

Armazenamento de Dados de Precipitação Utilizando Imagens de Radar do Estado do Paraná

Olavo Pereira do Nascimento¹, Daniel dos Santos Kaster¹

¹Departamento de Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

olavo.nascimento@uel.com, dskaster@uel.br

Abstract. *With the increasing use of climate data in multiple meteorological tools, the need for efficient and user-friendly datasets becomes evident. However, the state of Paraná lacks precipitation data that can be easily integrated into predictive models for academic and commercial use. This paper aims to use precipitation data captured through SIMEPAR radars, which are further processed through image segmentation and then stored in a way that minimizes the space used, without renouncing search efficiency. Thus aiming to create a database that can help other areas that lack appropriate data, especially agronomy.*

Resumo. *Com o crescente uso de dados climáticos em múltiplas ferramentas meteorológicas, se torna evidente a necessidade de conjuntos de dados eficientes e de fácil utilização. Contudo, o estado do Paraná não possui dados de precipitação para uso acadêmico e comercial que sejam facilmente integrados a modelos preditivos. Este trabalho tem como objetivo utilizar dados de precipitação capturadas através de radares do SIMEPAR, os quais são posteriormente processados através da segmentação das imagens e então armazenadas de forma a minimizar o espaço utilizado, sem renunciar a eficiência de busca. Visando a criação de um banco de dados que possa auxiliar outras áreas que carecem de dados apropriados, em especial a agrônômica.*

1. Introdução

A chegada da quarta revolução industrial é uma das maiores promessas do século XXI, entre diversos pontos de inovação um se destaca, inteligência artificial. Antes de tudo, é preciso estabelecer a importância de dados para o treinamento de modelos preditivos, os quais precisam de dados corretamente selecionados e preparados.

A rápida adoção de sistemas modernos para as previsões meteorológicas resultou em enormes bancos de dados para os campos de hidrologia e meteorologia [4]. Tais dados influenciam diretamente, não só essas áreas, como também os campos de ecologia e agronomia, que se tornaram dependentes de informações de precipitação com boa resolução espaço-temporal para alimentar ferramentas essenciais nos fluxos de análise climática [1][16].

Uma das principais entidades de coleta, análise e distribuição de dados climáticos no estado do Paraná é o SIMEPAR (Sistema de Tecnologia e Monitoramento Ambiental do Paraná), o qual utiliza três radares meteorológicos, posicionados em Teixeira Soares, Cascavel e Curitiba, assim como o uso de satélites para capturar dados atmosféricos.¹

¹http://www.simepar.br/prognozweb/simepar/radar_msc

A ideia de utilizar radares para mapear dados de precipitação em mapas é antiga, tendo origem na década de 1940 [12]. Na atualidade, uma época pós corrida espacial, onde ferramentas baseadas no uso satélites se tornaram algo comum no cotidiano de todos, dados de radares continuam relevantes. Um dos motivos para essa constatação é intrínseco à forma com que atuam. Enquanto satélites medem condições através da atmosfera, radares meteorológicos são capazes de medir as condições na superfície do planeta, permitindo extrair informações com maior resolução espaço-temporal [6].

Os radares utilizados operam em frequências entre 2,7GHz a 2,9GHz, sendo categorizadas como bandas de tipo S, o que pode ser verificado na tabela 1 (adaptada de [7]). A área de captura representa uma área de 480km ao redor dos radares, permitindo capturar não só todo o Paraná, como Santa Catarina, o sul do estado de São Paulo e o norte do Rio Grande do Sul [2].

Tabela 1. Frequência de Banda de Radares

Banda	Frequência (GHz)	Tamanho da Onda (cm)
Milímetro	40-100	0.75-0.30
Ka	26.5-40	1.1-0.75
K	18-26.5	1.7-11
Ku	12.5-18	2.4-17
X	8-12.5	3.75-2.4
C	4-8	7.5-3.75
S	2-4	15-7.5
L	1-2	30-15
UHF	0.3-1	100-30

Como evidenciado em [13], dados de precipitação com boa resolução espaço-temporal são necessários em escala global. Todavia, ainda que os radares capturem dados em um intervalo de 10 minutos, os dados gerados não são de fácil utilização para modelos preditivos, da mesma forma que a abundância de dados atmosféricos capturados diariamente pelos radares torna inviável a classificação manual [3].

Buscando mitigar esse problema em uma escala local, o objetivo deste trabalho de conclusão de curso é construir um banco de dados de precipitação a partir de dados de radar meteorológico com escopo limitado ao estado do Paraná. O processo de construção deste banco de dados inclui a coleta automatizada de dados disponibilizados publicamente, a estruturação dos dados de forma compacta e adequada, visando disponibilizar dados de qualidade apropriada para uso no treinamento de modelos preditivos empregados na área da agronomia. Espera-se que o banco de dados resultante beneficie a sociedade facilitando o acesso a dados de precipitação para uso acadêmico e comercial.

Na seção 2, é apresentada a fundamentação teórica-metodológica do trabalho, assim como os avanços na área de armazenamento de dados espaço-temporais. Na seção 3, são apresentados os objetivos deste trabalho. Na seção 4 são definidos os métodos que serão empregados para alcançar os objetivos propostos. Por fim, na seção 5, é definido um cronograma para o desenvolvimento de partes específicas do trabalho.

2. Fundamentação Teórico-Metodológica e Estado da Arte

2.1. Segmentação de Imagens

Segmentação de imagens se refere ao processo de partição de uma imagem em um conjunto de regiões que representam áreas de importância [11]. Na climatologia, esse conceito é utilizado para a análise de imagens climáticas, onde muitas vezes para facilitar a visualização dos dados são utilizadas diferentes cores para representar os valores capturados [9]. Exemplo disso são os dados atmosféricos distribuídos pelo SIMEPAR, o qual emprega um gradiente de cores para identificar diferentes valores de precipitação em um pixel da imagem.

Esta técnica é normalmente utilizada por conta de dois fatores. O primeiro é a decomposição da imagem em partes para análise aprofundada de certos aspectos. O segundo fator é a alteração da representação dos dados, visando obter dados que sejam mais significativos ou que permitam uma eficiência de análise superior aos dados brutos [11]. A figura 1 é segmentada seguindo o primeiro ponto apresentado anteriormente, decompor a imagem em diferentes partes de acordo com a situação meteorológica.

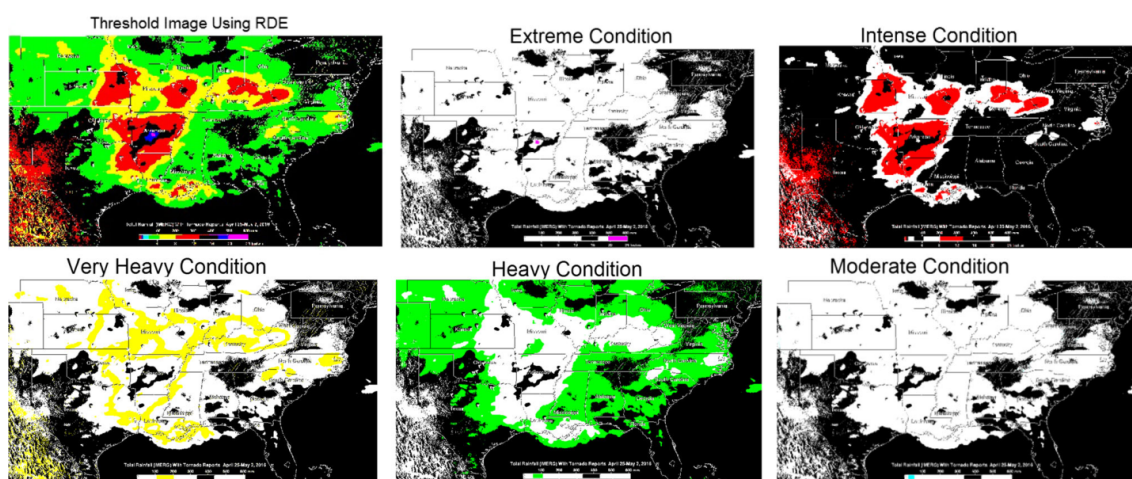


Figura 1. Exemplo de segmentação com base em condições meteorológicas [9]

2.2. Armazenamento de Imagens

2.2.1. Sistemas de Informações Geográficas

Sistemas de Informações Geográficas (SIG) são softwares que realizam diversas funções em dados geográficos, desde processamento e armazenamento até a análise e visualização dos dados. Através de novas tecnologias é possível integrar as múltiplas funções de um sistema de informações geográficas com ferramentas externas, intensificando ainda mais a influência desses sistemas em diversas áreas que lidam com dados geográficos [10].

Após o processamento dos dados meteorológicos, é essencial armazená-los de forma apropriada, trabalhando os principais benefícios desse tipo de sistema, como a rápida entrada de dados, armazenamento eficiente e manipulação de dados geográficos, facilitando consultas futuras. No âmbito de um menor custo espacial para armazenamento de dados, a maioria dos sistemas empregam arquivos binários que tem como função exclusiva registrar dados geográficos utilizando matrizes, as quais referenciam dados de um

região definida através de propriedades de projeção, proporcionando um uso de armazenamento até 51% menor que os dados de entrada [14].

Ao mesmo tempo, os sistemas disponíveis não fornecem uma forma simples para adicionar, armazenar e manipular dados com variação temporal, como dados de precipitação capturados por radares [4], sendo necessário estudar os dados que serão armazenados para chegar a possíveis soluções alternativas que se provem mais apropriadas para área abordada.

É necessário notar algumas adversidades particulares ao uso de bancos relacionais para o armazenamento de dados geográficos com uma dimensão temporal, exemplo disso é a execução de buscas baseadas em tempo, impondo uma desvantagem significativa para esse tipo de consulta devido ao fato que não existe nenhuma ordenação explícita entre relações [8].

Um método alternativo para armazenamento é processar os dados meteorológicos, gerar um arquivo XML com uma estrutura definida e armazenar essa informação em um banco que suporte nativamente essa opção, assim como apresentado em [5]. Contudo, isso resulta em um desempenho inferior para buscas que utilizam valores de várias tabelas, afetando a experiência do usuário ao realizar consultas em bancos maiores. É notável então que o equilíbrio entre os dois métodos deve ser encontrado e aplicado ciente dos pontos positivos e negativos de cada técnica.

Muitos dos trabalhos abordados sofrem com a difícil ingestão de dados nos sistemas de informações geográficas, onde os mesmos muitas vezes devem sofrer um longo pré-processamento antes de serem utilizados, dificultando a rápida aquisição de dados de precipitação [14]. Este trabalho busca permitir que os dados resultantes sejam de fácil acesso para outros membros da sociedade, diminuindo as barreiras de desenvolvimentos e expandindo o potencial de ferramentas baseadas em inteligência artificial na região estudada.

2.3. Consulta de Dados Espaço-Temporais

Para a realização de consultas de dados com base em uma dimensão temporal é primeiramente necessário entender a forma com que os dados armazenados são representados. Fenômenos geográficos podem ser categorizados de duas formas distintas: discretos ou contínuos. Dados discretos possuem posição e limites definidos, sendo definidos como entidades antes da leitura de seus atributos. Ao mesmo tempo, dados contínuos não possuem limites definidos e são distribuídos continuamente no espaço. Dados de precipitação apresentam características de ambas as categorias [15].

Com o tipo de informação definido existem diversas formas de representar a dimensão temporal dos dados geográficos, uma forma explorada em [8], onde os autores apresentam a ideia de *snapshots*. Onde cada imagem S_i simboliza um estado do mundo em uma posição no tempo t_i . Para calcular uma mudança é necessário encontrar a posição t_i desejada e analisar a diferença do valor de um pixel com o período t_{i-1} .

3. Objetivos

O objetivo geral deste projeto é criar um banco de dados de dados de precipitação para suportar aplicações em diversos campos de conhecimento. Para isso os dados finais devem

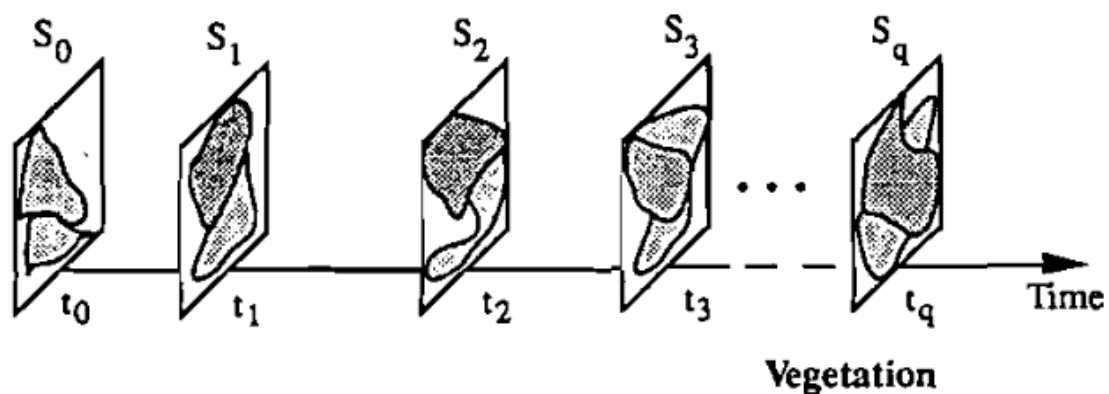


Figura 2. Evolução de uma snapshot conforme o tempo [8]

ser de simples distribuição e uso por terceiros, um ponto que pode ser visto neste trabalho, onde os dados utilizados se encontram em imagens de difícil aproveitamento por modelos preditivos. Além disso, é essencial que a busca por eficiência de armazenamento não sobreponha a velocidade de pesquisa dos dados resultantes, permitindo que ferramentas baseadas em consumo, atualização e consulta de dados continuem responsivas para o usuário.

4. Procedimentos metodológicos/Métodos e técnicas

Para alcançar os objetivos apontado na seção 3 são definidos alguns pontos relevantes para o desenvolvimento deste trabalho. Inicialmente, é preciso realizar uma análise aprofundada sobre o armazenamento de dados geográficos 2D com um dimensão adicional representando o tempo utilizando sistemas de informações geográficas.

Após a conclusão da revisão do estado da arte será necessário definir o método mais apropriado, conforme apontado pela literatura, para a ingestão dos dados de precipitação disponibilizados pelo SIMEPAR. Para isso, será necessário utilizar técnicas de segmentação de imagens para separar as informações definidas por diferentes cores e transformá-las em dados mais significativos que serão armazenados em um SGBD que possua suporte apropriado a tipos geográficos. É de suma importância balancear os benefícios e malefícios da estrutura relacional do banco de dados para alcançar uma performance apropriada para os objetivos definidos.

Por fim, os dados armazenados, assim como a modelagem de dados escolhida, devem ser avaliadas através da execução de consultas que explorem a natureza espaço-temporal do conjunto de dados criado. Averiguando se o resultado final é apropriado para a utilização por trabalhos futuros.

5. Cronograma de Execução

Atividades:

1. Revisão do estado da arte da área de armazenamento de dados espaço-temporais e segmentação de imagens;
2. Utilizar noções de segmentação para separar as camadas úteis das imagens de precipitação do SIMEPAR

3. Aplicação dos conceitos aprendidos para armazenar os dados de forma eficiente;
4. Definir métodos para execução de buscas espaço-temporais utilizando Sistemas de Informações Geográficas;
5. Avaliar a eficiência do *dataset* criado para aplicação em trabalhos futuros.
6. Escrita do Trabalho de Conclusão de Curso;

Tabela 2. Cronograma de Execução

	jan	fev	mar	abr	mai	set	out	nov	dez
Atividade 1						x			
Atividade 2							x	x	
Atividade 3								x	x
Atividade 4	x	x	x						
Atividade 5			x	x					
Atividade 6	x	x	x	x	x				x

6. Contribuições e/ou Resultados esperados

Através desse trabalho é esperado a criação de métodos e técnicas para o desenvolvimento e uso de bancos de dados de radares meteorológicos. Além disso, espera-se construir um banco de dados que possa ser utilizado por trabalhos futuros, auxiliando áreas que são dependentes de dados de precipitação para o estado, em especial a área da agronomia. Tanto os métodos de construção quanto o próprio banco de dados construído poderão beneficiar a sociedade como um todo garantindo previsibilidade a ferramentas climáticas ao fomentar o uso de dados específicos para a região estudada.

7. Espaço para assinaturas

Londrina, 12 de setembro de 2022.

Olavo P. do Nascimento

Aluno

Orientador

Referências

- [1] Oscar M Baez-Villanueva, Mauricio Zambrano-Bigiarini, Hylke E Beck, Ian McNamara, Lars Ribbe, Alexandra Nauditt, Christian Birkel, Koen Verbist, Juan Diego Giraldo-Osorio, and Nguyen Xuan Thinh. Rf-mep: A novel random forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment*, 239:111606, 2020.
- [2] Cesar Beneti, Itsumi Nozu, and Ernandes Saraiva. Monitoramento da precipitação e de eventos de tempo severo com radar meteorológico no estado do paraná. page 5, 1998.

- [3] David John Gagne, Amy McGovern, and Jerry Brotzge. Classification of convective areas using decision trees. *Journal of Atmospheric and Oceanic Technology*, 26(7):1341–1353, 2009.
- [4] Mathieu Lagrange, Hervé Andrieu, Isabelle Emmanuel, Gerard Busquets, and Stéphane Loubrié. Classification of rainfall radar images using the scattering transform. *Journal of Hydrology*, 556:972–979, 2018.
- [5] Xiaohua Liu and Youchuan Wan. Storing spatio-temporal data in xml native database. In *2010 2nd International Workshop on Database Technology and Applications*, pages 1–4. IEEE, 2010.
- [6] Timothy D Mitchell and Philip D Jones. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(6):693–712, 2005.
- [7] Michael Parker. *Digital Signal Processing 101, Second Edition: Everything You Need to Know to Get Started*. Newnes, USA, 2nd edition, 2017.
- [8] Donna J Peuquet and Niu Duan. An event-based spatiotemporal data model (estdm) for temporal analysis of geographical data. *International journal of geographical information systems*, 9(1):7–24, 1995.
- [9] Meera Ramadas, Millie Pant, Ajith Abraham, and Sushil Kumar. Segmentation of weather radar image based on hazard severity using rde: reconstructed mutation strategy for differential evolution algorithm. *Neural Computing and Applications*, 31(2):1253–1261, 2019.
- [10] Oscar Robayo. *Map to map: converting a NEXRAD rainfall map into a flood inundation map*. The University of Texas at Austin, 2005.
- [11] Linda G Shapiro, George C Stockman, et al. *Computer vision*, volume 3. Prentice Hall New Jersey, 2001.
- [12] Francisco J Tapiador, Francis J Turk, Walt Petersen, Arthur Y Hou, Eduardo García-Ortega, Luiz AT Machado, Carlos F Angelis, Paola Salio, Chris Kidd, George J Huffman, et al. Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104:70–97, 2012.
- [13] AC Turlapaty, VG Anantharaj, NH Younan, and FJ Turk. Precipitation data merging using a machine learning based fusion. In *EGU General Assembly Conference Abstracts*, page 4959, 2009.
- [14] Hongjie Xie, Xiaobing Zhou, Enrique R Vivoni, Jan MH Hendrickx, and Eric E Small. Gis-based nexrad stage iii precipitation database: automated approaches for data processing and visualization. *Computers & Geosciences*, 31(1):65–76, 2005.
- [15] May Yuan. Representing complex geographic phenomena in gis. *Cartography and Geographic Information Science*, 28(2):83–96, 2001.
- [16] Junmin Zhang, Jianhui Xu, Xiaoi Dai, Huihua Ruan, Xulong Liu, and Wenlong Jing. Multi-source precipitation data merging for heavy rainfall events based on cokriging and machine learning methods. *Remote Sensing*, 14(7):1750, 2022.