



Projeto de Iniciação Científica - PIBIC

INFERÊNCIA PARA DADOS ESPAÇO-TEMPORAIS COM RESPOSTAS CENSURADAS E FALTANTE: Uma aplicação em dados meteorológicos

Instituto de Matemática, Estatística e Computação Científica - IMECC

Departamento de Estatística

Orientador: Larissa Ávila Matos - larissam@unicamp.br

Bolsista: Nathan Brusamarello De Souto - n222854@dac.unicamp.br

Objetivos

- Coletar dados meteorológicos por meio do INMET (Instituto Nacional de Meteorologia) e CEPAGRI (Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura);
- Produzir análises exploratórias nas variáveis
- Construir modelos para ajustar os dados da precipitação nas mesorregiões do estado de São Paulo

Introdução

Há um interesse crescente em estudar dados que são indexados no espaço e no tempo simultaneamente. Assim, em muitas áreas, como agricultura, economia e biologia, modelos flexíveis estão sendo desenvolvidos para trabalhar com este tipo de dados com base nos modelos espaciais, uma vez que os dados espaciais são um caso particular de dados espaço-temporais quando o número de observações ao longo do tempo é igual a um. Uma vantagem deste tipo de modelo é que ele permite incluir interações entre os componentes espaciais e temporais.

Portanto, um dos objetivos do projeto foi fazer previsões sobre o comportamento das chuvas nas mesorregiões no Brasil. Porém, durante a realização da pesquisa, destinamos as análises somente para as mesorregiões do estado de São Paulo. Para isso, utilizamos dados das estações disponíveis pelo INMET e pelo CEPAGRI. Os dados são amostras do volume de chuva de 24 estações meteorológicas automáticas espalhadas pelo estado de São Paulo, que coletam chuva, temperatura, umidade, pressão atmosférica e intensidade do vento em intervalos de 15 minutos, e transmitem a informação em tempo real, onde algumas dessas variáveis apresentam censura devido ao método de coleta.



Conjunto de dados

No início do projeto, foram feitas análises com dados meteorológicos de 12 estações do INMET e 1 estação do CEPAGRI. Porém em Março de 2020 foram incorporadas 12 estações de cidades do estado de São Paulo. Na Figura 1, o gráfico a direita apresenta o estado de São Paulo com as cidades disponibilizadas pelo INMET e Campinas (CEPAGRI). Já no gráfico a esquerda temos as mesorregiões do estado de São Paulo.

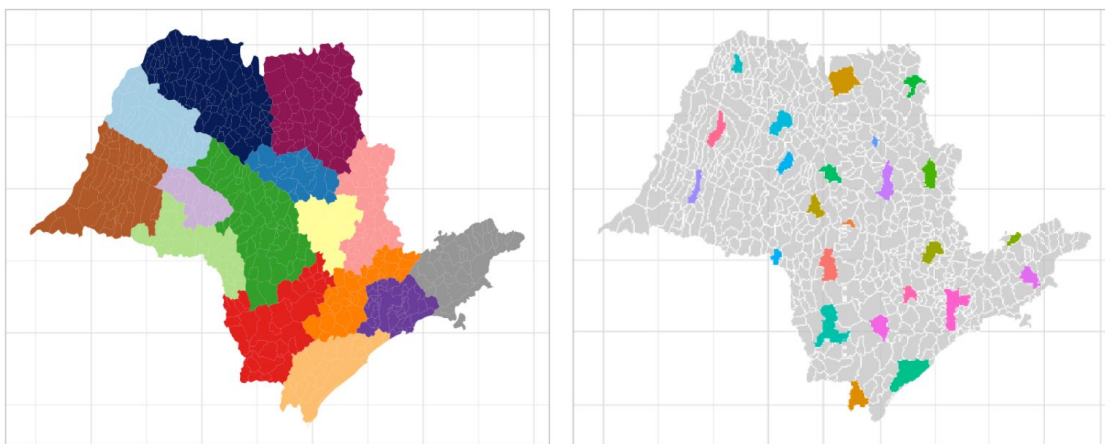


Figura 1: Regiões Metropolitanas das cidades disponibilizadas pelo INMET, com Campinas (CEPAGRI).

Durante a leitura dos dados, alguns desafios surgiram. Entre os arquivos disponibilizados pelo CEPAGRI, grande parte estavam com formatos diferentes e em muitos não tivemos êxito em realizar a leitura, devido a erros gerados pelo software e linguagem utilizados (RStudio/R). Com isso, limitamos nossa pesquisa em analisar os dados a partir do ano de 2011, pois foi a partir desse ano que conseguimos realizar a leitura de forma completa. Todos os processos de leitura e manipulação dos dados foram feitos através do software livre RStudio, utilizando a linguagem de programação R. A etapa de coleta/leitura dos dados foi a mais demorada em toda a pesquisa.

Resultados

Durante a etapa da análise exploratória, gráficos e tabelas foram produzidos para analisar os dados e resumi-los, facilitando a interpretação. Na Figura 2, é possível ver o comportamento da precipitação total (variável de interesse) nas cidades (representadas em cores diferentes em cada gráfico) que estão presentes em 13 das 15 mesorregiões.

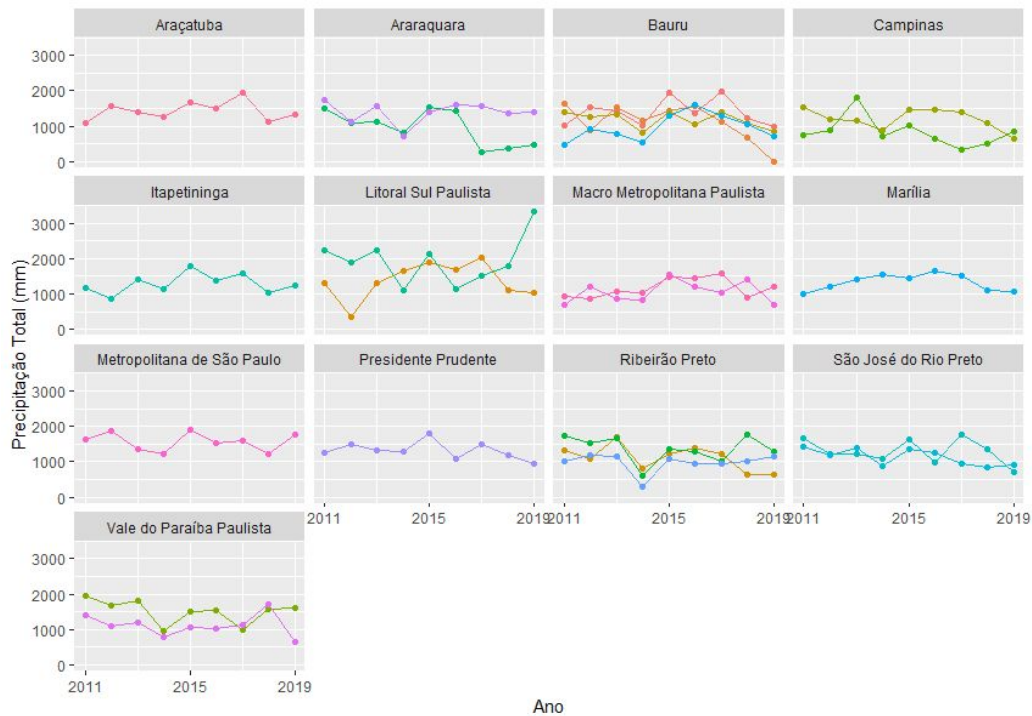


Figura 2: Precipitação total anual de cada cidade entre 2011 e 2019 representadas por gráficos de linhas (escalas diferentes para cada gráfico).

A Figura 3 também representa as médias anuais da precipitação diária das mesorregiões com uma visualização geográfica. É possível notar que os volumes de chuva aparentam estar diminuindo em todo estado de São Paulo com o decorrer dos anos, com exceção somente as regiões litorâneas.

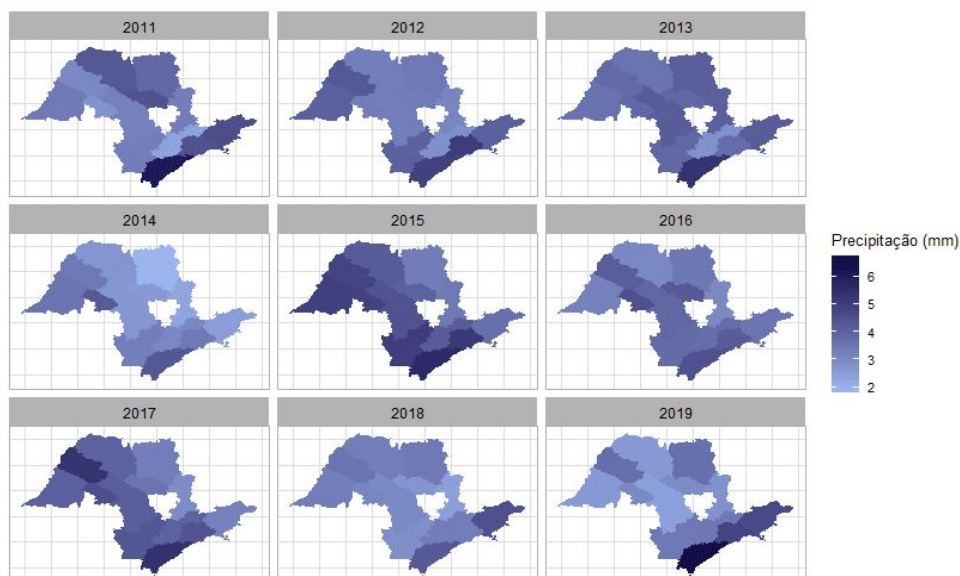




Figura 3: Média anual da precipitação diária (mm) das mesorregiões entre os anos 2011 e 2019 representada por gráficos de calor.

Foi feita uma análise exploratória detalhada para as outras variáveis analisadas e com essas análises partimos para a construção do modelo de previsão das precipitações médias mensais das mesorregiões. O modelo proposto foi:

$$Precipitação_i = \beta_0 + \beta_1 t_i + \sum_{a=2}^{11} \beta_a X_{ai} + Erro_i,$$

sendo que os β 's são os coeficientes, t_i é a tendência ou tempo, $X_a = I_{\{\text{mês}=a-1\}}$, e I é a função indicadora. Portanto, para o mês de janeiro temos que, $X_2 = 1$ (mês 1) e os outros meses são iguais a 0, isso é válido para todos os outros meses. O mês de Dezembro está no modelo também, consideramos a parametrização casela de referência, e o mês de Dezembro está no intercepto (β_0), ou seja, este modelo é identificável (não está superparametrizado). Além disso, consideramos que os erros seguem distribuição normal com média 0.

Para a mesorregião de Campinas, o modelo proposto se aproximou dos valores reais, como podemos ver na Figura 4. Alguns testes e análises foram feitos no modelo, e os mesmos trouxeram resultados validando as condições do modelo proposto. As mesorregiões Bauru, Araraquara e Araçatuba tiveram os resultados semelhantes.

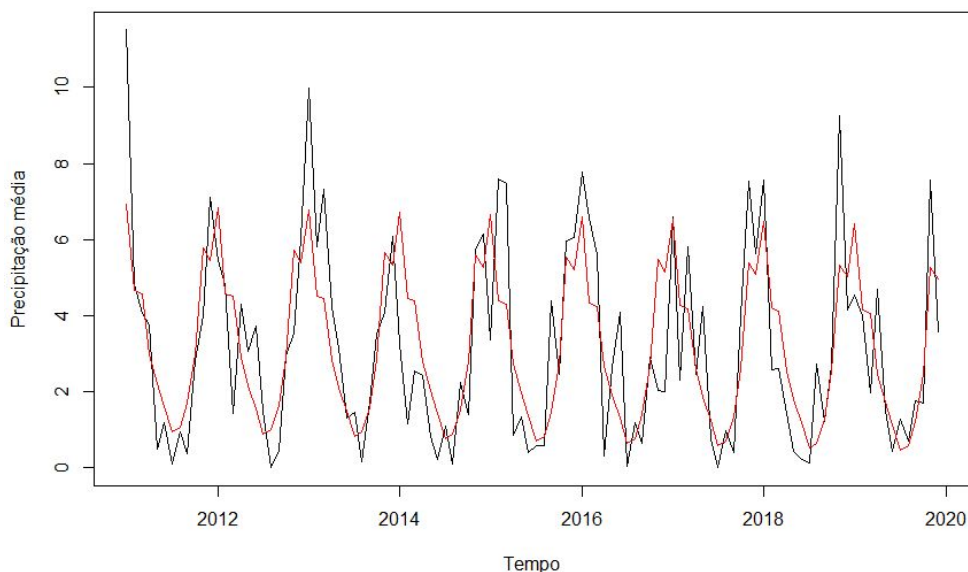


Figura 24: Série temporal da precipitação média da mesorregião Campinas sobreposta com os valores ajustados do modelo proposto.



Outras mesorregiões tiveram resultados diferentes, e os modelos não respeitaram as condições propostas. Para isso, será necessário a construção de um novo modelo, com a possibilidade de acrescentar as variáveis exploradas na primeira parte da pesquisa. Um possível novo modelo criado pode se ajustar bem aos dados das mesorregiões que já foram satisfeitas com o modelo simples, podendo até melhorar as conclusões.

Conclusão

Com as análises exploratórias foi possível notar indícios de que a precipitação está diminuindo com o passar do tempo, mas não foi possível concluir essa possibilidade com os testes de hipótese realizados.

Em relação ao comportamento das chuvas nos meses do ano, foi possível notar na análise exploratória que os meses do período do Verão tem precipitações maiores do que os meses do período do Inverno. Os testes que foram realizados nas mesorregiões Campinas, Bauru, Araraquara e Araçatuba, tiveram resultados diferentes para cada uma delas, sendo difícil especificar um comportamento comum no estado todo. Porém, a maioria dos testes de hipótese deram evidências para afirmarmos que os meses do Verão possuem precipitação médias maiores do que os meses do Inverno.

Devido à falta de tempo, não foi possível testar a influência das covariáveis na precipitação. Só foi possível ter uma análise exploratória separadamente de cada covariável, bem como analisar as correlações entre elas. A correlação entre a covariável Temperatura média com as covariáveis Temperatura mínima e Temperatura máxima foram muito altas. Para a construção de um modelo com covariáveis, é possível utilizar entre elas somente a Temperatura média.

Através da análise dos resíduos, foi possível concluir que o modelo não foi apropriado para os dados das mesorregiões Vale do Paraíba Paulista, Presidente Prudente, Ribeirão Preto, Litoral Sul Paulista, Macro Metropolitana Paulista, Marília e São José do Rio Preto. Provavelmente, através de modelos com mais covariáveis seja possível tirar outras conclusões sobre a precipitação média e fatores relacionados à ela.

Outra alternativa possível de ser explorada, é utilizar um modelo de regressão espacial, pois ele incorpora a questão da dependência espacial dos dados, podemos considerar a localização das estações.