

Redução de Dados Baseada em Amostragem para Redes de Sensores Sem Fio com Tolerância a Conexões Disruptivas

Israel Loureiro Cavalcante Vasconcelos¹, Andre Luiz Lins Aquino¹

¹Laboratório de Computação Científica e Análise Numérica
Universidade Federal de Alagoas (UFAL)
Código Postal 57072-900 – 91.501-970 – Maceió – AL – Brasil

{vasconcelos.isr, alla.lins}@gmail.com

Abstract. *This study aims to evaluate, on the disruptive connection networks scope, the quality of reconstruction about monitored environment using a sampling central algorithm to reduce the data streamed by sensors, comparing to other techniques already known in the literature. Using Monte Carlo methodology through software simulation, the approach under study showed satisfactory results.*

Resumo. *Este trabalho tem como objetivo avaliar, no escopo de redes com conexões disruptivas, a qualidade da reconstrução do ambiente monitorado utilizando uma técnica de redução de dados baseada em amostragem central, comparando seu desempenho com outras técnicas conhecidas na literatura. Utilizando ensaios Monte Carlo para experimentação através de simulação por software, a abordagem em estudo apresentou resultados satisfatórios.*

1. Introdução

Uma Rede de Sensores Sem Fio (RSSF) é um conjunto de dispositivos (conhecidos comumente como sensores ou nós sensores) independentes e que possuem comunicação direta entre si, o que caracteriza uma rede do tipo *Ad-Hoc*. Suas principais aplicações estão no monitoramento remoto de ambientes, onde os dados monitorados podem ser, por exemplo, temperatura, umidade, luminosidade, fluxo de fluidos, dentre outras. Esses nós sensores possuem, em geral, recursos computacionais e energéticos limitados. Os dados são transmitidos a uma estação base (*sink*) que é responsável por apresentar os resultados das operações realizadas e assim, reconstruir o fenômeno em estudo [Akyildiz et al. 2002].

Levando em consideração a limitação dos nós sensores nos mais diversos aspectos citados somados à alta demanda de recursos computacionais requerida pelos métodos de reconstrução de dados, diminuir a quantidade de informação recebida dos nós traz uma significativa contribuição na dinâmica de funcionamento da rede, principalmente no que compete a questões de armazenamento e reconstrução do fenômeno.

Num ambiente fechado, a comunicação sem fio está sujeita a uma série de interferências, como por exemplo da rede elétrica do local e, principalmente, obstáculos físicos. Assim como também em áreas de grande extensão territorial e hostis à presença humana, como por exemplo florestas e geleiras.

Alguns fatores como os citados anteriormente podem tornar a rede propensa a ocorrência de conexões disruptivas entre os nós que realizam o monitoramento, cenário que motivou a realização dos experimentos apresentados neste artigo.

Uma rede tolerante a conexões disruptivas, DTN (*Disruption Tolerant Network*, [Curran and Knox 2008]), é uma rede capaz de manter-se operante mesmo que não haja comunicação entre seus nós. Para darmos a nossa RSSF características de uma DTN, será utilizada a ideia de MULEs (*Mobile Ubiquity Lan Extension*, [Shah et al. 2003]), unidades móveis que visitam os nós para coletar dados.

No nosso caso, a MULE pode ser um robô com uma estação base móvel que percorre, em determinados intervalos de tempo, uma rota pré determinada ao longo do local monitorado, ou um usuário que recebe os dados através de uma aplicação para dispositivos móveis, a depender das condições e circunstâncias sob as quais estão determinadas este arcabouço.

Como principal fator crítico deste sistema, a capacidade de armazenamento e o poder de processamento das MULEs pode ser limitada, tornando necessário reduzir o conjunto de dados recebido dos nós sensores para otimizar o uso do armazenamento desses dispositivos e diminuir a demanda por recursos computacionais durante os processos de reconstrução, etapa que consumiu maior quantidade de tempo nos experimentos.

Agregando os fatores acima descritos, podemos delimitar o estudo tratado neste trabalho através do problema que segue:

Como reduzir o conjunto de dados coletados dos sensores de modo a comprometer minimamente a qualidade da informação reconstruída?

Dentro dessas condições, avaliaremos a abordagem de Redução por Amostragem Central [Aquino and Nakamura 2009], que utiliza uma série de artifícios e heurísticas para realizar esta redução de maneira eficiente, valendo-se destes artifícios afim de escolher dados mais significativos e buscar equilibrar da melhor forma o *trade off* entre quantidade de informações utilizadas e qualidade na reconstrução do fenômeno.

Para avaliar seu desempenho, comparamos a técnica de Redução por Amostragem Central com o método de Redução Baseada no Descarte de Pacotes [Luiz et al. 2012], cujo funcionamento se dá através de um pré estabelecimento de uma condição específica para realizar a amostragem e foi proposta para este cenário disruptivo com coleta via MULEs.

O experimento consiste em gerar, através de simulação, um ambiente que representa o cenário em estudo. Em seguida, utilizando Processos Pontuais [Baddeley 2007], descrevemos a distribuição dos sensores sobre o ambiente. Com os sensores distribuídos, nossa MULE irá visitar os nós sensores para coletar os dados, nesta etapa, repetiremos o processo para as duas soluções de amostragem citadas nos parágrafos anteriores afim de comparar seu desempenho.

Uma vez realizada a amostragem, o conjunto de dados reduzido (conhecido como *Sensor Stream*) será submetido a um processo de interpolação (Kriging Simples, [Krige 1951]) para termos como *output* uma reconstrução do campo inicialmente simulado. Para avaliar o erro, compararemos as medições dos campos reconstruídos.

Mostrando-se adequada ao escopo do problema, a técnica de Redução por Amostragem Central apresentou melhorias na ordem de 16% na qualidade da reconstrução dos ambientes estudados, a principal contribuição deste trabalho é a proposta de utilização da mesma para Redes de Sensores Sem Fio com Tolerância a Conexões

Disruptivas.

A seguir, será detalhada a estrutura com a qual encontra-se organizado este trabalho. Seção 2: É dada uma visão geral do experimento realizado, dividindo-o em etapas sequenciais e descrevendo o passo a passo de cada uma delas. Nesta seção também é construída uma fundamentação teórica acerca de conceitos importantes. Seção 3: Descreve a metodologia seguida pelo trabalho, abordando as ferramentas utilizadas e parâmetros considerados nos experimentos. Nas Seções 4 e 5 são apresentadas, respectivamente, a discussão dos resultados e conclusões obtidas.

2. Etapas do Experimento

2.1. Descrição do cenário

O ambiente estudado simula um cenário de floresta cuja aplicação consiste no monitoramento da intensidade de luz sob a sua vegetação, por exemplo, sombras formadas abaixo da copa das árvores, de modo que os pontos do campo (exemplificado na figura 1) representam a medição das variações dessa intensidade. A simulação foi realizada utilizando o software R-3.0.2 [R Development Core Team 2010], detalhado na seção 3.

Este ambiente será representado por um campo gaussiano aleatório de $100 \times 100 = 10^4$ pontos, com média $m = 100$ e variância $v = 10$. Segundo [Lopes 2010], um campo gaussiano é definido por um processo estocástico $S(x) : x \in d$, onde d representa a dimensão do espaço.

A estrutura adotada para descrever a correlação espacial entre pontos deste campo será o modelo de correlação de Matérn [Diggle and Ribeiro 2007], por tratar-se de um modelo amplamente utilizado na literatura. Sendo o mesmo dado pela equação 1:

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

onde u é a distância entre posições pontos, K_{κ} é a função de Bessel de ordem κ , $\kappa > 0$ é o parâmetro de suavização denominado "ordem", e $\phi > 0$ é o parâmetro de escala.

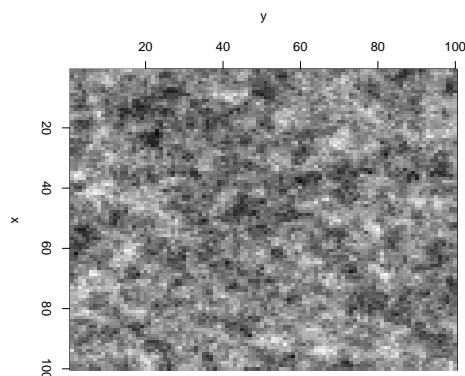
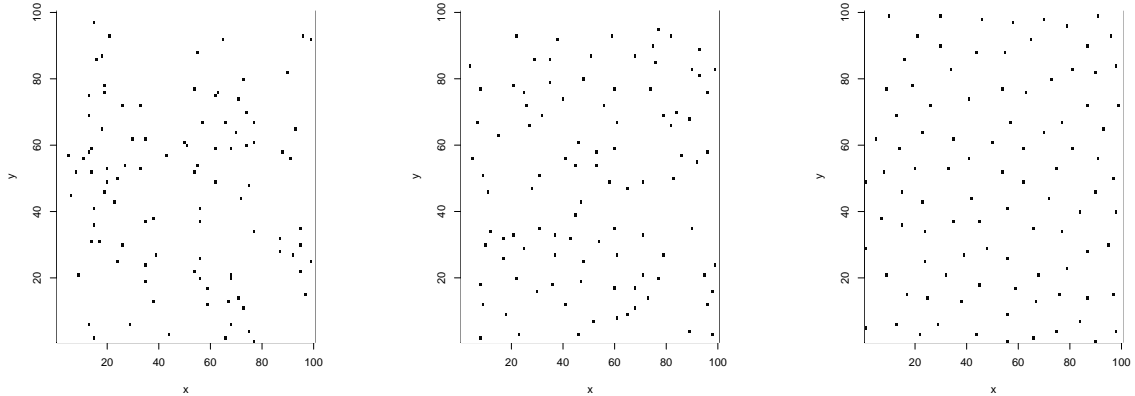


Figure 1. Campo Gaussiano simulando um ambiente, a intensidade das cores em cada ponto representam as medições no mesmo.



(a) Distribuição SSI com raio de repulsão $\alpha = 0$.

(b) Distribuição SSI com raio de repulsão $\alpha = 4$.

(c) Distribuição SSI com raio de repulsão $\alpha = 8$.

Figure 2. Distribuição SSI representando os nós da rede.

2.2. Distribuição espacial e Cobertura dos Sensores

Para representar a disposição espacial dos nós sensores sobre o campo, será empregado um processo SSI (*Simple Sequential Inhibition*, [Baddeley 2007]) que consiste em estipular um raio de repulsão ao redor de cada nó, de modo que não será possível introduzir nenhum outro nó na área coberta por este raio. Dessa forma, é possível controlar o espaçamento entre cada nó e o grau de aleatoriedade na distribuição. Um exemplo desta distribuição pode ser visto na figura 2.

Uma vez distribuídos os nós, é necessário delimitar a sub-área do campo da qual cada nó será responsável por cobrir. Para isso, será construído sobre o campo um diagrama de Voronoi, que é definido:

”Seja S um conjunto de n pontos (chamados *sítios* ou *sementes*), no plano. Para duas sementes distintas $p, q \in S$ a *dominância* de p sobre q é definida como o subconjunto (ou sub-área) do plano que está mais próximo de p do que de q , seguindo a expressão:

$$dom(p, q) = \{x \in R^2 | \rho(x, p) < \rho(x, q)\} \quad (2)$$

Onde ρ corresponde à função de distância euclidiana e x é um ponto qualquer do plano R^2 .” [Aurenhammer 1991]

Trazendo esta definição ao escopo do nosso problema, assumimos que as sementes do diagrama serão os nós sensores e que dominância é a sub-área coberta por cada nó. Esta sub-área é conhecida como *Célula de Voronoi*. A figura 3 mostra um diagrama de Voronoi representando as áreas de cobertura dos sensores sobre o campo.

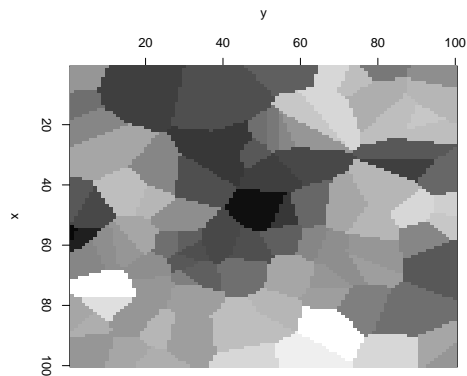


Figure 3. Campo Gaussiano simulando um ambiente, a intensidade das cores em cada ponto representam as medições no mesmo.

2.3. Soluções para Amostragem dos dados

2.3.1. Redução por Amostragem Central

Este algoritmo consiste em realizar uma triagem no conjunto de dados recebido dos sensores, agrupando-os em classes afim de selecionar os dados mais significativos. Para isso, o conjunto é ordenado em função das medições das amostras em ordem crescente e em seguida, dividido em 10 partes iguais, de modo que cada um desses intervalos representa uma classe do conjunto (a quantidade de classes pode ser diferente a depender da aplicação, sendo este valor escolhido de forma empírica). Esse processo é realizado para cada uma das células de Voronoi para conseguir, localmente, uma melhor aproximação.

Uma vez realizada classificação dos dados do conjunto seguindo os critérios descritos acima, o algoritmo coletará as amostras localizadas nos arredores da mediana de cada uma das classes, partindo do princípio de que os dados centrais de cada classe são mais representativos, propiciando uma aproximação mais precisa dos valores originais da classe. Esta abordagem é proposta e detalhada no trabalho de [Aquino and Nakamura 2009].

Por fim, após realizada esta amostragem, o algoritmo irá gerar um conjunto de dados reduzido para serem aplicados os processos de interpolação para reconstruir o ambiente em estudo.

O conjunto de dados reduzido resultante deste algoritmo segue exemplificado na figura 4.

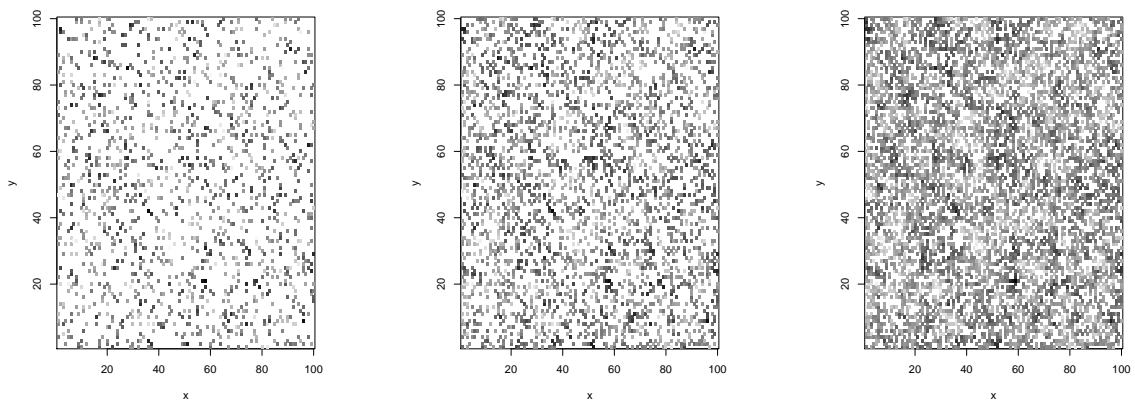
2.3.2. Políticas de Descarte de Pacotes

Nesta estratégia, deve ser definido previamente um critério para classificar prioridade no armazenamento de dados, sendo que aqueles não contemplados por estes critérios são simplesmente descartados em caso de esgotamento da capacidade de armazenamento do dispositivo que coleta estas amostras, o conjunto de dados gerado por estes algoritmos está ilustrado na figura 5. Estes algoritmos são descritos segundo o trabalho de [Luiz et al. 2012].

Descarte dos Primeiros Pacotes: Elimina os primeiros dados recebidos durante a coleta. Esta política de descarte sugere que as medições mais antigas são mais significativas que as recentes.

Descarte dos Últimos Pacotes: Elimina os últimos dados recebidos durante a coleta. Esta política de descarte sugere que as medições mais recentes são mais significativas que as antigas.

Descarte Aleatório de Pacotes: Elimina os dados recebidos de forma aleatória. Esta política de descarte sugere que todos os dados possuem mesma representatividade, realizando então uma amostragem aleatória para tentar aumentar a cobertura da rede.

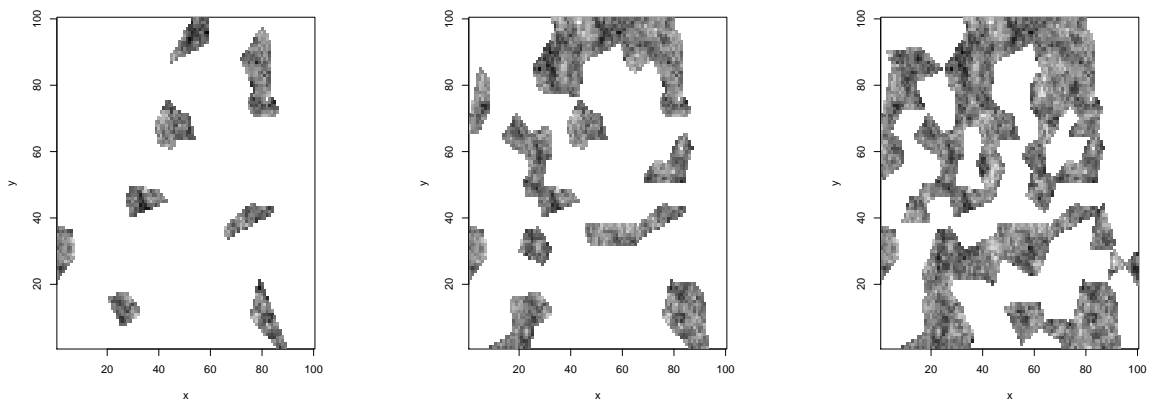


(a) Conjunto gerado utilizando 10% das amostras.

(b) Conjunto gerado utilizando 25% das amostras.

(c) Conjunto gerado utilizando 50% das amostras.

Figure 4. Conjuntos de dados reduzidos gerados pelo algoritmo de Redução por Amostragem Central.



(a) Conjunto gerado utilizando os dados de 10% dos sensores.

(b) Conjunto gerado utilizando os dados de 25% dos sensores.

(c) Conjunto gerado utilizando os dados de 50% dos sensores.

Figure 5. Conjuntos de dados reduzidos gerados pelo algoritmo de Descarte dos Primeiros Pacotes

2.4. Reconstrução e Avaliação do Erro

Após a etapa de amostragem dos dados, teremos em mãos um conjunto de pontos para reconstruir o campo. Para realizar a interpolação destes pontos, foi utilizado o método Kriging Simples, [Krige 1951].

Uma vez concluído o processo de reconstrução do campo, a métrica utilizada para avaliar a qualidade da mesma é o Valor Absoluto do Erro Relativo (VAER) [Frery et al. 2010], que é definido como:

$$\hat{\epsilon} = \frac{1}{10^4} \sum_{i,j=1}^{100} \left| \frac{f(i,j) - \hat{f}(i,j)}{f(i,j)} \right|, \quad (3)$$

em que f representa o campo gaussiano original e \hat{f} o campo gaussiano reconstruído. A Equação 3 está sempre bem definida, pois a probabilidade de observar $f(i,j) = 0$ é nula para qualquer (i,j) em virtude da natureza contínua do campo gaussiano.

3. Metodologia

3.1. Hardware e Ferramentas

Para realização dos experimentos, foi utilizada a plataforma R-3.0.2 [R Development Core Team 2010] utilizando os seguintes pacotes:

- SPATSTAT [Baddeley and Turner 2005], para a geração da disposição dos nós;
- RANDOMFIELDS [Schlather 2009], para gerar campos;
- GEOR [Ribeiro and Diggle 2001], para realizar a reconstrução utilizando a função KRIGR.CONV;

Sob o sistema operacional Linux e hardware nas seguintes especificações:

- Modelo SGI Rackable Standard-Depth Servers;
- 12 Cores em 2 sockets Intel Xeon Six-Core 5649 de 2.53-GHz;
- 144-GB de memória DDR3 1333 MHz (18 x 8 GB);
- 2 controladores NVIDIA Tesla C2075.

3.2. Espaço Paramétrico

Neste trabalho foi utilizado o seguinte espaço paramétrico:

1. Algoritmos de amostragem para redução de dados: Amostragem Central, Baseado em Descarte de Pacotes.
2. Percentual de uso dos dados amostrados: $\rho \in \{10, 25, 50, 75\}$
3. Escala: $\phi = 2$
4. Suavização: $\kappa = 0.5$
5. Repulsão do processo SSI: $\alpha = 4$

Foi escolhida uma semente única para cada ponto deste espaço paramétrico e realizadas 5 replicações independentes.

4. Resultados e discussões

Os erros médios produzidos na reconstrução em cada caso estão resumidos na tabela abaixo, as colunas representam o percentual de uso das amostras:

	10%	25%	50%	75%
Amostragem Central	0.0188	0.0155	0.0102	0.0049
Descarte dos Primeiros Pacotes	0.0230	0.0186	0.0125	0.0063
Descarte dos Últimos Pacotes	0.0222	0.0178	0.0116	0.0057
Descarte Aleatório de Pacotes	0.0222	0.0182	0.0119	0.0056

Podemos observar para o algoritmo de Redução por Amostragem Central uma diminuição no erro na ordem de 16%, em média. As imagens 4 e 5, podem dar uma ideia do porquê a estratégia de Amostragem Central apresentou melhores resultados em relação às políticas de Descarte de Pacotes.

A primeira técnica realiza uma série de manipulações no conjunto de dados (detalhadas na seção 2.3.1) de modo a conseguir selecionar aquelas amostras que melhor representam o conjunto, dessa forma, a quantidade de informação útil produzida neste algoritmo é mais significativa, resultando numa cobertura mais abrangente do campo e justificando os melhores resultados.

O gráfico da figura 6 exhibe, de forma mais detalhada, o comportamento das técnicas avaliadas, apresentando as médias com seus respectivos intervalos de confiança dos valores de erro produzidos por elas em 5 replicações independentes, a figura 7 ilustra o resultado da reconstrução dos dados utilizando cada uma delas.

5. Conclusões e Trabalhos Futuros

Levando em consideração o fato dos experimentos terem sido realizados simulando um ambiente sob condições controladas (sem contaminação) e com uma variância pequena entre os valores das medições, a técnica de Redução por Amostragem Central não apenas apresentou valores bem comportados em todas as situações avaliadas como também, quantitativamente, uma diminuição na ordem de, em média, 16% no erro da reconstrução.

A diminuição no erro somada à maior abrangência de dados significativos propiciadas pela abordagem de Amostragem Central proporcionam uma melhora expressiva na reconstrução do campo, de modo que consideramos satisfatórios os resultados produzidos neste experimento.

A figura 7 ilustra o desempenho dos algoritmos testados, tornando ainda mais evidente a expressividade dos resultados apresentados. As perspectivas futuras para o trabalho descrito neste artigo incluem avaliar em outros cenários as técnicas abordadas, além de adicionar elementos para tornar os experimentos mais realistas, como por exemplo, a inserção de ruído nas medições dos sensores e variações abruptas nas medidas de determinadas áreas do campo (para simular o aumento de temperatura causado por um incêndio em uma determinada área do mapa ou o uso de condicionadores de ar, por exemplo).

A construção de um arcabouço que propicie a realização deste experimento de maneira prática é pretendida, utilizando os resultados obtidos neste artigo como base para o monitoramento de um edifício real, no nosso caso, a sede do grupo de pesquisa onde

este projeto é desenvolvido. Monitoramento este que se dará seguindo a metodologia descrita neste trabalho, utilizando um *sink* móvel ou aplicação para celulares.

Não só o monitoramento do edifício é objetivado, como também a repetição deste experimento utilizando dados reais coletados a partir dos sensores e realização de um estudo comparativo avaliando os fatores abordados neste trabalho.

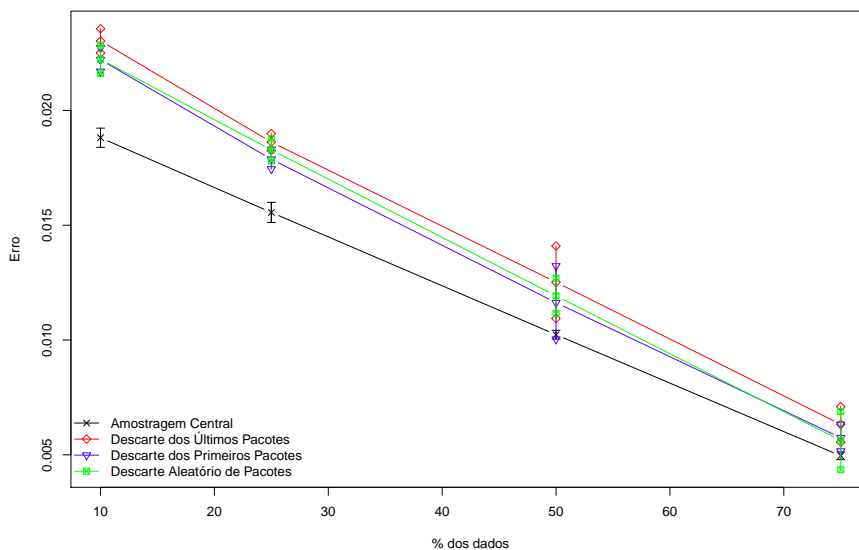
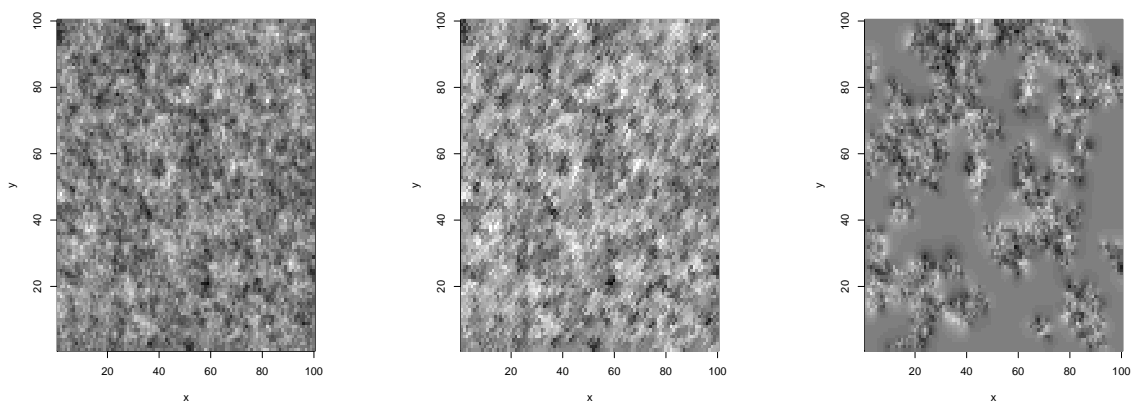


Figure 6. Gráfico comparativo entre os valores de erro na reconstrução dos 4 algoritmos avaliados.



(a) Campo Original

(b) Campo reconstruído utilizando Amostragem Central

(c) Campo reconstruído utilizando Descarte dos Primeiros Pacotes

Figure 7. Quadro comparativo entre as reconstruções dos algoritmos avaliados utilizando 50% das amostras coletadas.

References

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cyirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422.
- Aquino, A. L. L. and Nakamura, E. (2009). Data centric sensor stream reduction for real-time applications in wireless sensor networks. *MDPI AG, POSTFACH, CH-4005 BASEL, SWITZERLAND*, 9:9666–9688.
- Aurenhammer, F. (1991). Voronoi diagrams: A survey of a fundamental data structure. *ACM Computing Surveys*, 23:345 – 405.
- Baddeley, A. (2007). Spatial point processes and their application. In Weil, W., editor, *Stochastic Geometry*, volume 1892 of *Lecture Notes in Mathematics*, pages 1–75. Springer, Berlin.
- Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42. ISSN 1548-7660.
- Curran, K. and Knox, J. (2008). Disruption tolerant networking. *Computer and Information Science*.
- Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.
- Frery, A. C., Ramos, H. S., Alencar-Neto, J., Nakamura, E., and Loureiro, A. A. F. (2010). Data driven performance evaluation of wireless sensor networks. *Sensors*, 10(3):2150–2168.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52:119–139.
- Lopes, B. (2010). Reconstrução de sinais em redes de sensores sem fios com técnicas de geoestatística. Dissertação de mestrado, Universidade Federal de Alagoas, Maceió, AL, Brasil.
- Luiz, D. F., Figueiredo, C. M. S., and Nakamura, E. F. (2012). A coverage-based drop-policy in wireless sensor network with disruptive connections. In *ISCC*.
- R Development Core Team (2010). *The R Project for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ribeiro, P. J. and Diggle, P. J. (2001). geoR: a package for geostatistical analysis. *R-NEWS*, 1(2):14–18. ISSN 1609-3631.
- Schlather, M. (2009). *RandomFields: Simulation and Analysis of Random Fields*. R package version 1.3.41.
- Shah, R. C., Roy, S., Jain, S., and Brunette, W. (2003). Data mules: modeling and analysis of a three-tier architecture for sparse sensor networks. *Ad Hoc Networks*, 1:215 – 233.