

# Análise do impacto da Amostragem Preferencial

**Andreia Monteiro**

*Universidade do Minho, andreiaforte50@gmail.com*

**Raquel Menezes**

*Universidade do Minho, rmenezes@math.uminho.pt*

**Maria Eduarda Silva**

*Faculdade de Economia, Universidade do Porto, mesilva@fep.up.pt*

**Palavras-chave:** Amostragem Preferencial, Geoestatística, Séries Temporais, Tempo Contínuo

**Resumo:** Em inúmeros contextos são encontradas aplicações de modelos espaço-temporais, nomeadamente na agricultura, ciências ambientais, ecologia, entre outros.

Tradicionalmente a modelação temporal e espacial assume que os locais de amostragem (no tempo e no espaço) são seleccionados independentemente dos valores do processo em estudo. Por exemplo, os modelos para dados ambientais não têm habitualmente em consideração o facto de que os locais escolhidos para os monitores de poluição podem depender das concentrações hipotéticas nesses locais. Este fenómeno, conhecido como amostragem preferencial, está presente em muitos estudos sempre que o processo associado às localizações dos dados e o processo associado aos próprios dados são estocasticamente dependentes.

Diggle et al.(2010) [?] consideraram processos de Cox log-Gaussianos para modelar a dependência estocástica entre os locais de amostragem e a variável espacial em estudo e demonstraram que ignorar a natureza preferencial da amostragem pode levar a estimativas enviesadas e inferências enganosas.

O nosso objetivo principal é estender o conceito de amostragem preferencial à componente temporal, uma vez que frequentemente os dados de monitorização têm tanto uma estrutura espacial, determinada pelos locais onde são recolhidos os dados, como uma estrutura temporal determinada pela frequência com que essas observações são recolhidas.

No presente trabalho, através de estudos de simulação e utilizando modelos de séries temporais em tempo contínuo, mostra-se que o desenho amostral tem um impacto significativo na estimação e predição. Consideram-se quatro cenários distintos para a recolha dos dados que compõem a nossa amostra, desde o desenho determinístico (tempos equiespaçados) ao desenho estocástico. Os tempos amostrados são considerados como provenientes de uma distribuição de Poisson homogéneo ou de uma distribuição condicionada aos valores máximos da variável de interesse, incluindo a situação de amostragem

preferencial

### **Referências**

- [1] Diggle, P.J., Menezes, R., Su, T. Geostatistical inference under preferential sampling. *Applied Statistics*, 59(2), 1–20, 2010.