

Sistema FIEB



CENTRO UNIVERSITÁRIO SENAI CIMATEC
ESPECIALIZAÇÃO EM
DATA SCIENCE & ANALYTICS

GABRIEL CURVELO LEITE

**Orientador: Prof. Dr. Erick Giovani Sperandio
Nascimento**

**PREVISÃO DE DESLIZAMENTOS DE TERRA
INDUZIDOS POR PRECIPITAÇÃO EM SALVADOR/BA
UTILIZANDO REDES NEURAS ARTIFICIAIS**

Salvador (BA)
2020

Sistema FIEB



GABRIEL CURVELO LEITE

**Orientador: Prof. Dr. Erick Giovanni Sperandio
Nascimento**

**PREVISÃO DE DESLIZAMENTOS DE TERRA
INDUZIDOS POR PRECIPITAÇÃO EM SALVADOR/BA
UTILIZANDO REDES NEURAS ARTIFICIAIS**

Projeto apresentado ao CENTRO UNIVERSITÁRIO SENAI CIMATEC como requisito parcial para obtenção do título de Especialista em Data Science & Analytics.

Salvador (BA)
2020

CENTRO UNIVERSITÁRIO SENAI CIMATEC
ESPECIALIZAÇÃO EM DATA SCIENCE & ANALYTICS

ATA DE APRESENTAÇÃO DE PROJETO FINAL DE CURSO

Ata de apresentação do Projeto Final de Curso, “PREVISÃO DE DESLIZAMENTOS DE TERRA INDUZIDOS POR PRECIPITAÇÃO EM SALVADOR/BA UTILIZANDO REDES NEURAS ARTIFICIAIS”, submetido pelo aluno **Gabriel Curvelo Leite**, como parte dos requisitos para obtenção do Certificado de **Especialista em Data Science & Analytics** pelo Centro Universitário SENAI CIMATEC, às 16horas do dia 26 de novembro de 2020. Reuniu-se remotamente pela plataforma Teams, a Banca Examinadora designada pelo Prof Dr Erick Giovani Sperandio Nascimento – Orientador, constituída pelo Prof Dr Davidson Martins Moreira e pelo Prof Esp Flavio Santos Conterato. Ana Luiza Guimarães, Coordenadora da Especialização, deu início aos trabalhos com as devidas orientações, e a exposição foi realizada pelo estudante dentro do prazo de tempo estabelecido. Ao final da apresentação a banca reuniu-se atribuindo a seguinte nota: 9.0 (Nove).

A banca de avaliadores decidiu pela:

(X) Aprovação do trabalho

Caberá ao aluno apresentar em no máximo em 30 (trinta) dias a contar da data de assinatura desta Ata, uma cópia do trabalho em PDF com restrição de edição, constando as considerações pontuadas pela banca. A Ata de Apresentação do Projeto Final de Curso deve ser digitalizada e inserida na terceira página do PFC ou como anexo do artigo.

() Reprovação do trabalho

O aluno terá que se matricular novamente no TCC – Trabalho de Conclusão de Curso e ser submetido a uma banca avaliadora no semestre seguinte.

As ações consequentes ao status de Aprovação deverão obedecer ao prazo proposto acima sob pena do parecer final ser modificado para o status de Reprovado automaticamente e sem possibilidade de recurso.

Para constar, lavrou-se a presente ata que vai assinada por todos os membros da Banca. Por estarem cientes de suas obrigações estão de acordo com os termos desse documento:

Salvador, 26 de novembro de 2020.

Assinado digitalmente por Erick Giovani Sperandio Nascimento
DN: C=BR, S=Bahia, L=Salvador, O=SENAI/DR/BA, CN=Erick Giovani Sperandio Nascimento, E=erick.sperandio@fiob.org.br
Razão: Eu sou o autor deste documento
Localização: sua localização de assinatura aqui
Data: 2020.12.11 13:41:48 -03'00'
Foxit Reader Versão: 10.1.0

Prof Dr Erick Giovani Sperandio Nascimento
Orientador

Prof Prof Dr Davidson Martins Moreira
Membro da banca

Flavio Santos Conterato

Assinado digitalmente por Flavio Santos Conterato
DN: C=BR, S=Bahia, L=Salvador, O=SENAI/DR/BA, CN=Flavio Santos Conterato,
E=flavio.conterato@fiob.org.br
Razão: Eu estou aprovando este documento com minha assinatura de vinculação legal
Localização: SENAI CIMATEC
Data: 2020.12.10 01:24:55 -03'00'
Foxit Reader Versão: 10.1.0

Prof Esp Flavio Santos Conterato
Membro da banca

Assinado digitalmente por: Ana Luiza Medeiros Guimaraes Magalhaes
O tempo: 16-12-2020 14:21:07

Ana Luiza Medeiros Guimarães
Coordenadora Especialista

SUMÁRIO

1 INTRODUÇÃO	1
1.1 CONTEXTUALIZAÇÃO	1
1.1.1 O MUNICÍPIO DE SALVADOR	1
1.1.1.1 População	2
1.1.1.2 Dados socioeconômicos	3
1.1.1.3 Relevo	4
1.1.1.4 Áreas de risco	5
1.1.2 CLIMATOLOGIA DE SALVADOR	9
1.1.3 DESASTRES NATURAIS EM SALVADOR	10
1.1.3.1 Conceito	10
1.1.3.2 Movimentos de massa	10
1.1.3.3 Deslizamentos de terra em Salvador	11
1.2 BASES DE DADOS	13
1.2.1 HISTÓRICO DE OCORRÊNCIAS	13
1.2.2 DADOS METEOROLÓGICOS	14
1.2.3 MODELO NUMÉRICO DE PREVISÃO DO TEMPO (WRF)	16
1.2.3.1 Dados de entrada	16
1.2.3.2 Domínios	17
1.2.3.3 Extração de pontos de grade	17
1.2.3.4 Variáveis meteorológicas previstas	18

2 MOTIVAÇÃO.....	19
2.1 PROTOCOLOS EXISTENTES	19
2.2 EXISTÊNCIA DE DADOS.....	20
2.3 OUTROS ESTUDOS APLICADOS.....	20
3 OBJETIVOS	21
3.1 GERAL.....	21
3.2 ESPECÍFICOS.....	21
4 ESTADO DA ARTE	22
4.1 INTELIGÊNCIA ARTIFICIAL.....	22
4.1.1 DEFINIÇÃO.....	22
4.1.2 EVOLUÇÃO	23
4.1.3 SUBÁREAS.....	24
4.1.4 REDES NEURAIS ARTIFICIAIS	25
4.1.4.1 Conceituação.....	25
4.1.4.2 Elementos.....	25
4.1.4.3 Tipos.....	27
4.2 PÓS-PROCESSAMENTO DE MODELOS NUMÉRICOS DE PREVISÃO DO TEMPO.....	29
4.3 PREVISÃO DE DESLIZAMENTOS DE TERRA	31
5 METODOLOGIA.....	37

5.1 MODELO DE PREVISÃO DE PRECIPITAÇÃO (PÓS-PROCESSAMENTO DO WRF)	39
5.1.1 ANÁLISE EXPLORATÓRIA DE DADOS.....	40
5.1.2 ENGENHARIA DE DADOS.....	41
5.1.3 CONSTRUÇÃO DO MODELO DE INTELIGÊNCIA ARTIFICIAL.....	44
5.2 MODELO DE PREVISÃO DE DESLIZAMENTOS DE TERRA.....	46
5.2.1 ANÁLISE EXPLORATÓRIA DE DADOS.....	48
5.2.2 ENGENHARIA DE DADOS.....	49
5.2.3 CONSTRUÇÃO DO MODELO DE INTELIGÊNCIA ARTIFICIAL.....	52
6 RESULTADOS E DISCUSSÕES	55
6.1 ANÁLISE DE DADOS.....	55
6.1.1 DESLIZAMENTOS DE TERRA.....	55
6.1.2 PRECIPITAÇÃO	60
6.1.3 ASSOCIAÇÃO DOS DADOS	62
6.2 MÉTRICAS DOS RESULTADOS DE PRECIPITAÇÃO DO MODELO WRF	66
6.2.1 ANÁLISE DA RESOLUÇÃO ESPACIAL E DOS DOMÍNIOS (GRADES).....	67
6.2.2 ANÁLISE DOS AGRUPAMENTOS HORÁRIOS.....	70
6.2.3 ANÁLISE DA HORA DO DIA	74
6.2.4 ANÁLISE DO MÊS.....	75

6.2.5 ANÁLISE DA ESTAÇÃO DO ANO	76
6.3 MODELO DE PREVISÃO DE PRECIPITAÇÃO (PÓS-PROCESSAMENTO DO WRF)	77
6.4 MODELO DE PREVISÃO DE DESLIZAMENTOS DE TERRA.....	81
7 CONCLUSÃO.....	89
7.1 CONCLUSÕES.....	89
7.2 LIMITAÇÕES.....	90
7.3 TRABALHOS FUTUROS.....	91
REFERÊNCIAS BIBLIOGRÁFICAS	92

RESUMO

O município de Salvador-BA (Brasil) é historicamente marcado por deslizamentos de terra, que geram grandes impactos sociais e frequentemente causam mortes. A predisposição a estes deslizamentos de terra possui como condicionantes principais o relevo fortemente acidentado, características socioeconômicas da população e a incidência de eventos chuvosos.

Considerando o desafio de avaliação e previsão do risco de ocorrência de deslizamentos de terra de forma operacional, o presente trabalho tem como objetivo a criação e validação de um modelo que seja capaz de prever a quantidade de deslizamentos de terra no município de Salvador, com horizonte de previsão de 72 horas, de forma espacializada. Para isso, é criado, primeiramente, um modelo de previsão de precipitação, a partir dos resultados de um Modelo Numérico de Previsão do Tempo (WRF – Weather Research and Forecasting). Este modelo de previsão de precipitação tem seus resultados utilizados como entrada para o modelo de previsão de deslizamentos de terra. Os dois modelos são construídos utilizando-se Redes Neurais Artificiais.

O modelo de previsão de precipitação foi capaz de produzir resultados muito superiores aos dados originais do WRF. De acordo com a metodologia utilizada, as métricas dos resultados originais do WRF para a precipitação, com horizonte de 96 horas, foram MAE (Erro Absoluto Médio) = 1,04, R (Coeficiente de Correlação de Pearson) = 0,16, R^2 (Coeficiente de Determinação) = -0,24 e FAC2 (Fator de 2) = 0,34, enquanto o modelo desenvolvido obteve MAE = 0,48, R = 0,86, R^2 = 0,73 e FAC2 = 0,76.

Já o modelo de previsão de deslizamentos de terra apresentou como métricas médias de seus resultados: MAE = 0,01, R = 0,90, R^2 = 0,81 e FAC2 = 0,99.

A utilização de ambos os modelos trará ganhos expressivos à capacidade de atuação da Defesa Civil municipal para prevenção dos impactos de deslizamentos de terra.

1 INTRODUÇÃO

Considerando a problemática dos deslizamentos de terra em Salvador, este trabalho tem como proposta a modelagem de dados para a produção de um modelo de previsão de deslizamentos de terra, no âmbito da Defesa Civil de Salvador. Objetiva-se gerar um modelo capaz de prever o quantitativo de ocorrências com um horizonte de previsão de 72h, para diversas regiões da cidade.

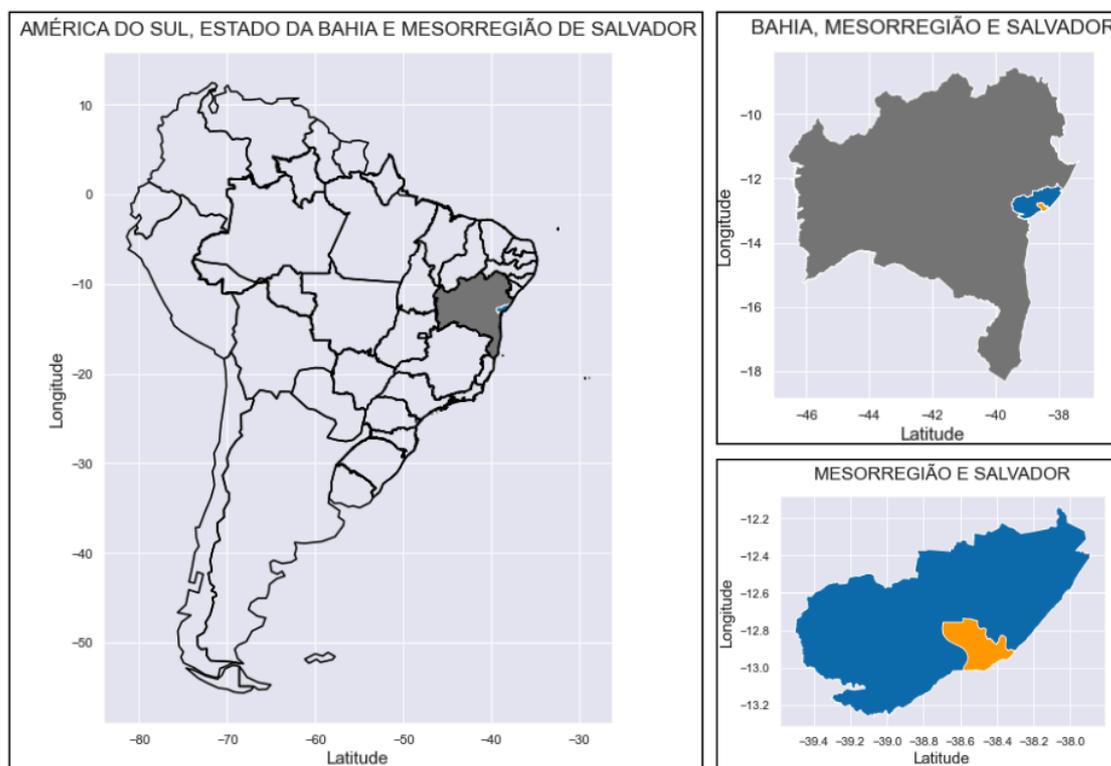
Nesta seção, são feitas considerações introdutórias que contextualizam a proposta do trabalho, apresentando informações relevantes sobre o município de Salvador e as bases de dados utilizadas. Em seguida, são apresentados a motivação para a solução proposta e seus objetivos.

1.1 CONTEXTUALIZAÇÃO

1.1.1 O MUNICÍPIO DE SALVADOR

O município de Salvador é a capital do Estado da Bahia (Brasil). Está localizado na região Nordeste do Brasil, fazendo parte da Mesorregião Metropolitana de Salvador, de acordo com definições do IBGE (Instituto Brasileiro de Geografia e Estatística). A Figura 1 mostra mapas com a localização do Estado da Bahia, da Mesorregião Metropolitana de Salvador e do Município de Salvador.

Figura 1 - Localização de Salvador (em laranja).



Fonte: Autor. Elaborado com dados do IBGE.

1.1.1.1 População

De acordo com dados do Censo Demográfico de 2010, realizado pelo IBGE (Instituto Brasileiro de Geografia e Estatística), o município de Salvador possuía, em 2010, a 3ª maior população do Brasil, com 2.675.656 pessoas, entre todos os municípios brasileiros. O município possui, ainda, uma área de 693,453 km², resultando em uma densidade demográfica de 3.859,44 hab./km², a 24ª maior do país (IBGE, 2010). Em 2019 foi estimada uma população de 2.872.347 pessoas. (IBGE, [s.d.]).

Estes dados evidenciam o tamanho do município, altamente povoado. A análise de alguns dados socioeconômicos permite traçar um perfil desta população, que, em grande parte, é carente de trabalho, renda, educação e moradia adequada.

1.1.1.2 Dados socioeconômicos

A Tabela 1, a seguir, mostra alguns indicadores sociais e econômicos referentes ao município de Salvador, divulgados pelo IBGE.

Tabela 1 - Dados socioeconômicos de Salvador.

Eixo	Indicador	Valor	Posição no ranking municípios brasileiros
Economia	PIB	R\$ 63 bilhões (aprox.)	9
	PIB per capita	R\$ 21.231,48	2106
Trabalho e Rendimento	Salário médio mensal dos trabalhadores formais	3,5 salários mínimos	70
	População ocupada	29,7%	505
	Percentual da população com rendimento nominal mensal per capita de até 1/2 salário mínimo	36,8%	3211
Educação	Taxa de escolarização de 6 a 14 anos de idade	95,9%	4637
Saúde	Mortalidade Infantil	14,56 óbitos por mil nascidos vivos	2011
	Internações por diarreia	0,2 internações por mil habitantes	4284

Fonte: Autor, elaborado com dados do IBGE.

Em 2017, o município apresentou Produto Interno Bruto (PIB) de quase R\$ 63 bilhões, o maior do estado da Bahia e o 9º maior do Brasil. Porém, o cálculo do PIB per capita evidencia a problemática social do município. Com um PIB per capita de R\$ 21.231,48, sua colocação entre todos os municípios brasileiros salta para a 2106º colocação, entre os 5570 municípios. (IBGE, [s.d.]).

Considerando-se os indicadores de trabalho e rendimento disponibilizados pelo IBGE, o salário médio mensal dos trabalhadores formais, de 3,5 salários mínimos, coloca o município na 70ª posição no ranking de municípios nacionais, enquanto a proporção da população ocupada, na 505ª posição. O Percentual da população com rendimento nominal mensal per capita de até 1/2 salário mínimo, de 36,8% chama a atenção para o contraste social na capital. Este indicador coloca o município de Salvador na posição 3211, entre os 5570 municípios.

Igualmente chocantes, são os resultados das taxas de escolarização de 6 a 14 anos de idade, de Mortalidade infantil e de Internações por diarreia, com as quais o município atinge as colocações 4637, 2011 e 4284, respectivamente.

Este conjunto de indicadores permite um entendimento das condições sociais no município. Apesar de ser uma das maiores metrópoles brasileiras, grande parcela de sua população ainda vive em condições altamente precárias.

1.1.1.3 Relevô

Salvador tem ainda como característica um relevo fortemente acidentado, sendo comum o contraste entre áreas altas e baixas. Dessa forma, a morfologia do município é predominantemente marcada por espigões e colinas, principalmente, além de um sistema de vales (PALMA, 2016).

Na cidade, é muito comum a existência de aclives acima de 10º de inclinação, existindo, ainda, diversas regiões densamente ocupadas que apresentam declividade superior a 45º (SANTANA, 2017).

1.1.1.4 Áreas de risco

Este contexto de grande densidade populacional, alta desigualdade e relevo acidentado deu origem a uma ocupação desordenada de áreas de encostas que, por si só, já ofereceriam riscos habitacionais (PALMA, 2016). Este fenômeno é evidenciado por diversos indicadores, entre eles a proliferação de aglomerados subnormais e de áreas consideradas de risco.

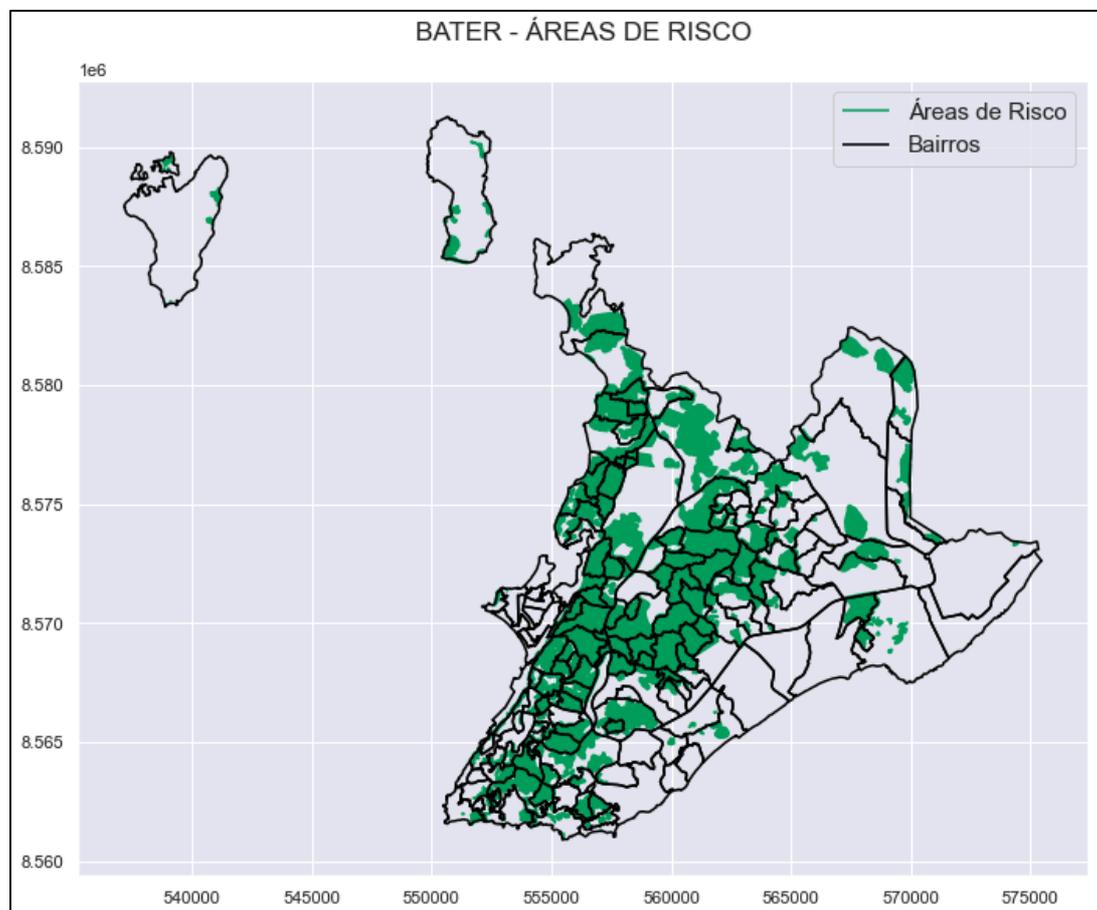
De fato, um estudo lançado em 2018, feito pelo IBGE em parceria com o CEMADEN (Centro Nacional de Monitoramento e Alertas de Desastres Naturais), permitiu o cruzamento de dados do Censo Demográfico de 2010 com áreas de risco mapeadas, através de metodologia inédita. Este estudo foi aplicado a 872 municípios monitorados pelo CEMADEN, considerados prioritários, e gerou informações intramunicipais sobre a população sujeita a riscos de inundações, enxurradas e movimentos de massa. A associação destes dados deu origem à Base Territorial Estatística de Áreas de Risco - BATER. (IBGE; CEMADEN, 2018).

Entre todos os 872 municípios objetos do estudo, Salvador se destacou como possuindo a maior população residente em áreas de risco, com 1.217.527 pessoas (45,5% de sua população total, referente ao ano 2010). A título de comparação, o segundo município com maior população em áreas de risco, São Paulo apresentou 674.329 pessoas (6,0% de sua população total), pouco mais da metade dos resultados de Salvador. (IBGE; CEMADEN, 2018).

Estes dados mostram a importância da produção de estudos e implementação de políticas públicas referentes à gestão dos riscos aos quais esta população está sujeita, bem como a prevenção, monitoramento e previsão desses desastres.

A Figura 2 mostra um mapa com as áreas de risco pertencentes à Base Territorial Estatística de Áreas de Risco (BATER). Já as Figuras 3, 4, 5 e 6 exemplificam situações de risco nas áreas de Salvador.

Figura 2 - Áreas de Risco – BATER (Base Territorial Estatística de Áreas de Risco).



Fonte: Autor. Elaborado com dados disponibilizados pelo IBGE.

Figura 3 - Áreas de risco em Salvador.

a) Baixa do Fiscal - Liberdade (Foto: Elói Corrêa/GOVBA) b) San Martin (Foto: Elói Corrêa/GOVBA)



Fonte: <http://www.secom.ba.gov.br/galeria/10534/100132/Areas-de-risco-atingidas-pela-chuva-em-Salvador.html#menu-galeria> (Acesso: 06/08/2020)

Figura 4 - Áreas de risco em Salvador.



Fonte: Codesal – Acervo.

Figura 5 - Áreas de risco em Salvador.



Fonte: Codesal – Acervo.

Figura 6 - Áreas de risco em Salvador.

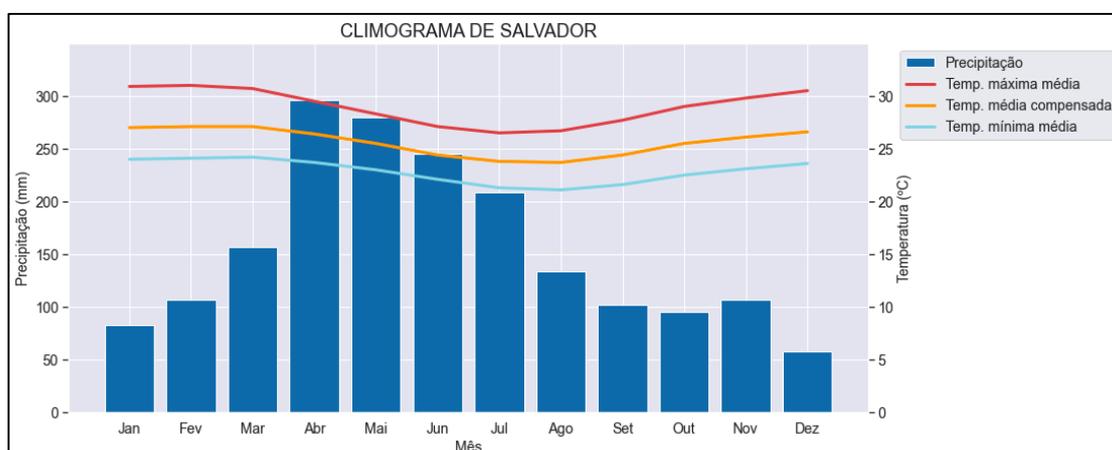


Fonte: Codesal – Acervo.

1.1.2 CLIMATOLOGIA DE SALVADOR

Em relação à climatologia, Salvador possui Normais Climatológicas de 1871,1 mm de precipitação anual, com temperatura média compensada variando entre 21,1°C e 27,1°C, referentes ao período de 1981-2010, como mostra a Figura 7. (INMET, 2018).

Figura 7 - Climograma de Salvador.



Fonte: Autor, elaborado com dados do INMET.

É possível observar que o município tem um período chuvoso bem definido, principalmente entre os meses de abril a julho. Neste período são mais frequentes eventos de precipitação intensa (SANTOS et al., 2016).

Os eventos de precipitação na cidade decorrem da atuação de alguns sistemas meteorológicos, como os ventos alísios impulsionados pela Alta Subtropical do Atlântico Sul (ASAS), o avanço de Sistemas Frontais (frentes frias) provenientes da região Sul do Brasil e Distúrbios Ondulatórios de Leste (DOL). A atuação da ASAS costuma proporcionar eventos chuvosos frequentes e que geram relativamente pequenos acumulados de chuva, principalmente entre abril e julho. Já a atuação de frentes frias e DOL costuma proporcionar eventos chuvosos mais expressivos.

Um dos sistemas meteorológicos mais preocupantes, as frentes frias atingem a região com mais frequência entre os meses de março a maio. Ou seja, é comum que a chegada de frentes frias ocorra em períodos nos quais o solo já se encontra úmido, devido à maior frequência das chuvas decorrentes dos ventos alísios. (BASTOS; FERREIRA, 2000; PALMA, 2016; SANTANA, 2017).

Este cenário de ocorrência de eventos chuvosos intensos, quando o solo já está úmido, reduz drasticamente a estabilidade de encostas, potencializando o risco de deslizamentos de terra.

1.1.3 DESASTRES NATURAIS EM SALVADOR

1.1.3.1 Conceito

O conceito de desastre natural compreende a produção de grandes danos a uma comunidade em decorrência de fenômenos naturais. No Brasil, os principais desastres naturais envolvem alagamentos e episódios de movimento de massa, sendo estes os que mais causam mortes. (CAMARINHA, 2016).

Particularmente no caso de Salvador, os desastres mais frequentes envolvem fenômenos de movimento de massa em encostas. (SANTANA, 2017).

1.1.3.2 Movimentos de massa

Movimentos de massa são desastres naturais do tipo geológico, nos quais ocorre rápido deslocamento de solo e rocha em uma encosta. Entre as diversas subdivisões existentes, no Brasil são mais característicos os deslizamentos de terra e as corridas de terra. Ambos têm sua deflagração relacionada com episódios de precipitação. (CAMARINHA, 2016; RIFFEL; GUASSELLI; BRESSANI, 2016).

As corridas de terra ou detrito geralmente ocorrem a partir de deslizamentos e envolvem volumes de material muito grandes. São muito mais

raros que os deslizamentos de terra, e têm potencial de impacto extremamente superior. (CAMARINHA, 2016; RIFFEL; GUASSELLI; BRESSANI, 2016).

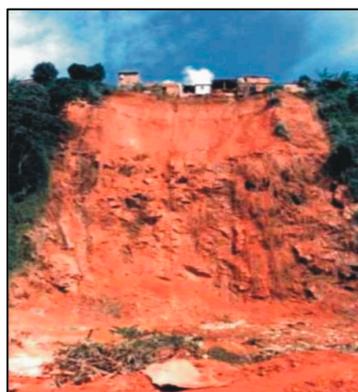
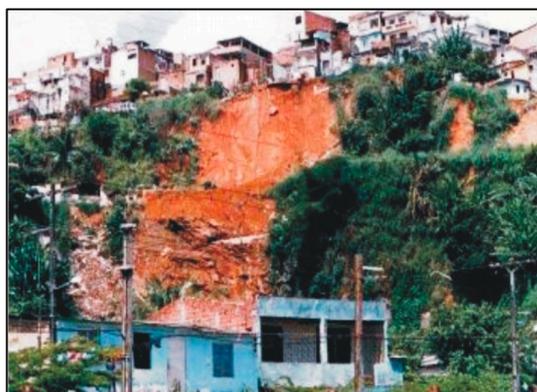
Já os deslizamentos de terra costumam ter impactos menores, devido a menores volumes de material deslocado, e são muito mais frequentes. (CAMARINHA, 2016; RIFFEL; GUASSELLI; BRESSANI, 2016).

1.1.3.3 Deslizamentos de terra em Salvador

A cidade de Salvador é historicamente marcada por deslizamentos de terra, devido à confluência de diversos fatores. As maiores condicionantes provavelmente envolvem o padrão de dispersão da população no espaço, suas características socioeconômicas (marcadas por grande desigualdade), a topografia naturalmente acidentada e a frequência e intensidade dos eventos de precipitação (principalmente no quadrimestre de abril a julho). (SANTANA, 2017; SANTOS et al., 2016).

Os episódios de movimentos de massa em encostas de Salvador com vítimas fatais não são incomuns na cidade e seus registros têm centenas de anos. Existem diversos registros de deslizamentos, muitos com vítimas, ao longo dos séculos XVIII, XIX, XX e XXI. (GEOHIDRO, 2003). A Figura 8 mostra exemplos de deslizamentos de terra registrados.

Figura 8 - Deslizamentos de terra em Salvador.



Fonte: GEOHIDRO, 2003.

Em sua maioria, os movimentos de massa em encostas de Salvador decorrem da perda estabilidade, devido, principalmente, a ações antrópicas. A ocupação desordenada de áreas inadequadas (como espigões, colinas e vales), via de regra, implica em impermeabilização de boa parte do solo, escoamento concentrado de redes de drenagem e esgotamento próximo a encostas, acúmulo de lixo nas encostas, cortes retos nas encostas, redução de vegetação e construção de habitações precárias. (COSTA; GIUDICE, 2012).

Essas alterações no ambiente natural fazem com que as encostas ocupadas percam estabilidade estrutural. Com isso, a incidência de precipitação tem maior potencial para provocar deslizamentos de terra, de forma que a precipitação geralmente é o deflagrador desses movimentos de massa, e não a causa em si.

1.2 BASES DE DADOS

Neste estudo, foram utilizadas bases de dados provenientes da Defesa Civil de Salvador (CODESAL), Instituto Nacional de Meteorologia (INMET) e Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN), sendo elas:

- Histórico de ocorrências (deslizamentos de terra). Fonte: CODESAL;
- Histórico de dados meteorológicos. Fonte: CODESAL, INMET, CEMADEN;
- Histórico de dados resultantes de um modelo numérico de previsão do tempo (WRF) operacionalizado pela CODESAL. Fonte: CODESAL;

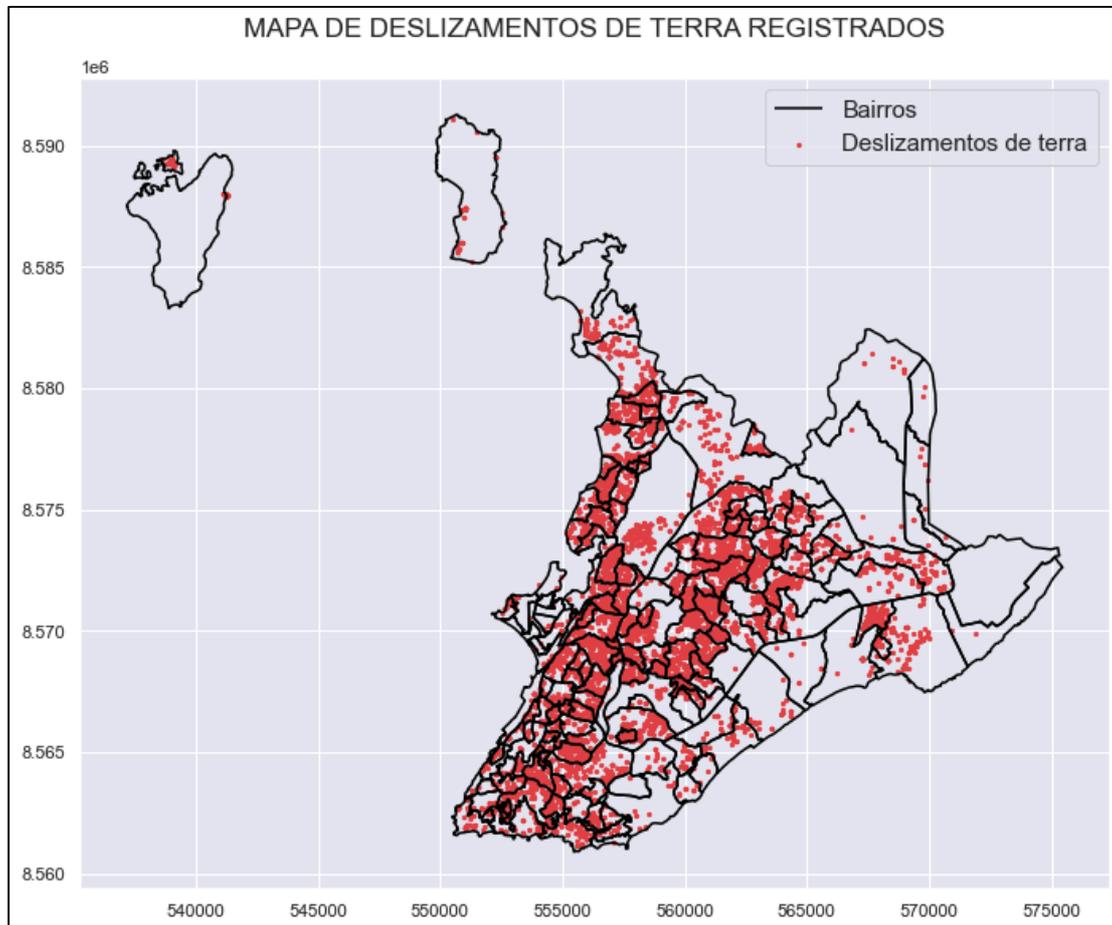
1.2.1 HISTÓRICO DE OCORRÊNCIAS

O histórico de ocorrências da CODESAL utilizado neste trabalho consiste em dados tabelados com informações referentes a vistorias realizadas pela CODESAL, desde 2004. Estes dados encontram-se armazenados em um banco de dado. Neste estudo foram selecionados apenas os registros referentes a Deslizamento de Terra.

Esses dados contêm informações sobre as solicitações feitas à Defesa Civil e sobre as vistorias realizadas pela equipe técnica do órgão. Entre as informações armazenadas, encontra-se o endereço do local da ocorrência. Como os logradouros do município estão em base de dados georreferenciada, as coordenadas geográficas podem ser facilmente obtidas.

Os registros utilizados possuem como menor granularidade os logradouros da cidade. Como aproximação da localização dessas ocorrências, são utilizadas as coordenadas geográficas (latitude e longitude) do centroide de cada logradouro. A Figura 9 mostra um mapa com a localização de todas as ocorrências utilizadas.

Figura 9 - Mapa de deslizamentos de terra registrados. (Sistema de Referência: SIRGAS 2000 / UTM zone 24S).



Fonte: Autor.

1.2.2 DADOS METEOROLÓGICOS

Os dados meteorológicos consistem em séries históricas coletadas por diversos sensores automáticos espalhados pela cidade. Alguns desses sensores são mais antigos, operantes desde o ano 2000, enquanto a maioria é mais recente. Desde o ano 2014, o município encontra-se em constante expansão da rede de monitoramento de dados meteorológicos.

Neste estudo, foram utilizados dados de 39 estações automáticas, sendo elas compostas por:

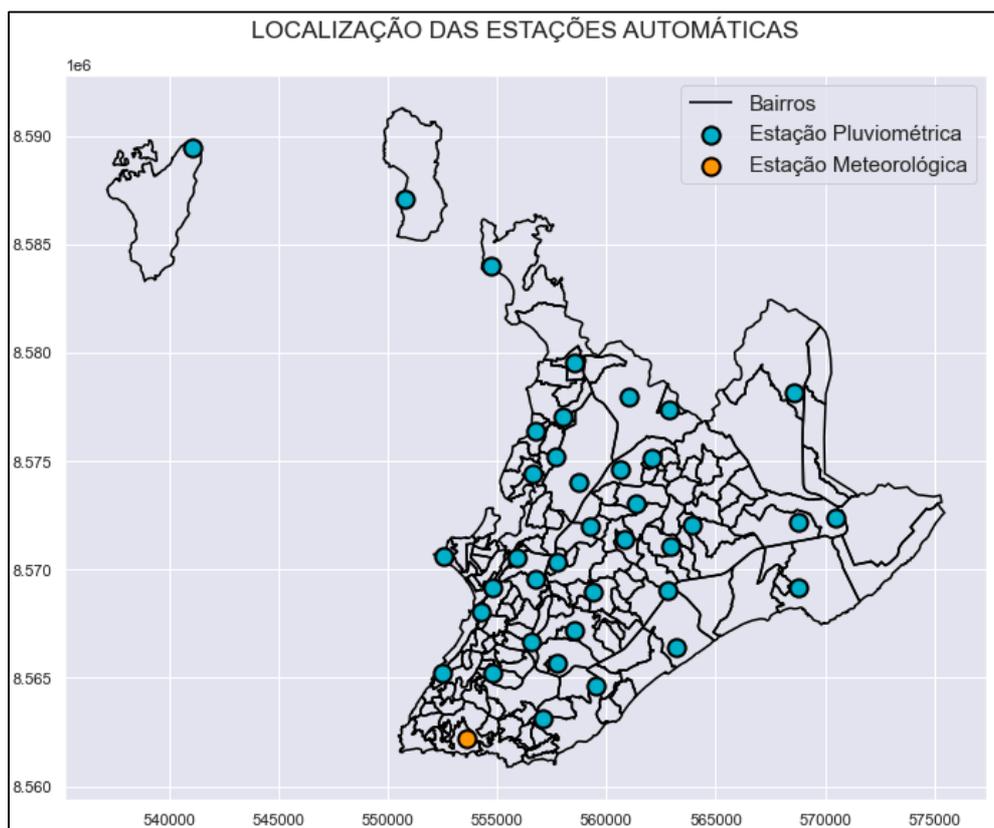
- 38 pluviômetros automáticos (dados de precipitação)

- 1 estação meteorológica automática (dados de precipitação, velocidade e direção do vento, temperatura, radiação, pressão e umidade relativa)

Os dados destes sensores são coletados automaticamente e armazenados em bancos de dados. Por serem dados de sensores automáticos que estão sujeitos a falhas, estes dados precisaram passar por tratamento e validação para serem utilizados.

A Figura 10 mostra um mapa com a distribuição das estações pluviométricas e meteorológica utilizadas.

Figura 10 - Mapa com a localização das estações utilizadas. (Sistema de Referência: SIRGAS 2000 / UTM zone 24S).



Fonte: Autor.

1.2.3 MODELO NUMÉRICO DE PREVISÃO DO TEMPO (WRF)

Visando atender à necessidade de se antever eventos chuvosos intensos, possibilitando a prevenção e redução de eventuais impactos, foi operacionalizado, na CODESAL, um modelo numérico de previsão do tempo. Utiliza-se o modelo WRF - Weather Research and Forecasting, um dos modelos numéricos de previsão do tempo mais utilizados no mundo, para as mais variadas aplicações, tanto para pesquisas quanto operacionalmente (POWERS et al., 2017).

Em cada rodada do modelo, são geradas previsões numéricas horárias e espacializadas para diversos parâmetros meteorológicos com horizonte de previsão de 96 horas à frente. Os dados resultantes dessas rodadas vêm sendo armazenados em bancos de dados desde janeiro de 2019.

Algumas características do modelo utilizado merecem atenção e serão descritas a seguir.

1.2.3.1 Dados de entrada

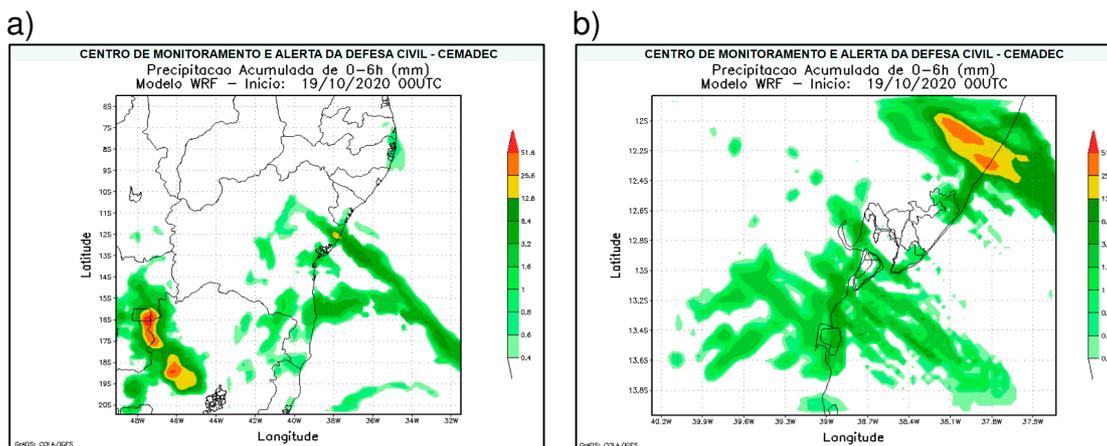
O modelo WRF utiliza, como entrada, dados do sistema numérico global de previsão do tempo GFS (Global Forecast System), fornecidos pelo National Centers for Environmental Prediction (NCEP), com resolução temporal de 6 horas. Todas as rodadas foram feitas com os dados referentes a 00UTC.

Ao longo do tempo, foram implementadas duas configurações do modelo, uma utilizando dados de entrada com resolução espacial de 1,0° de latitude e longitude e outra com resolução espacial de 0,25° de latitude e longitude. O resultado de ambas foi analisado.

1.2.3.2 Domínios

O modelo WRF utilizado na CODESAL foi configurado com 3 domínios aninhados entre si. O primeiro domínio está centralizado em -14.167° (latitude) e -40.281° (longitude), com 91×71 pontos. O segundo domínio está centralizado no Nordeste do Brasil, e conta com 97×85 pontos e espaçamento de 20 km. Já o terceiro domínio é centralizado no município de Salvador, com 81×61 pontos e espaçamento de 5 km. Neste trabalho, são utilizados apenas o segundo e o terceiro domínios. A Figura 11 mostra exemplos dos resultados para estes dois domínios.

Figura 11 - Exemplo de resultados do WRF, mostrando a precipitação prevista. a) Domínio com resolução de 20 km. b) Domínio com resolução 5 km.



Fonte: CODESAL.

1.2.3.3 Extração de pontos de grade

A partir dos resultados do modelo, são extraídos os valores de previsão meteorológica para os pontos de grade onde existem estações monitoradas pela CODESAL. Dessa forma, os dados produzidos pelo modelo WRF podem ser atribuídos a cada uma das estações e pode ser feita a correspondência entre os valores previstos e os valores observados por cada estação. Esta associação é importante para que os resultados do modelo possam ser validados, contrastando-se os dados previstos com os valores observados.

1.2.3.4 Variáveis meteorológicas previstas

Para cada ponto de grade em que é feita a extração explicada acima, são armazenados os resultados do modelo para as seguintes 8 variáveis meteorológicas:

- Precipitação na superfície
- Temperatura a 2 metros
- Umidade relativa a 2 metros
- Velocidade do vento – componente v (direção sul-norte) a 10 metros
- Velocidade do vento – componente u (direção oeste-leste) a 10 metros
- Pressão na superfície
- Temperatura do ponto de orvalho a 2 metros
- Radiação global na superfície

2 MOTIVAÇÃO

A seguir, são expostas as motivações para o presente trabalho, que envolvem a problemática dos protocolos atuais e o potencial em virtude da disponibilidade dos dados.

2.1 PROTOCOLOS EXISTENTES

Diversos estudos já foram feitos tentando estabelecer, para regiões específicas, limiares de risco de deslizamentos, principalmente em função da precipitação, obtendo indicadores de risco de movimentos de massa. Entretanto, a eficácia destes limiares (quando são estabelecidos com sucesso) é limitada. Primeiro porque os estudos são baseados em condições específicas, seja de características dos solos estudados, da dinâmica da precipitação etc. Nas condições reais, os cenários, ainda que similares entre si, possuem suas particularidades. É muito comum o solo em áreas de risco já não ser natural, mas emprestado ou fruto de aterro mal compactado, ou misturado com resíduos sólidos, ou ter sido feito um corte reto na encosta. Isto significa que, ainda que próximas, duas áreas de risco, ou mesmo duas encostas, podem apresentar comportamentos diferentes sob uma mesma perturbação.

Segundo porque, geralmente, obtêm-se limiares fixos e estáticos no tempo. Isto descarta o dinamismo intrínseco aos sistemas e fenômenos naturais. Na verdade, em um mesmo ponto, o limiar das condicionantes que deflagram os deslizamentos de terra pode não ser o mesmo para dois momentos distintos.

No município de Salvador, os protocolos, ainda que preventivos e conservadores, utilizam indicadores baseados, sobretudo, no quantitativo de chuva registrado nas últimas 72 horas. Estes indicadores são estáticos e utilizados em toda a cidade, sem distinção entre as diferentes localidades.

2.2 EXISTÊNCIA DE DADOS

Nos últimos anos, Salvador vem contando com expansão de rede coletora de dados hidrometeorológicos que podem ajudar a prever e mitigar os danos causados por desastres naturais, entre eles deslizamentos de terra. Visando aproveitar este potencial ainda não explorado, este trabalho propõe a utilização de dados existentes e disponíveis para produzir inteligência que venha a auxiliar nas capacidades de proteção e defesa civil, no que tange a deslizamentos de terra.

2.3 OUTROS ESTUDOS APLICADOS

Os estudos de previsões de deslizamentos de terra, seja através da modelagem do fenômeno dos deslizamentos em si, ou da modelagem utilizando-se técnicas da ciência de dados e aprendizado de máquina, são relativamente abundantes. Entretanto, geralmente resultam na produção de mapas de susceptibilidade a deslizamentos ou na previsão do deslocamento de terra em locais pontuais e monitorados, como mostra a revisão de literatura, na seção que analisa o estado da arte desta tecnologia.

A previsão de deslizamentos de terra em locais urbanizados, que são induzidos por intervenções humanas em conjunto com chuvas pode ser considerada inovadora e tem grande potencial, em virtude dos dados existentes.

3 OBJETIVOS

3.1 GERAL

Criar e validar um modelo preditivo que seja capaz de prever a quantidade de deslizamentos de terra esperada para as próximas 72 horas, de forma espacializada.

3.2 ESPECÍFICOS

1. Criar modelo de inteligência artificial capaz de processar os dados de resultados do WRF e gerar previsões de precipitação com menores erros.
2. Criar modelo de inteligência artificial capaz de prever, com horizonte de até 72h, o quantitativo de deslizamentos de terra, de forma espacializada, em função de:
 - Dados meteorológicos observados;
 - Dados meteorológicos previstos para as próximas 72h;
 - Histórico de ocorrências.

4 ESTADO DA ARTE

A pesquisa bibliográfica para este trabalho mostra que a Inteligência Artificial vem sendo amplamente aplicada para os mais diversos fins. Nesta seção será analisado o estado da arte da tecnologia, com enfoque para o domínio deste trabalho.

Inicialmente será feita uma introdução a respeito da Inteligência Artificial e das Redes Neurais Artificiais e, em seguida, serão analisados alguns estudos de aplicação dessas técnicas tanto na área de modelagem e previsão meteorológica, como na área de modelagem e previsão de deslizamentos de terra.

4.1 INTELIGÊNCIA ARTIFICIAL

4.1.1 DEFINIÇÃO

De acordo com RUSSELL e NORVIG (2003), a Inteligência Artificial trata do estudo de agentes inteligentes, bem como da criação dessas agentes. A área tem interseção com muitas outras disciplinas também interessadas no estudo de inteligência, tais como filosofia, psicologia e neurociência.

Para a conceituação dos agentes inteligentes, os autores destacam que existem diferentes abordagens, subdividindo-as em quatro grupos:

- i. Sistemas que *pensam como humanos*;
- ii. Sistemas que *agem como humanos*;
- iii. Sistemas que *pensam racionalmente e*
- iv. Sistemas que *agem racionalmente*.

As duas primeiras abordagens (sistemas que pensam como humanos e sistemas que agem como humanos) têm como primeiro desafio a formalização das capacidades humanas de pensar e agir. Já as duas últimas abordagens (sistemas que pensam racionalmente e sistemas que agem racionalmente) não se preocupam em emular condições e processos humanos, o foco está na

racionalidade. Independente da abordagem adotada, os agentes inteligentes devem ser capazes de tomar decisões racionais.

Já MONDAL (2020) define a Inteligência Artificial como a capacidade de máquinas ou artefatos de aprender, raciocinar e resolver problemas (não restritos a condições extremamente específicas, mas também problemas novos).

4.1.2 EVOLUÇÃO

RUSSELL e NORVIG (2003) trazem o histórico da Inteligência Artificial, cravando seu início na década de 1940, com a criação de um modelo de neurônio artificial, por MCCULLOCH e PITTS (1943), embora o termo Inteligência Artificial tenha se consolidado apenas em 1956, sendo creditado a John McCarthy. Os autores dividem a evolução histórica da seguinte forma:

- i. 1943 a 1956: A “Gestação e nascimento da Inteligência Artificial”. Os autores trazem que o início da área se dá com o modelo de neurônio artificial, baseado na célula nervosa humana, em trabalho de Warren McCulloch e Walter Pitts, de 1943, em que eles mostram que uma rede interconectada desses neurônios é capaz que aproximar funções e representar operadores lógicos.
- ii. O período de 1952 a 1969 é tratado como a germinação de um entusiasmo no campo da Inteligência Artificial, com a criação de algumas aplicações práticas que, apesar de relativamente simples, ajudaram a diminuir o ceticismo na ainda existente.
- iii. O período de 1966 a 1974 ficou marcado pela constatação de que o razoável sucesso que vinha sendo obtido com problemas simples não era traduzido para problemas mais complexos, seja por inaptidão dos sistemas elaborados ou pela capacidade computacional insuficiente.
- iv. O período de 1969 a 1979 ficou marcado por desenvolvimento de sistemas “especialistas”, baseados em conhecimentos específicos

de determinado domínio. Neste período, foram proeminentes trabalhos nas áreas de medicina diagnóstica e linguística.

- v. Posteriormente, a partir de 1980, houve maior incorporação da Inteligência Artificial na indústria, elevando o interesse geral na área e as aplicações da tecnologia.

Os autores trazem, ainda, a importância histórica da filosofia, matemática, psicologia, ciências da computação e linguística, dentre outras, para a fundamentação e desenvolvimento da Inteligência Artificial.

A análise da evolução histórica da Inteligência Artificial apresentada mostra a alternância de períodos de êxtase e desilusão, partes do processo de amadurecimento da disciplina. Hoje, a área de Inteligência Artificial está em pleno desenvolvimento e disseminação, sendo aplicada a cada vez mais áreas. Exemplo disto é o presente trabalho, que busca o emprego desta tecnologia para a previsão e prevenção de desastres naturais.

4.1.3 SUBÁREAS

A disciplina Inteligência Artificial não é restrita a um único eixo. Existem diversos desafios, tratados como subáreas, com enfoque variado, tais como:

- i. Processamento de linguagem natural (NLP – *Natural Language Processing*);
- ii. Capacidade de aprendizado (*Machine Learning*);
- iii. Representação de conhecimentos;
- iv. Habilidade de raciocínio;
- v. Visão computacional;

O presente trabalho está inserido na subárea de aprendizado de máquina (*machine learning*), pois objetiva a construção de sistemas capazes de aprender a partir de dados e tomar decisões futuras a partir da experiência anterior.

O processo de aprendizagem é definido por RUSSELL e NORVIG (2003) como a capacidade de um agente inteligente melhorar seu comportamento com base em experiências. Isso se dá através do chamado Aprendizado Induzido (*Inductive Learning*), no qual um agente é capaz de aprender a partir de exemplos apresentados, por meio de algoritmos específicos.

4.1.4 REDES NEURAIIS ARTIFICIAIS

4.1.4.1 Conceituação

Rede Neural Artificial é um tipo de modelo computacional capaz de representar funções complexas inspirado no sistema nervoso humano, que tem seu surgimento vinculado ao surgimento da Inteligência Artificial, em 1943. A ideia geral envolve uma rede interconectada de unidades que processam dados de entrada e geram uma saída. A saída de cada unidade pode ser entendida como seu estado de ativação, assim como em neurônios no sistema nervoso animal. (RUSSELL e NORVIG, 2003).

Desde seu surgimento, as Redes Neurais Artificiais geraram grande interesse por serem capazes de representar qualquer função computável. Essas redes têm enorme potencial de aplicação devido à existência de algoritmos que possibilitam um processo de aprendizagem de máquina (*machine learning*), tornando-as capazes de aprender a partir de exemplos (processo chamado de aprendizado induzido), mesmo em domínios altamente complexos. (RUSSELL e NORVIG, 2003).

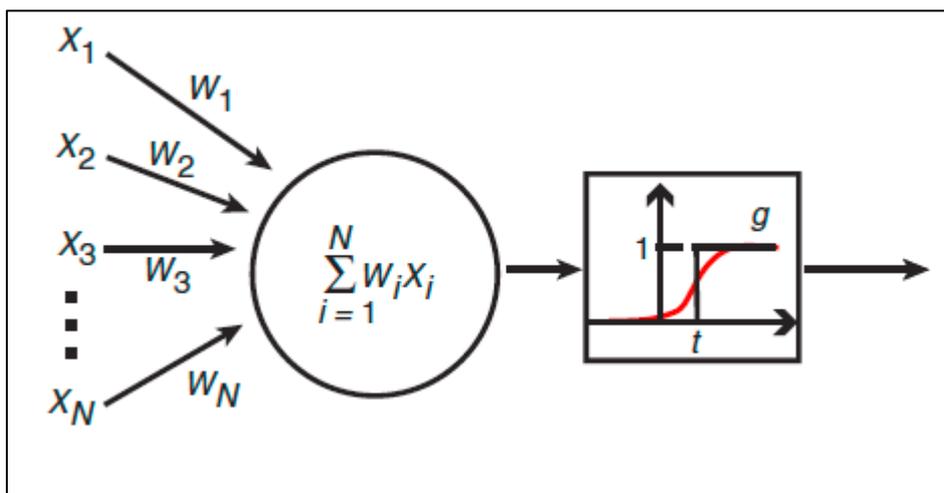
4.1.4.2 Elementos

As Redes Neurais Artificiais são constituídas por unidades de processamento, que possuem pesos e uma função de ativação. Para cada conjunto de dados de entrada, é computada a ativação de cada unidade. Este processo é dividido em duas etapas. Primeiramente, é computado o produto

escalar entre o vetor de dados de entrada e o vetor de pesos da unidade. Em seguida, o resultado dessa operação é utilizado na função de ativação, que quantifica a ativação da unidade.

A Figura 12, extraída de KROGH (2008) esquematiza este processo. A unidade de processamento computa o somatório do produto de cada dado de entrada ($X_1, X_2, X_3, \dots, X_N$) com seu respectivo peso ($W_1, W_2, W_3, \dots, W_N$). O resultado deste somatório é então utilizado na função de ativação, que determinará o grau de ativação da unidade. Um exemplo de função de ativação é a função sigmoide, que gera resultados na escala de 0 a 1.

Figura 12 - Esquema de uma unidade de processamento em uma rede neural.



Fonte: Extraído de KROGH (2008).

As Redes Neurais se tornam úteis devido ao processo de aprendizagem, no qual os pesos das unidades de processamento são ajustados, possibilitando aproximar funções. O método de aprendizado geralmente utilizado é a Retro-Propagação (*Back-Propagation*), que implementa a técnica de gradiente descendente (*gradient descent*) para redes com mais de uma camada. Enquanto o cálculo do gradiente descendente direciona o ajuste dos pesos, indicando o gradiente de cada peso para um mínimo local da função de erro, o

algoritmo de Retro-Propagação é utilizado como uma forma de distribuir a necessidade de ajuste ao longo da rede. (RUSSELL e NORVIG, 2003).

4.1.4.3 Tipos

Existem algumas variantes de Redes Neurais Artificiais, com distintas topologias e tipos de unidades de processamento. A Rede Neural mais simples é o Perceptron, uma rede formada por uma única unidade de processamento, que gera uma única saída. Na fase inicial da Inteligência Artificial, na década de 1950, os estudos de Redes Neurais foram focados no Perceptron, sendo demonstrado que era capaz de representar funções linearmente separáveis, como as funções lógicas “E” e “OU”. (RUSSELL e NORVIG, 2003).

No entanto, para aplicações mais complexas, se faz necessária a construção de redes com mais de uma unidade e, geralmente, organizadas em mais de uma camada. Esta topologia é chamada de Multilayer Perceptron (MLP), na qual, usualmente, cada unidade de uma camada está conectada a todas unidades da camada seguinte. Redes do tipo MLP são capazes de representar qualquer função booleana, por exemplo, porém trazem maiores dificuldades para o treinamento, processo de ajuste de pesos. (RUSSELL e NORVIG, 2003).

Por isso, inicialmente, não eram utilizadas redes MLP com muitas camadas, devido a dificuldades metodológicas e computacionais para ajuste dos pesos. No entanto, nas últimas décadas, se popularizou a utilização de redes muito maiores e mais complexas. Estas redes com maior número de camadas são chamadas de Redes Neurais Profundas (*Deep Neural Networks*), enquanto seu processo de aprendizado tem sido referenciado como aprendizado profundo (*Deep Learning*). Estas redes profundas têm capacidade de resolver problemas muito mais complexos, e seu emprego vem crescendo e se difundindo para os mais diversos domínios, de aplicações que buscam o reconhecimento de fala até a previsão meteorológica.

As Redes Neurais até então discutidas são do tipo Feed-forward Network, nas quais as conexões entre as unidades seguem um fluxo unidirecional. Em contraste, existem outras topologias, sendo a alternativa mais comum as Redes Recorrentes, que têm sua aplicação voltada para séries temporais. Nestas redes, o fluxo de informação não é unidirecional, podendo formar ciclos, de forma que o estado de ativação de determinada unidade é influenciado por estados anteriores ou mesmo por estados de ativação de unidades em camadas posteriores no arranjo. Assim, este tipo de topologia permite que as unidades guardem certa memória. Entre as principais variantes de Redes Recorrentes estão as Redes Neurais Recorrentes (RNN - *Recurrent Neural Network*) e as Redes Neurais Recorrentes com Memória Longa de Curto-Prazo (LSTM - *Long Short-term Memory*). Apesar do potencial, estas redes são difíceis de serem treinadas, sendo de certa forma sensíveis e instáveis, demandando bastante esforço até que se consiga uma rede com resultados satisfatórios (RUSSELL e NORVIG, 2003).

4.1.4.3.1 Redes Neurais Convolucionais

Outra variante das Redes Neurais Artificiais são as Redes Neurais Convolucionais (CNN - *Convolutional Neural Networks*). Estas redes são utilizadas sobretudo para o processamento de imagens ou vídeos. Através da operação matemática de convolução (utilizada em matrizes), as unidades deste tipo de rede são excelentes para identificação de padrões e reconhecimento de feições, não sendo dependentes da localização específica das feições no espaço dos dados de entrada. Por exemplo, uma Rede Neural Convolucional treinada para reconhecimento de rostos humanos em imagens aprende a identificar as feições do rosto independente de sua localização na imagem apresentada. Além disso, estas redes são muito tolerantes a alterações nos dados de entrada, como ruídos ou rotação em fotos, por exemplo. (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

Outra grande vantagem da CNN sobre MLP é que tendem a possuir muito menos parâmetros treináveis para uma mesma tarefa. Isto as torna muito mais simples e rápidas de serem treinadas.

Apesar de terem seu uso primordial aplicado ao processamento de imagens, as Redes Neurais Convolucionais têm encontrado grande sucesso no processamento de séries temporais. As séries temporais podem ser entendidas como imagens de apenas uma dimensão “espacial”, enquanto uma foto tem duas dimensões espaciais.

ZHAO et al. (2017) evidenciam a utilização de CNN para classificação de séries temporais, ressaltando resultados superiores a outras técnicas consideradas estado da arte para processamento e análise de séries temporais.

Devido ao enorme potencial das Redes Neurais Convolucionais (CNN), este trabalho busca sua utilização para análise e previsão de séries temporais multivariadas.

4.2 PÓS-PROCESSAMENTO DE MODELOS NUMÉRICOS DE PREVISÃO DO TEMPO

Redes Neurais Artificiais vêm sendo utilizadas nas mais diversas aplicações, com grande potencial de resultados promissores. Muitos estudos propõem e avaliam a performance destas redes, com as mais diversas configurações, aplicadas à modelagem meteorológica. Alguns usos envolvem a utilização destas redes de forma absoluta, modelando dados meteorológicos históricos diretamente, com resultados compatíveis e por vezes melhores que a modelagem meteorológica clássica e complexa, como a utilizada no modelo WRF (HEWAGE et al., 2020).

Já outras abordagens utilizam técnicas de inteligência artificial para a redução de erros dos modelos numéricos de previsão do tempo, o que pode ser entendido como um pós-processamento dos resultados da modelagem meteorológica.

Todos os estudos analisados apresentam ganhos de performance expressivos com a utilização dessas técnicas para o pós-processamento da modelagem meteorológica. Os estudos que tiveram como objetivo a própria modelagem e previsão meteorológica, de forma absoluta e independente de modelos numéricos de previsão do tempo também tiveram bons resultados, embora principalmente para horizontes de previsão mais curtos.

JABBARI e BAE (2018) utilizaram redes neurais artificiais para o pós-processamento e redução de viés das previsões de precipitação do modelo WRF, em um estudo de caso na Coréia do Sul. Além de avaliar a performance das previsões antes e após o pós-processamento, os autores investigaram o impacto no desempenho de modelos hidrológicos que utilizam os dados de previsão de precipitação para a previsão em tempo real de alagamentos.

Os resultados obtidos, tanto na performance das previsões de precipitação, como no desempenho da previsão de alagamentos, com a utilização do modelo de pós-processamento do WRF foram expressivos. Com isso, os autores concluem que o pós-processamento de modelos numéricos de previsão do tempo é essencial.

RASP e LERCH (2018) avaliaram o emprego de redes neurais artificiais para o pós-processamento de modelos numéricos de previsão do tempo, comparando com técnicas tradicionais. No estudo, fizeram a previsão de temperatura na superfície, comparando com dados meteorológicos históricos de estações espalhadas por toda a Alemanha.

A rede foi desenvolvida para fazer previsão probabilística, na qual se tenta obter a distribuição de probabilidade da variável de interesse. A técnica proposta foi contrastada com a técnica clássica Ensemble Model Output Statistics (EMOS) e a mais recente Quantile Regression Forests (QRF), que utiliza *Random Forests*.

Importante ressaltar que os autores utilizaram a informação da localidade através de embeddings, e não *one-hot-encoding*.

Os resultados demonstram que a técnica proposta superou técnicas de pós-processamento consideradas estado da arte, além de ser mais eficiente.

Os autores afirmam que o emprego de técnicas de *deep learning* para o pós-processamento de modelos numéricos de previsão do tempo terá uma importância marcante em um futuro próximo.

HEWAGE et al. (2020) utilizaram redes neurais profundas para a previsão de dados meteorológicos a partir de dados meteorológicos históricos. Os autores compararam a performance dos modelos propostos com modelos clássicos, incluindo o WRF e chegaram à conclusão de que as redes neurais utilizadas foram capazes, não só de modelar os dados meteorológicos, mas também de superar a performance do WRF, para um horizonte de previsão de até 12h.

Fica claro que a utilização de redes neurais artificiais aplicadas à previsão meteorológica vem ganhando espaço nos últimos anos, principalmente com a atual possibilidade do emprego de redes maiores e mais complexas, com o surgimento do chamado Aprendizado Profundo (*deep learning*). Entre as principais vantagens destas redes estão o baixo tempo de processamento e a versatilidade, já que são capazes de modelar os mais diversos fenômenos.

As alternativas de uso são variadas, como mostra a revisão de literatura. Estas redes podem ser utilizadas tanto para a previsão meteorológica diretamente, a partir de dados observados históricos, ou serem utilizadas para fazer um ajuste fino em resultados de Modelos Numéricos de Previsão do Tempo.

4.3 PREVISÃO DE DESLIZAMENTOS DE TERRA

Diversos estudos tentam estabelecer relações entre deslizamentos de terra e chuva, buscando uma capacidade de previsão desses deslizamentos. Parte-se do conhecimento de que a precipitação é um dos fenômenos indutores, como exposto por DAI e LEE (2001), CHOWDHURY e FLENTJE (2002) e extensiva literatura.

Além da chuva, os deslizamentos de terra são provocados pela interação de outros fatores, como as características do solo, a topografia e as intervenções antrópicas, principalmente. DAI e LEE (2001) reconhecem a importância destes fatores, mas ressaltam a dificuldade da inclusão destas informações, em virtude de sua inexistência, indisponibilidade ou mesmo incompatibilidade de escalas.

DAI e LEE (2001) modelaram a relação entre precipitação e deslizamentos de terra em Hong Kong, analisando a frequência dos deslizamentos e o volume de terra deslocado, em função da precipitação. Foram identificadas as características de chuva mais relevantes para a previsão de deslizamentos, sendo elas indicadoras da precipitação em curto espaço de tempo (12 horas e 24 horas) e da precipitação com maior antecedência (1 dia ou 30 dias), dependendo do propósito.

DAI e LEE (2001) ressaltam, ainda, que a capacidade de modelagem da relação entre precipitação e deslizamentos de terra é altamente dependente da quantidade e qualidade dos dados disponíveis. Isto indica a necessidade de regularmente se revisar os modelos propostos, conforme a base de dados é enriquecida.

KORUP e STOLLE (2014) analisaram a utilização crescente de técnicas relacionadas à ciência de dados, como mineração de dados e aprendizado de máquina, na previsão de deslizamentos de terra. Eles destacam o potencial destas técnicas para a detecção de padrões, principalmente considerando-se que não é incomum a existência de conjuntos de dados incompletos e com múltiplos parâmetros.

KORUP e STOLLE (2014) ressaltam que, apesar da grande quantidade de pesquisas aplicadas na área, a previsão de deslizamentos de terra, no tempo e no espaço, continua um desafio. No levantamento e avaliação de pesquisas existentes, os autores concluíram que a maioria das pesquisas nesta área envolvem a espacialização da susceptibilidade a deslizamentos, e não a previsão ou mensuração de riscos futuros, propriamente ditos.

KORUP e STOLLE (2014) destacam, ainda, que as abordagens utilizadas para a análise de susceptibilidade a deslizamentos costumam tentar a modelagem do fenômeno em si, resolvendo equações determinísticas que consideram as resistências e estabilidades das encostas. Em termos práticos, as limitações destas abordagens emergem da indisponibilidade de dados, incompatibilidade de escalas e simplificações dos contextos reais, principalmente. Além disso, os parâmetros estão sujeitos a elevado dinamismo no tempo.

Na análise da qualidade dos modelos propostos, KORUP e STOLLE (2014) afirmam que esta não aumenta consideravelmente com a complexidade dos modelos empregados. Alertam, ainda, que as taxas de sucesso apresentadas devem ser observadas com cautela, em virtude da especificidade dos modelos para cenários locais.

TIEN BUI et al. (2016) analisaram a eficácia de diferentes modelos para construção de mapas de susceptibilidade a deslizamentos, explorando técnicas para a adequada seleção de parâmetros.

Neste estudo, foram avaliados Support Vector Machines (SVM), Redes Neurais Artificiais (do tipo MLP e RBF), Kernel Logistic Regression (KLR) e Logistic Model Tree (LMT). Entre os 5 modelos avaliados, os autores concluíram que a Rede Neural com MLP apresentou melhor performance.

Os autores ressaltaram a importância da seleção de parâmetros que agreguem valor aos modelos, evitando a inserção de variáveis que causem “ruídos” ou muito correlacionadas com outras variáveis, e que resultem no aumento de complexidade e diminuição da robustez dos modelos.

LOGAR et al. (2017) avaliaram a capacidade de Redes Neurais Artificiais para prever o deslocamento de terra em dois locais diferentes, e com propriedades diferentes, um na Eslovênia (Macesnik) e outro no Reino Unido (Isle of Wight). O deslocamento de Macesnik possui um movimento médio de

100 mm por dia, enquanto o de Isle of Wight 5mm por ano. Ambos são considerados como induzidos por chuvas.

Os autores ponderam que, embora a modelagem da estabilidade de encostas seja eficaz, dificilmente é praticável operacionalmente, por limitação dos dados. Assim, implementaram um modelo de previsão utilizando apenas dados históricos de pluviometria e de deslocamentos das encostas.

LOGAR et al. (2017) concluíram que as Redes Neurais Artificiais são capazes de modelar deslizamentos de terra, mesmo apenas com dados de precipitação e deslocamentos anteriores.

TIEN BUI et al. (2017) produziram mapas de susceptibilidade para uma região do Vietnã, propondo a utilização de uma variante de Support Vector Machine (Least Squares Support Vector Machine - LSSVM), acoplado a um método para otimização de hiperparâmetros (Artificial Bee Colony - ABC).

Os autores concluíram que o método utilizado apresentou bons resultados e tem bom potencial para a previsão e mapeamento de susceptibilidade a deslizamentos de terra.

ZHU et al. (2017) tentaram, também, modelar o deslocamento de uma encosta (na China), a partir de dados históricos de deslocamentos e chuva. Os autores avaliaram o desempenho de dois métodos, ambos utilizando LSSVM (Least Squares Support Vector Machine). Em uma abordagem, utilizaram um algoritmo genético para a otimização do LSSVM. Na outra, utilizaram Double Exponential Smoothing (DES), uma técnica para a modelagem da tendência em séries temporais. No trabalho, chama a atenção a modelagem das componentes de tendência (longo prazo) e de periodicidade (curto prazo).

Eles concluíram que a chuva era, de fato, o desencadeador crucial do deslizamento, e que ambas as abordagens utilizadas apresentaram bons resultados. Ressalta-se que os modelos previram apenas um time-step (“passo”) a frente.

PHAM et al. (2018), também utilizando métodos de classificação para a produção de mapas de susceptibilidade, estudaram o potencial do método Aggregating One-Dependence Estimators (AODE), ainda não aplicado à previsão de deslizamentos de terra. Em comparação com modelos mais tradicionais, como SVM, Redes Neurais Artificiais e Regressão Logística. Os autores chegaram à conclusão que existe grande potencial do AODE nesta área, já que este modelo superou os demais no caso estudado.

YANG et al. (2019), assim como ZHU et al. (2017), fizeram modelagem da série temporal de dados de deslocamento em um deslizamento de terra decompondo os componentes de tendência e de periodicidade. Advogando a favor de uma modelagem dinâmica de séries temporais para a previsão de deslizamentos de terra, por meio de Redes Neurais do tipo LSTM (Long Short-Term Memory), em vez da utilização de abordagens consideradas estáticas pelos autores, como MLP, SVM e etc., eles concluíram que o modelo utilizado produziu resultados mais confiáveis.

HUANG et al. (2020) propuseram a utilização de uma Rede Neural do tipo Autoencoder para a extração de características das variáveis ambientais utilizadas. O resultado deste autoencoder foi, então, utilizado como entrada para um modelo do tipo SVM de classificação. O estudo teve como objetivo a produção de mapa de susceptibilidade a deslizamentos de terra, em função de diversas variáveis ambientais e um histórico de deslizamentos.

Os autores concluíram que a técnica proposta tem grande potencial de aplicação, já que na comparação feita com técnicas mais tradicionais, como apenas SVM e MLP, o modelo que utilizava autoencoder apresentou resultados superiores.

A pesquisa de literatura relacionada ao tema abordado neste trabalho deixa claro que existem diversas tentativas de se modelar e prever deslizamentos de terra, em vários países. Fica evidente, também, que esta área ainda se encontra em desenvolvimento, com grande potencial de melhoria e já aplicação em situações operacionais.

Foram encontrados basicamente dois tipos de propostas: um que tenta a previsão de deslocamentos lentos em deslizamentos de terras localizados e já monitorados, e outro que busca o mapeamento da susceptibilidade a deslizamentos de terra.

O presente trabalho utiliza uma abordagem diferente e, portanto, inovadora: a de previsão do quantitativo de deslizamentos de terra, em diferentes regiões da cidade de Salvador, delimitadas pela distância a estações pluviométricas existentes.

Além disso, os deslizamentos de terra constantes na base de dados utilizada neste trabalho possuem características peculiares, por se encontrarem em ambiente bem urbanizado, sendo comumente induzidos por ações antrópicas, além das condições ambientais.

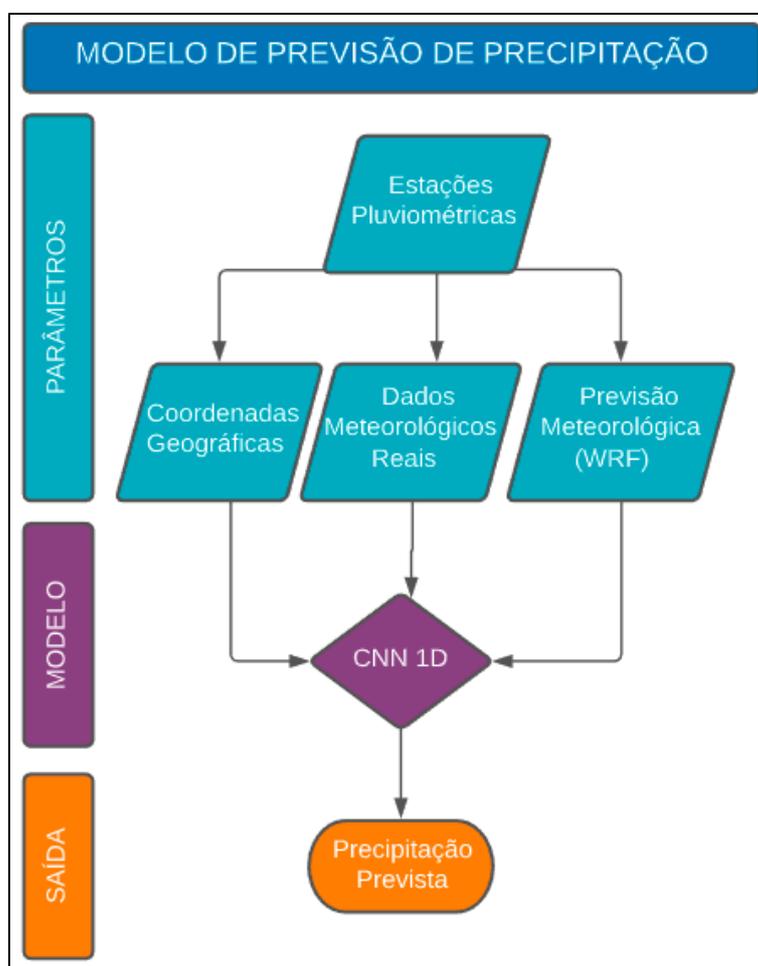
Ainda que os casos estudados pelos diferentes autores sejam diferentes do proposto neste trabalho, fica evidente que o fenômeno dos deslizamentos de terra pode ser modelado por Redes Neurais Artificiais.

Assim, opta-se pela utilização de Redes Neurais Artificiais para o desenvolvimento de um modelo de previsão de deslizamentos de terra. Esta tecnologia tem algumas características marcantes, como relativo baixo tempo de processamento, tanto para o processo de aprendizado quanto, principalmente, para a produção das previsões, além de ter alta versatilidade, sendo capaz de modelar os mais distintos fenômenos, complexos ou não.

5 METODOLOGIA

O presente trabalho foi dividido em duas partes. Primeiramente, foi desenvolvido um modelo de previsão de chuva, com horizonte de previsão de 96h, a partir dos resultados do modelo numérico de previsão do tempo WRF. A Figura 13 mostra um fluxograma que esquematiza este modelo de previsão de precipitação.

Figura 13 - Fluxograma do modelo de previsão de precipitação.

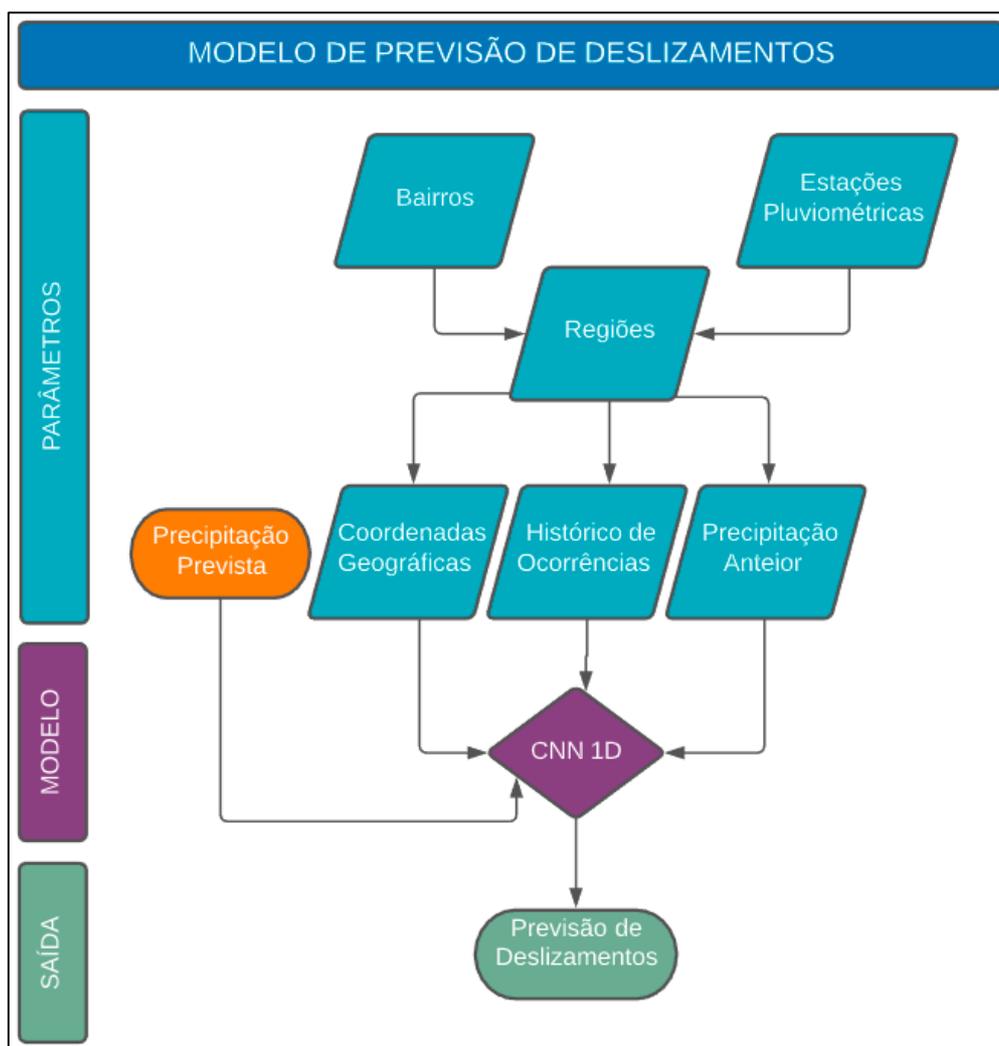


Fonte: Autor.

Em seguida, foi desenvolvido um segundo modelo, que, utilizando dados meteorológicos observados, um histórico de deslizamentos de terra e dados de

previsão de chuva (resultados do primeiro modelo desenvolvido), prevê a quantidade de deslizamentos de terra, com um horizonte de previsão de 72h, de forma espacializada. A Figura 14 mostra um fluxograma que esquematiza este modelo de previsão de deslizamentos de terra.

Figura 14 - Fluxograma do modelo de previsão de deslizamentos.



Fonte: Autor.

Todos os dados utilizados neste trabalho foram convertidos para horário UTC.

5.1 MODELO DE PREVISÃO DE PRECIPITAÇÃO (PÓS-PROCESSAMENTO DO WRF)

Para a previsão de chuva, foi construído um modelo que fizesse um pós-processamento dos resultados de precipitação do modelo numérico de previsão do tempo WRF, com o objetivo de reduzir seus erros. Para isso, utilizaram-se resultados históricos do WRF e dados meteorológicos observados históricos, medidos por estações pluviométricas e meteorológicas.

Primeiramente, foram calculadas métricas que indicassem a qualidade da previsão de precipitação do modelo WRF, analisando-se a influência de alguns fatores, como a resolução espacial dos dados de entrada e o mês do ano.

Em seguida, de posse desses resultados, foi identificada a resolução espacial dos dados de entrada ($0,25^\circ$ ou $1,0^\circ$ de latitude e longitude) que produzia melhores resultados, considerando-se que não fazia sentido usar ambas, por redundância e alto custo operacional para executar o modelo WRF.

Depois, foram identificadas as características que deveriam ser discriminadas para o modelo de pós-processamento proposto, como a hora do dia e o dia do ano, por exemplo. Finalmente, foi feita engenharia e transformação dos dados, adequando-os para a modelagem proposta.

O tipo de rede neural escolhida (CNN – *Convolutional Neural Network*), seus parâmetros e a utilização dos resultados do WRF com os dois domínios simultaneamente (resoluções de 5 km e de 20 km) decorreram de análises empíricas, buscando-se a estratégia que produzia os melhores resultados.

Para a análise dos resultados do modelo de pós-processamento proposto, foram utilizadas as seguintes métricas:

- Erro Absoluto Médio (MAE)
- Erro Quadrático Médio (MSE)
- Coeficiente de correlação de Pearson (R)
- Coeficiente de determinação (R^2)
- Fator de 2 (FAC2)

5.1.1 ANÁLISE EXPLORATÓRIA DE DADOS

Inicialmente, foi feita uma análise dos resultados do modelo WRF, focada na precipitação prevista. Esta análise dos resultados do WRF é essencial para a identificação dos parâmetros que influenciam a qualidade da previsão. Para isso, foram avaliadas métricas para cada um dos seguintes parâmetros:

- Resolução espacial dos dados de entrada do modelo (1º ou 0,25º de latitude e longitude)
- Domínio (grades com espaçamento de 5 km ou de 20 km)
- Diferentes agrupamentos horários (dados agrupados para cada 3 horas, 6 horas, etc.)
- Hora do dia
- Mês
- Estação do ano

5.1.2 ENGENHARIA DE DADOS

Como os resultados do WRF apresentaram melhores métricas com a resolução espacial de 0,25°, esta foi selecionada para a modelagem e pós-processamento. Os dados da resolução de 1° não serão utilizados.

O *dataset* utilizado para construir o modelo de pós-processamento do WRF foi então composto com os resultados do WRF para a resolução espacial de 0,25° (com dados de ambas as grades - 5 km e 20 km - e todas as 8 variáveis previstas), dados meteorológicos reais registrados por uma estação meteorológica automática e os dados de precipitação registrados pelas estações pluviométricas.

Devido à grande variação da qualidade dos resultados WRF ao longo do ano e de cada dia, as informações do dia do ano e da hora do dia foram incluídas como variáveis preditoras. Por terem caráter cíclico, estas variáveis foram transformadas, através das funções trigonométricas seno e cosseno (gerando duas variáveis cada).

Para reduzir a oscilação da qualidade dos dados, os dados horários foram agrupados em intervalos de 3 horas. Além disso, todas as variáveis foram escalonadas para o intervalo 0 a 1.

Os seguintes 31 atributos foram utilizados como preditores:

- *Timestep* da previsão do WRF (1 atributo);
- Seno e cosseno do dia do ano (2 atributos);
- Seno e cosseno da hora do dia (2 atributos)
- Latitude e longitude da estação para a qual é feita a previsão (2 colunas);
- Dados meteorológicos reais medidos por uma estação meteorológica automática (precipitação, pressão, radiação, temperatura do ar, umidade relativa do ar, velocidade do vento, seno da direção do vento e cosseno da direção do vento) (8 atributos);
- As 8 variáveis meteorológicas previstas pelo WRF para a grade de 5 km (8 atributos);

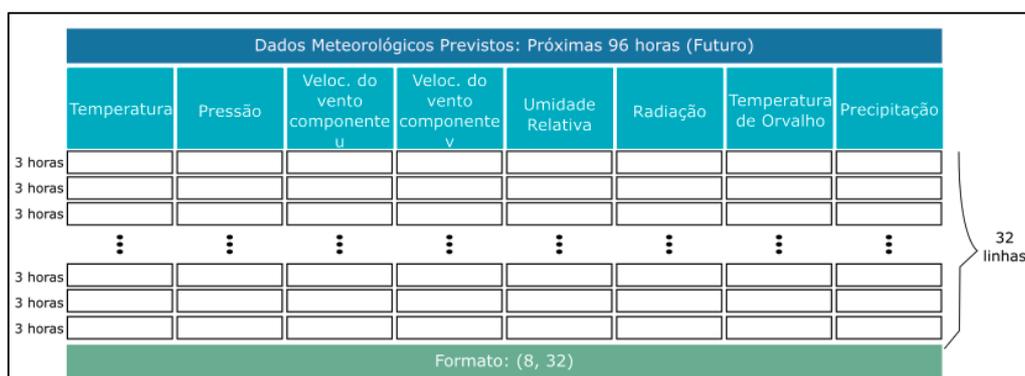
- As 8 variáveis meteorológicas previstas pelo WRF para a grade de 20 km (8 atributos);

Na rede convolucional, foi utilizada uma janela temporal de tamanho 32, de forma que cada instância com os dados do WRF contém a previsão das próximas 96 horas, subdivididas em 32 intervalos subsequentes (Figura 15) e cada instância com os dados meteorológicos reais contém dados das últimas 96 horas, subdivididas em 32 intervalos subsequentes (Figura 16)

Dessa forma, cada instância passada para o modelo tem 31 atributos, sendo que, para cada atributo, é passada uma série temporal, com o valor deste atributo em 32 intervalos subsequentes de tempo. No caso de variáveis que não têm evolução temporal, como as coordenadas da estação para a qual é feita a previsão, o mesmo valor é repetido as 32 vezes.

Com isso, a rede convolucional utilizada é capaz de captar e modelar a dinâmica temporal destas variáveis.

Figura 15 - Formato das instâncias com dados do WRF.



Fonte: Autor.

Figura 16 - Formato das instâncias com dados da estação meteorológica.

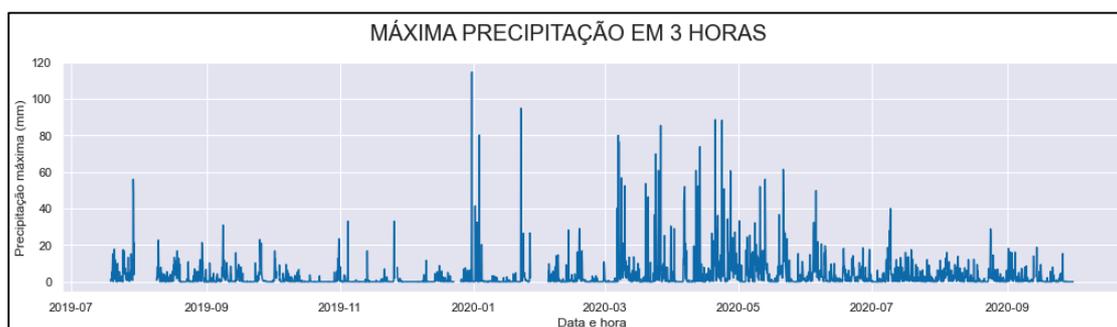
Dados Meteorológicos Reais: Últimas 96 horas (Passado)							
	Temperatura	Pressão	Velocidade do vento	Direção do vento	Umidade Relativa	Radiação	Precipitação
3 horas							
3 horas							
3 horas							
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3 horas							
3 horas							
3 horas							
Formato: (8, 32)							

32 linhas

Fonte: Autor.

Após todo o pré-processamento e engenharia de dados, o conjunto final de dados utilizados para a modelagem foi composto 11742 registros únicos. Estes dados compreendem o período entre 19/07/2019 a 30/09/2020. A Figura 17 mostra um gráfico da máxima precipitação registrada em intervalos de 3 horas, independente da estação pluviométrica, para todo o período utilizado.

Figura 17 - Máxima precipitação em 3 horas, ressaltando o período efetivamente utilizado para a modelagem.



Fonte: Autor.

5.1.3 CONSTRUÇÃO DO MODELO DE INTELIGÊNCIA ARTIFICIAL

Para a avaliação do modelo, foi utilizada a técnica de validação cruzada (*cross-validation*), com 5 iterações. Cada uma utilizou a seguinte proporção aproximada do conjunto de dados:

- Treino: 68 %
- Validação: 12 %
- Teste: 20 %

Para a modelagem de séries temporais, o cenário ideal envolveria que as subdivisões de treino, validação e teste fossem contínuas, de forma que, por exemplo, o conjunto de teste fosse composto por todos os dados dos últimos 3 meses da série temporal.

No entanto, o conjunto de dados efetivamente utilizado compreende um período muito curto, de pouco mais de um ano (possuindo dados de apenas um período chuvoso), e existe uma variação clara da precipitação ao longo do ano (como mostra a Figura 17). Se fosse utilizada essa estratégia, as instâncias do único período chuvoso do conjunto de dados não estariam representadas em todas as três subdivisões (treino, validação e teste), dificultando a modelagem.

Para contornar esta situação, foi adotada outra estratégia: optou-se por selecionar amostras aleatoriamente para cada subdivisão (treino, validação e teste), estratificando de acordo com o mês, de forma que cada subdivisão contivesse a mesma proporção de amostras de cada mês que o conjunto completo. Com isso, instâncias consecutivas em cada subdivisão (treino, validação e teste), não representam, necessariamente, intervalos subsequentes de tempo.

Vale ressaltar que, no momento da divisão para treino, validação e teste, cada instância já possuía uma janela temporal com 32 intervalos subsequentes de tempo, de forma que a seleção aleatória das instâncias não alterou a janela temporal já existente em cada uma dessas instâncias. Ou seja, mesmo após a

divisão para treino, validação e teste, cada instância continua contendo registros de 32 intervalos subsequentes de tempo.

O modelo de inteligência artificial foi construído utilizando-se o *framework* da biblioteca TensorFlow v 2.2.0.

Foi utilizada uma rede sequencial com 2 camadas do tipo Conv1D (cada uma com com 100 filtros e kernel de tamanho 2) e 1 camada do tipo Dense (com 32 unidades). Foi incluída camada de normalização dos *batches* e utilizou-se Relu como função de ativação nas duas primeiras camadas (do tipo Conv1D) e Softplus na camada de saída (com 32 unidades). Foi utilizado um *batch* de tamanho 45, por 65 épocas. O modelo foi treinado utilizando como métrica o Erro Quadrático Médio (MSE).

Tanto os hiperparâmetros como a quantidade de camadas e de unidades em cada camada foram ajustados empiricamente, mediante tentativa e erro. A Figura 18 mostra a arquitetura da rede utilizada.

Figura 18 - Rede utilizada para a previsão de chuva.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 32, 100)	6300
batch_normalization (Batch Normalization)	(None, 32, 100)	400
activation (Activation)	(None, 32, 100)	0
conv1d_1 (Conv1D)	(None, 32, 100)	20100
batch_normalization_1 (Batch Normalization)	(None, 32, 100)	400
activation_1 (Activation)	(None, 32, 100)	0
flatten (Flatten)	(None, 3200)	0
dense (Dense)	(None, 32)	102432
Total params: 129,632		
Trainable params: 129,232		
Non-trainable params: 400		

Fonte: Autor.

5.2 MODELO DE PREVISÃO DE DESLIZAMENTOS DE TERRA

Considerando-se a grande influência das chuvas nos deslizamentos de terra em Salvador, buscou-se desenvolver um modelo de previsão de deslizamentos de terra em função, principalmente, de dados pluviométricos. Para isso, foram acoplados dados históricos de deslizamentos de terra com dados meteorológicos no momento de seus registros (na verdade, no momento em que houve abertura de processo na Defesa Civil municipal).

Buscando-se não apenas modelar a relação entre chuva e deslizamentos de terra, mas sistematizar uma previsão de ocorrências, foram acoplados também dados de previsão meteorológica, utilizando-se o modelo de pós-processamento do WRF anteriormente desenvolvido neste mesmo trabalho.

Inicialmente, foi feita análise exploratória e limpeza dos dados. Alguns registros pluviométricos foram considerados inválidos, em função de possibilidade de problemas operacionais da estação pluviométrica. Como a equipe técnica da CODESAL faz análise manual diária do funcionamento dessas estações monitoradas, esta informação foi utilizada como um dos critérios para anular registros possivelmente incorretos. Foram também removidos valores absurdos e abruptos.

As estações pluviométricas utilizadas também possuem diferentes frequências de transmissão e medição de valores, algumas com frequência horária, outras com intervalos de 10 minutos e outras com intervalos de 5 minutos. Todos os dados foram, então, reamostrados para uma frequência horária.

Além disso, algumas estações pluviométricas foram descartadas, por estarem muito próximas entre si e serem consideradas redundantes.

Em seguida, foi feita análise dos dados de deslizamentos de terra. Para otimizar a capacidade de operacionalização deste modelo, optou-se por dividir o município em regiões, de forma que cada um dos 170 bairros de Salvador pertença a apenas uma região e cada região esteja associada a uma estação

pluviométrica. Esta estratégia permitiu a conexão entre os dados pluviométricos e os deslizamentos de terra registrados.

Foram gerados, então, dados de previsão de chuva para cada uma das estações pluviométricas, utilizando o modelo de pós-processamento do WRF também desenvolvido neste trabalho. Infelizmente, os resultados do WRF começaram a ser armazenados em julho de 2019, o que resulta em apenas cerca de 14 meses de dados de previsão meteorológica.

Todo o conjunto de dados foi, então, reamostrado para uma frequência de 6 horas. Finalmente, foi feita transformação dos dados, adequando-os para a modelagem proposta.

A escolha do tipo de rede neural (CNN – *Convolutional Neural Network*) e a configuração de seus parâmetros decorreram de testes empíricos.

Para a análise dos resultados do modelo desenvolvido, foram utilizadas as seguintes métricas:

- Erro Absoluto Médio (MAE)
- Erro Quadrático Médio (MSE)
- Coeficiente de correlação de Pearson (R)
- Coeficiente de determinação (R^2)
- Fator de 2 (FAC2)

5.2.1 ANÁLISE EXPLORATÓRIA DE DADOS

Primeiramente, fez-se uma análise espacial da distribuição dos deslizamentos de terra registrados, avaliando-se a quantidade de ocorrências por bairro. Em seguida, investigou-se a variação da quantidade de ocorrências registradas com a hora do dia, o dia da semana, o dia do ano, o mês do ano e o ano. Ressalta-se que a data e hora do registro referem-se ao momento em que foi aberto o processo na Defesa Civil municipal.

Neste momento, já se começou a avaliar a adequabilidade do agrupamento dos dados em intervalos de 3 ou de 6 horas. De fato, a análise da dinâmica temporal da quantidade de ocorrências registradas em intervalos de 1 hora evidenciou a existência de grande turbulência, enquanto o agrupamento dos dados em intervalos de 6 horas se mostrou mais suave e adequado à modelagem.

Já os dados de precipitação foram avaliados de acordo com a hora do dia, com o mês do ano e com o ano. Novamente, foi avaliado o potencial para agrupamento dos dados, principalmente em intervalos de 3 horas e de 6 horas.

A associação entre as ocorrências de deslizamentos de terra e os dados meteorológicos foi feita através da delimitação de regiões. Estas regiões são formadas a partir da associação de cada bairro ao pluviômetro mais próximo de seu centroide. A definição destas regiões é de extrema importância para a utilização do modelo de previsão de ocorrências de forma operacional, de forma que as previsões serão feitas para cada região.

Dessa forma, são definidas 39 regiões que contêm os 170 bairros de Salvador, de forma que cada bairro pertence a uma única região. Finalmente, foi calculada a correlação entre dados de precipitação e o quantitativo de deslizamentos de terra registrados, utilizando-se o coeficiente de correlação de Pearson (R), avaliando, assim, a variação deste indicador para acumulados de chuva em diferentes janelas temporais e dados agrupados em intervalos de 1, 3 e 6 horas.

5.2.2 ENGENHARIA DE DADOS

O modelo foi projetado para fazer a previsão de ocorrências com um horizonte de 72 horas, subdividido em 12 intervalos de 6 horas. Os dados estavam originalmente agrupados em intervalos de 3 horas, por esta ser a resolução de saída do modelo de pós-processamento do WRF, utilizado para a previsão de chuva. Para o modelo de previsão de deslizamentos de terra, todos os dados foram agrupados em intervalos de 6 horas. Durante esta agregação, o maior valor de chuva na resolução original (3 horas) foi incluído em uma nova coluna.

A precipitação que vinha sendo registrada em cada estação foi passada para o modelo com o cômputo de acumulados de chuva em diferentes janelas temporais: 6h, 72h, 120h e 240h.

De forma análoga, os deslizamentos de terra registrados anteriormente também foram utilizados, por meio de variáveis que indicassem a quantidade de ocorrências das últimas 6 horas, dos últimos 15 dias, 30 dias, 60 dias e 90 dias, além do total histórico de deslizamentos de terra já registrados naquela região.

Foram incluídas, também, colunas com informações do dia do ano, dia da semana e da hora do dia de cada ocorrência, em virtude da importância destas informações para a dinâmica das ocorrências registradas. Por serem cíclicas, estas variáveis foram transformadas, através das funções trigonométricas seno e cosseno (gerando duas variáveis cada).

Para incorporar a dimensão espacial, a latitude e a longitude do centroide de cada região também foram utilizadas como variáveis preditoras,

Todas as variáveis foram escalonadas para o intervalo 0 a 1.

Os seguintes 32 atributos foram utilizados como preditores:

- *Timestep* da previsão do WRF (1 atributo)
- Seno e cosseno do dia do ano (2 atributos)
- Seno e cosseno da hora do dia (2 atributos)
- Seno e cosseno do dia da semana (2 atributos)
- Latitude e longitude do centroide da região (2 atributos)
- Acumulados de chuva (5 atributos)
- Quantidade de deslizamentos de terra (6 atributos)
- Previsão de chuva para as próximas 72 horas, produzida pelo modelo de pós-processamento do WRF (12 atributos)

Para a rede convolucional, foi utilizada uma janela temporal de tamanho 19 (definida empiricamente), de forma que cada instância contém os 18 estados anteriores, além do atual.

Dessa forma, cada instância passada para o modelo tem 32 atributos, sendo que, para cada atributo, é passada uma série temporal, com o valor deste atributo em 19 intervalos subsequentes de tempo. No caso de variáveis que não têm evolução temporal, como as coordenadas da região para a qual é feita a previsão, o mesmo valor é repetido as 19 vezes. Com isso, a rede convolucional utilizada é capaz de captar e modelar a dinâmica temporal destas variáveis.

Após todo o pré-processamento e engenharia de dados, o conjunto final de dados utilizados para a modelagem foi composto 47128 registros únicos. Estes dados compreendem o período entre 19/07/2019 e 26/09/2020. A Figura 19 mostra um gráfico do período utilizado.

Figura 19 - Quantidade de deslizamentos de terra registrados em intervalos de 6 horas, ressaltando o período de dados efetivamente utilizado para a modelagem.



Fonte: Autor.

5.2.3 CONSTRUÇÃO DO MODELO DE INTELIGÊNCIA ARTIFICIAL

Para a avaliação do modelo, foi utilizada a técnica de validação cruzada (*cross-validation*), com 5 iterações. Cada uma utilizou a seguinte proporção aproximada do conjunto de dados:

- Treino: 72 %
- Validação: 13 %
- Teste: 15 %

Assim como explicado para o modelo de previsão de precipitação, para a modelagem de séries temporais, o cenário ideal envolveria que as subdivisões de treino, validação e teste fossem contínuas, de forma que, por exemplo, o conjunto de teste fosse composto por todos os dados dos últimos 3 meses da série temporal.

No entanto, o conjunto de dados efetivamente utilizado compreende um período muito curto, de pouco mais de um ano e existe uma variação clara da ocorrência de deslizamentos de terra ao longo do ano (como mostra a Figura 19). Se fosse utilizada essa estratégia, as instâncias do único período chuvoso do conjunto de dados (período que concentra as ocorrências dos deslizamentos de terra) não estariam representadas em todas as três subdivisões (treino, validação e teste), dificultando a modelagem.

Para contornar esta situação, foi adotada outra estratégia: optou-se por selecionar amostras aleatoriamente para cada subdivisão (treino, validação e teste), estratificando de acordo com a região, o dia da semana e a hora do dia, de forma que cada subdivisão contivesse a mesma proporção de amostras de desses atributos que o conjunto completo. Com isso, instâncias consecutivas em cada subdivisão (treino, validação e teste), não representam, necessariamente, intervalos subsequentes de tempo.

Vale ressaltar que, no momento da divisão para treino, validação e teste, cada instância já possuía uma janela temporal com os 19 intervalos subsequentes de tempo, de forma que a seleção aleatória das instâncias não

alterou a janela temporal já existente em cada instância. Ou seja, mesmo após a divisão para treino, validação e teste, cada instância continua contendo registros dos 18 intervalos de tempo anteriores, além do atual, totalizando uma janela temporal de 19 intervalos subsequentes de tempo.

O modelo de inteligência artificial foi construído utilizando-se o framework da biblioteca TensorFlow v 2.2.0.

Foi utilizada uma rede sequencial com 6 camadas do tipo Conv1D (cada uma com 200 filtros e kernel de tamanho 2) e uma camada do tipo Dense com 12 unidades. Foram incluídas camadas de normalização dos batches e utilizou-se Relu como função de ativação nas 6 camadas do tipo Conv1D e Softplus na camada de saída (com 12 unidades). Foi utilizado um batch de tamanho 150. O modelo foi treinado utilizando como métrica o Erro Quadrático Médio (MSE).

A Figura 20 mostra um resumo da rede utilizada. Tanto os hiperparâmetros como a quantidade de camadas e de unidades em cada camada foram ajustados empiricamente, mediante tentativa e erro.

Figura 20 - Rede utilizada para a previsão de ocorrências.

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 19, 200)	13000
batch_normalization (Batch Normalization)	(None, 19, 200)	800
activation (Activation)	(None, 19, 200)	0
conv1d_1 (Conv1D)	(None, 19, 200)	80200
batch_normalization_1 (Batch Normalization)	(None, 19, 200)	800
activation_1 (Activation)	(None, 19, 200)	0
conv1d_2 (Conv1D)	(None, 19, 200)	80200
batch_normalization_2 (Batch Normalization)	(None, 19, 200)	800
activation_2 (Activation)	(None, 19, 200)	0
conv1d_3 (Conv1D)	(None, 19, 200)	80200
batch_normalization_3 (Batch Normalization)	(None, 19, 200)	800
activation_3 (Activation)	(None, 19, 200)	0
conv1d_4 (Conv1D)	(None, 19, 200)	80200
batch_normalization_4 (Batch Normalization)	(None, 19, 200)	800
activation_4 (Activation)	(None, 19, 200)	0
conv1d_5 (Conv1D)	(None, 19, 200)	80200
batch_normalization_5 (Batch Normalization)	(None, 19, 200)	800
activation_5 (Activation)	(None, 19, 200)	0
flatten (Flatten)	(None, 3800)	0
dense (Dense)	(None, 12)	45612
Total params: 464,412		
Trainable params: 462,012		
Non-trainable params: 2,400		

Fonte: Autor.

6 RESULTADOS E DISCUSSÕES

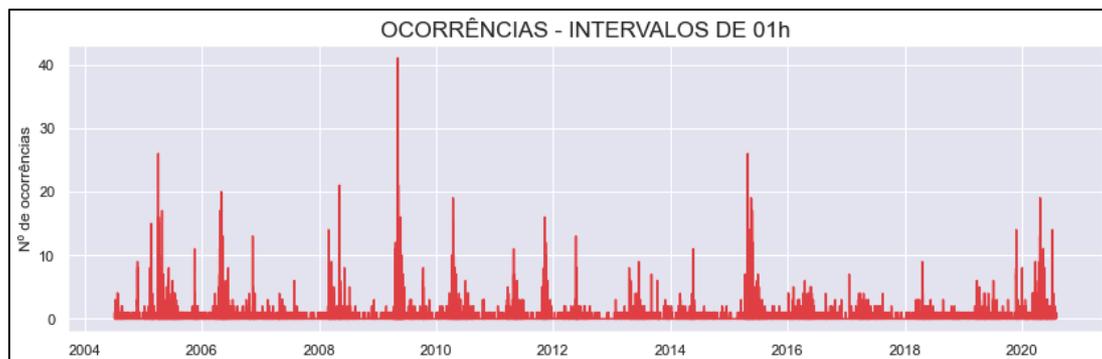
A seguir encontram-se os resultados obtidos durante os processos de análise dos dados de deslizamentos de terra e de precipitação, as métricas calculadas para a precipitação prevista pelo modelo WRF e os resultados dos dois modelos desenvolvidos neste trabalho, o de previsão de precipitação (pós-processamento do WRF) e o de previsão de deslizamentos de terra.

6.1 ANÁLISE DE DADOS

6.1.1 DESLIZAMENTOS DE TERRA

A Figura 21 mostra toda a série temporal com a quantidade de deslizamentos de terra registrados. Os dados estão agrupados em intervalos de 1 hora.

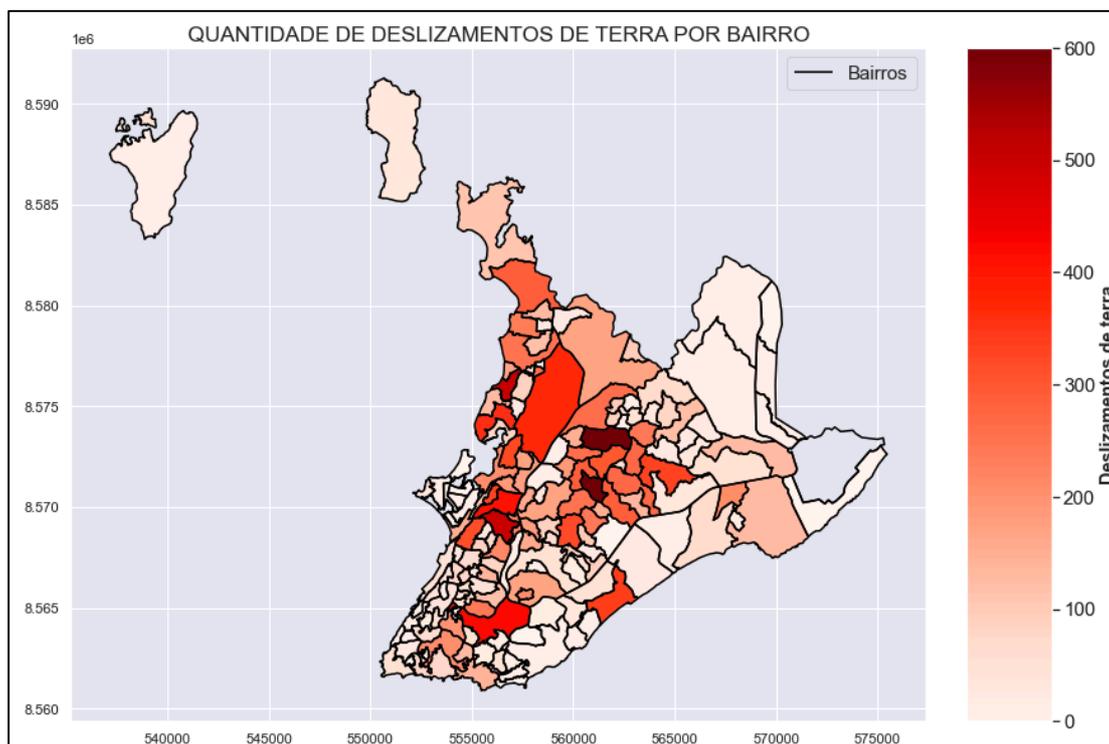
Figura 21 - Quantidade de deslizamentos de terra registrados (dados horários).



Fonte: Autor.

A Figura 22 mostra a quantidade de deslizamentos de terra por bairro, para toda a série analisada, de 2004 a setembro de 2020. Fica claro que algumas regiões concentram estas ocorrências.

Figura 22 - Quantidade de deslizamentos de terra por bairro. (Sistema de Referência: SIRGAS 2000 / UTM zone 24S).

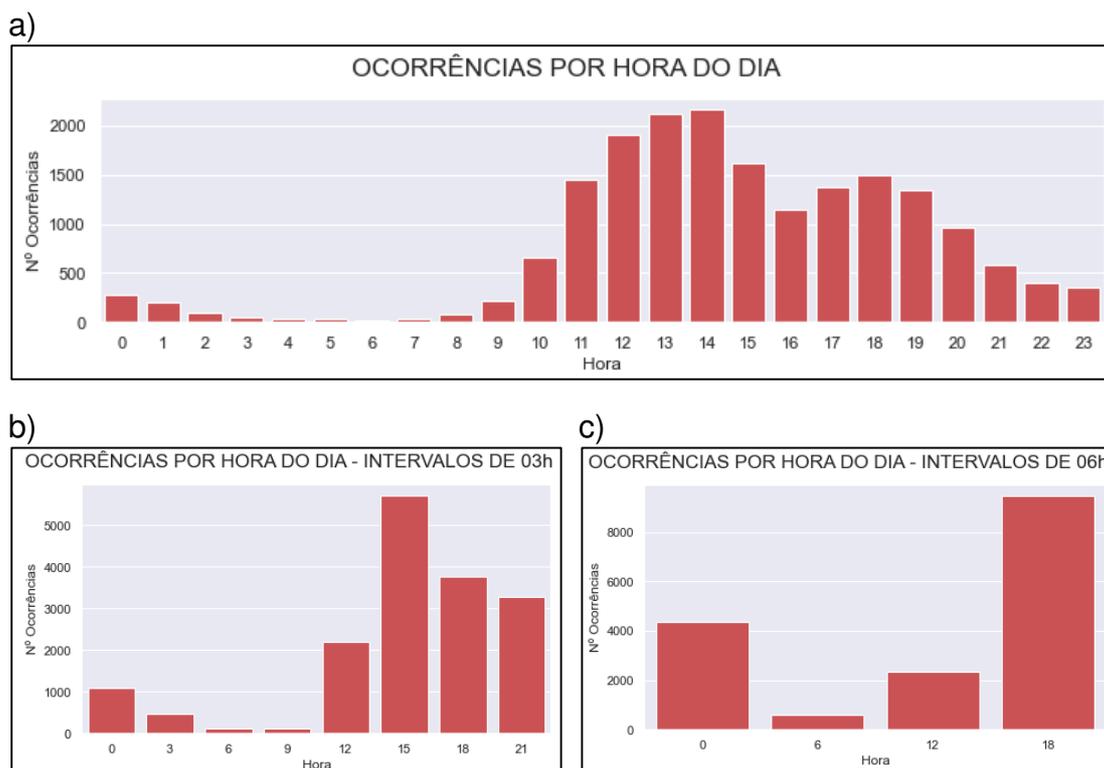


Fonte: Autor.

A Figura 23 mostra a distribuição das ocorrências de deslizamentos de terra por hora do dia. Os dados foram agrupados em intervalos de 1 hora (a), 3 horas (b) e 6 horas (c).

Fica evidente que existe um padrão de distribuição dos registros desses deslizamentos conforme a hora do dia. Novamente, cabe ressaltar que os dados são referentes ao momento em que houve abertura do processo na CODESAL, por meio de solicitação do interessado.

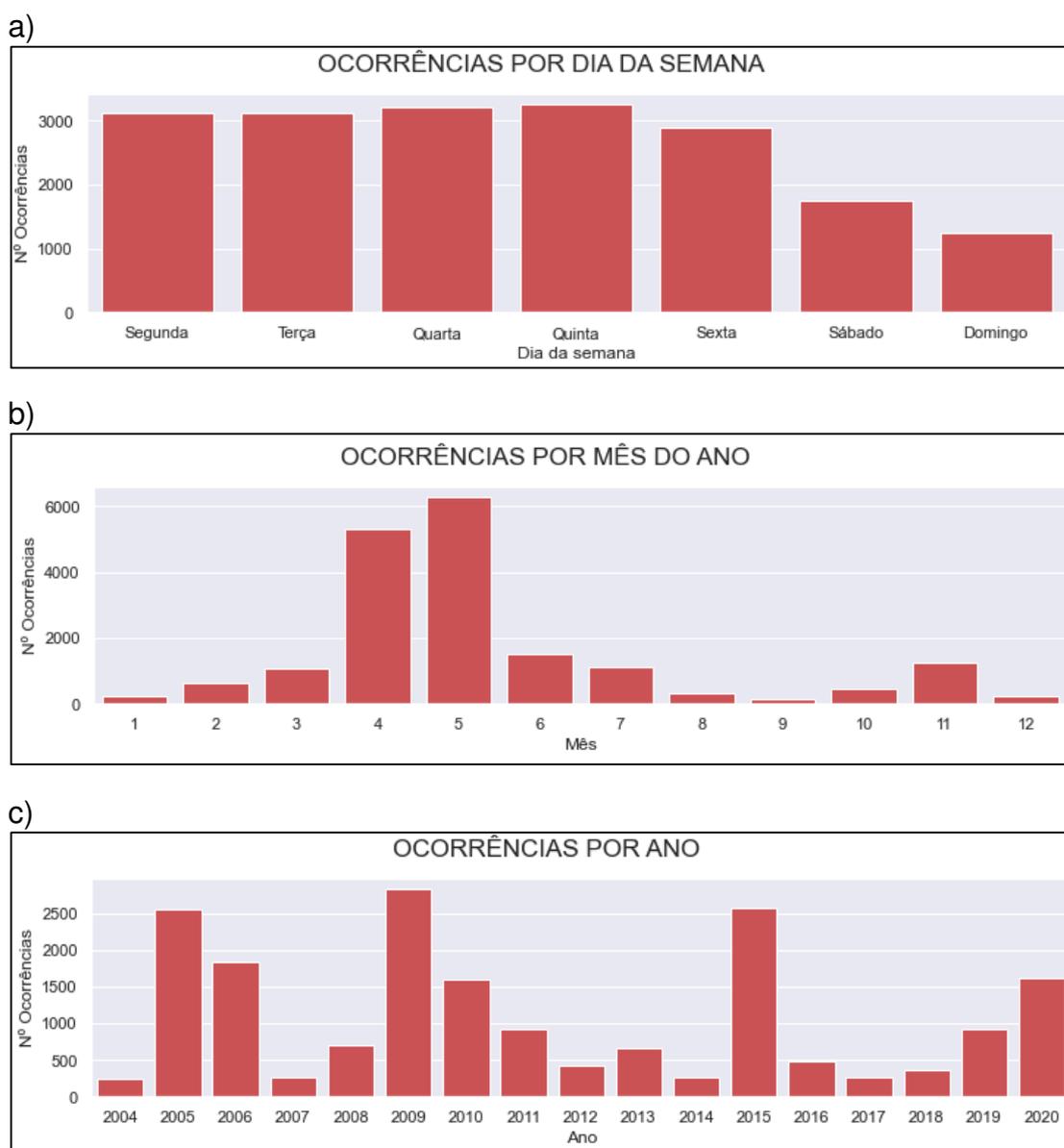
Figura 23 - Deslizamentos de terra registrados por hora do dia. a) Dados horários. b) Dados agrupados a cada 3 horas. c) Dados agrupados a cada 6 horas. Todos os dados estão em horário UTC.



Fonte: Autor.

A Figura 24 mostra a variação no registro dessas ocorrências conforme o dia da semana (a), o mês do ano (b) e ao longo dos anos (c). Observa-se que nos finais de semana há muito menos registros de deslizamentos. Além disso, existe uma grande concentração nos meses mais chuvosos, abril e maio.

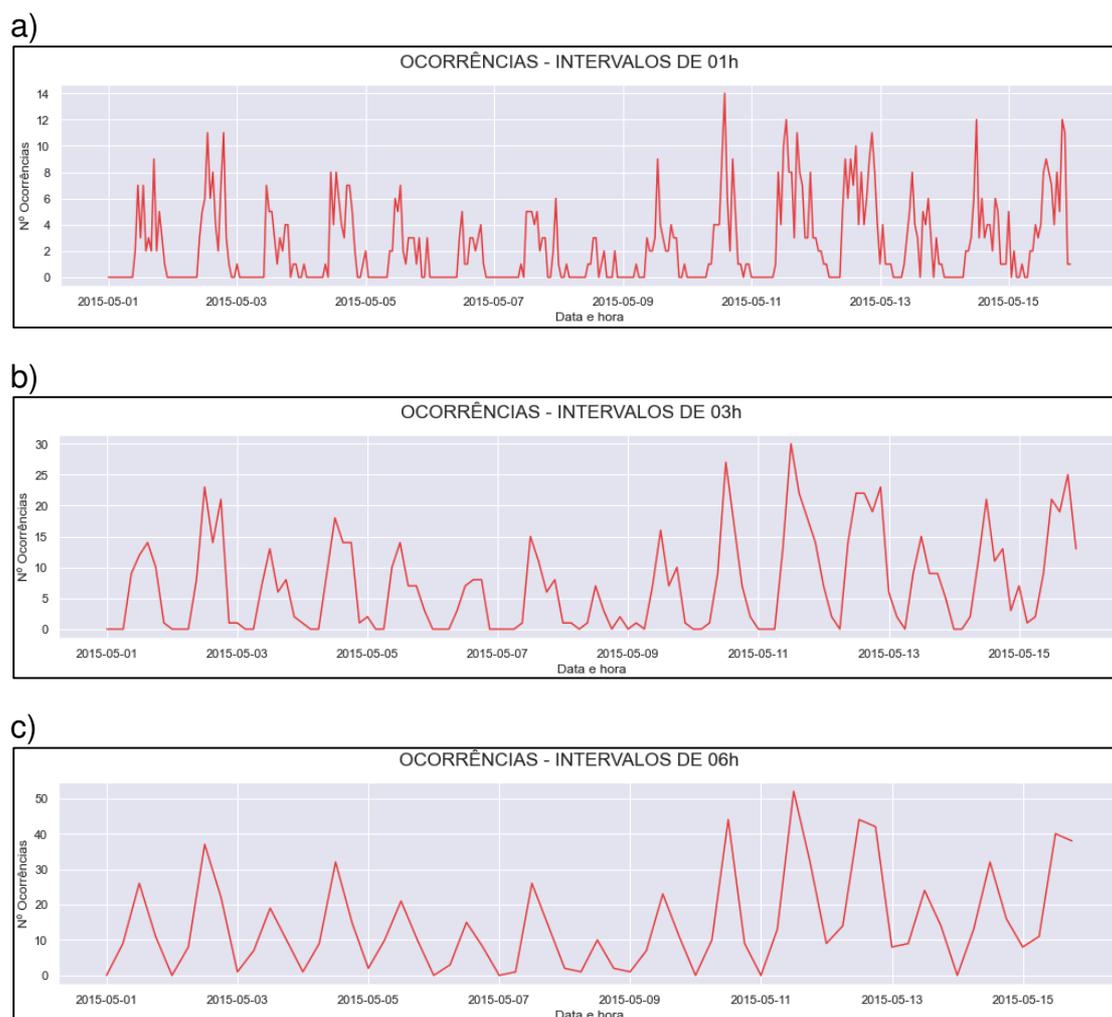
Figura 24 - Deslizamentos de terra registrados por: a) por dia da semana. b) Mês do ano. c) Ano. Todos os dados estão em horário UTC.



Fonte: Autor.

A Figura 25 apresenta, como exemplo, as ocorrências de deslizamentos de terra registradas entre os dias 01/05/2015 e 15/05/2015. Nota-se, na Figura (a), a existência de turbulência/oscilação na série temporal. Buscando-se atenuar estas oscilações, observou-se o efeito do agrupamento desses mesmos dados em intervalos de 3 horas (b) e 6 horas (c). No agrupamento de 3 horas, as oscilações ainda existem, embora sejam muito mais tênues. O agrupamento em intervalos de 6 horas ficou mais adequado à modelagem proposta, pela remoção destas oscilações.

Figura 25 - Deslizamentos de terra registrados entre os dias 01/05/2015 e 15/05/2015: a) Dados agrupados a cada 1 hora. b) Dados agrupados a cada 3 horas. c) Dados agrupados a cada 6 horas.

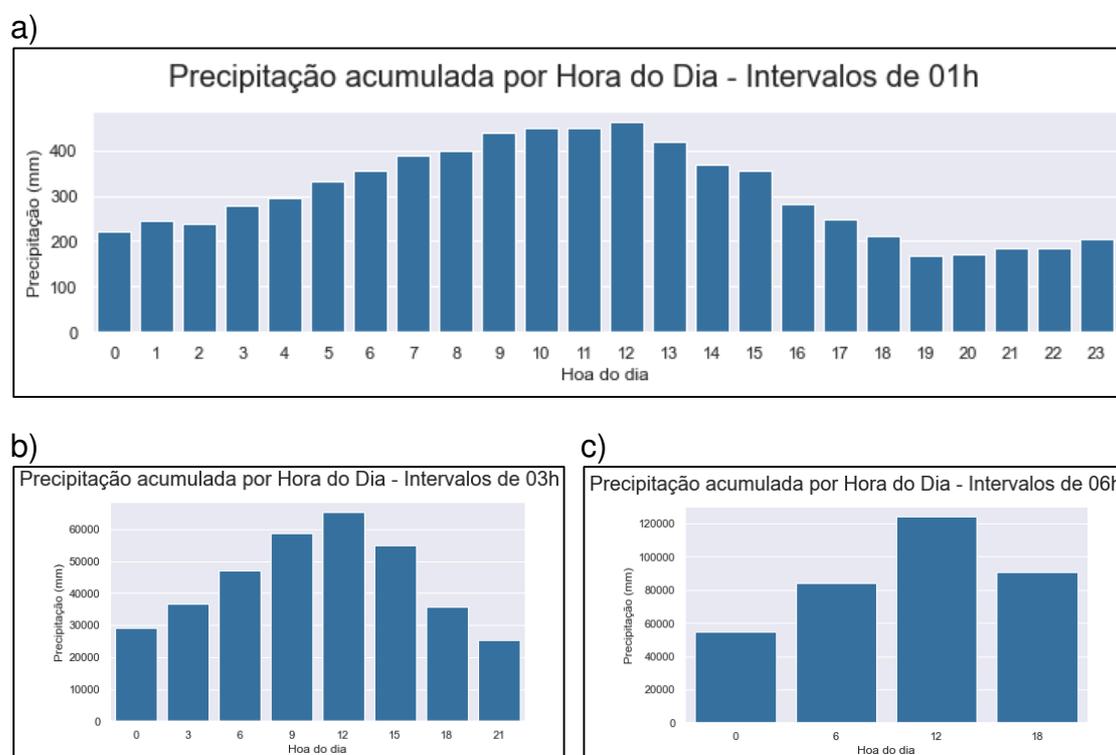


Fonte: Autor.

6.1.2 PRECIPITAÇÃO

A Figura 26 mostra a precipitação acumulada por hora do dia, utilizando a média por estações. Os dados foram agrupados em 1 hora (a), 3 horas (b) e 6 horas (c), com os intervalos fechados nos valores da direita.

Figura 26 - Deslizamentos de terra registrados entre os dias 01/05/2015 e 15/05/2015: a) Dados agrupados a cada 1 hora. b) Dados agrupados a cada 3 horas. c) Dados agrupados a cada 6 horas.

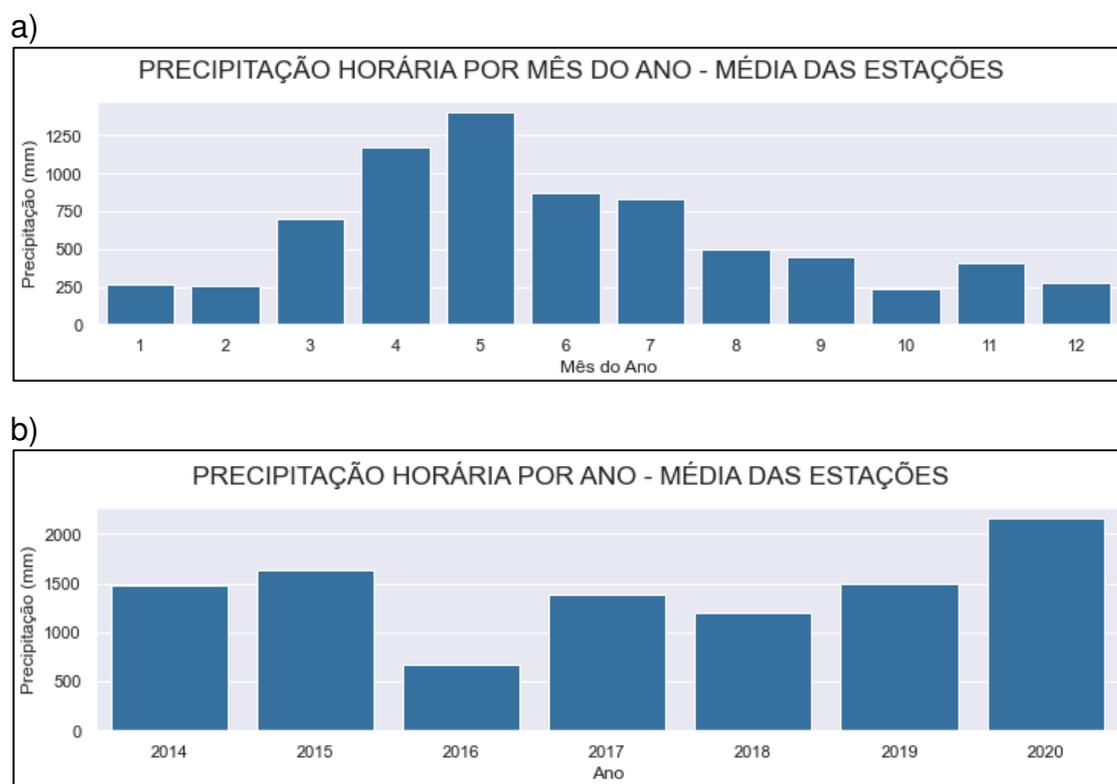


Fonte: Autor.

Nota-se que a distribuição das chuvas ao longo do dia possui um padrão bem definido, com maiores quantitativos entre as 5 e 15 horas UTC (2 e 12 horas no horário local). As agregações em 3 e 6 horas também evidenciam este comportamento e podem ser utilizadas para a modelagem.

A Figura 27 mostra a distribuição da precipitação ao longo dos meses (a) e dos anos (b). A distribuição ao longo dos meses está condizente com Normal Climatológica para a precipitação, como mostra o climograma, na Figura 7.

Figura 27 - Precipitação acumulada (média de todas as estações) por: a) Mês do ano. b) Ano. Todos os dados estão em horário UTC.

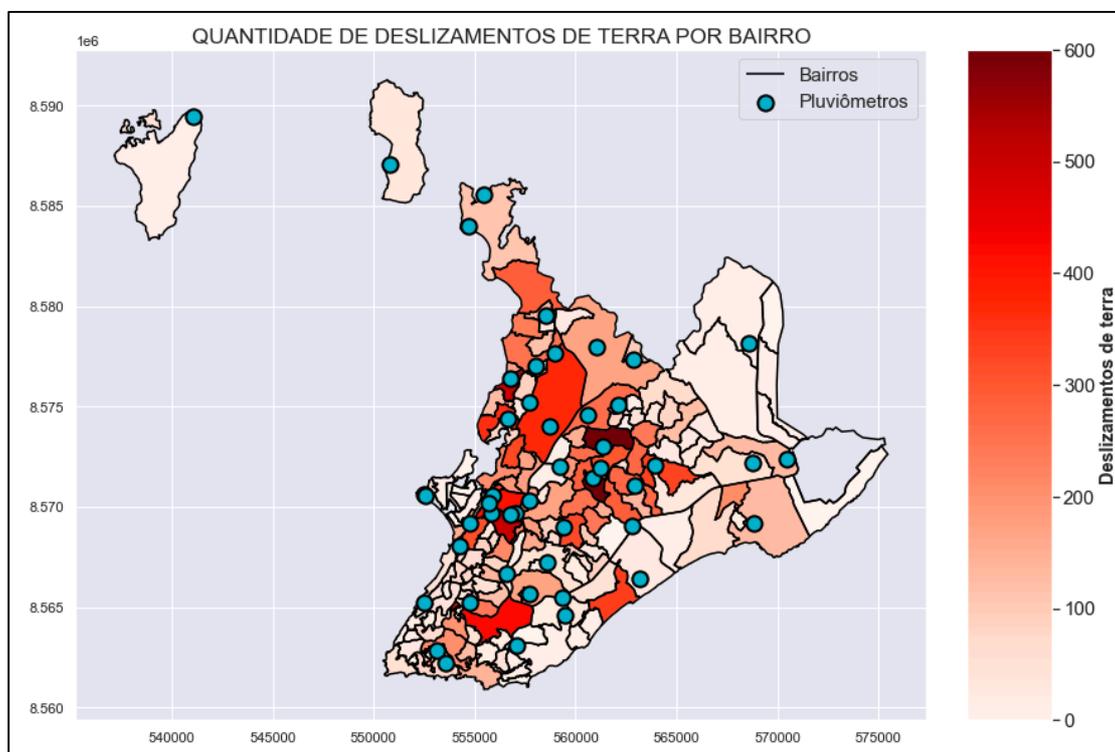


Fonte: Autor.

6.1.3 ASSOCIAÇÃO DOS DADOS

A Figura 28 mostra a quantidade de deslizamentos de terra por bairro, para toda a série analisada, de 2004 a setembro de 2020, destacando a localização das estações pluviométricas que existem atualmente.

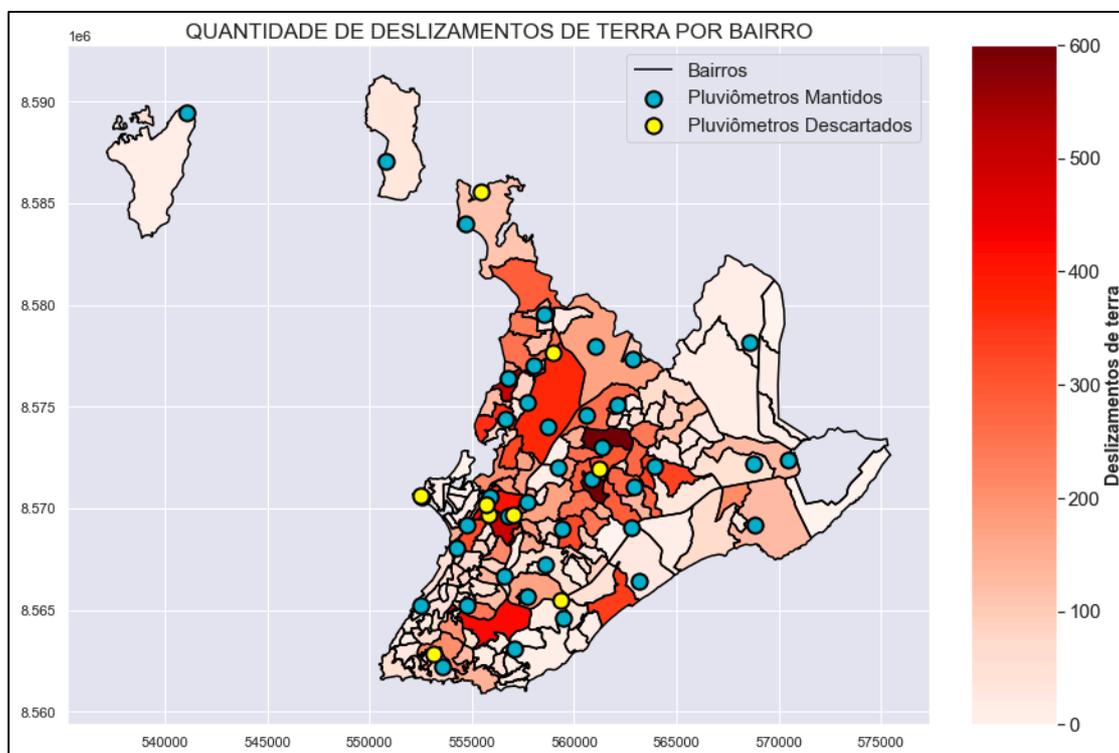
Figura 28 - Quantidade de deslizamentos de terra por bairro e localização dos pluviômetros analisados. (Sistema de Referência: SIRGAS 2000 / UTM zone 24S).



Fonte: Autor.

Nota-se que, em algumas regiões, existe uma certa sobreposição de estações pluviométricas, que estão muito próximas entre si. Assim, analisando-se, também, a qualidade da série de dados de cada estação, optou-se por desconsiderar algumas destas estações, como mostra a Figura 29.

Figura 29 - Quantidade de deslizamentos por bairro e localização dos pluviômetros mantidos e descartados. (Sistema de Referência: SIRGAS 2000 / UTM zone 24S).

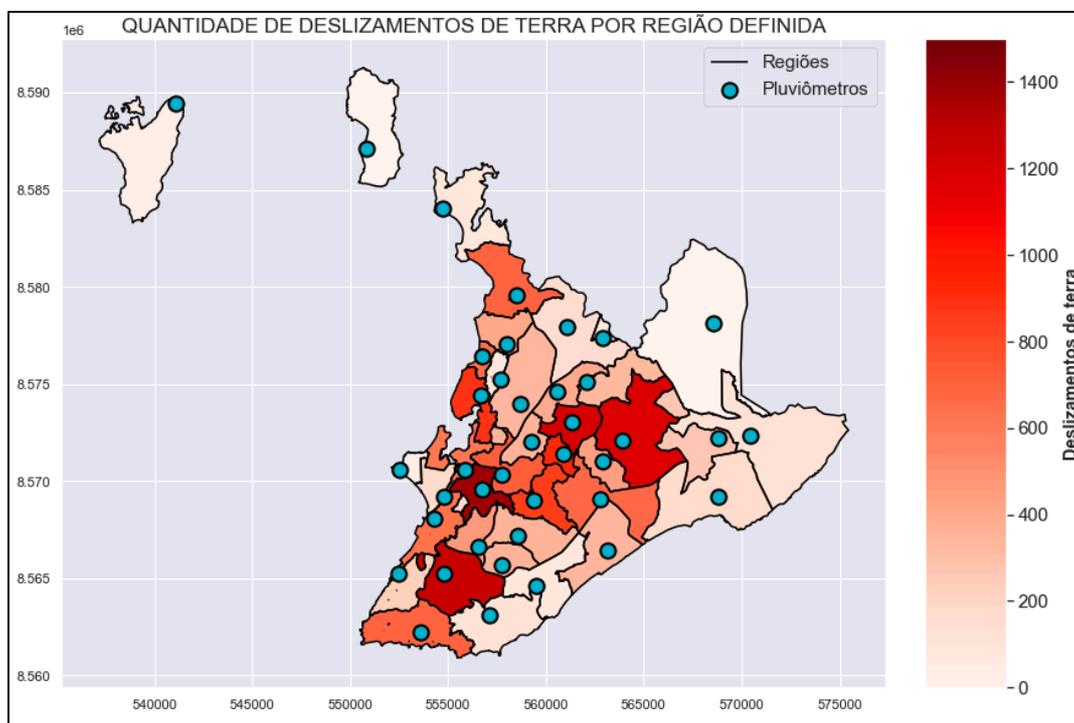


Fonte: Autor.

A Figura 30 mostra o mapa com as 39 regiões definidas (associando cada bairro à estação pluviométrica mais próxima de seu centroide), a quantidade de deslizamentos de terra em cada região e a localização das estações pluviométricas.

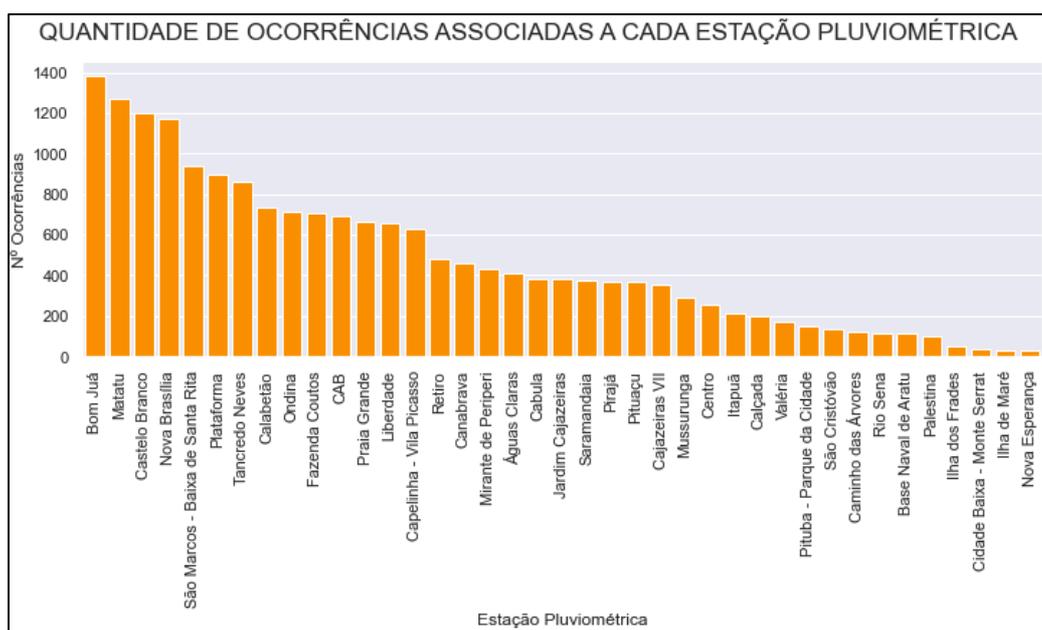
Já a Figura 31 mostra a quantidade de deslizamentos de terra registrados associados a cada estação pluviométrica, em função das regiões definidas

Figura 30 - Quantidade de deslizamentos de terra por região definida. (Sistema de Referência: SIRGAS 2000 / UTM zone 24S).



Fonte: Autor.

Figura 31 - Quantidade de deslizamentos de terra associados à região de cada estação pluviométrica.



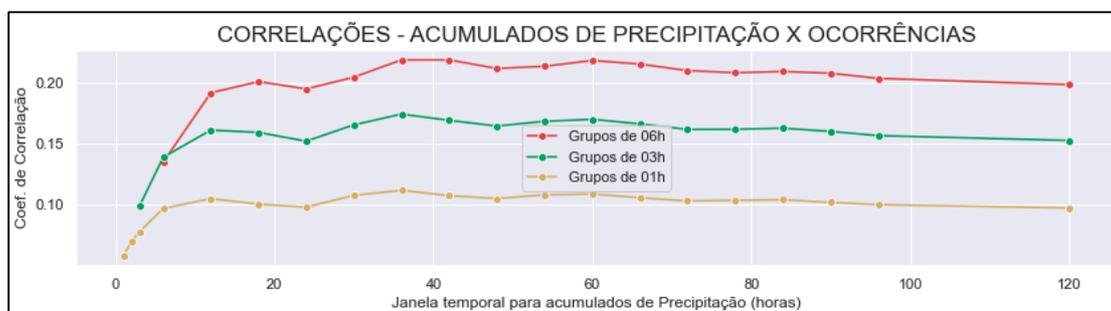
Fonte: Autor.

A Figura 32 mostra o coeficiente de correlação (R) de Pearson entre a precipitação acumulada em diferentes janelas temporais (1 hora a 120 horas) e o quantitativo de ocorrências. Os coeficientes de correlação foram calculados agrupando-se os dados em diferentes intervalos: 1 hora, 3 horas e 6 horas.

Nota-se, no gráfico, que, para os 3 agrupamentos utilizados (1h, 3h e 6h), R atinge os maiores valores para acumulados entre 40 e 60 horas. Ou seja, a precipitação acumulada em uma janela de 40 a 60 horas possui maior correlação com a quantidade de deslizamentos de terra.

Além disso, dados agrupados em 6 horas possuem maiores correlações entre precipitação e deslizamentos de terra do que os grupamentos em 1 e 3 horas.

Figura 32 - Coeficiente de correlação de Pearson entre a precipitação acumulada em diferentes janelas temporais (eixo x) e a quantidade de deslizamentos de terra. Dados agrupados em intervalos de 01h, 03h e 06h.



Fonte: Autor.

6.2 MÉTRICAS DOS RESULTADOS DE PRECIPITAÇÃO DO MODELO WRF

A seguir são mostradas as métricas calculadas para os resultados de precipitação do modelo WRF operacionalizado na CODESAL. Conforme exposto na seção de metodologia, foi analisada a qualidade da previsão de precipitação do modelo WRF em função dos seguintes fatores:

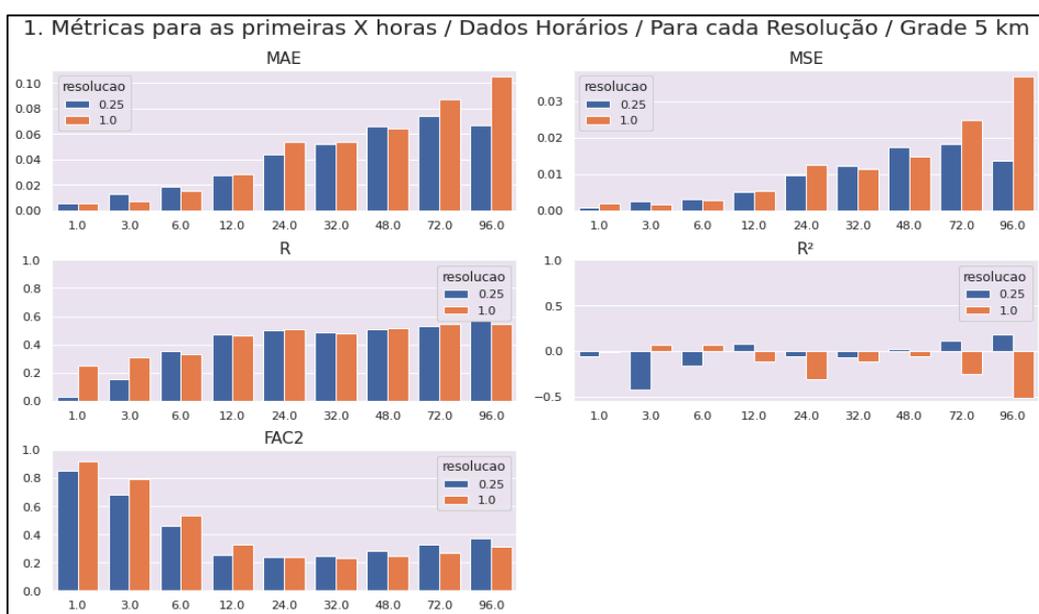
- Resolução espacial dos dados de entrada do modelo (1º ou 0,25º)
- Domínio (grade de 5 km ou de 20 km)
- Diferentes agrupamentos horários (dados agrupados para cada 3 horas, 6 horas, etc.)
- Hora do dia
- Mês do ano
- Estação do ano

6.2.1 ANÁLISE DA RESOLUÇÃO ESPACIAL E DOS DOMÍNIOS (GRADES)

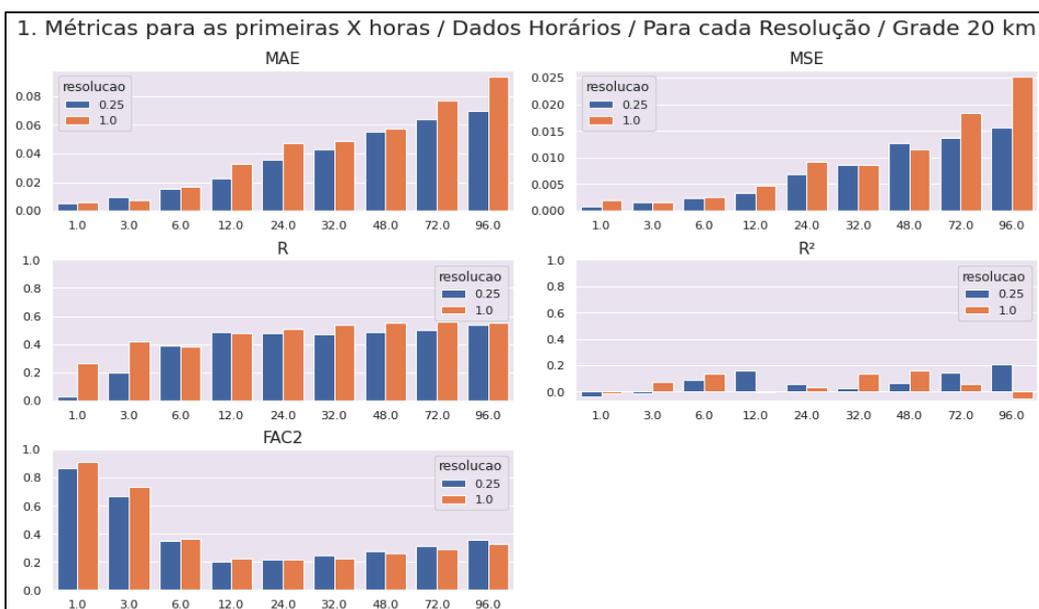
As Figuras 33 e 34 mostram as métricas por horizonte de previsão dos resultados do modelo WRF, para cada resolução espacial e tamanho de grade, respectivamente.

Figura 33 - Métricas para cada horizonte de previsão, por resolução temporal. a) Grade de 5 km. b) Grade de 20 km.

a)



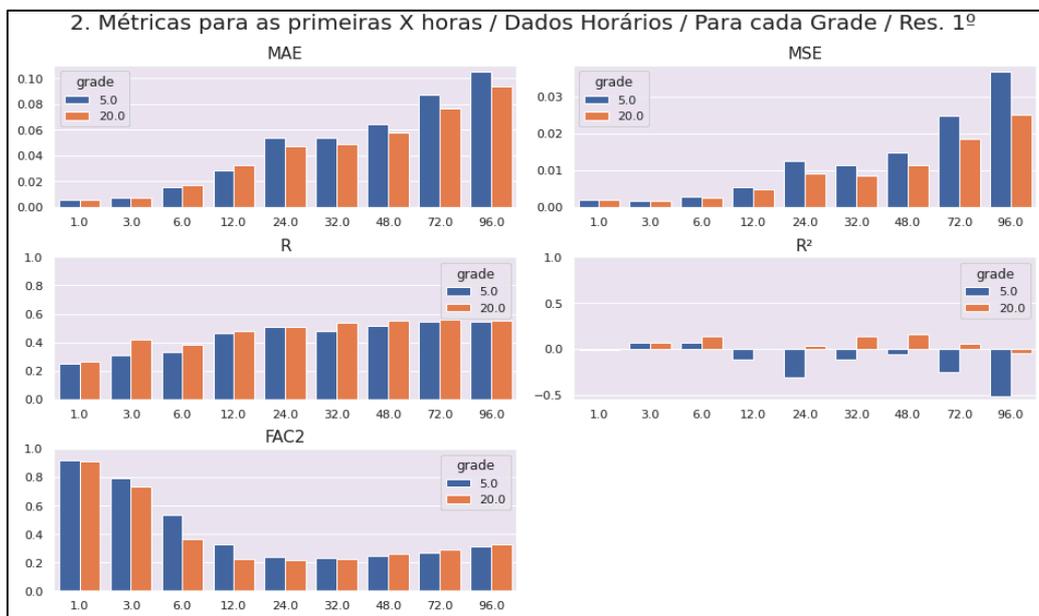
b)



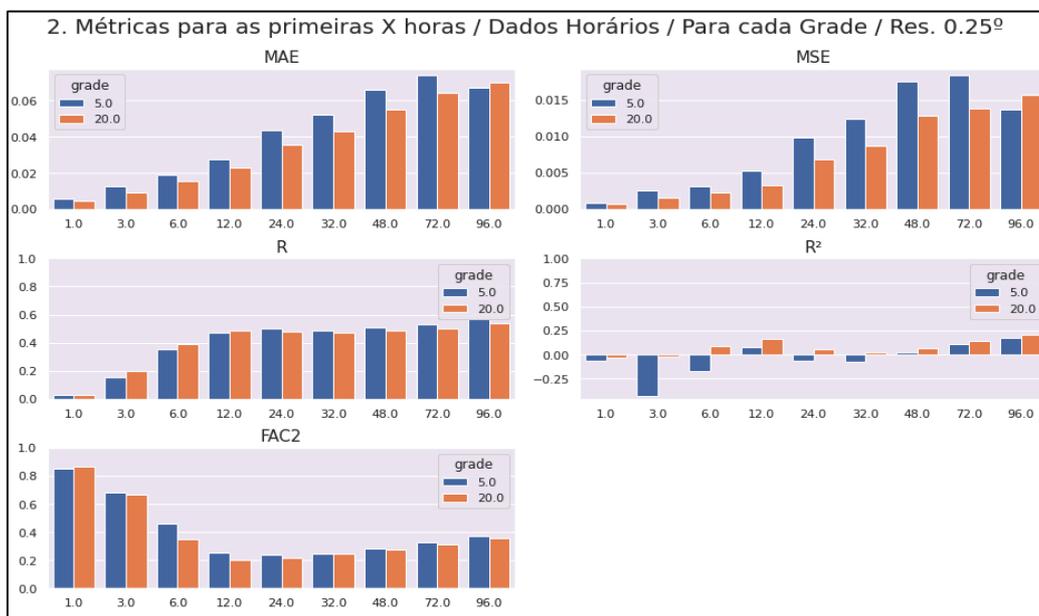
Fonte: Autor.

Figura 34 - Métricas para cada horizonte de previsão, por grade. a) Resolução de 1^o. b) Resolução de 0,25^o.

a)



b)



Fonte: Autor.

Nos gráficos acima, são contrastados, para cada resolução espacial e grade, as métricas em diferentes horizontes de previsão. São analisadas as primeiras: 01h, 03h, 06h, 12h, 24h, 32h, 48h, 72h, 96h de previsão do modelo.

A análise visual dos gráficos acima sugere que, para ambas as resoluções espaciais analisadas, os produtos da grade de 20 km possuem menores erros (MAE e MSE) que os da grade de 5 km em quase todos os horizontes analisados.

No caso da resolução de $0,25^\circ$, a grade de 20km chega a ter um MSE 46% menor que o da grade de 5km, para um horizonte de previsão de 12h.

Para a resolução de 1° , os resultados do coeficiente de correlação R são levemente superiores para a grade de 20km, embora não apresentem diferença significativa. Já para a resolução de $0,25^\circ$, a grade com melhor valor de R varia conforme o horizonte de previsão, chegando a apresentar diferença considerável, principalmente para um horizonte de 96h.

Já o coeficiente de determinação R^2 apresenta resultados controversos. Enquanto, para a resolução de $0,25^\circ$, os resultados, embora ainda longes de 1, chegam próximos a 0,25 para um horizonte de 96h na grade de 20km, para a resolução de 1° esta métrica chega a apresentar fortes valores negativos, indicando que o modelo teria performance inferior a uma previsão que sempre tivesse como resultado a média dos valores reais.

Os resultados de FAC2 não apresentam tanta diferença entre as resoluções ou grades. Como é de se esperar, FAC2 diminui com o aumento do horizonte de previsão, o que está de acordo com o aumento dos erros MAE e MSE.

Como as rodadas que utilizam dados de entrada com resolução espacial de $0,25^\circ$ apresentaram métricas claramente superiores para a qualidade da previsão de precipitação, optou-se por não utilizar as rodadas feitas com dados com resolução de 1° . As demais análises são feitas apenas para as rodadas com resolução espacial de $0,25^\circ$ de latitude e longitude.

6.2.2 ANÁLISE DOS AGRUPAMENTOS HORÁRIOS

As Figuras 35, 36 e 37 mostram as métricas para as 96 horas de previsão, das rodadas com que utilizam dados de entrada com resolução espacial de 0,25° de latitude e longitude, por tamanho de grade. Na Figura 35, os dados são originais, em intervalos de 1 hora. Na Figura 36, os dados estão agrupados em intervalos de 3 horas, e na Figura 37, em intervalos de 6 horas.

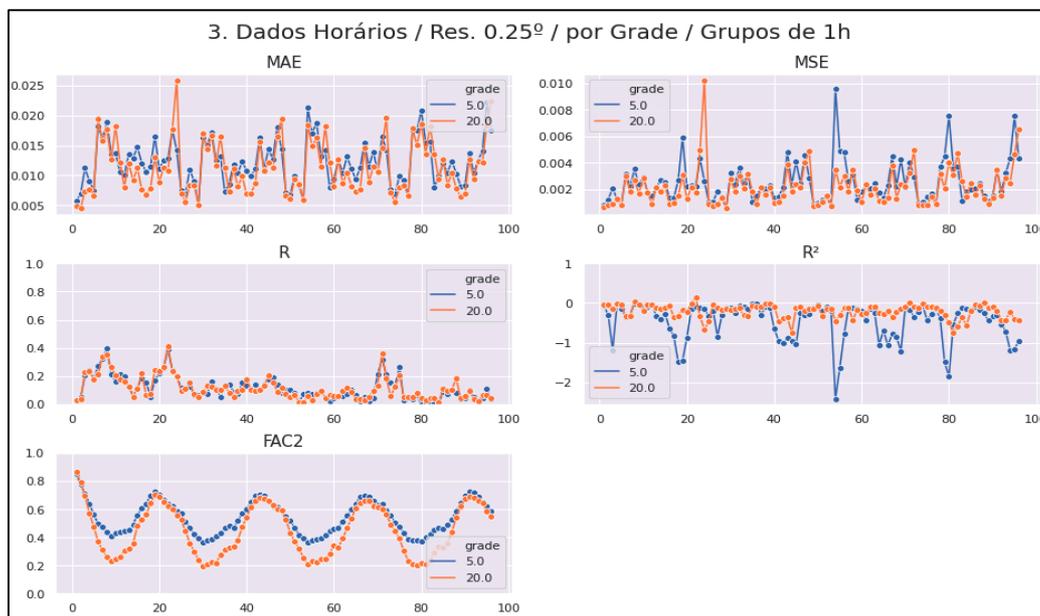
Nota-se, na Figura 35, uma turbulência nos valores das métricas (eixos verticais) ao longo das 96 horas de previsão (eixos horizontais), principalmente no caso do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE). Este comportamento indica fortes variações nos erros do modelo WRF para previsões horárias de precipitação, como é de se esperar, dada a incerteza intrínseca à modelagem deste fenômeno natural.

O coeficiente de correlação de Pearson (R) e o coeficiente de determinação (R^2) possuem uma leve tendência de queda ao longo dos 96 intervalos de previsão, o que indica que, de modo geral, as previsões das primeiras 24 horas tendem a ter uma melhor qualidade, embora esta variação seja pequena.

O Fator de 2 (FAC2) apresenta um comportamento senoidal ao longo dos 96 intervalos de previsão. Esta variação está presente nas métricas MAE e MSE, porém de forma muito mais obscura. Os picos desta métrica estão defasados entre si por cerca de 24 intervalos, assim como os vales. Este comportamento mostra a relação entre a hora do dia e a qualidade da previsão de precipitação.

Enquanto o MAE, o MSE e o R não evidenciam grandes diferenças entre cada uma das grades analisadas (com espaçamentos de 5 e de 20 km), R^2 sugere fortemente que as previsões da grade com resolução de 5 km têm qualidade mais baixa. De forma oposta, FAC2 apresenta resultados um pouco superiores para as previsões desta mesma grade.

Figura 35 - Métricas para cada uma das 96 horas previstas, por grade. Dados horários.



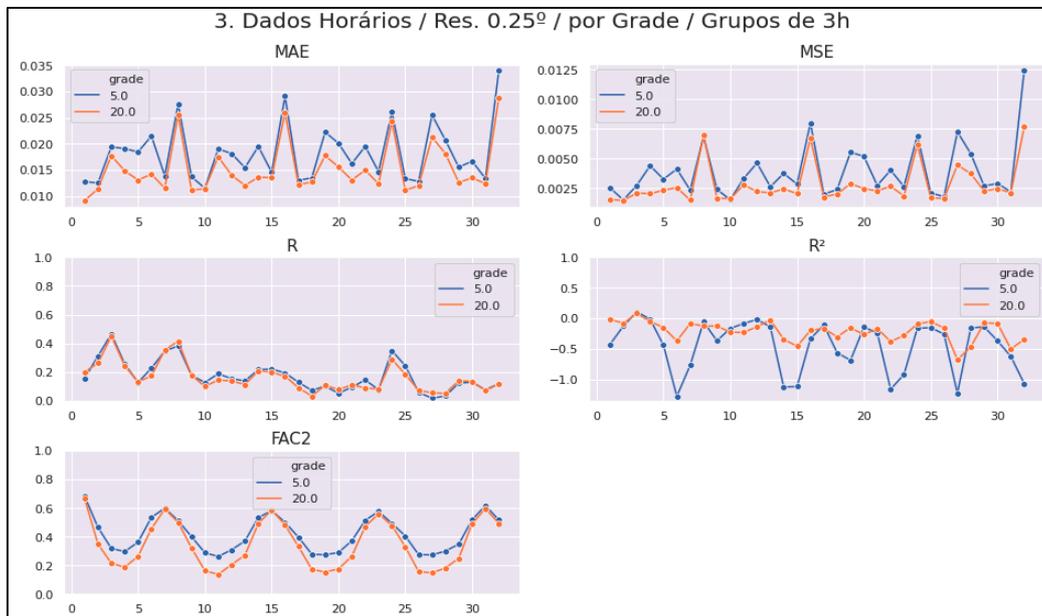
Fonte: Autor.

Considerando as métricas para os dados agrupados em intervalos de 3 horas (Figura 36), nota-se que boa parte da turbulência desaparece, o que facilita algumas análises.

MAE, MSE e R^2 passam a evidenciar a melhor qualidade das previsões feitas com a grade com espaçamento de 20 km, característica que não aparece em R. FAC2 não apresenta alterações substanciais com este agrupamento.

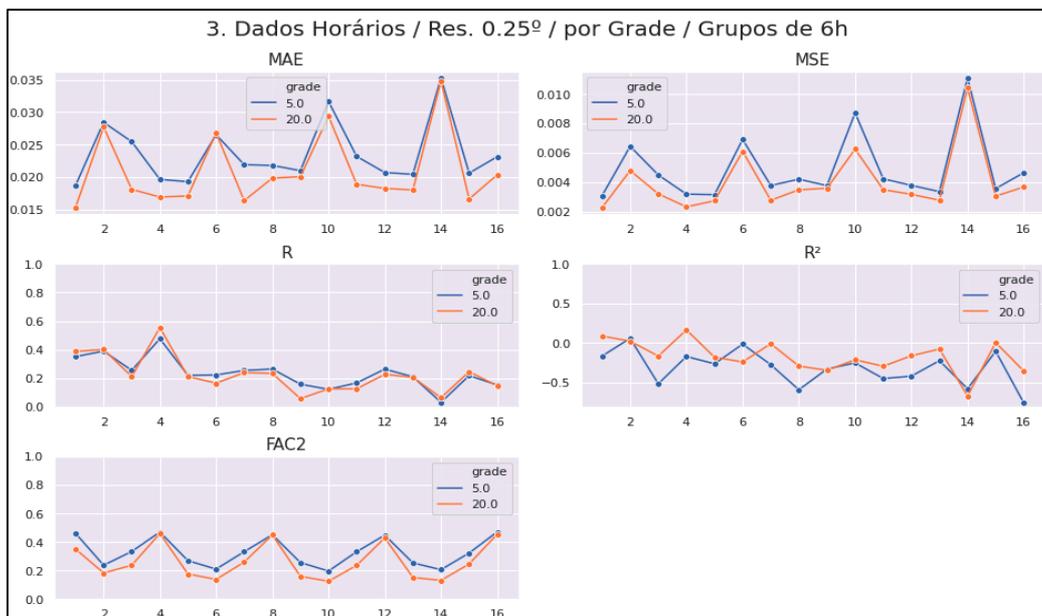
É importante notar que nenhuma das 5 métricas apresentou variações consideráveis na amplitude de seus valores, comparando-se os dados horários e o agrupamento em intervalos de 3 horas. Ou seja, o agrupamento da precipitação prevista em intervalos de 3 horas produz resultados tão bons quanto os dados horários, porém muito mais estáveis, sem a grande turbulência nos valores das métricas.

Figura 36 - Métricas para cada uma das 96 horas previstas, por grade. Dados em intervalos de 3 horas.



Fonte: Autor.

Figura 37 - Métricas para cada uma das 96 horas previstas, por grade. Dados em intervalos de 6 horas.



Fonte: Autor.

As métricas apresentadas na Figura 37, com os dados agrupados em intervalos de 6 horas, não possuem grande divergência das calculadas para o agrupamento de 3 horas, de forma que as mesmas características gerais estão presentes. Entretanto, nota-se uma redução em FAC2, que possui uma menor amplitude, mas mantém aproximadamente a mesma faixa de valores mínimos. Ou seja, para agrupamentos de 6 horas, uma proporção consideravelmente menor dos dados previstos está a um fator de 2 dos valores reais (entre metade e o dobro).

Em face do exposto, o agrupamento da precipitação prevista pelo modelo WRF em intervalos de 3 horas parece muito mais adequado do que em intervalos de 1 ou 6 horas, por produzir resultados um pouco melhores e razoavelmente estáveis.

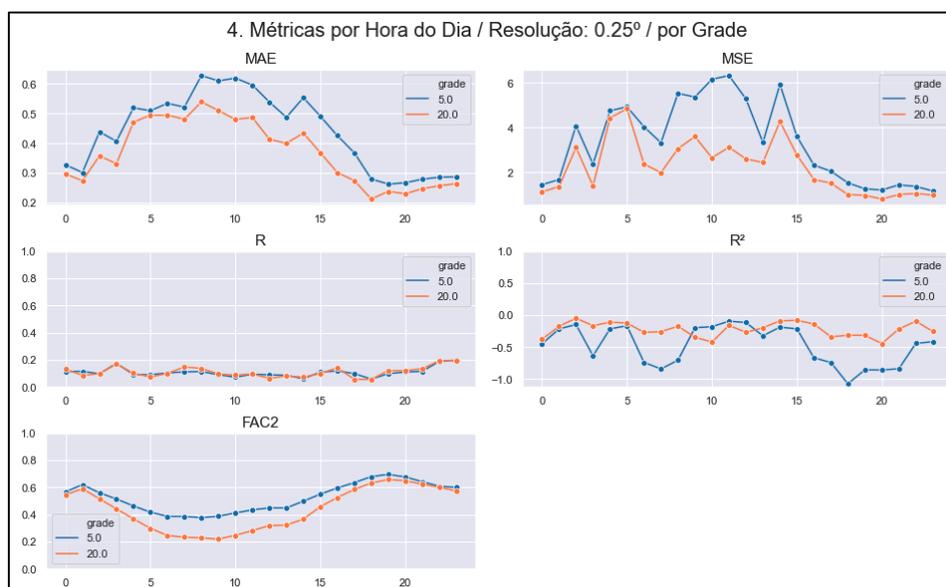
Chama a atenção, ainda, que, independentemente do agrupamento dos dados, R^2 apresenta valores negativos. Este dado preocupante pode decorrer da metodologia adotada para o cálculo das métricas e avaliação do modelo. Utiliza-se aqui uma resolução muito maior que a do modelo WRF estudado. Enquanto os dados de saída do WRF possuem resoluções espaciais de 5 ou de 20 km, as métricas são calculadas com valores medidos em espaçamentos muito menores, de forma que um mesmo valor de precipitação prevista pelo WRF pode ser contrastado com medições de vários pluviômetros inseridos numa mesma célula daquela grade.

Enquanto esta metodologia de avaliação pode não ser adequada para fins de otimização da parametrização do modelo WRF, é apropriada para o contexto aqui abordado, já que se busca uma modelagem espacializada e com alta resolução, dada a dinâmica espacial dos deslizamentos de terra em ambiente urbano e, particularmente, em Salvador;

6.2.3 ANÁLISE DA HORA DO DIA

A Figura 38 mostra as métricas por hora do dia, para as rodadas com dados de entrada com resolução espacial de $0,25^\circ$ de latitude e longitude. Para este cálculo, os dados foram agrupados independentemente do horizonte de previsão.

Figura 38 - Métricas por hora do dia, para cada grade.



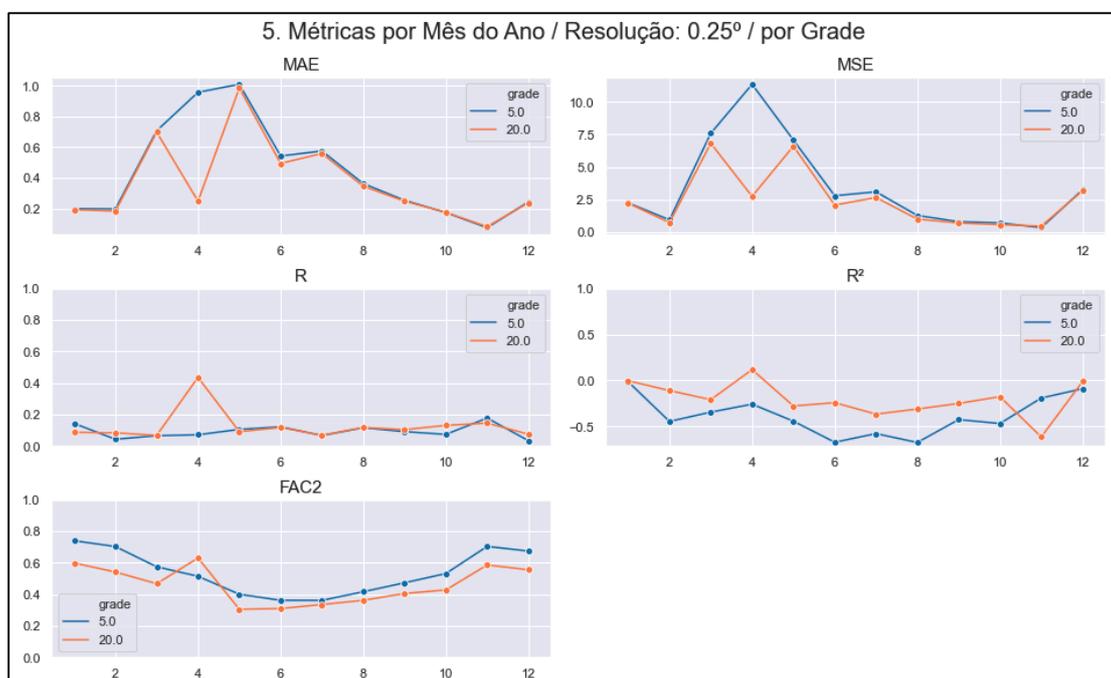
Fonte: Autor.

Na Figura 38, os gráficos de MAE, MSE e FAC2 mostram a grande variação dos erros conforme a hora do dia. Estes gráficos apresentam comportamento extremamente similar à variação da precipitação conforme a hora do dia (Figura 26). Isto indica que os erros tendem a ser maiores para valores de precipitação reais mais elevados. Já os gráficos de R e R^2 não apresentam um padrão claro de variação de acordo com a hora do dia. Isto sugere que a qualidade da previsão não varia tanto conforme a hora do dia, principalmente para a grade com resolução de 20 km. A análise conjunta deste comportamento (variação clara de MAE, MSE e FAC2, com R e R^2 razoavelmente estáveis) pode indicar que o modelo WRF consegue dar sinais de chuva, porém tem dificuldade em acertar a amplitude.

6.2.4 ANÁLISE DO MÊS

A Figura 39 mostra as métricas por mês do ano, para as rodadas com dados de entrada com resolução espacial de $0,25^\circ$ de latitude e longitude. Para este cálculo, os dados foram agrupados independentemente do horizonte de previsão

Figura 39 - Métricas para cada uma das 96 horas previstas, por grade. Dados em intervalos de 6 horas.



Fonte: Autor.

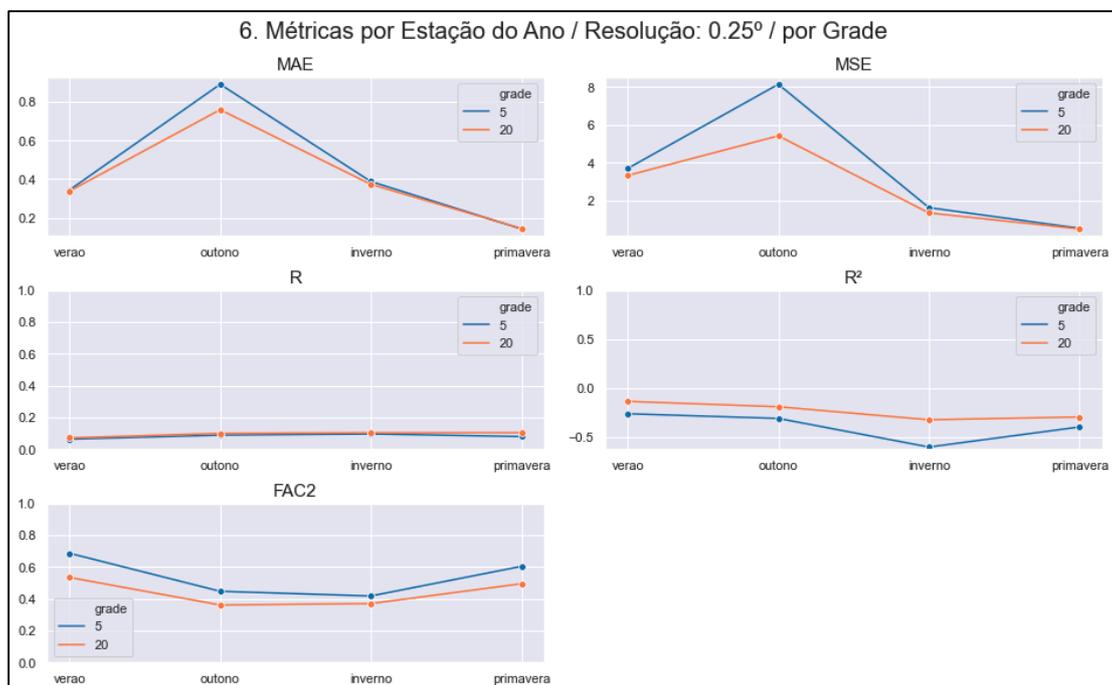
Na Figura, é possível notar que as métricas analisadas têm uma variação ao longo dos meses do ano muito parecida com a variação ao longo das horas do dia (Figura 38). De forma análoga à hora do dia, a amplitude do erro tende a ser mais elevada nos meses mais chuvosos, enquanto R e R² não têm um comportamento tão claro. No entanto, todas as 5 métricas para o mês 4 (abril) destoam das demais, no domínio (grade) com espaçamento de 20 km, quebrando este padrão sugerido.

6.2.5 ANÁLISE DA ESTAÇÃO DO ANO

Como poderia se inferir através da análise das métricas por mês do ano, a Figura 40 mostra as métricas por estação do ano. Embora o comportamento da performance do modelo WRF por estação do ano possa ser inferido através da análise da performance por mês do ano, esta agregação facilita a compreensão e comunicação por parte da equipe que interpreta os resultados do modelo.

Analisando-se a Figura 40, fica claro que o modelo WRF costuma apresentar maiores erros no Outono (que tende a ser mais chuvoso), com erros intermediários no verão e inverno e erros mais baixos na primavera.

Figura 40 - Métricas para cada uma das 96 horas previstas, por grade. Dados em intervalos de 6 horas.



Fonte: Autor.

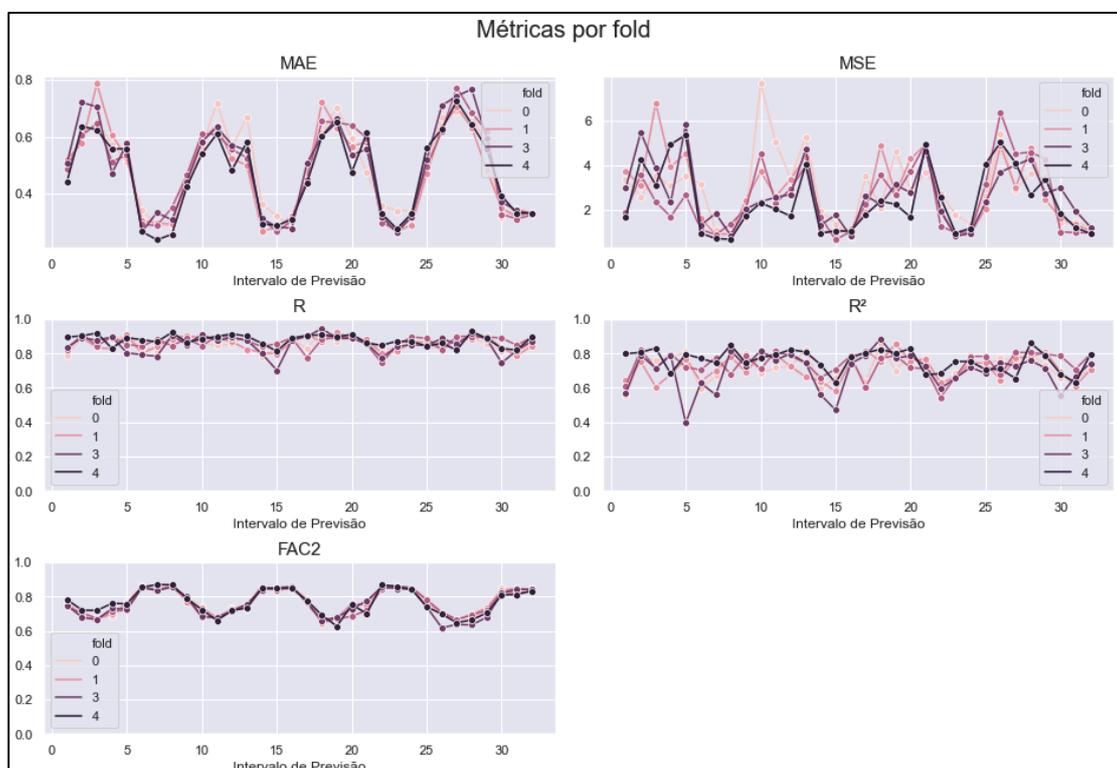
6.3 MODELO DE PREVISÃO DE PRECIPITAÇÃO (PÓS-PROCESSAMENTO DO WRF)

Os resultados do WRF contêm, originalmente, dados horários de previsão com um horizonte de 96 horas. Os dados de precipitação foram agregados para intervalos de 3 horas, configurando 32 intervalos de previsão.

O modelo de previsão de precipitação desenvolvido tem como saída 32 intervalos de previsão, cada um representando a precipitação em um intervalo de 3 horas, totalizando um horizonte de previsão de 72 horas.

A Figura 41 contém os gráficos de cada métrica por divisão (*fold*) e intervalo de previsão (*timestep*). A Tabela 2 mostra as métricas médias para cada uma das divisões (*folds*) utilizadas para a avaliação do modelo. Todas as métricas foram calculadas utilizando o conjunto de dados separado para testes.

Figura 41 - Métricas do modelo por fold e intervalo de previsão.



Fonte: Autor.

Tabela 2 - Métricas do modelo por fold e intervalo de previsão.

Fold	MAE	MSE	R	R²	FAC2
1	0,4897	2,8971	0,8591	0,7202	0,7689
2	0,4759	2,7664	0,8573	0,7169	0,7687
3	0,4813	2,5169	0,8656	0,7357	0,7610
4	0,4901	2,6453	0,8538	0,7000	0,7574
5	0,4687	2,4089	0,8775	0,7596	0,7665
Média	0,4811	2,6469	0,8627	0,7265	0,7645
Desvio-padrão	0,0082	0,1734	0,0084	0,0201	0,0045

Fonte: Autor.

A Tabela 3 mostra as métricas calculadas para os resultados originais do WRF e para os resultados pós-processados pela rede neural implementada, ambas para a precipitação. Os resultados apresentados são as médias dos 5 *folds*.

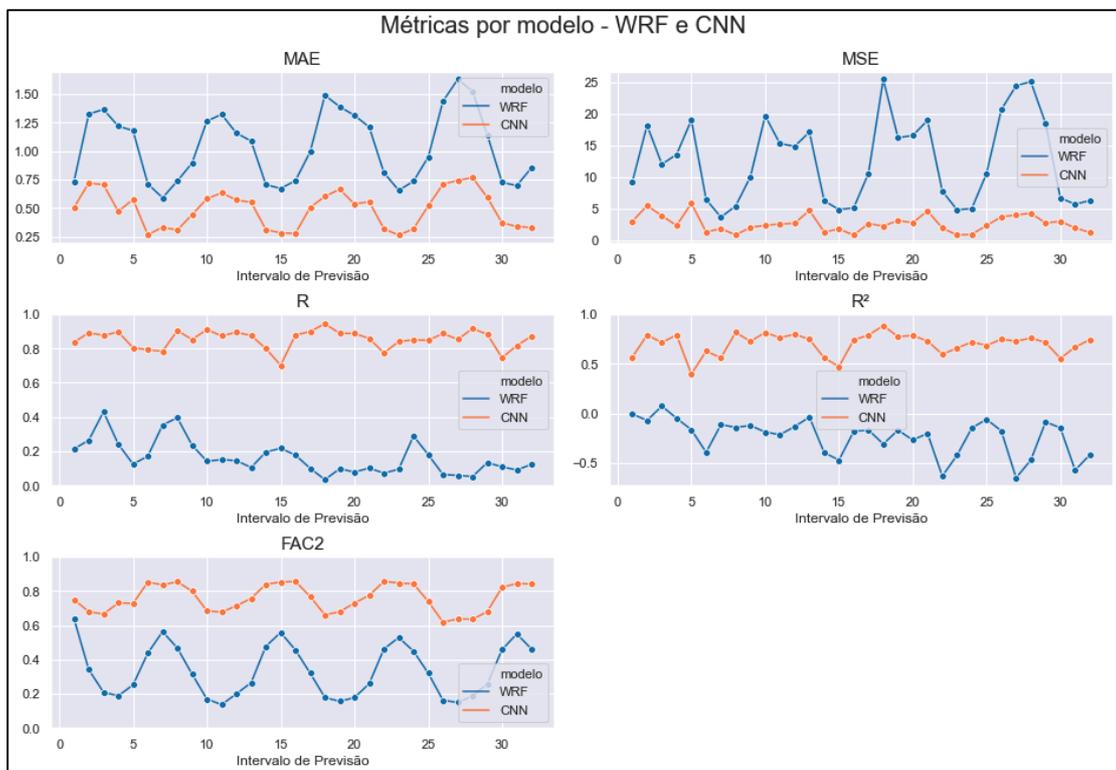
Tabela 3 - Métricas médias da previsão do WRF e do modelo de pós-processamento do WRF.

Modelo	MAE	MSE	R	R²	FAC2
WRF	1,0383	12,5894	0,1637	-0,2355	0,3357
Modelo de Pós-processamento do WRF	0,4811	2,6469	0,8627	0,7265	0,7645

Fonte: Autor.

A Figura 42 mostra gráficos com as métricas de cada modelo por intervalo de previsão. Fica evidente que o modelo de pós-processamento do WRF foi capaz de produzir resultados muito superiores aos resultados originais do WRF, para a precipitação. Não só cada uma das métricas apresentou resultados melhores, como também a oscilação ao longo dos intervalos de previsão diminuiu.

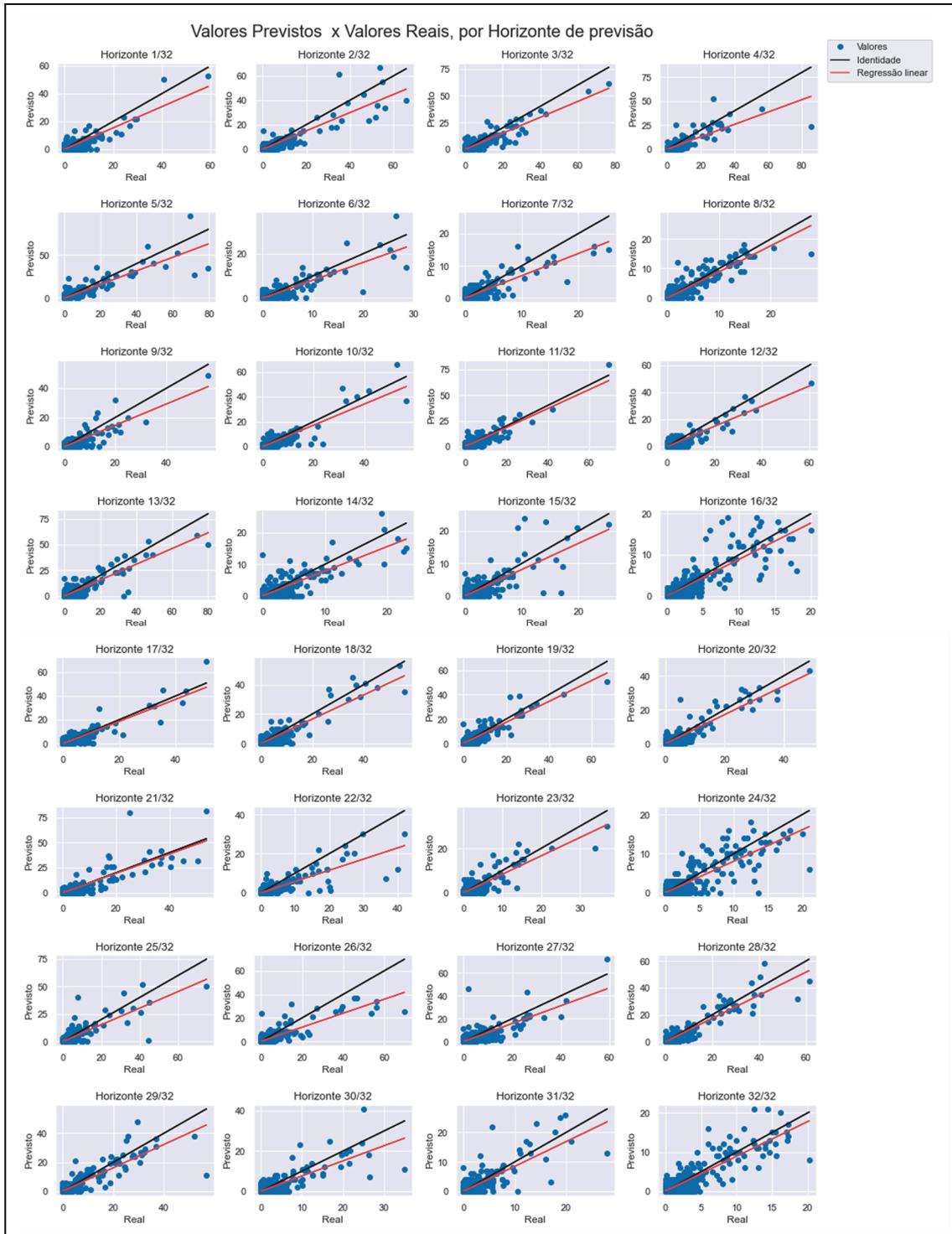
Figura 42 - Métricas de cada modelo para cada intervalo de previsão.



Fonte: Autor.

A Figura 43 contém gráficos de dispersão para cada um dos 32 intervalos de previsão, contrastando os valores previstos com os valores reais. Estes gráficos foram construídos utilizando-se os dados de teste, não vistos pelo modelo durante o treinamento. As linhas de regressão linear mostram que os valores previstos são razoavelmente próximos dos valores reais. No entanto, este modelo tem tendência a subestimar os valores de precipitação.

Figura 43 - Modelo de previsão de precipitação: valores previstos e valores reais, por intervalo de previsão, para os dados de teste.



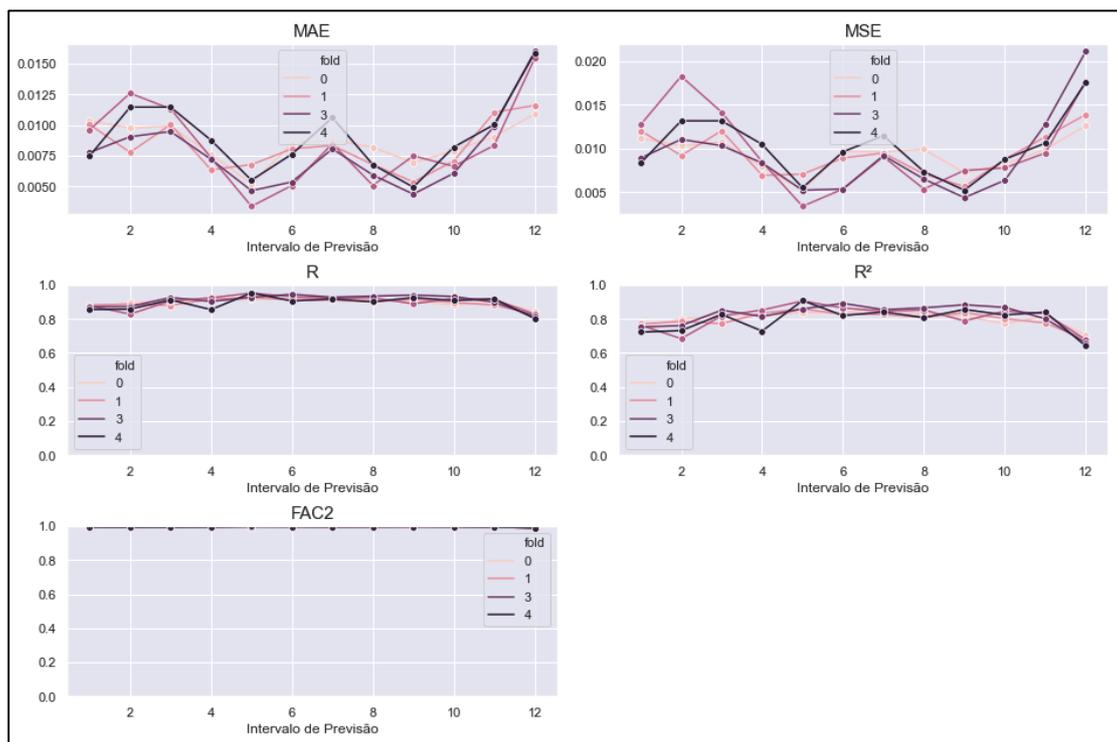
Fonte: Autor.

6.4 MODELO DE PREVISÃO DE DESLIZAMENTOS DE TERRA

O modelo de previsão de ocorrências foi projetado para ter como saída o quantitativo de deslizamentos de terra, para cada uma das 39 regiões definidas, com um horizonte de previsão de 72 horas, subdividido em 12 intervalos de 6 horas.

A Figura 44 contém os gráficos de cada métrica por *fold* e intervalo de previsão (*timestep*). A Tabela 4 mostra as métricas médias para cada uma das divisões (*folds*) utilizadas para a avaliação do modelo. Todas as métricas foram calculadas utilizando o conjunto de dados separado para testes

Figura 44 - Métricas do modelo por *fold* e intervalo de previsão.



Fonte: Autor.

A análise dos gráficos na Figura 44, que mostra as variações das métricas de cada *fold* ao longo dos 12 intervalos de previsão, evidencia a grande estabilidade entre os 5 *folds*. Nota-se que os gráficos de MAE e MSE apresentam oscilações ao longo dos intervalos de previsão. De forma não intuitiva, os primeiros intervalos de previsão apresentaram erros maiores, ou no mesmo patamar, dos últimos intervalos de previsão. Este comportamento pode ser visto nos gráficos de MAE, MSE e R². Apesar de diversas tentativas, não se conseguiu corrigir este padrão.

Os gráficos de R, R² e FAC2 apresentam bastante equilíbrio, indicando que a qualidade da previsão não decai fortemente ao longo dos 12 intervalos.

Tabela 4 - Métricas do modelo por *fold* e intervalo de previsão.

Fold	MAE	MSE	R	R²	FAC2
1	0,0086	0,0094	0,8983	0,8053	0,9943
2	0,0083	0,0094	0,8943	0,7981	0,9944
3	0,0084	0,0099	0,9013	0,8109	0,9947
4	0,0078	0,0091	0,9077	0,8218	0,9950
5	0,0091	0,0101	0,8925	0,7955	0,9944
Média	0,0084	0,0096	0,8988	0,8063	0,9946
Desvio-padrão	0,0004	0,0004	0,0054	0,0094	0,0002

Fonte: Autor.

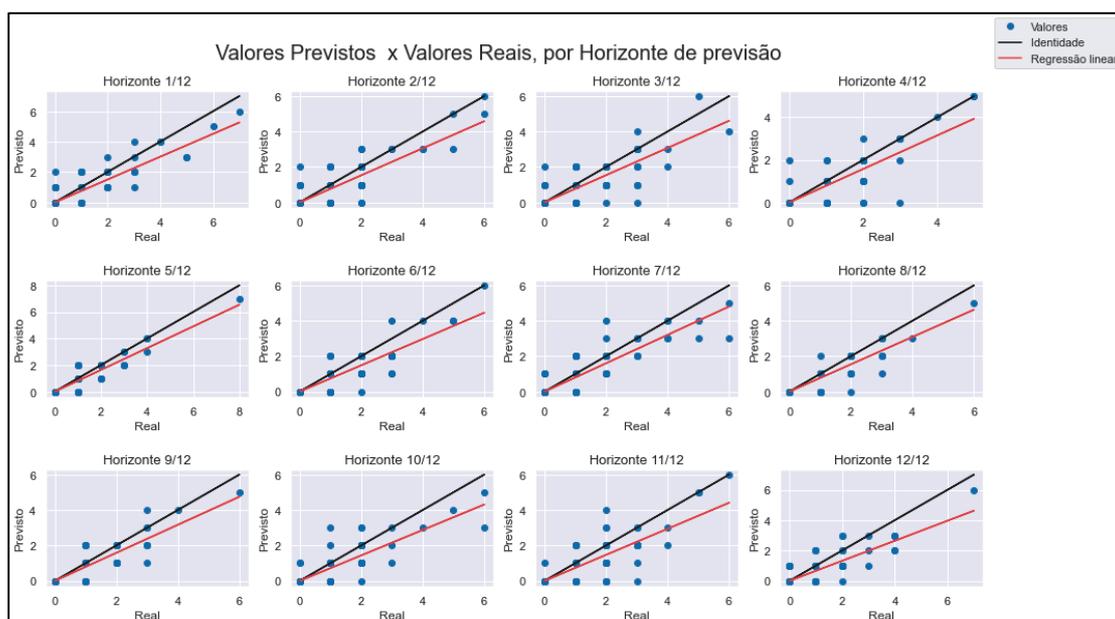
A análise da Tabela 4 mostra que o modelo desenvolvido apresentou grande estabilidade para os 5 *folds*, apesar de haver uma pequena variação entre eles.

A Figura 45 contém gráficos de dispersão de um dos *folds*, para cada um dos 12 intervalos de previsão, contrastando os valores previstos com os valores reais. Estes gráficos foram construídos utilizando-se os dados de teste, não vistos pelo modelo durante o treinamento. Assim como no caso do modelo de pós-processamento do WRF para a previsão de precipitação, este modelo tem tendência a subestimar os valores reais. Ainda assim, as linhas de regressão

linear mostram que os valores previstos são razoavelmente próximos dos valores reais.

Ressalta-se que a amplitude da quantidade de deslizamentos de terra é relativamente baixa, não ultrapassando 8 ocorrências registradas em 6 horas em uma das regiões. Os gráficos da Figura 45 contêm a representação de todos os 7070 registros presentes no conjunto de teste (não utilizados pelo modelo durante o treinamento). Como os valores são discretos e a amplitude é baixa, os pontos estão sobrepostos e não há muita dispersão.

Figura 45 - Modelo de previsão de deslizamentos: valores previstos e valores reais, por intervalo de previsão, para os dados de teste.



Fonte: Autor.

Um ponto positivo destes resultados é que este modelo tende a acertar a não ocorrência de deslizamentos, ou errar por pouco. Ou seja, quando não há deslizamentos de terra registrados (valor real igual a zero), a máxima quantidade prevista pelo modelo foi de dois deslizamentos, para estes dados de teste, em todos os 12 intervalos de previsão.

A Figura 46 contém 12 gráficos, que contrastam a quantidade de deslizamentos de terra reais e previstos pelo modelo, em cada intervalo de previsão, para cada instância dos dados separados para teste. Os marcadores indicam os deslizamentos de terra registrados.

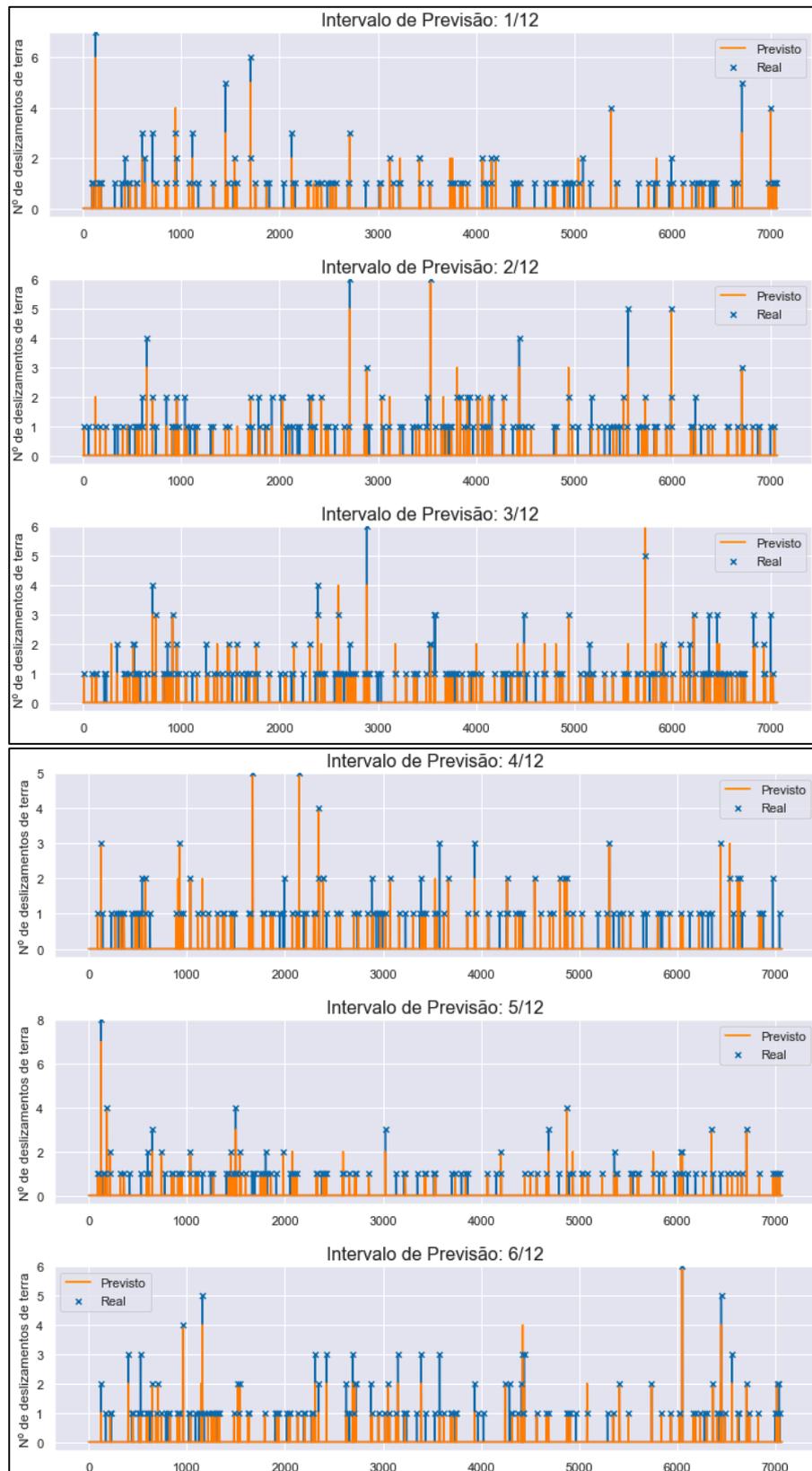
Nestes gráficos, fica visível a capacidade do modelo de representar este fenômeno complexo, visto que os dados previstos seguem um padrão parecido com o dos dados reais.

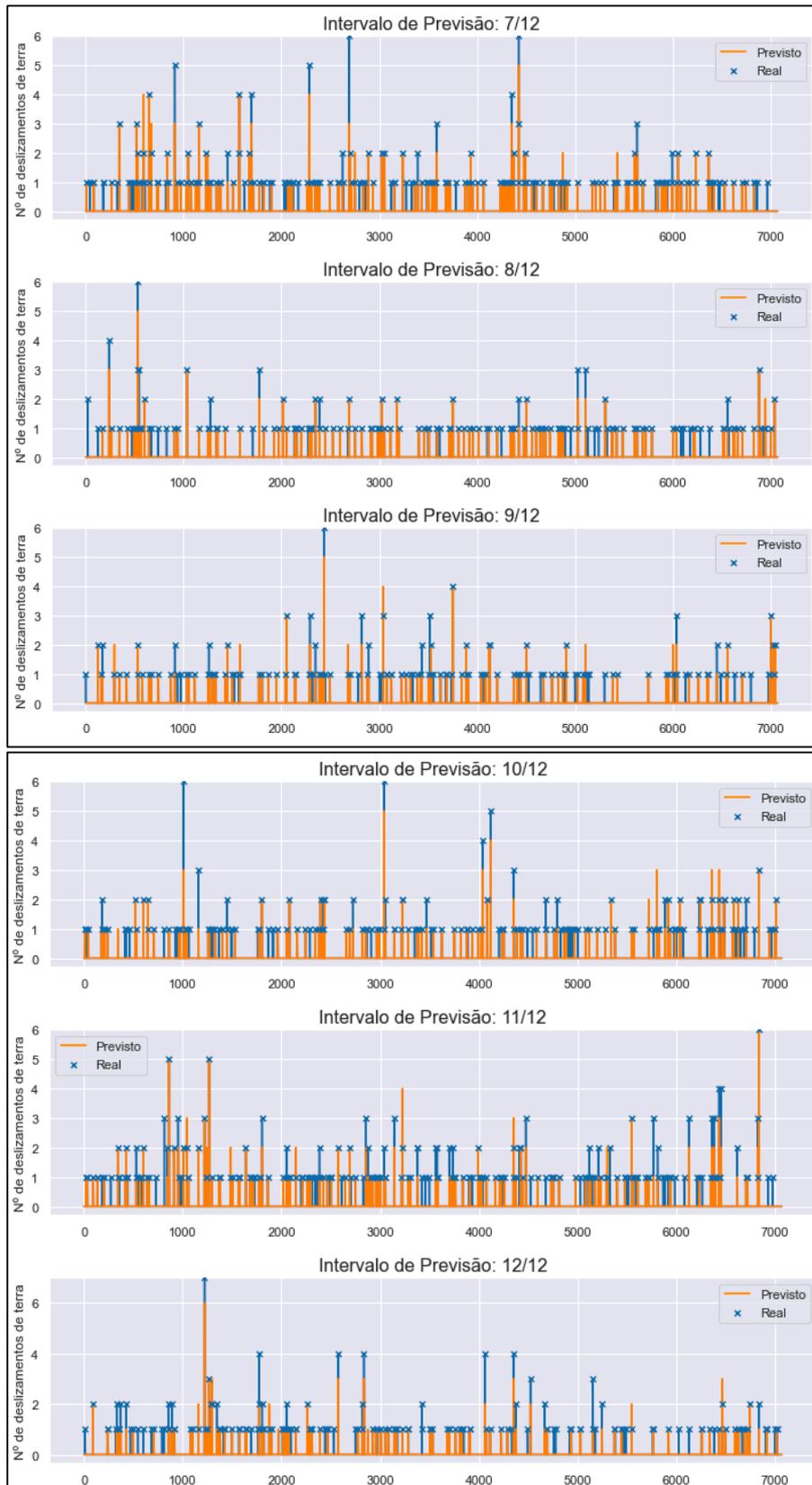
Como foi explicado na seção da metodologia utilizada para a construção dos modelos, a subdivisão do conjunto de dados em treino, validação e teste não foi feita utilizando períodos contínuos. Embora a utilização de períodos contínuos seja mais adequada para a modelagem de séries temporais, o período muito curto de dados efetivamente utilizados para a modelagem (menos de 15 meses) levou à decisão de selecionar amostras aleatoriamente para cada subdivisão (treino, validação e teste), estratificando de acordo com a região, o dia da semana e a hora do dia, de forma que cada subdivisão contivesse a mesma proporção de amostras desses atributos que o conjunto completo. Por isso, instâncias consecutivas em cada subdivisão não representam, necessariamente, intervalos subsequentes de tempo das mesmas regiões.

Vale ressaltar que, no momento da divisão em treino, validação e teste, cada instância já possuía uma janela temporal com os 19 intervalos subsequentes de tempo, de forma que a seleção aleatória das instâncias não alterou a janela temporal já existente em cada uma das instâncias. Ou seja, mesmo após a divisão para treino, validação e teste, cada instância passada para o modelo continha registros dos 18 intervalos de tempo anteriores, além do atual, totalizando uma janela temporal de 19 intervalos subsequentes de tempo.

Para facilitar a visualização dos dados reais e previstos, A Figura 47 contém uma ampliação dos gráficos apresentados na Figura 46, mostrando as instâncias 1000 a 2000 de cada intervalo de previsão.

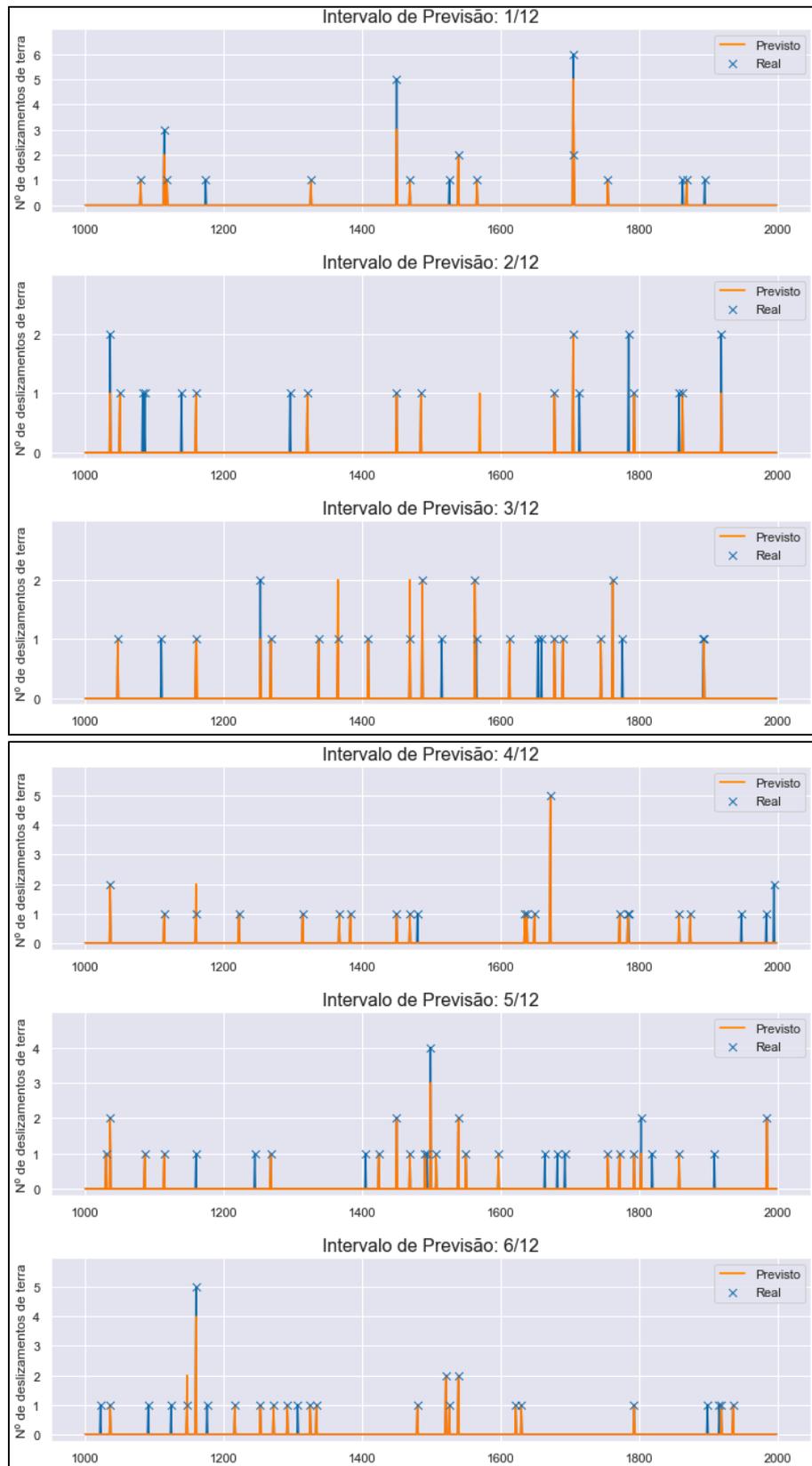
Figura 46 - Valores reais e previstos pelo modelo, para cada instância dos dados de teste.

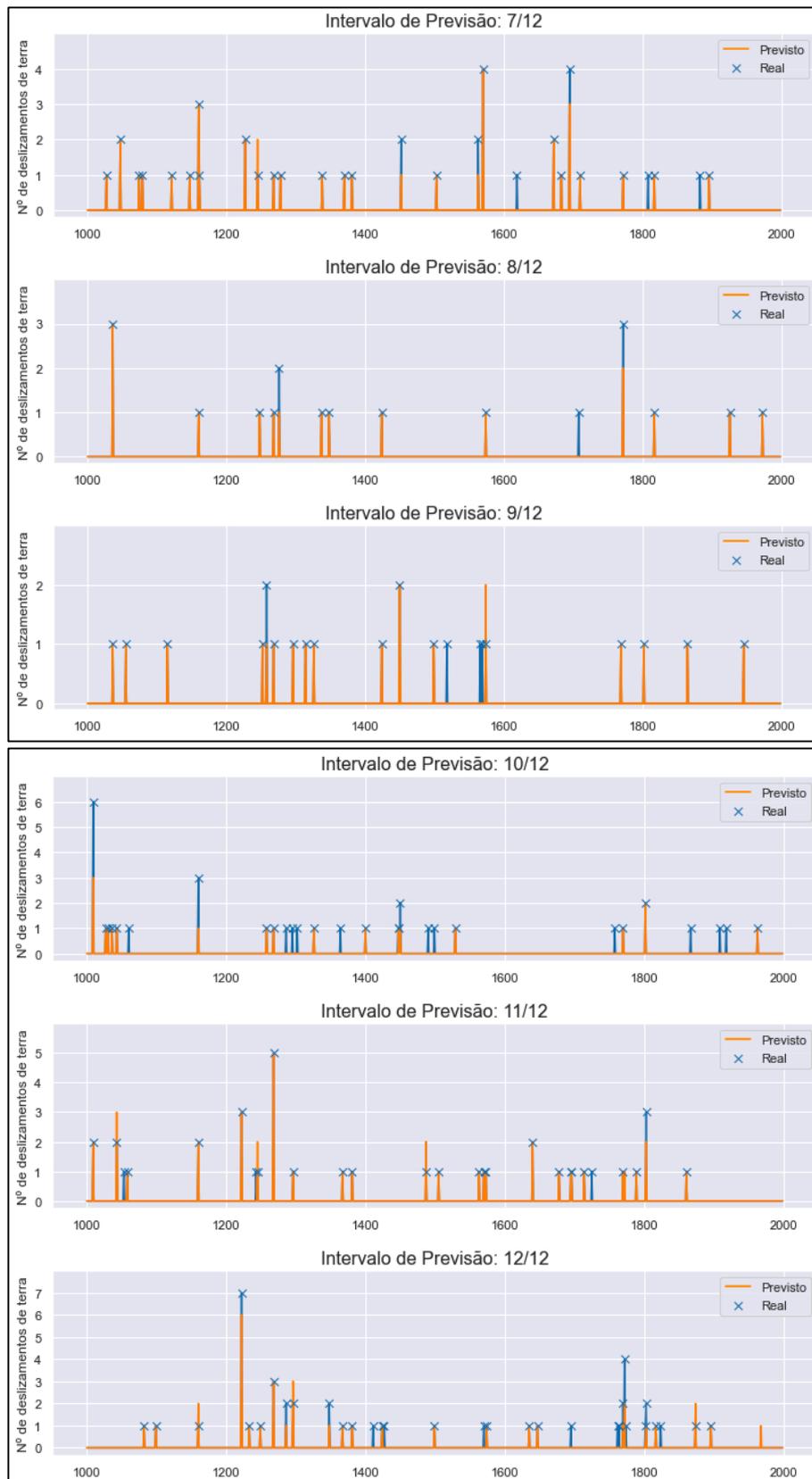




Fonte: Autor.

Figura 47 – Ampliação dos gráficos apresentados na Figura 46.





Fonte: Autor.

7 CONCLUSÃO

7.1 CONCLUSÕES

Com o objetivo de ampliar as capacidades de prevenção e mitigação dos impactos dos deslizamentos de terra no município de Salvador, foram desenvolvidos dois modelos utilizando inteligência artificial, que, juntos, são capazes de prever a ocorrência de deslizamentos de terra na cidade.

O primeiro modelo tem como objetivo a redução dos erros associados ao modelo numérico de previsão do tempo (WRF), para previsão de precipitação, operacionalizado na Defesa Civil de Salvador. As métricas analisadas mostram que, com o pós-processamento do modelo desenvolvido, a qualidade da previsão meteorológica foi muito superior.

O segundo modelo, que prevê quantidades de deslizamentos de terra para regiões delimitadas ao redor das estações pluviométricas, utilizou como entrada os resultados do primeiro modelo (precipitação prevista), os dados pluviométricos instantâneos e o histórico de deslizamentos de terra registrados.

Apesar de toda a complexidade do fenômeno estudado, o modelo de previsão de ocorrências apresentou resultados consideráveis, constituindo-se em um grande aperfeiçoamento das capacidades de defesa civil no município, referentes aos deslizamentos de terra.

Este modelo de previsão desenvolvido se mostrou capaz de modelar a ocorrência de deslizamentos de terra em Salvador, apesar de apresentar tendências de subestimação. De qualquer forma, seus resultados já podem ser utilizados como um indicador de risco futuro, se tornando na primeira ferramenta operacional que quantifique este risco de forma tangível, pelo menos no município de Salvador. Qualquer outra mensuração de risco de deslizamentos de terra feita hoje é largamente mais subjetiva que o modelo desenvolvido.

7.2 LIMITAÇÕES

Boa parte dos estudos existentes na área de previsão de deslizamentos de terra, ou mapeamento da susceptibilidade a deslizamentos de terra, trata de problemas razoavelmente determinísticos. Ou seja, identificam-se parâmetros físicos do contexto ambiental de determinado local e tenta-se prever os impactos de eventuais perturbações (chuva, remoção de vegetação, influências antrópicas) - no caso de previsão - ou inferir a susceptibilidade existente ou futura - no caso de identificação da susceptibilidade. Estes parâmetros são plenamente mensuráveis.

De forma diferente, o presente trabalho se propõe a prever o impacto de uma perturbação (chuva) a um sistema socioambiental complexo, com contexto caótico. A variável objeto de previsão indica a quantidade de deslizamentos de terra que foram verificados por equipes da Defesa Civil de Salvador e, conseqüentemente, inseridos em seu sistema.

É importante ressaltar que a Defesa Civil deve ser provocada para avaliar as situações de campo. Ou seja, depende primordialmente de solicitações da população. Dessa forma, é reconhecido que existem deslizamentos de terra que não são incorporados ao sistema da Defesa Civil, porque esta não teve conhecimento das ocorrências.

Assim, fica claro que, via de regra, o conhecimento da Defesa Civil sobre os deslizamentos de terra depende de ação humana (geralmente dos afetados, que entram em contato com o órgão). Esta dimensão humana é regida por livre arbítrio pessoal, e dificilmente pode ser modelada.

Além disso, existem outros contextos sociais e comunitários igualmente complexos. Por exemplo, no período chuvoso (abril a julho), o Poder Executivo local implanta planos de ação de órgãos relacionados à atividade de defesa civil e programas socioeconômicos, com incentivos financeiros, que visam a atenuar e prevenir o impacto à população. Esta característica, por si só, ainda que importantíssima, induz à proliferação de registros na Defesa Civil de Salvador,

por vezes inadequados. Existem até constatações de deslizamentos de terra deliberadamente provocados, com o intuito de recebimento de incentivos econômicos, embora estes casos sejam minoria.

Ainda que o objetivo maior deste trabalho seja a previsão do risco de deslizamentos de terra, podendo potencializar os esforços dos órgãos do Sistema Municipal de Proteção e Defesa Civil para redução de impactos, na verdade, estão sendo previstos aqui os deslizamentos de fato registrados na Defesa Civil.

Por isso, não se espera atingir uma modelagem plena, já que o sistema objeto de estudo possui as já citadas dimensões sociais, econômicas e naturais, que contêm complexidade exagerada, diferentemente dos outros trabalhos da área.

7.3 TRABALHOS FUTUROS

Além do aperfeiçoamento dos modelos implementados, o próximo passo lógico seria o desenvolvimento de um modelo de previsão do risco de deslizamentos para cada região. Isto seria um problema de classificação, enquanto o presente trabalho utilizou regressão.

Pretende-se, também, implementar modelos de pós-processamento dos resultados do WRF para as demais variáveis meteorológicas, principalmente velocidade e direção dos ventos, temperatura e pressão, que têm maior potencial de auxiliar nas capacidades de análise meteorológica, e podem gerar riscos diretos à população.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET). Anais... In: 2017 INTERNATIONAL CONFERENCE ON ENGINEERING AND TECHNOLOGY (ICET). ago. 2017
- BASTOS, C. C.; FERREIRA, N. J. Análise climatológica da Alta Subtropical do Atlântico Sul. In: XI CONGRESSO BRASILEIRO DE METEOROLOGIA. 2000
- CAMARINHA, P. I. M. Vulnerabilidade aos desastres naturais decorrentes de deslizamentos de terra em cenários de mudanças climáticas na porção paulista da Serra do Mar. Tese de Doutorado—São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2016.
- COSTA, J. DE J.; GIUDICE, D. S. Fenômenos naturais e ação antrópica: problemática de Salvador - BA. Revista Geonorte, v. 3, n. 5, p. 477 – 486–477 – 486, 12 nov. 2012.
- DAI, F. C.; LEE, C. F. Frequency–volume relation and prediction of rainfall-induced landslides. Engineering Geology, v. 59, n. 3, p. 253–266, 1 abr. 2001.
- GEOHIDRO. Plano diretor de encostas do município de salvador – Módulo I: Relatório do inventário das áreas de risco, 2003.
- HEWAGE, P. et al. Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications, 22 jun. 2020.
- HUANG, F. et al. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. Landslides, v. 17, n. 1, p. 217–229, 1 jan. 2020.
- IBGE. Censo Demográfico 2010, 2010. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html>>. Acesso em: 1 ago. 2020

IBGE. Cidades@, [s.d.]. Disponível em: <<https://cidades.ibge.gov.br/brasil/ba/salvador/panorama>>. Acesso em: 1 ago. 2020

IBGE; CEMADEN. População em áreas de risco no Brasil. 2018.

INMET. Instituto Nacional de Meteorologia. Normais Climatológicas do Brasil, 1981- 2010. Brasília, 2018.

JABBARI, A.; BAE, D.-H. Application of Artificial Neural Networks for Accuracy Enhancements of Real-Time Flood Forecasting in the Imjin Basin. *Water*, v. 10, n. 11, p. 1626, nov. 2018.

KORUP, O.; STOLLE, A. Landslide prediction from machine learning. *Geology Today*, v. 30, n. 1, p. 26–33, 2014.

KROGH, A. What are artificial neural networks? *Nature Biotechnology*, v. 26, n. 2, p. 195–197, fev. 2008.

LOGAR, J. et al. Prediction of rainfall induced landslide movements by artificial neural networks. *Natural Hazards and Earth System Sciences Discussions*, p. 1–18, 13 jul. 2017.

MONDAL, B. Artificial Intelligence: State of the Art. Recent Trends and Advances in Artificial Intelligence and Internet of Things, p. 389–425, 2020.

NASCIMENTO, G. C. Utilização do Modelo de Meso-Escala WRF em Áreas de Riscos no Município de Maceió. Trabalho de Conclusão de Curso (Graduação em Meteorologia) - Maceió: Universidade Federal de Alagoas, 2014.

PALMA, J. DOS S. Variabilidade das chuvas em Salvador e suas possíveis tendências espaço-temporais. 22 jul. 2016.

PHAM, B. T. et al. Spatial Prediction of Rainfall-Induced Landslides Using Aggregating One-Dependence Estimators Classifier. *Journal of the Indian Society of Remote Sensing*, v. 46, n. 9, p. 1457–1470, 1 set. 2018.

- RASP, S.; LERCH, S. Neural Networks for Postprocessing Ensemble Weather Forecasts. *Monthly Weather Review*, v. 146, n. 11, p. 3885–3900, 1 nov. 2018.
- RIFFEL, E. S.; GUASSELLI, L. A.; BRESSANI, L. A. Desastres associados a movimentos de massa: uma revisão de literatura. *Boletim Goiano de Geografia*, v. 36, n. 2, p. 301–301, 16 ago. 2016.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 2. ed. [s.l.] Pearson Education, 2003.
- SANTANA, N. S. *Análise da Vulnerabilidade Ambiental de Salvador: Um Subsídio ao Planejamento e à Gestão Territorial da Cidade*. 20 fev. 2017.
- SANTOS, A. P. P. DOS et al. Precipitação na Cidade de Salvador: Variabilidade Temporal e Classificação em Quantis. *Revista Brasileira de Meteorologia*, v. 31, n. 4, p. 454–467, 29 set. 2016.
- TIEN BUI, D. et al. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, v. 13, n. 2, p. 361–378, 1 abr. 2016.
- TIEN BUI, D. et al. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides*, v. 14, n. 2, p. 447–458, 1 abr. 2017.
- YANG, B. et al. Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides*, v. 16, n. 4, p. 677–694, 1 abr. 2019.
- ZHAO, B. et al. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, v. 28, n. 1, p. 162–169, fev. 2017.
- ZHU, X. et al. Comparison of two optimized machine learning models for predicting displacement of rainfall-induced landslide: A case study in Sichuan Province, China. *Engineering Geology*, v. 218, p. 213–222, 23 fev. 2017.