



Departamento de
Meteorologia

Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Instituto de Geociências

**MODELO DE PREVISÃO DE TURBULÊNCIA EM VOO BASEADO EM
APRENDIZADO DE MÁQUINA PARA A ROTA SANTIAGO, CHILE –
MENDOZA, ARGENTINA**

FILIPE MENEGARDO DE SOUZA

Rio de Janeiro

Março de 2021

MODELO DE PREVISÃO DE TURBULÊNCIA EM VOO BASEADO EM
APRENDIZADO DE MÁQUINA PARA A ROTA SANTIAGO, CHILE –
MENDOZA, ARGENTINA

FILIPE MENEGARDO DE SOUZA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Meteorologia do Instituto de Geociências do Centro de Ciências Matemáticas e da Natureza da Universidade Federal do Rio de Janeiro (PPGM-IGEO-CCMN-UFRJ), como parte dos requisitos necessários à obtenção do título de Mestre em Ciências (área: Meteorologia).

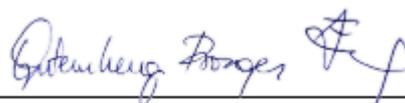
Orientadores: Gutemberg Borges França, *PhD*.
Wallace Figueiredo Menezes, *DSc*.

Rio de Janeiro
Março de 2021

MODELO DE PREVISÃO DE TURBULÊNCIA EM VOO BASEADO EM
APRENDIZADO DE MÁQUINA PARA A ROTA SANTIAGO, CHILE –
MENDOZA, ARGENTINA

FILIFE MENEGARDO DE SOUZA

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
PÓS-GRADUAÇÃO EM METEOROLOGIA DO INSTITUTO DE GEOCIÊNCIAS
DO CENTRO DE CIÊNCIAS MATEMÁTICAS E DA NATUREZA DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (PPGM-IGEO-CCMN-UFRJ)
COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA À OBTENÇÃO DO
GRAU DE MESTRE EM CIÊNCIAS (ÁREA: METEOROLOGIA).



Prof. GUTEMBERG BORGES FRANÇA, *PhD*
(orientador)



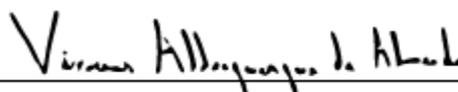
Prof. WALLACE FIGUEIREDO MENEZES, *DSc.*
(orientador)



Prof. CLAUDINE PEREIRA DERECZYNSKI, *DSc.*



Prof. GILBERTO FERNANDO FISCH, *DSc.*



VINÍCIUS ALBUQUERQUE DE ALMEIDA, *DSc.*

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2021

Menegardo-Souza, Filipe

Modelo de previsão de turbulência em voo baseado em aprendizado de máquina para a rota Santiago, Chile - Mendoza, Argentina/ Filipe Menegardo de Souza. - Rio de Janeiro: UFRJ/PPGM/IGEO/CCMN, 2021.

XXIV, 121 p.: il.; 29,7 cm.

Orientador: Gutemberg Borges França

Wallace Figueiredo Menezes

Dissertação (Mestrado) -

UFRJ/PPGM/IGEO/CCMN Programa de Pós-graduação em Meteorologia, 2021.

Referências Bibliográficas: p. 108-114.

1. CAT. 2. Andes. 3. Aprendizado de Máquina. 4. Meteorologia Aeronáutica. I. França, Gutemberg Borges *et al.* II. Universidade Federal do Rio de Janeiro, PPGM/IGEO/CCMN, Programa de Pós-graduação em Meteorologia. III. Título.

AGRADECIMENTOS

Ao orientador Dr. Gutemberg Borges França pelo direcionamento, dedicação e confiança depositada que permitiram o desenvolvimento deste trabalho.

Ao orientador Dr. Wallace Figueiredo Menezes, pelos ensinamentos e incentivos transmitidos durante o curso e ao longo da pesquisa.

Ao Departamento de Controle do Espaço Aéreo (DECEA) e ao Centro Integrado de Meteorologia Aeronáutica (CIMAER) por proporcionarem este Mestrado e acreditarem no caminho da capacitação profissional por meio da pesquisa.

À Universidade Federal do Rio de Janeiro (UFRJ) pela estrutura e oportunidade oferecida.

À LATAM Linhas Aéreas pela gentileza de disponibilizar os dados de aceleração vertical em voo de suas aeronaves.

Aos amigos do Laboratório de Meteorologia Aplicada (LMA), pelo convívio instrutivo e valiosas sugestões para a pesquisa.

À Dra. Fabiana Tenório Gomes da Silva, pela amizade, companheirismo e apoio nessa caminhada.

E agradeço a minha família, meus pais Paulo e Maria, minha irmã Michelle, pela compreensão e paciência durante os momentos em que não pude me dedicar a eles.

Resumo da Dissertação apresentada ao PPGM/IGEO/CCMN/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.).

MODELO DE PREVISÃO DE TURBULÊNCIA EM VOO BASEADO EM
APRENDIZADO DE MÁQUINA PARA A ROTA SANTIAGO, CHILE –
MENDOZA, ARGENTINA

Filipe Menegardo de Souza

Março/2021

Orientadores: Gutemberg Borges França
Wallace Figueiredo Menezes

Programa de Pós-graduação em Meteorologia

Nesta pesquisa avalia-se uma tradicional ferramenta de previsão de turbulência em voo para regiões montanhosas conhecida por Ábaco de Harrison e apresenta-se um conjunto de técnicas de aprendizado de máquina (AM) para gerar previsões de turbulência moderada ou severa (MOS) com até 24 horas de antecedência. As análises são realizadas usando medições de observações de turbulência *in situ* automatizadas da aceleração vertical em voo (VRTG) de aeronaves da LATAM, ocorridos no cruzamento da cordilheira dos Andes na rota Santiago, Chile – Mendoza, Argentina. A base de dados VRTG contém 1300 registros em 22 meses, de março de 2018 a dezembro de 2019. Na investigação foram identificados os parâmetros característicos associados aos eventos turbulentos, realizado treinamento e teste de vários algoritmos categóricos de AM e avaliados diferentes limiares aplicados ao Ábaco de Harrison. As análises indicam que 93% do total de casos turbulentos impacta os Níveis de Voo de 10000 ft a 30000 ft (*Flight Level*, FL100-300) ocorrendo de forma periódica com máximos no inverno e mínimos do verão em todos os setores do cruzamento da cordilheira. Os resultados da previsão da turbulência MOS demonstram que o

método baseado no Ábaco de Harrison possui performance limitada, com probabilidade de detecção do evento (POD_y) e do não-evento (POD_n) menores que 0,60 e *true skill statistic* (TSS) inferior a 0,25, enquanto os modelos com técnicas de AM conseguiram POD_y e POD_n de 0,80 e TSS acima de 0,60. Os modelos de previsão de turbulência MOS baseados em AM propostos apresentaram capacidade de generalização, mostrando-se ferramentas promissoras para o uso operacional, possibilitando previsões mais assertivas visando a segurança da navegação aérea.

Palavras-chave: CAT, Andes, aprendizado de máquina, meteorologia aeronáutica.

Abstract of Dissertation presented to PPGM/IGEO/CCMN/UFRJ as a partial fulfillment of requirements for the degree of Master of Science (M.Sc.).

IN-FLIGHT TURBULENCE FORECAST MODEL BASED ON MACHINE
LEARNING FOR THE SANTIAGO, CHILE – MENDOZA, ARGENTINA ROUTE

Filipe Menegardo de Souza

Março/2021

Advisors: Gutemberg Borges França
Wallace Figueiredo Menezes

Department: Meteorology

This research evaluates a traditional in-flight turbulence forecast tool for mountainous regions known as Harrison's Abacus and presents a set of machine learning techniques to generate moderate or severe turbulence 24-hour predictions. The analyzes are performed using measurements of automated *in situ* turbulence observations of the vertical acceleration in flight (VRTG) of LATAM aircraft that occurred at the Andes crossing, on the Santiago, Chile – Mendoza, Argentina route. The VRTG database contains 1,300 records in 22 months from March 2018 to December 2019. In the investigation, the characteristic parameters associated with turbulent events were identified, training and testing of several categorical machine learning algorithms were carried out and different thresholds applied to Harrison's Abacus were evaluated. The analyzes indicate that 93% of the total turbulent cases impact Flight Levels from 10000 ft to 30000 ft (FL100-300) occurring periodically with maximums in winter and minimums in summer in all sectors of the Andes crossing. The results of the moderate or severe turbulence forecast demonstrate that the Harrison's Abacus-based method has limited performance with probability of detecting the

event (POD_y) and the non-event (POD_n) below 0.60 and true skill statistic (TSS) below 0.25, while the models with machine learning techniques achieved POD_y and POD_n of 0.80 and TSS above 0.60. The proposed moderate or severe turbulence forecasting models based on machine learning presented generalization capacity, showing promising tools for operational use, allowing for more assertive forecasts aiming at the safety of air navigation.

keywords: CAT, Andes, machine learning, aviation weather.

SUMÁRIO

LISTA DE FIGURAS	XI
LISTA DE TABELAS.....	XVIII
LISTA DE ABREVIATURAS E SIGLAS.....	XX
1. INTRODUÇÃO	25
1.1 Motivação	28
1.2 Objetivo	32
1.3 Revisão Bibliográfica.....	32
2. FUNDAMENTAÇÃO TEÓRICA	36
2.1 Fontes de Turbulência	36
2.1.1 Ondas de Gravidade.....	37
2.2 Ábaco de Harrison	39
2.3 Número de Froude.....	40
2.4 Número de Richardson	42
2.5 Aprendizado de Máquina	43
3. DADOS E MÉTODO.....	47
3.1 Área de Estudo.....	47
3.2 Dados.....	49
3.3 Método	54
3.3.1 Métrica de Avaliação dos Resultados	60
4. RESULTADOS E DISCUSSÕES	62

4.1 Caracterização dos eventos turbulentos	62
4.2 Calibração e teste do Ábaco de Harrison	77
4.3 Treinamento e teste dos algoritmos de AM	78
4.4 Estudo de Caso	85
4.4.1 Caso I: 02 de março de 2018	86
4.4.2 Caso II: 06 de maio de 2018	91
4.4.3 Caso III: 28 de setembro de 2018.....	98
5. CONCLUSÕES	106
6. REFERÊNCIAS	108
APÊNDICE A - LISTA DE ATRIBUTOS.....	116
APÊNDICE B - ATRIBUTOS SELECIONADOS.....	120

LISTA DE FIGURAS

- Figura 1:** Estatística de acidentes relacionados a fatores meteorológicos, no período de 2000 a 2011, envolvendo companhias aéreas nos EUA. Fonte: traduzido de Eick (2014). 28
- Figura 2:** Carta SIGWX elaborada pelo CNMA com previsão de turbulência moderada na região dos Andes a partir do FL210 (em círculo vermelho, destaque do autor). Fonte: adaptado de REDEMET, disponível em <https://redemet.decea.gov.br/>. 31
- Figura 3:** Principais fontes de turbulência que impactam a aviação em nível de cruzeiro (FL250-450). Fonte: traduzido de Kim et al. (2018)..... 36
- Figura 4:** Representação da formação das ondas induzidas por cisalhamento (KHI). Perfil vertical do vento nas proximidades da corrente de jato (esquerda) e formação de nebulosidade num ambiente com KHI (direita). Fonte: adaptado de Wikipedia, disponível em: https://en.wikipedia.org/wiki/Kelvin%E2%80%93Helmholtz_instability 38
- Figura 5:** Representação da formação das ondas orográficas. Em tracejado retilíneo, o nível de equilíbrio da parcela numa atmosfera estável. Fonte: traduzido de Meted (2018). 38
- Figura 6:** Diagrama de intensidade da MWT sobre as Montanhas Rochosas em função da diferença de pressão entre os lados da montanha (DP) e a velocidade do vento normal ao relevo (V_n). Fonte: traduzido de Harrison (1957 apud WMO, 1973). 40
- Figura 7:** Fluxo idealizado sobre uma montanha em relação ao valor do número de Froude. Em a) $Fr= 0,1$; b) $Fr= 0,4$; c) $Fr= 1,0$; d) $Fr= 1,7$. Fonte: traduzido de Stull (1988). 41
- Figura 8:** Fluxo idealizado com $Fr=1$ formando um salto hidráulico. Fonte: traduzido de Stull (1988). 42

- Figura 9:** Área de estudo em destaque. Estão indicados os aeroportos de Santo Domingo (SCSN), Santiago (SCEL) e Mendoza (SAME). 47
- Figura 10:** Carta de rota do espaço aéreo superior. Em “A” o aeroporto de Santiago, em “B” o aeroporto de Mendoza. Em vermelho são indicadas as principais rotas para o cruzamento dos Andes. a) Mapa geral; b) rota norte; c) rota sul. Fonte: adaptado de Skyvector, disponível em <https://skyvector.com/>. 48
- Figura 11:** Esquema de início e término da Janela de Monitoramento para registro de VRTG. Fonte: traduzido de LATAM (2019) 51
- Figura 12:** Exemplo de medição de valores de VRTGMAX ou VRTGMIN na Janela de Detecção. Fonte: traduzido de LATAM (2019). 52
- Figura 13:** Fluxograma dos passos metodológicos da pesquisa. 55
- Figura 14:** Delimitação do retângulo de interesse (retângulo em vermelho) na latitude de 33,5°S, entre as longitudes de 68°W e 71°W e entre os níveis verticais de 700hPa 200hPa, representando o estado da atmosfera na área de estudo. 59
- Figura 15:** Configurações dos experimentos de treinamento e teste: a) modelo objetivo com Vn/DP (Ábaco de Harrison); b) algoritmos de AM. Fonte: adaptado de Parmezan, Lee e Chung (2012). 59
- Figura 16:** Esquema de um diagrama ROC. “D” resultado não-desejável, “B” resultado aleatório, “A” resultado desejável, “C” melhor resultado. A classificação perfeita fica no canto superior esquerdo. Fonte: adaptado de Wikipedia, disponível em https://pt.wikipedia.org/wiki/Característica_de_Operação_do_Receptor. 60
- Figura 17:** Tabela de Contingência 2 x 2. Verdadeiro positivo (VP); verdadeiro negativo (VN); falso positivo (FP); falso negativo (FN)..... 61
- Figura 18:** Posição dos 1300 registros de VRTG acima do FL100, de março de 2018 a dezembro de 2019, indicados de acordo com sua intensidade. Classe 1 (verde), Classe 2 (amarelo), Classe 3 (vermelho). Em círculo preto estão indicados os aeroportos de SCSN, SCEL e SAME. 62

Figura 19: Nível de voo versus registros de VRTG por classe.	63
Figura 20: Registros horários de VRTG por classe no período de março de 2018 a dezembro de 2019.	64
Figura 21: Registros horários de VRTG versus intensidade, no período de março de 2018 a dezembro 2019.	64
Figura 22: Total de pousos e decolagens da LATAM em SCEL ao longo do ano de 2019. A linha tracejada representa a média mensal de 7212 pousos e decolagens.....	66
Figura 23: Percentual dos registros horários de VRTG versus intensidade, no período de março de 2018 a dezembro de 2019, ponderado pelo total de voos mensal da LATAM em SCEL.	66
Figura 24: Média horária dos registros de VRTG por classe de turbulência no período de março de 2018 a dezembro de 2019. Horário local de Santiago (h) (Z-4).	68
Figura 25: Média horária de voos da LATAM em SCEL por dia, utilizando como referência os movimentos dos dias 28 de fevereiro de 2020 e 11 de março de 2020. Em tracejado, a média de 11 voos. Horário local de Santiago (h) (Z-4).	68
Figura 26: Média horária percentual dos registros de VRTG por classe de turbulência no período de março de 2018 a dezembro de 2019, ponderados pela média horária de voos da LATAM em SCEL. Horário local de Santiago (h) (Z-4).	69
Figura 27: Média mensal e desvio-padrão da componente zonal da velocidade do vento no FL180 (V_n) para os casos de VRTG.	70
Figura 28: Total dos registros horários de VRTG por classe de intensidade versus V_n	71
Figura 29: Percentual do total dos registros horários de VRTG por classe de intensidade versus V_n , ponderado pela incidência de V_n	72

Figura 30: Total dos registros horários de VRTG por classe de intensidade versus DP..... 73

Figura 31: Percentual do total dos registros horários de VRTG por classe de intensidade versus DP, ponderado pela incidência de DP. 73

Figura 32: Registros horários de VRTG de todas as classes versus Vn (ordenada em kt) e DP (abscissa em hPa). Os valores representam a quantidade de dados horários de turbulência para cada par de Vn e DP, identificados com o intervalo de cores estabelecido de acordo com a quantidade de registros de VRTG..... 74

Figura 33: Percentual dos registros horários de VRTG de todas as classes versus Vn (ordenada em kt) e DP (abscissa em hPa), ponderado pelas incidências de Vn e DP. Os valores representam a porcentagem de vezes que houve turbulência para cada par de Vn e DP, identificados com o intervalo de cores estabelecido. Em cinza significa que em nenhum momento no período de estudo foi atingido o valor correspondente ao parâmetro dicotômico..... 75

Figura 34: Registros horários de VRTG classes 2 e 3 versus Vn (ordenada em kt) e DP (abscissa em hPa). Os valores representam a quantidade de dados horários de turbulência MOS para cada par de Vn e DP, identificados com o intervalo de cores estabelecido de acordo com a quantidade de registros de VRTG..... 75

Figura 35: Percentual dos registros horários de VRTG classes 2 e 3 versus Vn (ordenada em kt) e DP (abscissa em hPa), ponderado pelas incidências de Vn e DP. Os valores representam a porcentagem de vezes que houve turbulência MOS para cada par de Vn e DP, identificados com o intervalo de cores estabelecido. Em cinza significa que em nenhum momento no período de estudo foi atingido o valor correspondente ao parâmetro dicotômico..... 76

Figura 36: Resultado dos experimentos de calibração e teste com o Ábaco de Harrison. Curva ROC de PODy versus 1-PODn e tabelas com os valores de PODy, PODn e TSS. a) curva ROC da calibração; b) curva ROC do teste; c) PODy, PODn e TSS da calibração; d) PODy, PODn e TSS do teste. Cada índice é um par com diferentes Vn e DP. O índice 1 é referente aos valores originais do

Ábaco de Harrison. Em realce verde o índice 5, que apresentou os melhores resultados. Valores de Vn em kt e DP em hPa..... 77

Figura 37: Curva ROC com valores de PODy versus 1-PODn dos resultados do treinamento dos algoritmos de AM com os dados originais. 79

Figura 38: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM balanceando artificialmente os dados com o ClassBalancer nas proporções de evento e não-evento: a) 50/50; b) 60/40; c) 70/30; d) 80/20; e) 90/10; f) 96/04..... 81

Figura 39: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM via matriz de custo, atribuindo pesos ao FN de: a) 10; b) 20; c) 30; d) 50; e) 100; f) 200; g) 380..... 82

Figura 40: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM balanceando os dados com a retirada aleatória de dados “NÃO”. RandomForest 2FP é o modelo utilizando matriz de custo com “peso” 2 atribuído ao Falso Positivo (FP)..... 83

Figura 41: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM balanceando os dados com a inserção aleatória de dados “SIM” baseados: a) média e um desvio-padrão de cada atributo; b) média e dois desvios-padrões de cada atributo; c) mediana e quartis inferior e superior de cada atributo. 84

Figura 42: Imagem GOES-16, canal 13 (banda termal) das 18Z do dia 02 de março de 2018. No retângulo preto em destaque, a área de estudo. Fonte: adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>. 87

Figura 43: Representação do estado básico da atmosfera no dia 02 de março de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt), temperatura potencial (K) em isolinhas vermelhas e velocidade vertical (m/s) em

hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z, e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque. 89

Figura 44: Representação do estado básico da atmosfera no dia 02 de março de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt) e cisalhamento vertical do vento ($10^{-3} s^{-1}$) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque. 91

Figura 45: Imagem GOES-16, canal 13 (banda termal das 18Z do dia 06 de maio de 2018. No retângulo preto em destaque, a área de estudo. Fonte: adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>. 93

Figura 46: Representação do estado básico da atmosfera no dia 06 de maio de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt), temperatura potencial (K) em isolinhas vermelhas e velocidade vertical (m/s) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z, e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque. 95

Figura 47: Representação do estado básico da atmosfera no dia 06 de maio de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt) e cisalhamento vertical do vento ($10^{-3} s^{-1}$) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque. 97

Figura 48: Imagem GOES-16, canal 13 (banda termal) das 18Z do dia 28 de setembro de 2018. No retângulo preto em destaque, a área de estudo. Fonte: adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>. 100

Figura 49: Representação do estado básico da atmosfera no dia 28 de setembro de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt),

temperatura potencial (K) em isolinhas vermelhas e velocidade vertical (m/s) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque..... 102

Figura 50: Representação do estado básico da atmosfera no dia 28 de setembro de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt) e cisalhamento vertical do vento ($10^{-3}s^{-1}$) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque. 104

LISTA DE TABELAS

Tabela 1: Acidentes aéreos em que a causa principal foi a turbulência em voo. Fonte: adaptado de Eick (2014).	29
Tabela 2: Conjunto de dados utilizado na pesquisa no período de estudo de março de 2018 a dezembro de 2019.	50
Tabela 3: A intensidade do VRTG em função da variação da aceleração vertical em voo em termos da aceleração gravitacional (g). (adaptado de LATAM, 2019).	50
Tabela 4: Resumo com os melhores resultados obtidos durante o treinamento dos algoritmos de AM.	85
Tabela 5: Registros horários de turbulência MOS para o dia 02 de março de 2028. Dados a cada 3h de acordo com as saídas do GFS.	86
Tabela 6: Resultados dos modelos preditivos ótimos de turbulência MOS para o dia 02 de março de 2018 em comparação com os dados observados. Em cinza ausência da turbulência, em vermelho presença.	91
Tabela 7: Registros horários de turbulência MOS para o dia 06 de maio de 2018. Dados a cada 3h de acordo com as saídas do GFS.	92
Tabela 8: Resultados dos modelos preditivos ótimos de turbulência MOS para o dia 06 de maio de 2018 em comparação com os dados observados. Em cinza ausência da turbulência, em vermelho presença.	98
Tabela 9: Registros horários de turbulência MOS para o dia 28 de setembro de 2018. Dados a cada 3h de acordo com as saídas do GFS.	99
Tabela 10: Resultados dos modelos preditivos ótimos de turbulência MOS para o dia 28 de setembro de 2018 em comparação com os dados observados. Em cinza ausência da turbulência, em vermelho presença.	105
Tabela 11: Parâmetros atmosféricos extraídos do retângulo de interesse.	116

Tabela 12: Lista de atributos selecionados para cada um dos conjuntos de dados de treinamento e teste dos algoritmos de AM.....	120
--	-----

LISTA DE ABREVIATURAS E SIGLAS

AIREP	Reporte em voo de aeronave (padrão OACI)
AIRMET	Informe meteorológico em rota para baixos níveis
AZUL	AZUL Linhas Aéreas Brasileiras
BIAS	Tendência de uma previsão
Brown	Diagnóstico de turbulência Brown
CAT	Turbulência de céu claro
CENIPA	Centro de Investigação e Prevenção de Acidentes Aeronáuticos
CIMAER	Centro Integrado de Meteorologia Aeronáutica
CIT	Turbulência induzida por convecção
CFS	<i>Correlation-based feature selection</i>
CMI	Centro Meteorológico Integrado
CNMA	Centro Nacional de Meteorologia Aeronáutica
CPTEC	Centro de Previsão do Tempo e Estudos Climáticos
DAL	<i>Delta Airlines</i>
DECEA	Departamento de Controle do Espaço Aéreo
desvpad_w	Desvio-padrão da velocidade vertical
DP	Diferença de pressão entre barlavento e sotavento
DW	Diferença entre o maior e o menor valores de w
ECT	Energia cinética turbulenta
EDR	Algoritmo estimador em voo de intensidade da turbulência
ETA	Modelo regional de previsão do tempo
EUA	Estados Unidos da América
Excel	Editor de planilhas produzido pela Microsoft

FAA	Administração Federal de Aviação dos EUA
FAB	Força Aérea Brasileira
FAR	Taxa de falso alarme
FIR	Região de Informação de Voo
FL	Nível de voo
FN	Falso Negativo
FP	Falso Positivo
Fr	Número de Froude
ft	Pés
g	Aceleração gravitacional
GAMET	Previsão de tempo significativo em baixos níveis
GFS	<i>Global Forecast System</i>
GOL	GOL Linhas Aéreas
GOES	<i>Geostationary Operational Environmental Satellite</i>
GTG	Guia gráfico de turbulência
GRADS	Sistema de análise e exibição de grade
H	Altura da montanha
hPa	Hectopascal
KHI	Instabilidade Kelvin-Helmholtz
kt	Nó
Km	Quilômetro
LATAM	LATAM Linhas Aéreas Brasil
m1desvpad	Intervalo considerando a média e um desvio-padrão
m2desvpad	Intervalo considerando a média e dois desvios-padrões
mediana	Intervalo considerando a mediana e quartis
METAR	Informe meteorológico regular de aeródromo

MOS	Moderada ou severa
MWT	Turbulência orográfica
N	Frequência Brunt-Vaisala
NCAR	Centro Nacional de Pesquisas Atmosféricas dos EUA
NCEP	Centros Nacionais de Previsão Ambiental dos EUA
NCT	Turbulência próxima de nuvens
NTSB	Conselho Nacional de Segurança nos Transportes dos EUA
OACI	Organização de Aviação Civil Internacional
ONU	Organização das Nações Unidas
PIREP	Reporte em voo de aeronave (padrão FAA)
PNT	Previsão Numérica de Tempo
POD _n	Probabilidade de detecção do não-evento
POD _y	Probabilidade de detecção do evento
QNH	Pressão atmosférica reduzida ao nível médio do mar
Ri _B	Richardson Bulk
Ri _B máx	Maior valor do Ri _B
Ri _B médio	Média do Ri _B
Ri _B mín	Menor valor do Ri _B
REDEMET	Portal Meteorológico do Comando da Aeronáutica
Ri	Número de Richardson
ROC	Curva Característica de Operação do Receptor
SAME	Aeroporto Internacional Governador Francisco Gabrielli
SAMF	Região de informação de voo de Mendoza
SBBS	Região de informação de voo de Brasília
SBCW	Região de informação de voo de Curitiba
SCEL	Aeroporto Internacional Arturo Merino Benítez

SCSN	Aeroporto de Santo Domingo, Chile
SIGMET	Informe meteorológico em rota
SIGWX	Tempo significativo
SISCEAB	Sistema de Controle do Espaço Aéreo Brasileiro
SPECI	Informe meteorológico especial de aeródromo
TEMP	Informe meteorológico em altitude, de estação terrestre
TI1	Algoritmo de diagnóstico de turbulência Ellrod-Knapp
TSS	<i>True skill statistic</i>
θ	Temperatura potencial
u	Velocidade do vento zonal
$u_{\text{máx}}$	Maior valor de u
UTC	Tempo Universal Coordenado
UAL	<i>United Airlines</i>
v	Velocidade do vento meridional
VN	Verdadeiro Negativo
V_n	Vento normal ao relevo no FL180
VP	Verdadeiro Positivo
VRTG	Aceleração vertical em voo
VRTGMAX	Maior valor do VRTG
VRTGMIN	Menor valor do VRTG
vws	Cisalhamento vertical do vento
vws máx	Maior valor do vws
vws médio	Média do vws
vws mín	Menor valor do vws
w	Velocidade vertical
$w_{\text{máx}}$	Maior valor de w

wmédio	Média de w
wmín	Menor valor de w
WAFC	Centro de previsão mundial de área
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
WRF	<i>Weather Research and Forecasting</i>
x	Dimensão horizontal
z	Dimensão vertical
Z	Hora Zulu

1. INTRODUÇÃO

A atmosfera é um fluido em movimento capaz de desenvolver distúrbios que podem ter tamanhos que variam desde centímetros até centenas de quilômetros apresentando-se com certa suavidade ou até mesmo na forma de turbulência severa.

Os voos realizados em áreas com turbulência podem ser perigosos constituindo uma das maiores preocupações para a aviação (SHARMAN *et al.*, 2012; SHARMAN & PEARSON, 2017; STORER, WILLIAMS & GILL, 2019; GULTEPE *et al.*, 2019), pois nem sempre o fenômeno está associado a um sinal visual e não pode ser detectado por sensores a bordo das aeronaves. Isso é um grande complicador, pois exige dos pilotos atenção redobrada em voo quanto aos primeiros sinais da entrada em áreas de risco, alertando a tripulação e passageiros quanto a necessidade de permanecerem em seus assentos e com o cinto de segurança afivelado. Como essa ação é extremamente subjetiva e rápida, por muitas vezes não se tem o tempo hábil necessário para o alerta. Súbitas oscilações de velocidade e altitude neste tipo de ambiente podem levar a danos estruturais nas aeronaves e, mais comumente, incidentes com passageiros e tripulantes (SHARMAN *et al.*, 2006; WOLFF & SHARMAN, 2008; SHARMAN *et al.*, 2014).

O Conselho Nacional de Segurança nos Transportes (*National Transportation Safety Board* - NTSB) registra que, por ano nos Estados Unidos da América (EUA), aproximadamente 58 pessoas sofrem lesões devido à turbulência (FAA, 2019a). Entre o período de 1980 a 2008, ocorreram 234 acidentes em que o fator principal foi a turbulência, resultando em 298 ferimentos graves e 3 fatalidades. Dos feridos, 184 envolveram comissários de bordo e 114 passageiros. De acordo com Sharman e Lane (2016) essas ocorrências geram um custo para as empresas aéreas dos EUA de duas formas: 1) por meio de indenizações pagas, equivalentes a mais de US\$ 10 milhões ao ano, e 2) por dias de afastamentos de trabalho da tripulação, que é superior a 7000 dias a cada ano. Outro custo significativo advém da impossibilidade das aeronaves realizarem o voo na sua rota ideal, levando a um maior consumo de combustível. De acordo com Search Technology (2000 *apud* STORER, WILLIAMS & GILL, 2019), estima-se que, em todo o mundo, isso afete 5% do total de voos, causando um gasto de até US\$ 16 milhões ao ano, no somatório das companhias aéreas.

Existem três fontes principais de turbulência que afetam a aviação (MARLTON, 2016; KIM *et al.*, 2018): 1) instabilidade de cisalhamento, tipicamente em regiões desprovidas de nebulosidade significativa, associada ao forte cisalhamento vertical e horizontal que ocorre nas proximidades das correntes de jato (*Clear Air Turbulence* – CAT) (DUTTON & PANOFSKY, 1970; WOLFF & SHARMAN, 2008; KIM & CHUN, 2010; STORER, WILLIAMS & GILL, 2019); 2) quebra das ondas orográficas na presença de relevo (DURRAN, 1990; DOYLE *et al.*, 2005; SHARMAN *et al.*, 2012) relacionado a propagação vertical da energia das ondas de gravidade em um ambiente estável (STULL, 1988), carregando *momentum* para a alta troposfera, proporcionando a quebra da onda em níveis de cruzeiro, acima e a sotavento da montanha (*Mountain Wave Turbulence* – MWT) (CARNEY *et al.*, 1995; TURNER, 1999; KIM *et al.*, 2018); e 3) quebra das ondas de gravidade induzidas por intensa convecção, categorizada dentro e fora da nuvem, dependendo de sua localização (*Convectively Induced Turbulence* – CIT). Quando fora da convecção, este evento é definido como *Near-Cloud Turbulence* (NCT) (KIM *et al.*, 2018; SHARMAN & TRIER, 2019).

Na América do Sul, a Cordilheira dos Andes atua como um obstáculo ao escoamento dos ventos predominantes de oeste em latitudes subtropicais e extratropicais (SILVA, 2004; MENEGARDO-SOUZA, 2018). Trata-se de um conjunto contínuo de montanhas que se estende na direção Norte-Sul por mais de 7000km, com largura variando de 200 a 700 km e altitude média de 4500 m (NORTE, 2015). Esses ventos, por muitas vezes associados às correntes de jato, ao interagir com a topografia provocam o surgimento de perturbações que podem chegar até a tropopausa e atingir milhares de quilômetros à jusante da montanha. Por esse motivo, companhias aéreas de todo o mundo possuem procedimentos operacionais específicos para o cruzamento dos Andes, sendo a condição de turbulência vital para o planejamento do voo, influenciando inclusive na escolha da rota a ser utilizada (MENEGARDO-SOUZA, 2018).

Os centros mundiais de previsão de área (*World Area Forecast Center* – WAFC), junto com os Centros Nacionais de Meteorologia Aeronáutica (CNMA), têm a responsabilidade de realizar as previsões de tempo significativo (SIGWX), incluindo turbulência, para os provedores do serviço de navegação aérea, companhias aéreas e Organização de Aviação Civil Internacional (OACI). Isso é feito por meio das cartas

SIGWX, em que as áreas de risco são indicadas de acordo com sua severidade e nível de voo. O WAFC confecciona a carta SIGWX de altos níveis (FL250/630 – *Flight Level*; Nível de Voo entre 25000 ft e 63000 ft) enquanto o CNMA a carta SIGWX dos níveis baixos (Superfície até FL250). Além disso, os serviços meteorológicos nacionais são responsáveis pela confecção de códigos meteorológicos específicos para as Regiões de Informação de Voo (*Flight Information Region* - FIR) sob sua jurisdição. Esses códigos (SIGMET, AIRMET e GAMET) são mensagens relativas à ocorrência ou previsão de determinados fenômenos que possam afetar a segurança das operações aéreas, incluindo turbulência em seus diferentes tipos. No Brasil, o serviço de meteorologia aeronáutica está todo centralizado no Centro Meteorológico Integrado (CMI), setor operacional pertencente ao Centro Integrado de Meteorologia Aeronáutica (CIMAER, <https://www2.fab.mil.br/cimaer/>), o qual é subordinado diretamente ao Departamento de Controle do Espaço Aéreo (DECEA, <https://www.decea.mil.br/>).

Por meio das imagens de satélite pode-se detectar a presença de nuvens lenticulares, rotoras e/ou outros padrões de nebulosidade típicos da presença da turbulência. Porém, nem sempre haverá umidade suficiente para a formação de nuvens, necessitando, assim, que o Meteorologista se utilize de outras ferramentas para sua análise, tradicionalmente, reporte de pilotos (AIREP/PIREP) e diagnósticos, que se constituem de métodos objetivos ou algoritmos de pós-processamento derivados das saídas dos modelos de previsão numérica do tempo (PNT). Esses diagnósticos podem ser de natureza empírica ou com base em suposições abrangendo aspectos físicos da atmosfera, que são tratados em parte ou sofrem simplificações, estando por isso, sujeitos a deficiências naturais da técnica empregada (HON, NG & CHAN, 2020).

Importante o fato que o clima está mudando e essa alteração pode ter impacto na turbulência no futuro (STORER, WILLIAMS & GILL, 2019). Pesquisas como as de Price e Rind (1994); Collins *et al.* (2013); Williams e Joshi (2013); Storer, Williams e Joshi (2017) corroboram com essa afirmação e sugerem que a turbulência aumentará em frequência e intensidade com as mudanças climáticas, com acréscimos de mais de 100% nas latitudes médias, ao redor das correntes de jato.

1.1 Motivação

De acordo com o levantamento realizado pela NTSB (Figura 1), para o período de 2000 a 2011, mais de 70% dos acidentes relacionados a fatores meteorológicos, envolvendo companhias aéreas nos EUA, tiveram como causa a turbulência em voo. Esse fenômeno causou mais ferimentos graves aos passageiros e tripulantes do que qualquer outro (EICK, 2014), sendo que a maior parte deles acontece acima do FL100, quando a aeronave está em nível de cruzeiro¹ e normalmente sem o aviso de atar o cinto de segurança (SHARMAN *et al.*, 2006).

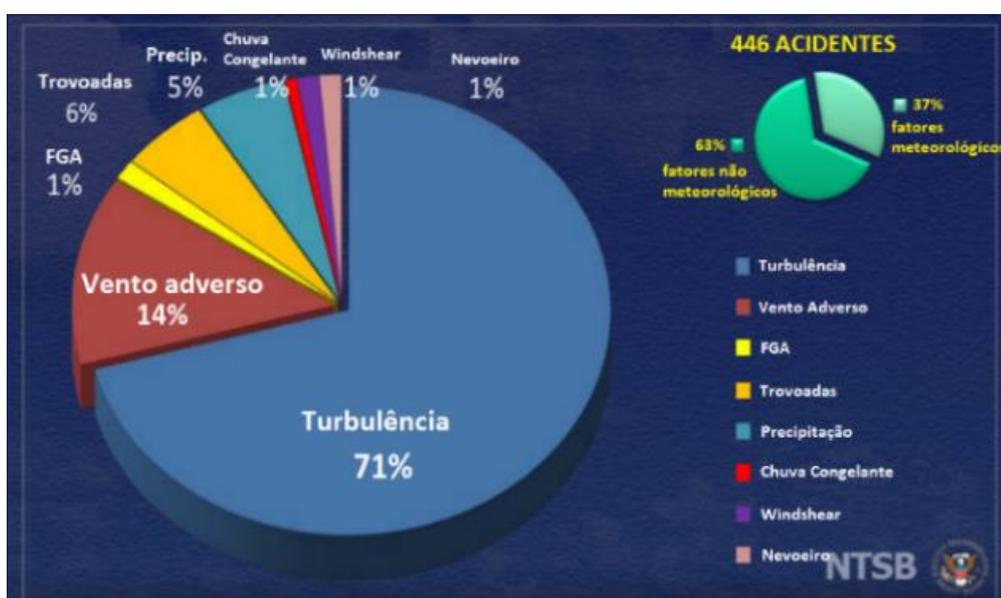


Figura 1: Estatística de acidentes relacionados a fatores meteorológicos, no período de 2000 a 2011, envolvendo companhias aéreas nos EUA. Fonte: traduzido de Eick (2014).

Apesar da turbulência raramente causar mortes, contudo como pode ser visto na Tabela 1, acidentes fatais já ocorreram. Em um estudo realizado por Eichenbaum (2000 *apud* WILLIAMS, 2014), identificou-se que cerca de 35% dos acidentes que tiveram como causa a turbulência, verificados pela NTSB, não tinham um código AIRMET ou SIGMET válidos para o horário e local do acidente e, por tanto, teriam sido inesperados pelas tripulações.

¹ Nível de voo que se mantém durante uma etapa considerável da viagem em virtude de, geralmente, a operação da aeronave ser favorecida com velocidade ótima e menor consumo de combustível (DECEA, 2021)

Tabela 1: Acidentes aéreos em que a causa principal foi a turbulência em voo. Fonte: adaptado de Eick (2014).

Óbitos	Aeronave	Local	Tipo	Data
1	B747	Pacífico	CAT	28 de dezembro de 1997
1	DC-9	Flórida	CIT	03 de outubro de 1990
17	F-28	Países Baixos	CIT	07 de outubro de 1981
2	B737	Índia	CIT	10 de maio de 1980
39	F-27	Alasca	MWT	02 de dezembro de 1968
85	L-188	Texas	CIT	03 de maio de 1968
124	B707	Japão	MWT	03 de março de 1966

Todas essas ocorrências, além do próprio custo humano, geram perdas financeiras da ordem de milhões de dólares ao ano para as companhias aéreas em forma de: indenizações, dispensa da tripulação lesionada e em maiores gastos de combustível com desvios de rota. Aliado a isso, as previsões de que as mudanças climáticas provoquem o aumento na frequência e intensidade das turbulências, principalmente nas regiões de latitudes médias (STORER, WILLIAMS & GILL, 2019; STORER, WILLIAMS & JOSHI, 2017), poderão impactar ainda mais as estatísticas de acidentes e os custos operacionais das companhias aéreas.

Todo planejamento de voo requer cuidados, porém aquele a ser realizado passando por regiões montanhosas necessita de atenção especial, visto que, o relevo tem a capacidade de penetrar profundamente na atmosfera (abaixo da altitude do topo do Monte Everest, 8848 m, está 65% de todo o ar da atmosfera), influenciando o tempo e o clima na Terra, ocasionado distúrbios nos ventos, alteração da distribuição da precipitação e formando ondas de gravidade que podem atingir até altos níveis (SMITH, 2019). Na região da rota Santiago (Chile) – Mendoza (Argentina) a Cordilheira dos Andes possui altitude média de 4500 m e atinge o ponto mais alto das Américas com 6962 m no Aconcágua. Fortes ventos associados às correntes de jato tornam a região, um local típico de intensas turbulências, necessitando para a

realização segura do voo, o conhecimento prévio da existência do fenômeno (MENEGARDO-SOUZA, 2018).

Pesquisando sobre a formação de turbulência em áreas montanhosas, Harrison (1957 *apud* WMO, 1973) desenvolveu um método objetivo de prognóstico de MWT para a região das Montanhas Rochosas, nos EUA (denominado de Ábaco de Harrison), relacionando os reportes dos pilotos (PIREP) com a máxima velocidade do vento perpendicular ao relevo abaixo de 20000 ft (6096 metros) e a diferença de pressão de cada lado da montanha (o método será apresentado e explicado com maiores detalhes no Capítulo 2, Seção 2.2). Apesar do diagrama ter sido construído inicialmente para a América do Norte, ele é amplamente utilizado por companhias aéreas de todo o mundo para o planejamento de voo no cruzamento dos Andes. Porém, não foi encontrado na literatura nenhum estudo específico sobre a validade e eficácia do método para a América do Sul.

O Brasil é membro da OACI, pertencente a um seleto grupo de apenas 11 países que representam o sistema da aviação mundial, possuindo capacidade de influência nos principais temas da aviação civil internacional (DECEA, 2020), servindo de padrão para o mundo no que se refere a prestação dos serviços de navegação aérea. O serviço de meteorologia aeronáutica no âmbito do Sistema de Controle do Espaço Aéreo Brasileiro (SISCEAB) é executado por intermédio do CIMAER, sendo este responsável pela confecção de prognósticos e gerenciamentos para a divulgação de informações meteorológicas com fins aeronáuticos tais como os códigos GAMET, AIRMET e SIGMET. O CIMAER também tem por atribuição elaborar e divulgar cartas SIGWX, válidas para 00Z, 06Z, 12Z e 18Z, da superfície até o FL250 na área compreendida entre as latitudes 12°N 40°S e longitudes 40°W 80°W (Figura 2), o que abrange quase a totalidade da América do Sul, incluindo a região da rota Santiago (Chile) – Mendoza (Argentina). As funções do CNMA foram absorvidas pelo CIMAER no dia 3 de novembro de 2019.

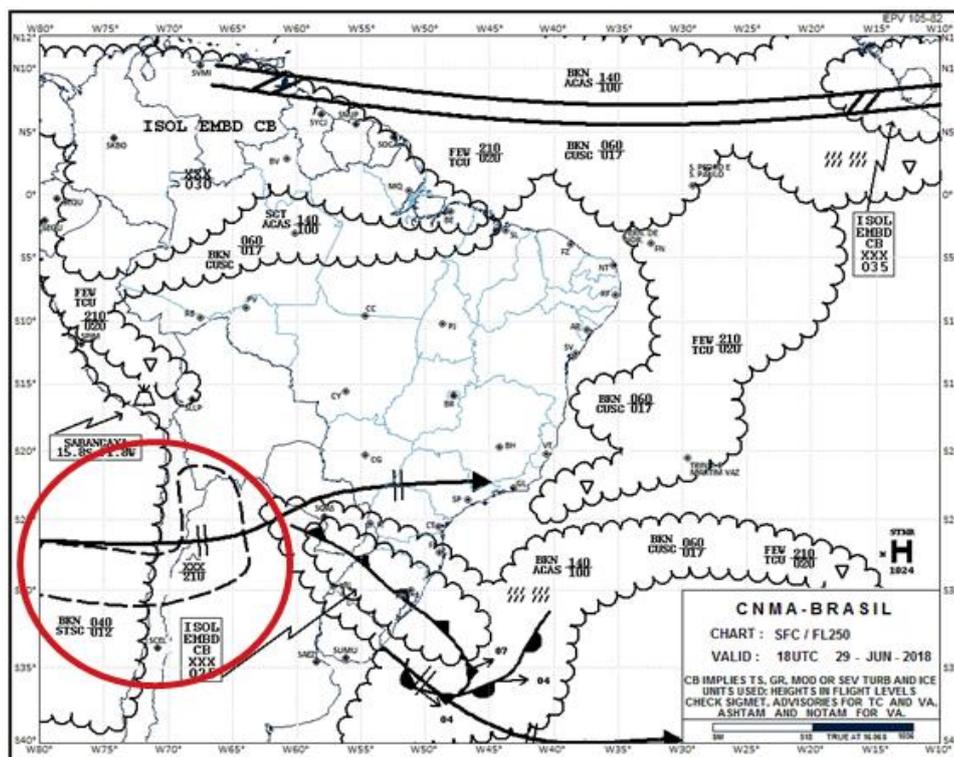


Figura 2: Carta SIGWX elaborada pelo CNMA com previsão de turbulência moderada na região dos Andes a partir do FL210 (em círculo vermelho, destaque do autor). Fonte: adaptado de REDEMET, disponível em <https://redemet.decea.gov.br/>.

Porém, conforme identificado por França, Almeida e Rosette (2016), a previsão do tempo no Brasil ainda é realizada pelos meteorologistas de forma subjetiva, integrando as diversas informações disponíveis em suas análises, para conseguirem estabelecer um modelo meteorológico conceitual. E, mesmo diante da importância do fenômeno, os poucos estudos relacionados ao tema na América do Sul, e em particular nos Andes, estão voltados aos processos de geração, formação e previsão do vento zonda² ou relacionando-os com as ondas de montanha.

Dessa forma, esta pesquisa além de contribuir para reduzir a grande lacuna nos estudos de turbulência atmosférica na América do Sul, procura atender as necessidades do serviço de navegação aérea, que diante do crescente fluxo aéreo (OACI, 2016) e da perspectiva do aumento da severidade e dos casos de turbulência por causa das mudanças climáticas (STORER, WILLIAMS & GILL, 2019), precisa de aprimoramento das ferramentas de apoio à tomada de decisão do meteorologista,

² O zonda é um vento forte, quente e muito seco associado à compressão adiabática na descida leste dos Andes no centro-oeste da Argentina, com maior frequência no inverno e primavera do Hemisfério Sul (Norte, 2015).

proporcionando a padronização de procedimentos, diminuindo a subjetividade e promovendo maior assertividade das previsões. Isso poderá levar a um melhor planejamento de voo, com rotas mais seguras, reduzindo os riscos de perdas humanas e os prejuízos financeiros.

1.2 Objetivo

Com este trabalho tem-se por objetivo principal desenvolver um modelo de previsão de turbulência moderada ou severa (MOS) para a rota Santiago, Chile – Mendoza, Argentina, com pós-processamento dos dados do *Global Forecast System* (GFS) utilizando aprendizado de máquina (AM).

Para atingir o objetivo principal, foram delimitados os seguintes objetivos específicos:

- 1: Identificar os parâmetros característicos associados aos casos de turbulência na rota Santiago, Chile – Mendoza, Argentina entre março de 2018 a dezembro de 2019;
- 2: Avaliar o Ábaco de Harrison como ferramenta objetiva para a previsão de turbulência MOS; e
- 3: Avaliar o uso de algoritmos de AM como modelos de previsão de turbulência MOS;

1.3 Revisão Bibliográfica

A validação da previsão e diagnóstico de turbulência é difícil, porque os conjuntos de dados objetivos de verificação são limitados (STORER, WILLIAMS & GILL, 2019). Tradicionalmente, as únicas observações rotineiras são aquelas fornecidas verbalmente pelos pilotos, conhecidas como PIREP, nos EUA, e como AIREP internacionalmente (SHARMAN *et al.*, 2014). O uso desses reportes em voo tem uma limitação, pois eles podem conter um erro substancial de posicionamento (latitude, longitude e nível de voo) e tempo (horário) em que ocorreu a turbulência, além de serem dependentes da subjetividade da tripulação e do tipo de aeronave (SHARMAN *et al.*, 2006). Para resolver essas deficiências, um algoritmo de notificação de turbulência *in situ* baseado em EDR (*Eddy Dissipation Rate*) vem sendo implementado em algumas companhias aéreas comerciais (SHARMAN *et al.*, 2014). Em janeiro de 2014, 200 aeronaves pertencentes as companhias aéreas *United Airlines* (UAL) e *Delta Airlines* (DAL) eram equipadas com esse sistema (SHARMAN

et al., 2014). Atualmente nenhuma das companhias aéreas do Brasil implementou o algoritmo, porém as empresas LATAM, GOL e AZUL realizam monitoramento da aceleração vertical sofrida pela aeronave (VRTG) em função da variação da aceleração gravitacional, com registros de latitude, longitude, nível de voo, horário e intensidade da turbulência.

Num dos primeiros estudos climatológicos, Harrison (1957 *apud* WMO, 1973) relacionou os dados de PIREP com a maior velocidade do vento normal (V_n – perpendicular ao relevo) abaixo de 20000 ft e a diferença de pressão entre os dois lados da montanha (Montanha Rochosas, EUA), servindo como método objetivo de previsão de MWT. Wolff e Sharman (2008) fizeram uma climatologia das turbulências entre FL180-600 nos EUA utilizando 12 anos de dados de PIREP. Eles concluíram que a MWT é uma importante fonte de turbulência na metade ocidental dos EUA, sendo sua ocorrência favorecida por fortes ventos de oeste, que são mais comuns no inverno, com direção dentro de 30° com a perpendicular da orientação da cordilheira e por relevos acima de 1500 m.

Na América do Sul, e em particular nos Andes, os trabalhos estão voltados aos processos de geração, formação e previsão do vento zonda ou relacionando-os com as ondas da montanha, tais como Silva (2004), Norte (2015), Vásquez e Falcón (2015), sendo que neste último os autores identificaram para os Andes Venezuelanos, que quando o fluxo de ar perpendicular à cordilheira for maior que 10 m/s, em condições de estabilidade, umidade superior a 70% até o cume e o número de Froude for ligeiramente maior que 1, o resultado é a presença de ondas orográficas associadas a nuvens lenticulares. Utilizando como base de dados os códigos SIGMET emitidos para a região de informação de voo de Mendoza (FIR-SAMF) no período de 2016 a 2017, Menegardo-Souza (2018) avaliou os resultados do *Ábaco* de Harrison como método objetivo de previsão de turbulência para a região da cordilheira dos Andes. Identificou-se que o *Ábaco* poderia ser um bom indicativo do fenômeno, apesar da baixa densidade de informação disponível no estudo. Uma climatologia de CAT usando dados de AIREP entre os anos de 2003 a 2014 nas regiões de informação de voo de Curitiba e Brasília (FIR-SBCW e FIR-SBBS) foi feito por Mello (2015) no Brasil. O autor verificou que a maior parte dos eventos (67,9%) ocorreu entre os meses de maio a outubro e que do total de turbulências, apenas de 2,7% são severas.

As perturbações atmosféricas podem ocorrer em diferentes escalas, mas apenas as da ordem de grandeza horizontal das aeronaves (aproximadamente 100 m) afetam a aviação. Para essa dimensão, não é possível prever direta e rotineiramente o movimento atmosférico (SHARMAN *et al.*, 2006), porém a maior parte da energia associada aos eventos turbulentos em escalas menores é proveniente dos eventos de grande escala (DUTTON & PANOFSKY, 1970). Como as maiores dimensões podem ser resolvidas pelos modelos de PNT, o problema da previsão de turbulência é o de identificar as características de grande escala que são propícias aos movimentos turbulentos em escala que afete as aeronaves (STORER, WILLIAMS & GILL, 2019). Dessa forma, algoritmos de pós-processamento aplicados à saída dos modelos de PNT são comumente utilizados para inferir regiões de turbulência significativa. Essas inferências são baseadas em diagnósticos derivados tipicamente de gradientes espaciais de variáveis do modelo (GULTEPE *et al.*, 2019).

Vários diagnósticos de pós-processamento foram propostos, e alguns são usados operacionalmente (GULTEPE *et al.*, 2019). Dutton e Panofsky (1970) verificaram que as ondas gravitacionais podem desestabilizar a atmosfera, reduzindo o número de Richardson (Ri) abaixo de 0,25, levando a instabilidade Kelvin-Helmholtz (KHI) e gerando ondas que quebram e formam turbulência. Brown (1973) desenvolveu o indicador de turbulência (Brown) baseado na taxa de mudança de Ri usando uma abordagem simplificada. Ellrod e Knapp (1992) desenvolveram um algoritmo (TI1) que usa as saídas dos campos de velocidade para derivar um diagnóstico baseado na deformação horizontal e no cisalhamento vertical do vento. Ellrod e Knox (2010) sugeriram uma mudança no TI1, adicionando o termo de tendência de divergência. Turner (1999) propôs um algoritmo para previsão de MWT a ser utilizado no WAFC de Londres. De acordo com o autor, os resultados mostraram-se promissores, indicando o seu uso como uma ferramenta de *first guess* naquele centro para análise do fenômeno. Em um estudo realizado no Brasil, Lyra, Chan e Dereczynski (2007) utilizaram os algoritmos Ri , TI1 e Brown no pós-processamento das saídas do modelo ETA de PNT para inferir as regiões de turbulência para dois dias selecionados. De acordo com os autores, os dados concordaram com as informações obtidas de cartas SIGWX. Kim e Chun (2010) analisaram nove casos de PIREP de CAT moderada/severa na Coreia do Sul utilizando cálculos de Ri e energia cinética turbulenta (*Turbulence Kinetic Energy* - TKE) no pós-processamento do modelo

Weather Research and Forecasting (WRF) de PNT com diferentes resoluções espaciais. Os autores encontraram que para a maioria dos casos, uma resolução de 10 km era adequada para capturar os eventos de CAT, exceto àqueles de origem orográfica (MWT).

Como cada diagnóstico individual é incapaz de prever todos os eventos de turbulência, Sharman *et al.* (2006) geraram o *Graphical Turbulence Guidance* (GTG). O produto utiliza *ensemble* de muitos diagnósticos, o que de acordo com os autores, parece melhorar o desempenho estatístico das previsões. No novo sistema GTG3, Sharman e Pearson (2017) usaram uma combinação de ventos de baixos níveis com diagnósticos de CAT produzindo 14 novos diagnósticos de MWT, utilizando como base os dados de PIREP e EDR.

A aplicação das técnicas de AM em estudos de turbulência em voo é relativamente recente. Williams (2014) produziu um prognóstico de *nowcasting* de CIT, baseado em diversas fontes de dados utilizando o algoritmo de AM denominado de florestas aleatórias (*random forest*). Muñoz-Esparza, Sharman e Deierling (2020) exploraram o uso de árvores de decisão no pós-processamento de modelo de PNT com alta resolução para prever turbulência acima de 20000 ft. Nos dois trabalhos, os autores fizeram uso dos dados de EDR para os EUA. Hon, Ng e Chan (2020) desenvolveram previsões de turbulência para a região da Ásia-Pacífico com técnicas de AM combinando uma coleção de diagnósticos de turbulência produzidos com o pós-processamento do modelo de PNT. De acordo com os autores, foram obtidos ganhos de performance de 3% a 17% em comparação com os índices de turbulência quando considerados individualmente.

2. FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo, apresenta-se de forma sucinta a teoria sobre turbulência, o Ábaco de Harrison, o número de Froude, o número de Richardson e os princípios do AM enfatizando os tipos de classificadores utilizados nos modelos preditivos deste trabalho.

2.1 Fontes de Turbulência

A turbulência atmosférica relevante para as aeronaves em nível de cruzeiro é categorizada em três diferentes tipos (Figura 3) baseado na localização e nos mecanismos de formação (MARLTON, 2016; KIM *et al.*, 2018): CAT, CIT e MWT.

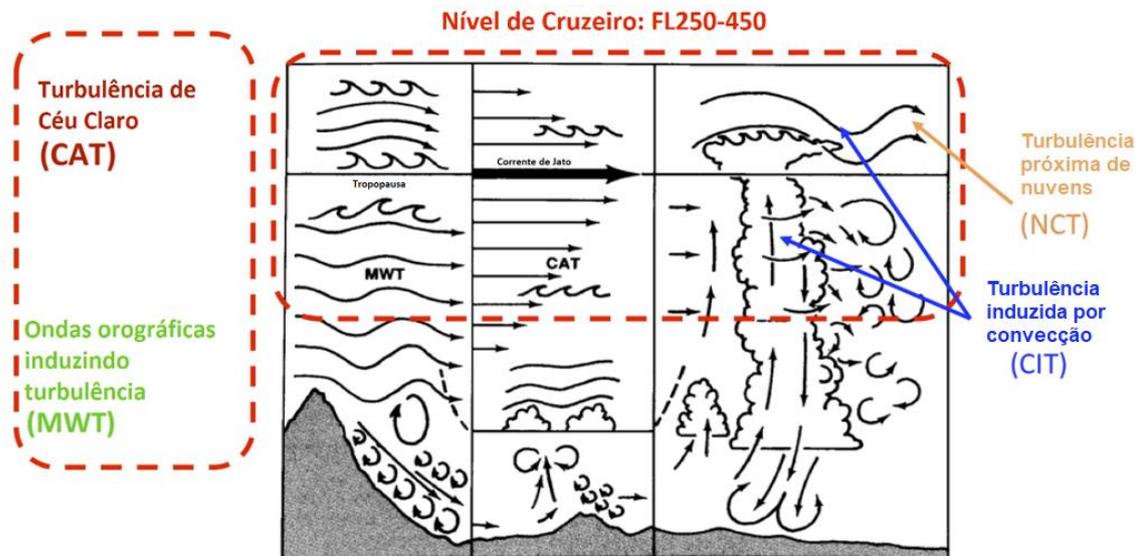


Figura 3: Principais fontes de turbulência que impactam a aviação em nível de cruzeiro (FL250-450). Fonte: traduzido de Kim *et al.* (2018)

A CAT é uma turbulência de altos níveis em regiões desprovidas de nebulosidade significativa, próxima da corrente de jato e de sistemas frontais. Longe de montanhas, ela é resultante do forte cisalhamento vertical do vento que favorece a instabilidade Kelvin-Helmholtz (KHI). (DUTTON & PANOFSKY, 1970; WOLFF & SHARMAN, 2008; KIM & CHUN, 2010; STORER, WILLIAMS & GILL, 2019).

Além da própria turbulência que ocorre dentro das nuvens convectivas (CIT) as intensas correntes de ar presentes podem formar fluxos que se propagam para além da nebulosidade, dando origem a ondas de gravidade. Essas ondas podem se quebrar formando a NCT (KIM *et al.*, 2018; SHARMAN & TRIER, 2019). De acordo

com Kaplan *et al.* (2005), cerca de 86% dos casos de turbulência severa examinados nos EUA ocorreram dentro de uma distância de até 100 km de uma convecção profunda. Isso ratifica uma das recomendações aos pilotos que é o afastamento lateral em pelo menos 20 milhas (32,19 km) e não passar por cima de tempestades (FAA, 2019b).

Ondas orográficas podem se formar quando o vento próximo à superfície passa por regiões montanhosas (DURRAN, 1990; DOYLE *et al.*, 2005; SHARMAN *et al.*, 2012). À medida que a onda se propaga verticalmente, sua amplitude aumenta devido à diminuição da densidade do ar, o que pode causar a quebra formando a MWT, acima e à sotavento da montanha (CARNEY *et al.*, 1995; TURNER, 1999; KIM *et al.*, 2018). As condições de propagação e quebra estão fortemente relacionadas ao escoamento e aos perfis de estabilidade (STULL, 1988; DOYLE *et al.*, 2005). A MWT está associada aos intensos ventos de oeste perpendiculares ao relevo verificados nas latitudes médias (STORER, WILLIAMS & GILL, 2019).

2.1.1 Ondas de Gravidade

Muitos distúrbios atmosféricos ocorrem como eventos periódicos, isto é, são ondas, com comprimento, período, velocidade de fase e amplitude (HOLTON, 2004). Enquanto as ondas atmosféricas em grande escala exibem principalmente movimento horizontal, nas ondas de menor escala a proporção do movimento vertical para o movimento horizontal é muito maior (CARNEY *et al.*, 1995). As ondas mais importantes que exibem essa propriedade são as ondas de gravidade, assim chamadas porque a força restauradora é a gravidade, e as ondas induzidas por cisalhamento (KHI).

Na KHI a quebra das ondas ocorre quando o cisalhamento vertical do vento excede um valor crítico (DUTTON & PANOFSKY, 1970). As ondas ficam cada vez mais íngremes até que o topo da onda ultrapasse a parte inferior. O fluxo transforma-se em turbulento, invertendo a estratificação preexistente, misturando ar frio com ar mais quente abaixo (Figura 4).

Dessa forma, as montanhas podem gerar ondas de gravidade (ondas orográficas) de grande amplitude que se propagam e quebram quando um valor crítico de cisalhamento vertical do vento é atingido (NASTROM & FRITTS, 1992), produzindo sérios riscos ao voo. Quanto mais alto ocorrer a propagação da onda de gravidade, maior a probabilidade dela se romper como consequência do aumento do cisalhamento do vento e da menor densidade do ar (STORER, WILLIAMS & GILL, 2019). As características dessas ondas, como amplitude e comprimento de onda, são altamente dependentes da topografia, fluxo de baixos níveis e estrutura vertical atmosférica (KIM *et al.*, 2018).

A frequência de flutuação dessas ondas, N (frequência Brunt-Vaisala), é definida em termos do gradiente vertical da temperatura potencial:

$$N^2 = \frac{g\partial\theta}{\theta\partial z}, \quad (01)$$

onde z é a dimensão vertical, g é a aceleração gravitacional, θ a temperatura potencial.

Esse parâmetro é uma medida da força de restauração da flutuabilidade quando uma parcela de ar é retirada de seu ponto de equilíbrio. Os valores típicos para N são $0,01 \text{ s}^{-1}$ na troposfera ou $0,02 \text{ s}^{-1}$ na estratosfera (SMITH, 2019).

2.2 Ábaco de Harrison

O Ábaco de Harrison é um método objetivo de prognóstico de MWT baseado em climatologia de PIREP, desenvolvido por Harrison (1957 *apud* WMO, 1973) para a região das Montanhas Rochosas, nos EUA. Na Figura 6 é apresentado o diagrama que indica a variação da intensidade da turbulência, considerando a diferença de pressão (hPa) entre os lados da montanha (DP) versus a velocidade do vento normal ao relevo (V_n) (nós e m/s).

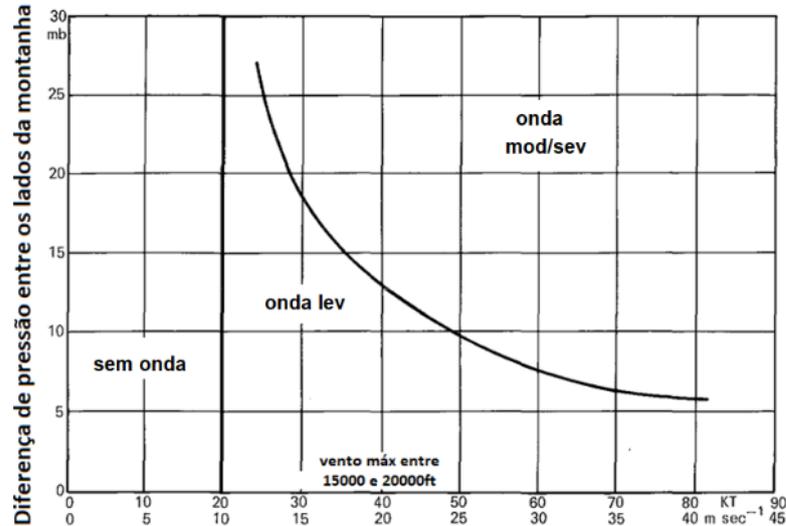


Figura 6: Diagrama de intensidade da MWT sobre as Montanhas Rochosas em função da diferença de pressão entre os lados da montanha (DP) e a velocidade do vento normal ao relevo (Vn). Fonte: traduzido de Harrison (1957 *apud* WMO, 1973).

Nesse diagrama assume-se que Vn é positivo, isto é, o vento flui de oeste e a DP é maior ou igual zero (a pressão à barlavento da montanha é maior ou igual a pressão à sotavento). De acordo com o Ábaco, para que haja turbulência, Vn deveria ser de pelo menos 20 kt e DP maior ou igual a zero.

É importante ressaltar que, apesar do diagrama ter sido construído inicialmente para as Montanhas Rochosas nos EUA, ele é amplamente utilizado por companhias aéreas para o planejamento de voo no cruzamento dos Andes. Para tal utiliza-se como Vn e DP, o vento normal em relação a cordilheira no FL180 à barlavento e a diferença da pressão reduzida ao nível médio do mar (QNH) extraídas do código METAR entre os aeroportos de Santiago (SCEL) e Mendoza (SAME).

2.3 Número de Froude

Uma questão muito importante para a previsão de MWT é saber se o ar fluirá ou não sobre uma montanha. A interação entre a estabilidade atmosférica, o fluxo de ar e a topografia pode gerar um amplo espectro de ondas de diferentes características. O número de Froude (Fr) relaciona esses três fatores num único parâmetro.

$$Fr = \frac{U}{NH}, \quad (02)$$

onde U é a velocidade do escoamento médio perpendicular ao relevo, N é a frequência Brunt-Vaisala e H a altura da montanha.

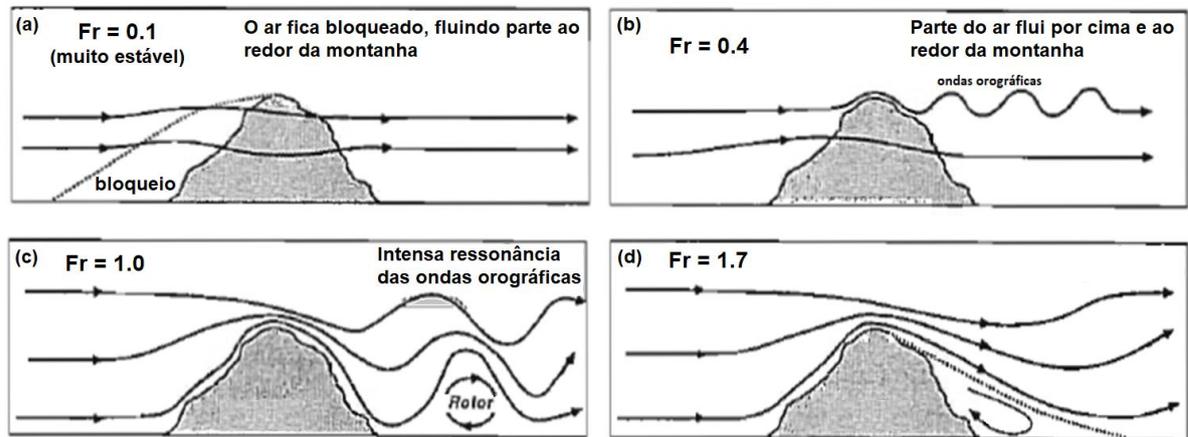


Figura 7: Fluxo idealizado sobre uma montanha em relação ao valor do número de Froude. Em a) $Fr = 0,1$; b) $Fr = 0,4$; c) $Fr = 1,0$; d) $Fr = 1,7$. Fonte: traduzido de Stull (1988).

De acordo com a Figura 7 tem-se:

- a) Froude = 0,1: situação de vento fraco e forte estabilidade. O escoamento é bloqueado e torna-se estagnado.
- b) Froude = 0,4: vento moderado e fraca estabilidade. Parte do escoamento percorre a montanha e o restante sobre ela. O ar no topo é perturbado ocorrendo a formação de ondas orográficas.
- c) Froude = 1,0: vento forte e fraca estabilidade. As ondas orográficas tornam-se mais intensas. Nessa situação poderá ocorrer o salto hidráulico (Figura 8) descrito por Stull (1988) e Durran (1990), que favorece a formação de fortes ventos catabáticos e turbulência severa em baixos níveis à sotavento.
- d) Froude = 1,7: vento intenso e fraca estabilidade. O ar fluirá livremente sobre a montanha sem formação de oscilações significativas. Cria-se uma região de cavidade com direção do vento contrária em superfície.

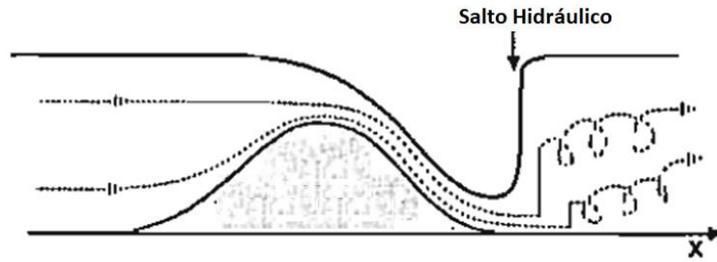


Figura 8: Fluxo idealizado com $Fr=1$ formando um salto hidráulico. Fonte: traduzido de Stull (1988).

Apesar de considerar a situação de uma montanha isolada, essa teoria é bastante consagrada e fornece os conceitos para descrever o comportamento e estrutura das ondas orográficas (Silva, 2004).

Smith (1989) utiliza a velocidade do vento perpendicular e a frequência Brunt-Vaisala constantes à montante. Porém a variabilidade de N e U com a altitude, dificulta a escolha de valores representativos (REINECKE & DURRAN, 2008). Para empregar a teoria, a estrutura atmosférica real deve ser simplificada estimando-se um efetivo vento de velocidade constante (U) e aproximando-se o perfil de estabilidade real como constante com a altura. Dois métodos de aproximação da estabilidade estática constante têm aparecido com frequência na literatura. O método da média, em que a frequência de Brunt-Vaisala é calculada pela média entre o solo e uma certa altura para produzir um N constante e o método de *bulk*, onde a mudança total da temperatura potencial entre o solo e uma certa altura é usada para definir N (REINECKE & DURRAN, 2008).

2.4 Número de Richardson

Em um ambiente estaticamente estável, movimentos verticais turbulentos agem contra a força restauradora da gravidade. Nesse ambiente, a flutuabilidade tende a suprimir a turbulência, enquanto o cisalhamento do vento tende a gerá-la (SILVA, 2004).

O escoamento em uma atmosfera estaticamente estável torna-se turbulento devido a magnitude do cisalhamento do vento iniciando a KHI. Para entender por que o cisalhamento causa turbulência, define-se o número de Richardson:

$$Ri = \frac{N^2}{(\partial U / \partial z)^2} = \frac{(g/\theta) \cdot (\partial \theta / \partial z)}{(\partial u / \partial z)^2 + (\partial v / \partial z)^2}, \quad (03)$$

onde N é a frequência Brunt-Vaisala, U é a velocidade do vento, z é a altura, g é a aceleração gravitacional e θ é a temperatura potencial.

O número de Richardson é adimensional, com o numerador representando a estratificação e o denominador representando o cisalhamento vertical do vento. Resulta de considerações teóricas que a instabilidade ocorre quando $Ri < 0,25$ favorecida pelo grande cisalhamento vertical do vento (denominador) e estratificação fraca (numerador) (SILVA, 2004).

Como raramente se conhece os gradientes precisos de velocidade do vento, normalmente são utilizadas aproximações calculadas a partir de medidas de velocidade em algumas alturas. Aproximando-se os gradientes pelas correspondentes diferenças, da equação (03) obtém-se o Richardson Bulk (Ri_B):

$$Ri_B = \frac{g\Delta\theta\Delta z}{\theta[(\Delta u)^2 + (\Delta v)^2]}, \quad (04)$$

2.5 Aprendizado de Máquina

O aprendizado de máquina (AM) é uma área da inteligência artificial que investiga como desenvolver sistemas capazes de adquirir conhecimento com base em dados, de forma a automaticamente reconhecer padrões complexos e com capacidade de melhorar com a experiência (HAN, KAMBER & PEI, 2011; BARNES, 2015).

A aquisição de conhecimento ocorre a partir do treinamento de algoritmos com dados pretéritos e que pode ser realizado de duas formas (BROWNLEE, 2016): 1) aprendizagem supervisionada, atribuindo variáveis de entrada e saída, “ x ” e “ y ”, o algoritmo aprende a mapear “ x ” em “ y ” a partir de exemplos “ x ” de um conjunto de treinamento em que os valores de “ y ” são conhecidos. Cada exemplo apresentado ao algoritmo de aprendizado possui um atributo especial que identifica o rótulo da classe à qual esse exemplo pertence (ROSSI, 2009); 2) aprendizagem não supervisionada, tem-se dados “ x ” e o objetivo é identificar de forma autônoma estruturas ou propriedades interessantes nos dados. O algoritmo de aprendizado não tem conhecimento dos rótulos das classes reais, dessa maneira agrupa os exemplos por meio de extração de padrões de seus atributos (ROSSI, 2009).

Um ponto importante dos algoritmos de AM é construir modelos (algoritmos treinados) que possuam boa capacidade de generalização, conseguindo prever com alta taxa de acerto, rótulos de classes para exemplos que não foram apresentados anteriormente (DOMINGOS, 2012; TAN *et al.*, 2018). Se os rótulos das classes forem discretos, o problema é conhecido como classificação, por outro lado, se forem contínuos, como regressão (HAN, KAMBER & PEI, 2011). As classes podem ser binárias ou multi-classes.

Depois do *overfitting*³, o maior problema no AM é a maldição da dimensionalidade (DOMINGOS, 2012). Como naturalmente com o passar do tempo, o conjunto de dados a ser analisado torna-se cada vez maior, não somente na quantidade de registros, mas também no de atributos, isso faz aumentar o custo computacional do tratamento das informações. Pode-se pensar que reunir mais recursos fornecerá uma melhor caracterização, mas na verdade seus benefícios podem ser superados pelos malefícios da maldição (DOMINGOS, 2012). O problema é que a quantidade de atributos impacta exponencialmente no número de dados para alcançar o conhecimento desejado, fazendo com que a partir de um determinado valor ocorra a degradação do desempenho do algoritmo. Num grande conjunto de atributos, alguns são ruins e outros são correlacionados o que atrapalha no aprendizado, sendo, portanto, necessário selecionar aqueles que são mais relevantes, de forma que se possa atingir a máxima performance do classificador (PARMEZAN, LEE & CHUNG, 2012).

Como por vezes os registros de um conjunto de dados apresentam a maioria de suas ocorrências voltadas a uma categoria específica, isso faz com que o algoritmo de classificação fique enviesado pendendo para a classe majoritária. Para evitar esse problema, é necessário a realização da correção de prevalência, equilibrando estatisticamente a ocorrência das classes (SOARES, 2007). Os métodos que implementam essa estratégia de correção são: 1) replicação ou eliminação aleatória de registros, de forma que o novo conjunto seja representativo dos dados originais; 2) matriz de custo, atribuindo-se um “peso” ao erro associado ao registro cuja classe seja menos numerosa (SOARES, 2007); 3) *ClassBalancer* (disponível na plataforma

³ *Overfitting* ou sobreajuste é um termo usado para descrever quando um modelo se ajusta muito bem ao conjunto de dados pretéritos, mas se mostra ineficaz para prever novos resultados.

WEKA), avalia as instâncias nos dados ajustando o “peso” de cada classe em relação ao conjunto total.

Para esta pesquisa, foram aplicados os classificadores disponibilizados na plataforma WEKA (*Waikato Environment for Knowledge Analysis*) (HALL *et al.*, 2009). O WEKA é um software livre desenvolvido pela Universidade de Waikato, Nova Zelândia que contém várias ferramentas integradas para tarefas de AM, sendo amplamente utilizado para aplicações de ensino e pesquisas. A seguir são apresentados os 5 tipos de classificadores que foram abordados neste trabalho, possuindo execuções distintas na tarefa de classificação.

a) **Árvore de Decisão:** consiste em uma classificação baseada em algoritmos de particionamento sucessivo, decorrente do princípio de dividir para conquistar (BOTELHO DE SOUZA, 2011). Possui uma estrutura semelhante a um fluxograma, sendo caracterizado pela presença de nós, ramos e folhas que resultam da separação de registros com características diferentes (HAN, KAMBER & PEI, 2011; TAN *et al.*, 2018). A partir da informação contida nos dados, cada nó denota um teste de atributo, cada ramificação representa um resultado do teste e cada folha contém um rótulo de classe.

b) **Bayesiano:** classificador probabilístico baseado no teorema de Bayes para a construção de modelos fundamentados em um princípio estatístico que combina o conhecimento prévio das classes com novas evidências colhidas dos dados (RUSSELL & NORVIG, 2009; HAN, KAMBER & PEI, 2011). Como frequentemente uma informação é apresentada na forma de probabilidade condicional, isto é, a probabilidade de ocorrer um evento B, na condição de que outro evento A já tenha ocorrido, o teorema de Bayes relaciona as probabilidades de A e B com suas respectivas probabilidades condicionadas. Assim, classificam um evento numa determinada classe, baseando-se na probabilidade desse evento pertencer a essa classe (BOTELHO DE SOUZA, 2011);

c) **Regra de Decisão:** é a técnica para classificar um evento baseada em um conjunto de regras “SE-ENTÃO” (HAN, KAMBER & PEI, 2011). Uma regra “R” cobre um dado “x” se a pré-condição de “R” corresponder aos atributos de “x”. A ordenação das regras pode ser implementada regra a regra ou classe a classe. O esquema de

ordenação baseado em regras garante que cada registro de teste seja classificado pela melhor regra que o cobrir. Na ordenação em classes, as regras que pertencem a mesma classe aparecem juntas no conjunto de regras e desde que uma das regras seja disparada, a classe será atribuída ao registro teste (BOTELHO DE SOUZA, 2011);

d) Redes Neurais Artificiais: são sistemas paralelos e distribuídos compostos por unidades de processamento (neurônios) que calculam determinadas funções matemáticas. Os neurônios estão dispostos em uma ou mais camadas e são interligados por um grande número de conexões que estão associados a pesos, os quais armazenam o conhecimento representado no modelo, ponderando a entrada recebida em cada neurônio da rede (MORAIS, 2010). O sinal de entrada segue para frente (*feedforward*) através da rede, camada por camada até chegar à saída, que retro propaga no sentido inverso (*backpropagation*) para que seja feito o ajuste dos pesos de todas as camadas de modo a minimizar o erro na camada de saída.

e) Regressão logística: consiste em um modelo linear generalizado, cuja classificação resultante é a estimativa da probabilidade de um evento ocorrer em função de um conjunto de variáveis preditoras. Resulta em uma classificação dicotômica em que os resultados de probabilidade ficam contidos no intervalo de zero a um (MONTGOMERY, PECK & VINING, 2012).

3. DADOS E MÉTODO

Neste Capítulo, apresenta-se a área de estudo com suas peculiaridades e importância, o conjunto de dados utilizados e a metodologia empregada nesta pesquisa.

3.1 Área de Estudo

A área de estudo é a região delimitada pelas latitudes 32,5°S e 34,5°S e longitudes 068°W e 071°W (polígono destacado na Figura 9), englobando a região dos Andes entre os aeroportos de Santiago (Aeroporto Internacional Arturo Merino Benítez - SCEL), e Mendoza (Aeroporto Internacional Governador Francisco Gabrielli - SAME), onde estão localizadas as principais aerovias para o cruzamento da cordilheira.

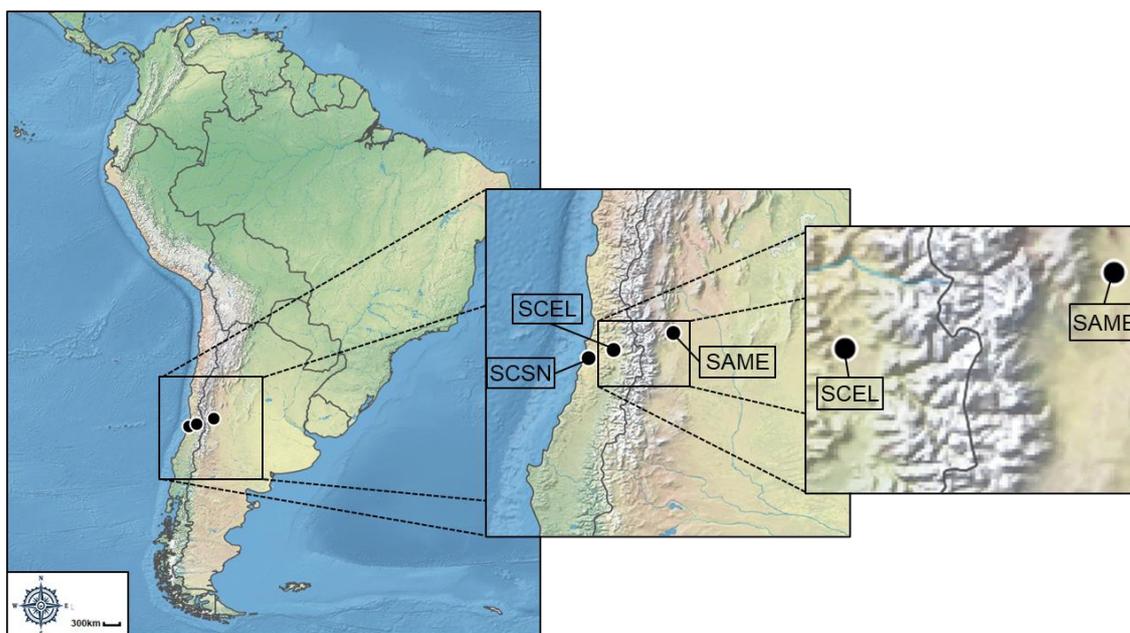


Figura 9: Área de estudo em destaque. Estão indicados os aeroportos de Santo Domingo (SCSN), Santiago (SCEL) e Mendoza (SAME).

Na Figura 10 apresenta-se a carta de rota do espaço aéreo superior com indicação das aerovias para o cruzamento da cordilheira dos Andes. Os destaques “A” e “B” representam as localizações de SCEL e SAME, respectivamente. A região é passagem preferencial das aeronaves que operam em SCEL (principal aeroporto da região de estudo, sendo o maior e mais movimentado do Chile) e têm como destino ou são provenientes de leste como Brasil, Argentina, Paraguai, Uruguai e países da Europa. De acordo com as condições de intensidade da turbulência, no planejamento

de voo são selecionadas as rotas mais propícias para a segurança, normalmente a rota norte (Figura 10b) ou a rota sul (Figura 10c).

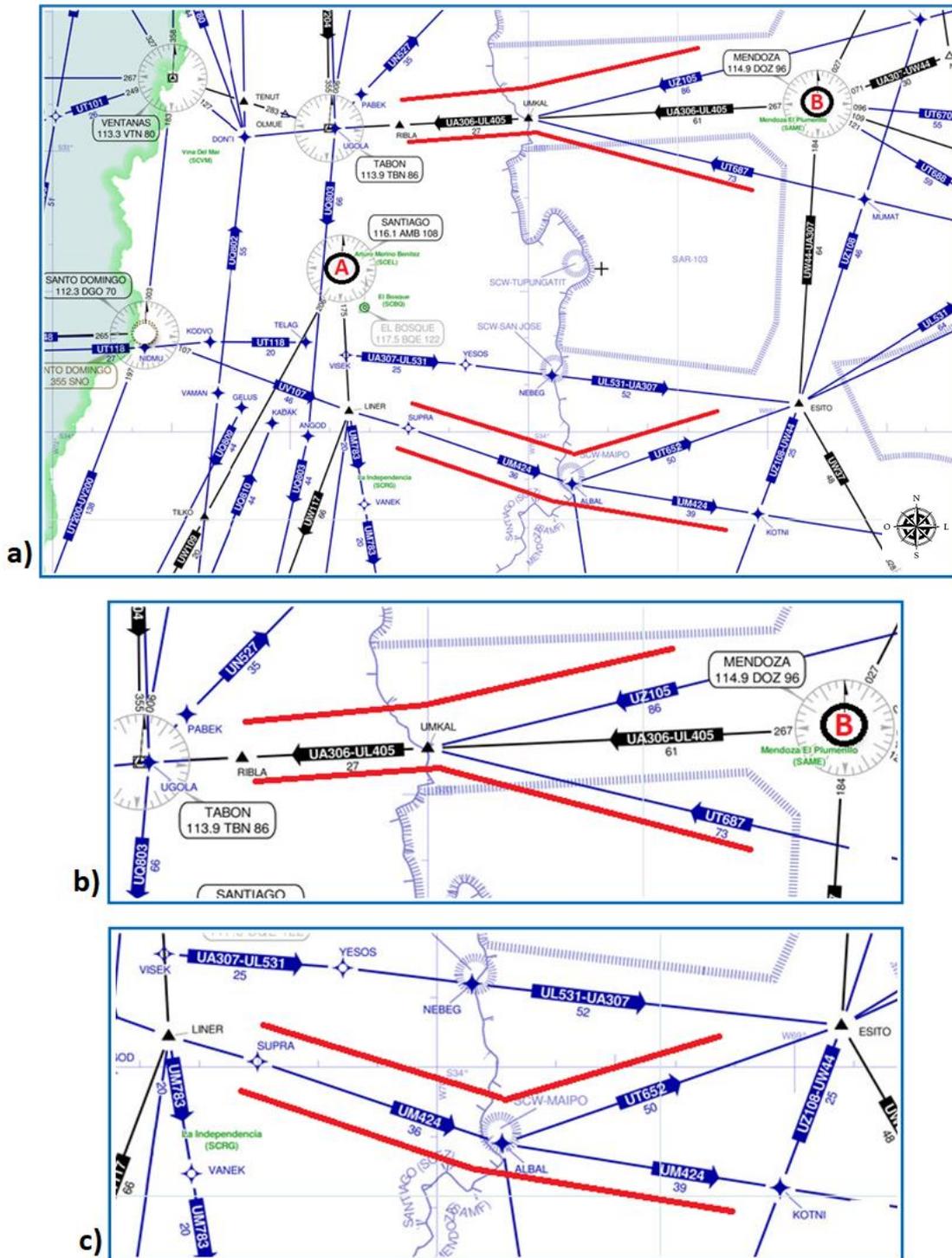


Figura 10: Carta de rota do espaço aéreo superior. Em “A” o aeroporto de Santiago, em “B” o aeroporto de Mendoza. Em vermelho são indicadas as principais rotas para o cruzamento dos Andes. a) Mapa geral; b) rota norte; c) rota sul. Fonte: adaptado de Skyvector, disponível em <https://skyvector.com/>.

O posicionamento meridional da Cordilheira dos Andes acaba sendo um obstáculo para o escoamento dos ventos predominantes de oeste das médias e altas latitudes, possuindo para isto elevações acima de 4500 m e chegando a 6962 m no Aconcágua. Esses ventos, influenciados pelas correntes de jato, interagem com o relevo gerando distúrbios que se propagam à sotavento e podem atingir até altos níveis da troposfera, causando intensas turbulências.

São comuns as ocorrências de turbulência em voo na região provocando situações de grande risco para a segurança da navegação aérea. No dia 15 de agosto de 2011, uma aeronave da Força Aérea Brasileira (FAB) enfrentou turbulência severa próxima a posição aeronáutica denominada de “UMKAL” (pode ser vista na Figura 10b), quando realizando aproximação para SCEL. Diante do acontecido o Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA) emitiu um comunicado com diversas recomendações de segurança quando do planejamento e realização de voo cruzando a cordilheira dos Andes (CENIPA, 2011). No dia 21 de julho de 2019 a aeronave com a delegação do clube paulista de futebol Palmeiras, em um voo direto de Fortaleza (Brasil) para Mendoza (Argentina), foi acometida por intensa turbulência quando aproximando para o aeroporto de Mendoza. O piloto após duas arremetidas em SAME, precisou desviar a rota e seguir para Buenos Aires (Argentina) (<https://veja.abril.com.br/placar/voo-do-palmeiras-a-argentina-arremete-duas-vezes-e-pouso-muda-de-lugar/>).

3.2 Dados

A Tabela 2 mostra os detalhes inerentes ao conjunto de dados usado (ou seja, tipo, frequência, informações específicas e fonte).

Tabela 2: Conjunto de dados utilizado na pesquisa no período de estudo de março de 2018 a dezembro de 2019.

Tipo	Frequência	Informação	Fonte
VRTG	variável	Aceleração vertical em voo (g)	LATAM Airlines Brazil
N° de voos	horário e mensal	Total de pousos e decolagens da LATAM em SCEL	LATAM Airlines Brazil
METAR	1h	Pressão reduzida ao nível médio do mar (QNH) de SCEL e SAME	REDEMET (https://redemet.decea.gov.br)
TEMP	12h	Vento no FL180 (500hPa)	REDEMET (https://redemet.decea.gov.br)
GFS 0,25°	3h	Análise e previsão em uma grade global de latitude-longitude de 0,25° por 0,25°	NCAR (https://rda.ucar.edu/datasets/ds084.1/)
GOES-16	15min	Banda termal canal 13	CPTEC (http://satellite.cptec.inpe.br/)

1 – Aceleração vertical em voo (VRTG) das aeronaves da empresa aérea LATAM registrados no período de março de 2018 a dezembro de 2019.

Aeronaves registram acelerações verticais em termos da aceleração gravitacional (g), que pode ser usada como um indicador de áreas turbulentas. Em um voo sem perturbações, o VRTG é igual a 1 g (LATAM, 2019). Se a aeronave sofrer uma ascensão ou queda repentina, os valores de VRTG serão maiores (g positivo) ou menores (g negativo), respectivamente. A Tabela 3 apresenta os limiares do VRTG usados para classificação dos eventos turbulentos.

Tabela 3: A intensidade do VRTG em função da variação da aceleração vertical em voo em termos da aceleração gravitacional (g). (adaptado de LATAM, 2019).

Classe 1		Classe 2		Classe 3	
-g	+g	-g	+g	-g	+g
0,4<g<=0,6	1,4<=g<1,6	0,2<g<=0,4	1,6<=g<1,8	g<=0,2	g >=1,8

Aqui, assume-se como turbulência leve quando os valores de VRTG correspondem aos da classe 1 que são menores ou iguais a 0,6 g e maiores que 0,4 g ou maiores

ou iguais a 1,4 g e menores que 1,6 g. Para turbulência moderada ou severa (MOS) os valores de VRTG correspondem aos da classe 2 ou classe 3 atingindo valores menores ou iguais a 0,4 g ou maiores ou iguais a 1,6 g. Vale ressaltar que embora apenas a LATAM tenha fornecido esses dados para a pesquisa, ela é a principal companhia aérea a realizar voos na região da rota Santiago, Chile – Mendoza, Argentina, realizando uma média mensal em 2019 de 7.212 pousos e decolagens apenas em SCEL, com variação mensal média menor que 10%.

As medições do VRTG ocorrem na chamada Janela de Monitoramento, que tem início 10 s após a decolagem e encerra 4 s antes do pouso (Figura 11).

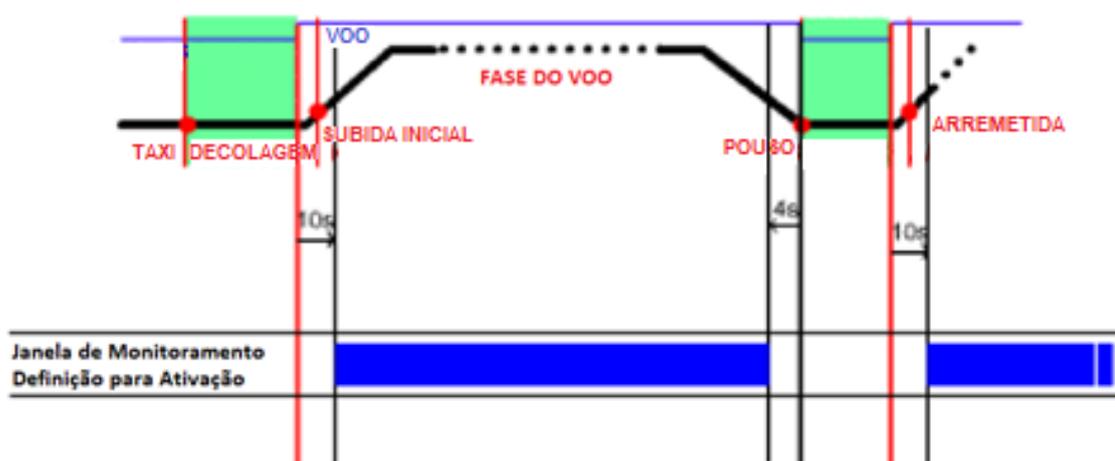


Figura 11: Esquema de início e término da Janela de Monitoramento para registro de VRTG. Fonte: traduzido de LATAM (2019)

O evento turbulento é considerado apenas quando os valores de VRTG estiverem dentro dos limites estabelecidos na Tabela 3, o que abrirá a Janela de Detecção, porém apenas o maior máximo (VRTGMAX) ou menor mínimo (VRTGMIN) será registrado (Figura 12). A Janela de Detecção encerra quando condições suaves ($VRTGMAX < 1,1 \text{ g}$ e $VRTGMIN > 0,9 \text{ g}$) são sentidas por mais de 30 s. Isso garante que apenas um evento é gerado em condições longas de turbulência (LATAM, 2019). Múltiplas ocorrências de um evento podem ocorrer dentro da Janela de Monitoramento, levando o sistema aguardar 300 s em acelerações normais ($VRTGMAX < 1,3 \text{ g}$ e $VRTGMIN > 0,7 \text{ g}$). Após esse tempo, um novo evento poderá ser gerado (LATAM, 2019).

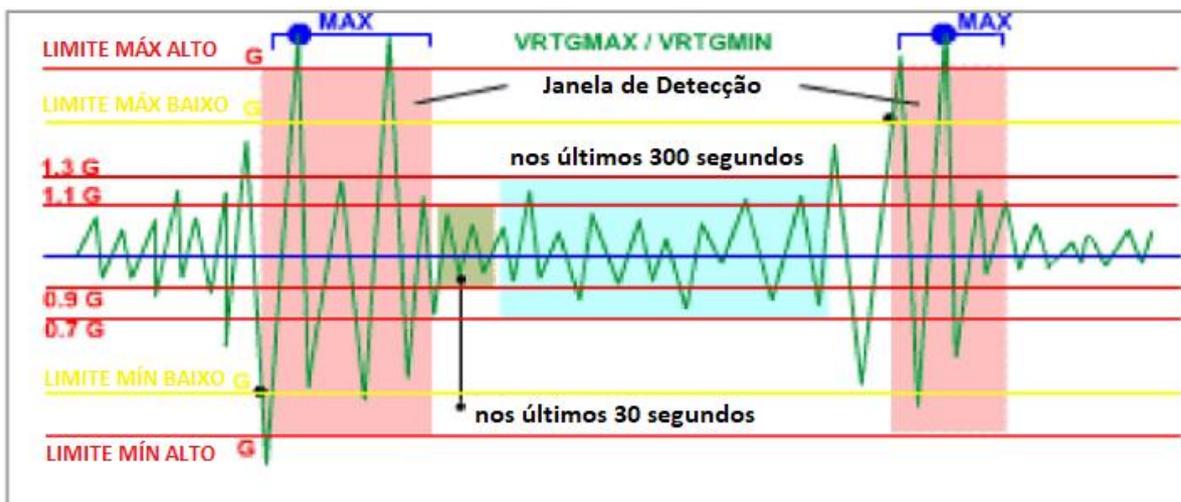


Figura 12: Exemplo de medição de valores de VRTGMAX ou VRTGMIN na Janela de Detecção.

Fonte: traduzido de LATAM (2019).

Como as aeronaves comerciais operam com bastantes limitações, apesar de estruturalmente estarem preparadas para cargas muito maiores, as componentes de VRTG quando a aeronave realiza uma curva são pequenas, dificilmente passam de $\pm 0,13$ g, ficando assim longe dos limites de disparo para abertura da Janela de Detecção. Alguns horários podem ter mais de um registro de VRTG, inclusive de classes distintas. Isso se deve ao fato de que mais de uma aeronave pode ter sobrevoado o local em níveis de voo diferentes, com ocorrência de turbulência maior em um do que no outro, ou uma mesma aeronave registrando VRTG em uma curta distância, e dentro do mesmo horário (os minutos não são contabilizados).

Importante destacar que os dados de VRTG não amostram sistematicamente o espaço aéreo, visto que são exclusivamente dependentes da realização do voo e do posicionamento da aeronave (latitude, longitude e nível de voo), e com base na análise da relação espacial entre EDR e PIREP/AIREP, a incerteza espacial de cada ponto de VRTG é da ordem de 150 km^3 (WANDISHIN *et al.*, 2011). Assim, a falta do registro de VRTG não significa necessariamente uma ausência de turbulência. Levando em conta esse fato e que turbulência atmosférica de interesse para a aviação é um fenômeno de escala sinótica, procedimentos de extrapolação são feitos para melhor aproximar o conjunto de dados da condição real (conforme detalhamento na Seção 3.3).

2 - Total mensal e horário de números de voos (pouso e decolagem) realizados pela LATAM em SCEL.

Como a quantidade de dados de VRTG é afetada diretamente pela quantidade de voos, esse fato poderia causar um viés nas análises. Dessa forma, o total mensal e horário de voos realizados pela LATAM em SCEL foram utilizados como um moderador daqueles dados. Os registros horários foram obtidos referentes a apenas dois dias do ano de 2020 (28 fevereiro e 11 março), não sendo possível a obtenção de dados pretéritos por não serem arquivados. Os dados a partir da segunda quinzena de março não puderam ser utilizados devido à queda brusca no número de voos causado pela pandemia do COVID-19. O total mensal de voos foi obtido referente ao ano de 2019.

3 - Dados horários de QNH dos aeroportos de SCEL e SAME obtidos do código METAR de março 2018 a dezembro 2019.

METAR e SPECI são códigos meteorológicos com fins aeronáuticos emitidos pelas estações meteorológicas de superfície dos aeródromos. Neles são registradas as informações significativas do tempo como vento, temperatura do ar e pressão atmosférica necessárias para as operações de pousos e decolagens das aeronaves. O METAR é confeccionado a cada hora e de forma regular, enquanto o SPECI só é gerado em situações de mudança das condições meteorológicas significativas ocorridas no intervalo entre dois códigos METAR. Os códigos SPECI foram descartados neste estudo pois para o uso do Ábaco de Harrison são necessárias as informações de QNH de SCEL e SAME para um mesmo horário.

5 - Direção e velocidade do vento no FL180 (500hPa) obtidos do código TEMP de SCSN no período de março 2018 a dezembro 2019.

O código meteorológico TEMP é usado para relatar o perfil atmosférico de pressão, temperatura, umidade e vento e é normalmente gerado diariamente em 00Z e 12Z. Aqui, os dados recuperados de uma estação meteorológica de altitude em SCSN (visto na Figura 9), são usados. Essa estação meteorológica foi escolhida por estar localizada a oeste dos Andes, próxima ao Oceano Pacífico. Assume-se que os perfis de vento obtidos em 12Z e 00Z são constantes nos períodos entre 07Z e 18Z e entre 19Z e 06Z, respectivamente. Na ausência de dados TEMP, são utilizados os dados

de previsão gerados pelo Centro de Previsão Mundial de Área (*World Area Forecast Center - WAFC*) (disponível no Portal Meteorológico do Comando da Aeronáutica - REDEMET).

6 - Análise do *Global Forecast System (GFS)* de março 2018 a dezembro 2019.

O GFS é um modelo global de previsão do tempo produzido pelo NCEP (*National Center for Environmental Prediction*) com resolução horizontal de 0,25° de latitude e longitude. Os dados desse modelo incluem análise e previsão em intervalos a cada 3 h com rodadas ocorrendo diariamente nos horários das 00Z, 06Z, 12Z e 18Z. As informações do GFS foram utilizadas para caracterizar o escoamento atmosférico e para extrair informações de variáveis meteorológicas necessárias nas análises. Em todos os casos, foram utilizadas as informações do GFS com a rodada das 00Z para cada dia.

7 - Imagens do satélite geoestacionário operacional ambiental (*Geostationary Operational Environmental Satellite – GOES 16*)

Imagens do GOES 16, canal 13 (banda termal centrada em 10,3 μm), são utilizadas para identificar a presença de células convectivas (disponível no Centro de Previsão do Tempo e Estudos Climáticos - CPTEC).

3.3 Método

Nesta seção é detalhada a metodologia desenvolvida para este estudo. Ela está esquematizada no diagrama de blocos da Figura 13, onde é apresentado o método adotado.

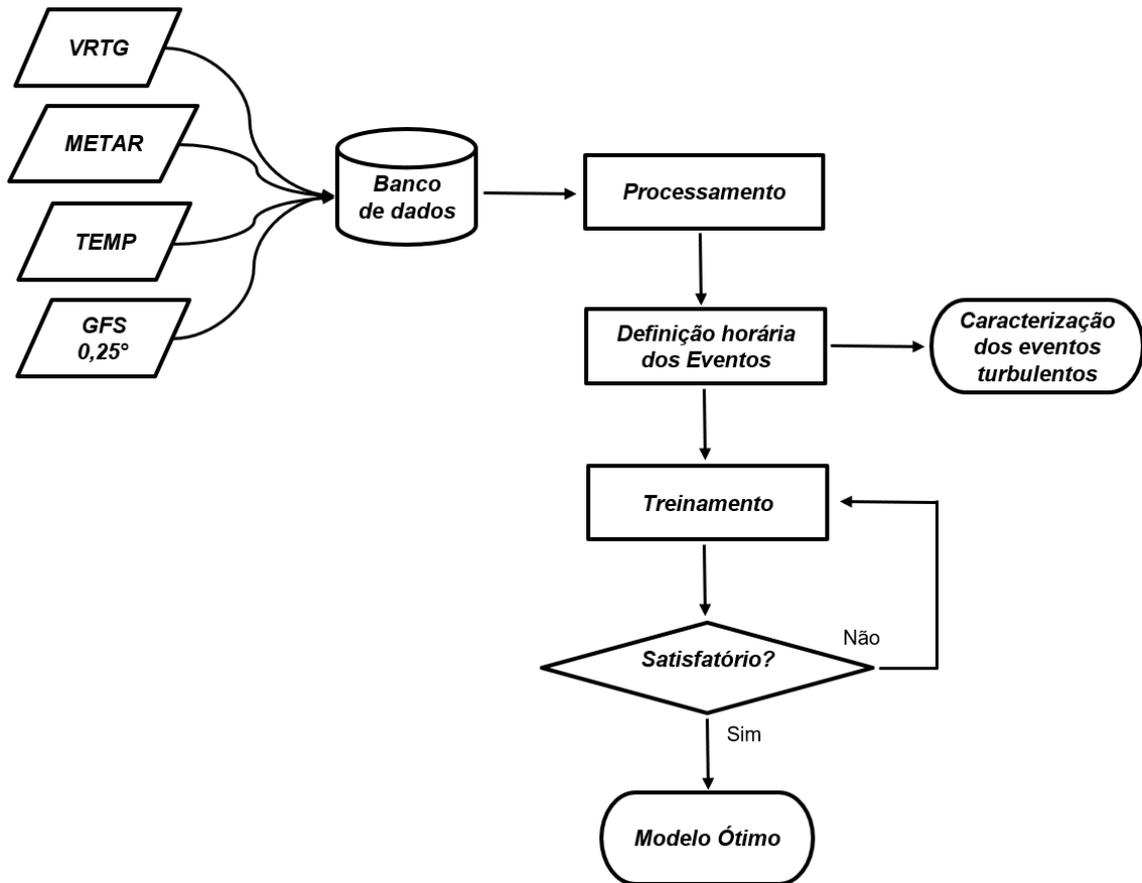


Figura 13: Fluxograma dos passos metodológicos da pesquisa.

Para atingir os objetivos delimitados na Seção 1.2, o método divide-se em 5 partes com 26 passos, a saber:

a) Processamento de dados

- i. Analisa-se a consistência dos dados de VRTG com descarte dos registros de VRTG abaixo do FL100 por serem oscilações de baixos níveis decorrentes de procedimentos de pouso ou decolagem e por estarem bem abaixo do nível médio do topo da cordilheira;
- ii. Em cada horário verificado com registros de VRTG, esses são contabilizados apenas uma única vez para cada classe estabelecida na Tabela 3, independentemente do FL em que foi registrado;
- iii. Define-se os horários dos eventos turbulentos levando em conta que cada dado de VRTG seja válido por um intervalo de tempo de 3h antes e 3h depois do horário registrado. Após isso, horários com VRTG são considerados SIM

para turbulência e os horários sem VRTG são considerados NÃO para turbulência;

- iv. Determina-se os valores de V_n para cada horário, baseado nos códigos TEMP, e DP com base na diferença dos valores de QNH do código METAR de SCEL e SAME, e sua estruturação em uma planilha do tipo Excel juntamente com os dados horários de VRTG do passo iii;

b) Caracterização dos eventos turbulentos

- v. Analisa-se a ocorrência da turbulência baseada nos dados de VRTG do passo i para distribuição geográfica (LAT/LON) e FL;
- vi. Analisa-se a ocorrência da turbulência conforme estabelecido do passo iii, com distribuição de frequências e números totais por classe de turbulência, registros mensais e horários. Pondera-se os valores dos gráficos de registros mensais e horários de VRTG com os dados dos números de voos efetuados pela LATAM em SCEL;
- vii. Calcula-se a média mensal e desvio-padrão dos dados de V_n dos horários com VRTG, do conjunto de dados do passo iv;
- viii. Analisa-se o comportamento de V_n e DP em relação ao VRTG do conjunto de dados do passo iv, considerando: a totalidade dos casos de turbulência, e a turbulência MOS;

c) Calibração e teste do Ábaco de Harrison

- ix. Ajusta-se a definição horária dos eventos turbulentos do passo iii, considerando apenas como SIM para os casos definidos como turbulência MOS;
- x. Particiona-se de forma aleatória o conjunto de dados horários do passo iv em 70% para calibração e 30% para teste, levando em conta o definido no passo ix;
- xi. Estabelece-se o fluxograma da Figura 15a para realização dos experimentos com o Ábaco de Harrison;
- xii. Seleciona-se diferentes valores de V_n e DP para construção dos modelos preditivos;

- xiii. Avalia-se o método considerando o conjunto teste com as métricas descritas na seção 3.3.1 e define-se o modelo ótimo;

d) Treinamento e teste dos algoritmos de AM

- xiv. Ajusta-se a definição horária dos eventos turbulentos do passo iii, considerando apenas como SIM para os casos definidos como turbulência MOS e com aproximação dos dados horários das ocorrências para o horário modelado mais próximo (os dados do GFS são a cada 3h a partir das 0Z);
- xv. Define-se o retângulo de interesse (limites indicados na Figura 14) de onde são obtidos os dados meteorológicos gerados pelo GFS $0,25^\circ$ para representar o estado da atmosfera na área de estudo;
- xvi. Extrai-se no retângulo de interesse os dados de velocidade zonal, velocidade vertical, temperatura potencial, variação zonal da temperatura potencial, variação zonal da velocidade vertical, cisalhamento vertical do vento e o número de Richardson Bulk (Ri_B) para cada grau de longitude e a cada 100hPa; velocidades verticais mínima, média, máxima e desvio-padrão; a diferença entre o maior e menor valores de velocidade vertical; o Ri_B em seus valores mínimo, médio, máximo e somatório; e o cisalhamento vertical do vento em seus valores mínimo, médio, máximo e somatório.
- xvii. Calcula-se o máximo vento zonal entre os níveis de 1000hPa e 500hPa na latitude de $33,5^\circ S$ e longitude de $71^\circ W$ (U da equação 2); a frequência Brunt-Vaisala (método bulk, Seção 2.3) e o número de Froude. Totalizando com os dados do passo xvi, 184 atributos (Tabela 11 no Apêndice A) para cada horário modelado do GFS, estruturado em uma planilha do tipo Excel juntamente com os dados horários de VRTG do passo xiv;
- xviii. Retira-se do conjunto de dados do passo xvii para posterior estudo de caso os 24 dados dos dias: 02 março 2018, 06 maio 2018 e 28 setembro 2018;
- xix. Estabelece-se o fluxograma da Figura 15b para a realização dos experimentos com os algoritmos de AM disponíveis na plataforma WEKA;
- xx. Obtêm-se cinco configurações do conjunto de dados do passo xviii da seguinte forma: 1) dados originais; 2) retirada aleatória de dados NÃO em quantidade suficiente para o balanceamento com os dados SIM; 3) inserção aleatória de dados SIM em quantidade suficiente para o balanceamento com os dados NÃO, pelos seguintes métodos que foram aplicados separadamente para

cada um dos 184 atributos: a) média e um desvio-padrão (m1desvpad); b) média e dois desvios-padrões (m2desvpad); c) mediana e quartis inferior e superior (mediana);

- xxi. Seleciona-se os atributos para cada uma das cinco configurações do conjunto de dados do passo xx, pelo método CFS (*Correlation-based feature selection*) e heurística *BestFirst-Search* com critério de parada de 5 consecutivos subconjuntos que não melhoram o método;
- xxii. Aplica-se a correção de prevalência da configuração dos dados originais do passo xx, após a seleção dos seus atributos do passo xxi, da seguinte forma: 1) varia-se artificialmente os pesos de SIM e NÃO (*ClassBalancer*) em 50% e 50%, 60% e 40%, 70% e 30%, 80% e 20%, 90% e 10%, 96% e 4%; 2) modifica-se os pesos atribuídos aos falsos negativos (Figura 17) via matriz de custo com os seguintes múltiplos: 10, 20, 30, 50, 100, 200 e 380;
- xxiii. Constrói-se os modelos mediante treinamento dos algoritmos de AM, utilizando a configuração padrão desses algoritmos estabelecida na plataforma WEKA, para cada uma das cinco configurações de dados do passo xxi e para cada uma das treze configurações do passo xxii;
- xxiv. Avalia-se os modelos via correlação cruzada de dez amostras do conjunto de treinamento usando as métricas descritas na Seção 3.3.1, definindo-se os modelos ótimos;

e) Estudo de caso

- xxv. Analisa-se o estado básico da atmosfera na área de estudo com as informações do GFS para os 24 dados dos dias selecionados em xviii;
- xxvi. Avalia-se os resultados dos modelos ótimos selecionados em xiii e xxiv em comparação com a análise feita em xxv;

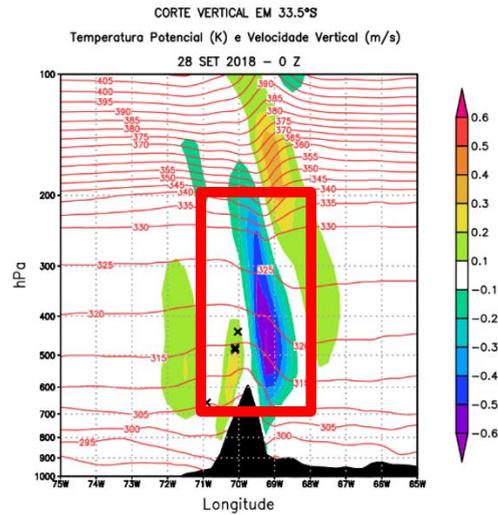
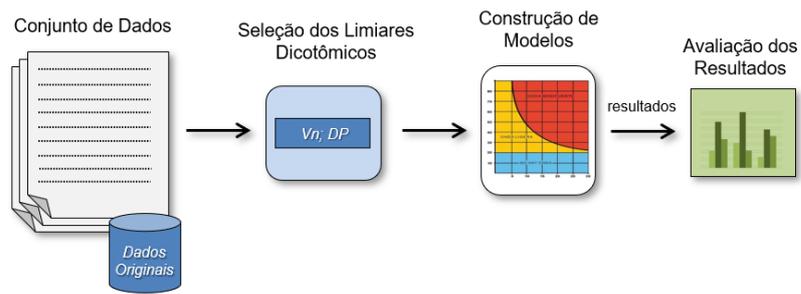
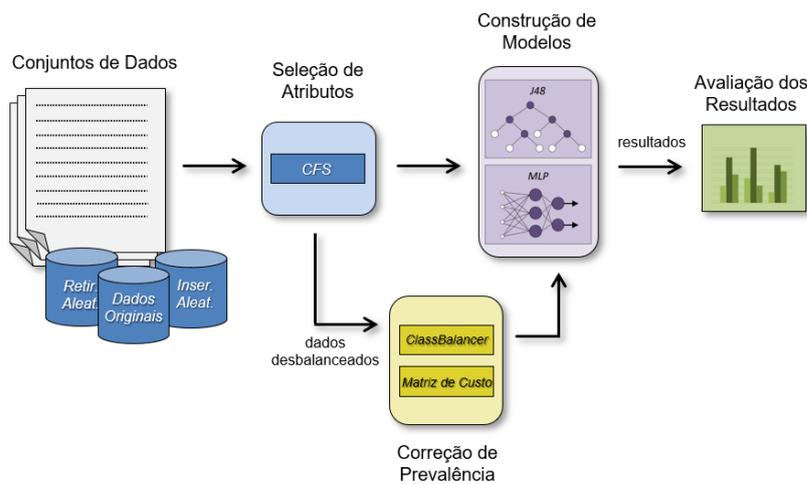


Figura 14: Delimitação do retângulo de interesse (retângulo em vermelho) na latitude de 33,5°S, entre as longitudes de 68°W e 71°W e entre os níveis verticais de 700hPa 200hPa, representando o estado da atmosfera na área de estudo.



(a)



(b)

Figura 15: Configurações dos experimentos de treinamento e teste: a) modelo objetivo com Vn/DP (Ábaco de Harrison); b) algoritmos de AM. Fonte: adaptado de Parmezan, Lee e Chung (2012).

3.3.1 Métrica de Avaliação dos Resultados

A curva Característica de Operação do Receptor (ROC) é uma representação gráfica que ilustra o desempenho de um sistema classificador binário à medida que o seu limiar de discriminação varia. Seu critério de mudança é resultado da operação de duas características: o verdadeiro positivo e o verdadeiro negativo. O melhor método de previsão possível produziria um ponto no canto superior esquerdo do gráfico (Figura 16), chamado de classificação perfeita, já uma informação aleatória produziria um ponto ao longo da diagonal do espaço ROC. Pontos acima da diagonal representam os melhores resultados, enquanto os pontos abaixo são resultados não-desejáveis. Diagramas ROC de POD_y versus $1-POD_n$ permitem determinar o valor limite ideal de um índice (*threshold*) que resulta no melhor POD_y com um valor aceitável baixo correspondente de $1-POD_n$ (KAY *et al.*, 2006).

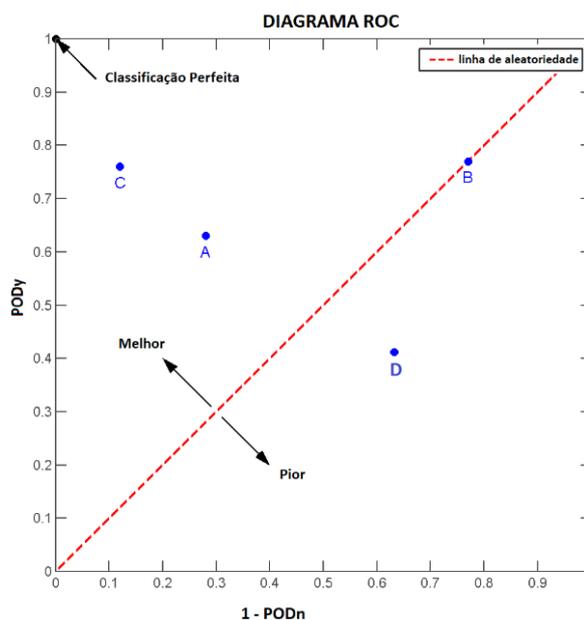


Figura 16: Esquema de um diagrama ROC. “D” resultado não-desejável, “B” resultado aleatório, “A” resultado desejável, “C” melhor resultado. A classificação perfeita fica no canto superior esquerdo.

Fonte: adaptado de Wikipedia, disponível em

https://pt.wikipedia.org/wiki/Característica_de_Operação_do_Receptor.

As medidas estatísticas utilizadas nessa pesquisa são derivadas da tabela de contingência (Figura 17).

		PREVISÃO	
		Evento	Não-Evento
OBSERVAÇÃO	Evento	VP	FN
	Não-Evento	FP	VN

Figura 17: Tabela de Contingência 2 x 2. Verdadeiro positivo (VP); verdadeiro negativo (VN); falso positivo (FP); falso negativo (FN)

Devido às limitações dos dados de VRTG que não amostram sistematicamente o espaço aéreo, assim como ocorre com o AIREP/PIREP, nem todas as estatísticas escalares como a taxa de falso alarme (FAR) e tendência de uma previsão (BIAS) são consideradas apropriadas de serem calculadas (BROWN & YOUNG, 2000). Por isso, nesse trabalho foram produzidas a POD_y (probabilidade de detecção do evento), POD_n (probabilidade de detecção do não-evento) e a TSS (*true skill statistic*) (Equações 5, 6 e 7).

$$POD_y = \frac{VP}{VP+FN} , \quad (05)$$

$$POD_n = \frac{VN}{VN+FP} , \quad (06)$$

$$TSS = POD_y + POD_n - 1 , \quad (07)$$

POD_y e POD_n não são consideradas probabilidades verdadeiras, mas a proporção do conjunto de dados observados corretamente categorizado pelas previsões (KAY *et al.*, 2006). O TSS mede a capacidade de um modelo de discriminar entre previsões “SIM” e “NÃO” (ELLROD & KNOX, 2010), variando de -1 a +1, onde +1 indica concordância perfeita e valores de zero ou menos indicam um desempenho não melhor que aleatório (ALLOUCHE, TSOAR & KADMON, 2006).

4. RESULTADOS E DISCUSSÕES

A seguir são apresentados os resultados e discussões de acordo com os passos estabelecidos na Seção 3.3.

4.1 Caracterização dos eventos turbulentos

Registraram-se originalmente 1300 dados de VRTG de março de 2018 a dezembro de 2019, dentro da área de estudo e acima do FL100 (3048 m), sendo 1190 de Classe 1, 91 de Classe 2 e 19 de Classe 3. A localização geográfica desses dados com a distinção entre as intensidades pode ser vista na Figura 18. Observa-se uma concentração dos registros em determinadas áreas devido ao alinhamento das aeronaves nas aerovias da região (como visto na Figura 10). Os eventos turbulentos ocorreram em todos os setores do cruzamento da cordilheira, não havendo área preferencial, afetando as aeronaves tanto à oeste, leste e acima dos Andes.

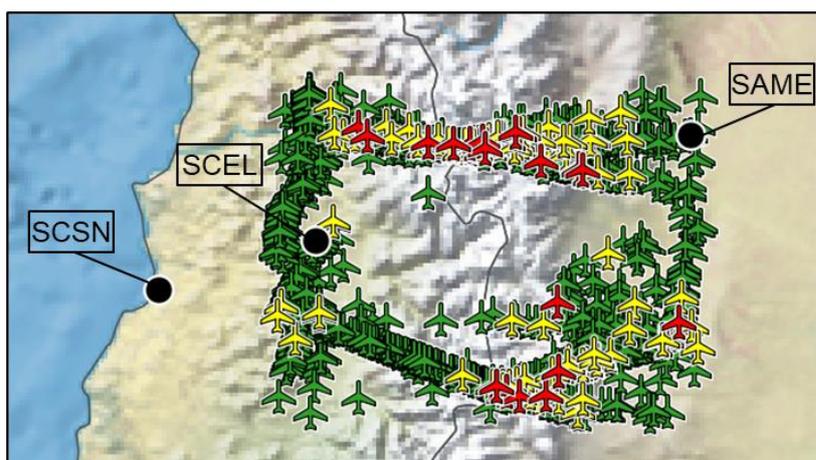


Figura 18: Posição dos 1300 registros de VRTG acima do FL100, de março de 2018 a dezembro de 2019, indicados de acordo com sua intensidade. Classe 1 (verde), Classe 2 (amarelo), Classe 3 (vermelho). Em círculo preto estão indicados os aeroportos de SCSN, SCEL e SAME.

Em relação aos níveis de voo mais impactados na área de estudo, observa-se na Figura 19 que os registros de VRTG são detectados em todos os níveis abaixo do FL400 (12192 m). Considerando os dados totais de VRTG, 93% (1210 ocorrências) estavam entre FL100-300, e o restante acima do FL300 (9144 m) (90 ocorrências).

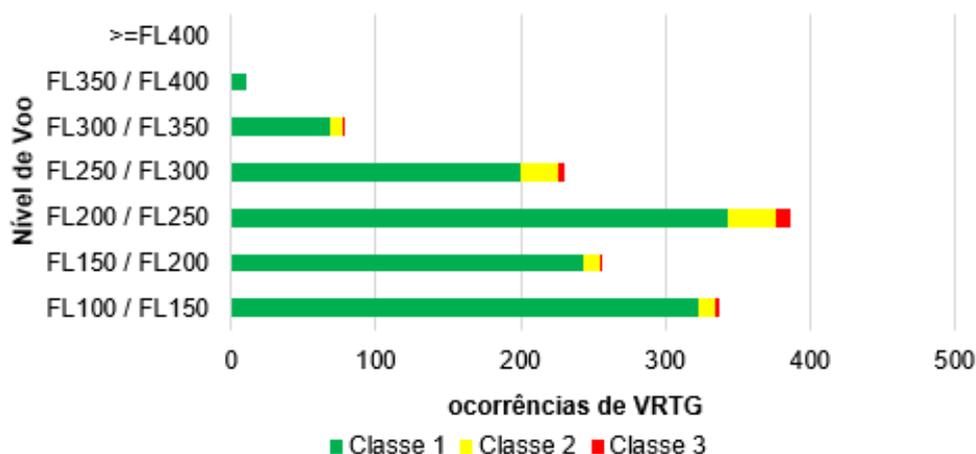


Figura 19: Nível de voo versus registros de VRTG por classe.

A comparação desses resultados com outros estudos não é tão simples, uma vez que os níveis de voo utilizados pelas aeronaves podem diferir para cada região com base em aspectos como topografia, tipo de aeronave, fase de voo e tempo de voo. Outro ponto a ser considerado é que as aeronaves costumam estar em níveis de cruzeiro, o que favoreceria maiores registros nesses níveis. Mello (2015) identificou que as maiores ocorrências de CAT no Brasil estão entre FL300-400, com cerca de 82,2% dos casos, não categorizando entre as intensidades das turbulências, diferindo bastante dos resultados aqui encontrados. Como as aeronaves na área de estudo estão próximas ao pouso ou decolagem do aeroporto de Santiago (SCEL), principal aeroporto da região, isso influencia no nível de voo a ser utilizado e, posteriormente, nos dados do VRTG, o que pode ter levado os níveis entre FL200-250 a terem os maiores registros de turbulência.

Considerando a definição horária da turbulência estabelecida nos passos i, ii e iii (Seção 3.3) ao longo dos 22 meses do período de estudo, 4164 registros horários de VRTG foram identificados, resultando em uma média de mais de 6 encontros com turbulência em voo por dia. Se apenas turbulência MOS for considerada, seria ao menos 1 por dia. A Figura 20 mostra os registros horários de VRTG versus intensidade de classe, de março de 2018 a dezembro de 2019. Pode-se notar que os registros de VRTG são predominantemente da classe 1 (83% ou 3469 registros), seguidos pelas classes 2 e 3 com 14% (575 registros) e 3% (120 registros), respectivamente.

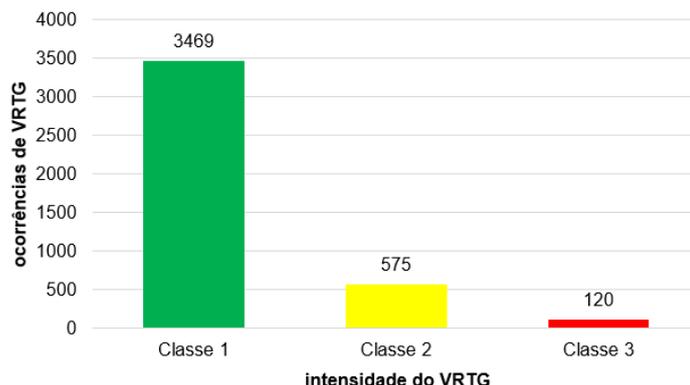


Figura 20: Registros horários de VRTG por classe no período de março de 2018 a dezembro de 2019.

Esses resultados são um tanto semelhantes aos obtidos na climatologia de CAT realizada por Mello (2015) no Brasil, quando utilizando dados de AIREP o autor constatou que apenas 2,7% eram turbulências severas. Por tanto, a turbulência que mais preocupa a aviação, a severa, atinge menos de 3% dos casos. Considerando a turbulência MOS, esse percentual ultrapassa os 15%.

Em relação a variabilidade mensal da turbulência, de acordo com a Figura 21, geralmente há mais registros de VRTG no mês de junho com mais de 300 ocorrências em 2018 e ultrapassando os 350 em 2019. Além disso, observa-se que de maio a novembro de 2018 são mais de 150 ocorrências mensais, enquanto em 2019 esse limiar foi vencido de abril a outubro.

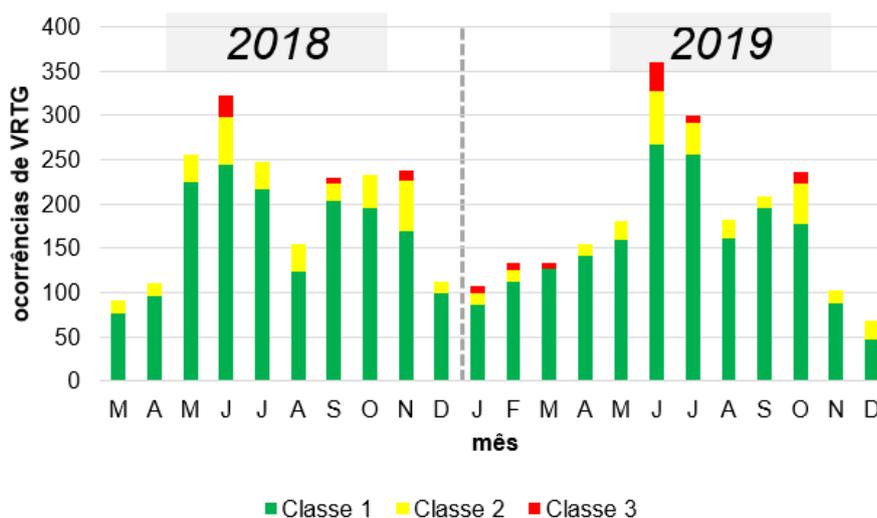


Figura 21: Registros horários de VRTG versus intensidade, no período de março de 2018 a dezembro 2019.

Existe um aparente comportamento periódico da turbulência, alternando entre os meses de maior ocorrência, tipicamente de maio a novembro com pico em junho (inverno) e o menor entre dezembro e abril com mínimo em dezembro (verão). A mesma variabilidade nos eventos de turbulência em relação às estações do ano foi encontrada em alguns estudos, embora tratem de outras regiões, tais como Mello (2015); Wolff e Sharman (2008); Jaeger e Sprenger (2007), sendo que neste último os autores realizaram uma climatologia global de CAT utilizando dados de reanálise com diferentes diagnósticos de turbulência. Os valores aparentemente baixos para o mês de agosto destoam dos meses próximos (julho e setembro). Isso pode ser devido ao próprio método da medição do VRTG, que, como visto na Seção 3.2, em situações de longas turbulências apenas registra um único valor (VRTGMÁX ou VRTGMÍN). Dessa forma em um período muito turbulento pode-se ter de forma paradoxal um reduzido número de registros de VRTG.

Como os valores de VRTG são dados automáticos registrados em voo, o fluxo de aeronaves poderia embutir um viés na análise do fenômeno. Como não foi possível verificar o total de aeronaves que sobrevoaram a área de estudo no período avaliado, um valor análogo encontrado e utilizado nesta pesquisa foi o total mensal de pousos e decolagens que a LATAM realiza em SCEL. Isso pôde ser feito pois a maior parte dos voos desta empresa aérea que decola ou tem por destino SCEL necessariamente realiza o cruzamento dos Andes.

Os dados mensais de voo da LATAM em SCEL do ano de 2019 que foram considerados válidos para ponderar todo o período de estudo são observados na Figura 22. Verifica-se que não há uma grande variação no número de voos entre os meses, permanecendo dentro de uma variação média inferior a 10%. A média mensal de voos no ano foi de 7212, com os maiores movimentos nos meses de janeiro, fevereiro, março e julho, e menores nos meses de abril, maio e junho.

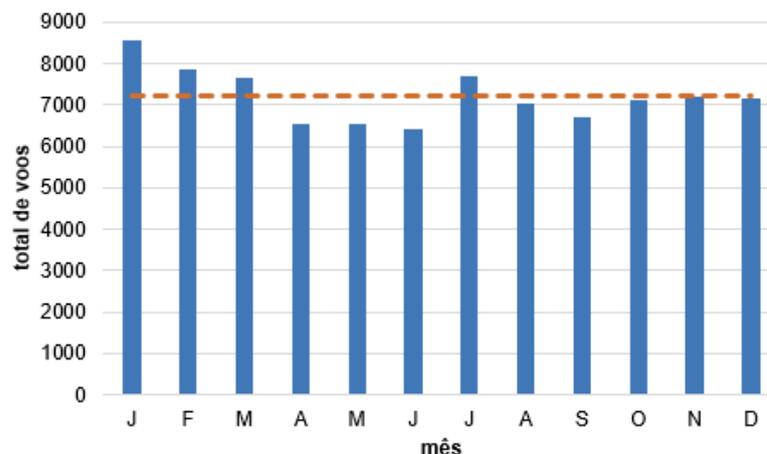


Figura 22: Total de pousos e decolagens da LATAM em SCEL ao longo do ano de 2019. A linha tracejada representa a média mensal de 7212 pousos e decolagens.

O ajuste dos dados mensais das ocorrências de VRTG pelo total mensal de movimentos da LATAM pode ser visto na Figura 23. Verifica-se que os gráficos das Figuras 21 e 23 possuem o mesmo comportamento, evidenciando que não houve alteração no padrão dos registros de VRTG ao longo dos meses pelo fluxo de aeronaves. Dessa forma, destaca-se que o período entre o final do outono até o começo da primavera, corresponde a época do ano mais favorável para a formação da turbulência. Esse fato pode ter relação com a intensificação dos ventos predominantes de oeste que ocorrem no inverno, que ao interagirem com a topografia podem gerar distúrbios atmosféricos que provocam a MWT.

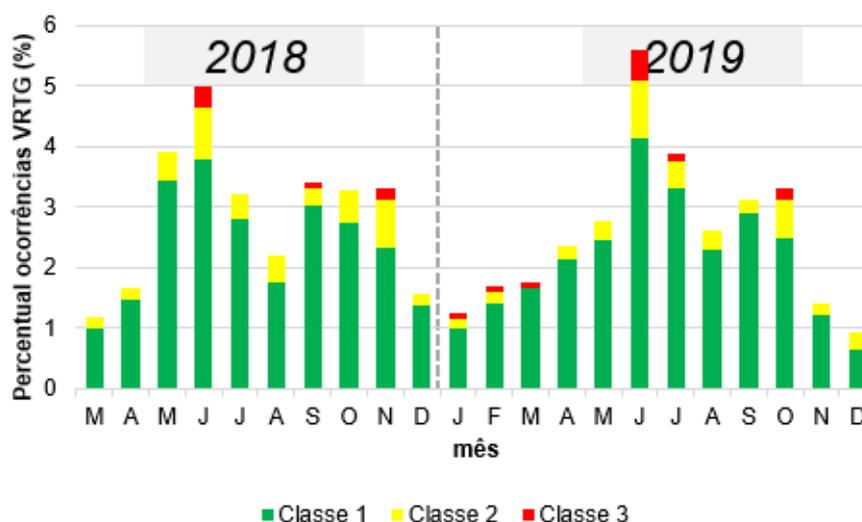


Figura 23: Percentual dos registros horários de VRTG versus intensidade, no período de março de 2018 a dezembro de 2019, ponderado pelo total de voos mensal da LATAM em SCEL.

Comparando os registros das turbulências das classes 1 e 2, verifica-se que há uma semelhança na variação ao longo dos meses. Porém um fato importante a ser destacado é o comportamento específico dos registros de VRTG classe 3. Embora seu máximo tenha coincidido com o máximo dos valores totais no mês de junho, foram significativos os registros de turbulência severa nos meses de janeiro, fevereiro e março de 2019, período com menores valores totais. Cerca de 18% (21 eventos) dos dados de VRTG classe 3 foram nesses três meses, em contraste com apenas 0,8% (353 ocorrências) do total de VRTG das classes 1 e 2. Essa característica não foi encontrada em nenhuma outra pesquisa da literatura.

Uma hipótese para esses registros de turbulência severa poderia ser a CIT. Esse tipo de turbulência ocorre dentro de nuvens convectivas, bem como no ar claro acima ou ao redor da nuvem e, às vezes, longe dos limites das nebulosidades (SHARMAN & TRIER, 2019). De acordo com Kaplan *et al.* (2005), cerca de 86% dos casos de turbulência severa examinados nos EUA estavam dentro de 100 km de uma convecção profunda. Mas as estimativas de frequência da CIT variam, dependendo da região e da estação favorável para a formação de uma tempestade. De acordo com Shi *et al.* (2000) e Reboita *et al.* (2010) a região de estudo tem um dos menores índices pluviométricos da América do Sul, o que indica que o aumento inesperado na ocorrência de turbulência severa no verão e início do outono provavelmente não está relacionado a CIT.

Analisando o perfil da turbulência em relação ao horário, observa-se na Figura 24 um aparente crescimento das ocorrências de VRTG ao longo do dia apresentando um máximo ao final da tarde, com mais de 250 registros às 16 h e um mínimo durante a madrugada, com valores próximos de 50 registros às 02 h. Essa tendência também é observada analisando cada classe de turbulência separadamente. Porém, como novamente o fluxo de aeronaves poderia embutir um viés na análise, verificou-se o total de voos por hora realizados pela LATAM em SCEL no período de estudo, utilizando como referência os movimentos realizados nos dias 28 de fevereiro de 2020 e 11 de março de 2020 para ponderar os dados horários de VRTG.

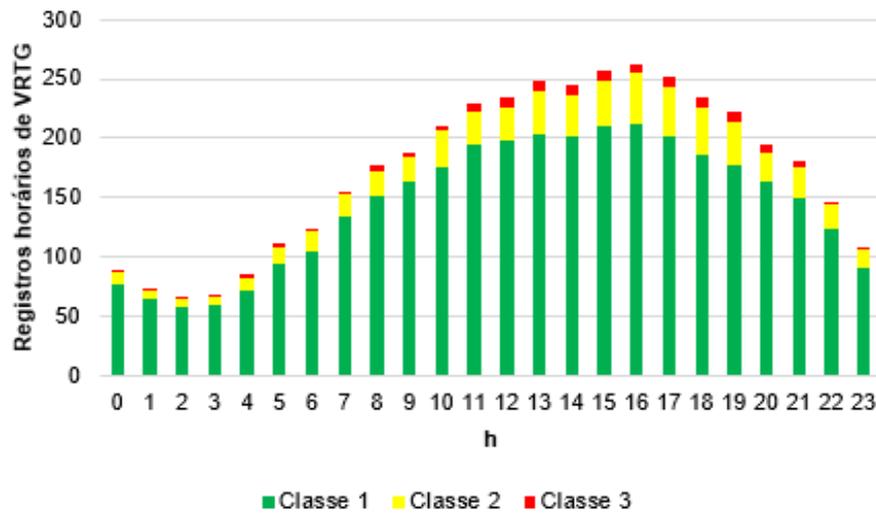


Figura 24: Média horária dos registros de VRTG por classe de turbulência no período de março de 2018 a dezembro de 2019. Horário local de Santiago (h) (Z-4).

Nota-se na Figura 25 uma grande variabilidade no número de voos quando analisados por hora, apresentando um mínimo de movimento durante a madrugada, das 0 h às 3 h e às 12 h, e um grande crescimento ao longo do dia, destacando máximos às 11 h, 17 h e 23 h, sendo que este último horário atingiu quase 22 pousos e decolagens. A média horária contabilizada foi de aproximadamente 11 voos.

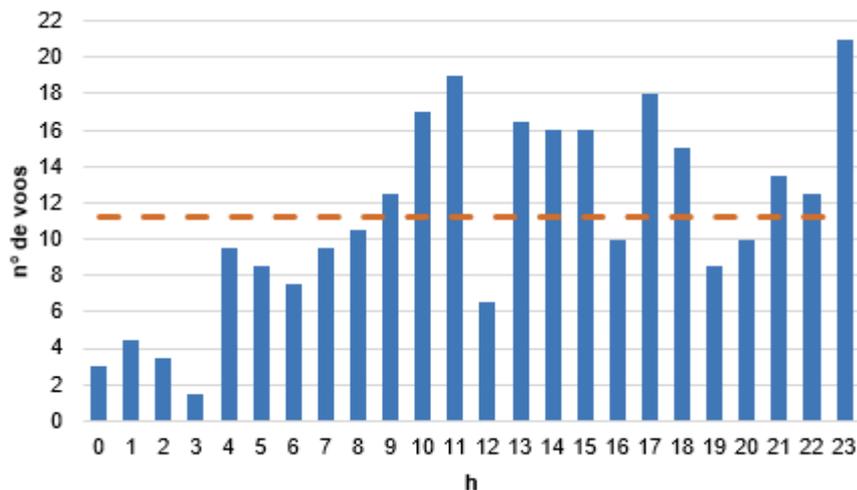


Figura 25: Média horária de voos da LATAM em SCEL por dia, utilizando como referência os movimentos dos dias 28 de fevereiro de 2020 e 11 de março de 2020. Em tracejado, a média de 11 voos. Horário local de Santiago (h) (Z-4).

O ajuste da média horária dos registros de VRTG pela média horária de voos da LATAM pode ser visto na Figura 26. Os gráficos das Figuras 24 e 26 não seguem o

mesmo padrão de distribuição das ocorrências de VRTG, evidenciando a influência do fluxo de aeronaves nos resultados. Houve uma redistribuição dos registros de VRTG ao longo do dia, identificando que a turbulência na área de estudo pode ocorrer independentemente da hora e sem um período preferencial. Os picos verificados em alguns poucos horários podem ter sido influenciados por pontuais ajustes sofridos na malha aérea, que não puderam ser abrangidos pelos dados da Figura 25.

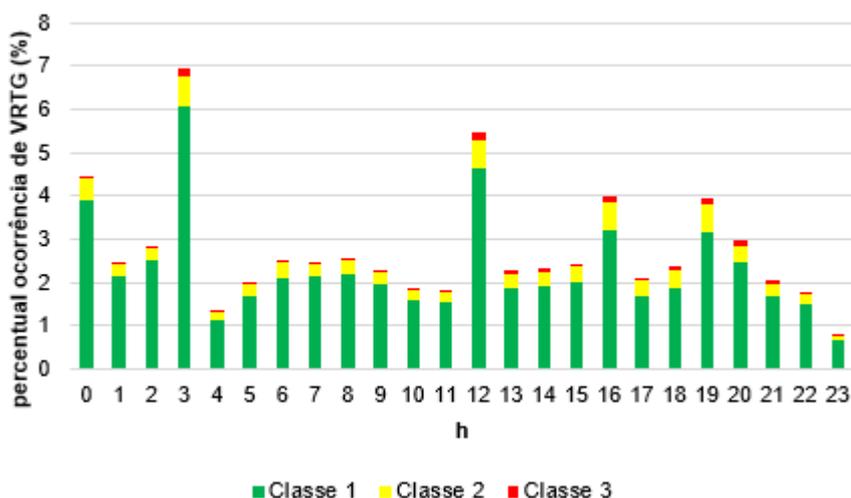


Figura 26: Média horária percentual dos registros de VRTG por classe de turbulência no período de março de 2018 a dezembro de 2019, ponderados pela média horária de voos da LATAM em SCEL. Horário local de Santiago (h) (Z-4).

Para compreender quais parâmetros podem influenciar na geração da turbulência nos Andes, foi analisada a média mensal da componente zonal da velocidade do vento no FL180 (5486 m) (V_n). A Figura 27 mostra que o V_n experimentou uma intensificação entre o final do outono e o início da primavera, permanecendo em grande parte desse período com uma velocidade média acima de 30 kt, atingindo um pico de quase 70 kt. O inverso ocorreu durante o verão e início do outono, quando V_n era menor, com velocidade média abaixo de 20 kt, chegando a atingir valores negativos (próximos a menos 10 kt), o que indica a presença de ventos de leste de baixa intensidade. O V_n possui uma grande variabilidade visto pelos desvios padrões atingindo valores entre 10 e 20 kt. Isso mostra que na área de estudo podem ocorrer desde fortes ventos de oeste até fracos ventos de leste, mas com a predominância dos primeiros.

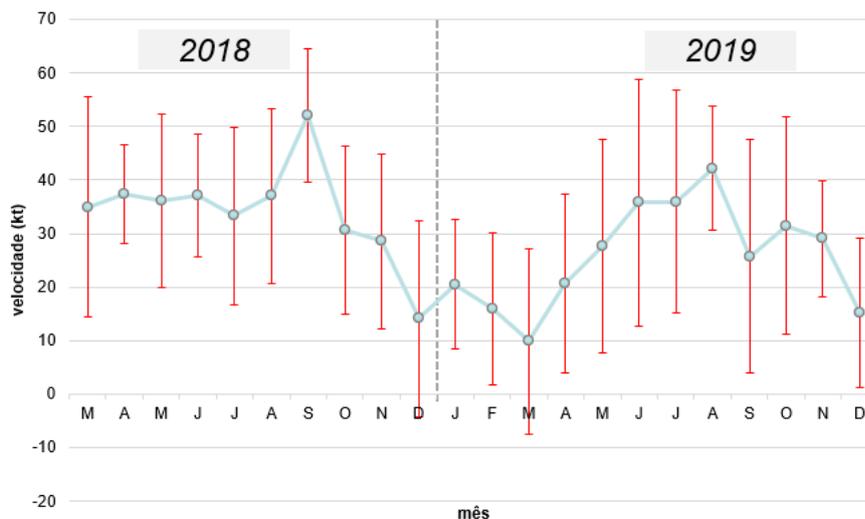


Figura 27: Média mensal e desvio-padrão da componente zonal da velocidade do vento no FL180 (V_n) para os casos de VRTG.

Então, o maior registro de turbulência identificado entre maio e novembro pode estar relacionado à intensificação dos ventos predominantes de oeste em latitudes médias durante o inverno, o que também foi verificado em vários estudos como Jaeger e Sprenger (2007), Mello (2015) e Wolff e Sharman (2008). Esse mecanismo é extremamente importante na região dos Andes, uma vez que o posicionamento específico da cordilheira, de orientação norte-sul, atua como um obstáculo ao fluxo atmosférico predominante de oeste com uma altitude média de 4500 m, proporciona a formação de ondas orográficas que podem quebrar e gerar MWT.

Uma característica semelhante foi observada por Wolff e Sharman (2008) para as Montanhas Rochosas nos EUA e também Vásquez e Falcón (2015) para os Andes venezuelanos. Os primeiros identificaram que na parte oeste dos EUA predominam as ocorrências de MWT devido aos fortes ventos de oeste de baixos níveis, que são mais comuns nos meses de inverno e pela presença de altitudes maiores ou iguais a 1500 m. Outro tipo de análise feita por Silva (2004), que estudando a dinâmica da atmosfera nos Andes próximos a SAME durante a ocorrência dos ventos zonda, identificou a importância da interação do relevo com o fluxo atmosférico, encontrando altos valores de TKE no topo da montanha e a presença de intensa e profunda corrente de jato entre os níveis de 100 e 650 hPa. O vento zonda tem maior frequência no inverno e primavera (NORTE, 2015). No entanto, valores mais baixos de V_n de dezembro de 2018 a abril de 2019, indicam que as ocorrências de

turbulência no verão e início do outono provavelmente não estão relacionadas aos mecanismos de MWT.

Analisando o comportamento dos registros de VRTG em relação ao Vn, verifica-se na Figura 28 que as turbulências ocorreram com Vn apresentando tanto valores negativos (ventos de leste) quanto valores positivos (ventos de oeste), variando de -40 kt a 90 kt. Dos 4164 registros horários de VRTG, a maior parte deles, cerca de 70% (2897), ocorreram com Vn de 10 a 50 kt, com um pico de 857 registros para Vn de 30 a 40 kt. Os dados de VRTG classe 1 foram verificados em todas as faixas de Vn, os de classe 2 entre -10 kt a 90 kt e os de classe 3 entre -10 kt a 70 kt. Verifica-se uma faixa preferencial de Vn centrada de 30 a 40 kt para a ocorrência de turbulência, e essa vai diminuindo de incidência com ventos mais intensos tanto positivos quanto negativos. Como a incidência da intensidade do vento poderia causar uma tendência nas análises, já que ventos de 80 a 90 kt, por exemplo, não são tão comuns para o FL180, os dados de VRTG versus Vn foram ponderados levando-se em conta a frequência dos valores de Vn.

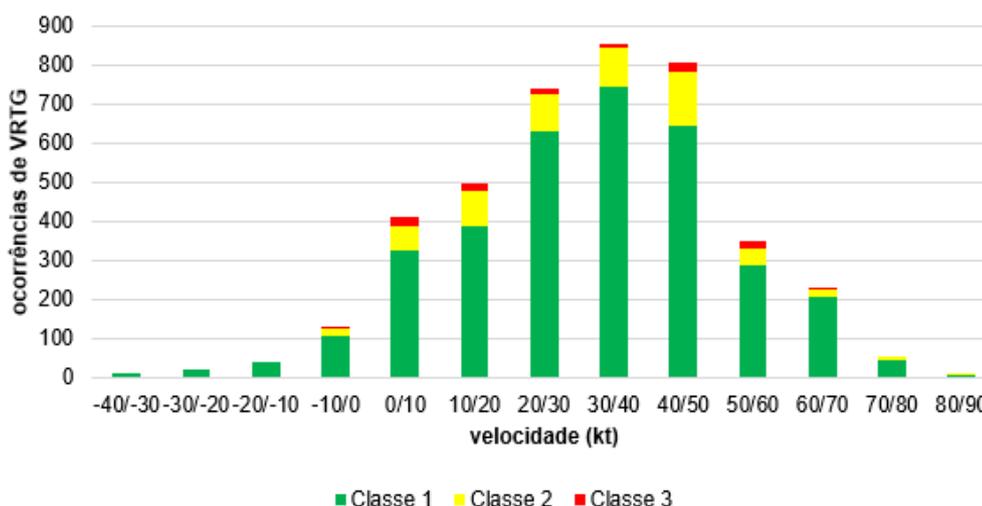


Figura 28: Total dos registros horários de VRTG por classe de intensidade versus Vn.

Nota-se na Figura 29, que ponderando os dados em relação a frequência de Vn ocorre uma mudança no comportamento dos registros de VRTG, apresentando agora um mínimo para Vn de 10 a 20 kt e um aumento na incidência da turbulência na direção dos extremos de intensidade do vento, tendência que também pode ser vista para cada classe de VRTG, principalmente em relação aos ventos de oeste. Apesar

da faixa de Vn de 30 a 40 kt ter registrado 857 casos de VRTG (visto na Figura 28), em apenas 27% das vezes que o vento atingiu tais velocidades resultou em turbulência. Em contraste, quando o Vn era de 80 a 90 kt, aconteceram somente 11 casos de VRTG, porém em 92% das vezes que o vento atingiu tal intensidade houve turbulência. Infere-se que quanto mais intenso em módulo o Vn, maior é a probabilidade e a severidade da turbulência.

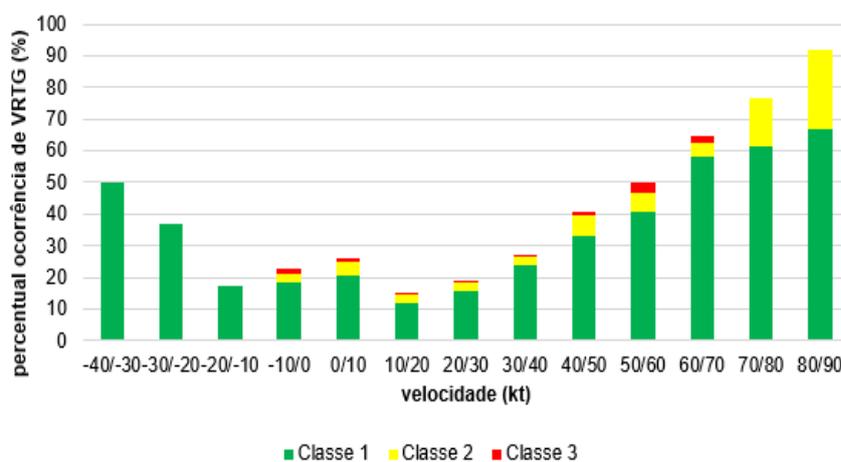


Figura 29: Percentual do total dos registros horários de VRTG por classe de intensidade versus Vn, ponderado pela incidência de Vn.

Analisando o comportamento do VRTG em relação a diferença de pressão entre SCEL e SAME (DP), verifica-se na Figura 30 que as turbulências ocorreram com DP apresentando tanto valores negativos (QNH maior em SAME) quanto valores positivos (QNH maior em SCEL), variando de - 20 hPa a 20 hPa. A maior parte dos registros horários de VRTG, cerca de 79% (3289), foram com DP entre - 4 hPa e 12 hPa, apresentando um pico de 992 registros para DP entre 4 hPa e 8 hPa. Os VRTG de classes 1 e 2 foram verificados em todas as faixas de DP e os de classe 3 entre - 16 hPa a 16 hPa. Nota-se uma faixa preferencial de DP centrada em 4hPa a 8hPa para a ocorrência de turbulência, e essa diminui de incidência com o aumento em módulo da diferença de pressão. Como também a incidência de DP poderia causar uma tendência nas análises, os dados de VRTG versus DP também foram ponderados levando-se em conta a frequência de ocorrência dos valores de DP.

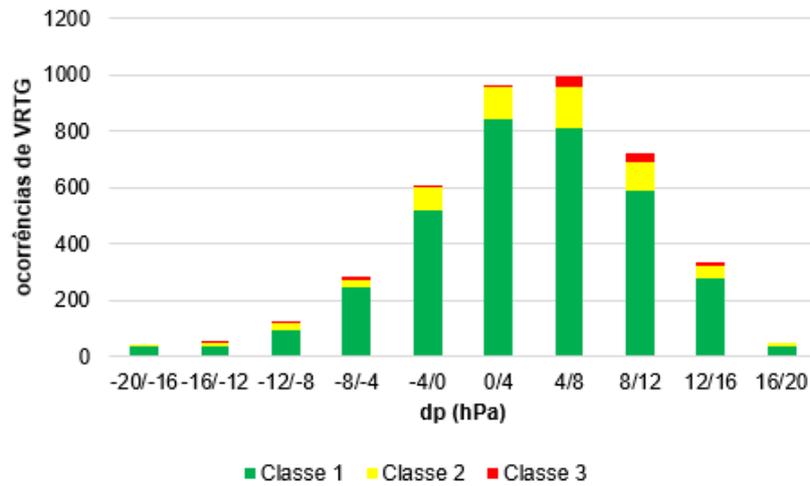


Figura 30: Total dos registros horários de VRTG por classe de intensidade versus DP.

Nota-se na Figura 31, que ponderando os dados em relação a incidência de DP ocorre uma mudança no comportamento dos registros de VRTG, apresentando agora um mínimo para DP de - 8 a - 4 hPa e um aumento na incidência da turbulência na direção dos extremos dos valores de DP, tendência esta que também pode ser vista para cada classe de VRTG. Apesar da faixa de DP de 4 hPa a 8 hPa ter concentrado 992 casos de VRTG (visto na Figura 30), em apenas 29% das vezes que a diferença de pressão atingiu tais valores resultou em turbulência. Em contraste, quando o DP era de 16 hPa a 20 hPa, aconteceram somente 49 casos de VRTG, porém em 78% das vezes que a DP atingiu tal limiar houve turbulência. Infere-se que quanto maior em módulo a DP, maior é a probabilidade e a severidade da turbulência.

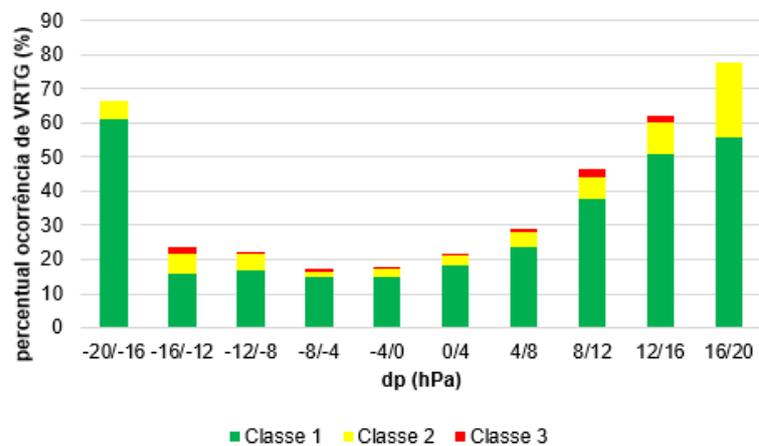


Figura 31: Percentual do total dos registros horários de VRTG por classe de intensidade versus DP, ponderado pela incidência de DP.

Na verificação do comportamento simultâneo de Vn e DP em relação aos registros horários de VRTG de todas as classes, nota-se na Figura 32 que a turbulência ocorre preferencialmente no primeiro quadrante, isto é, quando Vn e DP são positivos, englobando cerca de 72% (3016) dos casos. O segundo, terceiro e quarto quadrantes atingiram 23%, 4% e 1% respectivamente. A faixa preferencial para ocorrência de turbulência é com o par Vn de 30 a 40 kt e DP de 4 a 8 hPa (256 registros).

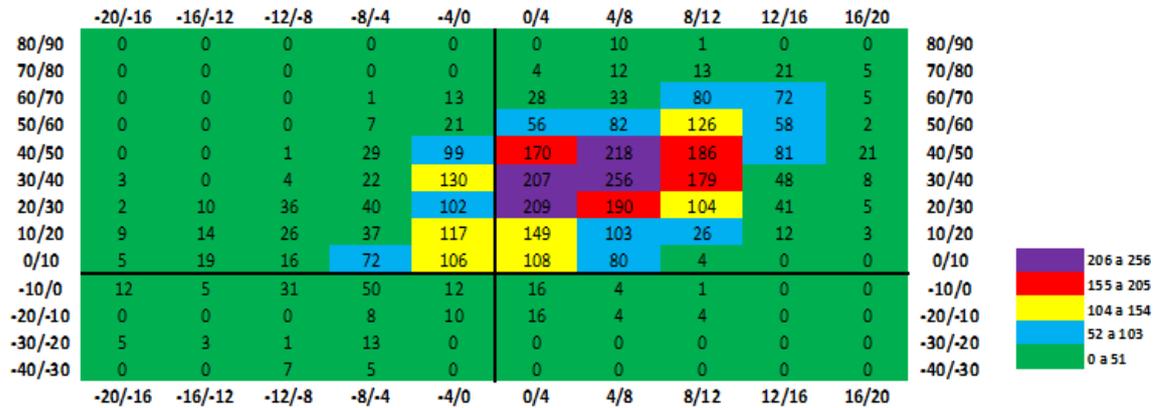


Figura 32: Registros horários de VRTG de todas as classes versus Vn (ordenada em kt) e DP (abscissa em hPa). Os valores representam a quantidade de dados horários de turbulência para cada par de Vn e DP, identificados com o intervalo de cores estabelecido de acordo com a quantidade de registros de VRTG.

Contudo, fazendo a ponderação dos dados de VRTG de todas as classes em relação a frequência dos valores de Vn e DP, tem-se na Figura 33 que as maiores probabilidades de acontecer uma turbulência no período de estudo situam-se nos extremos dos valores de Vn e DP, principalmente do primeiro quadrante. Apesar da faixa de Vn de 30 a 40 kt e DP de 4 a 8 hPa ter concentrado o maior número de registros de VRTG (visto na Figura 32), em apenas 28% das vezes que o par Vn e DP atingiram aqueles valores, ocorreu turbulência. Em contraste, quando Vn era de 70 a 80 kt e DP de 0 a 4 hPa, aconteceram somente 4 casos de VRTG, porém em 100% das vezes que o par Vn e DP atingiu tais limiares, houve turbulência.

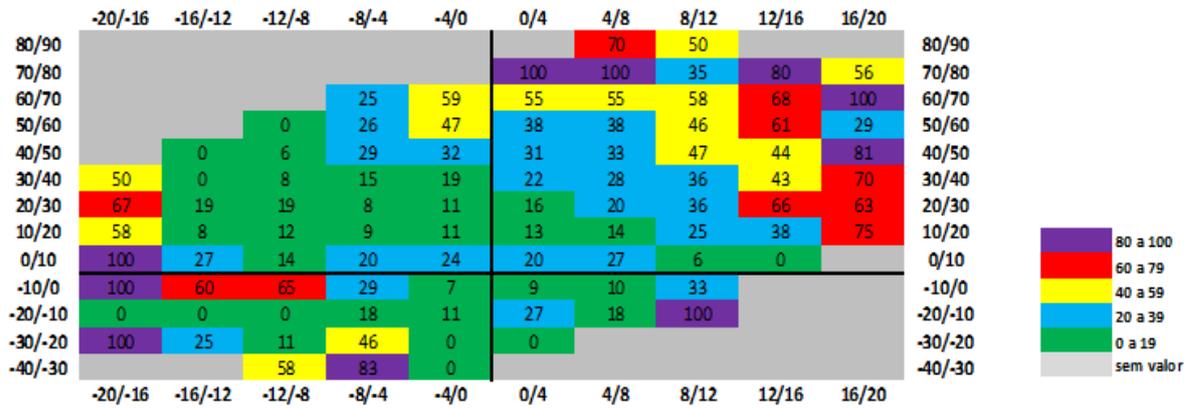


Figura 33: Percentual dos registros horários de VRTG de todas as classes versus Vn (ordenada em kt) e DP (abscissa em hPa), ponderado pelas incidências de Vn e DP. Os valores representam a porcentagem de vezes que houve turbulência para cada par de Vn e DP, identificados com o intervalo de cores estabelecido. Em cinza significa que em nenhum momento no período de estudo foi atingido o valor correspondente ao parâmetro dicotômico.

Na verificação do comportamento simultâneo de Vn e DP em relação aos registros horários de VRTG considerando apenas as classes 2 e 3, nota-se na Figura 34 que a turbulência MOS ocorre preferencialmente no primeiro quadrante, isto é, quando Vn e DP são positivos, englobando cerca de 74% (513) dos casos. O segundo, terceiro e quarto quadrantes atingiram 23%, 3% e 0% respectivamente. A faixa preferencial para ocorrência de turbulência MOS é com o par Vn de 40 a 50 kt e DP de 8 a 12 hPa (55 registros).

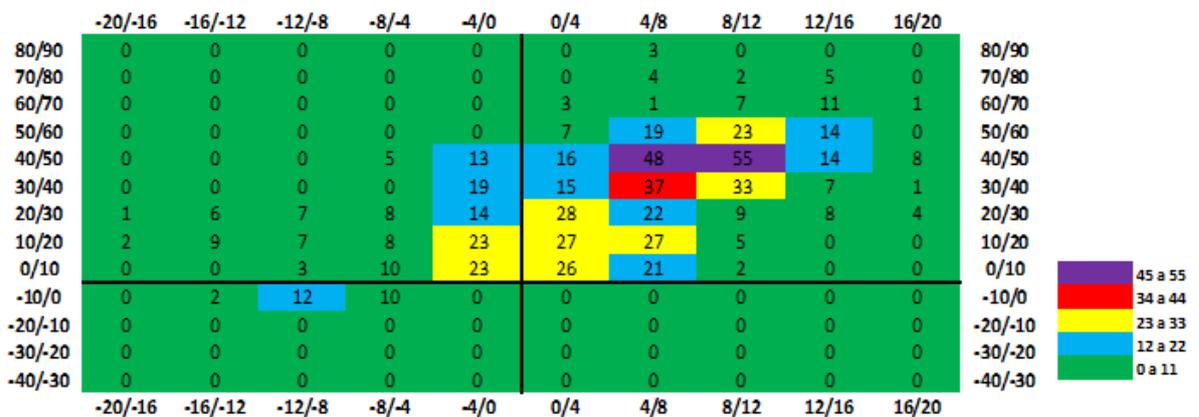


Figura 34: Registros horários de VRTG classes 2 e 3 versus Vn (ordenada em kt) e DP (abscissa em hPa). Os valores representam a quantidade de dados horários de turbulência MOS para cada par de Vn e DP, identificados com o intervalo de cores estabelecido de acordo com a quantidade de registros de VRTG.

Contudo, fazendo a ponderação dos dados de VRTG classes 2 e 3 em relação a frequência dos valores de Vn e DP, tem-se na Figura 35 que as maiores probabilidades de acontecer turbulência MOS no período de estudo situam-se nos extremos dos valores de Vn e DP, principalmente do primeiro quadrante. Apesar da faixa de Vn de 40 a 50 kt e DP de 8 a 12 hPa ter concentrado o maior número de registros de VRTG classe 2 e 3 (visto na Figura 34), em apenas 14% das vezes que o par Vn e DP atingiram aqueles valores, ocorreu turbulência MOS. Em contraste, quando Vn era de 70 a 80 kt e DP de 4 a 8 hPa, aconteceram somente 4 casos de VRTG classe 2 e 3, porém em 50% das vezes que o par Vn e DP atingiu tais limiares houve turbulência MOS.

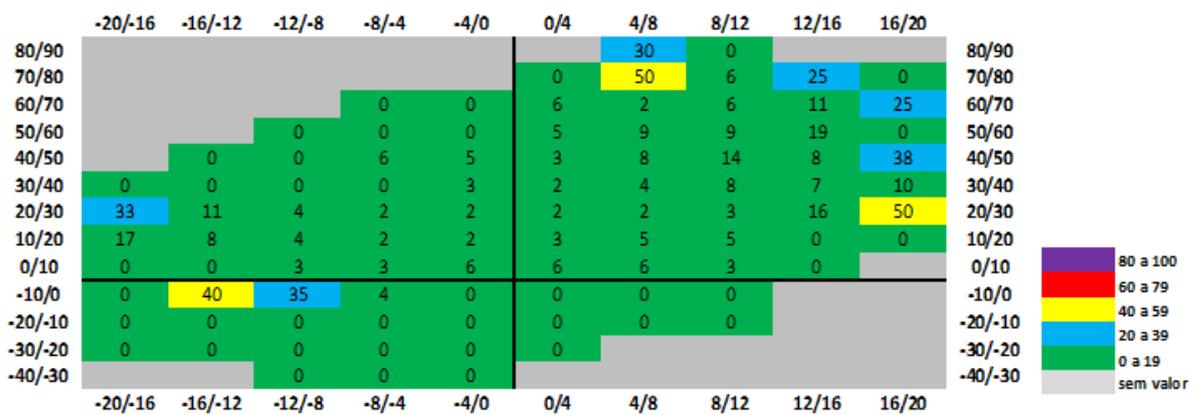


Figura 35: Percentual dos registros horários de VRTG classes 2 e 3 versus Vn (ordenada em kt) e DP (abscissa em hPa), ponderado pelas incidências de Vn e DP. Os valores representam a porcentagem de vezes que houve turbulência MOS para cada par de Vn e DP, identificados com o intervalo de cores estabelecido. Em cinza significa que em nenhum momento no período de estudo foi atingido o valor correspondente ao parâmetro dicotômico.

Dessa forma, utilizando como preditores de turbulência apenas os parâmetros Vn e DP, a maior parte dos registros de VRTG (tanto para a totalidade dos dados quanto para a MOS) acontece quando Vn e DP são positivos, isto é, o vento é de oeste e o QNH de SCEL é superior ao de SAME. Porém também são significativos os números de casos quando DP é negativo, isto é, o QNH de SAME é superior ao de SCEL, com Vn positivo (vento de oeste) ou negativo (vento de leste). Em termos de probabilidade, a turbulência tende a ocorrer quanto maiores forem em módulo os valores de Vn e DP, tanto para todas as classes quanto para a MOS.

4.2 Calibração e teste do Ábaco de Harrison

Ao longo dos 22 meses do período de estudo analisaram-se 16039 horas do conjunto de dados do passo iv (Seção 3.3), considerando a definição horária da turbulência estabelecida no passo ix (Seção 3.3), perfazendo um total 662 dados (a cada 1 h) identificadas como “SIM” para turbulência MOS e 15377 como “NÃO”.

Foram realizadas durante a fase de calibração diversas combinações de valores do par V_n e DP , sendo selecionadas oito delas, totalizando dezesseis experimentos realizados entre calibração e teste. Os resultados destes experimentos são vistos na Figura 36.

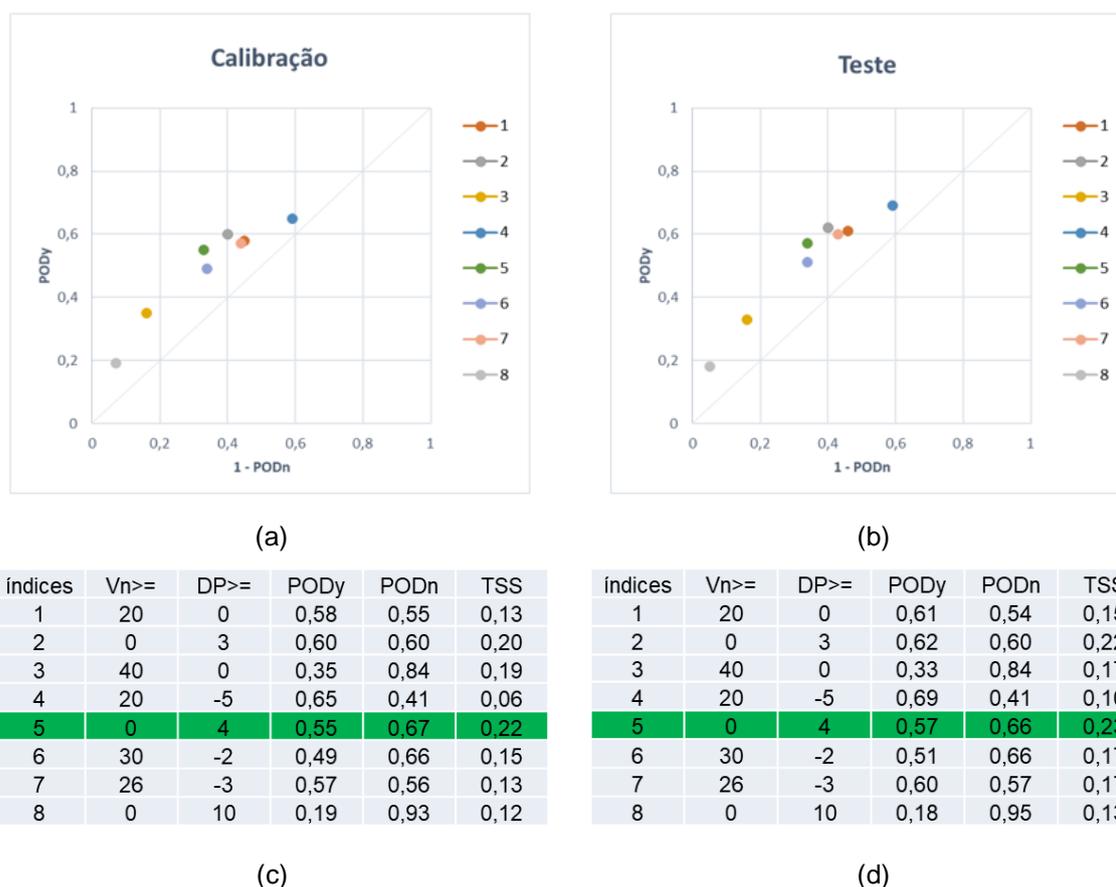


Figura 36: Resultado dos experimentos de calibração e teste com o Ábaco de Harrison. Curva ROC de POD_y versus 1-POD_n e tabelas com os valores de POD_y, POD_n e TSS. a) curva ROC da calibração; b) curva ROC do teste; c) POD_y, POD_n e TSS da calibração; d) POD_y, POD_n e TSS do teste. Cada índice é um par com diferentes V_n e DP . O índice 1 é referente aos valores originais do Ábaco de Harrison. Em realce verde o índice 5, que apresentou os melhores resultados. Valores de V_n em kt e DP em hPa.

O índice 1 ($V_n \geq 20$ kt e $DP \geq 0$ hPa), que se refere aos parâmetros originais do Ábaco de Harrison (seção 2.2), atingiu POD_y 0,61, POD_n 0,54 e TSS 0,15. O melhor modelo foi o índice 5 ($V_n \geq 0$ kt e $DP \geq 4$ hPa), conseguindo POD_y 0,57, POD_n 0,66, evidenciando com uma TSS de 0,23 mostrando um ganho de performance de 8% em relação aos limites originais do Ábaco de Harrison.

Verifica-se que o método objetivo considerando apenas como preditores os parâmetros V_n e DP possui uma limitação operacional no prognóstico da turbulência MOS, apresentando uma previsibilidade de no máximo 60%. O motivo para essa baixa performance está relacionado com os resultados apresentados nas Figuras 34 e 35, onde evidenciou-se uma clara disparidade entre os limiares de V_n e DP que indicam os números absolutos e os da probabilidade do evento, além da grande variabilidade de possibilidades de haver turbulência, quando se utiliza na análise somente o par V_n e DP . Dessa forma, outros métodos de maior complexidade e que fazem uso de mais parâmetros necessitam ser aplicados para que se obtenha um nível aceitável de previsibilidade.

4.3 Treinamento e teste dos algoritmos de AM

O conjunto de dados foi ajustado para o horário modelado mais próximo, em intervalos a cada 3 h, conforme descrito no passo xiii (Seção 3.3) e combinado com o passo xviii (Seção 3.3), totalizando ao longo dos 22 meses de estudo 5344 dados analisados, sendo 267 identificados como “SIM” para turbulência MOS e 5077 como “NÃO”. É importante salientar que, dos cinquenta e seis algoritmos disponíveis na plataforma WEKA, treze foram selecionados durante o treinamento para que pudessem ser utilizados com todos os métodos estabelecidos visando a posterior comparação de performance e seleção do modelo ótimo.

Considerando os treze algoritmos selecionados, as cinco configurações dos conjuntos de dados e as treze correções de prevalência, foram realizados um total de duzentos e trinta e quatro experimentos.

Os algoritmos foram treinados (via correlação cruzada com divisão da população dos dados em dez amostras aleatórias), com seleção dos atributos pelo método CFS e heurística *BestFirst-Search* com critério de parada de 5 consecutivos subconjuntos que não melhoram o método.

Os atributos do passo xvii (Seção 3.3), presentes na Tabela 11 do Apêndice A, selecionados para cada uma das cinco configurações dos conjuntos de dados durante o treinamento dos algoritmos de AM podem ser vistos na Tabela 12 do Apêndice B. Mesmo com a grande variabilidade de atributos selecionados para cada configuração dos conjuntos de dados, identifica-se que os parâmetros atmosféricos mais relacionados com a previsibilidade da turbulência MOS, foram as variações horizontais da temperatura potencial e da velocidade vertical em todos os setores da cordilheira, e o vento zonal, cisalhamento vertical do vento e Richardson Bulk imediatamente acima da montanha. Isso pode estar relacionado aos distúrbios ondulatórios causados pela MWT, fazendo tanto a temperatura potencial quanto a velocidade vertical oscilarem abruptamente em curtas distâncias, assim como a presença de fortes ventos que provocam intensos cisalhamentos verticais gerando CAT.

Verifica-se na Figura 37 a curva ROC com os resultados dos treze experimentos do treinamento dos algoritmos de AM utilizando os dados originais. Devido o problema do desbalanceamento dos registros de turbulência MOS (5% “SIM” e 95% “NÃO”) a maior parte dos modelos ficou enviesado para a classe majoritária, demonstrado pela concentração no canto inferior esquerdo da curva ROC, com alta taxa de PODn (acima de 0,90) e reduzida PODy (menor que 0,30). O melhor resultado foi obtido com o BayesNet encontrando-se uma PODy e PODn de 0,543 e 0,898 respectivamente.

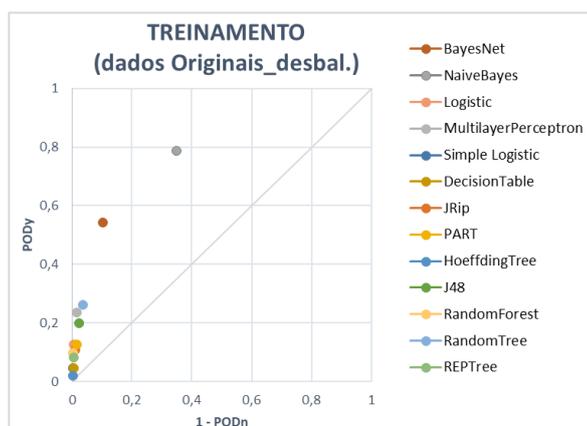
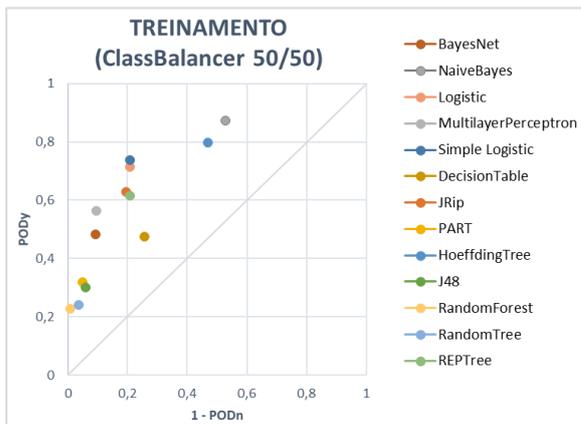
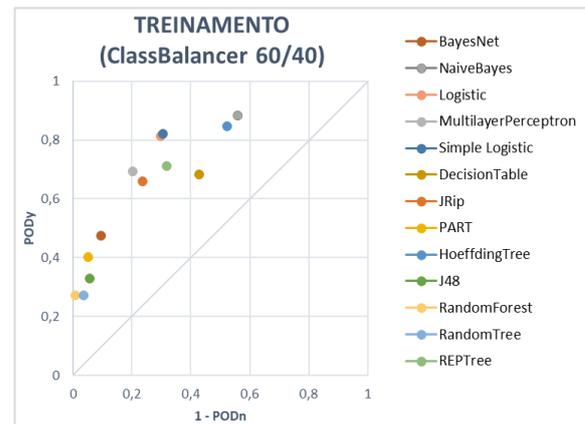


Figura 37: Curva ROC com valores de PODy versus 1-PODn dos resultados do treinamento dos algoritmos de AM com os dados originais.

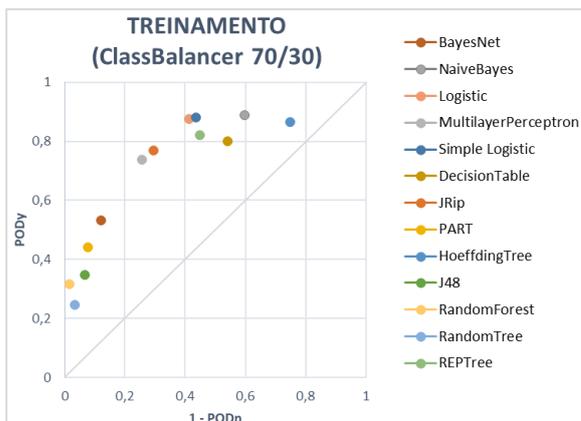
Realizando correção de prevalência dos dados modificando-se os “pesos” (*ClassBalancer*) dos registros de “SIM” e “NÃO”, para as proporções de 50/50, 60/40, 70/30, 80/20, 90/10 e 96/04, foram realizados setenta e oito experimentos. Observa-se na Figura 38 uma melhora na performance geral dos modelos em comparação aos resultados obtidos no treinamento utilizando os dados originais. Porém à medida que o “peso” atribuído aos registros “SIM” foram aumentando, os modelos gradativamente começaram a ficar tendenciosos a esse dado perdendo qualidade na detecção do “NÃO”, deslocando-se da parte inferior esquerda da curva ROC para o canto superior direito. O melhor resultado encontrado com esse método foi obtido utilizando a proporção “SIM”/“NÃO” de 96/04 (Figura 38f), em que o modelo com o RandomForest alcançou uma PODy de 0,824 e PODn de 0,789.



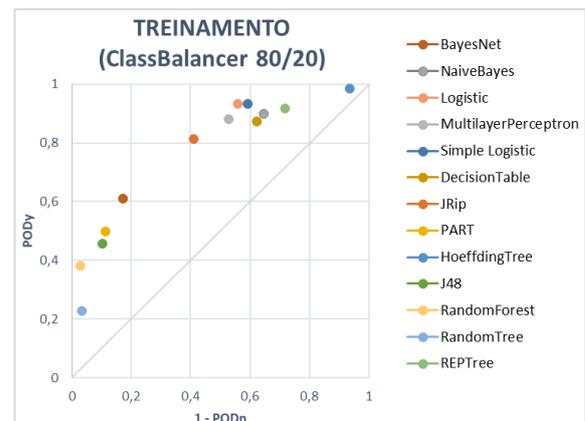
(a)



(b)



(c)



(d)

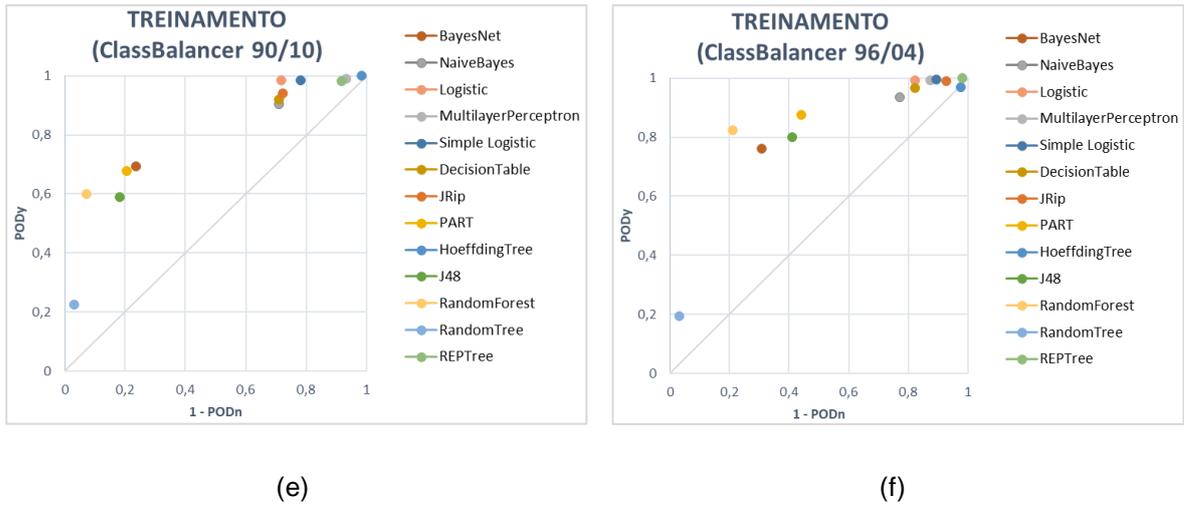
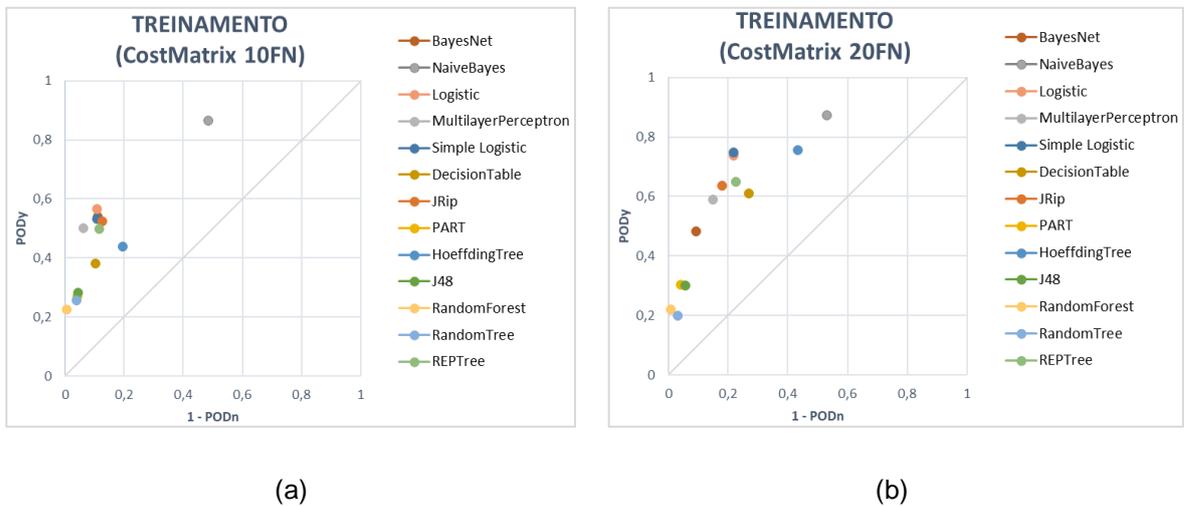
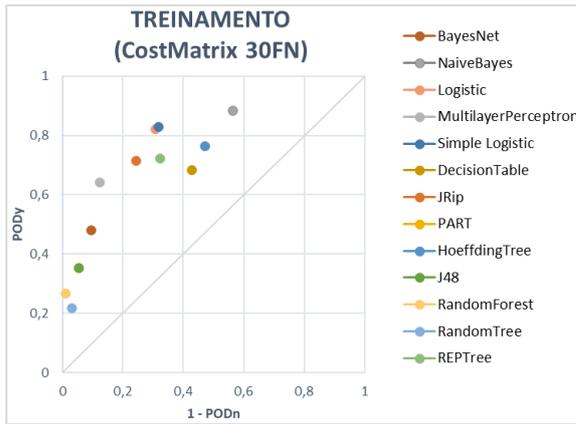


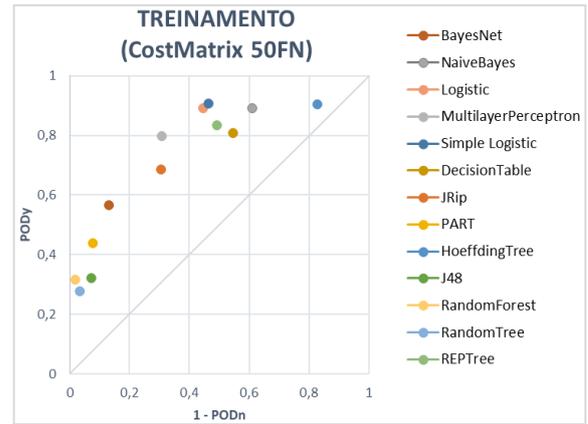
Figura 38: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM balanceando artificialmente os dados com o *ClassBalancer* nas proporções de evento e não-evento: a) 50/50; b) 60/40; c) 70/30; d) 80/20; e) 90/10; f) 96/04.

Utilizando a matriz de custo para a correção de prevalência nos dados, atribuindo-se “peso” ao falso negativo (FN) nas proporções de 10, 20, 30, 50, 100, 200 e 380, foram realizados noventa e um experimentos. Nota-se na Figura 39 uma tendência semelhante àquela apresentada com o uso do *ClassBalancer*, apresentando uma concentração inicial dos modelos na parte inferior esquerda da curva ROC e um deslocamento para o canto superior direito conforme incremento do “peso”, aumentando a PODy e diminuindo a PODn. O melhor resultado desse método foi obtido com um “peso” de 380 ao FN, em que o modelo com o RandomForest conseguiu uma PODy e PODn de 0,820 e 0,826, respectivamente (Figura 39g).

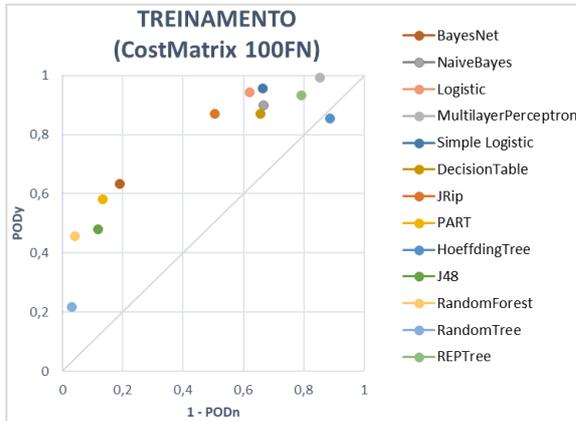




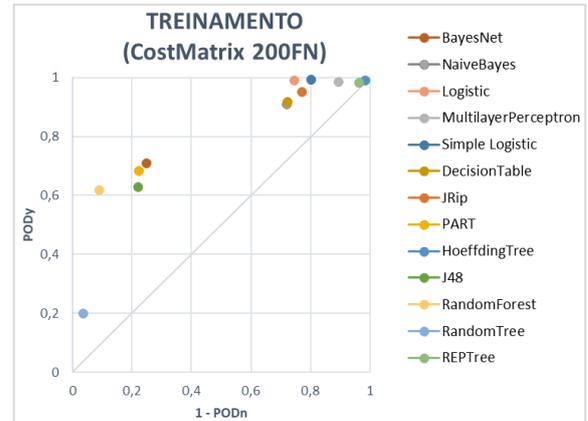
(c)



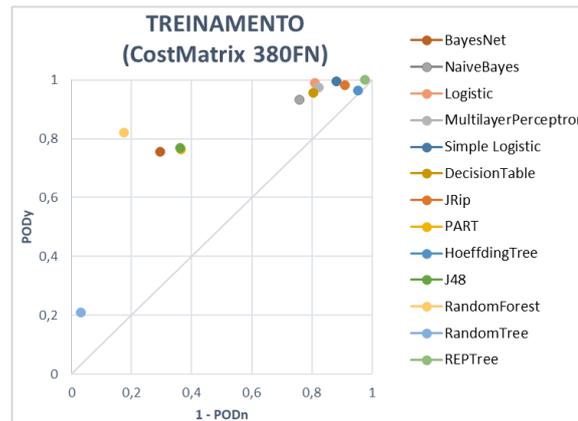
(d)



(e)



(f)



(g)

Figura 39: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM via matriz de custo, atribuindo pesos ao FN de: a) 10; b) 20; c) 30; d) 50; e) 100; f) 200; g) 380.

Realizando a retirada aleatória de dados da classe majoritária visando o equilíbrio estatístico das classes, foram removidos 4810 registros “NÃO”, restando um total de 534 dados horários, divididos igualmente entre “SIM” e “NÃO”. Verifica-se na Figura 40 que todos os modelos tiveram uma performance de PODy e PODn acima dos 70%, com valores muito próximos entre si. O melhor resultado obtido com esse método foi com o modelo utilizando RandomForest, atingindo uma PODy de 0,839 e PODn de 0,787.

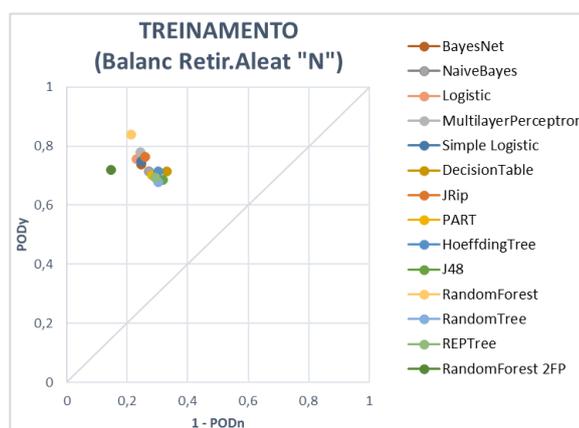
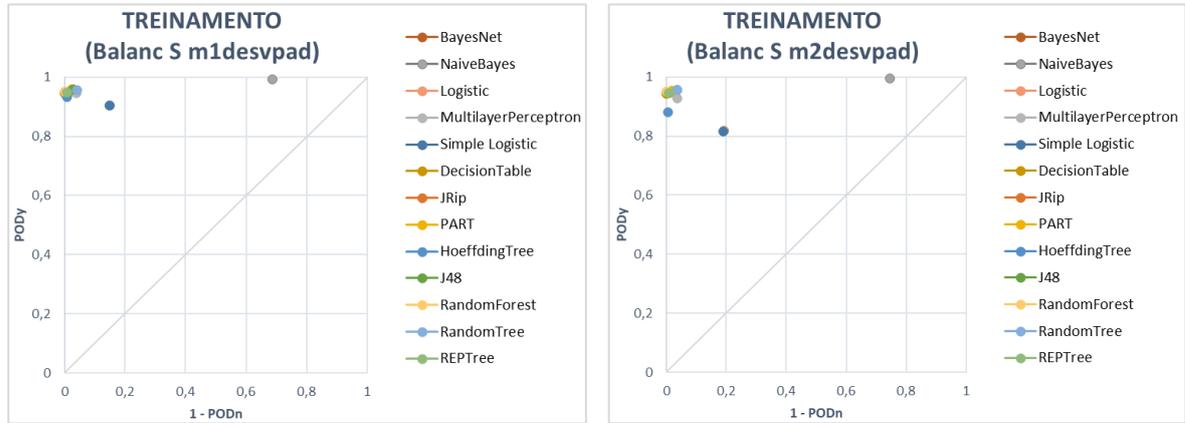


Figura 40: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM balanceando os dados com a retirada aleatória de dados “NÃO”. Random Forest 2FP é o modelo utilizando matriz de custo com “peso” 2 atribuído ao Falso Positivo (FP).

Para o equilíbrio estatístico das ocorrências mediante inserção de novos dados da classe minoritária, foram adicionados 4810 registros “SIM”, mediante três procedimentos (passo xx da Seção 3.3) que foram aplicados a cada um dos 184 atributos da Tabela 11 do Apêndice A, totalizando 10688 dados horários, divididos igualmente entre “SIM” e “NÃO”: 1) média e um desvio-padrão: foram inseridos novos valores de forma aleatória baseados na média e um desvio-padrão para mais e para menos de cada atributo; 2) média e dois desvios-padrões: foram inseridos novos valores de forma aleatória baseados na média e dois desvios-padrões para mais e para menos de cada atributo; 3) mediana e quartis: foram inseridos novos valores de forma aleatória baseados na mediana e quartis inferior e superior.

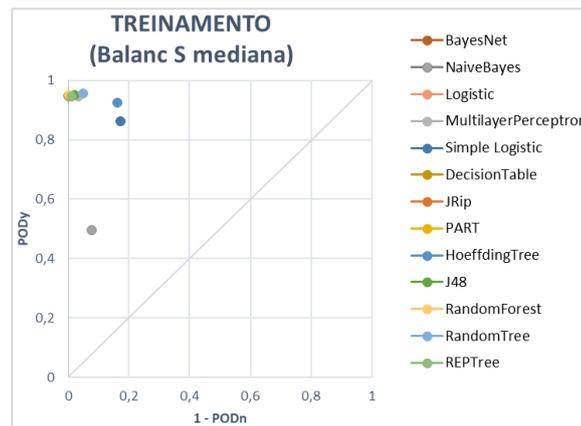
Verifica-se na Figura 41 que nas três formas de inserção de dados “SIM” os modelos apresentaram uma concentração dos resultados durante o treinamento no canto superior esquerdo da curva ROC, evidenciando altas taxas tanto de PODy quanto PODn, chegando muito próximo do que seria um modelo perfeito. Nas Figuras 41a-

b, os melhores resultados foram obtidos com o RandomForest. Na Figura 41c, dois modelos tiveram a mesma performance, BayesNet e RandomForest.



(a)

(b)



(c)

Figura 41: Curva ROC com valores de PODy versus 1-PODn com os resultados do treinamento dos algoritmos de AM balanceando os dados com a inserção aleatória de dados “SIM” baseados: a) média e um desvio-padrão de cada atributo; b) média e dois desvios-padrões de cada atributo; c) mediana e quartis inferior e superior de cada atributo.

Na Tabela 4 tem-se um resumo com os melhores resultados obtidos durante o treinamento dos algoritmos de AM, identificados para cada método, tipo de algoritmo, PODy, PODn e TSS.

Tabela 4: Resumo com os melhores resultados obtidos durante o treinamento dos algoritmos de AM.

Método	Algoritmo	PODy	PODn	TSS
Dados Originais	BayesNet	0,543	0,898	0,441
<i>ClassBalancer</i> 96/04	Random Forest	0,824	0,789	0,613
Matriz de Custo 380FN	Random Forest	0,820	0,826	0,646
Retir. Aleat. “NÃO”	Random Forest	0,839	0,787	0,626
“SIM” m1desvpad	Random Forest	0,952	1,000	0,952
“SIM” m2desvpad	Random Forest	0,951	0,999	0,950
“SIM” mediana	BayesNet	0,948	1,000	0,948
	Random Forest	0,950	0,998	0,948

Dentre os métodos que manipulam artificialmente o conjunto de dados, a melhor performance foi atingida utilizando a Matriz de Custo 380FN, seguida pelo *ClassBalancer* 96/04 e a pior com os dados originais. Modificando diretamente o conjunto de dados, o melhor resultado foi com a inserção de dados da classe minoritária “SIM” baseada na média e um desvio-padrão, seguido pela inserção de “SIM” baseada na média e dois desvios-padrões, inserção de “SIM” baseada na mediana e quartis e a pior com a retirada de dados da classe majoritária “NÃO”.

4.4 Estudo de Caso

Nesta seção são apresentadas vinte e quatro análises de casos referentes aos dias 02 de março de 2018, 06 de maio de 2018 e 28 de setembro de 2018, avaliando-se o comportamento dos modelos preditivos ótimos da Tabela 4 e do melhor método objetivo da Seção 4.2 na previsão da turbulência MOS.

Os casos foram selecionados de forma que os modelos pudessem ser avaliados em três situações distintas do estado termodinâmico da atmosfera: 1) dia com todos os horários não-favoráveis para turbulência; 2) dia em que houvesse uma mudança do estado termodinâmico da atmosfera, passando de não-favorável para favorável à turbulência; e 3) dia com todos os horários favoráveis para turbulência.

O dia 02 de março de 2018 foi selecionado como dia não-favorável, pois ele era o dia mediano de um total de três dias consecutivos caracterizados pela ausência de dados de VRTG em todos os horários. O dia 06 de maio de 2018 foi selecionado como dia em que houve mudança no estado termodinâmico, pois nos dois dias anteriores não foram registrados dados de VRTG em nenhum horário e nos dois dias posteriores sim. O dia 28 de setembro de 2018 foi selecionado como dia favorável, pois ele era o dia mediano de um total de três dias consecutivos caracterizados pela presença de dados de VRTG na maior parte dos horários.

Para caracterizar o estado termodinâmico da atmosfera nesses dias, foram utilizadas as imagens do GOES 16 (canal 13) e os dados do GFS realizando um corte vertical na latitude de 33,3°S entre as longitudes de 65°W a 75°W, variando de 1000 hPa a 100 hPa, apresentando: vento, temperatura potencial, velocidade vertical e cisalhamento vertical do vento.

4.4.1 Caso I: 02 de março de 2018

Na Tabela 5 estão os oito dados de turbulência MOS para o dia 02 de março de 2018. Nota-se que durante todo o período não foram registrados VRTG classe 2 ou 3, o que poderia indicar a não-ocorrência da turbulência MOS para todos os horários desse dia.

Tabela 5: Registros horários de turbulência MOS para o dia 02 de março de 2018. Dados a cada 3h de acordo com as saídas do GFS.

Hora (Z)	Turbulência MOS
0	NÃO
3	NÃO
6	NÃO
9	NÃO
12	NÃO
15	NÃO
18	NÃO
21	NÃO

Analisando-se as imagens do GOES-16, canal 13, verifica-se que não houve a formação de nuvens convectivas, excluindo-se a possibilidade de formação da CIT. Na Figura 42 tem-se a imagem das 18Z, onde constata-se a presença de poucas nuvens isoladas no setor norte da área de estudo de tonalidade azul (temperatura de brilho entre -30°C e -40°C).

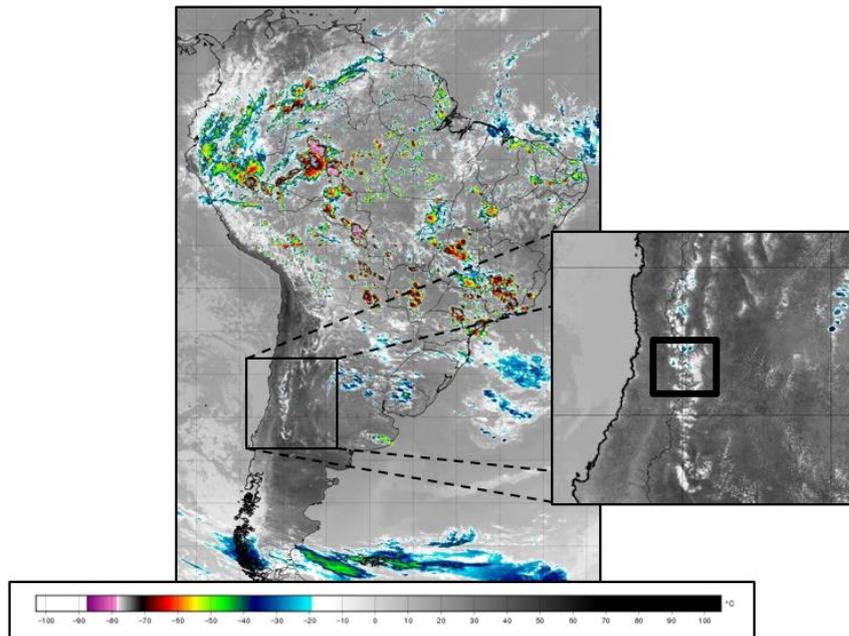
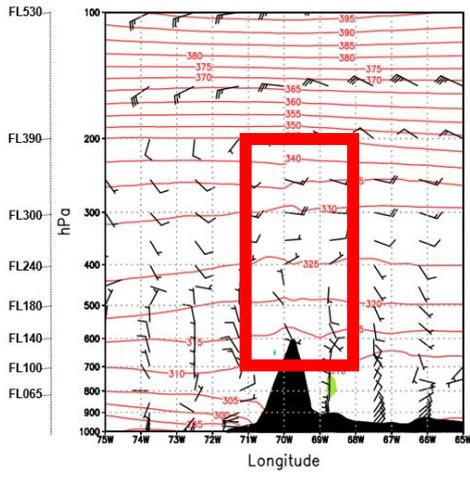
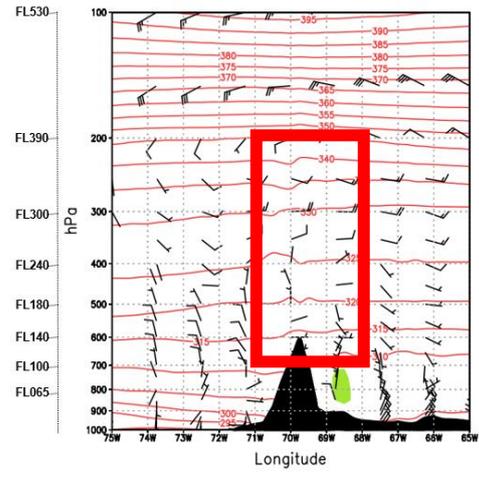


Figura 42: Imagem GOES-16, canal 13 (banda termal) das 18Z do dia 02 de março de 2018. No retângulo preto em destaque, a área de estudo. Fonte: adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>.

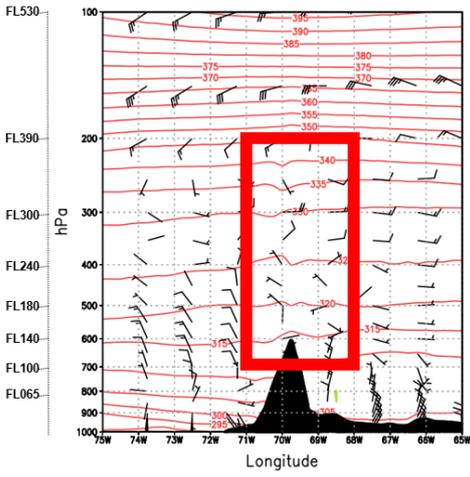
Analisando o comportamento termodinâmico da atmosfera, verifica-se na Figura 43 que dentro do retângulo de interesse o vento tem velocidade menor ou igual a 20 kt, variando de direção ao longo das horas. Como consequência tem-se um ambiente não favorável para a formação de ondas orográficas, visto que as isolinhas de temperatura potencial praticamente não apresentam distúrbios e a velocidade vertical não é significativa, com valores entre $\pm 0,1$ m/s. Esse padrão atmosférico manteve-se durante todos os horários do dia.



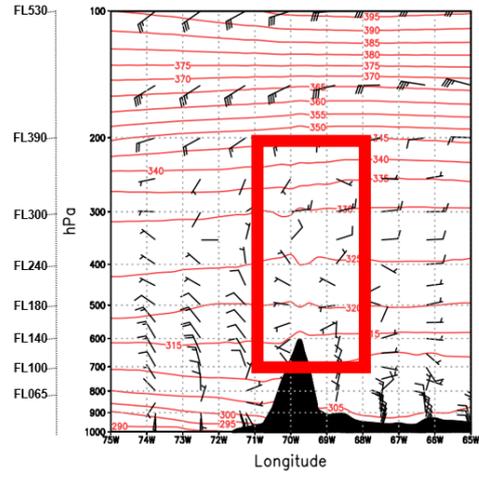
(a)



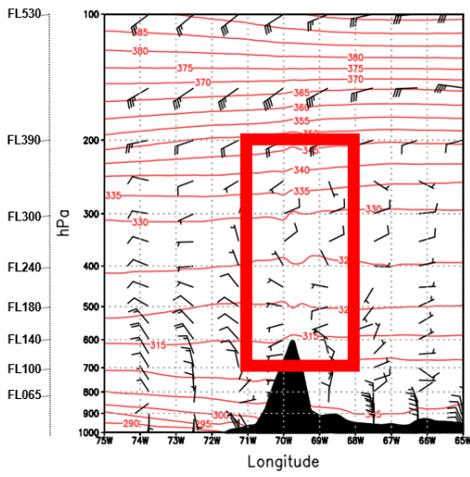
(b)



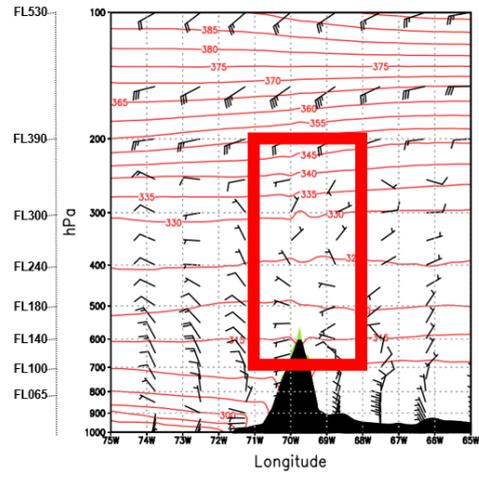
(c)



(d)



(e)



(f)

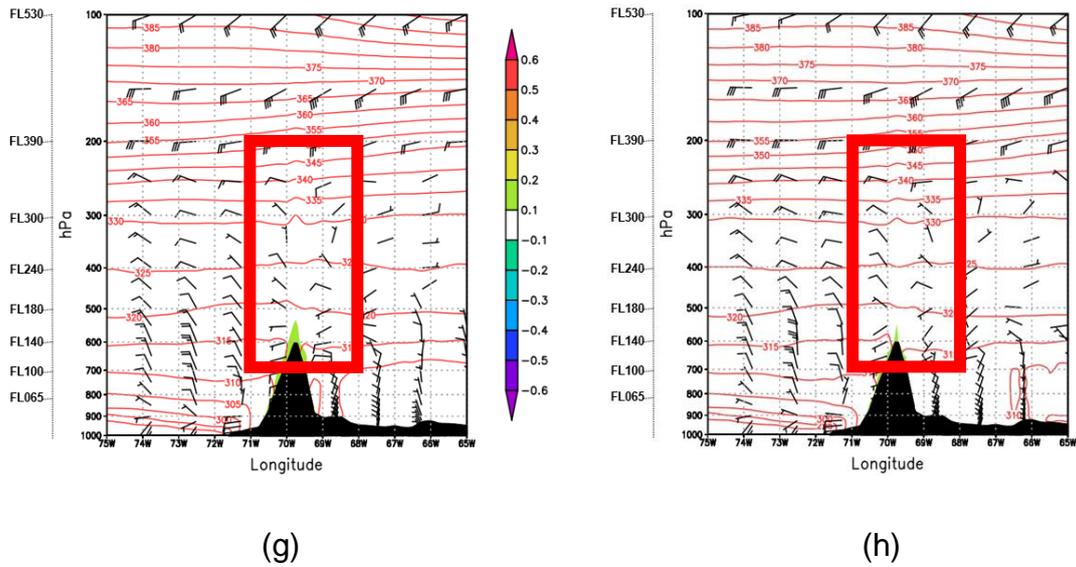
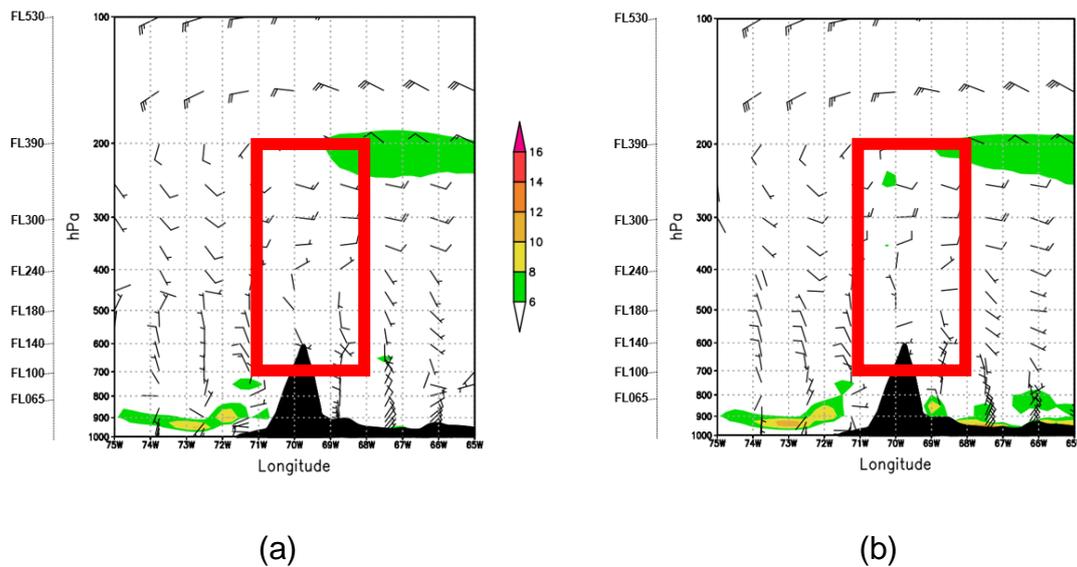
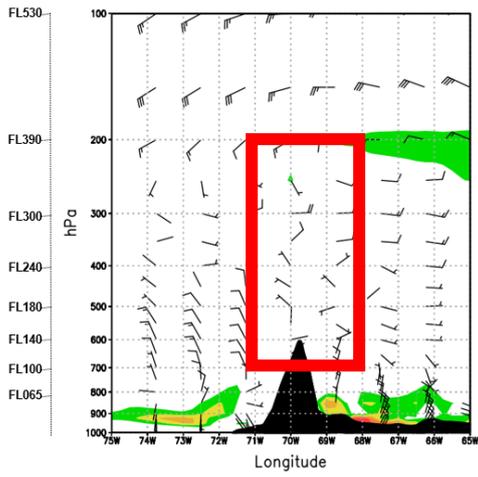


Figura 43: Representação do estado básico da atmosfera no dia 02 de março de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt), temperatura potencial (K) em isolinhas vermelhas e velocidade vertical (m/s) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z, e) 12Z; f) 15Z; g) 18Z; h) 21Z.

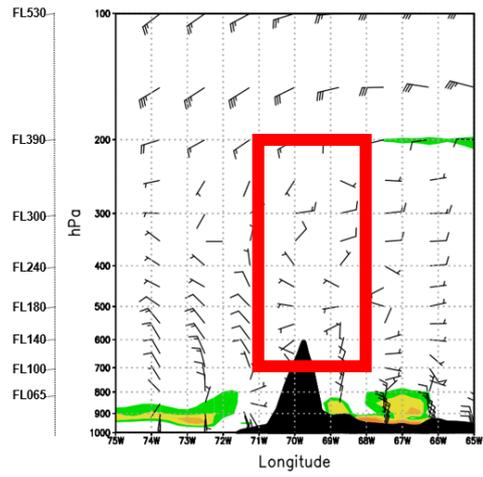
Retângulo de interesse em destaque.

Analisando o comportamento do cisalhamento vertical do vento, verifica-se na Figura 44 que dentro do retângulo de interesse ele não foi significativo, registrando valores menores que $6 \text{ (} 10^{-3} \text{ s}^{-1}\text{)}$, durante todos os horários do dia.

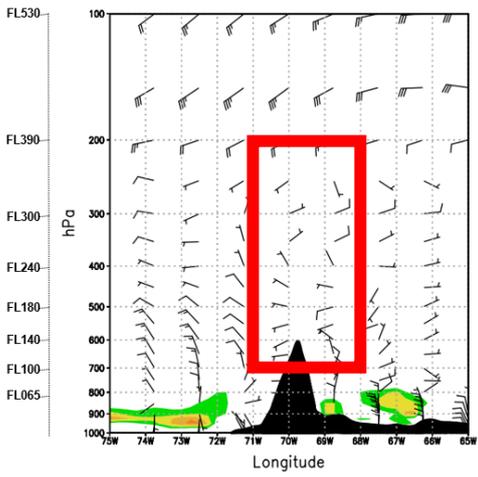




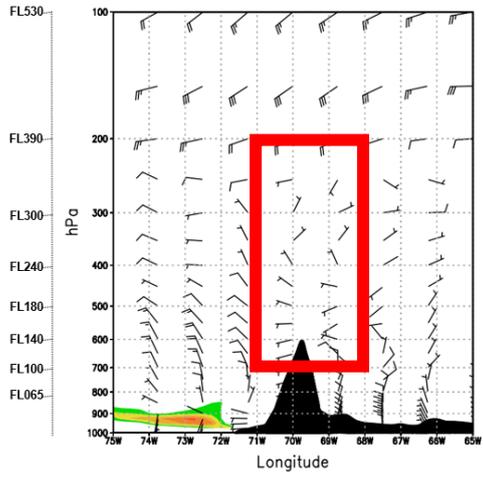
(c)



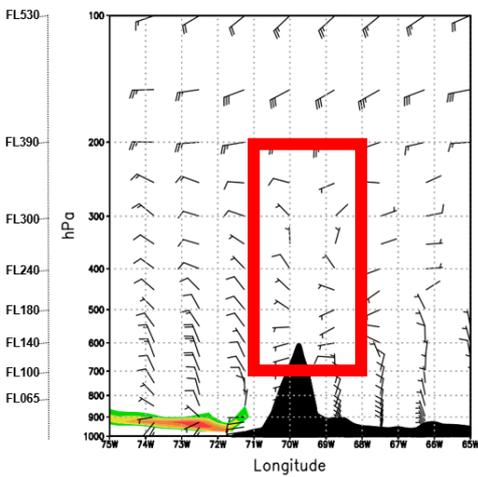
(d)



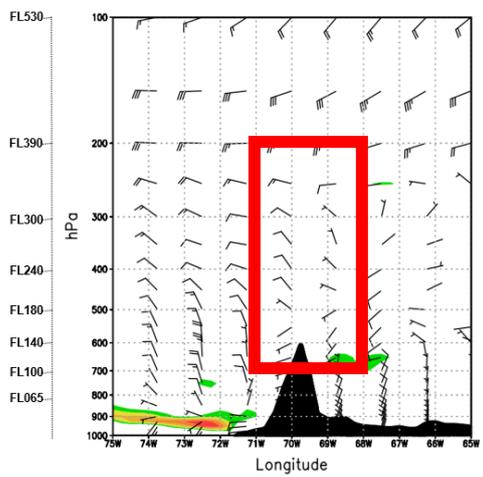
(e)



(f)



(g)



(h)

Figura 44: Representação do estado básico da atmosfera no dia 02 de março de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt) e cisalhamento vertical do vento ($10^{-3} s^{-1}$) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque.

Nota-se que o estado básico da atmosfera no dia 02 de março de 2018 não apresentou variações significativas ao longo das horas, indicando uma atmosfera não favorável para formação tanto de ondas orográficas quanto de cisalhamento vertical do vento. Por tanto, a ausência de dados de VRTG classe 2 e 3 durante todo o dia, pode significar realmente que não houve turbulência MOS.

Avaliando-se os modelos preditivos ótimos, desenvolvidos neste trabalho, de turbulência MOS para o dia 02 de março de 2018, tem-se na Tabela 6 que todos tiveram a mesma performance, prevendo que para o dia em questão não haveria a possibilidade da turbulência em nenhum horário, estando de acordo com os dados observados e a análise do estado básico da atmosfera.

Tabela 6: Resultados dos modelos preditivos ótimos de turbulência MOS para o dia 02 de março de 2018 em comparação com os dados observados. Em cinza ausência da turbulência, em vermelho presença.

	0Z	3Z	6Z	9Z	12Z	15Z	18Z	21Z
OBSERVADO								
Índice 5 ($V_n \geq 0kt$; $DP \geq 4hPa$)								
Dados Originais (BayesNet)								
ClassBalancer 96/04 (RandomForest)								
Matriz de Custo 380FN (RandomForest)								
Retir. Aleat. "NÃO" (RandomForest)								
"SIM" m1desvpad (RandomForest)								
"SIM" m2desvpad (RandomForest)								
"SIM" mediana (BayesNet)								
"SIM" mediana (RandomForest)								

4.4.2 Caso II: 06 de maio de 2018

Na Tabela 7 estão os oito dados de turbulência MOS para o dia 06 de maio de 2018. Nota-se que somente nos horários das 18Z e 21Z foram registrados VRTG classe 2

e 3, o que poderia indicar que para este dia a atmosfera estava inicialmente não-turbulenta, das 0Z às 15Z, passando a turbulenta ao final do período.

Tabela 7: Registros horários de turbulência MOS para o dia 06 de maio de 2018. Dados a cada 3h de acordo com as saídas do GFS.

Hora (Z)	Turbulência MOS
0	NÃO
3	NÃO
6	NÃO
9	NÃO
12	NÃO
15	NÃO
18	SIM
21	SIM

Analisando-se as imagens do GOES-16, canal 13, verifica-se que não houve a formação de nuvens convectivas, excluindo-se a possibilidade de formação da CIT. Na Figura 45 tem-se a imagem das 18Z, onde constata-se a presença de poucas nuvens esparsadas em tom branco (temperatura de brilho entre -10°C e -20°C) no setor sudoeste da área de estudo e uma camada de nuvens em tom cinza claro (temperatura de brilho entre 0°C e 10°C) no setor leste.

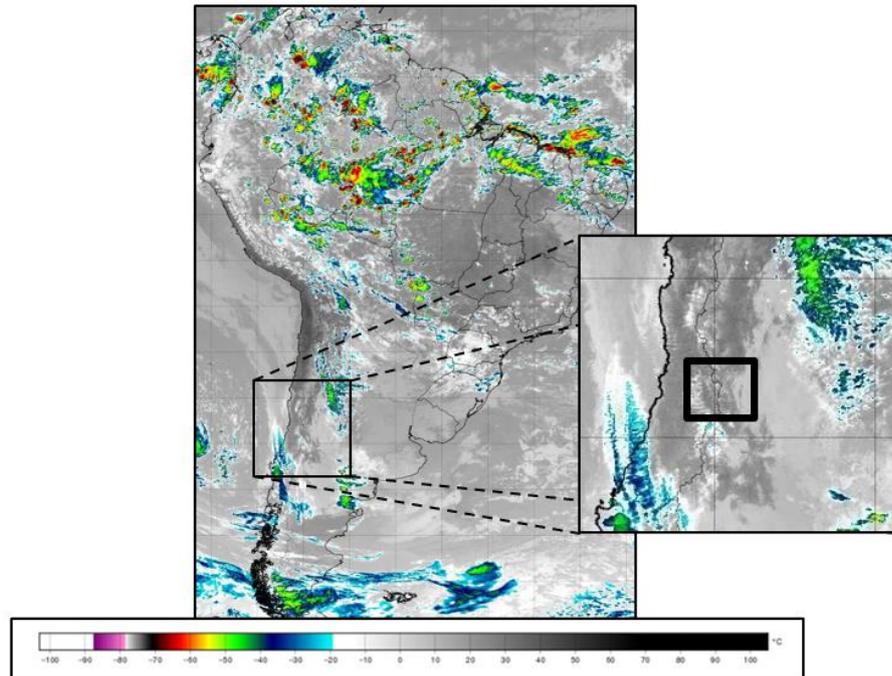
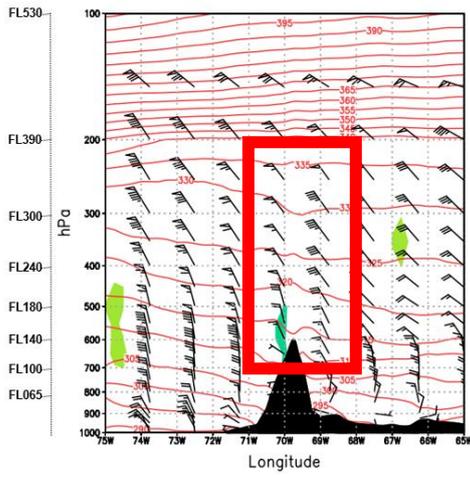
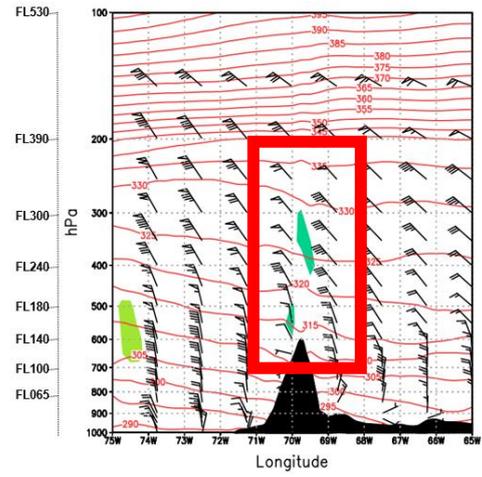


Figura 45: Imagem GOES-16, canal 13 (banda termal das 18Z do dia 06 de maio de 2018). No retângulo preto em destaque, a área de estudo. Fonte: adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>.

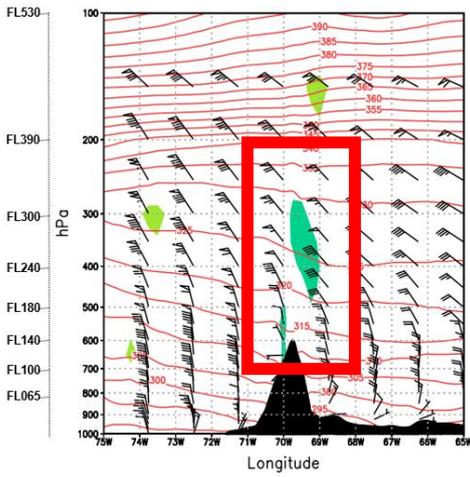
Analisando o comportamento termodinâmico da atmosfera, verifica-se na Figura 46 que dentro do retângulo de interesse o vento manteve uma direção predominante de noroeste, aumentando de intensidade com o passar das horas. Entre os FL100-180 a velocidade foi de até 30 kt, enquanto acima deste nível atingiu pico de 90 kt às 21Z. Por consequência, a componente zonal do vento sofreu acréscimo ao longo do dia, chegando ao maior valor ao final do período, quando então possibilitou a formação de ondas orográficas, que podem ser observadas pelos distúrbios nas linhas de temperatura potencial que ganharam amplitude a partir das 12Z, e principalmente nos horários das 18Z e 21Z, favorecidas pelo incremento da velocidade vertical alternando entre correntes ascendentes e descendentes em cima e à sotavento dos Andes com valores máximos de 0,2 m/s e - 0,3 m/s. Houve assim, uma mudança no padrão atmosférico ao longo do dia, passando de uma condição inicial não favorável para turbulenta com o passar das horas.



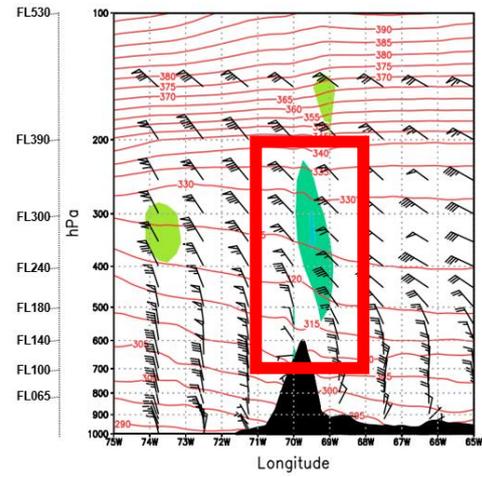
(a)



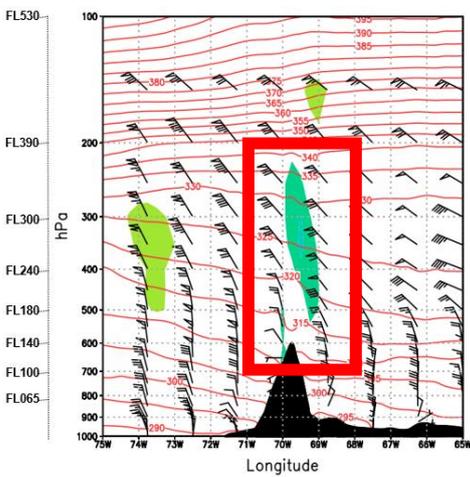
(b)



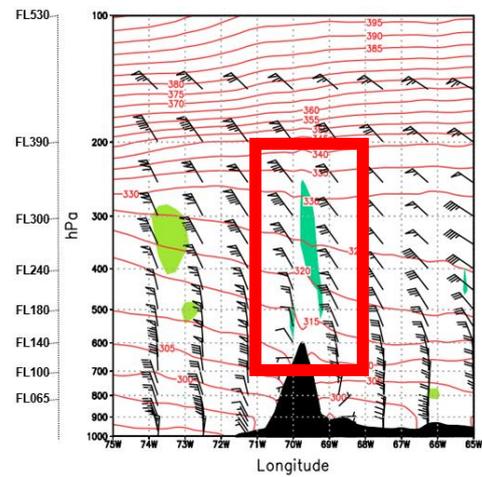
(c)



(d)



(e)



(f)

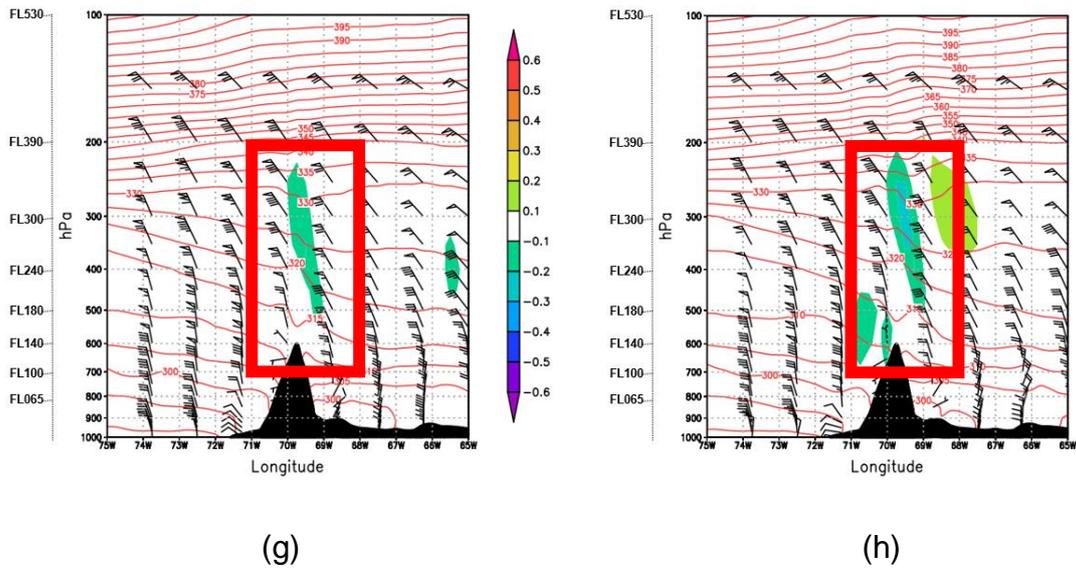
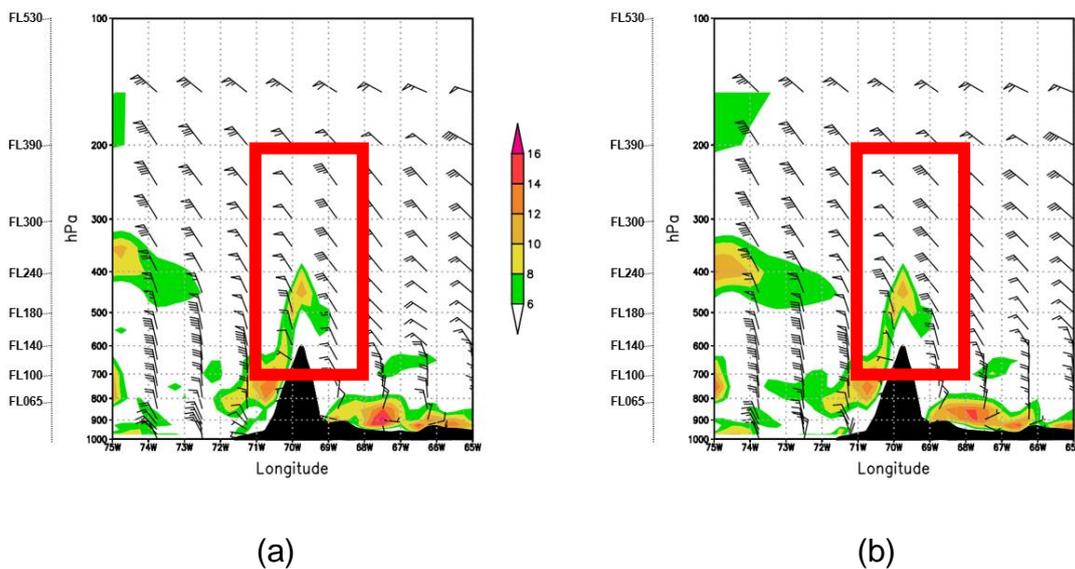
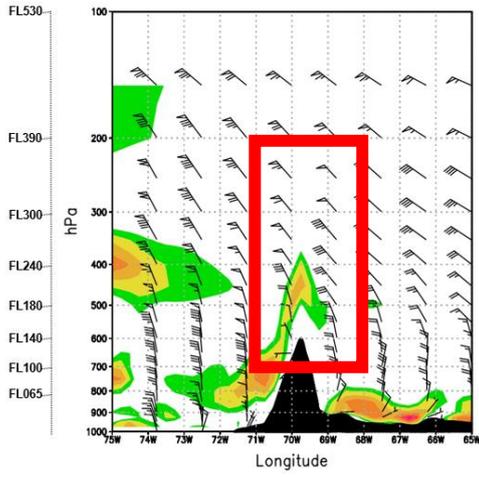


Figura 46: Representação do estado básico da atmosfera no dia 06 de maio de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt), temperatura potencial (K) em isolinhas vermelhas e velocidade vertical (m/s) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z, e) 12Z; f) 15Z; g) 18Z; h) 21Z.

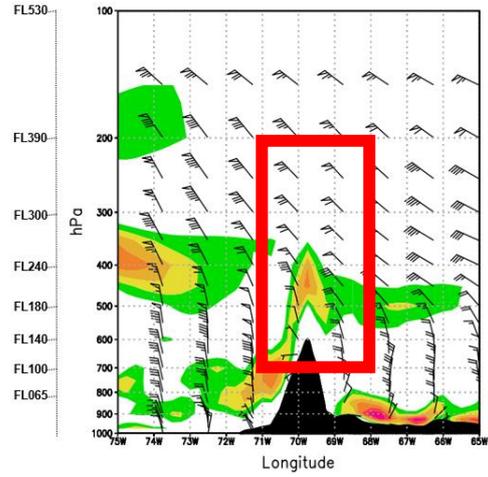
Retângulo de interesse em destaque.

Analisando o comportamento do cisalhamento vertical do vento, verifica-se na Figura 47 que dentro do retângulo de interesse ele foi mais significativo entre os FL140-240, apresentando valores maiores que $8 (10^{-3} \text{ s}^{-1})$ em todos os horários. Porém por consequência da intensificação dos ventos com o passar das horas, o cisalhamento vertical do vento também aumentou chegando a mais de $12 (10^{-3} \text{ s}^{-1})$ a partir das 12Z, atingindo o pico de mais de $14 (10^{-3} \text{ s}^{-1})$ nos horários das 18Z e 21Z.

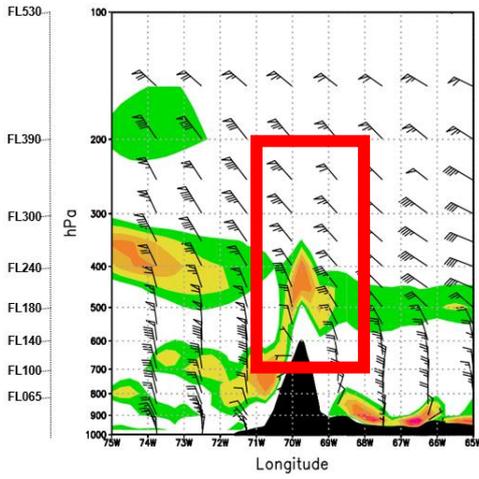




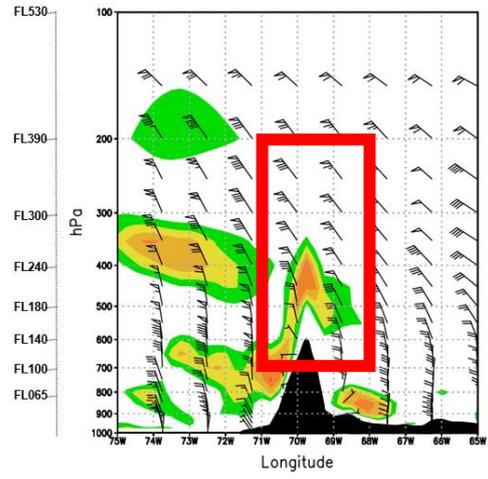
(c)



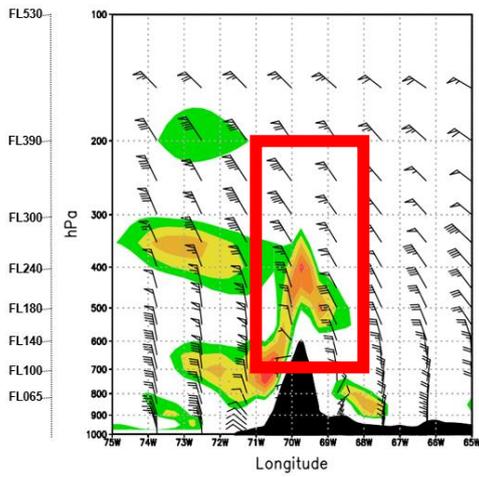
(d)



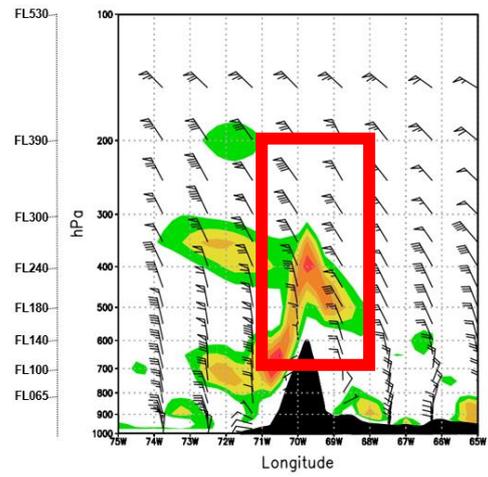
(e)



(f)



(g)



(h)

Figura 47: Representação do estado básico da atmosfera no dia 06 de maio de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt) e cisalhamento vertical do vento (10^{-3}s^{-1}) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque.

Nota-se que o estado básico da atmosfera no dia 06 de maio de 2018 sofreu alteração ao longo das horas devido a intensificação dos ventos, o que favoreceu a formação de ondas orográficas e o aumento do cisalhamento vertical do vento, saindo de uma condição não-turbulenta para turbulenta. Por tanto, a ausência de dados de VRTG classe 2 e 3 no período das 00Z às 09Z, pode significar realmente que não houve turbulência MOS. Porém a ausência daqueles dados nos horários das 12Z e 15Z pode significar que houve turbulência MOS, mas ela pode não ter sido registrada, seja pela não realização de um voo no horário, ou o voo foi realizado num FL menos propício para o grau de severidade da turbulência.

Tem-se na Tabela 8 que os quatro modelos preditivos baseados na inserção aleatória de dados “SIM” e o modelo objetivo (índice 5) não conseguiram identificar a tendência de variação do estado básico da atmosfera, de não-turbulento para turbulento, não prevendo a ocorrência da turbulência MOS em nenhum período do dia. Todos os demais modelos foram hábeis nessa indicação, porém diferenciando-se quanto ao horário de início da turbulência MOS. O modelo baseado na matriz de custo 380FN (RandomForest) previu que a turbulência teria início às 18Z, concordando 100% com os dados observados e com o estado básico da atmosfera. Já os modelos baseados nos dados originais (BayesNet), *ClassBalancer* 96/04 (RandomForest) e retirada aleatória de dados “NÃO” (RandomForest) anteciparam o horário de início para às 12Z, discordando dos dados observados, mas mostrando-se possível, mesmo que em menor grau, conforme análise das Figuras 46 e 47.

Tabela 8: Resultados dos modelos preditivos ótimos de turbulência MOS para o dia 06 de maio de 2018 em comparação com os dados observados. Em cinza ausência da turbulência, em vermelho presença.

	0Z	3Z	6Z	9Z	12Z	15Z	18Z	21Z
OBSERVADO								
Índice 5 ($V_n \geq 0kt$; $DP \geq 4hPa$)								
Dados Originais (BayesNet)								
<i>ClassBalancer</i> 96/04 (RandomForest)								
Matriz de Custo 380FN (RandomForest)								
Retir. Aleat. "NÃO" (RandomForest)								
"SIM" m1desvpad (RandomForest)								
"SIM" m2desvpad (RandomForest)								
"SIM" mediana (BayesNet)								
"SIM" mediana (RandomForest)								

4.4.3 Caso III: 28 de setembro de 2018

Na Tabela 9 estão os oito dados de turbulência MOS para o dia 28 de setembro de 2018. Nota-se que somente nos horários das 06Z, 09Z e 12Z não havia dados de VRTG classe 2 ou 3, o que poderia indicar que este foi um dia com condições propícias para a ocorrência da turbulência MOS, porém intercalado com um período de atenuação do fenômeno.

Tabela 9: Registros horários de turbulência MOS para o dia 28 de setembro de 2018. Dados a cada 3h de acordo com as saídas do GFS.

Hora (Z)	Turbulência MOS
0	SIM
3	SIM
6	NÃO
9	NÃO
12	NÃO
15	SIM
18	SIM
21	SIM

Analisando-se as imagens do GOES-16, canal 13, verifica-se que não houve a formação de nuvens convectivas, excluindo-se a possibilidade de formação da CIT. Na Figura 48 tem-se a imagem das 18Z, onde constata-se a presença de uma camada de nuvens com tonalidade verde e azul (temperatura de brilho entre -40°C e -50°C), cobrindo grande parte da área de estudo, indicando serem de estágio médio e alto.

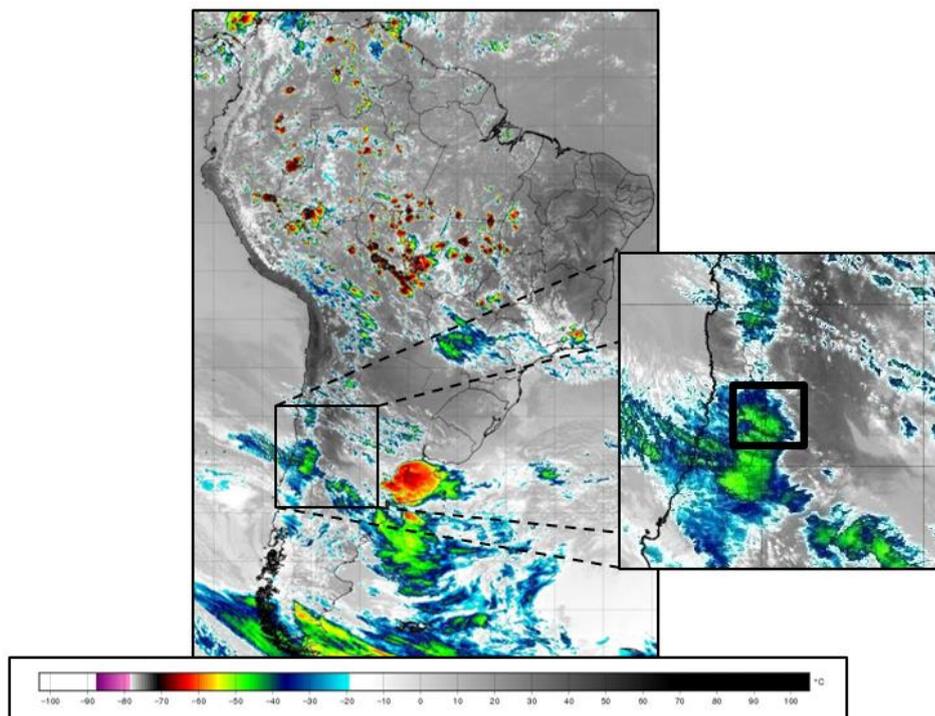


Figura 48: Imagem GOES-16, canal 13 (banda termal) das 18Z do dia 28 de setembro de 2018. No retângulo preto em destaque, a área de estudo. Fonte: adaptado de CPTEC, disponível em <http://satelite.cptec.inpe.br/home/index.jsp>.

Analisando o comportamento termodinâmico da atmosfera, verifica-se na Figura 49 que dentro do retângulo de interesse o vento acima do FL100 tem uma direção predominante de oeste com uma velocidade maior ou igual a 20 kt e picos de mais de 100 kt. Esses ventos com forte componente perpendicular ao relevo favorecem a formação de ondas orográficas, que podem ser observadas ao longo da seção vertical pelos distúrbios de grande amplitude apresentados pelas linhas de temperatura potencial, intensificados pela velocidade vertical alternando entre correntes ascendentes e descendentes em cima e à sotavento dos Andes com valores extremos superiores a $\pm 0,6$ m/s. Esse padrão atmosférico manteve-se durante todos os horários do dia.

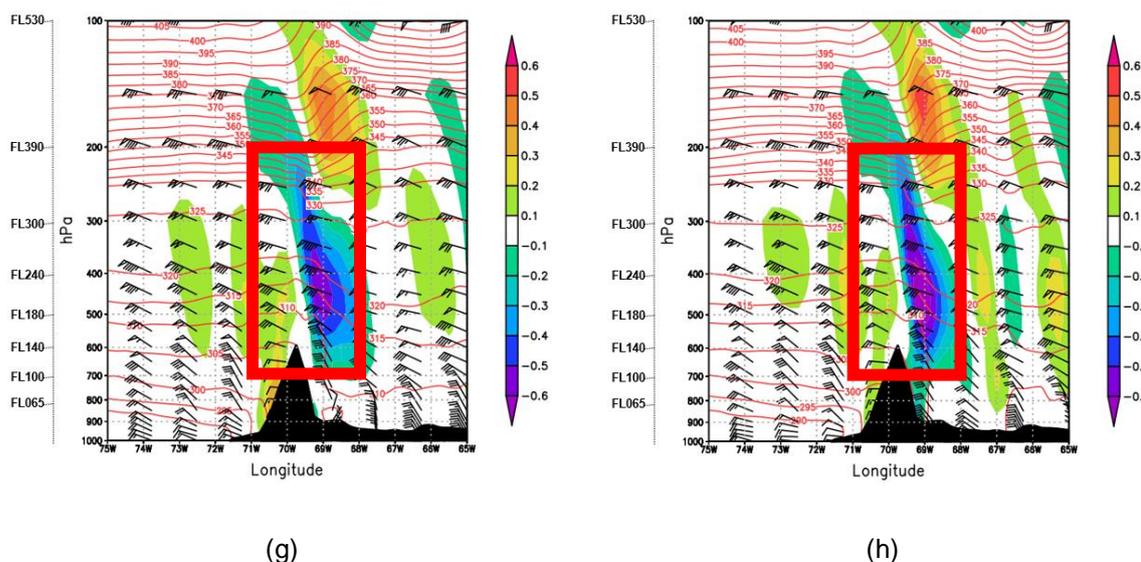
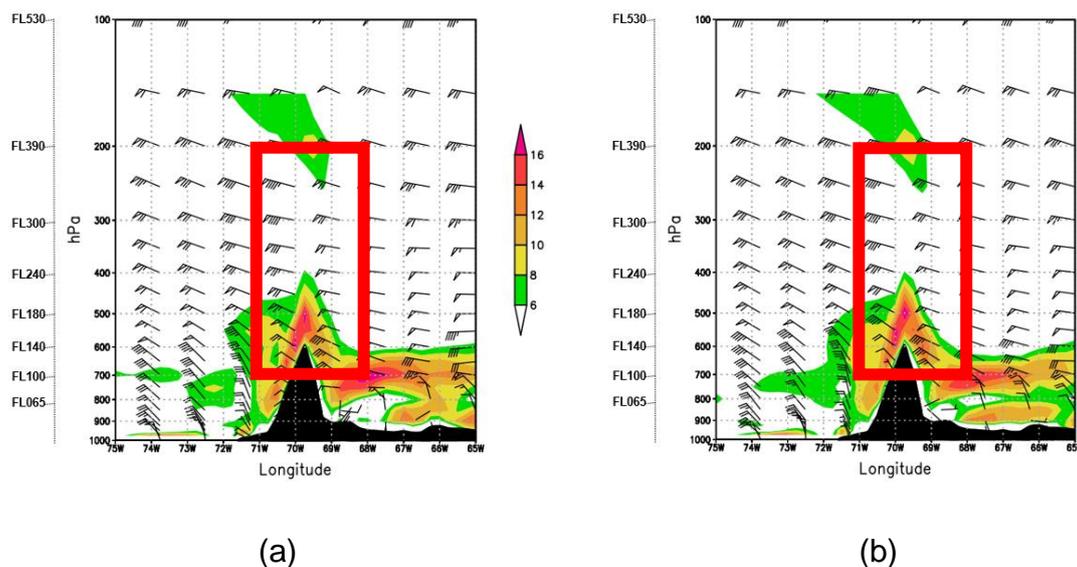
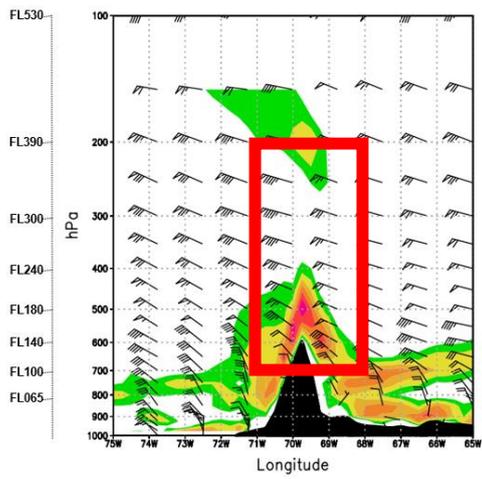


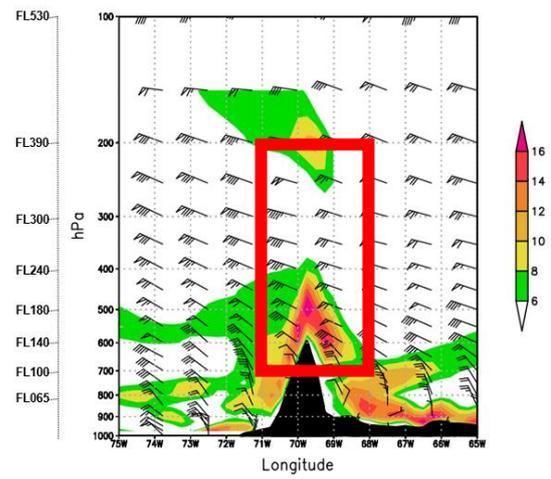
Figura 49: Representação do estado básico da atmosfera no dia 28 de setembro de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt), temperatura potencial (K) em isolinhas vermelhas e velocidade vertical (m/s) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque.

Analisando o comportamento do cisalhamento vertical do vento, verifica-se na Figura 50 que dentro do retângulo de interesse ele foi significativo próximo do FL390, variando de 6 a 12 (10^{-3} s^{-1}) e principalmente entre os FL100-240, onde apresentou valores maiores que 14 (10^{-3} s^{-1}) e máximos acima de 16 (10^{-3} s^{-1}) durante todos os horários do dia.

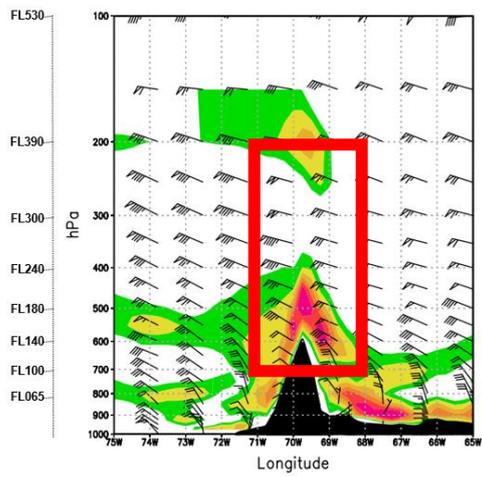




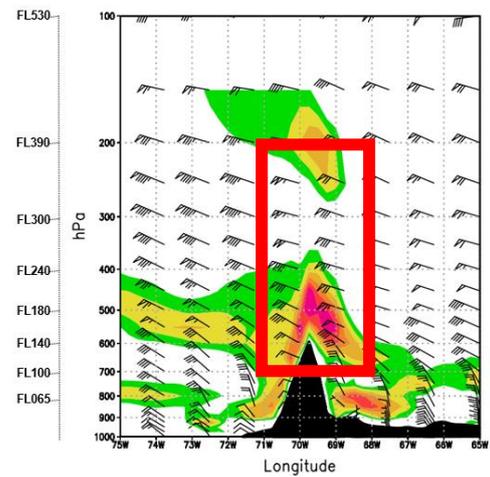
(c)



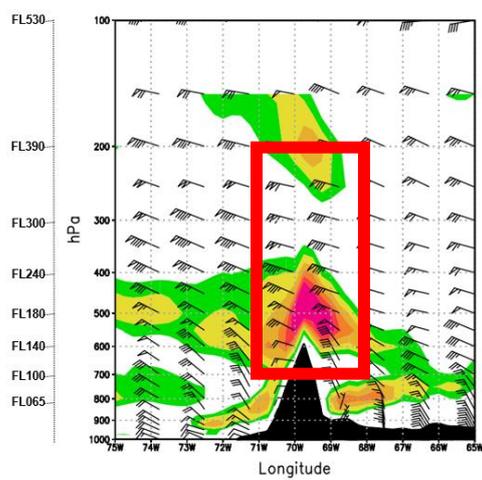
(d)



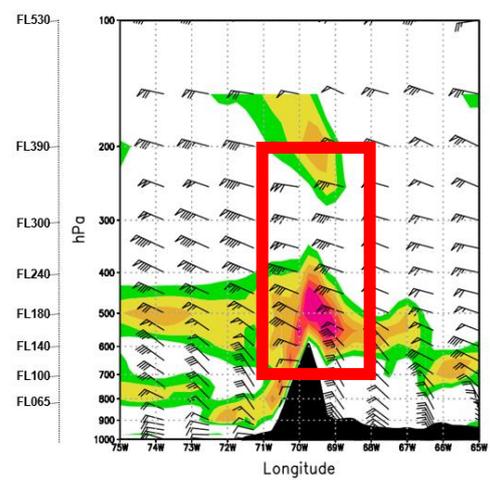
(e)



(f)



(g)



(h)

Figura 50: Representação do estado básico da atmosfera no dia 28 de setembro de 2018 com dados do GFS. Corte vertical em 33,3°S entre 65°W e 75°W e de 1000hPa a 100hPa, com correspondente FL, apresentando vento (kt) e cisalhamento vertical do vento (10^{-3}s^{-1}) em hachurado colorido para: a) 0Z; b) 3Z; c) 6Z; d) 9Z; e) 12Z; f) 15Z; g) 18Z; h) 21Z. Retângulo de interesse em destaque.

Nota-se que o estado básico da atmosfera no dia 28 de setembro de 2018 não apresentou variações significativas ao longo das horas, sendo influenciado tanto pelas ondas orográficas quanto pelo cisalhamento vertical do vento, com ambos concorrendo para a geração da turbulência MOS. Por tanto, a ausência de dados de VRTG classe 2 e 3 no período das 06Z às 12Z pode significar que houve turbulência MOS, mas que ela apenas não foi registrada, seja pela não realização de um voo no horário, ou o voo foi realizado num FL menos propício para o grau de severidade da turbulência.

Observa-se na Tabela 10 que os quatro modelos preditivos baseados na inserção aleatória de dados “SIM” não conseguiram identificar a ocorrência da turbulência MOS em nenhum horário. Todos os outros modelos foram efetivos na previsão da turbulência, diferenciando-se quanto a possibilidade de ter algum horário de atenuação do fenômeno ou não. O modelo objetivo (índice 5) previu que não haveria turbulência das 3 às 6Z e o baseado nos dados originais (BayesNet) que não teria somente às 3Z. Já os modelos baseados no *ClassBalancer* 96/04 (RandomForest), matriz de custo 380FN (RandomForest) e retirada aleatória de dados “NÃO” (RandomForest) previram que a turbulência MOS ocorreria durante todo o dia, apresentando as melhores performances, indo de acordo com a análise do estado básico da atmosfera (Figuras 49 e 50).

5. CONCLUSÕES

Neste trabalho são investigados os parâmetros característicos da turbulência em voo na rota Santiago (Chile) – Mendoza (Argentina), avalia-se a performance do Ábaco de Harrison como ferramenta objetiva de previsão de turbulência em voo para regiões montanhosas, e desenvolve-se um modelo de previsão de turbulência MOS com até 24 horas de antecedência, baseado em técnicas de AM. A pesquisa cobre um período de 22 meses, de março de 2018 a dezembro de 2019, utilizando registros de VRTG das aeronaves da LATAM. As principais conclusões são as seguintes.

Ao analisar os registros de VRTG, verifica-se que a turbulência atmosférica afeta principalmente a realização do voo entre os FL100-300, impactando as aeronaves em todos os setores do cruzamento da cordilheira dos Andes. Apresenta uma periodicidade, com máximos ocorrendo durante o inverno por influência da intensificação dos ventos predominantes de oeste nesta época, e ocorre ao longo do dia sem um aparente horário preferencial. A turbulência MOS é responsável por mais de 15% dos casos totais.

Os parâmetros atmosféricos que se mostraram mais significativos para a determinação do estado termodinâmico da atmosfera propício para a geração da turbulência MOS, foram as variações horizontais da temperatura potencial e da velocidade vertical, vento zonal, cisalhamento vertical do vento e Richardson Bulk. Tais atributos estão relacionados aos distúrbios ondulatórios causados pela MWT e com os intensos cisalhamentos verticais ocorridos em uma CAT.

O uso de modelo objetivo baseado em Vn e DP (Ábaco de Harrison) para prognóstico da turbulência MOS, mostrou-se pouco eficaz para o uso operacional. Isso, pois apesar da maior parte dos registros acontecerem com Vn e DP positivos, são significativos os números quando DP é negativo, tanto com vento de leste quanto de oeste, tornando inviável definir um limiar do par de preditores com aceitável performance de PODy, PODn e TSS, dada a enorme gama de possibilidades. Sua extrema simplicidade torna-o incompatível com a complexidade atmosférica associada à turbulência MOS e com o necessário grau de confiabilidade de uma previsão visando o planejamento e segurança de voo.

Já com o emprego de técnicas de AM foi possível ampliar a dimensionalidade de parâmetros analisados de modo a compatibilizar com o grau de complexidade necessário para a boa previsibilidade do fenômeno. Os melhores resultados encontrados com essas técnicas foram obtidos por meio da manipulação artificial do conjunto de dados com uso de “pesos” (*ClassBalancer* e Matriz de Custo) e com a retirada aleatória de dados da classe majoritária “NÃO”, visando o balanceamento de classes. Tais procedimentos responderam com desempenhos expressivos, tanto no treinamento quanto nos estudos de caso, sendo esses três modelos ferramentas hábeis para o uso operacional.

A técnica de correção de prevalência via inserção aleatória de novos dados da classe minoritária “SIM”, apesar de ter gerado modelos com as melhores performances durante o treinamento, verificou-se com os estudos de casos que eles eram tendenciosos. O método acabou alterando sobremaneira o conjunto de dados, mediante a inserção de 18 vezes mais a quantidade de dados “SIM”, fazendo com que o novo conjunto não fosse representativo dos dados originais. Com isso, o modelo acabou não adquirindo a necessária capacidade de generalização. Para este estudo, essa técnica de balanceamento de dados mostrou-se inapropriada.

Esta pesquisa demonstra os benefícios e o quanto promissora é a aplicação de técnicas de AM ao problema da previsão da turbulência em voo, com boa capacidade de generalização e resultados acima de tradicionais ferramentas utilizadas. Espera-se que esse conhecimento possa contribuir para reduzir a lacuna dos estudos de turbulência em voo na América do Sul e ajudar pesquisadores e meteorologistas a identificar mais facilmente as situações de risco de ocorrência do fenômeno, possibilitando previsões mais assertivas visando a segurança da navegação aérea.

Para trabalhos futuros, sugere-se usar um período de estudo mais longo e incluir os dados de EDR, se disponíveis, visando a consistência estatística dos resultados. Analisar a efetividade da aplicação da técnica em áreas que tenham pouca influência de relevo. Sugere-se também comparar a previsão dos modelos propostos com outros modelos de turbulência que são utilizados operacionalmente.

6. REFERÊNCIAS

ALLOUCHE, O.; TSOAR, A.; KADMON, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). **Journal of Applied Ecology**, v. 43, n. 6, p. 1223–1232, 2006.

BARNES, J. **Microsoft Azure Essentials Azure Machine Learning**. Microsoft Press, 2015.

BOTELHO DE SOUZA, P. **Uma estratégia baseada em algoritmos de mineração de dados para validar plano de operação de voo a partir de previsões de estados dos satélites do INPE**. D.Sc. São José dos Campos, SP: Instituto Nacional de Pesquisas Espaciais (INPE), 2011. Disponível em: <http://mtc-m16d.sid.inpe.br/col/sid.inpe.br/mtc-m19/2011/04.15.19.12/doc/publicacao.pdf>

BROWN, B. G.; YOUNG, G. S. **Verification of icing and turbulence forecasts: Why some verification statistics can't be computed using PIREPs**. In: NINTH CONFERENCE ON AVIATION, RANGE, AND AEROSPACE METEOROLOGY. Orlando, FL: American Meteorological Society, 2000.

BROWN, R. New indices to locate clear air turbulence. **Meteorological Magazine**, n. 102, p. 347–361, 1973.

BROWNLEE, J. **Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch**. Vermont, VIC, Austrália: Machine Learning Mastery, 2016.

CARNEY, T. Q. et al. Hazardous Mountain Winds and Their Visual Indicators. **Advisory Circular**, v. 00–57, n. FAA, 1995.

CENIPA. **Divulgação Operacional (DIVOP) n°003**. 2011. Disponível em: http://sistema.cenipa.aer.mil.br/cenipa/paginas/relatorios/prevencao/DIVOP_003_2011.pdf

COLLINS, M. et al. Long-term Climate Change: Projections, Commitments and Irreversibility. **Climate Change 2013 - The Physical Science Basis: Contribution**

of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, p. 1029–1136, 2013.

DECEA. **Notícias DECEA**. Disponível em: <<https://www.decea.mil.br/?i=midia-e-informacao&p=noticia>>. Acesso em: 10 mar. 2020.

DECEA. **SIRIUS BRASIL. A aviação do futuro já começou**. Disponível em: <<https://www.decea.mil.br/sirius/index.php/2011/06/14/nivel-de-cruzeiro-2/>>. Acesso em: 10 fev. 2021.

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78–87, 2012.

DOYLE, J. D. et al. Large-Amplitude Mountain Wave Breaking over Greenland. **Journal of the Atmospheric Sciences**, v. 62, n. 9, p. 3106–3126, 1 set. 2005.

DURRAN, D. R. Mountain Waves and Downslope Winds. **American Meteorological Society**, v. 23, n. 45, 1990.

DUTTON, J. A.; PANOFSKY, H. A. Clear air turbulence: a mystery may be unfolding. **Science (New York, N.Y.)**, v. 167, n. 3920, p. 937–944, 13 fev. 1970.

EICK, D. **Turbulence Related Accidents and Incidents**. . In: TURBULENCE IMPACT MITIGATION WORKSHOP 2. Washington, DC: NCAR, 3 set. 2014.

ELLROD, G. P.; KNAPP, D. I. An Objective Clear-Air Turbulence Forecasting Technique: Verification and Operational Use. **Weather and Forecasting**, v. 7, n. 1, p. 150–165, 1 mar. 1992.

ELLROD, G. P.; KNOX, J. A. Improvements to an Operational Clear-Air Turbulence Diagnostic Index by Addition of a Divergence Trend Term. **Weather and Forecasting**, v. 25, n. 2, p. 789–798, 1 abr. 2010.

FAA. **Turbulence: Staying Safe**. Disponível em: <https://www.faa.gov/travelers/fly_safe/turbulence/>. Acesso em: 9 nov. 2019. (2019a).

FAA. **Air Traffic Plans and Publications**. Disponível em: <https://www.faa.gov/air_traffic/publications/>. Acesso em: 9 nov. 2019. (2019b).

FRANÇA, G. B.; ALMEIDA, M. V. DE; ROSETTE, A. C. 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, p. 2335–2344, 26 maio 2016.

GULTEPE, I. et al. A Review of High Impact Weather for Aviation Meteorology. **Pure and Applied Geophysics**, v. 176, n. 5, p. 1869–1921, 1 maio 2019.

HALL, M. et al. The WEKA data mining software: an update. **ACM SIGKDD Explorations Newsletter**, v. 11, n. 1, p. 10–18, 2009.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3ª edição ed. Haryana, India; Burlington, MA: Morgan Kaufmann Publishers, 2011.

HOLTON, J. R. **An Introduction to Dynamic Meteorology**. 4. ed. Academic Press, 2004.

HON, K. K.; NG, C. W.; CHAN, P. W. Machine learning based multi-index prediction of aviation turbulence over the Asia-Pacific. **Machine Learning with Applications**, v. 2, p. 100008, 15 dez. 2020.

JAEGER, E. B.; SPRENGER, M. A Northern Hemispheric climatology of indices for clear air turbulence in the tropopause region derived from ERA40 reanalysis data. **Journal of Geophysical Research: Atmospheres**, v. 112, n. D20, 2007.

KAPLAN, M. L. et al. Characterizing the severe turbulence environments associated with commercial aviation accidents. Part 2: Hydrostatic mesoscale numerical simulations of supergradient wind flow and streamwise ageostrophic frontogenesis. **Meteorology and Atmospheric Physics**, v. 88, n. 3, p. 153–173, 1 abr. 2005.

KAY, M. P. et al. **Quality Assessment Product Development Team: Graphical Turbulence Guidance, Version 2.3**. In: AVIATION WEATHER TECHNOLOGY TRANSFER (AWTT) TECHNICAL REVIEW PANEL, 2006.

KIM, J.-H. et al. Improvements in nonconvective aviation turbulence prediction for the World Area Forecast System. **Bulletin of the American Meteorological Society**, p. 2295–2311, 2018.

KIM, J.-H.; CHUN, H.-Y. A Numerical Study of Clear-Air Turbulence (CAT) Encounters over South Korea on 2 April 2007. **Journal of Applied Meteorology and Climatology**, v. 49, n. 12, p. 2381–2403, 1 dez. 2010.

LATAM. In: **Vertical Acceleration High in Flight**. p. 485–486, 2019.

LYRA, A. DE A.; CHAN, C. S.; DERECZYNSKI, C. P. Indicadores de turbulência a partir de previsões do modelo regional ETA. **Revista Brasileira de Meteorologia**, v. 22, n. 2, p. 161–182, ago. 2007.

MARLTON, G. J. **On the development, characterisation and applications of a balloon-borne atmospheric turbulence sensor**. PhD. University of Reading, 2016. Disponível em: <http://centaur.reading.ac.uk/68910/>

MELLO, I. B. F. DE. **Climatologia e Estudo de Caso da Turbulência de Céu Claro a Partir de Registro de Aeronaves: análise de dados observacionais e de modelagem**. M.Sc. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2015. Disponível em: <http://mtc-m21b.sid.inpe.br/col/sid.inpe.br/mtc-m21b/2015/03.13.04.06/doc/publicacao>

MENEGARDO-SOUZA, F. **Avaliação dos Resultados do Ábaco de Harrion no Diagnóstico de Turbulência no Cruzamento da Cordilheira dos Andes**. Specialization Course—São José dos Campos: Instituto de Controle do Espaço Aéreo (ICEA), 2018. Disponível em: https://www.anac.gov.br/assuntos/setor-regulado/profissionais-da-aviacao-civil/meteorologia-aeronautica/arquivos/MET001V_Final_RevisaoBanca_2TMENEGARDO.pdf

METED. **Teaching and Training Resources for the Geoscience Community**. Disponível em: <https://www.meted.ucar.edu/mesoprim/mtnwave/>. Acesso em: 7 abr. 2018.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 5ª edição ed. Hoboken, NJ: Wiley, 2012.

MORAIS, E. C. **Reconhecimento de padrões e redes neurais artificiais em predição de estruturas secundárias de proteínas**. D.Sc. Rio de Janeiro, RJ:

Universidade Federal do Rio de Janeiro, 2010. Disponível em: <https://www.cos.ufrj.br/uploadfile/1277729485.pdf>

MUÑOZ-ESPARZA, D.; SHARMAN, R. D.; DEIERLING, W. Aviation Turbulence Forecasting at Upper Levels with Machine Learning Techniques Based on Regression Trees. **Journal of Applied Meteorology and Climatology**, v. 59, n. 11, p. 1883–1899, 13 nov. 2020.

NASTROM, G. D.; FRITTS, D. C. Sources of Mesoscale Variability of Gravity Waves. Part I: Topographic Excitation. **Journal of the Atmospheric Sciences**, v. 49, n. 2, p. 101–110, 1 jan. 1992.

NORTE, F. A. Understanding and Forecasting Zonda Wind (Andean Foehn) in Argentina: A Review. **Atmospheric and Climate Sciences**, v. 5, n. 3, p. 163–193, 15 maio 2015.

OACI. **Global Air Navigation Plan. DOC 9750**, 2016.

PARMEZAN, R. S.; LEE, H. D.; CHUNG, W. F. **Avaliação de métodos para seleção de atributos importantes para aprendizado de máquina supervisionado no processo de mineração de dados**. Foz do Iguaçu, PR: Universidade Estadual do Oeste do Parana, 2012.

PRICE, C.; RIND, D. Possible implications of global climate change on global lightning distributions and frequencies. **Journal of Geophysical Research: Atmospheres**, v. 99, n. D5, p. 10823–10831, 1994.

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

REINECKE, P. A.; DURRAN, D. R. Estimating Topographic Blocking Using a Froude Number When the Static Stability Is Nonuniform. **Journal of Atmospheric Sciences**, v. 65, p. 1035, 2008.

ROSSI, A. L. D. **Ajuste de parâmetros de técnicas de classificação por algoritmos bioinspirados**. M.Sc. São Carlos: Universidade de São Paulo, 2009. Disponível em: <https://teses.usp.br/teses/disponiveis/55/55134/tde-06052009-114528/pt-br.php>

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 3ª edição ed. Upper Saddle River: Prentice Hall, 2009.

SHARMAN, R. et al. An Integrated Approach to Mid- and Upper-Level Turbulence Forecasting. **Weather and forecasting**, v. 21, p. 268–287, 2006.

SHARMAN, R. et al. Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: A review. **Geophysical Research Letters**, v. 39, n. 12, 2012.

SHARMAN, R. et al. Description and Derived Climatologies of Automated In Situ Eddy-Dissipation-Rate Reports of Atmospheric Turbulence. **Journal of Applied Meteorology and Climatology**, v. 53, p. 1416–1432, 2014.

SHARMAN, R.; LANE, T. (EDS.). **Aviation Turbulence: Processes, Detection, Prediction**. Springer International Publishing, 2016.

SHARMAN, R.; PEARSON, J. M. Prediction of Energy Dissipation Rates for Aviation Turbulence. Part I: Forecasting Nonconvective Turbulence | Journal of Applied Meteorology and Climatology | American Meteorological Society. **Journal of Applied Meteorology and Climatology**, v. 56, p. 317–377, 2017.

SHARMAN, R.; TRIER, S. B. Influences of Gravity Waves on Convectively Induced Turbulence (CIT): A Review. **Pure and Applied Geophysics**, v. 176, n. 5, p. 1923–1958, 1 maio 2019.

SHI, W. et al. The Annual Cycle and Variability of Precipitation in Brazil. **NCEP/Climate Prediction Center Atlas**, v. 9, 2000.

SILVA, A. R. **Ondas Orográficas Sobre a Cordilheira dos Andes em Latitudes Subtropicais: estudo de caso**. M.Sc. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2004. Disponível em: <http://mtc-m16.sid.inpe.br/col/sid.inpe.br/jeferson/2004/05.27.09.45/doc/publicacao.pdf>

SMITH, R. B. Mountain-induced stagnation points in hydrostatic flow. **Tellus A**, v. 41A, n. 3, p. 270–274, 1989.

SMITH, R. B. 100 Years of Progress on Mountain Meteorology Research. **Meteorological Monographs**, v. 59, p. 20.1-20.73, 1 jan. 2019.

SOARES, J. DE A. **Pré-processamento em mineração de dados: um estudo comparativo em complementação**. D.Sc. Rio de Janeiro, RJ: Universidade Federal do Rio de Janeiro, 2007. Disponível em: <https://www.cos.ufrj.br/uploadfile/publicacao/1810.pdf>

STORER, L. N.; WILLIAMS, P. D.; GILL, P. G. Aviation Turbulence: Dynamics, Forecasting, and Response to Climate Change. **Pure and Applied Geophysics**, v. 176, n. 5, p. 2081–2095, 1 maio 2019.

STORER, L. N.; WILLIAMS, P. D.; JOSHI, M. M. Global Response of Clear-Air Turbulence to Climate Change. **Geophysical Research Letters**, v. 44, n. 19, p. 9976–9984, 2017.

STULL, R. B. **An Introduction to Boundary Layer Meteorology**. Springer Netherlands, 1988.

TAN, P.-N. et al. **Introduction to Data Mining**. 2ª edição ed. EUA: Pearson, 2018.

TURNER, J. Development of a mountain wave turbulence prediction scheme for civil aviation. **Met Office Forecasting Research Technical Report**, v. 265, 1999.

VÁSQUEZ, H.; FALCÓN, N. ONDAS DE MONTAÑA EN LA CAPA LÍMITE PLANETARIA DE LA REGIÓN ANDINA VENEZOLANA. **Revista Brasileira de Meteorologia**, v. 30, n. 3, p. 241–253, set. 2015.

WANDISHIN, M. S. et al. **Quality Assessment Report: Graphical Turbulence Guidance, Version 2.5**. 2011.

WILLIAMS, K. Using random forests to diagnose aviation turbulence. **Machine Learning**, p. 51–70, 2014.

WILLIAMS, P. D.; JOSHI, M. M. Intensification of winter transatlantic aviation turbulence in response to climate change. **Nature Climate Change**, v. 3, p. 644–648, 8 abr. 2013.

WMO. **The airflow over mountains**. WMO N°355, 1973. Technical Note N°127.

WOLFF, J. K.; SHARMAN, R. Climatology of Upper-Level Turbulence over the Contiguous United States. **Journal of Applied Meteorology and Climatology**, v. 47, n. 8, p. 2198–2214, 1 ago. 2008.

APÊNDICE A - LISTA DE ATRIBUTOS

Tabela 11: Parâmetros atmosféricos extraídos do retângulo de interesse.

Atributo	Informação	Atributo	Informação
1	u	071°W_700hPa	26 $\Delta w/\Delta x$ 071°W_400hPa
2	θ	071°W_700hPa	27 vws 071°W_400hPa
3	$\Delta\theta/\Delta x$	071°W_700hPa	28 R_{iB} 071°W_400hPa
4	w	071°W_700hPa	29 u 071°W_300hPa
5	$\Delta w/\Delta x$	071°W_700hPa	30 θ 071°W_300hPa
6	vws	071°W_700hPa	31 $\Delta\theta/\Delta x$ 071°W_300hPa
7	R_{iB}	071°W_700hPa	32 w 071°W_300hPa
8	u	071°W_600hPa	33 $\Delta w/\Delta x$ 071°W_300hPa
9	θ	071°W_600hPa	34 vws 071°W_300hPa
10	$\Delta\theta/\Delta x$	071°W_600hPa	35 R_{iB} 071°W_300hPa
11	w	071°W_600hPa	36 u 071°W_200hPa
12	$\Delta w/\Delta x$	071°W_600hPa	37 θ 071°W_200hPa
13	vws	071°W_600hPa	38 $\Delta\theta/\Delta x$ 071°W_200hPa
14	R_{iB}	071°W_600hPa	39 w 071°W_200hPa
15	u	071°W_500hPa	40 $\Delta w/\Delta x$ 071°W_200hPa
16	θ	071°W_500hPa	41 vws 071°W_200hPa
17	$\Delta\theta/\Delta x$	071°W_500hPa	42 R_{iB} 071°W_200hPa
18	w	071°W_500hPa	43 u 070°W_700hPa
19	$\Delta w/\Delta x$	071°W_500hPa	44 θ 070°W_700hPa
20	vws	071°W_500hPa	45 $\Delta\theta/\Delta x$ 070°W_700hPa
21	R_{iB}	071°W_500hPa	46 w 070°W_700hPa
22	u	071°W_400hPa	47 $\Delta w/\Delta x$ 070°W_700hPa
23	θ	071°W_400hPa	48 vws 070°W_700hPa
24	$\Delta\theta/\Delta x$	071°W_400hPa	49 R_{iB} 070°W_700hPa
25	w	071°W_400hPa	50 u 070°W_600hPa

Continuação

Atributo	Informação	Atributo	Informação		
51	θ	070°W_600hPa	76	vws	070°W_300hPa
52	$\Delta\theta/\Delta x$	070°W_600hPa	77	Ri _B	070°W_300hPa
53	w	070°W_600hPa	78	u	070°W_200hPa
54	$\Delta w/\Delta x$	070°W_600hPa	79	θ	070°W_200hPa
55	vws	070°W_600hPa	80	$\Delta\theta/\Delta x$	070°W_200hPa
56	Ri _B	070°W_600hPa	81	w	070°W_200hPa
57	u	070°W_500hPa	82	$\Delta w/\Delta x$	070°W_200hPa
58	θ	070°W_500hPa	83	vws	070°W_200hPa
59	$\Delta\theta/\Delta x$	070°W_500hPa	84	Ri _B	070°W_200hPa
60	w	070°W_500hPa	85	u	069°W_700hPa
61	$\Delta w/\Delta x$	070°W_500hPa	86	θ	069°W_700hPa
62	vws	070°W_500hPa	87	$\Delta\theta/\Delta x$	069°W_700hPa
63	Ri _B	070°W_500hPa	88	w	069°W_700hPa
64	u	070°W_400hPa	89	$\Delta w/\Delta x$	069°W_700hPa
65	θ	070°W_400hPa	90	vws	069°W_700hPa
66	$\Delta\theta/\Delta x$	070°W_400hPa	91	Ri _B	069°W_700hPa
67	w	070°W_400hPa	92	u	069°W_600hPa
68	$\Delta w/\Delta x$	070°W_400hPa	93	θ	069°W_600hPa
69	vws	070°W_400hPa	94	$\Delta\theta/\Delta x$	069°W_600hPa
70	Ri _B	070°W_400hPa	95	w	069°W_600hPa
71	u	070°W_300hPa	96	$\Delta w/\Delta x$	069°W_600hPa
72	θ	070°W_300hPa	97	vws	069°W_600hPa
73	$\Delta\theta/\Delta x$	070°W_300hPa	98	Ri _B	069°W_600hPa
74	w	070°W_300hPa	99	u	069°W_500hPa
75	$\Delta w/\Delta x$	070°W_300hPa	100	θ	069°W_500hPa

Continuação

Atributo		Informação		Atributo		Informação	
101	$\Delta\theta/\Delta x$	069°W_500hPa		126	Ri _B	069°W_200hPa	
102	w	069°W_500hPa		127	u	068°W_700hPa	
103	$\Delta w/\Delta x$	069°W_500hPa		128	θ	068°W_700hPa	
104	vws	069°W_500hPa		129	$\Delta\theta/\Delta x$	068°W_700hPa	
105	Ri _B	069°W_500hPa		130	w	068°W_700hPa	
106	u	069°W_400hPa		131	$\Delta w/\Delta x$	068°W_700hPa	
107	θ	069°W_400hPa		132	vws	068°W_700hPa	
108	$\Delta\theta/\Delta x$	069°W_400hPa		133	Ri _B	068°W_700hPa	
109	w	069°W_400hPa		134	u	068°W_600hPa	
110	$\Delta w/\Delta x$	069°W_400hPa		135	θ	068°W_600hPa	
111	vws	069°W_400hPa		136	$\Delta\theta/\Delta x$	068°W_600hPa	
112	Ri _B	069°W_400hPa		137	w	068°W_600hPa	
113	u	069°W_300hPa		138	$\Delta w/\Delta x$	068°W_600hPa	
114	θ	069°W_300hPa		139	vws	068°W_600hPa	
115	$\Delta\theta/\Delta x$	069°W_300hPa		140	Ri _B	068°W_600hPa	
116	w	069°W_300hPa		141	u	068°W_500hPa	
117	$\Delta w/\Delta x$	069°W_300hPa		142	θ	068°W_500hPa	
118	vws	069°W_300hPa		143	$\Delta\theta/\Delta x$	068°W_500hPa	
119	Ri _B	069°W_300hPa		144	w	068°W_500hPa	
120	u	069°W_200hPa		145	$\Delta w/\Delta x$	068°W_500hPa	
121	θ	069°W_200hPa		146	vws	068°W_500hPa	
122	$\Delta\theta/\Delta x$	069°W_200hPa		147	Ri _B	068°W_500hPa	
123	w	069°W_200hPa		148	u	068°W_400hPa	
124	$\Delta w/\Delta x$	069°W_200hPa		149	θ	068°W_400hPa	
125	vws	069°W_200hPa		150	$\Delta\theta/\Delta x$	068°W_400hPa	

Conclusão

Atributo	Informação	Atributo	Informação	
151	w	068°W_400hPa	176	DW
152	$\Delta w/\Delta x$	068°W_400hPa	177	ΣRi_B
153	vws	068°W_400hPa	178	Ri_B máx
154	Ri_B	068°W_400hPa	179	Ri_B mín
155	u	068°W_300hPa	180	Ri_B médio
156	θ	068°W_300hPa	181	Σvws
157	$\Delta\theta/\Delta x$	068°W_300hPa	182	vws máx
158	w	068°W_300hPa	183	vws mín
159	$\Delta w/\Delta x$	068°W_300hPa	184	vws médio
160	vws	068°W_300hPa		
161	Ri_B	068°W_300hPa		
162	u	068°W_200hPa		
163	θ	068°W_200hPa		
164	$\Delta\theta/\Delta x$	068°W_200hPa		
165	w	068°W_200hPa		
166	$\Delta w/\Delta x$	068°W_200hPa		
167	vws	068°W_200hPa		
168	Ri_B	068°W_200hPa		
169	$u_{m\acute{a}x}$ (1000-500hPa)			
170	N			
171	Fr			
172	wmáx			
173	wmín			
174	wmédio			
175	desvpad_w			

APÊNDICE B – ATRIBUTOS SELECIONADOS

Tabela 12: Lista de atributos selecionados para cada um dos conjuntos de dados de treinamento e teste dos algoritmos de AM.

Conjunto de dados	Atributos Selecionados
Dados originais	2, 5, 10, 12, 17, 18, 24, 32, 33, 37, 47, 50, 53, 63, 70, 73, 74, 79, 81, 87, 105, 108, 112, 125, 136, 138, 147, 150, 157 e 174
Retirada Aleatória de “NÃO”	3, 5, 7, 9, 10, 11, 12, 16, 24, 28, 30, 33, 38, 45, 47, 50, 53, 69, 71, 72, 73, 87, 90, 95, 98, 103, 105, 112, 121, 122, 125, 126, 136, 139, 140, 142, 150, 157, 164, 170, 172, 174, 175 e 179
Inserção Aleatória de “SIM” m1desvpad	3, 4, 5, 6, 7, 10, 12, 14, 17, 21, 24, 25, 26, 28, 31, 35, 38, 42, 47, 49, 51, 52, 56, 59, 62, 63, 66, 70, 73, 77, 79, 84, 85, 87, 88, 91, 94, 98, 101, 105, 108, 112, 114, 115, 119, 122, 125, 126, 129, 130, 131, 133, 136, 137, 138, 140, 143, 145, 147, 150, 152, 154, 156, 157, 159, 161, 164, 166, 168, 170, 180, 182, 183 e 184
Inserção Aleatória de “SIM” m2desvpad	1, 4, 5, 7, 9, 11, 12, 14, 17, 18, 19, 21, 24, 26, 28, 31, 32, 33, 35, 37, 38, 39, 40, 42, 47, 49, 53, 54, 56, 60, 61, 63, 67, 69, 70, 74, 75, 77, 82, 84, 85, 88, 89, 90, 91, 95, 96, 98, 102, 105, 109, 110, 111, 112, 116, 118, 119, 123, 124, 125, 126, 127, 130, 131, 132, 133, 136, 137, 138, 140, 143, 144, 145, 147, 150, 151, 152, 154, 157, 158, 159, 161, 164, 166, 168, 174, 176, 179, 180 e 183

Conclusão

Conjunto de dados	Atributos Selecionados
Inserção Aleatória de "SIM" mediana	3, 7, 10, 17, 24, 31, 38, 56, 73, 136, 143, 150, 157 e 164