

## LCE5700 - Geoestatística

**Aluna:** Izabela Regina Cardoso de Oliveira      **Nº USP:** 7417230      **Data:** 05/10/2011

**Resenha 2:** DE IACO, S., A new space-time multivariate approach for environmental data analysis. *Journal of Applied Statistics*, v.38, n.11, p. 2471-2483, 2011.

O artigo estudado trata de uma nova forma de análise de dados multivariados que possuem correlação espaço-temporal. A autora descreve como a técnica de análise de correlação canônica (CCA) pode ser combinada com métodos geoestatísticos espaço-temporais para analisar dois aspectos ambientais, quais sejam concentração de poluição no ar e condições meteorológicas. O uso de técnicas multivariadas na análise de dados com correlação espaço-temporal é, principalmente, devida à climatologistas. Porém, a proposta presente no artigo pode ser aplicada às diversas áreas do conhecimento.

As redes de monitoramento ambiental usualmente fornecem dados multivariados, coletados em diferentes estações e para um certo período de tempo. É fácil notar que tanto métodos multivariados quanto técnicas geoestatísticas espaço-temporais podem ser aplicados para interpretar e controlar a evolução das variáveis observadas.

De acordo com a autora, a co-krigagem é aplicada quando variáveis auxiliares podem melhorar a estimação de outras variáveis. A análise de krigagem fatorial tem sido desenvolvida em um contexto multivariado para determinar como a correlação espacial entre as variáveis muda de uma escala espacial para outra. Entretanto, até recentemente, essas ferramentas tem sido usadas apenas em um contexto espacial e não espaço-temporal.

A metodologia proposta no artigo é aplicada a um conjunto de dados ambientais, compostos por médias por hora de três poluentes ( $\text{NO}$ ,  $\text{NO}_2$  e  $\text{O}_3$ ) e três indicadores atmosféricos (temperatura, umidade e velocidade do vento), em janeiro e agosto de 2007, observados em várias estações de monitoramento localizadas no distrito de Milan e em sua vizinhança (Itália).

As variáveis  $\text{NO}$ ,  $\text{NO}_2$  e  $\text{O}_3$  representam uma ameaça real para a qualidade ambiental e as variáveis temperatura, umidade e velocidade do vento afetam a formação de nuvens de poluição, suas reações, estagnação e dispersão. Os meses de fevereiro e agosto estão entre os períodos mais críticos do ano em termos de concentração dessas nuvens.

Como um passo inicial do trabalho, várias estatísticas descritivas foram computadas. Os histogramas, matrizes de correlações e gráficos box-plots foram apresentados para as seis variáveis (3 contaminantes e 3 indicadores meteorológicos), separados nos dois meses. Na última etapa da análise exploratória foram apresentados mapas de contorno para as médias mensais das variáveis mensuradas. Os mapas foram obtidos usando o estimador de krigagem ordinário de

médias mensais, baseado nos modelos de variogramas espaciais das variáveis. Entretanto, as interações espaço-temporais entre poluentes e condições atmosféricas não podem ser ignoradas e uma nova metodologia de análise é sugerida.

Os variogramas amostrais das variáveis são a “ponte” entre as técnicas multivariadas e as geoestatísticas. O método geoestatístico multivariado conhecido como modelo de co-regionalização linear é baseado em combinações lineares de variogramas para descrever a estrutura multivariada dos dados, mas ele tem sido usado apenas em um contexto espacial. No trabalho a CCA é aplicada para relacionar o comportamento espaço-temporal (em termos de estrutura de correlação espaço-tempo) dos dois grupos de variáveis. As variáveis canônicas, que são combinações lineares de variogramas amostrais espaço-temporais, refletem as correlações subjacentes no espaço-tempo dos dois grupos de variáveis.

Para efetuar as análises, os dados de cada variável foram padronizados pela média mensal. Na fase de análise estrutural no espaço-tempo, foram criados dois vetores aleatórios, um associado às variáveis químicas (poluentes) e outro associado às variáveis meteorológicas. Cada componente do vetor é considerado uma variável aleatória intrínseca e sua correlação espaço-temporal é estimada computando o variograma espaço-temporal amostral. Nessa etapa, variogramas espaço-temporais são estimados para as variáveis pertencentes aos dois grupos sob estudo.

Na etapa seguinte a CCA é proposta com o objetivo de determinar relações simultâneas entre os variogramas amostrais dos dois grupos de variáveis. Sejam  $X$  e  $Y$  as matrizes de dados associadas, respectivamente, aos variogramas espaço-temporais amostrais para poluentes e aos variogramas espaço-temporais amostrais para medidas meteorológicas, computadas em diferentes distâncias espaciais e temporais. Cada matriz possui dimensão  $216 \times 3$ , cujo número de linhas é dado pelo número de distâncias espaço-temporais ( $3$  distâncias espaciais  $\times$   $72$  distâncias temporais).

Dada a matriz de covariâncias amostrais,  $\Sigma$ , de  $X$  e  $Y$ , a CCA é usada para encontrar  $\mathbf{a}$  e  $\mathbf{b}$  tal que a correlação entre as combinações lineares  $W = X\mathbf{a}$  e  $Q = Y\mathbf{b}$  seja maximizada. Dessa forma, a CCA é aplicada aos dois grupos de variogramas amostrais com o intuito de identificar uma relação simultânea entre eles e usar as variáveis canônicas mais importantes para descrever o comportamento espaço-temporal subjacente dos componentes dos dois grupos.

Nos resultados do trabalho são apresentados os variogramas espaço-temporais para as variáveis dos dois grupos, nos meses de agosto e setembro. Os resultados da CCA dos dois grupos de variogramas foram apresentados em uma tabela, separados por mês. Essa tabela contém os coeficientes canônicos e de correlação entre os grupos, considerando a primeira e segunda variáveis canônicas. Os resultados obtidos, tanto em relação à magnitude dos coeficientes quanto em

relação aos seus sinais, são coerentes com o fenômeno biológico estudado. Variogramas espaço-temporais foram construídos para as primeiras variáveis canônicas de cada grupo, nos dois meses. As novas superfícies fornecem uma visão simplificada do comportamento espaço-temporal das variáveis subjacentes.

Vale destacar que, além de recente e de possuir uma proposta bastante interessante, o artigo estudado foi muito bem escrito, de forma clara e detalhada. Dentre as sugestões para futuros estudos, a autora recomenda a aplicação de técnicas de co-krigagem e comparação do método proposto com os modelos de co-regionalização lineares no espaço-tempo.

Por fim, como já foi dito no início, a estrutura dos dados de poluentes e medidas meteorológicas, coletados nas estações de monitoramento (pontos fixados) em diferentes períodos de tempo, sugere uma análise geoestatística espaço-temporal. O processo subjacente é a associação entre os grupos de variáveis, baseada nas suas distribuições espaciais e temporais. Para conhecer como ocorre esse processo, o artigo propõe uma nova metodologia, utilizando análise de correlação canônica. Apresentar a metodologia proposta e sua eficiência são os principais objetivos científicos do trabalho. As variáveis resposta analisadas são contínuas, mas nem todas possuem distribuição normal. A covariável do estudo é o efeito de cada mês, que foi considerado na padronização das variáveis.