UNIVERSIDADE DE SÃO PAULO

RICARDO AUGUSTO DA SILVA JÚNIOR

GERAÇÃO DE MAPAS DE AMBIENTE DE RÁDIO EM SISTEMAS DE
COMUNICAÇÕES SEM FIO COM INCERTEZA DE LOCALIZAÇÃO

São Paulo
2019

RICARDO AUGUSTO DA SILVA JÚNIOR

GERAÇÃO DE MAPAS DE AMBIENTE DE RÁDIO EM SISTEMAS DE
COMUNICAÇÕES SEM FIO COM INCERTEZA DE LOCALIZAÇÃO

Tese de Doutorado apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Doutor em Ciências

Área de Concentração:

Sistemas Eletrônicos (3142)

Orientador:

Prof. Dr. Cristiano Magalhães Panazio

São Paulo

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Silva Júnior, Ricardo Augusto da

Geração de mapas de ambiente de rádio em sistemas de comunicações sem fio com incerteza de localização / R. A. Silva Júnior -- versão corr. -- São Paulo, 2019.

127 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1.Wireless 2.Processamento de Sinais 3.Geoestatística I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

Silva Júnior, Ricardo Augusto da. Geração de mapas de ambiente de rádio em sistemas de comunicações sem fio com incerteza de localização. / R. A. Silva Junior – ver corr. –São Paulo, 2019. 127 p. Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

Aprovado em:

Banca Examinadora

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Resumo

A geração e o uso dos mapas de ambiente de rádio (REM - *Radio Environment Map*) em sistemas de comunicações sem fio vêm sendo alvo de pesquisas recentes na literatura científica. Dentre as possíveis aplicações, o REM fornece informações importantes para os processos de predição e otimização de cobertura em sistemas de comunicações sem fio, pois é baseado em medidas coletadas diretamente da rede. Neste contexto, a geração do REM depende do processamento das medidas e suas localizações para a construção dos mapas, por meio de predições espaciais. Entretanto, a incerteza de localização das medidas coletadas pode degradar a acurácia das predições de forma significativa e, conseqüentemente, impactar as posteriores decisões baseadas no REM.

Este trabalho aborda o problema de geração do REM de forma mais realística, formulando um modelo de predição espacial que introduz erros de localização no ambiente de rádio de um sistema de comunicação sem fio. As investigações mostram que os impactos provocados pela incerteza de localização na geração do REM são significativos, especialmente nas técnicas de estimação utilizadas para a aprendizagem de parâmetros do modelo de predição espacial. Com isso, uma técnica de predição espacial é proposta e utiliza ferramentas da área geoestatística para superar os efeitos negativos causados pela incerteza de localização nas medidas.

Simulações computacionais são desenvolvidas para a avaliação de desempenho das principais técnicas de predição no contexto de geração do REM, considerando o problema da incerteza de localização. Os resultados de simulação da técnica proposta são promissores e mostram que levar em conta a distribuição estatística dos erros de localização pode resultar em predições com maior acurácia para a geração do REM. A influência de diferentes aspectos da modelagem do ambiente de rádio também é analisada e reforçam a ideia de que a aprendizagem de parâmetros do ambiente de rádio tem um papel importante na acurácia das predições espaciais, que são fundamentais para a geração confiável do REM. Finalmente, um estudo experimental do REM é realizado por meio de uma campanha de medidas, permitindo explorar o desempenho dos algoritmos de aprendizagem de parâmetros e predições desenvolvidos neste trabalho.

Palavras-Chave: Mapas de ambiente de rádio, geoestatística, predições espaciais, incerteza de localização e processos Gaussianos.

Abstract

The generation and use of radio environment maps (REM) in wireless systems has been the subject of recent research in the scientific literature. Among the possible applications, the REM provides important information for the coverage prediction and optimization processes in wireless systems, since it is based on measurements collected directly on the network. In this context, the REM generation process depends on the processing of the measurements and their locations for the construction of the maps through spatial predictions. However, the location uncertainty related to the measurements collected can significantly degrade the accuracy of the spatial predictions and, consequently, impact the decisions based on REM.

This work addresses the problem of the REM generation in a more realistic way, through the formulation of a spatial prediction model that introduces location errors in the radio environment of a wireless communication system. The investigations show that the impacts of the location uncertainty on the REM generation are significant, especially in the estimation techniques used to learn the parameters of the spatial prediction model. Thus, a spatial prediction technique is proposed, based on geostatistical tools, to overcome the negative effects caused by the location uncertainty of the REM measurements.

Computational simulations are developed for the performance evaluation of the main prediction techniques in the context of REM generation, considering the problem of location uncertainty. The simulation results of the proposed technique are promising and show that taking into account the statistical distribution of location errors can result in more accurate predictions for the REM generation process. The influence of different aspects of the radio environment modeling is also analyzed and reinforce the idea that the learning of radio environment parameters plays an important role in the accuracy of spatial predictions, which are fundamental for the reliable REM generation. Finally, an experimental study is carried out through a measurement campaign with the purpose of generating the REM in practice and to explore the performance of the learning and prediction algorithms developed in this work.

Key-Words: Radio environment map, geostatistics, spatial predictions, location uncertainty and Gaussian process.

À minha centenária avó, Maria de Lourdes

“Clarity is a consequence of handling your
confusion consciously.”

Jaggi Vasudev, Sadhguru

Agradecimentos

Talvez, esta seja a seção mais importante deste trabalho, pelo simples fato desta possibilitar a geração de todas as outras seções e capítulos. Talvez também seja por isso que é tão difícil escrevê-la...

Sem dúvida e primeiramente, gostaria de agradecer profundamente ao meu orientador, Professor Cristiano Panazio, pelas valiosas contribuições, todo o apoio e paciência ao longo de toda a jornada deste trabalho. Me sinto muito privilegiado em conhecê-lo Professor, e tenho certeza que sempre vou carregar comigo seus ensinamentos na busca pela verdadeira essência da ciência em nossas vidas.

Conhecer a Escola Politécnica da USP mudou minha vida e me permitiu conhecer Professores que contribuíram enormemente com a minha formação humana e acadêmica. Gostaria muito de agradecer aos professores da USP: Profa. Maria D. Miranda, Prof. Magno T. Silva, Prof. Vitor H. Nascimento e Prof. Marcio Eisencraft, além de todo o corpo docente e funcionários da Poli-USP. Fazer as viagens para São Paulo e assistir às aulas ampliou significativamente a minha visão sobre o aprendizado e me forneceu toda a motivação para continuar trabalhando sempre com alegria.

Gostaria de estender os agradecimentos aos meus companheiros do Instituto Nacional de Telecomunicações, que sempre me incentivaram nesta jornada de trabalhos e estudos. Agradeço em particular, ao meu amigo, Anderson Fagiani, pela várias e longas conversas sobre a teoria e a prática de propagação em sistemas de comunicações.

Um agradecimento especial dedico à minha querida Flávia Rennó e toda a sua família, que proporcionam momentos de pura alegria em minha vida. Agradeço a minha amiga-irmã Verônica Rennó por me apresentar ao campus da USP e fazer parte deste momento em minha vida.

Gostaria de dedicar todo o esforço deste trabalho e fazer um agradecimento carinhoso aos meus pais, Ricardo (oh meu amigo) e Mara (minha mãezinha), e aos meus irmãos, Renan e Fernanda, que, mesmo estando longe, me apoiam de forma única em todos os aspectos da minha vida. Sou eternamente grato à Deus, por ter vocês como família.

Finalmente, registro meu agradecimento a todos os cientistas, que nos fornecem a base de conhecimento necessária para que possamos entender um pouco mais sobre a natureza do universo.

Lista de Siglas

BLUP	<i>Best Linear Unbiased Predictor</i> (Melhor preditor linear não enviesado)
cPPP	<i>Cluster Poisson Point Process</i> (Processo pontual Poisson com aglomerados)
EMSPE	<i>Empirical Mean Square Prediction Error</i> (Erro quadrático médio de predição empírico)
ERB	<i>Radio Base Station</i> (Estação rádio base)
GPS	<i>Global Positioning System</i> (Sistema de posicionamento global)
HPPP	<i>Homogeneous Poisson Point Process</i> (Processo pontual Poisson homogêneo)
IDW	<i>Inverse Distance Weighting</i> (Ponderação da distância inversa)
KALE	<i>Kriging Adjusted for Location Error</i> (Kriging ajustado para os erros de localização)
KILE	<i>Kriging Ignoring the Location Error</i> (Kriging que ignora os erros de localização)
KU	<i>Universal Kriging</i> (Kriging universal)
KO	<i>Ordinary Kriging</i> (Kriging ordinário)
KS	<i>Simple Kriging</i> (Kriging Simples)
LBS	<i>Location Based Services</i> (Serviços baseados em localização)
LIvE	<i>Location Estimation based REM</i> (REM baseado em estimação de localização)
LOS	<i>Line of sight</i> (Linha de visada direta)
LS	<i>Least Squares</i> (Mínimos quadrados)
MCD	<i>Measurement Capable Devices</i> (Dispositivos capazes de realizar medidas)
MDT	<i>Minimization of Drive Tests</i> (Minimização de drive-tests)
MoM	<i>Method of Moments of Matheron</i> (Método dos momentos de Matheron)
MVUE	<i>Minimum Variance Unbiased Estimator</i> (Estimador não enviesado de mínima variância)
ML	<i>Maximum Likelihood</i> (Máxima verossimilhança)
MSE	<i>Mean Squared Error</i> (Erro quadrático médio)
MSPE	<i>Mean Squared Prediction Error</i> (Erro quadrático médio de predição)
nHPPP	<i>Non Homogeneous Poisson Point Process</i> (Processo pontual Poisson não homogêneo)
NLOS	<i>Non Line of sight</i> (Sem linha de visada direta)
NN	<i>Nearest Neighbour</i> (Vizinhança mais próxima)
NNI	<i>Natural Nearest Interpolation</i> (Interpolação natural de vizinhança)

OLS	<i>Ordinary Least Squares</i> (Mínimos quadrados ordinário)
PP	<i>Point Process</i> (Processos pontuais)
PPP	<i>Poisson Point Process</i> (Processos pontuais de Poisson)
REM	<i>Radio Environment Map</i> (Mapa de ambiente de rádio)
REML	<i>Residual Maximum Likelihood</i> (Máxima verossimilhança residual)
RMSPE	<i>Root Mean Squared Prediction Error</i> (Raiz quadrada do erro quadrático médio de predição)
SALE	<i>Semivariogram Adjusted for Location Error</i> (Semivariograma ajustado para os erros de localização)
STM	<i>Self-Tuning Method</i> (Método de auto-ajuste)
TALE	<i>Trend Adjusted for Location Error</i> (Tendência ajustada para os erros de localização)
TS-KALE	<i>Trend, Semivariogram and Kriging Adjusted for Location Error</i> (Tendência, semivariograma e Kriging ajustados para os erros de localização)
VR_{VR}	<i>Regionalized Variable Variance</i> (Variância da variável regionalizada)
VA	<i>Random Variable</i> (Variável aleatória)
VI	<i>Interpolation Variance</i> (Variância de interpolação)
VK	<i>Kriging Variance</i> (Variância Kriging)
VR	<i>Regionalized Variable</i> (Variável regionalizada)

Lista de Símbolos

α	Coeficiente de propagação
$\hat{\alpha}_{\text{adj}}$	Coeficiente de propagação ajustado do algoritmo TALE
β	Vetor de semivariâncias relacionado com a coordenada $\mathbf{s}_0$
γ	Função semivariograma do processo aleatório espacial P
γ_g	Função semivariograma do processo aleatório espacial P_g
$\hat{\gamma}_{\text{adj}}$	Vetor de semivariâncias ajustadas do algoritmo SALE
θ	Vetor de parâmetros de covariância
λ_ν	Vetor de pesos Kriging
λ_g	Vetor de pesos Kriging obtidos com procedimento de ajuste
μ	Potência média de recepção do processo P
μ_g	Potência média de recepção do processo P_g
ν	Multiplicador de Lagrange
ξ	Sombreamento log-normal do ambiente de rádio
ξ	Vetor de valores do sombreamento log-normal do ambiente de rádio
ϱ	Função intensidade do processo pontual de aglomerados Poisson
σ	Desvio padrão do sombreamento log-normal
σ_p	Desvio padrão dos erros de localização
σ_{KO}^2	Variância Kriging
σ_{VI}^2	Variância de interpolação
Γ	Matriz de semivariâncias formada a partir das coordenadas espaciais $\mathbf{s}$
A	Área de cobertura do domínio espacial
A_{Lag}	Conjunto de distâncias quantizadas do semivariograma experimental
$\mathbf{a}$	Vetor para armazenamento das estimativas $\hat{\alpha}$ do algoritmo TALE
C	Função covariância do processo aleatório espacial P
C_g	Função covariância do processo aleatório espacial P_g
$\mathbf{C}$	Matriz de covariância formada a partir das coordenadas espaciais $\mathbf{s}$
$\mathbf{C}_g$	Matriz de covariância ajustada formada a partir das coordenadas espaciais $\mathbf{s}$

$\mathbf{c}_0$	Vetor de covariância formado entre as coordenadas espaciais $\mathbf{s}$ e a coordenada $\mathbf{s}_0$
$\mathbf{c}_{g0}$	Vetor de covariâncias ajustado formado entre as coordenadas espaciais $\mathbf{s}$ e a coordenada $\mathbf{s}_0$
c_0	Variância <i>nugget</i>
D	Domínio espacial bidimensional
d_{A-B}	Distância entre dois pontos baseadas em coordenadas do sistema cartesiano
$\mathbf{d}_{adjTx}$	Vetor de distâncias ajustadas do algoritmo TALE
$\mathbf{D}_{adj}$	Matriz de distâncias ajustadas do algoritmo SALE
$\mathbf{d}_{Lag-adj}$	Vetor de distâncias lags ajustadas do algoritmo SALE
d_{adj}	Valor de distância escalar utilizado no algoritmo SALE
d_{corr}	Distância de decorrelação espacial do sombreamento log-normal
d_s	Distância à reta bissetriz do gráfico de dispersão
d_0	Distância de referência do modelo da perda por percurso média
e	Resíduo do método dos mínimos quadrados
F_p	Distribuição de probabilidade dimensional do processo P
f_k	Funções base do modelo da tendência
g	Distribuição de probabilidade dos erros de localização
$\mathcal{GP}$	Processo aleatório espacial Gaussiano
$\mathbf{h}$	Vetor de separação espacial
I	Função intensidade dos processos aleatórios pontuais
Lag	Distâncias lags do semivariograma experimental
m	Variância <i>sill</i>
MSE_α	Erro quadrático médio relacionado com a estimação do coeficiente de propagação
$\mathbf{M}_{Dg}$	Matriz de distâncias ajustadas do preditor KALE
$\mathbf{m}_{Dg0}$	Vetor de distâncias ajustadas do preditor KALE
$\mathcal{N}$	Distribuição densidade de probabilidade Gaussiana
N	Número total de medidas coletadas
N_h	Número de medidas com distância de separação espacial $\mathbf{h}$
N_{HPPP}	Número de eventos de localização do processo pontual Poisson homogêneo
$\mathbf{p}_e$	Vetor de erros de localização
$\mathbf{p}$	Vetor de medidas de potência de recepção

$\mathbf{p}_g$	Vetor de medidas de potência de recepção com erros de localização
P	Processo aleatório espacial
P_g	Processo aleatório espacial com incerteza de localização
$\bar{P}_L$	Perda por percurso média
$\bar{\bar{P}}_L$	Perda por percurso média em d_0
P_{tx}	Potência de transmissão
P_{rx}	Potência de recepção do processo P
P_{IDW}	Predição da potência de recepção com a técnica IDW
P_{NNI}	Predição da potência de recepção com a técnica NNI
$\mathbf{p}_{\text{adj}}$	Vetor de potência média de recepção ajustada do algoritmo TALE
P_{KO}	Predição da potência de recepção com a técnica KO
P_{KILE}	Predição da potência de recepção com a técnica KILE
P_{KALE}	Predição da potência de recepção com a técnica KALE
$\mathbf{P}_{\text{M}}$	Matriz para armazenamento dos valores de potência média de recepção do algoritmo TALE
$P_{\text{TS-KALE}}$	Predição da potência de recepção com a técnica TS-KALE
$\mathbf{r}$	Coordenadas espaciais realizadas
r	Distância <i>range</i> do sombreamento do canal sem fio
$\mathbf{s}$	Coordenadas espaciais estimadas
$\mathbf{s}_{\text{tx}}$	Localização fixa da ERB
$\mathbf{s}_0$	Coordenada espacial alvo das previsões
$\mathbf{u}, \mathbf{v}$	Vetores de erro de localização
$\mathbf{U}_x$ e $\mathbf{U}_y$	Matrizes com vetores de erros de localização utilizada no algoritmo TALE
VAR_{VR}	Variância do conjunto de valores regionalizadas
V_{NNI}	Subconjunto de medidas de vizinhança do preditor NNI
w	Pesos do preditor espacial IDW
$\mathbf{x}$	Vetor de funções base do ambiente de rádio
$\mathbf{x}_g$	Vetor de funções base ajustadas do ambiente de rádio
$\mathbf{z}$	Vetor de valores regionalizados sem erros de localização
$\mathbf{z}_g$	Vetor de valores regionalizados com erros de localização

$\mathbf{z}_{\text{adj}}$	Vetor de valores regionalizados e ajustados do algoritmo TALE
Z_{IDW}	Predição espacial do sombreamento com a técnica IDW
Z_{NNI}	Predição espacial do sombreamento com a técnica NNI
Z_{KO}	Predição espacial do sombreamento com a técnica KO
Z_{KILE}	Predição espacial do sombreamento com a técnica KILE
Z_{KALE}	Predição espacial do sombreamento com a técnica KALE
$Z_{\text{TS-KALE}}$	Predição espacial do sombreamento com a técnica TS-KALE
$\mathbb{E}\{\cdot\}$	Operador esperança matemática
$\text{Var}(\cdot)$	Operador de variância
$\text{cov}(\cdot)$	Operador de covariância
$(\cdot)^{\text{T}}$	Operador de transposição
$(\hat{\cdot})$	Operador de estimação
$(\bar{\cdot})$	Operador de cálculo da média

Lista de Tabelas

1.1	Métodos de Predição Espacial para a Geração do REM	6
2.1	Cenários de Propagação sem Fio e o Coeficiente de Propagação [96]	25
2.2	Parâmetros e Configurações da Simulação	39
2.3	Predições Espaciais - Métodos Kriging	43
3.1	Configurações de Simulação dos Cenários A - B - C	74
5.1	Parâmetros e Configurações do Experimento Prático	97

Lista de Figuras

1.1	Arquitetura geral do REM em sistemas de comunicações sem fio.	4
1.2	Ilustração da incerteza de localização em sistemas de comunicações sem fio.	8
2.1	Realização de um processo aleatório espacial Gaussiano.	15
2.2	Modelos analíticos: (a) função covariância e (b) função semivariograma.	22
2.3	Valores regionalizados obtidos a partir de um processo aleatório espacial Gaussiano.	23
2.4	Diagrama do modelo de geração do REM.	28
2.5	Interpretação geométrica do gráfico de dispersão.	32
2.6	Semivariograma experimental: (a) nuvem de semivariâncias (b) com agrupamento de medidas.	34
2.7	Modelos analíticos e partes experimentais: (a) covariância e (b) semivariograma.	37
2.8	Resultados de simulação: (a) potência média de recepção; (b) sombreamento do canal sem fio; (c) ambiente de rádio com amostragem espacial realizada a partir da densidade de probabilidade uniforme; (d) estimação da perda por percurso média; (e) semivariograma e (f) covariograma.	40
2.9	Resultados de simulação: (a) potência de recepção no ambiente de rádio; (b) REM; (c) mapa diferença: potência de recepção e o REM; (d) sombreamento log-normal do canal sem fio; (e) previsões espaciais via KO e (f) mapa diferença: sombreamento e previsões KO.	47
2.10	(a) variância Kriging; (b) desvio padrão Kriging e (c) variância de interpolação.	50
3.1	(a) ambiente de rádio gerado com erros de localização e (b) modelo de incerteza.	52
3.2	Influência dos erros de localização: (a) estimação da tendência e (b) remoção da tendência.	56
3.3	Influência dos erros de localização no: semivariograma experimental (a) e (c) e covariograma experimental (b) e (c)	57
3.4	Influência dos erros de localização na estimação dos parâmetros de covariância $\hat{\theta}$	58
3.5	Efeito dos erros de localização em relação à tendência do ambiente de rádio.	61
3.6	Efeito dos erros de localização em relação à covariância espacial do ambiente de rádio.	62

3.7	Diagrama de funcionamento do preditor espacial TS-KALE.	67
3.8	Desempenho da estimação ajustada da tendência: MSE_{α} (a)-(b)-(c) e dispersão da variável regionalizada (d)-(e)-(f).	70
3.9	Estimação do Semivariograma Experimental: (a) sem erros de localização; (b) com erros de localização e (c) com o ajuste SALE.	73
3.10	Histogramas estatísticos do cenário A: (a-b-c) estimativas de $\hat{m}$; (d-e-f) estimativas de $\hat{r}$	75
3.11	Histogramas estatísticos do cenário B: (a-b-c) estimativas de $\hat{m}$; (d-e-f) estimativas de $\hat{r}$	75
3.12	Histogramas estatísticos do cenário C: (a-b-c) estimativas de $\hat{m}$; (d-e-f) estimativas de $\hat{r}$	76
4.1	Fluxograma geral de simulação para análise de desempenho das predições espaciais.	79
4.2	Influência do número de medidas N no desempenho dos preditores KILE e KALE com estimação perfeita (Cenário A) e imperfeita (Cenário B) da perda por percurso média.	82
4.3	Influência dos erros de localização no desempenho dos preditores KILE e KALE com estimação perfeita (Cenário C) e imperfeita (Cenário D) da perda por percurso média.	83
4.4	Influência do sombreamento do canal sem fio no desempenho dos preditores KILE e KALE com estimação perfeita (Cenário E) e imperfeita (Cenário F) da perda por percurso média.	84
4.5	Influência da incerteza de localização e da estimação de θ no desempenho dos preditores KILE, KALE e TS-KALE, com perfeita (Cenário G) e imperfeita (Cenário H) estimação da tendência.	85
4.6	Estatísticas de EMSPE: (a) sem incerteza de localização; (b) com incerteza de localização.	88
4.7	Comparação de desempenho dos preditores Kriging e convencionais: (a) frente a incerteza de localização (b) em função do número de medidas.	89
4.8	Diferentes padrões de amostragem espacial: (a) completamente aleatória (uniforme); processo pontal de Poisson (b) homogêneo; (c) não homogêneo e (d) com aglomerados (<i>clusters</i>).	91
4.9	Estatísticas de EMSPE, considerando a amostragem espacial: (a) completamente aleatória (densidade de probabilidade uniforme); (b) HPPP; (c) nHPPP e (d) cPPP.	93
5.1	Diagrama em blocos da metodologia dos experimentos.	96
5.2	Área externa onde foi realizada a campanha de medidas.	98
5.3	Sistemas de transmissão e recepção utilizados na campanha de medidas.	99

5.4	Instrumento de medida com geolocalização e pontos espaciais de medições do experimento.	100
5.5	Resultados estatísticos relacionados com as medidas em condições de recepção LOS.	101
5.6	Resultados estatísticos relacionados com as medidas em condições de recepção NLOS.	102
5.7	Tratamento da tendência: (a) regressão do modelo log-distância e (b) histograma dos valores regionalizados, após a remoção da tendência.	104
5.8	Resultados da análise geoestatística: (a) histograma da variância <i>sill</i> ; (b) histograma do parâmetro <i>r</i> ; (c) semivariograma e (d) covariograma	106
5.9	Resultados de predição: (a) mapa de predição da potência média de recepção; (b) mapa de predição espacial Kriging do sombreado log-normal; (c) REM prático; (d) mapa da variância Kriging.	108
5.10	Visualização do REM prático no local do experimento da campanha de medidas.	109
5.11	Visualização das coordenadas espaciais sem e com os erros de localização.	111
5.12	Resultados de desempenho: (a) histograma dos erros quadráticos de predição e (b) curvas de EMSPE dos preditores KILE, KALE e TS-KALE, obtidas a partir das medidas práticas.	112

Conteúdo

Resumo	ii
Abstract	iii
Lista de Siglas	vii
List de Símbolos	ix
Lista de Tabelas	xiii
Lista de Figuras	xvi
1 Introdução	3
1.1 Contextualização e Estado da Arte	3
1.2 Definição do Problema	8
1.3 Escopo e Objetivos de Pesquisa	9
1.4 Contribuições	10
1.5 Organização da Tese	12
1.6 Publicações geradas	12
2 Modelo de Predição Espacial para a Geração do REM	13
2.1 Conceitos da Geoestatística	13
2.1.1 Processos Aleatórios Espaciais	14
2.1.2 Estacionariedade	17
2.1.3 Covariância e Semivariograma - Propriedades	19
2.1.4 Variáveis Regionalizadas	23
2.2 Modelagem do Ambiente de Rádio	24
2.2.1 Modelagem da Tendência	24
2.2.2 Modelagem do Processo Aleatório Espacial	26
2.3 Modelo de Predição Espacial para a Geração do REM	28
2.3.1 Tratamento da Tendência Espacial	29
2.3.2 Semivariograma e Covariograma Experimentais	32
2.3.3 Estimação dos Parâmetros de Covariância	37
2.3.4 Simulação da Fase I de Aprendizagem do Ambiente de Rádio	39
2.3.5 Preditor Kriging - Processos Gaussianos	42
2.3.6 Simulação da Fase II - Geração do REM	47
2.4 Conclusões - Capítulo 2	50

3	Mitigação da Incerteza de Localização	51
3.1	Modelo de Incerteza de Localização do Ambiente de Rádio	54
3.2	Impacto da Incerteza de Localização	55
3.3	Tratamento da Incerteza de Localização	59
3.3.1	Procedimentos de Ajuste do Primeiro e Segundo Momentos	60
3.3.2	Kriging Ajustado para os Erros de Localização - KALE	63
3.3.3	Kriging que Ignora os Erros de Localização - KILE	64
3.4	Preditor Espacial TS-KALE	67
3.4.1	Estimação Ajustada da Tendência do Ambiente de Rádio - TALE	68
3.4.2	Estimação Ajustada do Semivariograma Experimental - SALE	71
3.4.3	Formulação Matemática do Preditor TS-KALE	77
3.5	Conclusões - Capítulo 3	77
4	Desempenho de Predição Espacial do REM	78
4.1	Fluxograma de Simulação	79
4.2	Predições com Parâmetros de Covariância Conhecidos	81
4.2.1	Influência do Número de Medidas	81
4.2.2	Influência dos Erros de Localização	83
4.2.3	Influência do Sombreamento do Canal sem Fio	84
4.3	Predições com Estimação dos Parâmetros de Covariância	85
4.4	Comparações com Preditores Espaciais Convencionais	87
4.5	Influência da Amostragem Espacial	90
4.6	Conclusões - Capítulo 4	94
5	Estudo Experimental da Geração do REM	95
5.1	Metodologia	96
5.2	Campanha de Medidas	97
5.3	Resultados da Fase I - Aprendizagem de Parâmetros	103
5.3.1	Análise da Tendência	103
5.3.2	Análise Geoestatística	105
5.4	Resultados da Fase II - Predições do REM	107
5.5	Conclusões - Capítulo 5	113
6	Conclusões e Perspectivas	114
	Bibliografia	116
	Apêndice A Anexos - Artigos	127

Capítulo 1

Introdução

Este capítulo introduz os conceitos dos mapas de ambiente de rádio (REM - *Radio Environment Map*) com o propósito de definir o problema de pesquisa a ser abordado neste trabalho de doutorado. A contextualização da pesquisa contemplando o estado da arte é colocada, permitindo que o escopo e os objetivos de pesquisa sejam apresentados. Em seguida, uma seção dedicada às contribuições alcançadas com o trabalho de pesquisa é apresentada, além da descrição da estrutura do texto e do conteúdo dos capítulos.

1.1 Contextualização e Estado da Arte

A demanda por tráfego e a qualidade dos serviços relacionados com as modernas redes sem fio, especialmente as novas gerações das redes celulares, requerem sistemas de comunicações que operam com elevado desempenho na gerência e no uso dos recursos espectrais e energéticos [1]. Com isso, o planejamento eficiente das redes de comunicações sem fio é um ponto crítico para que os melhores desempenhos sejam alcançados envolvendo diferentes aspectos (*e.g.*, probabilidade de erro, alta capacidade, ampla cobertura, etc). Além dos esforços de planejamento para a utilização eficiente do espectro, há um interesse no uso mais inteligente das informações de localização relacionadas com o sistema de comunicação sem fio [2-6]. De fato, o impacto do conhecimento geolocalizado associado aos sinais de rádio experimenta um crescimento significativo, atingindo diferentes campos das áreas das comunicações e navegação (*e.g.*, futuras gerações de redes móveis, monitoramento e controle de ambientes, transportes inteligentes, redes de sensores e aplicações com robôs) [7-8].

Dentro deste contexto, recentemente tem havido um interesse da comunidade científica na geração e no uso dos mapas de ambiente de rádio [9-15]. Parte disto ocorre em razão da forma como o REM utiliza as informações geolocalizadas e os potenciais benefícios para os sistemas de comunicações sem fio, especialmente para os processos de planejamento e otimização de cobertura, importantes na área de comunicações móveis [16-18].

O conceito do REM, inicialmente proposto em [19-20], consiste em um banco de dados composto por informações geolocalizadas sobre o ambiente de rádio a serem utilizadas pelos sistemas de rádios cognitivos em cenários de acesso dinâmico ao espectro. Desde então, a pesquisa sobre o REM se expandiu, uma vez que informações relevantes sobre o ambiente de rádio podem ser extraídas a partir do REM, permitindo sua relação com diferentes aplicações envolvendo: planejamento, otimização e minimização de *drive tests* (MDT - *Minimization of Drive Tests*) em redes de telefonia celular [16-18], [21-25]; decisões sobre a alocação de recursos de rádio em sistemas de rádios cognitivos [26], combate à interferência e compartilhamento do espectro [27-28]; além de aplicações específicas relacionadas com transportes de alta velocidade (*e.g.*, metrô e trens) [29] e problemas de estimação de localização de fontes de transmissão em redes de sensores [30].

É importante mencionar que a forma como o REM se relaciona com cada tipo de aplicação depende fundamentalmente de qual é o tipo de grandeza que constitui o REM e como as informações de localização são utilizadas para contruí-lo. Este trabalho segue a abordagem em [16-18], na qual o REM consiste em um mapa que indica uma grandeza física do ambiente de rádio, sendo a mais comum a intensidade do sinal recebido, *i.e.*, a potência de recepção do sinal de rádio. Desta forma, a geração do REM ocorre através da realização de predições espaciais sobre o ambiente de rádio a partir de um conjunto finito de medidas coletadas. Em outras palavras, o processo de geração do REM e a acurácia dos seus resultados compõem um problema de predição espacial [31].

A Figura 1.1 apresenta a arquitetura geral do REM aplicada aos sistemas de comunicações sem fio, evidenciando o posicionamento de seus diferentes módulos de processamento [11], [21], [32]. O funcionamento da arquitetura é essencialmente baseado na coleta e armazenamento de medidas seguido do processamento das predições espaciais e da representação dos resultados.

Figura 1.1: Arquitetura geral do REM em sistemas de comunicações sem fio.

A capacidade de medição incorporada aos dispositivos da rede (*i.e.*, MCD - *Measurement Capable Devices*) possibilita a realização e o envio de medições para o REM-SA (*Storage and Aquisition*). O módulo de aquisição estabelece a sinalização de controle necessária para a comunicação com os dispositivos da rede. O conjunto de medidas capturado na rede de comunicação compõe os dados do módulo de armazenamento. Diferentes tipos de informações podem ser armazenadas (*e.g.*, níveis de potência de recepção e relação sinal-ruído, indicadores de tráfego [31], além de parâmetros específicos do canal sem fio [29]), refletindo a diversidade de aplicações onde o REM pode ser inserido. O REM-M (*Manager*) é o responsável pela gerência das informações coletadas e possui acesso direto às informações do REM-SA. O enfoque do trabalho está sobre o módulo de predição espacial, que consiste em um processo específico que realiza a geração do REM através de técnicas de predições espaciais. Desta forma, a interface de gerenciamento poderá atuar sobre os resultados fornecidos pelo REM-M para que as decisões possam ser realizadas na rede de comunicação sem fio.

É importante mencionar que os resultados de predições espaciais alcançados dependem da obtenção das medidas geolocalizadas do ambiente de rádio. Com isso, a aplicabilidade das predições espaciais aos sistemas de comunicações sem fio está diretamente relacionada com a caracterização do ambiente de rádio, que é complexa em função dos diferentes mecanismos de propagação do canal sem fio (*e.g.*, reflexão, difração, espalhamento e a perda por percurso média). Apesar da difusão significativa dos estudos e trabalhos científicos sobre estes mecanismos de propagação (*e.g.*, [33-35]), a necessidade de avanço e a evolução dos sistemas de comunicações sem fio impulsionam a pesquisa nesta área (*e.g.*, [36-38]), incluindo o interesse na precisão alcançada pelos modelos de propagação [39-40], bem como os efeitos do canal sem fio sobre as predições espaciais [41]. De fato, os mecanismos de propagação provocam efeitos que têm influência direta na potência de recepção e, consequentemente, podem afetar o processo de geração do REM. Com isso, a capacidade das técnicas de predição espacial em processar um conjunto de medidas geolocalizadas e prover predições espaciais confiáveis é um dos pontos críticos na pesquisa sobre o REM [13], [31].

Dentro deste contexto, iniciativas na literatura científica propõem diferentes preditores espaciais que visam alcançar alta acurácia, robustez e eficiência nas diferentes aplicações relacionadas com o REM [13], [31], [42-43]. A Tabela 1.1 apresenta um quadro com as principais técnicas de predição espacial utilizadas. Em geral, dois aspectos são considerados para categorizar os preditores espaciais: i) sua concepção determinística ou estatística e ii) o seu modo de funcionamento contemplando a predição direta e indireta.

Os preditores determinísticos consistem em interpoladores baseados em funções matemáticas que geram superfícies de predição espacial. Os critérios de otimização envolvidos e a forma como essas funções são utilizadas permitem obter diferentes variantes de interpoladores determinísticos. O preditor IDW (*Inverse Distance Weighting*) é fundamentado no cálculo do inverso da distância entre as medidas coletadas e o alvo das predições, ou seja, cada medida observada possui uma relevância que depende da distância obtida entre a respectiva medida e a localidade alvo da predição.

Tabela 1.1: Métodos de Predição Espacial para a Geração do REM

Preditor Espacial	Tipo	Referências
IDW	Direto e Determinístico	[11], [14], [42-45]
NN	Direto e Determinístico	[42], [46-47]
NNI	Direto e Determinístico	[42], [46-47]
Superfície de Tendência	Direto e Determinístico	[42-43]
LIVe	Baseado em Modelos de Propagação (Indireto)	[48-49]
STM	Baseado em Modelos de Propagação (Indireto)	[42], [50]
Kriging	Direto e Geostatístico	[11], [57]

O preditor NN (*Nearest Neighbour*) é um interpolador e atribui ao resultado de predição a medida observada que mais se aproxima da localidade alvo da predição. Embora de implementação relativamente fácil, o preditor NN não apresenta alta acurácia nas predições e depende significativamente da forma como a coleta de medidas é realizada. De forma semelhante ao NN, o preditor NNI (*Natural Nearest Interpolation*) forma um subconjunto de medidas próximo ao alvo da predição espacial. Assim, um processamento que combina as medidas de vizinhança coletadas permite obter a predição no alvo. Além disso, a região local do preditor NNI pode ser determinada através de diferentes técnicas como o uso de diagramas de Voronoi [42], [47]. O método de superfície de tendência é baseado na interpolação de funções polinomiais (ou radiais, *e.g.*, *Thin Plate Splines* [43]), promovendo a suavização dos resultados de predição espacial ao longo do REM, adequada para casos onde há baixa variação nos valores das medidas observadas.

As técnicas STM (*Self-Tuning Method*) e LIVe (*Location Estimation based REM*) são exemplos de métodos indiretos para a geração do REM. O modo de funcionamento indireto é caracterizado primeiramente pela obtenção dos parâmetros de transmissão (*e.g.*, potência de transmissão e características da antena), além da localização do transmissor. Em seguida, modelos de propagação são aplicados para a realização das predições espaciais. Com isso, é possível dividir os métodos indiretos em três categorias em que: i) a localização e os parâmetros de transmissão são conhecidos *a priori*; ii) somente os parâmetros de transmissão são conhecidos (não se sabe a localização do transmissor) ou somente a localização do transmissor é conhecida (não se sabe os parâmetros de transmissão) e iii) quando não há o conhecimento da localização nem dos parâmetros de transmissão.

A geração do REM no caso i) consiste na aplicação direta dos modelos de propagação a partir dos parâmetros de transmissão conhecidos. Esta abordagem é tradicional e amplamente difundida no planejamento de redes com sistemas de comunicações sem fio [51]. O segundo caso ii), dependendo das informações conhecidas *a priori*, requer a estimação da localização do transmissor ou a estimação dos parâmetros de transmissão, resultando no caso (i). A geração do REM no caso iii) é mais complexa, uma vez que requer a estimação dos parâmetros e da localização do transmissor. Ainda assim, em todos os casos, as predições são alcançadas através da aplicação dos modelos de propagação, que requerem uma calibração adequada e complexa em função do ambiente de rádio.

A técnica LIVe é um exemplo de método indireto em que a potência e a localização do transmissor são estimados primeiramente para que um modelo de propagação empírico seja utilizado. Em [49] resultados promissores desta técnica são apresentados. Apesar disso, argumenta-se em [42] sobre a falta de adaptabilidade dos modelos de propagação às condições reais do ambiente de rádio e que mais parâmetros de transmissão devem ser levados em conta no uso de métodos indiretos. Isso é proposto no método STM, no qual as características do sistema de irradiação (*e.g.*, diagrama de irradiação das antenas) é incluída na geração do REM. As análises e os resultados em [50] mostraram que levar em conta informações sobre os parâmetros de transmissão é um aspecto realístico importante, embora a complexidade computacional seja significativamente aumentada.

É fundamental ressaltar que ambas as abordagens (direta e indireta) são afetadas pelo comportamento aleatório do ambiente de rádio. Desta forma, enquanto o desempenho pretendido com os métodos indiretos depende da capacidade dos modelos de propagação em capturar os efeitos do canal sem fio, os preditores diretos se beneficiam das predições construídas diretamente a partir das medidas coletadas. É neste contexto que a geoestatística fornece ferramentas para que a geração do REM seja realizada de forma mais eficiente e menos complexa, comparada ao métodos indiretos. Embora relativamente recente, a utilização das técnicas da área geoestatística no campo das comunicações vem crescendo, sobretudo acerca do problema de geração do REM como em [11-12], [52-56]. É possível mencionar algumas razões para isto: i) a primeira está relacionada com a necessidade de análises e predições espaciais em problemas que envolvem a dimensão espacial em sistemas de comunicações sem fio e ii) o fato de que as técnicas geoestatísticas permitem explorar as estatísticas espaciais dos fenômenos analisados (*e.g.*, ambiente de rádio) a favor das predições. Dentre os métodos diretos, o Kriging (conhecido como predição em processos Gaussianos na teoria de aprendizagem de máquina [58], [31]) é uma das técnicas mais difundidas no campo da geoestatística¹.

No conjunto de iniciativas de pesquisa sobre a geração do REM, vários trabalhos científicos utilizam o Kriging para a realização das predições, especialmente nas aplicações voltadas para comunicações sem fio, redes móveis e de sensores [16-18], [31], [69], [70-73]. A principal vantagem do Kriging é que a sua concepção é fundamentada na minimização do erro quadrático médio de predição espacial, possibilitando a obtenção de predições com maior acurácia. De fato, dentre os preditores diretos e lineares, o Kriging é o BLUP (*Best Linear Unbiased Predictor*) [62-63]. Os resultados em [54] mostram que o Kriging atingiu melhor desempenho em termos de acurácia comparado aos modelos de propagação convencionais em comunicações sem fio, indicando o potencial de contribuição da geoestatística para as predições espaciais no contexto do REM [31], [38].

¹Por volta da década de 1950, o professor francês Georges Matheron [59] publicou uma série de estudos e resultados científicos de significativa relevância para teoria das variáveis regionalizadas, permitindo sua posição e reconhecimento como fundador da área geoestatística [60]. O nome Kriging foi escolhido por Georges Matheron em homenagem ao sul-africano Daniel Krige [61], engenheiro de minas e estatístico, pioneiro nos estudos do que viria a ser o campo da geoestatística. A literatura da área é ampla e trata de vários aspectos sobre análises e processos aleatórios espaciais [62-63]. Uma visão introdutória da geoestatística pode ser encontrada em [64-66], enquanto tópicos mais avançados (*e.g.*, estatísticas espaço-temporais e aplicação de predições Bayesianas) podem ser acessados em [67-68].

1.2 Definição do Problema

Um ponto em comum entre as técnicas de predição apresentadas está no uso das informações de localização dos dispositivos da rede para a realização das predições espaciais. No caso da arquitetura geral do REM (Figura 1.1), os módulos de localização e aquisição são responsáveis por obter a informação de posicionamentos dos dispositivos. Na prática, tais localizações são estimadas através de métodos que utilizam as medidas realizadas pelos dispositivos do sistema de comunicação sem fio [74], [109]. No entanto, os efeitos do canal de propagação (*e.g.*, múltiplos percursos e interferências) e o desempenho dos métodos de estimação de localização da rede não permitem obter absoluta acurácia e precisão nas localizações [74], [76-79]. Isso significa que erros de localização podem ocorrer e, com isso, uma incerteza é associada às estimativas de localização dos dispositivos.

A Figura 1.2 ilustra o problema da incerteza de localização, que se não combatido, pode degradar consideravelmente a geração do REM. Ainda, a relevância da incerteza de localização aumenta em função do uso crescente das informações de localização nas técnicas de gerenciamento e alocação de recursos de rádio em redes de telefonia celular [2], nos requisitos de acurácia de localização impostos pelas regulamentações dos serviços públicos e de emergência [74], além da difusão dos diversos serviços baseados em localização (LBS - *Location-based Services*) [75]. Neste contexto, há um intenso esforço de pesquisa com foco no problema de estimação de localização em comunicações sem fio e redes de sensores [4], [76-77]. No entanto, os impactos provocados pela incerteza de localização sobre o aspecto específico das predições espaciais ainda é alvo recente de pesquisa (*e.g.*, redes móveis e comunicações sem fio [78-80], aplicações com robôs e sensores [81-82]). De fato, apesar da aplicação dos processos Gaussianos (*i.e.*, Kriging na terminologia geoestatística) ser amplamente discutida no contexto das predições espaciais, a maioria dos trabalhos se volta para investigação de erros nas medidas realizadas e não em suas respectivas localizações [82-84]. Este trabalho de doutorado argumenta sobre a necessidade de incorporar a incerteza de localização ao problema de predição espacial e propõe verificar seus impactos na geração do REM.

Figura 1.2: Ilustração da incerteza de localização em sistemas de comunicações sem fio.

1.3 Escopo e Objetivos de Pesquisa

A acurácia da localização alcançada em um sistema de comunicação sem fio depende do tipo de método utilizado² e dos efeitos do canal de propagação nas condições de recepção. Como consequência direta, as estimativas realizadas podem resultar em erros de localização na faixa entre 5 m e 50 m para as técnicas baseadas na tecnologia GPS (*Global Positioning Systems*) [85-86] e até 300 m (ou mais) para as técnicas baseadas em rede [74], [78], [86-87]. Na prática (em operação), erros e falhas nos próprios métodos de localização podem resultar em incertezas de localização ainda maiores, exigindo a configuração e a calibração das aplicações de localização instaladas no sistema de comunicação sem fio. Com isso, apesar da notória evolução dos métodos de estimação de localização³, é razoável admitir que as estimativas de localização envolvidas nos sistemas de comunicações sem fio não são altamente acuradas, especialmente em razão dos efeitos adversos presentes no canal sem fio. Desta forma, é possível destacar o escopo e o objetivo geral de pesquisa, de acordo com:

- **Escopo:** Análise dos problemas de predições espaciais em comunicações sem fio, com enfoque nas consequências provocadas pela incerteza de localização na geração do REM;
- **Objetivo Geral:** Proposta de métodos e técnicas de predição que visam o tratamento dos efeitos negativos dos erros de localização nos processos de predição espacial utilizados no REM.

Especificamente neste trabalho, há um interesse nas técnicas de predição que apresentam menor complexidade e que, ao mesmo tempo, sejam robustas contra os efeitos do canal de propagação. Sobre este aspecto, há um debate científico intenso sobre duas abordagens que estão envolvidas nesta relação de compromisso [89]: Bayesiana e a clássica.

A abordagem Bayesiana possui uma característica interessante, pois permite que distribuições estatísticas conhecidas *a priori* possam ser incorporadas aos modelos de predição para que as distribuições *a posteriori* possam ser obtidas, incluindo diferentes tipos de incerteza (*e.g.*, erros de localização e estimação de parâmetros). No entanto, a realização prática da abordagem Bayesiana enfrenta importantes obstáculos [90]. Primeiramente, a necessidade da especificação adequada das distribuições *a priori* e a consequente complexidade analítica relacionada com as distribuições condicionais necessárias para a estimação de parâmetros e a realização das predições. Esta situação gera um segundo obstáculo, que consiste no alto custo computacional envolvido na realização das integrais exigidas para a conversão de um modelo *a priori* para a distribuição *a posteriori*, dificultando significativamente a busca por preditores ótimos em algum sentido estatístico [63], [68], [90].

²Os métodos de estimação de localização são categorizados de acordo com as atribuições de responsabilidade sobre as medições de sinal e o processamento de estimação de localização [74]. Em geral, é possível dividir os métodos em três categorias, nas quais os processamentos de estimação de localização podem ser realizados pelos (i) próprios dispositivos, (ii) pela rede ou (iii) de forma assistida, onde há uma combinação (híbrida) entre os dispositivos e a rede para a estimação de localização.

³Em [88], é possível encontrar informações detalhadas e propostas de técnicas de posicionamento em redes móveis. Um exemplo promissor é o método híbrido A-GPS (*Assisted-Global Positioning System*), que utiliza a infraestrutura de uma rede móvel para dar o suporte às tarefas de estimação de localização do GPS nos dispositivos, resultando em erros da ordem de 10 m a 50 m [74].

De outro modo, a praticidade da abordagem clássica combinada com os métodos geoestatísticos permite que ferramentas de estimação (*e.g.*, método dos mínimos quadrados e o semivariograma) possam ser utilizadas para a geração do REM. Neste caso, os parâmetros do modelo de predição são constantes desconhecidas e, portanto, devem ser estimadas, enquanto a abordagem Bayesiana assume que cada parâmetro possui uma distribuição estatística, exigindo pressuposições plausíveis para tais distribuições. O funcionamento mais objetivo da geoestatística torna os modelos de predição mais flexíveis, permitindo que soluções (*i.e.*, preditores) menos complexas possam ser obtidas e, por isso, é adotada neste trabalho. De fato, quando o modelo é caracterizado por processos Gaussianos, o preditor Bayesiano (*i.e.*, esperança condicional), que é ótimo no sentido de minimização do erro de predição espacial, converge para a técnica Kriging da geoestatística clássica [63], [68]. Ainda assim, é importante mencionar que a presença da incerteza de localização afeta a pressuposição Gaussiana utilizada na maioria dos modelos de predição tradicionais [78], [83-84]. Neste sentido, apesar da abordagem Bayesiana possibilitar a obtenção de preditores ótimos, porém mais complexos (*e.g.*, preditores não lineares), os esforços deste trabalho são direcionados ao uso de métodos geoestatísticos que combatam o problema da incerteza de localização através de soluções de menor complexidade (*e.g.*, preditores lineares), característica requerida no contexto das diferentes redes de sistemas de comunicações sem fio. Neste aspecto, ressalta-se que ambas as abordagens (clássica e Bayesiana) são formadas por etapas que contemplam uma fase de estimação ou aprendizagem de parâmetros, denotada neste trabalho como treinamento, e uma fase de predições espaciais. A busca por técnicas de predição mais eficientes envolve diversas particularidades nessas duas fases.

1.4 Contribuições

Esta seção apresenta as principais contribuições realizadas através deste trabalho de doutorado: Primeiramente, a formulação de um modelo de predição espacial para a geração do REM é desenvolvida no capítulo 2 (seção 2.3). Aspectos da sua modelagem são discutidos à luz das características específicas dos sistemas de comunicação sem fio. A principal contribuição consiste no uso de ferramentas da geoestatística, visando a extração de informações do ambiente de rádio e permitindo que predições baseadas nas estatísticas das próprias medidas sejam realizadas em sistemas de comunicações sem fio. Dentro deste contexto, uma expressiva parte dos trabalhos se volta para a investigação das predições espaciais, considerando erros nas medidas utilizadas e não em suas respectivas localizações [82-84]. Assim, a introdução dos erros de localização no caso específico dos sistemas de comunicações sem fio é realizada, permitindo tornar o modelo de predição do REM mais realístico. O modelo de incerteza de localização é discutido no capítulo 3 (seção 3.1) e seus impactos na aprendizagem de parâmetros constitui a principal contribuição desta etapa. Simulações baseadas na modelagem analítica do problema permitiram quantificar os impactos da incerteza de localização nas técnicas de estimação do modelo de predição do REM.

Frente ao problema da incerteza de localização, [91-93] apresentam uma metodologia para manipular quantidades analiticamente intratáveis quando há erros de localização nas medidas utilizadas pelos preditores Kriging. Tal metodologia é baseada em um procedimento de ajuste, que utiliza as estatísticas dos erros de localização (*i.e.*, KALE - *Kriging Adjusted for Location Error*), para que predições com maior acurácia sejam alcançadas [92-93] e, por isso, é incluída na formulação do modelo de predição espacial do REM. Contudo, as investigações realizadas mostraram que o impacto dos erros de localização na estimação de parâmetros relacionados ao Kriging compromete os ganhos de desempenho a favor do preditor KALE. Isto significa que a etapa de treinamento do modelo de predição do REM (*i.e.*, aprendizagem de parâmetros) precisa atingir melhores resultados, sobretudo quando o problema da incerteza de localização se junta à necessária estimação de parâmetros.

Este trabalho apresenta uma proposta de preditor espacial contemplando novas técnicas de estimação que expandem a metodologia em [91-93] para lidar com a incerteza de localização na aprendizagem de parâmetros, buscando predições espaciais com maior acurácia na geração do REM, considerando a incerteza de localização. Tais algoritmos de aprendizagem e predição são desenvolvidos (seção 3.3), levando em conta a densidade de probabilidade dos erros de localização para combater os seus problemas e constituem a principal contribuição deste etapa. Especificamente, dois algoritmos (TALE - *Trend Adjusted for Location Error* e SALE - *Semivariogram Adjusted for Location Error*) são combinados na fase de treinamento do modelo para a formulação da técnica de predição espacial proposta TS-KALE (*Trend, Semivariogram an Kriging adjusted for Location Error*). A avaliação de desempenho das predições é realizada com simulações computacionais (capítulo 4), contemplando os preditores espaciais analisados e a técnica de predição proposta. Diferentes fatores do ambiente de rádio (*e.g.*, número de medidas, severidade dos efeitos do canal de propagação e dos erros de localização) são investigados e uma análise de influência permite contribuir para o melhor entendimento de tais fatores e seus impactos nas predições. Curvas teóricas de desempenho são utilizadas para a validação dos algoritmos e um enfoque é dado à questão sobre o conhecimento *a priori* e a necessidade de estimação dos parâmetros do modelo de predição do REM. Além disso, uma seção aborda os diferentes tipos de amostragem espacial utilizados na modelagem da distribuição espacial dos dispositivos na rede dos sistemas de comunicação sem fio. Simulações computacionais destes modelos mostraram a influência do padrão espacial dos dispositivos e, conseqüentemente, das medidas capturadas para gerar as predições espaciais do REM.

Finalmente, um estudo experimental do REM é apresentado no capítulo 5. Uma campanha de medidas foi conduzida na faixa de frequência de ondas milimétricas com o propósito de aplicar os algoritmos de aprendizagem e predição desenvolvidos, permitindo a geração do REM na prática. A introdução de erros de localização nas medidas geolocalizadas foi realizada por meio de um experimento para testar a robustez das técnicas de predição. A análise dos resultados práticos consiste na principal contribuição desta etapa, permitindo que discussões sejam realizadas, na busca da compreensão mais aprofundada da aplicação das técnicas de predição em comunicações sem fio.

1.5 Organização da Tese

Os capítulos que compõem esta tese de doutorado estão organizados da seguinte forma:

- **Capítulo 2:** reúne os fundamentos relacionados com as ferramentas da geoestatística, que foram diretamente utilizadas na formulação do modelo de predição espacial para a geração do REM. Simulações sobre a aprendizagem de parâmetros e a aplicabilidade da teoria de processos Gaussianos ao caso dos sistemas de comunicações sem fio são discutidas neste capítulo.
- **Capítulo 3:** introduz a incerteza de localização ao modelo de sistema e faz uma análise dos seus impactos nas etapas do modelo de predição espacial, além de propor um preditor espacial composto por técnicas de estimação que visam combater os efeitos discutidos.
- **Capítulo 4:** realiza uma análise de desempenho das predições espaciais das principais técnicas de predição usadas no REM, bem como do preditor espacial proposto. Diferentes aspectos do ambiente de rádio e seus impactos nas predições são discutidos.
- **Capítulo 5:** apresenta um estudo experimental do REM, realizado com dados práticos obtidos a partir de uma campanha de medidas. Investigações e análises são realizadas a fim de explorar o comportamento e o desempenho dos algoritmos de aprendizagem e predição espacial desenvolvidos no trabalho.
- **Capítulo 6:** apresenta as conclusões e o apontamento de novas oportunidades de pesquisa.

1.6 Publicações geradas

No contexto de contribuição, o Anexo A apresenta as publicações científicas diretamente relacionadas com as pesquisas realizadas neste trabalho de doutorado:

- R. Augusto e C. M. Panazio, “Sobre a Geração de Mapas de Ambiente de Rádio com Kriging Ordinário,” in *Proc. XXXIV Brazilian Symposium on Telecommunications (SBrT)*, Ago. 2016.
- R. Augusto e C. M. Panazio, “On Geostatistical Methods for Radio Environment Maps Generation under Location Uncertainty,” *Journal of Communication and Information Systems (JCIS)*, vol. 33, no. 01, Abr. 2018.

Capítulo 2

Modelo de Predição Espacial para a Geração do REM

Este capítulo possui o objetivo de apresentar e discutir as características de cada etapa do modelo de predição espacial para a geração do REM. Trata-se de um modelo baseado na teoria de processos aleatórios espaciais fundamentada especificamente no domínio espacial, que utiliza ferramentas da geoestatística e dos processos Gaussianos para a geração das predições espaciais. Desta forma, entende-se que seria apropriado abordar os conceitos de análise espacial da geoestatística que foram diretamente aplicados ao modelo de predição do REM. Neste sentido, tópicos sobre a teoria de variáveis regionalizadas, pressuposições de estacionaridade, além das propriedades das funções covariância e semivariograma são abordados neste capítulo. Detalhes mais específicos sobre estes conceitos e ferramentas podem ser encontrados em [58], [60], [62-68].

2.1 Conceitos da Geoestatística

A geoestatística teve seu início na década de 1950 com D. Krige, G. Matheron, B. Matérn, A. N. Kolmogorov, além de outros pesquisadores que atuavam em diferentes áreas [60], [62-68], destacando as ciências da Terra e do clima, além da indústria de mineração, óleo e gás. Desde então, a evolução do campo da geoestatística foi marcada por fundamentações teóricas que possibilitaram a sua expansão para diversas áreas do conhecimento. Neste contexto, compreender estatisticamente os fenômenos espaciais por trás das várias aplicações consiste no principal propósito da geoestatística, para que problemas sobre predições espaciais, inferências e decisões possam ser solucionados de forma mais eficaz. A aplicabilidade destes conceitos na área de sistemas de comunicações sem fio se fundamenta necessariamente na caracterização do ambiente de rádio (*i.e.*, canal de propagação) como um processo aleatório espacial, cujas características são descritas adiante.

2.1.1 Processos Aleatórios Espaciais

O estudo sobre os processos aleatórios espaciais é baseado na escolha de modelos que representam de forma adequada o fenômeno espacial em análise. Especificamente, essas representações envolvem descrições matemáticas, estatísticas e visuais do processo aleatório espacial. O principal objetivo da geoestatística consiste em prover uma descrição estatística sobre a variabilidade espacial do fenômeno para que seja possível investigá-lo por meio de modelos geoestatísticos, ou seja, que se fundamentam essencialmente no domínio espacial.

Um modelo geoestatístico amplamente utilizado para descrever os processos aleatórios espaciais é formado por duas componentes definidas em todas as coordenadas espaciais $\mathbf{s} = (s_x, s_y)$, *i.e.*,

$$P(\mathbf{s}) = \mu(\mathbf{s}) + \xi(\mathbf{s}), \text{ com } \mathbf{s} \in D, \quad (2.1)$$

em que $\mu(\mathbf{s})$ consiste na média do processo aleatório espacial e $\xi(\mathbf{s})$, que representa as variações aleatórias do processo $P(\mathbf{s})$ sobre a média $\mu(\mathbf{s})$ ao longo do domínio espacial D , definido em duas dimensões neste trabalho (*i.e.*, $D \in \mathbb{R}^2$). Sob o ponto de vista estatístico, o processo $P(\mathbf{s})$ é composto por um conjunto infinito de variáveis aleatórias $P(\mathbf{s}_i)$, cujas realizações são governadas por mecanismos aleatórios, caracterizados pela distribuição de probabilidade i -dimensional F_P ,

$$F_P = \text{Prob}[P(\mathbf{s}_1) < p(\mathbf{s}_1), \dots, P(\mathbf{s}_i) < p(\mathbf{s}_i)], \quad (2.2)$$

em que $P(\mathbf{s}_i)$ e $p(\mathbf{s}_i)$ consistem na variável aleatória (VA) e a sua realização na coordenada espacial $\mathbf{s}_i$, respectivamente. A Figura 2.1 ilustra a realização de um processo aleatório espacial¹ que apresenta flutuações aleatórias ao longo do espaço, geradas a partir da distribuição de probabilidade Gaussiana, ou seja, cada VA que compõe este processo aleatório espacial é Gaussiana. Assim, $P(\mathbf{s})$ é dito Gaussiano se a distribuição conjunta de $\{P(\mathbf{s}_1), \dots, P(\mathbf{s}_i)\}$ para qualquer conjunto de coordenadas espaciais $\{\mathbf{s}_1, \dots, \mathbf{s}_i\}$ é do tipo Gaussiana multivariada [58].

Os processos Gaussianos multivariados são especificados por meio das funções média $\mu(\mathbf{s})$ e covariância $C(\mathbf{s}_i, \mathbf{s}_j)$, e denotados neste trabalho com a nomenclatura $\mathcal{GP}(\mu(\mathbf{s}), C(\mathbf{s}_i, \mathbf{s}_j))$. Mais adiante, será mostrado que esta caracterização estatística Gaussiana, bem como o comportamento das funções $\mu(\mathbf{s})$ e $C(\mathbf{s}_i, \mathbf{s}_j)$, são aspectos relevantes para a modelagem do ambiente de rádio aplicada ao REM em comunicações sem fio. A média de um processo aleatório espacial $\mu(\mathbf{s}) = \mathbb{E}\{P(\mathbf{s})\}$ consiste no valor esperado de $P(\mathbf{s})$ em uma posição espacial $\mathbf{s}_i$, *i.e.*, $\mu(\mathbf{s}_i) = \mathbb{E}\{P(\mathbf{s}_i)\}$. Assim, dependendo das características de $P(\mathbf{s})$, é possível que $\mu(\mathbf{s})$ assumam diferentes valores ao longo das coordenadas espaciais $\mathbf{s}$ no domínio D . Nestas circunstâncias, o processo aleatório espacial $P(\mathbf{s})$ exibe uma característica denotada como tendência espacial [62-63], [66].

¹Na literatura científica é comum encontrar as expressões campo espacial e campo aleatório para a denotação do processo aleatório espacial. Neste trabalho, não serão feitas distinções entre estas expressões, embora a última seja utilizada nas principais referências deste trabalho, além de amplamente conhecida na comunidade científica.

Figura 2.1: Realização de um processo aleatório espacial Gaussiano.

Na análise de processos aleatórios espaciais visando a realização de predições, a identificação da tendência $\mu(\mathbf{s})$ é um passo importante e, na maioria dos casos, requer algum tipo de tratamento estatístico. Em geral, este tratamento pode ser feito de duas formas: i) incorporando a modelagem selecionada para a tendência ao funcionamento do preditor espacial e ii) fazendo a estimação seguida da remoção da tendência dos dados para o posterior processamento e análise. Este trabalho discutirá o uso dessas abordagens, visto que o ambiente de rádio exibe um tipo comportamento que requer a modelagem da tendência espacial. No exemplo da Figura 2.1, o processo $P(\mathbf{s})$ não exibe tendência, pois foi gerado com média nula, *i.e.*, $\mu(\mathbf{s}) = \mathbb{E}\{P(\mathbf{s})\} = 0$. Ainda assim, é possível visualizar as variações espaciais ocorrendo na forma de aglomerados² ao longo do domínio D .

Um parâmetro importante dos processos aleatórios espaciais é a sua variância $\sigma^2 = \text{Var}\{P(\mathbf{s})\}$, definida como a variância associada a cada coordenada espacial, $\sigma^2(\mathbf{s}_i) = \text{Var}\{P(\mathbf{s}_i)\}$. Evidentemente, trata-se de uma medida de dispersão de cada VA e, portanto, quantifica a variabilidade contida em $P(\mathbf{s})$. Neste contexto, a heterocedasticidade³ é refletida em flutuações da variância de cada VA ao longo do espaço, enquanto a homocedasticidade implica na variância constante para o processo $P(\mathbf{s})$. Neste ponto, já se torna importante mencionar que as pressuposições de estacionariedade (discutidas adiante) utilizadas na análise geoestatística se relacionam com o comportamento destas medidas estatísticas (média e variância) de $P(\mathbf{s})$.

²Os aglomerados espaciais consistem em regiões nas quais os valores do processo aleatório são similares e sua identificação sugere (mas não constata) a presença de correlação espacial no processo. Assim, de acordo com os valores do processo aleatório, os aglomerados podem ser de alta intensidade (*hotspots*) ou de baixa intensidade (*coldspots*).

³Neste caso, o processo aleatório espacial não possui variância constante σ^2 nas coordenadas $\mathbf{s}$, afetando as análises realizadas (*e.g.*, cálculo de intervalos de confiança) [65]. Assim, a utilização de transformações nas variáveis se torna necessária para tratar o problema. Além disso, existem testes e ferramentas estatísticas que podem ser aplicados sobre os dados para que a heterocedasticidade seja identificada.

Com estas informações, nota-se que as propriedades estatísticas das VAs espaciais que compõem $P(\mathbf{s})$ determinam a relação de dependência entre tais variáveis aleatórias. Esta relação consiste na caracterização estatística espacial de $P(\mathbf{s})$, que é descrita por sua covariância espacial,

$$C(\mathbf{s}_i, \mathbf{s}_j) = \text{cov}(P(\mathbf{s}_i), P(\mathbf{s}_j)) = \mathbb{E}\{[P(\mathbf{s}_i) - \mu(\mathbf{s}_i)][P(\mathbf{s}_j) - \mu(\mathbf{s}_j)]\}. \quad (2.3)$$

Recorrendo ao modelo (2.1), verifica-se que as flutuações espaciais aleatórias $\xi(\mathbf{s})$ estão relacionadas com (2.3), que expressa a similaridade entre as VAs e requer o conhecimento da média do processo $P(\mathbf{s})$ nas coordenadas espaciais $\mathbf{s}_i$ e $\mathbf{s}_j$. De outro modo, a ideia de dissimilaridade espacial entre as VAs é descrita pela função semivariograma, amplamente difundida na área geoestatística⁴ e definida como a metade da variância das diferenças entre as VAs que compõem $P(\mathbf{s})$,

$$\begin{aligned} \gamma(\mathbf{s}_i, \mathbf{s}_j) &= \frac{1}{2} \text{Var}\{P(\mathbf{s}_i) - P(\mathbf{s}_j)\} \\ &= \frac{1}{2} \mathbb{E}\{[P(\mathbf{s}_i) - P(\mathbf{s}_j) - \mathbb{E}\{P(\mathbf{s}_i) - P(\mathbf{s}_j)\}]^2\}. \end{aligned} \quad (2.4)$$

Na literatura geoestatística, denota-se $2\gamma(\mathbf{s}_i, \mathbf{s}_j)$ como variograma e $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ como semivariograma, sendo que ambas as funções transmitem a ideia de dissimilaridade. A razão relacionada com o fator multiplicativo (2) é explicada em detalhes na subseção 2.3.2.

A covariância e o semivariograma de um processo aleatório espacial $P(\mathbf{s})$ são ditos estruturados se os valores não nulos de $C(\mathbf{s}_i, \mathbf{s}_j)$ e $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ apresentarem um comportamento interpretável e que possa ser descrito por modelos analíticos de covariância e semivariograma. Assim, a análise destas funções é importante para que a caracterização estatística espacial de $P(\mathbf{s})$ possa ser realizada, envolvendo a relação entre $\mu(\mathbf{s})$ e $\xi(\mathbf{s})$, enquanto a utilização destas funções permite que predições espaciais e inferências possam ser realizadas sobre o processo $P(\mathbf{s})$. Desta forma, a seleção de modelos apropriados para (2.3) e (2.4) em relação ao fenômeno estudado é essencial, pois permitirá atingir melhores resultados de predições sobre o processo $P(\mathbf{s})$, não somente no sentido tradicional, *i.e.*, predições enviesadas ou não enviesadas, mas em termos de acurácia, *i.e.*, com erros de predição menores. Isto é possível porque a formulação dos preditores geoestatísticos utiliza métodos que exploram a covariância espacial do processo aleatório a favor das predições espaciais.

Dentro deste contexto, o formato das funções $C(\mathbf{s}_i, \mathbf{s}_j)$ e $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ é um aspecto relevante e depende da distância entre as observações de $P(\mathbf{s})$ através de $\mathbf{h} = (h_x, h_y)$, definido como um vetor bidimensional de separação espacial entre as coordenadas espaciais $\mathbf{s}$, *i.e.*, $\mathbf{h} \equiv \mathbf{s}_i - \mathbf{s}_j$. Em outras palavras, a forma como estas funções se comportam em função de $\mathbf{h}$ indica que observações do processo $P(\mathbf{s})$ que estão relativamente próximas apresentam alta similaridade, enquanto observações distantes são ditas espacialmente descorrelacionadas (apresentam baixa similaridade ou alta dissimilaridade).

⁴Embora as ideias de similaridade (correlação) e dissimilaridade (descorrelação) possam ser extraídas de ambas as funções, o uso do semivariograma é mais difundido na geoestatística, em razão deste não exigir a estimação dos valores médios do processo aleatório no cômputo de cada diferença entre os pares de observações [67].

Este conceito sobre a covariação dos atributos das VAs espaciais na modelagem de $P(\mathbf{s})$ constitui a ideia central da geoestatística para que o uso das funções covariância e semivariograma possibilite a geração de predições espaciais com melhor acurácia. Na prática, as funções espaciais (2.3) e (2.4) são estimadas a partir de um conjunto de medidas coletado, para que as estatísticas calculadas possam quantificar a covariância e o semivariograma do ambiente de rádio. Com isso, é possível mencionar que a formulação do modelo de predição espacial do REM passa pela modelagem do ambiente de rádio, que é baseada na seleção adequada dos modelos da tendência, covariância, semivariograma e sobretudo das pressuposições de estacionariedade dos processos aleatórios espaciais. Neste ponto introdutório, é interessante visitar duas questões embasadoras: i) a relação das pressuposições de estacionariedade com as funções covariância e semivariograma e ii) quais são as propriedades mais importantes destas funções para o propósito de geração das predições espaciais em ambientes de rádio de sistemas de comunicações sem fio.

2.1.2 Estacionariedade

O estudo das hipóteses de estacionariedade consiste em um tópico amplamente difundido sobre a teoria de processos aleatórios. Ainda assim, a aplicação das pressuposições de estacionariedade aos processos aleatórios em domínios espaciais implica no surgimento de conceitos mais abstratos como a noção de equilíbrio estacionário⁵ e a ergodicidade⁶ relacionados a este caso espacial.

Neste contexto, um dos principais problemas da geoestatística consiste no acesso a várias realizações dos processos aleatórios espaciais [63], [67]. De fato, para que se tenha consistência estatística nas predições e inferências, mais realizações do processo aleatório espacial são necessárias. Entretanto, na maioria dos casos práticos, poucas realizações do processo aleatório são acessíveis. Assim, a aplicação das pressuposições de estacionariedade se torna fundamentalmente útil para a geoestatística, pois permite assumir que os mesmos mecanismos aleatórios governam as várias realizações do processo $P(\mathbf{s})$ e com isso, substitui-se as repetições inacessíveis das várias realizações por repetições ao longo do espaço. Desta forma, a noção de equilíbrio estacionário indica que a distribuição de probabilidade conjunta do processo aleatório é a mesma ao longo do domínio espacial D , ou seja, não depende especificamente das coordenadas espaciais $\mathbf{s}$. Isto significa que observações tomadas em diferentes localidades de um processo aleatório espacial estacionário podem ser interpretadas como sendo realizações do mesmo processo [63], [67], [89], [94].

⁵A noção de equilíbrio estacionário estabelece uma consistência estatística sobre as realizações dos processos aleatórios espaciais [94]. Trata-se da projeção das hipóteses de estacionariedade da estatística tradicional (*e.g.*, séries temporais) sobre o domínio espacial, indicando que a distribuição de probabilidade conjunta que governa o processo aleatório é invariante à translações espaciais, isto é, não depende especificamente da posição das coordenadas $\mathbf{s}$.

⁶A ergodicidade se relaciona com as hipóteses de estacionariedade e diz respeito à convergência entre as estatísticas calculadas, a partir de uma única realização do processo aleatório espacial e os valores esperados do processo, considerando mais realizações (*e.g.*, ergodicidade da média). Em outras palavras, um processo aleatório espacial é ergódico se as estatísticas computadas para uma única realização são representativas no sentido de obter a caracterização estatística de tal processo, considerando mais realizações. Assim, a informação a ser extraída a partir de uma única realização do processo é a mesma no comparativo com várias realizações. Detalhes específicos do conceito da ergodicidade aplicado ao caso espacial podem ser encontrados em [63, pp. 19].

Neste sentido, de acordo com a teoria de processos aleatórios, três hipóteses de estacionariedade podem ser mencionadas neste caso dos domínios espaciais:

- Hipótese de estacionariedade no sentido restrito: Estabelece que um processo aleatório espacial é estritamente estacionário se a família de VAs espaciais $P(\mathbf{s}_1), P(\mathbf{s}_2), \dots, P(\mathbf{s}_i)$ e $P(\mathbf{s}_1 + \mathbf{h}), P(\mathbf{s}_2 + \mathbf{h}), \dots, P(\mathbf{s}_i + \mathbf{h})$ é caracterizada estatisticamente pela mesma função distribuição de probabilidade conjunta F_P para todo i e $\mathbf{h}$ no domínio espacial D . Isto significa que a distribuição F_P é independente das posições espaciais específicas $\mathbf{s}$ e não é alterada por translações vetoriais provocadas pelo vetor de separação espacial $\mathbf{h}$ [63], [67].
- Hipótese de estacionariedade no sentido amplo: Esta hipótese é voltada para a estacionariedade de aspectos específicos da distribuição de probabilidade (*i.e.*, momentos) e estabelece que um processo aleatório espacial é estacionário no sentido amplo se o primeiro momento é constante, *i.e.*, $\mu(\mathbf{s}) = \mathbb{E}\{P(\mathbf{s})\} = \mu$ e a covariância espacial $C(\mathbf{s}_i, \mathbf{s}_j)$ é invariante às translações espaciais, ou seja, dependa somente do vetor de separação espacial $\mathbf{h}$ [63], [67].
- Hipótese de estacionariedade intrínseca: Esta hipótese se concentra na diferença entre as observações de $P(\mathbf{s})$ e estabelece que um processo aleatório espacial é intrinsecamente estacionário se, para qualquer vetor de distância de separação espacial $\mathbf{h}$, as diferenças $[P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i)]$ no domínio espacial D , são estacionárias no sentido amplo [63], [67].

Desta forma, a modelagem espacial do fenômeno sob análise e o entendimento de suas características são pontos relevantes para a aplicação das hipóteses de estacionariedade em processos aleatórios espaciais. Na prática, o uso da estacionariedade é ponderado por três aspectos: i) a presença da tendência; ii) a heterocedasticidade e iii) a presença de uma covariância espacial estruturada. Especificamente, no caso do ambiente de rádio deste trabalho, o comprometimento da estacionariedade⁷ ocorre somente em razão da tendência, ou seja, das flutuações da potência média de recepção ao longo das coordenadas $\mathbf{s}$, enquanto a variância do processo $P(\mathbf{s})$ é assumida constante (*i.e.*, homocedasticidade). Isto significa que o tratamento da tendência do ambiente de rádio se torna necessário, enquanto a covariância espacial de $P(\mathbf{s})$ é definida para todo o domínio espacial D . Assim, tendo em vista o objetivo de predições no REM, ignorar a presença de uma correlação estruturada no ambiente de rádio pode levar a resultados não enviesados, mas não necessariamente com boa acurácia. De outro modo, fazer o tratamento da tendência e capturar a variabilidade espacial contida no processo aleatório poderá contribuir para a realização de predições espaciais mais acuradas. As próximas subseções apresentam as principais propriedades das funções covariância e semivariograma, considerando a estacionariedade e que serão úteis para a modelagem do processo aleatório espacial que caracteriza o ambiente de rádio.

⁷Processos aleatórios espaciais cujas médias ou variâncias dependam das coordenadas espaciais $\mathbf{s}$ exibem características não estacionárias podem dificultar as análises e a aplicação dos métodos baseados nas funções covariância e semivariograma. Nestes casos, é recomendável recorrer a modelagem não estacionária destas funções [64].

2.1.3 Covariância e Semivariograma - Propriedades

A modelagem dos processos aleatórios espaciais é fundamentada na especificação da distribuição estatística das VAs que formam o processo $P(\mathbf{s})$. Com isso, a relação estatística entre tais VAs define a caracterização espacial do processo $P(\mathbf{s})$ e sua formulação é descrita por modelos analíticos para as funções covariância e semivariograma. Neste trabalho, tais modelos são paramétricos e apresentam certas propriedades. Esta subseção descreve as principais propriedades da função covariância e, em seguida, as informações sobre o semivariograma são colocadas, considerando a pressuposição de estacionariedade no sentido amplo. Isto permite descrever a relação direta entre estas duas funções, que também será útil na etapa de modelagem do ambiente de rádio.

Sobre a pressuposição de estacionariedade no sentido amplo, ressalta-se a invariância à translação espacial $\mathbf{h}$, a qual indica que a covariância entre quaisquer coordenadas espaciais de um processo $P(\mathbf{s})$ estacionário não dependerá das posições espaciais específicas, mas do vetor de separação entre as coordenadas, de acordo com

$$C(\mathbf{s}_i, \mathbf{s}_i + \mathbf{h}) = \mathbb{E}\{[P(\mathbf{s}_i) - \mu][P(\mathbf{s}_i + \mathbf{h}) - \mu]\} = C(\mathbf{h}). \quad (2.5)$$

É importante mencionar que o vetor de separação espacial $\mathbf{h}$ expressa o módulo (distância) e o ângulo (direção) de separação entre duas coordenadas espaciais $\mathbf{s}_i$ e $\mathbf{s}_j$. Com isso, se a covariância espacial for independente da direção, *i.e.*, depender somente da distância entre as coordenadas, $|\mathbf{h}| = |\mathbf{s}_i - \mathbf{s}_j|$, o modelo de covariância apresenta a propriedade de isotropia e, neste caso, o processo aleatório espacial é denotado como estacionário e isotrópico. De outro modo, a anisotropia surge quando a covariância depende da direção ou ângulo de separação entre as observações, o que exige uma modelagem anisotrópica para os modelos da covariância e do semivariograma, tornando-os, conseqüentemente, mais complexos [125]. Adiante, será discutido que o uso da propriedade de isotropia, além de simplificar os procedimentos envolvidos nas técnicas de estimação, também é adequada ao caso da modelagem do ambiente de rádio.

A importância da seleção de um modelo de covariância decorre da sua representatividade sobre o ponto de vista do comportamento do processo aleatório espacial $P(\mathbf{s})$ e da sua adequação em relação aos métodos de estimação a serem utilizados para capturar a variabilidade espacial das VAs. Neste sentido, $C(\mathbf{s}_i, \mathbf{s}_j)$ deve indicar como se dá a covariação de uma VA localizada em $\mathbf{s}_i$ com outra VA, posicionada em $\mathbf{s}_j$, do mesmo processo $P(\mathbf{s})$. Em grande parte dos modelos de covariância, a covariação indicada por $C(\mathbf{s}_i, \mathbf{s}_j)$ atingirá valores menores à medida em que $|\mathbf{h}| \rightarrow \infty$, ou seja, VAs do processo $P(\mathbf{s})$ que estão muito distantes entre si são ditas espacialmente descorrelacionadas, enquanto VAs próximas apresentam forte correlação espacial. Ainda, na modelagem do ambiente de rádio em comunicações sem fio, (2.5) também pode ser interpretada como uma função de autocovariância. Em todo caso, neste trabalho, não será feita uma distinção no uso dos termos autocovariância e covariância dos processos aleatórios para designação da função $C(\mathbf{s}_i, \mathbf{s}_j)$.

Uma característica importante da função covariância consiste na sua propriedade definida positiva, a qual pode ser descrita da seguinte forma

$$\sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) \geq 0, \quad \forall i, j \in \mathbb{N}^*, \quad \forall \lambda \in \mathfrak{R} \text{ e } \forall \mathbf{s} \in D. \quad (2.6)$$

Demonstração. Considere a combinação linear $\sum_i \lambda_i P(\mathbf{s}_i)$. O cálculo da variância desta combinação linear permite obter o seguinte resultado [67]:

$$\text{Var}\left\{\sum_i \lambda_i P(\mathbf{s}_i)\right\} = \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j), \text{ com } \text{Var}\left\{\sum_i \lambda_i P(\mathbf{s}_i)\right\} \geq 0. \quad (2.7)$$

□

Como a variância desta combinação é positiva para todo $\lambda \in \mathfrak{R}$ e $\forall \mathbf{s} \in D$, segue que a covariância $C(\mathbf{h})$ é necessariamente definida positiva. De fato, esta é a principal condição a ser atendida para que uma função arbitrária seja um modelo analítico de covariância válido [67]. Também é importante mencionar que esta propriedade possui reflexo direto na função semivariograma. Para compreender esta questão é necessário destacar a relação entre a função covariância e o semivariograma, baseando-se na estacionaridade no sentido amplo. Primeiramente, a função teórica do semivariograma em (2.4) pode ser expandida, resultando na seguinte expressão

$$\begin{aligned} \gamma(\mathbf{h}) &= \frac{1}{2} \text{Var}\{P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i)\} \\ &= \frac{1}{2} \mathbb{E}\{[P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i) - \mathbb{E}\{P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i)\}]^2\} \\ &= \frac{1}{2} \mathbb{E}\{[P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i)]^2\}. \end{aligned} \quad (2.8)$$

A equação (2.8) mostra que o semivariograma é determinado pela metade do valor esperado das diferenças quadráticas entre as observações espacialmente correlacionadas do processo $P(\mathbf{s})$. Esta definição auxiliará no entendimento dos métodos utilizados para a estimação do semivariograma por meio da média das diferenças quadráticas para um dado vetor de separação espacial $\mathbf{h}$. Assim, torna-se possível estabelecer a relação entre $\gamma(\mathbf{h})$ e $C(\mathbf{h})$, de acordo com [67]

$$\begin{aligned} \gamma(\mathbf{h}) &= \frac{1}{2} \text{Var}\{P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i)\} \\ &= \frac{1}{2} [\text{Var}\{P(\mathbf{s}_i + \mathbf{h})\} + \text{Var}\{P(\mathbf{s}_i)\} - 2\text{cov}\{P(\mathbf{s}_i + \mathbf{h}), P(\mathbf{s}_i)\}] \\ &= \frac{1}{2} \text{Var}\{P(\mathbf{s}_i + \mathbf{h})\} + \frac{1}{2} \text{Var}\{P(\mathbf{s}_i)\} - \text{cov}\{P(\mathbf{s}_i + \mathbf{h}), P(\mathbf{s}_i)\} \\ &= \frac{1}{2} \text{Var}\{P(\mathbf{s}_i + \mathbf{h})\} + \frac{1}{2} \text{Var}\{P(\mathbf{s}_i)\} - C(\mathbf{h}) \end{aligned} \quad (2.9)$$

O segundo momento (*i.e.*, variância σ^2) constante, que caracteriza a estacionaridade no sentido amplo, permite fazer o rearranjo dos termos $C(\mathbf{h})$ e $\gamma(\mathbf{h})$ e obter a relação direta entre as funções covariância e semivariograma, dada por

$$\begin{aligned}\gamma(\mathbf{h}) &= \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 - C(\mathbf{h}) \\ C(\mathbf{h}) &= \sigma^2 - \gamma(\mathbf{h}).\end{aligned}\tag{2.10}$$

A relação em (2.10) permite mencionar o reflexo da propriedade definida positiva da função covariância na propriedade definida negativa da função semivariograma. De fato, pela mesma razão expressa em (2.7), a função semivariograma deve ser tal que garanta valores não negativos para a variância do processo $P(\mathbf{s})$. De forma analítica, realiza-se a substituição da relação encontrada em (2.10) na propriedade em (2.7) levando a seguinte demonstração

$$\sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{h}) \leq 0, \quad \forall i, j \in \mathbb{N}^*, \quad \forall \lambda \in \mathbb{R}, \quad \forall \mathbf{s} \in D \text{ com } \sum_i \lambda_i = 0\tag{2.11}$$

Demonstração.

$$\begin{aligned}\text{Var}\left\{\sum_i \lambda_i P(\mathbf{s}_i)\right\} &= \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j), \\ &= \sum_i \sum_j \lambda_i \lambda_j \{\sigma^2 - \gamma(\mathbf{s}_i, \mathbf{s}_j)\} \\ &= \sigma^2 \sum_i \sum_j \lambda_i \lambda_j - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j) \\ &= \sigma^2 \sum_i \lambda_i \sum_j \lambda_j - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j).\end{aligned}\tag{2.12}$$

□

Nota-se em (2.12) que o condicionamento aplicado ao somatório dos pesos λ_i da combinação linear é importante, sobretudo quando o processo aleatório espacial é intrinsecamente estacionário (onde tem-se a possibilidade de indefinição da variância σ^2), impondo a condição citada, ou seja, a função semivariograma é condicionalmente definida negativa [67].

Conforme mencionado, a relação estatística entre as VAs do processo $P(\mathbf{s})$ determina o comportamento da covariância e do semivariograma. De forma similar ao caso de $C(\mathbf{h})$, a utilização de um modelo analítico para $\gamma(\mathbf{h})$ busca descrever uma relação entre a intensidade do semivariograma e a distância entre as coordenadas espaciais $\mathbf{s}$. A diferença é que, no caso do semivariograma, esta descrição deve refletir como a dissimilaridade entre as observações de $P(\mathbf{s})$ evolui em função da separação $\mathbf{h}$. Assim, se os valores não nulos de $\gamma(\mathbf{h})$ apresentarem um formato interpretável por meio de um modelo analítico, o semivariograma é dito espacialmente estruturado.

Um ponto importante é que estes modelos funcionais analíticos são caracterizados por um vetor de parâmetros θ , que deverá ser estimado a partir das observações coletadas. De forma geral, tem-se o interesse no comportamento de três parâmetros específicos dos modelos analíticos, *i.e.*, $\theta = [m \ r \ c_0]$: i) o parâmetro m , conhecido como variância *sill* na terminologia geoestatística, que expressa a variância σ^2 do processo aleatório espacial; ii) o parâmetro r , denotado como a distância *range*, que tem relação direta com o decaimento (covariância) e com a inclinação (semivariograma) dos modelos analíticos e iii) o parâmetro c_0 , que consiste na variância *nugget*, conhecida na geoestatística como efeito pepita⁸ e que indica uma descontinuidade na origem das funções covariância e semivariograma, ou seja, variações abruptas de $C(\mathbf{h})$ e $\gamma(\mathbf{h})$ em separações espaciais $\mathbf{h}$ muito pequenas.

A Figura 2.2 apresenta um exemplo de modelo exponencial para as funções covariância (a) e semivariograma (b). É possível encontrar diferentes modelos na literatura geoestatística como: esférico, exponencial, Gaussiano, J-Bessel, Matérn, linear, entre outros [62-67]. As particularidades de cada modelo estão relacionadas com o seu comportamento na origem ($|\mathbf{h}| \rightarrow 0$), com a sua estrutura (formato da função) e se existe um patamar alcançado em regime assintótico ($|\mathbf{h}| \rightarrow \infty$). Informações mais específicas sobre os modelos analíticos podem ser encontradas em [63], [67]. A seleção dos modelos de covariância/semivariograma adequados e a interpretação do significado de seus parâmetros são pontos que devem ser levados em conta para que a modelagem de sistema seja plausível, além de estar em conformidade com a formulação do problema a ser tratado. Em particular, o modelo analítico para a covariância/semivariograma do tipo exponencial será útil no desenvolvimento deste trabalho, em razão da sua adequação às características do ambiente do rádio, explicadas nas próximas seções.

Figura 2.2: Modelos analíticos: (a) função covariância e (b) função semivariograma.

⁸A expressão efeito pepita deriva das análises e investigações da geoestatística aplicadas na indústria da mineração na busca por minérios raros, cuja presença é indicada por este comportamento nos resultados de estimação da covariância e do semivariograma.

2.1.4 Variáveis Regionalizadas

A fundamentação teórica da geoestatística associa uma coordenada espacial $\mathbf{s}$ à cada VA que compõe o processo $P(\mathbf{s})$ e, com isso, a realização do processo $P(\mathbf{s})$ consiste nos valores assumidos por cada VA, *i.e.*, $P(\mathbf{s}_i)$. No campo da geoestatística, esta realização é denominada variável regionalizada (VR). A aplicação dos métodos da probabilidade e estatística à realização das VAs ao longo de um domínio espacial define a teoria das variáveis regionalizadas. Detalhes mais específicos sobre a fundamentação da VR podem ser acessados em [63-64], [66-67].

A Figura 2.3 mostra a realização de um processo aleatório espacial Gaussiano, gerado com média nula e covariância espacial do tipo exponencial. Um conjunto de coordenadas espaciais $\mathbf{s}$, gerado aleatoriamente a partir da densidade de probabilidade uniforme, é utilizado para realização da amostragem espacial sobre o processo $P(\mathbf{s})$. O resultado consiste no conjunto de medidas $\{P(\mathbf{s}_i), i = 1, \dots, N\}$, em que N consiste no número total de medidas coletadas. Na prática, isto significa que o conjunto de medidas capturado é somente uma parte observável da VR do processo $P(\mathbf{s})$. Por simplicidade, tal conjunto também é denotado como a VR neste trabalho, ou seja, os valores regionalizados formam as N medidas que os métodos geoestatísticos possuem para que a estimação das funções covariância e semivariograma sejam realizadas, visando as posteriores predições espaciais.

No ambiente de rádio, esta questão toma importância, pois depende da capacidade dos dispositivos e da própria rede em realizar as medidas de potência de recepção para a geração do REM. Além disso, a disposição dos pontos da amostragem espacial depende das aplicações dos sistemas de comunicações sem fio e, consequentemente, dos padrões espaciais dos dispositivos no ambiente de rádio. Com isso, a forma como estes conceitos são abordados em um sistema de comunicação sem fio retrata a necessidade de formulação de um modelo de sistema.

Figura 2.3: Valores regionalizados obtidos a partir de um processo aleatório espacial Gaussiano.

2.2 Modelagem do Ambiente de Rádio

O modelo de sistema utilizado neste trabalho é caracterizado pelo ambiente de rádio de um sistema de comunicação sem fio composto por vários dispositivos e uma estação rádio base (ERB). Assume-se que tais dispositivos são capazes de realizar as medidas de intensidade do sinal recebido, *i.e.*, potência de recepção, e enviá-las para a ERB, a qual coordena toda a operação da rede do sistema de comunicação sem fio. Isto significa que a ERB é responsável pelo processamento das medidas coletadas e pela geração do REM. Os níveis de potência de recepção nos dispositivos são influenciados diretamente pelos mecanismos de propagação do ambiente de rádio como as difrações, reflexões, refrações, bem como a perda por percurso média e, consequentemente, podem afetar a geração das predições espaciais do REM.

Neste trabalho, o ambiente de rádio consiste em um processo aleatório espacial $P(\mathbf{s})$ definido em todas as coordenadas espaciais $\mathbf{s} = (s_x, s_y)$ e composto por duas componentes, *i.e.*,

$$P(\mathbf{s}) = \mu(\mathbf{s}) + \xi(\mathbf{s}), \text{ com } \mathbf{s} \in D, \quad (2.13)$$

em que $\mu(\mathbf{s})$ é a tendência do modelo e representa a perda por percurso média do ambiente de rádio, enquanto o termo $\xi(\mathbf{s})$ expressa as flutuações espaciais aleatórias em torno da média e representa o sombreamento do canal sem fio ao longo da região de cobertura D atendida pela ERB. Neste trabalho, há um interesse nos efeitos em larga escala do ambiente de rádio, sobretudo no sombreamento do canal de propagação. A razão para isto é que os efeitos de pequena escala (*e.g.*, múltiplos percursos) apresentam variações significativas em distâncias muito pequenas (poucos metros ou até centímetros dependendo da frequência de operação). Este tipo de variação em pequena escala inviabiliza a aplicação de predições espaciais que se baseiam nas informações de localização das redes sem fio [79]. Por outro lado, já é possível mencionar que as variações do sombreamento apresentam correlação espacial de dezenas e até centenas de metros, dependendo do ambiente de propagação, justificando a seleção do sombreamento para a componente $\xi(\mathbf{s})$.

2.2.1 Modelagem da Tendência

A tendência espacial $\mu(\mathbf{s})$ do ambiente de rádio é baseada na modelagem da perda por percurso média do canal sem fio. Assim, $\mu(\mathbf{s})$ consiste na potência média de recepção (em dBm) e pode ser descrita por um conjunto de funções base com seus respectivos coeficientes, de acordo com

$$\mu(\mathbf{s}) = P_{\text{tx}} - \sum_{k=1}^p \alpha_k f_k(\mathbf{s}), \quad (2.14)$$

em que P_{tx} é a potência de transmissão utilizada no transmissor da ERB, $f_k(\mathbf{s})$ são as funções base do modelo e α_k formam o conjunto de p coeficientes desconhecidos da tendência espacial.

Neste contexto, ferramentas para a modelagem de propagação no planejamento de redes de sistemas de comunicações sem fio são amplamente investigadas na literatura e muitos artigos científicos discutem diferentes métodos e modelos sobre este tópico (*e.g.*, [33-36], [95-96]). Neste trabalho, o modelo log-distância é utilizado para a modelagem da perda por percurso média,

$$\bar{P}_L(d) = \bar{\bar{P}}_L(d_0) + 10 \log \left(\frac{d}{d_0} \right)^\alpha, \quad (2.15)$$

em que $\bar{P}_L(d)$ é a atenuação (em dB) sofrida pelos sinais transmitidos devido à perda por percurso média a uma dada distância d (metros), $\bar{\bar{P}}_L(d_0)$ é a atenuação (em dB) a uma distância de referência d_0 e α consiste no coeficiente de propagação do modelo de perda por percurso média [95-97]. A escolha deste modelo é feita em função de duas razões: i) primeiramente devido ao seu uso consolidado e difundido na literatura científica e ii) pela possibilidade de formulação linear do problema de estimação da tendência espacial, explicado com mais detalhes na seção 2.3.

É importante notar que o modelo em (2.15) depende diretamente da distância entre a ERB, com localização assumida fixa em $\mathbf{s}_{tx} = (s_{tx}, s_{yt})$, e os dispositivos da rede que estão associados às coordenadas espaciais estimadas $\mathbf{s}$, obtidas por meio do método de localização utilizado na rede. Sem perda de generalidade, algumas simplificações podem ser adotadas (com $\bar{\bar{P}}_L(d_0) = 0$ dB e $d_0 = 1$ m) e por meio de manipulações algébricas, é possível reescrever (2.15) para o caso espacial,

$$\begin{aligned} \bar{P}_L(\mathbf{s}) &= +10\alpha \log(d(\mathbf{s}_{tx}, \mathbf{s})) \quad \text{em que } \mathbf{s}, \mathbf{s}_{tx} \in D, \\ &= +10\alpha \log \sqrt{(s_{tx} - s_x)^2 + (s_{yt} - s_y)^2}, \end{aligned} \quad (2.16)$$

em que $d(\mathbf{s}_{tx}, \mathbf{s})$ é a função distância aplicada ao caso das coordenadas espaciais em duas dimensões. O modelo em (2.16) indica que a perda por percurso média aumenta com aumento do logaritmo da função distância e depende do valor do parâmetro α , que implica em diferentes inclinações para a curva da atenuação $\bar{P}_L(\mathbf{s})$ em função da distância. A Tabela I traz a associação com alguns tipos de ambiente de propagação e os valores para o coeficiente de propagação.

Tabela 2.1: Cenários de Propagação sem Fio e o Coeficiente de Propagação [96]

Tipo de Ambiente	Coeficiente de Propagação
Espaço-Livre	2
Rural (Plano)	3
Rural (Montanhoso)	3.5
Sub-urbano (cidade)	4
Urbano (cidade)	4.5

É possível aplicar o modelo da perda por percurso média em (2.16) nas funções base da modelagem da tendência em (2.14), considerando o parâmetro α (com $k = 1$), a fim de obter a potência média de recepção do ambiente de rádio, expressa de acordo com

$$\begin{aligned}\mu(\mathbf{s}) &= P_{\text{tx}} - \alpha f(\mathbf{s}) \\ &= P_{\text{tx}} - 10\alpha \log(d(\mathbf{s}_{\text{tx}}, \mathbf{s})).\end{aligned}\tag{2.17}$$

Neste ponto, é possível destacar duas informações importantes: i) a inspeção de (2.17) permite observar que o modelo log-distância é linear no parâmetro α e ii) embora linear em α , o modelo não é linear nas coordenadas espaciais $\mathbf{s}$, pois envolve as funções logaritmo e raiz quadrada em razão das distâncias a serem obtidas em duas dimensões. Estas características irão influenciar as etapas de aprendizagem de parâmetros (cujos valores são desconhecidos e constantes) e das previsões espaciais (cujos valores são desconhecidos e aleatórios).

2.2.2 Modelagem do Processo Aleatório Espacial

O modelo (2.13) indica que as flutuações espaciais aleatórias do sombreamento $\xi(\mathbf{s})$ ocorrem em torno da potência média de recepção, ou seja, $\xi(\mathbf{s})$ é responsável pela aleatoriedade espacial do processo $P(\mathbf{s})$. Em um sistema de comunicação sem fio, isto ocorre devido às construções e obstáculos que estão relativamente próximos aos dispositivos na rede de comunicação. De fato, o sinal transmitido por um canal sem fio experimenta flutuações aleatórias ao longo das obstruções até os receptores. Uma vez que a localização, o tamanho e as propriedades dielétricas dessas obstruções são desconhecidas, o uso de modelos estatísticos para descrever as variações aleatórias é fundamental. Um dos modelos mais usados na literatura científica para a modelagem destes efeitos em larga escala é o sombreamento log-normal [95-98]. Este modelo é caracterizado por um processo aleatório Gaussiano quando o nível da grandeza medida está na escala em dB e captura de forma satisfatória as variações da potência de recepção em ambientes de rádio propagação em função do sombreamento [95-98]. Desta forma, o modelo de sistema relacionado com o ambiente de rádio (2.13) pode ser expandido de acordo com

$$\begin{aligned}P(\mathbf{s}) &= \mu(\mathbf{s}) + \xi(\mathbf{s}) \\ &= P_{\text{tx}} - \underbrace{10\alpha \log d(\mathbf{s}_{\text{tx}}, \mathbf{s})}_{\text{Perda por Percurso Média}} + \underbrace{\xi(\mathbf{s})}_{\text{Sombreamento}},\end{aligned}\tag{2.18}$$

em que $\xi(\mathbf{s})$ representa o sombreamento log-normal do canal sem fio 0e consiste em um processo Gaussiano com média zero e com covariância espacial *i.e.*, $\xi(\mathbf{s}) \sim \mathcal{GP}(0, C(\mathbf{s}_i, \mathbf{s}_j))$. Portanto, verifica-se que o processo $P(\mathbf{s})$ que caracteriza o ambiente de rádio também é um processo aleatório Gaussiano espacialmente correlacionado, mas com média que depende das características da tendência, além da covariância espacial do sombreamento log-normal, *i.e.*, $P(\mathbf{s}) \sim \mathcal{GP}(\mu(\mathbf{s}), C(\mathbf{s}_i, \mathbf{s}_j))$.

Neste âmbito, a seleção do modelo de covariância espacial para a correlação do sombreamento log-normal é um aspecto importante do modelo de sistema. Sobre este tópico, uma ampla análise que aborda diferentes modelos de correlação do sombreamento do canal sem fio pode ser acessada em [98], incluindo discussões sobre a viabilidade matemática dos modelos, fundamentada na propriedade definida positiva, e sobre a plausibilidade física dos modelos, relacionada com a quantidade e a representatividade dos parâmetros utilizados. Com isso, o uso da propriedade da isotropia é adequada para a caracterização da covariância espacial do sombreamento log-normal. A razão para tal é que um dos modelos de correlação do sombreamento mais difundidos na literatura é o modelo de Gudmundson, o qual é baseado nas distâncias de separação $|\mathbf{h}|$ entre as medidas no ambiente de rádio [99]. De forma específica, o modelo de Gudmundson indica que a correlação espacial do sombreamento decresce exponencialmente em função do aumento da distância $|\mathbf{h}|$ entre as medidas [96-97]. De fato, vários estudos baseados em campanhas conduzidas em ambientes de propagação constataam a validade do modelo de Gudmundson [99-103], enquanto diversos artigos científicos propõem métodos de geração do sombreamento log-normal, considerando este modelo [104-108]. Baseado nestas informações, este trabalho seleciona o modelo de covariância exponencial e isotrópica para caracterizar a correlação espacial do sombreamento log-normal gerado nas simulações, que é expresso de acordo com

$$C(|\mathbf{h}|) = m \exp\left(-\frac{|\mathbf{h}|}{r}\right), \quad (2.19)$$

em que o parâmetro m (variância *sill*) expressa a variância do sombreamento, (*i.e.*, $m = \sigma^2$) e quantifica a dispersão das variações aleatórias de $\xi(\mathbf{s})$. O parâmetro r , indica o *range* do modelo de covariância e representa a distância de descorrelação ($r = d_{\text{corr}}$) do sombreamento do canal sem fio. Isto permite caracterizar as zonas de dependência espacial do ambiente de rádio, nas quais as variações de potência de recepção (*i.e.*, sombreamento) experimentadas pelos dispositivos são espacialmente correlacionadas. Deste modo, observações que se distanciam até d_{corr} tendem a apresentar alta correlação espacial, enquanto os valores observados de $P(\mathbf{s})$ que se distanciam além de d_{corr} tendem a possuir baixa correlação espacial. A relação entre a função covariância e o semivariograma permite obter o modelo equivalente isotrópico exponencial para o semivariograma,

$$\gamma(|\mathbf{h}|) = m \left\{ 1 - \exp\left(-\frac{|\mathbf{h}|}{r}\right) \right\}. \quad (2.20)$$

Verifica-se que os modelos (2.19) e (2.20) descrevem, de forma analítica, a similaridade e a dissimilaridade espacial do sombreamento do canal sem fio e a aprendizagem de seus parâmetros é um passo importante. Mais adiante, técnicas de estimação da área geoestatística serão utilizadas para estimar (2.19) e (2.20) e seus parâmetros, permitindo a extração de informação sobre $P(\mathbf{s})$, que caracteriza o ambiente de rádio.

2.3 Modelo de Predição Espacial para a Geração do REM

Esta seção apresenta o modelo de predição espacial utilizado para a geração do REM. A Figura 2.4 mostra o diagrama do modelo, que se fundamenta na abordagem geoestatística. O ambiente de rádio consiste em um processo aleatório espacial $P(\mathbf{s})$, gerado através da combinação entre a potência média de recepção $\mu(\mathbf{s})$, obtida a partir da perda por percurso média, e o sombreamento log-normal $\xi(\mathbf{s})$ modelado através de processos Gaussianos. Diferentes técnicas podem ser utilizadas para a geração do sombreamento log-normal do canal sem fio: soma de senóides [104]; decomposição Cholesky e filtragem autoregressiva de primeira ordem [105], [106]; transformações de Fourier dos modelos de correlação do sombreamento [107], [108]. Neste trabalho, o método de Cholesky é escolhido, uma vez que este é baseado na decomposição da função covariância do processo $P(\mathbf{s})$, que possui relação direta com a função semivariograma, principal ferramenta da geoestatística utilizada para a geração das predições espaciais.

O modelo contempla a Fase I, que consiste na aprendizagem dos parâmetros do ambiente de rádio, caracterizada por três etapas de estimação: i) o tratamento da tendência; ii) a estimação do semivariograma experimental e iii) a estimação dos parâmetros de covariância do sombreamento log-normal. O êxito na aprendizagem dos parâmetros é essencial para que as predições possam ser realizadas na Fase II por meio da técnica Kriging e o REM seja gerado. Ambas as fases contam com o conjunto finito de N medidas coletadas, a partir das medições realizadas pelos dispositivos da rede. Este processo caracteriza a amostragem espacial aplicada ao ambiente de rádio e permite formular modelo de sistema matricialmente, de acordo com

$$\mathbf{p} = \mathbf{x} \alpha + \boldsymbol{\xi}, \quad (2.21)$$

em que $\mathbf{p} = [P_{\text{rx}}(\mathbf{s}_1), \dots, P_{\text{rx}}(\mathbf{s}_N)]^T$ forma o vetor com o conjunto de medidas de potência de recepção, $\mathbf{x} = [-10 \log d(\mathbf{s}_{\text{tx}}, \mathbf{s}_1), \dots, -10 \log d(\mathbf{s}_{\text{tx}}, \mathbf{s}_N)]^T$ consiste no vetor de funções base com $P_{\text{tx}} = 0$ dBm, α é o parâmetro do modelo da tendência (perda por percurso média) e $\boldsymbol{\xi} = [\xi(\mathbf{s}_1), \dots, \xi(\mathbf{s}_N)]$ é o vetor aleatório Gaussiano de média zero, desvio padrão σ e matriz de covariância $\mathbf{C}$.

Figura 2.4: Diagrama do modelo de geração do REM.

Neste ponto, duas considerações iniciais são enfatizadas: primeiramente, é assumido que o modelo vetorial em (2.21) não é afetado por erros de localização, ou seja, não há incerteza sobre a localização das coordenadas espaciais $\mathbf{s}$. Isto significa que o conjunto $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ é obtido de forma perfeita pelos métodos de estimação de localização, refletindo as posições verdadeiras dos dispositivos e, por isso, é denotado como conjunto de coordenadas espaciais estimadas sem incerteza de localização. Segundo, assume-se que a realização e o envio das medições ocorre de forma ideal nos dispositivos da rede. Desta forma, os valores de potência de recepção que compõem o vetor de medidas $\mathbf{p}$ não contém erros de medição e somente refletem os efeitos em larga escala do ambiente de rádio. Em um momento posterior, a incerteza de localização será incorporada às coordenadas espaciais estimadas $\mathbf{s}$ com o objetivo de investigar quais são os impactos produzidos no desempenho das predições espaciais.

2.3.1 Tratamento da Tendência Espacial

A primeira etapa da fase de aprendizagem de parâmetros consiste no tratamento da tendência e a sua realização é importante por duas razões: i) a presença da tendência implica em flutuações da potência média de recepção e afeta a estacionaridade do processo $P(\mathbf{s})$; ii) a aplicação de técnicas de estimação (*e.g.*, semivariograma experimental), se realizada diretamente sobre as medidas coletadas (com a presença da tendência), pode resultar em estimativas enviesadas [63], impactando os resultados de predição espacial. Assim, sobre a perspectiva de uso dos preditores no caso específico do ambiente de rádio, o tratamento da tendência se torna necessário (detalhes específicos sobre este ponto também são discutidos na subseção 2.3.5).

É importante ressaltar que os valores de potência de recepção $\{P_{rx}(\mathbf{s}_i), i = 1, 2, \dots, N\}$ são formados por uma parcela média, $\boldsymbol{\mu} = \mathbf{x} \boldsymbol{\alpha}$, e contaminados pelas flutuações espaciais $\boldsymbol{\xi}$. Com isso, o tratamento da tendência é realizado para que os valores regionalizados de interesse sejam obtidos, ou seja, a VR desejada deve caracterizar as flutuações aleatórias associadas ao sombreamento $\boldsymbol{\xi}$ (sem a tendência). Isto significa que a VR é obtida somente após a estimação (*i.e.*, $\hat{\boldsymbol{\mu}}$) seguida da remoção da tendência das medidas. Este processo é conhecido como *detrending* e implica no uso das técnicas geoestatísticas sobre o resíduo obtido a partir da remoção da tendência, denotado como a VR [66].

Nota-se que o tratamento da tendência requer a estimação do coeficiente $\boldsymbol{\alpha}$, a partir de $\mathbf{p}$. Métodos de estimação baseados em máxima verossimilhança (ML - *Maximum Likelihood*) e mínimos quadrados (LS - *Least Squares*) são opções amplamente difundidas na literatura [62-63], [110-111] e úteis para estes tipos de problemas de estimação, sobretudo em sistemas de comunicações sem fio [17-19], [78]. Neste caso, o modelo da perda por percurso média adotado (log-distância) traz alguns benefícios, uma vez que este permite elaborar o problema de estimação de forma linear. Em outras palavras, o uso do modelo log-distância torna o modelo de sistema (2.21) linear no parâmetro $\boldsymbol{\alpha}$ e permite caracterizar o problema de estimação da tendência como sendo a extração da componente média a partir de medidas ruidosas, que são submetidas a uma forma de ruído colorido com uma covariância espacial $C(\mathbf{s}_i, \mathbf{s}_j)$.

Na situação em que o modelo de sistema é linear nos parâmetros e submetido à variações aleatórias Gaussianas, tal como o modelo (2.21), é possível encontrar o estimador não enviesado de mínima variância (MVUE - *Minimum Variance Unbiased Estimator* [111, pp. 85-86]), de acordo com

$$\hat{\alpha} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{p}. \quad (2.22)$$

Neste caso, o desempenho do estimador em (2.22) é estatisticamente especificado e dado por $\hat{\alpha} \sim \mathcal{GP}(\alpha, \sigma^2(\mathbf{x}^T \mathbf{x})^{-1})$, uma vez que $\hat{\alpha}$ é uma transformação linear do vetor Gaussiano $\mathbf{p}$ [111]. É importante ressaltar que (2.22) considera $\mathbf{p}$ como um vetor Gaussiano espacialmente descorrelacionado, *i.e.*, $\mathbf{p} \sim \mathcal{GP}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$. No entanto, o sombreamento log-normal do ambiente de rádio implica em uma correlação espacial no vetor $\mathbf{p}$ através da sua covariância, *i.e.*, $\mathbf{p} \sim \mathcal{GP}(\boldsymbol{\mu}, \mathbf{C})$, em que $\mathbf{C}$ é a matriz de covariâncias espaciais (com dimensões $N \times N$) relacionada com os N valores de potência de recepção que compõem as medidas do vetor $\mathbf{p}$. Com isso, ainda é possível encontrar o MVUE ([111, pp. 94-97]), mas levando em conta o conhecimento sobre a matriz $\mathbf{C}$, resultando em

$$\hat{\alpha} = (\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x})^{-1} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{p}. \quad (2.23)$$

Nota-se que o estimador (2.23) requer o conhecimento da matriz de covariância inversa ($\mathbf{C}^{-1}$). Porém, algumas dificuldades começam a surgir: primeiramente, os parâmetros de covariância na formação da matriz $\mathbf{C}$ não são necessariamente conhecidos *a priori* e, de fato, um aspecto realístico sobre o modelo de sistema consiste em assumir a falta de conhecimento sobre tais parâmetros. Outro ponto é que as etapas seguintes da Fase I têm o objetivo de estimar a covariância $\mathbf{C}$ do ambiente de rádio, levando a um círculo vicioso. Frente a esta situação e com o objetivo de ampliar as discussões sobre as particularidades da aprendizagem de parâmetros, é possível destacar duas abordagens que trazem diferentes métodos⁹ para a solução deste problema:

- (i) **Abordagem Sequencial:** A abordagem sequencial é baseada no processo de *detrending* seguido da estimação do semivariograma experimental para que um modelo analítico seja ajustado e os parâmetros de covariância possam ser obtidos. Diferentes técnicas de estimação (*e.g.*, LS) são aplicadas ao caso espacial, retratando a abordagem geoestatística [63], [67-68].
- (ii) **Abordagem Conjunta:** A abordagem conjunta é fundamentada em técnicas de máxima verossimilhança (*e.g.*, ML) e trata a estimação da tendência e dos parâmetros de covariância de forma conjunta [63, pp.116]. Assim, os métodos são aplicados diretamente sobre as medidas coletadas, ou seja, a remoção da tendência e a estimação do semivariograma experimental não são realizados nesta abordagem. Além disso, ressalta-se que o uso da técnica ML é baseado em pressuposições sobre a distribuição de probabilidade do processo aleatório espacial.

⁹Além das abordagens apresentadas, o uso de técnicas não paramétricas para a estimação da covariância é uma alternativa. Neste caso, não há um modelo analítico para a forma da função covariância (semivariograma) e, portanto, não há garantia de que os resultados de estimação obtidos sejam válidos no sentido de levar a uma estrutura definida positiva para a matriz de covariâncias obtida a partir de técnicas não paramétricas [67-68].

Sobre a abordagem conjunta, alguns pontos importantes podem ser citados. Primeiramente, a distribuição estatística Gaussiana do sombreamento do canal sem fio faz com que a técnica ML seja mais adequada ao problema da estimação de parâmetros. No entanto, em um momento posterior, será mostrado que a incerteza de localização sobre as estimativas de posicionamento dos dispositivos da rede afeta a característica estatística Gaussiana do processo aleatório espacial [78], [83], [93]. Ainda, a estimação ML pode gerar resultados enviesados e, nestas situações, o método de maximização da verossimilhança residual (REML - *Residual Maximum Likelihood*), baseado em transformações nos dados, torna-se uma opção [63]. Apesar disso e embora a técnica ML tenha desempenho superior, a estimação pode ser computacionalmente custosa para grandes quantidades de medidas, uma vez que o cálculo de cada verossimilhança requer a fatoração Cholesky da matriz de covariância $\mathbf{C}$ ou operações equivalentes, que são da ordem $O(N^3)$ [83].

Por outro lado, a abordagem sequencial baseada na geoestatística se apresenta de forma mais flexível por meio da técnica LS, adequada ao modelo da tendência, e na qual nenhuma pressuposição sobre a distribuição estatística dos dados é requerida, embora não se possa especificar seu desempenho e, com isso, argumentar em termos de otimalidade (exceto quando $\mathbf{p} \sim \mathcal{GP}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$) [111]. Desta forma, em virtude da praticidade presente nos métodos geoestatísticos adotados (baseados nas funções covariância e semivariograma) e da não pressuposição sobre a distribuição de probabilidade do processo aleatório, aliada às complexidades computacionais do ML, a abordagem sequencial é adotada no modelo de predição espacial neste trabalho.

A aplicação da técnica LS especificamente ao problema de estimação da tendência do ambiente de rádio resulta na estimativa do coeficiente de propagação $\hat{\alpha}$. Assim a estimativa $\hat{\alpha}$ é utilizada no modelo em (2.17) para obter os valores estimados da potência média de recepção em diferentes coordenadas espaciais, $\hat{\boldsymbol{\mu}}(\mathbf{s})$. Isto permite que a componente média das predições sejam obtidas nas diversas localidades do domínio espacial D para a geração do REM. Entretanto, a intensidade e as características estatísticas das flutuações aleatórias do sombreamento $\boldsymbol{\xi}$ afetam a estimativa $\hat{\alpha}$. Com isso, a remoção da tendência do vetor de medidas $\mathbf{p}$ é feita para que, por meio dos métodos geoestatísticos, as predições das variações do sombreamento $\boldsymbol{\xi}$ possam ser realizadas e incorporadas na geração do REM, visando maior acurácia nos resultados de predição. Analiticamente, este processo de remoção da tendência é dado por

$$\mathbf{z} = \mathbf{p} - \hat{\boldsymbol{\mu}}, \quad (2.24)$$

em que $\mathbf{p} = [P_{\text{rx}}(\mathbf{s}_1) \dots P_{\text{rx}}(\mathbf{s}_N)]^T$ é o vetor com os valores de potência de recepção coletados do ambiente de rádio, $\hat{\boldsymbol{\mu}} = [\hat{\boldsymbol{\mu}}(\mathbf{s}_1) \dots \hat{\boldsymbol{\mu}}(\mathbf{s}_N)]^T$ é o vetor com as estimativas de potência média de recepção, obtido a partir de $\hat{\alpha}$ e $\mathbf{z} = [Z(\mathbf{s}_1) \dots Z(\mathbf{s}_N)]^T$, que consiste no vetor com os valores regionalizados, após a remoção da tendência. Outro ponto relevante consiste em ressaltar que a estimação da tendência não é ideal ($\hat{\alpha} \neq \alpha$) e como esta é a primeira etapa da aprendizagem de parâmetros, há a possibilidade de algum resíduo contaminar o vetor $\mathbf{z}$ e, conseqüentemente, influenciar as etapas de estimação subsequentes, bem como as predições espaciais.

2.3.2 Semivariograma e Covariograma Experimentais

O semivariograma e o covariograma experimentais consistem nas estimativas empíricas das funções teóricas semivariograma e covariância, necessárias para a geração das previsões espaciais. Esta subseção descreve como as partes experimentais destas funções são obtidas, a partir do vetor $\mathbf{z}$ (que representa a VR). Primeiramente, o semivariograma é abordado devido ao seu uso mais difundido na área geoestatística. Em seguida, os detalhes do covariograma são apresentados.

A estimação do semivariograma é baseada em dois conceitos: i) a interpretação geométrica de um gráfico de dispersão e ii) o método dos momentos de Matheron (MoM) [62-67]. O gráfico de dispersão, mostrado na Figura 2.5, relaciona as observações $Z(\mathbf{s}_i)$ com $Z(\mathbf{s}_i + \mathbf{h})$, mediante uma determinada separação espacial $\mathbf{h}$. Um ponto de dispersão $\mathbf{T}$ associado à coordenada espacial $\mathbf{s}_1$ é posto no gráfico, com abscissa $Z(\mathbf{s}_1 + \mathbf{h})$ e ordenada $Z(\mathbf{s}_1)$, *i.e.*, posicionado em $(Z(\mathbf{s}_1 + \mathbf{h}), Z(\mathbf{s}_1))$. Este ponto de dispersão $\mathbf{T}$ é projetado na reta bissetriz do gráfico, resultando em dois valores de ordenadas, $Z(\mathbf{s}_1 + \mathbf{h})$ e $Z(\mathbf{s}_1)$. A magnitude da diferença entre tais ordenadas, dada por $|Z(\mathbf{s}_1 + \mathbf{h}) - Z(\mathbf{s}_1)|$, também é mostrada no gráfico de dispersão. Deste modo, a triangulação formada com a conexão à reta bissetriz (linha tracejada) permite equacionar a distância d_s , expressa por

$$d_s = |Z(\mathbf{s}_1 + \mathbf{h}) - Z(\mathbf{s}_1)| \cos\left(\frac{\pi}{4}\right). \quad (2.25)$$

Tomando o valor quadrático da distância à reta bissetriz, tem-se

$$d_s^2 = \frac{1}{2} |Z(\mathbf{s}_1 + \mathbf{h}) - Z(\mathbf{s}_1)|^2. \quad (2.26)$$

Figura 2.5: Interpretação geométrica do gráfico de dispersão.

Supondo a presença de um conjunto de N_h pontos de dispersão que distanciam de $\mathbf{h}$, é possível calcular a média das distâncias quadráticas à reta bissetriz, dada por

$$d_s^2 = \frac{1}{2N_h} \sum_{i=1}^{N_h} |Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)|^2. \quad (2.27)$$

A distância quadrática d_s^2 é denotada como momento de inércia e permite demonstrar a origem do fator (2) na relação entre o variograma $2\gamma(\mathbf{h})$ e o semivariograma $\gamma(\mathbf{h})$ [60], [67].

O aumento das diferenças quadráticas entre as observações (aumento do momento de inércia) implica em maior dispersão dos pontos na Figura 2.5, refletindo uma baixa correlação espacial entre as observações. De outro modo, se as diferenças quadráticas são mínimas, os pontos de dispersão ficam sobre a reta bissetriz (momento de inércia nulo), refletindo na máxima correlação espacial entre as observações. Assim, a distância d_s^2 é uma medida da variabilidade da diferença entre os pares de medidas que se distanciam de $\mathbf{h}$. Neste sentido, o gráfico de dispersão reúne detalhes importantes para o entendimento da estimação do semivariograma, que aplica este conceito para estimar a variância das diferenças entre as VAs do processo aleatório espacial $P(\mathbf{s})$.

Matematicamente, a estimação do semivariograma experimental $\hat{\gamma}(\mathbf{h})$ consiste no valor médio das diferenças quadráticas entre os valores regionalizados $\mathbf{z}$, considerando diferentes separações espaciais $\mathbf{h}$. O método dos momentos de Matheron (MoM) é o principal estimador utilizado para obter o semivariograma experimental, expresso de acordo com [62-67]

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} [Z(\mathbf{s}_j) - Z(\mathbf{s}_i)]^2, \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D, i, j = 1, 2, \dots, N. \quad (2.28)$$

em que N_h representa o número de pares de medidas que se distanciam de $\mathbf{h}$ tomadas no conjunto de valores regionalizados. O uso do método dos momentos é difundido na área da geoestatística e algumas questões tomam importância neste contexto: primeiramente, nota-se que o uso das diferenças quadráticas amplifica as variabilidades estimadas. Segundo, a verificação de não viés do estimador (2.28), *i.e.*, $\mathbb{E}\{\hat{\gamma}(\mathbf{h})\} = \gamma(\mathbf{h})$, demonstrada abaixo [66, pp. 53],

Demonstração.

$$\begin{aligned} \mathbb{E}\{\hat{\gamma}(\mathbf{h})\} &= \frac{1}{2N_h} \mathbb{E} \left\{ \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} [Z(\mathbf{s}_j) - Z(\mathbf{s}_i)]^2 \right\}, \\ &= \frac{1}{2N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} \mathbb{E} \left\{ [Z(\mathbf{s}_j) - Z(\mathbf{s}_i)]^2 \right\}, \\ &= \frac{1}{N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} \frac{1}{2} \text{Var} \left\{ [Z(\mathbf{s}_j) - Z(\mathbf{s}_i)]^2 \right\} = \frac{1}{N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} \gamma(\mathbf{h}) = \frac{1}{N_h} N_h \gamma(\mathbf{h}) = \gamma(\mathbf{h}). \end{aligned} \quad (2.29)$$

□

A demonstração em (2.29) sobre o estimador MoM traz duas informações importantes: i) indica que as semivariâncias $\hat{\gamma}(\mathbf{h})$ consistem em estimativas não enviesadas da função teórica semivariograma $\gamma(\mathbf{h})$ e ii) o número total de medidas N e, conseqüentemente, o número de pares de medidas N_h são fatores que afetam os resultados do estimador (2.28). De fato, para que as estimativas $\hat{\gamma}(\mathbf{h})$ tenham confiabilidade estatística, é necessário que N_h seja suficientemente grande, exigindo que o número de medidas totais N também seja elevado. Ainda, é importante mencionar que a obtenção de um conjunto de medidas que se distanciam exatamente de $\mathbf{h}$ em um domínio espacial contínuo nas coordenadas $\mathbf{s}$ não é uma tarefa trivial. Com isso, a continuidade espacial do processo aleatório e o fato do vetor $\mathbf{z}$ ser composto por um conjunto finito de N medidas capturadas na realização de $P(\mathbf{s})$ impõem desafios à estimação do semivariograma experimental.

Frente a esta situação prática, algum tipo de agrupamento deve ser realizado sobre as distâncias obtidas para que os resultados de estimação tenham maior confiança estatística. Este processo de agrupamento é denotado como *binning* e consiste em um dos principais aspectos na estimação do semivariograma experimental [66]. Ainda assim, variações nos resultados das estimativas $\hat{\gamma}(\mathbf{h})$ podem ocorrer, uma vez que a média do estimador MoM é baseada na quantidade N_h , que pode variar em razão da forma como o agrupamento é aplicado no conjunto total de medidas.

A Figura 2.6 mostra as duas principais formas de estimação do semivariograma experimental: (a) a nuvem do semivariograma, que é constituída diretamente, a partir dos valores das diferenças quadráticas tomadas entre todas as N medidas do vetor $\mathbf{z}$ e (b) o semivariograma experimental obtido com o agrupamento das distâncias em conjunto com os resultados do estimador MoM. Ressalta-se que cada ponto da nuvem de semivariância está relacionado com um par de observações, ou seja, nenhum tipo de agrupamento de distâncias é aplicado. Este formato da nuvem do semivariograma é uma ferramenta útil na verificação de *outliers* nas medidas coletadas, que consistem em semivariâncias muito afastadas das estimativas produzidas via método dos momentos de Matheron.

Figura 2.6: Semivariograma experimental: (a) nuvem de semivariâncias (b) com agrupamento de medidas.

De outro modo, nota-se que o agrupamento das distâncias de separação suavizou as estimativas do semivariograma experimental. De forma específica, tal agrupamento é obtido através de um processo de quantização aplicado à todas as distâncias $|\mathbf{h}|$ calculadas entre as medidas do vetor $\mathbf{z}$, resultando em um novo conjunto composto por distâncias denominadas lags [63-64]. Assim, todas as medidas cujas distâncias de separação foram quantizadas para uma determinada distância lag são incluídas no cômputo da respectiva semivariância. É possível verificar a quantidade de medidas utilizadas em cada agrupamento, mostradas nas caixas numéricas da Fig.2.6(b). Nota-se que algumas semivariâncias foram computadas com mais medidas do que outras. Isto significa que cada estimativa $\hat{\gamma}(\mathbf{h})$ é obtida com uma precisão diferente. Portanto, a forma como a quantização é aplicada juntamente com o estimador MoM é um ponto de relevância na implementação prática da estimação do semivariograma experimental, afetando a precisão das estimativas. Os passos mostrados adiante, baseados na técnica de agrupamento de distâncias [64-66] e implementadas em [112-113], descrevem o processo de quantização das distâncias $|\mathbf{h}|$ aplicados ao semivariograma experimental isotrópico¹⁰:

- (i) Primeiramente, realiza-se a geração de uma matriz triangular superior Δ , cujos elementos (i, j) são dados pelas distâncias de separação $|\mathbf{h}| = |\mathbf{s}_i - \mathbf{s}_j|$ entre todos¹¹ os pares de coordenadas espaciais que compõem o vetor $\mathbf{z}$, *i.e.*, para todo $i < j$, com $i = 1, 2, \dots, N$ e $j = 1, 2, \dots, N$.
- (ii) Uma distância lag inicial, denotada como Lag_1 , é escolhida a partir da matriz Δ . Em geral, a escolha de Lag_1 é baseada em um determinado critério [64-66]. Um primeiro critério consiste em optar pela menor distância encontrada na matriz de distâncias Δ . De outro modo, a distância Lag_1 pode ser calculada como a média dos mínimos valores de cada coluna da matriz Δ (exceto os valores nulos). Foi observado que esta segunda estratégia produziu maior suavização no semivariograma comparada com o primeiro critério e, portanto, foi escolhida para este passo.
- (iii) O agrupamento das distâncias é realizado através da quantização da matriz de distâncias Δ baseado no valor escolhido para Lag_1 . Cada agrupamento é representado por uma distância denotada como Lag_k . Assim, as distâncias $|\mathbf{h}_{ij}|$ da matriz Δ são quantizadas de acordo com a função: $Q(|\mathbf{h}_{ij}|) = \text{Lag}_k$ para k que minimiza a diferença $||\mathbf{h}_{ij}| - \text{Lag}_k|$. Após a aplicação da quantização ainda é possível definir o valor representativo de cada agrupamento Lag_k como o múltiplo da distância Lag_1 (*i.e.*, $\text{Lag}_k = k\text{Lag}_1$) ou o centro de massa (média das distâncias) de cada agrupamento [65-66]. Pela mesma razão citada acima (melhor suavização do semivariograma), a segunda estratégia é adotada neste passo.

¹⁰É importante mencionar que a anisotropia pode aumentar a complexidade da estimação do semivariograma $\hat{\gamma}(\mathbf{h})$ de forma significativa. De fato, a presença de anisotropia implica na inclusão de mais um aspecto (ângulo) no processo de quantização aplicado ao semivariograma. Assim, a verificação do tipo de anisotropia (*e.g.*, geométrica) é fundamental para a aplicação de transformações nas coordenadas espaciais, visando o tratamento da anisotropia, ou seja, a conversão para um espaço isotrópico. Detalhes sobre os diferentes tipos de anisotropia e os procedimentos de ajuste que incluem a modelagem dos parâmetros anisotrópicos podem ser acessados em [62-63], [66], [125].

¹¹Os cálculos realizados a partir de todos os pares de medidas implicam na complexidade $O(N^2)$ sobre as tarefas computacionais do semivariograma, que pode se tornar crítica em aplicações específicas onde N é extremamente elevado (*e.g.*, sensoriamento remoto e processamento de imagens médicas) [126-127].

- (iv) Finalmente, o estimador MoM é utilizado para calcular o semivariograma experimental a partir do resultado do agrupamento com as distâncias lags:

$$\hat{\gamma}(\text{Lag}_k) = \frac{1}{2\|\mathbf{A}_{\text{Lag}_k}\|} \sum_{|\mathbf{s}_i - \mathbf{s}_j| \in \mathbf{A}_{\text{Lag}_k}} [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2, \quad i, j = 1, 2, \dots, N. \quad (2.30)$$

em que $\|\cdot\|$ é o operador cardinalidade e $\mathbf{A}_{\text{Lag}_k}$ é o conjunto de pares de coordenadas espaciais, $(\mathbf{s}_i, \mathbf{s}_j)$, cujas distâncias $|\mathbf{h}| = |\mathbf{s}_i - \mathbf{s}_j|$, quando quantizadas, correspondem à distância Lag_k . O número total de distâncias lags dos agrupamentos realizados para a quantização das distâncias não necessariamente é fixo, pois depende do valor da distância lag inicial e de $\text{Lag}_{\max}$, que determinará o cálculo do último valor de semivariância e distância lag. Sobre este aspecto, é possível mencionar recomendações da literatura geoestatística para a escolha de $\text{Lag}_{\max}$, *e.g.*, metade da distância do domínio espacial em análise [63-64], [66]. Em geral, a partir de distâncias de separação $|\mathbf{h}|$ significativamente grandes, torna-se mais difícil obter um número de medidas elevado. Por isso, o tamanho do domínio espacial analisado (área de cobertura no caso do ambiente de rádio) e as características da amostragem espacial são pontos que têm impacto nos parâmetros utilizados para a estimação do semivariograma experimental.

O uso do estimador MoM na estimação da covariância do processo aleatório espacial permite obter o covariograma experimental $\hat{C}(\mathbf{h})$, dado por

$$\hat{C}(\mathbf{h}) = \frac{1}{2N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} [Z(\mathbf{s}_i) - \hat{\mu}(\mathbf{s}_i)][Z(\mathbf{s}_j) - \hat{\mu}(\mathbf{s}_j)], \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D, \quad i, j = 1, 2, \dots, N. \quad (2.31)$$

em que $\hat{\mu}(\mathbf{s}_i)$ e $\hat{\mu}(\mathbf{s}_j)$ consistem na estimativas dos valores esperados nas coordenadas espaciais $\mathbf{s}_i$ e $\mathbf{s}_j$, ou seja, as estimativas $\hat{C}(\mathbf{h})$ requerem o conhecimento do valor esperado para cômputo do covariograma. Ainda sobre o ponto de vista prático, tal como no caso do semivariograma experimental em (2.30), o covariograma também pode ser expresso, considerando o agrupamento através das distâncias lags, resultando em

$$\hat{C}(\text{Lag}_k) = \frac{1}{\|\mathbf{A}_{\text{Lag}_k}\|} \sum_{|\mathbf{s}_i - \mathbf{s}_j| \in \mathbf{A}_{\text{Lag}_k}} [Z(\mathbf{s}_i) - \bar{\mathbf{z}}][Z(\mathbf{s}_j) - \bar{\mathbf{z}}], \quad i, j = 1, 2, \dots, N. \quad (2.32)$$

em que $\bar{\mathbf{z}} = (1/N) \sum_{i=1}^N Z(\mathbf{s}_i)$ consiste no estimador da média do processo aleatório espacial, uma vez que, na prática, o vetor $\mathbf{z}$ é finito e composto por N medidas. Isto significa que o uso de $\bar{\mathbf{z}}$ pode adicionar um viés aos resultados do covariograma, sobretudo quando há a presença da tendência nas medidas [67]. Esta é a principal razão do uso mais difundido do semivariograma. No caso em que a remoção da tendência é realizada, o covariograma é dado pela média dos produtos de correlação $[Z(\mathbf{s}_i)Z(\mathbf{s}_j)]$. Finalmente, ambos os resultados $\hat{C}(\text{Lag}_k)$ e $\hat{\gamma}(\text{Lag}_k)$, bem como as distâncias lags são utilizados na aprendizagem dos parâmetros, descritos na próxima subseção.

2.3.3 Estimação dos Parâmetros de Covariância

Neste momento, é importante enfatizar que os modelos analíticos e suas partes experimentais são essencialmente diferentes. O semivariograma e o covariograma experimentais não são paramétricos e seus resultados não necessariamente produzem matrizes válidas, em termos de estrutura definida positiva para a covariância e definida negativa para o semivariograma. Esta é a principal razão do uso dos modelos analíticos, pois estes são funções paramétricas que fornecem matrizes de covariâncias e semivariâncias válidas e que são necessárias para o funcionamento dos preditores espaciais Kriging.

A Figura 2.7 mostra as diferenças entre as partes experimentais e seus respectivos modelos para a covariância (a) e para o semivariograma (b). O modelo analítico selecionado deve ser representativo no sentido de capturar o comportamento das estimativas $\hat{\gamma}(\mathbf{h})$ e $\hat{C}(\mathbf{h})$. A principal característica destes modelos, denotados como $\gamma(\mathbf{h}, \boldsymbol{\theta})$ para o semivariograma e $C(\mathbf{h}, \boldsymbol{\theta})$ para a covariância, é a presença do vetor de parâmetros $\boldsymbol{\theta} = [m \ r \ c_0]$. Com isso, a aprendizagem de parâmetros do ambiente de rádio consiste no processo de ajuste aplicado aos modelos analíticos envolvendo a estimação de $\boldsymbol{\theta}$. Especificamente, algoritmos de otimização são aplicados para que o ajuste dos modelos em relação aos dados experimentais seja alcançado de forma eficiente. Nota-se que, enquanto as estimativas empíricas são baseadas no conjunto finito de distâncias lags, a estimação de $\boldsymbol{\theta}$ permite que os modelos analíticos forneçam valores de semivariâncias e covariâncias para qualquer distância de separação espacial $|\mathbf{h}|$. Esta característica é fundamental para o funcionamento do preditor Kriging (discutido adiante). Portanto, o ajuste do modelo exponencial do semivariograma às estimativas experimentais de semivariâncias e distâncias lags permitirá a estimação do vetor de parâmetros $\boldsymbol{\theta}$ do sombreamento do canal sem fio. Este processo é interpretado neste trabalho como o treinamento do modelo de predição espacial e seu resultado, $\hat{\boldsymbol{\theta}}$, é essencial no funcionamento dos preditores Kriging, visando a geração do REM em sistemas de comunicações sem fio.

Figura 2.7: Modelos analíticos e partes experimentais: (a) covariância e (b) semivariograma.

Neste contexto, a técnica LS, amplamente conhecida dos problemas de regressão [111] e utilizada em sistemas de comunicações sem fio [41-42], [78], [79], é uma solução de baixo custo computacional para a estimação de $\boldsymbol{\theta}$, a partir de uma função custo e do modelo selecionado. Sua principal variante, o método dos mínimos quadrados ordinário (OLS - *Ordinary Least Squares* [67], [111]), fornece uma estimativa para o vetor $\boldsymbol{\theta}$, que minimiza o resíduo quadrático do modelo $\hat{\gamma}(\mathbf{h}) = \gamma(\mathbf{h}, \boldsymbol{\theta}) + e(\mathbf{h})$, de acordo com o seguinte problema de otimização:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} e^2(\mathbf{h}), \\ \hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} [\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}, \boldsymbol{\theta})]^T [\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}, \boldsymbol{\theta})],\end{aligned}\tag{2.33}$$

em que $e(\mathbf{h})$ consiste na função custo que expressa a diferença entre as estimativas do semivariograma $\hat{\gamma}(\mathbf{h})$ e os valores do modelo analítico sob teste $\gamma(\mathbf{h}, \boldsymbol{\theta})$, parametrizado pelo vetor $\boldsymbol{\theta}$. É importante ressaltar que, na prática, a estimação do semivariograma fornece um vetor composto por K semivariâncias (*i.e.*, $\hat{\gamma}(\text{Lag}_k)$, com $k = 1, \dots, K$) e K distâncias lags (*i.e.*, Lag_k , com $k = 1, \dots, K$) em que o valor de K depende do número de medidas N e das variáveis relacionadas ao agrupamento ou quantização de distâncias (distância lag inicial e $\text{Lag}_{\max}$). De forma específica, o problema de otimização em (2.33) é solucionado a partir de $\hat{\gamma}(\text{Lag}_k)$ e Lag_k com algoritmos de busca exaustiva, sendo possível destacar soluções computacionalmente eficientes como os algoritmos de Levenberg-Marquadt e da região de confiança (*Trust Region*) [114-115]. Ambos os algoritmos são amplamente utilizados em problemas de regressão não linear. O algoritmo Levenberg-Marquadt foi desenvolvido primeiro e tem sido utilizado por apresentar bons resultados de ajuste para diferentes tipos de funções não lineares, enquanto o algoritmo da região de confiança consiste em uma evolução da técnica Levenberg-Marquadt, sendo mais eficiente e adequado aos casos em que existem restrições nos valores dos parâmetros e, por estas razões, é escolhido para a estimação de $\boldsymbol{\theta}$ neste trabalho.

O tópico sobre os avanços nos procedimentos de otimização para o ajuste do semivariograma (antes realizado de forma subjetiva [66]) já vem sendo alvo da literatura científica [66], [110], [116]. Em geral, os procedimentos de ajuste são realizados, levando em conta somente alguns ou todos os parâmetros dos modelos, ou seja, é possível atribuir um valor fixo a um determinado parâmetro durante o procedimento de ajuste [114]. Ainda assim, a inicialização de cada parâmetro é requerida e a convergência dos algoritmos de busca pode ser afetada se os resultados iniciais do ajuste não forem adequados, o que ocorre quando a inicialização não é assertiva ou quando o modelo selecionado é incompatível com a covariância capturada pelos valores experimentais. Desta forma, as recomendações de ajuste são guiadas pelas características de modelagem e o conhecimento físico sobre o processo aleatório espacial, (*e.g.*, interesse específico nos efeitos em larga escala do ambiente de rádio). Informações específicas envolvendo a inicialização de parâmetros, discussões e recomendações sobre as técnicas de ajuste do semivariograma podem ser acessadas em [63], [66-68], [94], [110], [116].

2.3.4 Simulação da Fase I de Aprendizagem do Ambiente de Rádio

Esta subseção apresenta os resultados de simulação da Fase I de aprendizagem de parâmetros, destacando a estimação da tendência, do semivariograma e do covariograma experimentais, bem como do vetor de parâmetros θ por meio da técnica LS. O ambiente de rádio é caracterizado pelas transmissões de uma ERB que possui localização fixa em $\mathbf{s}_{tx} = (100, 100)$, potência de transmissão igual a 0 dBm e que, com uma antena omnidirecional, atende uma área de cobertura com dimensão espacial de 300 m $\times$ 300 m.

A Tabela 2.2 mostra os parâmetros do ambiente de rádio e as configurações utilizadas na simulação. Os valores escolhidos para as configurações buscam representar os cenários de comunicações sem fio, especialmente os sistemas de comunicações sem fio externos (principal alvo deste trabalho). Especificamente, as configurações mostradas retratam um ambiente de propagação urbano tipicamente caracterizado pelas faixas $3 \leq \alpha \leq 5$ e $5 \text{ dB} \leq \sigma \leq 13 \text{ dB}$ para o coeficiente de propagação e para o desvio padrão do sombreamento log-normal, respectivamente [95-97]. Investigações realizadas por meio de campanhas de medidas indicam diferentes valores para a distância de decorrelação espacial (d_{corr}) do sombreamento, de acordo com o tipo de ambiente de propagação, *e.g.*, 10 m para microcélulas urbanas, 50 m a 120 m para ambientes urbanos e 50 m a 400 m para ambientes suburbanos [117]. Nesta simulação, foram coletadas $N = 200$ medidas do ambiente de rádio considerando a ausência da incerteza de localização *i.e.*, as coordenadas espaciais $\mathbf{s}$ são estimadas de forma perfeita pelos métodos de localização. Além disso, é assumido que não ocorrem variações do processo aleatório (*e.g.*, superposição de ruído, erros de medidas) em pequena escala (*i.e.*, $c_0 = 0$), uma vez que existe o interesse somente nos efeitos em larga escala do ambiente de rádio.

Os resultados de simulação podem ser observados na Figura 2.8, onde o ambiente de rádio é composto por (a) mapa da potência média de recepção devida à perda por percurso média e por (b) mapa do sombreamento log-normal do canal sem fio com covariância exponencial e isotrópica. A combinação dos mapas (a) e (b) resulta no processo aleatório espacial $P(\mathbf{s})$, que caracteriza o ambiente de rádio, mostrado em (c). Observa-se que o sombreamento log-normal influencia diretamente os valores de potência de recepção que compõem o processo $P(\mathbf{s})$.

Tabela 2.2: Parâmetros e Configurações da Simulação

Parâmetros do Ambiente de Rádio	Configurações de Simulação
Modelo de Perda por Percurso	Log-Distância ($\alpha = 4$)
Sombreamento	Log-normal (Gaussiano)
Desvio Padrão do Sombreamento	$\sigma = 5 \text{ dB}$
Distância de decorrelação do Sombreamento	$d_{\text{corr}} = 50 \text{ m}$
Amostragem Espacial	Uniforme
Incerteza de Localização	Ausente
Número de Medidas Coletadas	$N = 200$
Parâmetros da Covariância Espacial	$\theta = [m \ r \ c_0] = [25 \ 50 \ 0]$

Figura 2.8: Resultados de simulação: (a) potência média de recepção; (b) sombreamento do canal sem fio; (c) ambiente de rádio com amostragem espacial realizada a partir da densidade de probabilidade uniforme; (d) estimação da perda por percurso média; (e) semivariograma e (f) covariograma.

Os pontos da amostragem espacial (*i.e.*, coordenadas espaciais estimadas *s* sem incerteza de localização) são gerados na área de cobertura da ERB, a partir da densidade de probabilidade uniforme, resultando no conjunto de potências de recepção, $\{P_{rx}(\mathbf{s}_1), \dots, P_{rx}(\mathbf{s}_{200})\}$. Dentro deste contexto, é importante mencionar que a densidade de probabilidade utilizada na modelagem da amostragem espacial depende de características específicas das aplicações dos sistemas de comunicações sem fio (*e.g.*, campanhas de *drive-test*, nas quais as rotas são previamente definidas). Neste sentido, a densidade de probabilidade uniforme é utilizada na amostragem espacial das medidas nos casos onde não há um controle *a priori* sobre a localização dos dispositivos da rede ou a falta de conhecimento sobre os padrões de movimentação espacial associados com tais dispositivos. Em um momento posterior deste trabalho (capítulo 4), os impactos dos diferentes tipos de amostragem espacial serão analisados, especialmente nos resultados de predição espacial.

O conjunto $\{P_{rx}(\mathbf{s}_1), \dots, P_{rx}(\mathbf{s}_{200})\}$ será utilizado na aprendizagem de parâmetros para que a variabilidade espacial do processo aleatório, provocada pelo sombreamento log-normal, possa ser capturada pelos métodos geoestatísticos, visando as predições do REM. O êxito nesta captura ocorre quando a amostragem espacial consegue acompanhar as variações espaciais do ambiente de rádio. Isto permite que o preditor Kriging explore a covariância estimada para alcançar melhores predições.

O primeiro resultado de aprendizagem de parâmetros é mostrado em (d) e consiste na regressão do modelo da tendência por meio da aplicação do estimador (2.22) sobre o conjunto de medidas $\{P_{rx}(s_1), \dots, P_{rx}(s_{200})\}$, para obtenção de $\hat{\alpha}$. As medidas de perda por percurso afetadas pelo sombreamento log-normal são mostradas a fim de formar um comparativo entre o modelo da tendência utilizado (log-distância) e os resultados da regressão do modelo a partir da estimação de $\hat{\alpha}$. No caso da simulação, o resultado alcançado $\hat{\alpha} = 4,005$ indica um desempenho relativamente satisfatório, pois $\alpha = 4$. A diferença entre $\hat{\alpha}$ e α ocorre em razão da variabilidade das medidas em virtude do sombreamento e do conjunto finito de medidas coletado. Ainda assim, são os requisitos de desempenho relacionados com as aplicações que venham utilizar as previsões do REM que permitem qualificar as estatísticas obtidas com o estimador $\hat{\alpha}$.

A partir dos resultados de estimação de α , é possível obter $\hat{\mu}$ e remover a tendência das medidas $\{P_{rx}(s_1), \dots, P_{rx}(s_{200})\}$ para que o vetor $\mathbf{z}$, que representa as variações espaciais do sombreamento, seja obtido e utilizado nas etapas posteriores da aprendizagem de parâmetros. Os resultados de estimação do semivariograma e do covariograma são mostrados na forma de pontos em (e) e (f). A quantidade de pares de observações usados para o cálculo de cada semivariância e covariância também é mostrada nas caixas numéricas. Sobre as estimativas, é possível observar a variação no número de observações agrupadas para a obtenção dos resultados experimentais. Os agrupamentos foram realizados até a distância de 200 m, pois apenas um pequeno número de observações coletadas se distanciava acima de 200 m. Agrupamentos de distâncias lags entre 100 e 150 metros resultaram em mais de 1000 combinações (pares de observações), enquanto os agrupamentos relacionados distâncias lags muito pequenas não se beneficiaram de muitas combinações.

Os comportamentos crescente para o semivariograma e decrescente para o covariograma refletem o aumento da dissimilaridade e a redução da similaridade à medida em que a distância entre os pares de observações é aumentada. Nota-se a formação de um patamar no semivariograma experimental que se aproxima da variância do sombreamento do canal sem fio ($\sigma^2 = 25$), mostrada com a linha tracejada em (e). Tal patamar é formado a partir de distâncias lags maiores que 50 m. Um resultado similar pode ser visualizado no comportamento do covariograma, onde uma queda mais significativa da covariância ocorre a partir de distâncias acerca de 50 m. Finalmente, as linhas sólidas em (e) e (f) consistem nos resultados de ajuste dos modelos analíticos exponenciais contemplando a estimação de $\hat{\theta} = [22,02 \ 43,46 \ 0]$ com a técnica LS. As diferenças entre $\hat{\theta} \neq \theta$ e $\hat{\alpha} \neq \alpha$ ocorrem em razão das características do ambiente de rádio, da distribuição e o número de medidas da amostragem espacial e, sobretudo do desempenho atingido pelas técnicas de estimação da abordagem geoestatística. O resultado obtido com o vetor $\hat{\theta}$ finaliza a fase de aprendizagem de parâmetros e se mostra essencial por duas razões: i) permite que informações possam ser extraídas sobre o ambiente de rádio e ii) permite que o preditor Kriging possa explorar a covariância espacial capturada a favor das previsões na geração do REM. A próxima subseção se dedica a mostrar os detalhes sobre o funcionamento do preditor Kriging, iniciando a fase de previsões do REM.

2.3.5 Preditor Kriging - Processos Gaussianos

A aprendizagem de parâmetros possibilita a predição espacial, mas tão importante quanto a realização das predições é a obtenção de um valor com elevada acurácia (*i.e.*, confiável) para a potência de recepção em coordenadas espaciais onde não foram realizadas as medidas, especialmente no caso do REM. A teoria de processos Gaussianos fornece o preditor Kriging (da terminologia geoestatística), que consiste no melhor preditor linear para o problema de predição no sentido de minimização do erro quadrático médio de predição espacial [58-68]. Antes de apresentar as particularidades do Kriging, é fundamental analisar o problema da predição de forma geral, envolvendo a minimização do erro quadrático médio de predição (MSPE - *Mean Square Prediction Error*).

Em um primeiro momento, para encontrar o preditor geral que minimiza o MSPE, considera-se o vetor de variáveis aleatórias observadas, que representam a VR, denotado como $\mathbf{z}$, e a variável T , que consiste na VA, cujos valores aleatórios devem ser preditos, a partir do vetor de medidas $\mathbf{z}$. O preditor espacial $\hat{T}$ é uma função das medidas observadas, *i.e.*, $\hat{T} = f(\mathbf{z})$ e o seu MSPE é dado por

$$\text{MSPE}(\hat{T}) = \mathbb{E}\{(T - \hat{T})^2\}, \quad (2.34)$$

em que $\mathbb{E}\{\cdot\}$ é o operador esperança, relacionado com a distribuição conjunta de T e $\mathbf{z}$ através de $\hat{T}$. A minimização do MSPE($\hat{T}$) é obtida quando o preditor consiste no valor esperado condicionado às observações, *i.e.*, esperança condicional, $\hat{T} = \mathbb{E}\{T|\mathbf{z}\}$, conforme o desenvolvimento [68, pp.134-135]:

Demonstração.

$$\mathbb{E}\{(T - \hat{T})^2\} = \mathbb{E}_{\mathbf{z}}\{\mathbb{E}_T\{(T - \hat{T})^2|\mathbf{z}\}\}, \quad (2.35)$$

em que $\mathbb{E}_{\mathbf{z}}\{\cdot\}$ e $\mathbb{E}_T\{\cdot\}$ são aplicados em relação à $\mathbf{z}$ e à T , respectivamente. Da definição da variância, o termo $\mathbb{E}_T\{(T - \hat{T})^2|\mathbf{z}\}$ pode ser reescrito de acordo com

$$\mathbb{E}_T\{(T - \hat{T})^2|\mathbf{z}\} = \text{Var}_T\{(T - \hat{T})|\mathbf{z}\} + [\mathbb{E}_T\{(T - \hat{T})|\mathbf{z}\}]^2. \quad (2.36)$$

Condicional à $\mathbf{z}$, a função do preditor $\hat{T} = f(\mathbf{z})$, resulta em um valor constante. Desta forma, tem-se que $\text{Var}_T\{(T - \hat{T})|\mathbf{z}\} = \text{Var}_T\{T|\mathbf{z}\}$ e $\mathbb{E}_T\{(T - \hat{T})|\mathbf{z}\} = \mathbb{E}_T\{T|\mathbf{z}\} - \hat{T}$, resultando em

$$\mathbb{E}_T\{(T - \hat{T})^2|\mathbf{z}\} = \text{Var}_T\{T|\mathbf{z}\} + [\mathbb{E}_T\{T|\mathbf{z}\} - \hat{T}]^2. \quad (2.37)$$

Aplicando o operador $\mathbb{E}_{\mathbf{z}}\{\cdot\}$ sobre a equação (2.37), temos a expressão do MSPE,

$$\mathbb{E}_{\mathbf{z}}\{\mathbb{E}_T\{(T - \hat{T})^2|\mathbf{z}\}\} = \mathbb{E}\{(T - \hat{T})^2\} = \mathbb{E}_{\mathbf{z}}\{\text{Var}_T\{T|\mathbf{z}\}\} + \mathbb{E}_{\mathbf{z}}\{[\mathbb{E}_T\{T|\mathbf{z}\} - \hat{T}]^2\}. \quad (2.38)$$

Verifica-se que somente o segundo termo de (2.38) depende da forma do preditor $\hat{T}$ e vai à zero se, e somente se, $\hat{T} = \mathbb{E}\{T|\mathbf{z}\}$ levando à minimização do MSPE. $\square$

É fundamental mencionar que a obtenção do preditor $\hat{T} = \mathbb{E}\{T|\mathbf{z}\}$ segue diretamente da escolha do MSPE como critério a ser otimizado, ou seja, o preditor dado pela esperança condicional é ótimo no sentido de minimização do MSPE, que não necessariamente é a melhor métrica para a avaliação de desempenho em aplicações específicas que se baseiam em previsões. Ainda assim, em função do interesse específico na acurácia e precisão das previsões espaciais, a métrica MSPE e a sua raiz quadrada (RMSPE - *Root Mean Square Prediction Error*) são amplamente utilizadas na análise de desempenho das previsões em sistemas de comunicações sem fio [16-18], [25-28], [49-50], [72-73], [78-82] e, por isso são adotadas neste trabalho. De forma similar ao caso da estimação da tendência, são os requisitos de desempenho das aplicações que permitem qualificar os resultados de MSPE alcançados com as previsões espaciais. Em todo caso, busca-se por preditores com elevada acurácia, robustos frente aos efeitos do canal e de complexidade não proibitiva, além de se adequarem ao funcionamento e a disponibilidade de recursos das aplicações.

O preditor Kriging visa atender as características citadas e, de fato, a linearidade da regressão de distribuições Gaussianas multivariadas permite demonstrar que a técnica Kriging converge para o preditor da esperança condicional quando o processo aleatório espacial é Gaussiano [63, pp. 638]. No ambiente de rádio, a característica Gaussiana deriva diretamente do sombreamento log-normal do canal sem fio. Com isso, diferentes variações do método Kriging podem ser utilizadas em função das características do problema de previsão, especialmente da média do processo aleatório $P(\mathbf{s})$.

A Tabela 2.3 mostra as variantes dos métodos Kriging, de acordo com o modelo da tendência. Nas circunstâncias em que o valor da média μ é conhecido *a priori*, o método Kriging simples (KS) é aplicado. Neste caso, não há um modelo de tendência para os dados, uma vez que a média μ é conhecida. Entretanto, não é comum assumir o conhecimento da média μ , sobretudo na prática, pois isto exige o acesso a várias realizações do processo $P(\mathbf{s})$. Particularmente, no caso dos sistemas de comunicações sem fio, o envio de medidas ao longo do tempo para a ERB (acesso a mais realizações do processo aleatório espacial), depende da capacidade da rede em prover os recursos de banda e energia necessários, além do controle para tais transmissões. Esta situação pode ser tornar crítica em aplicações que possuem restrições de recursos ou exigem consumo mínimo de energia nos dispositivos (*e.g.*, redes de sensores). De outro modo, o método Kriging ordinário (KO) assume a estacionaridade da média do processo aleatório $P(\mathbf{s})$, ou seja, μ é fixa e constante ao longo do domínio espacial D , mas precisa ser estimada [62-68]. Deste modo, o preditor KO se concentra nas variações espaciais em torno da média (estimada) do processo aleatório $P(\mathbf{s})$.

Tabela 2.3: Previsões Espaciais - Métodos Kriging

Método Kriging	Média	Modelo da Tendência
Kriging Simples (KS)	Conhecida	Nenhum
Kriging Ordinário (KO)	Desconhecida	Constante
Kriging Universal (KU)	Desconhecida	Função das Coordenadas Espaciais

O comprometimento da estacionaridade em função da presença da tendência leva ao uso do preditor Kriging universal (KU), que busca realizar a modelagem não estacionária da tendência, incorporando-a ao funcionamento do preditor espacial. A principal dificuldade na aplicação do KU está na estimação do semivariograma, que será afetado pela característica não estacionária da tendência e que possui parâmetros a serem estimados [146]. Além de comprometer o desempenho de predição, isto implica na necessidade do conhecimento *a priori* do semivariograma do processo aleatório espacial, além de restrições quanto às funções base da tendência, que devem ser linearmente independentes [62-63]. Geralmente, o KU considera monômios de baixo grau e superfícies suaves (cujas potências não excedam o grau polinomial dois) para a modelagem não estacionária da tendência [62], [67]. No caso específico do ambiente de rádio, a perda por percurso é modelada com funções não lineares nas coordenadas espaciais (raiz quadrada e logaritmos), incluindo expoentes que podem ser maiores que dois, o que pode elevar a complexidade da incorporação da tendência ao preditor KU. A situação se torna crítica nos casos onde mais funções base são utilizadas (*e.g.*, uso de modelos de perda por percurso mais complexos).

Frente a esta situação, este trabalho adota a estratégia *detrending*, que prioriza a estimação seguida da remoção da tendência das medidas capturadas para que os valores regionalizados possam ser obtidos e o preditor KO possa ser aplicado ao problema de predição. De forma analítica, a estimativa produzida pelo preditor KO consiste na combinação linear das medidas $\mathbf{z} = [Z(\mathbf{s}_1) \dots Z(\mathbf{s}_N)]^T$, de acordo com

$$Z_{\text{KO}}(\mathbf{s}_0) = \boldsymbol{\lambda}^T \mathbf{z} = \sum_{i=1}^N \lambda_i Z(\mathbf{s}_i), \quad (2.39)$$

em que $Z_{\text{KO}}(\mathbf{s}_0)$ é o resultado escalar da predição em uma coordenada espacial desconhecida $\mathbf{s}_0$, obtida por meio do vetor coluna de pesos Kriging $\boldsymbol{\lambda}$, com dimensões $N \times 1$, e do vetor coluna de medidas $\mathbf{z}$, com dimensões $N \times 1$. É importante mencionar que a posição desconhecida $\mathbf{s}_0$ em (2.39) é levada em conta através da obtenção dos pesos Kriging λ_i , cujos valores dependem das covariâncias relacionadas com a coordenada espacial $\mathbf{s}_0$, que é alvo da predição. É por esta razão que o modelo analítico do semivariograma e seus parâmetros $\boldsymbol{\theta}$ são essenciais para os preditores Kriging. As características do preditor espacial KO estão associadas a dois fatores importantes: i) a busca pelos pesos ótimos Kriging λ_i que minimizam o MSPE, que consiste na variância do erro de predição espacial, $\text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\}$ e ii) pela condição imposta aos pesos Kriging, assumindo a estacionariedade, de modo que (2.39) seja não enviesado, de acordo com:

$$\begin{aligned} \mathbb{E}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\} &= \mathbb{E}\left\{\sum_{i=1}^N \lambda_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0)\right\} = \sum_{i=1}^N \lambda_i \mathbb{E}\{Z(\mathbf{s}_i)\} - \mathbb{E}\{Z(\mathbf{s}_0)\} = 0 \\ &= \sum_{i=1}^N \lambda_i \mu - \mu = 0 \implies \sum_{i=1}^N \lambda_i = 1. \end{aligned} \quad (2.40)$$

A variância do erro de predição espacial é expressa por

$$\text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\} = \mathbb{E}\left\{\left(Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right)^2\right\} - \mathbb{E}\left\{\left(Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right)\right\}^2. \quad (2.41)$$

Considerando a condição de não viés para o segundo termo em (2.41) e usando a definição do preditor KO em (2.39), é possível expandir o primeiro termo de (2.41), resultando na seguinte expressão para a variância do erro de predição espacial

$$\begin{aligned} \text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\} &= \mathbb{E}\left\{Z_{\text{KO}}^2(\mathbf{s}_0) - 2Z_{\text{KO}}(\mathbf{s}_0)Z(\mathbf{s}_0) + Z^2(\mathbf{s}_0)\right\} \\ &= \mathbb{E}\left\{Z_{\text{KO}}^2(\mathbf{s}_0)\right\} + \mathbb{E}\left\{Z^2(\mathbf{s}_0)\right\} - 2\mathbb{E}\left\{Z_{\text{KO}}(\mathbf{s}_0)Z(\mathbf{s}_0)\right\} \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \mathbb{E}\{Z(\mathbf{s}_i)Z(\mathbf{s}_j)\} - 2 \sum_{i=1}^N \lambda_i \mathbb{E}\{Z(\mathbf{s}_i)Z(\mathbf{s}_0)\} + \mathbb{E}\{Z^2(\mathbf{s}_0)\}. \end{aligned} \quad (2.42)$$

Ressaltando as seguintes definições da função covariância espacial

$$\begin{aligned} \mathbb{E}\{Z(\mathbf{s}_i)Z(\mathbf{s}_j)\} &= C(\mathbf{s}_i, \mathbf{s}_j) + \mu^2; \\ \mathbb{E}\{Z(\mathbf{s}_i)Z(\mathbf{s}_0)\} &= C(\mathbf{s}_i, \mathbf{s}_0) + \mu^2; \\ \mathbb{E}\{Z^2(\mathbf{s}_0)\} &= C(0) + \mu^2, \end{aligned} \quad (2.43)$$

e substituindo-as em (2.42), torna-se possível encontrar a relação entre a variância do erro de predição e a função covariância, dada por

$$\text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\} = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - 2 \sum_{i=1}^N \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) + C(0). \quad (2.44)$$

Conforme mencionado, o problema central do preditor KO consiste em encontrar os valores de λ_i que minimizam (2.44) sob a condição de não viés em (2.40). A solução é encontrada com a aplicação do método dos multiplicadores de Lagrange através da seguinte função lagrangiana [63, pp. 163],

$$\phi(\lambda_i, \nu) = \text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\} + 2\nu \left\{ \sum_{i=1}^N \lambda_i - 1 \right\}, \quad (2.45)$$

em que ν é o multiplicador de Lagrange. Derivando a função (2.45) em relação aos pesos Kriging e ao operador de Lagrange, é possível obter

$$\begin{aligned} \frac{\partial \phi(\lambda_i, \nu)}{\partial \lambda_i} &= +2 \sum_{j=1}^N \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - 2 \sum_{i=1}^N C(\mathbf{s}_i, \mathbf{s}_0) + 2\nu = 0, \\ \frac{\partial \phi(\lambda_i, \nu)}{\partial \nu} &= +2 \left\{ \sum_{i=1}^N \lambda_i - 1 \right\} = 0, \quad \text{com } i, j = 1, \dots, N. \end{aligned} \quad (2.46)$$

Este resultado permite formular o sistema composto por $N + 1$ equações de covariâncias do preditor espacial KO, expresso de acordo com

$$\begin{cases} \sum_j \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) + \nu = C(\mathbf{s}_i, \mathbf{s}_0), & i, j = 1, \dots, N, \\ \sum_i \lambda_i = 1. \end{cases} \quad (2.47)$$

Reescrevendo o sistema de equações em (2.47) de forma matricial, tem-se $\mathbf{C}\boldsymbol{\lambda}_\nu = \mathbf{c}_0$, *i.e.*,

$$\begin{bmatrix} C(\mathbf{s}_1, \mathbf{s}_1) & \dots & C(\mathbf{s}_1, \mathbf{s}_N) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ C(\mathbf{s}_N, \mathbf{s}_1) & \dots & C(\mathbf{s}_N, \mathbf{s}_N) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \nu \end{bmatrix} = \begin{bmatrix} C(\mathbf{s}_1, \mathbf{s}_0) \\ \vdots \\ C(\mathbf{s}_N, \mathbf{s}_0) \\ 1 \end{bmatrix},$$

em que $\mathbf{C}$ consiste na matriz de covariâncias obtida entre todas as coordenadas espaciais $\mathbf{s}$ relacionadas com as medidas observadas, $\boldsymbol{\lambda}_\nu$ é o vetor de pesos Kriging (incluindo o multiplicador de Lagrange ν) e $\mathbf{c}_0$ é o vetor de covariâncias entre as coordenadas espaciais $\mathbf{s}$ e a coordenada espacial alvo da predição $\mathbf{s}_0$. É importante notar que a composição das matrizes $\mathbf{C}$ e $\mathbf{c}_0$ é realizada a partir da aplicação do modelo analítico de covariância e do vetor $\hat{\boldsymbol{\theta}}$ obtido na fase de aprendizagem de parâmetros. Assim, o cálculo dos pesos ótimos Kriging para a solução do sistema de equações envolve a inversão matricial

$$\boldsymbol{\lambda}_\nu = \mathbf{C}^{-1} \mathbf{c}_0. \quad (2.48)$$

A relação entre as funções covariância e semivariograma, $C(\mathbf{s}_i, \mathbf{s}_j) = C(0) - \gamma(\mathbf{s}_i, \mathbf{s}_j)$, pode ser aplicada em (2.47) para reescrever o sistema de equações em função do semivariograma, $\mathbf{\Gamma}\boldsymbol{\lambda}_\nu = \boldsymbol{\beta}$,

$$\begin{bmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_1) & \dots & \gamma(\mathbf{s}_1, \mathbf{s}_N) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_1) & \dots & \gamma(\mathbf{s}_N, \mathbf{s}_N) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ -\nu \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_0) \\ 1 \end{bmatrix}.$$

em que $\mathbf{\Gamma}$ é a matriz de semivariâncias entre as coordenadas espaciais $\mathbf{s}$ e $\boldsymbol{\beta}$ é o vetor de semivariâncias que relaciona as coordenadas espaciais $\mathbf{s}$ e a coordenada alvo da predição $\mathbf{s}_0$. A predição espacial $Z_{\text{KO}}(\mathbf{s}_0)$ é dada pela combinação linear entre as medidas do vetor $\mathbf{z}$ e os pesos Kriging λ_i obtidos em (2.48). Finalmente, a predição da potência de recepção na coordenada alvo $\mathbf{s}_0$ é obtida com a estimativa da potência média de recepção $\hat{\mu}(\mathbf{s}_0)$ e a predição $Z_{\text{KO}}(\mathbf{s}_0)$, de acordo com

$$\hat{P}_{\text{rx}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + Z_{\text{KO}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + \sum_{i=1}^N \lambda_i Z(\mathbf{s}_i). \quad (2.49)$$

2.3.6 Simulação da Fase II - Geração do REM

A Figura 2.9 mostra os resultados da Fase II para a geração do REM com a aplicação do preditor Kriging, a partir das medidas coletadas do processo aleatório espacial mostrado anteriormente (Figura 2.8). A Fig.2.9(a) apresenta o mapa verdadeiro da potência de recepção, submetido aos efeitos da perda por percurso média e do sombreamento log-normal, enquanto a geração do REM, obtida através da combinação da estimação da tendência com as previsões KO, é mostrada na Fig.2.9(b). Por meio do comparativo entre a Fig.2.9(d) e a Fig.2.9(e), é possível observar que o preditor KO foi capaz de capturar a variabilidade espacial contida nas medidas e que é necessária para as previsões sobre o ambiente de rádio. Ainda assim, os mapas da Fig.2.9(c) e da Fig.2.9(f) mostram que a magnitude das diferenças entre as previsões de potência de recepção e do sombreamento comparadas aos valores verdadeiros do ambiente de rádio flutuam espacialmente nos mapas. Embora as diferenças sejam relativamente baixas em uma parte significativa da área de cobertura da ERB, é possível observar algumas regiões nas quais as previsões não foram precisas. Estas diferenças refletem a necessidade de avaliar o desempenho esperado dos preditores a fim de verificar se os métodos geoestatísticos consistem em uma solução efetiva na geração confiável do REM. Sobre este aspecto, é importante ressaltar que o Kriging provê uma estimativa esperada da potência de recepção e que apresenta melhor acurácia, pois é obtida no sentido de minimização do MSPE.

Figura 2.9: Resultados de simulação: (a) potência de recepção no ambiente de rádio; (b) REM; (c) mapa diferença: potência de recepção e o REM; (d) sombreamento log-normal do canal sem fio; (e) previsões espaciais via KO e (f) mapa diferença: sombreamento e previsões KO.

Uma característica interessante dos métodos Kriging consiste no fornecimento de uma medida de dispersão das estimativas de predição produzidas, chamada de variância Kriging (VK). De forma analítica, a VK é denotada como $\sigma_{\text{KO}}^2(\mathbf{s}_0)$, e expressa a variância do erro de predição espacial, *i.e.*, $\sigma_{\text{KO}}^2(\mathbf{s}_0) = \text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\}$. O desenvolvimento analítico abaixo traz a demonstração para a obtenção da VK. Especificamente, σ_{KO}^2 é obtida através da multiplicação do sistema de N equações em (2.47), dado por $\sum_j \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - \nu = C(\mathbf{s}_i, \mathbf{s}_0)$, pelo fator $\sum_i \lambda_i$, com $\sum_i \lambda_i = 1$ e $i, j = 1, \dots, N$.

Demonstração. Desenvolvimento sobre o sistema de equações Kriging

$$\begin{aligned}
 \sum_j \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) + \nu &= C(\mathbf{s}_i, \mathbf{s}_0), \\
 \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) + \sum_i \lambda_i \nu &= \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0), \\
 \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) &= \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) - \sum_i \lambda_i \nu, \\
 \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) &= \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) - \nu.
 \end{aligned} \tag{2.50}$$

Realizando as substituições na variância Kriging mostrada em (2.44),

$$\begin{aligned}
 \sigma_{\text{KO}}^2(\mathbf{s}_0) &= \text{Var}\{Z_{\text{KO}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\} = \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - 2 \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) + C(0) \\
 &= \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) - \sum_i \lambda_i \nu - 2 \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) + C(0) \\
 &= C(0) - \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) - \nu.
 \end{aligned} \tag{2.51}$$

Utilizando a relação direta entre a covariância e o semivariograma, é possível reescrever $\sigma_{\text{KO}}^2(\mathbf{s}_0)$,

$$\begin{aligned}
 \sigma_{\text{KO}}^2(\mathbf{s}_0) &= C(0) - \sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) - \nu \\
 &= C(0) - \sum_i \lambda_i \{C(0) - \gamma(\mathbf{s}_i, \mathbf{s}_0)\} - \nu \\
 &= \sum_i \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}_0) - \nu = \boldsymbol{\beta}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\beta}.
 \end{aligned} \tag{2.52}$$

□

Os resultados mostrados em (2.51) e (2.52) indicam que a VK depende da relação de covariância (semivariância) entre as coordenadas espaciais $\mathbf{s}$ e $\mathbf{s}_0$, dos pesos Kriging λ_i e da variância $\sigma^2 = C(0)$. Com isso, a medida de dispersão sobre as predições em $\mathbf{s}_0$ é menor se a parcela $\sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) - \nu$ for significativamente grande comparada à $C(0)$ ou se $\sum_i \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}_0) - \nu = \boldsymbol{\beta}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\beta}$ for mínimo.

Deste modo, se a covariância espacial entre as medidas nas coordenadas $\mathbf{s}$ e na coordenada alvo $\mathbf{s}_0$ for pequena, a variância associada ao erro de predição tende a aumentar. Se tal covariância for alta, a variância dos erros de predição espacial tende a reduzir. Isto ressalta a influência da distância de separação entre as medidas ($\mathbf{h}$), que determina a relação de covariância entre as observações por meio das parcelas $\sum_i \lambda_i C(\mathbf{s}_i, \mathbf{s}_0)$ e $\sum_i \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}_0)$, e do próprio modelo da covariância (semivariograma) na dispersão associada aos resultados de predição.

Neste contexto, diversos trabalhos utilizam a VK em predições espaciais (*e.g.*, em comunicações sem fio [27-28], [72-73]), principalmente nas aplicações da geoestatística relacionadas com a indústria da mineração (*e.g.*, cálculo de intervalos de confiança para classificações em reservas minerais [60-63], [118-119]). No entanto, é importante notar que as equações da VK em (2.51) e (2.52) dependem fundamentalmente do modelo de covariância entre as coordenadas espaciais e não utilizam diretamente as medidas do vetor $\mathbf{z}$. Em outras palavras, o valor da VK não depende diretamente das medidas e pode ser calculado a partir do arranjo espacial entre as observações, descrito pelo modelo de covariância. Isto significa que conjuntos de medidas diferentes, mas com arranjos espaciais idênticos, possuem a mesma incerteza de predição, retratando uma característica homocedástica para a VK [60]. Frente a esta situação, Yamamoto propõe a variância de interpolação (VI) e argumenta que a VK não é o melhor indicador de incerteza das estimativas para problemas de classificação, pois falha na captura da dispersão local das predições em razão de ser exclusivamente baseada em uma descrição espacial global do processo aleatório (semivariograma) e não nas medidas coletadas [120-121]. De forma específica, a VI visa contornar esta característica da VK, levando em conta a dispersão local associada às medidas no cômputo da variância dos erros de predição, de acordo com

$$\sigma_{VI}^2(\mathbf{s}_0) = \sum_{i=1}^N \lambda_i [Z(\mathbf{s}_i) - Z_{KO}(\mathbf{s}_0)]^2. \quad (2.53)$$

Verifica-se que a variância de interpolação associada a uma coordenada espacial alvo $\mathbf{s}_0$ leva em conta as medidas do vetor $\mathbf{z}$ e o modelo de covariância espacial através dos pesos Kriging λ_i . A diferença $[Z(\mathbf{s}_i) - Z_{KO}(\mathbf{s}_0)]^2$ mede a dispersão das predições em relação ao conjunto de medidas utilizado para capturar a variabilidade espacial do processo aleatório $P(\mathbf{s})$ e, neste sentido, o cálculo da VI amplifica as diferenças quando os valores de predições se diferenciam das medidas capturadas.

A Figura 2.10 mostra os mapas: (a) da VK; (b) do desvio padrão Kriging e (c) da VI, todos relacionados com os resultados das predições espaciais mostradas na Figura 2.9. É possível observar que os menores valores da VK (e do desvio padrão Kriging) ocorrem nos pontos exatos da amostragem espacial usados na captura das medidas, pois se a distância entre $\mathbf{s}_0$ e $\mathbf{s}$ é pequena, a covariância $C(\mathbf{s}_i, \mathbf{s}_0)$ tende a ser alta, reduzindo a variância $\sigma_{KO}^2(\mathbf{s}_0)$. De outro modo, à medida em que a distância entre uma coordenada espacial alvo $\mathbf{s}_0$ e $\mathbf{s}$ é aumentada, a VK assume valores maiores, indicando maior incerteza associada às predições em $\mathbf{s}_0$. Trata-se de um comportamento fundamentado na relação de covariância entre as coordenadas espaciais do processo aleatório.

Figura 2.10: (a) variância Kriging; (b) desvio padrão Kriging e (c) variância de interpolação.

Os resultados da VI são diferentes e, de certa forma, acompanham a variabilidade espacial dos resultados de predição, indicando as regiões em que as predições espaciais foram significativamente altas ou baixas. Com isso, o funcionamento e consequentemente os resultados da VK e da VI são fundamentalmente distintos, retratando as discussões sobre a efetividade no uso destes indicadores de incerteza, especialmente nos problemas envolvendo classificações baseadas nas predições alcançadas. Ainda assim, é fundamental ressaltar que a discussão sobre a independência das medidas na VK deve ser ponderada por uma análise que inclui as pressuposições relacionadas com a concepção do preditor Kriging. Em específico, quando o processo aleatório espacial é Gaussiano, o sistema de equações Kriging converge para o preditor da esperança condicional $\hat{T} = \mathbb{E}\{T|\mathbf{z}\}$ e com isso, seu desenvolvimento resulta diretamente na expressão da VK [68, pp. 136]. Em outras palavras, a relação indireta da VK com as medidas consiste em uma característica específica da variância condicional (relacionado ao preditor da esperança condicional), quando a distribuição das observações é Gaussiana multivariada¹²[128]. Este trabalho está focado na geração das predições espaciais do ambiente de rádio sem considerar posteriores procedimentos de classificação. Ainda assim, detalhes que abrangem as discussões sobre o uso da VK e VI podem ser acessados em [60], [120-121], [128].

2.4 Conclusões - Capítulo 2

Este capítulo apresentou os detalhes da modelagem do ambiente de rádio e dos métodos geostatísticos utilizados no modelo de predição do REM, envolvendo os aspectos da aprendizagem de parâmetros e das predições espaciais. Ressalta-se que, até o presente momento, assumiu-se que as coordenadas espaciais $\mathbf{s}$ são perfeitamente estimadas pela ERB. Neste sentido, é intuitivo pensar que a presença de erros de localização das medidas poderá, de alguma forma, afetar os resultados de predição espacial. Com isso, o próximo capítulo incorpora a de incerteza de localização dos sistemas de comunicações sem fio e se propõe a investigar seus impactos na geração do REM.

¹²No caso em que o processo aleatório espacial não é Gaussiano, a variância associada às predições a partir da distribuição condicional não irá convergir para as expressões da VK e, dependerá do conjunto de medidas [128].

Capítulo 3

Mitigação da Incerteza de Localização

Este capítulo introduz o modelo da incerteza de localização nas predições espaciais do ambiente de rádio em um sistema de comunicação sem fio. Com isso, os impactos dos erros de localização sobre as etapas do modelo de predição do REM são discutidos e um preditor espacial é proposto, visando a robustez frente ao problema de incerteza de localização nas predições espaciais.

A precisão das estimativas de localização em sistemas de comunicações sem fio depende de fatores como o método de estimação de localização utilizado e os efeitos do canal sem fio nas condições de recepção. Diferentes métodos de localização se propõem a fornecer estimativas de localização precisas (*e.g.*, métodos baseados em rede, nos próprios dispositivos e técnicas envolvendo uma combinação entre os dispositivos e a rede [74]). O fato é que o funcionamento dos métodos e os recursos disponíveis para a estimação de localização (*e.g.*, largura de banda e energia) influenciam as estimativas de localização. Sobre a precisão dos principais métodos: as estimativas podem resultar em erros de localização na faixa entre 5 m e 50 m para as técnicas baseadas na tecnologia GPS [85-86] e até 300 m (ou mais) para as diferentes técnicas baseadas na rede de comunicação sem fio [74], [78], [86-87]. Neste contexto, apesar da evolução dos métodos de localização, é razoável admitir que erros de localização possam ocorrer, justificando a necessidade da modelagem estatística da incerteza de localização não somente em predições espaciais, mas também em outras tarefas que se baseiam na informação de localização obtida na rede.

Até o momento, foi assumido que as coordenadas espaciais $\mathbf{s}$ refletiam as posições exatas dos dispositivos da rede, ou seja, os valores de potência de recepção do conjunto $\{P_{\text{rx}}(\mathbf{s}_1), \dots, P_{\text{rx}}(\mathbf{s}_N)\}$ foram associados ao conjunto de coordenadas espaciais $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ estimadas sem erros de localização (*i.e.*, equivalente ao uso de métodos de localização ideais). Neste capítulo, as coordenadas espaciais do conjunto $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ são afetadas por uma incerteza de localização.

A Figura 3.1 mostra a realização de um processo aleatório espacial submetido ao problema dos erros de localização. Algumas pressuposições iniciais são importantes: i) os dispositivos da rede são distribuídos de acordo com a densidade de probabilidade uniforme na área de cobertura da ERB e estão posicionados nas coordenadas espaciais realizadas $\mathbf{r}$, que consistem em suas verdadeiras posições; ii) cada dispositivo é capaz de medir a intensidade dos sinais recebidos e enviar as medições para a ERB, formando o conjunto de potências de recepção coletadas $\{P_{rx}(\mathbf{r}_1), \dots, P_{rx}(\mathbf{r}_N)\}$ e iii) o método de localização utilizado na rede realiza a estimação da localização de cada dispositivo, resultando nas coordenadas espaciais estimadas $\mathbf{s}$. Entretanto, há uma incerteza sobre as estimativas das coordenadas espaciais $\mathbf{s}$, pois estas são afetadas pelos erros de localização presentes no sistema de comunicação sem fio. Isto significa que as N medidas coletadas do processo aleatório espacial mostrado na Fig.3.1(a) estão relacionadas com o conjunto de coordenadas espaciais estimadas $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, que possuem incerteza de localização, enquanto o conjunto de coordenadas espaciais realizadas $\{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ consiste no verdadeiro posicionamento dos dispositivos.

Deste modo, o modelo de incerteza de localização, ilustrado na Fig.3.1(b), pode ser descrito de forma analítica de acordo com

$$\mathbf{r} = \mathbf{s} + \mathbf{p}_e(\mathbf{s}), \quad (3.1)$$

em que $\mathbf{p}_e(\mathbf{s}) \equiv \mathbf{r} - \mathbf{s}$ consiste no vetor aleatório de erro de localização bidimensional, que segue a densidade de probabilidade $g(\cdot)$ e atua corrompendo as coordenadas espaciais $\mathbf{s}$ do modelo (3.1). Com isso, se não há incerteza sobre as estimativas de localização $\mathbf{s}$, *i.e.*, $\mathbf{p}_e(\mathbf{s}) = 0$, a verdadeira coordenada dos dispositivos $\mathbf{r}$ é conhecida pela ERB. Em outras palavras, a incerteza dos métodos de localização impede que as estimativas que resultam nas coordenadas $\mathbf{s}$ atinjam a posição espacial $\mathbf{r}$, enquanto a diferença depende das características de $\mathbf{p}_e(\mathbf{s})$. O problema da incerteza de localização surge porque os verdadeiros valores de potência de recepção $\{P_{rx}(\mathbf{r}_1), \dots, P_{rx}(\mathbf{r}_N)\}$ são equivocadamente associados ao conjunto de coordenadas espaciais estimadas $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ e não ao conjunto $\{\mathbf{r}_1, \dots, \mathbf{r}_N\}$.

Figura 3.1: (a) ambiente de rádio gerado com erros de localização e (b) modelo de incerteza.

A associação entre as medidas de potência de recepção e as coordenadas espaciais com erros de localização implica na observação de um novo processo aleatório espacial, corrompido pela incerteza de localização e denotado como $P_g(\mathbf{s})$. Como as técnicas de estimação e predição se baseiam nas medidas e distâncias calculadas por meio das informações de localização, é possível inferir que $\mathbf{p}_e(\mathbf{s})$ terá impacto nas etapas do modelo de predição espacial do REM. Com isso, a seleção de um modelo estatístico para a distribuição $g(\cdot)$ dos erros de localização $\mathbf{p}_e(\mathbf{s})$ é uma decisão importante. De fato, as diferenças de desempenho entre os métodos de localização utilizados nos sistemas de comunicações sem fio podem resultar em diferentes caracterizações estatísticas para a incerteza de localização. Um exemplo consiste no uso da tecnologia GPS, na qual o número de satélites disponíveis e a geometria formada para a obtenção das estimativas de posicionamento são dependentes da localização [80]. Em comunicações móveis, os efeitos do canal de propagação e o desempenho dos métodos de localização influenciam diretamente as estatísticas de $\mathbf{p}_e(\mathbf{s})$.

Neste contexto, dois modelos com diferentes distribuições de probabilidade para $\mathbf{p}_e(\mathbf{s})$ são evidenciados na literatura [91-93]: i) vetor de erro de localização bidimensional com densidade de probabilidade Gaussiana, $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_d)$, em que σ_p é o desvio padrão dos erros de localização, dado em metros e $\mathbf{I}_d$ é uma matriz identidade, em que d é a dimensão do domínio espacial; ii) vetor de erro de localização bidimensional com densidade de probabilidade uniforme, $\mathbf{p}_e(\mathbf{s}) \sim \text{Unif}(-\sqrt{3}\sigma_p, +\sqrt{3}\sigma_p)$. Em ambos os modelos, a severidade dos erros de localização é expressa pelo desvio padrão σ_p . Métodos de localização com desempenho ruim geram estimativas com grande incerteza, caracterizada por altos valores para σ_p . De outro modo, sistemas de comunicação sem fio com métodos de localização melhores oferecem estimativas com menor incerteza (*i.e.*, valores menores de σ_p). As diferentes faixas de erros de localização (5 m a 50 m para o GPS e 50 m a 300 m para os métodos baseados em rede) permitem ter uma ideia mais clara sobre os possíveis intervalos a serem considerados para a escolha do σ_p do modelo de incerteza.

Gabrosek, em [91], investigou as características do processo $P_g(\mathbf{s})$, apontando que os impactos de $\mathbf{p}_e(\mathbf{s})$ com distribuições uniforme e Gaussiana em diferentes modelos de covariância foram similares (efeitos que serão discutidos mais adiante). Em [84] e [93], Diggle e Cressie examinaram diferentes modelos estatísticos dos erros de localização e constataram que o parâmetro σ_p tem peso mais significativo no desempenho das predições Kriging do que o formato específico da distribuição $g(\cdot)$. Ainda assim, é fundamental que a escolha de $g(\cdot)$ esteja em conformidade com as fontes de incerteza de localização das aplicações em análise. Na prática, existem métodos de localização capazes de fornecer estatísticas associadas às estimativas de localização, mas a análise destas estatísticas no cenário de propagação não é uma tarefa trivial e, de fato, somente a partir destas informações a ERB poderá compor um modelo estatístico para $\mathbf{p}_e(\mathbf{s})$. Neste trabalho, estas circunstâncias são incorporadas, considerando o conhecimento prévio das estatísticas dos erros de localização, com $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$ (utilizada nas principais referências de predições em comunicações sem fio [78-82]), uma vez que há um interesse específico no uso de $g(\cdot)$ no modelo de predição espacial.

3.1 Modelo de Incerteza de Localização do Ambiente de Rádio

A introdução dos erros de localização ao modelo de sistema do ambiente de rádio em (2.18) é realizada por meio do modelo de incerteza (3.1), resultando em

$$\begin{aligned} P_g(\mathbf{s}) &= \mu(\mathbf{s} + \mathbf{p}_e(\mathbf{s})) + \xi(\mathbf{s} + \mathbf{p}_e(\mathbf{s})) \\ &= P_{tx} - \underbrace{10\alpha \log d(\mathbf{s}_{tx}, \mathbf{s} + \mathbf{p}_e(\mathbf{s}))}_{\text{Perda por percurso média}} + \underbrace{\xi(\mathbf{s} + \mathbf{p}_e(\mathbf{s}))}_{\text{Sombreamento}}, \end{aligned} \quad (3.2)$$

em que $P_g(\mathbf{s})$ é o processo aleatório espacial corrompido pelos erros de localização, sendo $\mathbf{s}, \mathbf{r} \in D$. Nota-se em (3.2) que a associação espacial equivocada é incorporada ao modelo no momento em que tanto a perda por percurso média, como o sombreamento log-normal são afetados pelos erros de localização $\mathbf{p}_e(\mathbf{s}) \equiv \mathbf{r} - \mathbf{s}$. Ainda, é assumido que a observação $Z_g(\mathbf{s})$ consiste em um valor regionalizado, obtido a partir da remoção da tendência e sujeito aos efeitos dos erros de localização do sistema de comunicação sem fio.

Neste contexto, algumas dificuldades começam a surgir, envolvendo a caracterização do processo $P_g(\mathbf{s})$ e a influência dos erros de localização na estimação dos parâmetros do ambiente de rádio. De fato, em vez de observar o processo aleatório espacial livre de erros, $P(\mathbf{s}) = \mu(\mathbf{s}) + \xi(\mathbf{s})$, tem-se a observação de um processo aleatório com incertezas de localização, $P_g(\mathbf{s}) = \mu(\mathbf{s} + \mathbf{p}_e(\mathbf{s})) + \xi(\mathbf{s} + \mathbf{p}_e(\mathbf{s}))$, cujas medidas coletadas são observações com distribuição de probabilidade diferente do processo aleatório do tipo Gaussiano livre de erros de localização [78]. Isto significa que a distribuição de probabilidade Gaussiana das observações $Z(\mathbf{s})$ é afetada em razão do erro nas coordenadas espaciais. A equação (3.3) mostra esta alteração, descrevendo a função distribuição cumulativa das observações com erros de localização $Z_g(\mathbf{s})$, permitindo verificar que a caracterização estatística das medidas é afetada pela incerteza de localização [93],

$$\begin{aligned} \Pr(Z_g(\mathbf{s}) \leq z_g) &= \int \Pr(Z(\mathbf{s} + \mathbf{u}) \leq z_g) g(\mathbf{u}) d\mathbf{u} \\ &= \int \int_{-\infty}^{z_g} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{z^2}{2\sigma^2}\right\} g(\mathbf{u}) dz d\mathbf{u}. \end{aligned} \quad (3.3)$$

em que $\mathbf{u} \equiv \mathbf{p}_e(\mathbf{s}) = (u_x, u_y)$ é o vetor de erro de localização bidimensional, cuja distribuição de probabilidade $g(\mathbf{u})$ é a parcela responsável por afetar a distribuição de probabilidade Gaussiana das observações $Z(\mathbf{s})$, mediante a inclusão de $\mathbf{p}_e(\mathbf{s})$. Como resultado, a covariância espacial das observações com os efeitos dos erros de localização se torna diferente, comparada à covariância espacial das observações sem incerteza de localização. Esta diferença terá influência no tratamento da tendência e na aprendizagem de parâmetros da Fase I e, conseqüentemente, poderá ter impacto na Fase II, voltada para as predições espaciais. Uma vez incorporado o erro de localização ao modelo de sistema, o próximo objetivo consiste em investigar quais são os impactos sobre as etapas do modelo de predição espacial do REM.

3.2 Impacto da Incerteza de Localização

A influência do erro de localização $\mathbf{p}_e(\mathbf{s})$ pode ser inicialmente analisada nas etapas de estimação e remoção da tendência. Para simplificação da análise, assume-se que $P_{tx} = 0$ dBm e ressalta-se que somente o conjunto de coordenadas espaciais estimadas $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ é conhecido pela ERB. Com isso, o processo $P_g(\mathbf{s})$ pode ser descrito de forma analítica de acordo com

$$P_g(\mathbf{s}) = \underbrace{-10\alpha \log d(\mathbf{s}_{tx}, \mathbf{s} + \mathbf{p}_e(\mathbf{s}))}_{\text{Perda por percurso média}} + \underbrace{\xi(\mathbf{s} + \mathbf{p}_e(\mathbf{s}))}_{\text{Sombreamento}}. \quad (3.4)$$

A amostragem espacial, caracterizada pela captura de N medidas do ambiente de rádio, permite formular o modelo de sistema em (3.4) de forma matricial, *i.e.*, $\mathbf{p}_g = \mathbf{x}\alpha + \boldsymbol{\xi}$, sendo $\mathbf{p}_g$ um vetor coluna $N \times 1$ com as medidas de potência de recepção com erros de localização. Nota-se que tanto $\boldsymbol{\xi}$ como o vetor de funções base $\mathbf{x} = -10 \log d(\mathbf{s}_{tx}, \mathbf{s} + \mathbf{p}_e(\mathbf{s}))$ são afetados por $\mathbf{p}_e(\mathbf{s})$ e, portanto, poderão influenciar o resultado da estimação do coeficiente de propagação $\hat{\alpha} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{p}_g$. Ainda assim, a influência dos erros de localização sobre $\hat{\alpha}$ depende de como o vetor $\mathbf{p}_e(\mathbf{s})$ afeta o logaritmo das distâncias entre a ERB e as coordenadas espaciais estimadas $\mathbf{s}$. Tal como a etapa de estimação do coeficiente $\hat{\alpha}$, a remoção da tendência também é afetada por $\mathbf{p}_e(\mathbf{s})$, pois as estimativas da potência média de recepção são obtidas no conjunto de coordenadas estimadas $\mathbf{s}$, $\hat{\boldsymbol{\mu}} = [-10\hat{\alpha} \log d(\mathbf{s}_{tx}, \mathbf{s}_1), \dots, -10\hat{\alpha} \log d(\mathbf{s}_{tx}, \mathbf{s}_N)]^T$, e subtraídas das medidas de potência de recepção coletadas do ambiente de rádio, resultando em

$$\mathbf{z}_g = \mathbf{p}_g - \hat{\boldsymbol{\mu}}, \quad (3.5)$$

em que $\mathbf{z}_g = [Z_g(\mathbf{s}_1), \dots, Z_g(\mathbf{s}_N)]^T$ consiste no vetor de valores regionalizados que foi recuperado, mas corrompido pelos erros de localização. Por meio de $\mathbf{z}_g$, os efeitos de $\mathbf{p}_e(\mathbf{s})$ podem se propagar pelo modelo de predição espacial, pois $\hat{\boldsymbol{\mu}}$ foi obtido a partir das coordenadas $\mathbf{s}$ e não $\mathbf{r}$. Uma simulação foi conduzida para verificar os impactos do vetor $\mathbf{p}_e(\mathbf{s})$ nas etapas relacionadas com o tratamento da tendência, considerando a alteração do desvio padrão σ_p , com $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$. O processo aleatório espacial gerado com $\alpha = 3$, sombreamento log-normal com $\sigma = 5$ dB e parâmetros $\boldsymbol{\theta} = [m \ r \ c_0] = [25, 10, 0]$, será afetado pela incerteza de localização. Com $L = 5000$ realizações, o método de Monte Carlo foi usado para a estimação do erro quadrático médio (MSE - *Mean Square Error*) de α e a variância dos valores regionalizados,

$$\begin{aligned} \text{MSE}_\alpha &= \frac{1}{L} \sum_{l=1}^L (\alpha - \hat{\alpha}_l)^2 \\ \text{VAR}_{\text{VR}} &= \frac{1}{L} \sum_{l=1}^L \text{Var}\{\mathbf{z}_g^l\} = \frac{1}{L} \sum_{l=1}^L \frac{1}{N-1} \sum_{i=1}^N (\mathbf{z}_g^l(i) - \bar{\mathbf{z}}_g^l)^2, \text{ com } \bar{\mathbf{z}}_g = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_g(i). \end{aligned} \quad (3.6)$$

Figura 3.2: Influência dos erros de localização: (a) estimação da tendência e (b) remoção da tendência.

A Figura 3.2 traz os resultados e mostra os comportamentos: (a) do MSE_{α} e (b) da VAR_{VR} , ambos em função do desvio padrão σ_p , que expressa a severidade dos erros de localização adicionados ao modelo de sistema. É possível observar que o aumento de σ_p levou a uma queda de desempenho do estimador $\hat{\alpha}$, *i.e.*, aumento do MSE_{α} . Esta queda de desempenho se mostrou mais acentuada quando o desvio padrão σ_p foi configurado com valores da ordem da distância de decorrelação do sombreamento do canal sem fio (expressa pelo parâmetro $r = 10$ m). Ainda assim, o impacto dos erros de localização foi mais significativo na variância dos valores regionalizados (recuperados a partir da remoção da tendência). A variabilidade contida em $\mathbf{z}_g$ deveria refletir a variância do processo aleatório espacial gerado ($\sigma = 5$ e $\sigma^2 = 25$). No entanto, observa-se o aumento significativo da VAR_{VR} a medida em que σ_p é aumentado. Isto significa que a remoção da tendência afetada por $\mathbf{p}_e(\mathbf{s})$ adicionou um viés à cada um dos elementos que compõem o vetor $\mathbf{z}_g$ e, provavelmente, terá influência nas etapas que utilizam as medidas em $\mathbf{z}_g$ com algum propósito.

O método dos momentos de Matheron, reescrito em (3.7), ocorre na sequência e será afetado pela incerteza de localização em dois aspectos: i) Os elementos de $\mathbf{z}_g$ apresentam um viés e, portanto, as diferenças quadráticas calculadas a partir destes valores também serão afetadas e não representarão as semivariâncias do processo aleatório espacial e ii) As distâncias calculadas entre as medidas $Z_g(\mathbf{s}_i)$, afetadas pelos erros de localização, são baseadas no conjunto de coordenadas espaciais estimadas com erros de localização $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, o que também afeta os resultados dos estimadores em (3.7).

$$\begin{aligned} \hat{\gamma}(\mathbf{h}) &= \frac{1}{2N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} [Z_g(\mathbf{s}_j) - Z_g(\mathbf{s}_i)]^2, \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D, \quad i, j = 1, 2, \dots, N, \\ \hat{\gamma}(\text{Lag}_k) &= \frac{1}{2\|\mathbf{A}_{\text{Lag}_k}\|} \sum_{|\mathbf{s}_i - \mathbf{s}_j| \in \mathbf{A}_{\text{Lag}_k}} [Z_g(\mathbf{s}_i) - Z_g(\mathbf{s}_j)]^2. \end{aligned} \quad (3.7)$$

Desta forma, o semivariograma obtido será diferente quando comparado ao caso em que não há incerteza de localização. Para investigar esta diferença, a Figura 3.3 apresenta os resultados de simulação que mostram os efeitos dos erros de localização no semivariograma e no covariograma experimentais. Foram coletadas $N = 1000$ medidas de um processo aleatório espacial gerado ($\alpha = 3$, $\sigma = 5$ dB e $\theta = [25, 10, 0]$), que foram corrompidas por erros de localização $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$, com $\sigma_p = 7$ m. Os estimadores do semivariograma e do covariograma processaram as medidas provenientes da remoção da tendência baseadas nas coordenadas estimadas $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. É possível notar uma descontinuidade próxima à origem dos resultados de ambos os estimadores. Isso ocorre porque os efeitos dos erros de localização são mais significativos para as medidas que estão relativamente próximas, causando uma elevação e um achatamento no semivariograma (c) e no covariograma (d), respectivamente. No comparativo com o semivariograma (a) e o covariograma (b) sem os erros de localização, constata-se que a presença de $\mathbf{p}_e(\mathbf{s})$ resulta em outra estrutura para as funções semivariograma $\gamma_g(\mathbf{s}_i, \mathbf{s}_j)$ e covariância $C_g(\mathbf{s}_i, \mathbf{s}_j)$. De fato, as funções $\gamma_g(\mathbf{s}_i, \mathbf{s}_j)$ e $C_g(\mathbf{s}_i, \mathbf{s}_j)$ do processo aleatório espacial com incerteza de localização $P_g(\mathbf{s})$ se diferenciam das funções, $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ e $C(\mathbf{s}_i, \mathbf{s}_j)$, do processo aleatório livre de erros de localização $P(\mathbf{s})$. Estes resultados corroboram as investigações realizadas em [83], [91-93], nas quais cita-se que não há uma estrutura não trivial para $\mathbf{p}_e(\mathbf{s})$ que ainda mantenha as exatas características da covariância do processo aleatório observado.

Figura 3.3: Influência dos erros de localização no: semivariograma experimental (a) e (c) e covariograma experimental (b) e (c)

Sob o ponto de vista das previsões espaciais, a busca pelo preditor ótimo no sentido de minimização do erro de previsão espacial passa pelo conhecimento da covariância espacial afetada pela incerteza de localização. Como a estimação dos parâmetros de covariância do ambiente de rádio antecede a Fase II de previsões, o semivariograma experimental afetado pelos erros de localização terá impacto direto na estimação do vetor θ , utilizado pelos preditores espaciais. Para investigar este tópico, a Figura 3.4 apresenta os histogramas das estimativas $\hat{\theta}$, obtidas a partir da aplicação da técnica LS ao semivariograma obtido, considerando $m = 25$, $r = 10$, $c_0 = 0$, $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$ com $\sigma_p = 7$ m e $L = 10.000$ realizações.

É possível observar que a estimação LS resulta em diferentes valores para a variância *sill* estimada ($\hat{m}$) e para o *range* estimado ($\hat{r}$). Ainda, nota-se que os histogramas (a) e (b), cujas medidas não apresentam incerteza de localização, refletem os valores esperados para os parâmetros do ambiente de rádio, enquanto o mesmo não ocorre para os histogramas (c) e (d), que realizam a estimação sobre as medidas corrompidas por $\mathbf{p}_e(\mathbf{s})$. A remoção da tendência e a dificuldade de ajuste do modelo analítico do semivariograma às semivariâncias experimentais afetadas pelos erros de localização (especialmente em relação ao parâmetro de inclinação r) são os principais fatores que provocaram as diferenças nas estatísticas do vetor θ . Resultados similares também foram observados para diferentes configurações, incluindo o uso da técnica ML e diversos valores do vetor θ . Ressalta-se que as variações em micro-escala espacial não foram consideradas, uma vez que há o interesse específico nos efeitos em larga escala do ambiente de rádio.

Figura 3.4: Influência dos erros de localização na estimação dos parâmetros de covariância $\hat{\theta}$.

3.3 Tratamento da Incerteza de Localização

O impacto da incerteza de localização na aprendizagem dos parâmetros e suas consequências sobre a acurácia das predições é um tópico relevante na análise realizada até o momento. Neste contexto, embora o uso da teoria de processos Gaussianos em predições espaciais seja amplamente discutido, ainda não há um número expressivo de trabalhos que têm como alvo a pesquisa sobre predições em conjunto com o tratamento dos problemas de incerteza de localização em processos aleatórios espaciais [83-84]. Ainda assim, iniciativas recentes vêm crescendo neste tópico e alguns trabalhos podem ser destacados como: [91-93], que apresentam uma metodologia para incorporar e tratar a incerteza de localização em processos Gaussianos, baseados na abordagem geoestatística; [41], [83-84], que realizam uma análise mais investigativa sobre os efeitos dos erros de localização nas estatísticas de predição espacial (*e.g.*, distribuição *a posteriori* das predições e intervalos de confiança) e além destes, trabalhos que utilizam as predições espaciais em diferentes aplicações, incluindo: [78], que leva em conta a incerteza de localização no problema de predição de cobertura em sistemas de comunicações móveis; [79-80], que fornecem uma visão unificada sobre os preditores espaciais (clássico e Bayesiano), investigando a influência dos erros de localização nas predições em comunicações sem fio e [81-82] que abordam os impactos da incerteza de localização em redes de sensores. Sobre estes últimos, a questão de limitação de recursos de energia é crítica e por isso, [82] também investiga a complexidade dos métodos para tratar quantidades analiticamente devido aos efeitos provocados pela incerteza de localização (*e.g.*, covariâncias e distribuições *a posteriori*).

Um ponto em comum entre estes trabalhos está no uso da distribuição de probabilidade dos erros de localização $g(\cdot)$ para combater seus efeitos negativos nas aplicações analisadas. Por outro lado, a principal diferença está na forma como os erros de localização são incorporados ao modelo de sistema e como suas estatísticas são utilizadas, além de quais são os métodos de estimação e predição selecionados. Este trabalho apresenta uma proposta de preditor espacial, com o objetivo de aplicá-lo para a geração do REM, considerando a presença de erros de localização em um sistema de comunicação sem fio. Especificamente, novas técnicas de estimação, que expandem a metodologia geoestatística em [91-93], são desenvolvidas para lidar com a incerteza de localização na fase de aprendizagem de parâmetros do ambiente de rádio. Estas técnicas se fundamentam em um procedimento de ajuste, que é aplicado sobre as covariâncias entre as medidas capturadas. A ideia é fundamentada na diferença existente entre a covariância dos processos aleatórios espaciais sem e com incerteza de localização $P(\mathbf{s})$ e $P_g(\mathbf{s})$. Assim, a necessidade do tratamento dos problemas causados pelos erros de localização aumenta à medida em que a diferença entre as covariâncias de $P(\mathbf{s})$ e $P_g(\mathbf{s})$ também se torna significativa. As técnicas desenvolvidas são descritas adiante e permitem a formulação do preditor espacial proposto. Em um primeiro momento, os detalhes sobre a metodologia em [93] são mostrados (subseções 3.3.1 e 3.3.2). Em seguida, as características das técnicas de estimação desenvolvidas são apresentadas (seção 3.4).

3.3.1 Procedimentos de Ajuste do Primeiro e Segundo Momentos

O procedimento de ajuste em [93] permite obter o preditor espacial Kriging ajustado pelas estatísticas dos erros de localização (KALE - *Kriging Adjusted for Location Error*). Na formulação do preditor KALE, torna-se necessário obter o primeiro e o segundo momentos (*i.e.*, média e covariância) do processo aleatório espacial com incerteza de localização $P_g(\mathbf{s})$. De fato, o entendimento sobre a obtenção destes momentos é fundamental para a aplicação do procedimento de ajuste. Inicialmente, o ajuste em relação ao primeiro momento é discutido e está relacionado com a tendência do ambiente de rádio. Em seguida, o procedimento de ajuste relacionado com a covariância espacial do sombreamento log-normal é abordado.

A ilustração da Figura 3.5 auxilia na discussão do procedimento de ajuste do primeiro momento. Nota-se que o vetor de erro de localização $\mathbf{u} = (u_x, u_y)$ diferencia a coordenada espacial estimada $\mathbf{s}_i = (s_{xi}, s_{yi})$ da verdadeira posição do dispositivo, dada pela coordenada realizada $\mathbf{r}_i = (r_{xi}, r_{yi})$, afetando a distância entre a ERB e os dispositivos da rede. Consequentemente, a potência média de recepção em $\mathbf{s}_i$, denotada como $\mu(\mathbf{s}_i)$, será diferente do valor verdadeiro da potência média de recepção na posição verdadeira $\mathbf{r}_i$, expressa de acordo com

$$\begin{aligned}\mu(\mathbf{r}_i) &= P_{tx} - \alpha f(\mathbf{r}_i) \\ &= P_{tx} - 10\alpha \log(d(\mathbf{s}_{tx}, \mathbf{r}_i)), \\ &= P_{tx} - 10\alpha \log \sqrt{(s_{tx} - s_{xi} + u_x)^2 + (s_{ytx} - s_{yxi} + u_y)^2}.\end{aligned}\tag{3.8}$$

Como a ERB possui somente o conhecimento da coordenada espacial estimada $\mathbf{s}_i$ (e não da coordenada realizada $\mathbf{r}_i$), a diferença entre os valores de potência média de recepção $\mu(\mathbf{r}_i)$ e $\mu(\mathbf{s}_i)$ retrata o impacto do erro de localização $\mathbf{u}$ na tendência, sendo que $\mu(\mathbf{r}_i) = \mu(\mathbf{s}_i)$ somente se $\mathbf{u} = \mathbf{0}$. Ainda, nota-se que a influência do vetor $\mathbf{u}$ depende do quão significativa são as parcelas u_x e u_y no modelo da tendência em (3.8). Com isso, para encontrar o primeiro momento do processo aleatório espacial $P_g(\mathbf{s})$, deve-se levar em conta as estatísticas dos erros de localização. Isto é realizado através da integral do valor esperado do processo aleatório livre de incertezas $P(\mathbf{s})$ considerando a distribuição $g(\cdot)$ nas regiões onde ocorrem os erros de localização,

$$\begin{aligned}\mu_g(\mathbf{s}) &= \int \mu(\mathbf{s} + \mathbf{u})g(\mathbf{u})d\mathbf{u} \\ &= P_{tx} - \alpha \int f(\mathbf{s} + \mathbf{u})g(\mathbf{u})d\mathbf{u},\end{aligned}\tag{3.9}$$

em que $\mathbf{u} \equiv \mathbf{p}_e(\mathbf{s})$ e $\mu_g(\mathbf{s}) \equiv \mathbb{E}\{P_g(\mathbf{s})\}$ consiste na média do processo aleatório com incerteza de localização $P_g(\mathbf{s})$. Os detalhes da demonstração de (3.9) podem ser acessados em [91, pp. 56-57]. A integral das funções base $f(\mathbf{s} + \mathbf{u})$ é ponderada por α , uma vez que o modelo é linear neste parâmetro. Este é o caso da tendência do ambiente de rádio quando o modelo log-distância é utilizado.

Figura 3.5: Efeito dos erros de localização em relação à tendência do ambiente de rádio.

A obtenção do segundo momento do processo aleatório espacial $P_g(\mathbf{s})$ é mais complexa em função de duas razões: i) da forma do modelo de covariância, ou seja, como este depende da distância $\mathbf{h}$ entre as medidas e ii) diferentemente do caso da tendência, cada coordenada espacial que compõem um par de observações é afetada por um vetor de erro de localização distinto. A Figura 3.6, adaptada de [91], auxilia no entendimento da relação entre estas duas razões, mostrando como os vetores de erros de localização afetam: (a) as distâncias entre as coordenadas espaciais estimadas $\mathbf{s}$ e (b) as distâncias entre a coordenada alvo da predição espacial $\mathbf{s}_0$ e as coordenadas espaciais estimadas $\mathbf{s}$. Neste ponto, ressalta-se que ambas as relações (a) e (b) são essenciais para que as predições Kriging possam ser realizadas, levando em conta as estatísticas dos erros de localização.

Na Fig.3.6(a), verifica-se que os dois vetores de erros de localização, $\mathbf{u} = (u_x, u_y)$ e $\mathbf{v} = (v_x, v_y)$, afetam as coordenadas espaciais $\mathbf{s}_i$ e $\mathbf{s}_j$, tornando-as diferentes das coordenadas realizadas $\mathbf{r}_i$ e $\mathbf{r}_j$. Da relação entre vetores em duas dimensões, tem-se que a separação espacial entre as coordenadas realizadas $\mathbf{r}_i$ e $\mathbf{r}_j$ é dada por $\boldsymbol{\varepsilon}_1 = \mathbf{h} + \mathbf{v} - \mathbf{u}$ e depende de ambos os vetores $\mathbf{u}$ e $\mathbf{v}$. Para o caso da covariância espacial isotrópica, existe o interesse em $|\boldsymbol{\varepsilon}_1|$, descrito analiticamente como

$$|\boldsymbol{\varepsilon}_1| = \sqrt{((s_{xj} + v_x) - (s_{xi} + u_x))^2 + ((s_{yj} + v_y) - (s_{yi} + u_y))^2}. \quad (3.10)$$

Diferente do tratamento da tendência, o impacto dos erros de localização na covariância decorre da forma como ambos os vetores $\mathbf{u}$ e $\mathbf{v}$ afetam a distância obtida entre as medidas. De outro modo, a Fig.3.6(b) é voltada para o aspecto de predição, pois mostra como o erro de localização $\mathbf{u}$ afeta a separação espacial entre a coordenada espacial alvo (fixa) da predição $\mathbf{s}_0$ e a coordenada espacial estimada $\mathbf{s}_i$, resultando em $\boldsymbol{\varepsilon}_0 = \mathbf{h} + \mathbf{u}$. Neste caso, a distância ou módulo do vetor de separação espacial entre a coordenada realizada $\mathbf{r}_i$ e $\mathbf{s}_0$ é descrita de acordo com

$$|\boldsymbol{\varepsilon}_0| = \sqrt{(s_{xi} - s_{x0} + u_x)^2 + (s_{yi} - s_{y0} + u_y)^2}. \quad (3.11)$$

Figura 3.6: Efeito dos erros de localização em relação à covariância espacial do ambiente de rádio.

O segundo momento do processo aleatório $P_g(\mathbf{s})$ é expresso pela covariância espacial entre as diferentes coordenadas estimadas $\mathbf{s}$, denotada como $C_g(\mathbf{s}_i, \mathbf{s}_j)$, e entre a coordenada alvo e as coordenadas estimadas, descrita por $C_{g0}(\mathbf{s}_i, \mathbf{s}_0)$. De forma específica, $C_g(\mathbf{s}_i, \mathbf{s}_j)$ e $C_{g0}(\mathbf{s}_i, \mathbf{s}_0)$ são obtidas através da integral da função covariância espacial sem incertezas de localização $C(\mathbf{h})$ sobre a distribuição de probabilidade $g(\cdot)$ nas regiões onde existem erros de localização, de acordo com

$$\begin{aligned} C_g(\mathbf{s}_i, \mathbf{s}_j) &= C_g(\mathbf{h}) = \int \int C(\mathbf{h} + \mathbf{v} - \mathbf{u}) g(\mathbf{u}) g(\mathbf{v}) d\mathbf{u} d\mathbf{v}, \\ C_{g0}(\mathbf{s}_i, \mathbf{s}_0) &= C_{g0}(\mathbf{h}) = \int C(\mathbf{h} + \mathbf{u}) g(\mathbf{u}) d\mathbf{u}. \end{aligned} \quad (3.12)$$

Os detalhes da demonstração de (3.12) podem ser acessados em [91, pp. 58-59]. É importante notar que as formas de $C_g(\mathbf{s}_i, \mathbf{s}_j)$ e $C_{g0}(\mathbf{s}_i, \mathbf{s}_0)$ dependem da combinação entre a covariância espacial sem incerteza de localização $C(\mathbf{h})$ e as distribuições densidade de probabilidade $g(\mathbf{u})$ e $g(\mathbf{v})$. Em [91], uma investigação sobre o comportamento de ε_1 e ε_0 , considerando diferentes distribuições $g(\cdot)$, aponta que a influência dos vetores $\mathbf{u}$ e $\mathbf{v}$ é significativa, sobretudo para pequenas distâncias $\mathbf{h}$. Em outras palavras, os efeitos dos erros de localização são significativos quando as medidas estão mais próximas, comparadas aos casos em que estão muito distantes. De fato, para que a covariância espacial entre medidas suficientemente distantes seja afetada, é necessário que erros de localização significativos ocorram. Conforme mencionado, a probabilidade de ocorrência de erros de localização mais severos nos sistemas de comunicações sem fio depende do desempenho dos métodos utilizados para a estimação de localização dos dispositivos. Logo, é possível argumentar que as diferenças entre as covariâncias dos processos aleatórios $P(\mathbf{s})$ e $P_g(\mathbf{s})$ e a possível degradação de desempenho das predições, em termos de acurácia, são argumentos que justificam a busca por técnicas que possam combater os efeitos negativos da incerteza de localização nas predições espaciais.

3.3.2 Kriging Ajustado para os Erros de Localização - KALE

A ideia central do preditor KALE consiste em utilizar as covariâncias do segundo momento em (3.12) na realização de previsões espaciais com a técnica Kriging, levando em conta as estatísticas dos erros de localização $g(\cdot)$. No entanto, a solução das integrais em (3.12) depende da combinação da densidade $g(\cdot)$ com os modelos da covariância espacial, dificultando significativamente a solução analítica do problema [91-93]. Frente a estas circunstâncias, o método de integração de Monte Carlo é utilizado neste trabalho com o propósito de solucionar as integrais (3.12) de forma numérica, a partir da geração de vetores aleatórios de acordo com a densidade $g(\cdot)$. Especificamente, o método consiste em um ajuste aplicado ao cálculo de distâncias entre as coordenadas espaciais, possibilitando a obtenção da matriz e do vetor de covariâncias ajustadas, $\mathbf{C}_g$ e $\mathbf{c}_{g0}$, do preditor KALE.

É importante mencionar que os autores em [93] apontam que o ajuste realizado no preditor KU em [92] não está correto, mas que o modelo para o KO, utilizado neste trabalho, procede. Além disso, ressalta-se que a dificuldade relacionada com a estimação do semivariograma no KU aumenta com o regime de incerteza de localização e a incorporação da tendência não estacionária. É por esta razão que a formulação do preditor KALE contempla o ajuste do segundo momento (*i.e.*, covariâncias) e pode ser descrito a partir dos seguintes passos: primeiramente, são gerados M vetores aleatórios, de acordo com a densidade de probabilidade $g(\cdot)$, *i.e.*, $\mathbf{u}_k = (u_x, u_y) \sim g(\mathbf{u})$ e $\mathbf{v}_k = (v_x, v_y) \sim g(\mathbf{v})$, com $k = 1, \dots, M$, em que M consiste no limite utilizado do método da integração de Monte Carlo. Com isso, as coordenadas espaciais $\mathbf{s}$ e os vetores $\mathbf{u}$ e $\mathbf{v}$ são utilizados para que a matriz e o vetor de covariâncias ajustados, $\mathbf{C}_g$ e $\mathbf{c}_{g0}$, sejam obtidos com o método da integração de Monte Carlo. O procedimento de ajuste do método é colocado em (3.13) e descrito nos algoritmos 1 e 2, mostrados adiante na forma de pseudo-código com notação em MATLAB.

$$\begin{aligned} \mathbf{C}_g &= \frac{1}{M^2} \sum_{k=1}^M \sum_{l=1}^M C(\mathbf{s}_i + \mathbf{v}_k - \mathbf{u}_l, \mathbf{s}_j + \mathbf{v}_k - \mathbf{u}_l), \\ \mathbf{c}_{g0} &= \frac{1}{M} \sum_{k=1}^M C(\mathbf{s}_i + \mathbf{u}_k, \mathbf{s}_0 + \mathbf{u}_k), \text{ com } i, j = 1, \dots, N. \end{aligned} \quad (3.13)$$

Inspecionando os algoritmos 1 e 2, nota-se que o ajuste tem início no cômputo da matriz de distâncias ajustadas $\mathbf{M}_{Dg}$, utilizada no algoritmo 1, e no vetor de distâncias ajustadas $\mathbf{m}_{Dg0}$, utilizado no algoritmo 2. Cada elemento de $\mathbf{M}_{Dg}$ é obtido por meio do ajuste aplicado ao cálculo das distâncias entre as coordenadas espaciais $\mathbf{s}$, levando em conta os vetores $\mathbf{u}$ e $\mathbf{v}$. De forma similar, o cálculo de distância de cada elemento do vetor $\mathbf{m}_{Dg0}$ também é ajustado. Os elementos de $\mathbf{M}_{Dg}$ e $\mathbf{m}_{Dg0}$ são usados na formação da matriz e do vetor de covariâncias ajustadas, $\mathbf{C}_g$ e $\mathbf{c}_{g0}$, através da aplicação do modelo da covariância exponencial, $C(|\mathbf{h}|) = m \exp(-|\mathbf{h}|/r)$, com os parâmetros m e r . A média calculada sobre os valores de covariância acumulados em M realizações permite obter as covariâncias ajustadas em cada elemento de $\mathbf{C}_g$ e $\mathbf{c}_{g0}$, considerando a densidade $g(\cdot)$.

O sistema de equações do KALE é construído a partir das covariâncias ajustadas em (3.13), $\mathbf{C}_g = \boldsymbol{\lambda}_g \mathbf{c}_{g0}$, que consistem em uma solução aproximada (via Monte Carlo) para (3.12) e cuja solução é obtida por inversão matricial, dada por $\boldsymbol{\lambda}_g = \mathbf{C}_g^{-1} \mathbf{c}_{g0}$. A predição da potência de recepção com o KALE é dada pela estimação da tendência, $\hat{\mu}(\mathbf{s}_0)$, em conjunto com a predição espacial do sombreamento $Z_{\text{KALE}}(\mathbf{s}_0)$, obtida pela combinação linear entre os pesos obtidos com ajuste $\boldsymbol{\lambda}_g$ e o conjunto de medidas $\mathbf{z}_g$ afetadas pelos erros de localização,

$$\hat{P}_{\text{KALE}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + Z_{\text{KALE}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + \sum_{i=1}^N \lambda_{gi} Z_g(\mathbf{s}_i). \quad (3.14)$$

3.3.3 Kriging que Ignora os Erros de Localização - KILE

A diferença entre os preditores KALE e KO (subseção 2.3.5, eq. 2.39) consiste no uso dos pesos $\boldsymbol{\lambda}_g$, obtidos a partir da solução do sistema de equações com as covariâncias ajustadas em (3.13). Com isso, o uso das medidas $\mathbf{z}_g$ em um sistema com as covariâncias sem o ajuste, *i.e.*, obtidas diretamente a partir das coordenadas estimadas $\mathbf{s}$, define o preditor Kriging que ignora os erros de localização (KILE - *Kriging Ignoring Location Error*), cujas predições são obtidas de acordo com

$$\hat{P}_{\text{KILE}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + Z_{\text{KILE}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + \sum_{i=1}^N \lambda_i Z_g(\mathbf{s}_i). \quad (3.15)$$

A principal diferença entre as predições Z_{KILE} e Z_{KALE} está no uso das covariâncias sem e com os ajustes, possibilitando a obtenção dos pesos Kriging $\boldsymbol{\lambda}$ e $\boldsymbol{\lambda}_g$. Neste sentido, dois aspectos devem ser considerados ao levar em conta as estatísticas dos erros de localização nas predições espaciais: i) os impactos da incerteza de localização na aprendizagem de parâmetros e ii) a degradação de desempenho das predições, em termos de acurácia.

Sobre o ponto i), os trabalhos que lidam com a incerteza de localização nas predições espaciais apontam as dificuldades na estimação de $\boldsymbol{\theta}$ nestas circunstâncias. Este tópico vem sendo investigado na literatura por meio de diferentes estratégias como: [93], que assume a normalidade (Gaussiana) para as medidas e propõe uma abordagem *pseudo-likelihood* para a estimação de $\boldsymbol{\theta}$; [79], em que um ajuste é proposto na técnica ML na fase de aprendizagem através de distribuições analíticas conhecidas *a priori* para as covariâncias ajustadas; [78], em que se propõe levar em conta a distribuição *a posteriori* das observações usando modificações baseadas em uma aproximação estocástica do algoritmo de maximização da esperança, possibilitando a obtenção de preditores não lineares (*e.g.*, preditor da esperança condicional). Diggle e Fanshawe em [84] relatam que lidar com a incerteza de localização na estimação de $\boldsymbol{\theta}$ é uma iniciativa plausível para que melhores predições possam ser alcançadas. Entretanto, a dificuldade na solução das integrais envolvidas na estimação de parâmetros quando há erros de localização no modelo de sistema pode aumentar a complexidade do problema de forma significativa [83-84].

Algoritmo 1: Procedimento de ajuste para obtenção da matriz de covariância $\mathbf{C}_g$ **Entrada:**Matriz $(N \times 2)$ de coordenadas espaciais composta pelos vetores $\mathbf{s}_x$ e $\mathbf{s}_y \Rightarrow \mathbf{s} = (\mathbf{s}_x, \mathbf{s}_y)$ Parâmetros do modelo de covariância $\Rightarrow \boldsymbol{\theta} = [m \ r]$ Densidade de probabilidade dos erros de localização $\Rightarrow g(\cdot)$ **Saída:**Matriz de covariâncias ajustadas $(N \times N)$ entre as coordenadas espaciais estimadas $\Rightarrow \mathbf{C}_g$

```

1  início
2   $\mathbf{M}_{Dg} \Rightarrow$  matriz de distâncias ajustadas com inicialização de zeros e dimensões  $(N \times N)$ 
3   $\mathbf{C}_g \Rightarrow$  matriz de covariâncias ajustadas com inicialização de zeros e dimensões  $(N \times N)$ 
4   $M \Rightarrow$  limite do somatório para o método de integração de Monte Carlo
5  for  $i = 1 \dots N$ 
6      for  $j = 1 \dots N$ 
7          if  $(i > j)$ 
8               $\mathbf{u} = (u_x, u_y) \sim g(\mathbf{u}) \Rightarrow$  vetor de erro de localização  $(M \times 1)$ 
9               $\mathbf{v} = (v_x, v_y) \sim g(\mathbf{v}) \Rightarrow$  vetor de erro de localização  $(M \times 1)$ 
10             for  $k = 1 \dots M$ 
11                 par de coordenadas espaciais estimadas atual
12                  $s_{xi} = \mathbf{s}_x(i)$ ,  $s_{yi} = \mathbf{s}_y(i)$ ,  $s_{xj} = \mathbf{s}_x(j)$  e  $s_{yj} = \mathbf{s}_y(j)$ 
13                 procedimento de ajuste no cálculo da distância entre as coordenadas atuais
14                 e armazenamento do valor obtido em cada posição  $(i, j)$  da matriz  $\mathbf{M}_{Dg}$ 
15                  $\mathbf{M}_{Dg}(i, j) = \sqrt{(s_{xj} + u_x(k) - v_x(k) - s_{xi})^2 + (s_{yj} + u_y(k) - v_y(k) - s_{yi})^2}$ 
16                 acúmulo dos valores da covariância exponencial ajustados
17                 em cada posição  $(i, j)$  da matriz  $\mathbf{C}_g$ 
18                  $\mathbf{C}_g(i, j) = \mathbf{C}_g(i, j) + m \exp\left\{-\frac{\mathbf{M}_{Dg}(i, j)}{r}\right\}$ 
19             end
20             obtenção do valor final de covariância em cada posição  $(i, j)$  da matriz  $\mathbf{C}_g$ 
21              $\mathbf{C}_g(i, j) = \frac{1}{M} \mathbf{C}_g(i, j)$ 
22         end
23     end
24 end
25 fim

```

Sobre o ponto ii), um trabalho sobre previsões climáticas da camada de ozônio do planeta Terra com dados submetidos à erros de localização é realizada em [93]. As conclusões apontam que ignorar a incerteza de localização pode resultar em previsões cujo viés e o erro quadrático médio tendem a crescer de forma significativa com o aumento da variância dos erros de localização.

Algoritmo 2: Procedimento de ajuste para obtenção do vetor de covariância $\mathbf{c}_{g0}$ **Entrada:**Matriz $(N \times 2)$ de coordenadas espaciais composta pelos vetores $\mathbf{s}_x$ e $\mathbf{s}_y \implies \mathbf{s} = (\mathbf{s}_x, \mathbf{s}_y)$ Parâmetros do modelo de covariância $\implies \boldsymbol{\theta} = [m \ r]$ Densidade de probabilidade dos erros de localização $\implies g(\cdot)$ Coordenada alvo da predição espacial $\implies \mathbf{s}_0 = (s_{x0}, s_{y0})$ **Saída:**Vetor de covariâncias ajustadas $(N \times 1)$ entre $\mathbf{s}$ e $\mathbf{s}_0 \implies \mathbf{c}_{g0}$

```

1  início
2       $\mathbf{m}_{Dg0} \implies$  inicialização com zeros do vetor  $(N \times 1)$  de distâncias ajustadas entre  $\mathbf{s}$  e  $\mathbf{s}_0$ 
3       $\mathbf{c}_{g0} \implies$  inicialização com zeros do vetor  $(N \times 1)$  de covariâncias ajustadas entre  $\mathbf{s}$  e  $\mathbf{s}_0$ 
4       $M \implies$  limite do somatório para o método de integração de Monte Carlo
5      for  $i = 1 \dots N$ 
6          for  $k = 1 \dots M$ 
7               $\mathbf{u} = (u_x, u_y) \sim g(\mathbf{u}) \implies$  vetor de erro de localização  $(M \times 1)$ 
8               $s_{xi} = \mathbf{s}_x(i)$  e  $s_{yi} = \mathbf{s}_y(i) \implies$  coordenada espacial estimada atual
9              procedimento de ajuste no cálculo de distância entre as coordenadas espaciais  $\mathbf{s}$  e  $\mathbf{s}_0$ 
10             e armazenamento do valor obtido em cada posição  $(i)$  do vetor  $\mathbf{m}_{Dg0}$ 
11              $\mathbf{m}_{Dg0}(i) = \sqrt{(s_{xi} + u_x(k) - s_{x0})^2 + (s_{yi} + u_y(k) - s_{y0})^2}$ 
12             acúmulo dos valores da covariância exponencial ajustados
13             em cada posição  $(i)$  do vetor  $\mathbf{c}_{g0}$ 
14              $\mathbf{c}_{g0}(i) = \mathbf{c}_{g0}(i) + m \exp\left\{-\frac{\mathbf{m}_{Dg0}(i)}{r}\right\}$ 
15         end
16         obtenção do valor final de covariância em cada posição  $(i)$  do vetor  $\mathbf{c}_{g0}$ 
17          $\mathbf{c}_{g0}(i) = \frac{1}{M} \mathbf{c}_{g0}(i)$ 
18     end
19 fim

```

Na mesma linha, um estudo de predições do clima conduzido em [83] apontou que a superioridade de desempenho a favor do preditor KALE em relação ao KILE está condicionada ao conhecimento dos parâmetros do vetor $\boldsymbol{\theta}$, ou seja, para que os ajustes das covariâncias resultem em melhores resultados para o KALE, é necessário que $\boldsymbol{\theta}$ seja conhecido. Isto significa que a estimação de $\boldsymbol{\theta}$ desempenha um papel importante nos resultados de predição e, de fato, a aprendizagem de parâmetros sob os efeitos dos erros de localização consiste em um problema realístico no contexto de predições espaciais do REM. Diante destas dificuldades, este trabalho propõe uma técnica de predição espacial que expande a metodologia em [93] para o caso do REM, visando obter predições com maior acurácia, considerando aprendizagem de parâmetros sob os efeitos dos erros de localização.

3.4 Preditor Espacial TS-KALE

A Figura 3.7 mostra o diagrama com o funcionamento da técnica de predição espacial proposta. A ideia central consiste no uso da densidade de probabilidade dos erros de localização $g(\cdot)$ em mais etapas do modelo de predição espacial do REM. Especificamente, duas técnicas são propostas em etapas que antecedem a estimação do vetor θ , contemplando o tratamento da tendência (TALE - *Trend Adjusted for Location Error*) e a estimação do semivariograma (SALE - *Semivariogram Adjusted for Location Error*), ambas levando em conta a densidade de probabilidade $g(\cdot)$. Como resultado, tem-se o preditor Kriging com a tendência e o semivariograma ajustados (TS-KALE - *Trend, Semivariogram and Kriging Adjusted for Location error*). O objetivo consiste em evitar o problema da alta complexidade envolvida no uso de $g(\cdot)$ nos processos de otimização da estimação do vetor θ , retratando uma solução alternativa ao problema dos erros de localização na fase de aprendizagem de parâmetros. Assim, o resultado de predição espacial, dado por $\hat{P}_{\text{TS-KALE}}(\mathbf{s}_0)$, é obtido a partir do vetor de parâmetros estimado $\hat{\theta}$ com as técnicas de ajuste propostas, TALE e SALE, que são descritas adiante (subseções 3.4.1 e 3.4.2).

Desta forma, o propósito específico do preditor TS-KALE consiste em fornecer resultados experimentais do semivariograma que favoreçam o seu ajuste (*i.e.*, estimação dos parâmetros) através do uso de $g(\cdot)$ nas etapas da tendência e do semivariograma do modelo de predição espacial. Diferentes testes foram realizados em cada etapa do modelo de predição a fim de verificar se as técnicas propostas trouxeram melhores resultados para a aprendizagem de parâmetros. Em um primeiro momento, os detalhes relacionados com a técnica de ajuste em relação à tendência (TALE) são mostrados e, em seguida, o ajuste na estimação do semivariograma experimental (SALE) é abordado. Finalmente, a formulação matemática do preditor espacial TS-KALE é apresentada.

Figura 3.7: Diagrama de funcionamento do preditor espacial TS-KALE.

3.4.1 Estimação Ajustada da Tendência do Ambiente de Rádio - TALE

A estimação ajustada da tendência (TALE) é baseada em um procedimento aplicado sobre o cálculo de distâncias realizado na estimação do coeficiente de propagação α e na remoção da tendência das medidas. O método de integração de Monte Carlo também é utilizado através de M realizações do algoritmo proposto, denotado como TALE. O resultado das M iterações permite obter uma estimativa do coeficiente de propagação ajustado, $\hat{\alpha}_{\text{adj}}$, que é utilizada para obtenção do conjunto de valores da potência média de recepção ajustados, $\mathbf{p}_{\text{adj}}$. Assim, na remoção da tendência, o vetor $\mathbf{p}_{\text{adj}}$, é subtraído do vetor de medidas de potência de recepção com erros de localização $\mathbf{p}_g$, permitindo obter vetor de valores regionalizados ajustados $\mathbf{z}_{\text{adj}}$. Desta forma, a ideia central do algoritmo TALE consiste em utilizar a densidade $g(\cdot)$ para obter o parâmetro, $\hat{\alpha}_{\text{adj}}$, e o conjunto de valores regionalizados, $\mathbf{z}_{\text{adj}}$, que sejam mais próximos dos verdadeiros valores, α e $\mathbf{z}$, mediante a presença de incerteza de localização nas medidas.

O algoritmo 3 apresenta os detalhes do ajuste da técnica TALE na forma de pseudo-código com notação em MATLAB. O ajuste tem início com o uso das matrizes $\mathbf{U}_x$ e $\mathbf{U}_y$, com dimensões $N \times M$, que são compostas por vetores de erros de localização gerados de acordo com $g(\cdot)$. Tais vetores são incorporados ao cálculo de distância entre as N coordenadas espaciais $\mathbf{s}$ e a localização fixa da ERB, resultando no vetor de distâncias ajustado, $\mathbf{d}_{\text{adjTx}}$, com dimensão $N \times 1$. Assim, a aplicação do modelo da tendência (log-distância) em cada elemento de $\mathbf{d}_{\text{adjTx}}$ permite obter o vetor de funções base ajustadas $\mathbf{x}_g$, necessário para que a estimação do coeficiente de propagação α seja feita, levando em conta as estatísticas da incerteza de localização.

A cada M -ésima iteração do algoritmo o coeficiente $\hat{\alpha}$ e os N valores de potência média de recepção com ajuste (obtidos a partir de $\mathbf{x}_g$ e $\hat{\alpha}$) são armazenados no vetor $\mathbf{a}$ e nas M colunas da matriz $\mathbf{P}_M$, respectivamente. Ao final do algoritmo, dois cálculos são realizados: i) a média dos M valores de $\hat{\alpha}$ armazenados no vetor $\mathbf{a}$ para obter o coeficiente de propagação ajustado $\hat{\alpha}_{\text{adj}}$ e ii) a média sobre os M valores de potência média de recepção da matriz $\mathbf{P}_M$, resultando em um vetor $(N \times 1)$ de potências médias de recepção ajustadas $\mathbf{p}_{\text{adj}}$. Finalmente, $\mathbf{p}_{\text{adj}}$ é utilizado para a remoção da tendência, ou seja, é subtraído das potências de recepção coletadas com erros de localização $\mathbf{p}_g$ para a obtenção do vetor de valores regionalizados ajustado, $\mathbf{z}_{\text{adj}}$. Logo, a saída do algoritmo TALE é dada pelo coeficiente $\hat{\alpha}_{\text{adj}}$ e o vetor $\mathbf{z}_{\text{adj}}$. Com estes resultados, é possível estimar a potência média de recepção de forma ajustada em diferentes coordenadas espaciais $\mathbf{s}_0$, de acordo com

$$\hat{\mu}_{\text{TALE}}(\mathbf{s}_0) = P_{\text{tx}} - 10\hat{\alpha}_{\text{adj}} \log(d(\mathbf{s}_{\text{tx}}, \mathbf{s}_0)). \quad (3.16)$$

Sobre as predições, o vetor $\mathbf{z}_{\text{adj}}$ obtido com a remoção da tendência das N medidas é utilizado nas etapas seguintes do modelo de predição do REM, começando com a estimação do semivariograma experimental. Antes, uma análise de desempenho dos resultados do TALE é realizada com o objetivo de verificar se o algoritmo trouxe benefícios no tratamento da tendência com incerteza de localização.

Algoritmo 3: TALE - Estimação ajustada da tendência do preditor TS-KALE**Entrada:**Matriz $(N \times 2)$ de coordenadas espaciais composta pelos vetores $\mathbf{s}_x$ e $\mathbf{s}_y \Rightarrow \mathbf{s} = (\mathbf{s}_x, \mathbf{s}_y)$ Vetor de localização fixa da ERB $\Rightarrow \mathbf{s}_{tx} = (s_{tx}, s_{ytx})$ Vetor $(N \times 1)$ de potência de recepção com erros de localização $\Rightarrow \mathbf{p}_g = [P_g(\mathbf{s}_1) \dots P_g(\mathbf{s}_N)]^T$ **Saída:**Coeficiente de propagação estimado com ajustes $\Rightarrow \hat{\alpha}_{adj}$ Vetor $(N \times 1)$ com os valores regionalizados obtidos com ajuste $\Rightarrow \mathbf{z}_{adj}$ **1 início****2** $\mathbf{U}_x$ e $\mathbf{U}_y \Rightarrow$ matrizes $(N \times M)$ com vetores de erros de localização gerados com $g(\cdot)$ **3** $u_x(i, k)$ e $u_y(i, k) \Rightarrow$ elementos da k -ésima coluna da i -ésima linha de $\mathbf{U}_x$ e $\mathbf{U}_y$ **4** $\mathbf{P}_M \Rightarrow$ matriz $(N \times M)$ para armazenamento dos valores de potência média de recepção**5** $\mathbf{p}_{adj} \Rightarrow$ vetor $(N \times 1)$ com os valores finais de potência média de recepção ajustada**6** $\mathbf{a} \Rightarrow$ vetor $(M \times 1)$ para armazenamento das estimativas $\hat{\alpha}$ **7** $M \Rightarrow$ limite do somatório para aplicação do método de Monte Carlo**8 for** $k = 1 \dots M$ **9 for** $i = 1 \dots N$ **10** coordenadas estimadas atuais $\Rightarrow s_{xi} = \mathbf{s}_x(i)$ e $s_{yi} = \mathbf{s}_y(i)$ **11** obtenção do vetor de distâncias ajustadas $(N \times 1)$ entre as coordenadas $\mathbf{s}$ e $\mathbf{s}_{tx}$ **12** $\mathbf{d}_{adjTx}(i) = \sqrt{(s_{xi} + u_x(i, k) - s_{tx})^2 + (s_{yi} + u_y(i, k) - s_{ytx})^2}$ **13 end****14** obtenção do vetor $(N \times 1)$ de funções base ajustadas por meio da aplicação**15** do modelo log-distância nos elementos do vetor $\mathbf{d}_{adjTx}$ **16** $\mathbf{x}_g = -10 \log(\mathbf{d}_{adjTx})$ **17** estimação do coeficiente de propagação a partir dos vetores $\mathbf{x}_g$ e $\mathbf{p}_g$ **18** $\hat{\alpha} = (\mathbf{x}_g^T \mathbf{x}_g)^{-1} \mathbf{x}_g^T \mathbf{p}_g$ **19** armazenamento do coeficiente de propagação estimado no k -ésimo elemento de $\mathbf{a}$ **20** $\mathbf{a}(k) = \hat{\alpha}$ **21** armazenamento dos valores de potência média de recepção ajustados**22** na k -ésima coluna da matriz $\mathbf{P}_M$ **23** $\mathbf{P}_M(k) = \hat{\alpha} \mathbf{x}_g = -10 \hat{\alpha} \log(\mathbf{d}_{adjTx})$ **24 end****25** obtenção do coeficiente de propagação ajustado**26** $\hat{\alpha}_{adj} = \frac{1}{M} \sum_{k=1}^M \mathbf{a}(k)$ (média dos M valores armazenados no vetor $\mathbf{a}$)**27** média calculada na i -ésima linha da matriz $\mathbf{P}_M$ considerando M colunas**28** $\mathbf{p}_{adj}(i) = \frac{1}{M} \sum_{k=1}^M \mathbf{P}_M(i, k)$, com $i = 1 \dots N$ **29** obtenção dos valores regionalizados com ajuste**30** $\mathbf{z}_{adj} = \mathbf{p}_g - \mathbf{p}_{adj}$ **31 fim**

Três cenários foram simulados, considerando diferentes valores para o coeficiente de propagação ($\alpha = 2, 3$ e 4), diferentes parametrizações para o vetor θ e áreas de cobertura com $500 \text{ m} \times 500 \text{ m}$, $1000 \text{ m} \times 1000 \text{ m}$ e até $3000 \text{ m} \times 3000 \text{ m}$. Os erros de localização foram introduzidos e sua severidade foi sendo modificada, contemplando situações de ausência de incerteza ($\sigma_p = 0$) até valores mais expressivos para o desvio padrão dos erros de localização ($\sigma_p = 200 \text{ m}$).

Os resultados, mostrados na Figura 3.8, são colocados em termos do MSE_α , relacionado com a estimação da tendência (a-b-c), e da dispersão dos valores regionalizados VAR_{VR} , após a remoção da tendência das medidas (d-e-f). Nota-se que a técnica TALE apresentou melhores resultados do que a estimação sem ajustes à medida em que os valores de σ_p foram aumentados. Este comportamento ocorreu tanto na estimação do coeficiente de propagação (menor MSE_α) como na redução da dispersão dos valores regionalizados recuperados (menor VAR_{VR}). Isto indica que fazer o ajuste levando em conta as estatísticas dos erros de localização nesta etapa é plausível, especialmente nas situações em que a incerteza de localização é mais severa. Além disso, a diferença de desempenho a favor do TALE também é observada quando σ_p é da ordem da distância de decorrelação (*i.e.*, parâmetro r) do sombreamento do canal sem fio. Isto é explicado pelo fato de que, quando σ_p é elevado (da ordem de r), as medidas capturadas se diferenciam dos valores verdadeiros de potência de recepção de forma bastante significativa, afetando a estimação de α (*i.e.*, regressão do modelo log-distância). De outro modo, para baixos valores de σ_p a técnica TALE não apresentou vantagem significativa. Em suma, as investigações de simulação apontam resultados promissores para o TALE, uma vez que, na média, um melhor desempenho no tratamento da tendência foi obtido.

Figura 3.8: Desempenho da estimação ajustada da tendência: MSE_α (a)-(b)-(c) e dispersão da variável regionalizada (d)-(e)-(f).

3.4.2 Estimação Ajustada do Semivariograma Experimental - SALE

Esta subseção apresenta os detalhes da estimação ajustada do semivariograma experimental, técnica denotada como SALE. De forma similar ao TALE, o procedimento de ajuste realizado no SALE também leva em conta as estatísticas dos erros de localização. O objetivo do SALE consiste em obter semivariâncias $\hat{\gamma}_{\text{adj}}$ e distâncias lags $\mathbf{d}_{\text{Lag-adj}}$ ajustadas que resultem em um desempenho melhor da posterior aprendizagem de parâmetros do vetor $\boldsymbol{\theta}$. Para isso, um procedimento de ajuste é incorporado ao cálculo de distâncias realizado antes da aplicação do estimador do método dos momentos (MoM) do semivariograma experimental. Assim, somente após o ajuste, o algoritmo de estimação do semivariograma experimental, descrito na subseção 2.3.2, é aplicado para obtenção das semivariâncias $\hat{\gamma}_{\text{adj}}$ e das distâncias lags $\mathbf{d}_{\text{Lag-adj}}$.

O conjunto de coordenadas espaciais estimadas $\mathbf{s}$, o vetor de valores regionalizados $\mathbf{z}$ (que também pode ser proveniente do TALE, *i.e.*, $\mathbf{z}_{\text{adj}}$) e a densidade de probabilidade $g(\cdot)$ são as variáveis de entrada do SALE. Os detalhes dos procedimentos da técnica são mostrados no algoritmo 4 na forma de pseudo-código com notação em MATLAB. O método de integração de Monte Carlo é utilizado, contemplando dois laços de iteração (*i.e.*, interno e externo), através dos limites M_1 e M_2 . O limite M_1 (externo) está relacionado com a estimação dos vetores de semivariâncias ($\hat{\gamma}_{\text{adj}}$) e distâncias lags ($\mathbf{d}_{\text{Lag-adj}}$) ajustados, enquanto o limite M_2 (interno) é usado para o cálculo dos elementos da matriz de distâncias ajustada ($\mathbf{D}_{\text{adj}}$), a partir das coordenadas espaciais estimadas $\mathbf{s}$.

O algoritmo se inicia com o cálculo de cada elemento da matriz $\mathbf{D}_{\text{adj}}(i, j)$, dado pela média dos M_2 valores escalares de distâncias, d_{adj} , que são calculados com o método de integração de Monte Carlo, considerando a introdução dos vetores de erros de localização gerados, $\mathbf{u}$ e $\mathbf{v}$. Em seguida, o processo de quantização do semivariograma experimental, descrito na subseção 2.3.2, é aplicado sobre $\mathbf{D}_{\text{adj}}$, resultando na matriz de distâncias lags com ajustes $\mathbf{D}_{\text{Lag-adj}}$, que são utilizadas para a estimação das semivariâncias via MoM. O resultado consiste nos k -ésimos vetores de semivariâncias e distâncias lags ajustadas, $\hat{\gamma}_{\text{adj}}^k$ e $\mathbf{d}_{\text{Lag-adj}}^k$, que são armazenados a cada iteração do algoritmo (M_1). Por fim, cada elemento dos vetores finais de semivariâncias e distâncias lags ajustadas, $\hat{\gamma}_{\text{adj}}(i)$ e $\mathbf{d}_{\text{Lag-adj}}(i)$, é obtido com o cálculo da média das componentes dos k vetores, $\hat{\gamma}_{\text{adj}}^k(i)$ e $\mathbf{d}_{\text{Lag-adj}}^k(i)$, armazenados nas iterações do algoritmo, de acordo com

$$\hat{\gamma}_{\text{adj}}(i) = \frac{1}{M_1} \sum_{k=1}^{M_1} \hat{\gamma}_{\text{adj}}^k(i) \quad \text{e} \quad \mathbf{d}_{\text{Lag-adj}}(i) = \frac{1}{M_1} \sum_{k=1}^{M_1} \mathbf{d}_{\text{Lag-adj}}^k(i) \quad (3.17)$$

A quantização realizada sobre $\mathbf{D}_{\text{adj}}$ ocorre de forma diferente a cada iteração do algoritmo (M_1) em razão da incorporação dos vetores aleatórios de erros de localização $\mathbf{u}$ e $\mathbf{v}$. O resultado disso, embora com pouca frequência, foi a obtenção dos vetores $\hat{\gamma}_{\text{adj}}^k$ e $\mathbf{d}_{\text{Lag-adj}}^k$ com comprimentos diferentes ao longo das k -ésimas iterações. Desta forma, optou-se por compor os vetores finais $\hat{\gamma}_{\text{adj}}$ e $\mathbf{d}_{\text{Lag-adj}}$ com os menores comprimentos verificados nos k vetores armazenados, para que o valor médio de cada posição dos vetores $\hat{\gamma}_{\text{adj}}$ e $\mathbf{d}_{\text{Lag-adj}}$ pudesse ser calculado, considerando as k iterações do algoritmo.

Algoritmo 4: SALE - Estimação ajustada do semivariograma do preditor TS-KALE**Entrada:**Matriz $(N \times 2)$ de coordenadas espaciais composta pelos vetores $\mathbf{s}_x$ e $\mathbf{s}_y \implies \mathbf{s} = (\mathbf{s}_x, \mathbf{s}_y)$ Vetor $(N \times 1)$ de valores regionalizados $\implies \mathbf{z}$ Densidade de probabilidade dos erros de localização $\implies g(\cdot)$ **Saída:**Vetor de semivariâncias experimentais ajustadas $\implies \hat{\gamma}_{\text{adj}}$ Vetor de distâncias lags ajustadas $\implies \mathbf{d}_{\text{Lag-adj}}$ **1 início**

```

2    $\mathbf{D}_{\text{adj}} \implies$  inicialização com zeros da matriz  $(N \times N)$  de distâncias ajustadas
3    $M_1$  e  $M_2 \implies$  limites dos somatórios (interno e externo) da integração de Monte Carlo
4   for  $k = 1 \dots M_1$ 
5       for  $i = 1 \dots N$ 
6           for  $j = 1 \dots N$ 
7                $\mathbf{u} = (u_x, u_y) \sim g(\mathbf{u}) \implies$  vetor  $(M_2 \times 1)$  de erros de localização
8                $\mathbf{v} = (v_x, v_y) \sim g(\mathbf{v}) \implies$  vetor  $(M_2 \times 1)$  de erros de localização
9               for  $t = 1 \dots M_2$ 
10                   $s_{xi} = \mathbf{s}_x(i)$ ,  $s_{yi} = \mathbf{s}_y(i)$ ,  $s_{xj} = \mathbf{s}_x(j)$  e  $s_{yj} = \mathbf{s}_y(j)$ 
11                  procedimento de ajuste no cálculo da distância entre as coordenadas atuais
12                  com os vetores de erros de localização  $\mathbf{u}$  e  $\mathbf{v}$ 
13                   $d_{\text{adj}} = \sqrt{(s_{xj} + u_x(t) - v_x(t) - s_{xi})^2 + (s_{yj} + u_y(t) - v_y(t) - s_{yi})^2}$ 
14                  acúmulo de valores de distância calculados ( $d_{\text{adj}}$ )
15                  em cada posição  $(i, j)$  da matriz  $\mathbf{D}_{\text{adj}}$ 
16                   $\mathbf{D}_{\text{adj}}(i, j) = \mathbf{D}_{\text{adj}}(i, j) + d_{\text{adj}}$ 
17               end
18               obtenção do valor final de distância em cada posição  $(i, j)$  da matriz  $\mathbf{D}_{\text{adj}}$ 
19                $\mathbf{D}_{\text{adj}}(i, j) = \frac{1}{M_2} \sum_{k=1}^{M_2} \mathbf{D}_{\text{adj}}(i, j)$ 
20           end
21       end
22        $\triangleright$  Estimação do semivariograma via quantização e MoM (subseção 2.3.2)
23        $\triangleright \mathbf{D}_{\text{Lag-adj}} \implies$  matriz de distâncias lags quantizadas a partir de  $\mathbf{D}_{\text{adj}}$ 
24        $\triangleright \hat{\gamma}_{\text{adj}}^k \implies k$ -ésimo vetor de semivariâncias ajustadas obtido a partir de  $\mathbf{D}_{\text{Lag-adj}}$ 
25        $\triangleright \mathbf{d}_{\text{Lag-adj}}^k \implies k$ -ésimo vetor de distâncias lag ajustadas obtido a partir de  $\mathbf{D}_{\text{Lag-adj}}$ 
26   end
27   vetor de semivariâncias com ajuste  $\implies \hat{\gamma}_{\text{adj}}(i) = \frac{1}{M_1} \sum_{k=1}^{M_1} \hat{\gamma}_{\text{adj}}^k(i)$ 
28   distâncias lag com ajuste  $\implies \mathbf{d}_{\text{Lag-adj}}(i) = \frac{1}{M_1} \sum_{k=1}^{M_1} \mathbf{d}_{\text{Lag-adj}}^k(i)$ 
29 fim

```

Uma simulação é realizada para verificar os resultados de desempenho atingidos com a técnica SALE. Um processo aleatório espacial é gerado com $\alpha = 3$ e com o vetor de parâmetros de covariância $\theta = [25, 10, 0]$ em uma área de cobertura de $100 \text{ m} \times 100 \text{ m}$. As coordenadas espaciais estimadas $\mathbf{s}$ são corrompidas por erros de localização $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$, com $\sigma_p = 17 \text{ m}$. Para que somente os resultados obtidos com a técnica SALE sejam analisados, a estimação e a remoção da tendência são realizadas de forma ideal, ou seja, os valores regionalizados retratam exatamente as variações do sombreamento log-normal. Os resultados podem ser observados na Figura 3.9, considerando o semivariograma experimental: (a) com a ausência de erros de localização no modelo de sistema; (b) com erros de localização ($\sigma_p = 17 \text{ m}$) e (c) com a aplicação do SALE.

Conforme discutido na seção 3.2, é possível observar em (b) que os erros de localização afetaram os resultados, especialmente os valores das semivariâncias relacionados com as distâncias lags menores (descontinuidade na origem), afetando o seu formato. Isto terá impacto direto na posterior estimação de θ . De outro modo, os resultados em (c) mostram uma suavização no semivariograma obtido com a técnica SALE. Esta suavização é explicada pela combinação de dois fatores: primeiramente, pelo cômputo dos resultados experimentais (semivariâncias e distâncias lags), que é feito por meio da média calculada sobre as M_1 iterações do algoritmo. Segundo, em razão da introdução dos vetores de erros de localização $\mathbf{u}$ e $\mathbf{v}$, que afeta especificamente as distâncias mínima e máxima, definidas na quantização do semivariograma experimental (descrito em 2.3.2). Isto influencia a quantidade de distâncias lags obtidas. De forma específica, observou-se que, no SALE, uma elevação dos valores de distância mínima ocorreu ao aplicar a quantização sobre as distâncias ajustadas pelos erros de localização, reduzindo o número de agrupamentos (*i.e.*, lags) formados a partir desta distância mínima. Este comportamento resultou na suavização dos pontos do semivariograma que, embora diferente do resultado em (a), foi observado para diversas configurações do ambiente de rádio.

Figura 3.9: Estimação do Semivariograma Experimental: (a) sem erros de localização; (b) com erros de localização e (c) com o ajuste SALE.

Baseado nestes resultados envolvendo a suavização alcançada pelo SALE, um conjunto de simulações foi realizado com o objetivo de investigar se a técnica proporciona melhores resultados para a posterior aprendizagem de parâmetros, ou seja, se o SALE traz benefícios para a estimação de θ . Com isso, diferentes configurações do sistema de comunicação sem fio foram simuladas (*e.g.*, diferentes tamanhos de área de cobertura, valores para os parâmetros de covariância e regimes de incerteza de localização). Dentre os cenários investigados, três configurações distintas para o ambiente de rádio foram selecionadas (A - B - C) e suas informações são sintetizadas na Tabela 3.1. Os resultados de simulação de cada cenário consistem em histogramas das estatísticas coletadas $\hat{\theta}$, obtidas com $L = 5000$ realizações dos processos aleatórios espaciais, cujas medidas são corrompidas pela incerteza de localização com densidade de probabilidade Gaussiana.

Como há o interesse na modelagem dos efeitos de propagação em larga escala (*i.e.*, perda por percurso média e o sombreamento do canal sem fio), o parâmetro *nugget*, que descreve as variações de micro-escala, é configurado com $c_0 = 0$ e, com isso, as estatísticas de $\hat{\theta}$ são separadas em histogramas relacionados com a estimação da variância *sill*, dada por $\hat{m}$, e do parâmetro *range*, dado por $\hat{r}$. Ressalta-se que $\hat{m}$ e $\hat{r}$ são estimativas diretas da variância e da distância de decorrelação espacial do sombreamento do canal sem fio (*i.e.*, $\hat{m} = \hat{\sigma}^2$ e $\hat{r} = \hat{d}_{\text{corr}}$).

Os histogramas dos cenários A, B e C são mostrados nas Figs. 3.10, 3.11 e 3.12, respectivamente. Os resultados são apresentados para cada cenário, considerando a ausência e a presença dos erros de localização, bem como o resultado alcançado com a técnica SALE, sendo que os histogramas (a-b-c) estão relacionados com as estatísticas de $\hat{m}$, enquanto os histogramas (d-e-f) estão relacionados com os resultados de $\hat{r}$. Conforme mencionado, assume-se que a estimação e a remoção da tendência são realizadas de forma ideal para que o desempenho da técnica SALE possa ser analisado.

É possível observar que, para todos os cenários, as estatísticas do parâmetro $\hat{r}$, relacionado com a inclinação do semivariograma, foram afetadas de forma bastante significativa, comparadas com as estatísticas de $\hat{m}$. Isto ocorre em razão da dificuldade em ajustar o modelo analítico da covariância (e sua inclinação através de r) às semivariâncias experimentais afetadas pelos erros de localização, especialmente em função da descontinuidade na origem do semivariograma. Neste sentido, nota-se que a suavização proporcionada pelo SALE gerou, na média, estatísticas que se aproximaram mais dos resultados obtidos sem os erros de localização. Isto significa que as semivariâncias do SALE facilitaram o ajuste dos modelos de covariância aos dados experimentais.

Tabela 3.1: Configurações de Simulação dos Cenários A - B - C

Parâmetros e Configurações	Cenário A	Cenário B	Cenário C
Dimensões Espaciais de Cobertura	100 m $\times$ 100 m	500 m $\times$ 500 m	2000 m $\times$ 2000 m
Parâmetros de Covariância	$\theta = [25 \ 10 \ 0]$	$\theta = [25 \ 50 \ 0]$	$\theta = [35 \ 200 \ 0]$
Número de Medidas Coletadas	$N = 100$	$N = 300$	$N = 500$
Incerteza de Localização	$\sigma_p = 10$ m.	$\sigma_p = 50$ m.	$\sigma_p = 220$ m.

Figura 3.10: Histogramas estatísticos do cenário A: (a-b-c) estimativas de $\hat{m}$; (d-e-f) estimativas de $\hat{r}$.

Figura 3.11: Histogramas estatísticos do cenário B: (a-b-c) estimativas de $\hat{m}$; (d-e-f) estimativas de $\hat{r}$.

Figura 3.12: Histogramas estatísticos do cenário C: (a-b-c) estimativas de $\hat{m}$; (d-e-f) estimativas de $\hat{r}$.

De forma similar ao caso da técnica TALE, foi observado que o desempenho a favor do SALE foi significativo em situações mais severas de incerteza de localização, *e.g.*, quando os valores de σ_p são da ordem do parâmetro r , ou seja, quando os erros de localização no sistema de comunicação sem fio resultam em erros de localização equivalentes às zonas de dependência espacial do ambiente de rádio, expressas pela distância de decorrelação espacial do sombreamento log-normal. De fato, como a técnica SALE também é fundamentada no cálculo de distâncias entre as coordenadas espaciais, a diferenciação entre as medidas capturadas e os valores verdadeiros de potência de recepção se torna significativa quando σ_p é elevado, especialmente da ordem do parâmetro r .

Apesar do interesse principal estar na modelagem dos efeitos em larga escala, algumas investigações de simulação que incluíram os efeitos em micro-escala (que não são mostradas) indicaram que a melhora proporcionada pela técnica SALE foi menos significativa comparada ao caso em que $c_0 = 0$. Isto significa que o SALE não foi eficiente na captura do efeito pepita no qual o semivariograma se torna extremamente descontínuo (*e.g.*, altos valores de *nugget* comparadas ao *sill*). Isto é explicado pelo fato do método atuar suavizando a descontinuidade na origem do semivariograma. Assim, argumenta-se que melhor desempenho alcançado com o SALE está relacionado a captura dos efeitos em larga escala dos processos aleatório espaciais, fornecendo semivariâncias e distâncias lags ajustadas que contribuem para a melhor estimação do vetor θ . Além disso, resalta-se que o método dos mínimos quadrados requer a inicialização de todos os parâmetros (m, r, c_0) e sua convergência pode não ser alcançada se tal inicialização não for adequada ou se houver incompatibilidade entre o semivariograma experimental estimado e o modelo analítico utilizado.

3.4.3 Formulação Matemática do Preditor TS-KALE

Esta subseção reúne as informações relacionadas com as técnicas TALE e SALE para que a formulação matemática do preditor espacial TS-KALE possa ser apresentada, permitindo compreender melhor o preditor proposto. Ressalta-se que a concepção do TS-KALE é fundamentada nos ajustes do KALE aplicado ao modelo de predição espacial para a geração do REM. Desta forma, as predições espaciais de potência de recepção do TS-KALE, em coordenadas espaciais desconhecidas $\mathbf{s}_0$, pode ser expressa analiticamente de acordo com

$$\begin{aligned}\hat{P}_{\text{TS-KALE}}(\mathbf{s}_0) &= \hat{\mu}_{\text{TALE}}(\mathbf{s}_0) + Z_{\text{TS-KALE}}(\mathbf{s}_0) \\ &= \hat{\mu}_{\text{TALE}}(\mathbf{s}_0) + \sum_{i=1}^N \lambda_{\text{TS-KALE}}^i Z_{\text{adj}}(\mathbf{s}_i)\end{aligned}\tag{3.18}$$

em que $\hat{\mu}_{\text{TALE}}(\mathbf{s}_0)$ é a estimativa ajustada da potência média de recepção, obtida com a técnica TALE, e $Z_{\text{TS-KALE}}(\mathbf{s}_0)$, que consiste na predição espacial do sombreado do canal sem fio, obtida a partir dos valores regionalizados ajustados $Z_{\text{adj}}(\mathbf{s}_i)$ e que é fundamentada nos ajustes propostos pelos algoritmos TALE e SALE. A combinação destas duas parcelas, $\hat{\mu}_{\text{TALE}}(\mathbf{s}_0)$ e $Z_{\text{TS-KALE}}(\mathbf{s}_0)$ define a predição espacial da potência de recepção $\hat{P}_{\text{TS-KALE}}(\mathbf{s}_0)$ do preditor TS-KALE.

A forma com os pesos são obtidos na combinação linear de $Z_{\text{TS-KALE}}(\mathbf{s}_0)$ é o que diferencia o preditor TS-KALE das técnicas Kriging convencionais (KILE e KALE). Especificamente, $\lambda_{\text{TS-KALE}}^i$ são os pesos Kriging obtidos com a solução do sistemas de semivariâncias/covariâncias construídos com o vetor de parâmetros $\boldsymbol{\theta}$, obtido com os resultados da tendência (TALE) e do semivariograma experimental (SALE), ambos ajustados. Em outras palavras, o resultado de predição espacial do TS-KALE é dado por uma parcela média, $\hat{\mu}_{\text{TALE}}(\mathbf{s}_0)$, e uma predição espacial $Z_{\text{TS-KALE}}(\mathbf{s}_0)$, derivada do ajuste do sistema de semivariâncias/covariâncias (*i.e.*, KALE), mas que é construído a partir dos parâmetros resultantes das técnicas TALE e SALE. Neste sentido, é possível interpretar a técnica proposta como sendo o preditor KALE dotado de um sistema de semivariâncias/covariâncias que utiliza técnicas de estimação mais aprimoradas para aprendizagem de parâmetros, considerando o problema da incerteza de localização.

3.5 Conclusões - Capítulo 3

Este capítulo apresentou o modelo de incerteza de localização aplicado ao ambiente de rádio. Seus impactos foram discutidos, especialmente na aprendizagem de parâmetros. Com isso, uma técnica de predição foi proposta (TS-KALE) e suas características (TALE e SALE) foram investigadas em diversas configurações do ambiente de rádio. Os resultados relacionados com a Fase I de aprendizagem se mostram promissores. Com isso, o próximo capítulo explora a Fase II, com o objetivo de avaliar o desempenho das técnicas de predição usadas na construção do REM.

Capítulo 4

Desempenho de Predição Espacial do REM

Este capítulo apresenta os resultados de simulação de desempenho dos preditores espaciais utilizados na Fase II do modelo de predição espacial do REM. A avaliação de desempenho é conduzida, buscando investigar o impacto de diferentes fatores na acurácia das predições.

Inicialmente, o fluxograma de simulação e as métricas de desempenho são apresentados, seguido da exposição e discussões sobre os resultados. Expressões teóricas sobre o desempenho dos preditores são usadas para validar os algoritmos de predição Kriging. A aprendizagem de parâmetros é investigada por meio de simulações, considerando o conhecimento *a priori* do vetor θ , bem como a sua estimação. Isto permite verificar como o conhecimento sobre os parâmetros do vetor θ influenciam os resultados das predições espaciais. Além disso, o impacto de diferentes fatores do ambiente de rádio é analisado, considerando: i) o número de medidas capturadas pelos dispositivos ao longo da área de cobertura do sistema de comunicação sem fio; ii) a estimação da perda por percurso média (tendência); iii) a introdução da incerteza de localização nas medidas e iv) a influência das flutuações aleatórias espaciais do sombreamento do canal sem fio.

O preditor TS-KALE é adicionado às simulações, com o propósito de verificar se os algoritmos de aprendizagem e predição desenvolvidos trouxeram benefícios, em termos de melhoria na acurácia das predições espaciais. Um comparativo com preditores determinísticos também é realizado e discussões sobre o funcionamento das técnicas são feitas, a partir da análise dos resultados. Isto possibilita o posicionamento das propostas do trabalho em relação às técnicas convencionais utilizadas na geração do REM. Finalmente, uma subseção aborda como diferentes modelos de distribuição espacial dos dispositivos podem ser utilizados para gerar diferentes padrões de amostragem espacial das medidas e, consequentemente, influenciar as predições em sistemas de comunicação sem fio.

4.1 Fluxograma de Simulação

A Figura 4.1 mostra o fluxograma utilizado nas simulações para a análise de desempenho, que são baseadas no método de Monte Carlo. O ponto inicial de cada realização de simulação consiste na geração das coordenadas espaciais relacionadas com o modelo de incerteza de localização. Especificamente, a densidade de probabilidade uniforme é utilizada para gerar o conjunto de coordenadas espaciais estimadas $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, que possuem uma incerteza de localização retratada pelo vetor $\{\mathbf{p}_e(\mathbf{s}_1), \dots, \mathbf{p}_e(\mathbf{s}_N)\}$. Isto permite obter o conjunto de coordenadas espaciais realizadas $\{\mathbf{r}_1, \dots, \mathbf{r}_N\}$, que descreve a posição exata dos dispositivos no ambiente de rádio e que são utilizadas na geração do processo aleatório espacial, *i.e.*, medidas de potência de recepção $\{P_{rx}(\mathbf{r}_1), \dots, P_{rx}(\mathbf{r}_N)\}$. A Fase I fornece os parâmetros do ambiente de rádio ($\hat{\alpha}$, $\hat{\gamma}$ e $\hat{\theta}$), necessários para as previsões da Fase II. Assim, o conjunto de medidas é processado pelos algoritmos de aprendizagem de parâmetros e previsão espacial (Fase I e II) e uma coordenada espacial alvo $\mathbf{s}_0$ é aleatoriamente escolhida juntamente com seu valor de potência de recepção, $P_{rx}(\mathbf{s}_0)$, para que o erro quadrático médio de previsão possa ser calculado em cada realização de simulação.

Figura 4.1: Fluxograma geral de simulação para análise de desempenho das previsões espaciais.

Nas simulações, os resultados de predição espacial na coordenada alvo, $\hat{P}_{\text{KILE}}(\mathbf{s}_0)$ e $\hat{P}_{\text{KALE}}(\mathbf{s}_0)$, são obtidos por meio da combinação entre a estimativa da potência média de recepção, $\hat{\mu}(\mathbf{s}_0)$, e a predições do sombreamento do canal sem fio, $Z_{\text{KILE}}(\mathbf{s}_0)$, $Z_{\text{KALE}}(\mathbf{s}_0)$. No caso do preditor TS-KALE, a estimativa da potência média de recepção ajustada, $\hat{\mu}_{\text{TALE}}(\mathbf{s}_0)$, é combinada com $Z_{\text{TS-KALE}}(\mathbf{s}_0)$, resultando na predição da potência de recepção $\hat{P}_{\text{TS-KALE}}(\mathbf{s}_0)$, a partir das técnicas TALE e SALE. Desta forma, cada preditor utiliza o conjunto $\{P_{\text{rx}}(\mathbf{r}_1), \dots, P_{\text{rx}}(\mathbf{r}_N)\}$, as coordenadas espaciais estimadas $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ e: (a) a covariância espacial sem o procedimento de ajuste, *i.e.*, preditor KILE; (b) a covariância espacial com o procedimento de ajuste, *i.e.*, preditor KALE e (c) a tendência, o semivariograma e a covariância espacial obtidos com o procedimento de ajuste, *i.e.*, preditor TS-KALE. Isto significa que, enquanto o preditor KILE ignora a distribuição $g(\cdot)$ os outros preditores levam em conta as suas estatísticas, sendo que o preditor KALE só utiliza $g(\cdot)$ no sistema de equações Kriging da Fase II e o preditor proposto TS-KALE utiliza a distribuição $g(\cdot)$ em ambas as fases (I e II) do modelo de predição.

A métrica utilizada para avaliar o desempenho dos preditores consiste no erro quadrático médio de predição espacial (MSPE - *Mean Square Prediction Error*),

$$\text{MSPE} = \mathbb{E}\{[\hat{P}_{\text{rx}}(\mathbf{s}_0) - P_{\text{rx}}(\mathbf{s}_0)]^2\}. \quad (4.1)$$

Esta métrica mede a variabilidade das predições em relação ao valor verdadeiro de potência de recepção do ambiente de rádio. A estimação empírica do erro quadrático médio (EMSPE - *Empirical Mean Square Prediction Error*) é realizada por meio das diferenças quadráticas (dB²) entre os valores de potência de recepção (dados em dBm), de acordo com

$$\text{EMSPE} = \frac{1}{L} \sum_{l=1}^L [\hat{P}_{\text{rx}}^l(\mathbf{s}_0) - P_{\text{rx}}^l(\mathbf{s}_0)]^2, \quad (4.2)$$

em que L é o número de realizações independentes de simulação. Embora seja uma métrica ótima em um sentido estatístico específico, o MSPE é amplamente utilizado na análise de desempenho das predições para a geração do REM, sobretudo em sistemas de comunicações sem fio [16-18], [72-73], [78-80], [83]. Além disso, as equações abaixo (extraídas de [93]) fornecem as expressões teóricas de MSPE, considerando os erros de localização nas predições, mas assumindo o perfeito conhecimento dos parâmetros $\boldsymbol{\theta}$ da covariância $\mathbf{C}$. Isto significa que tais expressões podem ser utilizadas para validar os algoritmos de predição Kriging desenvolvidos neste trabalho quando $\boldsymbol{\theta}$ é conhecido.

$$\begin{aligned} \text{MSPE KILE} &= \sigma^2 - 2\mathbf{c}_{g0}^T \mathbf{C}^{-1} \mathbf{c}_0 + \mathbf{c}_0^T \mathbf{C}^{-1} \mathbf{C}_g \mathbf{C}^{-1} \mathbf{c}_0 + (\mathbf{c}_0^T \mathbf{C}^{-1} (\mathbf{x}_g - \mathbf{x}) \alpha)^2 \\ \text{MSPE KALE} &= \sigma^2 - \mathbf{c}_{g0}^T \mathbf{C}_g^{-1} \mathbf{c}_{g0} + (\mathbf{c}_0^T \mathbf{C}^{-1} (\mathbf{x}_g - \mathbf{x}) \alpha)^2. \end{aligned} \quad (4.3)$$

4.2 Predições com Parâmetros de Covariância Conhecidos

A análise de desempenho dos preditores espaciais é iniciada assumindo o conhecimento do semivariograma γ e dos parâmetros de covariância θ do ambiente de rádio. Isto permite avaliar os resultados de predição das técnicas KILE e KALE em relação às expressões teóricas de desempenho do MSPE, mostradas em (4.3). O tratamento da tendência também é incluído, contemplando duas situações: i) o conhecimento perfeito de α e a remoção ideal para recuperação dos valores regionalizados, denotada como estimação perfeita da tendência e ii) a estimação do parâmetro α (*i.e.*, regressão do modelo log-distância) seguida da recuperação dos valores regionalizados, denotada como estimação imperfeita da tendência.

4.2.1 Influência do Número de Medidas

Esta subseção tem o objetivo de verificar se o uso das estatísticas dos erros de localização resulta em melhor desempenho para as predições, considerando a alteração do número de medidas capturadas do ambiente de rádio. Configurações típicas de sistemas de comunicações sem fio são selecionadas: área de cobertura de 500 m $\times$ 500 m; ERB com localização fixa em $\mathbf{s}_{tx} = (s_{tx}, s_{ytx}) = (100, 100)$; modelo do ambiente de rádio em larga escala, composto pela perda por percurso média, com $\alpha = 3$, e pelo sombreamento log-normal, com desvio padrão $\sigma = 7$ dB e distância de decorrelação espacial $d_{corr} = 100$ m, *i.e.*, $\theta = [49, 100, 0]$. O conjunto de medidas é corrompido por erros de localização $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$, com $\sigma_p = 40$ m.

Os resultados de simulação são mostrados na Figura 4.2. Primeiramente, verifica-se que as curvas de EMSPE estimadas convergiram para o desempenho teórico do MSPE, permitindo a validação dos algoritmos de predição Kriging desenvolvidos. Sobre o comparativo entre os preditores, as curvas da técnica KALE indicam que o uso das estatísticas da incerteza de localização produziu melhores resultados, em termos de acurácia (*i.e.*, menores valores de EMPSE). Os cenários A e B consistem nos casos em que o tratamento da tendência (*i.e.*, estimação e remoção) foi realizado de forma perfeita e imperfeita, respectivamente, permitindo verificar seus impactos no desempenho dos preditores. Por exemplo: para $N = 125$ medidas, ambos os preditores KILE e KALE tiveram uma piora de desempenho em torno de 4 dB² no EMSPE no comparativo entre os cenários A e B. Entre os preditores, para as situações em que $N > 100$, uma redução significativa (mais que 7 dB²) foi obtida pelo preditor KALE em relação ao KILE no cenário com estimação imperfeita da tendência.

Desta forma, os resultados apontam que o preditor KALE é mais acurado nas estimativas de predição do que o KILE com os mesmos recursos (número de medidas). De fato, embora o aumento de N tenha favorecido ambos os preditores, os erros de localização afetaram significativamente o desempenho do KILE, enquanto o KALE se beneficiou com mais medidas coletadas e apresentou melhor acurácia para todos os valores de N . Portanto, é possível mencionar que levar em conta as estatísticas dos erros de localização é mais eficiente que ignorá-las à medida que N é aumentado.

A habilidade dos preditores atingirem um bom desempenho com o menor número de medidas possível define a sua eficiência e consiste em um fator chave na escolha de um preditor espacial para a geração do REM, especialmente em sistemas cujas arquiteturas não permitem a realização e o envio de muitas medidas, devido as restrições de recursos do sistema de comunicação (*e.g.*, energia e largura de banda). Os principais exemplos neste sentido se relacionam com aplicações em redes de sensores, envolvendo o monitoramento e vigilância de diferentes tipos de ambientes, controle e gerência de tráfego, plantas de operação industriais, entre outras aplicações [80-81]. Na prática, os métodos de localização utilizados e a limitação de recursos nas redes de sensores podem resultar em erros de localização significativos, especialmente em redes de sensores com mobilidade, exigindo o tratamento dos consequentes efeitos negativos na acurácia das previsões [81-82].

Mesmo com métodos de localização mais sofisticados, estas situações de incerteza também ocorrem em sistemas de comunicações móveis. A razão para isto é que os métodos de localização fazem medidas de sinais e parâmetros específicos da rede, que são afetados por condições de recepção deterioradas, submetidas aos efeitos do canal de propagação. Isto significa que a aleatoriedade do canal sem fio pode amplificar a incerteza de localização no ambiente de rádio. Além disso, o funcionamento dos métodos envolve procedimentos de sinalização de mensagens na rede, o que também pode ocasionar falhas e a geração de informações de localização inválidas (*e.g.*, *outliers* de localização). Com isso, a capacidade dos preditores em fornecer previsões com alta acurácia em circunstâncias adversas do canal de propagação e da incerteza de localização, fundamenta a robustez das técnicas de previsão espacial. Como a geração do REM se relaciona com diversas aplicações de comunicações sem fio, argumenta-se que esta robustez é essencial, pois cada aplicação pode estar vinculada a um tipo de canal de propagação, além de utilizar métodos de localização específicos, resultando em diferentes regimes de incerteza de localização.

Figura 4.2: Influência do número de medidas N no desempenho dos preditores KILE e KALE com estimação perfeita (Cenário A) e imperfeita (Cenário B) da perda por percurso média.

4.2.2 Influência dos Erros de Localização

Dada a importância da robustez das predições, uma análise é realizada, considerando o aumento gradual da severidade dos erros de localização introduzidos no modelo de sistema, assumindo um conjunto fixo de $N = 100$ medidas coletadas para ambos os preditores KILE e KALE. Os parâmetros e configurações do sistema de comunicação sem fio são mantidos iguais ao caso da simulação anterior. Os resultados podem ser observados na Figura 4.3, na qual as estimações perfeita e imperfeita da tendência são representadas pelos cenários C e D, respectivamente.

A primeira conclusão obtida a partir dos gráficos aponta que o preditor KALE é mais robusto frente aos erros de localização, comparado com o preditor KILE. Nota-se também que os resultados de estimação do EMSPE seguem as curvas teóricas do MSPE dos preditores. No comparativo entre as técnicas, observa-se uma redução de 4 dB^2 no EMSPE, obtida pelo KALE quando $\sigma_p = 40 \text{ m}$ e mais de 6 dB^2 no cenário em que a estimação da tendência é considerada. Neste aspecto, verifica-se que ambos os preditores tiveram seu desempenho afetado pela estimação imperfeita da tendência. De fato, o que está ocorrendo é que à medida em que a severidade dos erros de localização é aumentada, o processo aleatório espacial observado com incerteza de localização se torna significativamente diferente do processo aleatório espacial sem erros de localização. Esta diferença é refletida nas covariâncias espaciais sem e com o procedimento de ajuste e, conseqüentemente, no sistema de equações de covariâncias do Kriging. Com isso, as estimativas produzidas pelos preditores espaciais KILE e KALE são distintas. Em diversas configurações analisadas, o preditor KALE se manteve a frente do KILE, fornecendo predições mais acuradas para o REM. Estes resultados reafirmam a conclusão de que ignorar os erros de localização ao tentar capturar a variabilidade espacial para a geração de predições não é uma decisão eficiente do ponto de vista de desempenho, em termos de acurácia das predições espaciais.

Figura 4.3: Influência dos erros de localização no desempenho dos preditores KILE e KALE com estimação perfeita (Cenário C) e imperfeita (Cenário D) da perda por percurso média.

4.2.3 Influência do Sombreamento do Canal sem Fio

Esta subseção investiga os impactos de um aspecto intrínseco do canal sem fio nos resultados de desempenho dos preditores. De forma específica, o desvio padrão do sombreamento log-normal σ é alterado nas simulações com o objetivo de variar a dispersão das flutuações espaciais da potência de recepção do ambiente de rádio. Esta característica é importante no âmbito dos sistemas de comunicações sem fio, uma vez que o valor de σ caracteriza diferentes tipos de ambiente de propagação (*e.g.*, para cenários de comunicações externos, o valor de σ varia entre 5 e 13 dB [95-97]). Além disso, as características estatísticas (*e.g.*, densidade de probabilidade e seus parâmetros) do sombreamento do canal são utilizadas na modelagem de problemas envolvendo a probabilidade de inoperância e a cobertura em sistemas de comunicações sem fio [96]. As configurações de simulação foram mantidas iguais aos casos anteriores, exceto que a raiz quadrada da variância *sill* σ^2 é alterada. Neste caso, o número de medidas coletadas foi mantido fixo e igual a $N = 100$, enquanto as coordenadas espaciais também são corrompidas por erros de localização Gaussianos com desvio padrão $\sigma_p = 40$ m.

Os resultados podem ser observados na Figura 4.4. Novamente, constata-se a superioridade de desempenho do KALE em relação ao KILE. Os casos de estimação perfeita e imperfeita da tendência são caracterizados pelos cenários E e F, respectivamente. É possível verificar a convergência com as curvas teóricas e que o aumento de σ afeta significativamente o desempenho de ambos os preditores. Isto indica que a variabilidade espacial dos valores da potência de recepção tem efeito direto na acurácia das predições baseadas nas técnicas Kriging. As expressões teóricas de MSPE em (4.3) também suportam o entendimento deste aspecto, permitindo inferir que os erros de predição dependem diretamente da variância do sombreamento do canal sem fio. Desta forma, as características da perda por percurso média e suas flutuações em função do sombreamento possuem um papel importante no estudo de desempenho das predições espaciais do REM.

Figura 4.4: Influência do sombreamento do canal sem fio no desempenho dos preditores KILE e KALE com estimação perfeita (Cenário E) e imperfeita (Cenário F) da perda por percurso média.

4.3 Predições com Estimação dos Parâmetros de Covariância

Até o momento, as análises de desempenho das predições apontam que o uso das estatísticas da incerteza de localização permite que melhores resultados sejam alcançados, considerando o completo conhecimento dos parâmetros do vetor θ do ambiente de rádio. Na prática, tais parâmetros não são necessariamente conhecidos e, portanto, sua estimação se faz necessária. De fato, considerar a falta de conhecimento sobre os parâmetros usados na modelagem do canal sem fio torna o modelo de sistema mais realístico. Atuando neste sentido, os resultados que seguem consideram a estimação do vetor θ , além da estimação perfeita e imperfeita da tendência (com isso, somente as curvas empíricas de EMSPE são mostradas). O objetivo consiste em constatar qual é o impacto da aprendizagem dos parâmetros no desempenho dos preditores e se os procedimentos de ajuste propostos através do preditor TS-KALE trazem efetivamente melhores resultados de predição.

Uma simulação é conduzida incorporando as estimativas $\hat{\alpha}$, $\hat{\gamma}$ e $\hat{\theta}$ com as configurações do sistema de comunicação sem fio mantidas iguais aos casos das simulações anteriores. Os resultados podem ser observados na Figura 4.5, na qual as situações de estimação perfeita e imperfeita da tendência são expressas pelos cenários G e H, respectivamente. Nota-se que quando o vetor θ é estimado, a diferença de desempenho entre os preditores KILE e KALE é reduzida. Isto pode ser verificado comparando-se as reduções no EMSPE obtidas pelo preditor KALE em relação ao KILE, nos casos em que θ é conhecido *a priori*. Em outras palavras, ainda é vantajoso realizar o procedimento de ajuste do KALE, mas a melhora de desempenho proporcionada não é substancial e depende de forma significativa da estimação de θ . Estes resultados corroboram as conclusões apontadas em [84] e mostram que a estimação dos parâmetros de covariância tem papel importante nas predições afetadas pela incerteza de localização.

Figura 4.5: Influência da incerteza de localização e da estimação de θ no desempenho dos preditores KILE, KALE e TS-KALE, com perfeita (Cenário G) e imperfeita (Cenário H) estimação da tendência.

Sobre a técnica proposta, o desempenho alcançado pelo preditor TS-KALE nas situações destacadas ($\sigma_p = 40$ m) aponta reduções da ordem de 5 dB² a 7 dB² no EMSPE em relação aos preditores KILE e KALE, indicando que os procedimentos de ajuste realizados na estimação da tendência e do semivariograma contribuíram para o alcance de previsões espaciais com maior acurácia. As investigações realizadas durante as simulações também revelaram o melhor desempenho obtido pelo TS-KALE em diferentes configurações do ambiente de rádio (*e.g.*, diversos tamanhos de área de cobertura, número de medidas, bem como diferentes valores para os parâmetros *sill* e *range*).

É importante mencionar que embora uma técnica alternativa para a estimação de α , γ e θ submetida aos erros de localização seja proposta, não se pode argumentar que tal técnica é ótima em algum sentido estatístico, pois o método LS utilizado na aprendizagem de parâmetros não pressupõe o conhecimento da distribuição de probabilidade das observações. Ainda assim, apesar das diferenças em relação ao modelo de sistema, os resultados alcançados corroboram as conclusões apresentadas em [78-79], [81-83], [93] no sentido de apontar os benefícios com o uso das estatísticas dos erros de localização nas previsões espaciais. Neste sentido, é importante ressaltar quais são as principais diferenças entre o preditor proposto e as técnicas utilizadas nas principais referências deste trabalho. Em relação à [79], é possível apresentar algumas diferenças como: i) o uso da técnica ML para estimação dos parâmetros do canal sem fio, considerando a incerteza de localização nesta etapa, enquanto nosso trabalho incorpora a incerteza de localização no modelo de previsão espacial geoestatístico baseado no semivariograma e na técnica LS; ii) o uso do modelo de covariância Gaussiano (para que uma expressão analítica fechada possa ser utilizada), enquanto nosso trabalho usa o método de integração de Monte Carlo a fim de permitir diferentes combinações entre modelos de covariância (*e.g.*, exponencial) e a densidade de probabilidade dos erros de localização $g(\cdot)$; iii) a análise de desempenho é baseada em ambientes unidimensionais, enquanto nosso trabalho considera a geração de mapas de ambiente de rádio contemplando duas dimensões.

Em relação à [78], este apresenta um modelo de sistema diferente devido à abordagem Bayesiana e a decomposição do sombreamento log-normal através de funções base espaciais para a realização do Kriging com rank fixo (*fixed rank* Kriging) [122]. O propósito do método de rank fixo visa contornar o problema da inversão matricial do Kriging para que seja possível fazer previsões a partir de um número significativamente grande de medidas (*e.g.*, longas campanhas de *drive-test*). A redução de complexidade alcançada é plausível, embora a acurácia das previsões espaciais do método seja afetada, caracterizando uma relação de compromisso entre complexidade e acurácia. Sobre a aprendizagem de parâmetros do canal sem fio, [78] propõe um algoritmo de aproximação estocástica ao método de maximização da esperança para que a técnica ML pudesse ser utilizada no preditor Kriging com rank fixo, levando em conta a incerteza de localização. Isto permitiu que o preditor ótimo da esperança condicional fosse obtido nos regimes de incerteza de localização, resultando em melhores previsões comparadas com o desempenho dos preditores lineares (BLUP) utilizados em nosso trabalho (embora com modelo de sistema diferente e mais flexível).

4.4 Comparações com Preditores Espaciais Convencionais

Com o propósito de expandir a análise das predições, um comparativo entre técnicas é realizado, incluindo os principais preditores usados na geração do REM. Os resultados mostrados até o momento constataam que a aprendizagem de parâmetros foi afetada pelos erros de localização e, de fato, qualquer incerteza na estimação dos parâmetros envolvidos nas etapas do modelo de predição espacial poderá refletir no desempenho dos preditores [123-124]. Por isso, a análise realizada inclui preditores que não necessitam do conhecimento de θ . O propósito consiste em verificar se a estimação dos parâmetros e funções α , γ e θ afeta os preditores Kriging de modo a comprometer o seu desempenho no comparativo com as técnicas convencionais. Isto também permite realizar o posicionamento em relação à diversos trabalhos sobre a geração do REM.

O preditor IDW é uma das técnicas de predição espacial mais citadas nos estudos para a geração do REM [11], [14], [42-45], pois consiste em um preditor determinístico de menor complexidade, baseado no inverso da distância entre as posições relacionadas com as medidas coletadas e a coordenada alvo $\mathbf{s}_0$, de acordo com

$$\begin{aligned}\hat{P}_{IDW}(\mathbf{s}_0) &= \hat{\mu}(\mathbf{s}_0) + Z_{IDW}(\mathbf{s}_0) \\ Z_{IDW}(\mathbf{s}_0) &= \frac{\sum_{i=1}^N w_i(\mathbf{s}_i, \mathbf{s}_0) Z_g(\mathbf{s}_i)}{\sum_{i=1}^N w_i(\mathbf{s}_i, \mathbf{s}_0)}, \text{ com } w_i(\mathbf{s}_i, \mathbf{s}_0) = \frac{1}{d(\mathbf{s}_i, \mathbf{s}_0)^p},\end{aligned}\quad (4.4)$$

em que $\hat{P}_{IDW}(\mathbf{s}_0)$ é o resultado da predição IDW, $Z_g(\mathbf{s}_i)$ consiste nas medidas coletadas (após a remoção da tendência) com erros de localização e $w_i(\mathbf{s}_i, \mathbf{s}_0)$ são os pesos que dependem do inverso da p -ésima potência da função distância $d(\mathbf{s}_i, \mathbf{s}_0)$, com IDW-1 ($p = 1$) e IDW-2 ($p = 2$).

O preditor NNI é mais um exemplo de técnica determinística de predição usada na geração do REM, que estima a potência de recepção por meio da média ponderada dos valores de vizinhança da coordenada alvo $\mathbf{s}_0$. Diferentes métodos podem ser utilizados na determinação da zona de vizinhança espacial (*e.g.*, diagramas de Voronoi e distâncias espaciais), resultando em um subconjunto V_{NNI} de medidas para compor o resultado de predição. Neste comparativo, a região de vizinhança é escolhida como o subconjunto V_{NNI} contendo as $\|V_{NNI}\|$ medidas mais próximas das coordenadas alvo em termos de distância espacial, de acordo com

$$\hat{P}_{NNI}(\mathbf{s}_0) = \frac{1}{\|V_{NNI}\|} \sum_{\mathbf{s}_i \in V_{NNI}} P(\mathbf{s}_i) \quad (4.5)$$

Conforme mencionado, essas duas técnicas foram escolhidas por duas razões principais: i) por serem convencionais nos estudos comparativos para a geração do REM e ii) por não dependerem da estimação de θ , bem como de pressuposições estatísticas sobre o processo aleatório espacial.

As técnicas IDW, com parâmetro $p = 2$ (escolhido por apresentar melhor desempenho que $p = 1$) e NNI, supondo um agrupamento de (no mínimo) três dispositivos mais próximos (*i.e.*, $V = 3$) foram adicionadas na análise. As configurações do ambiente de rádio foram mantidas iguais às simulações anteriores e mediante $L = 20.000$ realizações, os valores de EMSPE de cada preditor foram coletados, considerando as situações de ausência e presença de erros de localização.

A Figura 4.6 mostra os resultados estatísticos na forma de *boxplots*, cujas marcas centrais consistem nas medianas das distribuições dos valores de EMSPE de cada preditor. Também são evidenciados os limites do primeiro e do terceiro quartis (bordas do *boxplot*), os máximos e mínimos das distribuições (intervalos pontilhados), além das medidas que são classificadas como *outliers* do conjunto de dados (pontos flutuantes). No cenário sem incerteza de localização (a), é possível observar uma equivalência de desempenho entre os preditores KILE, KALE e TS-KALE, pois $\sigma_p = 0$ m. Os resultados dos *boxplots* mostram que, neste regime de ausência de erros de localização, a exploração da covariância espacial realizada nas técnicas Kriging faz sentido e resulta em predições com maior acurácia, comparada com os resultados dos preditores convencionais. Por outro lado, a incorporação dos erros de localização em (b) afetou as estatísticas no EMSPE de todas as técnicas. Especificamente, o desempenho do preditor KILE (41.7 dB^2) foi degradado e neste regime ($\sigma_p = 70$ m) apresentou piores resultados que o preditor IDW (39.89 dB^2), enquanto o preditor KALE (36 dB^2), que também foi afetado, conseguiu superar ambos os preditores. De fato, do que está ocorrendo é que a falha na captura da variabilidade espacial do ambiente de rádio (por ignorar a incerteza de localização) compromete significativamente o KILE, enquanto o KALE é afetado pelos impactos dos erros de localização na estimação de θ . Em ambas as situações, de ausência e presença de erros de localização, a técnica determinística NNI foi a que apresentou menor acurácia, enquanto o preditor TS-KALE atingiu o melhor desempenho dentre todos os preditores analisados, pois apresentou o menor valor (32 dB^2) para a estatística mediana dos erros de predição (refletindo maior acurácia).

Figura 4.6: Estatísticas de EMSPE: (a) sem incerteza de localização; (b) com incerteza de localização.

A Figura 4.7 adiciona mais resultados ao comparativo entre os preditores Kriging e as técnicas convencionais para a geração do REM. A Fig.4.7(a) resgata a análise de robustez dos preditores contra os efeitos da incerteza de localização. Novamente, as mesmas configurações do ambiente de rádio anteriores foram adotadas, considerando a estimação de α , γ e θ e uma variação maior do desvio padrão ($\sigma_p = [0, 100]$). No caso da Fig.4.7(b) o desempenho dos preditores é expresso em função do número de medidas, assumindo um valor fixo para $\sigma_p = 70$ m.

Verifica-se na Fig.4.7(a) que o impacto dos erros de localização foi significativo nas técnicas que, de alguma forma, ignoraram a incerteza de localização (*e.g.*, IDW, KILE e NNI). Por exemplo: quando $\sigma_p = 0$ m, apesar de apresentar pior desempenho, a diferença entre o preditor NNI e os demais é relativamente pequena. No entanto, à medida em que σ_p aumenta, esta diferença cresce de forma bastante significativa, comprometendo a utilização da técnica NNI. Uma situação similar ocorre com o preditor IDW, com a diferença de que este superou o desempenho do KILE, que foi mais afetado pelos erros de localização ao tentar explorar a covariância das medidas do que o preditor determinístico IDW. Observa-se que para toda a faixa de σ_p , a técnica TS-KALE apresenta melhor desempenho frente aos ajustes realizados no preditor KALE e as técnicas convencionais IDW e KILE. Conforme destacado (elipses) a diferença de desempenho, em termos de EMSPE, entre o preditor com maior (TS-KALE) e menor (NNI) acurácia supera 11 dB².

Um quadro equivalente de resultados é mostrado na Fig.4.7(b), ou seja, o preditor TS-KALE foi capaz de atingir melhor desempenho no comparativo com as outras técnicas, utilizando o menor número de medidas coletadas. Desta forma, argumenta-se que o TS-KALE se apresenta como uma técnica de predição, que é caracterizada por uma abordagem alternativa aplicada à fase de aprendizagem de parâmetros (TALE e SALE), que busca fornecer melhores resultados de predição, quando existe incerteza de localização no sistema de comunicação sem fio.

Figura 4.7: Comparação de desempenho dos preditores Kriging e convencionais: (a) frente a incerteza de localização (b) em função do número de medidas.

4.5 Influência da Amostragem Espacial

Até o momento, as predições espaciais foram baseadas na captura de medidas feita a partir da densidade de probabilidade uniforme (*i.e.*, amostragem completamente aleatória sobre o ambiente de rádio). Isto significa que as coordenadas espaciais das medidas coletadas são independentes e equiprováveis no domínio espacial D , que representa a área de cobertura do sistema de comunicação sem fio. Sobre este aspecto, a escolha de um modelo para a distribuição espacial dos dispositivos é um ponto relevante, pois depende de características específicas dos sistemas de comunicações sem fio e, neste sentido, a distribuição uniforme foi inicialmente utilizada baseando-se na pressuposição da falta de conhecimento *a priori* sobre as estatísticas que permitem modelar as localizações dos dispositivos na rede. Ainda assim, o resultado da amostragem espacial sobre o ambiente de rádio depende da distribuição espacial dos dispositivos, que não necessariamente é uniforme. Neste contexto, há um crescente interesse na literatura no uso de modelos de distribuição espacial para caracterizar as localizações dos dispositivos nas redes de comunicações sem fio, além do posicionamento das ERBs [129-131]. O estudo destes modelos se fundamenta na pesquisa da geometria estocástica aplicada a área de comunicações sem fio [131], [133].

A teoria de processos aleatórios pontuais (PP - *Point Process*)¹ é o principal objeto de estudo da geometria estocástica e fornece ferramentas adequadas para a modelagem de padrões espaciais dos dispositivos em redes sem fio. Sobre a definição, um processo aleatório pontual do tipo espacial consiste em um processo estocástico, cujas realizações resultam em um conjunto contável de pontos sobre o domínio D . Desta forma, as posições dos dispositivos (*i.e.*, coordenadas espaciais $\mathbf{s}$) são interpretadas como eventos de localização e suas realizações dão origem a um padrão espacial de pontos². A partir disto, estes conceitos são usados para modelar diferentes padrões de amostragem espacial das medidas, com o objetivo de investigar seus impactos nas predições do REM.

Um dos processos pontuais espaciais mais utilizados em comunicações sem fio é o processo pontual de poisson (PPP - *Poisson Point Process*), em específico, com característica homogênea (HPPP - *Homogeneous Poisson Point Process*). No HPPP, o número de pontos do padrão espacial N_{HPPP} é uma variável aleatória que segue a distribuição estatística de Poisson [131]. Uma característica importante do HPPP está na sua intensidade I , que consiste na razão entre o número de pontos e a área do domínio espacial (dada em km^2 ou m^2), *i.e.*, $I = N_{\text{HPPP}}/A$. Dessa forma, I expressa o número de pontos (dispositivos) por km^2 ou m^2 . Além disso, o valor constante de I implica na homogeneidade do HPPP, enquanto a sua variação implica em uma não homogeneidade ao PPP.

¹A principal vantagem no uso da geometria estocástica através da teoria de PP está na simplificação do tratamento matemático na solução de problemas complexos que envolvem o desempenho de redes de comunicações sem fio [131].

²A conexão entre a teoria de processos aleatórios pontuais e a geoestatística ocorre de duas formas. Primeiramente, o próprio processo aleatório pode se apresentar de forma pontual no espaço como em aplicações das ciências sociais (*e.g.*, índices de criminalidade e acidentes, indicadores econômicos e mapeamentos geográficos de doenças) e da Terra (*e.g.*, planejamentos ecológicos e análise de espécies). Segundo, o conjunto de localizações onde as medidas são realizadas pode ser interpretado como um processo aleatório pontual e, portanto, possui características e propriedades estatísticas [68]. Detalhes sobre a teoria de processos aleatórios pontuais e suas diferenças em relação aos processos aleatórios espaciais contínuos podem ser acessados em [68].

Com isso, a não homogeneidade do processo pontual de Poisson (nHPPP - *nonhomogeneous Poisson Point Process*) é utilizada para modelar situações nas quais a intensidade da distribuição espacial das medidas depende das coordenadas espaciais $\mathbf{s}$. Especificamente, a intensidade $I(\mathbf{s})$ não é constante e, de fato, consiste em uma função das coordenadas espaciais $\mathbf{s} = (s_x, s_y)$, *i.e.*, $I(\mathbf{s})$, indicando que a densidade de dispositivos por área e, conseqüentemente, das medidas capturadas, é mais ou menos intensa em certas regiões da área de cobertura³. O uso de algoritmos que utilizam as funções intensidade combinadas com a inicialização de centros de pontos espaciais na área de cobertura permite a geração de aglomerados de pontos espaciais, caracterizando o processo denotado como *clustering* (cPPP - *Cluster Poisson Point Process*). A Figura 4.8 mostra diferentes exemplos de amostragem espacial obtidos com a geração de processos aleatórios pontuais de Poisson. As explicações seguem logo adiante e detalhes sobre o processo de geração dos modelos podem ser acessados em [131-132].

Figura 4.8: Diferentes padrões de amostragem espacial: (a) completamente aleatória (uniforme); processo pontual de Poisson (b) homogêneo; (c) não homogêneo e (d) com aglomerados (*clusters*).

³O formato de $I(\mathbf{s})$ deve capturar o padrão espacial dos dispositivos, seja por uma proximidade com a ERB ou pelas localidades nas quais a probabilidade de ocorrência de eventos de localização (*i.e.*, presença dos dispositivos) é maior. Isto significa que o grau de homogeneidade expresso pela intensidade $I(\mathbf{s})$ não está vinculado estritamente aos aspectos de propagação no ambiente de rádio, mas sim com a realização do padrão espacial dos dispositivos.

A Fig.4.8(a) apresenta uma realização de amostragem espacial com $N = 50$ medidas, cujas coordenadas espaciais $\mathbf{s}$ são obtidas por meio da densidade de probabilidade uniforme. Este modelo é usado para a amostragem completamente aleatória sobre a área de cobertura do ambiente de rádio, ou seja, quando não há conhecimento *a priori* sobre as estatísticas da distribuição espacial dos dispositivos na rede. O padrão HPPP é mostrado na Fig.4.8(b) e foi simulado com intensidade constante de $I = 200$ medidas por km^2 . Neste processo, a área $A = 0.5 \text{ km} \times 0.5 \text{ km} = 0.25 \text{ km}^2$ determina a média da distribuição de Poisson, resultando em $N_{\text{HPPP}} = I \times A = 50$ eventos de localização na área de cobertura de 0.25 km^2 . Nota-se que, neste modelo, N_{HPPP} assume um valor diferente a cada realização sendo que, na média, são esperados 50 eventos de localização.

A Fig.4.8(c) apresenta o padrão nHPPP, que foi simulado com uma função intensidade específica, dada por $I(s_x, s_y) = \exp\{-(s_x - 100)^2 + (s_y - 100)^2\}$. Este formato específico escolhido para a função $I(s_x, s_y)$ intensifica o número de medidas do padrão espacial em torno de uma coordenada espacial $\mathbf{s} = (s_x, s_y) = (100, 100)$, permitindo simular uma concentração de dispositivos em uma determinada região do ambiente de rádio, por exemplo. Existem diferentes formas de geração do padrão espacial não homogêneo [131], [132]. Uma alternativa que simplifica a geração do nHPPP (utilizada neste trabalho) é baseada em um algoritmo de seleção conhecido como *thinning out*, no qual os pontos espaciais do HPPP são gerados e, em seguida, uma seleção é feita, a partir da aceitação de uma coordenada espacial $\mathbf{s}$ com probabilidade $I(s_x, s_y)/I$, em que $I(s_x, s_y)$ é a função intensidade não homogênea e I é a intensidade constante do HPPP [132]. Assim, a cada realização, o número de medidas do padrão espacial segue as estatísticas de Poisson, mas tais medidas são distribuídas de forma não homogênea no espaço, de acordo com o algoritmo de aceitação ou rejeição das coordenadas.

Finalmente, o padrão espacial cPPP é mostrado na Fig.4.8(d). O algoritmo de geração do cPPP começa com a inicialização das coordenadas espaciais dos aglomerados com o HPPP, ou seja, determina-se a quantidade inicial de aglomerados ($N_{\text{HPPP}} = 50$, neste caso de simulação) e suas posições na área de cobertura, obtidas a partir da densidade de probabilidade uniforme. Diferentemente dos outros padrões espaciais, uma função intensidade $\varrho(\mathbf{s})$ é usada para determinar como as próximas coordenadas espaciais serão geradas, a partir de cada aglomerado inicializado na área de cobertura. Funções do tipo exponenciais e Gaussianas são exemplos de modelos para $\varrho(\mathbf{s})$. O caso do padrão da Fig.4.8(d) segue o método em [132], no qual as coordenadas espaciais dos aglomerados são distribuídas de acordo com a função aleatória Gaussiana,

$$\varrho(s_x, s_y) = \frac{N_\varrho}{2\pi\sigma_\varrho^2} \exp\left\{-\frac{1}{2\pi\sigma_\varrho^2}(s_x^2 + s_y^2)\right\}, \quad (4.6)$$

cujas médias são as componentes (s_x, s_y) , ambas com desvio padrão σ_ϱ e a variável N_ϱ , que pondera a função Gaussiana, consiste na média do HPPP [132]. Isto significa que as novas coordenadas espaciais são distribuídas em torno das componentes (s_x, s_y) , de acordo com a função em 4.7, enquanto o número de novas coordenadas espaciais geradas segue a distribuição de Poisson com média N_ϱ .

Estes diferentes processos aleatórios pontuais são introduzidos na etapa de amostragem espacial do modelo de predição do REM. Com isso, uma simulação é realizada, levando em conta os padrões com amostragem completamente aleatória (uniforme), HPPP, nHPPP e cPPP. O objetivo consiste em investigar qual é o impacto dos diferentes tipos de amostragem espacial nas predições, considerando, além disso, a presença de erros de localização. As configurações do ambiente de rádio foram mantidas iguais aos casos das simulações anteriores (Figura 4.6). Ressalta-se que a incerteza de localização é incluída, com desvio padrão $\sigma_p = 70$ m. A intensidade dos processos PPP foi escolhida de modo que a quantidade média de eventos de localização na área de cobertura fosse equivalente ao caso da amostragem completamente aleatória com $N = 50$ medidas. A Figura 4.9 apresenta os resultados de desempenho, por meio das estatísticas dos preditores obtidas com o armazenamento dos valores de EMSPE, considerando 1000 realizações de simulação. Os resultados evidenciam que o padrão espacial das medidas coletadas tem impacto na acurácia das predições alcançadas, especialmente no cenário não homogêneo (caso em que os preditores foram mais afetados).

Figura 4.9: Estatísticas de EMSPE, considerando a amostragem espacial: (a) completamente aleatória (densidade de probabilidade uniforme); (b) HPPP; (c) nHPPP e (d) cPPP.

A análise dos resultados pode ser realizada à luz de três pontos importantes: i) o valor esperado (média) dos eventos de localização nos diferentes tipos de processos PPP; ii) o grau de homogeneidade (constante ou não constante) ao longo da área de cobertura e iii) a presença de erros de localização na captura de medidas do ambiente de rádio. A observação das estatísticas dos preditores permite constatar que o mesmo número médio de eventos de localização na amostragem traz resultados equivalentes para os casos uniforme, HPPP e cPPP. Entretanto, a não homogeneidade espacial do nHPPP impactou de forma significativa todos os preditores espaciais. No caso das técnicas determinísticas (IDW e NNI), aumentos da ordem de 6 dB^2 e 12 dB^2 no EMSPE ocorreram no comparativo entre os processos homogêneo e não homogêneo. O mesmo ocorreu entre as técnicas Kriging, com aumentos da ordem de 4 dB^2 e 8 dB^2 no EMSPE para os preditores KILE e KALE, respectivamente. Isto mostra que a disposição espacial dos dispositivos tem influência significativa na acurácia das previsões espaciais do REM em toda a área de cobertura.

Um ponto interessante a ser observado está nos resultados estatísticos do caso cPPP, que se aproximaram dos cenários uniforme e HPPP. A formação dos aglomerados favorece o aumento do número de medidas próximas (dentro dos aglomerados) e distantes (entre aglomerados) e, com isso, pode aprimorar as estimativas de semivariâncias experimentais. Em outras palavras, a quantidade e o espaçamento entre os aglomerados de distribuição espacial podem beneficiar a captura da variabilidade espacial do ambiente de rádio a partir das medidas, provendo melhores resultados de aprendizagem de parâmetros e previsões. De fato, as investigações de simulação indicaram que se o número de aglomerados inicializados no cPPP for pequeno, os resultados de previsão são afetados de forma significativa. No comparativo entre as técnicas, os resultados para o preditor proposto TS-KALE se mostraram promissores. Para todos os tipos de amostragem espacial, a técnica TS-KALE apresentou melhor desempenho que os preditores Kriging e as técnicas convencionais. É importante observar que a soma dos efeitos da não homogeneidade com a incerteza de localização também afetou o preditor TS-KALE, mas suas estatísticas de EMSPE (mediana, média, intervalos de primeiro e terceiro quartis, além dos *outliers*) se mostraram mais estáveis nos diferentes cenários considerados.

4.6 Conclusões - Capítulo 4

Este capítulo apresentou resultados estatísticos de desempenho dos preditores espaciais utilizados na geração do REM. A influência de diferentes aspectos do sistema de comunicação sem fio foi analisada nos resultados das previsões. Foi possível constatar que a incerteza de localização afetou de forma significativa a aprendizagem de parâmetros, comprometendo o desempenho dos preditores estatísticos (Kriging) na tentativa de capturar a variabilidade espacial das medidas. Neste cenário, o preditor proposto TS-KALE se provou mais robusto contra as adversidades do canal (*e.g.*, variações do sombreamento log-normal e erros de localização), apresentando resultados promissores, tanto no comparativo com os preditores Kriging convencionais e com as técnicas determinísticas.

Capítulo 5

Estudo Experimental da Geração do REM

Este capítulo apresenta um estudo experimental do REM, a partir de uma campanha de medidas. Dois pontos principais constituem os objetivos do estudo: i) explorar a geração do REM de forma prática e ii) testar os algoritmos de aprendizagem e predição desenvolvidos.

Inicialmente, a metodologia do estudo experimental é apresentada, bem como as informações iniciais sobre o experimento prático. Detalhes sobre os procedimentos de testes realizados durante e após os experimentos também são colocados. As iniciativas de estudos de propagação nas faixas de frequências de ondas milimétricas motivaram a realização do experimento nesta banda de frequências, considerando a frequência de 28 GHz [36], [39].

Uma análise das etapas da Fase I com a aprendizagem de parâmetros é realizada e seus resultados são discutidos. Após o treinamento do modelo de predição do REM, os algoritmos de predição espacial Kriging são aplicados sobre os dados processados. Os resultados de predição espacial da Fase II são apresentados, contemplando a visualização do REM obtido na prática. O método de validação cruzada é utilizado para que o desempenho dos preditores possa ser avaliado com a métrica empírica do erro quadrático médio de predição (EMSPE - *Empirical Mean Square Prediction Error*). Este método remove medidas específicas do conjunto coletado para a aprendizagem de parâmetros (*i.e.*, treinamento), reservando-as para o teste do modelo de predição. Desta forma, diferentes processamentos podem ser realizados para investigar a eficácia do modelo na captura da variabilidade espacial das medidas, visando melhores predições espaciais. Um experimento controlado, assumindo o conhecimento das estatísticas da incerteza de localização, permitiu introduzir erros nas coordenadas das medidas e investigar como os preditores espaciais reagiram à incorporação de incerteza de localização. Ao final, uma seção é dedicada às discussões sobre os resultados, além das limitações relacionadas com o experimento prático.

5.1 Metodologia

A Figura 5.1 ilustra a metodologia aplicada ao estudo experimental do REM. O processo tem início com a coleta de medidas do ambiente de rádio, realizado a partir de medições da potência de recepção com um equipamento analisador de espectro. As medições são executadas de forma geolocalizada e seus detalhes são explicados na próxima seção. Sobre a metodologia, o pré-processamento das medidas coletadas consiste em uma etapa importante, pois permite obter os dados que serão transferidos do instrumento de medida utilizado em campo para as etapas do modelo de predição do REM. Estas tarefas consistem no primeiro contato com as medidas, denotado como exploração dos dados. O processamento dos dados com o uso do modelo de predição espacial é a etapa central do experimento prático, envolvendo os algoritmos de aprendizagem de parâmetros (Fase I) e das predições (Fase II), para que o REM seja construído na prática.

Observando o diagrama, verifica-se que o modelo de predição é treinado na Fase I. Nesta etapa, análises sobre a tendência e os resultados geoestatísticos são essenciais na extração de informações sobre o processo aleatório espacial que gerou as medidas. O objetivo desta etapa consiste em realizar uma análise dos dados por meio de suas descrições estatísticas, buscando a estimação da covariância espacial, a partir das medições práticas. importante ressaltar que, neste momento, não há a pressuposição de incerteza de localização sobre as medições (discutida adiante). Em seguida, os algoritmos de predição Kriging são aplicados sobre os dados processados, para que o REM possa ser gerado na prática. Uma etapa de visualização do mapa é utilizada para verificar se as predições capturaram a variabilidade espacial do canal sem fio. Além disso, o método de validação cruzada, amplamente utilizado na avaliação de desempenho de predições espaciais [31], [63], [66], [78], utiliza medidas de teste para que os resultados de desempenho possam ser obtidos neste contexto prático.

Figura 5.1: Diagrama em blocos da metodologia dos experimentos.

5.2 Campanha de Medidas

A Figura 5.2 mostra a área externa de um instituto de pesquisa, local onde a campanha de medidas foi conduzida. O local engloba uma área de $100 \text{ m} \times 60 \text{ m}$, composta por uma pequena vegetação cercada por prédios e grandes árvores ao redor. Os procedimentos de testes e medidas foram realizados durante o dia e sem intensa movimentação no campus. Uma análise de sinais foi realizada em cada ponto de coleta de medidas, considerando as conexões de RF, o correto apontamento entre as antenas e a verificação das condições de recepção no espectro de frequências (visando a constatação de possíveis interferências). É importante ressaltar que cada medida deve ser executada de forma adequada para que o experimento seja controlado, a fim de formar a base de dados necessária para o processamento dos dados e a realização das previsões espaciais.

Dado a complexidade dos procedimentos, especialmente em razão da frequência de operação, um conjunto de 28 locais de medição foi considerado dentro da área de cobertura analisada. Em cada local ou ponto de medição, um tempo de varredura foi programado em um analisador de espectro para armazenar 1000 valores de potência de recepção. A Tabela 5.1 mostra os detalhes sobre os parâmetros técnicos relacionados com os sistemas de transmissão e recepção utilizados, enquanto a Figura 5.3 traz as fotos dos sistemas de transmissão e recepção. O gerador de sinais de RF Keysight (EXG N5173B), que opera até a frequência de 40 GHz, e uma antena omnidirecional compõem o sistema de transmissão, enquanto o sistema de recepção é composto por uma antena direcional do tipo corneta e um analisador de espectro FieldFox, capaz de analisar sinais até a frequência de 50 GHz. Os suportes utilizados nas antenas possuem a mesma altura (1,70 m). Sobre os posicionamentos, o sistema de transmissão é colocado em um local mais elevado (uma passarela a aproximadamente 5,70 m de altura), enquanto a recepção é feita a altura de 1,70 m ao longo da área externa do campus. Os mesmos tipos de cabos de RF, que operam em 28 GHz, são utilizados para as conexões entre o gerador de RF e a antena omnidirecional na transmissão e entre o analisador de espectro e a antena corneta na recepção.

Tabela 5.1: Parâmetros e Configurações do Experimento Prático

Parâmetros	Sistema de Transmissão	Sistema de Recepção
Equipamentos	Gerador de RF Keysight	Analisador de Espectro FieldFox
Antenas	Omnidirecional (28 GHz)	Corneta (28 GHz)
Ganho das antenas	12.2 dBi	12.6 dBi
Ângulos de largura de feixe	—	19° (V) e 20.7° (H)
Altura sobre o solo	5,70 m	1,70 m
Perda nos cabos de RF	15,84 dB	15,84 dB
Potência de transmissão	+ 19 dBm	—

Figura 5.2: Área externa onde foi realizada a campanha de medidas.

Figura 5.3: Sistemas de transmissão e recepção utilizados na campanha de medidas.

Os procedimentos de medidas foram conduzidos de forma geolocalizada durante o experimento. Conforme mencionado, cada local de medição é caracterizado por uma distribuição de 1000 medidas de potência de recepção. A Figura 5.4 mostra os locais de medição selecionados na campanha e a fonte de geolocalização do instrumento de medida (GPS acoplado ao analisador de espectro). Com isso, a base de dados obtida a partir do instrumento de medida é composta pelos pares de coordenadas (latitude, longitude) associados à cada distribuição de valores de potência de recepção capturada nos 28 locais de medição da campanha. Nota-se que algumas medidas foram realizadas próximo ao ponto de transmissão (Tx), enquanto outras foram capturadas em pontos mais distantes e obstruídos, impactando suas condições de recepção (*e.g.*, atrás dos prédios do campus e em árvores mais afastadas no perímetro da área externa analisada).

A Figura 5.5 apresenta alguns resultados estatísticos relacionados com um local de medição em condição de visada direta (LOS - *Line of Sight*), relativamente próximo do sistema de transmissão. O aspecto do histograma obtido sugere um comportamento normal (Gaussiano) para a distribuição dos valores de potência de recepção capturados neste local. Visando a complementação desta análise o teste de aderência de Anderson-Darling (*P-Value*) é aplicado sobre o conjunto de medidas, permitindo a verificação estatística da normalidade dos dados [134-135]. O alinhamento dos pontos com a linha tracejada do gráfico mostra a aderência das medidas com a distribuição Gaussiana. O valor de $P\text{-Value} = 0.441$ indica que o teste de Anderson-Darling falhou ao rejeitar a hipótese de que os dados não são normais (*i.e.*, há grande probabilidade das medidas serem Gaussianos neste caso).

Figura 5.4: Instrumento de medida com geolocalização e pontos espaciais de medições do experimento.

Figura 5.5: Resultados estatísticos relacionados com as medidas em condições de recepção LOS.

A série temporal de medidas mostra como ocorrem as flutuações em torno do valor médio calculado ($-88,85$ dBm) para a distribuição de valores capturada. Neste caso, as variações da potência de recepção medidas não são predominantemente causadas por intensas reflexões de múltiplos percursos, mas principalmente pelos efeitos sombreamento do canal sem fio, em função da presença dos prédios e árvores ao redor. Diversas varreduras de medições foram realizadas atestando o mesmo comportamento neste local de medição específico.

De outro modo, durante o experimento foi possível observar variações mais intensas da potência de recepção, especialmente nos pontos de medidas mais afastados ou em meio a locais com obstruções. Um exemplo pode ser observado na Figura 5.6, em que as variações de nível de recepção ocorrem de forma mais abrupta dentro do período de varredura do instrumento em uma condição NLOS (*Non Line-of-Sight*), sem visada direta com o sistema de transmissão. De forma similar ao caso anterior, inúmeras varreduras foram realizadas neste local específico a fim de verificar a presença de variações mais significativas do conjunto de medidas coletado. A constatação de que se trata de um ponto crítico de recepção é refletida no histograma obtido, que já não apresenta um aspecto similar ao caso anterior (Figura 5.5) e não sugere um comportamento Gaussiano para as medidas. Além disso, a falta de alinhamento das medidas no gráfico do teste de aderência também reforça esta análise, indicando que, neste ponto específico de medida, os dados não seguiram a distribuição Gaussiana (*i.e.*, não tiveram aderência significativa à densidade de probabilidade Gaussiana).

Figura 5.6: Resultados estatísticos relacionados com as medidas em condições de recepção NLOS.

A relevância desta análise ocorre em função do uso de estimadores que são baseados em pressuposições sobre a estatística Gaussiana dos dados, (*e.g.*, técnicas ML). Como consequência, os resultados de estimação e predição podem ser impactados em razão das diferenças na distribuição estatística das medidas. Neste contexto, o problema de estimação do valor médio da potência de recepção em uma localidade (também conhecido como média local) já é alvo de pesquisa na literatura [136-139]. O ponto crítico reside no processamento adequado das medições em uma localidade no sentido de suavizar os efeitos dos múltiplos percursos e, ao mesmo tempo, manter as variações lentas relacionadas com o somreamento do canal sem fio. Tal processamento consiste no cálculo médio das amostras do sinal capturadas em uma janela temporal ou espacial e que pode ser baseado em abordagens analíticas¹ ou empíricas². Neste trabalho, visando a investigação dos efeitos da variabilidade das medidas nos estimadores e preditores, somente a etapa de filtragem de recepção do instrumento de medidas é habilitada (configuração de recepção), ou seja, nenhum tipo de pós filtragem é aplicado sobre os dados coletados.

¹Um estudo analítico realizado a partir das distribuições estatísticas dos múltiplos percursos (*i.e.*, *Rayleigh* e *Rice*) determinou o comprimento espacial apropriado (da ordem de 40 vezes o comprimento de onda dos sinais transmitidos) e que vem sendo interpretado como uma recomendação acerca do tópico [33], [41], [139].

²Diversos trabalhos que realizam campanhas de medidas implementam diferentes soluções empíricas (*i.e.*, procedimentos de medições práticos), com diferentes janelas espaciais e temporais para suavizar os efeitos dos múltiplos percursos [35], [54], [99].

5.3 Resultados da Fase I - Aprendizagem de Parâmetros

Para que o REM seja construído com as medidas práticas, é necessário que um valor representativo da potência de recepção esteja associado a cada local de medição. Especificamente, o valor médio (em dBm) de cada distribuição capturada (nos 28 locais de medição) é calculado e elencado como a medida de potência de recepção nas respectivas localidades. Este processamento é feito somente para que o vetor de potência de recepção ao longo do espaço, $\mathbf{p}$, seja obtido e o treinamento do modelo de predição possa ser iniciado, visando a construção do REM.

Desta forma, a estimação da potência média de recepção é discutida na subseção da tendência. Em seguida, as ferramentas geoestatísticas do modelo de predição espacial são aplicadas, para que uma análise sobre os resultados práticos de estimação dos parâmetros de covariância seja realizada.

5.3.1 Análise da Tendência

O tratamento da tendência neste estudo experimental é caracterizado por duas etapas; primeiramente, a regressão do modelo log-distância (tendência), que é realizada para que o coeficiente de propagação do modelo seja estimado, a partir das medidas práticas. Com isso, a remoção da tendência permite a extração dos valores regionalizados, que somente representam as variações espaciais de potência de recepção que ocorrem em razão do sombreamento do canal sem fio. Desta modo, os preditores espaciais podem ser aplicados sobre os valores regionalizados, para que a potência média de recepção seja posteriormente incorporada às estimativas, compondo o resultado final de predição. O modelo da tendência, descrito anteriormente na subseção 2.2.1, é obtido pela combinação entre as equações 2.15 e 2.17, de acordo com

$$\mu(\mathbf{s}) = P_{\text{tx}} - \bar{\bar{P}}_L(d_0) - 10\alpha \log \left(\frac{d(\mathbf{s}_{\text{tx}}, \mathbf{s})}{d_0} \right). \quad (5.1)$$

A constante $\bar{\bar{P}}_L(d_0)$ expressa a perda por percurso média a uma distância de referência d_0 , que é associada à região de campo distante das antenas utilizadas nos experimentos [95-97]. O valor de $\bar{\bar{P}}_L(d_0)$ requer a seleção adequada da distância de referência, ou seja, como sendo maior do que a distância de campo próximo das antenas. Especificamente, $\bar{\bar{P}}_L(d_0)$ pode ser obtido de forma empírica, a partir de medições tomadas relativamente próximas ao transmissor [97]. Durante os experimentos, diversas medições foram realizadas (considerando $d_0 = 1$ m), permitindo a obtenção de um valor empírico de potência de recepção, $P_{\text{rx}} = -45$ dBm. Aplicando este valor no modelo 5.1, com $P_{\text{tx}} = +19$ dBm e $d(\mathbf{s}_{\text{tx}}, \mathbf{s}) = d_0$, tem-se $-45 = +19 - \bar{\bar{P}}_L(d_0) - 10\alpha \log(d_0/d_0)$. Assim, é possível determinar o resultado da perda por percurso média na distância de referência³, $\bar{\bar{P}}_L(d_0) = 64$ dB.

³Na ausência de informações sobre d_0 , o valor de $\bar{\bar{P}}_L(d_0)$ pode ser aproximado como a perda por percurso no espaço livre, considerando $d_0 = 1$ m [96-97]. A equação da perda por percurso no espaço livre (*i.e.*, $20 \log_{10} 4\pi d_0/(c/f)$), em que c corresponde à velocidade da luz, $f = 28$ GHz e $d_0 = 1$ m), resulta em um valor $\bar{\bar{P}}_L(d_0) = 61.38$ dB, relativamente próximo ao obtido de forma empírica através das medições.

A obtenção do valor empírico para $\bar{P}_L(d_0)$ consiste em uma forma de calibração do modelo da tendência. Conforme mencionado, um GPS é acoplado ao analisador de espectro e, com isso, as coordenadas obtidas nos locais de medição consistem em pares de latitude e longitude. Assim, devido diferença de altura entre o transmissor (posicionado em uma passarela a 5,70 m) e os locais de medição (com um suporte para antena a 1,70 m) nas distâncias envolvidas, os pares de coordenadas de latitude e longitude foram transformados para o sistema cartesiano, com $\mathbf{s} = (s_x, s_y, s_z)$, para verificar o impacto da altura⁴ no cálculos das distâncias, realizado de acordo com

$$d_{A-B} = \sqrt{(s_{xA} - s_{xB})^2 + (s_{yA} - s_{yB})^2 + (s_{zA} - s_{zB})^2}, \quad (5.2)$$

em que d_{A-B} consiste na distância entre as coordenadas espaciais (s_{xA}, s_{yA}, s_{zA}) e (s_{xB}, s_{yB}, s_{zB}) . Detalhes sobre a transformação dos pares de coordenadas de latitude e longitude para o sistema cartesiano podem ser acessados em [85, Apêndice C].

Com o modelo da tendência calibrado através de $\bar{P}_L(d_0)$ e o cálculo das distâncias com (5.2), é possível iniciar a regressão do modelo em (5.1), visando a estimação do coeficiente de propagação α . O estimador $\hat{\alpha} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{p}$ é aplicado sobre as medidas de potência de recepção e seu resultado, $\hat{\alpha}$, é utilizado no modelo da tendência para que os valores regionalizados possam ser recuperados. A Figura 5.7 mostra os resultados com (a) a regressão do modelo em conjunto com as medidas de perda por percurso observadas e (b) o histograma dos valores regionalizados.

Figura 5.7: Tratamento da tendência: (a) regressão do modelo log-distância e (b) histograma dos valores regionalizados, após a remoção da tendência.

⁴Embora não tenha tido um impacto alterando os resultados de estimação de forma significativa, a diferença de altura foi levada em conta, pois influencia diretamente as distâncias calculadas utilizadas nos algoritmos de estimação. Ainda assim, ambos os cálculos de distâncias, baseados em latitudes e longitudes e na transformação para o sistema cartesiano, são incorporados aos algoritmos das etapas do modelo de predição espacial do REM. Além disso, a sua validação é realizada comparando-se os resultados de distância obtidos com as distâncias medidas em *softwares* específicos de mapeamento de imagens de satélite (*e.g.*, *Google Earth* e *Google Maps*).

O resultado da estimativa, $\hat{\alpha} = 3, 4$, consiste em um valor considerado plausível para o parâmetro α no cenário de propagação do experimento. Contudo, é possível notar as diferenças no histograma obtido em relação à distribuição Gaussiana dos valores regionalizados, relacionada com o sombreamento log-normal. Esta diferença ocorre em razão dos efeitos provocados pelo canal sem fio nas medidas coletadas. Em outras palavras, o histograma estatístico com aspecto Gaussiano seria obtido se as medidas capturadas fossem afetadas somente pelo sombreamento log-normal do canal sem fio. Entretanto, conforme mostrado na Figura 5.6, vários locais de medição são caracterizados por valores que não têm aderência estatística com a distribuição Gaussiana. Este resultado é esperado, uma vez que nenhum tipo de pós-processamento foi aplicado sobre as medidas para suavizar os efeitos do canal em pequena escala e, de fato, as medidas de potência de recepção são influenciadas, de forma conjunta, por diferentes efeitos em grande e pequena escala. De certa forma, isto requer a robustez dos algoritmos de estimação em função de alterações nas estatísticas das medidas capturadas. Além disso, outro fator que impacta o histograma obtido é no número relativamente reduzido de locais de medição, influenciando o resultado de $\hat{\alpha}$ e, conseqüentemente, a remoção da tendência.

5.3.2 Análise Geoestatística

A análise geoestatística é realizada com o processamento do vetor de valores regionalizados, $\mathbf{z} = \mathbf{p} - \hat{\boldsymbol{\mu}}$, obtido com o tratamento da tendência, que é essencial para as próximas etapas do modelo de predição espacial (*i.e.*, estimação do semivariograma $\hat{\gamma}$ e do covariograma $\hat{C}$ experimentais). Sobre o funcionamento destas etapas, o algoritmo de estimação do semivariograma/covariograma, descrito na subseção 2.3.2, é aplicado sobre o vetor $\mathbf{z}$ extraído das medidas práticas $\mathbf{p}$. Com isso, o vetor de parâmetros $\boldsymbol{\theta}$ pode ser estimado por meio do ajuste dos modelos analíticos do semivariograma/covariância, concretizando o treinamento do modelo para a geração das predições.

A Figura 5.8 mostra os resultados da análise geoestatística, relacionada com a fase de aprendizagem de parâmetros do ambiente de rádio. É importante mencionar que as distribuições de potência de recepção coletadas nos pontos de medições são utilizadas nesta etapa, para que os histogramas estatísticos possam ser obtidos. Especificamente, o vetor $\mathbf{z}$, de 28 medidas de potência de recepção (ao longo do espaço), é processado pelos métodos geoestatísticos em um algoritmo que é iterado 1000 vezes, em razão das distribuições de valores de potência de recepção coletadas nos locais de medição. Isto permite armazenar as estimativas do vetor $\boldsymbol{\theta}$ para análise estatística dos resultados.

As Figs.5.8 (a) e (b) mostram os histogramas relacionados com as estatísticas da variância (*sill*), $\hat{m} = \hat{\sigma}^2$, e do parâmetro *range*, $\hat{r} = d_{\text{corr}}$, respectivamente ($c_0 = 0$, em função do interesse específico nos efeitos em largura escala). Os resultados dos histogramas indicam que, na média, a variância *sill* foi estimada com $\hat{\sigma}^2 = 114.09$, ou seja, as variações provocadas pelo sombreamento do canal sem fio são quantificadas pelo desvio padrão $\hat{\sigma} = 10.6$ dB. O parâmetro *range* foi estimado, na média, com $\hat{r} = 15,38$ m e quantifica a distância de decorrelação espacial do sombreamento do canal sem fio.

As Figs.5.8 (c) e (d) apresentam o semivariograma e o covariograma experimentais, associados à uma única realização das medidas capturadas. Nota-se que o aspectos de $\hat{\gamma}$ e $\hat{C}$ sugerem um tipo de comportamento exponencial para a covariância espacial estimada. As curvas sólidas mostradas consistem nos modelos analíticos do semivariograma e da covariância, ambos obtidos partir da estimativa $\hat{\theta}$. Diferente do semivariograma, que é baseado em diferenças quadráticas, é possível verificar estimativas negativas para as covariâncias experimentais (em função dos produtos calculados entre os valores regionalizados recuperados). Ainda assim, é importante mencionar que os modelo analíticos, relacionados com $\hat{\gamma}$ e $\hat{C}$, fornecem resultados positivos, em função da característica definida positiva requerida na formação e solução do sistema de covariâncias do preditor Kriging. A inclinação existente no semivariograma em (c) e o decaimento do covariograma em (d) permitem observar até quais distâncias existe uma significativa correlação espacial entre as medidas. Nota-se a formação de patamar nas estimativas de semivariâncias, que está relacionado com a variância *sill* estimada. Outro ponto importante é a finalização do decaimento do covariograma, que tal como a inclinação do semivariograma, ocorre em distâncias lags da ordem de 15 m. A observação deste comportamento e o resultado das estimativas obtidas (na média), $\hat{\theta} = [\sigma^2 \ r] = [114.09 \ 15.38]$, são fatores importantes, pois permitem a extração de informações sobre o ambiente de rádio, além de serem essenciais na geração das previsões.

Figura 5.8: Resultados da análise geoestatística: (a) histograma da variância *sill*; (b) histograma do parâmetro r ; (c) semivariograma e (d) covariograma

5.4 Resultados da Fase II - Predições do REM

O uso dos modelos analíticos na aprendizagem de parâmetros é fundamental para a formação do sistema matricial de semivariâncias do preditor Kriging ($\mathbf{\Gamma}\boldsymbol{\lambda}_\nu = \boldsymbol{\beta}$), mostrado na subseção 2.3.5,

$$\begin{bmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_1) & \dots & \gamma(\mathbf{s}_1, \mathbf{s}_N) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_1) & \dots & \gamma(\mathbf{s}_N, \mathbf{s}_N) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ -\nu \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_N, \mathbf{s}_0) \\ 1 \end{bmatrix}. \quad (5.3)$$

No caso da geração do REM prático, a aplicação dos modelos analíticos na composição da matriz e dos vetores em (5.3) é fundamentada nos cálculos de distâncias entre as coordenadas espaciais, obtidas com as medidas geolocalizadas. Os algoritmos de predição foram adaptados para ambos os sistemas de coordenadas, ou seja, considerando os pares de latitude e longitude e as coordenadas no sistema cartesiano. Assim, o uso das informações de localização, a precisão do instrumento de medida e o processamento geoestatístico dos dados coletados são aspectos importantes para que a solução do sistema matricial em (5.3) resulte nos pesos ótimos, $\boldsymbol{\lambda}_\nu = \mathbf{\Gamma}^{-1}\boldsymbol{\beta}$, na busca pela geração de predições espaciais com maior acurácia.

Desta forma, as coordenadas espaciais que são alvos da predição $\mathbf{s}_0$ são selecionadas a fim de formar uma zona espacial, composta por um conjunto de coordenadas que engloba toda a área da campanha de medidas. Neste sentido, os algoritmos de predição espacial são aplicados para construir o mapa de ambiente de rádio sobre a zona espacial do local do experimento. Em cada coordenada espacial $\mathbf{s}_0$, a predição da potência de recepção é composta pela estimação da tendência $\hat{\mu}(\mathbf{s}_0)$ e pelo resultado de predição Kriging $Z_{KO}(\mathbf{s}_0)$, de acordo com

$$\hat{P}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + Z_{KO}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + \sum_{i=1}^N \lambda_i Z(\mathbf{s}_i). \quad (5.4)$$

Os resultados de predição espacial, colocados em função das coordenadas de latitude e longitude, podem ser observados na Figura 5.9. O REM obtido na prática (c) consiste na combinação entre o mapa da potência média de recepção (a), relacionado com a tendência, e o mapa Kriging de predição do sombreamento do canal sem fio (b). O mapa da variância Kriging em (d) indica qual o grau de incerteza associado às predições nas diversas coordenadas espaciais. Algumas observações importantes podem ser colocadas; primeiramente, nota-se uma diferença visível entre o mapa de predição da potência média de recepção (a), obtido a partir da regressão do modelo log-distância (tendência), e o REM prático em (c), que inclui as variações espaciais das predições de sombreamento com o preditor Kriging. É possível observar que os algoritmos de predição buscam acompanhar as variações espaciais da potência de recepção em função das características do ambiente de rádio.

Figura 5.9: Resultados de predição: (a) mapa de predição da potência média de recepção; (b) mapa de predição espacial Kriging do sombreado log-normal; (c) REM prático; (d) mapa da variância Kriging.

Para complementar a análise dos resultados das predições espaciais, a Figura 5.10 apresenta os contornos de nível do REM prático sobre o local da campanha de medidas. A posição do transmissor do experimento (Tx) também é mostrada juntamente com os contornos traçados. É possível observar que as predições da potência de recepção com menor intensidade (menores que - 100 dBm) ocorrem em regiões com obstruções como prédios e árvores. Verifica-se que os resultados de predição buscam capturar a presença destas obstruções e, de fato, tal captura foi eficaz nas regiões próximas aos locais onde as medidas foram coletadas. Além disso, nota-se que os valores mais altos das predições de potência de recepção (maiores que -100 dBm) ocorrem próximos ao transmissor e nas áreas mais abertas do campus. Entretanto, é importante observar que, como a coleta de medidas foi realizada no ambiente externo às construções, existem predições de potência com intensidade relativamente alta dentro de prédios, por exemplo. Isto auxilia a coleta de medidas para novas campanhas, de acordo com a área de cobertura desejada para as predições. Neste sentido, a visualização do REM prático e a sua interpretação sobre o local do experimento são pontos relevantes para análise e trazem resultados plausíveis em um primeiro momento. Dentro deste contexto, dois pontos do experimento devem ser discutidos em função da metodologia utilizada: i) a forma como o desempenho das predições espaciais do REM pode ser estimada e ii) como a introdução da incerteza de localização pode ser realizada no experimento prático.

Figura 5.10: Visualização do REM prático no local do experimento da campanha de medidas.

Sobre o ponto i), a avaliação da acurácia das predições obtidas requer uma análise de desempenho dos resultados da predição. Isto é realizado por meio do método de validação cruzada, permitindo acessar o desempenho das predições, a partir das medidas práticas coletadas. Especificamente no algoritmo de predição, um valor de potência de recepção é removido do conjunto de medidas coletadas ao longo do espaço ($N = 28$) e utilizado para teste⁵, enquanto os valores restantes (27 medições) são usados para a aprendizagem de parâmetros (*i.e.*, treinamento do modelo de predição do REM). A distribuição dos valores de potência de recepção armazenada em cada um dos locais de medição permite fazer 1000 realizações do algoritmo de validação cruzada. Com isso, o erro quadrático médio de predição empírico (EMSPE) pode ser estimado, de acordo com

$$\text{EMSPE} = \frac{1}{1000 \times N} \sum_{k=1}^{1000} \sum_{i=1}^N [\hat{P}_k^i(\mathbf{s}_0) - P_k^i(\mathbf{s}_0)]^2 \quad (5.5)$$

em que N é o número de medidas ao longo do espaço, $\hat{P}_k^i(\mathbf{s}_0)$ consiste no resultado de predição, relacionado com a i -ésima medida referente aos locais de medição, que é comparado com a potência de recepção $P_k^i(\mathbf{s}_0)$, associada à coordenada alvo da predição, $\mathbf{s}_0$, utilizada para teste. Este processamento é realizado a cada k -ésima realização do algoritmo de predição espacial para que o EMSPE possa ser estimado.

⁵Embora a subdivisão do conjunto de dados em blocos de teste seja comum na prática, a validação cruzada é realizada com cada uma das 28 medidas (técnica conhecida como *Leave-one-out* [66]), devido ao número relativamente pequeno de medidas capturadas ao longo do espaço.

Sobre o ponto ii), a introdução dos erros de localização nas medidas é realizada com um experimento controlado, pois se apresenta de forma mais complexa em função de algumas razões. A principal razão consiste na incerteza intrínseca já existente no experimento, devido ao uso do GPS acoplado ao instrumento de medida. Embora o analisador de espectro seja um instrumento de alta precisão, as medidas geolocalizadas possuem precisão finita. Com isso, este experimento controlado considera duas pressuposições importantes: i) assume-se que as coordenadas espaciais de latitude e longitude capturadas formam o conjunto de localizações mais próximo do verdadeiro posicionamento das medições e ii) assume-se o conhecimento *a priori* da distribuição $g(\cdot)$, considerando erros de localização espacialmente descorrelacionados, *i.e.*, $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$.

As principais referências deste trabalho utilizam a distribuição Gaussiana para o modelo estatístico dos erros de localização $\mathbf{p}_e(\mathbf{s})$ (*e.g.*, em [80], argumenta-se sobre a compatibilidade deste modelo com vários sistemas de localização e navegação, enquanto [79] utiliza a distribuição Gaussiana baseada em medições práticas de erros de localização do sistema GPS) [74-84]. Contudo, é importante mencionar que a estimação de $g(\cdot)$ se torna necessária na prática⁶, não somente em virtude do aspecto da distribuição estatística, mas também da possível correlação espacial entre os erros de localização. No caso da tecnologia GPS previamente mencionado, o número e a geometria formada entre os satélites são dependentes do posicionamento e afetam as estimativas de localização, implicando na dependência espacial e na característica sistemática dos erros de localização. Algumas observações importantes sobre este tópico podem ser acessadas em [80], [82], embora um modelo específico para erros de localização sistemáticos e espacialmente correlacionados não seja colocado. De fato, a formulação de um modelo que descreve a correlação espacial de $\mathbf{p}_e(\mathbf{s})$ ao longo do espaço não é trivial, mas consiste em um aspecto realístico, sobretudo em razão das diferenças no funcionamento (e o consequente desempenho) dos diversos métodos de localização utilizados nas aplicações dos sistemas de comunicações sem fio (*e.g.*, métodos baseados em redes de telefonia móvel, tecnologia GPS/A-GPS e sistemas de localização em redes de sensores [74]).

Na análise das medidas práticas, constatou-se que a introdução de erros de localização (com distribuição Gaussiana) diretamente nas coordenadas de latitudes e longitudes provocou diferenças expressivas no posicionamento das coordenadas, requerendo uma faixa de valores adequada para σ_p . De fato, valores maiores de σ_p , se utilizados diretamente sobre as latitudes e longitudes, podem provocar erros nas primeiras casas decimais das coordenadas, deslocando-as para regiões distantes (muito além da área externa do campus ou até mesmo fora da cidade). Para solucionar este problema, o conjunto de latitudes e longitudes foi transformado para o sistema de coordenadas cartesianas, permitindo o uso de valores mais representativos para σ_p , dado em metros.

⁶A utilização de *softwares* baseados no mapeamento de imagens de satélite (*e.g.*, *Google Earth*), bem como o uso de dispositivos de localização com altíssima precisão, são ferramentas para indicação exata das coordenadas espaciais que podem dar suporte na estimação da distribuição dos erros de localização na prática. Além disso, diversos trabalhos nas áreas de mapeamento geográfico e cartografia geodésia investigam a precisão das imagens de satélite de alta resolução e apontam a plausibilidade do seu uso para diversas aplicações específicas que requerem alta precisão [140-141].

A Figura 5.11 mostra as coordenadas de latitude e longitude obtidas com o GPS em conjunto com as coordenadas espaciais afetadas pela incerteza de localização, introduzida de forma controlada no experimento através do modelo selecionado, $\mathbf{p}_e(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$. Ícones diferentes são utilizados para distinção das medidas sem e com os erros de localização. Neste exemplo, o desvio padrão foi configurado com $\sigma_p = 6$, provocando um erro, na média, de 6 metros entre as coordenadas espaciais do GPS e as corrompidas pela incerteza de localização.

Uma simulação é desenvolvida para que o desempenho dos preditores Kriging seja estimado, considerando a incerteza de localização introduzida no experimento. Ressalta-se que, embora as medidas de potência de recepção do ambiente de rádio sejam práticas, o modelo de incerteza é conhecido *a priori* e utilizado para testar os algoritmos de aprendizagem e predições. De acordo com a métrica em (5.5), a distribuição dos valores de potência de recepção armazenada e o método da validação cruzada aplicado sobre as 28 medidas ao longo do espaço, permitem fazer 1000 realizações de simulação, para que o EMSPE possa ser estimado. Os erros de localização $\mathbf{p}_e(\mathbf{s})$ com distribuição Gaussiana são introduzidos nas coordenadas de forma gradual por meio do aumento do desvio padrão na faixa $0 \text{ m} \leq \sigma_p \leq 9$ metros. Em cada realização, o treinamento do modelo de predição é conduzido com a aprendizagem de parâmetros e, em seguida, os preditores KILE, KALE e TS-KALE utilizam as medidas e os parâmetros estimados para realizar as predições espaciais.

Figura 5.11: Visualização das coordenadas espaciais sem e com os erros de localização.

Os resultados podem ser observados na Figura 5.12, que mostra: (a) as estatísticas dos erros quadráticos de predição Kriging na forma de histograma, sem a introdução dos erros de localização nas medidas práticas e (b) as curvas de desempenho (EMSPE) dos preditores KILE, KALE e TS-KALE para diferentes valores de σ_p . No histograma em (a), é possível notar a extensão na distribuição dos valores em função das realizações do algoritmo sendo que, na média, o erro quadrático de predição espacial foi estimado em 68.15 dB^2 . Já o aumento no EMSPE dos preditores em razão da incerteza de localização pode ser visualizado nas curvas em (b), destacando o melhor desempenho alcançado pelo TS-KALE frente aos preditores KILE e KALE. É possível visualizar diferenças da ordem de 7 dB^2 a 10 dB^2 entre os preditores KILE e TS-KALE. As investigações com a simulação também mostraram que valores mais elevados para σ_p (*e.g.*, maiores que 15 metros) afetam o semi-variograma e a estimação de θ de forma significativa, especialmente em razão do número limitado de medidas ao longo do espaço.

Estes resultados práticos corroboram as principais referências do trabalho (*e.g.*, [78-79], [83-84], [91-93]) no sentido de apontar os benefícios alcançados quando as estatísticas da incerteza de localização são utilizadas a favor das predições. Além disso, as investigações sobre o experimento prático e o processamento das medidas permitiram a análise de pontos importantes sobre o estudo experimental. Estes pontos consistem em aspectos do modelo de predição, que tiveram impacto nos resultados de predição, e limitações do próprio experimento prático: i) estimação da distribuição $g(\cdot)$ dos erros de localização na prática; ii) a quantidade de locais de medições e a distribuição espacial das medidas capturadas; iii) a proximidade da distribuição estatística dos dados ao caso Gaussiano e iv) o desempenho de treinamento do modelo de predição espacial.

Figura 5.12: Resultados de desempenho: (a) histograma dos erros quadráticos de predição e (b) curvas de EMSPE dos preditores KILE, KALE e TS-KALE, obtidas a partir das medidas práticas.

O primeiro ponto está relacionado com a estimação da distribuição dos erros de localização. Embora o conhecimento prévio da densidade de probabilidade $g(\cdot)$ seja considerado em diversas referências [78-82], a estimação de $g(\cdot)$ consiste em um aspecto plausível na prática. Ainda assim, é importante mencionar que a análise estatística das estimativas de localização em um cenário prático de propagação não é uma tarefa trivial, sobretudo em função dos diferentes tipos de métodos de localização. Isto reforça a busca por técnicas de estimação a serem aplicadas neste problema, além de modelos estatísticos realísticos para a incerteza de localização em diferentes aplicações.

O segundo ponto é dado pela forma como as medidas são capturadas, *i.e.*, amostragem espacial. Experimentações com os dados coletados, como a inclusão e a remoção de medidas, permitiram verificar as flutuações nos resultados de desempenho, reforçando as conclusões sobre os impactos da amostragem espacial discutidos no capítulo 4. De fato, a remoção dos dados de um local de medição consiste na marginalização de uma das dimensões do processo aleatório Gaussiano multivariado, usado para modelar o comportamento das variações da potência de recepção no ambiente de rádio, enquanto a inclusão de medidas incorpora mais informação para o modelo de predição.

A discussão entre a aprendizagem e as predições leva aos pontos iii) e iv). Sabe-se que a propriedade da linearidade da regressão de distribuições Gaussianas multivariadas faz com que as equações do Kriging convirjam para a esperança condicional, demonstrado em 2.35, considerando a Gaussianidade do ambiente de rádio [63, pp. 638]. Com isso, a proximidade estatística dos dados ao caso Gaussiano contribui para a minimização do MSPE. Esta questão toma relevância, uma vez que diversas técnicas estimação se baseiam em pressuposições estatísticas dos dados. A investigação prática reforça a inferência de que o comportamento não Gaussiano dos dados tem influência na aprendizagem de parâmetros, impactando as predições. Assim, a questão chave para o Kriging está na captura da variabilidade espacial da potência de recepção, a fim de encontrar alguma estrutura nos dados (*i.e.*, covariância), permitindo que predições com maior acurácia sejam alcançadas.

5.5 Conclusões - Capítulo 5

Este capítulo complementou as investigações deste trabalho, apresentando um estudo experimental do REM. A aplicação do modelo de predição espacial possibilitou a geração do REM na prática, enquanto as análises da aprendizagem de parâmetros permitiu que aspectos sobre o modelo de predição e do experimento prático fosse discutidos. A introdução de erros de localização com um modelo estatístico previamente conhecido também permitiu que os algoritmos de predição fossem testados com medidas práticas. O desempenho dos preditores Kriging foi estimado, e resultados promissores foram obtidos a favor do TS-KALE. O aprofundamento do estudo prático mostrou que o uso das estatísticas dos erros de localização faz sentido no momento em que a incerteza de localização é suficientemente elevada a ponto de diferenciar a covariância entre as medidas das coordenadas espaciais estimadas em relação à covariância entre as medidas em suas verdadeiras localizações.

Capítulo 6

Conclusões e Perspectivas

Esta tese de doutorado utilizou o preditor espacial Kriging da teoria de processos Gaussianos para obter o REM de um sistema de comunicação sem fio. Um modelo de predição foi formulado, a partir de uma abordagem geoestatística aplicada ao cenário das redes de comunicações sem fio. Os problemas da incerteza de localização dos dispositivos e a falta de conhecimento sobre os parâmetros do canal sem fio foram introduzidos e permitiram tornar o modelo mais realístico. Com isso, a análise dos impactos efetivos dos problemas considerados permitiu concluir que a regressão da tendência do ambiente de rádio, bem como a estimação dos parâmetros do canal sem fio, desempenham um papel importante na geração confiável do REM, especialmente quando há erros de localização nas medidas coletadas no sistema de comunicação sem fio.

Dentro das investigações sobre como combater os efeitos negativos provocados pela incerteza de localização no REM, uma técnica de predição espacial é proposta. Os algoritmos desenvolvidos, baseados na estimação ajustada da tendência (TALE) e do semivariograma (SALE), permitiram formular o preditor TS-KALE, que apresentou resultados promissores frente às variantes dos métodos Kriging (KILE e KALE) e as técnicas convencionais de predição (*e.g.*, IDW e NNI) utilizadas na geração do REM. Com isso, argumenta-se sobre dois pontos principais do TS-KALE: i) a sua flexibilidade e ii) a sua simplicidade em função das etapas onde o procedimento de ajuste foi realizado. O primeiro ponto i) ocorre diretamente em razão do uso das ferramentas da geoestatística, que são adequadas à captura de medidas geolocalizadas e que viabilizaram a introdução das estatísticas dos erros de localização em mais etapas do modelo de predição. O ponto ii) está relacionado à aplicação do ajuste nas etapas não envolvidas diretamente na estimação de θ , evitando o aumento significativo da complexidade. Ainda assim, é importante mencionar que não se pode alegar a otimalidade (em algum sentido estatístico) dos resultados de estimação em função técnicas utilizadas (MoM e LS). Desta forma, o potencial do TS-KALE reside em sua abordagem geoestatística e na melhor acurácia alcançada frente a outros preditores, considerando o mesmo treinamento do modelo de predição e submetido aos problemas de incerteza de localização e estimação de parâmetros.

Um estudo experimental foi incorporado a pesquisa e o REM foi obtido de forma prática. O uso de ferramentas da geoestatística possibilitou a aprendizagem de parâmetros do ambiente de rádio, a partir das medidas práticas coletadas. Assim, os algoritmos de predição desenvolvidos no trabalho foram aplicados, possibilitando a visualização do REM no local do experimento. Os resultados obtidos com o estudo experimental indicaram que a abordagem geoestatística do modelo foi útil na exploração da covariância espacial das medidas geolocalizadas a favor das predições.

A introdução da incerteza de localização foi implementada através de um experimento controlado, assumindo o conhecimento prévio das estatísticas dos erros de localização. As investigações feitas com as simulações computacionais, envolvendo medidas práticas e seus resultados levaram a uma discussão com recomendações sobre a fase de treinamento do modelo e sua relação com o desempenho das predições. A principal conclusão vem com a constatação do impacto da incerteza de localização na acurácia das predições baseadas em processos Gaussianos. Estes impactos são expressivos quando a aprendizagem de parâmetros é baseada em informações de posicionamento equivocadas em função da incerteza de localização. Desta forma, a principal conclusão no tratamento da incerteza de localização vem com o fato de que esta só faz sentido no momento em que os erros de localização são suficientemente expressivos a ponto de diferenciar a covariância entre as medidas das coordenadas espaciais estimadas em relação à covariância entre as medidas em suas verdadeiras localizações.

Algumas direções de pesquisa podem ser apontadas como perspectivas futuras: i) propostas e comparativos de outras técnicas de predição, uma vez que, para o TS-KALE, não se pode alegar a otimalidade (em algum sentido estatístico) dos resultados de estimação em função das técnicas utilizadas (MoM e LS); ii) investigação dos impactos provocados por alterações na distribuição estatística dos dados em relação à densidade Gaussiana, tanto na fase de aprendizagem de parâmetros como nos preditores; iii) estudo de diferentes abordagens e técnicas utilizadas na fase de aprendizagem de parâmetros [84], [142]; iv) introdução de erros sobre os valores de potência das medidas coletadas ([58, pp. 192]), além de mais transmissores no sistema de comunicação sem fio; v) investigação de modelos estatísticos mais realísticos para a incerteza de localização, considerando erros de posicionamento espacialmente correlacionados e sistemáticos; vi) considerar a estimação da densidade de probabilidade $g(\cdot)$ dos erros de localização e verificar seus impactos, tanto sobre a perspectiva do modelo de predição como em experimentos práticos e vii) estudos das diversas aplicações de comunicações sem fio e que se baseiam nos resultados de predição obtidos com o REM (*e.g.*, [21], [26-29], [79-80], [143-145]. Sobre esta última, é importante mencionar que o notório avanço das técnicas de comunicações digitais, que já operam com elevada eficiência espectral, impulsiona a busca por mais largura de banda e consequentemente, o uso das faixas de frequência de onda milimétricas (mmWaves) [1], [36], [39]. Desta forma, estudos sobre o REM em sistemas de comunicações que operam com ondas milimétricas podem contribuir para a compreensão da propagação nesta faixa de frequências, uma vez que a geração do REM requer a análise das medições e suas estatísticas ao longo do domínio espacial.

Bibliografia

- [1] M. Agiwal, A. Roy e N. Saxena, “Next generation 5G wireless networks: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617-1655, Feb. 2016.
- [2] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson e H. Wymeersch, “Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G,” *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102-112, Nov. 2014.
- [3] A. H. Sayed, A. Tarighat e N. Khajehnouri, “Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information,” *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24-40, Jul. 2005.
- [4] A. Tahat, G. Kaddoum, S. Yousefi, S. Valaee e F. Gagnon, “A look at the recent wireless positioning techniques with a focus on algorithms for moving receivers,” *IEEE Access*, vol. 4, pp. 6652-6680, Set. 2016.
- [5] S. Yarkan e H. Arslan, “Exploiting location awareness toward improved wireless system design in cognitive radio,” *IEEE Communications Magazine*, vol. 46, no. 1, pp. 128-136, Jan. 2008.
- [6] H. Celebi e H. Arslan, “Utilization of location information in cognitive wireless networks,” *IEEE Wireless Communications*, vol. 14, no. 4, pp. 6-13, Ago. 2007.
- [7] T. Q. Duong, M. Elkashlan, G. K. Karagiannidis, H. Wymeersch, Y. Mostofi e B. Shim, “Guest editorial location-awareness for radios and networks, part I,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1285-1287, Jul. 2015.
- [8] T. Q. Duong, M. Elkashlan, G. K. Karagiannidis, H. Wymeersch, Y. Mostofi e B. Shim, “Guest editorial location-awareness for radios and networks, part II,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2269-2271, Nov. 2015.
- [9] A. Kliks, P. Kryszkiewicz, A. Umbert, J. Pérez-Romero, F. Casadevall e L. Kulacz, “Application of radio environment maps for dynamic broadband access in TV bands in urban areas,” *IEEE Access*, vol. 5, pp. 19842-19863, Out. 2017.

- [10] S. Chaudhari, M. Kosunen, S. Mäkinen, C. Ramanathan, J. Oksanen, M. Laatta, J. Ryyänen, V. Koinunen e M. Valkama, "Spatial interpolation of cyclostationary test statistics in cognitive radio networks: methods and field measurements," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1113-1129, Fev. 2018.
- [11] H. B. Yilmaz, T. Tugcu, F. Alagöz e S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162-169, Dez. 2013.
- [12] J. Pérez-Romero, A. Zalonis, L. Boukhatef, A. Kliks, K. Koutlia, N. Dimitriou e R. Kurda, "On the use of radio environment maps for interference management in heterogeneous networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 184-191, Ago. 2015.
- [13] J. Li, G. Ding, X. Zhang e Q. Wu, "Recent advances in radio environment map: A survey," in *Proc. Second International Conference, Part II, MLICOM*, Weihai, China, Ago. 2017, pp. 247-257.
- [14] A. Umberto, F. Casadevall e E. G. Rodriguez, "An outdoor TV band radio environment map for a manhattan like layout," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, Poznan, 2016, pp. 399-403.
- [15] E. Dutkiewicz, Y. He, B. A. Jayawickrama e H. V. Abeywickrama, "Radio environment maps generation and spectrum sensing testbed for spectrum sharing in 5G networks," in *Proc. IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, Verona, 2017, pp. 33-36.
- [16] H. Braham, S. Ben Jemaa, G. Fort, E. Moulines e B. Sayrac, "Fixed rank kriging for cellular coverage analysis," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4212-4222, Mai. 2017.
- [17] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijärvi e P. Mähönen, "Automated coverage hole detection for cellular networks using radio environment maps," in *Proc. 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, Tsukuba Science City, 2013, pp. 35-40.
- [18] H. Braham, S. B. Jemaa, B. Sayrac, G. Fort e E. Moulines, "Low complexity spatial interpolation for cellular coverage analysis," in *Proc. 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Hammamet, 2014, pp. 188-195.
- [19] Y. Zhao, J. Gaeddert, K. K. Base e J. H. Reed, "Radio environment map enabled situation-aware cognitive radio learning Algorithms," in *Proc. SDR 06 - Technical Conference and Product Exposition*, Virginia Tech, 2006.
- [20] Y. Zhao, L. Morales, J. Gaeddert, K. K. Bae, J. S. Um e J. H. Reed, "Applying radio environment maps to cognitive wireless regional area networks," in *Proc. 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Dublin, 2007, pp. 115-118.
- [21] Reconfigurable Radio Systems (RRS); Use cases for building and exploitation of radio environment maps (REMs) for intra-operator scenarios, Eur. Telecommun. Stand. Inst., 3GPP, Tech. Rep. TR 102

- 947 V1.1.1, 2013. [Online]: http://www.etsi.org/deliver/etsi_tr/102900_102999/102947/01.01.01_60/tr_102947v010101p.pdf [acessado em Fev. 2019].
- [22] Study on Minimization of drive-tests in next generation networks; (Release 9), 3GPP, Tech. Rep. TR 36.805 v1.3.0 1, 2009. [Online]: <http://www.qtc.jp/3GPP/Specs/36805-900.pdf> [acessado em Fev. 2019].
- [23] J. Johansson, W. A. Hapsari, S. Kelley e G. Bodog, "Minimization of drive tests in 3GPP release 11," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 36-43, Nov. 2012.
- [24] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula e B. Sebire, "Minimization of drive tests solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28-36, Jun. 2012.
- [25] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijärvi e P. Mähönen, "Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks," in *Proc. 8th International Conference on Cognitive Radio Oriented Wireless Networks*, Washington, DC, 2013, pp. 37-42.
- [26] E. Dall'Anese, S. J. Kim e G. B. Giannakis, "Channel gain map tracking via distributed Kriging," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1205-1211, Mar. 2011.
- [27] K. Sato e T. Fujii, "Kriging-based interference power constraint: Integrated design of the radio environment map and transmission power," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 1, pp. 13-25, Mar. 2017.
- [28] K. Sato e T. Fujii, "Kriging-based interference power constraint for spectrum sharing based on radio environment map," in *Proc. IEEE Globecom Workshops*, San Diego, CA, 2015, pp. 1-6.
- [29] Z. Hou, Y. Zhou, L. Tian, J. Shi, Y. Li e B. Vucetic, "Radio environment map aided doppler shift estimation in LTE-railway," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4462-4467, Mai. 2017.
- [30] N. Ezzati, H. Taheri e T. Tugcu, "Optimised sensor network for transmitter localisation and radio environment mapping," *IET Communications*, vol. 10, no. 16, pp. 2170-2178, Jun. 2016.
- [31] J. Riihijärvi e P. Mahonen, "Machine learning for performance prediction in mobile cellular networks," *IEEE Computational Intelligence Magazine*, vol. 13, no. 1, pp. 51-60, Fev. 2018.
- [32] J. Van De Beek, T. Cai, S. Grimoud, B. Sayrac, P. Mahonen, J. Nasreddine e J. Riihijärvi, "How a layered rem architecture brings cognition to today's mobile networks," *IEEE Wireless Communications*, vol. 19, no. 4, pp. 17-24, Ago. 2012.
- [33] J. B. Andersen, T. S. Rappaport e S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 42-49, Jan. 1995.
- [34] M. F. Iskander e Zhengqing Yun, "Propagation prediction models for wireless communication systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 662-673, Mar. 2002.

- [35] A. Neskovic, N. Neskovic e G. Paunovic, "Modern approaches in modeling of mobile radio systems propagation environment," *IEEE Communications Surveys & Tutorials*, vol. 3, no. 3, pp. 2-12, Terceiro trimestre. 2000.
- [36] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios e J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks - with a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213-6230, Dez. 2017.
- [37] I. Trigui, S. Affes e B. Liang, "Unified stochastic geometry modeling and analysis of cellular networks in LOS/NLOS and shadowed fading," *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5470-5486, Dec. 2017.
- [38] C. Phillips, D. Sicker e D. Grunwald, "A survey of wireless path loss prediction and coverage mapping methods," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 255-270, Mar. 2012.
- [39] S. Sun, T. S. Rappaport, T. A. Thomas, A. Ghosh, H. C. Nguyen, I. Z. Kovács, I. Rodriguez, O. Koymen e A. Partyka, "Investigation of prediction accuracy, sensitivity, and parameter stability of large-scale propagation path loss models for 5G wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2843-2860, Mai. 2016.
- [40] C. Phillips, D. Sicker e D. Grunwald, "Bounding the error of path loss models," in *Proc. IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Aachen, 2011, pp. 71-82.
- [41] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 964-978, Mar. 2012.
- [42] M. Pesko, T. Javornik, A. Košir, M. Štular e M. Mohorčič, "Radio environment maps: The survey of construction methods," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 11, no. 11, Nov. 2014.
- [43] C. Singhal e S. De, *Resource allocation in next-generation broadband wireless access networks*, Advanced in Wireless Technologies and Telecommunication (AWTT) Book Series, 2017.
- [44] M. Angjelinoski, V. Atanasovski e L. Gavrilovska, "Comparative analysis of spatial interpolation methods for creating radio environment maps," in *Proc. 19th Telecommunications Forum (TELFOR)*, Belgrade, 2011, pp. 334-337.
- [45] D. Denkovski, V. Atanasovski, L. Gavrilovska, J. Riihijärvi, and P. Mähönen, "Reliability of a radio environment Map: Case of spatial interpolation techniques," in *Proc. 7th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, (CROWNCOM)*, Stockholm, 2012, pp. 248-253.
- [46] L. Bolea, J. Pérez-Romero e R. Agustí, "Received signal interpolation for context discovery in Cognitive Radio," in *Proc. International 14th Symposium on Wireless Personal Multimedia Communications (WPMC)*, Brest, 2011, pp. 1-5.

- [47] S. Ureten, "Single and multiple emitter localization in cognitive radio networks," Ph.D Dissertation, School of Electrical Engineering and Computer Science, University of Ottawa, Canada, 2016.
- [48] H. B. Yilmaz, "Cooperative spectrum sensing and radio environment map construction in cognitive radio networks," Ph.D Dissertation, Boğaziçi University, Turquia, 2012.
- [49] H. B. Yilmaz e T. Tugcu, "Location estimation-based radio environment map construction in fading channels," in *Wireless Communications and Mobile Computing*, vol.15, no.3, pp. 561-570, Mar. 2015.
- [50] M. Pesko, T. Javornik, L. Vidmar, A. Košir, M. Štular e M. Mohorčič, "The indirect self-tuning method for constructing radio environment map using omnidirectional or directional transmitter antenna," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, Dez. 2015.
- [51] IEEE VT-S Propagation Committee. Wireless Planning Tools for Mobility and Indoor Systems, Internet: <http://rsl.ece.ubc.ca/archive/planning-tools/home.html> [acessado em Fev. 2019].
- [52] M. Höyhty, A. Mämmelä, M. Eskola, M. Matinmikko, J. Kalliovaara, J. Ojaniemi, J. Suutala, R. Ekman, R. Bacchus e D. Roberson, "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2386-2414, Abr. 2016.
- [53] C. T. Phillips, "Geostatistical techniques for practical wireless network coverage mapping," Ph.D Dissertation, University of Colorado, Colorado, 2012.
- [54] C. Phillips, M. Ton, D. Sicker e D. Grunwald, "Practical radio environment mapping with geostatistics," in *Proc. IEEE International Symposium on Dynamic Spectrum Access Networks*, Bellevue, WA, 2012, pp. 422-433.
- [55] K. Sato, K. Inage e T. Fujii, "Spectrum database-assisted radio propagation prediction for wireless distributed networks: A geostatistical approach," in *Proc. IEEE 18th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, Valencia, 2016, pp. 1-6.
- [56] J. Riihijarvi, P. Mahonen, M. Wellens e M. Gordziel, "Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields," in *Proc. IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Dez. 2008, pp. 1-6.
- [57] N. Cressie, "Spatial prediction and ordinary Kriging," *Mathematical Geology*, vol. 20, no. 4, pp. 405-421, Mai. 1988.
- [58] C. Rasmussen e C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [59] G. Matheron, "Applications des méthodes statistiques á l'évaluation des gisements," *Annales del Mines*, no.12, pp. 50-75, 1955.
- [60] J. K. Yamamoto e P. M. B. Landim, *Geoestatística conceitos e aplicações*. São Paulo, Oficina de Textos, 2015.

- [61] D. G. Krige, "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, vol. 52, no. 6, pp. 119-139, Dez. 1951.
- [62] N. Cressie, *Statistics for spatial data*. Hoboken, NJ, USA: John Wiley & Sons, revised edition, 1993.
- [63] J. P. Chilès e P. Delfine, *Geostatistics modeling spatial uncertainty*. Hoboken, New Jersey: John Wiley & Sons, Second Edition, 2012.
- [64] H. Wackernagel, *Multivariate geostatistics: An introduction with applications*. Springer-Verlag Berlin Heidelberg, Second, Completely Revised Edition, 2003.
- [65] E. H. Isaaks e R.M. Srivastava, *An introduction to applied geostatistics*, Madison Avenue. New York: Oxford University Press, 1989.
- [66] M. A. Oliver e R. Webster, *Basic steps in geostatistics: The variogram and Kriging*. Springer Cham Heidelberg New York Dordrecht London: SpringerBriefs in Agriculture, 2015.
- [67] J. M. Montero, G. G. Avilés e J. Mateu, *Spatial and spatio-temporal geostatistical modeling and Kriging*. United Kingdom UK: John Wiley & Sons, 2015.
- [68] P. J. Diggle e P. J. Ribeiro, *Model-based geostatistics*. New York, USA: Springer Science +Business Media, LLC, 2007.
- [69] S. Yoo, "A spatial prediction theory for long-term fading in mobile radio communications," *ETRI Journal*, vol. 15, no. 3, pp. 27-34, Jan. 1994.
- [70] A. Konak, "Estimating path loss in wireless local area networks using ordinary Kriging," in *Proc. Winter Simulation Conference*, Baltimore, MD, 2010, pp. 2888-2896.
- [71] I. Nevat, G. W. Peters, F. Septier e T. Matsui, "Estimation of spatially correlated random fields in heterogeneous wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2597-2609, Mai. 2015.
- [72] G. Hernández-Peñaloza e B. Beferull-Lozano, "Field estimation in wireless sensor networks using distributed Kriging," in *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, ON, 2012, pp. 724-729.
- [73] G. Boccolini, G. Hernández-Peñaloza e B. Beferull-Lozano, "Wireless sensor network for spectrum cartography based on Kriging interpolation," in *Proc. IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sydney, NSW, 2012, pp. 1565-1570.
- [74] Guolin Sun, Jie Chen, Wei Guo e K. J. R. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12-23, Jul. 2005.

- [75] R. C. Shit, S. Sharma, D. Puthal e A. Y. Zomaya, "Location of Things (LoT): A review and taxonomy of sensors localization in IoT infrastructure," *IEEE Communications Surveys & Tutorials*, vol. PP, no. 99, pp. 1-1, Jan. 2018.
- [76] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs e E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1327-1346, Nov. 2016.
- [77] G. Han, J. Jiang, C. Zhang, T. Q. Duong, M. Guizani e G. K. Karagiannidis, "A survey on mobile anchor node assisted localization in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2220-2243, Mar. 2016.
- [78] H. Braham, S. B. Jemaa, G. Fort, E. Moulines e B. Sayrac, "Spatial prediction under location uncertainty in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7633-7643, Nov. 2016.
- [79] L. S. Muppirisetty, T. Svensson e H. Wymeersch, "Spatial wireless channel prediction under location uncertainty," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1031-1044, Feb. 2016.
- [80] M. Froehle, T. Charalambous, I. Nevat e H. Wymeersch, "Channel prediction with location uncertainty for ad-hoc networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. PP, no. 99, pp. 1-1, Mai. 2017.
- [81] M. Jadaliha, Y. Xu, J. Choi, N. S. Johnson e W. Li, "Gaussian process regression for sensor networks under localization uncertainty," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 223-237, Jan. 2013.
- [82] L. V. Nguyen, S. Kodagoda, R. Ranasinghe e G. Dissanayake, "Adaptive placement for mobile sensors in spatial prediction under locational errors," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 794-802, Feb. 2017.
- [83] D. Cervone e N. S. Pillai, "Gaussian process regression with location errors," from the arXiv database, arXiv: 1506.08256 [stat.ME], Jun. 2015.
- [84] T. R. Fanshawe e P. J. Diggle, "Spatial prediction in the presence of positional error," *Environmetrics*, vol. 22, no. 2, pp. 109-122, Mar. 2011.
- [85] M. S. Grewal, L. R. Weill e A. P. Andrews, *Global positioning systems, inertial navigation, and integration*. NJ, USA: Wiley, Hoboken, 2007.
- [86] S. Sand, A. Dammann e C. Mensing, *Positioning in wireless communications systems*. UK: John Wiley & Sons, 2014.
- [87] White Paper. (Fev. 2012). An overview of LTE positioning, accessed on 2018. [Online]: https://www.spirent.com/~media/white%20papers/mobile/lte_lbs_white_paper_2012.pdf [Acessado em Fev. 2019].

- [88] J. A. del Peral-Rosado, "Evaluation of the LTE positioning capabilities in realistic navigation Channels," Ph.D Dissertation, Universitat Autònoma de Barcelona, Bellaterra, 2014.
- [89] L. Anselin, "What is special about spatial data? Alternative perspectives on spatial data analysis," *Symposium on Spatial Statistics, Past, Present and Future*, NY, USA, Department of Geography, Syracuse University, NCGIA Technical Report, 1989.
- [90] M. Gaudard, M. Karson, E. Linder e D. Sinha, "Bayesian spatial prediction," *Environmental and Ecological Statistics*, vol. 6, no. 2, pp. 147-171, Jun. 1999.
- [91] J. G. Gabrosek, "The effect of locational uncertainty in geostatistics," Ph.D. Dissertation, Iowa State University, 1999.
- [92] J. Gabrosek e N. Cressie. "The effect on attribute prediction of location uncertainty in spatial data," *Geographical Analysis*, vol. 34, no. 3, pp. 262-285, Jul. 2002.
- [93] N. Cressie e J. Kornak, "Spatial statistics in the presence of location error with an application to remote sensing of environment," *Statistics Sciences*, vol. 18, no. 4, pp. 436-465, Nov. 2003.
- [94] L. Anselin, Class Lecture, Topic: "Spatial regression (Maximum likelihood estimation)." The Center of Spatial Data Science, University of Chigago, 2017.
- [95] D. A Guimarães, *Digital Transmission, A Simulation-aided introduction with VisSim/comm*. Springer Heidelberg Dordrecht London New York, 2009.
- [96] S. Haykin e M. Moher, *Modern wireless communication*, NJ, USA: Prentice-Hall, 2004.
- [97] A. Goldsmith, *Wireless communications*. New York, USA: Cambridge University Press, 2005.
- [98] S. S. Szyszkowicz, H. Yanikomeroğlu e J. S. Thompson, "On the feasibility of wireless shadowing correlation models," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4222-4236, Nov. 2010.
- [99] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145-2146, Nov. 1991.
- [100] M. Gudmundson, "Analysis of handover algorithms [microcellular radio]," in *Proc. 41st IEEE Vehicular Technology Conference Gateway to the Future Technology in Motion*, 1991, pp. 537-542.
- [101] D. Giancristofaro, "Correlation model for shadow fading in mobile radio channels," *Electronics Letters*, vol. 32, no. 11, pp. 958-959, 23 Mai. 1996.
- [102] ETSI, "Selection procedures for the choice of radio transmission technologies of the UMTS," Technical Report, 101.112 v3.2.0, Apr. 1998. [Online]: http://www.etsi.org/deliver/etsi-tr/101100_101199/101112/03.01.00_60/tr_101112v030100p.pdf [acessado em Fev. 2019].
- [103] P. Agrawal e N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4024-4036, Ago. 2009.

- [104] C. Xiaodong e G. B. Giannakis, "A two-dimensional channel simulation model for shadowing processes," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 6, pp. 1558-1567, Nov. 2003.
- [105] H. Claussen, "Efficient modelling of channel maps with correlated shadow fading in mobile radio systems," in *Proc. 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, 2005, pp. 512-516.
- [106] T. Ali e M. Saquib, "Numerically efficient generation of exponentially correlated two-dimensional shadowing," in *Proc. IEEE International Conference on Communications (ICC)*, London, 2015, pp. 4265-4271.
- [107] R. Fraile, J. F. Monserrat, J. Gozávez, and N. Cardona, "Mobile radio bi-dimensional large-scale fading modelling with site-to-site cross-correlation," *Transactions on Emerging Telecommunications Technologies*, vol. 19, no. 1, pp. 101-106, Feb. 2007.
- [108] J. F. Monserrat, R. Fraile, N. Cardona e J. Gozávez, "Effect of shadowing correlation modeling on the system level performance of adaptive radio resource management techniques," in *Proc. 2nd International Symposium on Wireless Communication Systems*, 2005, pp. 460-464.
- [109] R. S. Campos e L. Lovisolo, *RF positioning: Fundamentals, applications and tools (GNSS technology and applications)*, USA: Artech House GNSS Technologies and Series, 2015.
- [110] N. Cressie, "Fitting variogram models by weighted least squares," *Journal of the International Association for Mathematical Geology*, vol. 17, no. 5, pp. 563-586, Dez. 1984.
- [111] S. M. Kay, *Fundamentals of statistical signal processing: Estimation Theory*. Prentice Hall, Inc., 1993.
- [112] R. Augusto e C. M. Panazio, "On Geostatistical Methods for Radio Environment Maps Generation under Location Uncertainty," *Journal of Communication and Information Systems*, vol. 33, no. 01, Abr. 2018.
- [113] R. Augusto e C. M. Panazio, "Sobre a geração de mapas de ambiente de rádio com Kriging ordinário," in *Proc. XXXIV Brazilian Symposium on Telecommunications*, Ago. 2016.
- [114] J. J. Moré e D.C. Sorensen, "Computing a trust region step," *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 553-572, 1983.
- [115] J.J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in *Proc. Conference on Numerical Analysis of University of Dundee Sootland*, 1977.
- [116] X. Jian, R. A.Olea e Y.-S. Yu, "Semivariogram modeling by weighted least squares," *Computers & Geosciences*, vol. 22, no. 4, pp. 387-397, Mai. 1996.
- [117] J. Salo, L. Vuokko, H. M. El-Sallabi e P. Vainikainen, "An additive model as a physical basis for shadow fading," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 13-26, Jan. 2007.
- [118] R. Froidevaux, "Geostatistics and ore reserve classification," *CIM Bulletin*, vol. 75, no. 843, pp. 77-83, 1982.

- [119] M. E. Rossi e C. V. Deutsch, *Mineral Resource Estimation*, Dordrecht: Springer Science+Business Media, 2014.
- [120] J. K. Yamamoto, "An alternative measure of the reliability of ordinary Kriging estimates," *Mathematical Geology*, vol. 32, no. 4, pp. 489-509, Mai. 1999.
- [121] M. M. Rocha e J. K. Yamamoto, "Comparison between Kriging variance and interpolation variance as uncertainty measurements in the capanema iron mine, state of Minas Gerais - Brazil," *Natural Resources Research*, vol. 9, no. 3, Mar. 2000.
- [122] N. Cressie e G. Johannesson, "Fixed rank Kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 209-226, Jan. 2008.
- [123] B. François, "Parametric estimation of covariance function in Gaussian-process based Kriging models. Application to uncertainty quantification for computer experiments," Ph.D. Thesis, *Statistics* [math.ST]. Université Paris-Diderot - Paris VII, 2013.
- [124] J. Wågberg, D. Zachariah, T. B. Schön e P. Stoica, "Prediction performance after learning in Gaussian process regression," from the arXiv database, arXiv:1606.03865v3 [stat.ML], Mar. 2017.
- [125] M. Eriksson e P. P. Siska, "Understanding anisotropy Computations," *Mathematical Geology*, vol. 32, no. 6, pp. 683-700, Ago. 2000.
- [126] Y. S. Lagadapati, "Fast semivariogram computation using FPGA Architectures," M.A thesis, University of Texas at Tyler, Mai. 2015.
- [127] D. Marcotte, "Fast variogram computation with FFT," , *Computers & Geosciences* , vol. 22, no. 10, pp. 1175-1186, Jan. 1996.
- [128] G.B.M. Heuvelink e E.J. Pebesma, "Is the ordinary Kriging variance a proper measure of interpolation error?," , in *Proc. of the fifth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Jan. 2002, pp. 179-186.
- [129] Z. Wang, R. Schoenen, H. Yanikomeroglu e M. St-Hilaire, "The impact of user spatial heterogeneity in heterogeneous cellular networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, 2014, pp. 1278-1283.
- [130] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996-1019, Jun. 2013.
- [131] Wang HM e Zheng TX. "*Random Cellular Networks and Stochastic Geometry. In: Physical Layer Security in Random Cellular Networks*, SpringerBriefs in Computer Science. Springer, Singapore, 2016.
- [132] D. P. Kroese e Z. I. Botev, "Spatial process generation," from the arXiv database, arXiv:1308.0399v1 [stat.CO], Ago. 2013.

- [133] C. Saha, M. Afshang e H. S. Dhillon, "Poisson cluster process: Bridging the gap Between PPP and 3GPP hetnet models," from the arXiv database, arXiv:1702.05706v1 [cs.IT] , Fev. 2017.
- [134] T. W. Anderson e D. A. Darling, "Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes," *Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193-212, 1952.
- [135] T. W. Anderson e D. A. Darling, "A test of goodness of fit," *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 765-769, Dez. 1954.
- [136] W. C. Y. Lee, "Estimate of local average power of a mobile radio signal," in *IEEE Transactions on Vehicular Technology*, vol. 34, no. 1, pp. 22-27, Fev. 1985.
- [137] A. J. Goldsmith, L. J. Greenstein e G. J. Foschini, "Error statistics of real-time power measurements in cellular channels with multipath and shadowing," in *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 439-446, Ago. 1994.
- [138] D. Wong e D. C. Cox, "Estimating local mean signal power level in a Rayleigh fading environment," in *IEEE Transactions on Vehicular Technology*, vol. 48, no. 3, pp. 956-959, Mai. 1999.
- [139] P. Njemcevic e V. Lipovac, "Estimation of radio signal spatial local mean," 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 2016, pp. 1-5
- [140] D. Potere, "Horizontal Positional Accuracy of Google Earth's High-Resolution Imagery Archive," *Sensors*, vol. 08, no. 12, pp. 7973-7981, Dez. 2008.
- [141] S. Christian Benker e Richard P. Langford and Terry L. Pavlis, "Positional accuracy of the Google Earth terrain model derived from stratigraphic unconformities in the Big Bend region, Texas, USA," *Geocarto International*, vol. 26, no. 4, pp. 291-303, Abr. 2011.
- [142] M. Kasparick, R. L. G. Cavalcante, S. Valentin, S. Stańczak e M. Yukawa, "Kernel-Based Adaptive Online Reconstruction of Coverage Maps With Side Information," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5461-5473, Jul. 2016.
- [143] N. Perpinias, A. Palaos, J. Riihijärvi e P. Mähönen, "A measurement-based study on the use of spatial interpolation for propagation estimation," *IEEE International Conference on Communications (ICC)*, pp. 2715-2720, London, 2015.
- [144] V. Chowdappa, C. Botella, J. J. Samper-Zapater e R. J. Martinez, "Distributed Radio Map Reconstruction for 5G Automotive," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 36-49, Abr. 2018.
- [145] S. Bi, J. Lyu, Z. Ding e R. Zhang, "Engineering Radio Map for Wireless Resource Management," from the arXiv database, arXiv:1807.08235v1 [cs.IT], Jul. 2018.
- [146] M. Armstrong, "Problems with universal Kriging," *Mathematical Geology*, vol. 16, no. 1, pp. 101-108, Dez. 1984.

Apêndice A

Anexos - Artigos

On Geostatistical Methods for Radio Environment Maps Generation under Location Uncertainty

Ricardo Augusto and Cristiano Panazio

Abstract—Radio environment map (REM) can provide important information for designing and optimizing the performance of wireless communication networks. However, the location uncertainty related to the measurements used to build the REM can considerably deteriorate the accuracy of such map. This paper addresses this problem by proposing a modified approach of a classical geostatistical prediction tool, named Kriging method, which incorporates the location uncertainty and is able to improve the REM accuracy without adding significant complexity. Finally, we also show through simulation results that the average path loss and covariance parameter estimation play an important role and should be considered when the location errors occur in the wireless communication systems.

Index Terms—Spatial Predictions, Kriging, Location Errors, Geostatistics.

I. INTRODUCTION

RECENTLY, there has been growing interest in the generation and use of Radio environment map (REM) in wireless communications systems [1-7]. Part of this is due to the attempt to make better use of geolocation information such as efficient use of spectral resources in wireless communications systems [8-9]. Initially, the REM concept, proposed in [10-11], was defined as a database with geolocalized information applied to the dynamic spectrum access used in cognitive radio systems. Ever since, the focus on REM has expanded because of its potential applicability in telecommunications. In fact, REM can be used to extract relevant information about the behavior of the wireless channel as well as for making decisions related to different applications in wireless communications, such as: planning and optimization coverage tasks [3], [6-9]; minimization of drive tests (MDT) in mobile networks [37-40]; analysis and decision-making in cognitive radio systems involving: resource allocation [28], detection of coverage opportunities [5] and interference analysis in heterogeneous networks [29-30]; essential public safety services related to mobile networks such as 911 calls [8] and other specific applications, related to high speed trains [31-32], radar systems [33] and location-estimation-problems [34-36].

It is of fundamental importance to realize that the way each application relates to REM depends on how the location information is used in the system model and what kind of quantity constitutes the REM concept. In this work, we follow the approach considered in [3-5], in which REM consists

of a map that indicates a physical quantity of the radio environment, the most common being the signal intensity, i.e., radio signal power. Specifically, from a limited set of measurements collected in the radio environment, we will use spatial predictions for the REM generation. The applicability of spatial predictions to wireless communications systems is directly related to the spatial statistical characterization of the radio environment, which is complex because of the several propagation mechanisms of the wireless channel [21-27]. In fact, these mechanisms will influence the received signals and, consequently, may affect the REM generation process. Thus, the processing of geolocalized measurements for reliable generation of coverage maps presents itself as a challenge. Moreover, modern wireless communication systems require accurate and robust spatial predictions against adversities. On the latter, a problem that has recently been investigated consists in the study of the impacts of the location uncertainty on the performance of the spatial predictions [41-44].

In a wireless communication system, the location information depends mainly on the measurements sent by the network devices as well as the network-based location estimation methods, which gives an error between 50 and 300 meters, and the global positioning system (GPS), with an error between 5 and 30 meters [8], [41]. Despite the evolution of positioning techniques, due to channel adversities leading to imperfect location estimation as well as inaccuracy of measurements performed by network devices, it is reasonable to assume that the location information involved in the wireless communication system is not entirely accurate. Besides that, with the growth in the use of geolocation services, especially on the high accuracy required in public emergency services, we can mention that there is a fundamental need to incorporate the location uncertainty into the analysis of spatial predictions tools applied to the wireless communications systems.

In this context, we are looking for low-complexity spatial predictors (e.g., linear predictors), that are also robust to the channel adversities. On this aspect, a trade-off arises through two different approaches applied to the wireless spatial predictions: Bayesian and classical. In the Bayesian framework, a *priori* statistical distributions can be incorporated into the model so that the predictive posteriori distribution can be obtained, considering the uncertainties of the model [12-13], [16-17]. This is a promising and interesting feature, since it allows to incorporate different types of uncertainties into spatial predictions such as location uncertainty. However, complex integration methods are probably involved in the search for optimal predictors in some statistical sense (e.g., the conditional mean of the posterior distribution) and, conse-

quently, two important points emerge as obstacles to practical prediction Bayesian methods [17]: First, the requirement of a *priori* knowledge about the system model distributions as well as the analytical complexity of the parameters related to the conditional distributions necessary for the formulation of the Bayesian approach. Second, the high computational cost in the evaluation of the integrals to convert a specified model and prior to a posterior or Bayesian predictive distributions. A discussion about treatability is made in [42], in which a unified view of the Gaussian processes and the Bayesian approach is applied to the prediction problem with location uncertainty.

In an alternative way, the classical approach is characterized by the conventional estimation methods with semivariogram-based geostatistical framework used by Kriging spatial predictors, which allows us to find a flexible and low complexity solution for the REM generation. Indeed, when the model is characterized by Gaussian processes, the optimal conditional mean predictor of Bayes posterior distribution converges to the classical Kriging predictor. However, as we shall see later, the location uncertainty affects the Gaussianity of the model. In this sense, although the Bayesian approach can lead to non-linear predictors that may be more robust, our attention is focused on two specific points: the use of low complexity linear predictors considering the treatment of the location uncertainty; and dealing with the lack of knowledge about the covariance parameters related to the radio environment.

The main contributions of our work can be summarized as follows: We extend our work in [76] by applying a geostatistical model for REM generation and discussing the main peculiarities and limitations of this approach; We incorporate the location uncertainty into the radio environment model and use the methodology in [69-70] to manipulate intractable quantities that arise when location errors are considered in the spatial prediction problem. This allowed us to make adjustments in the conventional Kriging technique to take into account the statistics of the location errors in the spatial predictions, i.e., Kriging adjusted for location error (KALE). Moreover, we show that the covariance parameter estimation subject to location errors plays an important role and significantly affects the performance of KALE predictor. In this sense, we propose a new method that modifies the estimation of the semivariogram, (i.e., semivariogram adjusted for location error (SALE)) in order to improve the performance of the predictor KALE when there is lack of knowledge about the covariance parameters of the radio environment.

This paper is structured as follows. Sections II and III presents the system model of the radio environment and the geostatistical framework for REM generation. Section IV presents the Kriging spatial predictor. Section V incorporates the location uncertainty into the system model and propose a method for dealing with uncertainties. Numerical results and discussion are given in Section VI. Finally, Section VII points out the conclusions and proposals for future research.

¹Particularly, we are interested in the modeling of large-scale phenomena, especially the average path loss and shadowing of wireless channel, since the small-scale fading decorrelates within tens of centimeters (depending on the carrier frequency), making it infeasible to predict based on location information [42].

II. RADIO ENVIRONMENT MODEL

The radio environment model is based on a wireless communication system composed of mobile devices and a radio base station (RBS). It is assumed that such devices are capable of performing the measurements of the received signal level and send them to the RBS, which coordinates the entire operation of the wireless communication system network. This means that the RBS is responsible for the processing of the measurements collected and for the REM generation. Specifically, the propagation mechanisms of the radio environment, such as diffraction, reflection, refraction as well as average path loss, will influence the intensity of the signals received in the devices and consequently can affect the process and accuracy related to the REM generation.

In this work, we considered that the radio environment is interpreted as a spatial random process P , which is composed by two main components and is defined in all spatial locations $\mathbf{s} = (x, y)$, according to

$$P(\mathbf{s}) = \mu(\mathbf{s}) + \xi(\mathbf{s}), \text{ with } \mathbf{s} \in D, \quad (1)$$

where $\mu(\mathbf{s})$ consists of the spatial process drift, which is called trend in geostatistical terminology, $\xi(\mathbf{s})$ represents the zero-mean spatial random fluctuations of the model and D represents a coverage area of interest that compose the spatial domain in two dimensions.

In this model, the trend component is related to the wireless channel path loss while the spatial random fluctuations of the model depict shadowing effects of the wireless channel¹. In fact, the spatial random process $P(\mathbf{s})$ consists of a collection of random variables $P(\mathbf{s}_i)$ which are reception power measurements whose statistical properties depend on $\mu(\mathbf{s})$ and $\xi(\mathbf{s})$. Therefore, the main challenge on spatial predictions is to capture the effects of the components that form the model in (1) from measurements in the radio environment. First, the following subsections address the component related to the spatial trend. Then, the concepts about spatial random fluctuations are presented.

A. Trend Modeling

Trend modeling is characterized by the mean component $\mu(\mathbf{s})$ and is based on average path loss of the wireless channel. The mean component consists of the average reception power measurements and can be described by a set of base functions with corresponding coefficients, according to

$$\mu(\mathbf{s}) = P_{\text{tx}} - \sum_{k=1}^p \alpha_k f_k(\mathbf{s}), \quad (2)$$

where P_{tx} is the RBS transmission power (dBm), $f_k(\mathbf{s})$ are the base functions of the model and α_k are the unknown constant trend coefficients. Propagation tools for planning wireless systems and models for path loss have been extensively investigated in the literature and many papers discuss different models and methods on this subject (e.g., [45-48]). In this work, the model considered for average path loss is the well-known log-distance model, denoted as

$$\bar{P}_L(d) = \bar{P}_L(d_0) + 10 \log \left(\frac{d}{d_0} \right)^\alpha, \quad (3)$$

where $\bar{P}_L(d_0)$ is the attenuation (dB) at a reference distance d_0 (meters), also known as critical distance, $\bar{P}_L(d)$ is the attenuation (dB) suffered by the signals transmitted due to average the path loss at a given distance d (meters) and α which consists of the propagation coefficient of the path loss model [49]. This model was chosen due to its widespread use and since it allows to formulate the trending estimation as a linear estimation problem. It is important to note that the log-distance path loss model depends on the distance d between the RBS, with fixed location $\mathbf{s}_{tx} = (x_{tx}, y_{tx})$, and the devices in network. By making some simplifications without loss of generality, (with $\bar{P}_L(d_0) = 0\text{dB}$ and $d_0 = 1\text{m}$) and algebraic manipulations, it becomes possible to transporting this model to two dimensions as follows:

$$\begin{aligned}\bar{P}_L(\mathbf{s}) &= +10\alpha \log(d(\mathbf{s}_{tx}, \mathbf{s})) \quad \text{where } \mathbf{s}, \mathbf{s}_{tx} \in D, \\ &= +10\alpha \log \sqrt{(x_{tx} - x)^2 + (y_{tx} - y)^2},\end{aligned}\quad (4)$$

where $d(\mathbf{s}_{tx}, \mathbf{s})$ is the distance function with two dimensional coordinates. The model in (4) suggests how the average path loss of the radio environment fluctuates as a function of distance. In this sense, the value of the α parameter implies different decays of the logarithmic function of the path loss. Table I shows the association of some known scenarios and the respective values for the propagation coefficient [50].

TABLE I

Wireless Scenarios and Propagation Coefficient	
Environment Type	Propagation Coefficient
Free-Space	2
Rural (Plan)	3
Rural(Mountain)	3.5
Suburban (City)	4
Urban(City)	4.5

By using the base functions of the model in (2) and considering one-parameter α path loss model ($p = 1$), it is possible to apply the average path loss in the trend model in (2) and express the average power reception, according to

$$\begin{aligned}\mu(\mathbf{s}) &= P_{tx} - \alpha f(\mathbf{s}) \\ &= P_{tx} - 10\alpha \log(d(\mathbf{s}_{tx}, \mathbf{s})).\end{aligned}\quad (5)$$

At this point, it is possible to highlight two information that will be useful: i) An inspection of the log-distance model allows to observe that this model is linear in the α parameter and ii) Although it is linear in α , the model is not linear on the spatial coordinates $\mathbf{s}$. These characteristics will influence the parameter estimation and spatial predictions processes internal to the geostatistical methods.

B. Spatial Random Process Modeling

The model in (1) indicates that the spatial random fluctuations of $\xi(\mathbf{s})$ occur around the average reception power. In a wireless communication system, this is due to the large buildings, mountains and large-scale obstacles that adorn the receivers in the network. In fact, a signal transmitted by a wireless propagation channel undergoes random variations due to obstructions along the path of the receiver. Since the location, size and dielectric properties of these obstruction

objects are unknown, statistical models for describing such fluctuations become fundamentally necessary in this analysis. One of the most commonly used models in the literature is the log-normal shadowing [50-52]. This model is characterized by a Gaussian random process in the dB scale and has been confirmed empirically to accurately model the variation in received power in both outdoor and indoor radio propagation environments [51-52]. Thus, the log-normal shadowing is used to characterize the random fluctuations $\xi(\mathbf{s})$ and is incorporated into the system model by adding a Gaussian spatial random process to the average path loss trend (4), according to

$$\begin{aligned}P(\mathbf{s}) &= \mu(\mathbf{s}) + \xi(\mathbf{s}) \\ &= P_{tx} - \underbrace{10\alpha \log d(\mathbf{s}_{tx}, \mathbf{s})}_{\text{Path Loss}} + \underbrace{\xi(\mathbf{s})}_{\text{Shadowing}},\end{aligned}\quad (6)$$

where $\xi(\mathbf{s})$ is a zero-mean Gaussian spatial random process with standard deviation σ , representing the shadowing in the spatial position $\mathbf{s}$. Also, $\xi(\mathbf{s})$ is characterized by structured parametric spatial covariance $C(\mathbf{s}_i, \mathbf{s}_j)$. In other words, the spatial dependence relationship between the collections of random variables that form the spatial random process P arises in the form of a structured covariance $C(\mathbf{s}_i, \mathbf{s}_j)$. On the geostatistical perspective, these are important presuppositions of our work and favor the choice of the Kriging predictor. In fact, the central idea of geostatistics is to explore this relationship of spatial dependence to generate spatial predictions at unknown locations. Thus, the spatial predictors must be followed preceded by steps that have the objective of extracting the covariance parameters of $\xi(\mathbf{s})$. On this aspect, two spatial functions are of fundamental importance, namely covariance and semivariogram.

Formally, the covariance between two random variables in the coordinate pairs $C(\mathbf{s}_i, \mathbf{s}_j)$, $\forall \mathbf{s}_i, \mathbf{s}_j \in D$, is defined as

$$C(\mathbf{s}_i, \mathbf{s}_j) = \mathbb{E}\{[P(\mathbf{s}_i) - \mu(\mathbf{s}_i)][P(\mathbf{s}_j) - \mu(\mathbf{s}_j)]\} \quad (7)$$

where $\mu(\mathbf{s}_i)$ is the mean of the random variable $P(\mathbf{s}_i)$ and $\mathbb{E}\{\cdot\}$ is the expectation operator. The semivariogram function is defined as the variance of the differences of the random variables that compose $P(\mathbf{s})$,

$$\begin{aligned}\gamma(\mathbf{s}_i, \mathbf{s}_j) &= \frac{1}{2} \text{Var}\{P(\mathbf{s}_i) - P(\mathbf{s}_j)\} \\ &= \frac{1}{2} \mathbb{E}\{[P(\mathbf{s}_i) - P(\mathbf{s}_j) - \mathbb{E}\{P(\mathbf{s}_i) - P(\mathbf{s}_j)\}]^2\}.\end{aligned}\quad (8)$$

Two properties are of fundamental importance for these functions, especially when it is assumed that the spatial random process are wide sense stationary: First, the expected value of the spatial random process exists and is constant, i.e., $\mathbb{E}\{P(\mathbf{s})\} = \mu(\mathbf{s}) = \mu$. Note that in the case of the radio environment, the average reception power depends on the spatial position related to the RBS. Consequently, the presence of the trend affects the stationarity and, therefore, the trend must be estimated. Second, the covariance and semivariance for any pair of points in the spatial random process is finite and its value depends only on the separation vector $\mathbf{h}$, i.e., $C(\mathbf{s}_i, \mathbf{s}_i + \mathbf{h}) = C(\mathbf{h})$, where the vector $\mathbf{h} = (h_x, h_y)$ consists of a spatial separation distance vector, $\mathbf{h} \equiv \mathbf{s}_i - \mathbf{s}_j$, [12], [69]. Thus, in these circumstances the structured covariance

does not depend specifically on the spatial positions analyzed, but on the separation $\mathbf{h}$ of such coordinates. In this way, we can rewrite the theoretical expressions of covariance and semivariogram considering the invariance to the translations

$$\begin{cases} C(\mathbf{h}) = \mathbb{E}\{[P(\mathbf{s}_i + \mathbf{h}) - \mu][P(\mathbf{s}_i) - \mu]\}, \forall \mathbf{s}_i, \mathbf{s}_i + \mathbf{h} \in D, \\ \gamma(\mathbf{h}) = \frac{1}{2} \text{Var}\{P(\mathbf{s}_i + \mathbf{h}) - P(\mathbf{s}_i)\}, \forall \mathbf{s}_i, \mathbf{s}_i + \mathbf{h} \in D. \end{cases} \quad (9)$$

If the direction of the vector $\mathbf{h}$ does not influence the expected values in (9), the covariance and semivariogram will only depend on the distance, i.e., $|\mathbf{h}| = |\mathbf{s}_i - \mathbf{s}_j|$. This isotropic property is useful in the application of the covariance and semivariogram functions in the characterization of correlation model of the log-normal shadowing of the wireless channels. In fact, one of the most used models for shadowing correlation characterization is the Gudmundson model [51-55]. Specifically, this model belongs to the class of correlation models that take into account the separation distance $|\mathbf{h}|$ between measurement collected in a radio environment. Exhaustive tests and measurement campaigns indicate an exponential behavior for the decay of the spatial correlation as a function of distance of the log-normal shadowing of the wireless communication channels [51-56]. Therefore, in this paper, an option was made for isotropic-exponential structured covariance for the spatial random process generation, given by

$$C(|\mathbf{h}|) = m e^{-\frac{|\mathbf{h}|}{r}}, \quad (10)$$

where m represents the variance of the spatial random process and is called *sill variance* while the parameter r is known as *range* and reflects the exponential decay of the covariance function. Under wide sense stationarity assumption there is a direct relationship between the covariance function and the semivariogram function [12], $C(\mathbf{s}_i, \mathbf{s}_j) = C(0) - \gamma(\mathbf{s}_i, \mathbf{s}_j)$. Thus, the isotropic-exponential semivariogram model is given by

$$\gamma(|\mathbf{h}|) = m \left\{ 1 - e^{-\frac{|\mathbf{h}|}{r}} \right\}. \quad (11)$$

Obtaining reliable estimates of reception power $\hat{P}(\mathbf{s})$ at unknown spatial positions will depend on the ability of the spatial predictors to capture the spatial variations of the radio environment. Thus, the semivariogram and covariance functions are fundamental for the spatial prediction problem of the REM generation process.

III. RADIO ENVIRONMENT SPATIAL PREDICTION MODEL

In this section, we present the spatial prediction method based on geostatistical approach applied to the radio environment modeling. Fig. 1 shows the diagram of the proposed model. First, the trend modeling is performed to generate the average reception power measurements as a function of the path loss model in (4). The spatially correlated shadowing generation characterizes the spatial random process modeling. Different techniques can be used to generate log-normal shadowing, and it is possible to highlight the techniques based on Cholesky decomposition, autoregressive filtering, sum of sinusoids and Fourier transformations of the shadowing correlation models (with higher computational cost) [22], [57-62]. In this work, the Cholesky decomposition is chosen since it is directly based on the covariance matrix of the spatial random process and has a direct relation with the semivariogram. The combination of the trend with spatial random process forms the large-scale component of the wireless communication channel.

In a wireless communication system, only an observable part of the radio environment can be captured. Specifically, a spatial sampling is performed and N network devices send the reception power measurements P_{rx} to the RBS composing a set of N power measurements. The spatial collection of measurements from the network devices characterizes the spatial sampling process applied to the radio environment. For simplicity, it is assumed that $P_{tx} = 0\text{dBm}$. Thus, based on the system model (6), we can write

$$\mathbf{P} = \mathbf{X}\alpha + \boldsymbol{\xi}, \quad (12)$$

where $\mathbf{P} = [P_{rx}(\mathbf{s}_1) \dots P_{rx}(\mathbf{s}_N)]^T$ are the set of the power measurements, $\mathbf{X} = [-10 \log d(\mathbf{s}_{tx}, \mathbf{s}_1) \dots -10 \log d(\mathbf{s}_{tx}, \mathbf{s}_N)]^T$ are the vector of base functions, α is the path loss exponent and $\boldsymbol{\xi} = [\xi(\mathbf{s}_1) \dots \xi(\mathbf{s}_N)]$ is the zero-mean spatial correlated Gaussian random vector. Initially, we emphasize an important assumption: In this section, the vector model (12) is location-error-free as location uncertainty will be incorporated later on. This means that the measurements of the radio environment are taken exactly in the spatial coordinates ($\mathbf{s}$) of the two-dimensional space and will be processed by geostatistical methods to generate spatial predictions.

The geostatistical approach used in the work is characterized by three estimation stages: i) trend estimation; ii) experimental semivariogram (covariogram) estimation and iii) estimation of covariance parameters. These stages form the well-known learning phase of the channel parameters and precede the

Fig. 1: Diagram with the model for generating spatial predictions based on the geostatistical approach.

spatial prediction phase of the radio environment. In other words, the estimates of spatial predictions depend directly on the information about wireless channel parameters. Thus, we are particularly interested in proposing techniques that provide support for spatial predictors when there is location uncertainty in wireless system as well as the lack of knowledge about wireless channel parameters.

The first process to be performed is the trend estimation and it is important for two reasons: i) the presence of a trend in the radio environment affects the stationarity of the spatial random process influencing the results of the spatial predictors and ii) the experimental estimates of the semivariogram or the covariogram, if performed directly on the measurements, can lead to biased results [12]. In this sense, the objective of the trend estimation process is to obtain the regionalized variable (RV) [12], [15], [18], which only characterizes the spatial random variations of the wireless channel shadowing. The log-distance model allows us to formulate the estimation problem (α estimation) linearly allowing the use of the minimum variance unbiased estimator through methods such as least squares (LS) techniques. Thus, after trend estimation of the radio environment (performed with least squares fitting), RV is obtained by removing the estimation results from the trend of the measurements collected, i.e., $\mathbf{Z} = \mathbf{P} - \hat{\boldsymbol{\mu}}$ characterizing the detrending process.

The next stage of the learning phase is the experimental semivariogram estimation process that is obtained from the RV. The experimental semivariogram (and covariogram) consists of the empirical estimate of the theoretical semivariogram (and covariance) function required for the spatial predictions. The method of moments of Matheron [12-15], [23] is used to obtain the experimental semivariogram from the regionalized measurements and can be described by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} [Z(\mathbf{s}_j) - Z(\mathbf{s}_i)]^2, \forall \mathbf{s}_i \in D, i = 1, 2, \dots, N, \quad (13)$$

where N_h represents the number of pairs of points distant of $\mathbf{h}$ obtained from RV. Hence, the semivariogram estimates are obtained through the mean of the semivariance values of all pairs of points that are distanced themselves from $\mathbf{h}$. In practice, for the calculated experimental semivariance to have a statistical meaning, it is necessary that N_h should be sufficiently large, i.e., a reasonable number of measurements must be taken for each distance $\mathbf{h}$. However, the RV is composed of a finite set consisting of N measurements. Facing this practical situation, the semivariogram estimation is based on pre-calculated separation distances, denoted as lag distances.

This work makes the semivariogram estimation through the methodology of [76] which is based on the quantization of the separation distances and Matheron method (details can be seen in [12], [14-18]). Fig. 2 shows a conceptual example of the parametric-analytical semivariogram model (solid curve) and the results of experimental semivariogram estimation (points). The analysis of the semivariogram estimation aims to provide a description of the spatial variability of the spatial random process. It is important to note that the analytical models and experimental parts have different definitions. The experimental

Fig. 2: Semivariogram analytical model (solid curve) experimental semivariogram (points).

semivariogram (covariogram) has no parameters while the semivariogram (covariance) analytical model is parametric. Also, there is an interest in the estimation of three parameters: the upper bound parameter m , called *sill*, the r parameter, called *range* and the *nugget* c_0 parameter, which is related to the nugget effect [66]. These parameters are grouped in the form of a vector $\boldsymbol{\theta} = [m, r, c_0]$.

In the wireless communications context, the parameter m consists of the variance of the spatial random process generated for wireless shadowing characterization and, therefore, can be interpreted as a measure of the shadowing severity. The parameter r is related to the slope of the semivariogram (decay of the covariance model) and allows to measure the separation distances from which there is a low spatial correlation between the captured measurements. Then, it can be interpreted as a measure of the de-correlation distance d_{corr} of the wireless shadowing. Specifically, the distances between measurements that are smaller than r present a higher degree of spatial correlation, resulting in spatial dependence zones. Otherwise, measurements of the radio environment that distances beyond the *range* are weakly correlated characterizing the zones of spatial independence. The parameter c_0 is related to the nugget effect, which implies an abrupt change in the spatial variability of the random process at small separation distances (less than the smaller spatial sampling intervals considered) and may result from measurements inaccuracies, short-range variations, or superimposed noise [23-24].

The covariance parameter estimation can be performed using different techniques, e.g., Maximum Likelihood (ML) and Least Squares (LS) [12], [15-20]. The Gaussian nature of the wireless channel shadowing makes the ML method better suited for the estimation problems. However, as we will see later, the location uncertainty affects the Gaussianity of the model [41], [69], [72]. Also, although the ML method has superior performance, the estimation can be computationally challenging for large data sets, as each likelihood evaluation requires a Cholesky factorization of covariance matrix or equivalent operations, which is $O(N^3)$ [72]. On the other hand, the least squares estimators uses the separation distances $\mathbf{h}$ and the experimental semivariogram to find the values of $\boldsymbol{\theta}$ that lead to the minimization a cost function by using search algo-

gorithms (e.g., Trust Region and Levenberg-Marquardt [67-68]). In fact, the LS estimation presents a practical geostatistical-based solution in which no probabilistic assumptions are made about the measurements and, for this reason, (associated with ML convergence problems) it is used in this work.

A. Radio Spatial Prediction Model Simulation

Figure 3 presents the initial simulation results of radio environment spatial prediction model related to the learning phase. Typical configuration settings were used and the radio environment simulation is composed of: the average received power due to the average path loss effects, depicted in Fig.3(a), and the wireless channel shadowing with a isotropic-exponential covariance analytical model, shown in Fig.3(b). The combination of maps in Figs. 3(a) and 3(b) results in the spatial random process presented in Fig.3(c). The RBS is characterized by a transmitter with Equivalent Isotropically Radiated Power (EIRP = 0dBm) positioned in the location $s_{tx} = (20, 20)$, within a coverage area with size given by $100m \times 100m$. The propagation coefficient of the path loss model was simulated with the value $\alpha = 3$. The log-normal shadowing of the radio environment has standard deviation $\sigma = 5dB$ ($m = 25$) and zones of spatial dependence equal to 20 meters ($r = 20$). Also, we assume that abrupt spatial variations do not occur at small separation distances ($c_0 = 0$), since we are particularly interested in modeling the large-scale effects of the channel only with average path loss and shadowing. The spatial sampling is performed by capturing $N = 200$ power measurements which were uniformly distributed.

Particularly on this aspect, the spatial sampling pattern is conditioned to the type of application analyzed (e.g., drive-tests campaign or measurement report systems). In a planning for the execution of drive-tests, there is more control over the locations where the measurements will be captured through routes that are predefined and usually accompany the roads and streets of the cities. The spatial sampling with routes is common in works where some practical drive-test activity was performed as in [41]. Otherwise, an operation scenario where the measurements are sent from users devices, there are no predefined routes but a spatial pattern movement of the devices in the network, which will determine the type of the spatial distribution. In fact, the use of uniform distribution in REM occurs in applications related to distributed sensors networks [43], cognitive radio networks [29-30] and also in wireless communications systems [5-6]. In this context, the uniform spatial sampling hypothesis is plausible in scenarios where there is no *a priori* control over measurement locations and, therefore, is used in this work. Even so, it is important to note that the minimum control in the execution of the sampling strategy is enough to find alternatives to choose the most appropriate strategies (e.g., [80-82]). Moreover, for analysis simplification, we also consider the use of omnidirectional antennas as in [41-42]. The incorporation of directional antenna modeling is possible to make the model more practical, but it increases the simulation complexity. Actually, it is important to mention that the directional antenna transmission impacts the radio parameter estimation and can consequently affect the performance of spatial predictions [79-80].

Fig. 3: Simulation results: (a) average received power; (b) wireless shadowing; (c) radio environment realization with uniform spatial sampling; (d) path loss trend estimation results; (e) experimental semivariogram; (f) experimental covariogram.

The trend estimation is presented in Fig.3(d), which shows a comparison between the theoretical average path loss model and the result of trend estimation along with average path loss measurements contaminated with shadowing of the wireless channel. The experimental semivariogram and covariogram are presented in Figs.3(e) and 3(f). The amount of combinations of pairs of points associated with each lag distance is also shown (boxes). It is possible to note that the semivariogram and the covariogram reveal important information about the parameters used in the generation process of the radio environment. First, an upper bound is revealed by the estimates of semivariance in Fig.3(e). It is verified that such an upper bound occurs from distances of the order of the *range*. This result can also be seen in the behavior of the covariogram, where a more pronounced covariance drop occurs around 20 meters of separation. The solid lines in Figs.3(e) and 3(f) are the semivariogram and covariance analytical functions with the LS estimated parameters $\hat{\theta} = [m, r, c_0] = [31.3, 23.8, 0]$. The difference between the estimated covariance and the true parameters $\theta = [m, r, c_0] = [25, 20, 0]$ is due to the estimation performance of the sequential methods (i.e., trend, semivariogram and covariance parameter estimation) used in this geostatistical approach.

IV. SPATIAL PREDICTIONS - KRIGING METHOD

The formulation of the spatial prediction problem in wireless communications is to obtain a reliable value of received power in a coverage area from a limited set of measurements sent by the network devices. A spatial predictor is intended to solve this problem and Kriging is a best linear unbiased predictor (BLUP) in the sense that it minimizes the variance of the prediction error [12], [14-20].

Depending on the characteristics of the problem, different Kriging methods are better suited to perform the spatial prediction. In circumstances where the specific value of the mean μ is known *a priori*, the simple Kriging (SK) method is used. In the case where μ is not necessarily known, but constant, the ordinary Kriging (OK) method is used, since it requires only the constant behavior of μ along the D domain. In practice, the mean of the spatial random process can vary along the space (as in the radio environment), compromising the stationarity assumption. In this situation, the universal Kriging (UK) method can be used if the trend can be decomposed into a set of base functions represented by low-degree monomials [12]. In other words, if the trend is linear at coordinates or with powers that normally do not exceed the power of two, the UK can be used. Despite this, the average path loss trend in the radio environment is modeled with non-linear functions in the coordinates (such as square root and logarithms) with powers that are usually greater than two indicating that the use of the UK in this case requires complex transformations for the decomposition of base functions to be achieved. Then, in order to overcome this problem, the OK predictor combined with detrending process is the chosen tool to perform spatial predictions in this work, where

$$Z_{OK}(s_0) = \lambda^T \mathbf{Z} = \sum_{i=1}^N \lambda_i Z(s_i), \quad (14)$$

is the scalar result of spatial prediction in a given unknown spatial coordinate s_0 , obtained through the linear combination with the Kriging weights column vector λ , with size $N \times 1$, and column vector of measurements $\mathbf{Z}$, with size $N \times 1$.

It is important to mention that the unknown locality s_0 is taken into account indirectly in the OK predictor (14) through the calculation of the Kriging weights λ_i . Specifically, the values λ_i are calculated with the objective of minimizing the variance of the prediction error, i.e., $\text{Var}\{Z_{OK}(s_0) - Z(s_0)\}$, based on the condition that the predictor in (14) is unbiased. In this sense, the optimal Kriging weights consists in the search for the solution of a system of equations, given by

$$\lambda = \Gamma^{-1} \beta, \quad (15)$$

where Γ is the semivariance matrix between the measurements that is based on covariance parameters and β is the semivariance vector between the RV measurements and the target coordinate of spatial prediction s_0 . The details of the Lagrange mathematical procedure for obtaining the Kriging system of equations as well as the optimal weights can be found in [12], [14-20], [76-78]. It is worth noting that the matrix inversion may lead to some numerical problems, especially in application where the amount of data is significantly high [78], due to ill-conditioning. One way to overcome this if it may happen is to use a fixed-rank Kriging method [77], which decomposes the matrix into low order matrices.

The spatial prediction $Z_{OK}(s_0)$ generated by OK predictor is a result of the linear combination of the Kriging weights with the trend-removed vector of measurements $\mathbf{Z}$. Thus, the sum of $Z_{OK}(s_0)$ with the the average received power estimation in s_0 , denoted as $\hat{\mu}(s_0)$, allows to obtain the prediction value of the received power at unknown location according to

$$\begin{aligned} \hat{P}(s_0) &= \hat{\mu}(s_0) + Z_{OK}(s_0) \\ &= \hat{\mu}(s_0) + \sum_{i=1}^N \lambda_i Z(s_i). \end{aligned} \quad (16)$$

Fig. 4 shows the spatial prediction results, i.e., REM, based on the measurements captured from the spatial random process previously shown in Fig. 3. In Fig. 4(a) the true reception power map subject to the average path loss and shadowing effects is present while the REM generation, composed of trend estimation and spatial predictions, is shown in Fig. 4(b). The comparison between the wireless shadowing and OK spatial predictions can be observed in Figs. 4(d) and 4(e). Through a single realization and difference maps, it is possible to verify that the Kriging spatial predictor was able to capture the spatial variability of the radio environment. However, it is noted in Figs.4(c) and 4(f) that the difference between the values of spatial predictions and the true values of the radio environment fluctuates spatially in the map. Investigations carried out in our simulations also indicated that as important as the number of measurements is the representativeness of such measurements, i.e., how efficient these measurements are in capturing the spatial variability of the random process that characterizes the radio environment. Therefore, the spatial sampling pattern and density (i.e., the relationship of N with the size of the coverage area) and the de-correlation distance of

Fig. 4: Simulation results of REM: (a) true power reception; (b) REM obtained from trend estimation and OK predictions; (c) difference map: power reception and REM; (d) wireless shadowing; (e) OK predictions; (f) difference map: shadowing and OK prediction.

the wireless shadowing are factors that have a direct influence on the prediction accuracy of REM.

Furthermore, it is important to highlight that these results refer to a single realization of the spatial random process (Section VI brings the performance analysis with more realizations). At this point, it is intuitive to think that the incorporation of uncertainties on the devices positions can influence spatial prediction results. In the next section, we introduce the location error in our system model and propose a spatial predictor that can mitigate its harmful impacts.

V. DEALING WITH LOCATION UNCERTAINTY

In this work, the uncertainty location uncertainty model proposed by Cressie and Gabrosek in [69-71] is applied to the wireless communication system. It is assumed that the network devices are uniformly distributed within the coverage area and that each device is capable of making reception power measurements. Figure 5 illustrates the problem of location uncertainty in the wireless environment showing how Gaussian location errors can affect the positions reported by devices in the wireless network and by consequence how it will imply in some perturbations when creating the REM. Through network location methods, the devices send their location information to RBS, forming the set of intended sites $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, marked as small circles in Fig.5. However, the incorporation of location uncertainty corrupts this information, since the devices are positioned in the set of realized sites $\{\mathbf{r}_1, \dots, \mathbf{r}_N\}$, which are the true unknown positions, marked as small crosses. Thus, the reception power measurements

are taken in $\mathbf{r}$ and mistakenly associated with $\mathbf{s}$ coordinates. This mismatch between the power measurements and the set $\mathbf{s}$ characterizes the location uncertainty model,

$$\mathbf{r} = \mathbf{s} + \mathbf{p}(\mathbf{s}), \quad (17)$$

where $\mathbf{p}(\mathbf{s}) \equiv \mathbf{r} - \mathbf{s} = (u_x, u_y)$ consists of the two dimensional location error vector related to the spatial coordinates and follows a given probability distribution $g(\cdot)$. In this model, it is assumed that the location error $\mathbf{p}(\mathbf{s})$ is independent in the spatial coordinates, that is, a correlation between location errors in the radio environment is not assumed *a priori*. Two types of statistical distributions for the location error are evidenced in [69-71]:

- (i) *i.i.d.* normal location errors, $\mathbf{p}(\mathbf{s}) \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_{dN})$, where σ_p is a standard deviation of location error, measured in meters, $\mathbf{I}_{dN}$ is identity matrix with size of $dN \times dN$ and d is the dimension of spatial domain ($d = 2$).
- (ii) *i.i.d.* uniform location errors that is the random vector $\mathbf{p}(\mathbf{s}) \sim \text{Unif}(-\sqrt{3}\sigma_p, +\sqrt{3}\sigma_p)$.

Gabrosek in [70] investigated the influence of location errors with Gaussian and uniform distributions on several spatial covariance models and pointed out that both distributions caused similar effects on the covariance of the spatial random process. In this work, we chose the Gaussian distribution to characterize the random effects of location errors of uncertainty model, since it is plausible to assume this distribution to model location errors in the context of wireless communications systems [41-42].

Fig. 5: Spatial random process generated with $\theta = [25, 10, 0]$, $\alpha = 3$ and with the location uncertainty.

In this way, by incorporating the uncertainty model (17) into the system model of the radio environment (6), it becomes possible to write

$$\begin{aligned} P_g(\mathbf{s}) &= \mu(\mathbf{r}) + \xi(\mathbf{r}) \\ &= \mu(\mathbf{s} + \mathbf{p}(\mathbf{s})) + \xi(\mathbf{s} + \mathbf{p}(\mathbf{s})) \\ &= P_{tx} - \underbrace{10\alpha \log d(\mathbf{s}_{tx}, \mathbf{s} + \mathbf{p}(\mathbf{s}))}_{\text{Path Loss}} + \underbrace{\xi(\mathbf{s} + \mathbf{p}(\mathbf{s}))}_{\text{Shadowing}}, \end{aligned} \quad (18)$$

where we denote the set of received power measurements with location errors as $\{P_g(\mathbf{s}_1), \dots, P_g(\mathbf{s}_N)\}$ and $\mathbf{s}, \mathbf{r} \in D$. Also, it is assumed that the set $\{Z_g(\mathbf{s}_1), \dots, Z_g(\mathbf{s}_N)\}$ represents the RV obtained from the removed trend estimation process subject to location errors. In this context, some difficulties arise, especially in aspects related to the characterization of the spatial random process with location uncertainty and the influence of such errors in the trend and covariance parameters estimation. To verify that the spatial random process of the radio environment has its Gaussian characteristics affected when subjected to location uncertainty, the equation below shows the cumulative distribution function of the location corrupted process $Z_g(\mathbf{s})$,

$$\begin{aligned} \Pr(Z_g \leq z) &= \int \Pr(Z(\mathbf{s} + \mathbf{u}) \leq z) g(\mathbf{u}) d\mathbf{u} \\ &= \int \int_{-\infty}^z (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{t^2}{2\sigma^2}\right\} g(\mathbf{u}) dt d\mathbf{u}, \end{aligned} \quad (19)$$

where $g(\mathbf{u})$ is the probability density distribution of location error vector $\mathbf{p}(\mathbf{s}) = \mathbf{u} = (u_x, u_y)$, which is responsible for affecting the Gaussian characteristic of the spatial random process. This means that the covariance of the spatial random process subject to location uncertainty is different when compared to the covariance of the location-error-free random process and, consequently this difference will have some influence on the estimation of the covariance parameters.

Two important effects of the location uncertainty are shown in Fig. 6. First, the estimation and removal of the trend is influenced by location errors. This can be observed in Fig.6(a) where the dispersion (variance) of the RV increases as the severity of the location errors also increases. As a consequence, some residue of the trend contaminates the RV

that will be processed by the next estimation steps. The effects of location errors on the spatial covariance are shown through experimental semivariograms in Figs. 6(b) and 6(c), obtained from the measurements $Z(\mathbf{s})$ (without location errors) and $Z_g(\mathbf{s})$ (with location errors), respectively. Through the simulations, a change in the structured covariance was observed as well as a lifting effect in the vicinity of the semivariogram origin (flattening effect on the covariogram origin). These observations agree with the results presented in [69-73].

Specifically in the works [69-71], the influence of the location errors on the structured covariance is also investigated and an adjustment procedure to take into account the location uncertainty in the spatial predictions is proposed. In this work, we propose a method to deal with location uncertainties and that complements the methodology of [69], with the objective of applying it to the radio environment of wireless communications systems. The methodology of [69] is based on the adjustment of the covariances of the Kriging system through the use of the location error statistics $g(\cdot)$. Thus, the adjusted covariances between intended sites $\mathbf{s}$, denoted as $C_g(\mathbf{s}_i, \mathbf{s}_j)$, and intended sites $\mathbf{s}$ and target coordinate $\mathbf{s}_0$, denoted as $C_g(\mathbf{s}_i, \mathbf{s}_0)$, are obtained through

$$\begin{aligned} C_g(\mathbf{s}_i, \mathbf{s}_j) &= C_g(\mathbf{h}) = \int \int C(\mathbf{h} + \mathbf{v} - \mathbf{u}) g(\mathbf{u}) g(\mathbf{v}) d\mathbf{u} d\mathbf{v}, \\ C_g(\mathbf{s}_i, \mathbf{s}_0) &= C_{g0}(\mathbf{h}) = \int C(\mathbf{h} + \mathbf{u}) g(\mathbf{u}) d\mathbf{u}, \end{aligned} \quad (20)$$

where $\mathbf{u} = (u_x, u_y)$ is a location error vector in position $\mathbf{s}$, and $\mathbf{v} = (v_x, v_y)$ is a location error vector in $\mathbf{s} + \mathbf{h}$. These integrals are analytically intractable, even for simple combinations of covariance and statistical distributions of location errors. In an attempt to solve this problem, as pointed out in [69] [72], we adopted the Monte Carlo Integration approach to evaluate (20) and reach approximations for adjusted covariances.

The practical implementation to solve such integrals is accomplished through an approximation of the covariances between the measurements from the generation of M *i.i.d.* location error sample vectors with distribution $g(\cdot)$ and described by $\mathbf{u}_i = (u_x, u_y) \sim g(\mathbf{u})$ and $\mathbf{v}_i = (v_x, v_y) \sim g(\mathbf{v})$, with $i = 1, \dots, M$, according to

$$\begin{aligned} C_g(\mathbf{s}_i, \mathbf{s}_j) &\approx \frac{1}{M^2} \sum_{k=1}^M \sum_{l=1}^M C(\mathbf{s}_i + \mathbf{v}_k - \mathbf{u}_l, \mathbf{s}_j + \mathbf{v}_k - \mathbf{u}_l) \\ C_{g0}(\mathbf{s}_i, \mathbf{s}_0) &\approx \frac{1}{M} \sum_{k=1}^M C(\mathbf{s}_i + \mathbf{u}_k, \mathbf{s}_0 + \mathbf{u}_k), \text{ with } i, j = 1, \dots, N. \end{aligned} \quad (21)$$

The use of data with location errors $\{Z_g(\mathbf{s}_1), \dots, Z_g(\mathbf{s}_N)\}$ combined with Kriging weights obtained from covariance $C(\cdot)$ without adjustments and applied to the spatial predictor in (14) characterize the Kriging ignoring location error (KILE) predictor. In another way, the use of data with location errors combined with Kriging weights that were obtained from the adjusted covariance $C_g(\cdot)$ in the system of Kriging equations is denoted as a Kriging adjusted for location error (KALE). The system of adjusted Kriging equations is constructed from adjusted covariances, given by $\Gamma_g \lambda_g = \beta_g$. Thus, the KALE optimal weights are obtained from the matrix inversion

Fig. 6: Location error effects: (a) dispersion of RV due to the location errors; (b) semivariogram without location errors; (c) semivariogram with location errors; (d) semivariogram adjusted for location error (SALE).

Fig. 7: Histogram of covariance parameter estimation: (a) *sill* estimation with location errors; (b) *sill* estimation based on SALE method; (c) *range* estimation with location errors; (d) *range* estimation based on the SALE method.

calculation $\lambda_g = \Gamma_g^{-1} \beta_g$. The compact form of both predictors can be described according to:

$$\begin{aligned} Z_{KILE}(s_0) &= \lambda^T Z_g, \\ Z_{KALE}(s_0) &= \lambda_g^T Z_g. \end{aligned} \quad (22)$$

With spatial predictors in (22) and the average received power estimation, the prediction of the received power in unknown coordinates can be obtained according to

$$\begin{aligned} \hat{P}_{KILE}(s_0) &= \hat{\mu}(s_0) + Z_{KILE}(s_0), \\ \hat{P}_{KALE}(s_0) &= \hat{\mu}(s_0) + Z_{KALE}(s_0). \end{aligned} \quad (23)$$

On the spatial predictions in (22-23), two important points should be highlighted: i) the covariance adjustments of the KALE predictor are performed in the system of Kriging equations and ii) the covariance parameters used in this system of equations can be perfectly known or estimated, and in the latter case, they are affected by the location uncertainty. This means that the lack of knowledge about covariance parameters θ will have some impact on the Kriging systems and, consequently, on the spatial predictions. In this sense, this work proposes a method that complements the KALE methodology considering the location uncertainty knowledge when the covariance parameters are unknown. Specifically, we propose that the statistical distribution $g(\cdot)$ is also taken into account in the method of moments related to the experimental semivariogram, which we called the semivariogram adjusted for location error (SALE) method. In other words, we proposed additional modifications to the experimental semivariogram estimation procedure to include location error statistics. With the knowledge of the probability distribution

of location errors, $g(\cdot)$, the location error sample vectors are used to perform a random perturbation in the distance matrix used in the estimation of the experimental semivariogram. The algorithm and quantization procedures for obtaining the Lag distances, based on [76], are applied on the disturbed distances matrix, resulting in adjusted Lag distances and semivariances. The results using the SALE method are shown in Fig.6(d) and although the adjusted semivariogram is still different from the semivariogram obtained without location uncertainty, it is more smoothed when compared to the semivariogram that was corrupted by the location errors, shown in Fig.6(c).

Our main motivation for using the SALE method combined with the KALE predictor, namely S-KALE predictor, is presented in Fig.7. The histograms of Figs.7(a) and 7(b) show that the *sill* estimation was not significantly affected by the location uncertainty while the LS estimation related to the *range* parameter, shown in Fig.7(c), was significantly affected. On the other hand, it is possible to observe in the histogram of Fig.7(d) that the SALE method allowed to reach a result closer to the optimal value for the *range* parameter estimation. Similar results were obtained with several radio environment configurations (e.g., coverage sizes) and different values for the *sill* and *range* parameters. This motivates us to use the S-KALE predictor in the circumstances where covariance parameter estimation is required. Although we are considering only the large-scale effects ($c_0 = 0$), it was observed that when the spatial random process has higher *nugget* values, comparable to the values of the sill variance (i.e., pure nugget effect with extremely discontinuous semivariograms), the SALE method performs equivalent to the case where the estimated semivariogram is not adjusted.

VI. SIMULATION RESULTS

This section presents the Monte Carlo simulation results related to the performance of KILE, KALE and S-KALE spatial predictors. The simulations are based on $L = 10,000$ independent realizations through the following steps:

- (i) For each realization, an initial spatial sampling strategy with uniform distribution is generated with the intended sites $\{s_1, \dots, s_N\}$. Location errors $\mathbf{p}(s)$ are generated according to a Gaussian distribution with location error standard deviation σ_p and corrupt the intended locations resulting in the set of realized sites $\{\mathbf{r}_1, \dots, \mathbf{r}_N\}$.
- (ii) For each realization, the radio environment is generated by combining the trend (with average path loss effects) and spatial random variations (shadowing effects) in the realized spatial coordinates $\mathbf{r}$ and a target coordinate s_0 is randomly chosen for the performance evaluation of the spatial predictors.
- (iii) For each realization, the Kriging method is used to predict the channel shadowing value in s_0 , using: (a) the intended sites and isotropic-exponential covariance without adjustment, i.e., the KILE predictor; (b) the intended sites and the adjusted isotropic-exponential covariance, i.e., the KALE predictor and (c) the intended sites with adjusted semivariogram and isotropic-exponential covariance, i.e., the S-KALE predictor.
- (iv) The average received power estimation is added to the spatial predictions (i.e., Z_{KILE} , Z_{KALE} and Z_{S-KALE}) and, therefore, the results can be compared with the true values of the spatial random process according to

$$\text{MSPE} = \frac{1}{L} \sum_{l=1}^L [\hat{P}^l(s_0) - P^l(s_0)]^2. \quad (24)$$

where MSPE is the mean square prediction error and it is empirically estimated through the average over L independent realizations. The objective is to verify the performance of the spatial predictors under analysis in relation to the number of

measurements required to achieve a given performance as well as the robustness to the location uncertainty. Although it is not necessarily an optimal metric, MSPE is widely used in the context of performance analysis of REM spatial predictions, especially in wireless communications systems as in [29-30], [41-42]. In fact, MSPE is a mean indicator of the dispersion of spatial prediction errors in REM generation process.

First, the *a priori* knowledge of the covariance parameters θ of the wireless shadowing is assumed and a performance analysis is conducted to verify whether taking into account the location uncertainty in spatial predictions can yield better results. Therefore, the first performance analysis is made on the KILE and KALE spatial predictors. Typical wireless communications settings are chosen with average path loss with $\alpha = 3$, wireless shadowing with known parameters $\theta = [49, 100, 0]$ and coverage size of $500\text{m} \times 500\text{m}$.

Initially, the variable to be analyzed consists of the number of measurements N while the location errors are incorporated into the simulation with Gaussian distribution and fixed standard deviation, $\sigma_p = 40\text{m}$. The perfect and imperfect average path loss estimation are considered, characterizing Scenarios A and B, respectively. The results are shown in Fig.8 in which it can be seen that taking into account the location error statistics on spatial predictions produced better performance. Although the increase in N favored both predictors, the location errors significantly affected KILE performance, while KALE benefited from more measurements. For situations where $N > 100$, the MSPE squared reduction of KALE over KILE was more than 7dB with imperfect trend estimation. Therefore, we can mention that taking into account the location error statistics is more efficient than ignoring them as N increases. Moreover, the ability of predictors to achieve reasonable performance with the least number of measurements as possible (i.e., predictor efficiency) is the key factor in choosing the predictor for REM generation, especially in systems whose architecture does not allow to send many measurements due to bandwidth or energy resources constraints.

Fig. 8: Influence of N on the performance of KILE and KALE spatial predictors with perfect (Scenario A) and imperfect (Scenario B) average path loss estimation and known covariance parameters.

Another important aspect is the robustness of the spatial predictor against the severity of the location errors. In fact, despite the notorious evolution of network-based location estimation methods, channel adversities make it difficult to obtain accurate location estimates, resulting in an uncertainty in reported positioning of network devices including the possibility of generating outliers and invalid location information in the wireless network. Thus, the robustness of the spatial predictor against the negative effects of the location errors is also an important feature for REM design.

In this sense, a comparative analysis is performed considering the gradual increase of the standard deviation of the location errors σ_p , assuming a fixed set consisting of $N = 100$ measurements collected in the radio environment for both KILE and KALE predictors. The same configuration settings as the previous simulation were chosen. The results are shown in Fig. 9 and the cases in which the perfect and imperfect trend estimation occurs are represented by the Scenarios C and D, respectively. The first conclusion that can be drawn from the plots is that KALE has proved to be more robust against location errors when compared to KILE. It is possible to notice MSPE squared reductions of the order of 4dB when the location errors have $\sigma_p = 40\text{m}$ and more than 5dB when the path loss trend estimation is imperfect. This reaffirms that ignoring location errors while attempting to capture the spatial variability for the Kriging predictor is not an efficient decision from the point of view of the accuracy of the predictions. In fact, what is occurring is that the difference between the spatial covariance with and without location uncertainty increases as long as the location error severity is also increased. Therefore, ignoring location uncertainty fails significantly to capture the spatial variability of the radio environment.

At this point, it is important to note that the results obtained consider the complete knowledge about covariance parameters of the wireless communication channel. In practice, the lack of knowledge about θ consists of a more realistic aspect to be applied on the system model. A simulation is conducted

incorporating the covariance parameter estimation through the semivariogram and least squares estimation to verify the impact on both KILE and KALE predictors as well as the performance of proposed S-KALE predictor. In addition, for comparative purposes with other REM method, we compare the Kriging methods to IDW technique [1], [4], which is essentially based on the deterministic calculation of the inverse of the distance between the measurements collected and the target coordinate s_0 . The results are shown in Fig.10 and the two cases of perfect and imperfect average path loss estimation are represented by the Scenarios E and F, respectively. It is possible to note that when the covariance parameters are estimated first, the difference in performance between KALE and KILE is reduced. Also, the difference between the Scenarios E and F shows that the trend estimation affected the performance of all the predictors. In fact, both KILE and KALE were significantly affected and the IDW technique had the worst performance, except for KILE in a regime of high location uncertainty. For instance, in Scenario F with $\sigma_p = 40\text{m}$, the KILE is 3dB worse than the IDW and KALE has about the same MSPE performance, while the S-KALE is 4dB better than it. In other words, it is still advantageous to proceed with the adjustment on KALE, but the performance improvement is not substantial and depends significantly on the results of the covariance parameter estimation. These results corroborate the investigations carried out in [72] and show that the covariance parameter estimation plays an important role on spatial predictions subject to location uncertainty. The performance results point out that the use of the location error statistics in the semivariogram estimation (SALE) related to the S-KALE overcomes the performance of KALE in the circumstances where the covariance estimation is required. Our simulation analyzes also revealed the superiority of S-KALE in different configurations of the radio environment (e.g., different coverage areas and number of measurements) as well as different values of the parameters *range* and *sill* of the wireless communication channel.

Fig. 9: Influence of location uncertainty on the performance of KILE and KALE spatial predictors with perfect (Scenario C) and imperfect (Scenario D) average path loss estimation and known covariance parameters.

Fig. 10: Influence of location uncertainty on the performance of KILE, KALE and S-KALE spatial predictors with perfect (Scenario E) and imperfect (Scenario F) average path loss estimation and unknown covariance parameters.

Although the different approach in the estimation process of covariance parameters, these results corroborates the conclusions presented in references [41] and [42]. Specifically, regarding our work, [42] presents some differences, including: i) the use of the ML technique to estimate the wireless channel parameters considering location uncertainty at this step, whereas our work incorporates the location uncertainty in the classical geostatistical framework based on the semivariogram and the least squares estimators; ii) the use of closed-form expressions for Gaussian covariance functions with location uncertainty, whereas our work uses the Monte Carlo integration methods to accommodate different spatial covariance combinations with statistical distributions of location errors in the Kriging predictor and iii) all performance analyzes are based on one-dimensional environments, while our work considers two-dimensional radio environments maps.

Still in relation to other works, the system model of [41] presents different characteristics in relation to our work due to the Bayesian approach and the shadowing decomposition of the system model with several spatial base functions to perform the fixed rank Kriging method. The ML estimation is also used and a new strategy is proposed taking into account the posterior distribution of the observations using modifications based on the stochastic approximated Expectation-Maximization (EM) algorithm for parameter estimation. This brought the possibility of the conditional expectation predictor being obtained, which is non-linear and more complex, and presents better results compared to the BLUP predictor used in our work (albeit with a different system model), especially in a regime with high severity for location errors. Indeed, dealing with the location uncertainty in the covariance parameter estimation is a plausible initiative, but as pointed out by Diggle and Fanshawe [73] and in [74-75], the intractability of the integrals involved in parameter estimation implies analytical and computational challenges when the positioning uncertainty is present in the system model.

In this sense, the S-KALE predictor proposed in this work points out an alternative and low complexity solution applied for large-scale radio environments, since the performance results were considerably better in several cases when compared with deterministic techniques, such as IDW, and conventional Kriging methods. In other words, the potential of our work lies in the use of techniques that reduce the complexity of the channel parameter estimation in the geostatistical model when there is no *a priori* knowledge of the covariance parameters of wireless channel subject to location uncertainty.

VII. CONCLUSIONS

In this manuscript, we have applied the Kriging spatial predictor, a well-known technique in geostatistics, to obtain a REM. After formulating the spatial prediction problem, we have shown the necessary steps to generate the REM through the geostatistical based framework. We introduce the location errors in the system model to make the prediction problem more realistic. Moreover, we consider an important assumption on the lack of knowledge about the covariance parameters of the wireless channel. In this sense, it was possible to verify that the average path loss trend and covariance parameter estimation processes plays an important role when subjected to location errors and significantly affect the KILE and KALE performance. To deal with these problems, we propose the the S-KALE predictor, which took into account the location error statistical distribution in parameter estimation, through the adjusted semivariogram procedure. The simulation results showed that the S-KALE predictor was able to provide superior performance when compared to KALE and KILE predictors when covariance parameter estimation is required, subject to location uncertainty. On future research directions, practical research on REM (e.g., [4], [26], [83-84]) and investigations on the influence of location uncertainty on structured spatial covariance are important points in spatial prediction problems applied to the wireless communications.

REFERENCES

- [1] H. B. Yilmaz, T. Tugcu, F. Alagöz and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," in *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162-169, December 2013, doi: 10.1109/MCOM.2013.6685772.
- [2] P. R. Jordi, Z. Andreas, B. Lila, K. Adrian, K. Katerina, D. Nikos, K. Reben, "On the use of radio environment maps for interference management in heterogeneous networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 184-191, August 2015, doi: 10.1109/MCOM.2015.7180526.
- [3] H. Braham; S. Ben Jemaa; G. Fort; E. Moulines; B. Sayrac, "Fixed Rank Kriging for Cellular Coverage Analysis," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4212-4222, May 2017, doi: 10.1109/TVT.2016.2599842.
- [4] A. Umberto, F. Casadevall and E. G. Rodriguez, "An outdoor TV band Radio Environment Map for a Manhattan like layout," *2016 International Symposium on Wireless Communication Systems (ISWCS)*, Poznan, 2016, pp. 399-403, doi: 10.1109/ISWCS.2016.7600936.
- [5] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijärvi and P. Mähönen, "Automated coverage hole detection for cellular networks using radio environment maps," *2013 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, Tsukuba Science City, 2013, pp. 35-40.
- [6] H. Braham, S. B. Jemaa, B. Sayrac, G. Fort and E. Moulines, "Low complexity spatial interpolation for cellular coverage analysis," *2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Hammamet, 2014, pp. 188-195, doi: 10.1109/WIOPT.2014.6850298.
- [7] S. Sodagari, "A Secure Radio Environment Map Database to Share Spectrum," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1298-1305, Oct. 2015, doi: 10.1109/JSTSP.2015.2426132.
- [8] A. H. Sayed, A. Tarighat and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," in *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24-40, July 2005, doi: 10.1109/MSP.2005.1458275.
- [9] H. Celebi and H. Arslan, "Utilization of Location Information in Cognitive Wireless Networks," in *IEEE Wireless Communications*, vol. 14, no. 4, pp. 6-13, August 2007, doi: 10.1109/MWC.2007.4300977.
- [10] Y. Zhao, J. Gaedert, K. K. Base, J. H. Reed, "Radio Environment Map Enabled Situation-Aware Cognitive Radio Learning Algorithms," *Proceeding of the SDR 06 - Technical Conference and Product Exposition, Wireless@Virginia Tech*, 2006.
- [11] Y. Zhao, L. Morales, J. Gaedert, K. K. Bae, J. S. Um and J. H. Reed, "Applying Radio Environment Maps to Cognitive Wireless Regional Area Networks," *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 115-118, Dublin, 2007, doi: 10.1109/DYSPAN.2007.22.
- [12] J. P. Chilès and P. Delfine, *Geostatistics Modeling Spatial Uncertainty*, Second Edition, John Wiley & Sons, 2012.
- [13] H. Omre and H. B. Kjetil, "The Bayesian bridge between simple and universal kriging", *Mathematical Geology*, vol. 21, no. 7, pp. 767-786, 1989, doi: 10.1007/BF00893321.
- [14] M. A. Oliver and R. Webster, *Basic Steps in Geostatistics: The Variogram and Kriging*, SpringerBriefs in Agriculture, 2015, doi: 10.1007/978-3-319-15865-5.
- [15] J. M. Montero, G. G. Avilés, J. Mateu, *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*, University of Castilla-La Mancha, Spain and University Jaume I of Castellón, Spain, Wiley & Sons, 2015, doi: 10.1002/9781118762387.
- [16] N. Cressie, *Statistics for Spatial Data*, Wiley, revised edition, 1993, doi: 10.1002/9781119115151.
- [17] P. J. Diggle, and P. J. Ribeiro, *Model-based Geostatistics*, Springer Science Business Media, LLC, 2007, doi: 10.1007/978-0-387-48536-2.
- [18] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*, Springer-Verlag Berlin Heidelberg, 2003, doi: 10.1007/978-3-662-03550-4.
- [19] E. H. Isaaks and R.M. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, 1989.
- [20] P. K. Kitanidis, *Introduction to Geostatistics: Applications in Hydrogeology*, Cambridge University Press, 1997.
- [21] M. Malmirchegini and Y. Mostofi, "On the Spatial Predictability of Communication Channels," in *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 964-978, March 2012, doi: 10.1109/TWC.2012.012712.101835.
- [22] S. S. Szyszkowicz, H. Yanikomeroglu and J. S. Thompson, "On the Feasibility of Wireless Shadowing Correlation Models," in *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4222-4236, Nov. 2010, doi: 10.1109/TVT.2010.2082006.
- [23] J. Riihijärvi and P. Mähönen, "Estimating wireless network properties with spatial statistics and models," *10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, Paderborn, Germany, 2012, pp. 331-336.
- [24] J. Riihijärvi, P. Mahonen, M. Wellens and M. Gordziel, "Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields," *IEEE 19th International Symposium on Personal, Indoor and M. R Communications*, pp. 1-6, Cannes, 2008, doi: 10.1109/PIMRC.2008.4699912.
- [25] M. Mirahsan, R. Schoenen, S. S. Szyszkowicz and H. Yanikomeroglu, "Measuring the spatial heterogeneity of outdoor users in wireless cellular networks based on open urban maps," *IEEE International Conference on Communications (ICC)*, pp. 2834-2838, London, 2015, doi: 10.1109/ICC.2015.7248756.
- [26] N. Perpinias, A. Palaos, J. Riihijärvi and P. Mähönen, "A measurement-based study on the use of spatial interpolation for propagation estimation," *IEEE International Conference on Communications (ICC)*, pp. 2715-2720, London, 2015, doi: 10.1109/ICC.2015.7248736.
- [27] A. Konak, "Estimating path loss in wireless local area networks using ordinary kriging," *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, MD, 2010, pp. 2888-2896, doi: 10.1109/WSC.2010.5678983.
- [28] E. Dall'Anese, S. J. Kim and G. B. Giannakis, "Channel Gain Map Tracking via Distributed Kriging," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1205-1211, March 2011, doi: 10.1109/TVT.2011.2113195.
- [29] K. Sato and T. Fujii, "Kriging-Based Interference Power Constraint: Integrated Design of the Radio Environment Map and Transmission Power," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 1, pp. 13-25, March 2017, doi: 10.1109/TCCN.2017.2653189.
- [30] K. Sato and T. Fujii, "Kriging-Based Interference Power Constraint for Spectrum Sharing Based on Radio Environment Map," *IEEE Globecom Workshops (GC Wkshps)*, pp. 1-6, San Diego, CA, 2015, doi: 10.1109/GLOCOMW.2015.7414008.
- [31] Z. Hou, Y. Zhou, L. Tian, J. Shi, Y. Li, B. Vucetic, "Radio Environment Map Aided Doppler Shift Estimation in LTE-Railway," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp.1-1, doi: 10.1109/TVT.2016.2599558.
- [32] Y. Zhou, "Radio environment map based maximum a posteriori Doppler shift estimation for LTE-R," *High Mobility Wireless Communications (HMWC)*, International Workshop on, Beijing, 2014, pp. 5-5.
- [33] F. Paisana, Z. Khan, J. Lehtomäki, L. A. Da Silva and R. Vuoltoniemi, "Exploring radio environment map architectures for spectrum sharing in the radar bands," *23rd International Conference on Telecommunications*, 2016, doi: 10.1109/ICT.2016.7500412.
- [34] A. Landström and J. V. de Beek, "Transmitter localization for 5G mmWave REMs by stochastic generalized triangulation," *23rd International Conference on Telecommunications (ICT)*, Thessaloniki, 2016, pp. 1-5, doi: 10.1109/ICT.2016.7500413.
- [35] M. T. Ozsahin and T. Tugcu, "Transmitter location estimation for radio environment map construction using software defined radio," *IEEE 21st Symposium on Communications and Vehicular Technology*, Benelux (SCVT), Delft, 2014, pp. 94-99, doi: 10.1109/SCVT.2014.7046715.
- [36] V. Rakovic, M. Angjelinoski, V. Atanasovski and L. Gavrilovska, "Location estimation of radio transmitters based on spatial interpolation of RSS values," *Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 7th International ICST Conference, Stockholm, 2012, pp. 242-247, doi: 10.4108/icst.crowncom.2012.248454.
- [37] 3GPP TR 36.805 v1.3.0 1, "Study on minimization of drive-tests in next generation networks (release 9)," tech. rep., 3rd Generation Partnership Project, 2009.
- [38] J. Johansson, W. A. Hapsari, S. Kelley and G. Bodog, "Minimization of drive tests in 3GPP release 11," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 36-43, November 2012, doi: 10.1109/MCOM.2012.6353680.
- [39] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula and B. Seibre, "Minimization of drive tests solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28-36, June 2012, doi: 10.1109/MCOM.2012.6211483.
- [40] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijärvi and P. Mähönen, "Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks," *8th International Conference on Cognitive Radio Oriented Wireless Networks*, Washington, DC, 2013, pp. 37-42, doi: 10.1109/CROWNCom.2013.6636791.

- [41] H. Braham, S. B. Jemaa, G. Fort, E. Moulines and B. Sayrac, "Spatial Prediction Under Location Uncertainty in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7633-7643, Nov. 2016, doi: 10.1109/TWC.2016.2605676.
- [42] L. S. Muppirisetty, T. Svensson and H. Wymeersch, "Spatial Wireless Channel Prediction under Location Uncertainty," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1031-1044, Feb. 2016, doi: 10.1109/TWC.2015.2481879.
- [43] M. Jadhaliha, Y. Xu, J. Choi, N. S. Johnson and W. Li, "Gaussian Process Regression for Sensor Networks Under Localization Uncertainty," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 223-237, Jan. 15, 2013, doi: 10.1109/TSP.2012.2223695.
- [44] T. Farnham, "Radio environment map techniques and performance in the presence of errors," *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Valencia, 2016, pp. 1-6, doi: 10.1109/PIMRC.2016.7794911.
- [45] J. B. Andersen, T. S. Rappaport and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 42-49, Jan 1995, doi: 10.1109/35.339880.
- [46] M. F. Iskander and Zhengqing Yun, "Propagation prediction models for wireless communication systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 662-673, Mar 2002, doi: 10.1109/22.989951.
- [47] A. Neskovic, N. Neskovic and G. Paunovic, "Modern approaches in modeling of mobile radio systems propagation environment," *IEEE Communications Surveys & Tutorials*, vol. 3, no. 3, pp. 2-12, Third Quarter 2000, doi: 10.1109/COMST.2000.5340727.
- [48] IEEE VT-S Propagation Committee. Wireless Planning Tools for Mobility and Indoor Systems.
- [49] D. A. Guimarães, *Digital Transmission, A Simulation-Aided Introduction with VisSim/Comm*. Signals and Communications Technology. Springer, 2009, doi: 10.1007/978-3-642-01359-1.
- [50] S. Haykin and M. Moher, *Modern Wireless Communication*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA. 2004.
- [51] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145-2146, 7 Nov. 1991, doi: 10.1049/el:19911328.
- [52] A. Goldsmith, *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005, doi: 10.1017/CBO9780511841224.
- [53] M. Gudmundson, "Analysis of handover algorithms [microcellular radio]," *Vehicular Technology Conference, Gateway to the Future Technology in Motion*, 41st IEEE, St. Louis, MO, 1991, pp. 537-542, doi: 10.1109/VETEC.1991.140549.
- [54] D. Giancristofaro, "Correlation model for shadow fading in mobile radio channels," *Electronics Letters*, vol. 32, no. 11, pp. 958-959, 23 May 1996, doi: 10.1049/el:19960655.
- [55] ETSI TR 101.112 v3.2.0, "Selection procedures for the choice of radio transmission technologies of the UMTS," April 1998.
- [56] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4024-4036, August 2009, doi: 10.1109/TWC.2009.071293.
- [57] C. Xiaodong and G. B. Giannakis, "A two-dimensional channel simulation model for shadowing processes," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 6, pp. 1558-1567, Nov. 2003, doi: 10.1109/TVT.2003.819627.
- [58] H. Claussen, "Efficient modelling of channel maps with correlated shadow fading in mobile radio systems," *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, 2005, pp. 512-516, doi: 10.1109/PIMRC.2005.1651489.
- [59] T. Ali and M. Saquib, "Numerically efficient generation of exponentially correlated two-dimensional shadowing," *IEEE International Conference on Communications (ICC)*, London, 2015, pp. 4265-4271, doi: 10.1109/ICC.2015.7248993.
- [60] R. Fraile, J. F. Monserrat, J. Gozálviz, and N. Cardona, "Mobile radio bi-dimensional large-scale fading modelling with site-to-site cross-correlation," *Eur. Trans. Telecomm.*, 2008, pp.101-106, doi: 10.1002/ett.1179.
- [61] J. F. Monserrat, R. Fraile, N. Cardona and J. Gozálviz, "Effect of Shadowing Correlation Modeling on the System Level Performance of Adaptive Radio Resource Management Techniques," *2nd International Symposium on Wireless Communication Systems*, Siena, 2005, pp. 460-464, doi: 10.1109/ISWCS.2005.1547743.
- [62] I. Forkel, M. Schinnenburg, and M. Ang, "Generation of two-dimensional correlated shadowing for mobile radio network simulation," *WPMC*, vol.2, Sept. 2004, pp. 314-319.
- [63] N. Cressie, "Fitting variogram models by weighted least squares," *Journal of the International Association for Mathematical Geology*, vol. 17, no. 5, pp. 563-586, 1985, doi: 10.1007/BF01032109.
- [64] R. A. Olea, *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, 1999, doi: 10.1007/978-1-4615-5001-3.
- [65] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Inc., 1993.
- [66] F. F. Pitard, "Exploration of the 'Nugget Effect'," *Geostatistics for the Next Century: An International Forum in Honour of Michel David's, Contribution to Geostatistics*, Vol. 6, pp 124-136, Montreal, 1993, doi: 10.1007/978-94-011-0824-9_16.
- [67] J. J. Moré, D.C. Sorensen, "Computing a Trust Region Step," *SIAM Journal on Scientific and Statistical Computing*, Vol. 3, pp 553-572, 1983, doi: 10.1137/0904038.
- [68] J.J. Moré, "The Levenberg-Marquardt Algorithm: Implementation and Theory." *Numerical Analysis*, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer Verlag, 1977, pp. 105-116.
- [69] N. Cressie and J. Kornak, "Spatial statistics in the presence of location error with an application to remote sensing of environment," *Statistics Sciences*, vol. 18, no. 4, pp. 436-465, Nov. 2003, doi: 10.1214/ss/1081443228.
- [70] J. G. Gabrosek, "The effect of locational uncertainty in geostatistics," Ph.D. Dissertation, Retrospective Theses and Dissertations. 12128, Digital Repository @ Iowa State University, 1999.
- [71] J. Gabrosek, and N. Cressie. "The effect on attribute prediction of location uncertainty in spatial data," *Geographical Analysis*, vol. 34, no. 3, The Ohio State University, July, 2002, doi: 10.1111/j.1538-4632.2002.tb01088.x.
- [72] D. Cervone and N. S. Pillai, "Gaussian process regression with location errors," from the arXiv database, arXiv: 1506.08256 [stat.ME], June, 2015.
- [73] T. R. Fanshawe, and P. J. Diggle, "Spatial prediction in the presence of positional error," *Environmetrics*, 22: 109-122, 2011, doi: 10.1002/env.1062.
- [74] J. J. Barber, A. E. Gelfan and J. A. Silander, "Modelling map positional error to infer true feature location," *Canadian Journal of Statistics*, 2006, doi: 10.1002/cjs.5550340407.
- [75] B. François, "Parametric estimation of covariance function in Gaussian-process based Kriging models. Application to uncertainty quantification for computer experiments," Ph.D. Thesis, *Statistics [math.ST]*. Université Paris-Diderot - Paris VII, 2013.
- [76] R. Augusto, C. M. Panazio, "Sobre a Geração de Mapas de Ambiente de Rádio com Kriging Ordinário," XXXIV Brazilian Symposium on Telecommunications (SBrT), August, 2016, Santarém, Pará, Brazil.
- [77] N. Cressie, G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70: 209-226, 2008, doi: 10.1111/j.1467-9868.2007.00633.x.
- [78] J. R. Bradley, N. Cressie, T. Shi, "A comparison of spatial predictors when datasets could be very large," *Statistics Surveys*, 10, 100-131, 2016, doi: 10.1214/16-SS115.
- [79] M. Pesko, T. Javornik, L. Vidmar, A. Košir, M. Štular and M. Mohorčič, "The indirect self-tuning method for constructing radio environment map using omnidirectional or directional transmitter," *EURASIP Journal on Wireless Communications and Networking*, , Nov. 2015, doi: 10.1186/s13638-015-0297-2.
- [80] J. M. V. García, "Localization in Wierless Sensor Networks Using a Mobile Robot," Ph.D Thesis, Aalto University School of Electrical Engineering, Finland, 2016.
- [81] S. Grimoud, B. Sayrac, S. Ben Jemaa and E. Moulines, "Best Sensor Selection for an Iterative REM Construction," 2011 IEEE Vehicular Technology Conference (VTC Fall), San Francisco, CA, 2011, pp. 1-5, doi: 10.1109/VETECF.2011.6093004.
- [82] Zhenghao Zhang, Husheng Li and Changxing Pei, "Optimum experimental design for estimating spatial variogram in cognitive radio networks," 44th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, 2010, pp. 1-6, doi: 10.1109/CISS.2010.5464833.
- [83] E. Dutkiewicz, Y. He, B. A. Jayawickrama and H. V. Abeywickrama, "Radio environment maps generation and spectrum sensing testbed for spectrum sharing in 5G networks," *IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, Verona, 2017, pp. 33-36, doi: 10.1109/APWC.2017.8062233.
- [84] A. Kliks, P. Kryszkiewicz, A. Umbert, J. Pérez - Romero, F. Casadevall and L. Kulacz, "Application of Radio Environment Maps for Dynamic Broadband Access in TV Bands in Urban Areas," *IEEE Access*, vol. 5, pp. 19842-19863, 2017, doi: 10.1109/ACCESS.2017.2751138.

Ricardo Augusto received Engineer's and Master's degrees in Telecommunication Engineering from the National Institute of Telecommunication (Inatel), Brazil, in 2010 and 2013, respectively. Currently, he is a PhD student at Escola Politécnica of the University of São Paulo (USP), Brazil. Since 2010, he has been working as a specialist in communications systems in the training and consulting area of Inatel. His research interests include wireless communications systems, statistical signal processing, geolocation, and machine learning techniques.

Cristiano Panazio received Engineer's and Master's degrees in Electrical Engineering in 1999 and 2001, respectively, from State University of Campinas and a PhD degree in Telecommunications in 2005 from the Conservatoire National des Arts et Métiers. Currently, he is assistant professor at the Laboratory of Communications and Signals of the Escola Politécnica of the University of São Paulo. He is a member of the Brazilian Telecommunications Society (SBrT) and has served as an Associate Editor of the IEEE Transactions on Circuits and Systems - I from 2015 to 2017. His research interests are adaptive equalization (data aided and blind), synchronization techniques, mobile communications, multicarrier modulation and space-time transceivers.

Sobre a Geração de Mapas de Ambiente de Rádio com Kriging Ordinário

Ricardo Augusto e Cristiano Panazio

Resumo — Este artigo apresenta e discute importantes ferramentas da geoestatística utilizadas para a geração de mapas de ambiente de rádio (REM – *Radio Environment Maps*) em sistemas de comunicações sem fio. A partir de um processo de amostragem espacial, o método Kriging Ordinário (KO) foi utilizado para gerar o REM. Resultados de simulação mostram a influência de parâmetros relacionados à modelagem do ambiente de rádio na precisão de geração do REM. Sobre esta perspectiva, são apontados desafios e limitações relacionados à geração e uso do REM em aplicações de sistemas de comunicações sem fio.

Palavras-Chave – Kriging, Geoestatística, REM, Semivariograma.

Abstract — This article presents and discusses important tools of geostatistics used for the generation of Radio Environment Maps (REM) in wireless communications systems. From a spatial sampling process, Ordinary Kriging method was used to generate REM. Simulation results show the influence of parameters related to the modelling of radio environment on the accuracy of the generation of REM. On this perspective, challenges and limitations related to the generation and use of REM in wireless communications systems are point out.

Keywords – Kriging, Geostatistics, REM, Semivariogram.

I. INTRODUÇÃO

O mapa de ambiente de rádio (REM, em inglês), conceito proposto por [1], consiste em um mapa que indica uma grandeza física do ambiente de rádio, sendo a mais comum, a intensidade dos sinais no ambiente de propagação. O REM é gerado a partir de um conjunto de medidas coletadas do ambiente e pode ser consultado para tomadas de decisão relacionadas aos problemas de cobertura em sistemas de comunicações sem fio [2]. Sobre esta perspectiva, o REM pode ser usado em aplicações como: Tarefas de planejamento e monitoramento de cobertura em sistemas de comunicações sem fio [2]; Minimização de *Drive-Tests* (MDT) [3]; Análise e tomada de decisões em sistemas de acesso dinâmico ao espectro [4]; Sensoriamento espectral de faixas licenciadas e não licenciadas em sistemas de rádios cognitivos [5].

Tem-se como principal desafio do REM a obtenção de mapas da forma mais fidedigna possível sobre toda a região analisada, usando um número relativamente pequeno de medidas [6], pois o desempenho obtido nas aplicações subsequentes depende das informações contidas nos mapas. Diferentes métodos, com abordagens determinísticas e estocásticas, são usados para gerar o REM [5]. Neste contexto, a geoestatística se apresenta como um dos métodos mais promissores [3] [4].

Este trabalho possui o objetivo de apresentar, de forma resumida, as ferramentas da geoestatística necessárias para o tratamento do problema de geração do REM, apontando peculiaridades e limitações que podem ser úteis para futuras pesquisas.

Sobre a estrutura do trabalho, a seção II aborda os conceitos elementares da geoestatística. A seção III traz os aspectos da modelagem e geração do REM bem como os resultados de simulações. Finalmente, as conclusões e propostas para trabalhos futuros são mostradas na seção IV.

II. CONCEITOS DE GEOESTATÍSTICA

A geoestatística, por meio dos métodos Kriging, visa realizar previsões espaciais para reproduzir um fenômeno espacial em estudo. Isso é feito a partir da caracterização espacial do fenômeno analisado. Tal fenômeno pode ser descrito por um processo aleatório $Z(s)$ em um domínio D ,

$$\{Z(s), s \in D\} \quad (1)$$

em que s é uma coordenada no plano (x,y) .

A realização da função aleatória espacial, denotada por $z(s)$, é conhecida como variável regionalizada (VR), sendo $\{z(s_i)\}$, com $i=1, 2, \dots, N$ o conjunto de valores regionalizados. Especificamente, tais valores regionalizados são o resultado das realizações de uma coleção de variáveis aleatórias espaciais $Z(s_i)$ que formam o processo aleatório $Z(s)$. Assim, a VR é obtida por intermédio de um processo de amostragem dos valores da realização do campo espacial. Este processo é chamado de regionalização. Neste sentido, a regionalização é o primeiro passo para se obter a caracterização espacial necessária para a posterior previsão espacial.

A. Semivariograma Teórico e Experimental

O raciocínio principal da geoestatística é baseado na ideia de que existe uma relação de dependência espacial entre a coleção de variáveis aleatórias que formam o processo aleatório. Esta relação surge na forma de uma correlação espacial estruturada no campo espacial.

As funções covariância $C(s_i, s_j)$ e semivariograma $\gamma(s_i, s_j)$ são usadas para caracterizar a estrutura espacial dos processos aleatórios espaciais. A covariância entre duas variáveis aleatórias nos pares de coordenadas (s_i, s_j) é definida como

$$C(s_i, s_j) = E\{(Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j))\}, \forall s_i, s_j \in D \quad (2)$$

em que $\mu(s_i)$ é a média da variável aleatória $Z(s_i)$. A função semivariograma de um campo aleatório é definida como sendo a variância das diferenças das variáveis aleatórias que compõem $Z(s)$, i.e.,

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2} \text{Var} \left\{ \left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j) \right] \right\} \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D$$

$$= \frac{1}{2} E \left\{ \left[\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j) \right) - E \left\{ \left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j) \right) \right\} \right]^2 \right\} \quad (3)$$

Neste âmbito, um ponto importante da teoria de processos aleatórios é sua relação com a hipótese de estacionariedade [7] [8]. Especificamente, considerando a estacionariedade no sentido amplo, a caracterização estatística das variáveis aleatórias que formam $Z(\mathbf{s})$ não depende da posição espacial específica ($\mathbf{s}$), mas sim de um vetor de translação espacial ($\mathbf{h}$) entre os pontos espaciais. Como resultado, a função semivariograma torna-se

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var} \left\{ \left[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}) \right] \right\} \quad (4)$$

É importante notar que o semivariograma em (4) retrata a relação entre dois pontos espaciais considerando uma determinada direção espacial. Logo, é possível que esta relação mude com a direção retratando um fenômeno anisotrópico [8]. No caso em que não ocorrem alterações no semivariograma em relação à direção, tem-se o semivariograma isotrópico que depende somente de $|\mathbf{h}|$. Tratar o caso anisotrópico foge do escopo deste trabalho, que assumirá apenas as hipóteses mais simplificadoras de estacionariedade e isotropia.

Neste contexto, ressalta-se que a realização das predições espaciais depende do conhecimento do semivariograma. Contudo, o que se dispõe na prática, é de um conjunto de medidas coletadas que formam a VR, tornando necessária a estimação de (4). Tal estimação é feita aproximando (4) por uma função analítica cujos parâmetros são ajustados de acordo com o semivariograma experimental que será mostrado a seguir.

Em geral, o cálculo do semivariograma experimental é baseado em distâncias de separação pré-definidas, denotadas como Lags [7-9]. Nota-se que é possível fazer o cálculo do semivariograma experimental diretamente com as distâncias de separação entre as medidas, ou seja, sem aproximações, mas isto requer o uso de métodos não convencionais de regressão não-paramétrica¹, que não necessariamente geram resultados melhores do que a técnica descrita a seguir [9].

Os passos abaixo descrevem os pontos mais importantes no cálculo do semivariograma experimental baseado nas distâncias Lags:

(1) É gerada uma matriz Δ , triangular superior, cujos elementos (i,j) são dados pelas distâncias de separação $|\mathbf{h}_{ij}| = |\mathbf{s}_i - \mathbf{s}_j|$ entre todos os pares de pontos que compõem a VR, i.e., para todo $i < j$, sendo $i, j = 1, \dots, N$.

(2) Em seguida, uma distância de separação inicial, denominada Lag_1 é escolhida a partir de Δ . Em geral, a escolha de Lag_1 é baseada em dois critérios [7], sendo a segunda estratégia utilizada neste trabalho:

- ✓ Adota-se Lag_1 como a menor distância de separação espacial encontrada em Δ .
- ✓ Adota-se Lag_1 como a média dos valores mínimos de cada coluna de Δ (excetuando os valores nulos).

(3) Define-se $\text{Lag}_m = m\text{Lag}_1$, com m natural e maior que um, de modo a se realizar uma quantização uniforme das

distâncias $|\mathbf{h}_{ij}|$. Assim, as distâncias $|\mathbf{h}_{ij}|$ são quantizadas com $Q(\mathbf{h}_{ij}) = \text{Lag}_k$ para k que minimiza o valor $|\mathbf{h}_{ij}| - \text{Lag}_k|$.

(4) O método dos momentos de *Matheron* [7] é usado para calcular o valor do semivariograma experimental

$$\gamma_{\text{exp}}(\text{Lag}_m) = \frac{1}{2 \|A_{\text{Lag}-m}\|} \sum_{|\mathbf{s}_i - \mathbf{s}_j| \in A_{\text{Lag}-m}} \left[z(\mathbf{s}_i) - z(\mathbf{s}_j) \right]^2 \quad (5)$$

em que $\|\cdot\|$ denota o operador cardinalidade e $A_{\text{Lag}-m}$ é o conjunto de pares de pontos cujas distâncias quantizadas correspondem à distância de separação Lag_m . O número total m de distâncias de separação depende de $\text{Lag}_{\text{máx}}$, que normalmente é escolhida como metade da distância do mapa analisado [8] [9]. Em geral, têm-se poucos pares de pontos para grandes distâncias de separação. Isso impacta na confiabilidade dos valores do semivariograma experimental para grandes distâncias. Em [8] [9], encontram-se informações detalhadas sobre a forma de cálculo do semivariograma experimental em relação às distâncias de separação e a direção (anisotropia).

Os resultados de estimação do semivariograma experimental, obtidos a partir da (5), são usados em um processo de ajuste a fim de obter um modelo analítico para o semivariograma dos dados. A Figura 1 mostra um exemplo de um modelo analítico ajustado – curva sólida –, a partir de um semivariograma experimental (pontos).

Figura 1: Semivariograma analítico (curva cheia) e experimental (pontos).

Em geral, três parâmetros são ajustados: 1) a variância *sill* retratada pelo patamar ($C + C_0$) do semivariograma; 2) a distância r , indicando a distância de separação a partir da qual se tem uma independência espacial no processo aleatório; 3) o efeito pepita (C_0), que reflete a ocorrência de uma descontinuidade na origem do semivariograma. Em [7], o comportamento dos parâmetros de ajuste do semivariograma analítico é visitado em mais detalhes bem como a apresentação dos modelos analíticos mais utilizados na geoestatística (ex: esférico, seno cardinal e Gaussiano). Neste trabalho, o modelo analítico Gaussiano foi utilizado [7] [8]. O melhor desempenho de ajuste do modelo baseado no semivariograma experimental é o principal objetivo da análise variográfica, uma vez que tal modelo representará a variabilidade espacial do fenômeno em estudo.

B. Predições Espaciais – Método Kriging Ordinário

Na geoestatística, os diferentes métodos Kriging, usados para realizar as predições espaciais, são baseados no modelo analítico do semivariograma e nas pressuposições relacionadas à hipótese de estacionariedade das funções aleatórias. Dependendo destas hipóteses e do conhecimento a

¹ O cálculo do semivariograma experimental sem aproximações de distância de separação resulta na nuvem do semivariograma, comumente utilizada para identificação de *outliers* nos dados [9].

priori sobre a média do processo aleatório, diferentes preditores Kriging são utilizados.

Sobre a pressuposição de estacionariedade no sentido amplo, ressalta-se duas informações: 1) A média do processo aleatório existe e é constante, ou seja, $\mu = \mu(\mathbf{s}) = E\{Z(\mathbf{s})\}$. 2) A covariância do processo aleatório depende somente do vetor de translação $\mathbf{h}$. Assim, quando o valor específico da média μ é conhecido inicialmente, o método Kriging Simples (KS) é utilizado. Entretanto, somente com várias realizações do processo aleatório é possível obter μ . No caso em que μ não é necessariamente conhecida, mas constante (estacionariedade no sentido amplo), o método Kriging Ordinário é utilizado [7].

Dentro deste aspecto, um ponto importante consiste na verificação da dependência de μ quanto à posição espacial ($\mathbf{s}$), retratando a presença de uma tendência $\mu(\mathbf{s})$ no processo aleatório espacial. Neste caso, a média do processo aleatório não é constante e consequentemente, a abordagem da estacionariedade fica comprometida. Sobre estas circunstâncias, o preditor Kriging Universal (KU) pode ser utilizado, desde que, a tendência possa ser decomposta em um conjunto de funções base representadas por monômios com grau que normalmente não excede a potência de dois [7]. Apesar disso, a tendência presente no ambiente de rádio – descrição que é feita na seção III – é modelada com funções não lineares com potências que normalmente são maiores que dois. Logo, optou-se neste trabalho pela realização da estimação e remoção da tendência do processo aleatório antes do processo de Krigagem. Uma vez removida a tendência dos dados, o método KO pode ser utilizado. Especificamente, o preditor Kriging Ordinário, adotado neste trabalho, é definido por

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^N \lambda_i z(\mathbf{s}_i) \quad (6)$$

em que $Z^*(\mathbf{s}_0)$ é o resultado da predição espacial em uma determinada coordenada espacial desconhecida $\mathbf{s}_0$, obtido a partir das medidas $z(\mathbf{s}_i)$ e dos pesos Kriging (λ_i). Estes pesos são calculados sobre a suposição de estacionariedade em segunda ordem (sentido amplo) e, além disso, sob duas condições importantes [7] [8]: 1) que o estimador seja não viesado e 2) que a variância de predição seja mínima.

1) Condição de não viés do KO:

A equação (7) mostra o cálculo relacionado à condição de não viés do preditor KO. O valor esperado da diferença entre a predição espacial $Z^*(\mathbf{s}_0)$ e o valor verdadeiro do processo aleatório no referido ponto $Z(\mathbf{s}_0)$, é zero. Para isso a soma dos valores dos pesos Kriging deve ser unitária.

$$\begin{aligned} E\{Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\} &= E\left\{\sum_{i=1}^N \lambda_i z(\mathbf{s}_i) - Z(\mathbf{s}_0)\right\} = 0 \\ &= \sum_{i=1}^N \lambda_i E\{z(\mathbf{s}_i)\} - E\{Z(\mathbf{s}_0)\} = 0 \\ &= \sum_{i=1}^N \lambda_i \mu - \mu = 0, \text{ logo, } \sum_{i=1}^N \lambda_i = 1 \end{aligned} \quad (7)$$

em que μ é a média (constante) do processo aleatório.

2) Variância de Predição do KO

A variância de predição (também chamada de variância Kriging) é dada por definição como

$$\begin{aligned} \text{Var}\{Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\} &= E\left\{\left[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\right]^2\right\} \\ &\quad - E\left\{\left[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\right]\right\}^2 \end{aligned} \quad (8)$$

Considerando a condição de não viés para o segundo termo e usando a definição do estimador KO, é possível expandir o primeiro termo da equação (8), resultando em

$$\begin{aligned} \text{Var}\{Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\} &= E\left\{\left[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\right]^2\right\} \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j E\{Z(\mathbf{s}_i) Z(\mathbf{s}_j)\} \\ &\quad - 2 \sum_{i=1}^N \lambda_i E\{Z(\mathbf{s}_i) Z(\mathbf{s}_0)\} + E\{Z^2(\mathbf{s}_0)\} \end{aligned} \quad (9)$$

Ressaltando as seguintes definições da covariância

$$\begin{aligned} E\{Z(\mathbf{s}_i) Z(\mathbf{s}_j)\} &= C(\mathbf{s}_i, \mathbf{s}_j) + \mu^2 \\ E\{Z(\mathbf{s}_i) Z(\mathbf{s}_0)\} &= C(\mathbf{s}_i, \mathbf{s}_0) + \mu^2 \\ E\{Z^2(\mathbf{s}_0)\} &= C(\mathbf{0}) + \mu^2 \end{aligned} \quad (10)$$

e substituindo-as em (9), encontra-se a relação entre a variância de predição e a função de covariância, dada por

$$\begin{aligned} \text{Var}\{Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\} &= \sum_{i=0}^N \sum_{j=0}^N \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) \\ &\quad - 2 \sum_{i=0}^N \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) + C(\mathbf{0}). \end{aligned} \quad (11)$$

3) Sistema de Equações do KO

O problema central do KO consiste em encontrar os pesos que minimizam a variância de predição em (9) sob a condição de não viés em (7). A solução se dá com a aplicação do método de Lagrange utilizando a seguinte função lagrangiana

$$\phi(\lambda_i, \beta) = \text{Var}\{Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\} - \beta \left\{ \sum_{i=1}^N \lambda_i - 1 \right\}. \quad (12)$$

em que β é o multiplicador de Lagrange.

O resultado da diferenciação da função em (12) em relação aos pesos Kriging (λ_i) e ao multiplicador de Lagrange, permite obter o sistema de equações da Krigagem Ordinária

$$\sum_{j=1}^N \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - \beta = C(\mathbf{s}_i, \mathbf{s}_0) \text{ e } \sum_{i=1}^N \lambda_i = 1, \quad i = 1, \dots, N \quad (13)$$

Este sistema de equações relaciona a função de covariância entre os pontos espaciais ($\mathbf{s}_i$ e $\mathbf{s}_j$) que formam o campo espacial e o ponto desconhecido ($\mathbf{s}_0$). Substituindo-se a relação do sistema em (13) na expressão (11), obtém-se a variância de predição

$$\sigma_{OK}^2(\mathbf{s}_0) = C(\mathbf{0}) - \sum_{i=1}^N \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) + \beta. \quad (14)$$

Dada a relação entre a função covariância e a função semivariograma $C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{0}) - \gamma(\mathbf{s}_i, \mathbf{s}_j)$, é possível reescrever o sistema de equações Kriging em função do semivariograma

$$\sum_{j=1}^N \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j) + \beta = \gamma(\mathbf{s}_i, \mathbf{s}_0) \text{ e } \sum_{i=1}^N \lambda_i = 1, \quad i = 1, \dots, N \quad (15)$$

A solução do sistema de equações KO consiste nos pesos Kriging λ_i para aplicação de (6). Além disso, associado a cada predição, é possível calcular a variância de predição em função do semivariograma, dada por:

$$\sigma_{KO}^2(\mathbf{s}_0) = \sum_{i=1}^N \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}_0) + \beta. \quad (16)$$

III. MODELAGEM E GERAÇÃO DO REM

Apresentado o método de predição espacial Kriging, esta seção aborda os aspectos de modelagem do ambiente de rádio que será alvo do processo de predição espacial.

A modelagem do ambiente de rádio é baseada em um sistema de comunicação sem fio composto por dispositivos móveis e uma estação rádio base. Supõe-se que tais dispositivos são capazes de realizar as medidas de intensidade do nível de sinal recebido e enviá-las para a estação rádio-base, que coordena toda a operação da rede do sistema de comunicação sem fio.

Interno à estação rádio-base, o centro de predição espacial é responsável pelo processamento das medidas coletadas e pela geração do REM. Especificamente, o ambiente de rádio é caracterizado por um canal de propagação que apresenta uma série de mecanismos como difrações, reflexões e refrações, além da perda por percurso e absorção de energia. Esses mecanismos influenciam a intensidade dos sinais recebidos e consequentemente afetam a precisão do processo de geração do REM. Assim, o processo aleatório espacial referente ao ambiente de rádio pode ser interpretado como

$$Z(s) = \mu(s) + \xi(s) \quad (17)$$

em que $\mu(s)$ é a tendência do campo espacial, representada pelo o modelo de perda por percurso (*Path Loss Model*) e $\xi(s)$ é a parcela aleatória do campo espacial, representada pelos efeitos do sombreamento (*Shadowing*) do canal.

De forma analítica, a modelagem do ambiente de rádio é descrita pela seguinte equação:

$$\bar{P}_L(d) = \underbrace{\bar{P}_L(d_0)}_{\text{Perda por Percurso}} + 10 \log \left(\frac{d}{d_0} \right)^\alpha + \underbrace{X_{\sigma_s}(d)}_{\text{Sombreamento}} \quad [\text{dB}] \quad (18)$$

em que $\bar{P}_L(d)$ representa a atenuação (dB) a uma determinada distância d (metros) e $\bar{P}_L(d_0)$ retrata a atenuação a uma

distância de referência (d_0) desconsiderando os efeitos do sombreamento. A variável aleatória Gaussiana X_{σ_s} (na escala em dB) denota o sombreamento do modelo com média igual a zero e desvio padrão σ_s . Em particular para o sombreamento, tem-se a característica distância de correlação (d_c) que é uma medida espacial que reflete o grau de correlação do campo espacial. Finalmente, o parâmetro α é o coeficiente de propagação do modelo de perda por percurso. Para cenários urbanos e suburbanos (característicos dos sistemas de comunicações sem fio [10]), é possível considerar a variação $2 < \alpha < 6$.

O primeiro passo para a geração do REM via KO consiste na verificação e estimação da tendência nos dados da VR. Nas simulações deste trabalho, um processo de regressão não linear, baseado no método dos mínimos quadrados, foi realizado a fim de identificar e estimar a perda por percurso do processo aleatório. Nesta metodologia, assume-se o conhecimento a priori de duas informações importantes: (1) a localização do transmissor; (2) as distâncias envolvidas entre os dispositivos da rede e o transmissor. Isso significa que as incertezas sobre estas informações influenciam o tratamento da tendência dos dados necessário para o KO, afetando, portanto, todo o processo de predição espacial. Em seguida, a tendência é subtraída das medidas coletadas para posterior estimação do semivariograma experimental. O processo de ajuste utilizando o semivariograma experimental permite obter o modelo analítico necessário para o preditor KO.

A Figura 2 mostra os resultados de simulação para a realização de uma função aleatória (c) composta por: (a) perda por percurso e (b) sombreamento com as seguintes configurações: $\alpha = 3,5$, $d_0 = 100\text{m}$, $\sigma_s = 5\text{dB}$ e grid $500 \times 500\text{m}$. Dois casos foram simulados para a mesma realização: (d) com $N = 30$ medidas; (e) com $N = 100$ medidas coletadas com estratégia de amostragem com distribuição uniforme.

Figura 2: Resultados de simulação: (a) perda por percurso, (b) sombreamento do canal, (c) realizações do processo aleatório e da amostragem espacial (d) predição espacial com $N = 30$, (e) predição espacial com $N = 100$ e (f) semivariograma experimental e analítico.

Os resultados de estimação e ajuste usando o semivariograma experimental são mostrados em (f) para os dois casos de simulação. Nota-se que, embora ambos os casos tenham a mesma separação r para independência espacial, têm-se patamares diferentes para cada um deles. Ainda, no caso em que $N = 30$ amostras, fica clara a baixa verossimilhança do mapa gerado em relação à verdadeira realização do sombreamento do ambiente de rádio. Isso significa que inferências espaciais e decisões tomadas com base nestas circunstâncias podem comprometer o desempenho das aplicações subsequentes. De outro modo, com $N = 100$ amostras, tem-se um resultado de predição que caracteriza melhor precisão na geração do REM.

É intuitivo notar que o número de amostras coletadas (N) e, principalmente, o padrão espacial de amostragem afetam o desempenho da predição. Entretanto, ressalta-se que os resultados acima estão relacionados a uma realização particular do processo aleatório. De fato, a aleatoriedade do processo de amostragem e do próprio processo aleatório requer que a análise de desempenho seja feita para mais realizações. Para buscar resultados que levem consistência estatística, a avaliação da precisão é feita utilizando a métrica raiz do erro quadrático médio (RMSE – *Root Mean Square Error*) que capta a magnitude do erro, na forma de diferenças quadráticas, entre os valores reais do processo aleatório e as predições espaciais, de acordo com

$$RMSE = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} [Z(s_i) - Z^*(s_i)]^2} \quad (19)$$

em que N_T é o número total de pontos do mapa, ou seja, as comparações são feitas em todo o grid 500m x 500m, com resolução de 1m.

A Figura 3 mostra os resultados de simulação na forma de histograma estatístico para a métrica RMSE referente a 1000 realizações dentro de um cenário com $\alpha=3,5$, $d_d=100$ m, $\sigma_s=5$ dB e $N=100$ amostras.

Figura 3: Resultados de simulação para métrica estatística RMSE

Ressalta-se que as realizações dos processos aleatórios bem como os processos de amostragem são independentes. O objetivo seria observar, para estas condições de campo espacial, qual seria o comportamento esperado (médio) do desempenho das predições KO. Assim, sobre estas circunstâncias de propagação, nota-se no histograma que, em média, espera-se um RMSE da ordem de 6dB. Isso significa que, o preditor Kriging é capaz de reproduzir a estrutura do sombreamento do canal com erros que em média, são da ordem de 6dB. Ressalta-se que valores baixos para o RMSE indicam boa precisão na reprodução alcançada pelas

predições espaciais. Ainda assim, são os requisitos de desempenho relacionados às aplicações e a forma como estas venham a utilizar o REM que determinam o grau de precisão exigido nos resultados de geração das predições espaciais.

IV. CONCLUSÕES E PROPOSTAS

As principais ferramentas da geoestatística, como o semivariograma experimental e o método KO, foram apresentadas e utilizadas na modelagem e na geração do REM. A influência do número de medidas, do padrão espacial de amostragem e o tratamento da tendência dos dados experimentais foram levantados como pontos importantes na geração REM, sobretudo quando se considera o aspecto funcional dos sistemas de comunicações sem fio. A métrica estatística RMSE foi usada para a verificação e análise do desempenho (médio) do preditor KO na reprodução do sombreamento do ambiente de rádio.

Sobre possíveis desafios a serem tratados e superados, é possível citar: (1) Estudos de influência acerca dos parâmetros e modelos de propagação específicos na precisão de geração do REM; (2) Consideração de incertezas sobre as medidas e localizações dos dispositivos que compõem o sistema de comunicação sem fio; (3) Uso dos resultados da predição espacial Kriging para aplicações em comunicações como sistemas de acesso dinâmico ao espectro; (4) Estudos sobre a utilização de diferentes estratégias de amostragem espacial; (5) Consideração das questões relacionadas à anisotropia na estimação do semivariograma experimental.

Referências

- [1] Zhao, Y. ; Gaedert, J. ; Base K, K. ; Reed, J. H and Wireless@Virginia Tech, "Radio Environment Map Enabled Situation-Aware Cognitive Radio Learning Algorithms". *Proceeding of the SDR 06 - Technical Conference and Product Exposition*, 2006.
- [2] 3GPP TR 36.805 v1.3.0 1, "Study on minimization of drive-tests in next generation networks; (release 9)," tech. rep., 3rd Generation Partnership Project, 2009.
- [3] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijärvi and P. Mähönen, "Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks," *Cognitive Radio Oriented Wireless Networks (CROWNCOM)*, 2013 8th International Conference on, Washington, DC, 2013, pp. 37-42.
- [4] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijärvi and P. Mähönen, "Automated coverage hole detection for cellular networks using radio environment maps," *Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt)*, 2013 11th International Symposium on, Tsukuba
- [5] H. B. Yilmaz, T. Tugcu, F. Alagöz and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," in *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162-169, December 2013.
- [6] S. Grimoud, B. Sayrac, S. Ben Jemaa and E. Moulines, "Best Sensor Selection for an Iterative REM Construction," *Vehicular Technology Conference (VTC Fall)*, 2011 IEEE, San Francisco, CA, 2011, pp. 1-5.
- [7] Chiles's, J. P. ; Delfine, P. ; *Geostatistics Modeling Spatial Uncertainty*, Second Edition, Copyright © 2012 by John Wiley & Sons, Inc. All rights reserved.
- [8] Yamamoto J. K. ; Landim, P. M. B. ; "Geoestatística: Conceitos e Aplicações", Oficina de Textos, ano: 2013.
- [9] Diggle, P. J. ; Ribeiro, P. J. ; "Model-based Geostatistics", *Springer Science Business Media*, LLC, 2007
- [10] Cho, Y. S. ; Kim, J. ; Yang, W. Y. ; Kang, C. "MIMO-OFDM Wireless Communications Systems with Matlab", John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop, # 02-01, Singapore 129809, 2010.