

**MODELAGEM
COMPUTACIONAL
DE CONHECIMENTO**

Dissertação de Mestrado

Reconstrução de Sinais em Redes de Sensores sem Fios com Técnicas de Geoestatística

Bruno Lopes Vieira
blopesvieira@gmail.com

Orientador:
Alejandro C. Frery

Maceió, abril de 2010

Bruno Lopes Vieira

Reconstrução de Sinais em Redes de Sensores sem Fios com Técnicas de Geoestatística

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Curso de Mestrado em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Alejandro C. Frery

Maceió, abril de 2010

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária Responsável: Helena Cristina Pimentel do Vale

V658r Vieira, Bruno Lopes.

Reconstrução de sinais em redes de sensores sem fios com técnicas de Geoestatística / Bruno Lopes Vieira, 2010.

xii, 80 f. : il.

Orientador: Alejandro C. Frery.

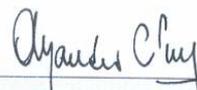
Dissertação (mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2010.

Bibliografia: f. 72-80.

1. Geoestatística. 2. Redes de computação. 3. Sensores sem fio. 4. Reconstrução de sinais. 5. Algoritmo de agrupamento. I. Título.

CDU: 004.738.5

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional de Conhecimento pelo Programa Multidisciplinar de Pós-Graduação em Modelagem Computacional de Conhecimento, da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina:



Prof. Dr. Alejandro Cesar Frery

UFAL – Instituto de Computação

Orientador



Profa. Dra. Eliana Silva de Almeida

UFAL – Instituto de Computação

Examinadora



Prof. Dr. Renato José de Sobral Cintra

UFPE – Departamento de Estatística

Examinador

Maceió, maio de 2010.

Resumo

As Redes de Sensores sem Fios (RSsF) são conjuntos de dispositivos que obtêm amostras de fenômenos ambientais, sejam eles naturais (como, por exemplo, temperatura, pressão atmosférica, intensidade de iluminação, concentração de substâncias em cursos d'água) ou antrópicos (qualidade do ar em sinais de trânsito, pressão ao longo de um oleoduto). Esses dispositivos têm despertado muito interesse, tanto pelas suas potenciais aplicações quanto pelos desafios teóricos e tecnológicos que seu uso otimizado oferece. O objetivo deste trabalho trata da análise da reconstrução de sinais nessas redes, com base em técnicas de geostatística. Analisam-se três processos de *kriging*: simples, ordinário e bayesiano. Ao simples, analisam-se três abordagens encontradas na literatura para estimação ou informação do parâmetro da média e ao bayesiano propõe-se uma variante capaz de reduzir o tempo de processamento necessário, estimando a média por mínimos quadrados generalizados, sendo uma constante na inferência bayesiana. Leva-se em consideração o processo de agrupamento dos nós sensores, com simulações sem agrupamento e com os sensores agrupados pelos algoritmos *LEACH* e *SKATER*. O algoritmo de *kriging* bayesiano apresenta os melhores resultados qualitativos na maioria dos casos, mas se torna inviável para sistemas que necessitem de respostas rápidas. Nesses casos, recomenda-se o algoritmo de *kriging* ordinário. A variante proposta para o *kriging* bayesiano reduz o tempo de computação, mas não o suficiente para sistemas de tempo real.

Abstract

Wireless sensor networks are formed by mobile devices that collect and process *data* from an environment, and transmit them to a data center which is responsible for taking decisions. This work aims to analyze the signal reconstruction in these networks using geostatistic techniques. Three processes of kriging are used: simple, ordinary and Bayesian. Three approaches to simple kriging were found in the literature, according to the way the mean of the *data* is estimated, were assessed them all. A new Bayesian approach is proposed: use general least square to estimate the mean, and set it as a constant into the Bayesian inference. The effect of clustering techniques is assessed, namely without clusters and with clusters formed by *LEACH* and *SKATER* algorithms. Bayesian kriging presents the best qualitative results in almost all scenarios, but it is not available to systems that require fast answers; in this case we recommend ordinary kriging. The proposed variant of Bayesian kriging reduces the time required, without hampering the quality of the reconstructed signal, but the time reduction is not enough for real-time systems.

Agradecimentos

Inicialmente a meus pais, que, diante de todas as dificuldades, puderam me proporcionar esta vivência acadêmica.

Agradeço ao Prof. Alejandro, amigo e eterno orientador, por todo o apoio nesta etapa.

À Profa. Eliana, que sempre se fez presente em minha vida acadêmica, tanto como professora quanto como amiga.

Aos colegas e amigos do LaCCAN, em especial a Bruno, David, Emanuelle, Evellyn, Ivan, Paulo, Rian, Tamer e o Prof. Leonardo.

Agrado à recepção que me foi dada no grupo *Sensornets* da UFMG e de Belo Horizonte, em especial a Guilherme, Heitor, Karina, Leandro e Letícia e ao Prof. Loureiro, pela acolhida em seu laboratório.

A Baltazar pelo apoio no uso do *cluster*.

A Alyson, Lorena e Marcelo, pelo tempo que passamos juntos e pelos trabalhos que pudemos desenvolver.

Aos amigos que sempre estiveram ao meu lado: Herlí, Lorena e Valdiran.

A Henrique, por todo o apoio e companheirismo nesta etapa.

Ao CNPq, pelo financiamento do projeto.

Agradecemos a ANP, a Petrobras e ao Laboratório de Computação Científica e Visualização da Universidade Federal de Alagoas por garantir acesso aos recursos computacionais do *cluster* GradeBR/UFAL da Rede Galileu.

Todos os que me ajudaram, direta ou indiretamente, meus mais sinceros agradecimentos.

Bruno Lopes

*“Quem, de três milênios,
Não é capaz de se dar conta
Vive na ignorância, na sombra,
À mercê dos dias, do tempo.”*

Johann Wolfgang von Goethe

Conteúdo

I	Introdução	1
II	Geoestatística, <i>Kriging</i> e as Redes de Sensores sem Fios	4
2.1	Definições iniciais	4
2.2	Modelos de correlação	8
2.2.1	Modelo de correlação Matérn	8
2.2.2	Modelo de correlação exponencial	9
2.2.3	Modelo de correlação esférico	10
2.2.4	Modelo de correlação <i>wave</i>	11
2.3	Variograma	12
2.3.1	Variograma teórico	12
2.3.2	Variograma empírico	13
2.4	<i>Kriging</i>	13
2.4.1	<i>Kriging</i> simples e ordinário	15
2.4.2	<i>Kriging</i> bayesiano	18
2.5	Redes de Sensores sem Fios	20
2.5.1	Agrupamento pelo algoritmo <i>LEACH</i>	22
2.5.2	Agrupamento pelo algoritmo <i>SKATER</i>	23
III	Modelagem proposta	27
3.1	Modelo formulado	27
3.1.1	Sinal de origem amostrado	27
3.1.2	Processos pontuais e a distribuição dos sensores	28
3.1.3	Reconstrução do sinal	35
3.1.4	Validação do modelo	37
3.2	Implementação de simulações Monte Carlo na literatura	37
3.3	Critérios para pesquisa reproduzível	38
3.4	Implementação na plataforma R	40
3.5	Ambiente de execução	41
3.6	Resultados esperados	44
IV	Análise dos resultados	45
4.1	Dados obtidos com a simulação	45
4.2	Análise da qualidade do sinal reconstruído	46
4.2.1	Análise do comportamento dos parâmetros do modelo	46
4.2.2	Análise do campo gaussiano reconstruído	49
4.3	Análise do tempo necessário à reconstrução do sinal	53
4.4	Considerações acerca dos dados obtidos	54

V Resultados e discussões	70
----------------------------------	-----------

Referências bibliográficas	72
-----------------------------------	-----------

Lista de Figuras

2.1	Exemplos de curvas da correlação Matérn	9
2.2	Exemplos de curvas da correlação exponencial	10
2.3	Exemplos de curvas da correlação esférica	11
2.4	Exemplos de curvas da correlação <i>wave</i>	12
2.5	Construção de um variograma teórico	14
2.6	Exemplo de reconstrução por <i>kriging</i>	18
2.7	Elementos fundamentais de uma Rede de Sensores sem Fios	20
2.8	Modelo de uma Rede de Sensores sem Fios	21
2.9	<i>Clusters</i> de sensores	21
2.10	Grupos gerados pelo algoritmo <i>LEACH</i>	23
2.11	Sensores espalhados sobre uma superfície	24
2.12	Grafo construído a partir de 12 sensores	25
2.13	Árvore geradora mínima	25
2.14	<i>Clusters</i> formados pelo algoritmo <i>SKATER</i>	26
3.1	Exemplos de campos aleatórios gaussianos	28
3.2	Exemplos de processo pontual Poisson	30
3.3	Exemplos de processo pontual Poisson não Homogêneo	31
3.4	Perspectiva da função de intensidade	32
3.5	Exemplos de processo pontual <i>SSI</i>	33
3.6	Exemplos do processo composto <i>C</i>	34
3.7	Exemplos do processo M^2P^2	36
3.8	Diagrama ilustrativo da reconstrução dos campos gaussianos	37
3.9	Diagrama ilustrativo do sistema de simulações	42
4.1	Erro quadrático médio do estimador de ϕ pelo <i>kriging</i> simples	47
4.2	Erro quadrático médio do estimador de ϕ pelo <i>kriging</i> ordinário	48
4.3	Erro quadrático médio do estimador de ϕ pelo <i>kriging</i> bayesiano	49
4.4	Erro de reconstrução por <i>kriging</i> simples para $\phi = 1$	50
4.5	Erro de reconstrução por <i>kriging</i> simples para $\phi = 5$	51
4.6	Erro de reconstrução por <i>kriging</i> ordinário para $\phi = 1$	52
4.7	Erro de reconstrução por <i>kriging</i> ordinário para $\phi = 5$	53
4.8	Erro de reconstrução por <i>kriging</i> bayesiano para $\phi = 1$	54
4.9	Erro de reconstrução por <i>kriging</i> bayesiano para $\phi = 5$	55
4.10	Processo de reconstrução por <i>kriging</i>	57
4.11	Processo de reconstrução por <i>kriging</i> com <i>LEACH</i>	58
4.12	Processo de reconstrução por <i>kriging</i> com <i>SKATER</i>	59
4.13	Erro de reconstrução por parâmetros inferidos para $\phi = 1$	60
4.14	Erro de reconstrução por parâmetros inferidos para $\phi = 5$	61
4.15	Processo de reconstrução pelos parâmetros estimados	62

4.16 Erro de reconstrução por <i>kriging</i> simples pelos parâmetros considerando a média para $\phi = 1$	64
4.17 Erro de reconstrução por <i>kriging</i> simples pelos parâmetros considerando a média para $\phi = 5$	65
4.18 Erro de reconstrução por <i>kriging</i> ordinário pelos parâmetros considerando a média para $\phi = 1$	66
4.19 Erro de reconstrução por <i>kriging</i> ordinário pelos parâmetros considerando a média para $\phi = 5$	67
4.20 Erro de reconstrução por <i>kriging</i> bayesiano pelos parâmetros considerando a média para $\phi = 1$	68
4.21 Erro de reconstrução por <i>kriging</i> bayesiano pelos parâmetros considerando a média para $\phi = 5$	69

Lista de Tabelas

3.1	Plataformas de planilhas eletrônicas avaliadas	39
3.2	Geradores de números pseudo-aleatórios de planilhas eletrônicas	40
4.1	Viés e EQM dos estimadores do parâmetro ϕ	46
4.2	Média dos erros absolutos relativos do <i>kriging</i>	56
4.3	Média dos erros absolutos relativos por reconstrução	60
4.4	Média dos erros absolutos relativos por reconstrução com média	63

Lista de Códigos

2.1	Algoritmo de <i>kriging</i>	17
3.1	Algoritmo para paralelização das simulações	42
4.1	Reconstrução do campo gaussiano original com base na estimativa	51



Introdução

“You need to prepare your students for the future, not the past.”

Spencer Graves

R-help (novembro de 2006)

A necessidade de se monitorar áreas nas quais não é possível a supervisão humana em toda sua extensão tem se tornado cada vez maior. Isso se dá, por exemplo, no monitoramento de incêndios em florestas, onde as áreas são suficientemente vastas de forma a inviabilizar a disposição de pessoas suficientes para cobrir toda a área. Na medição de níveis de concentração de gases tóxicos, uma vez que há risco à vida de pessoas que estejam presentes neste ambiente (bem como em diversos outros ambientes insalubres). Quando a escala não permite intervenção humana, como no monitoramento da válvula mitral do coração ou de tumores, utilizando nanosensores. Quando a presença de um agente monitorador humano pode interferir no resultado, como se dá no estudo da vocalização de anfíbios, que podem alterar seu comportamento em decorrência da presença do pesquisador (ser artificialmente inserido no ambiente, incapaz de passar despercebido por estes animais).

Uma das mais convenientes formas de se obter a automação desses e diversos outros tipos de monitoramentos é o uso de Redes de Sensores sem Fios (RSsF, [Akyildiz et al., 2002](#); [Szewczyk et al., 2004](#); [Sarangapani, 2007](#); [Sohraby et al., 2007](#); [Wu & Tseng, 2007](#)). Cada RSsF consistem num conjunto de nós sensores independentes, munidos de dispositivos de sensoramento e da capacidade de se comunicar, transmitindo os dados, tipicamente com algum tipo de agregação, a um nó especial que, por sua vez, os transmitirá a uma central de processamento.

O estudo das RSsF é uma vasta área de pesquisa, com diversos problemas de interesse da comunidade científica mundial ainda em aberto. Dentre estes, destacam-se questões como a minimização do consumo de energia pelos sensores, uma vez que normalmente

não há fonte de alimentação além das baterias das quais são munidos, de forma a garantir a maior durabilidade da RSsF (Akyildiz et al., 2002). Entretanto, essa preocupação com a minimização no consumo de energia não pode interferir demasiadamente na qualidade dos dados obtidos. Uma vez que uma RSsF é implantada com um propósito específico, este deve ser cumprido.

Essas RSsF obterão amostras do fenômeno que se deseja monitorar. Sendo possível possível distribuir os sensores no espaço de forma a se obter amostras de intervalos regulares pode-se utilizar a metodologia proposta por Shannon (1949) e obter todos os valores do fenômeno de origem. Mas nos cenários em que as RSsF são comumente utilizadas não é possível distribuir sensores de forma regular. Essas redes são normalmente utilizadas para monitorar vastas áreas (e.g. florestas) ou ambientes em que não se possui total controle da localização dos sensores (e.g. furacões, ambientes marinhos móveis, dentro da corrente sanguínea) e estes sensores são dispostos através de aviões que sobrevoam a região lançando-os em cenários descritos por Alencar-Neto (2008).

O problema tratado neste trabalho é: seja um fenômeno de interesse, representado por uma função que denota a informação de interesse neste fenômeno (dito sinal), obtém-se uma quantidade finita de amostras que são processadas através de alguma técnica de reconstrução para se obter todos os valores do sinal, numa tentativa de obter valores os mais próximos o possível dos originais.

Seguindo conexões apontadas por autores, como Frery et al. (2010); Alencar-Neto (2008); Aquino et al. (2008); Camilli et al. (2007); Frery et al. (2008); Jedermann & Lang (2009); Ni et al. (2009) e Umer et al. (2008), este trabalho visa avaliar a relação entre as RSsF e a geoestatística no que concerne à reconstrução de sinais.

A geoestatística (Bivand & Pebesma, 2008; Diggle & Ribeiro Jr., 2007; Le & Zidek, 2006) é um ramo da estatística espacial em que os dados estão relacionados ao seu posicionamento geográfico. Adota-se, então, a hipótese de que o sinal é estacionário em relação à covariância dos dados e, a partir desse pressuposto, avalia-se a sua reconstrução por alguns processos de *kriging* (Krige, 1951; Diggle & Ribeiro Jr., 2007), com uma de suas versões modificada.

Este trabalho encontra-se estruturado na forma que segue.

Capítulo 2: Apresenta as principais definições necessárias ao entendimento do processo, bem como as noções algorítmicas de *kriging* e conceituações sobre as RSsF. Além da introdução das técnicas de reconstrução encontradas na literatura é definida uma variante proposta neste trabalho. A qualidade dos dados resultantes dessas técnicas de reconstrução serão avaliadas através de um ensaio Monte Carlo de força bruta (Metropolis & Ulam, 1949), analisando-se viés, variância e erro quadrático médio apresentados (definidos no capítulo) bem como a viabilidade por tempo computacional necessário.

Capítulo 3: Neste capítulo apresentam-se os cenários utilizados para amostragem nas si-

mulações, disposição dos sensores e os detalhes técnicos de implementação. São apresentados resultados da análise da precisão numérica de ferramentas utilizadas para ensaios Monte Carlo e especifica-se o ferramental adotado.

Capítulo 4: Expõe a análise dos dados obtidos nas simulações efetuadas e os resultados obtidos. Os resultados nas métricas de qualidade são apresentados em tabelas numéricas e graficamente, através do erro quadrático médio e de uma medida de erro neste capítulo definida para o sinal resultante.

Capítulo 5: As principais conclusões obtidas através da análise efetuada no capítulo anterior são apresentadas, juntamente com alguns trabalho a serem desenvolvidos no futuro.

De uma forma específica, as etapas desenvolvidas neste trabalho são: (i) especificar um ferramental capaz de reconstruir sinais amostrados por RSsF; (ii) simular sinais que representem diferentes cenários; (iii) simular a disposição de sensores sobre cada um destes cenários na maior diversidade possível de técnicas de disposição; (iv) aplicar algoritmos de tipicamente utilizados em RSsF para economia de energia; (v) reconstruir o sinal utilizando as técnicas inicialmente especificadas; (vi) analisar a qualidade dos dados obtidos.



Geoestatística, *Kriging* e as Redes de Sensores sem Fios

“What we have is nice, but we need something very different.”

Robert Gentleman

Statistical Computing 2003 (junho de 2003)

DENTRE as principais técnicas para reconstrução de sinais da geoestatística, destaca-se o *kriging* (Krige, 1951), originado dos trabalhos de Daniel G. Krige e Georges Matheron. Sua relação com as Redes de Sensores sem Fios tem sido discutida em diversos trabalhos, como os de Alencar-Neto (2008); Camilli et al. (2007); Frery et al. (2008); Jedermann & Lang (2009); Ni et al. (2009) e Umer et al. (2008).

Neste capítulo serão explanados os conceitos necessários ao entendimento do *kriging* bem como suas variações utilizadas nesse trabalho: simples, ordinário e bayesiano. Em seguida, apresentam-se as definições necessárias ao entendimento das Redes de Sensores sem Fios e os algoritmos de agrupamento (clusterização) utilizados nos experimentos.

2.1 Definições iniciais

Nesta seção são explanados conceitos básicos de probabilidades que fundamentam o trabalho que se segue, com base nos textos de James (2006) e Diggle & Ribeiro Jr. (2007).

Definição 1 (Experimento Aleatório). *Um experimento aleatório consta de uma observação repetível não controlada, com um conjunto finito de resultados observáveis.*

Definição 2 (Espaço Amostral). Dado um experimento aleatório qualquer, o espaço amostral dito conjunto Ω consiste no conjunto de todos os resultados que podem ser observados por este experimento.

Definição 3 (Evento). Seja um conjunto $\omega \subset \Omega$ que denota um evento de Ω . Se ω é um conjunto unitário, diz-se que ω é um evento simples. Se $\omega = \Omega$, tem-se que $\Pr(\Omega) = 1$ e o evento é dito certo; caso $\omega = \emptyset$, o evento é impossível, isto é, $\Pr(\emptyset) = 0$.

Definição 4 (Variável Aleatória Real). Uma variável aleatória real X é uma função $X: \Omega \rightarrow \mathbb{R}$ que mapeia os resultados de um experimento aleatório na reta, em que Ω representa o espaço amostral.

A partir deste ponto, sempre que for referida alguma variável aleatória, tratar-se-á de uma variável aleatória real.

Definição 5 (Ocorrência de uma Variável Aleatória). Seja $X: \Omega \rightarrow \mathbb{R}$ uma variável aleatória. Uma ocorrência desta variável aleatória será um elemento de \mathbb{R} mapeado de Ω que denota um resultado do experimento aleatório que X mapeia.

Uma variável aleatória será caracterizada pelo modo como seus valores se distribuem, ou seja, sua distribuição. Conhecer a distribuição de uma variável aleatória X é ser capaz de calcular a probabilidade dela estar num conjunto \mathcal{A} formado por um número arbitrário de uniões e intersecções.

Definição 6 (Função de Distribuição Acumulada). Uma função de distribuição acumulada é uma função $F(t) = \Pr(X \leq t)$, $t \in \mathbb{R}$ que caracteriza cada distribuição. Possui as seguintes propriedades:

1. é não decrescente, isto é, se $t_1 < t_2$, então $F(t_1) \leq F(t_2)$;
2. é contínua à direita, isto é, se $t_n \downarrow t$ (t_n como valor limitante para t) quando $n \rightarrow \infty$, então $F(t_n) \rightarrow F(t)$;
3. $\lim_{t \rightarrow -\infty} F(t) = 0$ e $\lim_{t \rightarrow \infty} F(t) = 1$.

Definição 7 (Variável Aleatória Discreta). Uma variável aleatória X é dita discreta se ela possui uma quantidade finita ou infinita enumerável de valores possíveis. É descrita por uma função de probabilidade $p(t_i) = \Pr(X = t_i)$, $i = 1, 2, \dots$, onde $\sum_{i=1}^n p(t_i) = 1$ e $0 \leq p(t_i) \leq 1$.

Definição 8 (Variável Aleatória Contínua). Seja X uma variável aleatória e $F(t)$ sua função de distribuição acumulada. A função de densidade de X é dada por $f(t) = \partial F(t) / \partial t$, que define uma variável aleatória contínua, se $f(t)$ existir. Suas propriedades são:

1. *é sempre positiva, ou seja $f(t) \geq 0$;*
2. *sua integral na reta vale 1, então $\int_{\mathbb{R}} f(t) dt = 1$.*

Definição 9 (Variáveis Aleatórias Distribuídas de forma Conjunta). *Num experimento aleatório em que o interesse está no comportamento conjunto de duas ou mais variáveis aleatórias, a distribuição conjunta dessas variáveis aleatórias será dada, no caso discreto, por $\Pr(X_1 = t_1 \text{ e } X_2 = t_2 \text{ e } \dots \text{ e } X_n = t_n) = p_{X_1 X_2 \dots X_n}(t_1, t_2, \dots, t_n)$, com o valor do somatório $\sum_{i=1}^n p_{X_1 X_2 \dots X_n}(t_1, t_2, \dots, t_{n_i}) = 1$. No caso contínuo, sejam as variáveis aleatórias (X_1, X_2, \dots, X_n) . Sua densidade é caracterizada por uma função $f_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n)$, onde:*

- *f_{X_1, X_2, \dots, X_n} é não negativa;*
- *$\int \dots \int f_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n = 1$.*

Neste trabalho, serão tratadas apenas de variáveis aleatórias contínuas.

Definição 10 (Esperança de uma Variável Aleatória). *Também conhecida por média ou valor esperado, a esperança de uma variável aleatória X é uma média ponderada onde os pesos são dados pelas probabilidades $\Pr(X = t)$. De uma forma genérica, é definida como $\mathbb{E}[X] = \int_{\mathbb{R}} t f(t) dt$. No caso contínuo, especificamente, $\mathbb{E}[X] = \int t f(t) dt$, se a integral existir.*

Definição 11 (Esperança de uma Função). *Seja X uma variável aleatória e $\psi(X): \mathbb{R} \rightarrow \mathbb{R}$ uma função qualquer. A esperança de $\psi(X)$ é dada por $\mathbb{E}[\psi(X)] = \int_{\mathbb{R}} \psi(t) f(t) dt$, se a integral existir.*

Definição 12 (Esperança do Produto de Transformações de Variáveis Aleatórias). *Sejam X_1, \dots, X_n variáveis aleatórias, a esperança de $\psi(X_1) \dots \psi(X_n)$, dita $\mathbb{E}[\psi(X_1) \dots \psi(X_n)]$, será a integral $\int \psi(t_1) \dots \psi(t_n) f(t_1) \dots f(t_n) dt_1 \dots dt_n$, se a integral existir.*

Definição 13 (Probabilidade Condicional). *Sejam X e Y variáveis aleatórias. A probabilidade condicional $\Pr(X \leq x | Y = y)$ é dada pela fração*

$$\frac{\Pr(X \leq x \text{ e } Y = y)}{\Pr(Y = y)}.$$

Definição 14 (Esperança Condicional). *A esperança condicional é definida como $\mathbb{E}[X | Y] = \int_{\mathbb{R}} x f(x | Y = y) dx$, onde, caso X e Y possuam distribuição conjunta,*

$$f(x | Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Definição 15 (Variância de uma Variável Aleatória). *Sendo X uma variável aleatória, sua variância é o valor $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$, se $\mathbb{E}[X^2]$ existir e a diferença estiver bem definida.*

Definição 16 (Covariância dentre Variáveis Aleatórias). Sendo X, Y variáveis aleatórias, é definida por $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, se as esperanças existirem.

Definição 17 (Correlação dentre Variáveis Aleatórias). A correlação dentre duas variáveis aleatórias X, Y é dada por $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$.

Definição 18 (Distribuição Gaussiana). Uma variável aleatória X segue uma distribuição gaussiana $N(\mu, \alpha)$ de média μ e variância σ^2 se possuir função de densidade

$$f(t | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(t - \mu)^2}{2\sigma^2}\right\}, \quad (2.1)$$

$(\mu, \sigma) \in \Theta$, onde $\Theta = \mathbb{R} \times \mathbb{R}_+$ é o espaço paramétrico, com $t \in \mathbb{R}$.

Definição 19 (Distribuição Gaussiana Multivariada). Dada uma coleção de variáveis aleatórias X_1, \dots, X_n , elas seguem uma distribuição gaussiana multivariada de média $\mu = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ e matriz de covariância G simétrica positiva se sua densidade conjunta for

$$f(t | \mu, G) = \frac{1}{(2\pi)^{n/2} |G|^{1/2}} \exp\left\{-\frac{1}{2}(t - \mu)G^{-1}(t - \mu)\right\}, \quad (2.2)$$

onde $|\cdot|$ representa o valor do determinante da matriz apresentada neste.

Definição 20 (Processo Estocástico). Um processo estocástico $\{S(x) : x \in \mathbb{R}^d\}$, $d \geq 1$ consta de uma coleção de variáveis aleatórias, onde $x \in \mathbb{R}^d$ representa a indexação de cada variável aleatória $S(\cdot)$.

Definição 21 (Processo Estocástico Gaussiano). Para que um processo estocástico $\{S(x) : x \in \mathbb{R}^d\}$ seja gaussiano, a distribuição conjunta de $S(x)$, $x \in \mathbb{R}^d$ deve ser gaussiana multivariada. Este processo é completamente especificado por suas funções de média e covariância, respectivamente:

$$\mu(x) = \mathbb{E}[S(x)],$$

$$\text{Cov}\{S(x_i), S(x_j)\} = \mathbb{E}[S(x_i)S(x_j)] - \mathbb{E}[S(x_i)]\mathbb{E}[S(x_j)].$$

Definição 22 (Processo Estocástico Estacionário em Covariância). Um processo estocástico diz-se estacionário em relação à covariância se sua média for constante igual a μ e a covariância depender apenas da diferença vetorial entre x_i e x_j : $\text{Cov}\{S(x_i), S(x_j)\} = \gamma(u)$, $u = x_i - x_j$, assim $\gamma(u) = \text{Cov}\{S(z), S(z + u)\}$; como $\text{Cov}\{S(z_i), S(z_j)\} = \text{Cov}\{S(z_i + h), S(z_j + h)\}$, pode-se utilizar qualquer valor para z dentro do domínio admissível.

Sempre que nesse trabalho se referenciar um processo estocástico, tratar-se-á de um processo estocástico estacionário em covariância. E quando a diferença vetorial dentre duas variáveis aleatórias é dada pela distância euclidiana, o processo estocástico é dito isotrópico. Assim, $\gamma(x_i, x_j) = \gamma(\|u\|)$, com $\|\cdot\|$ denotando distância euclidiana.

Definição 23 (Campos Aleatórios). Um campo gaussiano é definido por um processo estocástico $\{S(x): x \in \mathbb{R}^d\}$, onde d representa a dimensão do espaço. São comumente utilizados para representar fenômenos naturais, como iluminação, umidade, temperatura etc. Sua caracterização se dá pela distribuição conjunta de suas variáveis aleatórias.

Definição 24 (Estimador). Um estimador $\hat{\theta}$ é uma função de variáveis aleatórias $\hat{\theta}(x_1, \dots, x_n) \rightarrow \mathbb{R}$ tal que x_1, \dots, x_n são amostras de valores gerados por uma variável aleatória da qual se deseja obter um valor próximo ao real de um de seus parâmetros, que denota uma tentativa de propiciar essa aproximação.

Neste trabalho sempre que for referido um campo gaussiano tratar-se-á de um campo aleatório cuja distribuição conjunta de suas variáveis aleatórias é gaussiana multivariada.

2.2 Modelos de correlação

Um modelo de correlação é uma função paramétrica $\rho(u)$ definida para o domínio \mathbb{R}^+ . Seu objetivo é especificar o modo como se comporta a correlação dentre variáveis aleatórias de um mesmo processo estocástico. Se ela é válida para \mathbb{R}^d , será também para \mathbb{R}^m , $m < d$, mas não necessariamente para dimensões maiores.

Essas funções costumam decrescer quando $u = \|x_i - x_j\|$ aumenta em modelos estacionários e é interessante que possuam diferentes graus de suavização. Ou seja, que haja fatores que suavizem suas curvas. Seguem informações sobre alguns modelos de correlação, aplicados a um processo estocástico gaussiano estacionário isotrópico $\{S(x): x \in \mathbb{R}^2\}$.

2.2.1 Modelo de correlação Matérn

O modelo de correlação Matérn é um dos mais utilizados na literatura. Ele apresenta duas propriedades extremamente interessantes aos estudos de geostatística: depende apenas do posicionamento geográfico dos dados e possui um parâmetro κ de suavização da curva entre dois pontos (Diggle & Ribeiro Jr., 2007).

Sua representação é dada na forma

$$\rho(u) = \{2^{(\kappa-1)}\Gamma(\kappa)\}^{-1} \left(\frac{u}{\phi}\right)^\kappa K_\kappa\left(\frac{u}{\phi}\right), \quad (2.3)$$

onde $K_\kappa(\cdot)$ representa a função de Bessel de ordem κ , $\phi > 0$ é um parâmetro de escala, responsável por dimensionar a distância entre os valores, e $\kappa > 0$ é o parâmetro responsável pela curva de suavização dentre os valores, denomina-se ordem. Exemplos de curvas geradas pelo modelo de correlação Matérn estão na figura 2.1.

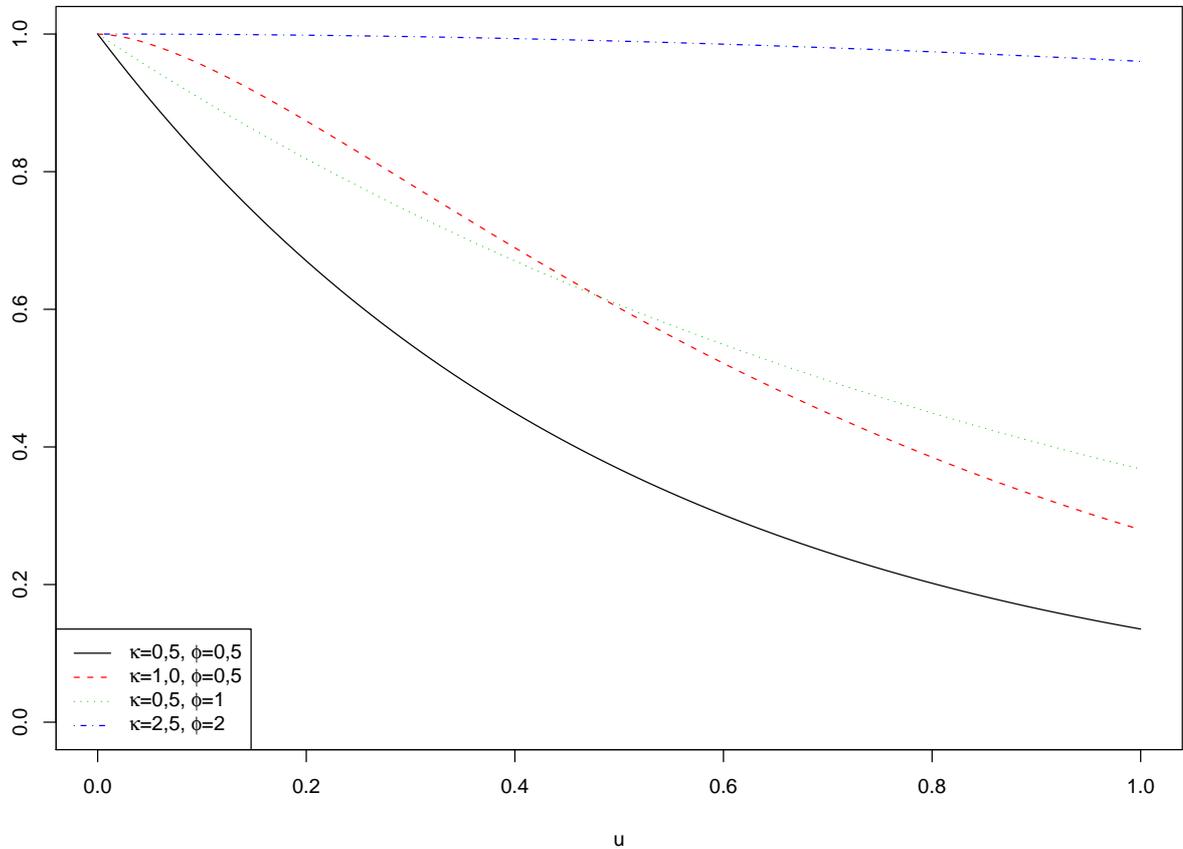


Figura 2.1: Exemplos de curvas da correlação Matérn

Para $\kappa = 0,5$, a correlação Matérn reduz-se a exponencial (seção 2.2.2), $\rho(u) = \exp\{-u/\phi\}$, e para $\kappa \rightarrow \infty$, $\rho(u) \rightarrow \exp\{-(u/\phi)^2\}$, também conhecida como função de correlação gaussiana (Handcock & Wallis, 1994).

Vale ressaltar que correlações de ordem ou escala diferentes não são comparáveis. A partir daí, sugere-se uma reparametrização da função para $\alpha = 2\phi\sqrt{\kappa}$.

2.2.2 Modelo de correlação exponencial

Assim como o modelo Matérn, o modelo exponencial apresenta um parâmetro κ de suavização e um ϕ de escala (Diggle & Ribeiro Jr., 2007). Entretanto, ϕ está limitado a $0 < \phi \leq 2$. Sua função é

$$\rho(u) = \exp\left\{-\frac{u\kappa}{\phi}\right\}. \quad (2.4)$$

Apesar do processo $S(x)$ ser contínuo, ou seja, suas variáveis aleatórias possuem uma distribuição contínua, este modelo possui limitações: ele não será diferenciável (não possui derivada) para todos os valores de ϕ . Entretanto ele será infinitamente diferenciável para $\phi = 2$, o que dificulta a obtenção de estimativas sob este modelo, uma vez que se utilizam

procedimentos de otimização que necessitam de diferenciações.

Esse caso extremo, contudo, costuma representar uma situação irreal, onde a observação de uma pequena quantidade de dados seria suficiente para avaliar todos os demais. Isso se equivale ao modelo extremo na correlação Matérn ($\kappa \rightarrow \infty$). Exemplos de curvas provenientes do modelo exponencial estão na figura 2.2.

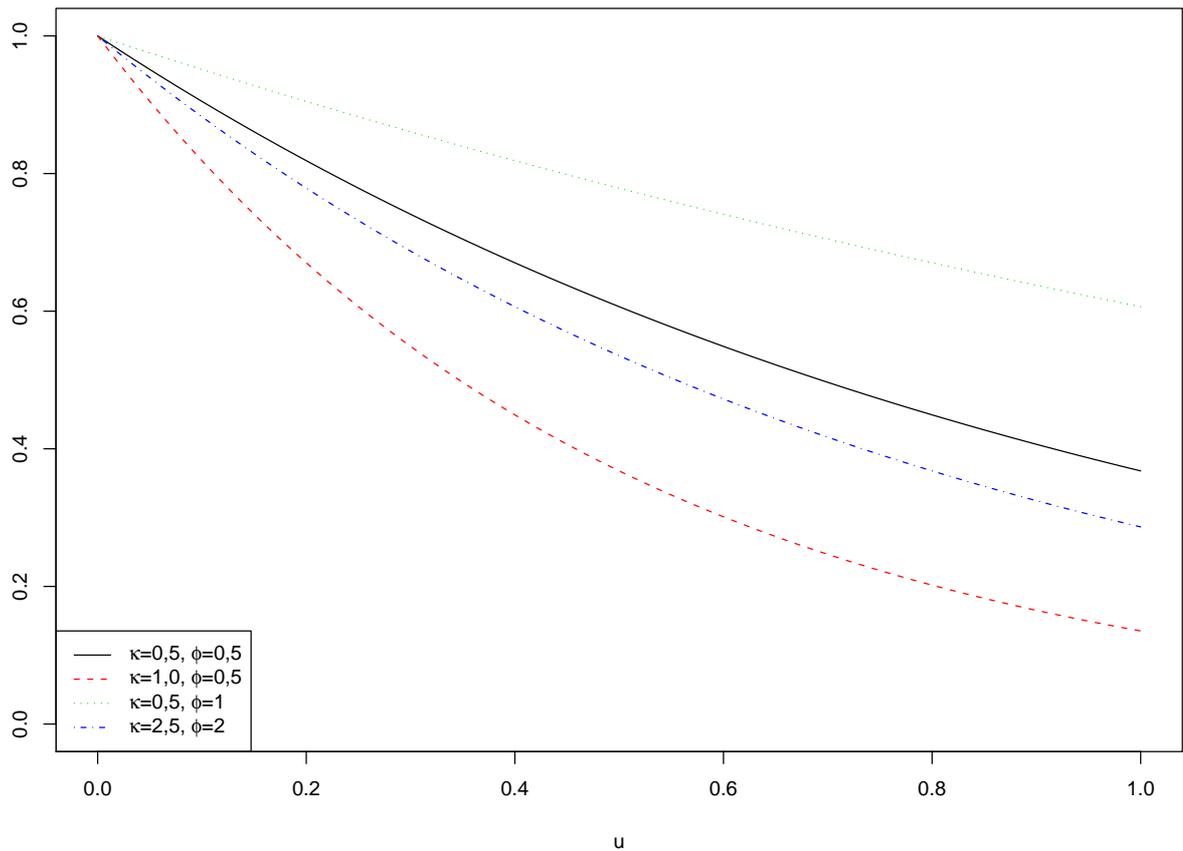


Figura 2.2: Exemplos de curvas da correlação exponencial

2.2.3 Modelo de correlação esférico

Diferentemente dos modelos anteriores, o modelo esférico não apresenta parâmetro de suavização. Seu único parâmetro é a escala de distância (ϕ), conforme a função

$$\rho(u) = \begin{cases} 1 - \frac{3}{2} \left(\frac{u}{\phi} \right) + \frac{1}{2} \left(\frac{u}{\phi} \right)^3, & \text{se } 0 \leq u \leq \phi \\ 0, & \text{se } u > \phi \end{cases} \quad (2.5)$$

Consoante não possuir parametrização para suavização da curva, seu modelo apresenta uma grande vantagem: um limitante finito de distância. Para valores suficientemente gran-

des de distância, leia-se $u > \phi$, o valor da correlação será $\rho(u) = 0$. Amostras de suas curvas estão expostos na figura 2.3.

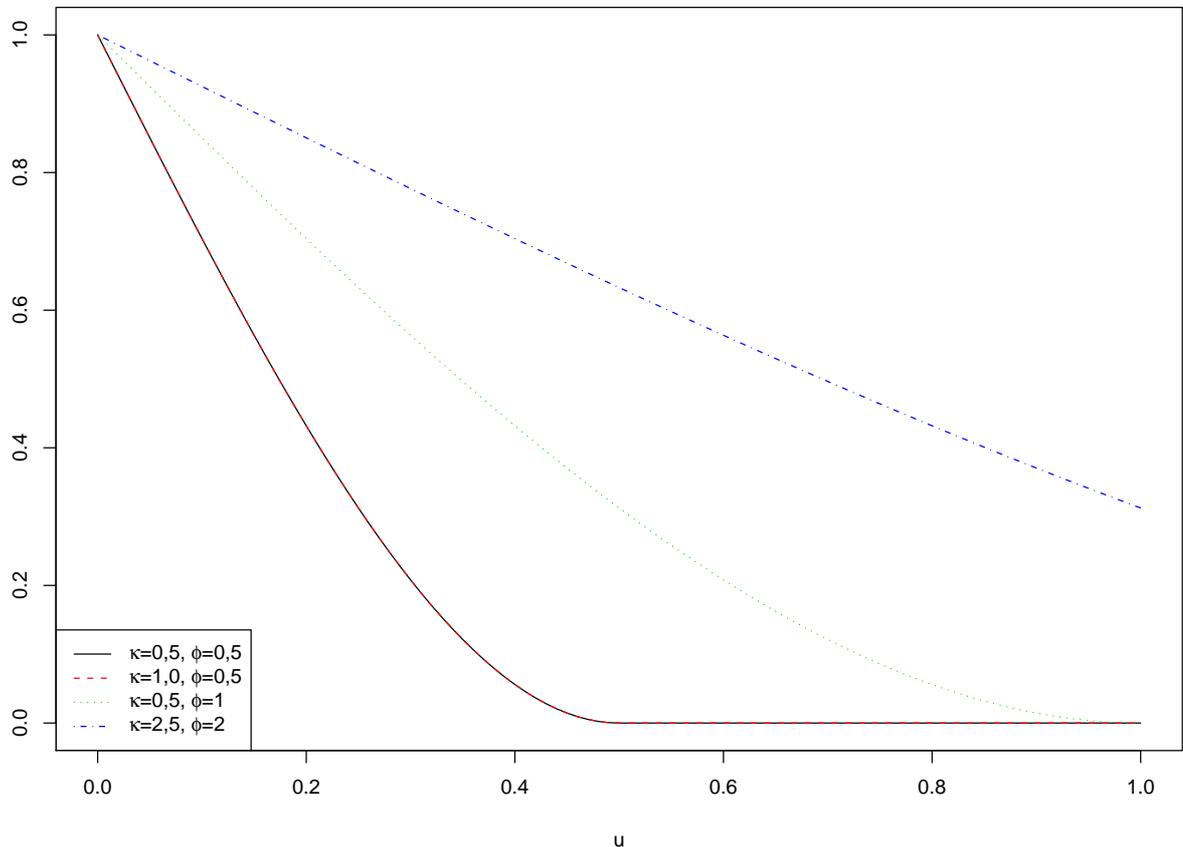


Figura 2.3: Exemplos de curvas da correlação esférica

Além de perder flexibilidade pela parametrização, um outro problema presente é o fato dessa função de correlação ser apenas uma vez diferenciável para $u = \phi$. Isto proporciona dificuldades na estimação por máxima verossimilhança.

2.2.4 Modelo de correlação *wave*

Modelos não monótonos, (sempre crescente ou sempre decrescente) de correlação são raros. Um exemplo é a função *Wave*,

$$\rho(u) = \left(\frac{u}{\phi}\right)^{-1} \sin\left(\frac{u}{\phi}\right). \quad (2.6)$$

Assim como o esférico, ele também é uniparametrizado pela escala (ϕ). Seu comportamento pode ser observado na figura 2.4.

Outros modelos de correlação podem também ser encontrados na literatura (ver, por

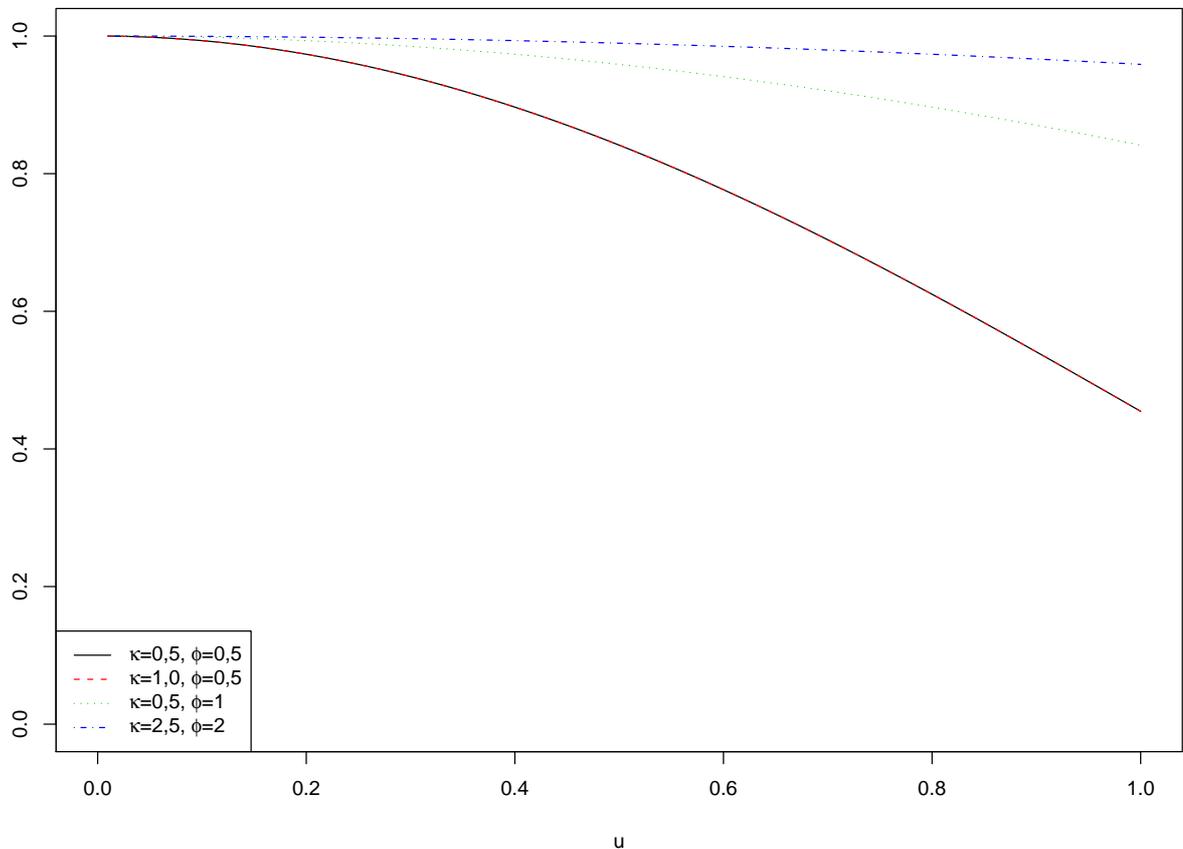


Figura 2.4: Exemplos de curvas da correlação *wave*

exemplo, [Gneiting, 1997](#); [Schlather, 1999](#)). Entretanto, esses apresentados são os mais utilizados em geoestatística.

2.3 Variograma

Também conhecido por semivariograma ou semivariâncias, o variograma é uma ferramenta muito importante à geoestatística, sendo definido pela distância dentre os valores amostrados em cada ponto. Ele descreve naturalmente estruturas de dependência num modelo gaussiano e pode ser útil como ferramenta de diagnóstico nos demais.

2.3.1 Variograma teórico

Num processo estocástico, o variograma teórico é dado pela função $V(x_i, x_j) = \frac{1}{2}\text{Var}\{S(x_i) - S(x_j)\}$. No caso de um processo estacionário (assim como exposto na definição 22), o variograma se reduz a uma função de $u = \|x_i - x_j\|$ e $\gamma(u) = \text{Cov}\{S(x), S(x - u)\}$ equivalente, expressa por $V_Y(u) = \gamma(0) - \gamma(u) = \sigma^2\{1 - \rho(u)\}$, onde σ^2 é a variância de $S(x)$ e

$\rho(u) = \text{Corr}\{S(x), S(x-u)\}$. Como a média é constante:

$$V_Y(u) = \frac{1}{2} \mathbb{E} [\{S(x) - S(x-u)\}^2]. \quad (2.7)$$

2.3.2 Variograma empírico

Dado um conjunto de variáveis aleatórias $Y_i: i = 1, \dots, n$ que modelam os dados de um fenômeno sendo medido e $y_i: i = 1, \dots, n$ ocorrências dessas variáveis, os valores $v_{ij} = \frac{1}{2}(Y_i - Y_j)^2$ correspondem a uma estimativa não viesada da semivariância, também denominada variograma ordinário.

O variograma teórico V_T será os pares de distâncias e variogramas ordinários correspondentes $(u_{ij}, v_{ij}): j > i$. Graficamente, ele corresponde a plotagem dos valores v_{ij} pelas distâncias correspondente (ver exemplo na figura 2.5b, construído a partir dos pontos amostrados no campo gaussiano da figura 2.5a).

2.4 Kriging

Dado um processo estocástico que modela um sinal ocorrido, porém não observado, $S(\cdot)$, pretende-se efetuar predição sobre ocorrências de uma variável aleatória $T = \mathcal{F}(S)$, onde T denota o conjunto de variáveis aleatórias que serão preditas, S representa o conjunto de todos os valores a partir de $S(x)$ e $\mathcal{F}(\cdot)$ é a função de predição.

Sejam amostras observadas $Y = (y_1, \dots, y_n)$, onde cada y_i representa uma amostra, com possível ruído, de seu correspondente $S(x)$, num processo estocástico gaussiano estacionário $\{S(x): x \in \mathbb{R}^2\}$ com x representando o posicionamento geográfico. S possui distribuição gaussiana multivariada de vetor de médias $\mu \mathbf{1}$, tal que $\mathbf{1}$ é um vetor em que todos os elementos são 1. Sua matriz de variância é $\sigma^2 R$, onde R é $n \times n$ tal que $r_{ij} = \rho(\|x_i - x_j\|)$.

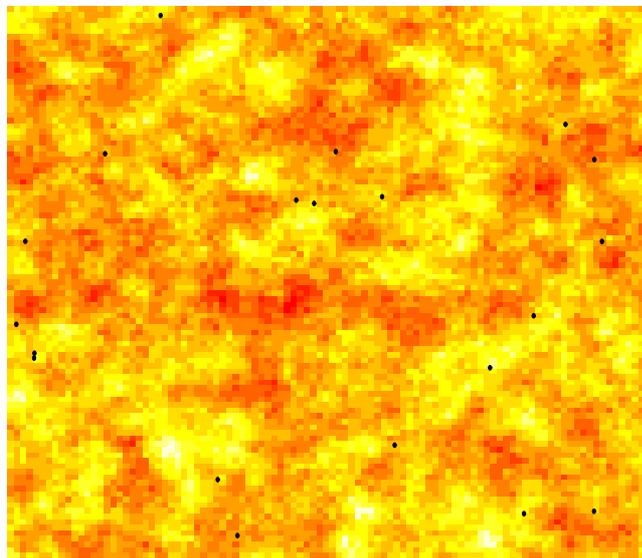
Da mesma forma Y possui distribuição gaussiana multivariada com vetor de médias $\mu \mathbf{1}$ e matriz de variância

$$\begin{aligned} \sigma^2 V &= \sigma^2(R + v^2 I) \\ \sigma^2 V &= \sigma^2 R + \tau^2 I, \end{aligned} \quad (2.8)$$

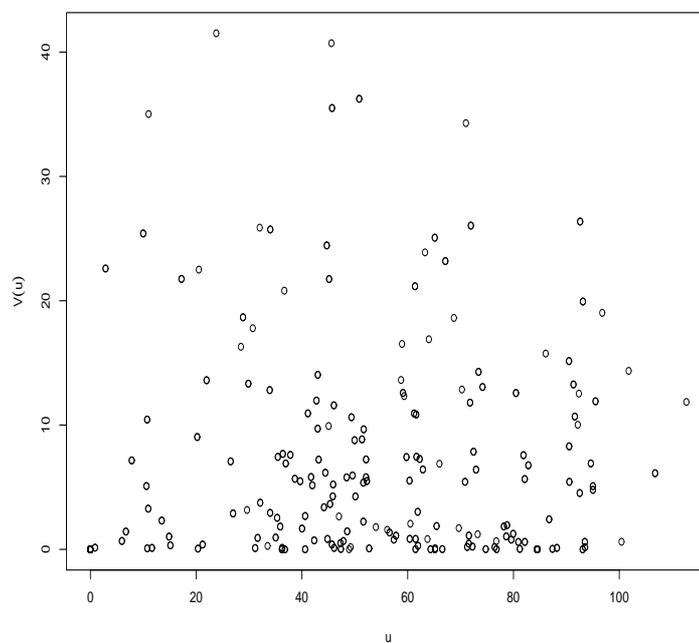
com I sendo a matriz identidade.

Seja um preditor \hat{T} uma função que estima valores ainda não ocorridos de uma variável aleatória, utiliza-se qualquer função t de Y tal que $\hat{T} = t(Y)$ o erro do preditor pelos mínimos quadrados será:

$$\text{MSE}(\hat{T}) = \mathbb{E} [(\hat{T} - T)^2]. \quad (2.9)$$



(a) Campo gaussiano com pontos em destaque amostrados



(b) Variograma empírico

Figura 2.5: Campo gaussiano com pontos amostrados em 2.5a, a partir dos quais se constrói o variograma empírico em 2.5b

A forma que minimiza esta quantidade é dada quando $\hat{T} = \mathbb{E}[T | Y]$. Daí, tem-se que:

$$\mathbb{E}[(T - \hat{T})^2] = \mathbb{E}_Y[\text{Var}(T | Y)]. \quad (2.10)$$

Neste trabalho, são analisadas algumas variantes do processo de *kriging* (Krige, 1951), que, de uma forma geral, consistem no resultado de

$$\hat{S}(x) = \int_W w_i S(x) dx, \quad (2.11)$$

onde $S(x)$ denota o conjunto de amostras, $W \subset \mathbb{R}^2$ é a janela de \mathbb{R}^2 que o processo estocástico modela e $w_i: i = 1, \dots, n$ um vetor de pesos, construídos a partir das funções de correlação, que suavizam a interpolação dentre os pontos (Diggle & Ribeiro Jr., 2007).

2.4.1 *Kriging* simples e ordinário

Dentre todos os modelos de *kriging*, o simples apresenta-se como o modelo matemático menos complexo. Nele se considera que médias de grupos dos valores do processo se aproximam da média global. Relaxando-se essa suposição, de forma a permitir que as médias locais possam se distanciar da global, o *kriging* ordinário usa uma estimação diferente para a média.

Ou seja:

- no *kriging* simples a média é informada e todos os demais parâmetros são tratados como conhecidos, entretanto, é necessário acoplar estimadores dos demais parâmetros ao preditor;
- no *kriging* ordinário, a média é tratada como parâmetro desconhecido, sendo substituída pelo estimador por mínimos quadrados

$$\hat{\mu} = (\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{V}^{-1}\mathbf{Y}. \quad (2.12)$$

Alguns autores, entretanto, adotam que no *kriging* simples a média não é informada, mas computada através dos mínimos quadrados ordinários dos valores amostrados, dito \bar{y} (Diggle & Ribeiro Jr., 2007, pp. 137). Entretanto, essa abordagem não foi encontrada em resultados de experimentos da literatura. Uma alternativa a esta abordagem comumente utilizada (Ribeiro Jr. & Diggle, 2001) é utilizar os mínimos quadrados generalizados, levando em consideração os valores de covariância dentre as variáveis indexadas pelas coordenadas amostradas.

O preditor de *kriging* para $\hat{T} = \hat{S}(x)$ é dado por

$$\hat{T} = \mu + r'V^{-1}(Y - \mu\mathbf{1}), \quad (2.13)$$

tal que r é o vetor tal que $r_i = \rho(\|x - x_i\|)$, $i = 1, \dots, n-1$ e V é dado por (2.8). A variável aleatória Y possui distribuição gaussiana multivariada de média $\mu \mathbf{1}$ e matriz de variância dada por (2.8).

Daí, segue que

$$\hat{S}(x) = \mu + \sum_{i=1}^n a_i(x)(Y_i - \mu) = \left[1 - \sum_{i=1}^n a_i(x)\right] \mu + \sum_{i=1}^n a_i(x) Y_i. \quad (2.14)$$

O preditor por *kriging* ordinário pode ser expresso como $\hat{S}(x) = \sum_{i=1}^n w_i(x) Y_i$, onde w é o vetor de pesos do *kriging* para o ponto x_i . Ou seja, dado um ponto não amostrado x_i o qual deseja-se reconstruir com base no conjunto de amostras Y , será computado um vetor de pesos w , com a soma de seus elementos $\sum_{i=1}^n w_i = 1$. Esse vetor de pesos w será único para cada i .

Se $V_Y(\cdot)$ é a função de semivariância (semivariograma) de um processo estocástico gaussiano estacionário $\{S(x) : x \in \mathbb{R}^2\}$, tal que x denota o posicionamento geográfico dos pontos, então o vetor de pesos para um ponto x_i será dado por

$$w_1 V_Y(u_{x_1}) + \dots + w_n V_Y(u_{x_n}) = V_Y(u_{x_i}), \quad (2.15)$$

no *kriging* simples.

Para o *kriging* ordinário, insere-se na equação uma variável β responsável por tornar os valores proporcionais (normalizá-los). Assim, a equação que representa o vetor de pesos w para o *kriging* ordinário é

$$w_1 V_Y(u_{x_1}) + \dots + w_n V_Y(u_{x_n}) + \beta = V_Y(u_{x_i}), \quad (2.16)$$

onde β é constante.

Em ambos os casos o vetor de pesos é definido pela resolução de um sistema de equações baseadas na semivariância. Daí nota-se o impacto do modelo de correlação adotado no *kriging*, que é parte da definição de semivariância e a construção dos pesos de forma a prover curvas de variação suaves, determinadas por seus fatores de suavização (κ).

Como os modelos de *kriging* simples e ordinário diferem apenas na estimação da média por este último, o *kriging* simples é também conhecido como *kriging* ordinário com média (Diggle & Ribeiro Jr., 2007).

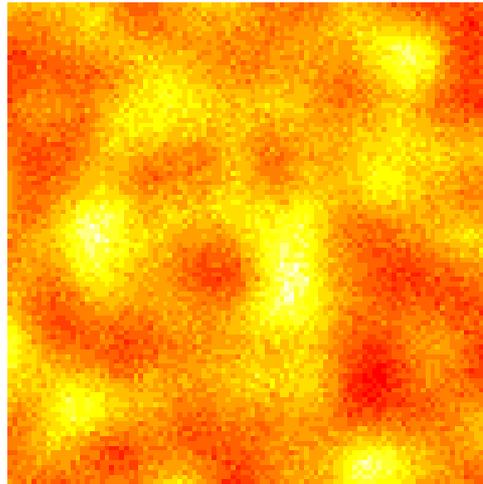
A escolha de um modelo que represente adequadamente os dados é crucial para que se obtenha um bom resultado. Da mesma forma, encontrar valores de κ que se aproximem do ideal proporcionará uma reconstrução o mais acurada o possível.

Uma descrição algorítmica do processo de *kriging* está no código 2.1. Como ilustração dos resultados que se podem obter através das duas técnicas a figura 2.6 apresenta um campo gaussiano em 2.6a com 20 pontos amostrados e reconstruídos por *kriging* simples e

ordinário, respectivamente 2.6b e 2.6c.

```
1 Entrada: dados amostrados
2 Saida: sinal reconstruido
3
4 S <- amostra()
5 Ponto <- S
6 Para cada ponto x a ser predito
7 {
8     w <- ComputaPesos()
9     T <- 0
10    Para cada ponto y amostrado
11    {
12        T <- T + (w[y]*S[y])
13    }
14    Ponto[x] <- T
15 }
```

Código 2.1: Algoritmo de *kriging*



(a) Campo aleatório original

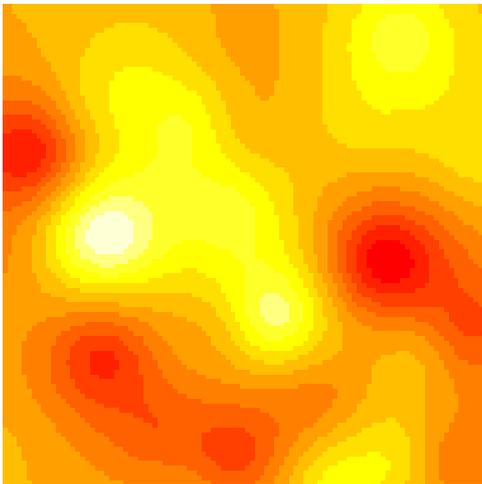
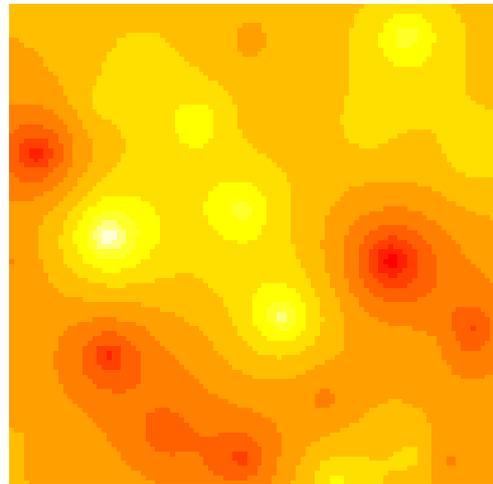
(b) Reconstrução por *kriging* simples(c) Reconstrução por *kriging* ordinário

Figura 2.6: Dada uma região de interesse 2.6a, amostram-se alguns pontos através de 20 sensores dispostos aleatoriamente e reconstrói-se a partir desses dados amostrados por *kriging* simples 2.6b e ordinário 2.6c

2.4.2 *Kriging* bayesiano

Diferentemente dos modelos anteriores de *kriging*, o bayesiano não faz nenhuma distinção formal entre os pontos não observados e os parâmetros do modelo a serem estimados. Conforme Diggle & Ribeiro Jr. (2007), ambos são tratados como variáveis aleatórias não observadas.

Seus principais requisitos são:

- conhecimento da distribuição *a priori* de $\theta \in \Theta$, onde Θ é o espaço paramétrico do processo estocástico modelado que, como nos demais, é gaussiano estacionário $\{S(x): x \in \mathbb{R}^2\}$;
- capacidade computacional para a especificação de um modelo com seus dados *a pri-*

ori a serem computados *a posteriori*.

Um dado *a priori* é uma informação parametrizada ao modelo. Por sua vez, um dado *a posteriori* é inferida no modelo através da informação *a priori*.

A capacidade computacional citada, muito superior à necessária para o *kriging* simples e ordinário (Gaudard et al., 1999), trata de resolver a seguinte equação:

$$\Pr(\theta | y) = \frac{\ell(\theta; y)\pi(\theta)}{\int \ell(\theta; y)\pi(\theta)d\theta}, \quad (2.17)$$

onde as funções utilizadas são a de verossimilhança, $\ell(\cdot; \cdot)$, e a distribuição *a priori*, $\pi(\cdot)$, dada pelo teorema de Bayes (2.18). O parâmetro y corresponde aos dados amostrados, ocorrências de $S(x)$.

O Teorema de Bayes diz que

$$\Pr(A | B) = \frac{\Pr(B | A)\Pr(A)}{\Pr(B)}. \quad (2.18)$$

Neste caso, $\Pr(A)$ e $\Pr(B)$ seriam as probabilidades *a priori* e $\Pr(A | B)$ e $\Pr(B | A)$ as probabilidades *a posteriori*.

No enfoque bayesiano, o modelo pode ser especificado como $[Y, \theta]$, onde Y representa as variáveis aleatórias que descrevem o fenômeno a ser reconstruído (o sinal), θ são todos os parâmetros do modelo a serem inferidos e $[\cdot]$ denota a distribuição das variáveis aleatórias que lhe são parametrizadas. Assim, com os valores *a priori* de θ pode-se calcular a $[Y, \theta] = [Y | \theta][\theta]$, onde a notação $|$ denota condicionalidade dentre as variáveis aleatórias. Pelo Teorema de Bayes, apresentado na equação (2.18), chega-se a

$$[\theta | Y] = \frac{[Y | \theta][\theta]}{[Y]}, \quad (2.19)$$

onde $[Y]$ pode ser obtido por

$$[Y] = \int [Y | \theta][\theta]d\theta, \quad (2.20)$$

induzido pela função de verossimilhança. As inferências acerca do parâmetro são feitas por computação de dados *a posteriori*, com $[\theta | Y]$.

Modelando o proposto neste trabalho com o enfoque bayesiano tem-se três elementos: um sinal não observado, S ; um conjunto de valores medidos, Y ; e os parâmetros do modelo, θ . A especificação do modelo, então, se dá por

$$[Y, S, \theta] = [\theta][S | \theta][Y | S, \theta]. \quad (2.21)$$

Os valores *a priori* de θ podem ser encarados como a opinião do analista dos dados. Sob a

aplicação do Teorema de Bayes, equação (2.18),

$$[S | Y] = \int [S | Y, \theta][\theta | Y] d\theta. \quad (2.22)$$

Algoritmicamente, iniciando pelo valor inicial arbitrado, serão computados os valores inferidos. Com estes calculados, os parâmetros do modelo são computados (inferência sobre θ) e o processo se repete até um limite de convergência preestabelecido.

Nesse trabalho serão avaliadas computações com o valor de média conhecido e inferido por média ponderada, levando em consideração valores inferidos de correlação.

2.5 Redes de Sensores sem Fios

As Redes de Sensores sem Fios são conjuntos de dispositivos que obtêm amostras de fenômenos ambientais, sejam eles naturais (como, por exemplo, temperatura, pressão atmosférica, intensidade de iluminação, concentração de substâncias em cursos d'água – ver figura 2.7) ou antrópicos (qualidade do ar em sinais de trânsito, pressão ao longo de um oleoduto). Uma vez com essas informações captadas, elas são transmitidas a um nó especial, denominado *sink* que tipicamente efetua algum tipo de agregação de dados, numa comunicação através de redes *ad-hoc* (Rappaport, 2002; Sarangapani, 2007; Sohraby et al., 2007; Wu & Tseng, 2007). Este se responsabiliza por transmitir as informações coletadas através de algum meio como a Internet para um centro de tratamento de dados e tomada de decisões, o qual pode efetuar consultas na rede. Seu funcionamento pode ser modelado conforme o diagrama da figura 2.8, modelado por Frery et al. (2010).

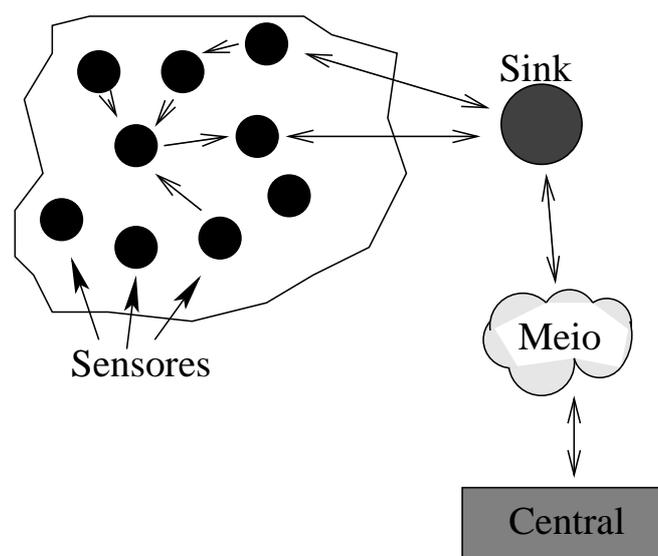


Figura 2.7: Elementos fundamentais de uma Rede de Sensores sem Fios

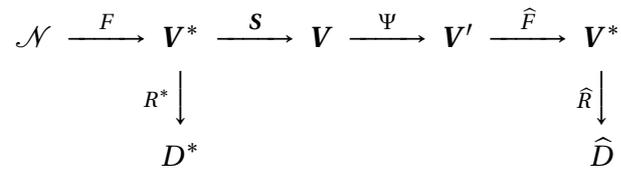


Figura 2.8: Modelo de uma Rede de Sensores sem Fios

No diagrama supracitado, \mathcal{N} representa a natureza, com toda sua infinidade de fatos, F é o fenômeno de interesse (o sinal, i.e.: iluminação, temperatura, umidade etc.), com todos os seus dados em \mathbf{V}^* . Se fosse possível observar F em toda a sua complexidade, construir-se-ia um conjunto de regras ideais R^* com as quais seria possível tomar decisões ideais D^* . Em vez disso, utiliza-se um conjunto de sensores $\mathbf{S} = \{S_1, \dots, S_n\}$ que disponibiliza um conjunto de valores \mathbf{V} , sob os quais efetuam-se transformações Ψ (i.e.: agregação, fusão de dados etc.) que geram um conjunto de dados \mathbf{V}' . Nesse último conjunto de dados, aplica-se uma função de reconstrução do sinal \hat{F} que permite idealmente reconstruir o conjunto de valores \mathbf{V}^* a partir do qual pode-se construir um conjunto de regras \hat{R} com as quais se pode tomar decisões \hat{D} . Este conjunto \hat{D} deve ser o mais próximo possível de D^* (idealmente iguais).

Tipicamente, aplica-se algum algoritmo para reduzir a quantidade de dados transmitidos (Nakamura et al., 2007). Dentre os principais motivos para que se efetue esta operação, está que a transmissão de dados comumente é a operação mais custosa em termos de consumo de energia (Akyildiz et al., 2002). Assim, minimizando essas transmissões pode-se prolongar o tempo de vida útil da rede, mas não se pode comprometer demasiadamente a qualidade dos dados.

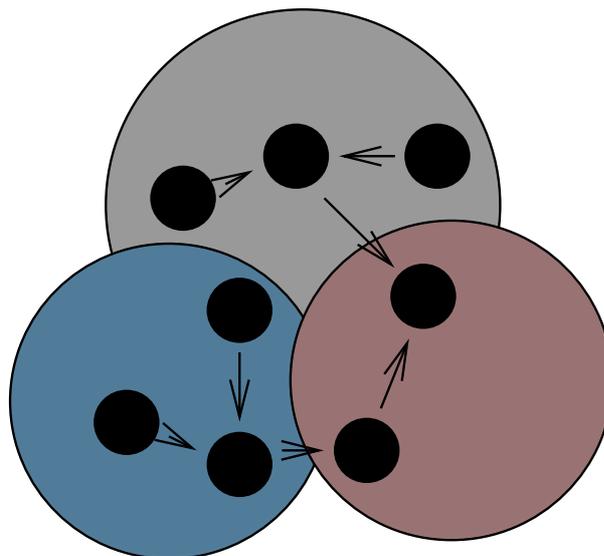


Figura 2.9: Clusters de sensores

Há diversos outros aspectos tecnológicos, como autoconfiguração, adaptação na ocor-

rência de falhas e limitações de energia, aguardando por soluções satisfatórias. Uma das formas de otimizar o consumo de energia é utilizar agrupamentos (*clusters*) de sensores. A técnica consiste em formar grupos de sensores que transmitem seus dados a um nó dentro do grupo, para que somente este transmita as informações ao destino. Os grupos são definidos e é eleito um dito “representante” (*clusterhead*) responsável por receber as informações dos demais sensores e transmiti-lo (ver figura 2.9).

Essa é a ideia dos algoritmos *LEACH* e *SKATER* (seções 2.5.1 e 2.5.2), dentre diversos outros (ver [Akyildiz et al., 2002](#); [Heinzelman et al., 2002](#); [Younis & Fahmy, 2004](#); [Assunção et al., 2006](#); [Younis et al., 2006](#); [Yoon & Shahabi, 2007](#); [Abbasi & Younis, 2007](#)). O objetivo é sempre o mesmo: analisar globalmente o estado da rede para formar grupos de sensores e estratégias de fusão de informações, diminuindo a redundância nos dados transmitidos e otimizando o consumo de energia. Dessa forma o consumo é minimizado nos sensores que apenas transmitem para este representante do grupo. Além disso, pode-se aplicar alguma transformação nesses dados no nível dos *clusters*, como por exemplo no caso de medição de temperatura transmitir a média de cada grupo, em vez da informação de cada sensor.

As principais referências que tratam deste assunto não empregam explicitamente a analogia existente entre esse problema e o problema de amostragem e reconstrução de sinais. [Alencar-Neto \(2008\)](#) e [Frery et al. \(2008\)](#) estabelecem formalmente essa analogia, e propõem medidas de erro na reconstrução do sinal empregando diferentes técnicas de amostragem e de reconstrução (ambas utilizando células de Voronoi).

A dissertação de mestrado de [Alencar-Neto \(2008\)](#) generaliza esses resultados, considerando (além da amostragem e da reconstrução por células de Voronoi) amostragem baseada em funções características dos sensores (dando, assim, maior realismo ao estudo) e *kriging* simples na reconstrução do sinal. Esta última mostrou produzir resultados significativamente melhores do que os gerados pela reconstrução por Voronoi na maioria dos casos.

2.5.1 Agrupamento pelo algoritmo *LEACH*

O algoritmo *LEACH* (*Low Energy Adaptive Clustering Hierarchy* – [Heinzelman et al., 2000, 2002](#)) propõe a divisão dos nós sensores numa quantidade pré-definida de *clusters*. Escolhem-se aleatoriamente os líderes de cada grupo (*clusterheads*) e os integrantes de cada grupo são definidos através de uma função que minimiza a distância entre os membros de um *cluster* e seu *clusterhead* (ver figura 2.10).

Esse processo ocorre numa etapa de configuração, prévia ao início do sensoriamento. Determina-se uma janela de tempo em que essa configuração permanece até que haja a reconfiguração da rede, onde os *clusterheads* se alternam, numa tentativa de maximizar o tempo de vida útil desses sensores, visto que eles promovem mais comunicação, gastando mais energia.

Neste protocolo, há uma série de fatores não especificados e ainda há a premissa de que

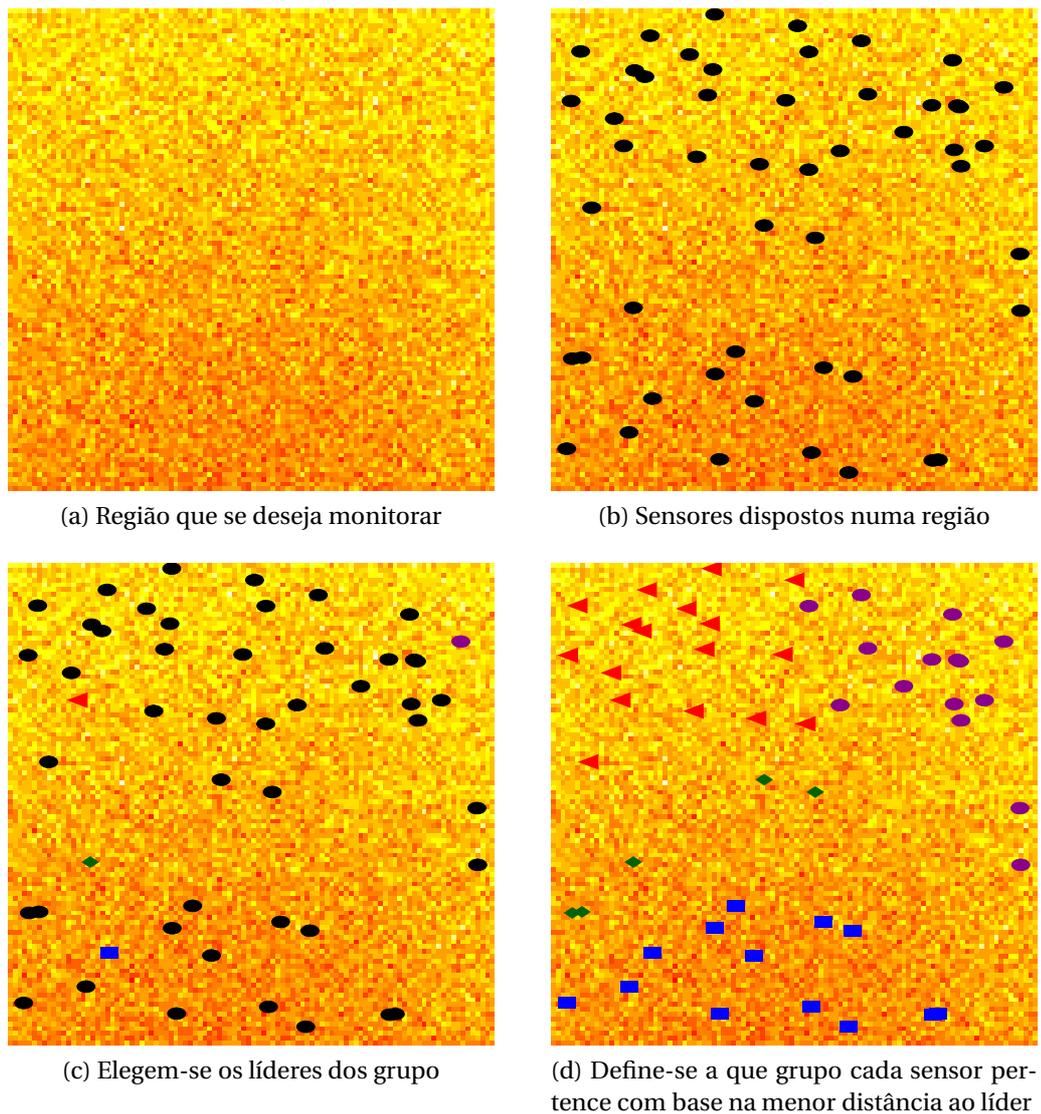


Figura 2.10: Grupos gerados pelo algoritmo *LEACH*

cada líder comunica-se diretamente com o *sink*, sem necessitar transmitir dados para outros sensores e que estes os repassam (comunicação *single-hop*). Entretanto, em diversos casos, esse escopo não retratará uma situação real. Dentre as diversas variações deste protocolo encontradas na literatura, muitas que desconsideram essa restrição.

2.5.2 Agrupamento pelo algoritmo *SKATER*

Uma outra alternativa para o processo de agrupamento de sensores é o protocolo *SKATER* (*Spatial 'K'cluster Analysis by Tree Edge Removal* – Assunção et al., 2006). A abordagem adotada é prover grupos com dados o mais correlacionados o possível.

Os sensores são organizados num grafo onde cada sensor determina um vértice. As arestas são definidas pelo raio de comunicação (ver figura 2.12, que apresenta o grafo correspondente aos sensores da figura 2.11).

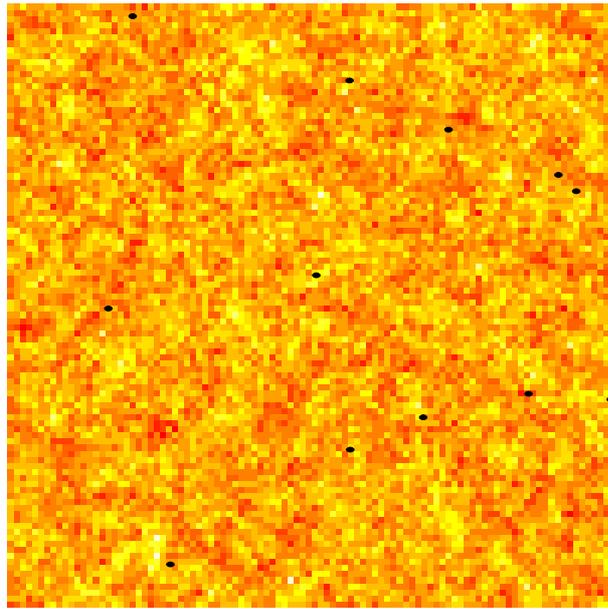


Figura 2.11: Sensores espalhados sobre uma superfície

Dado que um sensor consegue se comunicar com outro, atribui-se um peso a aresta que representa a distância entre os dados amostrados naquele instante pelos sensores. Considerando esses pesos calcula-se uma árvore geradora (subgrafo contendo todos os vértices do grafo original, entretanto apenas um caminho de um nó a outro, conforme a figura 2.13) de custo mínimo (a soma dos pesos das arestas para se percorrer de um vértice a outro possuirá o menor valor possível).

Dessa árvore subtraem-se as n arestas de maior peso, onde n representa a quantidade desejada de *clusters*. Os subgrafos resultantes conexos definem cada um dos grupos (ver figura 2.14). Note-se que os *clusterheads* não possuem fator determinante na formação dos *clusters*.

Ou seja, as etapas de construção dos grupos são as seguintes:

- constrói-se o grafo onde cada vértice representa um sensor e o peso das arestas denota a distância dentre os valores coletados por cada sensor;
- obtém-se a árvore geradora mínima do grafo construído;
- removem-se as $n - 1$ arestas de maior peso, onde n é a quantidade de grupos desejadas e os grupos serão os subgrafos conexos obtidos.

Assim, os grupos deverão conter os sensores de dados amostrados mais correlacionados. Bem como no protocolo *LEACH*, há etapas de configuração e reconfiguração, para minimizar o desgaste energético nos *clusterheads*.

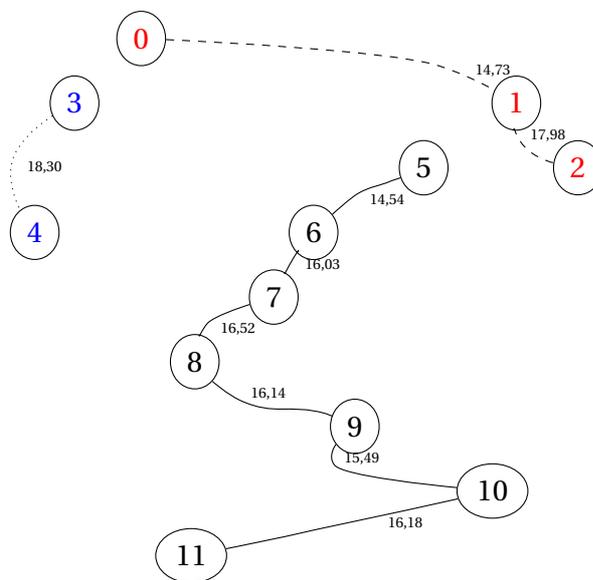
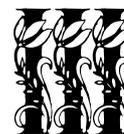


Figura 2.14: Divisão dos sensores da figura 2.12 em 3 *clusters* pelo algoritmo *SKATER*



Modelagem proposta

“R may be the wrong tool for the job, but it’s the wrong job.”

Rolf Turner

R-help (maio de 2008)

A avaliação dos algoritmos de reconstrução de sinais apresentados no capítulo 2 foi efetuada através de simulações Monte Carlo implementadas na plataforma R. Neste capítulo seguem os detalhes da modelagem e implementação do sistema de simulações e de como os dados foram gerados.

3.1 Modelo formulado

O cenário de simulações foi montado através de um processo estocástico gaussiano estacionário isotrópico. Os dados são amostrados na forma (x, y) , onde y representa uma ocorrência da variável aleatória $S(x)$.

Seguem novas definições necessárias a compreensão da implementação do modelo.

3.1.1 Sinal de origem amostrado

Os dados utilizados nas simulações são provenientes de campos aleatórios gaussianos $\{S(x) : x \in \mathbb{R}^2\}$ gerados pelos métodos apresentados por [Chan & Wood \(1997\)](#); [Lantuejoul \(2002\)](#) e [Schlather \(1999\)](#).

Neste trabalho, adotou-se o modelo Matérn (ver seção 2.2.1), citado como modelo de correlação mais utilizado em geostatística ([Bivand & Pebesma, 2008](#); [Diggle & Ribeiro Jr., 2007](#); [Le & Zidek, 2006](#)), para as variáveis aleatórias do campo gaussiano. Assim, pode-se variar os parâmetros deste modelo de forma a gerar as situações encontradas na natureza.

Os parâmetros variados são média, variância, coeficiente de suavização (κ) e escala (ϕ). Este último permite modelar, dentre outras aplicações, a intensidade da luz que chega ao solo de uma floresta, em função da densidade da folhagem das árvores (ver [Alencar-Neto, 2008](#)).

A partir deste ponto, sempre que for referido um campo gaussiano ele terá média zero e variância unitária. Exemplos de campos aleatórios gaussianos com parâmetros variados podem ser vistos na figura 3.1.

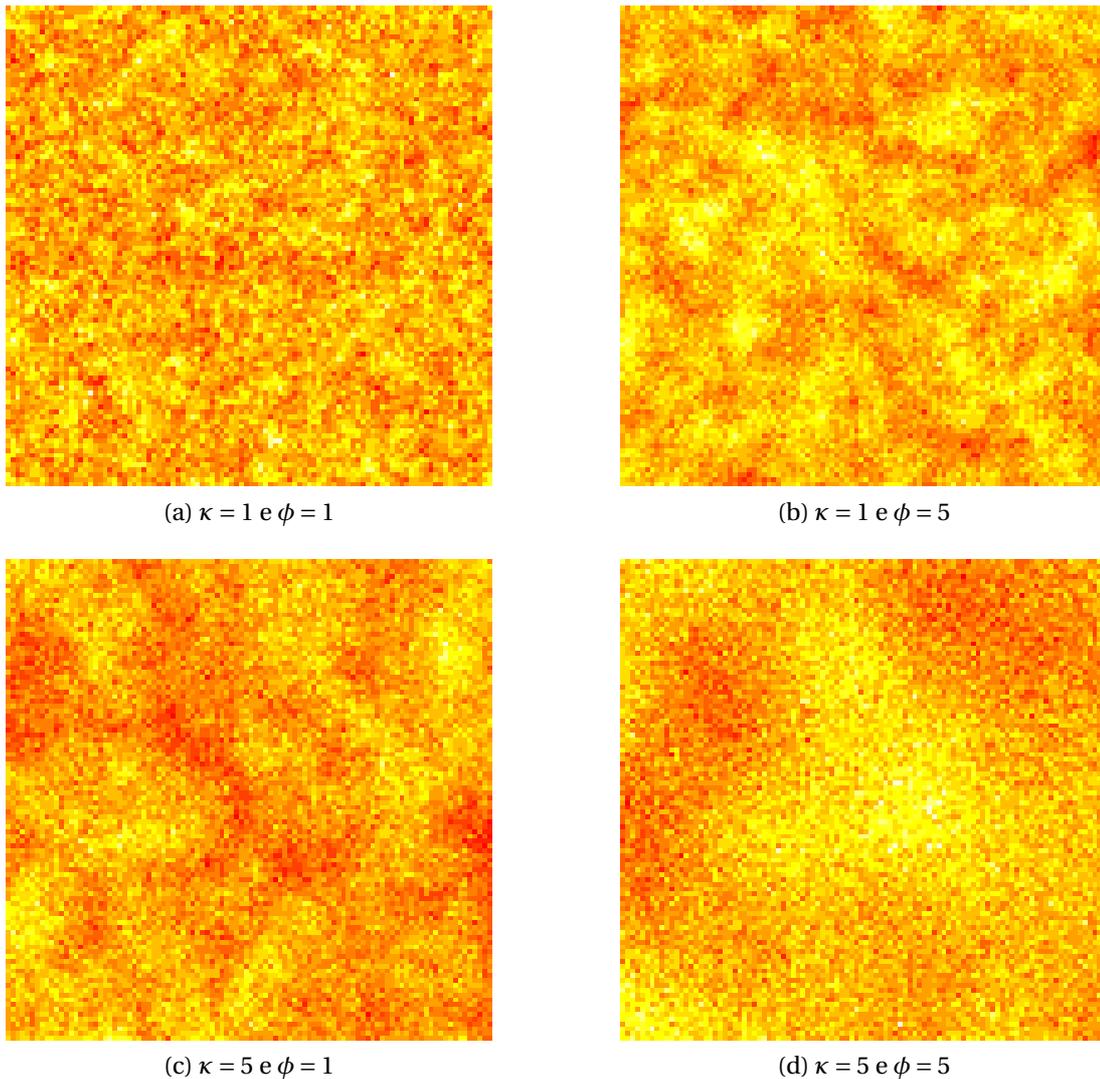


Figura 3.1: Exemplos de campos aleatórios gaussianos de 100×100 pontos, média 0 e variância 1, variando parâmetros de suavização (κ) e escala (ϕ), onde utiliza-se uma escala de cores para representar a mudança de valores

3.1.2 Processos pontuais e a distribuição dos sensores

A distribuição dos sensores segue o modelo proposto por [Alencar-Neto \(2008\)](#); [Frery et al. \(2008\)](#), utilizando processos pontuais ([Baddeley, 2006](#)) para distribuir os nós sensores. Estes, são modelos estocásticos que descrevem a distribuição de dados no espaço. Para a compre-

ensão do modelo adotado, é necessário incluir algumas definições com base nos textos de James (2006); Alencar-Neto (2008), a saber.

Definição 25 (Processo Estocástico Poisson). Sendo $E_{t,\Delta t}^k$ a notação para k eventos ocorridos no intervalo de tempo $(t, t + \Delta t]$, a distribuição de um processo de Poisson pode ser definida como $\Pr(E_{0,t}^k) = (\lambda t)^k e^{-\lambda t} / k!$ Sua caracterização se dá pelo parâmetro λ , relação com a intensidade, parâmetro η na forma $\eta(t) = \lambda t$, que denota a quantidade esperada de eventos por unidade de tempo. As seguintes hipóteses seguem da probabilidade acima:

1. $\Pr(E_{t,t+\Delta t}^k) = \Pr(E_{0,\Delta t}^k), \forall k, t \text{ e } \Delta t;$
2. $\Pr(E_{t_1,\Delta t_1}^{k_1} \cap E_{t_2,\Delta t_2}^{k_2}) = \Pr(E_{t_1,\Delta t_1}^{k_1}) \Pr(E_{t_2,\Delta t_2}^{k_2})$ se $(t_1, \Delta t_1] \cap (t_2, \Delta t_2] = \emptyset;$
3. $\lim_{t \rightarrow 0} \frac{\Pr(E_{0,t}^a)}{\Pr(E_{0,t}^b)} = 0$, com $a \geq 2$ e $b \geq 1$.

Definição 26 (Processos pontuais Homogêneos com Independência). São processos pontuais onde a quantidade de eventos esperados (parâmetro η) é constante, ou seja, dado um processo $\{S(x) : x \in \mathbb{R}^d\}$, $\eta(x) = \lambda > 0, \forall x \in \mathbb{R}^d$.

A partir deste ponto, considere-se todo processo estocástico com dimensão 2.

Definição 27 (Distribuição Uniforme). Uma variável aleatória contínua $X : \Omega \rightarrow \{x \in \mathbb{R} : \alpha \leq x \leq \beta\}$ segue a distribuição uniforme se sua densidade for dada pela função $f(t) = (\beta - \alpha)^{-1} \mathbb{1}_{(a,b)}(t)$, com $-\infty < \alpha < \beta < \infty$.

Definição 28 (Distribuição Poisson). Uma variável aleatória segue a distribuição de Poisson se sua função de probabilidade for dada por $\Pr(t) = \lambda^t e^{-\lambda} / t!$, com $\lambda > 0$ e $t \geq 0$ inteiro.

Definição 29 (Processo pontual Binomial). Um processo pontual $\{S(x) : x \in \mathbb{R}^d\}$ é dito binomial $B(n)$ para um suporte retangular se as coordenadas dos pontos que por ele foram gerados forem ocorrências de variáveis aleatórias uniformes independentes, onde $n \geq 1$ é o número de pontos.

Definição 30 (Processo pontual Poisson). Seja um processo pontual Poisson de intensidade $\eta > 0$, $t > 0$ e $\lambda > 0$. Nele, p_1, p_2, \dots, p_n representam os pontos deste processo em um subconjunto $A \subset \mathbb{R}^2$, onde o número de pontos de qualquer conjunto $B \subset A$, denominado $N(B)$ possuirá as seguintes propriedades:

1. $N(B)$ segue uma distribuição de Poisson de média $\mu(B)$ definida por $\int_B \eta(u) du$, tal que $\eta(u)$ é constante para todo $u \in B$;
2. as variáveis aleatórias $N(B_1), \dots, N(B_n)$ são coletivamente independentes se os conjuntos B_1, \dots, B_n forem disjuntos.

A partir daí, um processo pontual binomial será um processo pontual de Poisson condicionado ao número de pontos n . Exemplos de processos pontuais Poisson podem ser observados na figura 3.2. Estes podem ser encarados como exemplos de processos pontuais binomiais com n dado em 3.2a, 3.2b, 3.2c e 3.2d.

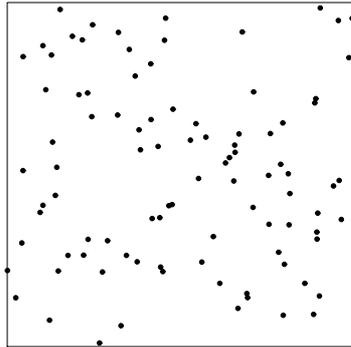
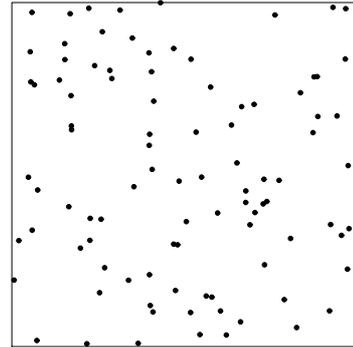
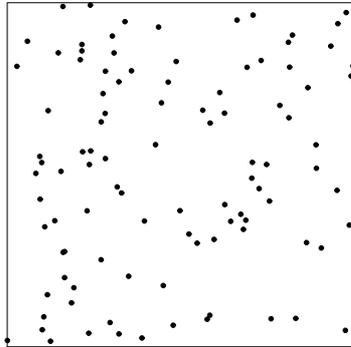
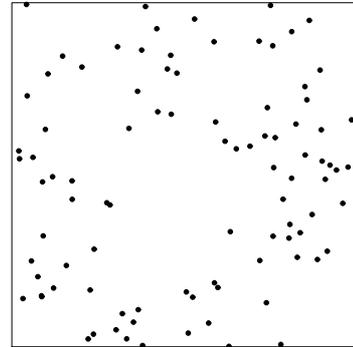
(a) $n = 94$ pontos(b) $n = 95$ pontos(c) $n = 100$ pontos(d) $n = 94$ pontos

Figura 3.2: Exemplos de um processo pontual de Poisson com $\lambda = 100$

Definição 31 (Processos pontuais não homogêneos). Num processo pontual não homogêneo a intensidade do parâmetro η pode variar de acordo com a área onde o processo é aplicado. Assim é possível construir processos onde a probabilidade de se obter n pontos varie em decorrência da área de A em questão. Pode-se construir um processo desse tipo definindo sua intensidade η como uma função tal que $\eta: A \rightarrow \mathbb{R}_+$ com $\int_B \eta < \infty$, qualquer que seja $B \subset A$

Definição 32 (Processo pontual Poisson não homogêneo). É um conjunto de pontos p_1, \dots, p_n localizados num conjunto $A \subset \mathbb{R}^2$ tal que o número de pontos $N(B)$ de qualquer $B \subset A$ possui as seguintes propriedades:

1. $N(B)$ segue uma distribuição de Poisson de média $\mu(B)$ tal que $\eta(u)$ varia de acordo com a função de intensidade $\eta: A \rightarrow \mathbb{R}_+$ qualquer que seja $u \in B$;
2. as variáveis aleatórias $N(B_1), \dots, N(B_n)$ são coletivamente independentes se os conjuntos B_1, \dots, B_n forem disjuntos.

Exemplos de um processo de Poisson não homogêneo podem ser observados na figura 3.3 com $\eta(x, y) = xe^{\alpha x}$, cujo comportamento pode ser observado na figura 3.4.

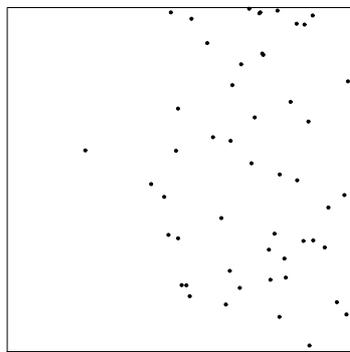
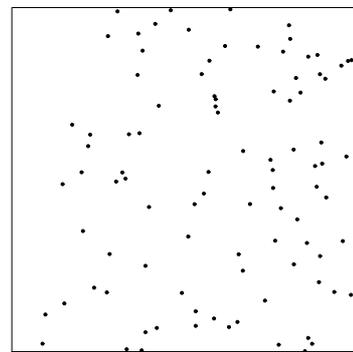
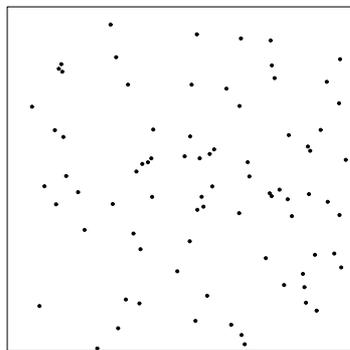
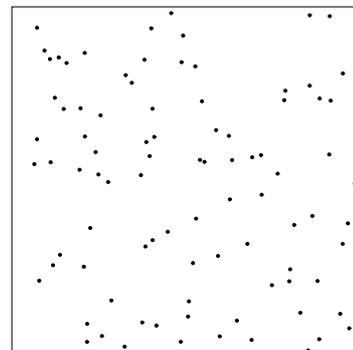
(a) $\alpha = 10$ (b) $\alpha = 50$ (c) $\alpha = 75$ (d) $\alpha = 100$

Figura 3.3: Exemplos de um processo pontual de Poisson não Homogêneo com $\eta(x, y) = xe^{\alpha x}$

De forma iterativa, o processo pontual *SSI* (*Matérn's Simple Sequential Inhibition*) apresenta-se com tentativas de por n pontos através de uma distribuição uniforme de média $2n$ numa região A . Com um parâmetro de distanciamento obrigatório (raio de inibição) r , insere um primeiro ponto em A e efetuam-se até T tentativas para cada um dos $n - 1$ pontos restantes, sendo estas bem sucedidas se nenhuma das distâncias entre este último ponto

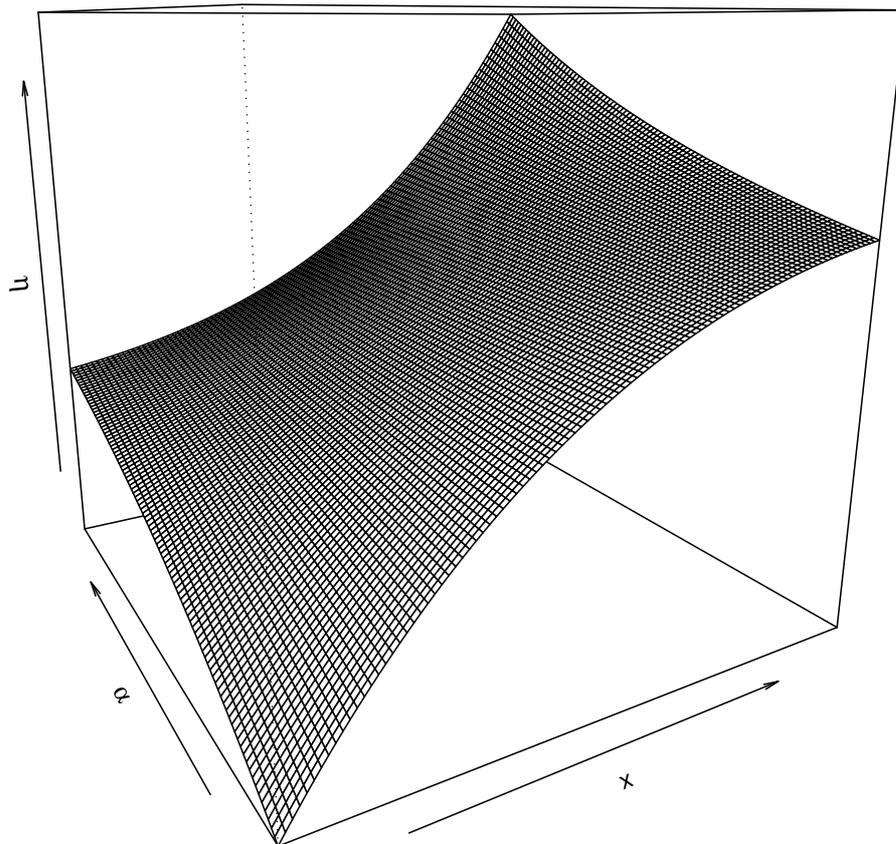


Figura 3.4: Perspectiva da função de intensidade $\eta(x, y) = xe^{\alpha x}$

e quaisquer outro for menor ou igual a r , caso contrário, o ponto é descartado. Na figura 3.5 pode-se observar exemplos de um processo SSI de $n = 100$, $T = 1.000$ e raio de inibição variando em $r = \{0,01, 0,05, 0,10, 0,11\}$.

Um dos usos comuns aos processos pontuais é modelar a disposição dos sensores numa RSsF. Diversas situações foram modeladas por Alencar-Neto (2008); Frery et al. (2008), através do processo composto $C(n, \alpha)$, em que n denota a quantidade de sensores a serem dispostos e α um coeficiente de atratividade. São consideradas situações de disposição num plano $A \subset \mathbb{R}^2$, conforme segue nomeado pela característica atribuída ao processo composto C pela valoração do parâmetro α .

Independente: processo pontual Binomial definido na região A , na forma $B[n, A]$.

Atrativo: composição de dois processos de Poisson, um com intensidade $\alpha\eta$ sobre uma área alvo $A' \subset A$ e outro com intensidade η , sobre a área $A \setminus A'$, na forma $S[n, \alpha]$.

Repulsivo: processo pontual SSI com raio de inibição $r_{\max} = n^{-1/2}$, na forma $SSI[n, r_{\max}, A]$.

Regular: situação de repulsão extrema, quando $\alpha = -\infty$.

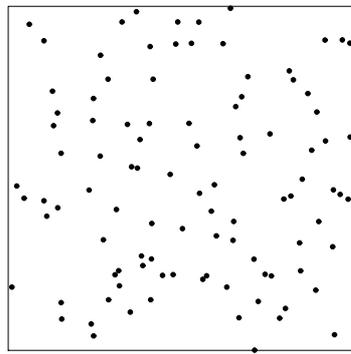
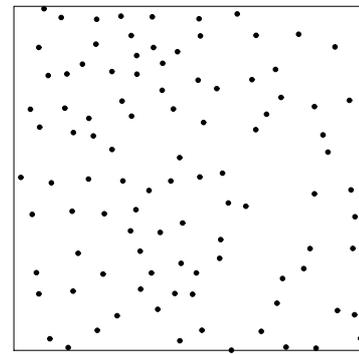
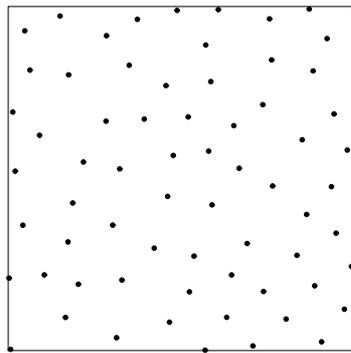
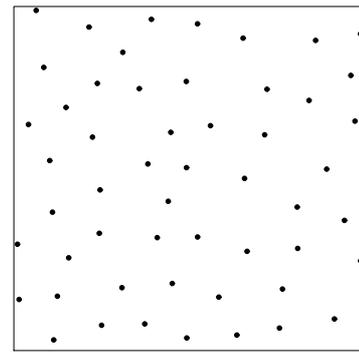
(a) $r = 0,01$ (b) $r = 0,05$ (c) $r = 0,10$ (d) $r = 0,11$

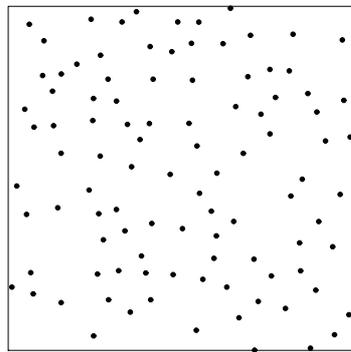
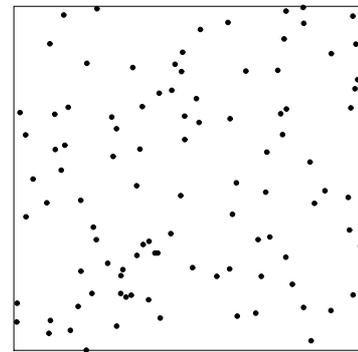
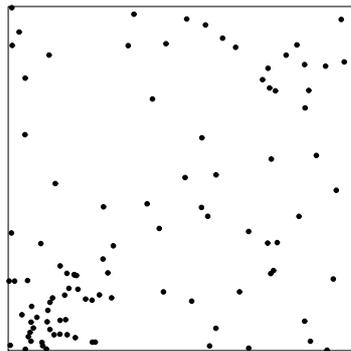
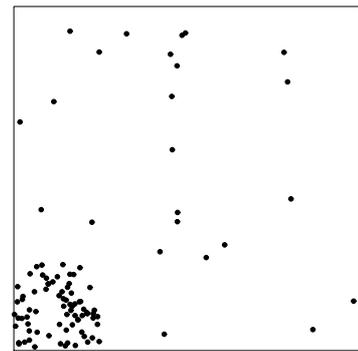
Figura 3.5: Exemplos de um processo pontual de *SSI* com $n = 100$ e $T = 1.000$

Assim, este processo será definido como

$$C(n, \alpha) = \begin{cases} SSI(n, r_{\max}(1 - e^{-\alpha})), & \text{se } \alpha < 0 \\ B(n), & \text{se } 0 \leq \alpha \leq 1 \\ S(n, \alpha), & \text{se } \alpha > 1. \end{cases} \quad (3.1)$$

Na figura 3.6 pode-se observar exemplos do processo composto C .

Dado que para se distribuir sensores sobre grandes regiões o meio mais comum é o uso de aviões, os quais depõem os nós sensores, modela-se este procedimento através do processo C . Sob um plano de voo de altura constante, os sensores são dispostos gradativamente, num processo repulsivo (figura 3.6a). Caso sejam lançados muitos sensores de uma única vez, nós tendem a se aglomerar numa parte da região, conforme um processo atrativo (figuras 3.6c e 3.6d). Quando não há controle a disposição e esta se dá de forma totalmente aleatória, caracteriza-se um processo independente (figura 3.6b).

(a) $\alpha = -10$: processo repulsivo(b) $\alpha = 0$: processo independente(c) $\alpha = 10$: processo atrativo(d) $\alpha = 50$: processo fortemente atrativoFigura 3.6: Exemplos do processo composto C com $n = 100$

Este processo composto C será um caso particular de um processo mais complexo, o M^2P^2 (Ramos et al., 2010), que modela mais cenários, conforme a definição 33.

Definição 33 (Multilevel Marked Point Process). O processo $M^2P^2(n, \alpha, m, r_c, r_i)$ representa diversas possibilidades de disposição de sensores, onde n representa a quantidade de sensores e α o coeficiente de atratividade, m a quantidade de H-sensores (um grupo pequeno de sensores com quantidade menor que a metade do total), r_c o raio máximo de comunicação dentre os L-sensores (sensores além dos H-sensores), r_i o raio de inibição dentre os H-sensores. Seja um número pequeno $m > 1$ de H-Sensores num grupo $n > m$ de L-Sensores num círculo de raio $r_c > 0$ centrado em cada H-Sensor. Os H-Sensores são dispostos e em seguida os L-Sensores são dispostos “próximos” aos H-Sensores. Seja um processo pontual de Poisson de função de intensidade

$$\eta(x, y) = \begin{cases} \alpha, & \text{se } d((x, y), (h_{x_i}, h_{y_i})) \leq r_c, 1 \leq i \leq m, \\ 1, & \text{caso contrário,} \end{cases}$$

onde d é uma medida de distância (neste trabalho, a euclidiana), $\{(h_{x_1}, h_{y_1}), \dots, (h_{x_m}, h_{y_m})\}$ são as coordenadas dos H-Sensors e r_c é o raio máximo de comunicação dos L-Sensors. Denote-se este processo por $\Lambda(n - m, \alpha, r_c, h)$. O processo M^2P^2 é um processo composto de m amostras de $H(m, r_i)$, os H-Sensors, e $n - m$ amostras de $\Lambda(n - m, \alpha, r_c, h)$, os L-Sensors. O processo corresponde as seguintes configurações de α :

$$\alpha \begin{cases} = -1 & \text{processo totalmente binomial,} \\ = 0 & \text{processo repulsivo para os H-Sensors e binomial para os L-Sensors,} \\ > 0 & \text{processo repulsivo para os H-Sensors e atrativo em torno dos} \\ & \text{H-Sensors para os L-Sensors.} \end{cases} \quad (3.2)$$

Exemplos dessas situações podem ser observados na figura 3.7.

3.1.3 Reconstrução do sinal

O sinal será reconstruído utilizando os algoritmos de *kriging* definidos nas seções 2.4.1 e 2.4.2 além de algumas variantes. São estas:

- *kriging* simples com média constante;
- *kriging* simples com média computada através de mínimos quadrados ordinários;
- *kriging* simples com média computada por mínimos quadrados generalizados;
- *kriging* ordinário;
- *kriging* bayesiano;
- *kriging* bayesiano modificado (média computada por mínimos quadrados generalizados).

Uma vez com os campos devidamente reconstruídos, utilizar-se-ão os valores computados pelas funções de *kriging* dos parâmetros κ e ϕ para reconstruir os campos gaussianos com a mesma função que os gerou na primeira etapa do processo. Esse processo está representado no diagrama da figura 3.8. Da natureza (\mathcal{N}), representa-se numericamente um fenômeno de interesse (F representado em V). Desse fenômeno coletam-se 50 valores (\mathbb{R}_{50}) em pontos correspondentes às localizações de sensores dispostos (S), nos quais por um algoritmo de agrupamento (Ψ) serão compostos 6 grupos, totalizando 6 valores de média aritmética de cada grupo (\mathbb{R}^6). Com esses valores reconstrói-se o sinal de origem com a computação dos parâmetros dos campos gaussianos do sinal de origem (κ e ϕ , representados por $\hat{\theta}$). A partir desses valores reconstróem-se os valores numéricos do campo gaussiano de origem ($\mathbb{R}_{100 \times 100}^*$).

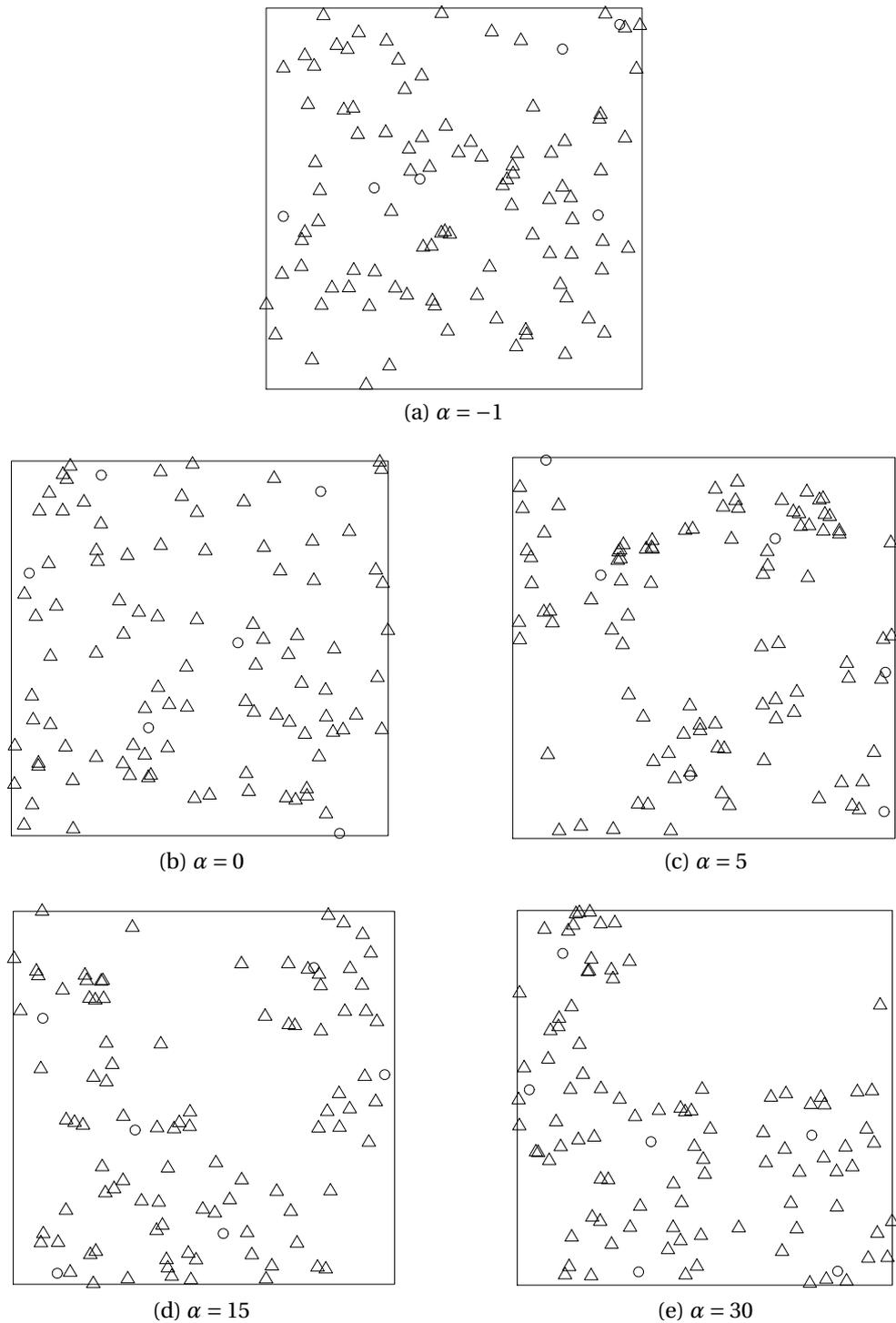


Figura 3.7: Exemplos do processo M^2P^2 , com raio de comunicação 20, raio de sensoriamento 5 e 6 H -Sensors, para 100 sensores, variando o valor de α , onde cada \triangle denota um L -sensor e \circ um H -sensor

$$\mathcal{N} \xrightarrow{F} \mathbb{R}^{100 \times 100} \xrightarrow{S} \mathbb{R}_{50} \xrightarrow{\Psi} \mathbb{R}_6 \xrightarrow{\hat{\theta}} \mathbb{R}^{100 \times 100}$$

Figura 3.8: Diagrama ilustrativo da reconstrução dos campos gaussianos com base nos valores inferidos aos parâmetros κ e ϕ pelos algoritmos de *kriging*

3.1.4 Validação do modelo

O modelo será validado através do Método de Monte Carlo (Metropolis & Ulam, 1949; Murray, 1953; Robert & Casella, 1999), que consiste na repetição exaustiva do procedimento utilizando dados diferentes, porém provenientes do mesmo processo. Para aferir a qualidade dos algoritmos, utilizam-se três métricas de qualidade: o viés (definição 34), a variância (definição 15) e o erro quadrático médio (definição 35).

Definição 34 (Viés). O viés do estimador $\hat{\theta}$, denotado $\mathbb{B}[\hat{\theta}]$, é a diferença entre o valor esperado do estimador e o verdadeiro valor do parâmetro: $\mathbb{B}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta$.

Um estimador cujo viés seja igual a zero é dito não viesado.

Definição 35 (Erro Quadrático Médio – EQM). Define-se como $\text{EQM}(\hat{\theta}) = \mathbb{B}^2[\hat{\theta}] + \text{Var}(\hat{\theta})$. Sua utilização permite quantificar unificadamente o quão distante (em média) do valor real estão as estimativas e o quanto as estimativas estão distantes umas das outras. Para um estimador não viesado, o EQM reduz-se à variância do estimador.

3.2 Implementação de simulações Monte Carlo na literatura

A literatura sobre simulações de Monte Carlo é extremamente vasta. Uma busca simples no *ISI Web of Knowledge/Web of Science* com a chave “monte carlo” retorna mais de 100.000 resultados.

Encontram-se disponíveis diversos estudos que implementam simulações de Monte Carlo utilizando planilhas eletrônicas. Dentre diversos outros, pode-se citar trabalhos, como o de Li & Low (2010), que apresenta uma nova metodologia para análise de riscos na saúde pública; Tatone & Grasselli (2010) com uma modelagem probabilística para o tombamento de blocos com implementação baseada em planilhas eletrônicas; Schilstra & Martin (2009) afirma que resultados interessantes de aplicações simples sempre podem ser obtidos por simulação de Monte Carlo em planilhas eletrônicas; Rivard et al. (2009) efetua análises de *benchmark* para um equipamento de braquiterapia; Oscar (2009) modela a vida e crescimento da salmonela.

Um maior destaque se dá a trabalhos com grande número de citações, como o de Thompson et al. (1992), que estende modelagens da análise de riscos na saúde pública;

Lindqvist & Westoo (2000) modelam uma avaliação de riscos ecotoxicológicos em bacias; e o de Vose (1998), que apresenta métodos para análise da contaminação de gêneros alimentícios. Esses três últimos trabalhos possuem mais de 40 citações no *ISI Web of Knowledge/Web of Science*, sendo os atualmente mais citados em uma busca com a chave “spreadsheet monte carlo”.

O uso das planilhas eletrônicas é bastante difundido devido a sua facilidade de uso, como pode ser visto no sítio da [European Spreadsheet Risks Interest Group \(2010\)](#). Nele podem ser encontrados diversos casos de uso bem como histórias de desastres por erros em planilhas.

Ao mesmo tempo, há diversos artigos na literatura que criticam o uso de planilhas em aplicações estatísticas devido a problemas numéricos que lhes são recorrentes (ao exemplo de, dentre diversos outros, [McCullough, 2008a,b, 2004, 1999, 1998](#); [McCullough & Heiser, 2008](#); [McCullough & Wilson, 2005, 2002, 1999](#); [Nash, 2008, 2006](#)). Entretanto, no início deste trabalho apenas foram encontradas disponíveis na literatura análises do Microsoft Excel até a versão 2003 e GNU Gnumeric; nada foi encontrado sobre outras planilhas eletrônicas, a exemplo do OpenOffice.org Calc, a mais difundida dentre a comunidade *Open Source*. Isso motiva a uma séria avaliação das propriedades numéricas das planilhas eletrônicas antes de seu uso.

3.3 Critérios para pesquisa reproduzível

Para que a pesquisa possa ser validada por outros pesquisadores, a mais importante de suas características é a reprodutibilidade. Só através da capacidade de reproduzir os experimentos e verificar a consistência dos dados se podem avaliar a corretude dos resultados e a pertinência das conclusões.

Diversos são os trabalhos que trazem à tona a problemática da falta de precisão numérica em programas científicos. Dentre diversos outros, podemos citar os trabalhos de [Almiron et al. \(2009, 2010\)](#); [Altman & McDonald \(2001\)](#); [Bustos & Frery \(2006\)](#); [Kennedy & Gentle \(1980\)](#); [Callaert \(2003\)](#); [Altman et al. \(2007\)](#); [Oliveira & Stewart \(2006\)](#), que não só avaliam o erro de plataformas numéricas como apresentam diversos erros que lhes são recorrentes, inclusive trabalhos como o de [Andel & Yasinac \(2006\)](#), que avalia imprecisões nos resultados obtidos através de simuladores de Redes de Sensores sem Fios. Sem a certeza de que os resultados são realistas, uma análise torna-se inviável.

O modelo adotado para suprir esta propriedade está detalhado por [Koenker & Zeileis \(2009\)](#), além de uma vasta quantidade de informações disponibilizadas por [Vandewalle et al. \(2009\)](#). As principais características necessárias às experiências Monte Carlo são:

- independência de plataforma de sistema operacional;
- independência de plataforma de *hardware*;

- capacidade de reproduzir a sequência de números pseudo-aleatórios;
- boas propriedades numéricas.

Efetuiu-se esta análise nas planilhas eletrônicas mais difundidas, cujo conteúdo detalhado pode ser obtido no trabalho de [Almiron et al. \(2010\)](#). As plataformas avaliadas, cujas versões e arquiteturas de sistema estão apresentadas na tabela 3.1, foram:

- OpenOffice.org Calc (<http://www.openoffice.org/product/calc.html>);
- Microsoft Excel (<http://www.microsoft.com/Excel>);
- GNU Gnumeric (<http://www.gnome.org/gnumeric>);
- NeoOffice NeoCalc (<http://neowiki.neooffice.org/index.php/NeoCalc>);
- GNU Oleo (<http://www.gnu.org/software/oleo>).

Tabela 3.1: Plataformas de planilhas eletrônicas avaliadas

Plataforma		Calc		Excel		Gnumeric		NeoOffice		Oleo
<i>Hardware</i>	SO	2.4.1	3.0.1	2007	2008	1.8.3	1.9.1	2.2.5	3.0	1.99.16
i386	Windows	✓	✓	✓			✓			
	Ubuntu	✓	✓			✓				✓
amd64	Mac OS		✓		✓			✓	✓	

Para avaliar os geradores de números pseudo-aleatórios, adotaram-se os princípios de [Ripley \(1990, 1987\)](#), que diz que bons geradores de números pseudo-aleatórios devem prover sequências numéricas com as seguintes propriedades:

1. devem seguir a distribuição uniforme;
2. vetores de dimensão moderada de variáveis aleatórias diferentes, mas do mesmo gerador, devem ser coletivamente independentes;
3. devem ser reproduzíveis através de poucos parâmetros simples de especificar (a semente), independente do ambiente computacional (*hardware*, sistema operacional, linguagem de programação);
4. devem ser produzidas rapidamente;
5. devem possuir longos períodos.

Apenas através destes parâmetros, simulações Monte Carlo já se tornam inviáveis em planilhas eletrônicas, visto que apenas numa delas há documentação e possibilidade de informar a semente (Excel 2007). De acordo com sua documentação, o gerador de números pseudo-aleatórios implementado seria o de [Wichmann & Hill \(1982\)](#), entretanto, seguindo a mesma metodologia de [McCullough \(2008a\)](#), pôde-se verificar que essa informação está incorreta. O gerador também não corresponde a nova versão deste gerador proposta por [Wichmann & Hill \(2006\)](#). A situação dos geradores dessas planilhas está resumida na tabela 3.2.

Tabela 3.2: Documentação e disponibilidade de informar a semente em geradores de números pseudo aleatórios de planilhas eletrônicas

	Calc	Excel 2007	Excel 2008	Gnumeric	NeoOffice	Oleo
Documentação	N	S-incorreta	N	S	N	N
Semente	N	S	N	N	N	N

No ambiente adotado, o R, há diversos geradores de números pseudo aleatórios implementados. O gerador atualmente padrão e utilizado neste trabalho é o Mersenne-Twister ([Matsumoto & Nishimura, 1998](#)), que possui todas as propriedades supracitadas, com um período de $2^{19937} - 1$.

3.4 Implementação na plataforma R

R ([R Development Core Team, 2009](#)) é uma linguagem e um ambiente de computação funcional com um vasto ferramental de suporte à estatística computacional. É um projeto da GNU (<http://www.gnu.org>) sob a licença GPL ([Michaelson, 2004](#); [Ueda, 2005](#)), fazendo parte dos programas de código aberto denominados FLOSS (*Free/Libre Open Source Software*). A qualidade de suas propriedades numéricas pode ser aferida no trabalho de [Almiron et al. \(2009\)](#), garantindo que os valores resultantes das simulações sejam coerentes com o modelo adotado (sem incoerências por erro numérico).

Em seus repositórios oficiais (<http://cran.r-project.org/mirrors.html>) é possível encontrar uma vasta quantidade de pacotes que possibilitam incluir novas funções.

A base inicial das ferramentas de geoestatística utilizadas nesse trabalho provém do pacote `geoR` ([Ribeiro Jr. & Diggle, 2001](#)), que implementa os algoritmos de *kriging* simples, ordinário e bayesiano. Uma nova versão do algoritmo de *kriging* bayesiano é também proposta modificando o já existente. Nesta versão modificara, a média é computada por mínimos quadrados generalizados, levando em conta os parâmetros de covariância sendo mantida constante (não será ocorrência de uma variável aleatória, conforme definido na seção 2.4.2).

Para a implementação do algoritmo de agrupamento *SKATER* utiliza-se o pacote `igraph`

(Csardi & Nepusz, 2006), que implementa diversas operações em grafos, além de uma conveniente estrutura de dados para armazenamento dos grafos.

Os dados do sinal de origem provém de funções para gerar campos aleatórios gaussianos do pacote RandomFields (Schlather, 2009) e o ferramental para uso de processos pontuais é implementado pelo pacote spatstat (Baddeley & Turner, 2005).

A implementação do processo composto M^2P^2 é a provida por Ramos et al. (2010).

O sistema de simulações, então, efetua as etapas enunciadas abaixo, com elementos referenciados do diagrama da figura 2.8.

Geração do sinal de origem: gera-se um campo aleatório gaussiano (V^*).

Disposição dos sensores: os sensores são dispostos (S).

Agrupamento dos sensores: executa-se um algoritmo de agrupamento (Ψ).

Amostragem: os dados são coletados na forma (x, y) , onde x denota o posicionamento geográfico e y é uma ocorrência da variável aleatória Y , que representa os dados amostrados no ponto x (V'). Ao se utilizar um algoritmo de agrupamento, cada *clusterhead* efetua uma fusão dos dados de seu *cluster* e os valores coletados seguem com o posicionamento geográfico do *clusterhead*; apenas a informação fundida é transmitida, da mesma forma utilizada por Nordio et al. (2010).

Reconstrução do sinal: o sinal é reconstruído por um algoritmo de *kriging* (\hat{F}).

Para que os resultados sejam os mais realistas o possível, são analisados sensores dispostos através do processo M^2P^2 com coeficiente de atratividade $-1, 0$ e 30 . Em cada campo gaussiano, os sensores são dispostos em suas três configurações; em cada uma das configurações executam-se os dois algoritmos de agrupamento (seções 2.5.1 e 2.5.2), além de uma versão sem considerar agrupamentos; executam-se os três algoritmos de *kriging* (seções 2.4.1 e 2.4.2); no algoritmo de *kriging* simples efetua-se a média informada, computada por mínimos quadrados ordinários e computada por mínimos quadrados generalizados, considerando os valores de covariância inferidos; no *kriging* bayesiano efetua-se o algoritmo em sua forma tradicional, com média informada e com média computada por mínimos quadrados generalizados considerando os valores de covariância inferidos. O diagrama da figura 3.9 ilustra o modo como as simulações se processam, considerando que cada transição (representa pelas setas) deve contemplar todos os parâmetros.

3.5 Ambiente de execução

Simulações Monte Carlo tipicamente possuem alta complexidade computacional (polinomial, com expoente crescendo de acordo com a quantidade de parâmetros variados). Para

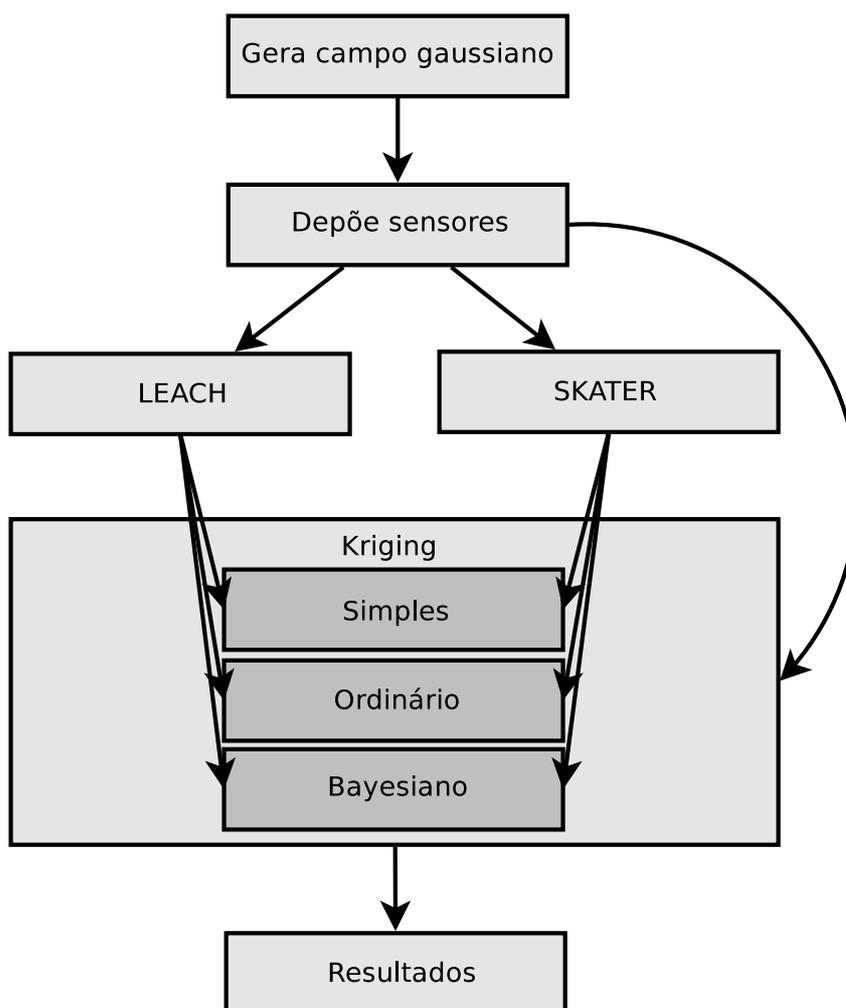


Figura 3.9: Diagrama ilustrativo do sistema de simulações

tornar viável a execução de um estudo Monte Carlo com as variáveis supracitadas utiliza-se de um ambiente de computação de alto desempenho.

O ambiente em questão é o *cluster* de computadores integrado à GradeBR-UFAL, integrado à Rede Galileu (integração dentre supercomputadores de quatro instituições: UFAL, UFRJ, PUC-Rio e USP) do Laboratório de Computação Científica e Visualização da Universidade Federal de Alagoas. Este *cluster* é composto de nós com dois processadores, cada um destes com quatro núcleos de processamento Intel Nehalem (Intel Corporation, 2008), e 24GB de memória RAM, tendo sido disponibilizadas doze computadores para este trabalho.

O sistema de paralelização das simulações foi desenvolvido em linguagem C (algoritmo exposto no código 3.1), utilizando a interface de paralelização MPI (Walker & Dongarra, 1994). A implementação desta interface é a OpenMPI (Gabriel et al., 2004), portátil à diversas plataformas (GNU Linux, MacOS, Microsoft Windows, BSD e diversos modelos baseados em Unix). O sistema operacional utilizado é o GNU Linux CentOS (<http://centos.org>) versão 5.4, *kernel* versão 2.6.18-164.el5 compilado para x86_64 (arquitetura amd64). O compilador utilizado é o Intel Compiler versão 11 (<http://software.intel.com/en-us/intel-compilers>).

```
1 Entrada: arquivos .R para simular
2 Saida: arquivos com dados da simulacao
3 cont <- 0
4 incompleto <- Verdadeiro
5 rank <- identificador do processador
6 tam <- quantidade de nucleos livres
7 Enquanto incompleto
8 {
9     Se rank = 0
10    {
11        Para cada nucleo livre n
12        {
13            cont <- cont + 1
14            Envia cont para n
15        }
16    }
17    Se nao
18    {
19        Recebe o valor de cont
20        Executa a simulacao de numero cont
21    }
22    Se rank = 0
23    {
24        Se cont >= Total de arquivos
25            incompleto <- FALSO
26        Para cada nucleo livre n
27            Envia incompleto para n
28    }
29    Se nao
30        Recebe o valor de incompleto
31 }
```

Código 3.1: Algoritmo para paralelização das simulações

A troca de mensagens dentre os nós se dá via SSH (*Secure SHell*, Lehtinen & Lonvick, 2006), na implementação OpenSSH (<http://openssh.com>), e os arquivos estão armazenados de forma distribuída em rede através dos sistemas de arquivos NFS (Callaghan, 2000) e LustreFS (Oracle Corporation, 2010), este para grandes e aquele para pequenos arquivos.

3.6 Resultados esperados

Com os sinais devidamente reconstruídos, serão avaliados no capítulo 4 a qualidade dos valores computados através dos critérios do viés, da variância e do erro quadrático médio. Serão avaliados os campos resultantes dos processos de *kriging* pelo pacote `geoR` e os reconstruídos através do pacote `RandomFields` utilizando os valores dos parâmetros estimados durante o processo de *kriging*.

No capítulo 5, serão discutidos os resultados da análise e apontados os trabalhos futuros.



Análise dos resultados

“Will Frank Harrell or someone else please explain to me a real application in which this is not fast enough?”

Brian D. Ripley

R-devel (dezembro de 2004)

UMA vez com os dados da simulação modelada no capítulo 3, procedeu-se a análise estatística. Como as simulações foram efetuadas através do método de Monte Carlo por Força Bruta (Metropolis & Ulam, 1949; Murray, 1953; Robert & Casella, 1999), analisaram-se o viés, a variância e o erro quadrático médio dos estimadores.

4.1 Dados obtidos com a simulação

Os dados resultantes do ensaio de Monte Carlo foram os campos gaussianos reconstruídos, o valor dos parâmetros κ e ϕ e o valor estimado para a média, nos casos em que esta não foi informada.

Para garantir a fidelidade nos valores a serem comparados, uma vez com os parâmetros estimados os campos gaussianos são reconstruídos através da mesma função que os gerou. O valor das médias é também utilizado na reconstrução e analisado isoladamente. Essa última análise se torna relevante uma vez que, à exceção do *kriging* ordinário que possui estimador próprio para a média, os algoritmos são flexíveis no modo como a média é computada. Tanto ela pode ser informada como estimada por padrão através de mínimos quadrados generalizados. Essa operação pode ter custo computacional inviável à sistemas de tempo real.

4.2 Análise da qualidade do sinal reconstruído

Conforme apontado por Diggle & Ribeiro Jr. (2007), o modelo de correlação Matérn é o mais utilizado por apresentar comportamento de interesse a diversos cenários e possuir parametrização não só de distanciamento (ϕ) como também de suavização de suas curvas (κ). Assim sendo, as análises deste trabalho consideram esse modelo de correlação. Em todas as tabelas os valores apresentados foram arredondados para duas casas decimais.

Os valores do parâmetros de atratividade (α) utilizados na distribuição dos sensores, feita através do processo M^2P^2 foram $\alpha \in \{-1, 0, 30\}$, com 11, 23 e 13 replicações respectivamente para cada ponto de α para $\phi = 1$ e $\alpha = \{0, 30\}$ com 15 e 92 replicações para $\phi = 5$. Em todos os casos, o parâmetro κ teve valor igual a 0,5.

4.2.1 Análise do comportamento dos parâmetros do modelo

Para a estimação dos parâmetros ϕ e κ , é necessário informar um ponto de partida para o algoritmo. Em ambos os casos, o valor informado foi 0,5, que coincidiu com o valor real do parâmetro κ . O valor encontrado pelo estimador do parâmetro κ foi este informado como partida (valor real). Os resultados das estimativas do parâmetro ϕ encontram-se na tabela 4.1.

Tabela 4.1: Viés e EQM dos estimadores do parâmetro ϕ para o valor real de $\kappa = 0,5$

Agrupamento	<i>Kriging</i>						ϕ	α
	Simples		Ordinário		Bayesiano			
	Viés	EQM	Viés	EQM	Viés	EQM		
—	0,75	0,88	0,04	1,67	0,08	0,01	1,0	-1
LEACH	10,15	120,87	3,24	20,46	5,32	35,14	1,0	-1
SKATER	4,50	40,81	2,35	17,88	3,18	15,75	1,0	-1
—	0,44	1,16	0,26	1,65	0,07	0,03	1,0	0
LEACH	5,86	40,76	1,97	10,01	3,14	14,33	1,0	0
SKATER	3,40	21,65	2,35	17,88	2,05	9,93	1,0	0
—	0,88	2,14	0,36	1,88	0,11	0,02	1,0	30
LEACH	9,85	153,98	7,83	115,31	3,48	16,10	1,0	30
SKATER	3,12	79,21	11,84	201,76	2,01	10,31	1,0	30
—	0,12	8,12	1,93	1,56	0,03	0,01	5,0	0
LEACH	4,12	55,55	5,89	98,02	1,20	3,47	5,0	0
SKATER	2,55	49,31	7,04	134,41	1,03	2,12	5,0	0
—	0,28	11,26	1,29	2,10	0,09	0,01	5,0	30
LEACH	7,32	133,45	8,78	172,21	2,20	4,76	5,0	30
SKATER	3,50	51,20	297,451	$\approx 87 \times 10^9$	1,69	8,87	5,0	30

Nos cenários avaliados, o modo como a média foi informada não influenciou na estimativa do parâmetro ϕ . O *kriging* bayesiano apresentou resultado qualitativo muito superior

aos demais. O estimador utilizado pelo *kriging* ordinário apresenta desvio muito grande, se comparado aos outros algoritmos. As figuras 4.1, 4.2 e 4.3 apresentam gráficos com os valores da tabela 4.1, agrupados por técnica de agrupamento utilizada.

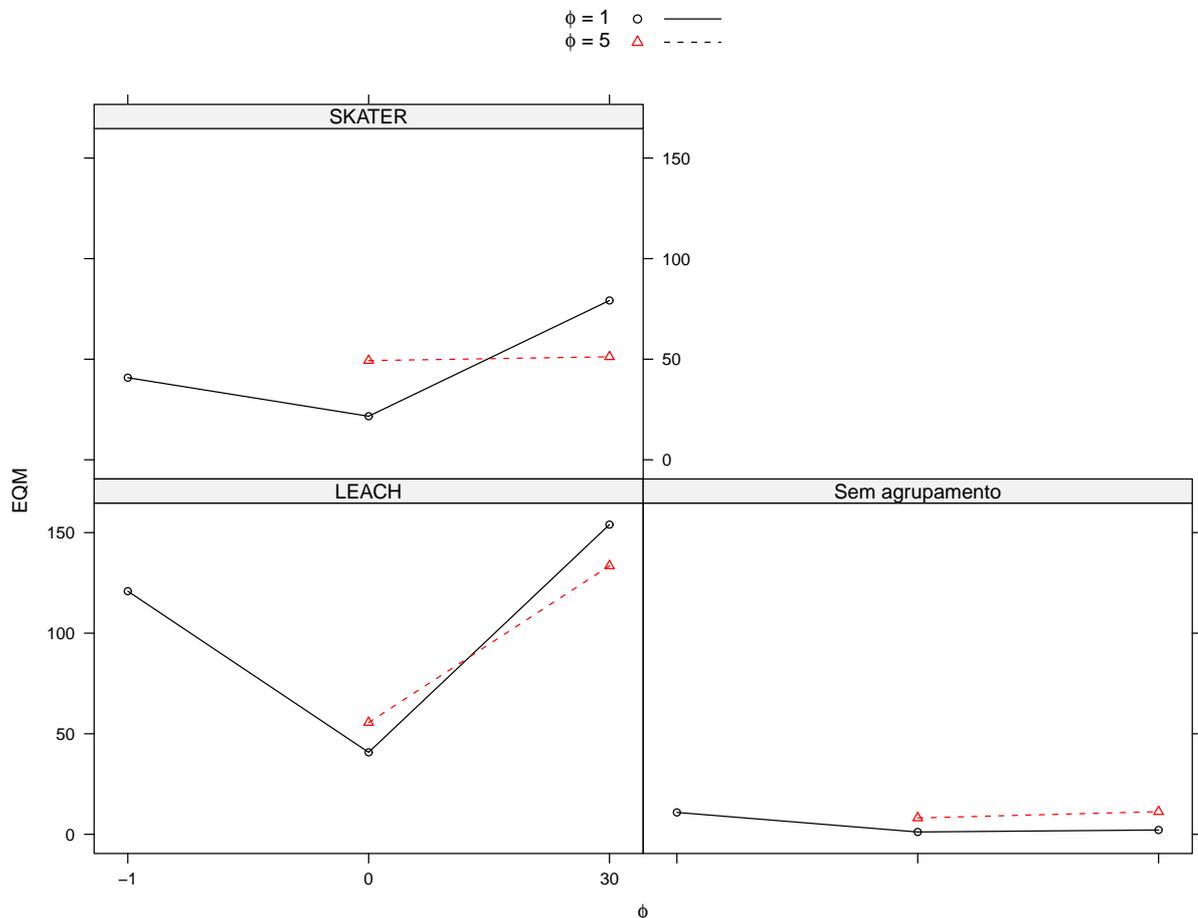


Figura 4.1: Erro quadrático médio do estimador de ϕ pelo *kriging* simples

No caso do algoritmo de agrupamento *LEACH*, a ordem de qualidade se repete, conforme a tabela. As diferenças qualitativas no estimador de ϕ são bastante acentuadas, reforçando a qualidade superior nos dados providos pelo *kriging* bayesiano. Quando não se utiliza algoritmo de agrupamento, o *kriging* bayesiano apresenta estimativas muito próximas do valor verdadeiro do parâmetro nos três cenários avaliados de atratividade para a distribuição dos sensores.

Os resultados do algoritmo de agrupamento *SKATER* são superiores qualitativamente aos do *LEACH*, mas consideravelmente inferiores aos sem agrupamento, à exceção do *kriging* ordinário, que apresenta uma estimativa desviada do valor correto neste cenário. Em testes menores averiguando outros valores para o parâmetro ϕ , o *kriging* ordinário apresenta bom comportamento.

Dentre os três tipos de *kriging* avaliados quanto a este estimador do parâmetro ϕ , o *kriging* bayesiano apresentou os melhores resultados. Estes resultados são melhores ainda ao

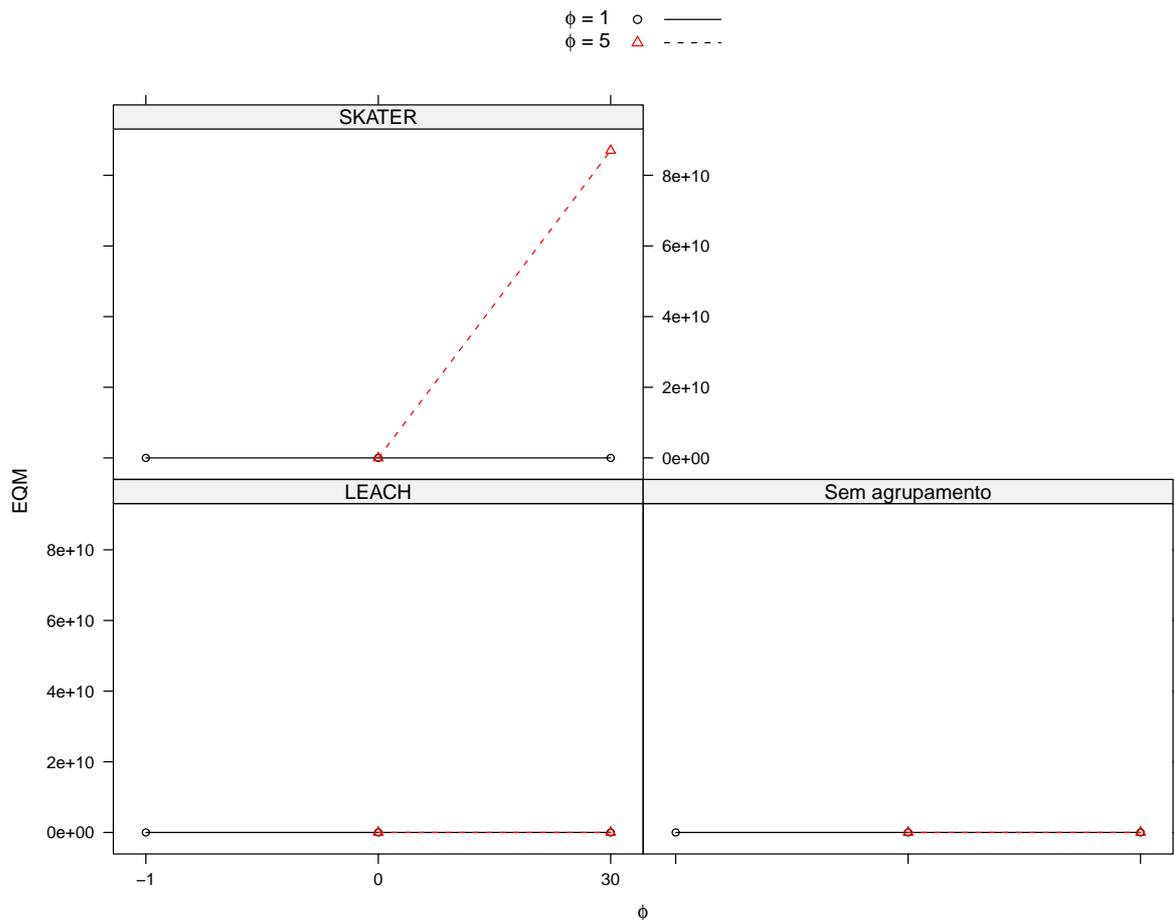


Figura 4.2: Erro quadrático médio do estimador de ϕ pelo *kriging* ordinário

não se utilizar algoritmos de agrupamento. Para o caso com algoritmo de agrupamento, o *SKATER* apresenta os melhores resultados, à exceção do *kriging* ordinário, que apresenta um viés extremamente alto, indicando possíveis problemas do estimador com o cenário avaliado. Note-se que o comportamento do estimador utilizado pelo *kriging* bayesiano sem uso de algoritmo de agrupamento, mesmo sendo inferior aos demais, não apresenta viés muito alto.

Como forma de se obter uma medida unificada de qualidade, utiliza-se o Erro Quadrático Médio (EQM – ver definição 35). Neste, obtém-se uma conceituação que leva em conta o viés e a variância.

Em todos os casos, o estimador utilizado pelo *kriging* bayesiano apresenta resultados qualitativamente muito superiores aos demais algoritmos. O destaque é ainda maior quando não se usa algoritmo de agrupamento, onde o EQM reduz-se para menos de 10^{-2} . À exceção dos resultados sem uso de algoritmo de agrupamento, onde o *kriging* ordinário supera os do *kriging* simples, este apresenta resultados superiores aos daquele. Os resultados do algoritmo *SKATER* são superiores aos do algoritmo *LEACH*, mas bastante inferiores se comparados aos sem uso de algoritmo de agrupamento.

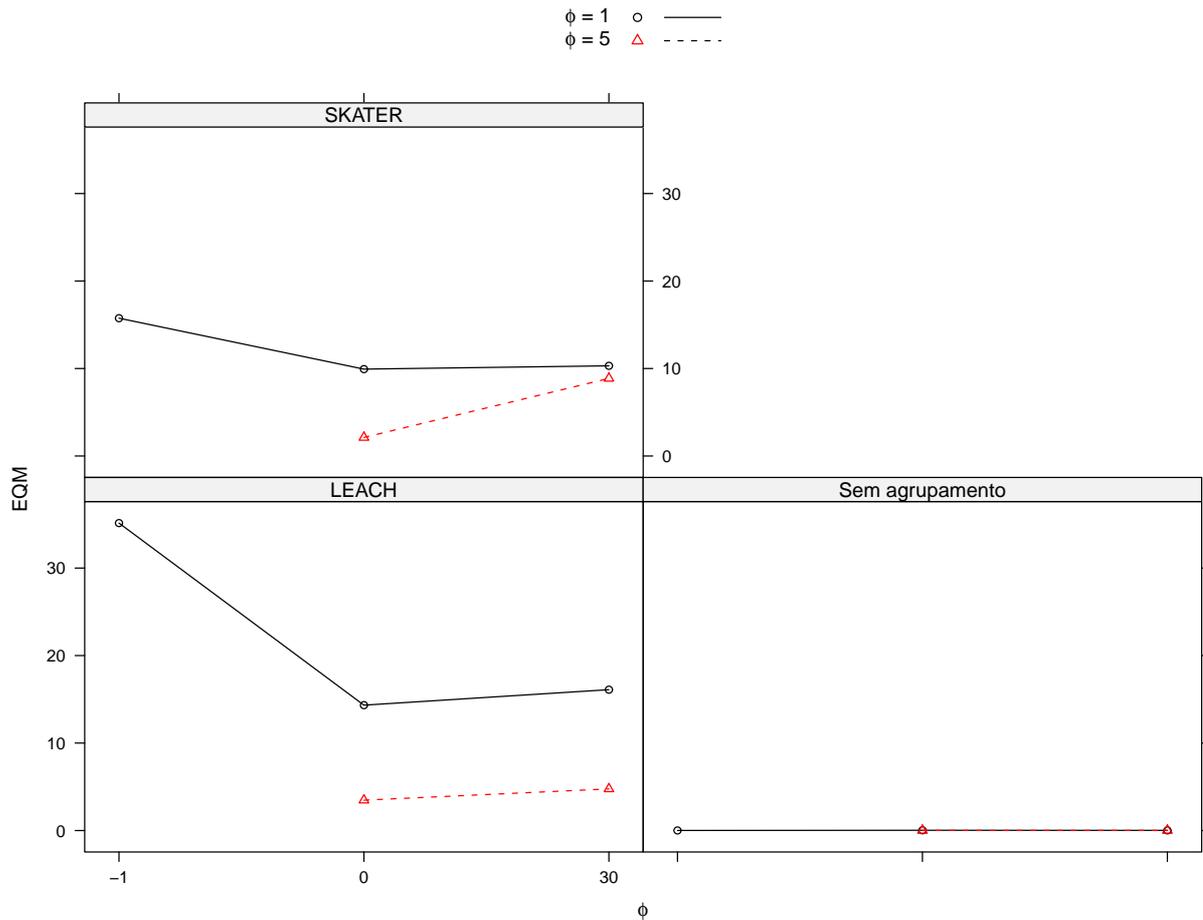


Figura 4.3: Erro quadrático médio do estimador de ϕ pelo *kriging* bayesiano

4.2.2 Análise do campo gaussiano reconstruído

A análise dos campos gaussianos reconstruídos divide-se em duas fases. Na primeira, analisam-se os dados de interpolação provenientes das funções de *kriging*; na outra, os campos são reconstruídos com base nas estimativas dos parâmetros.

A medida de qualidade adotada neste trabalho é a mesma usada por [Alencar-Neto \(2008\)](#): o erro absoluto relativo, descrito como

$$\hat{\epsilon} = \frac{1}{10^4} \sum_{i,j=1}^{100} \left| \frac{g(i,j) - \hat{g}(i,j)}{g(i,j)} \right|, \quad (4.1)$$

onde g representa o campo gaussiano original e \hat{g} o campo gaussiano reconstruído. A equação (4.1) está sempre bem definida, pois a probabilidade de $g(i,j) = 0$ é nula para qualquer (i,j) pela natureza contínua do campo gaussiano.

Campos reconstruídos por *kriging*

A média dos erros absolutos relativos oriundos dos campos reconstruídos pelo processo de *kriging* em cada um dos cenários avaliados estão expostas na tabela 4.2. Exemplos de casos reconstruídos podem ser observados nas figuras 4.10, 4.11 e 4.12. Os valores da tabela 4.2 estão expostos nos gráficos das figuras 4.4 e 4.5 para o *kriging* simples, figuras 4.6 e 4.7 para o *kriging* ordinário e 4.8 e 4.9 para o *kriging* bayesiano.

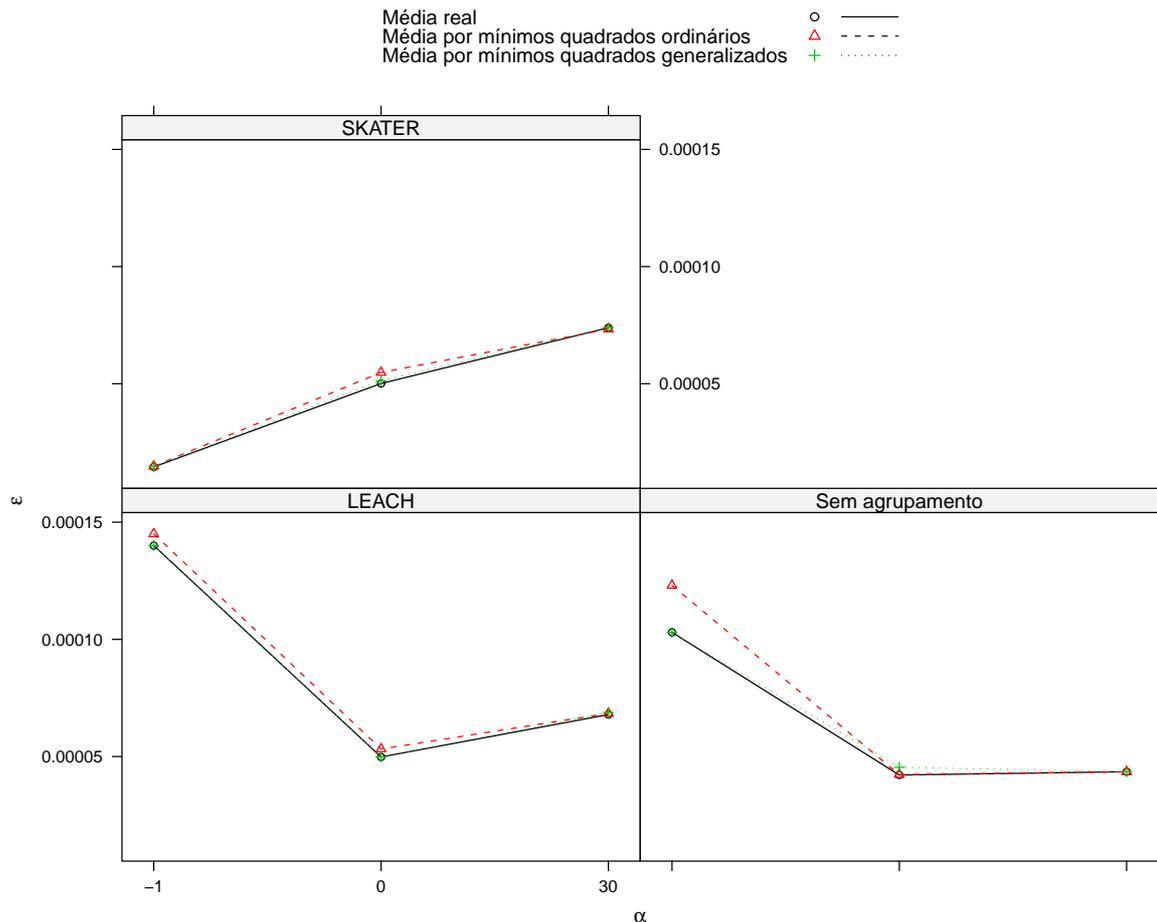


Figura 4.4: Erro de reconstrução por *kriging* simples para $\phi = 1$

Os três tipos de *kriging* apresentam resultados de $\hat{\epsilon}$ próximos. Sem o uso de algoritmo de agrupamento e para o algoritmo *LEACH*, o *kriging* ordinário apresenta os melhores resultados qualitativos no sinal reconstruído. Para o algoritmo *SKATER*, o *kriging* bayesiano apresenta os melhores resultados.

Concordantemente aos resultados anteriores, os menores valores de $\hat{\epsilon}$ são obtidos sem o uso de algoritmo de agrupamento. Dentre os dois algoritmos avaliados, o *LEACH* apresenta os melhores resultados.

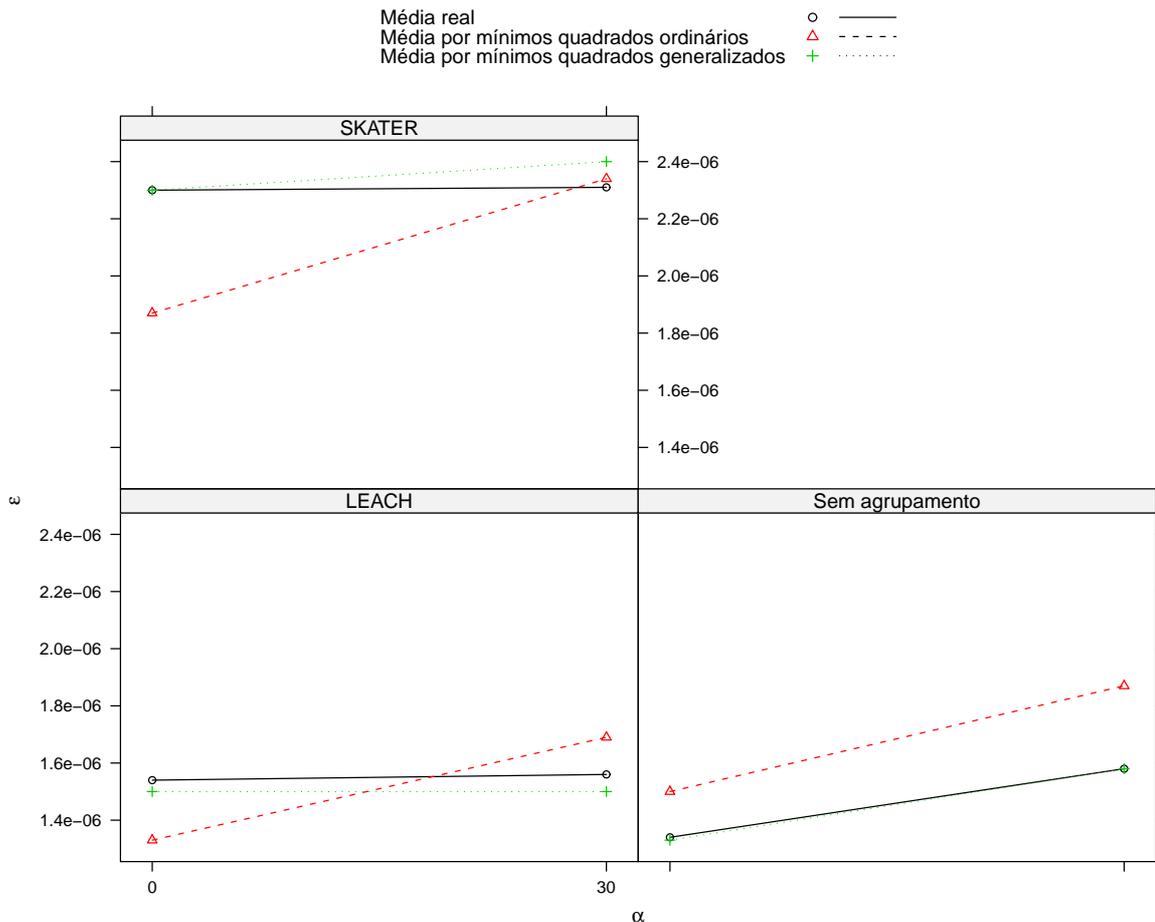


Figura 4.5: Erro de reconstrução por *kriging* simples para $\phi = 5$

Campos reconstruídos a partir dos parâmetros estimados

Com os valores dos parâmetros estimados, o campo gaussiano original foi reconstruído através da mesma função que o gerou, conforme o algoritmo apresentado no código 4.1.

```

1 Entrada: valores dos parametros estimados nas simulacoes
2 Saida: campo reconstruido
3 Para cada simulacao efetuada
4 {
5     campo <- G(theta)
6     estimativaTheta <- RecuperaValorDosParametros(campo)
7     estimativaG <- G(estimativaTheta)
8 }

```

Código 4.1: Reconstrução do campo gaussiano original com base na estimativa

Se g é a função que gera o campo gaussiano com os valores reais do parâmetro (θ), a mesma função com os parâmetros estimados ($\hat{\theta}$) constituirá a função \hat{g} , estimador de $g(\theta)$. Para \hat{g} computam-se as mesmas análises efetuadas nos parâmetros (viés, variância e erro

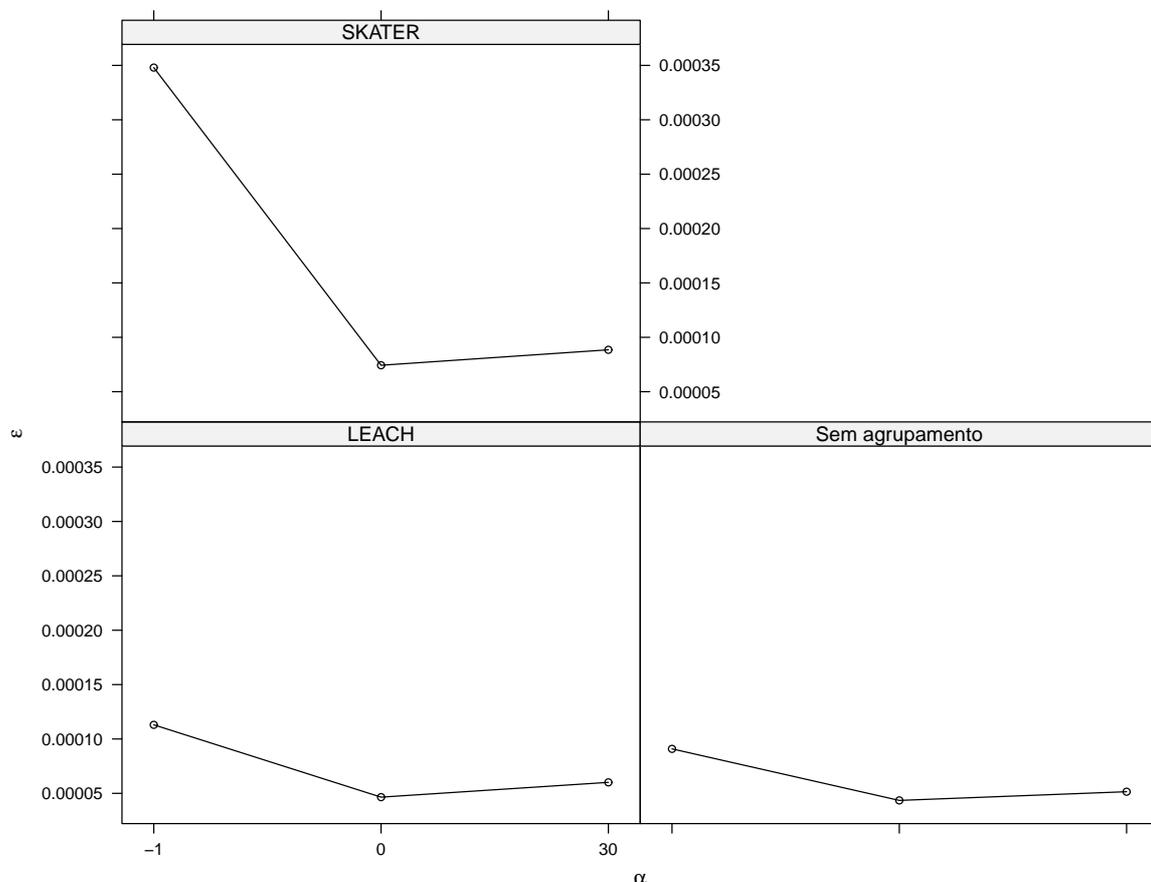


Figura 4.6: Erro de reconstrução por *kriging* ordinário para $\phi = 1$

quadrático médio). O erro absoluto relativo destes campos reconstruídos pode ser observado na tabela 4.3. Os resultados estão nos gráficos das figuras 4.13 e 4.14.

Neste caso os melhores resultados foram obtidos pelo algoritmo de *kriging* ordinário, à exceção do uso do algoritmo de agrupamento *LEACH*, onde o *kriging* simples apresenta os melhores valores do sinal reconstruído. O algoritmo *LEACH* apresentou os melhores resultados nos sinais reconstruídos, à exceção do algoritmo de *kriging* simples, onde o caso em que não se utilizou algoritmo de agrupamento foi qualitativamente superior.

Levando-se em consideração o valor estimado (ou informado) do parâmetro de média para a construção do campo gaussiano, os erros relativos absolutos ocorrem em média conforme denota a tabela 4.4. Graficamente, estes resultados podem ser observados nas figuras 4.16 e 4.17 para o *kriging* simples, 4.18 e 4.19 para o *kriging* ordinário e 4.20 e 4.21 para o *kriging* bayesiano.

Os melhores resultados sem uso de algoritmo de agrupamento foram obtidos pelo *kriging* ordinário e pelo bayesiano. O algoritmo *LEACH* obteve resultados qualitativamente superiores aos do *SKATER*, especialmente para o *kriging* simples, que obteve os melhores resultados para este algoritmo de agrupamento. No que tange ao algoritmo de agrupamento

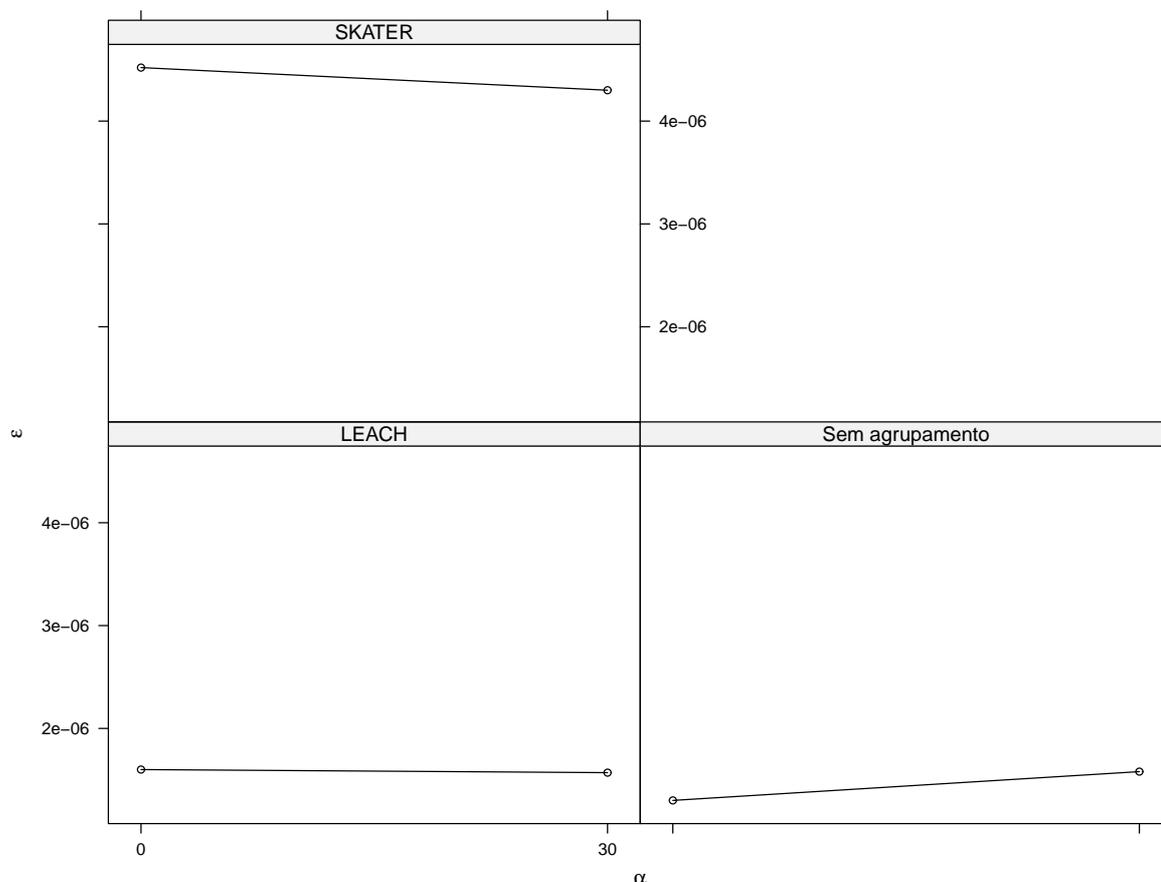


Figura 4.7: Erro de reconstrução por *kriging* ordinário para $\phi = 5$

SKATER, os melhores resultados foram obtidos pelo *kriging* bayesiano.

4.3 Análise do tempo necessário à reconstrução do sinal

Redes de Sensores sem Fios são utilizadas nos mais diversos cenários e, dentre diversos outros, muitos envolvem detecção de riscos iminentes, bem como incêndios em florestas, concentração de gases tóxicos (Akyildiz et al., 2002; Sha et al., 2006; Lloret et al., 2009). Nessas aplicações, o tempo necessário à reconstrução do sinal é crucial, dado que a função do sistema é o combate ao fator, que em pouco tempo se propaga atingindo proporções cada vez maiores.

Assim sendo, não basta que a técnica utilizada para a reconstrução do sinal seja eficiente na qualidade dos dados, mas que seja eficaz no tempo de reconstrução. Este problema está referenciado em diversos trabalhos, dentre os quais, os de Sha et al. (2006); Lloret et al. (2009); Fierens (2009) e Tsow et al. (2009).

É importante ressaltar que isso se dá a depender dos recursos computacionais. Mesmo que atualmente não haja alternativa disponível, é possível que em alguns anos, com o au-

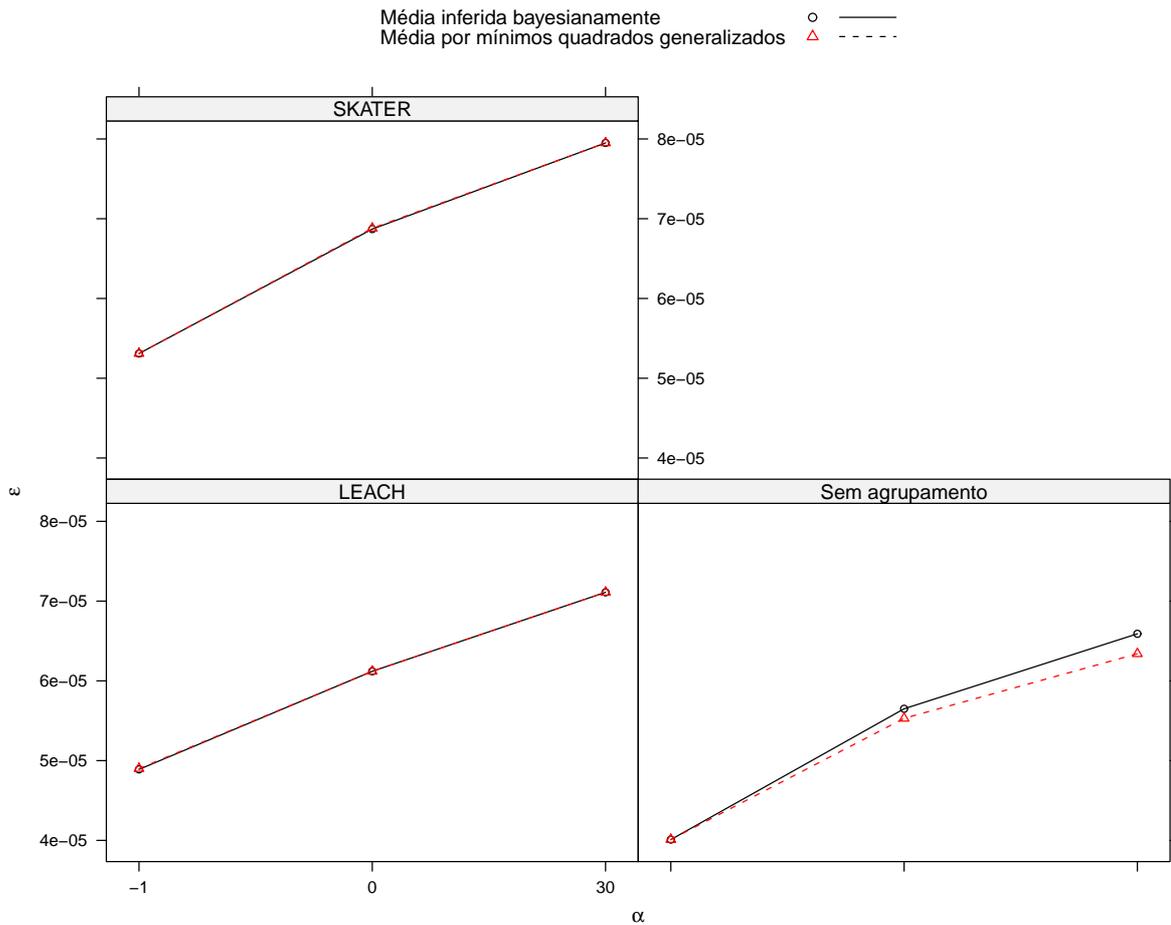


Figura 4.8: Erro de reconstrução por *kriging* bayesiano para $\phi = 1$

mento do poder computacional disponível, esse tópico seja preterido.

4.4 Considerações acerca dos dados obtidos

Em virtude da alta demanda computacional requerida pelo processo de *kriging* bayesiano, uma vez que a regressão de um modelo gaussiano sob inferência bayesiana possui complexidade computacional de $O(n^3)$, com n sendo a quantidade de pontos a serem estimados e os cenários neste trabalho simulados possuírem 100×100 pontos (Zhu et al., 1997; Gaudard et al., 1999), foram realizadas simulações de poucos cenários. Pelo mesmo motivo, não foi executada uma quantidade mais expressiva de iterações da simulação de Monte Carlo. Entretanto, encontra-se em andamento, na data de publicação deste trabalho, a simulação de diversos outros cenários que farão parte de um trabalho resultante desta pesquisa, a ser publicado.

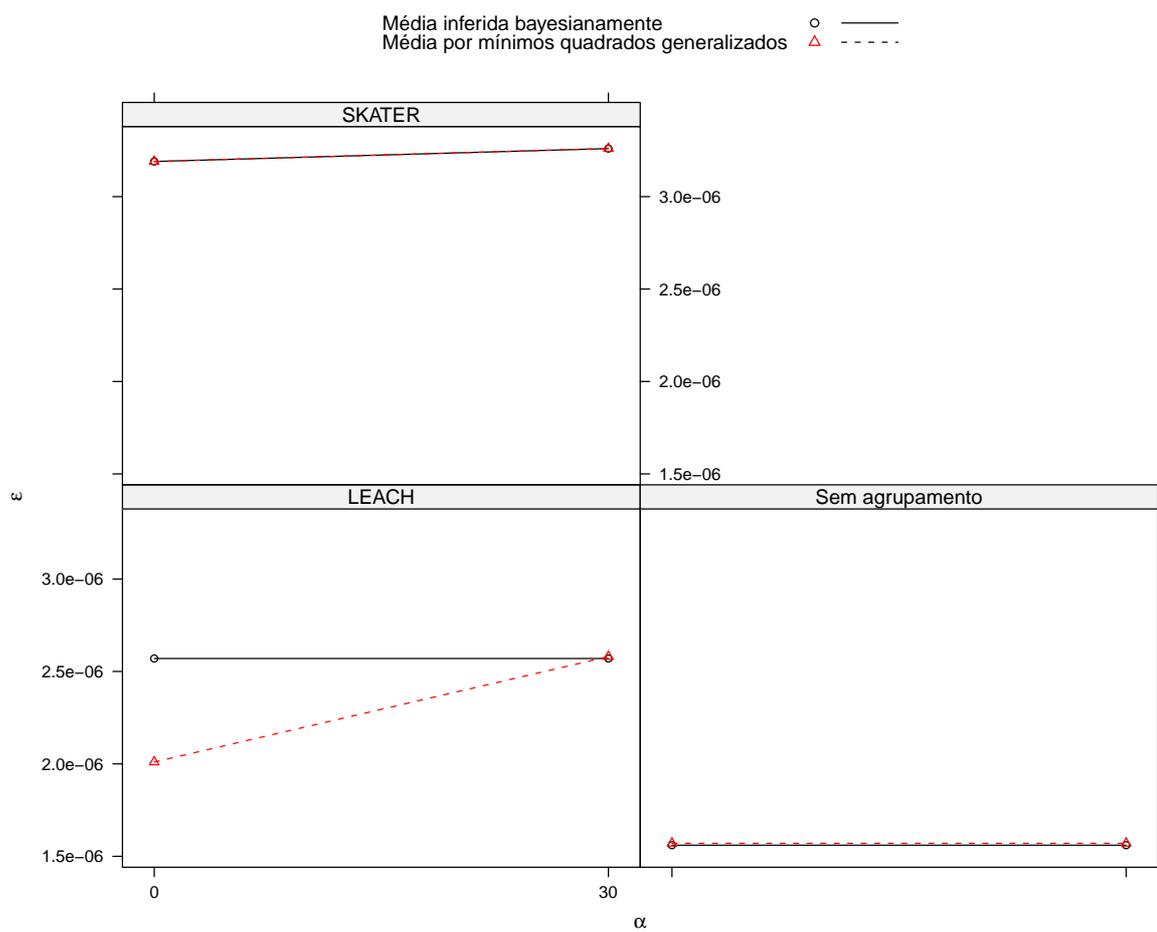
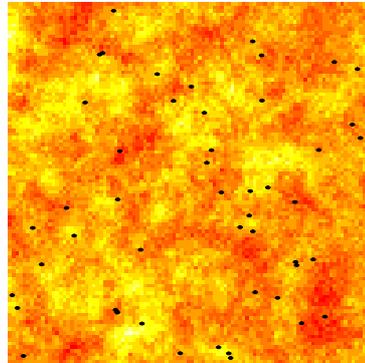
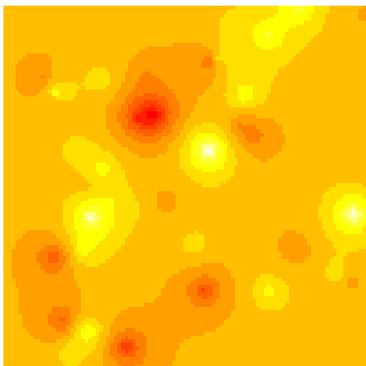
Figura 4.9: Erro de reconstrução por *kriging* bayesiano para $\phi = 5$

Tabela 4.2: Média dos erros absolutos relativos dos campos reconstruídos pelo processo de *kriging* com $\kappa = 0,5$

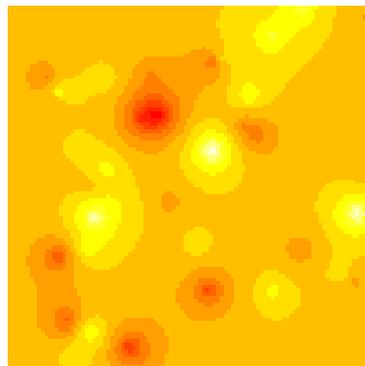
<i>Kriging Simples com média:</i>	<i>Sem cluster</i>	<i>LEACH</i>	<i>SKATER</i>	ϕ	α
Informada	1,23e-04	1,45e-04	1,48e-05	1,0	-1
Mínimos Quadrados Ordinários	1,03e-04	1,40e-04	1,47e-05	1,0	-1
Mínimos Quadrados Generalizados	1,03e-04	1,40e-04	1,45e-05	1,0	-1
Informada	4,23e-05	5,32e-05	5,48e-05	1,0	0
Mínimos Quadrados Ordinários	4,54e-05	4,98e-05	5,13e-05	1,0	0
Mínimos Quadrados Generalizados	4,21e-05	4,98e-05	5,01e-05	1,0	0
Informada	4,35e-05	6,84e-05	7,33e-05	1,0	30
Mínimos Quadrados Ordinários	4,35e-05	6,88e-05	7,40e-05	1,0	30
Mínimos Quadrados Generalizados	4,35e-05	6,79e-05	7,39e-05	1,0	30
Informada	1,50e-06	1,33e-06	1,87e-06	5,0	0
Mínimos Quadrados Ordinários	1,33e-06	1,50e-06	2,30e-06	5,0	0
Mínimos Quadrados Generalizados	1,34e-06	1,54e-06	2,30e-06	5,0	0
Informada	1,87e-06	1,69e-06	2,34e-06	5,0	30
Mínimos Quadrados Ordinários	1,58e-06	1,50e-06	2,40e-06	5,0	30
Mínimos Quadrados Generalizados	1,58e-06	1,56e-06	2,31e-06	5,0	30
<i>Kriging Ordinário com média:</i>					
Inferida	9,09e-05	1,13e-04	3,48e-04	1,0	-1
Inferida	4,35e-05	4,65e-05	7,43e-05	1,0	0
Inferida	5,15e-05	6,01e-05	8,85e-05	1,0	30
Inferida	1,30e-06	1,60e-06	4,52e-06	5,0	0
Inferida	1,58e-06	1,57e-06	4,30e-06	5,0	30
<i>Kriging Bayesiano com média:</i>					
Inferida	4,01e-05	4,90e-05	5,31e-05	1,0	-1
Mínimos Quadrados Generalizados	4,01e-05	4,89e-05	5,31e-05	1,0	-1
Inferida	5,53e-05	6,12e-05	6,88e-05	1,0	0
Mínimos Quadrados Generalizados	5,65e-05	6,12e-05	6,87e-05	1,0	0
Inferida	6,34e-05	7,11e-05	7,95e-05	1,0	30
Mínimos Quadrados Generalizados	6,59e-05	7,11e-05	7,95e-05	1,0	30
Inferida	1,57e-06	2,01e-06	3,19e-06	5,0	0
Mínimos Quadrados Generalizados	1,56e-06	2,57e-06	3,19e-06	5,0	0
Inferida	1,57e-06	2,58e-06	3,26e-06	5,0	30
Mínimos Quadrados Generalizados	1,56e-06	2,57e-06	3,26e-06	5,0	30



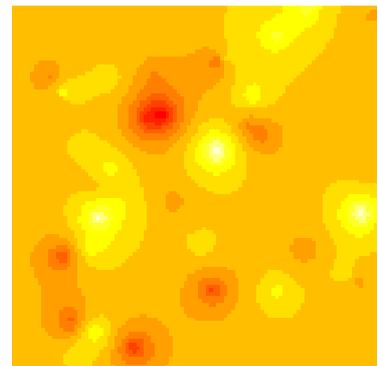
(a) Campo gaussiano original com sensores distribuídos



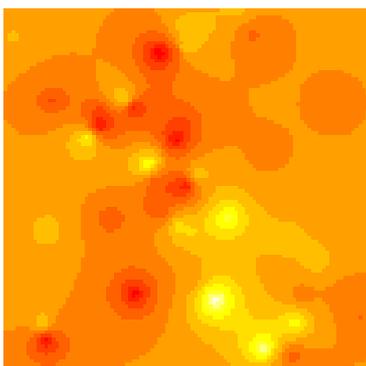
(b) *Kriging* simples com média real



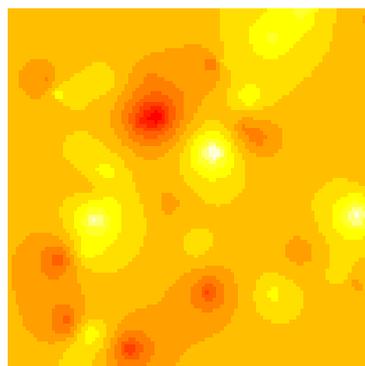
(c) *Kriging* simples com média por mínimos quadrados ordinários



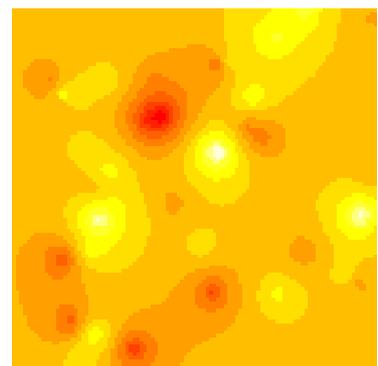
(d) *Kriging* simples com média por mínimos quadrados generalizados



(e) *Kriging* ordinário



(f) *Kriging* bayesiano



(g) *Kriging* bayesiano com média por mínimos quadrados generalizados

Figura 4.10: Exemplo de uma replicação do processo de reconstrução por *kriging* para 100 nós sensores com $\kappa = 0,5$ e $\phi = 5$, sem agrupamentos, com $\alpha = 0$

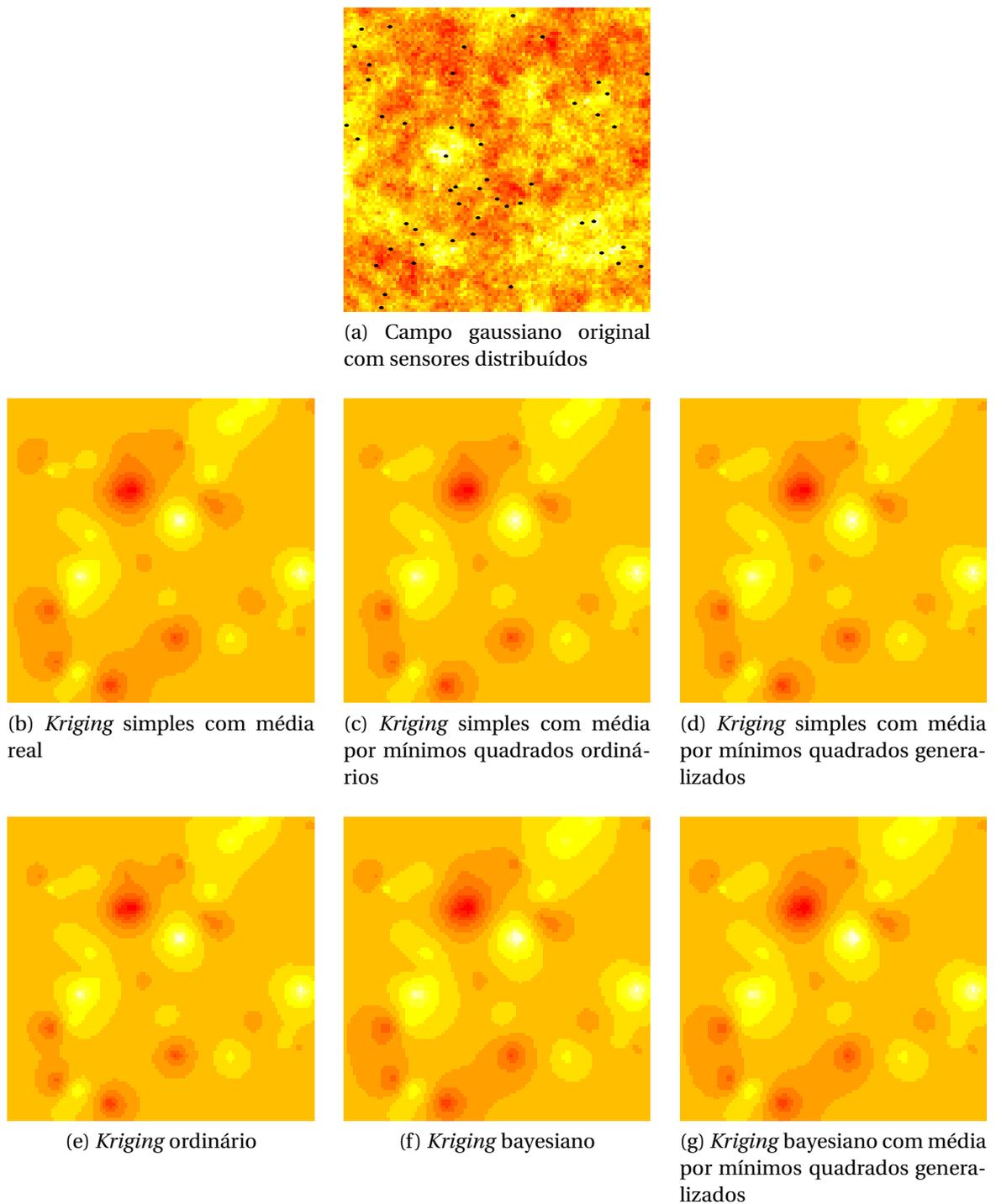
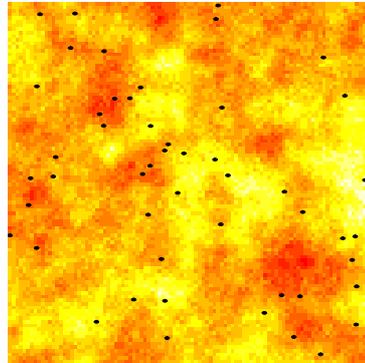
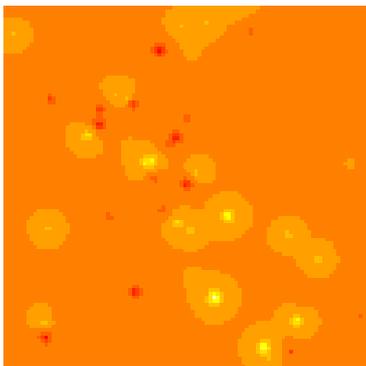


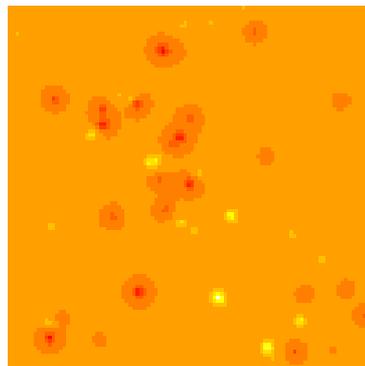
Figura 4.11: Exemplo de uma replicação do processo de reconstrução por *kriging* para 100 nós sensores e 5 grupos com $\kappa = 0,5$ e $\phi = 5$, com agrupamento por *LEACH* e $\alpha = 30$



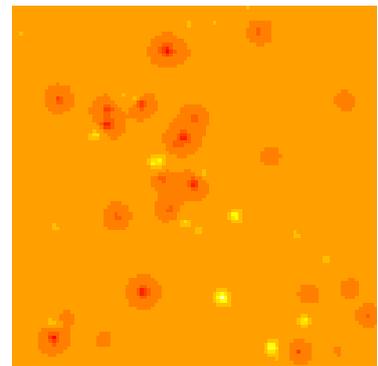
(a) Campo gaussiano original com sensores distribuídos



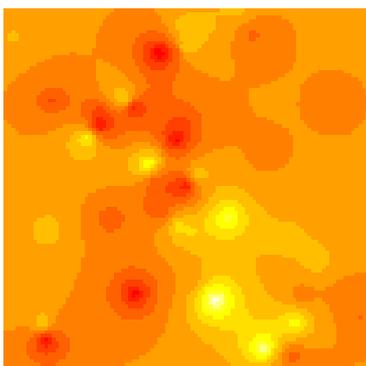
(b) *Kriging* simples com média real



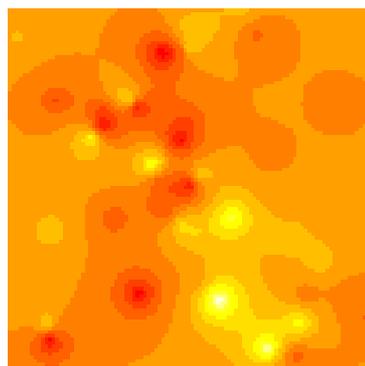
(c) *Kriging* simples com média por mínimos quadrados ordinários



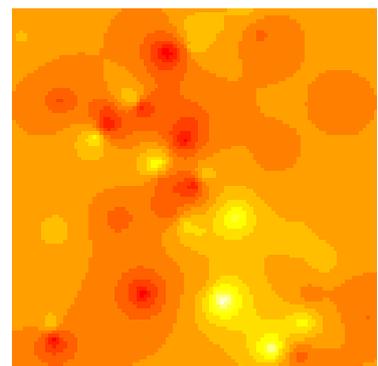
(d) *Kriging* simples com média por mínimos quadrados generalizados



(e) *Kriging* ordinário



(f) *Kriging* bayesiano

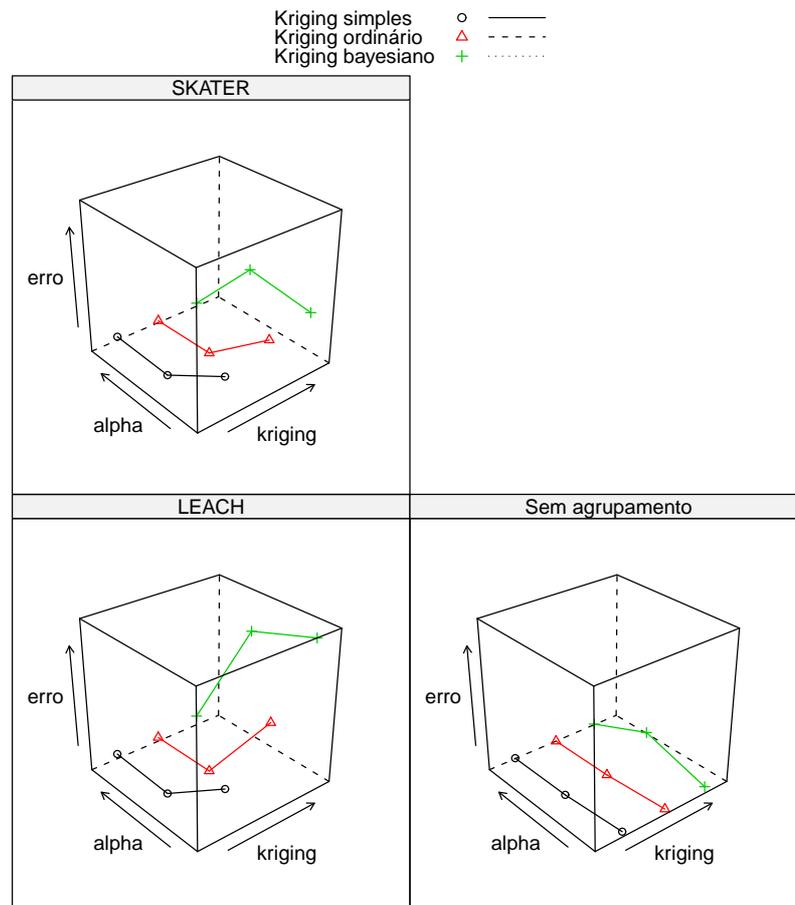


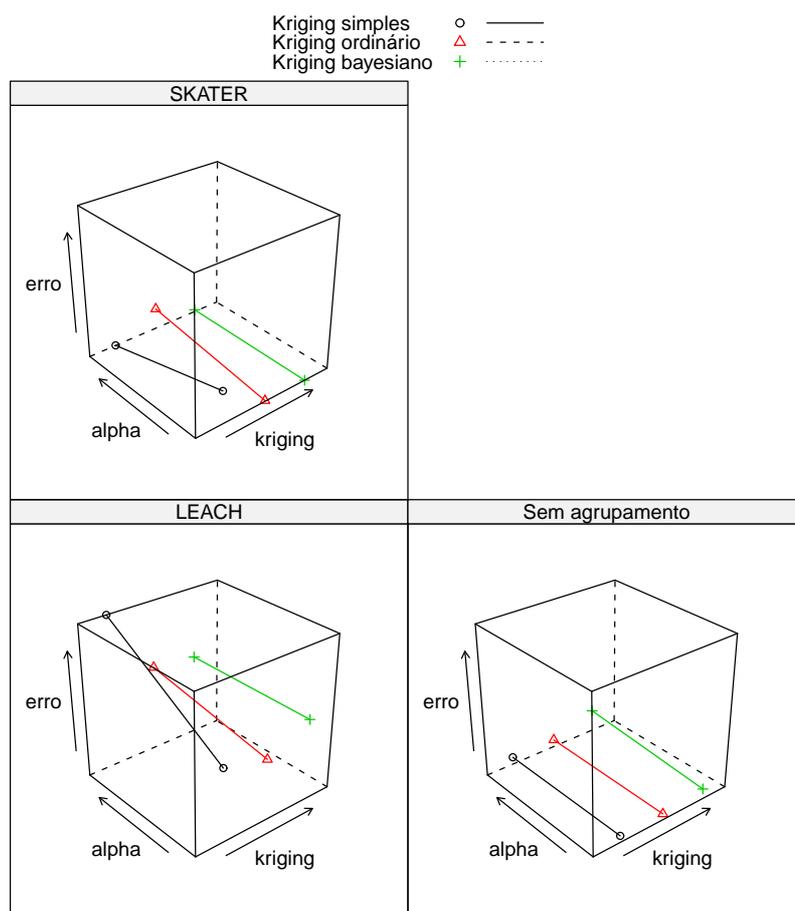
(g) *Kriging* bayesiano com média por mínimos quadrados generalizados

Figura 4.12: Exemplo de uma replicação do processo de reconstrução por *kriging* para 100 nós sensores e 5 grupos com $\kappa = 0,5$ e $\phi = 1$, com agrupamento por *SKATER* e $\alpha = 0$

Tabela 4.3: Média dos erros absolutos relativos dos campos reconstruídos pelos parâmetros inferidos para $\kappa = 0,5$

<i>Kriging</i>	<i>Sem cluster</i>	<i>LEACH</i>	<i>SKATER</i>	ϕ	α
Simples	1,34e-05	2,34e-04	1,01e-04	1,0	-1
Ordinário	1,12e-05	1,42e-04	8,96e-05	1,0	-1
Bayesiano	1,10e-05	7,62e-05	6,71e-05	1,0	-1
Simples	4,35e-05	2,01e-04	1,15e-04	1,0	0
Ordinário	3,34e-06	9,97e-06	9,07e-06	1,0	0
Bayesiano	3,12e-06	5,32e-06	5,11e-06	1,0	0
Simples	3,35e-06	1,76e-05	7,75e-06	1,0	30
Ordinário	2,34e-06	8,72e-06	5,63e-06	1,0	30
Bayesiano	3,12e-06	1,02e-05	8,15e-06	1,0	30
Simples	2,34e-07	1,77e-06	6,97e-09	5,0	0
Ordinário	1,44e-07	1,34e-06	2,34e-08	5,0	0
Bayesiano	1,55e-07	1,60e-06	7,35e-07	5,0	0
Simples	4,87e-07	1,77e-06	4,89e-08	5,0	30
Ordinário	1,69e-07	1,87e-06	4,73e-07	5,0	30
Bayesiano	1,69e-07	3,30e-06	9,06e-09	5,0	30

Figura 4.13: Erro de reconstrução por parâmetros inferidos para $\phi = 1$

Figura 4.14: Erro de reconstrução por parâmetros inferidos para $\phi = 5$

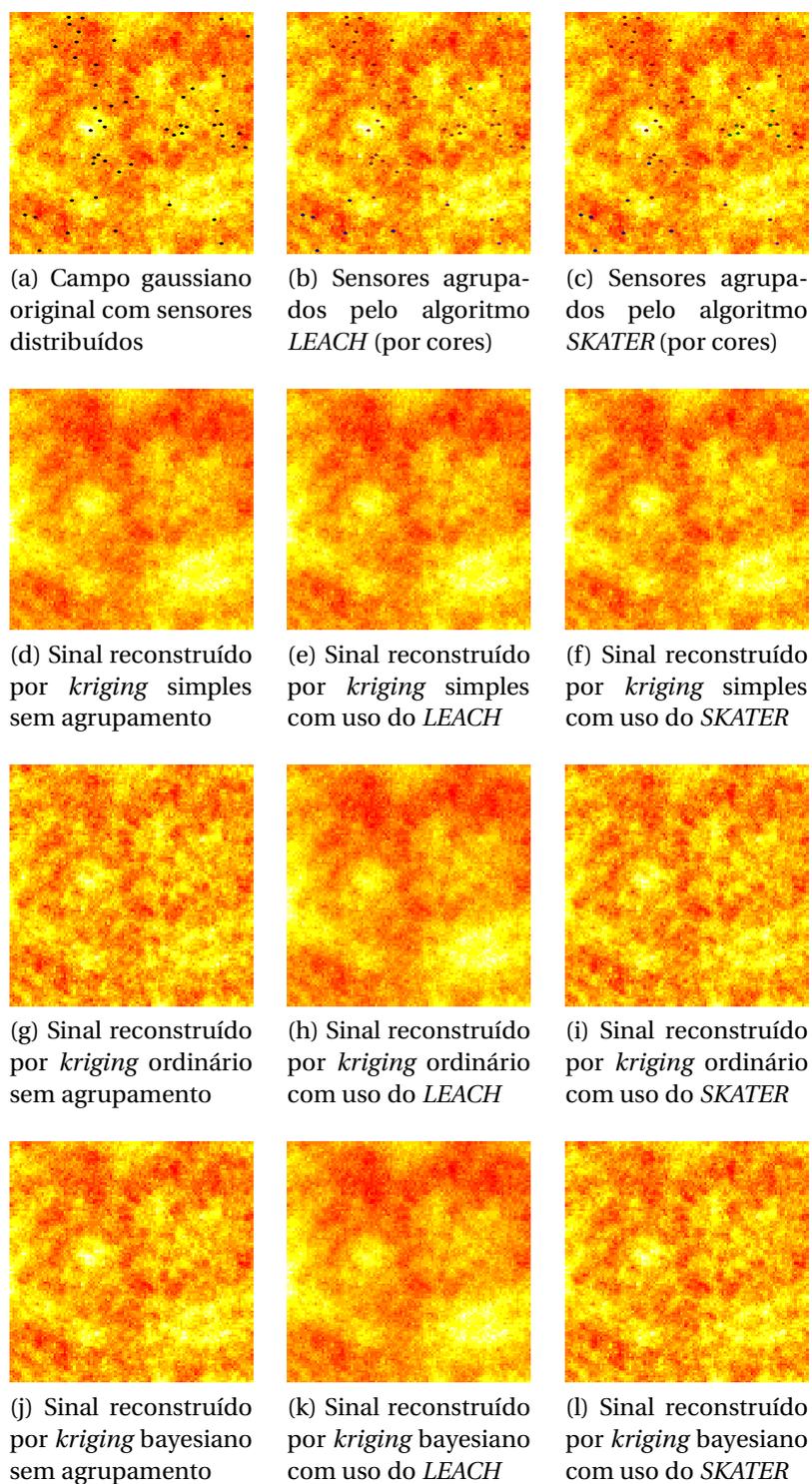


Figura 4.15: Exemplo de uma replicação do processo de reconstrução utilizando os parâmetros estimados para 100 nós sensores e 5 grupos com $\kappa = 0,5$ e $\phi = 5$

Tabela 4.4: Média dos erros absolutos relativos dos campos reconstruídos pelos parâmetros inferidos para $\kappa = 0,5$ com média

<i>Kriging Simples com média:</i>	<i>Sem cluster</i>	<i>LEACH</i>	<i>SKATER</i>	ϕ	α
Mínimos Quadrados Ordinários	7,14e-08	5,78e-06	6,87e-07	1,0	0
Mínimos Quadrados Generalizados	4,34e-08	3,86e-06	4,01e-07	1,0	0
Mínimos Quadrados Ordinários	5,35e-08	1,43e-06	1,88e-07	1,0	0
Mínimos Quadrados Generalizados	4,01e-08	9,79e-07	1,13e-07	1,0	0
Mínimos Quadrados Ordinários	3,87e-08	1,83e-07	5,58e-06	5,0	30
Mínimos Quadrados Generalizados	1,34e-08	1,85e-07	4,34e-06	5,0	30
Mínimos Quadrados Ordinários	4,74e-08	1,15e-06	8,00e-07	5,0	0
Mínimos Quadrados Generalizados	2,44e-07	1,49e-06	7,69e-07	5,0	0
Mínimos Quadrados Ordinários	4,75e-08	1,55e-06	8,08e-07	5,0	30
Mínimos Quadrados Generalizados	2,47e-07	1,55e-06	7,74e-07	5,0	30
<i>Kriging Ordinário com média:</i>					
Inferida	1,11e-07	1,69e-06	6,32e-07	1,0	-1
Inferida	1,20e-07	1,86e-06	7,11e-07	1,0	0
Inferida	1,77e-07	2,33e-06	7,57e-07	1,0	30
Inferida	1,70e-07	1,24e-06	7,09e-07	5,0	0
Inferida	1,67e-07	1,86e-06	8,08e-07	5,0	30
<i>Kriging Bayesiano com média:</i>					
Inferida	1,66e-07	2,01e-06	1,91e-07	1,0	-1
Mínimos Quadrados Generalizados	1,65e-07	2,01e-06	1,13e-06	1,0	-1
Inferida	1,59e-07	1,98e-06	1,86e-07	1,0	0
Mínimos Quadrados Generalizados	1,88e-07	1,98e-06	1,23e-06	1,0	0
Inferida	1,78e-07	2,21e-06	2,04e-07	1,0	30
Mínimos Quadrados Generalizados	1,78e-07	2,21e-06	1,57e-06	1,0	30
Inferida	1,59e-07	2,40e-06	1,67e-07	5,0	0
Mínimos Quadrados Generalizados	1,67e-07	3,40e-06	1,01e-06	5,0	0
Inferida	1,67e-07	3,30e-06	1,87e-07	5,0	30
Mínimos Quadrados Generalizados	1,67e-07	3,30e-06	1,37e-06	5,0	30

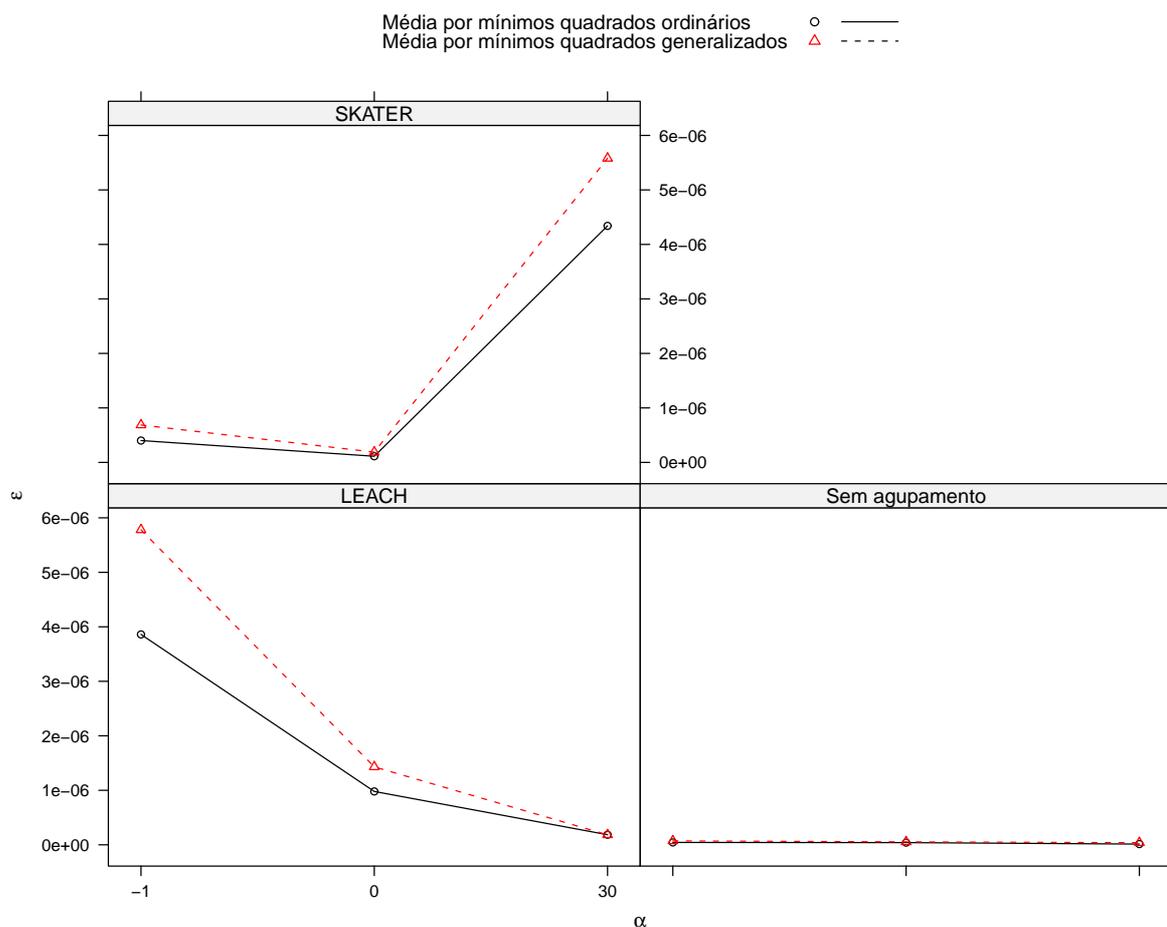


Figura 4.16: Erro de reconstrução por *kriging* simples pelos parâmetros considerando a média para $\phi = 1$

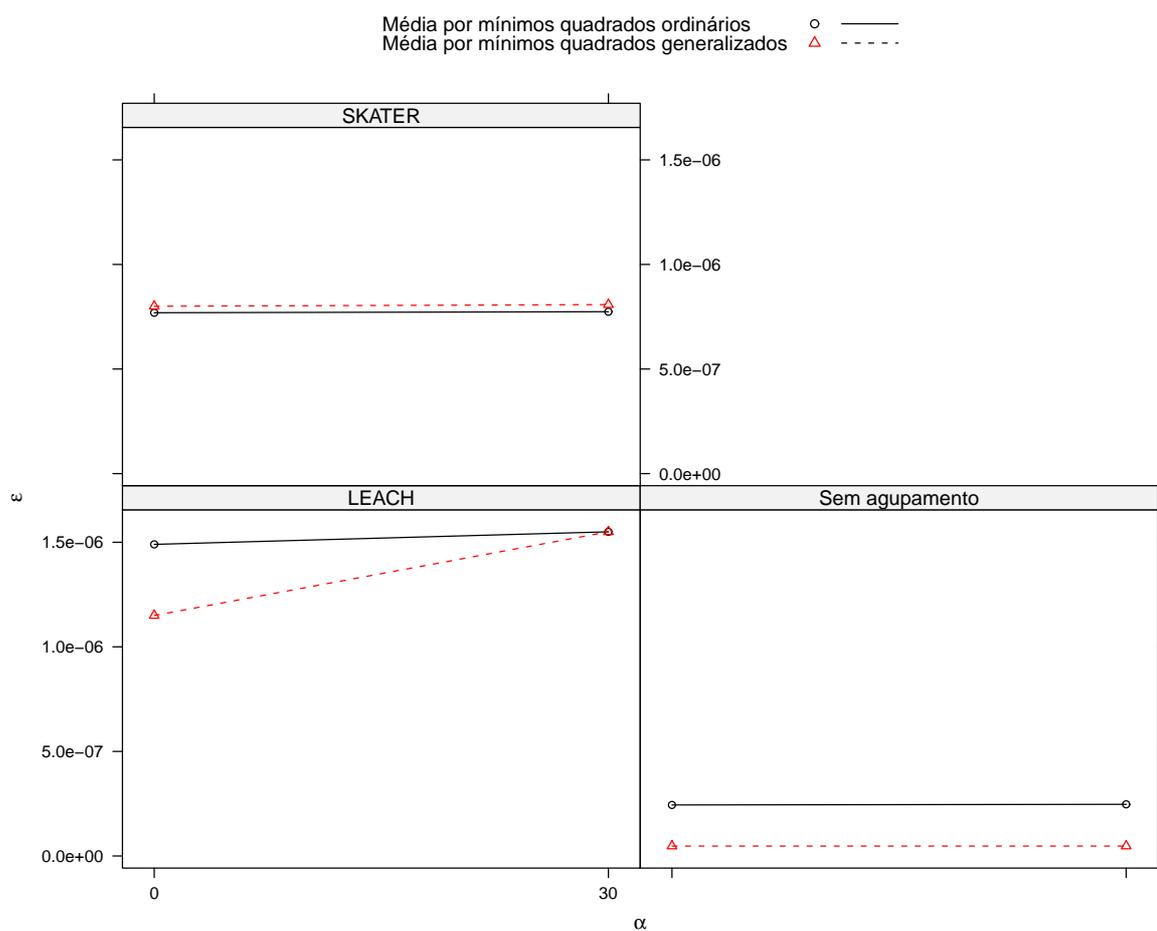


Figura 4.17: Erro de reconstrução por *kriging* simples pelos parâmetros considerando a média para $\phi = 5$

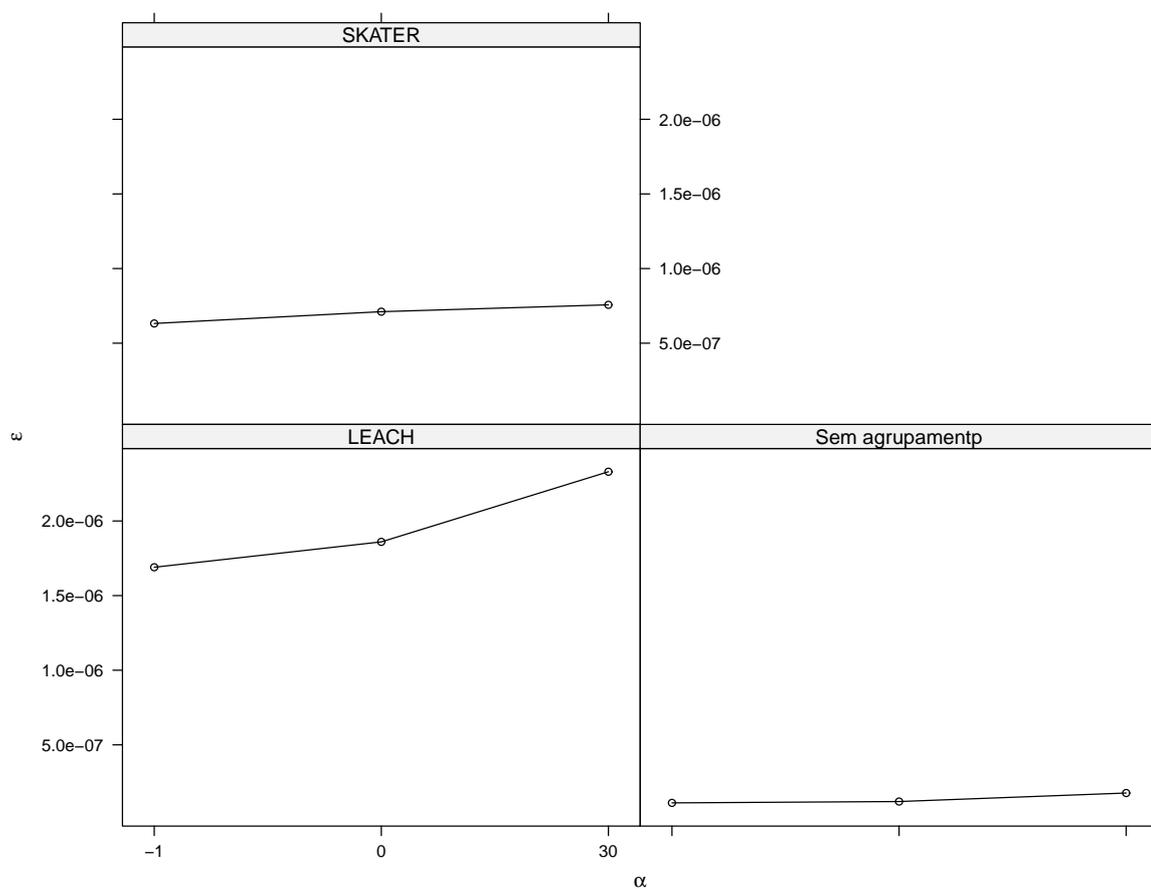


Figura 4.18: Erro de reconstrução por *kriging* ordinário pelos parâmetros considerando a média para $\phi = 1$

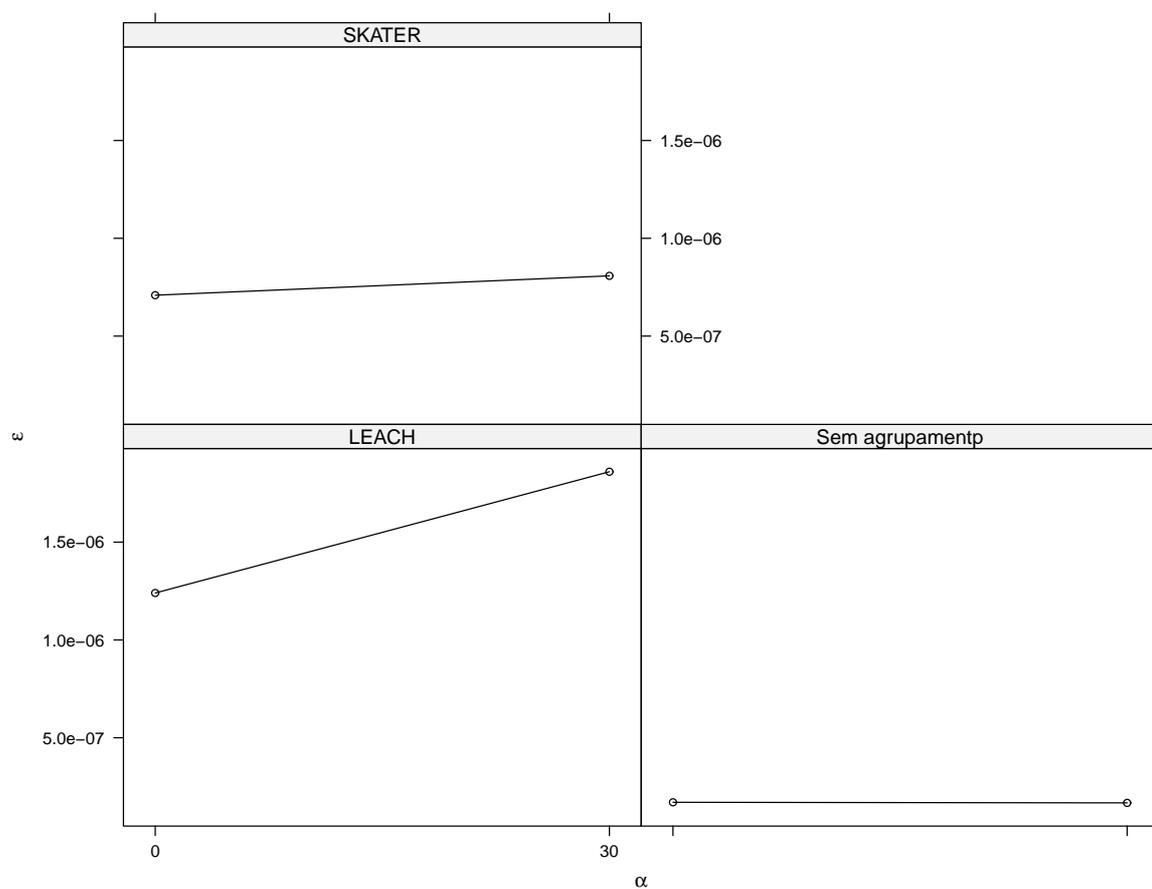


Figura 4.19: Erro de reconstrução por *kriging* ordinário pelos parâmetros considerando a média para $\phi = 5$

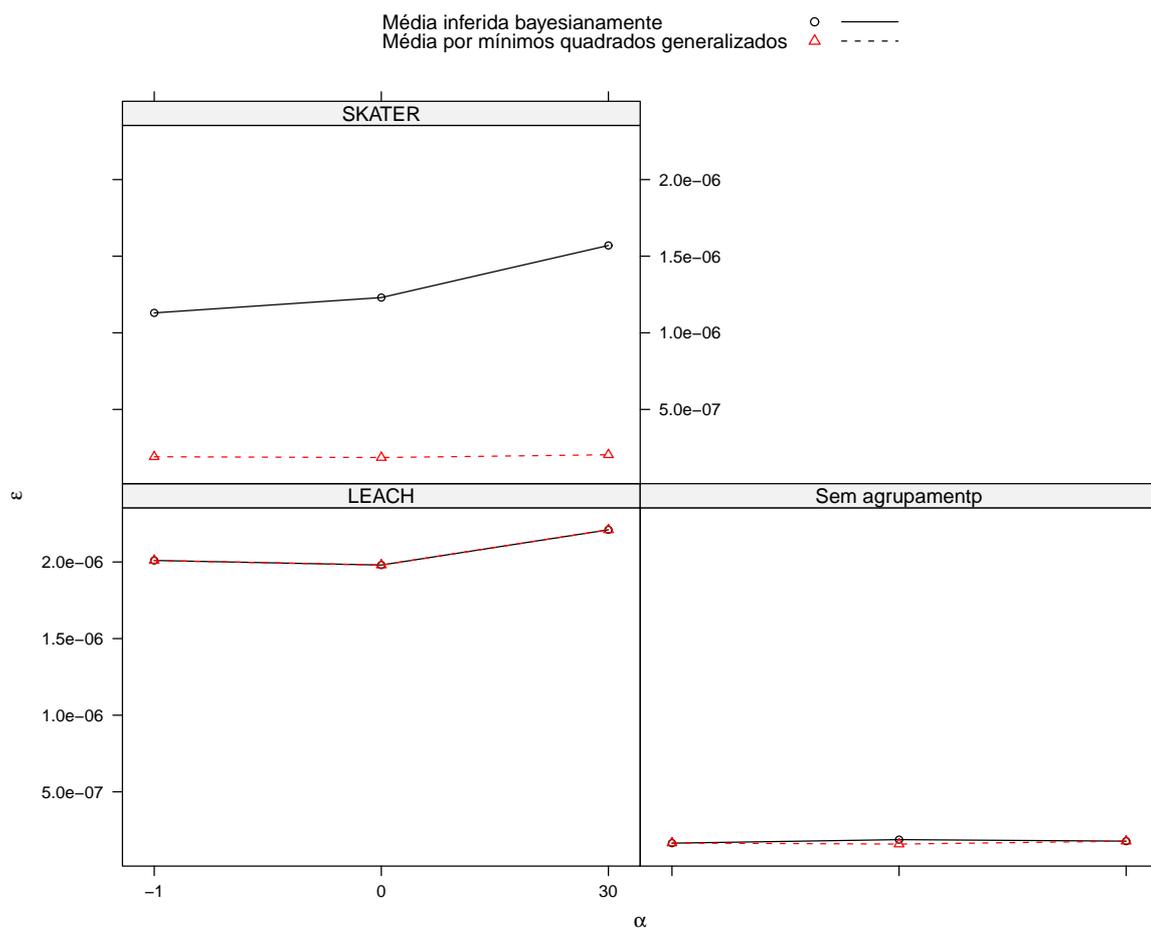


Figura 4.20: Erro de reconstrução por *kriging* bayesiano pelos parâmetros considerando a média para $\phi = 1$

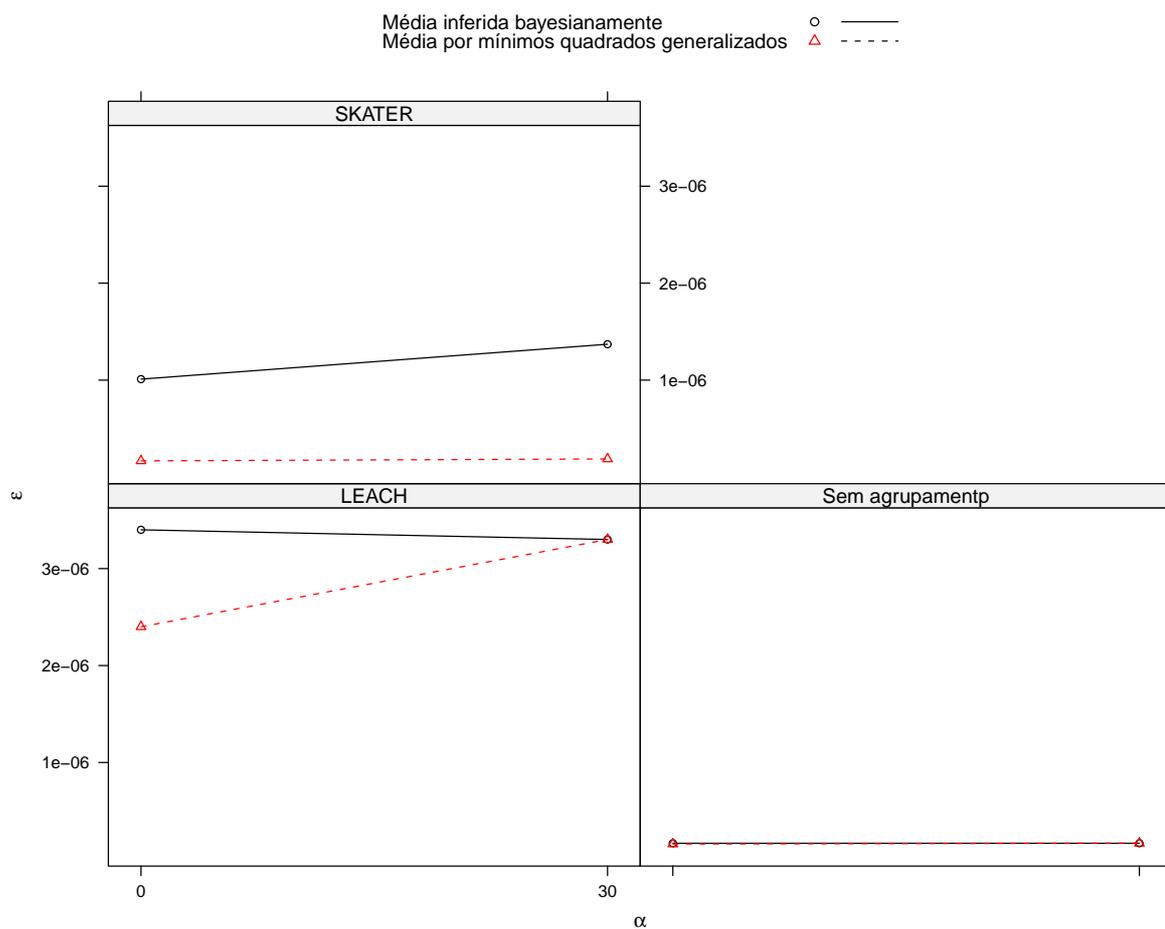


Figura 4.21: Erro de reconstrução por *kriging* bayesiano pelos parâmetros considerando a média para $\phi = 5$



Resultados e discussões

“It was simple, but you know, it’s always simple when you’ve done it.”

Simone Gabriellini

R-help (agosto de 2005)

A análise dos dados do capítulo 4 permitiu observar a maior qualidade dos dados gerados pelo algoritmo de *kriging* bayesiano na maioria dos cenários avaliados. Entretanto ele não apresenta bom comportamento com o algoritmo *LEACH*. Neste cenário o *kriging* simples apresentou os melhores resultados.

Algo a se considerar é que nem sempre sua utilização é viável em decorrência de sua necessidade de alto poder computacional (ver seção 4.3). À exemplo, sistemas de tempo real não são suscetíveis a computações de longos períodos de tempo; é necessário que a reação seja imediata à solicitação.

O sinal reconstruído diretamente pelos algoritmos de *kriging* apresenta maior qualidade que os gerados pela mesma função utilizada para a construção dos campos gaussianos de amostras, informando as estimativas dos parâmetros ϕ e κ . Esta última situação, à exceção da em que se utiliza o *kriging* bayesiano, necessita de menor tempo de computação, uma vez que a interpolação não necessita ser efetuada, basta estimar os parâmetros e informá-los à função que gera campos gaussianos.

O *kriging* que necessita de menor tempo para estimação dos parâmetros e reconstrução do campo gaussiano é o ordinário. Os parâmetros κ e ϕ estimados não apresentam resultados superiores ao *kriging* simples, mas os resultados da reconstrução dos campos gaussianos é superior ao *kriging* simples em todos os cenários avaliados que não o do algoritmo *LEACH*.

O *kriging* ordinário, então, apresenta os melhores resultados na relação qualidade dos parâmetros estimados / tempo de execução, uma vez que cada reconstrução levou em torno

de três horas, numa das estações especificadas no capítulo 3. Mesmo os resultados do *kriging* bayesiano sendo em geral superiores qualitativamente, ele só pode ser utilizado quando houver grande poder computacional e/ou não houver necessidade de resultados rápidos.

Para o *kriging* bayesiano foi avaliada uma variante proposta neste trabalho, que consiste na inclusão da informação *a priori* do parâmetro de média, transformada numa constante para o sistema de reconstrução. Essa abordagem possibilitou redução no tempo necessário de processamento, uma vez que a computação da média deixa de ser por inferência bayesiana e passa a ser efetuada por mínimos quadrados generalizados. Trata-se de uma alternativa minimizadora dos problemas decorrentes do tempo necessário à computação do *kriging* bayesiano; entretanto, ainda assim ele não se torna adequado a sistemas que necessitem de respostas rápidas. Qualitativamente só se apresentaram diferenças significativas ao *kriging* bayesiano convencional na reconstrução por parâmetros inferidos com o uso da média estimada, onde essa versão foi superada por esta em uma ordem de grandeza.

Uma tarefa a ser desenvolvida em trabalhos futuros e a análise de uma diversidade maior de cenários de simulação, que já se encontra em desenvolvimento no LCCAN. Para que os resultados sejam ainda mais realistas, trabalhos vindouros devem avaliar métricas de distâncias não euclidianas (mais detalhes no trabalho de Diggle & Ribeiro Jr., 2007) ou ainda a execução de algoritmos de *kriging* em espaços do \mathbb{R}^3 , levando em conta relevos acidentados que podem interferir na comunicação.

Simulações mais realistas devem constar de modelagens para o raio de comunicação, que neste trabalho foi considerado como circular (aproximação mais utilizada na literatura), mas que na verdade vai depender dos diversos obstáculos que houverem no local onde os nós sensores forem depositos. Comparações com outras técnicas de reconstrução não derivadas do trabalho de Krige (1951) são também de grande valia, bem como técnicas que sigam o modelo proposto por Baraniuk et al. (2010).

Referências bibliográficas

- Abbasi, A. A. & Younis, M. (2007), 'A survey on clustering algorithms for wireless sensor networks', *Computer Communications* **30**(14–15), 2826–2841. DOI [10.1016/j.comcom.2007.05.024](https://doi.org/10.1016/j.comcom.2007.05.024). ISSN 0140-3664.
- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. & Cyirci, E. (2002), 'Wireless sensor networks: A survey', *Computer Networks* **38**(4), 393–422. DOI [10.1016/S1389-1286\(01\)00302-4](https://doi.org/10.1016/S1389-1286(01)00302-4). ISSN 1389-1286.
- Alencar-Neto, J. (2008), Análise do erro em redes de sensores sem fio: A influência da distribuição espacial de nós nas técnicas de agrupamento, Dissertação de mestrado, Universidade Federal de Alagoas, Maceió, AL, Brasil.
- Almiron, M. G., Almeida, E. S. & Miranda, M. N. (2009), 'The reliability of statistical functions in four software packages freely used in numerical computation', *Brazilian Journal of Probability and Statistics* **23**(2), 107–119. DOI [10.1214/08-BJPS017](https://doi.org/10.1214/08-BJPS017). ISSN 0103-0752.
- Almiron, M. G., Lopes, B., Oliveira, A. L. C., Medeiros, A. C. & Frery, A. C. (2010), 'On the numerical accuracy of spreadsheets', *Journal of Statistical Software* **34**(4), 1–29. ISSN 1548-7660. URL <http://www.jstatsoft.org/v34/i04>.
- Altman, M. & McDonald, M. P. (2001), 'Choosing reliable statistical software', *Ps-Political Science & Politics* **34**(3), 681–687. ISSN 1049-0965.
- Altman, M., Gill, J. & McDonald, M. P. (2007), 'Accuracy: Tools for accurate and reliable statistical computing', *Journal of Statistical Software*. ISSN 1548-7660. URL <http://www.jstatsoft.org/v21/i01>.
- Andel, T. R. & Yasinac, A. (2006), 'On the credibility of manet simulations', *Computer* **39**(7), 48–54. DOI <http://dx.doi.org/10.1109/MC.2006.242>. ISSN 0018-9162.
- Aquino, A. L. L., Figueiredo, C. M. S., Nakamura, E. F., Frery, A. C., Loureiro, A. A. F. & Fernandes, A. O. (2008), Sensor stream reduction for clustered wireless sensor networks, in 'ACM Symposium on Applied Computing', ACM, Fortaleza, CE, Brasil.

- Assunção, R. M., Neves, M. C., Câmara, G. & Freitas, C. C. (2006), 'Efficient regionalisation techniques for socio-economic geographical units using minimum spanning trees', *International Journal of Geographical Information Science* **20**(7), 797–811.
- Baddeley, A. (2006), Spatial point process and their applications, in 'Lecture Notes in Mathematics', Vol. 1892 of *Stochastic Geometry*, pp. 1–75. DOI [10.1007/3-540-38174-0](https://doi.org/10.1007/3-540-38174-0). ISSN 1617-9692.
- Baddeley, A. & Turner, R. (2005), 'Spatstat: an R package for analyzing spatial point patterns', *Journal of Statistical Software* **12**(6), 1–42. ISSN 1548-7660. URL www.jstatsoft.org.
- Baraniuk, R., Cevher, V., Duarte, M. & Hegde, C. (2010), 'Model-based compressive sensing', *IEEE Transactions on Information Theory* **56**(4), 1982–2001. DOI [10.1109/TIT.2010.2040894](https://doi.org/10.1109/TIT.2010.2040894). ISSN 0018-9448.
- Bivand, R. S. & Pebesma, E. J. (2008), *Applied Spatial Data Analysis with R*, Use R!, Springer.
- Bustos, O. H. & Frery, A. C. (2006), 'Statistical functions and procedures in IDL 5.6 and 6.0', *Computational Statistics & Data Analysis* **50**(2), 301–310. DOI [10.1016/j.csda.2004.08.011](https://doi.org/10.1016/j.csda.2004.08.011). ISSN 0167-9473.
- Callaert, H. (2003), 'Comparing statistical software packages: The case of the logrank test in statxact', *American Statistician* **57**(3), 214–217. DOI [10.1198/0003130031900](https://doi.org/10.1198/0003130031900). ISSN 0003-1305.
- Callaghan, B. (2000), *NFS Illustrated*, Addison-Wesley Professional Computing Series, Addison-Wesley.
- Camilli, A., Cugnasca, C. E., Saraiva, A. M., Hirakawa, A. R. & Correa, P. L. P. (2007), 'From wireless sensors to field mapping: Anatomy of an application for precision agriculture', *Computers and Electronics in Agriculture* **58**(1), 25–36. DOI [10.1016/j.compag.2007.01.019](https://doi.org/10.1016/j.compag.2007.01.019). ISSN 0168-1699.
- Chan, G. & Wood, A. T. A. (1997), 'AS 312 – an algorithm for simulating stationary Gaussian random fields', *Applied Statistics – Journal of The Royal Statistical Society Series C* **46**(1), 171–181. DOI [10.1111/1467-9876.00057](https://doi.org/10.1111/1467-9876.00057). ISSN 0035-9254.
- Csardi, G. & Nepusz, T. (2006), 'The igraph software package for complex network research', *InterJournal Complex Systems*, 1695. URL <http://igraph.sf.net>.
- Diggle, P. J. & Ribeiro Jr., P. J. (2007), *Model-based Geostatistics*, Springer Series in Statistics, Springer.
- European Spreadsheet Risks Interest Group (2010), 'European spreadsheet risks interest group'. URL <http://www.eusprig.org>, última consulta em abril de 2010.

- Fierens, P. I. (2009), 'Number of sensors versus time to detection in wildfires', *International Journal of Wildland Fire* **18**(7), 825–829. DOI [10.1071/WF07137](https://doi.org/10.1071/WF07137). ISSN 1049-8001.
- Frery, A. C., Ramos, H., Alencar-Neto, J. & Nakamura, E. F. (2008), Error estimation in wireless sensor networks, in 'ACM Symposium on Applied Computing', ACM, Fortaleza, CE, Brasil.
- Frery, A. C., Ramos, H., Alencar-Neto, J., Nakamura, E. & Loureiro, A. A. F. (2010), 'Data driven performance evaluation of wireless sensor networks', *Sensors* **10**(3), 2150–2168. DOI [10.3390/s100302150](https://doi.org/10.3390/s100302150). ISSN 1424-8220.
- Gabriel, E., Fagg, G. E., Bosilca, G., Angskun, T., Dongarra, J. J., Squyres, J. M., Sahay, V., Kambadur, P., Barrett, B., Lumsdaine, A., Castain, R. H., Daniel, D. J., Graham, R. L. & Woodall, T. S. (2004), Open MPI: Goals, concept, and design of a next generation MPI implementation, in 'Proceedings, 11th European PVM/MPI Users' Group Meeting', Budapest, Hungria, pp. 97–104. ISSN 0302-9743.
- Gaudard, M., Karson, M., Linder, E. & Sinha, D. (1999), 'Bayesian spatial prediction', *Environmental and Ecological Statistics* **6**(2), 147–171. DOI [10.1023/A:1009614003692](https://doi.org/10.1023/A:1009614003692). ISSN 1573-3009.
- Gneiting, T. (1997), Symmetric Positive Definite Functions with Applications in Spatial Statistics, Tese de doutorado, Universitat Bayreuth.
- Handcock, M. S. & Wallis, J. R. (1994), 'An approach to statistical spatial temporal modeling of meteorological fields (with discussion)', *Journal of the American Statistical Association* **89**, 368–390.
- Heinzelman, W. B., Chandrakasan, A. & Balakrishnan, H. (2002), 'An application-specific protocol architecture for wireless microsensor networks', *IEEE Transactions on Wireless Communication* **1**, 660–670. DOI [10.1109/TWC.2002.804190](https://doi.org/10.1109/TWC.2002.804190).
- Heinzelman, W. R., Chandrakasan, A. & Balakrishnan, H. (2000), Energy-efficient communication protocol for wireless microsensor networks, in 'Proceedings of the Hawaii International Conference on System Sciences'.
- Intel Corporation (2008), First the tick, now the tock: Next generation Intel[®] microarchitecture (nehalem), Relatório técnico, Intel Corporation.
- James, B. R. (2006), *Probabilidade: um curso em nível intermediário*, Projeto Euclides, 3 ed., IMPA, Rio de Janeiro, RJ, Brasil.
- Jedermann, R. & Lang, W. (2009), The minimum number of sensors — interpolation of spatial temperature profiles in chilled transports, in 'EWSN '09: Proceedings of the 6th

- European Conference on Wireless Sensor Networks', Springer-Verlag, Berlin, Heidelberg, pp. 232–246. DOI 10.1007/978-3-642-00224-3_15.
- Kennedy, W. J. & Gentle, J. E. (1980), *Statistical Computing*, Marcel Dekker, New York, EUA.
- Koenker, R. & Zeileis, A. (2009), 'On reproducible econometric research', *Journal of Applied Econometrics* **24**, 833–847. DOI 10.1002/jae.1083.
- Krige, D. G. (1951), 'A statistical approach to some basic mine valuation problems on the Witwatersrand', *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52**, 119–139. ISSN 0038-223X.
- Lantuejoul, C. (2002), *Geostatistical simulation*, Springer, Nova York.
- Le, N. D. & Zidek, J. V. (2006), *Statistical Analysis of Environmental Space-Time Processes*, Springer Series in Statistics, Springer.
- Lehtinen, S. & Lonvick, C. (2006), 'Request for comments: 4250'. URL <http://tools.ietf.org/html/rfc4250>.
- Li, H. & Low, B. K. (2010), 'Reliability analysis of circular tunnel under hydrostatic stress field', *Computers and Geotechnics* **37**(1-2), 50–58. DOI 10.1016/j.compgeo.2009.07.005. ISSN 0266-352X.
- Lindqvist, R. & Westoo, A. (2000), 'Quantitative risk assessment for listeria monocytogenes in smoked or gravad salmon and rainbow trout in sweden', *International Journal of Food Microbiology* **58**(3), 181–196. DOI 10.1016/j.scitotenv.2009.09.053. ISSN 0168-1605.
- Lloret, J., Garcia, M., Bri, D. & Sendra, S. (2009), 'A wireless sensor network deployment for rural and forest fire detection and verification', *Sensors* **9**(11), 8722–8747. DOI 10.3390/s91108722. ISSN 1424-8220.
- Matsumoto, M. & Nishimura, T. (1998), 'Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator', *ACM Transactions on Modeling and Computer Simulation* **8**(1), 3–30. DOI 10.1145/272991.272995. ISSN 1049-3301.
- McCullough, B. D. (1998), 'Assessing the reliability of statistical software: Part I', *American Statistician* **52**(4), 358–366. ISSN 0003-1305.
- McCullough, B. D. (1999), 'Assessing the reliability of statistical software: Part II', *American Statistician* **53**(2), 149–159. ISSN 0003-1305.

- McCullough, B. D. (2004), 'Fixing statistical errors in spreadsheet software: the cases of gnumeric and Excel', *Computational Statistics & Data Analysis Statistical Software Newsletter* pp. 1–10. URL <http://www.csdassn.org/reportdetail.cfm?ID=508>.
- McCullough, B. D. (2008a), 'Microsoft Excel's 'Not The Wichmann-Hill' random number generators', *Computational Statistics & Data Analysis* **52**(10), 4587–4593. DOI [10.1016/j.csda.2008.03.006](https://doi.org/10.1016/j.csda.2008.03.006). ISSN 0167-9473.
- McCullough, B. D. (2008b), 'Special section on Microsoft Excel 2007', *Computational Statistics & Data Analysis* **52**(10), 4568–4569. DOI [10.1016/j.csda.2008.03.009](https://doi.org/10.1016/j.csda.2008.03.009). ISSN 0167-9473.
- McCullough, B. D. & Heiser, D. A. (2008), 'On the accuracy of statistical procedures in Microsoft Excel 2007', *Computational Statistics & Data Analysis* **52**(10), 4570–4578. DOI [10.1016/j.csda.2008.03.004](https://doi.org/10.1016/j.csda.2008.03.004). ISSN 0167-9473.
- McCullough, B. D. & Wilson, B. (1999), 'On the accuracy of statistical procedures in Microsoft Excel 97', *Computational Statistics & Data Analysis* **31**(1), 27–37. DOI [10.1016/S0167-9473\(99\)00004-3](https://doi.org/10.1016/S0167-9473(99)00004-3). ISSN 0167-9473.
- McCullough, B. D. & Wilson, B. (2002), 'On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP', *Computational Statistics & Data Analysis* **40**(4), 713–721. DOI [10.1016/S0167-9473\(02\)00095-6](https://doi.org/10.1016/S0167-9473(02)00095-6). ISSN 0167-9473.
- McCullough, B. D. & Wilson, B. (2005), 'On the accuracy of statistical procedures in Microsoft Excel 2003', *Computational Statistics & Data Analysis* **49**(4), 1244–1252. DOI [10.1016/j.csda.2004.06.016](https://doi.org/10.1016/j.csda.2004.06.016). ISSN 0167-9473.
- Metropolis, N. & Ulam, S. (1949), 'The Monte Carlo method', *Journal of the American Statistical Association* **44**(247), 335–341. ISSN 0162-1459.
- Michaelson, J. (2004), 'There's no such thing as a free (software) lunch', *Queue* **2**(3), 40–47. DOI <http://doi.acm.org/10.1145/1005062.1005066>. ISSN 1542-7730.
- Murray, F. J. (1953), 'Some applications of the Monte Carlo method', *Journal of the Operations Research Society of America* **1**(2), 75.
- Nakamura, E. F., Loureiro, A. A. F. & Frery, A. C. (2007), 'Information fusion for wireless sensor networks: Methods, models, and classifications', *ACM Computing Surveys* **39**(3), 9/1–55. DOI [10.1145/1267070.1267073](https://doi.org/10.1145/1267070.1267073).
- Nash, J. C. (2006), 'Spreadsheets in statistical practice - another look', *American Statistician* **60**(3), 287–289. DOI [10.1198/000313006X126585](https://doi.org/10.1198/000313006X126585). ISSN 0003-1305.

- Nash, J. C. (2008), 'Teaching statistics with Excel 2007 and other spreadsheets', *Computational Statistics & Data Analysis* **52**(10), 4602–4606. DOI [10.1016/j.csda.2008.03.008](https://doi.org/10.1016/j.csda.2008.03.008). ISSN 0167-9473.
- Ni, K., Ramanathan, N., Chehade, M. N. H., Balzano, L., Nair, S., Kohler, S. Z. E., Pottie, G., Hansen, M. & Srivastava, M. (2009), 'Sensor network data fault types', *ACM Transactions on Sensor Networks* **5**(3), 1–29. DOI [10.1145/1525856.1525863](https://doi.org/10.1145/1525856.1525863). ISSN 1550-4859.
- Nordio, A., Chiasserini, C. & Muscariello, A. (2010), 'Signal reconstruction in sensor networks with flat and clustered topologies', *Computer Networks* **54**(6), 991–1004. DOI [10.1016/j.comnet.2009.09.023](https://doi.org/10.1016/j.comnet.2009.09.023). ISSN 1389-1286.
- Oliveira, S. & Stewart, D. (2006), *Writing Scientific Software: a guide to good style*, Cambridge University Press.
- Oracle Corporation (2010), 'Lustre file system'. URL <http://lustre.org>, última consulta em abril de 2010.
- Oscar, T. P. (2009), 'General regression neural network and Monte Carlo simulation model for survival and growth of salmonella on raw chicken skin as a function of serotype, temperature, and time for use in risk assessment', *Journal of Food Protection* **72**(10), 2078–2087. ISSN 0362-028X.
- R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, ISBN 3-900051-07-0.
- Ramos, H., Guidoni, D., Loureiro, A. A. F., Frery, A. C. & Nakamura, E. F. (2010), Stochastic deployment models for heterogeneous wireless sensor networks, submetido ao SECOM 2010.
- Rappaport, T. S. (2002), *Wireless Communications Principles And Practice*, Prentice Hall Communications Engineering and Emerging Technologies Series, 2 ed., Prentice Hall.
- Ribeiro Jr., P. & Diggle, P. (2001), 'geoR: a package for geostatistical analysis', *R-News* **1**(2), 15–18. ISSN 1609-3631. URL <http://cran.R-project.org/doc/Rnews>.
- Ripley, B. D. (1987), *Stochastic Simulation*, Wiley, New York, EUA.
- Ripley, B. D. (1990), 'Thoughts on pseudorandom number generators', *Journal of Computational and Applied Mathematics* **31**, 153–163.
- Rivard, M. J., Melhus, C. S., Wazer, D. E. & Bricault Jr., R. J. (2009), 'Dosimetric characterization of round HDR Ir-192 applicators for breast brachytherapy', *Medical Physics* **36**(11), 5027–5032. DOI [10.1118/1.3232001](https://doi.org/10.1118/1.3232001). ISSN 0094-2405.

- Robert, C. P. & Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer-Verlag.
- Sarangapani, J. (2007), *Wireless Ad Hoc and Sensor Networks: protocols, performance, and control*, Control Engineering Series, CRC Press.
- Schilstra, M. J. & Martin, S. R. (2009), Simple stochastic simulation, in 'Methods in Enzymology: Computer Methods, Part B', Vol. 467 of *Methods in Enzymology*, Elsevier Academic Press, pp. 381–409.
- Schlather, M. (1999), Introduction to positive definite functions and to unconditional simulation of random fields, Relatório técnico, Department of Maths and Statistics, Lancaster University.
- Schlather, M. (2009), *Simulation and Analysis of Random Fields*.
- Sha, K., Shi, W. & Watkins, O. (2006), Using wireless sensor networks for fire rescue applications: requirements and challenges, in 'Proceedings of IEEE International Conference on Electro/information Technology', East Lansing, MI, EUA, pp. 239–244. DOI [10.1109/EIT.2006.252145](https://doi.org/10.1109/EIT.2006.252145).
- Shannon, C. E. (1949), 'Communication in the presence of noise', *Proceedings of the IRE* **37**(1), 10–21. ISSN 0096-8390.
- Sohraby, K., Minoli, D. & Znati, T. (2007), *Wireless sensor networks: technology, protocols, and applications*, Wiley.
- Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A. & Estrin, D. (2004), 'Habitat monitoring with sensor networks', *Communications of the ACM* **47**(6), 34–40.
- Tatone, B. S. & Grasselli, G. (2010), 'ROCKTOPPLE: A spreadsheet-based program for probabilistic block-toppling analysis', *Computers & Geosciences* **36**(1), 98–114. DOI [10.1016/j.cageo.2009.04.014](https://doi.org/10.1016/j.cageo.2009.04.014). ISSN 0098-3004.
- Thompson, K. M., Burmaster, D. E. & Crouch, E. A. C. (1992), 'Monte-Carlo techniques for quantitative uncertainty analysis in public-health assessments', *Risk Analysis* **12**(1), 53–63. ISSN 0272-4332.
- Tsow, F., Forzani, E., Rai, A., Wang, R., Tsui, R., Mastroianni, S., Knobbe, C., Gandolf, A. J. & Tao, N. J. (2009), 'A wearable and wireless sensor system for real-time monitoring of toxic environmental volatile organic compounds', *IEEE Sensors Journal* **9**(12), 1734–1740. DOI [10.1109/JSEN.2009.2030747](https://doi.org/10.1109/JSEN.2009.2030747). ISSN 1530-437X.

- Ueda, M. (2005), Licenses of open source software and their economic values, *in* 'SAINT-W '05: Proceedings of the 2005 Symposium on Applications and the Internet Workshops', IEEE Computer Society, Washington, DC, EUA, pp. 381–383.
- Umer, M., Kulik, L. & Tanin, E. (2008), Kriging for localized spatial interpolation in sensor networks, *in* 'SSDBM '08: Proceedings of the 20th international conference on Scientific and Statistical Database Management', Springer-Verlag, Berlin, Heidelberg, pp. 525–532. DOI 10.1007/978-3-540-69497-7_34.
- Vandewalle, P., Kovacevic, J. & Vetterli, M. (2009), 'Reproducible research'. URL <http://www.reproducibleresearch.net>, última consulta em abril de 2010.
- Vose, D. J. (1998), 'The application of quantitative risk assessment to microbial food safety', *Journal of Food Protection* **61**(5), 640–648. ISSN 0362-028X.
- Walker, D. W. & Dongarra, J. J. (1994), MPI: A message-passing interface standard, Relatório técnico, Tennessee University, Knoxville, TN, EUA.
- Wichmann, B. A. & Hill, I. D. (1982), 'Algorithm AS 183: An efficient and portable pseudo-random number generator', *Applied Statistics* **31**(2), 188–190. ISSN 0035-9254. URL <http://www.jstor.org/pss/2347988>.
- Wichmann, B. A. & Hill, I. D. (2006), 'Generating good pseudo-random numbers', *Computational Statistics & Data Analysis* **51**(3), 1614–1622. DOI 10.1016/j.csda.2006.05.019. ISSN 0167-9473.
- Wu, S. & Tseng, Y., eds (2007), *Wireless Ad-Hoc Networking: personal-area, local-area and the sensor-area network*, Wireless Networks and Mobile Communication Series, Auerbach Publications.
- Yoon, S. & Shahabi, C. (2007), 'The Clustered AGgregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks', *ACM Transactions on Sensor Networks*. DOI 10.1145/1210669.1210672. ISSN 1550-4859.
- Younis, O. & Fahmy, S. (2004), 'HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks', *IEEE Transactions on Mobile Computing* **3**(4), 366–379. DOI 10.1109/TMC.2004.41. ISSN 1536-1233.
- Younis, O., Krunz, M. & Ramasubramanian, S. (2006), 'Node clustering in wireless sensor networks: Recent developments and deployment challenges', *IEEE Network* **20**(3), 20–25. DOI 10.1109/MNET.2006.1637928. ISSN 0890-8044.
- Zhu, H., Williams, C. K. I., Rohwer, R. & Morcinec, M. (1997), Gaussian regression and optimal finite dimensional linear models, Technical report, Neural Computing Research

Group, Department of Computer Science & Applied Mathematics, Aston University,
Birmingham, Reino Unido.

Este trabalho foi redigido em $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ utilizando uma modificação do estilo IC-UFAL. As referências bibliográficas foram preparadas no JobRef e administradas pelo $\text{BIB}_{\text{E}}\text{X}$ com o estilo LoCCAN. O texto utiliza fonte Kerkis e os elementos matemáticos a família tipográfica Fourier-GUTenberg, ambas em corpo de 12 pontos. A numeração dos capítulos segue com a família tipográfica Art Nouveau Initialen.

As citações que iniciam cada capítulo estão disponíveis no pacote fortunes do R.

