



Vinícius Basseto Félix

Geoestatística espaço-temporal: modelagem de fenômenos naturais no espaço-tempo

Maringá – Paraná
28 de janeiro de 2019

Vinícius Basseto Félix

Geoestatística espaço-temporal: modelagem de fenômenos naturais no espaço-tempo

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá como requisito final para obtenção do título de Mestre em Bioestatística.

Orientadora: Prof.^a Dr.^a Eniuce Menezes de Souza

Coorientador: Prof.^o Dr.^o Diogo Francisco Rossoni

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Mestrado Acadêmico em Bioestatística

Maringá – Paraná

28 de janeiro de 2019

Vinícius Basseto Félix

Geoestatística espaço-temporal: modelagem de fenômenos naturais no espaço-tempo/ Vinícius Basseto Félix. – Maringá – Paraná, 28 de janeiro de 2019-
121 p.: 30 cm.

Orientadora: Prof.^a Dr.^a Eniuce Menezes de Souza

Coorientador: Prof.^o Dr.^o Diogo Francisco Rossoni

Dissertação (Mestrado) – Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Mestrado Acadêmico em Bioestatística, 28 de janeiro de 2019.

1. Geoestatística. 2. Modelagem espaço-temporal. 3. Altimetria. I. Prof.^a Dr.^a Eniuce Menezes de Souza. II. Universidade Estadual de Maringá. III. Departamento de Estatística. IV. Geoestatística espaço-temporal: modelagem de fenômenos naturais no espaço-tempo

Vinícius Basseto Félix

Geoestatística espaço-temporal: modelagem de fenômenos naturais no espaço-tempo

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá como requisito final para obtenção do título de Mestre em Bioestatística.

Orientadora: Prof.^a Dr.^a Eniuce Menezes de Souza

Coorientador: Prof.^o Dr.^o Diogo Francisco Rossoni

Prof.^a Dr.^a Eniuce Menezes de Souza
Orientadora

Prof.^o Dr.^o Diogo Francisco Rossoni
Coorientador

Professor convidado 1
Professor convidado 1

Professor convidado 2
Professor convidado 2

Maringá – Paraná
28 de janeiro de 2019

Dedico este trabalho aos meus pais que sempre incentivaram minha educação.

AGRADECIMENTOS

Agradeço aos meus orientadores que foram pacientes durante todo o desenvolvimento desta dissertação, visto todos os obstáculos encontrados, além do acompanhamento por toda minha jornada acadêmica desde o primeiro ano da graduação e que se encerra agora, apesar de estar deixando a linha de pesquisa, devo muito a eles pois seus ensinamentos foram além do pesquisador tornando-me uma pessoa melhor.

*A statistical analysis, properly conducted, is a delicate dissection of uncertainties, a surgery
of suppositions.
(M. J. Moroney)*

RESUMO

Esta dissertação visa destacar a importância da modelagem espaço-temporal, em particular, a abordagem Geoestatística, em que traz um levantamento de modelos desenvolvidos na literatura, bem como a teoria, propriedades e limitações. Identifica-se diversas dificuldades no processo de análise dos dados, seja no método, dados ou na ferramenta computacional. Portanto, propõe-se melhorias, tal que se solucione ou amenize estas problemáticas, em especial com aplicação em dados de altimetria de coordenadas GNSS no Brasil, visto que modelagem é um problema aberto na área, e pode gerar inúmeros benefícios, resultante de sua ampla e imensa aplicação em diversos campos, com destaque para agricultura.

Palavras-chave: Geoestatística. Modelagem Espaço-temporal. Altimetria. GNSS.

ABSTRACT

This dissertation aims to highlight the importance of space-time modelling, in particular, a geostatistical approach, in which it brings a survey of models developed in the literature, as well as a theory, properties and limitations. It identifies several difficulties in the process of data analysis, in the method, data or in the tool. Therefore, it is planned to carried out an application, solving these problems, specially for GNSS coordinate altimetry data in Brazil, as the modelling of this data is an open problem in the area, and can generate numerous benefits, because of its wide and immense application in several fields, such as on agriculture.

Keywords: Geostatistics. Space-time modelling. Altimetry. GNSS.

LISTA DE ILUSTRAÇÕES

Figura 1 – Possibilidades de aplicação no espaço-tempo.	18
Figura 2 – Sub-áreas da análise espaço-temporal. Traduzido e adaptado de Hengl (2009).	19
Figura 3 – Exemplo de semivariograma com patamar.	27
Figura 4 – Exemplo de semivariograma com patamar e seus respectivos parâmetros. Fonte: Rossoni (2011).	27
Figura 5 – Ilustração, traduzida e adaptada de Gneiting, Genton e Guttorp (2006), da relação entre as propriedades: separabilidade, similitude total, estacionariedade e suporte compacto para funções de covariância estacionárias ou não.	30
Figura 6 – Ilustração traduzida e adaptada da evolução da linguagem <i>R</i> no contexto espacial e espaço-temporal até 2010. Fonte: Apresentação “ <i>Modelling spatio-temporal data with R</i> ”, de Edzer Pebesma no evento GEOINFO 2010.	59
Figura 7 – Tradução do fluxograma de análise de <i>Hadley Wickham</i>	59
Figura 8 – Ilustração da estrutura de dados <i>tidy</i>	61
Figura 9 – Ilustração da estrutura de dados <i>gstat</i>	61
Figura 10 – Geometrias distintas para dados simulados de uma distribuição normal padrão ($n = 10$).	63
Figura 11 – Geometrias distintas para dados simulados de uma distribuição normal padrão ($n = 100$).	63
Figura 12 – Geometrias distintas para dados simulados de uma distribuição normal padrão ($n = 1000$).	64
Figura 13 – Geometrias distintas para dados simulados de uma distribuição exponencial ($n = 1000$).	64
Figura 14 – Explicação da escala logarítmica em relação à escala linear.	69
Figura 15 – Composição da EEA, destacada pelos países coloridos (verde ou azul). Fonte: Wikipedia.	78
Figura 16 – Disposição das estações.	79
Figura 17 – Função de autocorrelação temporal para estação DESH001.	79

Figura 18 – Função de autocorrelação temporal de todas as estações.	80
Figura 19 – Função de autocorrelação temporal de todas as estações.	81
Figura 20 – Concentração de PM10 nas estações no último dia de cada ano.	82
Figura 21 – Semivariograma de cada último dia de cada ano.	82
Figura 22 – Mapa de calor das semivariâncias para cada dia.	83
Figura 23 – Mapa de calor das semivariâncias (<i>log</i>) para cada dia.	83
Figura 24 – Semivariograma empírico 3D.	84
Figura 25 – Semivariograma empírico 2D.	85
Figura 26 – Estações brasileiras de monitoramento contínuo (RMBC). (Fonte: IBGE) .	89
Figura 27 – Exemplo de estação na Universidade Federal de Pelotas. (Fonte: IBGE) .	89
Figura 28 – Mapa do Brasil com as estações utilizadas.	90
Figura 29 – Gráfico adimensional da coordenada vertical.	91
Figura 30 – Gráfico temporal das diferentes estações.	92
Figura 31 – Gráfico temporal das diferentes estações, no ano de 2018.	93
Figura 32 – Mapa de calor da função de autocorrelação.	94
Figura 33 – Mapa de calor da semivariância espacial.	95
Figura 34 – Semivariância espaço-temporal amostral em linhas.	96
Figura 35 – Semivariância espaço-temporal amostral em formato de mapa de calor. . .	96
Figura 36 – Semivariância espaço-temporal amostral e ajuste do modelo exponencial- exponencial.	97
Figura 37 – Predição do modelo exponencial-exponencial.	97
Figura 38 – Ajuste dos modelos 2 e 3.	98
Figura 39 – Ajuste do modelo 4.	99
Figura 40 – Estimativa pontual das estações, obtida do modelo 4.	99
Figura 41 – Comparativo entre predito e observado.	100
Figura 42 – Interpolação espacial, obtida do modelo 4.	100
Figura 43 – Exemplo de WMCC aplicado em 5 estações de GNSS.	115
Figura 44 – Exemplo de WMCC aplicado em 5 estações de GNSS.	116
Figura 45 – Representação do sistema de coordenadas ENU. (Fonte: Wikipedia)	117
Figura 46 – Representação do sistema de coordenadas ECEF e ENU. (Fonte: Navipedia)	118

LISTA DE ABREVIATURAS E SIGLAS

ACF	Função de autocorrelação
BFGS	<i>Broyden-Fletcher-Goldfarb-Shannon;</i>
CG	<i>conjugate gradients;</i>
ECEF	<i>Earth Centred Earth Fixed</i>
EEA	<i>European Economic Area</i>
ENU	<i>East-North-Up</i>
GLONASS	<i>Globalnaya navigatsionnaya sputnikovaya sistema</i>
<i>Gau</i>	Gaussiano
GEOINFO	<i>Brazilian Symposium on Geoinformatics</i>
GNSS	<i>Global Navigation Satellite System</i>
GNSS-SP	Rede GNSS do Estado de São Paulo
GPS	<i>Global Navigation System</i>
iid	independente e identicamente distribuído
IGS	<i>International GNSS Service</i>
MPSE	<i>Minimum mean squared prediction error</i>
RBMC	Rede Brasileira de Monitoramento Contínuo
SIRGAS-CON	Sistema de Referência Geocêntrico para as Américas

LISTA DE SÍMBOLOS

Z	campo aleatório;
$Z(\cdot)$	campo aleatório espacial;
$Z(\cdot; \cdot)$	campo aleatório espaço-temporal;
$N(\cdot)$	conjunto de todos os pares pontos para qual diferença entre dois pontos é uma determinada distância/ <i>lag</i> ;
$N(\cdot; \cdot)$	conjunto de todos os pares pontos para qual diferença entre dois pontos é uma determinada distância e um intervalo de tempo;
θ_s	constante real;
θ_t	constante real;
a	constante real;
\bar{a}	complexo conjugado de uma constante real;
$\rho_{st}(\cdot; \cdot)$	correlação espaço-temporal;
$C_{st}(\cdot; \cdot)$	covariância espaço-temporal;
$C_s(\cdot)$	covariância puramente espacial;
$C_t(\cdot)$	covariância puramente espacial;
D	domínio espacial;
T	domínio temporal;
\mathbb{R}	espaço real;
$E(\cdot)$	esperança matemática;

$\hat{\gamma}_s(\cdot)$	estimador clássico de <i>Matheron</i> para semivariância puramente espacial;
$W(\theta)$	expressão do método de mínimos quadrado ponderados;
$\varphi(\cdot)$	função contínua completamente monótona para valores positivos;
$\cos(\cdot)$	função cosseno;
$\text{cov}(\cdot)$	função covariância;
\exp	função exponencial;
$\Phi(\cdot)$	função positiva com derivada completamente monótona;
$Pr(\cdot)$	função probabilidade;
var	função variância;
n	inteiro positivo;
h_s	<i>lag</i> espacial;
h_t	<i>lag</i> temporal;
Σ	matriz de covariância do campo aleatório;
μ_s	média “espacial”;
d	número de dimensões no espaço;
θ	parâmetro da função covariância espaço-temporal;
s	ponto no espaço;
t	ponto no tempo;
$Y^*(\cdot; \cdot)$	preditor de krigagem simples;
$Y(\cdot; \cdot)$	processo espaço-temporal;
$\text{tol}(\cdot)$	região de tolerância especificada em torno de \cdot ;
$\gamma_{st}(\cdot; \cdot)$	semivariância espaço-temporal;
$\gamma_s(\cdot)$	semivariância puramente espacial;
$\gamma_t(\cdot)$	semivariância puramente temporal.
$\varepsilon(\cdot; \cdot)$	termo de erro espaço-temporal;

SUMÁRIO

Introdução	17
0.1 Objetivos	20
I O estado da arte na geoestatística espaço-temporal	22
1 Geoestatística Espacial	23
1.1 História	23
1.2 Propriedades e Conceitos	24
1.3 Semivariograma espacial	26
2 Geoestatística Espaço-temporal	29
2.1 Propriedades e Conceitos	29
2.1.1 Função positiva definida	30
2.1.2 Estacionariedade	31
2.1.3 Representação espectral	33
2.1.4 Separabilidade	34
2.1.5 Simetria total	35
2.1.6 Isotropia e Anisotropia	36
2.1.7 Outras propriedades	36
2.2 Covariância	36
2.2.1 Covariâncias separáveis	40
2.2.1.1 Modelo <i>Sum</i>	40
2.2.1.2 Modelo <i>Metric</i>	41
2.2.1.2.1 Sum-metric	41
2.2.1.3 Modelo <i>Product</i>	42
2.2.1.4 Modelo <i>Product-Sum</i>	43
2.2.1.4.1 <i>Exponential Product-Sum</i>	48
2.2.1.4.2 <i>Integrated Product-Sum</i>	48
2.2.1.4.3 <i>Product-sum via Cressie e Huang</i>	49
2.2.1.5 Modelo de Ma	49
2.2.2 Covariâncias não-separáveis	50
2.2.2.1 Modelo de Cressie e Huang	50
2.2.2.2 Modelo de Gneiting	52
2.3 Krigagem espaço-temporal	53

2.3.1	Krigagem simples	54
2.3.2	Krigagem ordinária	55
II	Metodologia	57
3	Desenvolvimento computacional	58
4	Importação e Estruturação	60
5	Análise exploratória	62
5.1	Análise adimensional:	62
5.2	Análise temporal:	65
5.2.1	Autocorrelação	65
5.2.2	Correlação cruzada	65
5.2.3	Entropia	68
5.2.4	Granulosidade e Estabilidade	68
5.2.5	Outras métricas	68
5.3	Análise espacial:	69
5.4	Análise espaço-temporal:	69
6	Modelagem	71
6.1	Estimação	71
6.2	Krigagem	73
6.3	Diagnóstico e seleção	74
III	Aplicações	76
7	Poluição do ar	78
7.1	Definição do problema	78
7.2	Descritiva	79
7.3	Discussão	85
8	Coordenadas GNSS	86
8.1	Definição do problema	86
8.2	Descritiva	90
8.3	Modelagem	97
8.4	Discussão	101
IV	Considerações finais	102
	Referências	105

Apêndices	114
APÊNDICE A WMCC	115
APÊNDICE B Sistema local de coordenadas	117
Anexos	120
ANEXO A Prova do Teorema 3	121
ANEXO B Prova do Teorema 4	122
ANEXO C Prova do Teorema 5	124

INTRODUÇÃO

O estudo de séries temporais ou espaciais cresce cada vez mais, vide que ambos estão atrelados a inúmeras áreas de extrema relevância, como: Agronomia ([SCHEPERS et al., 2004](#)), Astronomia ([SCARGLE, 1982](#)), Batimetria ([LEJOT et al., 2007](#)), Ciências Sociais ([QUINN; ROHRBAUGH, 1983](#)), Economia ([BECK; KATZ, 2011](#)), Meteorologia ([WIGLEY; BRIFFA; JONES, 1984](#)), Saúde ([YANOSKY et al., 2008](#)), entre outras.

O motivo desta diversidade deve-se que, conceitualmente, todos os dados são espaço-temporais, já que se mensura ou coleta qualquer observação em um determinado tempo e espaço. O problema é que estas informações são normalmente negligenciadas, seja por não serem vistas como relevantes, ou pela incorporação destas na análise ser difícil, e talvez a mais problemática de todas, a coleta deste tipo de dados em ambos os domínios ser inviável.

Entretanto, com o avanço tecnológico a coleta de dados no tempo e espaço é cada vez mais comum ([GRÄLER; PEBESMA; HEUVELINK, 2016](#)), sendo necessário a construção de modelos espaço-temporais, a fim de atender a esta demanda. Por exemplo, um dos autores mais renomados da Estatística Espacial, *Noel Cressie*, dedicou somente 4 páginas, em seu livro de quase 900 ([CRESSIE, 1992](#)), aos modelos espaços-temporais, sendo que anos depois escreveu um livro só para técnicas no espaço-tempo ([CRESSIE; WIKLE, 2015](#)).

Assim, é possível encontrar na literatura cada vez mais trabalhos diversos e recentes com a abordagem espaço-temporal, como: a modelagem do efeito de memória longa no espaço-tempo para turbinas eólicas na Irlanda ([HASLETT; RAFTERY, 1989](#)), abordagem espaço-temporal na modelagem meteorológica ([HANDCOCK; WALLIS, 1994](#)), modelos estocásticos espaço-temporais para precipitação ([RODRIGUEZ-ITURBE et al., 1998](#)) e a análise de dados atmosféricos com funções de correlação ([GNEITING, 1999](#)), a reconstrução da dispersão global inicial do vírus H1N1pdm no espaço e tempo, via abordagem bayesiana ([LEMEY; SUCHARD; RAMBAUT, 2009](#)), o uso de modelagem para melhoria de operações de combate no Iraque ([BRAITHWAITE; JOHNSON, 2012](#)), previsões espaciais acuradas de localidades não monitoradas com o uso de modelos autorregressivos hierárquicos bayesianos ([SAHU; BAKAR, 2012](#)), prevenção de geadas com uso de modelos bayesianos no espaço-tempo no sudeste

australiano (CRIMP et al., 2015), análise espaço-temporal da severidade de acidentes com passageiros em táxis em Hong Kong com modelos logísticos espaço-temporais bayesianos (MENG et al., 2017), e um modelo multivariado no espaço-tempo para analisar a taxa de mortes por doença cardíaca considerando raça, sexo e local nos últimos 50 anos (QUICK; WALLER; CASPER, 2018).

Logo, vê-se que dentro do espaço-tempo há uma grande variedade de situações, de forma que diferentes vertentes de modelos são aplicados aos respectivos problemas (Figura 1).

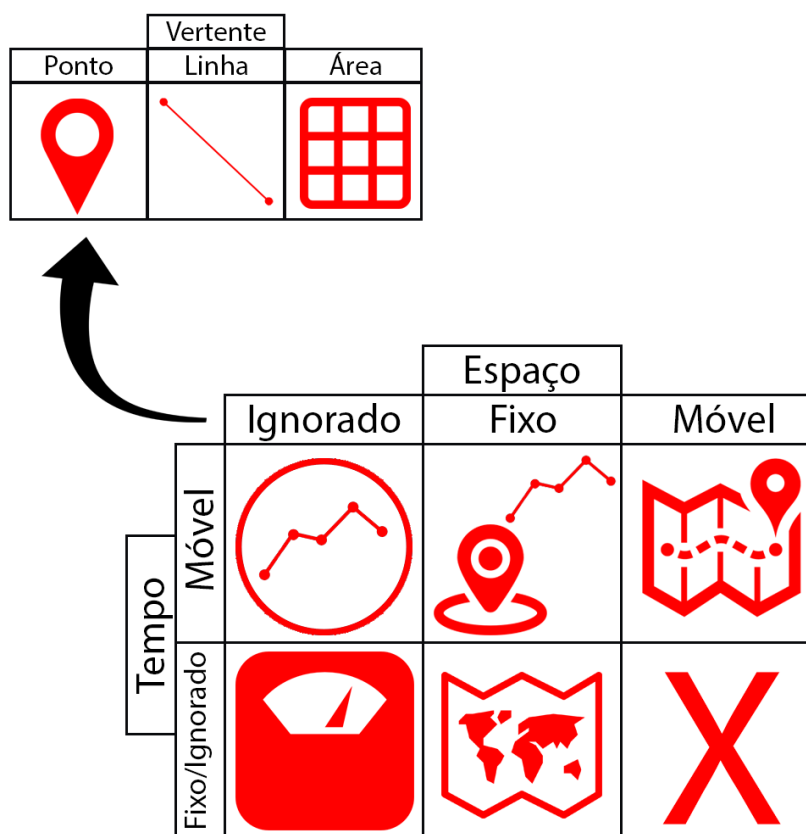


Figura 1 – Possibilidades de aplicação no espaço-tempo.

Na Figura 1 observa-se as diferentes combinações de aplicações, sendo que para um tempo móvel e espaço ignorado tem-se uma série temporal, por exemplo, a temperatura diária ao longo do ano; ao adicionar a informação que esta série pertence à uma cidade específica o espaço agora é fixo; já se ambos fossem móveis, um exemplo seria da migração de animais, que se dá pelo seu deslocamento ao longo do tempo e espaço. Ainda, para tempo fixo ou ignorado há situação do espaço ignorado, ou seja, a maioria dos estudos, isto é, um estudo de peso qualquer, em que só a variável de interesse é considerada, também o espaço pode estar fixo, sendo o caso dos estudos puramente espaciais e por fim o último caso é impossível, pois mover-se no espaço estando fixo no tempo é ilógico, de forma que a mobilidade no espaço é dependente do tempo. Importante ressaltar que algumas vertentes podem ser adotadas para estas situações, e estas são referentes à disposição dos dados no espaço, sendo elas: ponto,

linha ou área.

Assim, com a gama de situações possíveis criaram-se sub-áreas para lidar com cada uma delas, como apresentado na Figura 2.

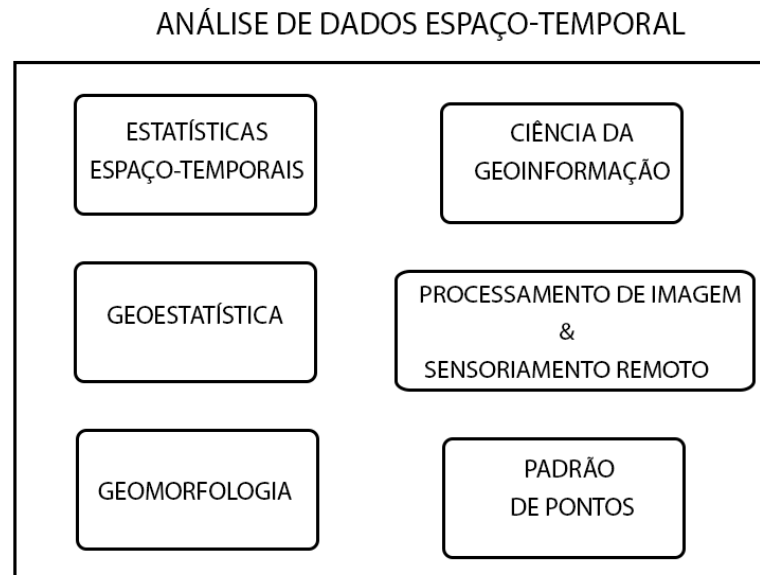


Figura 2 – Sub-áreas da análise espaço-temporal. Traduzido e adaptado de Hengl (2009).

Na Figura 2 nota-se os campos mais notáveis da análise espaço-temporal, sendo que estes consistem em:

- **Geoestatística:** análise de dados contínuos georreferenciados;
- **Geomorfologia:** análise de dados quantitativos de topografia;
- **Ciência da Geoinformação:** estudo computacional e estrutural de dados geográficos;
- **Processamento de Imagem & Sensoriamento remoto:** estudo de imagens e dados obtidos via sensores distantes;
- **Padrão de pontos:** análise de objetos, seja em linha e/ou ponto.

Com tantos campos possíveis, conseqüentemente existem diversas técnicas para solução respectiva de cada cenário, Cressie e Wikle (2015) resumem alguns dos tipos modelos desenvolvidos, como: modelo de efeitos regionais sem (REINSEL et al., 1981) e com correlação temporal (BLOOMFIELD et al., 1983), modelo de covariância invariante por tempo (STEIN, 1986) e modelo de mistura (OEHLERT, 1993).

Havendo tantos modelos é importante saber distinguir cada um deles, bem como suas propriedades. A priori, pode-se dividir as técnicas em 3 tipos:

- **Modelos determinísticos:** estes se pautam em efeitos conhecidos dos fenômenos estudados, podendo ser conhecidos também como modelos: fenomenológicos, mecânicos ou ainda biológicos. Tal que estes, em geral, baseiam-se em leis físicas para sua construção através de equações que representem evento de interesse;
- **Modelos empíricos:** estes se pautam em probabilidade, através do ajuste dos dados. Não requerendo assim o conhecimento aprofundado do fenômeno em si, porém em contrapartida possuem suposições estritas a serem cumpridas;
- **Modelos *data-driven*:** estes se pautam, em geral, em técnicas matemáticas visando a predição dos dados e não sua interpretação. Para tal utilizam-se em boa parte técnicas de *machine learning*.

Neste trabalho, em particular, faz-se o estudo da geoestatística espaço-temporal (Parte I), ou seja, quando a variável de interesse está atrelada à um ponto no espaço-tempo. Logo, estudou-se as classes de modelos de covariância e semivariância espaço-temporais, bem como suas propriedades e diferentes tipos, como será abordado na Seção 2.2.

Além disso, segundo Myers (2004) é comum e de certa forma fácil coletar dados temporais regulares, porém o mesmo não acontece no espaço. Ainda, alguns dados, por exemplo, coletados por equipamentos automáticos como monitoramento da qualidade do ar costumam conter muitos pontos distintos no tempo e poucos no espaço, esta situação torna-se problemática no ajuste (GNEITING; GENTON; GUTTORP, 2006). Assim, uma prática comum é a seleção de alguns modelos, como também critérios para escolha do que será utilizado para os fins propostos.

Desta forma, planeja-se uma aplicação a fim de compreender o processo de modelagem na geoestatística espaço-temporal. Para tal faz-se aplicações em dados de poluição do ar e coordenadas GNNS.

0.1 Objetivos

Dado que há diversos problemas, divide-se os objetivos em 3 tipos. Em termos da Geoestatística espaço-temporal, pretende-se:

- Levantar os modelos desenvolvidos ao longo da literatura, e elencar suas propriedades, vantagens e desvantagens;
- Utilizar análises descritivas, principalmente formas diferentes de visualização gráfica para entender o comportamento dos dados, visando: identificação de efeitos, visualização de dados em pequenos e longos períodos de tempo e/ou pontos no espaço, e indicativos para modelos e seus “chutes iniciais”.

Já em termos dos dados modelados para coordenadas GNSS, visa-se

- Identificação de modelo que permita a adequação dos dados irregularmente distribuídos pelo território brasileiro;
- Prover modelos com previsões, a fim de fazer a interpolação de pontos, para utilização em áreas que dependam destas coordenadas, e.g., Agricultura.

E por fim, na implementação computacional, objetiva-se

- Criar funções que permitam a visualização dos dados sem o constante processo de estruturação, visto a necessidade particular de formato inerente a cada pacote ou etapa;

Parte I

O estado da arte na geoestatística espaço-temporal

CAPÍTULO 1

GEOESTATÍSTICA ESPACIAL

Segundo [Cesare, Myers e Posa \(2001\)](#) uma solução para lidar com um fenômeno espaço-temporal é considerá-lo uma realização de um campo aleatório definido em \mathbb{R}^{d+1} em que d é o espaço físico dimensional, desta forma estende-se o domínio espacial para o espaço-temporal. Assim, antes de entender como a Geoestatística funciona no espaço-tempo, é necessário entender sua origem, ou seja, saber seu funcionamento somente no espaço.

1.1 História

A Geoestatística inicialmente pode ser considerada uma das vertentes da Estatística Espacial, em que objetivamente designa o estudo estatístico de fenômenos naturais, através da incorporação da dependência espacial ([ROSSONI, 2014](#)).

O pai da Geoestatística, *George Matheron*, inspirado nos trabalhos do professor *Hendrica Johanna de Wijs* sobre a concentração de minérios ([WIJS, 1951](#); [WIJS, 1953](#)), bem como em *Daniel Gerhardus Krige* e seu estudo da concentração de ouro em *Witwatersrand*, África do Sul, em que constatou a ineficácia do uso de técnicas que contam somente com a variância ([KRIGE, 1951](#)), não levando em conta a posição das observações, conseqüentemente, ele propôs uma forma de adequar o estudo dos fenômenos de modo que incorporasse suas respectivas posições no espaço.

Assim, [Matheron \(1971\)](#) consagrou a teoria das variáveis regionalizadas, ou seja, tal que se leve em conta a posição relativa do fenômeno. Porém, com o avanço das técnicas, a abrangência do termo Geoestatística foi além da original ([MATHERON, 1971](#)), de tal forma esta pode ser dada com uma subárea da Estatística visando o estudo das variáveis regionalizadas, em que se considera a característica estrutural do fenômeno ([YAMAMOTO; LANDIM, 2014](#)).

1.2 Propriedades e Conceitos

Seja $Z = \{Z(s), z \in D \subseteq \mathbb{R}^d\}$ um campo aleatório espacial real definido no domínio D do espaço \mathbb{R}^d d -dimensional, em que $d \leq 3$, e s consiste na posição, tal que

$$Z(s) = \mu(s) + \varepsilon'(s) + \varepsilon''(s), \quad (1.1)$$

em que

- $\mu(s)$: componente estrutural determinístico;
- $\varepsilon'(s)$: termo estocástico, com variação local e espacialmente correlacionado;
- $\varepsilon''(s)$: ruído aleatório, não correlacionado.

Assim, o conjunto de variáveis aleatórias $\{Z(s_i), i = 1, \dots, n\}$ correlacionadas entre si constitui uma função aleatória, de forma que não é possível a determinação de estatísticas no ponto s_i da função (SOARES, 2006). Assim, uma solução possível está na suposição de estacionariedade em diferentes níveis da função, isto é, esperança constante:

$$E[Z(s_1)] = E[Z(s_2)] = \dots = E[Z(s_n)] = E[Z(s)] = \mu_s. \quad (1.2)$$

Utilizando-se da expressão (1.2), em que a média “espacial” (μ_s) torna-se independente da localização, sendo obtida como a média aritmética das realizações das variáveis aleatórias (SOARES, 2006 apud YAMAMOTO; LANDIM, 2014), isto é,

$$\mu_s = E[Z(s)] = \frac{1}{n} \sum_{i=1}^n Z(s_i). \quad (1.3)$$

Entretanto, esta suposição requer que os valores do fenômeno estudado sejam homogêneos. Como esta é uma ocorrência rara (YAMAMOTO; LANDIM, 2014), é necessária a verificação tanto da distribuição quanto da variabilidade.

Definição 0.1 (Estacionariedade espacial de 2ª ordem) *Considerando $E[Z(s)]$, a esperança matemática do processo espacial é uma função que existe e não dependa de s , em que a mesma depende somente do lag/distância espacial (h_s), ou seja, esta esperança é independente da localização, tal como sua covariância. Logo, a função aleatória $Z(s)$ possui estacionariedade espacial de 2ª ordem ou fraca, se e somente, satisfaça*

$$E[Z(s)] = \mu_s, \quad \forall s \in \mathbb{R}^d, \quad (1.4)$$

ou ainda que $F_s(z) \equiv Pr(Z(s) \leq z)$ não dependa de s (CRESSIE, 1992).

Além de também satisfazer

$$\text{cov}[Z(s_1); Z(s_2)] = C_s(s_1 - s_2), \quad (s_1, s_2) \in \mathbb{R}^d, \quad (1.5)$$

para estimação de preditores lineares ótimos.

Como o interesse está em estudar a distância, é comum utilizar-se da covariância espacial, estatística que mensura a relação entre os valores de uma mesma variável, em que estes são obtidos em pontos separados por h_s , dada uma direção específica. Logo, a escolha da direção pode alterar a covariância calculada. Quando este fato ocorre, há indicativo de uma característica conhecida como anisotropia espacial (YAMAMOTO; LANDIM, 2014), e isotropia, caso contrário, ou seja, a covariância espacial independe da direção escolhida.

Assim, a covariância espacial é expressa por

$$C_s(h_s) = E\{[Z(s + h_s) - \mu_s][Z(s) - \mu_s]\}. \quad (1.6)$$

Ainda, a função de semivariância espacial pode ser entendida como variância do incremento, sendo dada por,

$$2\gamma_s(h_s) = E\{[Z(s + h_s) - Z(s)]^2\}. \quad (1.7)$$

Entretanto, Journel e Huijbregts (1978) apontam que a suposição de estacionariedade de 2ª ordem assume uma variância existente e finita, só que há possibilidade de fenômenos espaciais com dispersão infinita, sendo assim impossível o cálculo de estatísticas como variância e covariância, mas sendo possível a obtenção de uma semivariância.

Uma alternativa é a suposição de estacionariedade intrínseca, esta não requer a média constante ou a variância finita para o processo $Z(s)$, e sim que apenas os incrementos da função $[Z(s + h_s) - Z(s)]$ sejam estacionários de 2ª ordem (GOOVAERTS, 1997 apud YAMAMOTO; LANDIM, 2014), ou seja,

$$E[Z(s + h_s) - Z(s)] = \mu_s, \quad (1.8)$$

e

$$\text{var}[Z(s + h_s) - Z(s)] = 2\gamma_s(h_s). \quad (1.9)$$

Porém, na Geoestatística a suposição da estacionariedade de 2ª ordem é fraca, pois requer somente que os incrementos sejam estacionários de 2ª ordem. De tal forma, utilizando-se da expressão (1.9), a semivariância espacial é dada em termos da variância do incremento

$$\gamma_s(h_s) = \frac{1}{2} \text{var}[Z(s + h_s) - Z(s)], \quad \forall s, s + h_s \in D. \quad (1.10)$$

No caso em que Z é estacionário de segunda ordem, tem-se que

$$\gamma_s(h_s) = C_s(0) - C_s(h_s). \quad (1.11)$$

Além disso, se a covariância ou semivariância é uma função de comprimento h_s , pode-se denominá-la isotrópica.

1.3 Semivariograma espacial

Viu-se na expressão (1.7) que a função de semivariância mensura a variância do incremento, ou seja, a dispersão entre pontos separados por um *lag* espacial (h_s). Assim, deve-se haver uma proporcionalidade direta entre distância e variância, em que é muito comum a estabilização da dispersão a partir de uma distância determinada, categorizando os como semivariogramas com patamar (Figura 3).

Assim, é possível o ajuste de modelos teóricos visando este comportamento, em que Olea (2012) apresenta alguns deles, como os modelos: esférico, exponencial, gaussiano, entre outros.

Além dos modelos é necessário um estimador que permita o cálculo da variação do fenômeno regionalizado no espaço (JOURNAL; HUIJBREGTS, 1978). Ao longo do tempo alguns estimadores de semivariância foram desenvolvidos (FÉLIX et al., 2016), em que o mais popular e utilizado é o estimador clássico de Matheron (1962), dado por

$$\hat{\gamma}_s(h_s) = \frac{1}{2|N(h_s)|} \sum_{N(h_s)} [Z(s_i) - Z(s_j)]^2, \quad (1.12)$$

em que

- $Z(s_i)$ é o valor da variável no ponto s_i ;
- $Z(s_j)$ é o valor da variável no ponto s_j ;
- $N(h_s)$ é o conjunto de todos os pares de pontos (s_i, s_j) para os quais $|s_i - s_j| = h_s$, e o número de pares separados por uma determinada distância h_s ;
- $|N(h_s)|$ é a cardinalidade deste conjunto.

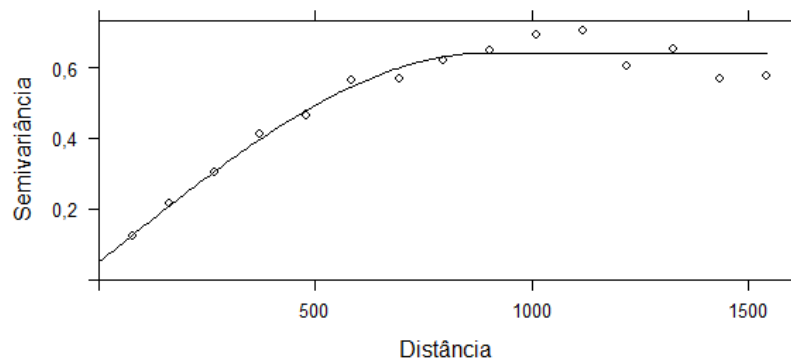


Figura 3 – Exemplo de semivariograma com patamar.

Desta forma, um semivariograma com patamar possui características intrínsecas, ou seja, parâmetros a serem estimados.

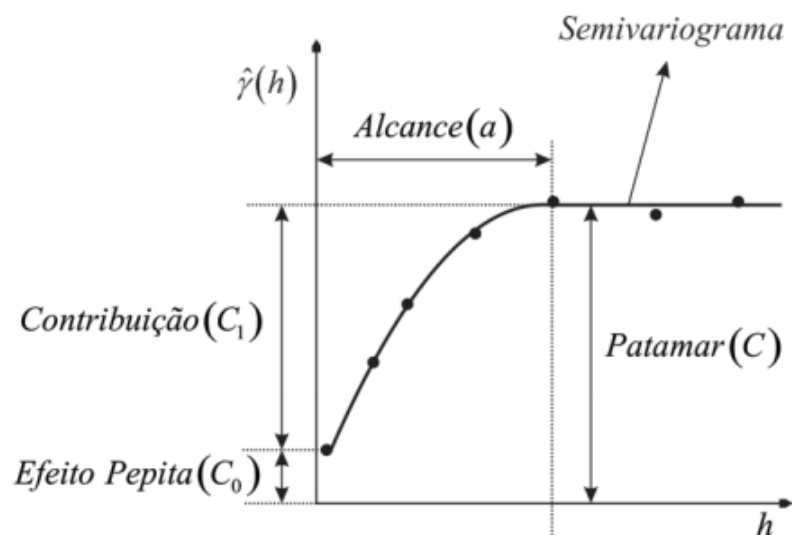


Figura 4 – Exemplo de semivariograma com patamar e seus respectivos parâmetros. Fonte: Rossoni (2011).

Estes são definidos, como:

- **Alcance:** distância máxima anterior a estabilização da semivariância, ou seja, amplitude em que há correlação espacial;
- **Patamar:** semivariância equivalente ao alcance;
- **Efeito Pepita:** teoricamente nulo, porém na prática mensura variabilidade desconhecidas;
- **Contribuição:** diferença entre Patamar e Efeito Pepita.

Ainda, se a covariância ou a semivariância não é isotrópica, um fenômeno espacial pode apresentar o efeito de anisotropia, ou seja, seu variograma altera-se conforme a direção escolhida no cálculo da semivariância. Existem diversos tipos de anisotropia, em que as 3 mais populares são:

- **Geométrica:** patamar fixo e alcances diferentes;
- **Zonal:** alcance fixo e patamares diferentes;
- **Mista:** alcances e patamares diferentes.

conforme a mudança da direção. Outro efeito é a tendência espacial, que pode ser indicada pelo crescimento constante do semivariograma empírico, tal que pode prejudicar o ajuste, e também a predição, pois possivelmente será a causa da não-estacionariedade ou até a não-normalidade dos resíduos.

CAPÍTULO 2

GEOESTATÍSTICA ESPAÇO-TEMPORAL

A abordagem da Geoestatística no espaço-tempo modela as observações como uma realização parcial espaço-temporal (GNEITING; GENTON; GUTTORP, 2006), tipicamente uma variável aleatória gaussiana dada por:

$$Z(s; t), \quad (s; t) \in \mathbb{R}^{d+1}, \quad (2.1)$$

que é indexada pelo espaço por $s \in \mathbb{R}^d$ e pelo tempo por $t \in \mathbb{R}$, logo $(s; t) \in \mathbb{R}^{d+1}$. Assumindo que os segundos momentos existem e são finitos, tem-se que a predição depende da especificação correta da estrutura de covariância e/ou semivariância espaço-temporal. Seja $Z(s; t)$ um campo aleatório espaço-temporal, em que $(s; t) \in \mathbb{R}^{d+1}$, tem-se que a covariância entre $Z(s_1; t_1)$ e $Z(s_2; t_2)$ depende, geralmente, das coordenadas $(s_1; t_1)$ e $(s_2; t_2)$, e nenhuma outra estrutura existe (GNEITING; GENTON; GUTTORP, 2006). Assim, na prática, a estimação e modelagem necessitam de simplificações de suposições, como: estacionariedade, separabilidade e simetria total, que são abordadas na Seção 2.1.

2.1 Propriedades e Conceitos

Na estatística espaço-temporal algumas propriedades são indispensáveis, principalmente no processo de modelagem, pois algumas delas são condições intrínsecas para utilização de determinados modelos. Assim, nesta seção aborda-se as principais propriedades e conceitos.

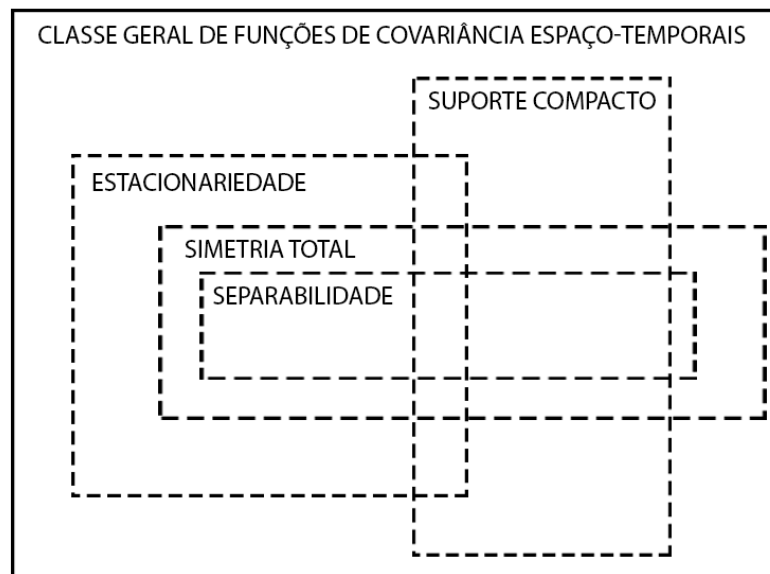


Figura 5 – Ilustração, traduzida e adaptada de Gneiting, Genton e Guttorp (2006), da relação entre as propriedades: separabilidade, simetria total, estacionariedade e suporte compacto para funções de covariância estacionárias ou não.

A Figura 5 ilustra as principais propriedades das funções de covariância espaço-temporal, nela vê-se que nenhuma propriedade é disjunta de outra. Ainda, nota-se que para uma função ser separável é necessário possuir simetria total.

Assim, seja o campo aleatório real $Z = \{Z(s; t), (s; t) \in \mathbb{R}^{d+1}\}$, detalha-se as propriedades a seguir.

2.1.1 Função positiva definida

Sabe-se que uma função de covariância é não-negativa definida, ou seja,

Definição 0.2 (Função não-negativa definida) Uma função qualquer $\{f(u, v) : u, v \in \mathbb{R}\}$ definida em \mathbb{R}^2 é dita uma função não-negativa definida, se para quaisquer números complexos $\{a_i : i = 1, \dots, n\}$, qualquer $\{u_i : i = 1, \dots, n\}$ em \mathbb{R} , e qualquer inteiro n , tem-se que

$$\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j f(u_i, u_j) \geq 0, \tag{2.2}$$

em que \bar{a}_j denota o conjugado complexo de a_j .

Assim, para uma função qualquer, $f(\cdot; \cdot)$, ser positiva definida é necessário que a inequação (2.2) seja estritamente positiva, sempre que $a \equiv (a_1, \dots, a_n)'$ é um vetor não nulo.

Visto, que um dos focos deste trabalho é a de modelos para covariância no espaço-tempo, esta propriedade torna-se uma suposição essencial, e para o contexto espaço-temporal, tem-se que

- $Z(s; t)$, em que $(s; t) \in \mathbb{R}^{d+1}$, é um campo aleatório espaço-temporal;
- n um inteiro positivo;
- $(s_1; t_1), \dots, (s_n; t_n)$ coordenadas espaço-temporais em \mathbb{R}^{d+1} .

Se o processo é estacionário com função de covariância espaço-temporal, C_{st} , definida em \mathbb{R}^{d+1} , sua matriz pode ser expressa por

$$[C_{st}(s_i - s_j; t_i - t_j)]_{i,j=1,\dots,n}. \quad (2.3)$$

Então, a função C_{st} é positiva definida se a matriz (2.3) é não-negativa definida para todas as possíveis coordenadas espaço-temporais. E pode-se reescrever a inequação (2.2) em termos da função de covariância espaço-temporal, de forma que a Definição 0.2 torna-se

$$\sum_{i=1}^m \sum_{j=1}^n a_i \bar{a}_j C_{st}(s_i - s_j; t_i - t_j) \geq 0, \quad (2.4)$$

para qualquer $\{a_i\}$, $\{(s_i; t_i)\}$, e n .

Ainda, segundo Gneiting, Genton e Guttorp (2006) há dificuldade em checar se uma função é positiva definida, tornando-se uma das complicações na modelagem paramétrica da covariância espaço-temporal.

2.1.2 Estacionariedade

Outra propriedade essencial é a de estacionariedade, muito presente na modelagem temporal (MORETTIN; TOLOI, 2006a) e também na modelagem espacial (JOURNEL; HUIJBREGTS, 1978; CRESSIE, 1992).

Definição 0.3 (Estacionariedade) *Uma função de covariância espaço-temporal, definida em \mathbb{R}^{d+1} , é estacionária, se satisfaz a Definição 0.2 e pode ser escrita como:*

$$cov(s, h_s; t, h_t) \equiv C_{st}(Z(s, t); Z(h_s, h_t)), \quad s, h_s \in \mathbb{R}^d, t, h_t > 0. \quad (2.5)$$

Se um processo aleatório $Z(\cdot; \cdot)$ possui esperança e covariância constante, tem-se que sua função de covariância é estacionária de 2ª ordem (estacionariedade fraca), isto é,

Definição 0.4 (Estacionariedade de 2ª ordem) Uma função de covariância espaço-temporal, definida em \mathbb{R}^{d+1} , é estacionária de 2ª ordem, se satisfaz

$$E[Z(s_1; t_1)] = E[Z(s_2; t_2)], \quad (2.6)$$

e

$$C_{st}[s_1, s_2; t_1, t_2], \quad (2.7)$$

depende somente das diferenças, $s_1 - s_2$ e $t_1 - t_2$, ou seja os lags espacial (h_s) e temporal (h_t), respectivamente.

Além disso, outra estacionariedade possível é a intrínseca, em que esta baseia-se na semivariância, visto que para alguns fenômenos as suposições para covariância não são alcançadas (MA, 2003).

Definição 0.5 (Estacionariedade intrínseca) Um processo $Z(s; t)$ possui estacionariedade intrínseca no espaço-tempo, se

$$E[Z(s_1; t_1) - Z(s_2; t_2)] = 0, \quad (2.8)$$

e

$$\gamma_{st}[s_1, s_2; t_1, t_2], \quad (2.9)$$

depende somente das diferenças, $s_1 - s_2$ e $t_1 - t_2$, ou seja os lags espacial (h_s) e temporal (h_t), respectivamente.

Então, sob suposição de estacionariedade intrínseca, considerada fraca (MYERS, 2004), a semivariância é expressa por

$$\gamma_{st}(h_s; h_t) = \frac{\text{var}[Z(s + h_s; t + h_t) - Z(s; t)]}{2}, \quad (2.10)$$

em que depende somente do vetor dos lags h_s e h_t , e não de uma localização ou um tempo.

Já sob uma suposição mais forte, a de estacionariedade de 2ª ordem do processo, que garante também estacionariedade intrínseca (MA, 2003), tem-se que

$$\gamma_{st}(h_s; h_t) = C_{st}(0; 0) - C_{st}(h_s; h_t). \quad (2.11)$$

Sendo a expressão crucial para obtenção de modelos de semivariância via covariância espaço-temporal. Em que um resultado de (SCHOENBERG, 1938 apud MA, 2003) implica que $\gamma(s; t)$ é um semivariograma intrínseco, se e somente se, $\exp\{-\theta\gamma_{st}(s; t)\}$ é uma função de covariância para todo \mathbb{R}_+^1 .

Por fim, Ma (2003) mostra que é possível verificar a estacionariedade através do espaço ou tempo, de forma que

Teorema 1 *Sejam θ_s e θ_t constantes em \mathbb{R} . Se C_{st_0} é uma função de covariância espaço-temporal estacionária em \mathbb{R}^{d+1} , então*

$$C_{st}(s; t) = C_{st_0}(s + \theta_s t; t + \theta_t s), \quad (s, t) \in \mathbb{R}^{d+1}. \quad (2.12)$$

é uma função de covariância espaço-temporal estacionária em \mathbb{R}^{d+1} .

Corolário 1.1 *Se $C_s(s)$ é uma função de covariância espacial estacionária em \mathbb{R}^d , então*

$$C_{st}(s; t) = C_s(s + \theta_s t). \quad (s; t) \in \mathbb{R}^{d+1}, \quad (2.13)$$

é uma função de covariância espaço-temporal estacionária em \mathbb{R}^{d+1} .

Corolário 1.2 *Se $C_t(t)$ é uma função de covariância temporal estacionária em \mathbb{R}_+ , então*

$$C_{st}(s; t) = C_t(t + \theta_t s). \quad (s; t) \in \mathbb{R}^{d+1}, \quad (2.14)$$

é uma função de covariância espaço-temporal estacionária em \mathbb{R}^{d+1} .

2.1.3 Representação espectral

Ver-se-á na sub-seção 2.1.4 a separabilidade presente entre algumas classes de modelos de covariância, porém é possível a construção de modelos não-separáveis, em que uma das abordagens famosas é dada pela representação espectral (CRESSIE; HUANG, 1999).

Segundo o Teorema de Bochner (1955), uma função contínua é positiva definida, se e somente se, é uma transformação de Fourier (BRACEWELL, 1986) de uma medida finita não-negativa.

Teorema 2 (Teorema de Bochner) *Suponha C_{st} uma função contínua e simétrica em \mathbb{R}^{d+1} . Então, C_{st} é uma função de covariância se e somente pode ser escrita na forma*

$$C_{st}(h_s; h_t) = \int \int e^{i(h'_s \omega + h'_t \tau)} dF(\omega, \tau), \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.15)$$

em que F é uma medida finita, não-negativa e simétrica definida em \mathbb{R}^{d+1} .

Ou seja, a classe de funções de covariância espaço-temporais estacionárias é idêntica a classe de transformações de *Fourier* de medidas finitas, não-negativas e simétricas, no domínio \mathbb{R}^{d+1} . A medida F em (2.15) é comumente chamada de medida espectral, e se C_{st} é integrável, a medida espectral é contínua com densidade de *Lebesgue* (MATTILA, 1999)

$$f(\omega; \tau) = (2\pi)^{-(d+1)} \int \int e^{-i(h'_s \omega + h_t \tau)} C_{st}(h_s; h_t) dh_s dh_t, \quad (\omega, \tau) \in \mathbb{R}^{d+1}, \quad (2.16)$$

e f é chamada de densidade espectral (BROWN; HWANG, 1997). Se f existe, a representação (Teorema 2) é reduzida para

$$C_{st}(h_s; h_t) = \int \int e^{i(h'_s \omega + h_t \tau)} f(\omega, \tau) dh_s dh_t, \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.17)$$

em que C_{st} e f podem ser obtidos via transformação de *Fourier*.

Por fim, é possível avaliar a estacionariedade da função de covariância de forma separada, para o tempo e espaço. Assim,

$$\text{cov}(Z(s; t); Z(h_s; h_t)) \equiv C_{st}(s - h_s; t), \quad (2.18)$$

é uma função de covariância espaço-temporal espacialmente estacionária. Analogamente,

$$\text{cov}(Z(s; t); Z(h_s; h_t)) \equiv C_{st}(s; t - h_t), \quad (2.19)$$

é uma função de covariância espaço-temporal temporalmente estacionária.

2.1.4 Separabilidade

Inicialmente, muitos modelos partiram desta propriedade para construção de modelos de covariância separáveis. Uma função de covariância espaço-temporal estacionária é separável se existem funções de covariância puramente estacionárias para o espaço (C_s) e tempo (C_t). Para tal, como visto em Iaco, Myers e Posa (2002), sejam $C_1(h_1)$ e $C_2(h_2)$ duas covariâncias, utiliza-se alguns resultados como:

- **Convexidade:** $C_1(h_1) + C_2(h_2)$ é uma função de covariância;
- **Convexidade:** Seja uma constante $b > 0$, $b \times C_1(h_1)$ é uma função de covariância;
- **Estabilidade:** $C_1(h_1) \times C_2(h_2)$ é uma função de covariância.

Logo, no contexto espaço-temporal tal que se satisfaça

$$C_{st}(h_s; h_t) = C_s(h_s)C_t(h_t), \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.20)$$

é uma função de covariância separável, em que

- $C_s(h_s)$ é uma função de covariância puramente espacial;
- $C_t(h_t)$ é uma função de covariância puramente temporal.

De forma equivalente, C_{st} é separável, desde que seja possível sua fatoração na seguinte forma

$$C_{st}(h; h_t) = \frac{C_{st}(h_s; 0)C_{st}(0; h_t)}{C_{st}(0; 0)}, \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.21)$$

sendo assim o processo de separação obtida de forma conjunta (MITCHELL; GENTON; GUMPERTZ, 2005).

Outra implicação da separabilidade é a de que a função de correlação espaço-temporal satisfaz a seguinte equação

$$\rho_{st}(h_s; h_t) = \rho_{st}(h_s; 0)\rho_{st}(0; h_t), \quad (h_s; h_t) \in \mathbb{R}^{d+1}. \quad (2.22)$$

2.1.5 Simetria total

A simetria total é dada de tal forma que se a densidade espectral existe, e seus fatores são produtos dos domínios espaciais e temporais. Uma função de covariância espaço-temporal estacionária é totalmente simétrica se

$$C_{st}(h_s; h_t) = C_{st}(h_s; -h_t) = C_{st}(-h_s; h_t) = C_{st}(-h_s; -h_t), \quad (h_s, h_t) \in \mathbb{R}^{d+1}. \quad (2.23)$$

No contexto puramente espacial esta propriedade é conhecida como simetria relativa ao eixo (SCACCIA; MARTIN, 2005) ou simetria refletida (LU; ZIMMERMAN, 2005). Ainda, pode-se especificar o Teorema 2, equação (2.15), para atender esta propriedade.

Teorema 3 (Teorema de Bochner para simetria total) *Suponha que C_{st} é uma função contínua definida em \mathbb{R}^{d+1} . Então, C_{st} é uma função espaço-temporal contínua estacionária e totalmente simétrica, se e somente se, pode ser escrita na forma*

$$C_{st}(h_s; h_t) = \int \int \cos(h'_s \omega) \cos(h_t \tau) dF(\omega, \tau), \quad (h_s, h_t) \in \mathbb{R}^{d+1}, \quad (2.24)$$

em que F é uma medida finita, não-negativa definida em \mathbb{R}^{d+1} .

O resultado do Teorema 3 é apresentado no Anexo A. Se a função C_{st} é totalmente simétrica, a densidade espectral f existe, então f é também totalmente simétrica, ou seja,

$$f(\omega; \tau) = f(-\omega; \tau) = f(\omega; -\tau) = f(-\omega; -\tau), \quad (\omega; \tau) \in \mathbb{R}^{d+1}. \quad (2.25)$$

Ainda, se a função de covariância espaço-temporal possui estruturas adicionais, por exemplo, simetria esférica (EINSTEIN, 1939), pode-se especializar ainda mais a expressão (2.15) do Teorema 2 para comportar demais especificidades, como feito no Teorema 3.

Todos os resultados apresentados são aplicados em funções contínuas, porém funções de covariâncias ajustadas, muitas vezes, envolvem um efeito pepita, ou seja, uma descontinuidade na origem (GNEITING; GENTON; GUTTORP, 2006). No contexto, espaço-temporal, este efeito pode ser: puramente espacial, puramente temporal, ou ainda espaço-temporal.

2.1.6 Isotropia e Anisotropia

De forma semelhante ao caso da estacionariedade espacial (2.18), a isotropia espacial dá-se por

$$\text{cov}(Z(s; t); Z(hs; h_t)) \equiv C_{st}(|s - h_s|; t). \quad (2.26)$$

Ainda, a anisotropia espacial geométrica é dada por

$$\text{cov}(Z(s; t); Z(hs; h_t)) \equiv C_{st}(\{p_1(s_1 - h_{s1})^2 + \dots + p_d(s_d - h_{sd})^2\}^{1/2}; t, h_t), \quad (2.27)$$

em que $p_1 > 0, \dots, p_d > 0$. Já a anisotropia espaço-temporal geométrica é dada por

$$\text{cov}(Z(s; t); Z(hs; h_t)) \equiv C_{st}(\{p_1(s_1 - h_{s1})^2 + \dots + p_d(s_d - h_{sd})^2 + q(t - h_t)^2\}^{1/2}), \quad (2.28)$$

em que $p_1 > 0, \dots, p_d > 0$ e $q > 0$.

Para modelos anisotrópicos são feitas construções como combinações lineares de modelos isotrópicos, em que os “comprimentos” são obtidos por transformações *affine* (DONG; GALATSANOS, 2002).

2.1.7 Outras propriedades

2.2 Covariância

Como no caso espacial, o principal objetivo de modelar a semivariância espaço-temporal é quantificar a dependência contida nos dados. O modelo resultante é utilizado para predi-

ção temporal e/ou interpolação espacial. Porém, algumas limitações surgem, por exemplo, diferente da aplicação puramente espacial onde a estimação é o passo mais difícil, um semivariograma espaço-temporal precisa ser uma função condicionalmente negativa definida (DYN; GOODMAN; MICCHELLI, 1986), que pode ser uma característica de difícil verificação em muitos casos (MYERS, 2004).

Ainda, técnicas de redução de dimensionalidade não funcionam bem para modelos espaços-temporais, pois se o “tempo” é meramente considerado mais uma dimensão seria necessário uma métrica apropriada no espaço-tempo (BILONICK, 1985), visto que toda a estrutura de dependência poderia ser desconsiderada ou até desconfigurada. Assim, uma alternativa simples é uma técnica de separabilidade, visto que uma maneira mais fácil de modelar o espaço-tempo é separar as dependências espaciais e temporais (MYERS, 2004), entretanto esta prática traz algumas desvantagens, como a falta de um termo que considere o espaço e tempo de forma conjunta.

É importante notar que, em geral, a soma das semivariâncias espacial e temporal não será uma função positiva estritamente definida e neste caso a matriz das equações de krigagem pode não ser invertível para alguns pontos (MYERS; JOURNEL, 1990). Enquanto o produto de 2 covariâncias será uma covariância, em geral, o produto de 2 semivariâncias não será uma semivariância (CESARE; MYERS; POSA, 2001).

Desta forma, Cesare, Myers e Posa (2001) expõem algumas das classes de modelos de covariância espaço-temporais amplamente utilizadas na literatura. Assim, seja $Z = \{Z(s; t), (s; t) \in D \times T\}$ um campo aleatório espaço-temporal estacionário de segunda ordem, em que $D \subset \mathbb{R}^d$ e $T \subset \mathbb{R}_+^1$, com valor esperado expresso por

$$E(Z(s; t)) = 0. \quad (2.29)$$

Neste caso, a covariância espaço-temporal é dada por

$$C_{st}(h_s; h_t) = Cov[Z(s + h_s; t + h_t), Z(s; t)], \quad (2.30)$$

em que $(h_s; h_t), (s; s + h_s) \in D^2$ e $(t; t + h_t) \in T^2$, sendo h_s e h_t incrementos no espaço e tempo, respectivamente.

Além disso, funções marginais podem ser calculadas, analogamente as marginais de funções de probabilidade, sendo $C_{st}(h_s; 0)$ e $C_{st}(0; h_t)$ as respectivas covariâncias marginais espacial e temporal, do mesmo modo $\gamma_{st}(h_s, 0)$ e $\gamma_{st}(0, h_t)$ são as respectivas semivariâncias marginais espacial e temporal.

A função de covariância espaço-temporal em (2.30) necessita satisfazer a condição de ser uma função positiva definida, a fim de ser uma função de covariância válida, ou seja, para

todo $(s_i, t_i) \in D \times T$, $a_i \in \mathbb{R}$, $i = 1, \dots, n$, e todo inteiro positivo n , $C_{st}(h_s; h_t)$ precisa satisfazer a expressão (2.4).

Para funções contínuas, ser positiva definida é equivalente ao processo ter uma função de distribuição espectral (MATÉRN, 1960).

Então, as funções de covariância e semivariância espaço-temporais precisam ser positivas definidas e estritamente negativas condicionalmente definidas, respectivamente. Assim, necessita-se de modelos que satisfaçam estas restrições.

Seja $Z(s_i; t_i)$ as observações coletadas no espaço-tempo, o estimador de semivariância empírico é dado por

$$2\hat{\gamma}_{st}(h_s; h_t) \equiv \frac{1}{N(h_s; h_t)} \sum_{i,j,t,t' \in N(h_s; h_t)} [Z(s_i; t) - Z(s_j; t)]^2, \quad (2.31)$$

em que

- $N(h_s; h_t) \equiv \{(i, j, t, t') : s_i - s_j \in \text{tol}(h_s); |t - t'| = h_t; i, j = 1, \dots, n\}$;
- $\text{tol}(h_s)$ é uma região de tolerância especificada em torno de h_s ;
- $|N(h_s; h_t)|$ é o número de elementos distintos em $N(h_s; h_t)$.

Então, o parâmetro θ da função de covariância espaço-temporal pode ser estimada ao ajustar o estimador (2.31) à semivariância espaço-temporal,

$$\begin{aligned} 2\gamma(h_s; h_t | \theta) &\equiv \text{var}(Z(s + h_s; t + h_t)) - Z(s; h_t) \\ &\equiv C_{st}(0; 0 | \theta) - C_{st}(h_s; h_t | \theta); \quad (h_s; h_t) \in \mathbb{R}^{(d+1)}. \end{aligned} \quad (2.32)$$

E pode-se utilizar o método de mínimos quadrados ponderados (CRESSIE, 2015), ao minimizar

$$W(\theta) \equiv \sum_{i,j,t,t' \in N(h_s; h_t)} |N(h_s; h_t)| \left(\frac{\hat{\gamma}(h_s; h_t)}{\gamma(h_s; h_t) | \theta} - 1 \right)^2. \quad (2.33)$$

Não só isso, como Zimmerman e Zimmerman (1991) mostram em estudo de simulação que ao se comparar o método (2.33) com métodos baseados em verossimilhança, o baseado em mínimos quadrados ponderados é às vezes o melhor, e nunca obteve uma performance ruim, além de ser computacionalmente mais rápido. Além disso, comparado com o método de mínimos quadrados ordinários, este é em geral superior para os diversos cenários simulados.

Seja então $Z = \{Z(s; t), (s; t) \in D \times T\}$ um campo aleatório espaço-temporal observado no espaço-tempo com n coordenadas $(s_1; t_1), \dots, (s_n; t_n)$. Assume-se que um processo espaço-temporal $Z(\cdot; \cdot)$ satisfaz a condição de regularidade, ou seja,

$$\text{var}(Z(s; t)) < \infty, \quad \forall s \in D, \quad t \geq 0. \quad (2.34)$$

Assim, pode-se definir sua média como

$$\mu(s; t) \equiv E(Z(s; t)), \quad (2.35)$$

e sua função de covariância como (2.5). Desta forma, o preditor linear ótimo (TOU-TENBURG, 1982) de $Z(s_0; t_0)$ é dado por

$$Z^*(s_0; t_0) = \mu(s_0; t_0) + c(s_0; t_0)' \Sigma^{-1} (Z - \mu), \quad (2.36)$$

em que

- $\Sigma \equiv C_{st}(Z)$;
- $c(s_0; t_0) \equiv C_{st}(Z(s_0; t_0), Z)$;
- $\mu(s; t) \equiv E(Z(s; t))$;
- $c(s_0; t_0)' \Sigma^{-1} c(s_0; t_0)$ é o erro quadrado médio mínimo da predição, do inglês, *minimum mean squared prediction error* (MSPE).

Também, assume-se que a função de covariância é estacionária no tempo e espaço, assim

$$\text{cov}(s, r; t, q) = C_{st}(s - h_s; t - h_t), \quad (2.37)$$

esta suposição é usualmente feita a fim de garantir que a covariância possa ser estimada dos dados. Logo, esta função C_{st} precisa satisfazer a condição de ser positiva definida para ser uma função de covariância válida.

Só assim (2.36) é um preditor espaço-temporal válido de $Z(s_0; t_0)$ com MSPE não-negativo.

A fim de garantir a condição de positividade definida, assume-se que C_{st} é uma função de covariância que pertence à uma família paramétrica, cujo os membros são positivos definidos, ou seja,

$$\text{cov}(Z(s; t), Z(s + h_s; t + h_t)) = C_{st}(h_s; h_t | \theta), \quad (2.38)$$

em que C_{st} satisfaz (2.4) para todo $\theta \in \Theta \subset \mathbb{R}^p$.

Uma condição semelhante existe para o semivariograma espaço-temporal, que é uma função negativa definida, tal que a seguinte inequação seja satisfeita

$$\sum_{i=1}^n \sum_{j=1}^n \gamma_{st}(s_i - s_j; t_i - t_j) a_i a_j \leq 0, \quad (2.39)$$

para qualquer inteiro $n \geq 2$, qualquer $(s_1; t_1), \dots, (s_n; t_n) \in \mathbb{R}^{d+1}$, e reais a_1, \dots, a_n , tal que $\sum_{i=1}^n a_i = 0$.

Como visto, para garantir que $C_{st}(\cdot; \cdot)$ seja uma função de covariância espaço-temporal positiva definida, é usual especificar que pertença à uma família paramétrica, tal que seus membros sejam positivos definidos. Assim, algumas vertentes são possíveis, uma delas consiste em modelo separáveis, ou seja, entender a covariância espaço-temporal como combinação das funções de covariâncias puramente espacial e temporal, podendo ser feito via adição, produto ou derivações destas operações (subseção 2.2.1). Ainda, algumas vertentes não-separáveis são possíveis e abordadas na subseção 2.2.2, sendo que Cressie e Huang (1999) deram início, ou pelo menos popularizaram esse tipo de modelo com sua representação espectral, como detalhado na subsubseção 2.2.2.1.

2.2.1 Covariâncias separáveis

2.2.1.1 Modelo *Sum*

Talvez o mais simples, o modelo *sum*, também conhecido como: *linear* ou *zonal*, consiste em outro uso da técnica de separabilidade, porém de forma aditiva, ou seja,

$$C_{st}(h_s; h_t) = C_t(h_t) + C_s(h_s), \quad (2.40)$$

e

$$\gamma_{st}(h_s; h_t) = \gamma_t(h_t) + \gamma_s(h_s). \quad (2.41)$$

Este modelo considera o tempo e espaço de forma completamente separada, assim alguns problemas surgem, por exemplo, as matrizes de covariância com certas configurações espaço-temporais são singulares, neste caso, a função de covariância é somente positiva semidefinida e não produz uma precisão ótima (CESARE; MYERS; POSA, 2001), o que não é satisfatório para (2.36).

Uma das propriedades do modelo *sum* é que dado sua semivariância (2.41), qualquer uma de suas marginais pode não ter limites, desta forma como é um modelo de classe separável uma anisotropia geométrica pode ser incorporada no componente espacial. Porém, há risco da matriz coeficiente no processo de *krigagem* (Seção 2.3) ser não-invertível (MYERS, 2004)

Entretanto, este modelo é próximo de ser aceitável, pois só as vezes falha na suposição de ser uma função estritamente positiva definida. Assim, uma questão que surge é: existe uma função $G(u, v)$ em \mathbb{R}^2 , tal que

$$\gamma_s(h_s) + \gamma_t(h_t) + G[\gamma_s(h_s); \gamma_t(h_t)], \quad (2.42)$$

seja estritamente negativa condicionalmente definida? Se sim, é possível a construção de um modelo aceitável, um exemplo, é o modelo *sum-metric* (Parágrafo 2.2.1.2.1).

2.2.1.2 Modelo *Metric*

Seja $C_{st}(\cdot)$ uma função estritamente positiva definida em \mathbb{R}^{d+1} e $\gamma(\cdot)$ uma função estritamente negativa condicionalmente definida em \mathbb{R}^{d+1} . Se $\mathbb{R}^d \times \mathbb{R}$ é considerado como um espaço $d + 1$ dimensional, então

$$C_{st}(h_s; h_t) = C(|h_s| + \alpha|h_t|), \quad (2.43)$$

em que α é um termo de anisotropia. E através de (2.11),

$$\gamma_{st}(h_s; h_t) = \gamma(|h_s| + \alpha|h_t|), \quad (2.44)$$

são modelos válidos em $\mathbb{R}^d \times \mathbb{R}$, tal que $\alpha > 0$.

Ainda, segundo Dimitrakopoulos e Luo (1994) ao invés do uso do termo $|h_s| + \alpha|h_t|$ nas expressões (2.43) e (2.44) pode-se utilizar de forma equivalente $|h_s|^2 + \alpha|h_t|^2$.

Independente da notação, uma das restrições presentes no modelo *metric* está que ambos seus patamares marginais serão iguais, em que só o alcance muda. Assim, torna-se muito suscetível a escolha de uma melhor alternativa, como sua derivação, o modelo *sum-metric*.

2.2.1.2.1 Sum-metric

Partindo da ideia apresentada na expressão (2.42), em que o modelo *sum* é atribuído com a função $G(u, v)$. Logo, a alternativa para melhora do modelo *sum* é combiná-lo com o modelo *metric*, ou seja,

$$\gamma_{st}(h_s; h_t) = \underbrace{\gamma_t(h_t) + \gamma_s(h_s)}_{\text{modelo sum}} + \underbrace{\gamma(|h_s| + \alpha|h_t|)}_{\text{modelo metric}}, \quad (2.45)$$

ou seguindo a notação de [Dimitrakopoulos e Luo \(1994\)](#),

$$\gamma_{st}(h_s; h_t) = \underbrace{\gamma_t(h_t) + \gamma_s(h_s)}_{\text{modelo sum}} + \underbrace{\gamma(|h_s|^2 + a|h_t|^2)}_{\text{modelo metric}}. \quad (2.46)$$

2.2.1.3 Modelo *Product*

O modelo *product* consiste também na ideia de separar as covariâncias em duas dependências. Assim, seja $Z_1(s)$ e $Z_2(t)$ variáveis aleatórias não-correlacionadas definidas em \mathbb{R}^d e \mathbb{R} , respectivamente, com funções de covariância estritamente positivas definidas. Desta forma, a função de covariância de $Z(s, t) = Z_1(s)Z_2(t)$, é dada por

$$C_{st}(h_s; h_t) = C_s(h_s)C_t(h_t). \quad (2.47)$$

Se $C_t(h_t)$ e $C_s(h_s)$ são funções estritamente positivas definidas em $\mathbb{R}^d \times \mathbb{R}$ o modelo dado por (2.47) pode ser escrito em termos da semivariância espaço-temporal

$$\gamma_{st}(h_s; h_t) = C_t(0)\gamma_s(h_s) + C_s(0)\gamma_t(h_t) - \gamma_s(h_s)\gamma_t(h_t), \quad (2.48)$$

em que

- γ_{st} é a semivariância espaço-temporal;
- γ_t é a semivariância puramente temporal;
- γ_s é a semivariância puramente espacial;
- C_t é a covariância puramente temporal;
- C_s é a covariância puramente espacial.

Isto é possível, pois C_s é uma função positiva definida em \mathbb{R}^d e C_t é também uma função positiva definida em \mathbb{R}_+ .

Como os modelos de covariância temporais e espaciais já estão disponíveis, pode-se utilizar dos modelos para covariância espaciais e temporais já conhecidos na literatura ([MATÉRN, 1960](#); [YAMAMOTO](#); [LANDIM, 2014](#); [CRESSIE, 2015](#)) e combiná-los como visto em (2.48).

Ainda, em termos marginais, pode-se reescrever a semivariância, equação (2.48), como:

$$\gamma_{st}(h_s; h_t) = \gamma_{st}(h_s; 0) + \gamma_{st}(0; h_t) - \left[\frac{\gamma_{st}(h_s; 0)\gamma_{st}(0; h_t)}{C_t(0)C_s(0)} \right]. \quad (2.49)$$

Entretanto, o modelo *product* é limitado, pois para cada par de coordenadas espaciais a função de covariância de duas séries temporais tem a mesma forma. Assim, para dois *lags* espaciais fixos, h_{s1} e h_{s2} , tem-se que

$$C_{st}(h_{s1}; h_t) \propto C_{st}(h_{s2}; h_t). \quad (2.50)$$

Analogamente, um resultado similar acontece na situação contrária para um par de pontos no tempo e a função de covariância cruzada de dois processos espaciais, logo para dois *lags* temporais fixos, h_{t1} e h_{t2} , tem-se que

$$C_{st}(h_s; h_{t1}) \propto C_{st}(h_s; h_{t2}). \quad (2.51)$$

2.2.1.4 Modelo *Product-Sum*

O modelo *product-sum* combina os modelos *product* (subsubseção 2.2.1.3) e *sum* (subsubseção 2.2.1.1) mesclando os produtos e somas, sendo obtido pela expressão seguinte

$$C_{st}(h_s; h_t) = \underbrace{k_1 C_s(h_s) C_t(h_t)}_{\text{modelo product}} + \underbrace{k_2 C_s(h_s) + k_3 C_t(h_t)}_{\text{modelo sum}}, \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.52)$$

em que $k_1 \geq 0$, $k_2 \geq 0$ e $k_3 \geq 0$. Novamente, se as covariâncias marginais são funções estritamente positivas definidas, logo, a função conjunta também é. Analogamente, pode-se escrever a semivariância da seguinte forma

$$\gamma_{st}(h_s; h_t) = [k_1 + k_2 C_t(0)] \gamma_s(h_s) + [k_1 + k_3 C_s(0)] \gamma_t(h_t) - k_1 \gamma_s(h_s) \gamma_t(h_t), \quad (2.53)$$

em que

- γ_{st} é a semivariância espaço-temporal;
- γ_t é a semivariância puramente temporal;
- γ_s é a semivariância puramente espacial;
- C_t é a covariância puramente temporal;
- C_s é a covariância puramente espacial.

Ainda, pode-se reescrever a equação (2.53) na forma

$$\gamma_{st}(h_s; h_t) = \gamma_{st}(h_s; 0) + \gamma_{st}(0; h_t) - k \gamma_{st}(h_s; 0) \gamma_{st}(0; h_t). \quad (2.54)$$

E como mostrado por Iaco, Myers e Posa (2001) a condição suficiente para k é dada pelo Teorema 4.

Teorema 4 *Seja $Z(s; t)$ um campo aleatório espaço-temporal estacionário de segunda ordem. Suponha que C_{st} é uma função de covariância espaço-temporal contínua e possui a forma dada por (2.52). Assim, supondo o semivariograma espaço-temporal definido por (2.54). Então, $k_1 > 0$, $k_2 \geq 0$ e $k_3 \geq 0$, se e somente, k satisfaz a seguinte inequação:*

$$0 < k \leq \frac{1}{\max(\text{sill}[\gamma_{st}(h_s; 0)], \text{sill}[\gamma_{st}(0; h_t)])}, \quad (2.55)$$

em que $\text{sill}(\cdot)$ é função que retorna o patamar do semivariograma.

Portanto, se k satisfaz (2.55), então o semivariograma expresso por (2.53) é válido, pois se não há limites não é possível calcular k , como vistos nos corolários 4.1 e 4.2.

Corolário 4.1 *Se tanto $\gamma_{st}(h_s; 0)$ ou $\gamma_{st}(0; h_t)$ não possuem limites, não há escolha de k que satisfaça a expressão (2.55), tal que o semivariograma dado por (2.53) é um semivariograma espaço-temporal inválido.*

Corolário 4.2 *Se tanto $\gamma_s(h_s)$ ou $\gamma_t(h_t)$ não possuem limites, o semivariograma dado por (2.69) é um semivariograma espaço-temporal inválido, para quaisquer valores dos coeficientes k_1 , k_2 e k_3 .*

Para modelagem e estimação deste modelo, tem-se uma condição implícita da transformação da covariância para semivariância, que consiste em

$$k_1 C_s(0) C_t(0) + k_2 C_s(0) + k_3 C_t(0) = C_{st}(0; 0). \quad (2.56)$$

Além disso, da expressão (2.53), sabe-se que

$$\gamma_{st}(h_s; 0) = [k_2 + k_1 C_t(0)] \gamma_s(h_s) = k_s \gamma_s(h_s), \quad (2.57)$$

$$\gamma_{st}(0; h_t) = [k_3 + k_1 C_s(0)] \gamma_t(h_t) = k_t \gamma_t(h_t). \quad (2.58)$$

Também, assume-se que

$$k_2 + k_1 C_t(0) = 1, \quad (2.59)$$

e

$$k_3 + k_1 C_s(0) = 1, \quad (2.60)$$

em ordem de estimar e modelar $\gamma_s(h_s)$ e $\gamma_t(h_t)$ por $\gamma_{st}(h_s; 0)$ e $\gamma_{st}(0; h_t)$, respectivamente.

Por fim, vê-se que da expressão (2.52) que $k_1 \geq 0$, $k_2 \geq 0$ e $k_3 \geq 0$ são condições suficientes para garantir que a função seja positiva definida. Logo, das equações (2.56), (2.59) e (2.60) obtêm-se

$$k_1 = [k_s C_s(0) + k_t C_t(0) - C_{st}(0, 0)] / C_s(0) C_t(0). \quad (2.61)$$

$$k_2 = [C_{st}(0, 0) - k_t C_t(0)] / C_s(0). \quad (2.62)$$

$$k_3 = [C_{st}(0, 0) - k_s C_s(0)] / C_t(0). \quad (2.63)$$

E na modelagem de semivariâncias temporais e espaciais separadas é necessário garantir que os patamares sejam escolhidos de forma que os numeradores satisfaçam as condições das expressões (2.61), (2.62) e (2.63).

Além disso, como visto em Iaco, Myers e Posa (2001), o modelo *product-sum* possui resultados interessantes quanto ao valor dos patamares. Utilizando da ligação entre a semivariância e sua correspondente covariância (2.11), segue que

$$\gamma_{st}(h_s; 0) = C_{st}(0; 0) - C_{st}(h_s; 0), \quad (2.64)$$

e

$$\gamma_{st}(0; h_t) = C_{st}(0; 0) - C_{st}(0; h_t). \quad (2.65)$$

Teorema 5 *Seja Z um campo aleatório espaço-temporal estacionário de segunda ordem. Suponha que a função de covariância espaço-temporal possui a forma expressa em (2.52) e é contínua. Então, utilizando-se dos resultados (2.57) e (2.58), tem-se que*

$$\lim_{h_s \rightarrow \infty} \lim_{h_t \rightarrow \infty} \gamma_{st}(h_s; h_t) = C_{st}(0; 0), \quad (2.66)$$

$$\lim_{h_s \rightarrow \infty} \lim_{h_t \rightarrow \infty} \gamma_{st}(h_s; 0) = k_s C_s(0), \quad (2.67)$$

e

$$\lim_{h_s \rightarrow \infty} \lim_{h_t \rightarrow \infty} \gamma_{st}(0; h_t) = k_t C_t(0). \quad (2.68)$$

Assim, examinando os comportamentos assintóticos de $\gamma_{st}(h_s; h_t)$, $\gamma_{st}(h_s; 0)$ e $\gamma_{st}(0; h_t)$ pelo Teorema 5 vê-se que os semivariogramas têm patamares distintos (em que a prova do Teorema 5 encontra-se no Anexo C).

Ainda, é possível a simplificação da semivariância, através das expressões (2.52), (2.57) e (2.58), isto é,

$$\gamma_{st}(h_s; h_t) = \gamma_{st}(h_s; 0) + \gamma_{st}(0; h_t) - k[\gamma_{st}(h_s; 0)\gamma_{st}(0; h_t)], \quad (2.69)$$

em que

$$k = \frac{k_1}{k_t k_s}. \quad (2.70)$$

Assim, aplicando a expressão (2.61) em (2.70) tem-se que

$$k = \frac{[k_s C_s(0) + k_t C_t(0) - C_{st}(0, 0)]}{C_s(0)C_t(0)k_t k_s}. \quad (2.71)$$

Portanto, no processo de estimação os termos separáveis já serão obtidos pelos semivariogramas espacial e temporal, restando assim o patamar espaço-temporal ($C_{st}(0, 0)$) para estimação de k . Iaco, Myers e Posa (2001) estabelecem o teorema e a condição necessária e suficiente, em termos de fronteira em k , para garantir a admissibilidade do semivariograma na forma simplificada como mostrado em (2.54).

Por fim, como este é o modelo separável mais complexo e utilizado, explicita-se a modelagem, visto que a mesma é aplicável nos modelos anteriores. Assim, seja H um conjunto de coordenadas, dado por

$$H = (s_i; t_j), \quad i = 1, \dots, n_s, \quad j = 1, \dots, n_t, \quad (2.72)$$

a semivariância espacial no vetor é expressa por

$$\hat{\gamma}_s(h_s) = \hat{\gamma}_{st}(h_s, 0) = \frac{1}{2|N(h_s)|} \sum_{N(h_s)} [Z(s + h_s; t) - Z(s; t)]^2, \quad (2.73)$$

em que

- $N(h_s) = \{(s+h_s; t) \in H \text{ e } (s; t) \in H \text{ é o conjunto espacial, tal que } \|s-h_s\| < \text{tol}(h_s)\}$;

- $|N(h_s)|$ é a cardinalidade do conjunto espacial;
- $\text{tol}(h_s)$ é a tolerância espacial.

Analogamente,

$$\hat{\gamma}_t(h_t) = \hat{\gamma}_{st}(0; h_t) = \frac{1}{2|N(h_t)|} \sum_{N(h_t)} [Z(s + h_s; h_t) - Z(s, t)]^2, \quad (2.74)$$

em que

- $N(h_t) = \{(s; t + h_t) \in H \text{ e } (s; t) \in H, \text{ tal que } \|t - h_t\| < \text{tol}(h_t)\}$;
- $|N(h_t)|$ é a cardinalidade do conjunto temporal;
- $\text{tol}(h_t)$ é a tolerância temporal.

Posteriormente, faz-se a seleção de modelos válidos para semivariâncias espacial e temporal, que possuam patamar, e assim as estimativas para $k_s C_s(0)$ e $k_t C_t(0)$ são obtidas.

Após, o semivariograma espaço-temporal amostral (GETHING et al., 2007), análogo ao caso especial (1.7), é expresso por,

$$\hat{\gamma}_{st}(h_s; h_t) = \frac{1}{|N(h_s; h_t)|} \sum_{N(h_s; h_t)} [Z(s + h_s; t + h_t) - Z(s; t)]^2, \quad (2.75)$$

em que

- $N(h_s; h_t) = \{(s + h_s, t + h_t) \in H, (s, t) \in H : \|s - h_s\| < \text{tol}(h_s) \text{ e } \|t - h_t\| < \text{tol}(h_t)\}$ é o conjunto espaço-temporal;
- $|N(h_s; h_t)|$ é a cardinalidade do conjunto espaço-temporal.

Assim, com patamares puramente temporal e espacial estimados, faz-se o cálculo da estimativa do patamar espaço-temporal, ao se ajustar o semivariograma amostral ao teórico (2.53), neste caso o patamar $C_{st}(0; 0)$ pode ser estimado ao minimizar $W(C_{st}(0; 0))$, valor dado pelo método de mínimos quadrados ponderados, (2.33), isto é,

$$W(C_{st}(0; 0)) = \sum_s^{L_s} \sum_t^{L_t} |L(h_s; h_t)| \left[\frac{\hat{\gamma}_{st}(s, h_s; t, h_t) - \gamma_{st}(h_s; h_t | C_{st}(0; 0))}{\gamma_{st}(h_s; h_t | C_{st}(0; 0))} \right]^2, \quad (2.76)$$

em que

- L_s é o número de *lags* espaciais;
- L_t é o número de *lags* temporais.

2.2.1.4.1 *Exponential Product-Sum*

Um caso especial do modelo *product-sum* é quando a covariância é expressa em forma da função exponencial em um modelo *metric*. Assim, na equação do modelo *metric* expressa em (2.43), seja $C_{st}(\cdot) = \exp(-\cdot)$, tem-se que

$$\begin{aligned} C_{st}(h_s; h_t) &= \exp[-(|h_s| + a|h_t|)] \\ &= \exp(-|h_s| - a|h_t|) \\ &= \exp(-|h_s|) \exp(-a|h_t|) \\ &= C_{st}(|h_s|)C_{st}(a|h_t|). \end{aligned} \quad (2.77)$$

Ou seja, o modelo *metric* é o produto de dois modelos exponenciais. Assim, pode-se aplicar a generalização da semivariância do modelo *product* expresso por (2.48), isto é,

$$\gamma_{st}(h_s; h_t) = A[1 - \exp(-|h_s|)] + B[1 - \exp(-a|h_t|)] - KAB[1 - \exp(-|h_s|)][1 - \exp(-a|h_s|)]. \quad (2.78)$$

Ainda, se $A = B$ e $K = 1/A$ é o modelo *metric*, pois

$$\begin{aligned} \gamma_{st}(h_s; h_t) &= A[1 - \exp(-|h_s|)] + B[1 - \exp(-a|h_t|)] - KAB[1 - \exp(-|h_s|)][1 - \exp(-a|h_s|)] \\ &= A[1 - \exp(-|h_s|)] + A[1 - \exp(-a|h_t|)] - \frac{A^2}{A}[1 - \exp(-|h_s|)][1 - \exp(-a|h_s|)] \\ &= A[1 - \exp(-|h_s|) + 1 - \exp(-a|h_t|)] - A[1 - \exp(-|h_s|)][1 - \exp(-a|h_s|)] \\ &= A[2 - \exp(-|h_s|) - \exp(-a|h_t|)] - A[1 - \exp(-|h_s|)][1 - \exp(-a|h_s|)]. \end{aligned} \quad (2.79)$$

Analogamente, o mesmo vale para o modelo *metric* dado pela notação de [Dimitrakopoulos e Luo \(1994\)](#).

2.2.1.4.2 *Integrated Product-Sum*

Já uma extensão do modelo *product-sum* está na integração com uma função de probabilidade. Assim, seja $f(u)$ uma função densidade de probabilidade em \mathbb{R}^k e $\gamma_{st}(h_s; h_t|\alpha)$ uma função estritamente negativa condicionalmente definida em \mathbb{R}^{d+1} para cada α no suporte de $f(u)$. Então

$$\int \gamma_{st}(h_s; h_t|\alpha) f(\alpha) d\alpha, \quad (2.80)$$

é também uma uma função estritamente negativa condicionalmente definida em \mathbb{R}^{d+1} .

2.2.1.4.3 Product-sum via Cressie e Huang

Com a representação de [Cressie e Huang \(1999\)](#) explicitada na Sub-seção 2.2.2.1, assume-se que a função de covariância espaço-temporal é integrável, a fim de garantir uma representação com a transformada de *Fourier*. Garantindo essa nova condição, é possível mostrar que $k_2 \neq 0$ ou $k_3 \neq 0$ então, independente do modelo escolhido para os modelos temporais e espaciais, o modelo espaço-temporal não é integrável. Entretanto, é possível generalizar o modelo *product-sum* na seguinte forma

$$\gamma_{st}(h_s; h_t) = [k_2 + k_1 C_t(0)]\gamma_s(h_s) + [k_3 + k_1 C_s(0)]\gamma_t(h_t) - k_1 \gamma_s(h_s)\gamma_t(h_t) + k_4 \gamma_{1s}(h_s) + k_5 \gamma_{2t}(h_t), \quad (2.81)$$

em que

- $\gamma_s(h_s)$ é um semivariograma espacial válido, com fronteiras;
- $\gamma_t(h_t)$ é um semivariograma temporal válido, com fronteiras;
- $\gamma_{1s}(h_s)$ é um semivariograma espacial válido, não necessariamente com fronteiras;
- $\gamma_{2t}(h_t)$ é um semivariograma temporal válido, não necessariamente com fronteiras;
- $k_i, \quad i = 1, \dots, 5$, são coeficiente não-negativos.

2.2.1.5 Modelo de Ma

[Ma \(2002\)](#) propôs um modelo de misturas em termos das funções de correlação. Seja $\rho_s(h_s) \times \rho_t(h_t)$ as funções correspondentes as funções de covariâncias expressa em (2.47). O modelo proposto é dado por

$$\rho_{st}(h_s; h_t) = \sum_{[i=0, \dots]} \sum_{[j=0, \dots]} [\rho_s(h_s)]^i \times [\rho_t(h_t)]^j p_{ij}, \quad (2.82)$$

em que p_{ij} é a densidade de probabilidade bivariada nos positivos inteiros. Ainda, a expressão (2.82) pode ser considerada um caso especial de (2.80) dado as seguintes substituições

$$[\rho_s(h_s)]^i = \left[1 - \frac{\gamma_s(h_s)}{C_s(0)} \right]^i, \quad (2.83)$$

e

$$[\rho_t(h_t)]^j = \left[1 - \frac{\gamma_t(h_t)}{C_t(0)} \right]^j. \quad (2.84)$$

Alternativamente, pode-se considerar

$$\begin{aligned}\rho_{st}(h_s; h_t | (i, j)) &= [\rho_s(h_s)]^i \times [\rho_t(h_t)]^j \\ &= \left[1 - \frac{\gamma_{st}(h_s; h_t | (i, j))}{[C_s(h_s)]^i [C_t(h_t)]^j} \right].\end{aligned}\quad (2.85)$$

Como $\rho_s(h_s)$ e $\rho_t(h_t)$ podem ser assumidos como funções estritamente positivas definidas, $\rho_{st}(h_s, h_t; (i, j))$ será uma função também estritamente positiva definida para cada par $(i, j) = \alpha$. Logo, o semivariograma correspondente $\gamma_{st}(h_s, h_t; (i, j))$ será uma função estritamente negativa condicionalmente definida para cada par (i, j) , atendendo as condições necessárias para a o modelo. Por fim, neste caso a expressão (2.80) seria reduzida a um somatório duplo.

2.2.2 Covariâncias não-separáveis

Modelos de covariância derivados de separação são utilizados muito mais por conveniência, já que satisfazem (2.4), do que pelo próprio ajuste, visto que possuem diversas limitações (CRESSIE; HUANG, 1999). Por exemplo, uma das maiores falhas destes modelos está em não possuírem um termo que seja função de h_s e h_t , tal que não sejam nulos, visto que pelo ponto de vista empírico não há plausibilidade.

Assim, visto as limitações dos modelos separáveis (Subseção 2.2.1) cria-se uma demanda por modelos não-separáveis.

Jones e Zhang (1997) utilizando-se de simples equações diferenciais parciais estocásticas para o espaço-tempo desenvolveram uma família de densidade espectral com 4 parâmetros, que implicitamente suporta as funções de covariância estacionárias, mesmo que não em forma fechada.

2.2.2.1 Modelo de Cressie e Huang

Cressie e Huang (1999) desenvolveram uma nova metodologia com novas classes de modelos não-separáveis, sob a suposição adicional de integrabilidade.

Teorema 6 *Suponha que C_{st} é uma função de covariância espaço-temporal contínua, limitada, integrável e simétrica em \mathbb{R}^{d+1} . Então, C_{st} é uma covariância estacionária, se e somente se,*

$$\rho(\omega; h_t) = \int e^{-ih'_s \omega} C_{st}(h_s; h_t) dh_t, \quad h_t \in R, \quad (2.86)$$

é positiva definida para quase todo $\omega \in \mathbb{R}^d$.

Utilizando-se do Teorema 6, Cressie e Huang (1999) construíram funções de covariância espaço-temporais estacionárias com forma fechada pela transformação inversa de *Fourier* em \mathbb{R}^{d+1} . Para tal, considerou-se funções expressas por

$$C_{st}(h_s; h_t) = \int e^{ih'_s \omega} \rho(\omega; h_t) d\omega, \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.87)$$

em que $\rho(\omega; h_t)$, $h_t \in \mathbb{R}^d$ é uma função contínua positiva definida para todo $\omega \in \mathbb{R}^d$.

Assim, considere a função de covariância estacionária dada por (2.37), também se assume que C_{st} é contínua, como sua função de distribuição espectral, logo possui densidade espectral expressa por

$$g(\omega; \tau) > 0. \quad (2.88)$$

Isto é, pelo Teorema 2,

$$C_{st}(h_s; h_t) = \int \int e^{ih'_s \omega + ih_t \tau} g(\omega; \tau) d\omega d\tau. \quad (2.89)$$

Além disso, se $C_{st}(\cdot; \cdot)$ é integrável, então

$$\begin{aligned} g(\omega; \tau) &= (2\pi)^{-d-1} \int \int e^{-ih'_s \omega - ih_t \tau} C_{st}(h_s; h_t) dh_s h_t \\ &= (2\pi)^{-1} \int e^{-ih_t \tau} H(\omega; u) du, \end{aligned} \quad (2.90)$$

em que

$$\begin{aligned} H(\omega; \tau) &\equiv (2\pi)^{-d} \int e^{-ih'_s \omega} C_{st}(h_s; h_t) dh_s \\ &= \int e^{ih_t \tau} g(\omega; \tau) d\tau. \end{aligned} \quad (2.91)$$

Para construção de $C(\cdot)$ e $g(\cdot)$ especifica-se modelos apropriados para $H(\omega; h_t)$. Assim, assume-se que

$$H(\omega; h_t) = \rho(\omega; h_t) K(\omega), \quad (2.92)$$

logo, as seguintes condições são satisfeitas:

- (C1) para cada $\omega \in \mathbb{R}^d$, $\rho(\omega, \cdot)$ é uma função de autocorrelação contínua, tal que $\int \rho(\omega; h_t) dh_t < \infty$, e $K(\omega) > 0$;

(C2) a função $K(\omega)$ satisfaz:

$$\int K(\omega)d\omega < \infty. \quad (2.93)$$

De tal forma, a expressão (2.90) torna-se

$$g(\omega; \tau) \equiv (2\pi)^{(-1)} K(\omega) \int e^{ih_t\tau} \rho(\omega; h_t) dh_t > 0, \quad (2.94)$$

por (C1). E não só isso, como

$$\int \int g(\omega; \tau) d\tau d\omega = \int K(\omega) d\omega < \infty, \quad (2.95)$$

por (C2). Assim, assumindo que $H(\omega; h_t)$ é dado por (2.92) de tal forma que as condições (C1) e (C2) sejam satisfeitas, tem-se que

$$C_{st}(h_s; h_t) = \int e^{ih'_s\omega} \rho(\omega; h_t) K(\omega) d\omega, \quad (2.96)$$

é uma função de covariância espaço-temporal contínua estacionária positiva definida em \mathbb{R}^{d+1} .

Pode-se então reescrever qualquer função de covariância espaço-temporal contínua estacionária integrável, que obedeças as condições (C1) e (C2), definindo as expressões

$$\rho(\omega; h_t) \equiv \frac{H(\omega; h_t)}{\int g(\omega; \tau) d\tau}, \quad (2.97)$$

e

$$K(\omega) \equiv \int g(\omega; \tau) d\tau. \quad (2.98)$$

em (2.92).

Para enfim construir as famílias de covariância espaço-temporais estacionárias não-separáveis [Cressie e Huang \(1999\)](#) utilizam das funções de densidade espectrais dadas por [Matérn \(1960\)](#) para exemplificar situações de como algumas formas fechadas da transformação de *Fourier* podem ser utilizada para este fim.

2.2.2.2 Modelo de Gneiting

[Gneiting \(2002\)](#) propôs um critério baseado na construção dada por [Cressie e Huang \(1999\)](#), porém este não depende da forma fechada da inversão de *Fourier* e também não requer a integrabilidade.

Para tal, lembra-se que uma função contínua $\varphi(t)$ definida para $t > 0$ ou $t \geq 0$ é completamente monótona se possui derivadas $\varphi(t)^{(n)}$ para todas ordens e $(-1)^n \varphi(t)^{(n)} \geq 0$, para $t > 0$ e $n = 0, 1, 2, \dots$

Do teorema de [Bernstein \(1929\)](#), uma função completamente monótona, $\varphi(t)$ definida para $t > 0$, é

$$\varphi(t) = \int_0^\infty \exp(-tv) dF(v), \quad t > 0, \quad (2.99)$$

em que $F(\cdot)$ é uma função não-decrescente.

Como, segundo ([GNEITING, 2002](#)) funções completamente monótonas e funções de covariância isotrópicas são relacionadas, tem-se que

Teorema 7 *Seja $\varphi(t), t \geq 0$, uma função completamente monótona, e seja, $\Phi(t), t \geq 0$, uma função positiva com derivada completamente monótona. Então,*

$$C_{st}(h_s; h_t) = \frac{\sigma^2}{\Phi(|h_t|^2)^{d/2}} \varphi\left(\frac{|h_s|^2}{\Phi(|h_t|^2)}\right), \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.100)$$

é uma função de covariância espaço-temporal.

Teorema 8 *Suponha que $\varphi(r), r > 0$, é uma função completamente monótona, e que $\psi(r), r \geq 0$, é uma função positiva com derivada completamente monótona. Então*

$$C_{st}(h_s; h_t) = \frac{1}{\psi(h_t^2)^{d/2}} \varphi\left(\frac{\|h_s\|^2}{\psi(h_t^2)}\right), \quad (h_s; h_t) \in \mathbb{R}^{d+1}, \quad (2.101)$$

é uma função de covariância espaço-temporal estacionária em \mathbb{R}^{d+1} .

Então, uma covariância espaço-temporal na forma (2.101) é além de estacionária, totalmente simétrica.

2.3 Krigagem espaço-temporal

Após a definição das funções de covariância espaço-temporais, pode-se utilizá-las para predição de um campo aleatório espaço-temporal. Para este fim, pode utilizar da krigagem, técnica que não requer a suposição de estacionariedade ([CRESSIE; WIKLE, 2015](#)).

Sejam os dados expressos por

$$Z(s_i; t_j) = Y(s_i; t_{ij}) + \varepsilon(s_i; t_{ij}), \quad j = 1, \dots, T_i, i = 1, \dots, m, \quad (2.102)$$

em que $\varepsilon(s_i; t_{ij})$ é o termo de erro, assumido como independente e identicamente distribuído (iid), com média 0 e variância σ_ε^2 .

2.3.1 Krigagem simples

O preditor de krigagem simples, $Y^*(s_0; t_0)$, nada mais é que uma combinação linear dos dados na forma

$$Y^*(s_0; t_0) \equiv \sum_{i=1}^m \sum_{j=1}^{T_i} l_{ij} Z(s_i; t_{ij}) + c, \quad (2.103)$$

em que l_{ij} e c são valores a serem obtidos, de forma que minimizem o erro quadrado médio da predição. Ainda, pode-se escrever (2.103) na forma vetorial

$$Y^*(s_0; t_0) \equiv \mathbf{I}'\mathbf{Z} + c, \quad (2.104)$$

em que $\mathbf{Z} \equiv (Z'_1, \dots, Z'_n)'$ e \mathbf{I}' é o vetor a ser obtido após otimização.

Defini-se então uma linguagem em termos matriciais (**escritos em negrito**), a fim de facilitar a notação das equações a seguir, sendo que

- $\mathbf{1}$ é um vetor positivo T -dimensional de valores 1;
- $\mathbf{Z} \equiv (Z'_1, \dots, Z'_n)'$;
- $\mathbf{C}_Z \equiv \text{var}(\mathbf{Z})$;
- $\mathbf{c}_0 \equiv \text{cov}(Y(s_0; t_0); \mathbf{Z})$.

Seguindo, se o processo $Y(\cdot; \cdot)$ e o termo de erro ($\varepsilon(\cdot; \cdot)$) são assumidos como processos gaussianos, pode-se escrever (2.102) na seguinte forma

$$\mathbf{Z} = \mathbf{Y} + \varepsilon, \quad (2.105)$$

e dado a suposição de normalidade, tem-se que

$$\begin{bmatrix} Y(s_0; t_0) \\ \mathbf{Z} \end{bmatrix} \sim \text{Gau} \left(\begin{bmatrix} \mu_{st}(s_0; t_0) \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} C_{st}(0, 0) & \mathbf{c}'_0 \\ \mathbf{c}_0 & \mathbf{C}_Z \end{bmatrix} \right), \quad (2.106)$$

em que

- $\mu_{st}(s_0; t_0) \equiv E[(Y(s_0; t_0))]$;
- $\boldsymbol{\mu} \equiv (\mu_{st}(s_i; t_j) : j = 1, \dots, T_i, i = 1, \dots, n)'$;

- $\mathbf{C}_Z = \Sigma_Y + \Sigma_\varepsilon$, em que Σ_Y e Σ_ε são as respectivas matrizes de covariância de \mathbf{Y} e ε .

Além disso, a distribuição posteriori é dada por (e.g., (RENCHEER, 2003 apud CRESSIE; WIKLE, 2015)),

$$Y(s_0; t_0) | \mathbf{Z} \sim \text{Gau} \left[\mu_{st}(s_0; t_0) + \mathbf{c}'_0 \mathbf{C}_Z^{-1} (\mathbf{Z} - \mu), C_{st}(0, 0) - \mathbf{c}'_0 \mathbf{C}_Z^{-1} \mathbf{c}_0 \right]. \quad (2.107)$$

Sob a suposição conjunta de normalidade (2.106), tem-se que o preditor de krigagem simples é a média posteriori de (2.104) (CRESSIE; WIKLE, 2015), isto é,

$$Y^*(s_0; t_0) = E[Y(s_0; t_0) | \mathbf{Z}] = \mu_{st}(s_0; t_0) + \mathbf{c}'_0 \mathbf{C}_Z^{-1} (\mathbf{Z} - \mu). \quad (2.108)$$

Assim, a variância da krigagem simples (ks) pode ser vista como a minimização do erro quadrado médio da predição, $E[Y(s_0; t_0) - \mathbf{l}'\mathbf{Z} - c]^2$, ou ainda sob a suposição de normalidade conjunta (2.106), a variância posteriori, ou seja,

$$\sigma_{ks}^2(s_0; t_0) \equiv E[Y(s_0; t_0) - Y^*(s_0; t_0)]^2 = C_{st}(0, 0) - \mathbf{c}'_0 \mathbf{C}_Z^{-1} \mathbf{c}_0. \quad (2.109)$$

Por fim, segundo Cressie e Wikle (2015) mesmo sem a normalidade, tanto (2.108) e (2.109) são apropriadas.

2.3.2 Krigagem ordinária

O preditor de krigagem ordinária é interessante no caso em que $Y(\cdot; \cdot)$ possui uma média constante desconhecida. Assim, utilizando-se do estimador de mínimos quadrados generalizados,

$$\hat{\mu}_{st_{gls}} \equiv (\mathbf{1}' \mathbf{C}_Z^{-1} \mathbf{1})^{-1} \mathbf{1}' \mathbf{C}_Z^{-1} \mathbf{Z}, \quad (2.110)$$

em que

- $\mathbf{1}$ é um vetor positivo T -dimensional de valores 1;
- $\mathbf{Z} \equiv (Z'_1, \dots, Z'_m)'$;
- $\mathbf{C}_Z \equiv \text{var}(Z)$.

De tal forma, o preditor de krigagem ordinária é dado por

$$\hat{Y}(s_0; t_0) \equiv \hat{\mu}_{st_{gls}} + \mathbf{c}'_0 \mathbf{C}_Z^{-1} (\mathbf{Z} - \hat{\mu}_{st_{gls}} \mathbf{1}), \quad (2.111)$$

ou ainda na forma matricial,

$$\hat{Y}(s_0; t_0) \equiv \lambda' \mathbf{Z}, \quad (2.112)$$

em que $\lambda' \equiv \{\mathbf{c}_0 + \mathbf{1}(1 - \mathbf{1}'\mathbf{C}_Z^{-1}\mathbf{c}_0)/(\mathbf{1}'\mathbf{C}_Z^{-1}\mathbf{1})\}$.

Portanto, a variância da krigagem ordinária (ko) é expressa como,

$$\sigma_{ko}^2 \equiv C_{st}(0, 0) - \mathbf{c}_0' \mathbf{C}_Z^{-1} \mathbf{c}_0 + (1 - \mathbf{1}'\mathbf{C}_Z^{-1}\mathbf{c}_0)^2 / (\mathbf{1}'\mathbf{C}_Z^{-1}\mathbf{1}). \quad (2.113)$$

Parte II

Metodologia

CAPÍTULO 3

DESENVOLVIMENTO COMPUTACIONAL

Posterior a definição teórica das técnicas possíveis, estabelece-se quais os ferramentais computacionais serão utilizados para aplicação em dados.

[Pebesma e Wesseling \(1998\)](#) mostram o *software gstat*, escrito em C, o programa de 1993 possuía funções para realização de operações de modelagem e krigagem.

Outro marco foi o desenvolvimentos das bibliotecas *geoS* ([RIBEIRO; DIGGLE, 1999](#)) e *geoR* ([JR; DIGGLE et al., 2001](#)) que expandiram os métodos existentes em suas respectivas ferramentas para modelagem de dados espaciais.

Em [Cesare, Myers e Posa \(2002\)](#) trouxeram algoritmos para modelagem espaço-temporal com as classes de modelos *product* e *product-sum* na linguagem *Fortran*, considerada por muitos como obsoleta.

Posteriormente, em 2003 um grupo de pesquisadores iniciaram o desenvolvimento de métodos para análise espacial em R, assim [Pebesma \(2004\)](#) cria um pacote na linguagem R, visando uma interface de suas funções do *software gstat*. Porém tamanho o sucesso que o projeto original foi descontinuado.

Tomando este indício, viu-se que o R além de gratuito e *open source*, ou seja, permite a colaboração entre pesquisadores, visto que os códigos são abertos. Também, é muito flexível, e poderoso, fazendo integração com praticamente qualquer base de dados e até outras linguagens (C, Java, Python, etc.). Além disso, possui uma infinidade de pacotes desenvolvidos por colaboradores do mundo inteiro, tanto que seu maior repositório (CRAN) possui atualmente mais de 12.000 pacotes.

E não é só na Estatística como um todo, mas justamente no contexto espaço-temporal o R cresce muito também. Em sua apresentação no *Brazilian Symposium on Geoinformatics (GEOINFO) 2010*, o pesquisador e autor *Edzer Pebesma* apresentou a evolução da linguagem em relação ao âmbito espacial e espaço-temporal (Figura 6).

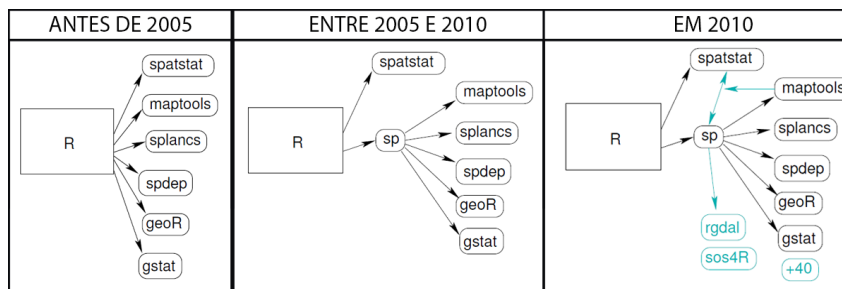


Figura 6 – Ilustração traduzida e adaptada da evolução da linguagem R no contexto espacial e espaço-temporal até 2010. Fonte: Apresentação “Modelling spatio-temporal data with R”, de Edzer Pebesma no evento GEOINFO 2010.

Observa-se na Figura 6 que até 2005 haviam alguns pacotes, posteriormente o pacote *sp* (PEBESMA et al., 2012) surge, de forma a otimizar o trabalho com quase todos os pacotes existentes, com exceção ao *spatstat*. Já em 2010, nota-se que há uma comunicação entre todos os pacotes, e não só isso como mais de 40 pacotes novos. E essa evolução continuou, em 2017 há centenas de bibliotecas tanto para tempo, espaço ou espaço-tempo.

Assim, escolhe-se o R como ferramenta, todavia devido a quantidade de bibliotecas é importante a determinação de quais pacotes serão utilizados, pois por mais que seja interessante ter tantas escolhas, a variedade de autores também implica em inconsistências quanto a sintaxe e integração dos diferentes códigos, visto que ainda há uma fraca compatibilidade intra pacotes. Assim, toma-se como referência uma representação gráfica do processo de uma análise de dados, para elencar os passos no contexto espaço-temporal no *software* R.

Fluxograma de análise de dados (Hadley Wickham)

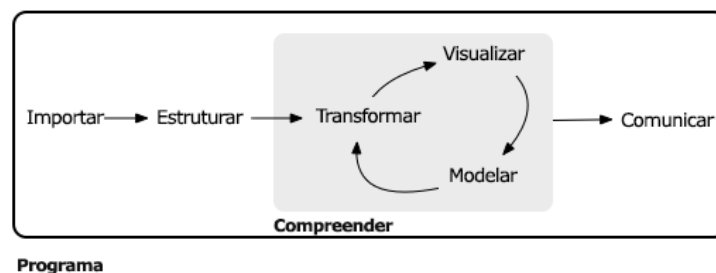


Figura 7 – Tradução do fluxograma de análise de Hadley Wickham.

Após visualização da Figura 7 foram traçadas as etapas de análise, bem como os pacotes que melhor se adequam as respectivas etapas no contexto geoestatístico espaço-temporal.

CAPÍTULO 4

IMPORTAÇÃO E ESTRUTURAÇÃO

Para importação dos dados são utilizadas ferramentas nativas do *R*, com exceção aos dados das coordenadas GNSS, pois como as séries estão disponíveis na *Internet*, e são extensas e variadas utilizar-se-á de ferramentas de *web scraping* (MUNZERT et al., 2014), de forma que o processo seja automatizado.

Os pacotes utilizados no *R* necessitam dos dados em um formato bem particular, assim será necessário tanto uma estruturação quanto manipulação dos dados (WICKHAM, 2016a; WICKHAM, 2016c). Isto se deve pois há diversos formatos dos dados, no contexto espaço-temporal pode-se destacar 3 ((NETWORK et al., 2016)):

- *time-wide*: diferentes colunas refletem em diferentes momentos no tempo;
- *space-wide*: diferentes colunas refletem em diferentes momentos no espaço;
- *long*: cada linha reflete num observação do espaço-tempo.

Uma vez conhecendo as possibilidades, utiliza-se a estrutura *tidy* (WICKHAM et al., 2014) (Figura 8), visto que está é base para pacotes do conjunto *tidyverse*, sendo então utilizados nas etapas de análise descritiva.

Sendo

- x : coordenada x ;
- y : coordenada y ;
- n : número de observações no espaço;
- m : número de observações no tempo;

- v : número de covariáveis.
- p_i : covariável i , $i = 1 \dots v$;
- t_i : tempo i , $i = 1 \dots m$;

Banco tidy (n.m) x (3.v)

	x	y	t	p_1	...	p_v
(n.m)						

(3.v)

Figura 8 – Ilustração da estrutura de dados *tidy*.

Porém, a estrutura da Figura 8 não é utilizada pelo pacote *gstat*, tal que a mesma requer um formato em que cada tempo é uma coluna (*time-wide*, ou seja,

Banco gstat (n) x (2.v.m)

	x	y	p_1	...	p_v	t_1	...	t_m
(n)								

(2.v.m)

Figura 9 – Ilustração da estrutura de dados *gstat*.

Portanto, para fazer a transição entre os formatos utiliza-se do pacote *tidy* e do pacote *gstat* para transformação das variáveis para seus respectivos tipos, e.g., coordenada, tempo, etc.

CAPÍTULO 5

ANÁLISE EXPLORATÓRIA

Como os modelos geoestatísticos são bem particulares aos fenômenos dos dados, é imprescindível conhecê-los bem, de forma que se possa determinar as funções adequadas para seu ajuste, assim algumas ferramentas de visualização e sumarização são essenciais (PEBESMA; BIVAND, 2005; WICKHAM, 2016b).

Para realização da exploração dos dados pode-se seguir alguns estágios. Dentro de cada estágio, pode-se utilizar inúmeras representações a fim de partir para análises no espaço-tempo. Portanto, a visualização apesar de ser considerada trivial em muitas outras análises acaba se tornando um ponto de preocupação e destaque no contexto espaço-temporal, sendo que Cressie e Wikle (2015) destaca um capítulo inteiro para discussão deste tópico.

Apesar de auxiliarem, estas ferramentas também apresentam vantagens e desvantagens dependendo do contexto dos dados analisados, como se verá nas aplicações (Parte III). Logo, apresenta-se algumas alternativas visuais para exploração dos dados.

5.1 Análise adimensional:

Ao desconsiderar as dependências, busca-se identificar valores perdidos, variabilidade e valores aberrantes.

Na análise estatística exploratória tradicional dois dos gráficos mais utilizados para avaliação da dispersão são o gráfico de pontos e o *boxplot* (WICKHAM; STRYJEWSKI, 2011), porém ambos podem ser bons ou não como identificadores de *outliers*, dependendo da quantidade de dados e variabilidade. Assim, avalia-se algumas alternativas para identificação da distribuição dos dados.

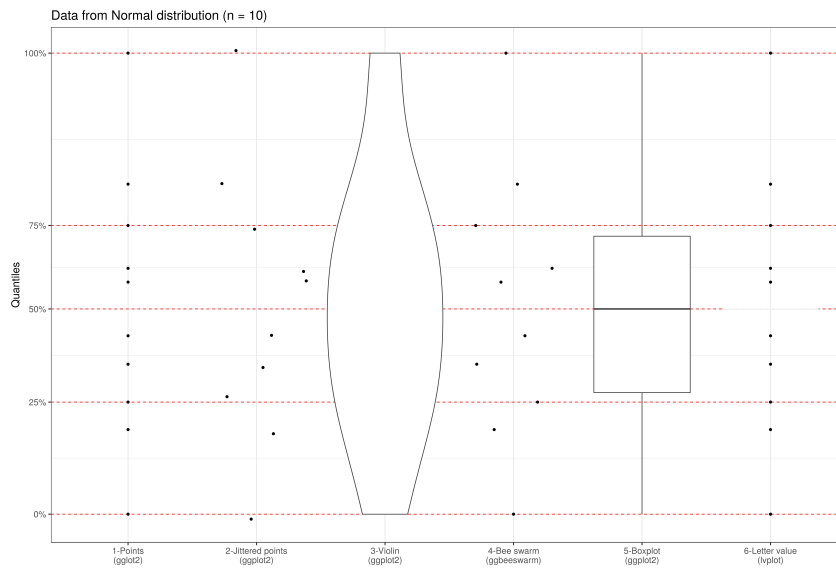


Figura 10 – Geometrias distintas para dados simulados de uma distribuição normal padrão ($n = 10$).

Vê-se na Figura 10 que no cenário de poucas observações a melhor forma visualização é o gráfico de pontos, pois neste caso não há sobreposição, enquanto as demais podem levar a suposições errôneas.

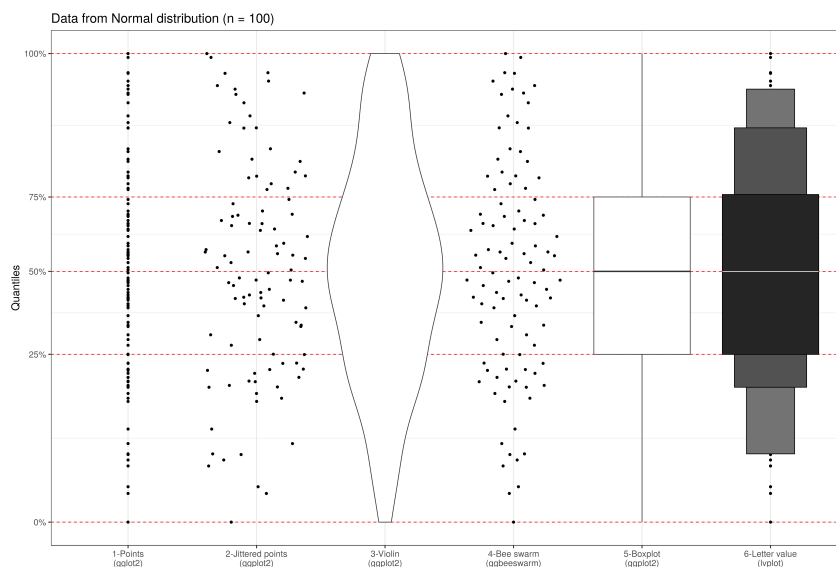


Figura 11 – Geometrias distintas para dados simulados de uma distribuição normal padrão ($n = 100$).

Na Figura 11 nota-se que o gráfico de pontos deixa de ser uma boa escolha, vide a superposição dos pontos tal que os pontos dispersos podem-se tornar uma alternativa viável. Além disso, o gráfico *letter value* (HOFMANN; WICKHAM; KAFADAR, 2017) identifica *outliers*, já que possui um critério robusto para muitos dados, mas sensível para poucas observações.

Este gráfico parte de um critério que a maior caixa contém 50% dos dados, a segunda maior 25% e assim por diante.

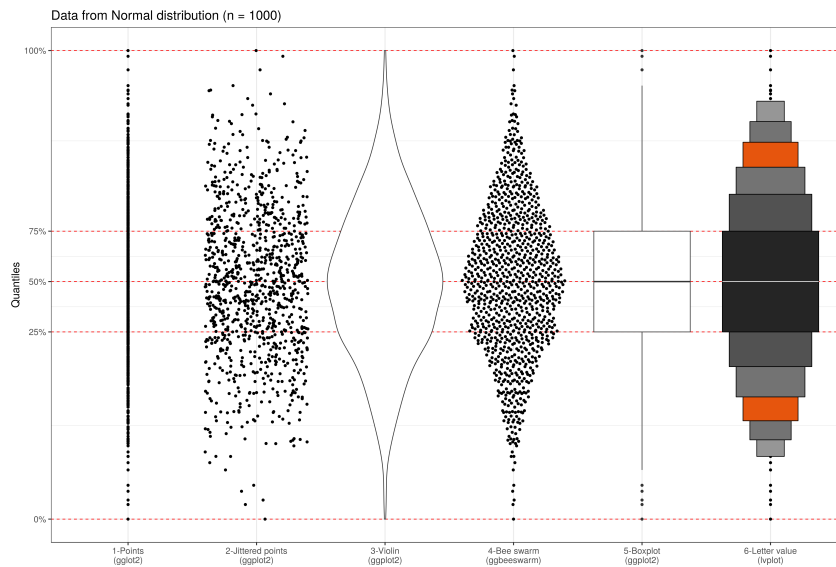


Figura 12 – Geometrias distintas para dados simulados de uma distribuição normal padrão ($n = 1000$).

Já na Figura 12 observa-se o cenário com mais dados em que os pontos dispersos também ficam sobrepostos, sendo que as demais alternativas tornam-se de fácil visualização, entretanto o *letter value* ainda apresenta mais *outliers* que o *boxplot*, mesmo com mais dados. Isto se deve, pois os dados simulados são normais, assim mostra-se um cenário com outra distribuição.

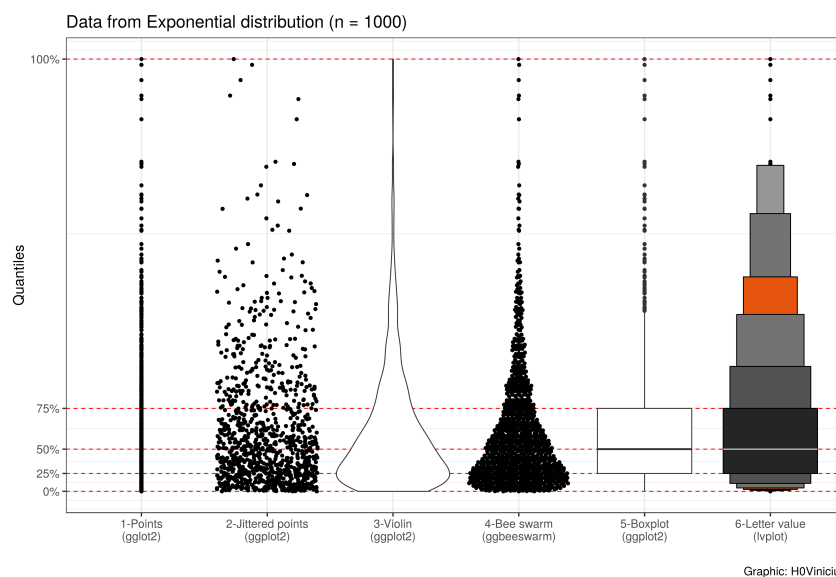


Figura 13 – Geometrias distintas para dados simulados de uma distribuição exponencial ($n = 1000$).

No caso com a distribuição exponencial vê-se que o boxplot apresenta inúmeros *outliers*, diferente do *letter value*. Mostrando-se uma forma visual interessante para casos em que o número de observações é alto e os dados são não normalmente distribuídos.

5.2 Análise temporal:

Para análise temporal, fixa-se o espaço, tal que se pode avaliar dois tipos de efeitos:

- **intra-séries:** sazonalidade, memória longa, tendência, autocorrelação, entre outros;
- **inter-séries:** correlação, covariância, entre outros.

5.2.1 Autocorrelação

Para tal inicia-se com a função de autocorrelação (MORETTIN; TOLOI, 2006a), em que nesta se faz a correlação da série com a própria, porém dado uma determinada defasagem (*lag*). Logo, aplica-se a correlação linear de *Pearson* entre os diferentes tempos, seja X_t um processo aleatório temporal estacionário e t um índice temporal, tem-se que a função de autocorrelação (ACF, do inglês *autocorrelation function*) é dada por:

$$\begin{aligned} A\rho_t(X_t) &= \rho_t[X, \text{lag}(X, t)] \\ &= \frac{C_t[X, \text{lag}(X, t)]}{\text{sd}(X)\text{sd}[\text{lag}(X, t)]} \\ &= \frac{C_t[X, \text{lag}(X, t)]}{\text{sd}(X)\text{sd}[\text{lag}(X, t)]} \\ &= \frac{C_t[X, \text{lag}(X, t)]}{\text{var}[X, \text{lag}(X, t)]}, \end{aligned}$$

em que o operador de retardo é expresso por

$$\text{lag}(X_i, t) = X_{i-t}. \quad (5.1)$$

5.2.2 Correlação cruzada

Enquanto a autocorrelação permite a análise de um efeito intra-série, esta torna-se ineficaz no caso multivariado. Desta forma, uma alternativa está no uso da correlação cruzada (CCF, do inglês *cross-correlation function*) que permite avaliar a correlação entre duas séries temporais distintas. Porém, a CCF parte de um *lag* fixo, de forma que se pode utilizar a

correlação cruzada *wavelet* (WCC, do inglês *wavelet cross-correlation*) (SOUZA; FÉLIX, 2018) para avaliação de duas séries em escalas variadas, sendo WCC dada por

$$C\rho_t(\lambda_j) = \frac{C_t^w(X_t^1, X_t^2, \lambda_j)}{\text{var}(X_t^1, \lambda_j)\text{var}(X_t^2, \lambda_j)}, \quad (5.2)$$

em que

- $\text{var}(X_t^1, \lambda_j)$ é a variância de X_t^1 associada com a escala λ_j ;
- $\text{var}(X_t^2, \lambda_j)$ é a variância de X_t^2 associada com a escala λ_j ;
- $C_t^w(X_t^1, X_t^2, \lambda_j)$ é a covariância *wavelet* entre X_t^1 e X_t^2 associada com a escala λ_j .

Porém, esta métrica é deficiente no sentido que ela é calculada de forma pareada, de forma que para muitas séries há um problema na visualização, por exemplo, se tivéssemos n séries para avaliação de todos os pares possíveis sem necessidade de ordenação teria-se de usar a fórmula da combinação, dada por

$$C_s^m = \binom{n}{s} = \frac{n!}{s!(n-s)!}, \quad (5.3)$$

logo, aplicando a equação (5.3) para os pares de n séries, tem-se

$$\begin{aligned} C_s^m &= \binom{n}{2} \\ &= \frac{n!}{2!(n-2)!} \\ &= \frac{n(n-1)(\cancel{n-2}!)!}{2(\cancel{n-2}!)!} \\ &= \frac{n(n-1)}{2}, \end{aligned}$$

geraria-se então $n(n-1)/2$ gráficos de correlação *wavelet* e $jn(n-1)/2$ vezes gráficos de correlação cruzada, em que j é ordem da transformada *wavelet*.

Assim, pode-se calcular a correlação cruzada *wavelet* múltipla, do inglês *wavelet multiples cross-correlation* (WMCC), considerando $\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^m)$ um processo estocástico multivariado e $\mathbf{W}_{jt} = (W_{1jt}, \dots, W_{mjt})$ os respectivos coeficientes *wavelet* numa escala λ_j obtidos através da aplicação da MODWT para cada processo X_t^i .

Logo, pensando inicialmente, no caso em que W_1 é uma variável dependente, obtém-se o seguinte modelo

$$W_1 = \beta_{12}W_2 + \beta_{13}W_3 + \dots + \beta_{1n}W_n + \varepsilon_1, \quad (5.4)$$

em que

- $\beta_{12}, \dots, \beta_{1n}$ são parâmetros desconhecidos;
- ε_1 é um ruído aleatório.

Portanto, a WMCC pode ser escrita como correlação entre os coeficientes observados e ajustados, ou seja,

$$\rho(W_{ijt}, \hat{W}_{ijt}) = \frac{Cov(W_{ijt}, \hat{W}_{ijt})}{\sqrt{\text{var}(W_{ijt})\text{var}(\hat{W}_{ijt})}}, \quad (5.5)$$

em que

- W_{ij} é escolhido para maximizar $\varphi_X(\lambda_j)$;
- \hat{W}_{ij} são os valores ajustados na regressão de W_{ijt} no resto dos coeficientes *wavelets* na escala λ_j .

Como visto em [Tiwari et al. \(2013\)](#), tomando dois processos estocásticos (X_{1t} e X_{2t}), sabe-se que para o desenvolvimento da correlação *wavelet* é necessária a construção das variâncias de $\{X_{1t}\}$ e $\{X_{2t}\}$ em diferentes escalas e da covariância $\{X_{1t}, X_{2t}\}$.

A variância *wavelet* então é estimada a partir da *maximal overlap discrete wavelet transform* (MODWT) e seus coeficientes para escala $\lambda_j = 2^{j-1}$, sendo expressa por

$$\hat{\sigma}_x^2 = \frac{1}{\hat{n}_j} \sum_{k=L_j-1}^{n-1} (\hat{W}_{j,k})^2, \quad (5.6)$$

em que

- $\hat{W}_{j,k}$ é o coeficiente *wavelet* da MODWT da variável X na escala λ_j ;
- $\hat{n}_j = L_j + 1$ é o número de coeficientes não afetados pela fronteira;
- $L_j = (2^j - 1)(L - 1) + 1$ é o tamanho do filtro *wavelet* na escala λ_j .

Ressalta-se que o uso da WMCC é de difícil de interpretação e recomendado para um número baixo ou médio de séries temporais, que não são os casos das aplicações deste trabalho, mas para exemplificação desta métrica tão distinta, faz-se uma aplicação para exemplo em 5 estações com coordenadas GNSS, as mesmas encontram-se no Apêndice A.

5.2.3 Entropia

A entropia de *Shannon* (JOST, 2006) é uma métrica que busca avaliar a “previsibilidade” de um sinal, tal que valores baixos indicam uma alta razão sinal/ruído, consequentemente altos valores uma baixa razão, isto é, maior dificuldade, visto que há mais ruído. Assim, a entropia é expressa por:

$$-\int_{-\pi}^{\pi} \hat{f}(X_t) \log \hat{f}(X_t) dX_t, \quad (5.7)$$

em que $\hat{f}(X_t)$ é a estimativa da densidade espectral (MORETTIN; TOLOI, 2006b) da série temporal X_t , sendo que está pode não ser estimável para algumas séries.

5.2.4 Granulosidade e Estabilidade

Utilizando-se do conceito do mesmo princípio de médias móveis, calcula-se a média da série temporal em intervalos distintos (d), assim a granulosidade (HYNDMAN, 2019) se dá pela variância destas médias:

$$\text{var}(\mu_d), \quad (5.8)$$

em que μ_d é o vetor de médias em intervalos distintos da série temporal X_t .

Similarmente, a estabilidade (HYNDMAN, 2019) se dá de forma parecida, em que ao invés da médias, calcula-se a variância das variâncias em intervalos distintos, isto é,

$$\text{var}(\sigma_d^2), \quad (5.9)$$

em que σ_d^2 é o vetor de variâncias em intervalos distintos da série temporal X_t .

5.2.5 Outras métricas

Com base no pacote *tsfeatures* (HYNDMAN, 2019) utiliza-se outras métricas a fim de identificar efeitos que possam auxiliar na modelagem.

- Coeficiente de *Hurst*: esta métrica é utilizada primariamente como uma medida de efeito de memória longa, ou seja, um efeito da própria que implica em uma dependência de um período maior de tempo (HASLETT; RAFTERY, 1989);
- Tendência: métrica derivada de uma extensão da decomposição STL, do inglês *seasonal trend decomposition* (CLEVELAND et al., 1990), de uma série, tal que valores maiores implica em uma maior indicativo da presença de tendência (HYNDMAN; KOEHLER, 2006);

- Linearidade: medida de linearidade baseada num ajuste polinomial quadrático;
- Curvatura: medida de curvatura baseada num ajuste polinomial quadrático.

5.3 Análise espacial:

Fixa-se o tempo e pode-se avaliar como se dão os efeitos no espaço, e.g, tendência, uma prática está no uso da geoestatística espacial. Para avaliação do espaço faz-se o cálculo do semivariograma espacial empírico em diferentes tempos, mas acontece o caso das diferenças entre as magnitudes da semivariância ser tão alta que a análise fica inviável. Assim, sugere-se o uso da aplicação da escala logarítmica.

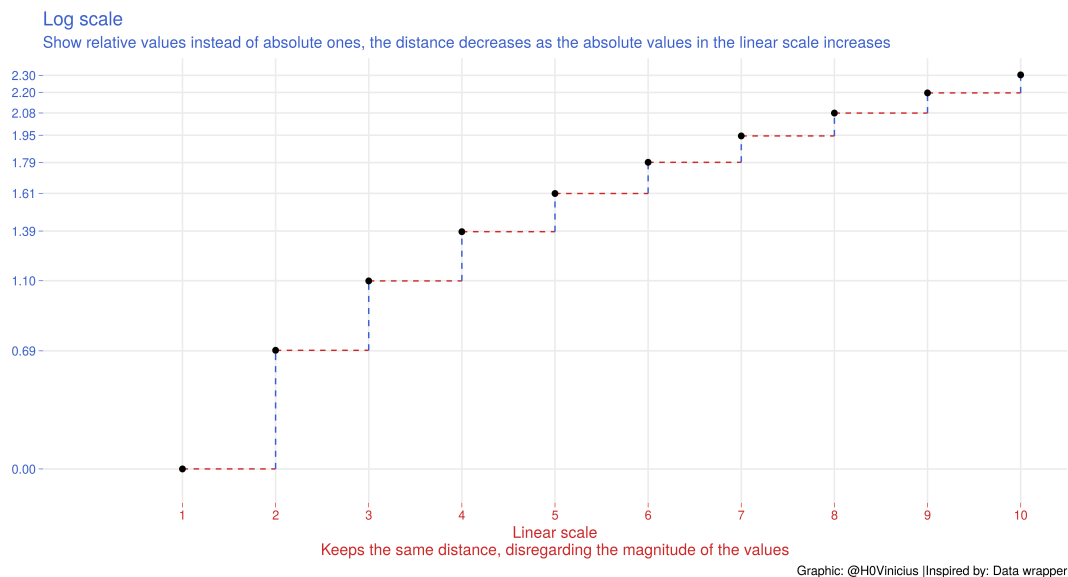


Figura 14 – Explicação da escala logarítmica em relação à escala linear.

Vê-se que aplicação da função \log faz-se com que a diferença entre valores menores torne-se mais evidente, enquanto a escala linear mantém as distâncias invariante a magnitude.

5.4 Análise espaço-temporal:

Avaliações visuais descritivas de como os dados se comportam no espaço-tempo, algumas possibilidades são:

- **Animação:** dado o contexto, a visualização em ambas as dimensões pode ser complicada, e a animação nada mais é que imagens (espaço) em movimento (tempo), sendo a

forma mais natural de representação, porém enfrenta a barreira de ser inviável em meios de comunicação estáticos como um artigo ou como esta própria dissertação, sendo assim não utilizada aqui;

- **Gráficos *Space (1-D)/time***: este gráfico consiste numa representação bidimensional, em que a variável de interesse representa-se por cor, tal qual um *heat map* e as dimensões ficam nos eixos;
- **Gráficos em multipaneis**: consiste em representações espaciais que ficam em diferentes painéis, em que cada um representa um tempo.

Ainda, uma análise essencial é a da semivariância espaço-temporal, tal que 3 possibilidades de visualização são possíveis:

- **Lag**: mesma representação do caso puramente espacial, porém cada tempo implica em uma cor, ou seja, como se fossem múltiplos semivariogramas espaciais;
- **Map**: solução semelhante ao *Space (1-D)/time*, em que a cor agora representa a semivariância, enquanto os eixos x e y implicam no tempo e distância;
- **Wireframe**: representação tri-dimensional, em que tempo, distância e semivariância são plotados em um cubo.

CAPÍTULO 6

MODELAGEM

Após avaliar os dados de forma descritiva, de tal forma que se possa entender seu comportamento, escolher-se-á um modelo espaço-temporal que satisfaça suas características.

Para modelagem é possível o uso de 3 principais pacotes: *gstat*, *CompRandFld* e *RandomFields* (NETWORK et al., 2016).

Utiliza-se como principal o pacote *gstat* (PEBESMA, 2004; GRÄLER; PEBESMA; HEUVELINK, 2016), que a princípio traz a classe de modelos separáveis, sendo eles: *product-sum*, *metric* e *sum-metric*.

6.1 Estimação

Escolhido um modelo, é necessário na etapa de estimação uma rotina de otimização, para resolução de problemas na forma numérica. Para o pacote *gstat*, a rotina *optim* é usada, do *R base*, em que possui os métodos:

- **BFGS** (do inglês, Broyden-Fletcher-Goldfarb-Shannon)(BROYDEN, 1970; FLETCHER, 1970; GOLDFARB, 1970; SHANNO, 1970);
- **CG** (do inglês, *conjugate gradients*) (FLETCHER; REEVES, 1964);
- **L-BFGS-B** (do inglês, *Limited-memory BFGS*)(BYRD et al., 1995);
- **SANN** (do inglês, *Simulated-annealing*)(BÉLISLE, 1992);
- **Brent** (BRENT, 1971).

Gräler, Pebesma e Heuvelink (2016) sugerem o uso de otimização que permita a imposição de limites e proveja limites sensíveis. Sem definição, o método *default* consiste no

L-BFGS-B, tal que o menor limite inferior e maior limite superior são suportados para o modelo.

Uma vez escolhido a classe, modelo e método de otimização, a estimação do semivariograma espaço-temporal depende do semivariograma amostral empírico, assim utiliza-se o método de estimação de mínimo quadrados ponderados entre a diferença do modelo com a superfície do semivariograma amostral, ou seja,

$$\Theta^* = \min_{\theta} \sum_i \sum_j \omega_{ij} |\gamma_{st\theta} - \hat{\gamma}_{ij}|^2, \quad (6.1)$$

em que ω_{ij} são os pesos, $\hat{\gamma}_{ij}$ é o variograma empírico e $\gamma_{st\theta}$ semivariograma que depende do vetor de parâmetros Θ , tal que se resolva este problema numericamente pelo método de otimização escolhido dentre as opções prévias.

Entretanto, resta a escolha dos pesos utilizados, por *default* pesos idênticos são atribuídos aos valores, mas o pacote conta com diversas opções (apresentadas na Tabela 1.

Método de ajuste	Pesos
0	Sem ajuste
1 e 3	$N(\cdot)$
2 e 4	$N(\cdot)/\gamma_{st}(\bar{h}_s; \bar{h}_t)^2$
5	exclusivo para REML
6	igual para todos
7	$N(\cdot)/(\bar{h}_s^2 + \alpha^2 \times \bar{h}_t^2)$
8	$N(\cdot)/\bar{h}_s^2$
9	$N(\cdot)/\bar{h}_t^2$
10	$1/\gamma_{st}(h_s; h_t)^2$
11	$1/(\bar{h}_s^2 + \alpha^2 \times \bar{h}_t^2)$
12	$1/\bar{h}_s^2$
13	$1/\bar{h}_t^2$

Tabela 1 – Método de ajuste e respectivos pesos do pacote *gstat*.

Em que

- $N(\cdot)$ é o número de pares;
- \bar{h}_s é a distância espacial média;
- \bar{h}_t é a distância temporal média;
- j é o número de intervalos;
- α é um escalar representando a anisotropia.

Tal que o uso de pesos idênticos implica no método de mínimos quadrados ordinários, enquanto uma opção como a número 2/4 implicam na correção da variância (CRESSIE, 1992), vide seu respectivo denominador.

Ainda, uma outra possibilidade seria o uso da verossimilhança, porém esta requer a inversão da matriz de covariância, e sendo o número de observações superior a 5.000 esta se torna inviável (NETWORK et al., 2016), sendo então preterida.

6.2 Krigagem

Após a estimação tem-se o processo de krigagem, em que se levanta a questão de tempo computacional. O pacote *gstat* possui dois métodos:

Método de krigagem	Comando
<i>Standard kriging</i>	krigeST
<i>Trans Gaussian kriging</i>	krigeSTTg

Tabela 2 – Métodos de krigagem e respectivos comandos do pacote *gstat*.

Dado a complexidade do cálculo, o uso da krigagem espaço-temporal pode ser computacionalmente inviável, de forma que se utiliza o artifício da krigagem local, de forma que é necessário determinar um critério de seleção dos vizinhos mais próximos no espaço-tempo. No pacote seleciona-se este critério através do algoritmo de falsos vizinhos mais próximos, com o pacote FNN (BEYGELZIMER et al., 2013).

Feita a modelagem, tem-se como descrito por Wickham (2007) 4 níveis de informação, são eles:

Nível 1 **Modelo:** consiste em estatísticas de ajuste.

Nível 2 **Coeficiente:** consiste nas estimativas de coeficientes, estatísticas de teste e valores p, respectivos as estimativas;

Nível 3 **Residual:** consiste nos resíduos e medidas de influência;

Nível 4 **Observacional:** consiste nos dados originais mais comportamento dos resíduos .

Portanto, após o ajuste de um modelo segue-se para seu diagnóstico, seleção, predição e conseqüente avaliação da mesma.

6.3 Diagnóstico e seleção

Por fim, para avaliação dos modelos ajustados, através de técnicas de diagnóstico e critérios de modelagem, para escolha daquele que melhor prediga os dados. Logo, algumas possíveis avaliações para seleção de modelos são possíveis.

- **Comparação visual:** avaliação visual entre as funções de covariância/semivariância empírica e ajustada, é essencial, porém é a mais subjetiva;
- **Métricas:** medidas resumo que sumarizam determinadas propriedades do modelo ajustado, objetivando a comparação entre modelos. Seja então:
 - n : o número de localidades no espaço;
 - m : o número de momentos no tempo;
 - $Z(s_i; t_j)$: os valores observados;
 - $\hat{Z}(s_i; t_j)$: os valores preditos.

As métricas mais usuais consistem em:

- r^2 : consiste na coeficiente de correlação entre os valores observados e preditos;
- erro médio quadrado, do inglês, *mean square error* (MSE):

$$\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m \left(Z(s_i; t_j) - \hat{Z}(s_i; t_j) \right)^2. \quad (6.2)$$

- erro médio quadrado normalizado, do inglês, *normalized mean square error* (NMSE):

$$\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{Z(s_i; t_j) - \hat{Z}(s_i; t_j)}{\sigma_{ij}} \right)^2. \quad (6.3)$$

- raiz do erro médio quadrado, do inglês, *root mean square error* (RMSE):

$$\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m \sqrt{\left(Z(s_i; t_j) - \hat{Z}(s_i; t_j) \right)^2}. \quad (6.4)$$

- erro médio absoluto, do inglês, *mean absolute error* (MAE):

$$\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m \left| Z(s_i; t_j) - \hat{Z}(s_i; t_j) \right|. \quad (6.5)$$

- erro médio absoluto percentual, do inglês, *mean absolute percentage error* (MAPE):

$$\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m \left| \frac{Z(s_i; t_j) - \hat{Z}(s_i; t_j)}{Z(s_i; t_j)} \right| \%. \quad (6.6)$$

– escore logarítmico:

$$\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m \left[\frac{1}{2} \log(2\pi\sigma_{ij}^2) + \frac{1}{2} \left(\frac{Z(s_i; t_j) - \hat{Z}(s_i; t_j)}{\sigma_{ij}} \right)^2 \right]. \quad (6.7)$$

Nota-se que as métricas utilizadas são versões dos casos puramente espacial, temporal ou mesmo adimensional.

- **Validação cruzada:** técnica que visa a quantificação da performance de um modelo preditivo. De forma simples implica na partição dos dados, tal que se modela um subconjunto e se valida na partição complementar.

Uma prática comum é o uso da validação cruzada, sendo que existem diversos métodos, por exemplo, [Bivand et al. \(2008\)](#) traz diversas possibilidades. Em que as mais comuns são:

- ***k-fold cross-validation***: dados são divididos em k partes, tal que cada parte é utilizada para validação;
- ***leave-one-out cross-validation***: cada valor da amostra é utilizado para validação.

A validação também se dá como nas métricas, de forma que técnicas comuns são utilizadas no caso espaço-temporal.

Parte III

Aplicações

Uma vez levantado as propriedades teóricas e recursos computacionais, modela-se dados reais a fim de mostrar as dificuldades e benefícios das técnicas apresentadas. Entretanto, faz-se um auxiliar que permita a identificação de propriedades para facilitar o processo de modelagem como um todo:

1. Dimensionamento: dimensão do tempo e espaço, uma vez que há um alto impacto na visualização e modelagem, e visto que dados com curto período de tempo não são nem indicados para este tipo de modelagem;
2. Regularidade espacial e temporal: a distribuição de dados não regular pode tornar alguns modelos não estimáveis, sendo um facilitador na modelagem;
3. *Missing data*: a falta de dados é outro problema, porém pode ser contornado por métodos de interpolação, porém vale ressaltar que dados e dependências métodos “simples” podem ser nocivos;
4. Frequência temporal: a frequência é interessante para avaliação de feitos cíclicos e sazonais, além de impactar na interpretação do resultado.
5. Efeitos da série: avalia-se aqui diversos efeitos a fim de buscar qualquer indicativo que auxilie na escolha da classe de modelos;
6. Efeitos entre as séries: este é um tipo de efeito de difícil identificação, sendo mais simples de identificar quando há um número baixo de séries temporais;
7. Verificação de separabilidade: um ponto crucial, visto que as classes de modelos do tipo separável aproveitam-se dos modelos espaciais, enquanto os não-separáveis necessitam de funções mais complexas, ressalta-se que além da verificação visual recente iniciativas estão sendo desenvolvidas com o uso teste de hipótese.

Assim, faz-se duas aplicações:

- **Poluição do ar**: nesta aplicação busca-se o uso das formas visuais exploratórias propostas, visto que este é um banco amplamente modelado em artigos e exemplos didáticos, e muitas vezes a exploração é ignorada;
- **Análise de coordenadas GNSS**: visa-se aqui um cenário com alta dimensionalidade e com o objetivo de oferecer um auxílio na modelagem através dos gráficos e métricas, e por fim a interpolação espacial.

CAPÍTULO 7

POLUIÇÃO DO AR

7.1 Definição do problema

Para demonstração das dificuldades de análise descritiva pré-processo de modelagem, faz-se uma aplicação adaptada dos dados de qualidade do ar, compilados por *Benedict Graeler* no ambiente *R*, e utilizados como exemplo no pacote *spacetime* (PEBESMA, 2012).

Originalmente, os dados foram obtidos da base *AirBase2*, plataforma pública sobre qualidade do ar da *European Economic Area* (EEA).

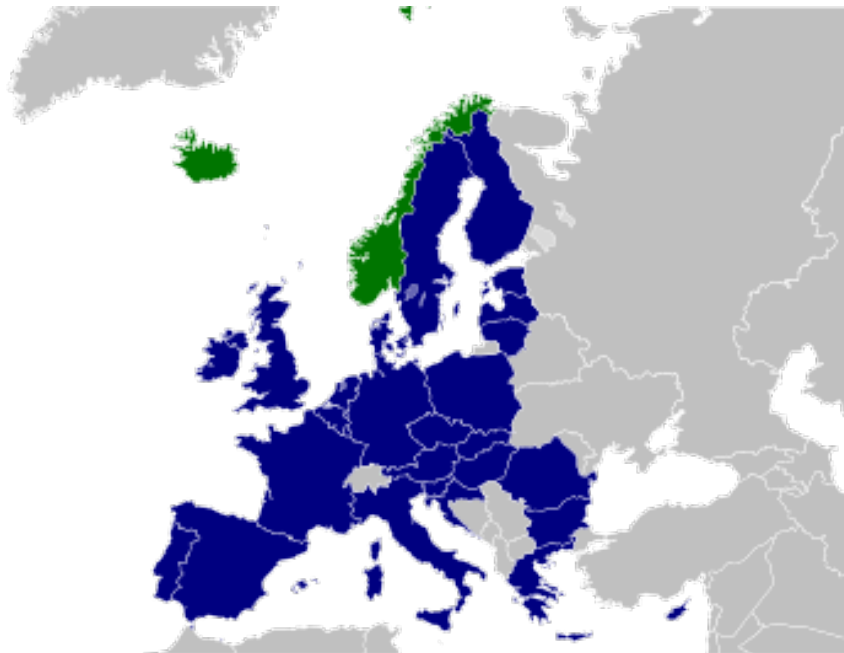


Figura 15 – Composição da EEA, destacada pelos países coloridos (verde ou azul). Fonte: Wikipedia.

Para o exemplo, utiliza-se de dados com médias diárias da concentração de PM10 (tipo

de partículas inaláveis, de diâmetro inferior a 10 micrômetros (μm) por metro cúbico, o qual é um indicador da qualidade do ar, sendo que a análise é feita com os dados coletados dos anos 2004 a 2009, em 65 estações rurais na Alemanha (Figura 16).

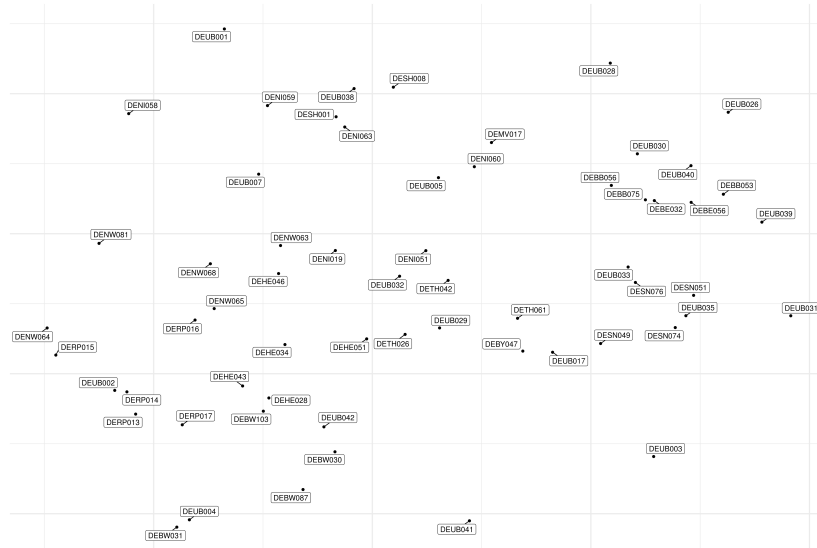


Figura 16 – Disposição das estações.

7.2 Descritiva

Inicialmente, para uma análise descritiva pode-se avaliar como se dá o comportamento dos dados fixando tanto o tempo quanto espaço. A priori, faz-se a avaliação fixando o espaço, ou seja, para cada estação uma descrição temporal.

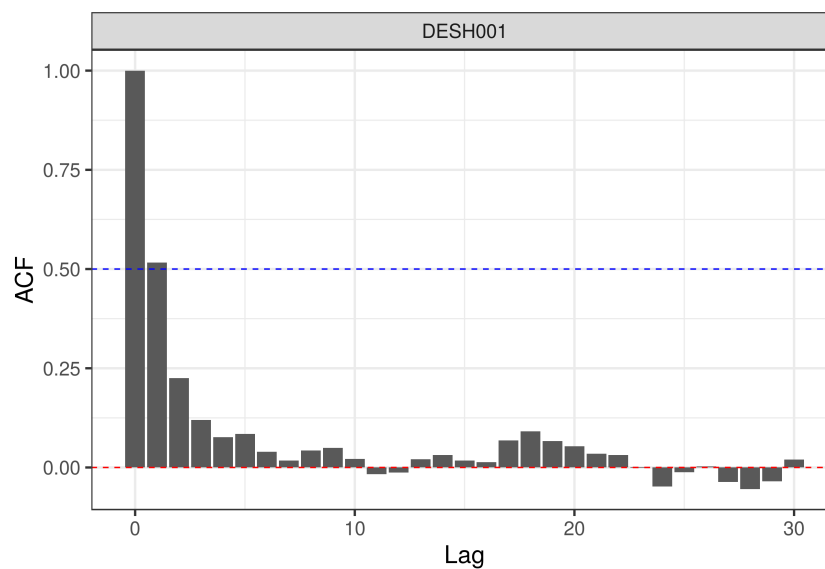


Figura 17 – Função de autocorrelação temporal para estação DESH001.

Como se vê na Figura 17, calcula-se a função de autocorrelação (ACF) (MORETTIN; TOLOI, 2006a). Observa-se que os valores são, em suma, positivos, sendo que o *lag* 1 possui a maior correlação (0,67), visto que a autocorrelação no *lag* nulo é sempre 1, já que se dá pela correlação da série com a própria sem defasagem alguma. Ainda, no início do gráfico da ACF vê-se um decrescimento e mesmo que a partir do 16º *lag* tenha correlações um pouco maiores, nenhuma outra correlação expressiva se destaca. Analisando todas as 65 estações nota-se que, em geral, os comportamentos são próximos, com exceção de algumas estações que possuem mais correlações negativas em determinados *lags*.

Porém é necessário visualizar as autocorrelações de todas as estações, no caso há 61 estações que poderiam ser visualizadas como na forma individual.

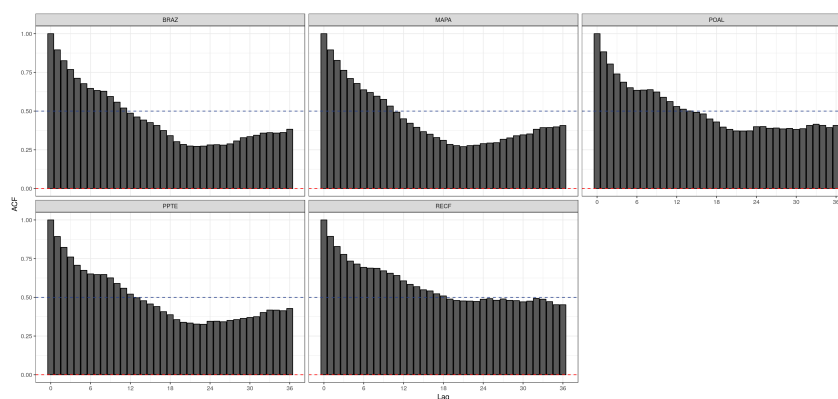


Figura 18 – Função de autocorrelação temporal de todas as estações.

Vê-se na Figura 18 que a avaliação torna-se difícil para um maior número de estações. Desta forma, pode-se alterar o método comum de visualização adotando o princípio do mapa de calor.



Figura 19 – Função de autocorrelação temporal de todas as estações.

Na Figura 19 pode-se avaliar que todas as autocorrelações são positivas, em que a maioria decresce constante com o passar dos *lags*.

Analogamente, pode-se fixar o tempo, para exemplificação faz-se a visualização do último dia (31/12) de cada ano.

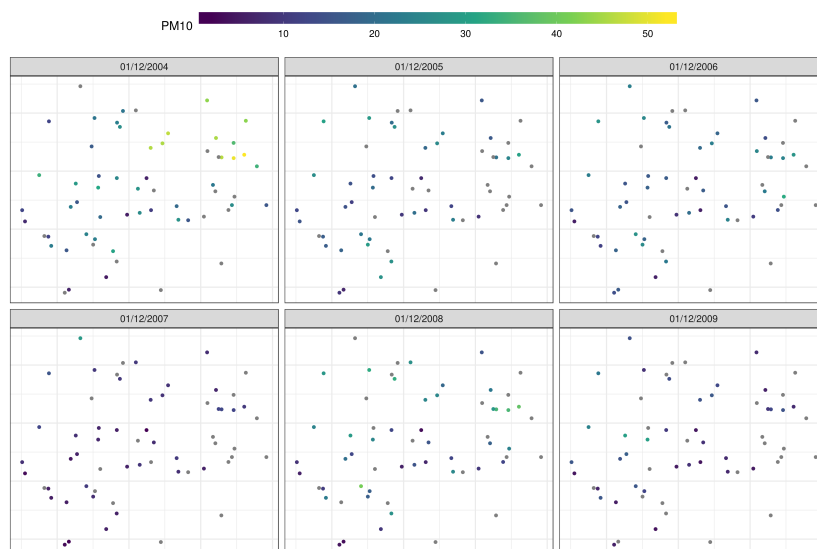


Figura 20 – Concentração de PM10 nas estações no último dia de cada ano.

Pela Figura 20 fica difícil a visualização de algum efeito, uma forma é o cálculo da semivariância espacial para cada tempo.

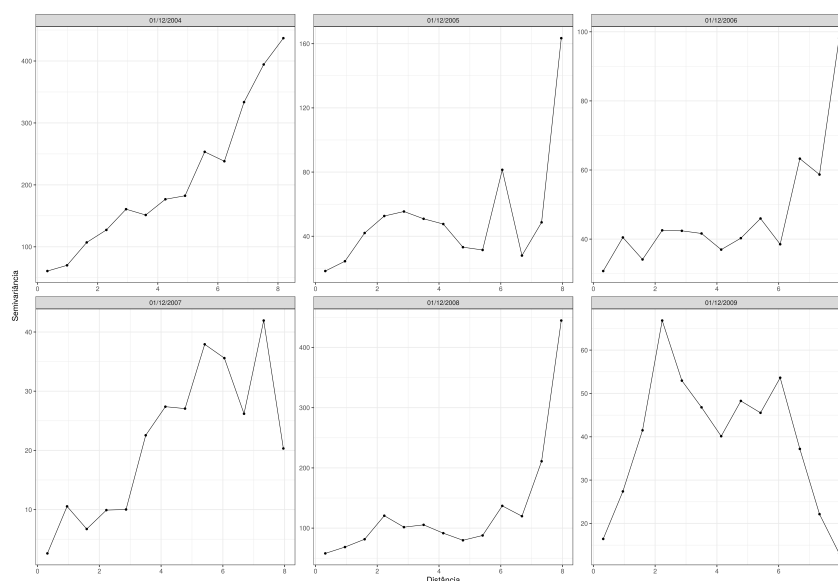


Figura 21 – Semivariograma de cada último dia de cada ano.

Vê-se na Figura 21 que os semivariogramas são muito distintos, e isto para 6 dias selecionados arbitrariamente. Além disso, se ao fixar o espaço para as 65 estações já era quase inviável a análise visual, ao fixar tempo tem-se 2192 dias, ou seja, a visualização um a um é impraticável, uma alternativa é o uso do gráfico mapa de calor (do inglês, *heatmap*).

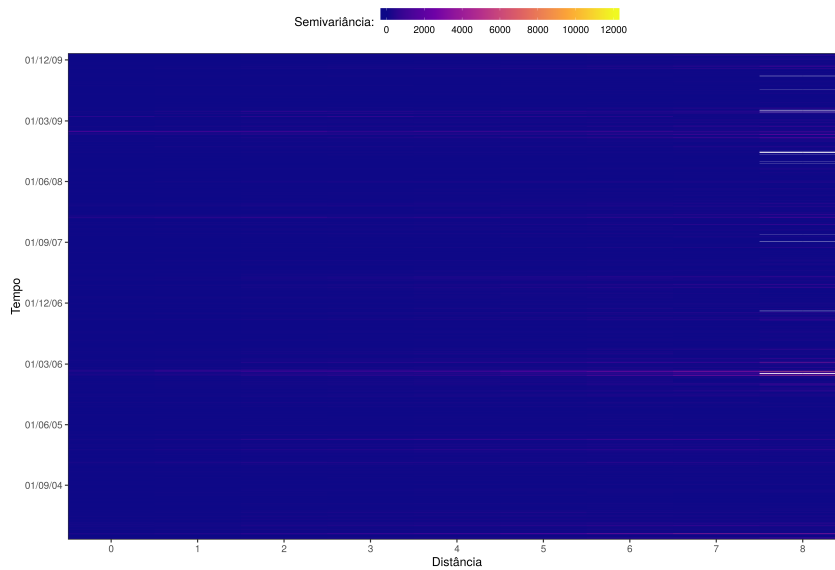


Figura 22 – Mapa de calor das semivariâncias para cada dia.

Na Figura 22 tem-se 2192 semivariogramas, em que além da alta quantidade de informação há a problemática da magnitude da semivariância, visto que em alguns dias há altas estimativas impossibilitando a visualização do comportamento, visto que uma vez que haja um valor discrepante os demais valores menores aparentam ter a mesma cor no mapa de calor, impossibilitando uma interpretação precisa. Entretanto, é importante ressaltar que todos os semivariogramas aqui foram calculados com a mesma distância, no caso a distância máxima entre dois pontos.

Como alternativa, pode-se fazer o mesmo gráfico com a escala logarítmica.

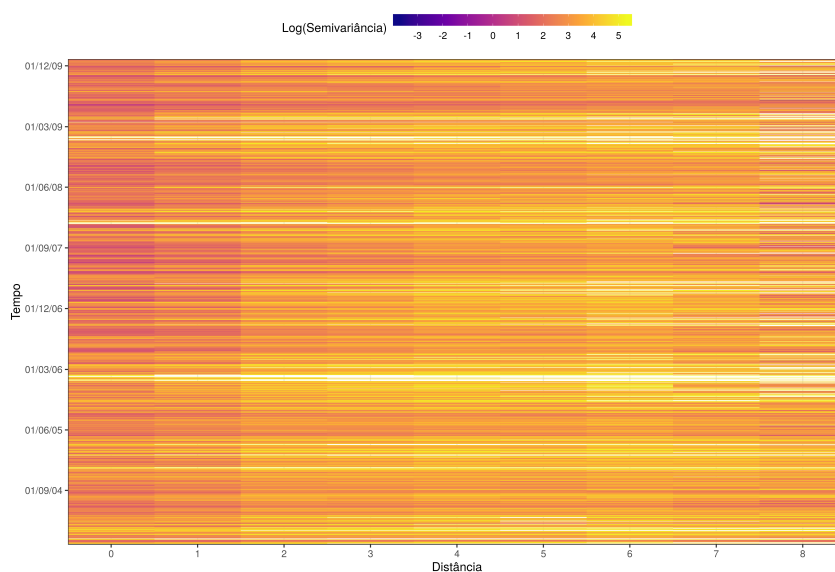


Figura 23 – Mapa de calor das semivariâncias (*log*) para cada dia.

Nota-se na Figura 22 que com a aplicação do logaritmo é possível notar alguns efeitos,

por exemplo, como a semivariância aumenta e aparentemente se estabiliza para alguns tempos.

Por fim, faz-se o gráfico do semivariograma espaço-temporal empírico.

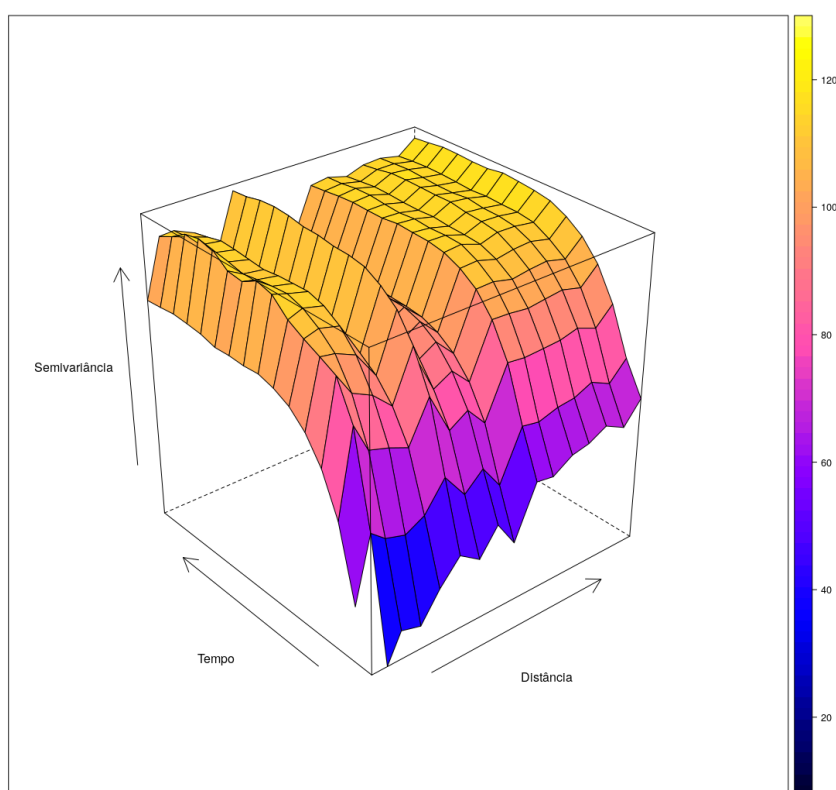


Figura 24 – Semivariograma empírico 3D.

Pela Figura 24 observa-se o comportamento do semivariograma, em que se nota que em determinadas distâncias há uma flutuação na semivariância, enquanto no tempo o comportamento mantém-se próximo ao caso exclusivamente espacial.

Ainda, é possível a avaliação do semivariograma na forma de mapa de calor.

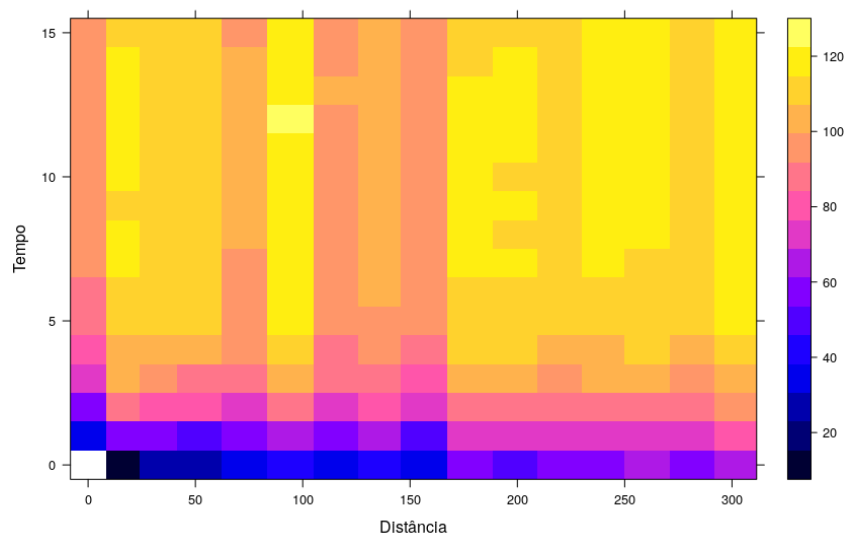


Figura 25 – Semivariograma empírico 2D.

O semivariograma empírico 2D permite a identificação dos mesmos efeitos, visto que semivariância aumenta com o tempo e alterna entre as distâncias, como visto na Figura 25.

7.3 Discussão

No dados aplicados nota-se que as alternativas gráficas propostas facilitam a identificação de efeitos e entendimento dos dados.

CAPÍTULO 8

COORDENADAS GNSS

8.1 Definição do problema

A aplicação em dados reais será em coordenadas de estações de monitoramento contínuo GNSS (*Global Navigation Satellite System*), a fim de investigar efeitos em altimetria. Assim define-se os conceitos relacionados, bem com os problemas a serem resolvidos.

O conceito de GNSS, em especial, o *Global Navigation System* (GPS), consiste em um conjunto de satélites que emitem sinais com informações essenciais para o posicionamento de veículos, pessoas ou qualquer objeto de interesse na superfície terrestre. Entretanto, tais informações podem estar comprometidas por influências externas, as quais podem ocorrer nos satélites, na propagação do sinal e no próprio local de recebimento do sinal, chamado de estação GNSS, onde fica o receptor/antena (MONICO, 2000). Assim, uma modelagem adequada de tais efeitos é fundamental (BRASSAROTE, 2014), visto que este sistema é essencial para diversas áreas, como: agricultura, transporte, meteorologia e militar. Algumas aplicações incluem a umidade do solo e altimetria de aeronaves (MASTERS, 2004), propagação de satélites ao longo de uma via férrea (MARAIS; LEFEBVRE; BERBINEAU, 2004), sistemas de monitoramento de terra e água (EGIDO et al., 2009), meteorologia (POLI et al., 2008), precisão em pouso (KELLY; DAVIS, 1994) e controle de tráfego aéreo (PILLEY, 1996).

Vê-se que além de inúmeras, estas aplicações envolvem áreas de suma importância que exigem um alto grau de precisão e acurácia. Para tal, utiliza-se de um sistema de posicionamento relativo em que coordenadas de pontos desconhecidos são determinados em função de pontos com coordenadas conhecidas, os quais fazem parte das chamadas redes de estações de referência. Diversas redes GNSS de monitoramento contínuo estão disponíveis, das quais podemos citar as redes IGS (do inglês, *International GNSS Service*), SIRGAS-CON (Sistema de Referência Geocêntrico para as Américas), RBMC (Rede Brasileira de Monitoramento Contínuo), GNSS-SP (Rede GNSS do Estado de São Paulo), dentre outras em âmbito global ou regional. Devido ao crescente avanço e popularização do GPS, cada vez mais usuários utili-

zam os dados de tais redes. Mesmo as redes brasileiras, tais como a RBMC, desempenham importante papel nas diferentes aplicações em nível nacional e internacional, além de pesquisas científicas, garantindo aos usuários alto nível de precisão nas suas coordenadas e maior produtividade nos levantamentos, o que leva a custos menores.

Na área científica, a permanente coleta de observações permite a estimação contínua das coordenadas das estações constituintes, monitorando, deste modo, as deformações da crosta terrestre, dentre outros efeitos. Consequentemente, a rede permite quantificar a variação temporal das coordenadas proporcionando um referencial constantemente atualizado aos usuários (PEREIRA; APARECIDO; FORTES, 2004).

Assim, é de extrema importância a modelagem de tais séries temporais para que erros na determinação de tais coordenadas não sejam propagados para todas as aplicações envolvidas. No que se refere à planimetria, modelos existentes conseguem entregar resultados bastante satisfatórios. Entretanto, a modelagem da componente altimétrica ainda permanece um problema em aberto, além de diversos fenômenos naturais que justificam a importância de modelagem dos efeitos causados na componente vertical das séries temporais de coordenadas GNSS.

FREITAS e CORDINI (1998) destacam a importância do estudo da maré terrestre na correta determinação do posicionamento geodésico, principalmente no que se refere à componente altimétrica. A maré terrestre é um fenômeno que consiste em uma deformação visco-elástica do globo terrestre causada pela ação gravitacional da Lua e do Sol. A Terra, a Lua e o Sol participam de um processo gravitacional interativo do qual resultam esforços diferenciais significativos desde o interior até as partes mais superficiais do planeta. As variações de distância do Sol e da Lua em relação à Terra produzem uma alteração na força da gravidade, cujo fenômeno é chamado de força de maré. Assim, a Terra, ao invés de um corpo rígido, é deformável, de modo que a ação da força de maré em cada porção do planeta causa deformações com características viscoelásticas, variáveis espaciais e temporais (FREITAS; CORDINI, 1998). Dentre os tipos de maré terrestre, a gravimétrica é a que mais afeta a componente vertical, chegando a causar deformações de até 50 cm nesta componente e 15 cm na horizontal (PEREIRA; FREITAS; MANTOVANI, 2009).

Assim, observa-se a necessidade de consideração além dos efeitos observados no tempo, mas também espacialmente. Outros efeitos reforçam tal preocupação com o espaço, tais como, o movimento das placas tectônicas ou litosféricas e a geotectônica de bacias sedimentares, as quais preservam um registro detalhado do ambiente e dos processos tectônicos que deram forma à superfície da Terra ao longo do tempo. Tais bacias também servem como importante repositório de recursos naturais, como água subterrânea, petróleo e recursos minerais diversos.

As deformações constantes na superfície terrestre também podem causar mudanças bruscas no nível médio das séries temporais de coordenadas das estações GNSS, efeito chamado de "jump geofísico", o qual pode ser permanente ou temporário, passando por um período

de acomodação. Dada a instabilidade temporal, existe uma dificuldade de usar componentes periódicas na modelagem. Entretanto, não se pode descartar a existência de ciclos hidrológicos ocorrido pelo movimento contínuo da água presente nos oceanos, continentes (superfície, solo e rocha) e na atmosfera. Esse movimento é alimentado pela força da gravidade e pela energia do Sol, que provocam a evaporação das águas dos oceanos e dos continentes.

Do exposto, parece natural que a variabilidade das séries temporais de coordenadas GNSS dependa da região geográfica em que se encontra a estação. Pode-se citar, por exemplo, as estações localizadas entre placas, tais como Concepción, no Chile, as quais podem apresentar um deslocamento horizontal (descontinuidade) das coordenadas em razão de terremotos. Já nas estações da Região Amazônica, percebe-se uma variação sazonal na componente vertical, cujas altitudes sofrem variação de poucos centímetros devido à hidrologia continental (ROCHA et al., 2017).

Verifica-se assim a necessidade de desenvolvimento de um modelo espaço-temporal das séries de coordenadas de estações GNSS para que os efeitos causados pela dinâmica natural Terra-Sol-Lua possam ser considerados. Uma vez que um modelo espaço-temporal esteja disponível, estimativas referentes à componente altimétrica poderão ser estimadas para localidades dentro da extensão modelada que não possuem estação GNSS, melhorando muito a qualidade do posicionamento, além da redução de custos nos levantamentos.

Portanto, pretende-se com este trabalho desenvolver um modelo espaço-temporal que represente a variabilidade das coordenadas altimétricas de estações GNSS no Brasil, que permita a estimação de tal variabilidade para coordenadas em qualquer instante e local dentro do território brasileiro.

Para modelagem utilizar-se-á as estações pertencentes à RBMC dos Sistemas GNSS (Figura 26), estas são fixadas com pinos de centragem forçada, especialmente projetados, e cravados em pilares estáveis e isolados (Figura 27). A maioria dos receptores da rede possui a capacidade de rastrear satélites GPS e GLONASS (do russo, *Globalnaya navigatsionnaya sputnikovaya sistema*). Esses receptores coletam e armazenam observações no tempo e espaço transmitidos pelos satélites destes sistemas (GPS ou GLONASS).



Figura 26 – Estações brasileiras de monitoramento contínuo (RMBC). (Fonte: IBGE)

Na Figura 26 vê-se que as estações são poucas em determinadas regiões, em relação à extensão do Brasil. Logo, a modelagem dos dados temporais de tais estações a partir de uma abordagem espaço-temporal seria de grande valia para que tais informações fossem disponibilizadas principalmente em regiões em que não se tem uma estação por perto, fato que conduz à redução da acurácia e precisão de qualquer posicionamento que utilize dados de tais estações.



Figura 27 – Exemplo de estação na Universidade Federal de Pelotas. (Fonte: IBGE)

Pela Figura 27 observa-se como é uma estação, estas apesar de isoladas e bem fixadas, acabam sendo afetadas pelas deformações constantes na superfície terrestre. Ainda, as

coordenadas das estações utilizam de um sistema local, para uma maior precisão e obtenção da altimetria (Para mais detalhes, checar o Apêndice B).

8.2 Descritiva

Os dados selecionados são de Estações brasileiras concentradas em sua maioria na região Sul do país, como se vê na Figura 28.



Figura 28 – Mapa do Brasil com as estações utilizadas.

Posteriormente, explora-se como se comporta a variável de interesse, desconsiderando as dependências.

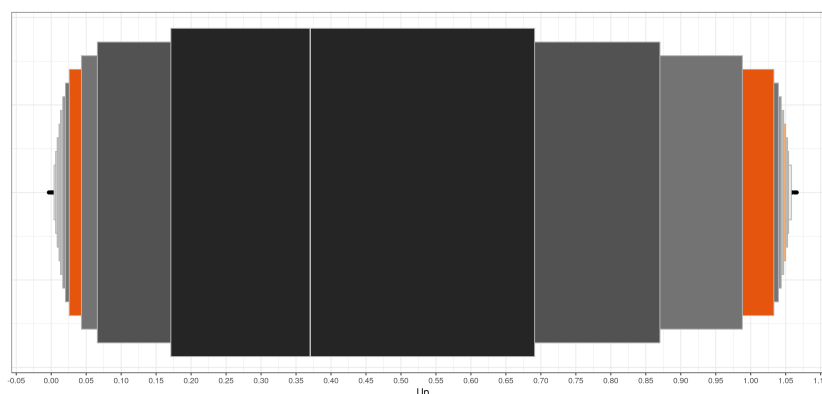


Figura 29 – Gráfico adimensional da coordenada vertical.

O gráfico *letter-value* é similar ao *boxplot*, tal que a caixa maior (cinza escuro) representa 50% dos dados, a segunda maior 25% e assim por diante. Logo, pela Figura 29 fica visível a assimetria dos dados, bem como a presença de alguns possíveis valores discrepantes, visto que este método possui um critério mais robusto na presença de uma maior quantidade de dados.

Visto o comportamento adimensional, segue-se então para análise exploratória temporal.

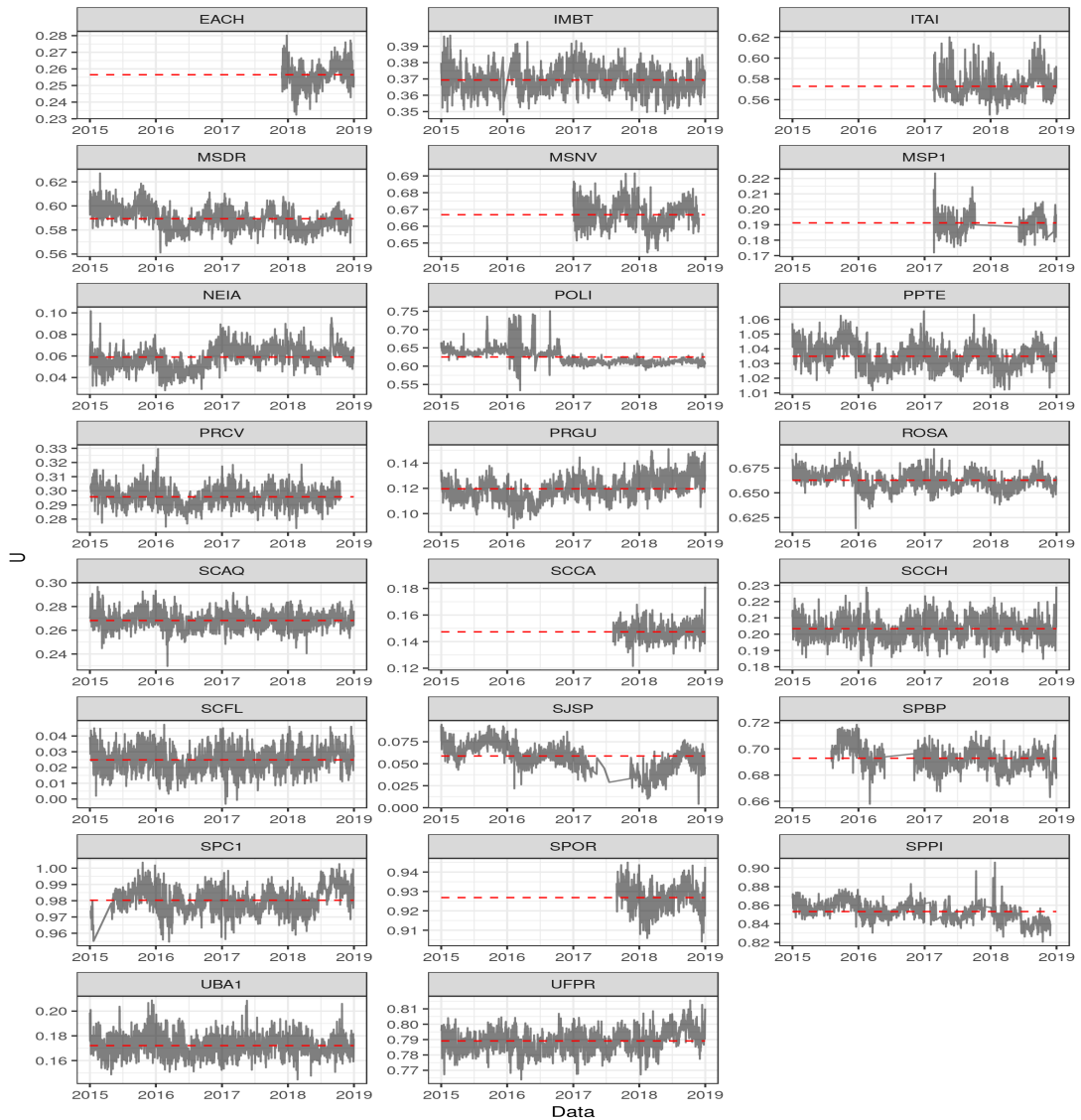


Figura 30 – Gráfico temporal das diferentes estações.

Observa-se na Figura 30 que os dados coletados vão de 2015 a 2019, porém algumas estações não existiam durante todo o período (e.g., EACH), assim considerou-se somente o período de 2018.

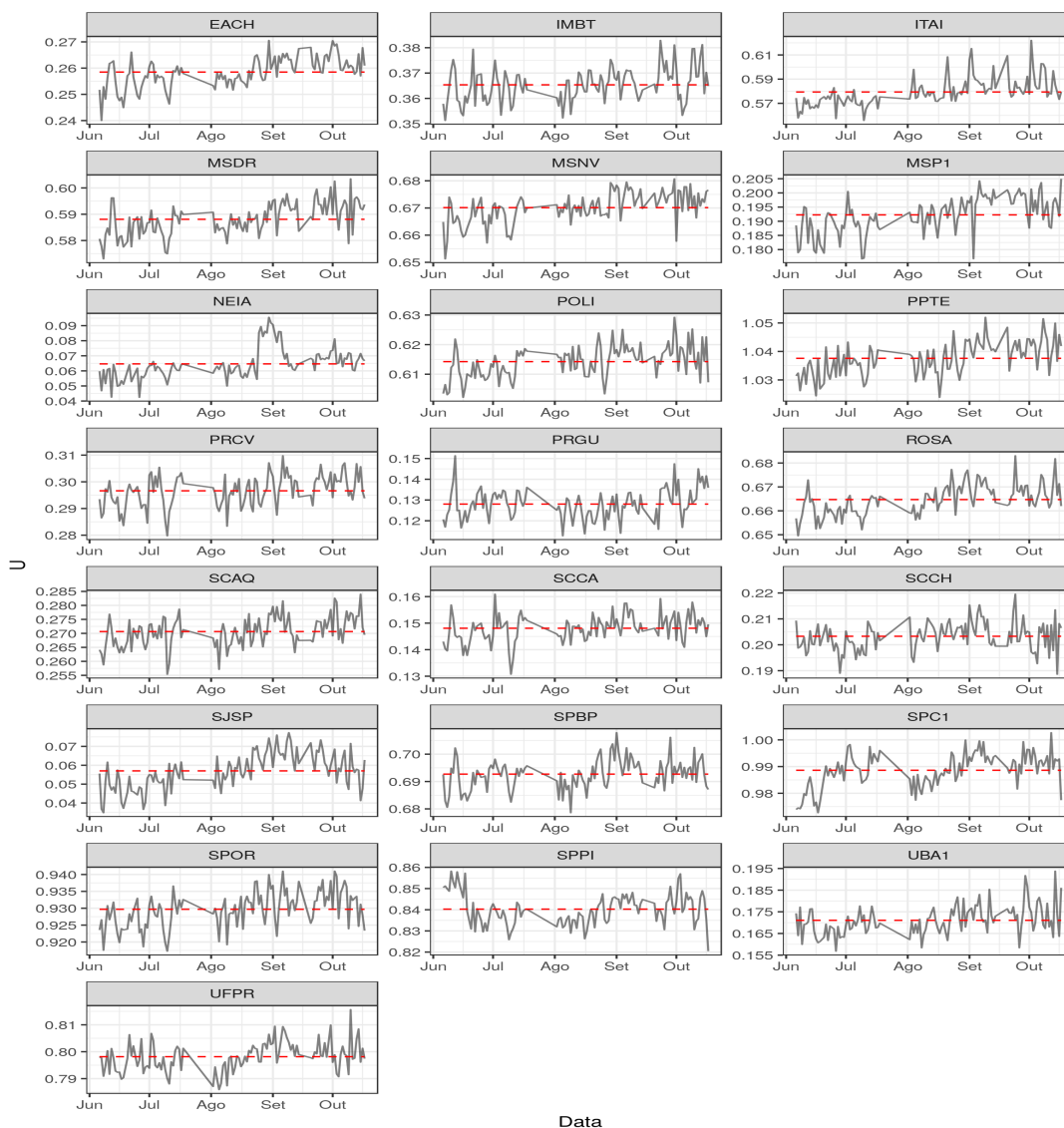


Figura 31 – Gráfico temporal das diferentes estações, no ano de 2018.

Ao se analisar a Figura 31 notam-se indicativos de diversos efeitos nas séries, como possíveis tendência (e.g., MSNV) ou até sazonalidade (e.g., ITAI). Busca-se assim avaliar a presença destes efeitos através de algumas métricas. Não só isso, mas como em um período que antecede Agosto há um ausência de dados.

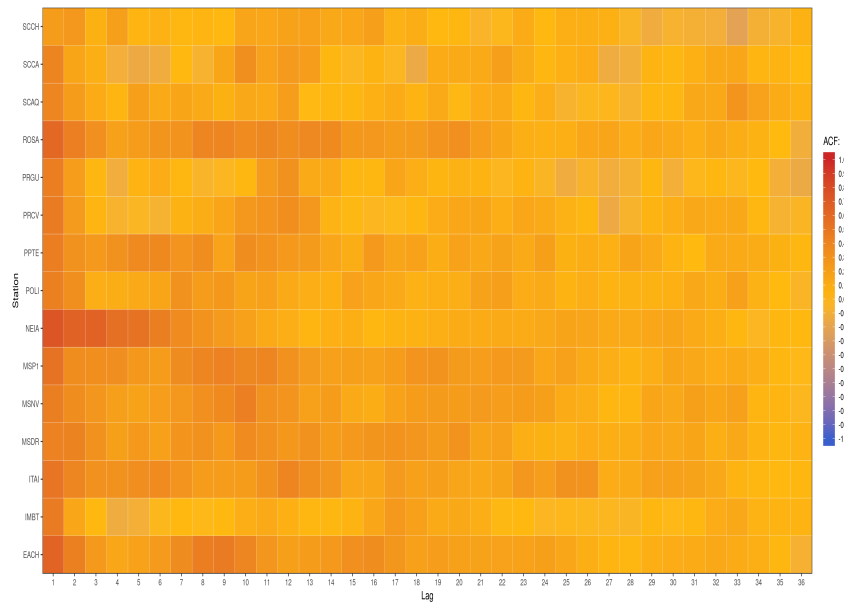


Figura 32 – Mapa de calor da função de autocorrelação.

Observa-se pela Figura 32 que todas as séries possuem uma ACF positiva que decresce lentamente conforme o *lag* aumenta.

Estação	Estab.	Granulosidade	Hurst	Tend.	Curva.	Linear.
EACH	0.4128	0.3027	0.8240	0.1888	3.7114	1.2531
IMBT	0.3917	0.2338	0.8347	0.1604	3.3318	4.1804
ITAI	0.3210	0.4758	0.8259	0.4268	-0.7844	2.4477
MSDR	0.5758	0.1302	0.8360	0.2882	9.8111	-0.2444
MSNV	0.4910	0.1537	0.8499	0.4187	-7.6616	-1.5080
MSP1	0.3531	0.7250	0.7982	0.5113	0.5599	-9.3594
NEIA	0.5525	0.1471	0.8492	0.4750	-1.3212	3.1580
POLI	0.5921	2.0716	0.9567	0.1836	2.2764	-0.8994
PPTTE	0.5200	0.1055	0.8424	0.1706	1.5474	-0.4496
PRCV	0.3733	0.2566	0.7902	0.1651	2.7449	-1.3367
PRGU	0.5913	0.0688	0.8690	0.1752	3.4646	3.2060
ROSA	0.5484	0.2448	0.8340	0.4004	-3.9488	-4.0677
SCAQ	0.2945	0.3049	0.7463	0.3129	-0.3854	-3.9113
SCCA	0.2430	0.3015	0.7385	0.2537	5.6727	-0.3692
SCCH	0.2939	0.2414	0.7482	0.3169	-2.0805	-0.0911
SCFL	0.3308	0.1983	0.7865	0.1280	5.2786	2.8204
SJSP	0.7401	0.0405	0.8830	0.1489	-0.0572	2.7337
SPBP	0.5136	0.1242	0.8385	0.2550	2.5364	2.9601
SPC1	0.5085	0.1445	0.8473	0.3412	-0.2779	-0.2541
SPOR	0.4026	0.1619	0.7850	0.2409	-1.2326	-4.5516
SPPI	0.6523	0.1781	0.8902	0.1356	-0.1810	-2.2334
UBA1	0.2355	0.2864	0.6891	0.2102	5.9681	-0.3124
UFPR	0.4500	0.1469	0.8344	0.3424	3.2709	4.7128

Tabela 3 – Métricas para séries temporais.

Pela Tabela 3 nota-se as métricas descritivas para todas as estações, tal que se destaca a estabilidade da estação SJSP, como também sua granularidade, entretanto este método utiliza intervalos pré-definidos para todas as séries, podendo gerar resultados capciosos. Além disso, nota-se que a o coeficiente de *Hurst* foi razoavelmente próximo entre as séries, e como maior que 0,50 é um indicativo de “persistência”. A métrica de tendência mostrou-se bem variada, porém quando comparada a Figura 31 é satisfatória na identificação da tendência, o que surpreende uma vez que sendo medida única identifica um efeito que em uma série pode estar presente uma ou mais vezes em diferentes janelas de tempo. Por fim, as métricas de curvatura e linearidade tornam-se de difícil avaliação ao comparar com a avaliação visual.

Avaliado o problema sob o aspecto temporal, analisa-se agora pelo viés espacial, em que se utiliza da semivariância espacial, visto que essa é a forma mais utilizada na exploratória espacial para este tipo de problema, mas conta com diversas dificuldades, uma vez que se faz o ajuste automático para todo tempo.

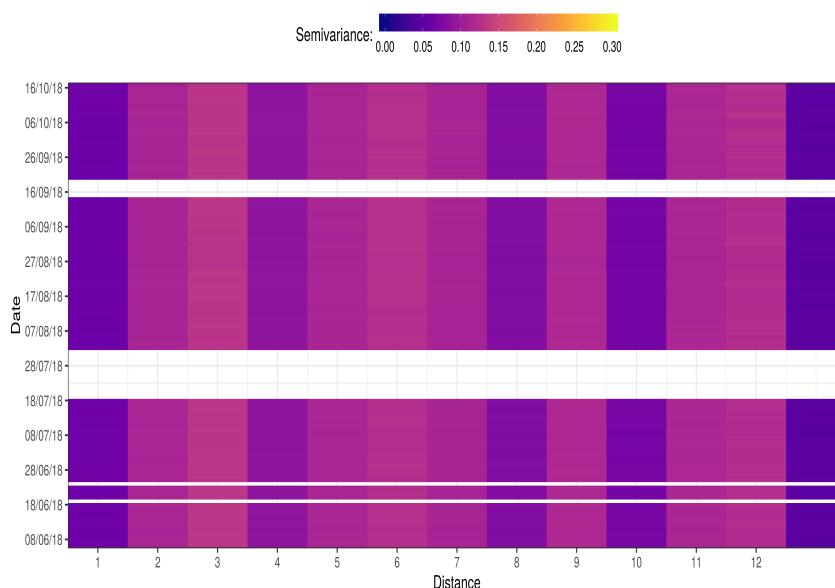


Figura 33 – Mapa de calor da semivariância espacial.

Na Figura 33 nota-se que a semivariância não possui uma magnitude tão grande, como no caso da poluição, logo não se aplica logaritmo. Entretanto ao se avaliar o mapa de calor observa-se que a semivariância aparenta alternar-se com a distância, o que não é um comportamento desejado uma vez que se espera a estabilização da mesma.

Feita as análises exploratórias unidimensionais viu-se que o comportamento espacial indica uma possível problemática, já a análise temporal mostrou que as séries apesar de possuírem uma função de ACF próxima possuem demais efeitos de forma bem distinta. Portanto, segue-se para avaliação espaço-temporal.

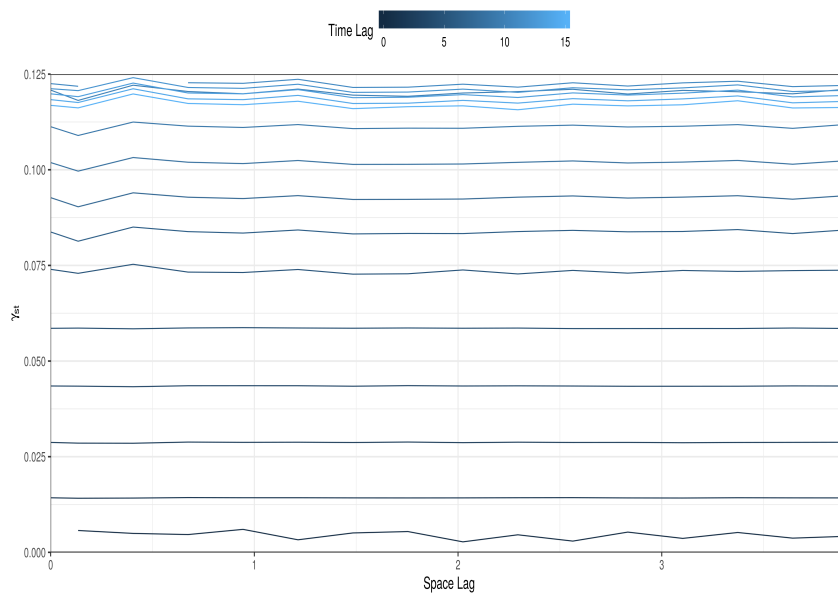


Figura 34 – Semivariância espaço-temporal amostral em linhas.

Ao analisar a Figura 34 nota-se alguns efeitos preocupantes como alguns tempos em que a semivariância é aparentemente constante, sendo que para tempos maiores há maior indício de um comportamento da semivariância, logo alterna-se o formato.

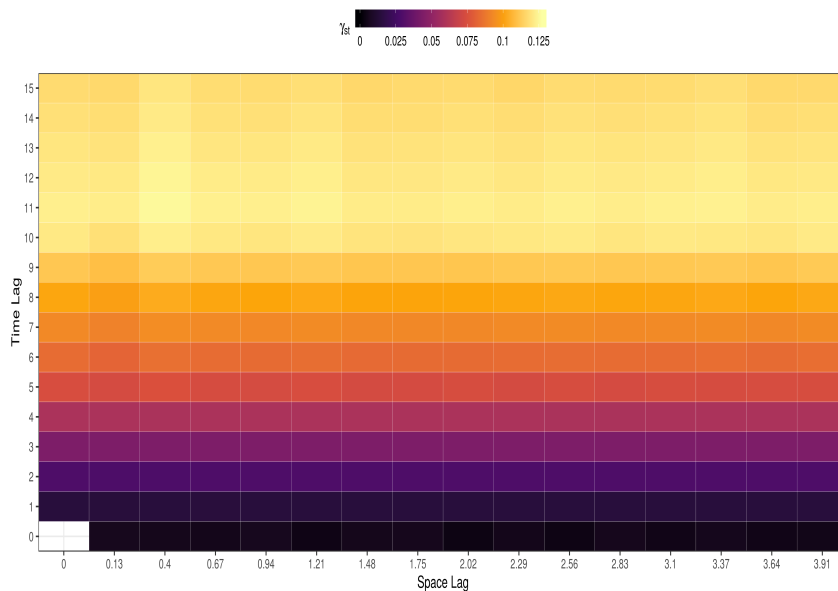


Figura 35 – Semivariância espaço-temporal amostral em formato de mapa de calor.

Utilizando a alternativa do mapa de calor na Figura 35 evidencia-se a suspeita da Figura 34 uma vez que fica nítido como a semivariância aumenta em função do tempo e praticamente mantém-se constante em função do espaço, sendo um indicativo de separabilidade.

8.3 Modelagem

Visto o indicativo de separabilidade faz-se um ajuste inicial de um modelo exponencial-exponencial, do tipo “metric”, por um método de ajuste automático do pacote *CompRandFld*.

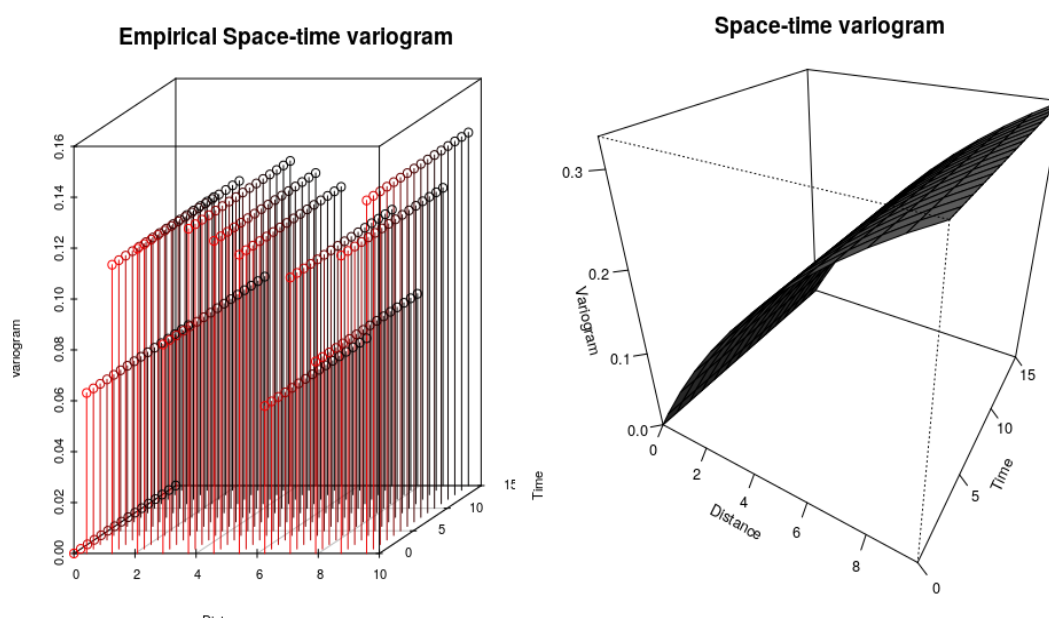


Figura 36 – Semivariância espaço-temporal amostral e ajuste do modelo exponencial-exponencial.

Na Figura 36 é plotado tanto o semivariograma empírico quanto o ajuste, nota-se que o modelo não captura o efeito presente.

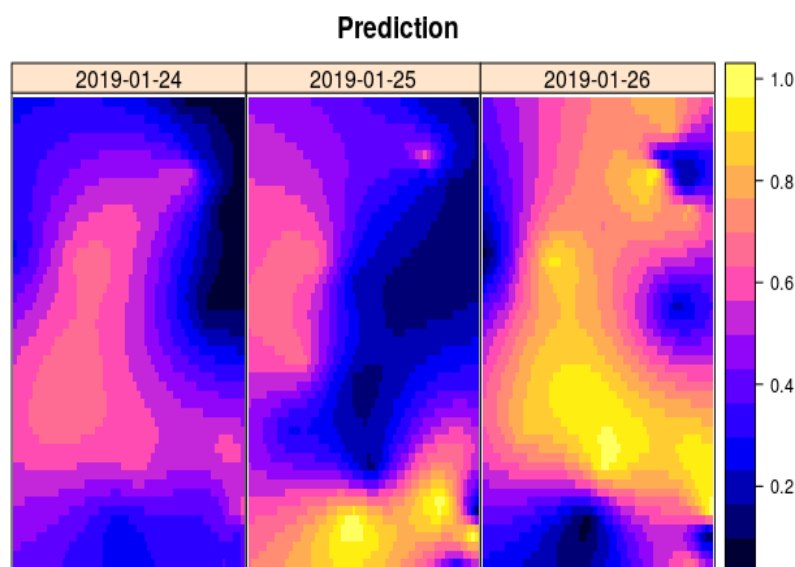


Figura 37 – Predição do modelo exponencial-exponencial.

Após este ajuste faz-se uma modelagem customizada a fim de capturar os efeitos da

série. Viu-se que os dados possuem um indicativo forte de separabilidade, assim utiliza-se á esta classe de modelo. Como visto na revisão, o modelo *product-sum* é muitas escolhido por sua sofisticação, porém o mesmo conta com um termo de produto, o que pode não ser interessante visto a fraca interação espaço-tempo.

Assim, ajustam-se 2 modelos:

- Modelo 2: classe *metric*, tipo exponencial-exponencial com chutes iniciais, em que patamar, alcance, pepita e anisotropia espaço-temporal são 0,88; 90, 0 e 1;
- Modelo 3: classe *product-sum*, tipo esférico-exponencial com chutes iniciais, em que patamar, alcance, pepita e k são 0,20; 100, 0 e 17.

Em que os chutes e modelos escolhidos foram feitos via tentativa erro, sendo escolhidos de forma visual ao comparar com o semivariograma empírico.

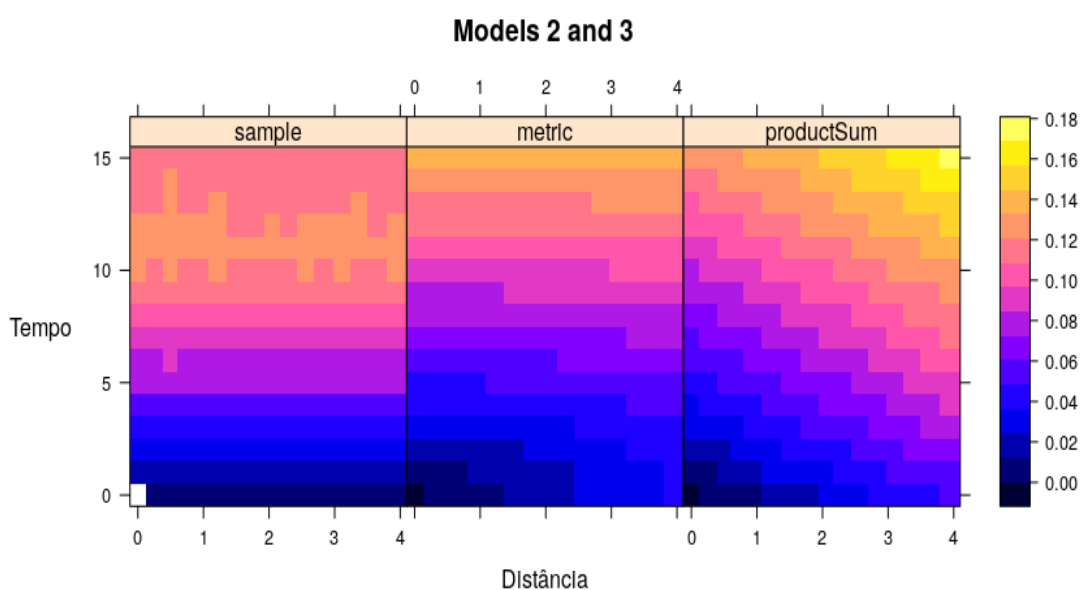


Figura 38 – Ajuste dos modelos 2 e 3.

Pela Figura 38 comprova-se a suspeita do modelo *Product-sum*, visto que o modelo possui um efeito diagonal por considerar uma interação forte entre espaço e tempo. Já o modelo 2 mostra-se mais adequado, pois foi construído de forma a ser mais estável em função do espaço, entretanto como já havia sido notado o semivariograma empírico possui uma semivariância que aumenta com o tempo até um determinado *lag* e ao de se estabilizar a mesma decresce, assim obtém-se um novo semivariograma empírico com corte temporal no tempo 10.

Assim, faz-se um ajuste de novo modelo da classe *metric* uma vez que seu comportamento assemelhasse aos dados, sendo necessária outra avaliação visual para obtenção de

chutes iniciais, em que o modelo 4 é da classe *metric*, tipo exponencial-exponencial com chutes iniciais, em que patamar, alcance, pepita e anisotropia espaço-temporal são 1; 150, 0 e 2.

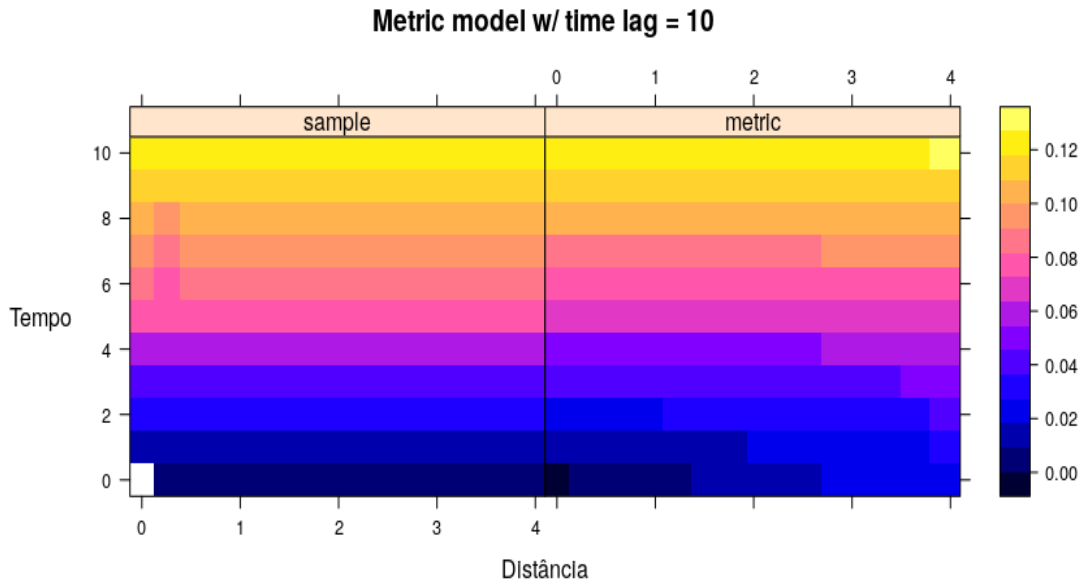


Figura 39 – Ajuste do modelo 4.

Como se vê na Figura 39 o semivariograma agora se estabiliza e o ajuste está visualmente próximo do empírico. Obtendo-se os valores -0,397 e 0,9936 para as métricas r^2 , NMSE, respectivamente.

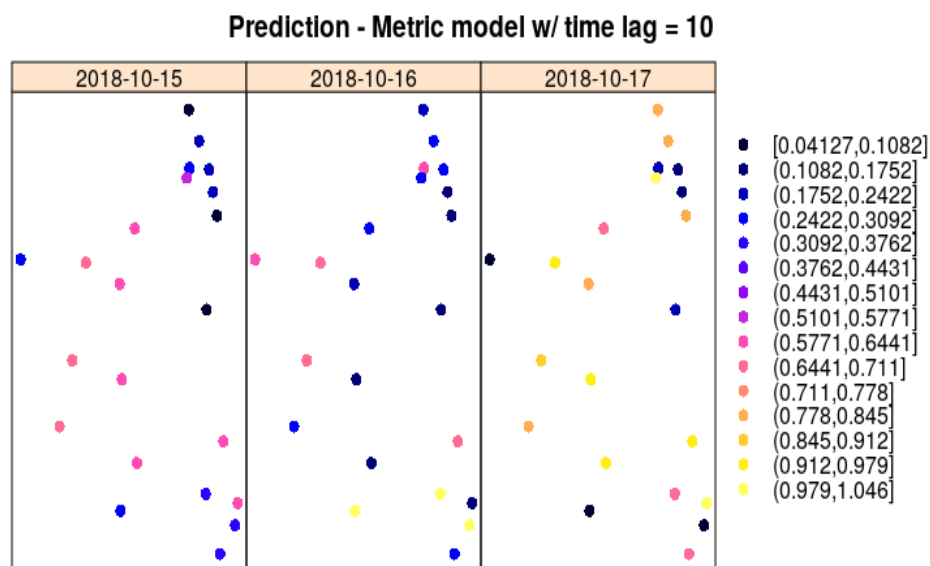


Figura 40 – Estimativa pontual das estações, obtida do modelo 4.

O modelo apesar de aparentar “satisfatório” ao utilizar o critério visual, mostra-se ruim quanto as métricas, uma vez que r^2 é baixo, assim como NSME que é normalizado também

possui valor moderado. mostra-se os valores obtidos das estações nas últimas observações do banco de dados.

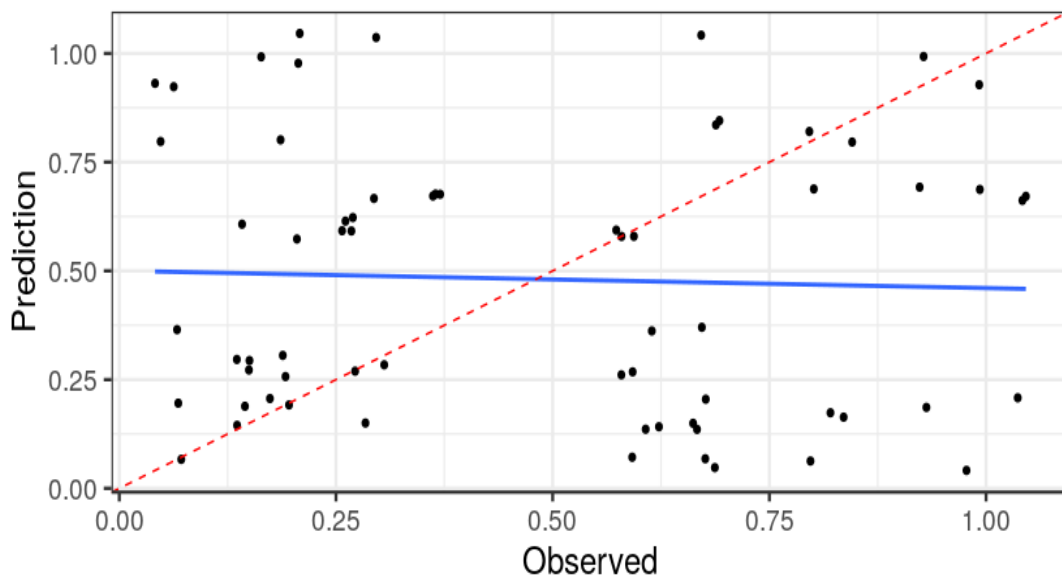


Figura 41 – Comparativo entre predito e observado.

Nota-se na Figura 41 que de fato o modelo está longe do ideal, ainda mais em um cenário que exige alta precisão. Ainda sim, faz-se a interpolação espacial para fim de avaliar os resultados.

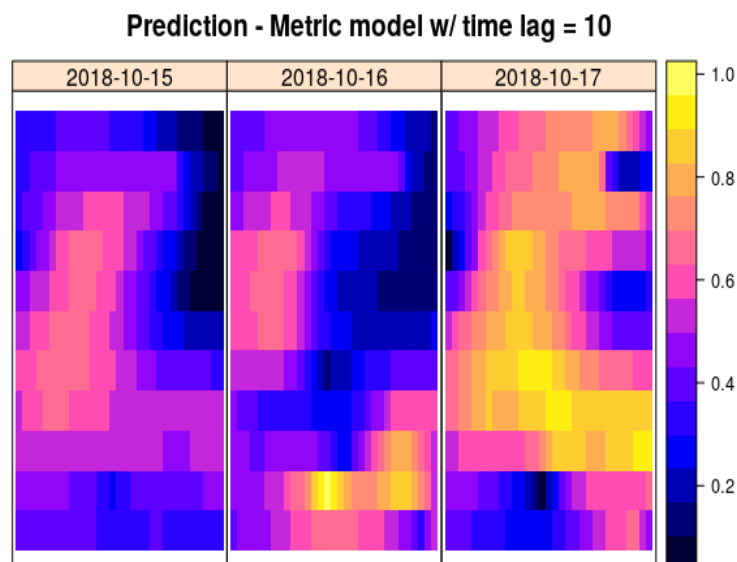


Figura 42 – Interpolação espacial, obtida do modelo 4.

Na Figura 42 tem-se a predição de pontos não existentes nos últimos 3 dias observados, nota-se que a mudança ao longo dos dias é drástica, em que se aumenta o valor da variável.

8.4 Discussão

Viu-se que a análise exploratória facilitou a modelagem, uma vez que indicou a presença de efeitos para escolha de modelos. Além disso, os dados apresentam uma alta dependência temporal na presença do modelo espaço-temporal, o que na análise só temporal não aparentava de forma tão explícita. Porém, o modelo não se mostrou excelente, visto que possui indicativos de uma performance medíocre por suas métricas, além disso sua interpolação mostra uma alta dependência espacial com a variável de interesse, visto seu comportamento na Figura 42, o que não era tão esperado vide o semivariograma empírico, não só isso mas o tempo implicou em uma crescente alta de um dia para o outros, sendo este um comportamento não tão rápido quando comparadas as séries na análise descritiva temporal.

Parte IV

Considerações finais

O trabalho apresenta modelos com diferentes vertentes, mostrando algumas de suas características. Em suas aplicações, traz-se as dificuldades envolvidas na modelagem, ao seguir as indicações encontradas na literatura, sendo elas:

- Métodos visuais ineficientes tanto na identificação de efeitos, como em lidar com quantidades razoáveis de regiões e/ou tempos;
- Pacotes estatísticos com baixa compatibilidade ou sem automatização de funções;
- Alto tempo de compilação para estimação do semivariograma amostral, visto que para o simples exemplo apresentado o tempo foi de aproximadamente 10 minutos;
- Ausência de indicativos para escolha dos modelos separáveis;
- E apesar de omitidos do exemplo, destaca-se que também há falta de insumo para interpretação do semivariograma espaço-temporal e de medidas para diagnóstico do ajuste e validação da predição.

Assim, foram realizadas:

- Novas formas de exploração dos efeitos no tempo e espaço, que provejam indicativos para escolha do modelo e seus respectivos parâmetros, como uso da função de autocorrelação parcial e de decomposições espectrais e *wavelets*;
- Funções que estruturam os dados automaticamente, e permitam a análise descritiva e a preparação dos dados para modelagem, sendo funções otimizadas e paralelizadas, vide o tempo longo de execução das funções atuais;
- Formas de apresentação dos parâmetros e superfícies estimadas aliada à presença de efeitos teóricos contidos nos dados, a fim de facilitar a modelagem.

Além do que, apesar de não utilizados na prática ressalta-se que os modelos não-separáveis são importantes, porém possuem uma maior dificuldade da modelagem, uma vez que necessitam de funções mais complexas, tal que para sua definição é necessário um maior entendimento do comportamento dos dados, sendo indicado em casos que há uma modelagem fenomenológica conhecida.

Conclui-se então que o trabalho fornece um processo e arsenal de ferramentas para modelagem em diferentes casos, considerando diferentes cenários, como: dimensionalidade, separabilidade, variabilidade, entre outros. Porém o método para ajuste de um modelo tanto pesquisado, quanto aplicado ainda se mostra extremamente manual e customizado, assim para trabalhos futuros sugere-se:

- Trabalho de simulação extensivo dos modelos existentes a fim de mostrar características empíricas, combinando diferentes classes, tipos e parâmetros de modelos;
- Investigação dos modelos de mistura e correlação, uma vez não foram utilizados aqui e possuem um desenvolvimento em trabalhos recentes;
- Desenvolvimento em outras ferramentas para integração de métodos de modelagem e visualização, uma vez que os pacotes atuais do *R* mostraram-se extremamente difíceis de manusear, visto a não integração entre as funções;
- Possibilidade de integrar práticas da abordagem de *Machine Learning* para automatização na escolha de parâmetros, bem como melhora da predição;
- Melhora nas métricas e forma de validação, pois ainda se mostram em um estado inicial de desenvolvimento, já que se utilizam as mesmas do caso espacial.

REFERÊNCIAS

- BECK, N.; KATZ, J. N. Modeling dynamics in time-series–cross-section political economy data. *Annual Review of Political Science*, Annual Reviews, v. 14, p. 331–352, 2011. Citado na página 17.
- BÉLISLE, C. J. Convergence theorems for a class of simulated annealing algorithms on rd. *Journal of Applied Probability*, Cambridge University Press, v. 29, n. 4, p. 885–895, 1992. Citado na página 71.
- BERNSTEIN, S. Sur les fonctions absolument monotones. *Acta Mathematica*, Springer, v. 52, n. 1, p. 1–66, 1929. Citado na página 53.
- BEYGELZIMER, A.; KAKADET, S.; LANGFORD, J.; ARYA, S.; MOUNT, D.; LI, S. Fnn: fast nearest neighbor search algorithms and applications. *R package version*, v. 1, 2013. Citado na página 73.
- BILONICK, R. A. The space-time distribution of sulfate deposition in the northeastern united states. *Atmospheric Environment (1967)*, Elsevier, v. 19, n. 11, p. 1829–1845, 1985. Citado na página 37.
- BIVAND, R. S.; PEBESMA, E. J.; GOMEZ-RUBIO, V.; PEBESMA, E. J. *Applied spatial data analysis with R*. [S.l.]: Springer, 2008. v. 747248717. Citado na página 75.
- BLOOMFIELD, P.; OEHLERT, G.; THOMPSON, M. L.; ZEGER, S. A frequency domain analysis of trends in dobson total ozone records. *Journal of Geophysical Research: Oceans*, Wiley Online Library, v. 88, n. C13, p. 8512–8522, 1983. Citado na página 19.
- BOCHNER, S. *Harmonic analysis and the theory of probability*. Berkeley and Los Angeles: University of California Press/Courier Corporation, 1955. Citado na página 33.
- BRACEWELL, R. N. *The Fourier transform and its applications*. [S.l.]: McGraw-Hill New York, 1986. v. 31999. Citado na página 33.
- BRAITHWAITE, A.; JOHNSON, S. D. Space–time modeling of insurgency and counterinsurgency in iraq. *Journal of Quantitative Criminology*, Springer, v. 28, n. 1, p. 31–48, 2012. Citado na página 17.
- BRASSAROTE, G. d. O. N. Análise multiescala de séries temporais do efeito da cintilação ionosférica nos sinais de satélite gps a partir de wavelets não decimadas. Universidade Estadual Paulista (UNESP), 2014. Citado na página 86.

- BRENT, R. P. An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, Oxford University Press, v. 14, n. 4, p. 422–425, 1971. Citado na página 71.
- BROWN, R. G.; HWANG, P. Y. Introduction to random signals and applied kalman filtering: with matlab exercises and solutions. *Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions*, by Brown, Robert Grover.; Hwang, Patrick YC New York: Wiley, c1997., 1997. Citado na página 34.
- BROYDEN, C. G. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, Oxford University Press, v. 6, n. 1, p. 76–90, 1970. Citado na página 71.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, SIAM, v. 16, n. 5, p. 1190–1208, 1995. Citado na página 71.
- CESARE, L. D.; MYERS, D.; POSA, D. Estimating and modeling space–time correlation structures. *Statistics & Probability Letters*, Elsevier, v. 51, n. 1, p. 9–14, 2001. Citado 3 vezes nas páginas 23, 37 e 40.
- CESARE, L. D.; MYERS, D.; POSA, D. Fortran programs for space-time modeling. *Computers & Geosciences*, Elsevier, v. 28, n. 2, p. 205–212, 2002. Citado na página 58.
- CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. StI: A seasonal-trend decomposition. *Journal of Official Statistics*, v. 6, n. 1, p. 3–73, 1990. Citado na página 68.
- CRESSIE, N. Statistics for spatial data. *Terra Nova*, Wiley Online Library, v. 4, n. 5, p. 613–617, 1992. Citado 4 vezes nas páginas 17, 24, 31 e 73.
- CRESSIE, N. *Statistics for spatial data*. [S.l.]: John Wiley & Sons, 2015. Citado 2 vezes nas páginas 38 e 42.
- CRESSIE, N.; HUANG, H.-C. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 94, n. 448, p. 1330–1339, 1999. Citado 6 vezes nas páginas 33, 40, 49, 50, 51 e 52.
- CRESSIE, N.; WIKLE, C. K. *Statistics for spatio-temporal data*. [S.l.]: John Wiley & Sons, 2015. Citado 5 vezes nas páginas 17, 19, 53, 55 e 62.
- CRIMP, S.; BAKAR, K. S.; KOKIC, P.; JIN, H.; NICHOLLS, N.; HOWDEN, M. Bayesian space–time model to analyse frost risk for agriculture in southeast australia. *International Journal of Climatology*, Wiley Online Library, v. 35, n. 8, p. 2092–2108, 2015. Citado na página 18.
- DIMITRAKOPOULOS, R.; LUO, X. Spatiotemporal modeling: covariances and ordinary kriging systems. *Geostatistics for the next century*. Kluwer, Dordrecht, p. 88–93, 1994. Citado 3 vezes nas páginas 41, 42 e 48.
- DONG, P.; GALATSANOS, N. P. Affine transformation resistant watermarking based on image normalization. In: IEEE. *Image Processing. 2002. Proceedings. 2002 International Conference on*. [S.l.], 2002. v. 3, p. 489–492. Citado na página 36.

- DYN, N.; GOODMAN, T.; MICCHELLI, C. A. Positive powers of certain conditionally negative definite matrices. In: ELSEVIER. *Indagationes Mathematicae (Proceedings)*. [S.l.], 1986. v. 89, n. 2, p. 163–178. Citado na página 37.
- EGIDO, A.; DELAS, M.; GARCIA, M.; CAPARRINI, M. Non-space applications of gnss-r: From research to operational services. examples of water and land monitoring systems. In: IEEE. *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*. [S.l.], 2009. v. 2, p. II–170. Citado na página 86.
- EINSTEIN, A. On a stationary system with spherical symmetry consisting of many gravitating masses. *Annals of Mathematics*, JSTOR, p. 922–936, 1939. Citado na página 36.
- FÉLIX, V. B.; JÚNIOR, O. A. G.; ROSSONI, D. F.; HENRIQUES, M. J. Estimadores de semivariância: Uma revisão. *Ciência e Natura*, Universidade Federal de Santa Maria, v. 38, n. 3, 2016. Citado na página 26.
- FLETCHER, R. A new approach to variable metric algorithms. *The computer journal*, Oxford University Press, v. 13, n. 3, p. 317–322, 1970. Citado na página 71.
- FLETCHER, R.; REEVES, C. M. Function minimization by conjugate gradients. *The computer journal*, Oxford University Press, v. 7, n. 2, p. 149–154, 1964. Citado na página 71.
- FREITAS, S. R. de; CORDINI, J. Implicações geodinâmicas no posicionamento geodésico. *Boletim de Ciências Geodésicas*, v. 3, n. 1, 1998. Citado na página 87.
- GETHING, P. W.; ATKINSON, P. M.; NOOR, A. M.; GIKANDI, P. W.; HAY, S. I.; NIXON, M. S. A local space–time kriging approach applied to a national outpatient malaria data set. *Computers & geosciences*, Elsevier, v. 33, n. 10, p. 1337–1350, 2007. Citado na página 47.
- GNEITING, T. Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society*, Wiley Online Library, v. 125, n. 559, p. 2449–2464, 1999. Citado na página 17.
- GNEITING, T. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, Taylor & Francis, v. 97, n. 458, p. 590–600, 2002. Citado 2 vezes nas páginas 52 e 53.
- GNEITING, T.; GENTON, M. G.; GUTTORP, P. Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability*, CHAPMAN & HALL, v. 107, p. 151, 2006. Citado 7 vezes nas páginas 9, 20, 29, 30, 31, 36 e 121.
- GOLDFARB, D. A family of variable-metric methods derived by variational means. *Mathematics of computation*, v. 24, n. 109, p. 23–26, 1970. Citado na página 71.
- GOOVAERTS, P. *Geostatistics for natural resources evaluation*. [S.l.]: Oxford University Press on Demand, 1997. Citado na página 25.
- GRÄLER, B.; PEBESMA, E.; HEUVELINK, G. Spatio-temporal interpolation using gstat. *R Journal*, v. 8, n. 1, p. 204–218, 2016. Citado 2 vezes nas páginas 17 e 71.

- HANDCOCK, M. S.; WALLIS, J. R. An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association*, Taylor & Francis, v. 89, n. 426, p. 368–378, 1994. Citado na página 17.
- HASLETT, J.; RAFTERY, A. E. Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Applied Statistics*, JSTOR, p. 1–50, 1989. Citado 2 vezes nas páginas 17 e 68.
- HENGL, T. *A practical guide to geostatistical mapping*. [S.l.]: Hengl, 2009. v. 52. Citado 2 vezes nas páginas 9 e 19.
- HOFMANN, H.; WICKHAM, H.; KAFADAR, K. Value plots: Boxplots for large data. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 26, n. 3, p. 469–477, 2017. Citado na página 63.
- HYNDMAN, R. *tsfeatures*. [S.l.], 2019. Citado na página 68.
- HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International journal of forecasting*, Elsevier, v. 22, n. 4, p. 679–688, 2006. Citado na página 68.
- IACO, S. D.; MYERS, D. E.; POSA, D. Space-time analysis using a general product-sum model. *Statistics & Probability Letters*, Elsevier, v. 52, n. 1, p. 21–28, 2001. Citado 5 vezes nas páginas 44, 45, 46, 122 e 124.
- IACO, S. D.; MYERS, D. E.; POSA, D. Nonseparable space-time covariance models: some parametric families. *Mathematical Geology*, Springer, v. 34, n. 1, p. 23–42, 2002. Citado na página 34.
- JONES, R. H.; ZHANG, Y. Models for continuous stationary space-time processes. *Modelling longitudinal and spatially correlated data*, TG Gregoire et al., Springer, Berlin, v. 122, p. 289–298, 1997. Citado na página 50.
- JOST, L. Entropy and diversity. *Oikos*, Wiley Online Library, v. 113, n. 2, p. 363–375, 2006. Citado na página 68.
- JOURNEL, A. G.; HUIJBREGTS, C. J. *Mining geostatistics*. [S.l.]: Academic press, 1978. Citado 3 vezes nas páginas 25, 26 e 31.
- JR, P. J. R.; DIGGLE, P. J. et al. geor: a package for geostatistical analysis. *R news*, London, v. 1, n. 2, p. 14–18, 2001. Citado na página 58.
- KELLY, R.; DAVIS, J. Required navigation performance (rnp) for precision approach and landing with gnss application. *Navigation*, Wiley Online Library, v. 41, n. 1, p. 1–30, 1994. Citado na página 86.
- KRIGE, D. G. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, Southern African Institute of Mining and Metallurgy, v. 52, n. 6, p. 119–139, 1951. Citado na página 23.
- LEJOT, J.; DELACOURT, C.; PIÉGAY, H.; FOURNIER, T.; TRÉMÉLO, M.-L.; ALLEMAND, P. Very high spatial resolution imagery for channel bathymetry and topography from an unmanned mapping controlled platform. *Earth Surface Processes and Landforms*, Wiley Online Library, v. 32, n. 11, p. 1705–1725, 2007. Citado na página 17.

- LEMEY, P.; SUCHARD, M.; RAMBAUT, A. Reconstructing the initial global spread of a human influenza pandemic: a bayesian spatial-temporal model for the global spread of h1n1pdm. *PLoS currents*, Public Library of Science, v. 1, 2009. Citado na página 17.
- LU, N.; ZIMMERMAN, D. L. Testing for directional symmetry in spatial dependence using the periodogram. *Journal of Statistical Planning and Inference*, Elsevier, v. 129, n. 1, p. 369–385, 2005. Citado na página 35.
- MA, C. Spatio-temporal covariance functions generated by mixtures. *Mathematical geology*, Springer, v. 34, n. 8, p. 965–975, 2002. Citado na página 49.
- MA, C. Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference*, Elsevier, v. 116, n. 2, p. 489–501, 2003. Citado 2 vezes nas páginas 32 e 33.
- MARAI, J.; LEFEBVRE, S.; BERBINEAU, M. Satellite propagation path model along a railway track for gnss applications. In: IEEE. *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*. [S.I.], 2004. v. 6, p. 4066–4070. Citado na página 86.
- MASTERS, D. Surface remote sensing applications of gnss bistatic radar: Soil moisture and aircraft altimetry. *Boulder: PhD Dissertation, University of Colorado*, 2004. Citado na página 86.
- MATÉRN, B. Spatial variation: Meddelanden fran statens skogsforskningsinstitut. *Lecture Notes in Statistics*, v. 36, p. 21, 1960. Citado 3 vezes nas páginas 38, 42 e 52.
- MATHERON, G. *Traité de géostatistique appliquée. 1 (1962)*. [S.I.]: Editions Technip, 1962. v. 1. Citado na página 26.
- MATHERON, G. Theory of regionalized variables and its applications. *Cah. Centre Morphol. Math.*, v. 5, p. 211, 1971. Citado na página 23.
- MATTILA, P. *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*. [S.I.]: Cambridge university press, 1999. v. 44. Citado na página 34.
- MENG, F.; XU, P.; WONG, S.; HUANG, H.; LI, Y. Occupant-level injury severity analyses for taxis in hong kong: A bayesian space-time logistic model. *Accident Analysis & Prevention*, Elsevier, v. 108, p. 297–307, 2017. Citado na página 18.
- MITCHELL, M. W.; GENTON, M. G.; GUMPERTZ, M. L. Testing for separability of space-time covariances. *Environmetrics*, Wiley Online Library, v. 16, n. 8, p. 819–831, 2005. Citado na página 35.
- MONICO, J. F. G. *Posicionamento pelo Navstar-GPS*. [S.I.]: Unesp, 2000. Citado na página 86.
- MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. [S.I.]: Blucher, 2006. Citado 3 vezes nas páginas 31, 65 e 80.
- MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. [S.I.]: Blucher, 2006. Citado na página 68.
- MUNZERT, S.; RUBBA, C.; MEISSNER, P.; NYHUIS, D. *Automated data collection with R: A practical guide to web scraping and text mining*. [S.I.]: John Wiley & Sons, 2014. Citado na página 60.

- MYERS, D. E. Estimating and modeling space-time variograms. In: *Proceedings of the joint meeting of TIES-2004 and ACCURACY-2004*. [S.l.: s.n.], 2004. Citado 4 vezes nas páginas 20, 32, 37 e 40.
- MYERS, D. E.; JOURNEL, A. Variograms with zonal anisotropies and noninvertible kriging systems. *Mathematical Geology*, Springer, v. 22, n. 7, p. 779–785, 1990. Citado na página 37.
- NETWORK, R. et al. Analyzing spatio-temporal data with r: Everything you always wanted to know—but were afraid to ask. *Submitted to Journal de la Société Française de Statistique*, 2016. Citado 3 vezes nas páginas 60, 71 e 73.
- OEHLERT, G. W. Regional trends in sulfate wet deposition. *Journal of the American Statistical Association*, Taylor & Francis, v. 88, n. 422, p. 390–399, 1993. Citado na página 19.
- OLEA, R. A. *Geostatistics for engineers and earth scientists*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 26.
- PEBESMA, E. spacetime: Spatio-temporal data in r. *Journal of Statistical Software*, v. 51, n. 7, p. 1–30, 2012. Citado na página 78.
- PEBESMA, E.; BIVAND, R.; PEBESMA, M. E.; RCOLORBREWER, S.; COLLATE, A. Package sp. *The Comprehensive R Archive Network*, Citeseer, 2012. Citado na página 59.
- PEBESMA, E.; BIVAND, R. S. S classes and methods for spatial data: the sp package. *unpublished report*, 2005. Citado na página 62.
- PEBESMA, E. J. Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, Elsevier, v. 30, n. 7, p. 683–691, 2004. Citado 2 vezes nas páginas 58 e 71.
- PEBESMA, E. J.; WESSELING, C. G. Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, Elsevier, v. 24, n. 1, p. 17–31, 1998. Citado na página 58.
- PEREIRA, K. D.; APARECIDO, J.; FORTES, L. P. S. *RBMC: Sete anos fornecendo Referência a posicionamentos GPS no Brasil e Exterior*. 2004. Blog Mundo Geo. Disponível em: <<http://mundogeo.com/blog/2004/10/01/rbmc-sete-anos-fornecendo-referencia-a-posicionamentos-gps-no-brasil-e-exterior>>. Citado na página 87.
- PEREIRA, R. A. D.; FREITAS, S. R. C. D.; MANTOVANI, M. S. Fatores gravimétricos no brasil para um modelo anelástico de terra. *Boletim de Ciências Geodésicas*, Universidade Federal do Paraná, v. 15, n. 2, 2009. Citado na página 87.
- PILLEY, H. R. *Airport control/management system using GNSS-based methods and equipment for the control of surface and airborne traffic*. [S.l.]: Google Patents, 1996. US Patent 5,574,648. Citado na página 86.
- POLI, P.; PAILLEUX, J.; DUCROCQ, V.; MOLL, P.; RABIER, F.; MAUPRIVEZ, M.; DUFOUR, S.; GRONDIN, M.; LECHAT-CARVALHO, F.; LATOUR, A. D. et al. Weather report: meteorological applications of gnss from space and on the ground. *InsideGNSS*, v. 3, n. 8, p. 30–39, 2008. Citado na página 86.

- QUICK, H.; WALLER, L. A.; CASPER, M. A multivariate space–time model for analysing county level heart disease death rates by race and sex. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 67, n. 1, p. 291–304, 2018. Citado na página 18.
- QUINN, R. E.; ROHRBAUGH, J. A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management science*, INFORMS, v. 29, n. 3, p. 363–377, 1983. Citado na página 17.
- REINSEL, G.; TIAO, G.; WANG, M.; LEWIS, R.; NYCHKA, D. Statistical analysis of stratospheric ozone data for the detection of trends. *Atmospheric Environment (1967)*, Elsevier, v. 15, n. 9, p. 1569–1577, 1981. Citado na página 19.
- RENCHER, A. C. *Methods of multivariate analysis*. [S.l.]: John Wiley & Sons, 2003. v. 492. Citado na página 55.
- RIBEIRO, J.; DIGGLE, P. geos: A geostatistical library for s-plus. *Department of Mathematics and Statistics, Lancaster University, Lancaster*, 1999. Citado na página 58.
- ROCHA, R. S. M. da; JEREZ, G. O.; BRASSAROTE, G. d. O. N.; MONICO, J. F. G. Avaliação do efeito da cintilação ionosférica e de diferentes intervalos de tempo de coleta de dados no posicionamento por ponto preciso na sua forma on-line. *Revista Brasileira de Geomática*, v. 5, n. 2, p. 251–276, 2017. Citado na página 88.
- RODRIGUEZ-ITURBE, I.; MARANI, M.; D'ODORICO, P.; RINALDO, A. On space-time scaling of cumulated rainfall fields. *Water resources research*, Wiley Online Library, v. 34, n. 12, p. 3461–3469, 1998. Citado na página 17.
- ROSSONI, D. F. Análise de variância para experimentos com dependência espacial. UNIVERSIDADE FEDERAL DE LAVRAS, 2011. Citado 2 vezes nas páginas 9 e 27.
- ROSSONI, D. F. Testes bootstrap para detecção de anisotropia espacial em fenômenos contínuos. UNIVERSIDADE FEDERAL DE LAVRAS, 2014. Citado na página 23.
- SAHU, S. K.; BAKAR, K. S. Hierarchical bayesian autoregressive models for large space–time data with applications to ozone concentration modelling. *Applied Stochastic Models in Business and Industry*, Wiley Online Library, v. 28, n. 5, p. 395–415, 2012. Citado na página 17.
- SCACCIA, L.; MARTIN, R. Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning and Inference*, Elsevier, v. 131, n. 1, p. 19–39, 2005. Citado na página 35.
- SCARGLE, J. D. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, v. 263, p. 835–853, 1982. Citado na página 17.
- SCHEPERS, A. R.; SHANAHAN, J. F.; LIEBIG, M. A.; SCHEPERS, J. S.; JOHNSON, S. H.; LUCHIARI, A. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. *Agronomy Journal*, American Society of Agronomy, v. 96, n. 1, p. 195–203, 2004. Citado na página 17.
- SCHOENBERG, I. J. Metric spaces and completely monotone functions. *Annals of Mathematics*, JSTOR, p. 811–841, 1938. Citado na página 33.

- SHANNO, D. F. Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, v. 24, n. 111, p. 647–656, 1970. Citado na página 71.
- SOARES, A. Geoestatística para as ciências da terra e do ambiente. 2ª edição. *Instituto Superior Técnico, Lisboa*, 2006. Citado na página 24.
- SOUZA, E.; FÉLIX, V. B. Wavelet cross-correlation in bivariate time-series analysis. *TEMA (São Carlos)*, SciELO Brasil, v. 19, n. 3, p. 391–403, 2018. Citado na página 66.
- STEIN, M. A simple model for spatial-temporal processes. *Water Resources Research*, Wiley Online Library, v. 22, n. 13, p. 2107–2110, 1986. Citado na página 19.
- TIWARI, A. K.; DAR, A. B.; BHANJA, N.; SHAH, A. Stock market integration in asian countries: Evidence from wavelet multiple correlations. *Journal of Economic Integration*, JSTOR, p. 441–456, 2013. Citado na página 67.
- TOUTENBURG, H. *Prior information in linear models*. [S.l.]: John Wiley & Sons, 1982. Citado na página 39.
- WICKHAM, H. Meifly: Models explored interactively. *Website ASA Sections on Statistical Computing and Graphics (Student Paper Award Winner 2007)*, Citeseer, 2007. Citado na página 73.
- WICKHAM, H. *dplyr: A Grammar of Data Manipulation*. [S.l.], 2016. Citado na página 60.
- WICKHAM, H. *ggplot2: elegant graphics for data analysis*. [S.l.]: Springer, 2016. Citado na página 62.
- WICKHAM, H. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. [S.l.], 2016. Citado na página 60.
- WICKHAM, H. et al. Tidy data. *Journal of Statistical Software*, Foundation for Open Access Statistics, v. 59, n. 10, p. 1–23, 2014. Citado na página 60.
- WICKHAM, H.; STRYJEWSKI, L. 40 years of boxplots. *Am. Statistician*, 2011. Citado na página 62.
- WIGLEY, T. M.; BRIFFA, K. R.; JONES, P. D. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. *Journal of climate and Applied Meteorology*, v. 23, n. 2, p. 201–213, 1984. Citado na página 17.
- WIJS, H. D. Statistics of ore distribution. part i: frequency distribution of assay values. *Journal of the Royal Netherlands Geological and Mining Society*, v. 13, p. 365–375, 1951. Citado na página 23.
- WIJS, H. de. Statistics of ore distribution. part ii: Theory of binomial distribution applied to sampling and engineering problems. *Journal of the Royal Netherlands Geological and Mining Society*, v. 15, p. 125–24, 1953. Citado na página 23.
- YAMAMOTO, J. K.; LANDIM, P. M. B. *Geoestatística: conceitos e aplicações*. [S.l.]: Oficina de Textos, 2014. Citado 4 vezes nas páginas 23, 24, 25 e 42.
- YANOSKY, J. D.; PACIOREK, C. J.; SCHWARTZ, J.; LADEN, F.; PUETT, R.; SUH, H. H. Spatio-temporal modeling of chronic pm 10 exposure for the nurses health study. *Atmospheric Environment*, Elsevier, v. 42, n. 18, p. 4047–4062, 2008. Citado na página 17.

ZIMMERMAN, D. L.; ZIMMERMAN, M. B. A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, Taylor & Francis Group, v. 33, n. 1, p. 77–91, 1991. Citado na página 38.

Apêndices

APÊNDICE A

WMCC

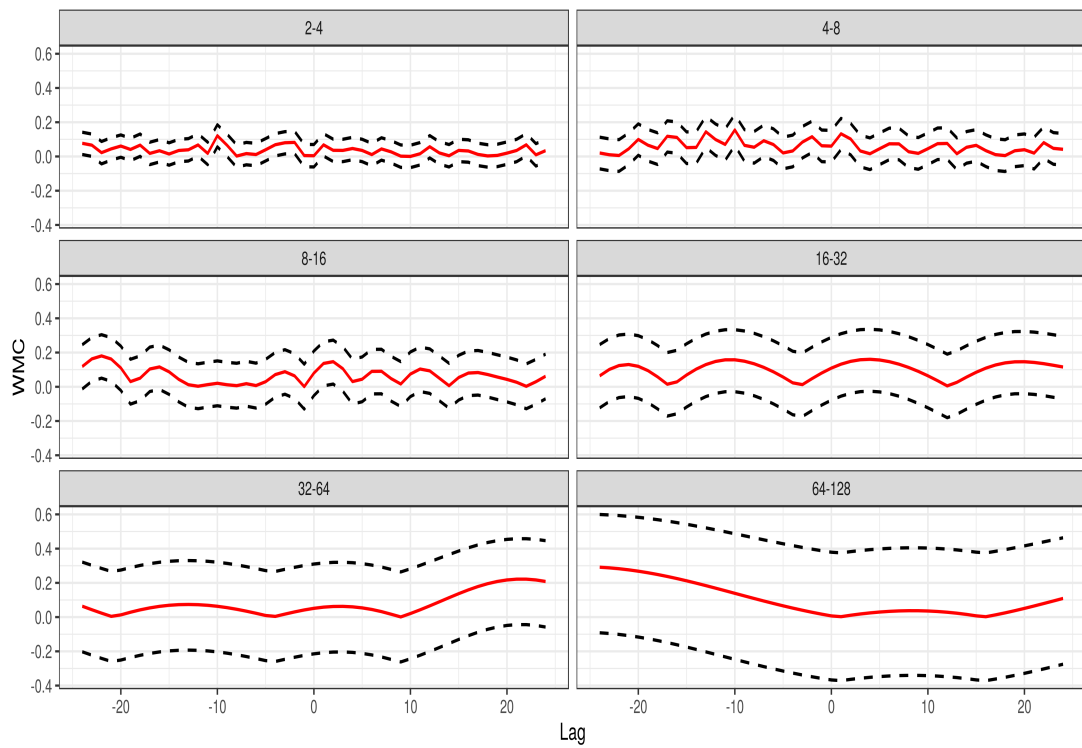


Figura 43 – Exemplo de WMCC aplicado em 5 estações de GNSS.

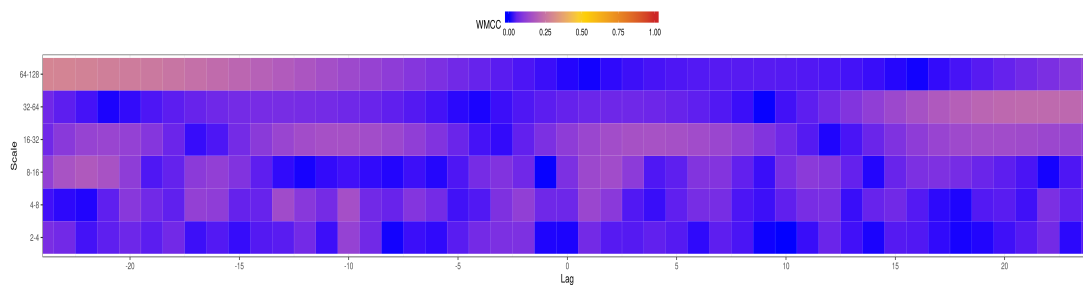


Figura 44 – Exemplo de WMCC aplicado em 5 estações de GNSS.

APÊNDICE B

SISTEMA LOCAL DE COORDENADAS

Como as estações GNSS provêm sinais no tempo e espaço, estes dados estão em um sistema de coordenadas, a fim de analisar a altimetria e ainda prover uma maior precisão, utiliza-se então um sistema local de coordenadas, que será posteriormente modelado tanto no tempo quanto no espaço.

Devido a necessidade de uma sistema local de coordenadas, as estações utilizam-se de um sistema local, no caso o ENU (do inglês, *East-North-Up*), como se vê representado na Figura 45.

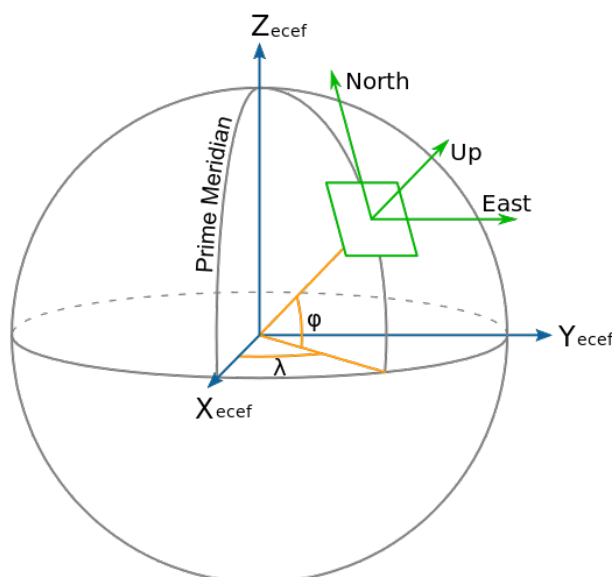


Figura 45 – Representação do sistema de coordenadas ENU. (Fonte: Wikipedia)

Na Figura 45 observa-se que o sistema ENU dá-se por 3 coordenadas respectivas, *East* (Leste), *North* (Norte) e *Up* (Vertical), em determinado espaço. As coordenadas Norte e Leste compõem a planimetria, ou seja, representam o espaço, enquanto a coordenada vertical representa a elevação em relação à superfície.

Para uma melhor exemplificação deste sistema, pode-se mostrar a relação deste com o sistema ECEF (do inglês, *Earth Centred Earth Fixed*), como se vê na Figura 46, que nada mais é que as coordenadas em (x, y, z) fixas no centro da Terra, como se vê na Figura 46, bem como em 45.

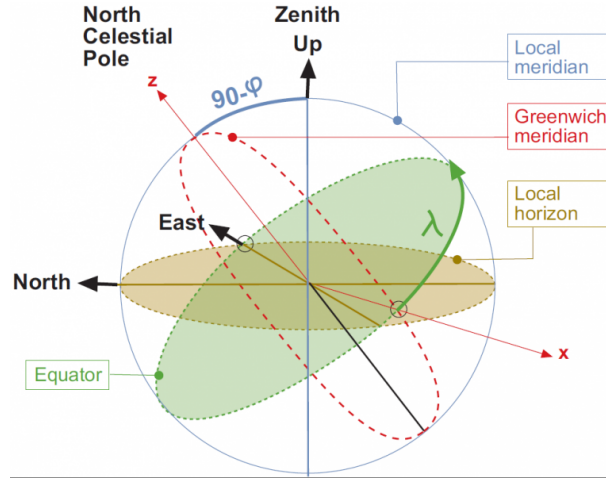


Figura 46 – Representação do sistema de coordenadas ECEF e ENU. (Fonte: Navipedia)

Pela Figura 46 vê-se que a transformação pode ocorrer de duas formas:

1. No sentido horário sobre o eixo leste em um ângulo de $90 - \varphi$ para alinhar o eixo vertical com o eixo z , ou seja, $R_1[-(\pi/2 - \varphi)]$;
2. No sentido horário sobre o eixo z em um ângulo de $90 + \lambda$ para alinhar o eixo leste com o eixo x , ou seja, $R_3[-(\pi/2 + \lambda)]$.

Em que,

$$R_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix}, \quad (B.1)$$

e

$$R_3 = \begin{bmatrix} \cos \theta & \sin \theta & 1 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (B.2)$$

Portanto, pode-se escrever a relação entre os sistemas de tal forma que

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R_3[-(\pi/2 + \lambda)]R_1[-(\pi/2 - \varphi)] \begin{bmatrix} E \\ N \\ U \end{bmatrix}. \quad (B.3)$$

Ou ainda, pode-se escrever ENU em função do sistema ECEF, na seguinte forma:

$$\begin{bmatrix} E \\ N \\ U \end{bmatrix} = R_1[-(\pi/2 - \varphi)]R_3[-(\pi/2 + \lambda)] \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (\text{B.4})$$

Assim, pode-se obter a coordenada vertical (up), a fim de se trabalhar com a altimetria, como objetivado.

Anexos

ANEXO A

PROVA DO TEOREMA 3

Suponha que C_{st} é uma função de covariância espaço-temporal estacionária e totalmente simétrica. [Gneiting, Genton e Guttorp \(2006\)](#) mostram que expandindo o denominador da integral de (2) como um produto de $\cos(h'_s\omega) + i \sin(h'_s\omega)$ e $\cos(h'_t\tau) + i \sin(h'_t\tau)$. Como C_{st} é uma função real, pode-se reescrever (2) como

$$C_{st}(h_s; h_t) = \int \int [\cos(h'_s\omega) \cos(h'_t\tau) - \sin(h'_s\omega) \sin(h'_t\tau)] dF(\omega; \tau), \quad (h_s; h_t) \in \mathbb{R}^{d+1}. \quad (\text{A.1})$$

Analogamente,

$$C_{st}(h_s; -h_t) = \int \int [\cos(h'_s\omega) \cos(h'_t\tau) - \sin(h'_s\omega) \sin(h'_t\tau)] dF(\omega; \tau), \quad (h_s; h_t) \in \mathbb{R}^{d+1}. \quad (\text{A.2})$$

Como a simetria total implica que $C_{st}(h_s; h_t) = C_{st}(h_s; -h_t)$, ou seja, (A.1) e (A.2) satisfazem a representação (2.24). Assim, qualquer função C_{st} na forma (2.24) é totalmente simétrica e admite a representação original de [Bochner \(2.15\)](#).

ANEXO B

PROVA DO TEOREMA 4

A fim de provar o Teorema 4, Iaco, Myers e Posa (2001) mostram que se C_{st} é expressada por um modelo *product-sum* de covariâncias puramente temporais e espaciais, com os coeficientes $k_1 > 0$, $k_2 \geq 0$ e $k_3 \geq 0$, tem-se que para k expresso em (2.71), segue que

$$k_s = k_2 + k_1 C_t(0) > 0, \quad (\text{B.1})$$

$$k_t = k_3 + k_1 C_s(0) > 0, \quad (\text{B.2})$$

ou seja, $k > 0$. Assim, as seguintes inequações são satisfeitas pela suposição:

$$k_1 > 0 \Rightarrow C_{st}(0; 0) < k_s C_s(0) + k_t C_t(0), \quad (\text{B.3})$$

$$k_2 \geq 0 \Rightarrow C_{st}(0; 0) \geq k_t C_t(0), \quad (\text{B.4})$$

e

$$k_3 \geq 0 \Rightarrow C_{st}(0; 0) \geq k_s C_s(0). \quad (\text{B.5})$$

Ainda, as inequações (B.4) e (B.5) são simultaneamente satisfeitas quando

$$C_{st} \geq \max\{k_s C_s(0), k_t C_t(0)\}. \quad (\text{B.6})$$

Já para garantir (B.3) e (B.6), tem-se que

$$k \leq \frac{1}{\max\{k_s C_s(0); k_t C_t(0)\}}. \quad (\text{B.7})$$

Logo,

$$k \in \left(0; \frac{1}{\max\{k_s C_s(0); k_t C_t(0)\}}\right]. \quad (\text{B.8})$$

As fronteiras de k e as inequações em (B.3) e (B.6) implicam que $k_1 > 0$, $k_2 \geq 0$ e $k_3 \geq 0$.

ANEXO C

PROVA DO TEOREMA 5

A fim de provar o Teorema 5, Iaco, Myers e Posa (2001) partem da continuidade de C_{st} a partir de (2.66):

$$\lim_{h_s \rightarrow \infty} \lim_{h_t \rightarrow \infty} C_{st}(0; 0) = 0. \quad (\text{C.1})$$

Então, devido à expressão (2.11) segue que

$$\lim_{h_s \rightarrow \infty} \lim_{h_t \rightarrow \infty} \gamma_{st}(h_s; h_t) = C_{st}(0; 0). \quad (\text{C.2})$$

Utilizando da relação

$$C_{st}(h_s; 0) = (k_1 C_t(0) + k_2) C_s(h_s) + k_3 C_t(0), \quad (\text{C.3})$$

que é implícita ao modelo *product sum*, ainda segue que

$$\lim_{h_s \rightarrow \infty} C_{st}(h_s; 0) = k_3 C_t(0). \quad (\text{C.4})$$

Portanto, aplicando (2.63), pode-se provar (2.67)

$$\lim_{h_s \rightarrow \infty} \gamma_{st}(h_s; 0) = C_{st}(0; 0) - \lim_{h_s \rightarrow \infty} C_{st}(h_s; 0) = k_s C_s(0). \quad (\text{C.5})$$

De forma similar, para provar (2.68), parte-se de que

$$C_{st}(0; h_t) = (k_1 C_s(0) + k_3) C_t(h_t) + k_2 C_s(0),$$

e de (2.63), tem-se então que

$$\lim_{h_t \rightarrow \infty} \gamma_{st}(0; h_t) = C_{st}(0; 0) - \lim_{h_t \rightarrow \infty} C_{st}(0; h_t) = k_t C_t(0).$$