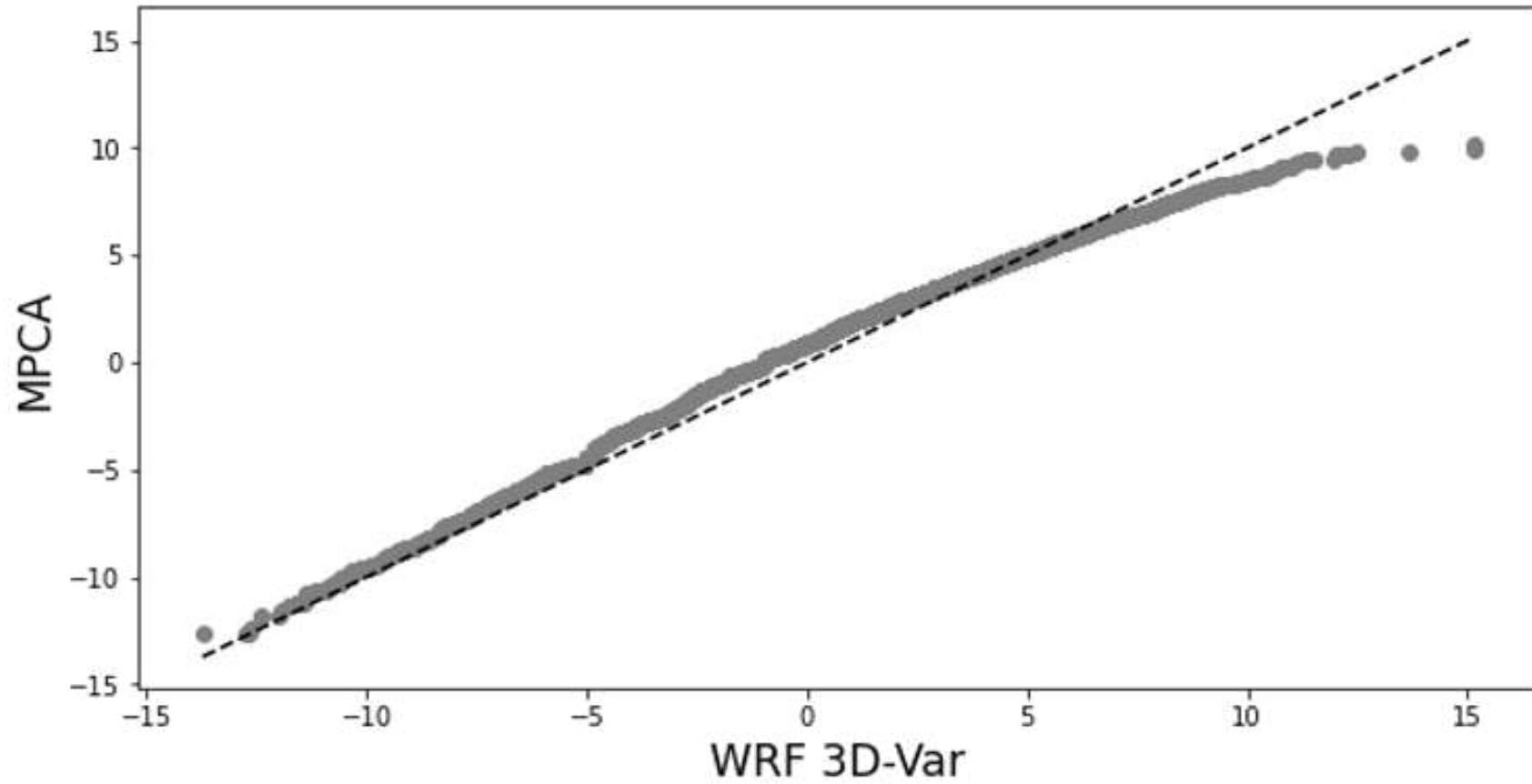


Figura 5.13: Gráfico quantil-quantil comparando a distribuição de probabilidade da componente u do vento gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.

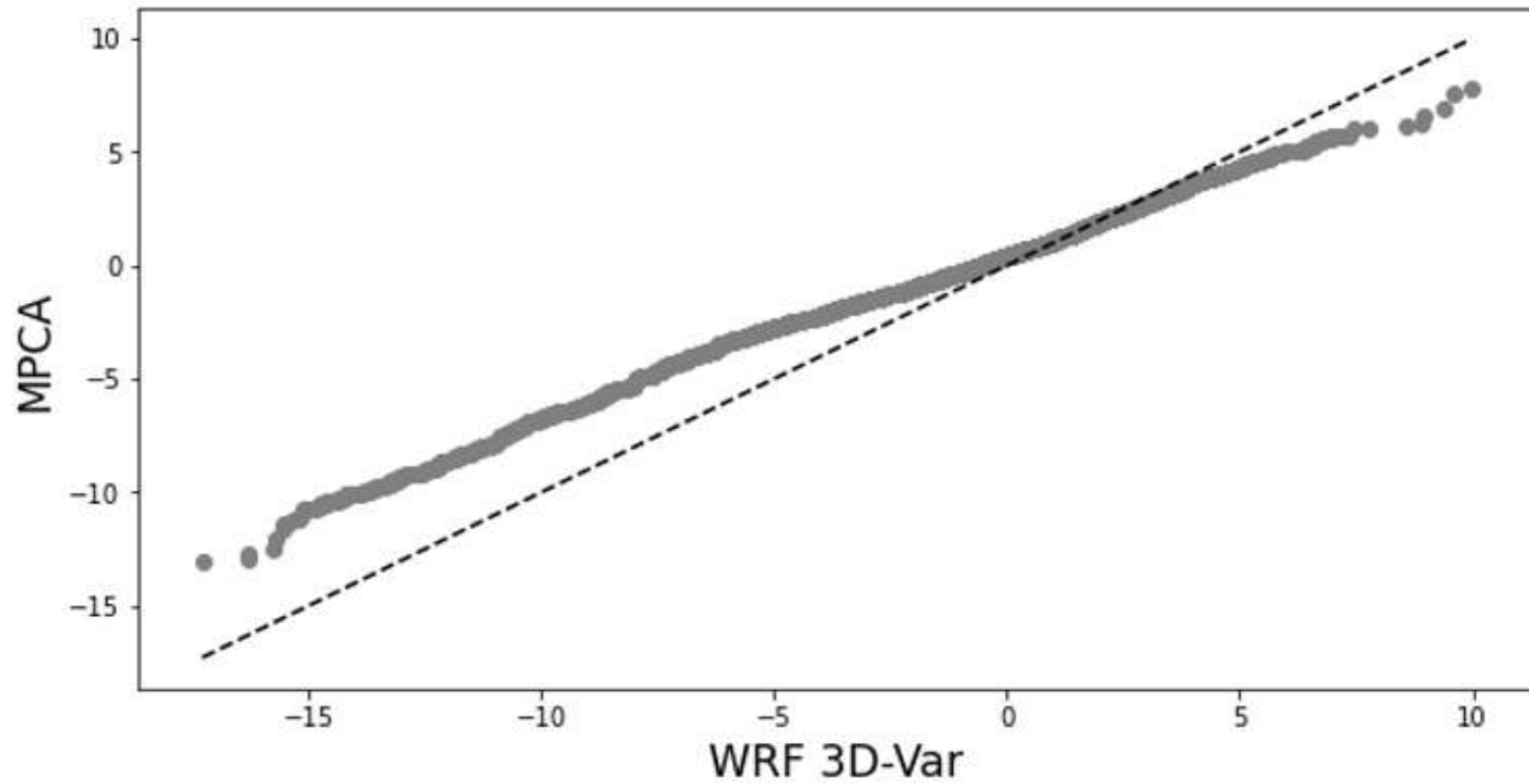


Figura 5.14: Gráfico quantil-quantil comparando a distribuição de probabilidade da componente v do vento gerada pela aplicação do método de assimilação de dados 3D-Var e o algoritmo MPCA aplicado a previsões de 6 horas para os anos de 2014 e 2015.

A figura 5.15 mostra um caso para 01/fev/2014 06 UTC para o campo de previsão de 6 horas sem assimilação (controle) (fig. 5.15a), a previsão de 6 horas com assimilação de dados 3D-Var (fig. 5.15b), a previsão de 6 horas com assimilação de dados a partir da MLP-NN otimizada (fig. 5.15c) e a diferença entre o campo de previsão de 6 horas com assimilação de dados por MLP-NN e 3D-Var (fig. 5.15d), considerando a temperatura do ar a 1000 hPa. Aqui, o campo de controle é a integração do modelo em 6 horas, que é considerado no campo de referência.

Comparando as figuras 5.15b e 5.15c com 5.15a, pode-se notar que há um aumento de valores nos arredores das localizações das estações (pontos vermelhos). Como esperado, embora o processo de assimilação remova grande parte da perturbação do ruído branco nos dados, parte dele ainda altera o campo da variável. A figura 5.15d representa a diferença quadrática entre a assimilação realizada pela técnica 3D-Var e os resultados do modelo treinado pela MPCA. A diferença entre os dois processos é inferior a 3K para todas as regiões, cerca de 1% da magnitude dos dados de temperatura do ar assimilados. As pequenas diferenças são esperadas uma vez que os modelos MPCA foram treinados a partir dos resultados do 3D-Var.

O tempo médio de execução de cada ciclo de assimilação do 3D-Var foi 00:01:11 (1 minuto e 11 segundos), enquanto o tempo médio de execução do modelo de rede neural foi próximo a 00:00:01 (cerca de 1 segundo). Portanto, o método MLP-NN foi (pelo menos) 71 vezes mais rápido que o 3D-Var, produzindo campos muito semelhantes. Destacamos a relevância de se ter uma técnica eficaz e rápida para assimilação de dados, permitindo incluir mais observações em uma resolução mais fina do modelo.

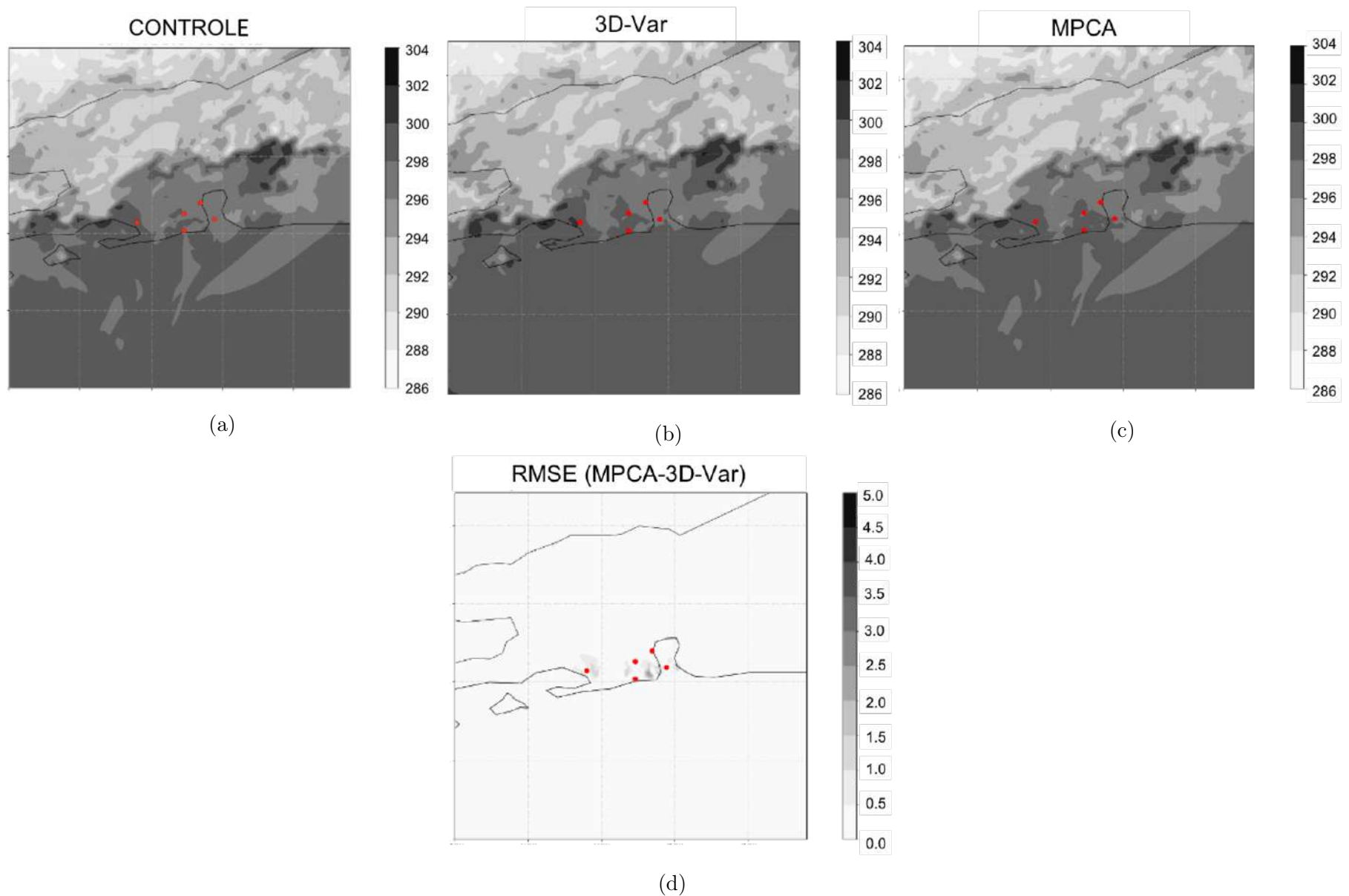


Figura 5.15: Campo de temperatura do ar em 01/fev/2014 06 UTC para: (a) previsão de 6 horas; (b) previsão de 6 horas com assimilação 3D-var; (c) previsão de 6 horas com assimilação MPCA; e (d) diferença de (c) e (b). Os pontos vermelhos representam a localização das estações de superfície onde os dados assimilados foram medidos.

5.2.3 Assimilação de dados das estações com redes neurais

Na seção anterior foram apresentados resultados da utilização de redes neurais artificiais para assimilação de dados a partir da emulação do método 3D-Var. Contudo, como estes resultados foram obtidos apenas a partir da assimilação de dados sintéticos uma questão ainda permanece a ser analisada: a performance do processo de assimilação de dados observados extraídos das estações meteorológicas de superfície dos aeroportos (e não gerados sinteticamente) se manterá? Desta forma, esta seção busca responder a este questionamento, apresentando os resultados do teste para o ano de 2019 de um modelo treinado utilizando dados observados de superfície e perfil vertical dos aeroportos – ver figura 3.2, para o período entre 2015 e 2018.

O processo de treinamento de redes neurais é um problema de otimização de múltiplos parâmetros a fim de encontrar topologias adequadas ao problema estudado.

A tabela 5.6 apresenta a topologia final e outras informações pertinentes sobre as redes neurais treinadas no TensorFlow e Weka. Naturalmente, muitas topologias e diferentes configurações foram testadas e exploradas, mas por uma questão de simplicidade, apenas as configurações que produziram os melhores resultados para a aplicação atual são apresentadas.

Além disso, o processo de treinamento é uma tarefa demorada e a tabela 5.7 apresenta o tempo de CPU consumido para treinamento de redes neurais - tanto no TensorFlow quanto no Weka - usando o conjunto de dados explorado neste trabalho.

A figura 5.16 apresenta o mapa de temperatura do ar de 2 m em 1º de fevereiro de 2019 12 UTC para o campo de previsão de 6 horas, previsão de 6 horas com assimilação de dados 3D-Var e previsão de 6 horas com assimilação de dados por redes treinadas em TensorFlow (NN-TensorFlow) e Weka (NN-Weka). Esta data foi escolhida aleatoriamente a partir do conjunto de dados de teste a ser usado aqui para ilustrar o desempenho da assimilação de dados por NN em comparação aos dados de controle e objetivo (3D-Var). O mapa de previsão de 6 horas sem assimilação (fig. 5.16a) apresenta o campo mais suave, com maiores temperaturas na região metropolitana.

O campo de previsão com assimilação de dados 3D-Var (fig. 5.16b) mostra uma intensificação das temperaturas mais altas na região metropolitana, de acordo com os dados de superfície observados (não apresentados, mas disponíveis em <https://www.redemet.aer.mil.br>) que apresentou temperatura do ar 2 m mais alta do que o campo de previsão de 6 horas sem assimilação nas localizações das estações. O campo mostra o comportamento esperado para saídas de assimilação de dados, onde um campo ótimo é gerado pelo ajuste entre o campo de previsão de 6 horas e as observações disponíveis em uma janela de tempo em específico.

Tabela 5.6: Topologia final e outras características de redes neurais nos experimentos do TensorFlow e Weka.

Parâmetro	NN-TensorFlow	NN-Weka
Versão	2.0.0	3.9.3
Número de Camadas	3	2
Número de unidades escondidas (em cada camada)	30	25
Função de ativação (camadas intermediárias)	ReLU	Sigmoid
Função de ativação (saída)	linear	linear
Otimizador	Adam ¹	Backpropagation ²
Taxa de aprendizado	0.001 (default)	0.3 (default)
Momentum	0.9 (default)	0.2 (default)
Épocas	1000	500

¹<https://keras.io/api/optimizers/adam/>

²<https://weka.sourceforge.io/doc.dev/weka/classifiers/functions/MultilayerPerceptron.html>

Tabela 5.7: Tempo de treinamento para redes neurais. A hora é formatada como hh:mm:ss, em que hh representa as horas, mm representa os minutos e ss representa os segundos ou fração.

Plataforma	Tempo
TensorFlow	08:57:38
Weka	09:10:05

As figuras 5.16c-d apresentam os resultados dos campos de previsão de 6 horas com assimilação de dados a partir dos modelos NN-TensorFlow e NN-Weka, respectivamente. Interessante notar que ambos os métodos de redes neurais são capazes de identificar temperaturas mais altas em relação ao campo de previsão de 6 horas sem assimilação de dados, mostrando sensibilidade aos dados de observação, ou seja, aprendendo o processo de ajuste entre o campo de previsão de 6 horas e os campos observados. Comparado com o campos de previsão de 6 horas com assimilação de dados 3D-Var, as variáveis alvo no processo de treinamento, ambos os resultados do campo de previsão com assimilação de dados a partir de NN apresentam valores mais altos indicando uma perturbação maior nos arredores das localizações das estações.

O campo de previsão de 6 horas com assimilação de dados 3D-Var calcula uma matriz de inovação na grade de observação e, em seguida, interpola os dados de volta para a grade do modelo usando o operador H (ver eq. (2.1)). No treinamento das NN o processo é o oposto, a influência das observações é computada na grade do modelo para treinamento e nenhuma interpolação é necessária para gerar o campo ótimo. Portanto, o comportamento do campo 3D-Var para o mapa de temperatura do ar de 2 m (5.16 b) está provavelmente relacionado a um efeito de interpolação.

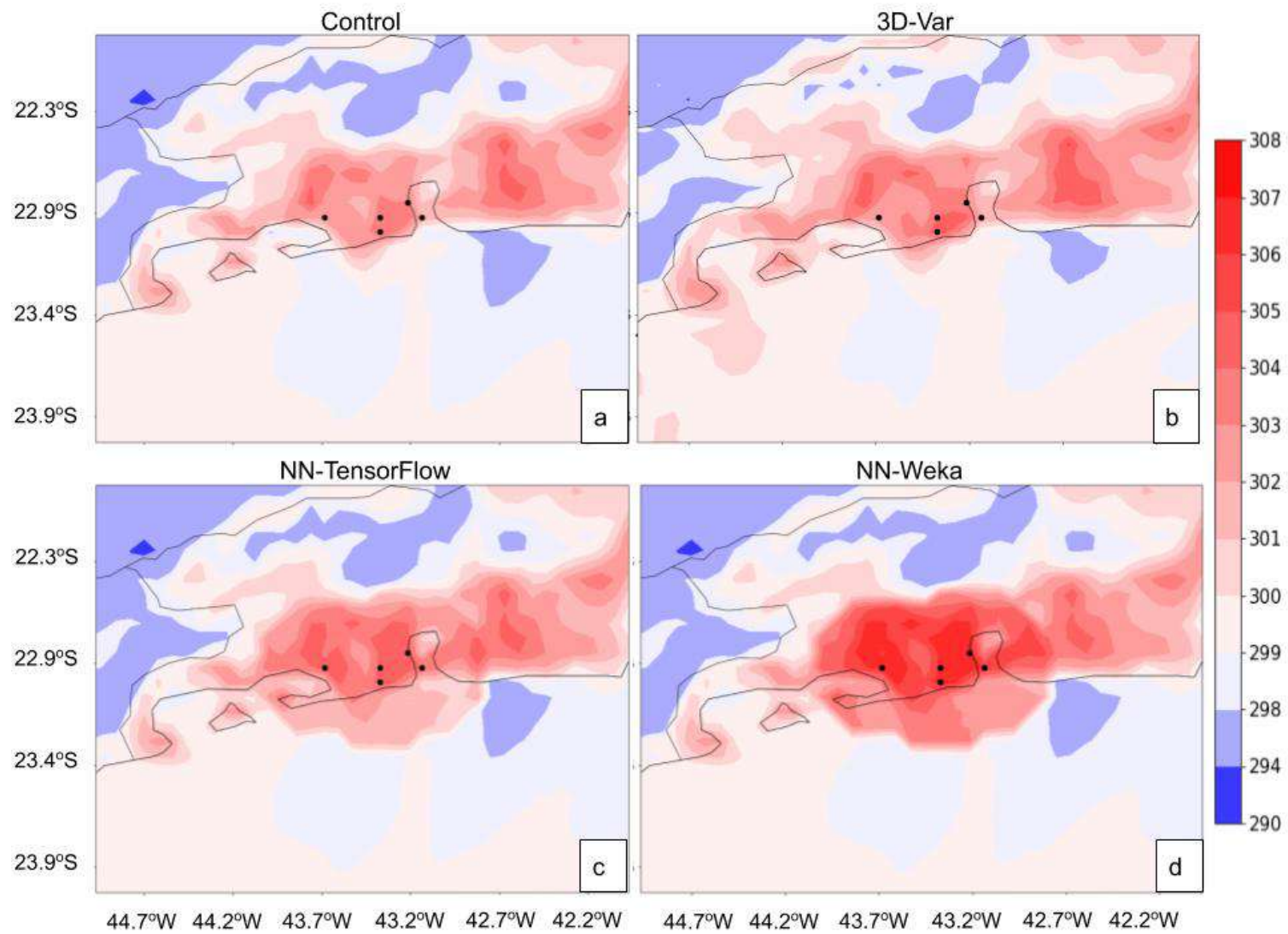


Figura 5.16: Mapa de temperatura do campo (a) previsão de 6 horas sem assimilação de dados (controle), (b) previsão de 6 horas com assimilação de dados 3D-Var e previsão de 6 horas com assimilação de dados por redes neurais - treinado no (c) TensorFlow e (d) Weka - em 1° de fevereiro de 2019 12 UTC.

Na figura 5.16 uma análise subjetiva foi exibida entre os resultados das redes neurais e 3D-Var, mas uma quantificação objetiva das diferenças entre o campo de previsão de 6 horas com assimilação de dados e o campo de observação é necessária para avaliar se as diferenças entre as redes neurais e 3D-Var indicam uma diminuição da qualidade no processo de assimilação de dados ou, ao contrário, uma melhoria.

A figura 5.17 apresenta o mapa de erro de temperatura em 1° de fevereiro de 2019 12 UTC (mesmo dia analisado na figura 5.16 e escolhido aleatoriamente do conjunto de dados de teste para ilustração da presente discussão) para o campo de previsão de 6 horas com assimilação 3D-Var (5.17a), previsão de 6 horas com assimilação NN-TensorFlow (5.17b) e previsão de 6 horas com assimilação NN-Weka (5.17c) em relação ao mapa observado. Os erros mostram que as redes neurais são capazes de reduzir o erro geral na área de estudo (no entorno dos dados da estação). A razão para tal melhoria não é clara à primeira vista, uma vez que o campo de previsão de 6 horas com assimilação de dados 3D-Var é usado como uma variável alvo e, portanto, espera-se que as redes neurais sejam limitadas em sua qualidade pela qualidade da variável alvo (3D-Var neste caso).

Conforme explicado na análise da figura 5.16, uma das possíveis explicações para essas melhorias está no cálculo do impacto dos dados da estação na vizinhança. O método 3D-Var realiza uma interpolação dos dados da grade do modelo para as localizações das observação com objetivo de executar o cálculo de incremento. Em seguida, uma nova interpolação é realizada mais para o domínio da grade original e assim é gerado o campo final. O conjunto de dados para treinamento de redes neurais foi criado de forma diferente. Pseudo-observações foram criadas em uma região de raio de 5 pontos da grade centralizada em cada localização da estação, computando a influência dos dados da estação já no domínio da grade. Desta forma, a rede neural já foi treinada na grade final e os possíveis erros de interpolação e suavização existentes no 3D-Var não foram inseridos nas redes neurais treinadas.

Importante observar que diferentemente do método 3D-Var (fig. 5.17a) e NN-TensorFlow (fig. 5.17b) que mostraram uma subestimação geral do campo de temperatura do ar de 2 m, os resultados do NN-Weka (fig. 5.17c) mostram regiões de superestimação próximas ao sudeste da região metropolitana do Rio de Janeiro, onde há maior densidade de estações.

Como observação final na análise da figura 5.17, o mapa de erros dos modelos NN se assemelha ao padrão 3D-Var - mostrando a capacidade dos algoritmos de aprendizagem - e pode até melhorar o desempenho do método 3D-Var, amplamente utilizado em muitos centros operacionais.

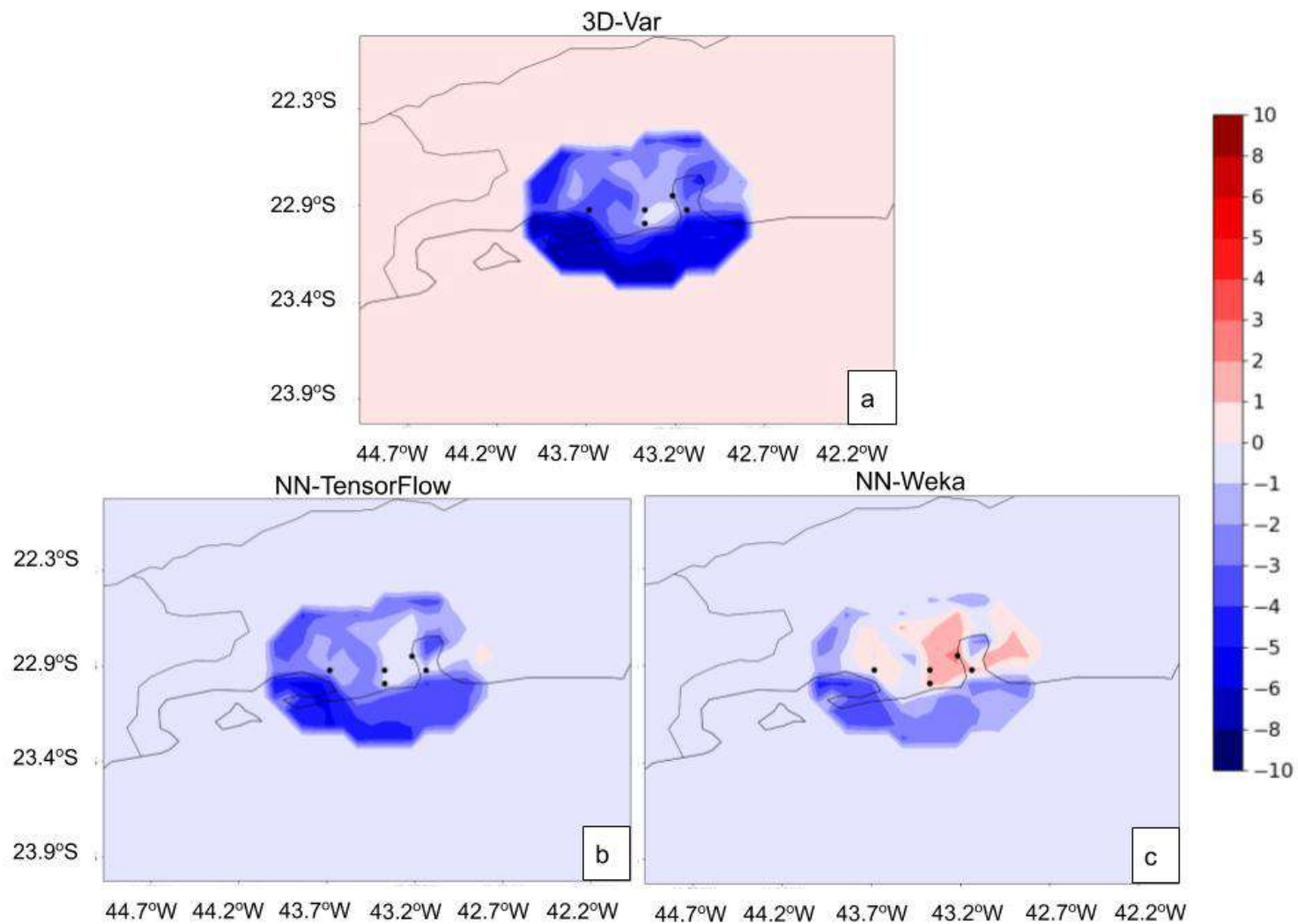


Figura 5.17: Mapa de erro de temperatura dos campos de previsão de 6 horas com assimilação de dados (a) 3D-Var e redes neurais - treinados em (b) TensorFlow e (c) Weka - em relação ao mapa observado em 1° de fevereiro de 2019 12 UTC.

Além dos dados de superfície, perfis de perfis de temperatura do ar, umidade e vento (velocidade e direção) também foram assimilados neste trabalho, obtidos a partir da estação do aeroporto SBGL.

A figura 5.18 apresenta o perfil de erro de temperatura média do ar de 1° de fevereiro de 2019 a 8 de fevereiro de 2019 para o perfil de previsão de 6 horas sem assimilação de dados, previsão de 6 horas com assimilação de dados 3D-Var (curva 3dvar-obs), previsão de 6 horas com assimilação de dados NN-TensorFlow (curva tf-obs) e previsão de 6 horas com assimilação de dados NN-Weka (curva weka-obs) em relação ao perfil observado da superfície até 20 km, na estratosfera. Além disso, também são apresentadas as diferenças entre os perfis médios dos perfis de previsão de 6 horas com assimilação de dados por redes neurais NN-TensorFlow (curva tf-3dvar) e NN-Weka (curva weka-3dvar) em relação aos perfis de previsão de 6 horas com assimilação de dados usando o método 3D-Var.

Importante ressaltar que diferentemente dos dados de superfície, os dados de perfil no SBGL estão disponíveis apenas duas vezes ao dia, às 00 e 12 UTC, e assim, a figura 5.18 apresenta uma média de 14 perfis (2 sondagens x 7 dias).

Pela figura 5.18 é possível identificar duas regiões diferentes com maiores erros: na superfície até 7,5 km (baixa a média troposfera) e na estratosfera (17,5 a 20 km). O efeito geral do processo de assimilação de dados é reduzir o erro médio do perfil, principalmente nas regiões acima mencionadas onde a saída de controle apresenta erros maiores (por exemplo, abaixo de 2,5 km o erro médio na saída de controle atinge um máximo de 3 K enquanto a assimilação de dados as saídas não excedem 2 K).

A análise dos resultados do modelo NN mostra uma estreita concordância com a curva 3D-Var com uma redução do erro médio (em grande parte do perfil com algumas exceções perto da tropopausa) e resultados ligeiramente melhores para NN-TensorFlow (NN-TF) em comparação com NN-Weka. É importante notar que a região da alta troposfera e baixa estratosfera é caracterizada por dados da rodada de previsão de 6 horas sem assimilação de dados, com melhores resultados do que as curvas de perfis de previsão de 6 horas com assimilação de dados.

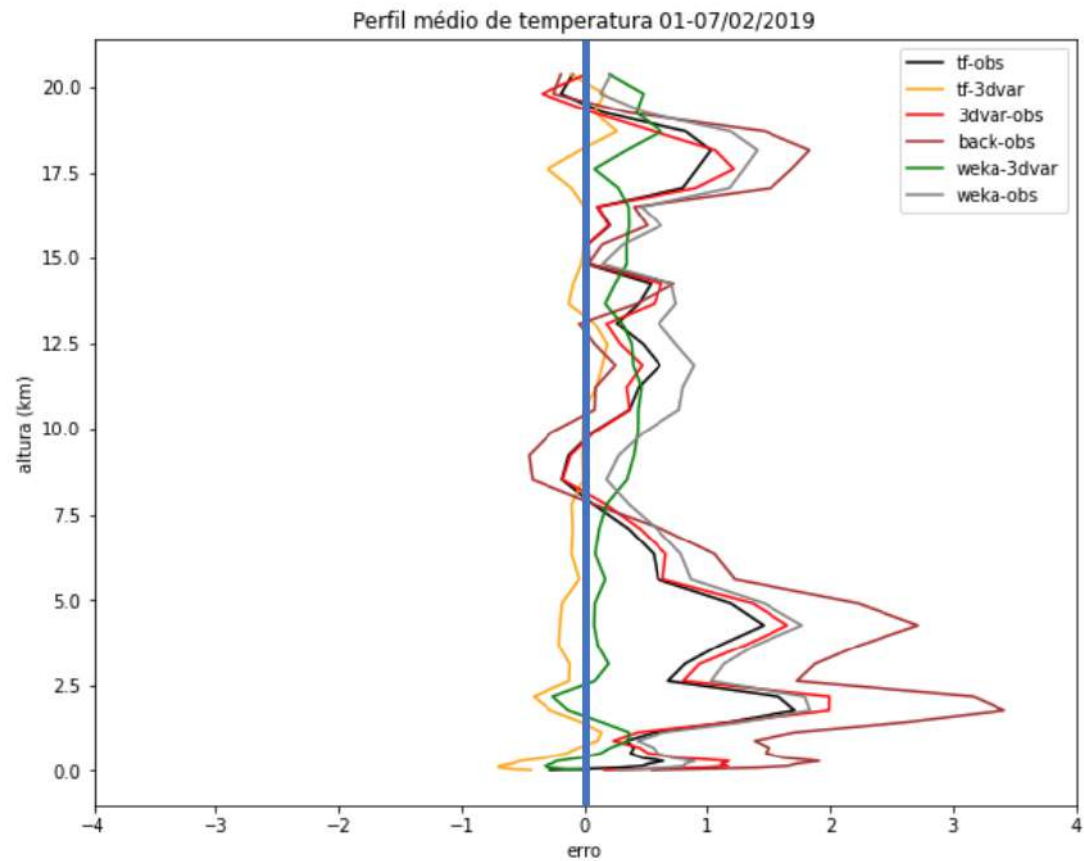


Figura 5.18: Perfil no período 1 a 7/2/2019 para os erros médios de previsão de 6 horas entre: perfil de previsão de 6 horas com assimilação de dados usando o *tensorflow* e dados observados (tf-obs); perfil de previsão de 6 horas com assimilação de dados usando o *tensorflow* e perfil de previsão de 6 horas com assimilação de dados usando o 3D-Var (tf-3dvar); perfil de previsão de 6 horas com assimilação de dados usando 3dvar e perfil médio observado (3dvar-obs); previsão de 6 horas sem assimilação de dados e perfil médio observado (back-obs); previsão de 6 horas com assimilação de dados usando o *weka* e previsão de 6 horas com assimilação de dados usando o 3dvar (weka-3dvar); previsão de 6 horas com assimilação de dados usando o *weka* e perfil médio observado (weka-obs). O eixo y contém alturas da superfície até cerca de 20 km.

Capítulo 6

Conclusões e Trabalhos Futuros

6.1 Conclusões

Neste trabalho inicialmente foram desenvolvidas técnicas para a previsão de eventos convectivos meteorológicos na área terminal de voo do Rio de Janeiro usando diferentes algoritmos de inteligência computacional, dados de descarga atmosférica obtidos a partir de sensoriamento remoto para classificar a ocorrência e severidade das tempestades e índices de instabilidade atmosférica. As principais conclusões desta parte do trabalho são resumidas da seguinte forma:

- A análise dos resultados mostra que o modelo de aprendizagem profunda totalmente conectado apresentou um desempenho levemente superior para previsão e severidade de tempestades, respectivamente, considerando os valores POD, FAR e CSI;
- A análise de erros mostrou que a maioria das previsões incorretas estavam relacionadas a eventos de baixa intensidade. Portanto, os modelos aqui apresentados podem ser usados para eventos de alta intensidade, que têm o maior potencial para interromper as operações de aviação;
- A avaliação de testes de 6 anos mostrou que o modelo pode ser utilizado para aplicação operacional pelo setor da aviação;
- O modelo pode ser uma ferramenta potencial para elaborar as mensagens meteorológicas de previsão de tempo significativo (com seis horas de antecedência) para a aviação na área terminal do Rio de Janeiro;
- Os resultados apresentaram um desenvolvimento de nosso trabalho anterior (FRANÇA *et al.* (2016)) com aumento (diminuição) do POD (FAR) para previsão de tempestade e incluíram previsões de severidade das tempestades;

- É importante destacar a importância dos dados medidos remotamente (por exemplo, descargas atmosféricas) para o desenvolvimento de modelos melhorados para previsões meteorológicas alinhadas com os fins da aviação;
- Foi possível relacionar EMCs com outros fenômenos meteorológicos de impacto à sociedade e navegação: convergência em baixos níveis, rajadas de vento, precipitação, densidade de descargas atmosféricas;
- Importante destacar que o processo de assimilação pode contribuir de forma muito positiva para a previsão refinada de EMC – horário de início e fim e local de ocorrência – após o alerta inicial feito pelo modelo de aprendizado de máquina.

O segundo tema neste trabalho foi o desenvolvimento de técnicas de assimilação de dados de superfície e de altos níveis em campos de previsão de 6 horas na área terminal de voo do Rio de Janeiro usando redes neurais artificiais. As principais conclusões desta parte do trabalho são resumidas da seguinte forma:

- Os resultados da rotina de assimilação de dados aplicadas a campos de previsão de 6 horas foram capazes de ajustar o campo inicial do perfil de temperatura e umidade relativa do aeroporto, atenuando a perturbação do ruído branco da observação sintética;
- Os experimentos realizados mostraram o impacto positivo da assimilação no desempenho geral do modelo. O método de assimilação em campos de previsão de 6 horas pode ser eficaz para a janela atual de tempo de *nowcasting* (previsão de curto prazo), abaixo de 24 horas, removendo a perturbação do ruído presente em observações e também ajustando os campos meteorológicos às informações locais;
- Os resultados mostraram similaridades entre os resultados de assimilação de dados entre 3D-Var e MLP-NN;
- A assimilação de dados com redes neurais emulando o 3D-Var com dados reais de estações gerou em média uma *análise* de qualidade relativamente superior;
- Neste estudo, o método de assimilação de dados neurais foi de 71 a 100 vezes mais rápido (tempo de CPU) que a abordagem 3D-Var nos experimentos usando dados sintéticos e com observações reais coletadas de sensores nos aeroportos.

6.2 Trabalhos Futuros

Sugestões de trabalho futuro para previsão de eventos convectivos incluem:

- Aperfeiçoamento do modelo de previsão de EMC para prever as áreas mais prováveis (ou setor da área terminal) a serem afetadas pelos EMCs;
- Análise completa das condições sinóticas relacionadas aos eventos mais intensos deve ser estudada para desenvolver um modelo conceitual de gênese e intensificação da tempestade;
- Uso de perfis atmosféricos obtidos por sensoriamento remoto e índices de instabilidade atmosférica derivados de dados de satélite de alta frequência podem ser útil para aplicar essa técnica de previsão em rotas de voo aéreas nas quais não há sondagens disponíveis.

Sugestões de trabalho futuro para assimilação de dados usando redes neurais artificiais incluem:

- Treinamento de redes neurais artificiais para emular métodos de assimilação mais sofisticados como o 4D-Var e os método híbrido que leva em consideração a evolução temporal da matriz de erros do modelo de previsão;
- Assimilação de outros tipos de dados, como: perfiladores acústico (SODAR) e ótico (LIDAR); dados de refletividade e vento extraídos de radares; radiação e perfis de temperatura e umidade obtidos a partir de satélites; dados de descargas atmosféricas; e dados obtidos a partir de sensores à bordo de aeronaves;
- Aplicação dos métodos de assimilação propostos no presente trabalho para outras áreas terminais importantes do Brasil;

6.3 Contribuições

A tabela 6.1 apresenta a lista de contribuições científicas realizadas como consequência da pesquisa desenvolvida nesta tese. Destaca-se que alguns dos trabalhos ainda estão em fase de publicação ou revisão e em breve estarão disponíveis para a comunidade científica.

Tabela 6.1: Resumo da contribuição científica realizada no âmbito da tese.

Título	Revista	Status
Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges	<i>International Journal of Remote Sensing</i>	Publicado ¹
Data Assimilation for Nowcasting in the Terminal Control Area of Rio de Janeiro	Anais do XI Workshop Brasileiro de Micrometeorologia	Publicado ²
Data assimilation for nowcasting in the terminal area of Rio de Janeiro	Ciência & Natura	Publicado ³
Artificial Neural Network for Data Assimilation by WRF model in Rio de Janeiro	Revista Brasileira de Geofísica	Aceito
Data Assimilation using WRFDA over the terminal area of Rio de Janeiro	Revista Brasileira de Meteorologia	Publicado ⁴
Neural networks for data assimilation of surface and upper-air data in Rio de Janeiro	Geoscientific Model Development	Em revisão

¹<https://doi.org/10.1080/01431161.2020.1717669>

²<https://www.even3.com.br/anais/micrometeorologia2019>

³<https://doi.org/10.5902/2179460X53224>

⁴<https://doi.org/10.1590/0102-77863610001>

Referências Bibliográficas

- ABADI, M., AGARWAL, A., BARHAM, P., et al., 2015. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”. Disponível em: <<https://www.tensorflow.org/>>. Software available from tensorflow.org.
- ALMEIDA, M. V., 2009, *Aplicação de Técnicas de Redes Neurais Artificiais na Previsão de curtiíssimo prazo da Visibilidade e Teto para o Aeroporto de Guarulhos SP*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ALMEIDA, V. A., FRANÇA, G. B., CAMPOS VELHO, H. F. D., 2020, “Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges”, *International Journal of Remote Sensing*, v. 41, n. 11, pp. 4372–4388. doi: 10.1080/01431161.2020.1717669. Disponível em: <<https://doi.org/10.1080/01431161.2020.1717669>>.
- ANOCHI, J. A., 2015, *CLIMATE PRECIPITATION PREDICTION BY SELF-CONFIGURED NEURAL NETWORKS*. Tese de D.Sc., INPE, São José dos Campos, SP, Brasil.
- ANOCHI, J. A., CAMPOS VELHO, H. F. D., 2014, “Optimization of feedforward neural network by multiple particle collision algorithm”. pp. 128–134.
- BARKER, D., HUANG, W., GUO, Y., et al., 2004, “A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results”, *Monthly Weather Review - MON WEATHER REV*, v. 132 (04). doi: 10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2.
- BARKER, D., HUANG, X.-Y., LIU, Z., et al., 2012, “The Weather Research and Forecasting Model’s Community Variational/Ensemble Data Assimilation System: WRFDA”, *Bulletin of The American Meteorological Society - BULL AMER METEOROL SOC*, v. 93 (06), pp. 831–843. doi: 10.1175/BAMS-D-11-00167.1.
- BATAN, L. J., 1973, *Radar Observations of the Atmosphere*. 1 ed. Chicago, University of Chicago Press.

- BREIMAN, L., 2001, “Random Forests”, *Machine Learning*, v. 45, pp. 5–32. doi: <https://doi.org/10.1023/A:1010933404324>.
- CINTRA, R. S., 2010, *Data assimilation with artificial neural networks in atmospheric general circulation model*. Tese de D.Sc., INPE, São José dos Campos, SP, Brasil.
- CINTRA, R., CAMPOS VELHO, H. F., 2012, “Global Data Assimilation Using Artificial Neural Networks In Speedy Model”. .
- CONIGLIO, M. C., ELMORE, K. L., KAIN, J. S., et al., 2010, “Evaluation of WRF Model Output for Severe Weather Forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment”, *Weather and Forecasting*, v. 25, n. 2, pp. 408–427. doi: 10.1175/2009WAF2222258.1.
- COOK, L., WOOD, B., KLEIN, A., et al., 2009, “Analyzing the Share of Individual Weather Factors Affecting NAS Performance Using the Weather Impacted Traffic Index”. In: *9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*. doi: 10.2514/6.2009-7017. Disponível em: <<https://arc.aiaa.org/doi/abs/10.2514/6.2009-7017>>.
- DALEY, R., 1991, “Atmospheric Data Analysis”. .
- DANIELS, T. S., TSOUCALAS, G., ANDERSON, M., et al., 2004, *Tropospheric airborne meteorological data reporting (TAMDAR) sensor development*. Conference Paper Paper 7.6. Disponível em: <<https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040161135.pdf>>.
- DE FREITAS, J. H. V., FRANÇA, G. B., MENEZES, W. F., 2019, “Deep Convection Forecasting Using Decision Tree in Rio de Janeiro Metropolitan Area”, *Anuário Do Instituto De Geociências*, v. 42, n. 1, pp. 127–134. doi: http://dx.doi.org/10.11137/2019_1_127_134. Disponível em: <http://www.anuario.igeo.ufrj.br/2019_01/2019_1_127_134.pdf>.
- DECEA, 2011, *Concepção Operacional ATM Nacional*. Relatório Técnico DCA 351-2. Disponível em: <<https://publicacoes.decea.gov.br/?i=publicacao&id=3678>>.
- DINES, W. H., 1917, “METEOROLOGY AND AVIATION”, *Monthly Weather Review*, v. 45, n. 8 (08), pp. 401–401. ISSN: 0027-0644. doi: 10.1175/1520-0493(1917)45<401b:MAA>2.0.CO;2. Disponível em: <[https://doi.org/10.1175/1520-0493\(1917\)45<401b:MAA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1917)45<401b:MAA>2.0.CO;2)>.

- DIXON, M., WIENER, G., 1993, “TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology”, *Journal of Atmospheric and Oceanic Technology*, v. 10, n. 6, pp. 785–797. doi: 10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.
- EIBE FRANK, M. A. H., WITTEN, I. H., 2016, *The WEKA Workbench. Online Appendix for 'Data Mining: Practical Machine Learning Tools and Techniques'*. Relatório técnico. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf>.
- FAA, 2016, *The Future of the NAS'*. Relatório técnico. Disponível em: <<https://www.faa.gov/nextgen/media/futureofthenas.pdf>>.
- FRANÇA, G. B., DE ALMEIDA, M. V., ROSETTE, A. C., 2016, “An automated nowcasting model of significant instability events in the flight terminal area of Rio de Janeiro, Brazil”, *Atmospheric Measurement Techniques*, v. 9, n. 5, pp. 2335–2344. doi: 10.5194/amt-9-2335-2016.
- FRANK, E., HALL, M. A., WITTEN, I. H., 2016, “The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"”. In: *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann.
- FRANÇA, G. B., DE ALMEIDA, M. V., BONNET, S. M., et al., 2018, “Nowcasting model of low wind profile based on neural network using SODAR data at Guarulhos Airport, Brazil”, *International Journal of Remote Sensing*, v. 39, n. 8, pp. 2506–2517. doi: 10.1080/01431161.2018.1425562.
- GALWAY, J. G., 1956, “The Lifted Index as a Predictor of Latent Instability”. .
- GEORGE, J. J., 1961, “Weather forecasting for aeronautics. By J. J. George. New York and London Academic Press, 1960. Pp. ix, 673; 592 Figures; 43 Tables. \$15”, *Quarterly Journal of the Royal Meteorological Society*, v. 87, n. 371, pp. 120–120. doi: 10.1002/qj.49708737120. Disponível em: <<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49708737120>>.
- GOODFELLOW, I., BENGIO, Y., COURVILLE, A., 2016, *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- GULTEPE, I., HEYMSFIELD, A. J., FIELD, P. R., et al., 2017a, “Ice-Phase Precipitation”, *Meteorological Monographs*, v. 58 (03), pp. 6.1–6.36. ISSN: 0065-9401. doi: 10.1175/AMSMONOGRAPHS-D-16-0013.1. Disponível em: <<https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0013.1>>.

- GULTEPE, I., SHARMAN, R., WILLIAMS, P. E. A., 2019, “A Review of High Impact Weather for Aviation Meteorology”, *Pure Appl. Geophys*, v. xxx, pp. 1–53. doi: <https://doi.org/10.1007/s00024-019-02168-6>.
- GULTEPE, I., HEYMSFIELD, A. J., GALLAGHER, M., et al., 2017b, “Ice Fog: The Current State of Knowledge and Future Challenges”, *Meteorological Monographs*, v. 58 (10), pp. 4.1–4.24. ISSN: 0065-9401. doi: 10.1175/AMSMONOGRAPHS-D-17-0002.1. Disponível em: <<https://doi.org/10.1175/AMSMONOGRAPHS-D-17-0002.1>>.
- HAN, L., FU, S., ZHAO, L., et al., 2009, “3D Convective Storm Identification, Tracking, and Forecasting—An Enhanced TITAN Algorithm”, *Journal of Atmospheric and Oceanic Technology - J ATMOS OCEAN TECHNOL*, v. 26 (04). doi: 10.1175/2008JTECHA1084.1.
- HARTER, F., 2004, *Reccurent Neural Networks for data assimilation in nonlinear dynamics*. Tese de D.Sc., INPE, São José dos Campos, SP, Brasil.
- HAYKIN, S., 1999, *Neural Networks: A Comprehensive Foundation*. 2 ed. New Jersey, USA, Prentice-Hall. ISBN: 0-13-273350-1.
- HOUZE, R., 2014, “Copyright”. In: Houze, R. A. (Ed.), *Cloud Dynamics*, v. 104, *International Geophysics*, Academic Press, p. iv. doi: <https://doi.org/10.1016/B978-0-12-374266-7.09984-4>. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780123742667099844>>.
- ICAO, 2016, *GLOBAL AIR NAVIGATION PLAN (GANP)*. Relatório Técnico 9750. Disponível em: <<https://www.icao.int/airnavigation/Documents/GANP-2016-interactive.pdf>>.
- ISAAC, G., BAILEY, M., COBER, S., et al., 2006, “Airport Vicinity Icing and Snow Advisor (AVISA)”, *Collection of Technical Papers - 44th AIAA Aerospace Sciences Meeting*, v. 20 (01). doi: 10.2514/6.2006-1219.
- ISAAC, G., BAILEY, BOUDALA, F., et al., 2011, “Decision Making Regarding Aircraft De-Icing and In-Flight Icing Using the Canadian Airport Nowcasting System (CAN-Now)”. 06. doi: 10.4271/2011-38-0029.
- ISAAC, G., BAILEY, M., BOUDALA, F., et al., 2014, “The Canadian Airport Nowcasting system (CAN-Now)”, *Meteorological Applications*, v. 21 (01). doi: 10.1002/met.1342.
- JOHNSON, J., MACKEN, P., WITT, A., et al., 1998, “The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm”,

Weather and Forecasting - WEATHER FORECAST, v. 13 (06), pp. 263–276. doi: 10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2.

KAIN, J., JANISH, P., WEISS, S., et al., 2003, “Collaboration between Forecasters and Research Scientists at the NSSL and SPC: The Spring Program”, *Bulletin of The American Meteorological Society - BULL AMER METEOROL SOC*, v. 84 (12). doi: 10.1175/BAMS-84-12-1797.

KAIN, J., XUE, M., CONIGLIO, M., et al., 2010, “Assessing Advances in the Assimilation of Radar Data and Other Mesoscale Observations within a Collaborative Forecasting-Research Environment”, *Weather and Forecasting - WEATHER FORECAST*, v. 25 (10). doi: 10.1175/2010WAF2222405.1.

KALNAY, E., 2003a. “Atmospheric Modeling, Data Assimilation and Predictability”. a.

KALNAY, E., 2003b. “Atmospheric Modeling, Data Assimilation and Predictability”. b.

KAMANGIR, H., COLLINS, W., TISSOT, P., et al., 2020, “A deep-learning model to predict thunderstorms within 400 km² South Texas domains”, *Meteorological Applications*, v. 27, n. 2, pp. e1905. doi: 10.1002/met.1905. Disponível em: <<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1905>>.

KESSINGER, C., HERZEGH, P., BLACKBURN, G., et al., 2016, “The FAA AWRP Oceanic weather program development team”, *12th aviation, range and aerospace meteorology conference, AMS, Atlanta, GA, 30 Jan–2 Feb 2006*.

KIRKPATRICK, S., GELATT, C., VECCHI, M., 1983, “Optimization by Simulated Annealing”, *Science (New York, N.Y.)*, v. 220 (06), pp. 671–80. doi: 10.1126/science.220.4598.671.

KOSKINEN, J., POUTIAINEN, J., SCHULTZ, D., et al., 2011, “The Helsinki Testbed: A Mesoscale Measurement, Research, and Service Platform”, *BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY*, v. 92 (03), pp. 325–342. doi: 10.1175/2010BAMS2878.1.

KUNCHEVA, L. I., 2002, “A theoretical study on six classifier fusion strategies”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 2, pp. 281–286.

- LECUN, Y., BENGIO, Y., HINTON, G., 2015, “Deep learning”, *Nature*, v. 521, n. 7553, pp. 436–444. doi: 10.1038/nature14539. Disponível em: <<https://doi.org/10.1038/nature14539>>.
- LIU, Y., WARNER, T., BOWERS, J., et al., 2008, “The Operational Mesogamma-Scale Analysis and Forecast System of the U.S. Army Test and Evaluation Command. Part I: Overview of the Modeling System, the Forecast Products, and How the Products Are Used”, *Journal of Applied Meteorology and Climatology - J APPL METEOROL CLIMATOL*, v. 47 (04), pp. 1077–1092. doi: 10.1175/2007JAMC1653.1.
- MASS, C., 2012, “NOWCASTING The Promise of New Technologies of Communication, Modeling, and Observation”, *Bulletin of the American Meteorological Society*, v. 93 (06), pp. 797–809. doi: 10.1175/BAMS-D-11-00153.1.
- MILLER, R. C., 1972, “Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central”, , n. 200(R).
- NASCIMENTO, E., 2005, “Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil?” *Revista Brasileira de Meteorologia*, v. 20 (01), pp. 121–140.
- NASCIMENTO, M. M. S., 2018, *Estudo do padrão sazonal e interanual de raios na região sudeste do Brasil e avaliação da representatividade da previsibilidade através de inteligência computacional (árvore de decisão)*. Graduação em meteorologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.
- NOWOSAD, A. G., 2001, *NEW APPROACHES TO METEOROLOGIC DATA ASSIMILATION*. Tese de D.Sc., INPE, São José dos Campos, SP, Brasil.
- NTSB, 2010, “NASDAC Review of National Transportation Safety Board (NTSB) weather-related accidents (2003–2007)”, Disponível em: <<http://www.asias.faa.gov/>>.
- PACHECO DA LUZ, E., BECCENERI, J., CAMPOS VELHO, H., 2011, “Multiple Particle Collision Algorithm Applied to Radiative Transference and Pollutant Localization Inverse Problems”. pp. 347 – 351, 06. doi: 10.1109/IPDPS.2011.171.
- PAULUCCI, T., FRANÇA, G., LIBONATI, R., et al., 2019, “Long-Term Spatial–Temporal Characterization of Cloud-to-Ground Lightning in the Me-

- tropolitan Region of Rio de Janeiro”, *Pure and Applied Geophysics*, (05), pp. 1–15. doi: 10.1007/s00024-019-02216-1.
- POLIFKE, F., SILVA, M. G. A. J., MENEZES, W. F., et al., 2015, “Atmospheric Indicators Evaluation Using WRF Numerical Model in Rainfall Events in the City of Rio de Janeiro”, *Anuário Do Instituto De Geociências (Ufrj. Impresso)*, pp. 81–90. doi: 10.11137/2015_2_81_90.
- PRISTO, M., DERECHYNSKI, C., REGOTO, P., et al., 2018, “Climatologia de Chuvas Intensas no Município do Rio de Janeiro”, *Revista Brasileira de Meteorologia*, v. 33 (12), pp. 615–630. doi: 10.1590/0102-7786334005.
- PURDOM, J. F. W., 1976, “Some Uses of High-Resolution GOES Imagery in the Mesoscale Forecasting of Convection and Its Behavior”. .
- QUINLAN, J. R., 1993, “C4.5: Programs for Machine Learning”. In: *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers.
- ROTHFUSZ, L. P., MCLAUGHLIN, M. R., RINARD, S. K., 1998, “An Overview of NWS Weather Support for the XXVI Olympiad”. .
- SACCO, W., OLIVEIRA, C., CLAUDIO, P., 2006, “Two stochastic optimization algorithms applied to nuclear reactor core design”, *Progress in Nuclear Energy*, v. 48 (08), pp. 525–539. doi: 10.1016/j.pnucene.2005.10.004.
- SACCO, W., FILHO, H., CLAUDIO, P., 2007, “COST-BASED OPTIMIZATION OF A NUCLEAR REACTOR CORE DESIGN: A PRELIMINARY MODEL”, (01).
- SACCO, W., LAPA, C., CLAUDIO, P., et al., 2008, “A Metropolis Algorithm applied to a Nuclear Power Plant Auxiliary Feedwater System surveillance tests policy optimization”, *Progress in Nuclear Energy*, v. 50 (01), pp. 15–21. doi: 10.1016/j.pnucene.2007.09.004.
- SACCO, W. F., DE OLIVEIRA, C. R. E., 2005, “A New Stochastic Optimization Algorithm based on a Particle Collision Metaheuristic”. .
- SCHROEDER, A., STAUFFER, D., SEAMAN, N., et al., 2006, “An Automated High-Resolution, Rapidly Relocatable Meteorological Nowcasting and Prediction System”, *Monthly Weather Review - MON WEATHER REV*, v. 134 (04), pp. 1237–1265. doi: 10.1175/MWR3118.1.
- SCOFIELD, WEISS, 1977, *A report on the Chesapeake Bay region nowcasting experiment*. Relatório técnico. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf>.

- SHOWALTER, A. K., 1953, “A Stability Index for Thunderstorm Forecasting”, *Bulletin of the American Meteorological Society*, v. 34, n. 6, pp. 250–252. ISSN: 00030007, 15200477. Disponível em: <<http://www.jstor.org/stable/26242131>>.
- SILVA, W., ALBUQUERQUE, F., FRANÇA, G., et al., 2016, “Conceptual model for runway change procedure in Guarulhos International Airport based on SODAR data”, *The Aeronautical Journal*, v. 120 (05), pp. 725–734. doi: 10.1017/aer.2016.33.
- SKAMAROCK, W. C., KLEMP, J. B., DUDHIA, J., et al., 2019, *A Description of the Advanced Research WRF Version 4*. Ncar tech. note. Disponível em: <<http://dx.doi.org/10.5065/1dfh-6p97>>.
- SUN, J., XUE, M., WILSON, J., et al., 2013, “Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges”, *Bulletin of the American Meteorological Society*, v. 95 (07), pp. 409–426. doi: 10.1175/BAMS-D-11-00263.1.
- WILKS, D. S., 2011, *Statistical methods in the atmospheric sciences*. Amsterdam; Boston, Elsevier Academic Press. ISBN: 9780123850225 0123850223. Disponível em: <https://www.amazon.com/Statistical-Atmospheric-Sciences-International-Geophysics/dp/0123850223/ref=pd_bxgy_14_img_3?_encoding=UTF8&psc=1&refRID=ESPQQOR2PB1TP1VJSGCZ>.
- WILSON, J., WILK, K., 1981, “Nowcasting applications of Doppler radar (invited paper)”, v. 165 (05), pp. 123.
- WILSON, J., CROOK, N., MUELLER, C., et al., 1998, “Nowcasting Thunderstorms: A Status Report”, *Bulletin of The American Meteorological Society - BULL AMER METEOROL SOC*, v. 79 (10), pp. 2079–2100. doi: 10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2.
- WILSON, J., 1966, *Movement and Predictability of Radar Echoes*. ESSA technical memorandum NSSL. U.S. Department of Commerce, Environmental Science Services Administration, Institutes for Environmental Research, National Severe Storms Laboratory. Disponível em: <<https://books.google.com.br/books?id=FuVOPwAACAAJ>>.
- XUE, M., WANG, D., GAO, J., et al., 2003, “The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assi-

milation”, *Meteorology and Atmospheric Physics*, v. 82 (01), pp. 139–170.
doi: 10.1007/s00703-001-0595-6.

Apêndice A

Sistema NOW-BR

O projeto do Sistema Brasileiro de *Nowcasting* (NOW-BR) para a aviação tem origem em 2018 com a formalização da parceria entre a Universidade Federal do Rio de Janeiro (UFRJ) e a Universidade da Força Aérea (UNIFA) para estruturação de uma Cátedra de Meteorologia Aeronáutica com bolsas de estudo de Iniciação Científica, Mestrado, Doutorado e Pós-Doutorado com objetivo de melhoria do entendimento e previsão de fenômenos de interesse à aviação. Este projeto pode ser considerado como uma consolidação de diversos trabalhos que têm sido desenvolvidos nos últimos anos no Laboratório de Meteorologia Aplicada (LMA) no Departamento de Meteorologia da UFRJ.

O principal objetivo do NOW-BR é o preenchimento da lacuna de ferramenta, objetiva e confiável, de previsão dos eventos meteorológicos impactantes à aviação a ser utilizada pelo Departamento de Controle do Espaço Aéreo (DECEA) no gerenciamento do tráfego de aeronaves dentro das áreas terminais dos principais aeroportos do país bem como nas rotas dos voos. Estas pesquisas em curso no âmbito da Cátedra de Meteorologia Aeronáutica estão na fronteira do conhecimento, pois as questões tratadas aqui estão ainda em aberto e sendo pesquisadas atualmente pela comunidade científica internacional: aprimoramento e/ou integração de técnicas/modelos que auxiliem no *nowcasting* (previsão de curtíssimo prazo) de forma objetiva e operacional, a fim de minimizar os efeitos danosos de eventos meteorológicos na segurança e otimização da navegação aérea.

Um esquema ilustrativo do fluxo de informações do NOW-BR encontra-se na figura A.1. É possível visualizar os tipos de dados disponíveis de entrada (coleta de dados), a aplicação de modelos de diferentes tipos e a geração de previsões de curto prazo para utilização do setor operacional dos aeroportos para suporte às ações de controle do tráfego aéreo e para sua segurança. O sistema prevê como entrada uma ampla variedade de dados, como: Estações Meteorológicas de Superfície (EMS), Estações Meteorológicas de Altitude (EMA), dados de radar, dados de Previsão Numérica do Tempo, dados de satélite, dados de descargas atmosféricas, dados

de *Tropospheric Airborne Meteorological Data Reporting* (TAMDAR) (DANIELS *et al.* (2004)), dados climatológicos e dados de sensores como o *Sonic Detection and Ranging* (SODAR) e *Light Detection and Ranging* (LIDAR). Os dados de entrada são processados e integrados no banco de dados central do NOW-BR onde passam por uma série de algoritmos científicos para controle de qualidade e preparação dos dados para serem utilizados pelos modelos de previsão implementados no sistema.

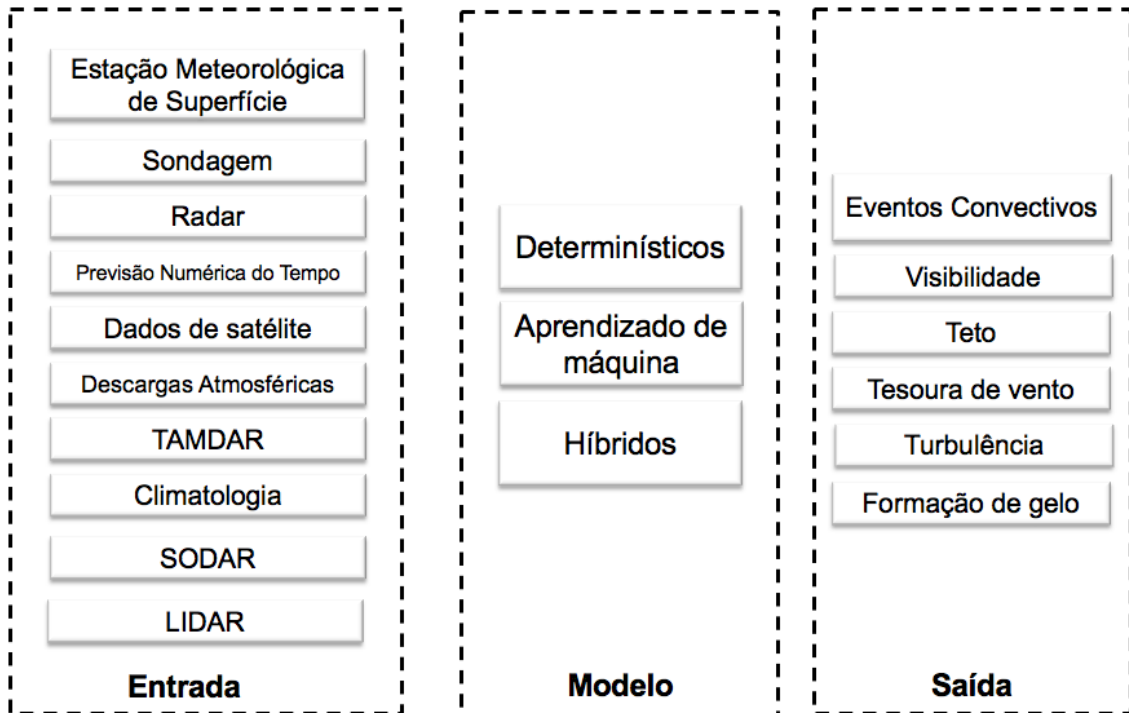


Figura A.1: Esquema ilustrativo dos tipos de dado de entrada, modelos e fenômenos contemplados pelo NOW-BR.

Entre os métodos de *nowcasting* atualmente disponíveis no sistema destacam-se: previsões de modelos atmosféricos regionais, análise multicritério das saídas dos modelos, climatologia para definição de região e horário de maior probabilidade dos fenômenos e modelos de aprendizado de máquina. Importante destacar que futuramente pretende-se incluir outras técnicas para *nowcasting*, como: previsão de ocorrência de raios, evolução dos núcleos convectivos em imagens de radar utilizando técnicas de *deep learning* e ciclos rápidos de assimilação. A distribuição das previsões será feita a partir de sistemas executados em servidores locais nos aeroportos, mas pretende-se que em uma versão futura do sistema seja construída uma interface web que possa facilitar o acesso à informação em servidores/computadores e também em dispositivos móveis (ex: *smartphones*, *tablets*).