

Luís Philipe C. Mendes¹, Rafael S. Erbisti²

¹Universidade Federal Fluminense, Niterói

²Instituto de Matemática e Estatística, Universidade Federal Fluminense

Introdução

- A exposição a poluentes atmosféricos, como dióxido de enxofre (SO₂), material particulado fino (PM_{2.5}) e ozônio (O₃), está associada a uma série de problemas de saúde, incluindo doenças respiratórias, cardiovasculares e até mesmo câncer.
- A compreensão sobre a dinâmica espacial e temporal desses poluentes é importante na execução de medidas proativas de controle da poluição.
- Particularmente, os modelos espaciais para dados georreferenciados são capazes de caracterizar a dinâmica da dependência espacial de desfechos medidos em distintos pontos no território, obtendo, assim, o comportamento da variável de interesse ao longo do espaço.
- Os modelos ajustados foram implementados no software R, a inferência sobre os parâmetros foi realizada sob a ótica bayesiana e a amostra da distribuição a posteriori gerada a partir dos métodos de aproximação de Monte Carlo via Cadeias de Markov (MCMC).

Objetivos

- Este trabalho utiliza modelos de geoestatística com objetivo de descrever o comportamento da concentração de PM_{2.5}.
- Serão comparados os modelos exponencial e cauchy quanto ajuste e previsão do poluente.

Metodologia

Área de estudo e dados

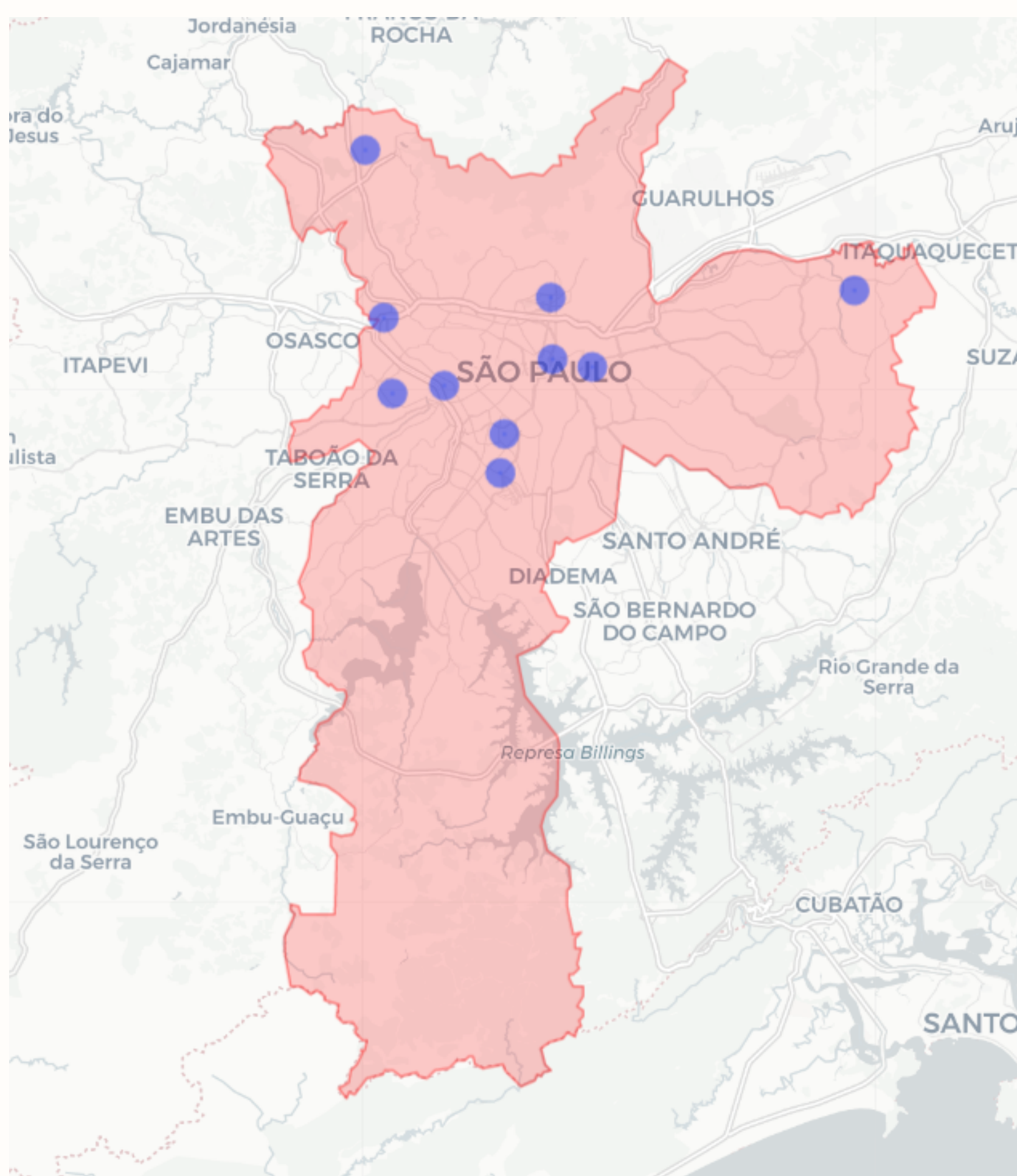


Figura 1. Cidade de São Paulo - SP, os pontos azuis representam as estações onde foram coletados os registros de PM_{2.5}

- Dados diários de 10 estações meteorológicas obtidos na CETESB, no ano de 2022.
- Dados agregados em 53 semanas do ano de 2022, em cada estação.

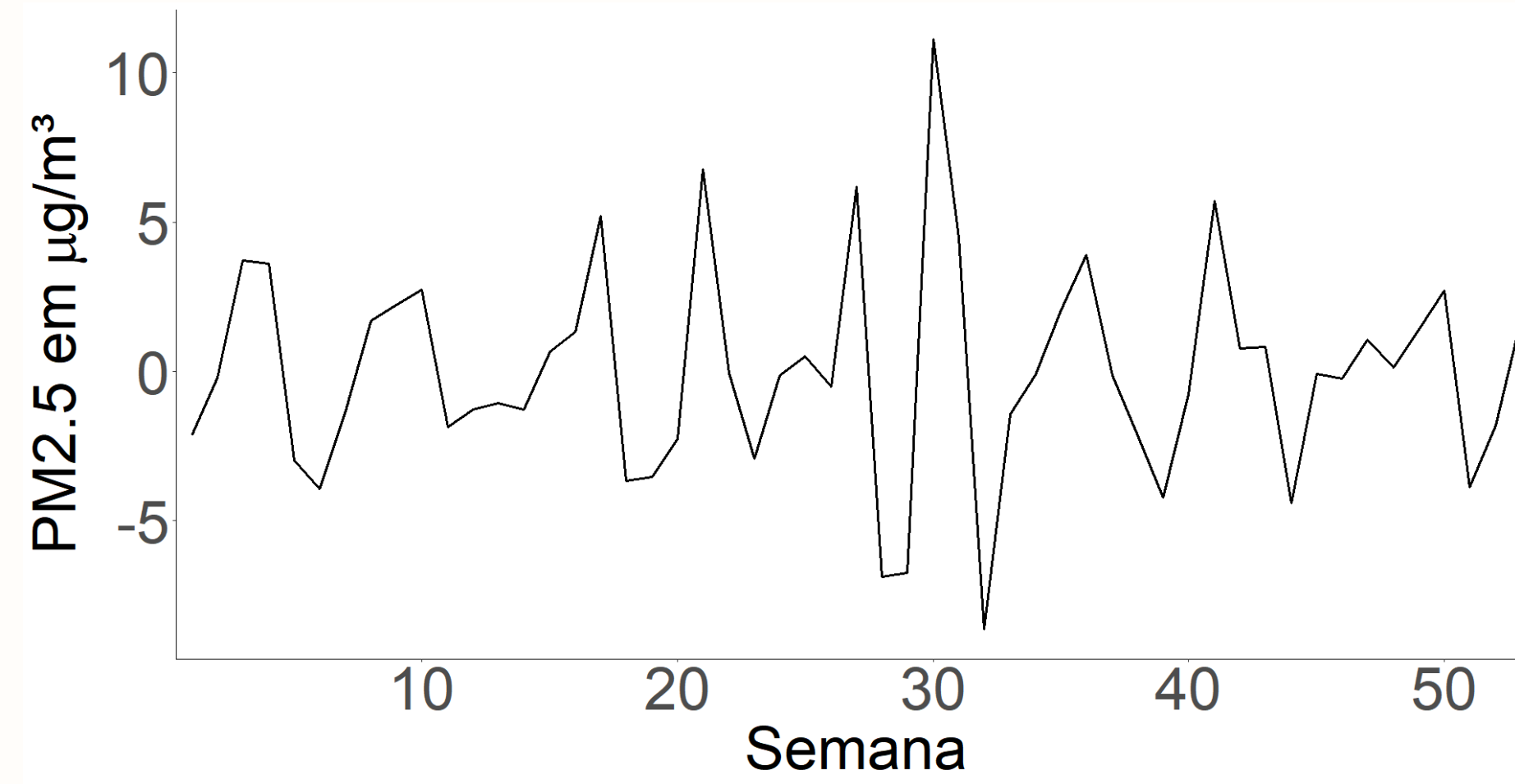


Figura 2. Série temporal da média, por semana, do PM_{2.5} subtraído da média do modelo com harmônicas, na cidade de São Paulo - SP em 2022.

- Utilizamos a média do modelo com harmônicas para retirar a sazonalidade dos dados.

Modelo espacial com réplicas independentes

Seja \mathbf{Y}_i o vetor da concentração de PM_{2.5} das 10 estações na i -ésima semana de 2022, $i = 1, \dots, 53$. Considere o modelo de geoestatística abaixo para representar o comportamento espacial do PM_{2.5} na cidade de São Paulo-SP:

$$\mathbf{Y}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i, \quad \varepsilon_i \sim \text{NMV}(\mathbf{0}, \Sigma)$$

- X_1 e X_2 são, respectivamente, as latitudes e longitudes das estações;
- β_0 , β_1 e β_2 são o intercepto e os coeficientes associados a latitude e longitude, respectivamente;
- ε é o termo de erro espacialmente estruturado, cuja matriz de covariância é dada por

$$\Sigma = \sigma^2 \rho(h)$$

com $\rho(h)$ sendo uma função de correlação espacial válida

Modelos de correlação espacial

Exponencial Cauchy

$$\rho(h) = \exp\{-h/\phi\} \quad \rho(h) = (1 + h/\phi)^{-1}$$

com h sendo a distância euclidiana entre as estações e (ϕ) o alcance espacial.

O procedimento de inferência foi feito sob o enfoque bayesiano e foi assumido independência a priori, isto é,

$$p(\beta, \sigma^2, \phi) = p(\beta)p(\sigma^2)p(\phi)$$

com

- $\beta = (\beta_0, \beta_1, \beta_2) \sim \text{NMV}(0, \mathbf{I})$, onde \mathbf{I} é a matriz identidade de dimensão 3×3 .
- $\sigma^2 \sim \text{GamaInversa}(1, 0.01)$.
- $\phi \sim \text{Gama}(1, \text{mediana}(h))$.

Resultados

A geração da amostra a posteriori foi feita a partir dos Métodos de Monte Carlo via Cadeias de Markov (MCMC), considerando:

- Tamanho da amostra a posteriori: 20 mil valores
- Burn-in*: 1000 valores
- Thinning* modelos exponencial e cauchy: 8 amostras

- Tamanho da amostra efetiva foi de 1900 valores para ambos modelos.

Abaixo está um gráfico com os resultados obtidos. Note que foram estimados a mediana e o intervalo de credibilidade de 95% (limite inferior e superior) calculado via *Highest Posterior Density Interval* (HPDI).

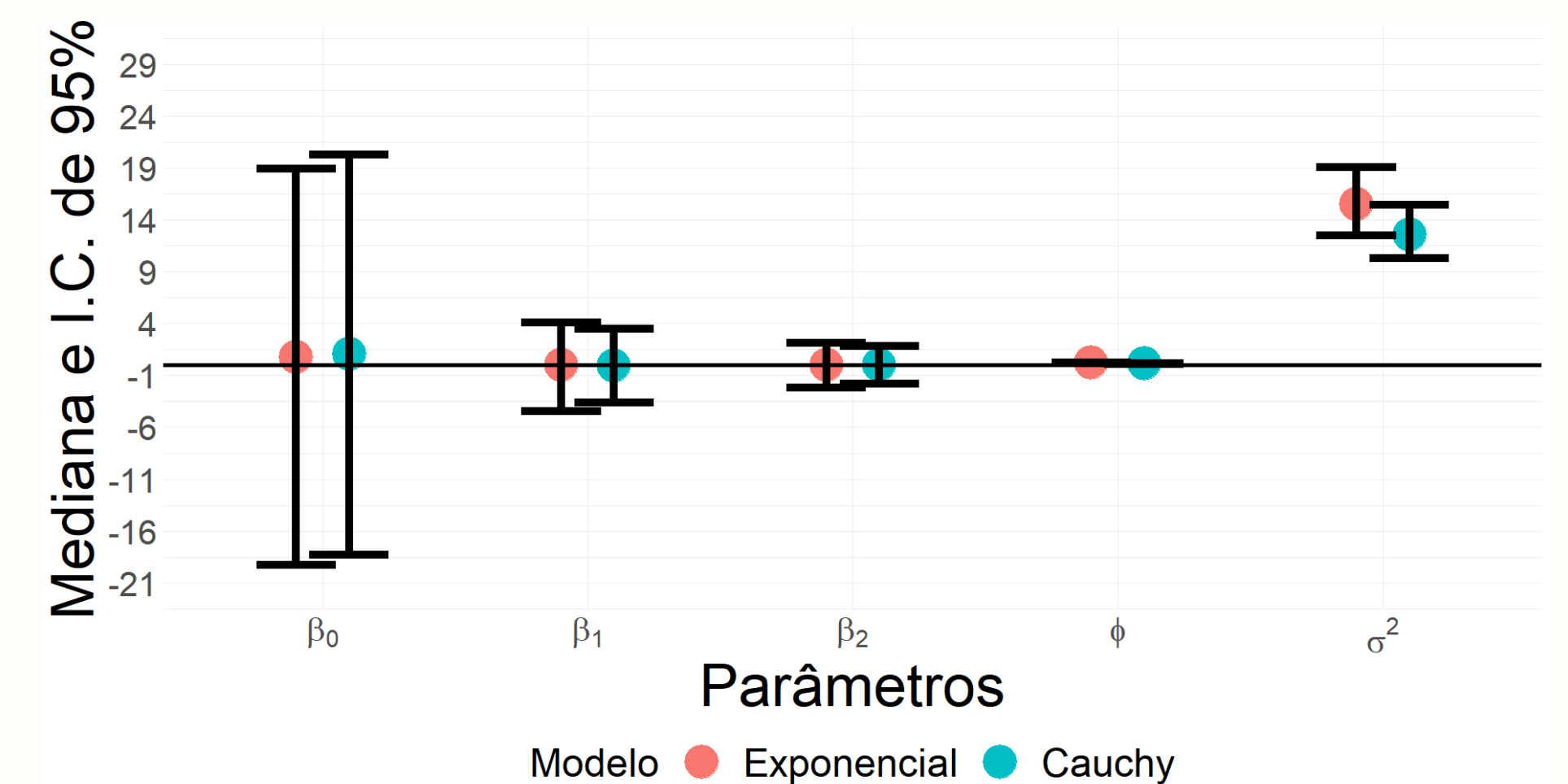


Figura 3. Mediana a posteriori e intervalos de credibilidade para os parâmetros estimados.

A figura abaixo apresenta as estimativas pontual e intervalar da curva de correlação espacial a posteriori para cada um dos modelos estimados.

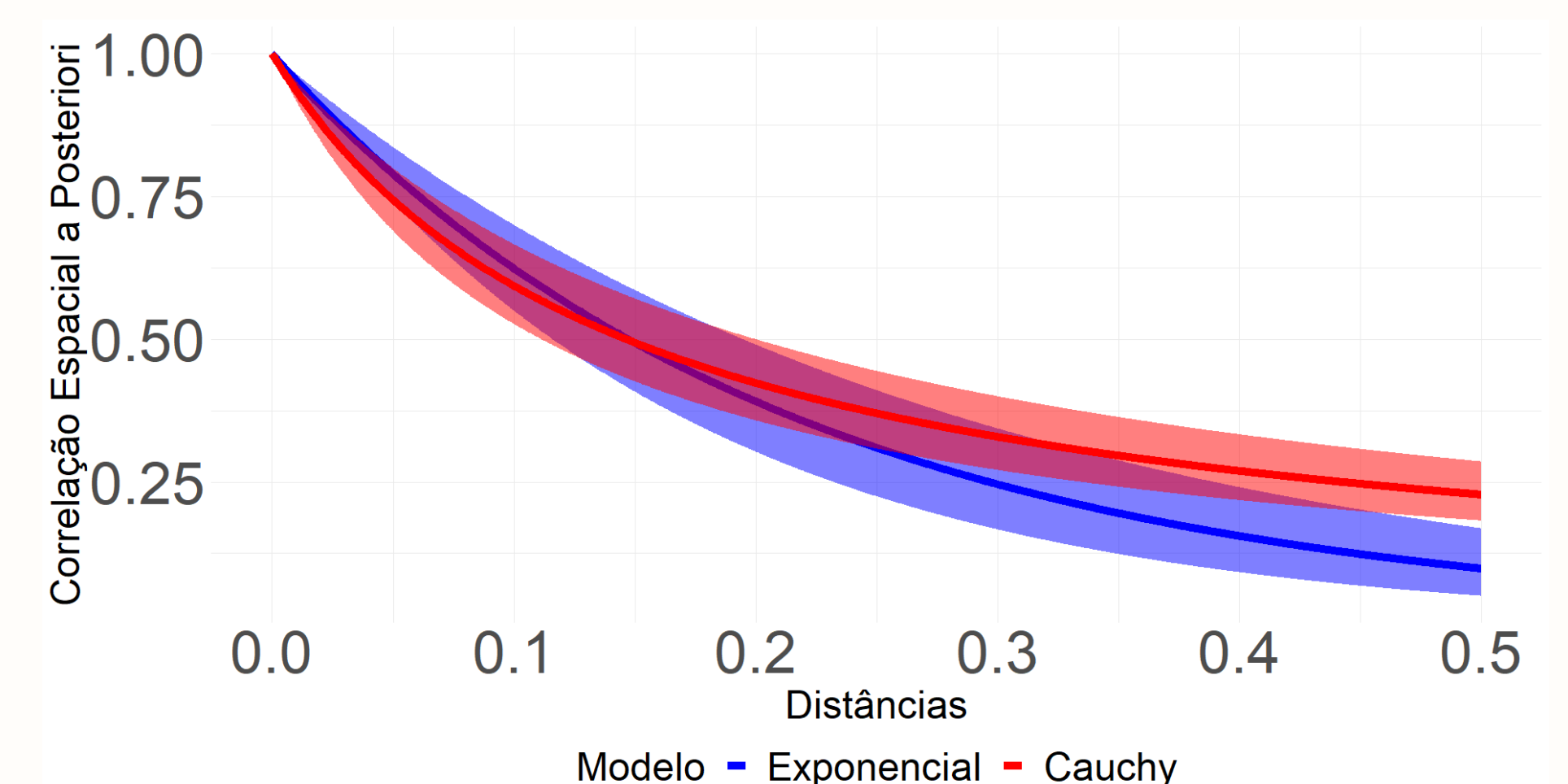


Figura 4. Estimativa dos modelos de correlação espacial juntamente com intervalo de credibilidade de 95%.

Já na tabela, segue a medida de desempenho dos modelos, mais especificamente o Critério de Informação de Akaike (AIC). Ela é uma medida de comparação de ajuste do modelo aos dados. É preferível que essa métrica seja a menor possível.

Tabela 1. Medidas de desempenho dos modelos.

Modelo	AIC
Exponencial	2471,973
Cauchy	2433,133

Conclusão

- Os dados envolveram 10 estações meteorológicas e 53 réplicas independentes, entretanto, uma limitação deste trabalho é não considerar a clara dependência temporal que há no dado na estrutura do modelo.
- Pela estimação dos parâmetros, não é possível verificar relação da latitude e longitude com a concentração do material particulado fino.
- Espera-se que esse feito varie ao longo do tempo e o modelo utilizado é limitado.
- Dentre os modelos estimados, o com função de correlação espacial cauchy foi o que apresentou a melhor medida de ajuste e a melhor capacidade de mensuração da correlação espacial do PM_{2.5}.