

# Uma investigação em uma base de dados geoespacial com uso de técnicas de *Big Data*

Jessica de O. Viana<sup>1</sup>, César A. da Silva<sup>1</sup>, Melissa M. P. Zanatta<sup>1</sup>, Bruno C. Vani<sup>1</sup>,  
Linnyer B. Ruiz Aylon<sup>3</sup>

<sup>1</sup>Instituto Federal de São Paulo (IFSP)  
19.470-000 – Presidente Epitácio – SP – Brasil

<sup>2</sup>Departamento de Informática – Universidade Estadual de Maringá (UEM)  
Maringá – PR - Brasil

jessicaoviana@yahoo.com.br, {cesar, melissa, brunovani}@ifsp.edu.br,  
linnyer@gmail.com

**Abstract.** *In global satellite navigation systems, a large volume of data is produced during the monitoring of radio waves propagated in the ionosphere. This monitoring is based on index calculations from data obtained continuously through receivers. This paper presents an investigation, using Big Data techniques, to explore the data obtained by the receivers, in order to identify the moment when the satellite signal can be obtained with better quality.*

**Resumo.** *Em sistemas globais de navegação por satélites, um grande volume de dados é produzido durante o monitoramento dos sinais de rádio propagadas na ionosfera. Esse monitoramento é baseado em cálculos de índices a partir de dados obtidos continuamente por meio de receptores. Este trabalho apresenta uma investigação, com uso de técnicas de Big Data, para explorar os dados obtidos pelos receptores, a fim de identificar o momento em que o sinal do satélite consegue ser obtido com melhor qualidade.*

## 1. Introdução

Os Sistemas Globais de Navegação por Satélites (*Global Navigation Satellite Systems – GNSS*) são compostos por conjuntos de satélites que permitem determinar as coordenadas da antena de um receptor (MONICO, 2008). Esse tipo de tecnologia é empregada em diversos segmentos, como a navegação, agricultura de precisão, topografia, aeronáutica, dentre outros. Nesse contexto, busca-se investigar a interferência provocada nos sinais dos satélites durante a passagem pela ionosfera.

A ionosfera é uma camada situada entre a baixa atmosfera e a magnetosfera, onde as ondas eletromagnéticas, utilizadas para comunicação, são refletidas (KOCHER, 2015). A variação rápida de amplitude e/ou fase dos sinais de ondas de rádio propagadas na ionosfera é chamada de cintilação ionosférica e pode influenciar diretamente na acurácia de aplicações que utilizam GNSS (INPE, 2011).

Os receptores GNSS utilizados para monitoramento de cintilação ionosférica fornecem estimativas de cintilação para cada satélite rastreado. Normalmente, esses receptores chegam a obter 50 amostras por segundo, gerando um grande volume de dados para serem armazenados (PEREIRA e CAMARGO, 2016). A manipulação de uma enorme base de dados em busca de informações relevantes faz parte do conceito utilizado na área de informática conhecido como *Big Data*. Esse conceito pode ser definido como um grande conjunto de dados que precisam de ferramentas especiais para exportar, organizar e transformar em informações que permitam fazer uma análise ampla e em tempo hábil (WU et al., 2014).

Este trabalho apresenta uma investigação, em uma base de dados geoespacial, com a finalidade de analisar e identificar o momento que ocorre a comunicação entre um receptor GNSS e um satélite com melhor qualidade.

## 2. Metodologia

O desenvolvimento deste trabalho é baseado no método de pesquisa experimental com características do tipo quantitativa. O estudo contempla o aprendizado de ferramentas e técnicas de *Big Data* para extrair informações relevantes em uma base de dados geoespacial que armazenam um grande volume de dados, podendo ultrapassar o volume de bilhões de registros. A base de dados utilizada nesta pesquisa é uma cópia parcial da base de dados CIGALA/CALIBRA, que foi solicitada junto à Faculdade de Ciências e Tecnologia (FCT/UNESP), câmpus Presidente Prudente, e está hospedada em servidor de alta capacidade de armazenamento e processamento localizado nas dependências da Coordenadoria de Tecnologia da Informação (CTI) do IFSP - Câmpus Presidente Epitácio (PEP).

As linguagens de programação mais utilizadas atualmente em *Big Data* são: R e Python. Ambas são *open source*, gratuitas e têm-se aprimorado no uso de enorme base de dados (CHIAVEGATTO, 2015). As ferramentas estudadas foram Jupyter Notebook e Google Colab. Estas ferramentas possuem diversas bibliotecas e proporcionam um ambiente de estudo em ciência de dados, análise preditiva e aprendizagem de máquina.

Com o uso da ferramenta Google Colab foi realizada a importação e análise dos dados a fim de compreender cada atributo que uma estação de monitoramento GNSS coleta e armazena. Inicialmente, foi escolhido um satélite e um período de 15 dias de coleta para realizar os estudos preliminares. Os dados desse período correspondem à 190239 registros e 66 atributos.

## 3. Processo de investigação e Resultados

A pesquisa busca investigar características no sinal de um satélite em relação à estação de monitoramento. A avaliação do sinal do satélite é feita, principalmente, pelo índice de cintilação denominado S4. Esse índice indica o nível de severidade de cintilações (ou flutuações) que afetaram a intensidade do sinal em um determinado período (no caso da base de dados utilizada, a cada minuto). Quanto menor esse valor, melhor é a qualidade do sinal. O índice S4 é obtido pela Equação 1, onde  $I$  se refere à intensidade do sinal amostrada a 50 Hz num intervalo de um minuto,  $std$  é o cálculo do desvio padrão e  $mean$  é o cálculo da média.

$$S4 = \frac{std(I)}{mean(I)} \quad (1)$$

Uma relação espacial entre a estação de monitoramento e o satélite é o ângulo denominado azimute. O valor do azimute é dado em graus a partir do Norte e varia de 0 a 360 graus. A elevação também é medida em graus e define a altura do satélite em relação a linha do horizonte.

A ferramenta Google Colab foi utilizada para importar os dados armazenados no servidor e iniciar a análise exploratória dos dados. Uma estação de monitoramento de satélites coleta diversas informações dos satélites, por isso, inicialmente, foi realizada uma análise dos dados e foram excluídos os atributos desnecessários. A Figura 1 mostra os cinco primeiros registros da base de dados com os atributos selecionados para a investigação.

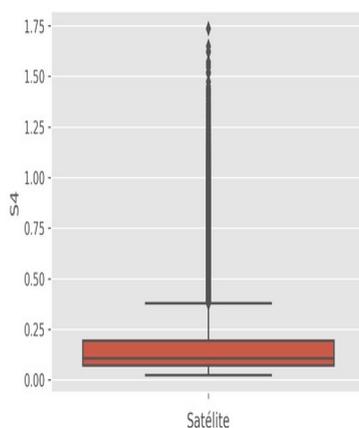
```
data.head()

  s4  azim
0 0.212 138.0
1 0.104 108.0
2 0.080 63.0
3 0.644 305.0
4 0.432 5.0
```

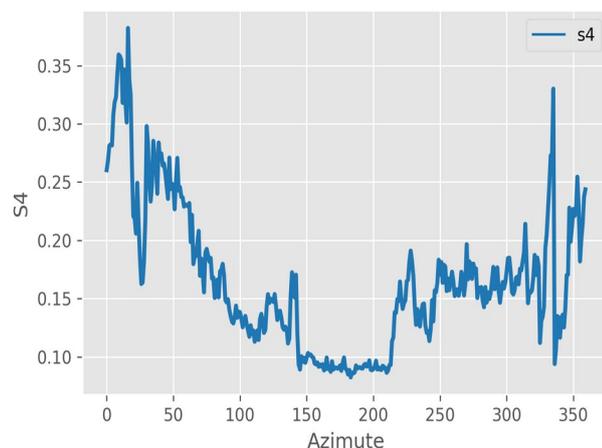
**Figura 1: Base de dados após a seleção dos atributos.**

Após a seleção dos atributos, os registros com valores indefinidos (*Not a Number - NaN*) foram excluídos e iniciou a análise exploratória dos dados. Na análise de dispersão dos valores de S4 foi gerado um gráfico do tipo *boxplot*, como pode ser visualizado na Figura 2. A partir desse gráfico, pode-se dizer que a maioria dos valores do índice S4, 75% dos dados, estão abaixo de 0.20, porém existe uma parte dos dados que são valores discrepantes (*outliers*).

Agora, deseja-se saber o momento que se consegue obter o melhor sinal do satélite. A Figura 3 apresenta o gráfico de distribuição do índice de cintilação S4 em relação ao valor do azimute. Pode-se observar que o sinal do satélite é recebido com menor indicativo de cintilações quando está posicionado entre 140 e 220 graus do azimute.



**Figura 2: Visualização da dispersão dos valores de S4.**



**Figura 3: Distribuição do índice S4 em relação ao Azimute.**

## 4. Conclusões

As aplicações de GNSS têm gerado uma enorme quantidade de dados, porém, organizar e extrair informações relevantes não é uma tarefa trivial. Este trabalho está em busca de aplicar as técnicas de *Big Data*, em uma base de dados com bilhões de registros, com objetivo de investigar a dinâmica do fenômeno e seus efeitos no posicionamento dos satélites. O desenvolvimento deste trabalho contribui para a extração de informações relevantes para o contexto de sistemas globais de navegação por satélites, como por exemplo, identificar o horário mais comum em que um determinado satélite sofre mais influência na ionosfera em seu caminho de propagação até o receptor de monitoramento.

## Agradecimentos

Os autores agradecem ao IFSP – Câmpus Presidente Epitácio pela infraestrutura e suporte fornecidos. Ao INCT – Sistemas Micro e Nanoeletrônicos (Nमितec) e ao INCT - Tecnologia GNSS no Suporte à Navegação Aérea (GNSS NavAer), financiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte à realização desta pesquisa.

## Referências

- Chiavegatto, A. D. P. (2015) “Uso de big data em saúde no Brasil: perspectivas para um futuro próximo”, *Epidemiologia e Serviços de Saúde* [online], v. 24, n. 2, p. 325-332.
- INPE (2011). “Cintilação”, Instituto Nacional de Pesquisas Espaciais. Disponível em: <<http://www2.inpe.br/climaespacial/SWMonitorUser/faces/about.xhtml>>. Acesso em: 03 set. 2019.
- Kocher, J. M. (2012) “Ionosfera:os caminhos da descoberta”, In: *Scientiarum Historia V Filosofia, Ciências e Artes: conexões interdisciplinares*, Rio de Janeiro. Livro de Anais *Scientiarum Historia V*. Rio de Janeiro: UFRJ, p. 1-8.
- Monico, J.F.G. (2008) “Posicionamento pelo GNSS”, São Paulo: Editora Unesp, p.477.
- Pereira, V. A. S., Camargo, P. O. (2016) “Programa científico para monitoramento em tempo real ou pós-processado das irregularidades ionosféricas e cintilação dos sinais GNSS”, *Boletim de Ciências Geodésicas*, 22(2), 282-302.
- Wu, X. et al. (2014) “Data mining with big data”, *IEEE Trans. Knowl. Data Eng.*, 26 (1), p. 97-107.