

# Redução de dados em redes de sensores sem fio tolerantes a atrasos

Israel L. C. Vasconcelos<sup>1</sup>, Andre L. L. Aquino<sup>1</sup>

<sup>1</sup>Instituto de Computação  
Universidade Federal de Alagoas – Maceió, AL – Brazil

{israel.vasconcelos, alla}@laccan.ufal.br

**Abstract.** *This paper presents a delay-tolerant sensor network model for environmental applications where a data-aware drop strategy is applied to improve the phenomenon coverage. Our model was designed for applications that monitor the forest temperature incidence for a wildlife observation. The proposed solution modeling is comprised of i. Phenomenon generation based on Gaussian random field; ii. Sensing nodes with a mobile sink; iii. Data processing based on a data-aware drop strategy; and iv. Phenomenon reconstruction based on simple kriging interpolation. Besides the satisfactory application of our model, the results show that the performance of data-aware strategy is approximately twice as better than traditional approaches in all evaluated scenarios.*

**Resumo.** *Este artigo apresenta um modelo de rede de sensores tolerante a atraso para aplicações de monitoramento ambiental onde uma solução de amostragem sensível aos dados é utilizada para melhorar a cobertura do fenômeno. O modelo proposto compreende uma aplicação de monitoramento de temperatura em regiões florestais, a implementação consiste em i. Simular regiões de interesse; ii. Distribuição dos nós sensores; iii. Realizar amostragem e processamento dos dados; iv. Reconstruir o fenômeno a partir das amostras; Os resultados obtidos apontaram que a abordagem sensível aos dados é capaz de reconstruir o fenômeno aproximadamente duas vezes mais próximo da realidade em relação às técnicas tradicionais, em todos os cenários avaliados.*

## 1. Introdução

No presente trabalho, utilizamos as Redes Tolerantes a Atrasos de Conexões Intermitentes (ICDT-WSNs da expressão em inglês) [Curran and Knox 2008] que corresponde a aplicação do conceito de Redes Tolerantes a Atrasos (DTN) [Curran and Knox 2008] em cenários de Redes de Sensores Sem Fio (WSNs) [Akyildiz et al. 2002]. Essas redes são utilizadas com sucesso em aplicações de monitoramento de vida selvagem, de locais de difícil acesso, de risco, de desastres naturais, como também de áreas urbanas.

Por conta do grande volume de dados gerado nessas redes ICDT-WSNs, recomenda-se o uso de técnicas de redução de dados para manter a confiabilidade e aumentar o tempo de vida da rede, sem comprometer o comportamento das aplicações. O consumo de energia é proporcional à quantidade de dados transmitidos, apresentando um comportamento previsível em situações onde o percentual de redução é fixado a priori. As abordagens tradicionais na literatura, para redução de tráfego em ICDT-WSNs, utilizam técnicas não sensíveis aos dados. Por exemplo, o descarte de pacotes por critério aleatório

ou ordem de chegada [Li and Bartos 2014]. Tais abordagens comprometem a qualidade dos dados das aplicações. Neste trabalho utilizamos uma redução de tráfego sensível aos dados. Nosso objetivo é tornar mais eficiente a utilização do armazenamento disponível mantendo o compromisso de preservar a qualidade dos dados.

Na solução utilizada [Aquino and Nakamura 2009], a redução é requisitada quando se esgota a memória de armazenamento do sorvedouro. Ela consiste em manipular as amostras coletadas pelos sensores a fim de selecionar as mais significativas, realizando os seguintes passos: i) ordena de forma crescente o conjunto de dados coletados em função dos valores das amostras; ii) divide o conjunto de dados ordenado em dez subconjuntos com intervalos igualmente espaçados, montando um histograma; iii) seleciona as amostras ao redor da mediana de cada subconjunto.

Neste trabalho montamos um arcabouço para modelagem e simulação de experimentos voltados a amostragem de dados em ICDT-WSNs. Tal arcabouço, define as principais diretrizes para experimentação e validação das soluções propostas dentro do contexto da avaliação. Este projeto, por intermédio dos esforços do bolsista de Iniciação Científica Israel Vasconcelos, já gerou as seguintes contribuições científicas: **Eventos nacionais:** Duas escolas regionais ERBASE 2014 e 2015 [Vasconcelos and Aquino 2014, Vasconcelos et al. 2015b]; **Eventos internacionais:** Dois eventos de alta qualidade da área de estudo proposta NTMS (Qualis-2016 B1) e ISCC (Qualis-2016 A2) [Vasconcelos et al. 2015a, Vasconcelos et al. 2015c]; **Submissão para revista:** Um artigo em processo de avaliação para o Journal of Network and Computer Applications (Qualis-2016 A2, JCR 3.5); **Premiações:** Obtivemos o prêmio de 2º melhor trabalho de Iniciação Científica da Sociedade Brasileira de Matemática Aplicada à Computação - SBMAC em 2017 (prêmio Beatriz Neves); recebemos Excelência Acadêmica por três anos consecutivos (2014, 2015 e 2016) no Encontro de Iniciação Científica da UFAL; recebemos o certificado de melhor artigo do ERBASE 2014 [Vasconcelos and Aquino 2014]. Este trabalho **apresenta um considerável avanço** frente aos anteriores supracitados pois utiliza um modelo que **incorpora nove parâmetros de simulação de rede** (apresentados na tabela 1), enquanto **os anteriores consideravam apenas os dois primeiros** daquela tabela, admitindo os demais como fixos ou ideais. Essas novidades aumentaram drasticamente a sofisticação do modelo de rede e a complexidade das simulações, bem como a abrangência dos resultados obtidos.

## 2. Modelo de Aplicação utilizando ICDT-WSN

Para modelar uma ICDT-WSN, inicialmente, representamos o comportamento ideal dessas redes como

$$\mathcal{N} \xrightarrow{P} \mathbf{V}^* \xrightarrow{R} D,$$

onde  $\mathcal{N}$  é o ambiente do qual é observado o fenômeno de interesse  $P$ , e  $\mathbf{V}^*$  a representação de  $P$  no domínio espacotemporal. Considerando uma observação ideal sobre os dados, sem perdas, é possível definir regras ( $R$ ) que conduzem a decisões ( $D$ ) ideais. O conjunto de dados  $\mathbf{V}^*$  foi modelado como um Campo Gaussiano que é definido como um processo estocástico

$$\mathbf{V}^*(x) : x \in d,$$

onde  $d$  representa a dimensão do domínio e  $x$  é a variável aleatória.

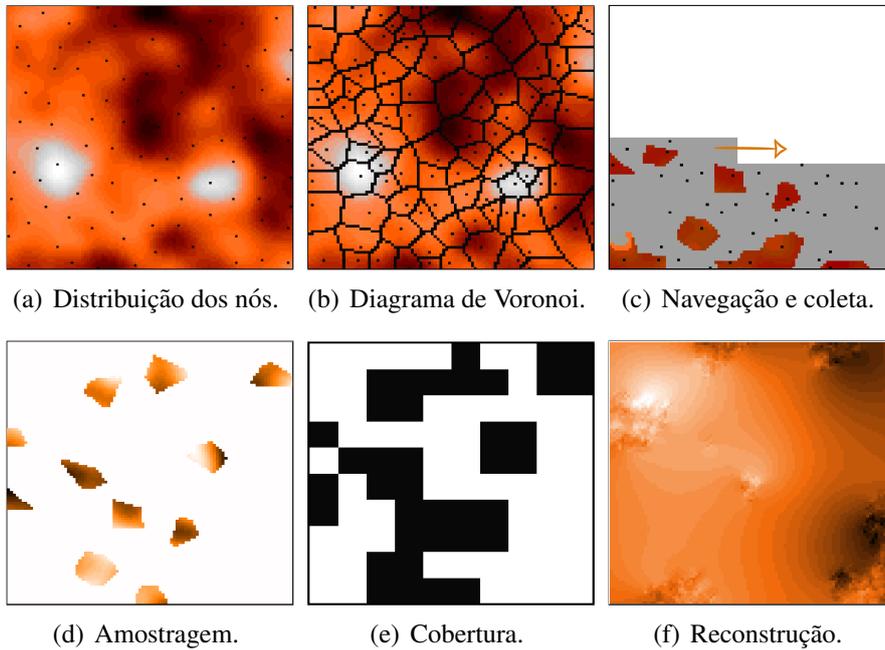
Como estudo de caso, em nossas simulações, consideramos uma região de interesse simulada na forma de uma matriz  $100 \times 100$ , representando uma área de  $10^4$  unidades quadradas, sendo seus pontos preenchidos por amostras de uma Gaussiana com média ( $\mu$ ) = 25 e variância ( $v$ ) = 64 (Campo Gaussiano Aleatório, [Wood and Chan 1994]) e correlacionados espacialmente em função do Modelo de Covariância de Matérn [Diggle and Ribeiro 2007]. Os valores definidos para média e variância representam um comportamento de temperatura,  $25^\circ C$ , variando  $\pm 8^\circ C$ .

Uma vez definidas as características físicas é possível realizar a modelagem dos atributos da rede detalhados a seguir. A atividade dos sensores é modelada como

$$\mathcal{N} \xrightarrow{P} \mathbf{V}^* \xrightarrow{S} \mathbf{V}.$$

Consideramos um conjunto de sensores  $s$ , onde  $S = (S_1, \dots, S_s)$  e cada sensor reporta uma série de medições  $V_i$  com  $1 \leq i \leq s$  do fenômeno. Assim, o conjunto das medições reportadas por todos os sensores é dado por  $\mathbf{V} = (V_1, \dots, V_s)$ .

A distribuição dos nós sensores  $S$  segue um processo parcialmente aleatório (Simple-Sequential Inhibition - SSI [Baddeley 2007]), onde o controle parcial das posições torna mais realista a modelagem. A Figura 1(a) ilustra a área de interesse gerada a partir do SSI combinada com a distribuição dos nós.



**Figure 1. Configuração da Rede, Coleta e Processamento dos Dados.**

Para delimitar a área pela qual cada nó é responsável por cobrir é construído sobre o campo um diagrama de Voronoi, [Aurenhammer 1991]. Onde as sementes do diagrama serão os nós-sensores e a dominância é a sub-área coberta por cada nó. Esta sub-área é conhecida como Célula de Voronoi. Cada sensor será capaz de medir apenas dados do fenômeno contidos em sua respectiva célula de Voronoi. Este procedimento é ilustrado na Figura 1(b).

Uma vez definido a posição e a área de cobertura de cada nó sensor, modelamos o comportamento da aplicação de amostragem de dados (Figura 1(d)), descrito como

$$\mathcal{N} \xrightarrow{P} \mathbf{V}^* \xrightarrow{S} \mathbf{V} \xrightarrow{\Psi} \mathbf{V}' .$$

Conforme definido anteriormente,  $\mathbf{V}$  representa o conjunto de dados coletados pelos sensores em sua totalidade. Dada a necessidade de redução por diversas limitações de recursos, o sorvedouro utiliza uma técnica amostragem  $\Psi$  para reduzir  $\mathbf{V}$ , resultando no conjunto ( $\mathbf{V}'$ ).

Como estudo de caso, consideramos uma aplicação que consiste em um nó sorvedouro móvel que **visita os nós sensores estáticos**, utilizando um algoritmo de movimento aleatório (Random Walk) ou de rota predefinida. Este procedimento é ilustrado na Figura 1(c). A transmissão é realizada de forma oportunista em apenas um salto (*single-hop*) no momento em que este entra no raio de transmissão dos dos nós estáticos, nesse momento a amostragem **sensível aos dados** -  $\Psi$  ([Aquino and Nakamura 2009]) é realizada durante a navegação do sorvedouro.

Após a amostragem de dados é necessário a aplicação de um processo de reconstrução modelado como

$$\mathbf{V}' \xrightarrow{\omega} \widehat{\mathbf{V}} \xrightarrow{R} \widehat{D} .$$

Nessa etapa, temos um conjunto de dados reduzidos  $\mathbf{V}'$  que servirá como entrada para executar o processo de reconstrução  $\omega$ , no nosso caso utilizamos o método de interpolação Kriging Simples, [Hengl 2009]. Após a reconstrução obtemos os dados finais  $\widehat{\mathbf{V}}$  utilizados pela aplicação (Figura 1(f)).

Como métrica de avaliação de robustez da nossa técnica, utilizamos o VAER (Valor Absoluto do Erro Relativo -  $\hat{\epsilon}$  [Frery et al. 2010]), definido como

$$\hat{\epsilon} = \frac{1}{10^4} \sum_{i,j=1}^{100} \left| \frac{\mathbf{V}^*(i,j) - \widehat{\mathbf{V}}(i,j)}{\mathbf{V}^*(i,j)} \right| ,$$

onde  $\mathbf{V}^*$  é o campo original, simulado inicialmente;  $\widehat{\mathbf{V}}$  é a reconstrução do campo a partir das amostras coletadas. Esta expressão é sempre bem definida pois  $\mathbf{V}^*(i,j) \neq 0$ .

Por fim, utilizamos como métrica de avaliação a identificação da cobertura global resultante de cada uma das estratégias. Para isso, uma matriz de  $10 \times 10$  foi montado sobre o campo, criando um *grid* de 100 setores, onde cada setor cobre uma área de  $10 \times 10$ . Se um setor possui pelo menos um ponto de informação dentro de sua área, este setor é considerado coberto (Figura 1(e)). A cobertura global é calculada simplesmente a partir da contagem de quantos setores foram cobertos ao final da execução de cada uma das técnicas sob as condições descritas no parágrafo anterior.

### 3. Discussão e Resultados

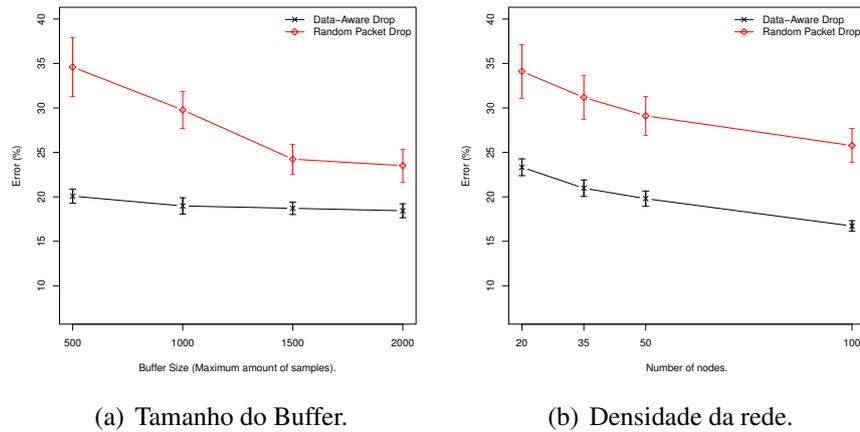
Nas simulações, foram executadas 30 replicações independentes do fenômeno para cada cenário avaliado, utilizando semente única. O erro médio é calculado com intervalo de confiança de 95%. Em todos os casos, nós variamos um único parâmetro, fixando todos

**Table 1. Parâmetros de Simulação**

Parâmetro	Valores
Capacidade do Buffer (Qtd. de Amostras)	{500, <b>1000</b> , 1500, 2000}
Perturbações ao Fenômeno	<b>Sem Eventos Externos</b> e Com Eventos
Quantidade de Nós	{20, 35, <b>50</b> , 100}
% de Amostragem dos Dados	{10%, <b>25%</b> , 50%}
Potência do Rádio (dBm)	{0, -10, -15, <b>-25</b> }
Velocidade do sorvedouro Móvel (m/s)	{2, 3, 4, 5, 6}
Tempo de Navegação do sorvedouro Móvel (s)	{250, 1000, 2500, <b>5000</b> , 10000}
Modelo de Navegação do sorvedouro Móvel	<b>Realista (Rota Aleatória)</b> e Ideal (Rota Linear)
Modelo de Entrega de Pacotes	<b>Realista (Com Falhas)</b> e Ideal (Sem Perdas)

os outros em seus valores padrão previamente definidos, a lista com os parâmetros e seus respectivos *defaults* (em negrito) é mostrada na Tabela 1.

As Figuras 2(a) e 2(b) apresentam a flutuação dos valores de  $\hat{\epsilon}$  (Erro de Reconstrução) em função da variação do tamanho do Buffer e da quantidade de nós da rede. É possível observar que a abordagem sensível (Data-Aware, linha preta) aos dados apresentou menor erro em todos os cenários, com valores até 42% inferiores em relação ao Estado da Arte (Random Packet Drop, linha vermelha).

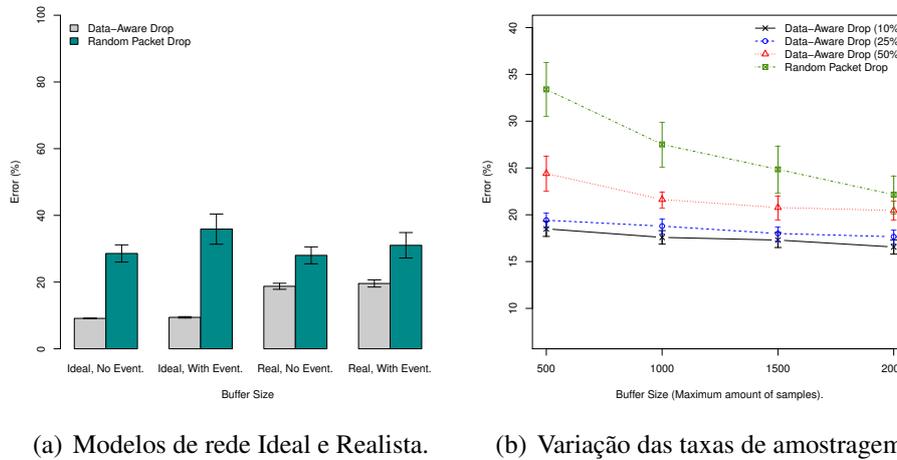


**Figure 2. Avaliação do Erro de Reconstrução.**

De forma análoga aos gráficos anteriores, avaliamos o  $\hat{\epsilon}$  (Erro de Reconstrução) à variação da **Potência do Rádio, Velocidade e Tempo de Navegação do sorvedouro**. Assim como as avaliações anteriores, a abordagem proposta apresentou valores de erro inferiores a 35%.

A Figura 3(a) evidencia o quanto a complexidade do modelo de rede interfere nos resultados. O **Modelo Ideal** não considera a perda de pacotes, a navegação do sorvedouro é ideal, largura de banda infinita e transmissão instantânea de pacotes. Por outro lado, o **Modelo Realista** considera a perda de pacotes por distância, baixa potência de rádio ou colisão de pacotes, movimento do sorvedouro com rota aleatória e largura de banda limitada.

Na **ausência do Evento Externo**, o comportamento das leituras segue os parâmetros definidos de média e variância definidos no início da seção 2. Em cenários

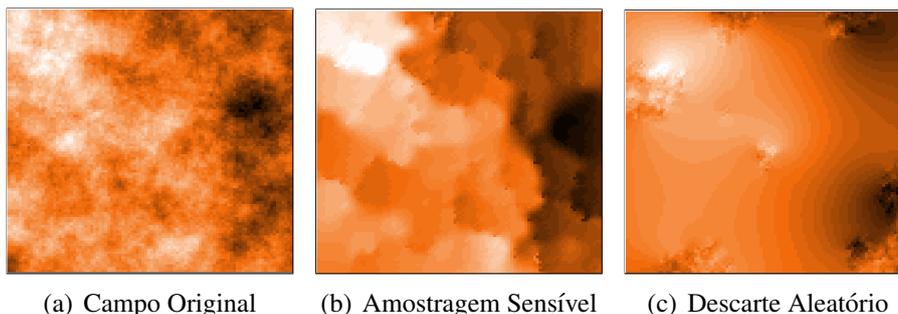


**Figure 3. Avaliação do Erro de Reconstrução.**

que consideram o evento, as leituras sofrerão variações bruscas em determinadas regiões do campo, adicionando complexidade ao fenômeno para a etapa de reconstrução. Nos resultados é possível observar uma redução no erro na ordem de 75% para o **Modelo Ideal** e 37% para o **Modelo Realista**.

A Figura 3(b) apresenta um comparativo dos percentuais de amostragem utilizados no algoritmo. Considerando que esta técnica monta um histograma a partir dos dados e realiza a amostragem em uma porcentagem pré definida deste conjunto, este experimento busca avaliar o impacto deste percentual na cobertura da rede, ou seja, o objetivo é definir o valor ótimo da porcentagem de dados amostrados em relação ao total. Neste experimento, consideramos como ótimo o percentual de amostragem fixado em 10% (redução de 90%) em virtude dos valores de **erro de reconstrução menores** em todos os casos mesmo em cenários com tamanho de Buffer limitado.

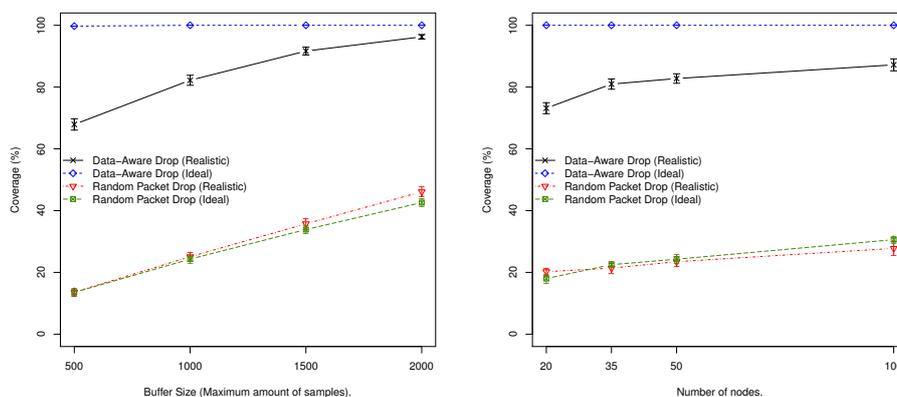
Finalmente, observando as Figuras 4(a), 4(b) e 4(c), é possível fazer uma comparação visual entre os resultados das técnicas avaliadas, essa comparação fortalece as conclusões obtidas através da análise dos dados. A Figura 4(a) apresenta o campo original, enquanto a Figura 4(c) apresenta grandes áreas contínuas com coloração constante resultantes da ausência de dados gerado pelo descarte aleatório de pacotes, enquanto o algoritmo de amostragem consegue preservar melhor seus dados (Figura 4(b)).



**Figure 4. Comparativo visual entre as técnicas de amostragem e descarte.**

Por fim, foi avaliado o desempenho das técnicas sob a métrica apresentada anteriormente. Os parâmetros de rede e níveis de complexidade dos modelos serão avaliados de forma análoga às avaliações anteriores.

Os gráficos apresentados na Figura 5 mostram o comportamento das técnicas quando colocadas à prova frente ao aumento de complexidade inerente aos ambientes reais, considerando uma quantidade significativamente maior de variáveis de interferência nos parâmetros em estudo, comparando este comportamento quando confrontado em relação ao modelo ideal, podendo além de visualizar com clareza as limitações deste modelo, também termos uma noção mais acurada do comportamento dos algoritmos em situações práticas (não simuladas). As linhas azul e preta correspondem a algoritmo sensível aos dados, onde é possível observar **resultados sempre superiores** em todos os cenários avaliados, **com larga diferença** (área de cobertura resultante **até 400% superior**) em relação à abordagem clássica (linhas em verde e vermelho).



(a) Variação do Buffer, quantidade de nós (b) Variação da quantidade de nós, Buffer fixo em 1000 amostras.

**Figure 5. Avaliação de Cobertura comparando os modelos Realista e Ideal.**

#### 4. Conclusão

Este trabalho apresentou uma solução baseada em amostragem para redução de dados de forma a minimizar perda de informação e otimizar a cobertura global da rede, apresentando resultados satisfatórios nos experimentos e **superando o estado da arte**.

Considerando essa modelagem sofisticada e a grande quantidade de cenários avaliados nos experimentos, os resultados são **conclusivos, consistentes e de larga abrangência**, trazendo informações resumidas de grande utilidade para adoção de **diretrizes para a execução de experimentos** em ambiente real, poupando tempo, esforços e custos de implementação por ter em mãos um grande leque de resultados disponíveis.

O projeto completo com guia para reprodução dos experimentos está disponível em repositório aberto para download no link <https://github.com/SensorNet-UFAL/DTN-Model4EnvApplications>. Finalmente, as perspectivas futuras incluem a incorporação do *know-how* desenvolvido nesta pesquisa para projeto e experimentação prática, desenvolvimento de novas aplicações e aprimoramento constante das soluções em estudo.

## References

- [Akyildiz et al. 2002] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cyirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422.
- [Aquino and Nakamura 2009] Aquino, A. L. L. and Nakamura, E. (2009). Data centric sensor stream reduction for real-time applications in wireless sensor networks. *Sensors*, 9:9666–9688.
- [Aurenhammer 1991] Aurenhammer, F. (1991). Voronoi diagrams: A survey of a fundamental data structure. *ACM Computing Surveys*, 23:345–405.
- [Baddeley 2007] 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.
- [Curran and Knox 2008] Curran, K. and Knox, J. (2008). Disruption tolerant networking. *Computer and Information Science*, 1(1):69–71.
- [Diggle and Ribeiro 2007] Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based Geostatistics*. Springer.
- [Frery et al. 2010] 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.
- [Hengl 2009] Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping*.
- [Li and Bartos 2014] Li, Y. and Bartos, R. (2014). A survey of protocols for intermittently connected delay-tolerant wireless sensor networks. *Journal of Network and Computer Applications*, 41(1):411–423.
- [Vasconcelos and Aquino 2014] Vasconcelos, I. L. C. and Aquino, A. L. L. (2014). Redução de dados baseada em amostragem para redes de sensores sem fio com tolerância a conexões disruptivas. In *XIV ERBASE, Workshop de Trabalhos de Iniciação Científica e Graduação (WTICGBASE)*.
- [Vasconcelos et al. 2015a] Vasconcelos, I. L. C., Lima, D. H. S., Figueiredo, C. M. S., and Aquino, A. L. L. (2015a). A coverage algorithm for intermittently connected delay tolerant wireless sensor networks. In *Proceedings of the 7th IFIP International Conference on New Technologies, Mobility and Security*.
- [Vasconcelos et al. 2015b] Vasconcelos, I. L. C., Lima, D. H. S., Figueiredo, C. M. S., and Aquino, A. L. L. (2015b). Estratégia de redução de dados baseada em amostragem para redes de sensores sem fio tolerantes a atrasos com conexões intermitentes. In *XV ERBASE, Workshop de Trabalhos de Iniciação Científica e Graduação (WTICGBASE)*.
- [Vasconcelos et al. 2015c] Vasconcelos, I. L. C., Lima, D. H. S., Figueiredo, C. M. S., and Aquino, A. L. L. (2015c). A sampling algorithm for intermittently connected delay tolerant wireless sensor networks. In *Proceedings of the 20th IEEE Symposium on Computers and Communications*.
- [Wood and Chan 1994] Wood, A. T. A. and Chan, G. (1994). Simulation of stationary gaussian processes in  $[0, 1]^d$ . *Journal of Computational and Graphical Statistics*, 3(4):409–432.