# MODELOS GAUSSIANOS GEOESTATÍSTICOS ESPAÇO-TEMPORAIS E APLICAÇÕES

Alexandre Sousa da SILVA[1]
Paulo Justiniano RIBEIRO JUNIOR[2]

▪ RESUMO: A especificação de funções de covariância espaço-temporais é uma das possíveis estratégias para modelagem de processos dos quais observações são tomadas em diferentes posições do espaço e do tempo. Tais funções podem definir processos separáveis ou não separáveis e na sua especificação deve-se garantir que são funções de covariância válidas atendendo a condição de serem positiva definidas. Entre estratégias para obtenção de tais funções estão as de Cressie e Huang (1999) e Gneiting (2002). A primeira se baseia na idéia de obter funções em um espaço de dimensão aumentada a partir de funções válidas no espaço original e necessita de operações no domínio da freqüência. Alternativamente a segunda proposta utiliza combinação de funções completamente monótonas e estritamente crescentes, evitando inversão de representações espectrais. Há ainda poucos relatos de uso e avaliações comparativas das diferentes propostas. Neste trabalho considerou-se a metodologia proposta por Gneiting, com diferentes valores do parâmetro que indica a força da interação entre o espaço e o tempo. Diferentes modelos foram aplicados à dois conjuntos de dados, um referente a estoques de peixe na costa de Portugal, e outro referente à armazenagem de água em um solo com citros. Utilizou-se a implementação no pacote *RandomFields* do programa R, revisando-se a metodologia e investigando-se a implementação computacional. Na aplicação o modelo de covariância separável se mostrou adequado para descrever o comportamento das observações disponíveis sendo a escolha do modelo determinada por ajustes de máxima verossimilhança.

▪ PALAVRAS-CHAVE: Campos Aleatórios; Geoestatística; Modelos Espaço-Temporias; Pacote *RandomFields*

[1]Departamento de Ciências Exatas, Universidade de São Paulo - ESALQ/USP, CEP 13418-900 Piracicaba, SP, Brasil. E-mails: *assilva@esalaq.usp.br*

[2]Departamento de Estatística, Universidade Federal do Paraná, CEP 81531-990 Curitiba, PR, Brasil. E-mails: *paulojus@ufpr.br*

## 1 Introdução

Dados espaço-temporais são caracterizados pela variabilidade no tempo e no espaço, e o objetivo da análise estatística deste tipo de dado está em descrever a incerteza não somente sobre as estimativas das quantidades de interesse mas, também, estimar valores em locais e/ou tempos não amostrados. Na literatura estatística existe um grande número de publicações referentes a processos puramente espaciais ou puramente temporais, e mais recentemente começam a ser apresentadas propostas para modelagem conjunta no espaço e tempo. Esta análise conjunta é o interesse dos modelos espaço-temporais.

Modelos espaço-temporais têm ganho uma crescente popularidade na última década, o que pode ser explicado por um lado pela numerosa aplicabilidade, principalmente em ciências ambientais e da saúde; e por outro lado pelo crescente desenvolvimento de recursos computacionais.

Gneiting e Schlather (2002), em uma revisão sobre estratégias de modelagem, consideram duas possíveis abordagens para a modelagem espaço-temporal: a abordagem geoestatística e a abordagem baseada em modelos.

Na abordagem geoestatística, geralmente considerada para campos aleatórios gaussianos, em que o processo é totalmente especificado pelo vetor de médias e a matriz de covariância. Em geral o vetor de médias é facilmente especificado a partir de informações contextuais, isto não acontece com a matriz de covariância, que é o componente chave desta metodologia. Entretanto, nesta abordagem, esta matriz possui elementos dados por uma função de covariância válida, que assegure a condição de matriz positiva definida.

Na abordagem baseada em modelos é enfatizada a adoção de modelos estocásticos e nesta metodologia a função de covariância não é a única estrutura utilizada para especificar o modelo, sendo assim, este método pode ir de uma função analítica simples a uma função intratável e somente implicitamente induzida pelo modelo adotado. Neste contexto incluí-se os modelos Bayesianos flexíveis, dinâmicos e/ou homogêneos e métodos baseados na técnica do filtro de Kalman.

Neste trabalho será considerado a abordagem geoestatística, para campos aleatórios espaço-temporais gaussianos. Uma forma intuitiva de produzir modelos válidos para a função de covariância espaço-temporal é através da combinação de funções válidas puramente espacial e puramente temporal. Uma propriedade assegura que o produto ou a soma de funções de covariâncias válidas tem como resultado uma função de covariância válida. As funções de covariância espaço-temporais obtidas desta forma são ditas separáveis já que não contemplam a possibilidade de interação entre o componente espacial e temporal. Por isso, em geral não são realísticas já que assumem a independência dos processos espaciais e temporais.

Alternativamente as funções de covariância espaço-temporais não separáveis consideram a interação entre o componente espacial e temporal, mas apresentam como grande desvantagem a dificuldade em se construir funções de covariância que sejam válidas. Métodos matemáticos sofisticados devem ser aplicados para garantir a validade destas funções e em geral se valem do teorema de Bochener (BOCHNER, 1959), que utiliza a representação espectral para garantir a validade das funções de covariância espaço-temporais. Nesta linha, Cressie e Huang (1999), Gneiting (2002) e Stein (2005), por exemplo, propõem famílias de modelos de covariância não separáveis válidas para processos espaço-temporais gaussianos.

## 2 Campos aleatórios

### 2.1 Campos aleatórios espaciais

Um campo aleatório ou uma função aleatória é um processo estocástico definido no espaço $G \subset \mathbf{R}^d$, ou seja, uma função cujos valores são realizações de variáveis aleatórias em qualquer ponto do domínio (SCHIMIDT; SANSÓ, 2006), ou em outras palavras, uma família ou coleção de variáveis aleatórias, em que cada um dos seus membros podem ser identificados ou localizados de acordo com a mesma métrica (SCHABENBERGER; GOTWAY, 2005). Um campo aleatório é definido como:

$$\{Z(\mathbf{s}) : \mathbf{s} \in G \subset \mathbf{R}^d\},$$

em que $Z(\mathbf{s})$ é o valor do atributo $Z$ na localização $\mathbf{s}$ e $d \geq 1$ é a dimensão do campo aleatório.

Segundo Schmidt e Sansó (2006) e Le e Zidek (2006), a descrição de um campo aleatório é obtida através das distribuições acumuladas finito-dimensionais $F$ se, para qualquer conjunto de pontos nas localizações $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n$ pertencentes à região $G$ e qualquer inteiro n,

$F_{\mathbf{s}_1,\mathbf{s}_2,...,\mathbf{s}_n}(z_1, z_2, ..., z_n) \equiv P\{Z(\mathbf{s}_1) \leq z_1, Z(\mathbf{s}_2) \leq z_2, ..., Z(\mathbf{s}_n) \leq z_n\}$.

Um campo aleatório gaussiano corresponde ao caso particular em que tais distribuições finito-dimensionais são gaussianas, ou seja, um campo aleatório é gaussiano se a distribuição para qualquer conjunto de índices $i = 1, 2, ..., n$ é uma gaussiana n-variada. Conseqüentemente, por propriedades da distribuição gaussiana multivariada, cada $Z(\mathbf{s}_i)$ é uma variável aleatória gaussiana univariada.

O processo gaussiano é particularmente importante na análise de campos aleatórios, já que estes são completamente especificados pelo vetor de médias e sua matriz de covariância (GNEITING; SCHLATHER, 2002). O vetor de médias é especificado pelo conhecimento de covariáveis, isto é, efeitos fixos, que influenciam a variável de interesse. A matriz de covariância precisa ser positiva definida para ser considerada válida, para tanto cada um de seus elementos devem ser dados por uma função de covariância que garanta a condição de positiva definida. Entretanto determinar se esta função é positiva definida é, em geral, não trivial. Uma forma freqüentemente utilizada para verificar a validade de uma função de covariância é verificar se a representação espectral desta função satisfaz o teorema de Bochner (BOCHNER, 1959).

### 2.1.1 Propriedades da função de covariância

A função de covariância $C(\mathbf{h})$ de um campo aleatório estacionário de segunda ordem deve satisfazer as seguintes propriedades :
(i) $Cov[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{0})] = Var[Z(\mathbf{s})] = C(\mathbf{0}) \geq 0$;
(ii) $C(\mathbf{h}) = C(\textbf{-h})$;
(iii) $C(\mathbf{0}) \geq \mid C(\mathbf{h}) \mid$;
(iv) $C(\mathbf{h}) = Cov[Z(\mathbf{s}), Z(\mathbf{s}\mathbf{+}\mathbf{h})] = Cov[Z(\mathbf{0}), Z(\mathbf{h})]$;
(v) Se $C_j(\mathbf{h})$ com $j = 1, 2, ..., k$, são funções de covariância válidas, então $\sum_{j=1}^{k} \mathbf{b}_j C_j(\mathbf{h})$ é uma função de covariância válida, se $\mathbf{b}_j \geq 0 \forall j$;
(vi) Se $C_j(\mathbf{h})$ com $j = 1, 2, ..., k$, são funções de covariância válidas, então $\prod_{j=1}^{k} C_j(\mathbf{h})$ é uma função de covariância válida;
(vii) Se $C(\mathbf{h})$ é uma função válida no $R^d$, então ela também será uma função de covariância válida em $R^p$, com $p < d$.

De acordo com Schabenberger e Gotway (2005), as propriedades (i) e (ii) são imediatas, a (iii) segue a forma da inequação de Cauchy-Schwarz, a (iv) caracteriza a falta de importância da coordenada absoluta, (v) assegura que combinações lineares de funções de covariância válidas são também funções de covariância válidas e (vi) assegura que o produto de funções de covariância válidas tem como resultado uma função de covariância válida.

Em campos aleatórios espaço-temporais a propriedade (ii) nem sempre é satisfeita, o que dá origem aos modelos de covariância espaço-temporais assimétricos. Já as propriedades (v) e (vi) dão suporte à construção de modelos de covariância espaço-temporais separáveis.

A matriz de covariância para $(Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n))$, denotada por $\Sigma$ com dimensão $n \times n$, será uma matriz simétrica, em que $\Sigma_{ij} = C(\mathbf{s}_i - \mathbf{s}_j)$, ou seja, cada valor de $\Sigma$ é igual a covariância entre respostas em duas localizações correspondentes. Para que uma função de covariância de um campo aleatório estacionário seja considerada válida é necessário e suficiente que $C$ satisfaça a condição de ser positiva definida, em outras palavras, para qualquer vetor não zero $\mathbf{a}$, a forma quadrática $\mathbf{a}'\Sigma\mathbf{a}$ deve ser maior igual a zero. Sendo assim esta condição pode ser escrita nas seguintes formas:

$$\mathbf{a}'\Sigma\mathbf{a} = \sum_{i=1}^{k}\sum_{j=1}^{k} \mathbf{a}_i\mathbf{a}_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0 = \sum_{i=1}^{k}\sum_{j=1}^{k} \mathbf{a}_i\mathbf{a}_j C(\mathbf{h}_{ij}) \geq 0,$$

em que $\mathbf{a}_i$ e $\mathbf{a}_j$ são elementos do vetor $\mathbf{a}$ e $\mathbf{h}_{ij}$ denota o vetor de separação entre $\mathbf{s}_i$ e $\mathbf{s}_j$.

Uma ferramenta de uso comum para investigação da dependência espacial em geoestatística é o semivariograma. Para campos aleatórios intrinsecamente estacionários define-se $2\gamma(\mathbf{h}) = E\{[\mathbf{Z}(s_i) - \mathbf{Z}(s_j)]^2\}$. Portanto, uma estimativa empírica do semivariograma é dada pelo estimador de momentos:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} \{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2,$$

em que $|N(\mathbf{h})|$ é a quantidade de pontos separados pela distância $\mathbf{h}$ de separação entre $s_i$ e $s_j$. De acordo com Schabenberger e Gotway (2005), o estimador $\hat{\gamma}(\mathbf{h})$ é não viesado para $\gamma(\mathbf{h})$ se $Z(\mathbf{s})$ for intrinsecamente estacionário. No caso do processo em que assume-se a estacionariedade de segunda ordem, o variograma informa também sobre a função de correlação.

Isto justifica uma prática comum em geoestatística de se utilizar o semivariograma empírico para inferir sobre parâmetros do modelo, através do ajuste de uma função válida para o semivariograma teórico.

Um outro paradigma para estimação de parâmetros, sob a pressuposição de estacionariedade forte, é dado pelo método de máxima verossimilhança. A estimação dos parâmetros de campos aleatórios pelo método de máxima verossimilhança requer pressuposições sobre a distribuição do processo. Dado $\mathbf{Z} = [Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n)]'$, que denota o vetor de observações e assumindo $\mathbf{Z}(\mathbf{s}) \sim G(\mu\mathbf{1}, \Sigma(\theta))$, o método da máxima verossimilhança consiste em maximizar o logaritmo da distribuição gaussiana n-variada:

$$L(\theta; Z_1, Z_2, ..., Z_n) = -\frac{1}{2}\{\ln(|\Sigma(\theta)|) + n\ln(2\pi) + (Z(\mathbf{s}) - \mathbf{1}\mu)'\Sigma(\theta)^{-1}(Z(\mathbf{s}) - \mathbf{1}\mu)\}. \tag{1}$$

Estimadores de máxima verossimilhança são muito utilizados em estatística, por assegurarem uma série de propriedades, já que sobre condição de regularidade padrão estes estimadores são assintoticamente gaussianos e eficientes. Na prática, em geoestatística, este método de estimação pode requerer muito tempo computacional, ou até mesmo ser proibitivo, devido à dimensão da matriz de covariância.

Diggle, Ribeiro e Christensen (2003) e Diggle e Ribeiro (2006) discutem em detalhes métodos baseados na verossimilhança para estimação de parâmetros do processos geoestatísticos espaciais, incluindo métodos bayesianos.

### 2.2 Campos aleatórios Espaço-temporais

Processos ambientais e geofísicos como concentração de poluentes, precipitação e superfície do vento são exemplos de fenômenos que podem ser modelados por campos aleatórios espaço-temporais ou simplesmente processos espaço-temporais. Este tipo de processo é caracterizado pela variabilidade na dimensão do espaço e do tempo. De acordo com Schabenberger e Gotway (2005), para o estudo desta variabilidade existem três possibilidades:

- análise espacial para cada processo temporal;
- análise temporal para cada processo espacial;
- análise espacial e temporal conjunta.

As duas primeiras possibilidades isolam a parte temporal ou a espacial e aplica técnicas padrões para o tipo de processo resultante. A terceira possibilidade considera o processo espacial e temporal conjuntamente e é a considerada neste trabalho. No processo de construção desta análise conjunta, a aplicação das técnicas padrões no estudo da variabilidade espacial e temporal separadamente pode ser utilizada como uma ferramenta exploratória e auxiliar na construção de modelos adequados.

Um campo aleatório espaço-temporal é definido como

$$\{Z(\mathbf{s}, t), \mathbf{s} \in R^d, t \in R\},$$

em que $Z(\mathbf{s}, t)$ é valor de atributo $Z$ no espaço $\mathbf{s} \in R^d$ e no tempo $t \in R$.

Através desta definição, percebe-se intuitivamente que o domínio natural do processo é $R^d \times R$. Neste trabalho o componente espacial será considerado bidimensional, ou seja, $d = 2$, por ser a situação comum na prática, entretanto ressalta-se que a dimensão do processo pode ser qualquer número finito positivo.

Tempo e espaço podem não ser diretamente comparáveis, já que as unidades das coordenadas dos dois processos apresentam grandezas diferentes. Fisicamente existe uma clara diferença entre as dimensões espaciais e temporais e os modelos precisam considerá-las. Segundo Gneiting et. al (2006) funções de covariância ou mesmo a krigagem se utilizam de espaços Euclidianos e são diretamente aplicados à problemas espaço-temporais. Sendo assim a simples separação vetorial do domínio espacial e temporal tem importantes implicações.

De acordo com Gneiting e Schlather (2002), na modelagem de processos espaço-temporais duas formas para especificação do modelo podem ser consideradas:

• *Especificação geoestatística.* Nesta metodologia é necessário o conhecimento de uma distribuição, geralmente gaussiana, para função aleatória $Z(\mathbf{s}, t)$ que é definida em toda coordenada espaço-temporal. A função de covariância é definida na dimensão espaço-temporal contínua, caracterizando o modelo da função aleatória, que pode ser utilizada na predição de qualquer localização espacial e qualquer instante de tempo. O ajuste da função de covariância é de importância central nesta metodologia, e portanto expressões com forma fechada para função de covariância são essenciais.

• *Especificação baseada em modelos.* Esta técnica enfatiza a adoção de modelos estocásticos para solução de problemas práticos nos quais a função de covariância não é especificada explicitamente, mas sim induzida pelo modelo. Esta forma de especificação espaço-temporal pode entretanto ir de uma função com forma analítica simples até uma intratável e somente implicitamente definida. Esta metodologia engloba desde modelos hierárquicos bayesianos espaço-temporais até a técnica do filtro de Kalman, com a característica de apresentar predições computacionalmente eficientes.

A especificação geoestatística é conceitualmente simples, particularmente quando combinado com a adoção do modelo gaussiano, que é completamente especificado por sua média e estrutura de covariância. Além disto, a krigagem, que geralmente é o objetivo da análise de campos aleatórios, requer a especificação apropriada da função de covariância. Entretanto esta forma de especificação assume estacionariedade espaço-temporal da estrutura de covariância, que portanto não pode assumir formas flexíveis como as induzidas pela especificação baseada em modelos. Este trabalho será direcionado à especificação geoestatística aplicada a processos espaço-temporais gaussianos.

Assumindo as condições de regularidade do campo aleatório espaço-temporal, $Var(Z(\mathbf{s},t)) < \infty$ para todo $\mathbf{s} \in \mathbf{R}^2, t \geq 0$, (CRESSIE; HUANG, 1999), define-se a média e a função de covariância por:

$$\mu(\mathbf{s},t) = E(Z(\mathbf{s},t))$$

$$Cov(Z(\mathbf{s}_1,t_1), Z(\mathbf{s}_2,t_2)) = C(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2).$$

De acordo com Gneiting et.al (2006), na prática o interesse inicial das análises é verificar a presença ou ausência de três características: estacionariedade, simetria completa e separabilidade. Após este estudo será possível decidir sobre a complexidade da modelagem.

Um campo aleatório espaço-temporal é estacionário se sua esperança não depende das coordenadas espaço-temporais e sua função de covariância só dependa de um vetor de separação dos pontos em $R^d \times R$. Portanto um campo aleatório espaço-temporal apresenta covariância estacionária no espaço se $Cov(Z(\mathbf{s}_1,t_1,\mathbf{s}_2,t_2))$ depende somente do vetor de separação $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$, e apresenta estacionariedade temporal se $Cov(Z(\mathbf{s}_1,t_1,\mathbf{s}_2,t_2))$ depende somente do vetor de separação $u = t_1 - t_2$. Desta forma se o campo aleatório espaço-temporal possui covariância estacionária espacial e temporalmente, este é considerada uma covariância espaço-temporal estacionária dada por:

$$Cov(Z(\mathbf{s}_1,t_1), Z(\mathbf{s}_2,t_2)) = c(\mathbf{h},u)$$

O campo aleatório espaço-temporal apresenta simetria completa se

$$Cov(Z(\mathbf{s}_1,t_1), Z(\mathbf{s}_2,t_2)) = Cov(Z(\mathbf{s}_1,t_2), Z(\mathbf{s}_2,t_1)),$$

para todas as coordenadas espaço-temporais $\mathbf{s}_1, t_1$ e $\mathbf{s}_2, t_2$ em $R^2 \times R$. No caso da covariância ser estacionária e completamente simétrica tem-se:

$$C(\mathbf{h},u) = C(\mathbf{h},-u) = C(-\mathbf{h},u) = C(-\mathbf{h},-u),$$

para todo $(\mathbf{h},u) \in R^2 \times R$.

A classe mais simples de funções de covariância espaço-temporal é dada pelos modelos separáveis. Funções de covariância espaço-temporais separáveis podem ser definidas com base nas propriedades de aditividade e multiplicabilidade (2.1.1, vi e vii). Desta forma poderão ser decompostas entre uma função de covariância puramente espacial e outra puramente temporal, que no caso aditivo pode ser escrita como:

$$Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = Cov(Z(\mathbf{s}_1, \mathbf{s}_2)) + Cov(Z(t_1, t_2))$$

e no caso multiplicativo por:

$$Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = Cov(Z(\mathbf{s}_1, \mathbf{s}_2))Cov(Z(t_1, t_2)), \tag{2}$$

em que em ambos os casos $\mathbf{s}_1, t_1$ e $\mathbf{s}_2, t_2 \in \mathbf{R}^2 \times \mathbf{R}$ .

Esta classe de modelos de covariância espaço-temporal não considera a interação entre o espaço e o tempo, mas é muito utilizada na prática por ser computacionalmente tratável e uma forma intuitiva e simples de obtenção de funções de covariância válidas.

Todos os modelos separáveis são também completamente simétricos. Para a demonstração desta relação considere a eq.(2), considere também as variáveis aleatória espaço-temporais $Z(\mathbf{s}_1, t_2)$ e $Z(\mathbf{s}_2, t_1)$, a função de covariância separável será dada por:

$$Cov(Z(\mathbf{s}_1, t_2), Z(\mathbf{s}_2, t_1)) = Cov(Z(\mathbf{s}_1, \mathbf{s}_2))Cov(Z(t_2, t_1)),$$

desta forma concluí-se que:

$$Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = Cov(Z(\mathbf{s}_1, t_2), Z(\mathbf{s}_2, t_1)),$$

que é a definição da propriedade de campos aleatórios espaço-temporais completamente simétricos.

Um exemplo simples de função de covariância espaço-temporal separável dado pela combinação de modelos de covariância exponenciais é:

$$Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = \sigma_1^2 exp(-A_1||(\mathbf{s}_1 - \mathbf{s}_2)||) + \sigma_2^2 exp(-A_2||(t_1 - t_2)||),$$

em que $A_1$ e $A_2$ são as denominadas *matrizes de anisotropia*.

As matrizes de anisotropia são utilizadas para permitir a representação combinada das dimensões espaciais e temporais. Neste exemplo, suponha um processo isotrópico no espaço tem-se:

$$A_1 = \begin{pmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \phi_2 \end{pmatrix}$$

Qualquer função de covariância espaço-temporal que não for escrita nas duas formas separáveis anteriormente mencionadas e nem combinações destas, serão ditas funções não separáveis. As funções de covariância separáveis, embora facilmente obtidas, são em geral insatisfatórias para descrever processos naturais. O que gera à necessidade de se especificar funções não separáveis. Kryriakidis e Journel (1999), revisam e discutem abordagens para especificação de processos geoestatísticos espaço-temporais. Cressie e Huang (1999), propõem classes de funções de covariância não separáveis válidas cuja obtenção não se baseiam no domínio das observações, mas sim no domínio da freqüência, uma vez que a representação espectral e a densidade espectral podem ser utilizadas para descrever as propriedades do processo. Gneiting (2002), em uma abordagem alternativa, obtém modelos válidos de função de covariância através da definição de funções monótonas e positivas.

### 2.2.1 Representação de Gneiting

A representação de Cressie e Huang (1999) constrói uma família de funções de covariância espaço-temporais estacionária e não separáveis através da inversão de Fourier na forma fechada, o que restringe esta família a um pequeno grupo de funções com soluções de forma fechada da inversão de Fourier. Gneiting (2002) propõe uma flexível e elegante família de funções de covariância espaço-temporais, cuja obtenção não requer operações no domínio espectral. A construção de funções de covariância válidas se dá através de componentes elementares em que sua validade é facilmente verificada.

A representação de Gneiting (2002) considerara qualquer função monótona $\phi(x)$, definida em $x \geq 0$ e qualquer função positiva $\psi(x)$, definida em $x \geq 0$ com derivadas completamente monótonas , e desta forma,

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(\psi(u)^2)^{d/2}} \phi\Big(\frac{||\mathbf{h}||^2}{\psi(|u|^2)}\Big) \tag{3}$$

é uma função de covariância espaço-temporal válida em $R^d \times R$, em que $\mathbf{h} = ||h||$, $u$ representa o vetor de distâncias no tempo e $d$ deve ser maior ou igual que a genuína dimensão espacial.

De acordo com Gneiting (2002) uma função contínua $\phi(x)$ é completamente monótona se possui derivadas $\phi^{(n)}$ de todos as ordens e $(-1)^n \phi^{(n)}(x) \geq 0$, em que $x > 0, n = 0, 1, 2, \ldots$.

Assim, Gneiting propõe uma família de funções de covariância espaço-temporal muito geral que não dependem da inversão de Fourier na forma fechada e não necessita de integrabilidade. De acordo com Elmatzoglou (2006) nesta família os componentes $\phi(x)$ e $\psi(x)$ podem ser associados com a estrutura espacial e temporal respectivamente. Gneiting (2002), apresenta possíveis escolhas para cada função, que são reproduzidas nas Tabelas 1 e 2.

Tabela - 1: Funções completamente monótonas $\phi(x)$,$x \geq 0$

| |
|---|
| $\phi(x) = exp(-cx^{\gamma}), c > 0, 0 < \gamma \leq 1$ |
| $\phi(x) = (1 + cx^{\gamma})^{\nu}, c > 0, 0 < \gamma \leq 1, \nu > 0$ |
| $\phi(x) = (2^{\nu-1}\Gamma(\nu))^{-1}(cx^{1/2})^{\nu}\mathbf{K}_{\nu}(cx^{1/2}), c > 0, \nu > 0$ |
| $\phi(x) = 2^{\nu}(exp(cx^{1/2}) + exp(-cx^{1/2}))^{\nu}, c > 0, \nu > 0$ |

Tabela - 2: Funções positivas $\psi(x)$, $x \geq 0$

| |
|---|
| $\psi(x) = (ax^{\alpha} + 1)^{\beta}, a > 0, 0 < \alpha \leq 1, 0 \leq \beta \leq 1$ |
| $\psi(x) = ln(ax^{\alpha} + b)/ln(b), a > 0, b > 1, 0 < \alpha \leq 1$ |
| $\psi(x) = (ax^{\alpha} + b)/(b(ax^{\alpha} + 1)), a > 0, 0 < b \leq 1$ |

Um exemplo simples da família de covariância espaço-temporal não separável proposta por Gneiting é dada como segue:

Considere a primeira linha das Tabelas 1 e 2, em que $\phi(x) = exp(-cx^{\gamma}), c > 0, 0 < \gamma \leq 1$ e $\psi(x) = (ax^{\alpha} + 1)^{\beta}, a > 0, 0 < \alpha \leq 1, 0 \leq \beta \leq 1$ são funções completamente monótona e positiva respectivamente. Substituindo na eq.(3) tem-se:

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|t|^{2\alpha} + 1)^{\frac{\beta d}{2}}} exp\Big\{\frac{c||\mathbf{h}||^{2\gamma}}{(a|t|^{2\alpha} + 1)^{\beta\gamma}}\Big\}. \quad (4)$$

Nesta expressão, nota-se que para $\beta = 0$ a covariância não depende do tempo. Multiplicando-se a eq.(4) por uma função de covariância puramente temporal, $C(|t|) = (a|t|^{\alpha} + 1)^{-\delta}$, tem-se como resultado:

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|t|^{2\alpha} + 1)^{\delta + \frac{\beta d}{2}}} exp\Big\{\frac{c||\mathbf{h}||^{2\gamma}}{(a|t|^{2\alpha} + 1)^{\beta\gamma}}\Big\},$$

no caso em que $\beta = 0$, a expressão da função de covariância se reduz a

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|t|^{2\alpha} + 1)^{\delta}} exp\{c||\mathbf{h}||^{2\gamma}\}$$

que é função de covariância separável na forma,

$$C(\mathbf{h}, u) = \sigma^2 C(u) C(\mathbf{h})$$

em que $C(u) = (a|t|^{2\alpha} + 1)^{-\delta}$ é a função puramente temporal, enquanto $C(\mathbf{h}) = exp\{c||\mathbf{h}||^{2\gamma}\}$ é a função puramente espacial. Esta representação sugere que o parâmetro $\beta$ pode ser utilizado para testar a hipótese de separabilidade.

## 3 Resultados e discussão

Muitos modelos tem sido sugeridos na literatura e o fato de terem sido adotados não significa necessariamente que estes representem melhor a realidade. A escolha de uma determinada classe de modelos é uma tarefa difícil e que requer um profundo conhecimento da realidade, por parte dos pesquisadores.

Neste trabalho, o interesse não é ajustar o melhor modelo e sim verificar o comportamento de alguns dos modelos discutidos no capítulo anterior. Para tanto, será considerado um conjunto referente a capacidade de armazenagem de água em um solo com citros.

A água é um dos mais importantes fatores para o adequado desenvolvimento de qualquer cultura agrícola e o conhecimento da sua dinâmica no solo é indispensável para o aprimoramento de práticas de manejo que visam à otimização da produtividade da cultura (MORETI, 2006).

Moreti (2006) fornece um estudo da capacidade de armazenagem de água em um Latossolo Vermelho Amarelo Argissólico cultivado com citros, em uma parcela experimental na qual foram coletados dados em duas transeções com 20 pontos de observação cada, com espaçamento de 4,0 m entre os pontos amostrados e 7,0 m entre as transeções. Para quantificar a armazenagem de água no solo foi instalado, em cada ponto amostral, um tubo de acesso à sonda de nêutros até a profundidade de 1,20 m, localizado no centro da distância entre duas plantas ao longo da linha.

A cultura de citros foi implantada em março de 1991, e as amostras foram coletadas de 2001 a 2004, totalizando 98 coletas de dados não eqüidistantes no tempo. Neste trabalho utilizou-se dados de 25 momentos de tempo tomados um em cada quatro, aos quais ajustou-se modelos espaço-temporais. A restrição de tomar apenas 25 dentre os 98 períodos de tempo visa contornar problemas computacioanais. Uma vez que o objetivo aqui é investigar a metodologia de modelagem e não fornecer resultados definitivos de análises destes dados.

### 3.0.2 Construindo modelos de covariância espaço-temporais

Tem-se a seguinte função de covariância espaço-temporal separável ajustada, com os parâmetros estimados pelo método da máxima verossimilhança:

$$C(\mathbf{h},u) = \frac{44.12}{(0.76|u|^{1.65}+1)^{0.08}} exp\{-0.001||\mathbf{h}||^{0.85}\} + \frac{0.25}{(0.76|u|^{1.65}+1)^{0.08}}. \tag{5}$$

Já para a função de covariância espaço-temporal não separável, com os parâmetros também estimados pelo método de máxima verossimilhança, tem-se que o modelo ajustado é dado por:

$$C(\mathbf{h},u) = \frac{1}{(0.76|u|^{1.65}+1)^{0.12}}\Big(\frac{44.12}{(0.76|u|^{3.3}+1)^{0.04}}$$
$$exp\Big\{-0.001\Big[\frac{||\mathbf{h}||}{(0.76|u|^{3.3}+1)^{0.04}}\Big]^{1.65}\Big\}\Big) + \frac{0.25}{(0.76|u|^{1.65}+1)^{0.08}}. \tag{6}$$

A distinção entre os dois modelos ajustados anteriormente é dada pelo parâmetro $\beta$ que determina se existe ou não interação entre os componentes puramente espacial e puramente temporal, ou seja, se a correlação espacial independe do tempo e vice-versa. Então o valor do parâmetro $\beta$ pode ser visto como o grau de interação espaço-temporal, em que valores grandes, próximos de 1, induz em um forte efeito de interação espaço-temporal, enquanto que valores pequenos, próximo de 0, induz em um efeito fraco de interação espaço-temporal. Isto sugere uma forma de testar a separabilidade da função de covariância, além de quantificar a intensidade da interação espaço-temporal.

O valor estimado para o parâmetro $\beta = 0.04$ é muito próximo de zero, o que indica que o modelos separável se ajustaria bem aos dados, em outras palavras, não há evidência clara de interação entre o espaço e o tempo. Ressaltando aqui novamente que a estimação deste parâmetro pode ser numericamente instáveis e em uma análise detalhada destes dados, métodos numéricos devem ser cuidadosamente calibrados.

Para verificar o efeito da suposição de separabilidade espaço-temporal foi considerado também o erro médio de predição, para tanto foi realizada a krigagem simples para diferentes valores de $\beta$. Os valores preditos para o tempo 26 foram comparados com os verdadeiros valores observados para este tempo. A Figura 1, ilustra estas predições e por este gráfico percebe que o modelo com $\beta = 0$ é o que mais se aproxima do valor real e ainda percebe-se que o aumento do valor de $\beta$, causa um diminuição na média e na variância. A Tabela 3 apresenta os erros quadráticos médios, a média e variância da predição, e através dela é possível verificar que o menor valor do erro quadrático médio ocorre quando $\beta = 0$, ou seja, o modelo separável é o mais indicado para o conjunto de dados considerado.

Na Figura 2, são ilustrados os mapas de krigagem simples com parâmetros $\beta = 0, 0.25, 0.95$. As predições foram realizados para o tempo 26, que não foi considerado na estimação. Estas imagens ilustram uma pequena diferença no comportamento das "manchas", mas uma diferença um pouco maior nas escalas. Através das legendas percebe a clara diminuição do valor médio considerado.

Tabela - 3: Erro quadrático médio de predição para funções de covariância espaço-temporais com diferentes valores do parâmetro $\beta$

| | EQM | Média | Variância |
|---|---|---|---|
| Modelo separável $\beta = 0$ | 5.561 | 17.114 | 0.753 |
| Modelo não separável $\beta = 0.25$ | 12.674 | 15.913 | 0.494 |
| Modelo não separável $\beta = 0.5$ | 26.060 | 14.371 | 0.309 |
| Modelo não separável $\beta = 0.75$ | 45.661 | 12.720 | 0.191 |
| Modelo não separável $\beta = 0.95$ | 64.436 | 11.451 | 0.132 |

## Considerações Finais

Processos ambientais e geofísicos como concentração de poluentes, campos de precipitação e superfície do vento são caracterizados pela variabilidade contínua no espaço e no tempo. Tais campos espaço-temporais são geralmente tratados como aleatórios, uma vez que as leis que governam seu comportamento não são completamente compreendidas e sua complexidade não segue uma descrição determinística precisa. Modelos estocásticos, por sua vez, são baseados tipicamente em um pequeno número de parâmetros que podem ser modelados e inferidos a partir de observações parciais dos processos.

O objetivo da modelagem de tais processos é informar sobre padrões do comportamento espaço-temporal. Parte-se do princípio de que não existe um modelo que se ajuste perfeitamente aos dados e o que ocorre, é que alguns

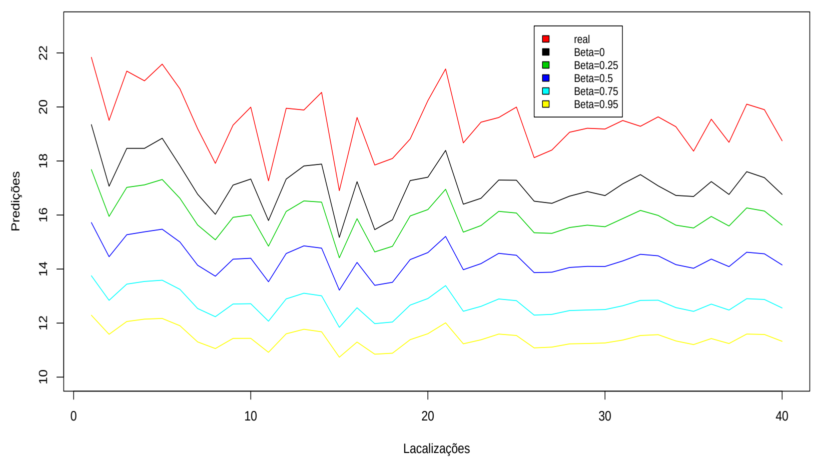


Figura - 1: Análise da anisotropia geométrica pela forma elíptica

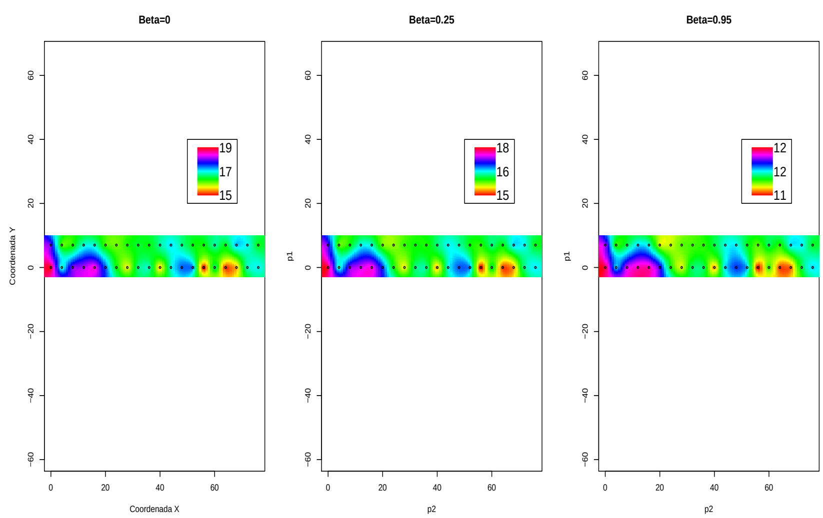


Figura - 2: Análise da anisotropia geométrica pela forma elíptica

modelos são mais capazes, que outros, de descrever certas situações. Através da modelagem é possível resumir de forma simples um conjunto de dados complexo, testar hipóteses cientificamente relevantes e fazer predições dos processos em posições não observadas do espaço e/ou tempo.

Fenômenos espaciais e temporais são caracterizados por suas unicidade e não reproducibilidade. Assumindo-se então que é uma amostra parcial simples da realização de um particular mecanismo estocástico. Conseqüentemente, inferências só podem ser feitas a partir da adoção de simplificações e suposições. A princípio a estratégia de análise espaço-temporal apresentada aqui poderiam ser utilizadas simplesmente como uma extensão da análise puramente espacial, pois simplesmente acrescenta uma dimensão referente ao tempo. Entretanto, a diferença física entre espaço e tempo e a ordenação natural dos tempos induzem características particulares e dificuldades adicionais.

O tempo e o espaço são duas noções completamente diferentes, o que dificulta a especificação de funções de covariância espaço-temporais válidas. Além disso estas precisam satisfazer certas condições para serem consideradas válidas. Uma especificação muito comum para a função de covariância espaço-temporal é assegurada por propriedades que dão suporte à família de funções de covariância separáveis. Estas podem ser separadas através da soma e/ou produto de funções de covariância válidas. A grande desvantagem deste método é que ele não considera na modelagem a interação entre o espaço e o tempo.

Uma alternativa às funções de covariância separáveis são as funções de covariância espaço-temporais não separáveis e neste caso a interação entre o espaço e o tempo é considerada na modelagem. Na literatura existem algumas possibilidades de funções não separável, no presente trabalho foram consideradas a especificação de Cressie e Huang e a família de Gneting.

A especificação de Cressie e Huang se baseia na análise da integral de Fourier o que requer operações utilizando representação Espectral da função de covariância e a inversão de volta do domínio da frequência com o das observações só é obtida em casos particulares. As funções de covariância espaço-temporais propostas por Gneting consideram a combinação de funções completamente monótonas e estritamente crescentes que asseguram funções de covariância válidas.

Nas aplicações deste trabalho foram consideradas funções de covariância espaço-temporais pertencente à família de Gneiting, estacionárias e completamente simétricas, que são uma escolha conveniente, visto que esta especificação não requer qualquer tipo de integrabilidade e possui implementação disponível no pacote RandomFields.

Para os dois conjuntos de dados foi ajustado inicialmente um modelo com estrutura de covariância separável, composta pelo produto entre uma função *cauchy generalizada*, para o componente puramente temporal, e uma função *stable*, para o componente espacial. Ainda para a covariância separável foi adicionado um efeito pepita composto pelo produto entre uma constante e uma função *cauchy generalizada*. Depois de estimados todos os valores para os parâmetros do modelo separável, foi ajustado um modelo de covariância não separável mas com a mesma estrutura de anisotropia estimada para o modelo separável. O modelo não separável considerado é composto pelo produto entre uma função de covariância *cauchy generalizada*, que depende apenas do tempo, e uma função não separável proposta por Gneiting, que depende do tempo e do espaço conjuntamente. Para o efeito pepita foi considerada a mesma estrutura estimada para o modelo separável. É importante ressaltar que o modelo separável proposto é uma caso particular do modelo não separável, quando o parâmetro $\beta = 0$.

A utilização de conjunto de dados reais levantou dificuldades práticas na condução das análises. A primeira dificuldade foi a depuração das linhas de comandos do pacote "*RandomFields*", que ainda se encontra em desenvolvimento e poucas são as publicações que se utilização deste software, um outro grande problema relacionado à parte computacional é a demanda por computadores com grande capacidade de memória e processamento, já que o número de observações para análise espaço-temporal cresce muito rapidamente com o aumento de obsevações na coordenada do tempo ou no espaço.

Para a análise do conjunto de dados foram considerados os mesmos pontos em todos os tempos, sendo isto uma obrigatoriedade da implementação disponível no pacote, já que não existe a possibilidade de se especificar coordenadas diferentes para cada tempo. Por conta destas limitações, algumas observações tiveram de ser descartadas da estimação e predição.

Foram encontradas dificuldades na estimação do parâmetro $\beta$ da função de covariância proposta por Gneiting, e portanto neste trabalho foram considerados outras formas de assegurar a validade dos valores estimados para este parâmetro. Na análise foi encontradas estimativa do parâmetro $\beta$ próximas de zero, apontando para uma fraca interação entre o espaço e o tempo.

Foi considerada na análise a predição de um tempo futuro e comparado com o verdadeiro valor observado para este tempo. O interesse neste caso foi verificar através do erro quadrático médio de predição a validade do ajuste da função de covariância. Na Tabela 3 existe uma grande diferença no erro quadrático médio, na média e na variância de predição. Através desta tabela pode-se concluir que a estrutura de covariância separável é a que mais se aproxima dos verdadeiros valores observados, o que pode-se perceber também é a diminuição da média e da variância de predição à medida que o valor do parâmetro $\beta$ se aproxima de 1.

No desenvolvimento deste trabalho foram feitas ainda várias simulações e análises que serviram para a melhor compreensão do pacote. Em uma destas simulações foi estimado um campo espaço-temporal com um determinado valor do parâmetro $\beta$, em seguida foi realizada a estimação dos parâmetros, e verificou-se que a estimativa de verossimilhança não se aproximou de forma satisfatória do valor estimado. Estes resultados são preliminares, e considera-se que estudos desta natureza devem ser efetuados de forma mais detalhada e portanto optou-se por não reportar aqui os resultados deste estudo de simulação. Uma condição que poderia ser considerada é a realização de várias estimações e então seria calculada a média das estimativas. Espera-se que a média dos parâmetros estimados se aproxime do valor dos parâmetros simulados. Com um número suficientemente grande de valores estimados pode-se ter uma melhor idéia da variabilidade de cada parâmetro e com isto a determinação de intervalos de confiança e cobertura poderiam ser obtidos, assim como poderia se avaliar o impacto em predições.

Estas são apenas algumas entre as muitas possibilidades em se aprofundar os estudos nesta área. Neste trabalho foi considerada apenas um resumo das novas metodologias de análise espaço-temporal e com o avanço computacional e o apelo prático desta metodologia fica cada vez mais fácil de se descrever melhor as leis da natureza. Modelos que utilizam a especificação da função de covariância, embora promissores e com fácil interpretação intuitiva, apresentam dificuldades e restrições para uso em dados reais. Estudos complementares são ainda necessário para auxiliar na compreensão das características de identificabilidade e aplicabilidade em situações reais, bem como as exigências e restrições de implementação.

## Agradecimentos

Este trabalho é parte da dissertação de mestrado do primeiro autor do Departmento de Ciências Exatas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

Agradeço a Dolorice Moreti e o Prof° Dr. Paulo Leonel Libardi do

Departamento de Física de Solos, por ter gentilmente cedido os dados da aplicação.


- *ABSTRACT: The specification of space-time covariance functions is one of the possible strategies to model processes observed at different locations and time points. Such functions can define separable and non-separable processes and must attend the condition of positive-definiteness. Among the strategies to obtain such valid functions are the ones suggested by Cressie and Huang (1999) and by Gneiting (2002). The former is based on the idea of obtaining valid functions in a space of increased dimension from valid functions on the primary dimension and requires operations in the frequency domain. Alternatively, the latter combines increasing monotone functions avoiding the inversion of spectral representations. There are still few reports of usage and comparisons of the strategies. This work follows Gneiting's proposals with different values for the space-time interaction parameter. Models were applied for the analysis of two real data sets, one about fish stocks in the Portuguese coast and a second on soil water storage. The implementation on the R package RandomFields was used, with methodology and computational implementation being reviewed. For both case the separable model provided a satisfactory fit, based on maximum likelihood estimation.*



- *KEYWORDS: Keywords: Random Fields; Geostatistic; Space-Time Models;* RandomFields *Package*


## Referências


BOCHNER, S. **Lectures on Fourier Integrals**. Princeton, N.J.: Princeton University Press, 1959.

CRESSIE, N.; HUANG, H-C. Classes of Non-Seperable, Spatio-temporal stationary covariance functions. **Journal of the American Statistical Association**, v.94, p.1330-1340, 1999.

DIGGLE, P.J.; RIBEIRO Jr., P.J.; CHRISTENSEN, O.F. **Spatial Statistics and Computational Methods**, Springer, New York, 2003.

DIGGLE, P.J.; RIBEIRO Jr., P.J. **Model-Based Geoestatistics**, Springer, 2006.

ELMATZOGLOU, I. *Spatio-temporal geoestatistical models,with an aplication in fish stock*, Submited for the degree of master in statistics at Lancaster University, Lancaster, 2006.

GNEITING, T. Nonseperable, Stationary Covariance Functions for Space-Time Data. **Journal of the American Statistical Association**, v.97, No.458, 2002.

GNEITING, T.; SCHLATHER, M. Stochastic models wich separate fractral dimension and Hurst effect. Technical Report Series, NRCSE-TRS no. 069, 2001.

GNEITING, T.; GENTON, M.G.; GUTTORP, P. Geostatistical Space-Time Models, Stationarity, Seperability and Full Symmetry. Technical Report no.475 Department of Statistics University of Washington, 2006.

KYRIAKIDIS, P.C.; JOURNEL, A.G. Geostatistical Space-Time Models: a review. **Mathematical Geology**, v.31, p.651-684, 1999.

LE, D.N.; ZIDEK, J.V. **Statistical Analysis of Environmental Space-time processes**, Springer, New York, 2006.

MORETI, D. *Avaliação espaço-temporal de processos do balanço de água em solo com citros*, 2006. 138p. Tese Doutorado - Universidade de São Paulo, Piracicaba, 2006.

SCHABENBERGER, O.; GOTWAY, C.A. **Statistical Methods for Spatial Data Analysis**, Chapman & Hall / CRC, 2005.

SCHMIDT, A.M.; SANSÓ, B. Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais. 17$^o$ SINAPE, 2006.

STEIN, M.L. Space-time covariance functions. Technical Report No. 4, University of Chicago, 2004.

Recebido em .

Aprovado após revisão em .