



# ANÁLISE DA CORRELAÇÃO ENTRE ÍNDICE DE CINTILAÇÃO E O CONTEÚDO ELETTRONICO TOTAL DA IONOSFERA NO PICO DA ANOMALIA MAGNÉTICA.

Pedro Alexandre dos Santos

Monografia para Exame de Qualificação do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Stephan Stephany, aprovada em 11 de dezembro de 2018.

URL do documento original:  
<http://urlib.net/xx/yy>

INPE  
São José dos Campos  
2018

**PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE  
Gabinete do Diretor (GB)  
Serviço de Informação e Documentação (SID)  
Caixa Postal 515 - CEP 12.245-970  
São José dos Campos - SP - Brasil  
Tel.:(012) 3945-6923/6921  
Fax: (012) 3945-6919  
E-mail: [pubtc@sid.inpe.br](mailto:pubtc@sid.inpe.br)

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO  
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):**

**Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

**Membros:**

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr<sup>a</sup> Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

**BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

**REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

**EDITORAÇÃO ELETRÔNICA:**

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



# ANÁLISE DA CORRELAÇÃO ENTRE ÍNDICE DE CINTILAÇÃO E O CONTEÚDO ELETTRONICO TOTAL DA IONOSFERA NO PICO DA ANOMALIA MAGNÉTICA.

Pedro Alexandre dos Santos

Monografia para Exame de Qualificação do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Stephan Stephany, aprovada em 11 de dezembro de 2018.

URL do documento original:  
<http://urlib.net/xx/yy>

INPE  
São José dos Campos  
2018



Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

Informar aqui sobre marca registrada (a modificação desta linha deve ser feita no arquivo publicacao.tex).

## LISTA DE FIGURAS

	<u>Pág.</u>
3.2 A subfigure . . . . .	13
3.3 A subfigure . . . . .	13
3.4 A subfigure . . . . .	15
3.5 A subfigure . . . . .	15
3.6 A subfigure . . . . .	18
3.7 A subfigure . . . . .	19
3.8 A subfigure . . . . .	20
3.1 A subfigure . . . . .	24
3.9 A subfigure . . . . .	25



## **LISTA DE TABELAS**

Pág.



## **LISTA DE ABREVIATURAS E SIGLAS**

WETAMC	– Campanha de Mesoescala Atmosférica na Estação Úmida
IBGE	– Instituto Brasileiro de Geografia e Estatística
MC	– Método das Covariâncias
EDO	– Equações Diferenciais Ordinárias
EDP	– Equações Diferenciais Parciais
ECT	– Energia Cinética Turbulenta
FDP	– Função de Distribuição de Probabilidade
PR	– Plot de Recorrência
FFT	– Fast Fourier Transform
tS1200	– Temperatura medida no nível superior às 12 horas
tS2300	– Temperatura medida no nível superior às 23 horas
tM1200	– Temperatura medida no nível médio às 12 horas
tM2300	– Temperatura medida no nível médio às 23 horas
tI1200	– Temperatura medida no nível inferior às 12 horas
tI2300	– Temperatura medida no nível inferior às 23 horas
wS1200	– Velocidade vertical do vento medida no nível superior às 12 horas



## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO . . . . .</b>	<b>1</b>
<b>2 IONOSFERA . . . . .</b>	<b>7</b>
2.1 Anomalias na ionosfera . . . . .	8
2.2 TEC, VTEC e S4 . . . . .	9
2.3 Bolhas de Plasma . . . . .	9
<b>3 Metodologia . . . . .</b>	<b>11</b>
3.1 . . . . .	11
3.2 Seleção inicial dos dados de S4 e estações . . . . .	11
3.3 Gerando a série espaço-temporal para os dados de VTEC . . . . .	11
3.4 Gerando um mapa das estações com os meridianos magnéticos . . . . .	12
3.5 Suavização da sérire temporal de S4 . . . . .	12
3.6 Extração da série temporal de VTEC, para algumas estações . . . . .	13
3.7 Seleção fina dos dados de S4 . . . . .	16
3.8 Visualizando S4 e VTEC . . . . .	21
3.9 Análises: S4×VTEC em São José dos Campos . . . . .	21
<b>REFERÊNCIAS BIBLIOGRÁFICAS . . . . .</b>	<b>27</b>



## 1 INTRODUÇÃO

A necessidade de localização no espaço levou a humanidade ao desenvolvimento de diversas ferramentas, tais como os sistemas de coordenadas, a bússola, os mapas, e mais recentemente o Sistema de Posicionamento Global (GPS). Desenvolvido pelos norte americanos se tornou completamente operacional em 1995, com um custo estimado de 10 bilhões de dólares. Consiste de uma constelação de 24 satélites, cada um circulando a Terra duas vezes ao longo do dia, em uma configuração em que ao menos 4 satélites sejam visíveis de qualquer ponto da Terra. O receptor do sinal utiliza a informação enviada pelo satélite para calcular sua distância a cada um destes utilizando a informação entre o instante de recebimento e o instante de transmissão.

Em 1991, a associação internacional de Aviação Civil utilizou pela primeira vez o termo sistema de navegação por satélite (GNSS), para denominar todo e qualquer sistema semelhante ao GPS, que atualmente é usado para denominar o sistema americano, também conhecido como Navstar GPS. Outro sistema que se encontra completamente operacional é o russo GLONASS. O sistema chinês COMPASS e o europeu GALILEO se encontram em fase de implementação. Considerando a diversidade de sistemas é possível notar sua grande relevância, e como todo sistema está sujeito a pertubações e interferências que devido aos impactos precisam ser estudados, entre elas se pode citar a cintilação ionosférica.

A cintilação ionosférica, medida por meio do índice S4, é definida como o desvio padrão da intensidade do sinal de GPS em um intervalo de um minuto, com 50 amostras por segundo, portanto, quanto maior a perturbação do sinal, maior o valor deste índice. Uma observação que pode ser extraída desta definição, é que dada uma causa para a perturbação do sinal de GPS, é possível que outros sinais de radiofrequência, em regiões relativamente próximas do espectro apresentam interferência semelhante, logo, o índice S4 pode ser interpretado como uma medida de perturbação em sinais de radiofrequência.

Os fenômenos de cintilação, de interesse para este trabalho, decorrem da trajetória do sinal percorrer regiões da ionosfera com baixa densidade de elétrons, com períodos de duração de algumas horas. Estas regiões migram na ionosfera, geralmente no sentido do Sul e do Leste magnético, podendo se expandir ou contrair, sendo denominadas de bolhas ionosféricas. Estas surgem no Hemisfério Sul e sua ocorrência começa a se intensificar de outono a novembro, apresentando picos ao longo do verão, a atividade começa a reduzir em Março. Sua formação começa entre 22:00 UT e 23:00 UT na

região do equador magnético, encerrando-se entre as 04:00 e 05:00 UT. Além da atividade mais intensa durante o verão é possível observar uma grande variabilidade dia a dia no decorrer do ano, o que torna difícil sua previsão, (TAKAHASHI et al., 2016).

A ionosfera, camada de gás da atmosfera ionizada pela radiação solar, é a região onde o fenômeno de interesse se manifesta. Este, assim como outros fenômenos, como a aurora boreal, apresentam um certo nível de organização devido às linhas de campo magnético da Terra. Isto ocorre pois partículas ionizadas ou carregas podem ser mover livremente ao longo das linhas magnéticas mas não entre elas. Assim, o estudo do campo magnético da Terra se faz relevante para um grande número de aplicações, (LAUNDAL; RICHMOND, 2017). Um resultado imediato deste estudo é que o norte geográfico e magnético não coincidem, e que simplificadamente o campo magnético poderia ser descrito por um dipolo magnético, com centro comum ao da Terra, porém inclinado em relação a linha que liga o norte e sul geográfico. Atualmente, existe vários sistemas de coordenadas magnéticas cujo propósito dependem da região, da aplicação e da faixa de altitude de interesse, para uma revisão entre os sistemas mais comuns consulte a referência (LAUNDAL; RICHMOND, 2017). Para este trabalho foi adotado o sistema AACGM, pois é mais adequado à altura ionosférica, contudo ele pertence a classe de sistemas não-ortogonais.

A ionosfera devido a vários fatores, não somente o campo magnético da Terra, mas toda a interação com o sistema solar, onde o Sol é um dos principais atores, uma vez que é a principal fonte de radiação ionizante, apresenta uma riqueza de fenômenos. Agora, levando em consideração a inclinação do plano de rotação da Terra, em relação ao plano de translação é evidente que em certas regiões há a formação de anomalias. Dentre estas uma de particular interesse, na região Sul, é a anomalia da ionização equatorial que consiste na formação de uma região de alta densidade de elétrons entre 15 e 20 graus de latitude magnética, logo após ao por do Sol. O máximo dessa densidade é denominado de pico da anomalia magnética e se manifesta na região do vale do Paraíba.

Atualmente, não existe um modelo matemático, que expresse a física do fenômeno, capaz de modelar e predizer o surgimento e a evolução das bolhas, de forma, que a predição do decorrente fenômeno de cintilação também não é possível. Todavia, os impactos decorrentes desse fenômeno, como a perda de sistemas de navegação, em uma sociedade que apresenta cada vez mais um consumo por este tipo de informação, seja em sistemas de produção como na agricultura (STAFFORD, 2000), ou na

aviação, ou para o simples uso pessoal ao percorrer uma cidade, podem ser catastróficas, implicando, por exemplo, em perdas de vidas humanas, ou na redução na produção de alimentos. No caso da aviação, há uma tendência mundial de se utilizar unicamente navegação e procedimentos de pouso/decolagem baseados em GNSS.

Assim, faz-se desejável uma abordagem que permita prever a formação e a evolução das bolhas ionosféricas, bem como a intensidade da cintilação dada pelo índice S4, entretanto note que esta é uma quantificação da intensidade de falha do GPS, e não da bolha ionosférica, logo é necessário um valor numérico que forneça informação com enfoque na ionosfera. O conteúdo eletrônico total vertical pode ser este número, visto que é a quantidade de elétrons, em uma seção unitária, ao longo de dois pontos da atmosfera, sua unidade de medida é o TECU=  $10^{16}$ elm<sup>2</sup>. Neste caso, a bolha pode ser compreendida como uma depressão com 30 à 50 TECU de diferença em relação a sua vizinhança.

Na falta de uma modelagem matemática que possa simular um fenômeno tão complexo, a alternativa é a utilização de modelos direcionados por dados, isto é, modelos matemáticos que são ajustados (será denominado de aprendizagem, no contexto de aprendizado de máquina), por exemplo, por processo de optimização, para um conjunto de dados. Em um primeiro momento, este termo e sua colocação podem aparecer estranhos, entretanto, eles estão presentes desde o inicio da ciência, um exemplo, são as três leis de Kepler da astronomia, pois estas foram elaboradas por Johannes Kepler extraindo informações de um extenso conjunto de observações catalogadas por Tycho Brahe, ao longo, de sua vida.

Note, que no exemplo desenvolvido, tem-se um base de dados, isto é, uma grande massa de dados, e um novo conhecimento foi extraído desta: foi realizada o que hoje se chama uma mineração de dados, ou seja, foi extraído, ou inferido conhecimento novo que não era evidente em uma grande base de dados. A mineração também é adequada para descrever de maneira concisa os dados, vindo talvez colaborar com a predição de eventos ([HAN et al., 2011](#)).

Existem termos que se confundem devido a má utilização, por exemplo, peso e massa, e aqui mineração de dados e descoberta de conhecimento em base de dados, esta é a descoberta de conhecimento, pela extração de padrões representando conhecimento armazenado em um conjunto de dados, enquanto aquela é a extração de padrões em um conjunto de dados. A diferença fica mais clara sumarizando os processos contidos na abordagem de descoberta de conhecimento em base de dados:

- a) **Limpeza de dados** para a remoção de ruídos, dados inconsistentes e ou incompletos;
- b) **Integração de dados**, onde múltiplas fontes de dados podem ser integradas;
- c) **Seleção de dados**, onde dados relevantes a tarefa de análise são extraídos do banco de dados;
- d) **Transformação de dados**, onde os dados são transformados e consolidados em uma forma mais adequada para mineração realizando operações de sumarização ou agregação;
- e) **Mineração de dados**, que é um processo essencial onde métodos inteligentes, como redes neurais, árvores de decisão são aplicados para a extração de padrões;
- f) **Avaliação de padrões** para identificar os padrões verdadeiramente interessantes;
- g) **Apresentação do conhecimento**, onde técnicas de visualização e representação dos dados são usadas para exibir o conhecimento extraído aos usuários.

A maior parte do tempo em um processo de descoberta de conhecimento é gasto na preparação dos dados, o que inclui seleção, pré-processamento e transformação de dados, contudo os processos são interativos, por exemplo, após minerar pode ser observado a necessidade de uma nova transformação nos dados, pois, isso explica o grande tempo consumido na fase de preparação.

A mineração de dados é realizada utilizando os dados preparados segundo algum algoritmo. Atualmente, pode-se dizer que estes podem ser agrupados em dois conjuntos maiores os supervisionados, e os não supervisionados, entretanto na última década começou a aparecer algoritmos autosupervisionados, semisupervisionados, entre outras variações. Um algoritmo é supervisionado, quando seu dado de entrada é dividido em um conjunto de exemplo e seu correspondente rótulo, tal que o algoritmo busca aproximar uma função que mapeia as variáveis no alvo. Seja  $\mathcal{X}$  o domínio das observações,  $\mathcal{Y}$  o domínio dos rótulos,  $P(\mathbf{x}, y)$  a probabilidade conjunta sobre as observações e rótulos  $\mathcal{X} \times \mathcal{Y}$ , um algoritmo supervisionado tem por objetivo treinar uma função  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , tal que  $f(\mathbf{x})$  consiga predizer o rótulo  $y$  correto para uma

dada observação  $\mathbf{x}$ , a partir de um conjunto de treino  $\{(x_i, y_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} P(\mathbf{x}, y)$  (ZHU; GOLDBERG, 2009).

A classificação e a regressão, são dois exemplos de algorítimos supervisionados que visam predizer as saídas lidando com valores discretos e contínuos, respectivamente. Algoritmos não supervisionados, por sua vez, são utilizados quando se deseja encontrar algum padrão na estrutura do conjunto de dados. Comumente métodos como, clusterização, detecção de novidade e redução da dimensionalidade são aplicados neste tipo de paradigma, onde o conjunto de treino  $\{x_i\}_{i=1}^n$  não diz nada a respeito sobre os rótulos de cada observação.

Considerando-se um escopo mais ambicioso, de predição da ocorrência de cintilação, o objetivo deste trabalho restringe-se à utilização de técnicas de mineração e visualização de dados históricos da ionosfera, de forma a analisar a evolução espaço-temporal do índice S4 e do TEC e sua correlação. Este trabalho objetiva complementar trabalhos anteriores relacionados à predição da ocorrência de cintilação (REZENDE, 2009; LIMA et al., 2014; LIMA et al., 2015), além de trabalhos relacionados ao objetivo específico aqui abordado, como por exemplo a correlação entre o gradiente do TEC e a cintilação (RAGHAVARAO et al., 1998; RAY et al., 2006), ou então a correlação entre a derivada temporal do TEC e a cintilação (RAGHUNATH SWAPNA ; VENKATA RATNAM, 2016).

Os dados utilizados neste trabalho se restringem ao período de atividade solar máxima de um ciclo solar, no caso, de 01/12/2013 até 28/02/2014. Devido a problemas técnicos e de equipamentos/sensores, existem muitas lacunas nestes dados, exigindo cuidados especiais no seu processamento.



## 2 IONOSFERA

A ionosfera é uma região ionizada da alta atmosfera, estendendo-se entre 60Km até 10000Km de altitude, assim, engloba partes da mesosfera, termosfera, e exosfera. Esta camada constitui-se de íons, elétrons livres criados primariamente por processo de fotoionização, e alguma porção de gás neutro. A fotoionização ionosférica consiste de um processo físico-químico, onde alguma espécies químicas presentes na atmosfera ganham ou perdem elétrons decorrentes da absorção de radiação solar predominantemente nas faixas mais altas do ultravioleta e raios-X ([Rishbeth; Garriott, 1969](#); [NEGRETI, 2012](#)). A ionização, também, pode ocorrer devido a colisões com partículas altamente energéticas, provindas do meio solar ou galácticas, são mais facilmente observadas em altas latitudes, e na região da Anomalia Magnética do Atlântico Sul.

A ionosfera pode ser dividida em regiões, faixas de altitudes, as quais se diferenciam pelos processos físicos e químicos que governam o comportamento daquela faixa, além disso, estes mesmos processos podem variar devido à quantidade de radiação solar recebida, logo, vai-se observar diferenças entre a noite e o dia. Ao longo da noite, a camada F é a única que apresenta uma ionização significativa, enquanto as camadas E e D apresentam um valor extremamente baixo de ionização. Durante o dia, a camada D e E se tornam mais ionizadas, assim como a camada F, que se divide em duas regiões, F1 que é mais fracamente ionizada, e F2 que é mais intensamente ionizada. A camada F2 existe durante a noite e durante o dia, sendo a principal responsável pela reflexão e refração dos sinais de rádio.

A camada D é a mais interna, estando entre 60Km à 90Km acima da superfície da Terra. Sua ionização é devida à radiação do hidrogênio ionizado na série de Lyman-alpha no comprimento de onda de 121.6nm ionizando o óxido nítrico, *NO*, presente na camada. Além disso, raios X altamente energéticos, com comprimento de onda inferior de 1nm podem ionizar as moléculas de *N<sub>2</sub>* e de *O<sub>2</sub>*. A camada D apresenta alta taxa de recombinação, de modo que existem mais moléculas neutras do que íons. Tem uma taxa de absorção considerável para ondas de rádio de média e alta baixa frequências, e baixas frequências apresentam elevada atenuação, principalmente, devido à absorção de energia pelo elétrons livres, o que aumenta as chances de colisão. Este efeito desaparece durante à noite, devido a uma menor ionização. Pode apresentar valores elevados de ionização em altas latitudes em decorrência de erupções solares com grandes quantidades de matéria hadrônica, prótons, em sua maioria, com uma duração de 24 à 48 horas.

A camada E é a intermediária e está situada entre 90Km à 150Km acima da superfície da Terra. A ionização decorre principalmente devido ao espalhamento de raio-X leve (entre 1 e 10nm) e ultravioleta distante (UV) provindos do Sol com moléculas de oxigênio. A estrutura vertical da camada E é determinada em sua maior parte pela competição entre efeitos de ionização e de recombinação. É importante pela presença de correntes elétricas que nela fluem e interagem com o campo magnético (KIRCHHOFF, 1991). A noite, a camada E quase desaparece, pois sua fonte primária de ionização não está presente.

A camada F se estende de 150Km à mais de 500Km acima da superfície da Terra. Apresenta a maior concentração de elétrons, portanto, sinais que são capazes de penetrar até esta são capazes de escapar para o espaço. Predominam, nesta, a ionização de átomos de oxigênio por meio de radiação solar no espectro do extremo ultravioleta, entre, 10nm e 100nm. A camada é subdividida em duas regiões a F2 que está presente durante o dia e a noite, e a F1 que aparece somente durante o dia.

A subcamada F2 se inicia aproximadamente a 300Km de altitude, englobando toda a região superior da ionosfera, inclusive a região de pico da densidade de elétrons. Este máximo no perfil vertical de ionização decorre do balanço entre os processos de transporte de plasma e os processos físico-químicos. Acima deste pico, a ionosfera se encontra em equilíbrio difusivo, ou seja, o plasma se distribui com a sua própria escala de altura. A presença do campo magnético contribui para a distribuição da ionização.

## 2.1 Anomalias na ionosfera

A ionosfera apresenta várias anomalias, ou seja, várias irregularidades na distribuição de elétrons. Este trabalho tem interesse na anomalia equatorial. Esta aparece aproximadamente entre 15 e 20 graus de latitude magnética, tanto no hemisfério norte, quanto no hemisfério sul, na camada F2. Consiste na formação de uma região de alta densidade eletrônica, e é uma anomalia, pois a densidade de plasma deveria ser maior em regiões equatoriais, e não em latitudes magnéticas mais altas.

Sua origem decorre da deriva vertical do plasma da camada F na região equatorial: o processo de ionização da camada F faz surgir um campo elétrico, apontando para leste, enquanto o campo magnético aponta para o norte, considerando então  $\vec{E} \times \vec{B}$ , tem-se o surgimento de uma força perpendicular ao campo magnético e ao campo elétrico, o que neste caso, aponta para cima, deslocando o plasmas para regiões de mais alta altitudes. Agora, quando em altas altitudes, o plasma for efeito gravita-

cional e diferença de pressão é trazido de volta à altitudes mais baixas, porém este movimento de descida é mais eficiente ao longo das linhas de campo magnético, levando a um aumento na densidade de plasma em regiões de médias latitudes.

A distribuição do plasma também pode ser alterada pela ação de outras variáveis, como o vento. São José dos Campos, encontra-se na região da anomalia equatorial, mais especificamente no pico da anomalia, ou seja, na região onde densidade de plasma em altas altitudes atinge seu valor máximo.

## 2.2 TEC, VTEC e S4

TEC, ou conteúdo total de elétrons é uma quantidade descritiva utilizada para avaliar a densidade do plasma ionosférico. É o número total de elétrons integrado entre dois pontos, ao longo de um tubo com seção de  $1\text{ m}^2$ , ou seja, é uma densidade numérica de elétrons, ao longo, de uma coluna. Geralmente é reportada em unidades de TEC (TECU), definido com  $\text{TECU} = 10^{16}\text{ el/m}^2$ . É importante para determinar a cintilação e os atrasos de fase e de grupo em ondas de rádio no meio.

VTEC, ou conteúdo total de elétrons vertical é projeção do TEC, ao longo de uma linha normal a superfície da Terra, em outras palavras, ela fornece o conteúdo de elétrons ao longo da normal para cada ponto da superfície da Terra.

S4 corresponde ao desvio padrão da intensidade do sinal de GPS, e é utilizado para medir cintilação ionosférica, que é a perda do sinal. Este fenômeno tem sua origem em não homogeneidades na distribuição do plasma ionosférico. Dentro das diversas possíveis irregularidades, que levam a ocorrência de cintilação, existe um interesse particular nas bolhas de plasma.

## 2.3 Bolhas de Plasma

As bolhas ionosféricas podem ser definidas como regiões de baixa densidade de plasma ionosférico quando comparadas com a sua vizinhança. Utilizando medidas de VTEC é possível definir essa diferença como 30-50 TECU (TAKAHASHI et al., 2016).

São originadas na região equatorial, após a rápida elevação do plasma, devido a anomalia equatorial, isto é, o plasma ao acender cria regiões de baixa densidade. Após sua formação podem evoluir para altas altitudes (centenas de quilômetros), estendendo-se ao longo das linhas de campo magnético (milhares de quilômetros)

nas direções norte-sul, alcançado em torno de 20 graus de latitude magnética.

### **3 Metodologia**

#### **3.1**

Este trabalho usa o índice S4, uma lista de estações que realizam medidas deste e o VTEC. O trabalho também adota a ideia de pesquisa reproduzível, com códigos documentados, para tal se faz uso da tecnologia de “notebooks” em Python. Além disso, os “notebooks” estão disponíveis em <https://github.com/retiarus/tese>, juntamente com um conjunto de dados.

Os dados de S4 foram originalmente disponibilizados em formato de texto, juntamente com uma lista de estações ao longo do território brasileiro. Cada arquivo contém uma lista de medidas ordenadas pelo tempo, com múltiplas medidas por minuto cada uma associada a um ao longo da linha de visada da estação satélite. A etapa inicial consistiu em armazenar e organizar estes dados em um banco de dados, pois então, técnicas como filtragem permite uma rápida seleção dos dados por estação e elevação. Também foi adicionado ao banco um tabela com dados para as estações.

Os notebooks estão organizados tal que os primeiros dois dígitos indiquem estabelecem uma ordem de execução, por exemplo, o notebook com inicial 00 precisa ser executado antes do notebook com inicial 01. Existem notebooks com os mesmos valores de dígitos, isto é, podem existir dois notebooks com os dígitos 02, isto significa que um não apresenta dependência em relação ao outro e podem ser executados ao mesmo tempo.

#### **3.2 Seleção inicial dos dados de S4 e estações**

Este notebook realiza uma consulta a tabela de estações no banco de dados. Utilizando, então, a lista de estações retorna faz iterativamente uma consulta para cada estação selecionando apenas medidas cuja elevação é superior a 30.0. Somente são mantidas as estações que contém medidas, gerando assim, uma tabela final com os dados das estações e um conjunto de arquivos com dados de S4, um para cada estação válida.

#### **3.3 Gerando a série espaço-temporal para os dados de VTEC**

Os dados de VTEC estão inicialmente organizados em arquivos de texto, as duas primeiras linhas são de cabeçalho, onde a primeira, denota o instante da configuração, a segunda fornece o significado de cada coluna. A primeira indica a longitude, a

segunda a latitude e a terceira o VTEC. Medidas de VTEC 999.000 nestes arquivos denotam ausência de valor. Finalmente, cada arquivo constitui um mapa de VTEC, onde cada linha fornece o valor de VTEC para um ponto deste. Assim, o papel do notebook `00_generate_vtec_dataframe.ipynb` é o de converter este conjunto de arquivos em um tabela, indexada pelo tempo, onde cada linha contém uma matriz, que realiza o papel do mapa. Pode-se observar esta estrutura também como uma série espaço-temporal, no qual um índice de tempo qualquer indexa um mapa de VTEC. Este notebook também realiza um ordenamento temporal nos dados, o que será utilizado posteriormente para o cálculo de derivadas temporais.

### 3.4 Gerando um mapa das estações com os meridianos magnéticos

Este notebook realiza um ordenamento na tabela de estações por estado e cidade. Seguindo de um agrupamento por cidade, uma vez que existem cidades com mais de um estação e neste trabalho se utilizou apenas uma. Então, utiliza os dados de localização das estações, a altitude destas e uma data de referência para utilizando modelo do sistema de coordenadas geomagnética AACGM, por meio do pacote Python AACGMV2, gerar as latitudes e longitudes geomagnéticas para os dados. Utilizando, esses dados plotou-se um mapa contendo as estações, assim, como os meridianos magnéticos que passam por estas. Lembrando-se que as bolhas ionosféricas evoluem ao longo dos meridianos magnéticos se tem que este mapa fornece uma visualização das estações que apresentam maior correlação entre si.

### 3.5 Suavização da sérire temporal de S4

O notebook `02_preprocessing_s4_data.ipynb` é utilizado para o pré-processamento dos dados de S4. Primeiramente, toma-se o valor médio de todas as medidas realizada por minuto, isto é, realiza-se uma reamostragem combinando as medidas de diferentes satélites no mesmo minuto, tal que se tenha um valor de S4 por minuto. Após esta etapa é feita uma interpolação spline de ordem 4 para tratar instantes sem amostras. Em seguida, é feita uma reamostragem para intervalos de 10 min, novamente utilizando o valor médio, este intervalo de tempo foi escolhido, pois é o mesmo do dados de VTEC. Nas figuras [3.2](#) e [3.3](#) é possível observar em azul uma amostra do sinal S4, pode-se notar que este apresenta rápidas flutuações e ruído. Assim, optou-se por aplicar uma função de suavização. Em [3.2](#), tem-se a utilização do filtro de Savitzky–Golay com janela de 9 pontos e polinômio interpolador de terceira ordem, enquanto na figura [3.3](#) foi utilizado um filtro baseado em média móvel com pesos gaussianos com uma janela também de 9 pontos.

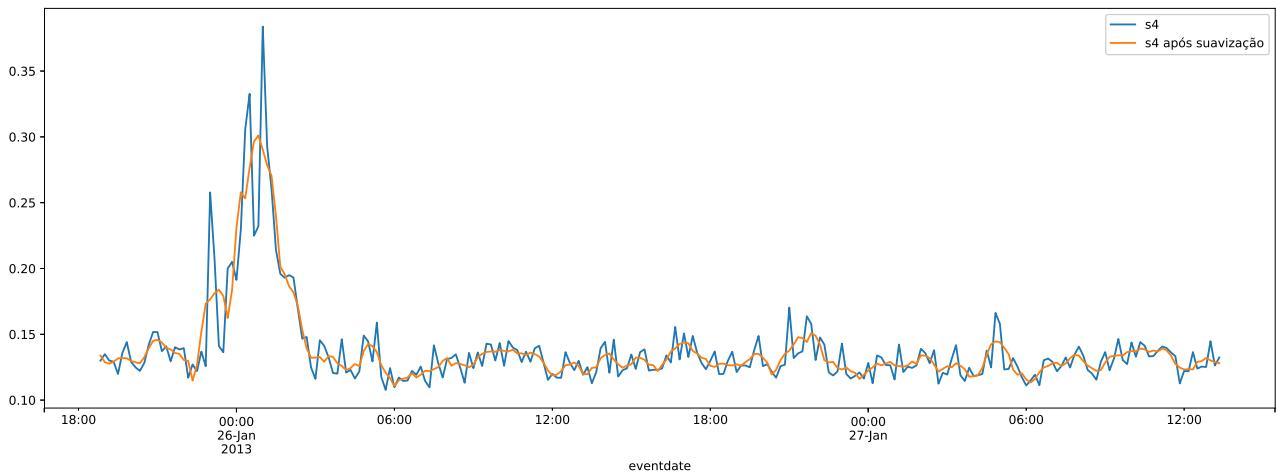


Figura 3.2 - A subfigure

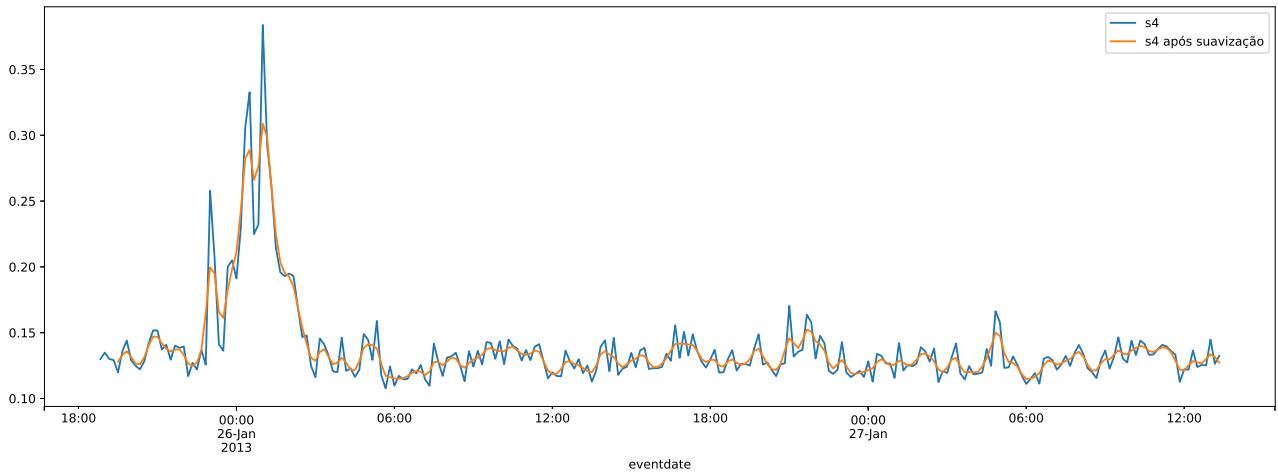


Figura 3.3 - A subfigure

Finalmente, optou-se por utilizar uma combinação das duas técnicas de suavização aplicando primeiro do filtro de Savitzky–Golay seguido da média móvel com pesos gaussianos, ambos com os parâmetros especificados no parágrafo anterior.

Este notebook gera também uma tabela com todos os dados de S4, em passos de 10 min, onde as colunas representam o conjunto inicial de estações.

### 3.6 Extração da série temporal de VTEC, para algumas estações

O papel do notebook 03\_extract\_vtec\_stations.ipynb é o de extrair da série espaço-temporal do VTEC, séries temporais para as estações onde o índice S4 é medido.

Estes dados então são preprocessados aplicando o mesmo processo de suavização utilizados nos dados de S4. Na figura 3.4 é possível visualizar uma amostra do VTEC para São José dos Campos, juntamente com a aplicação separa das duas técnicas de suavização. Na figura 3.5 há uma amostra do VTEC para o conjunto inicial de estações, observe que no intervalo entre às 9:00 e 15:00 UT o VTEC é aproximadamente igual entre as diversas estações, e que a partir das 18:00 UT os valores começam a divergir entre si, apresentando grandes diferenças após 00:00 até 06:00 UT, onde então começa a se agrupar novamente.

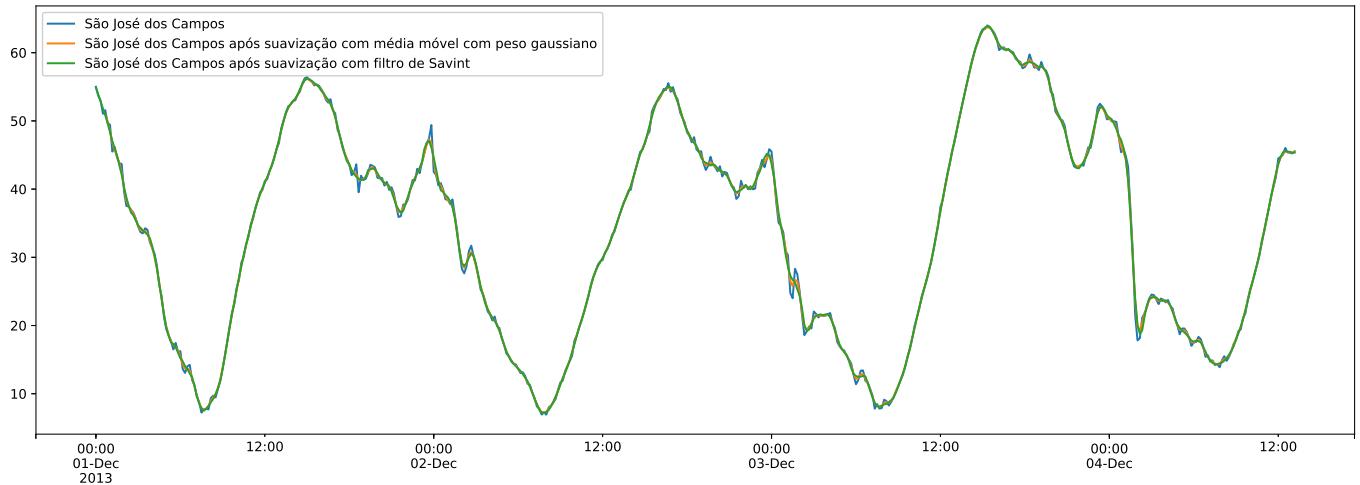


Figura 3.4 - A subfigure

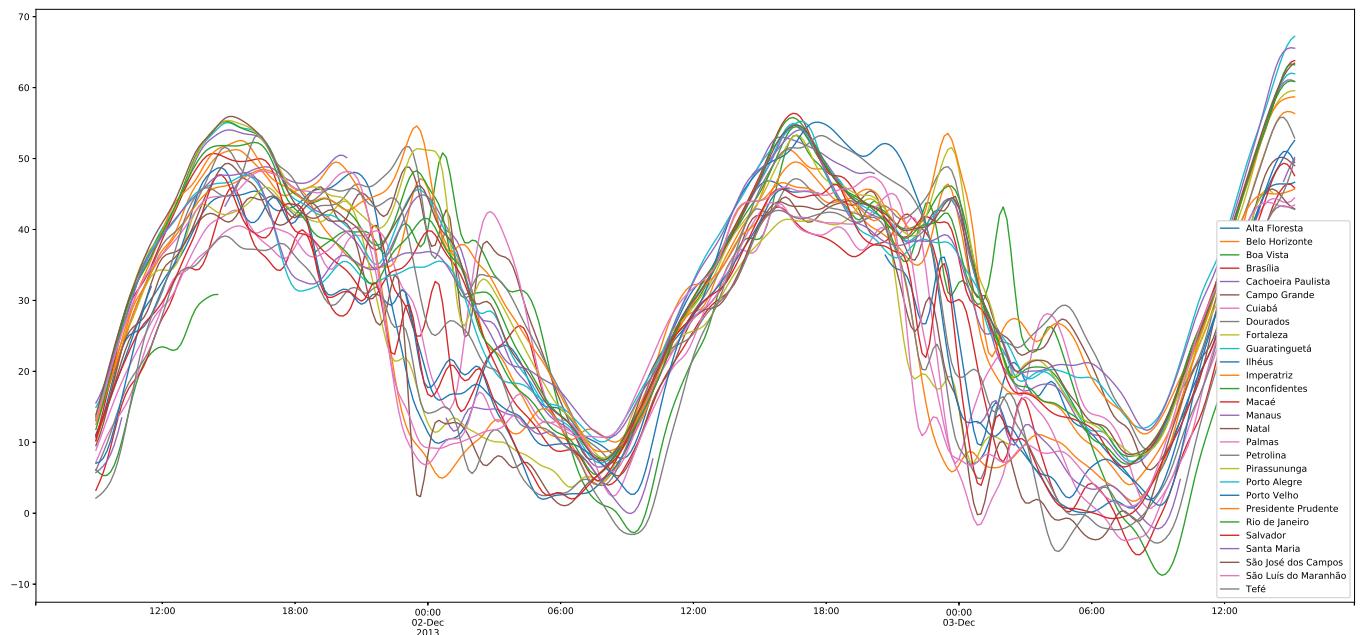


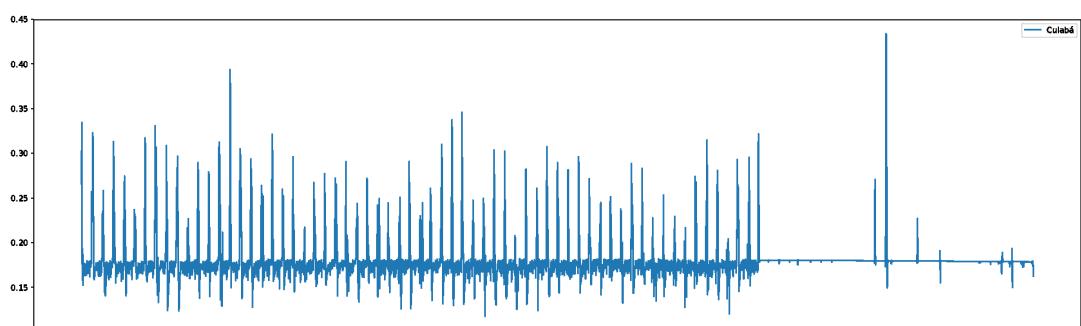
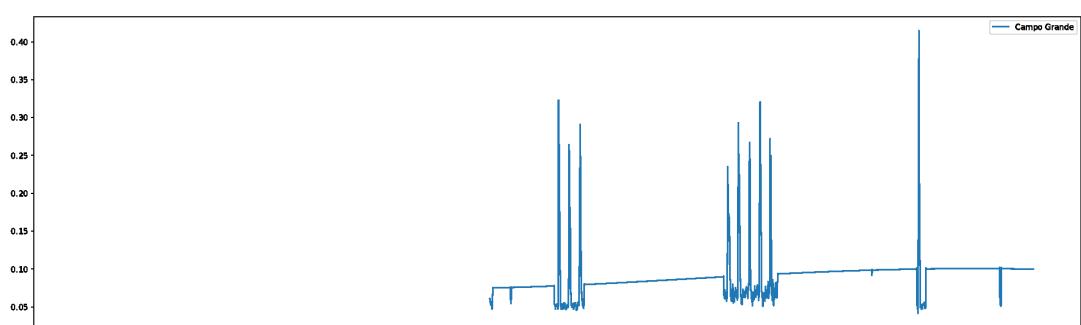
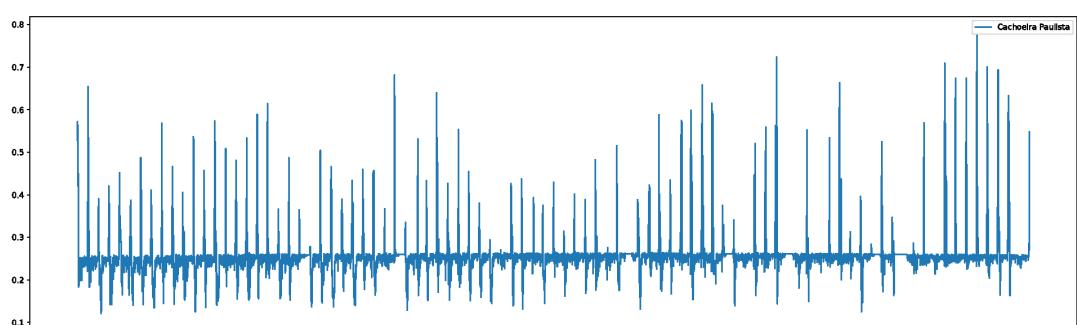
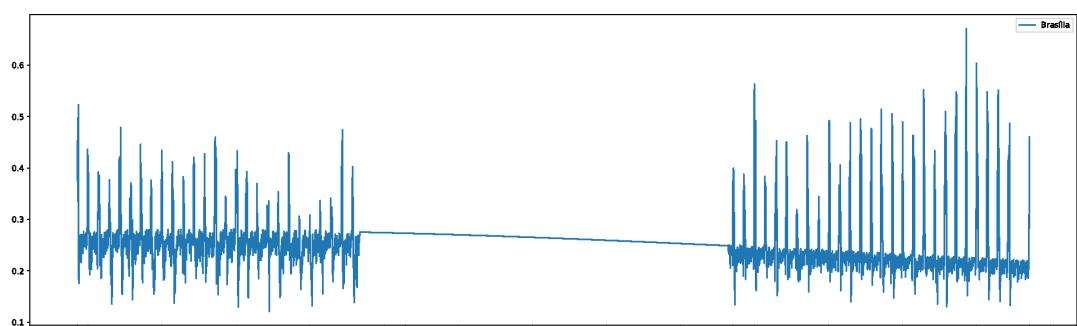
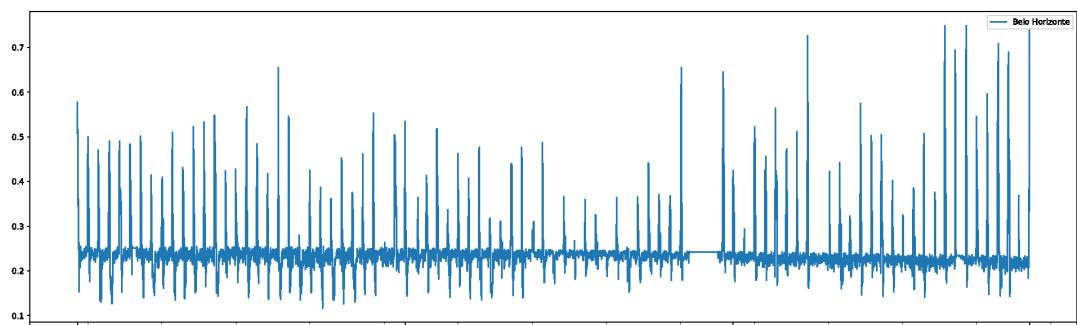
Figura 3.5 - A subfigure

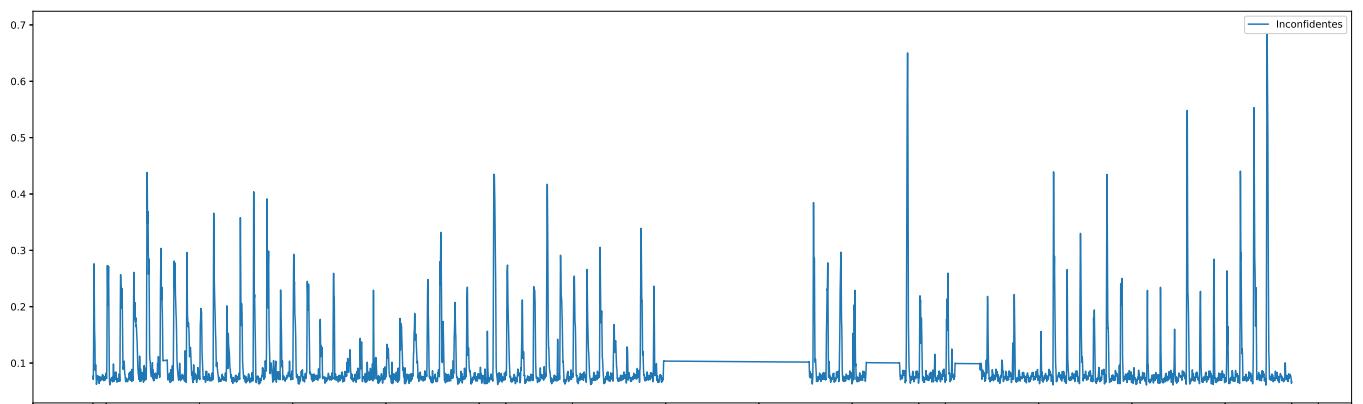
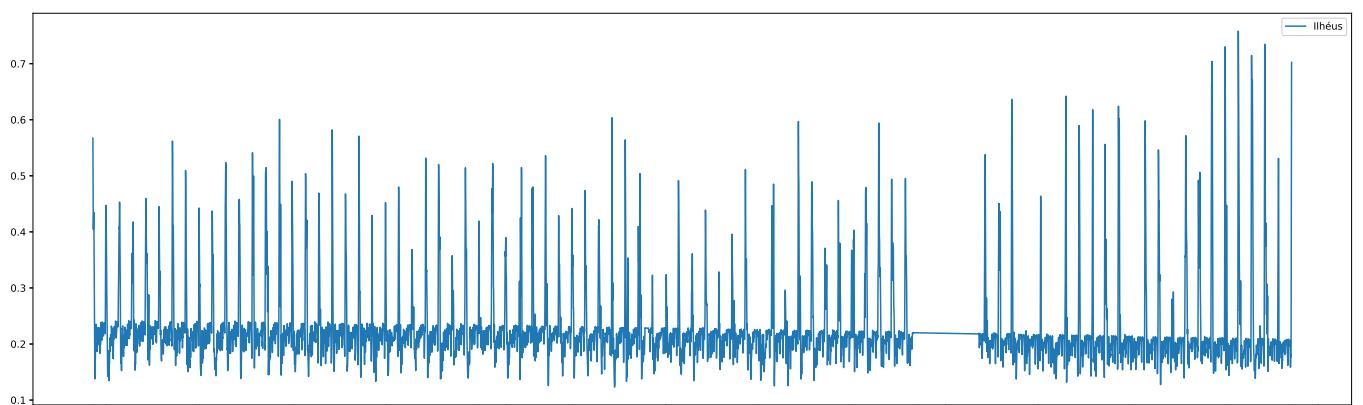
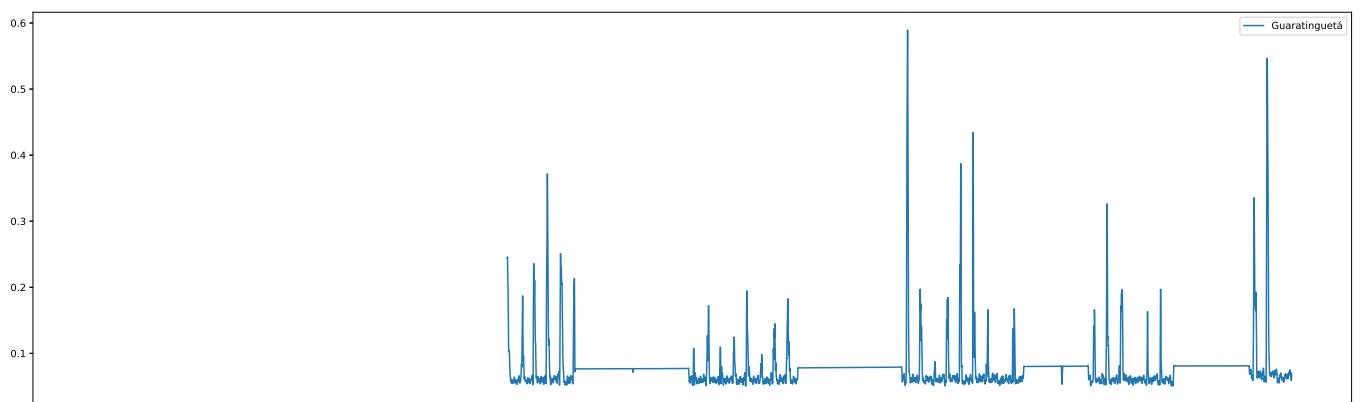
Em sua fase final este notebook gera uma tabela, onde cada coluna representa uma estação diferente com os dados de VTEC indexados pelo tempo, de modo que, pode-se falar em séries temporais para os dados de VTEC.

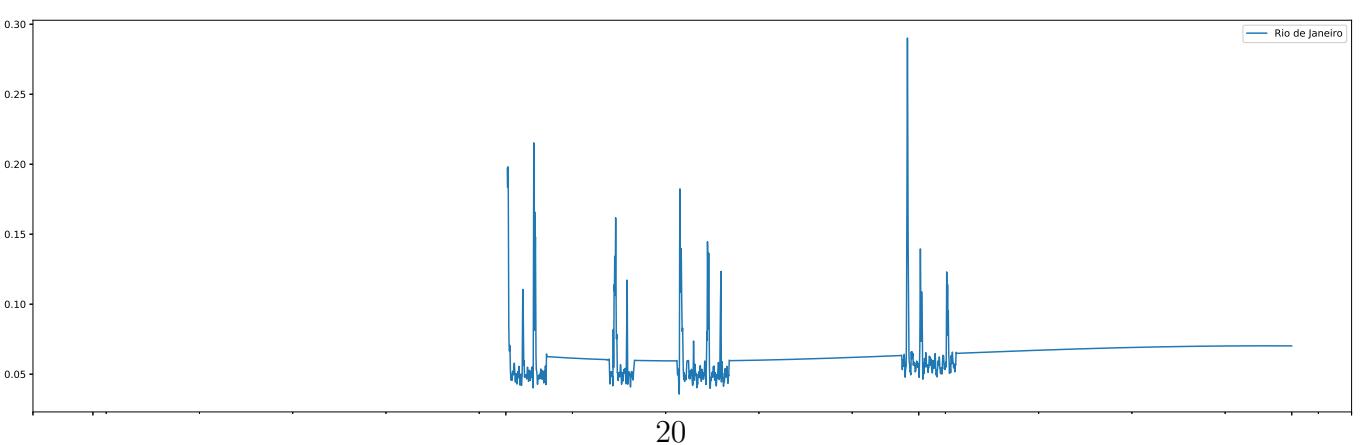
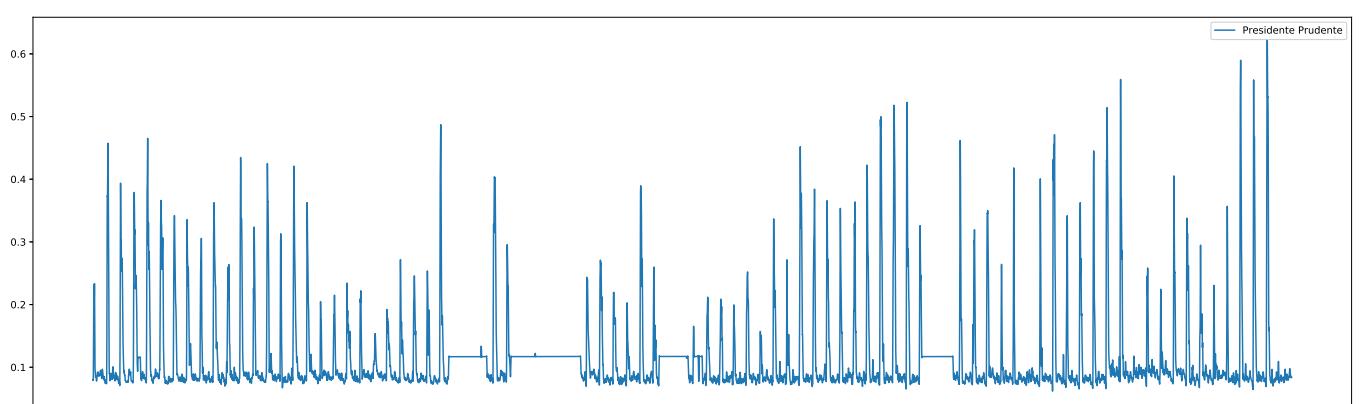
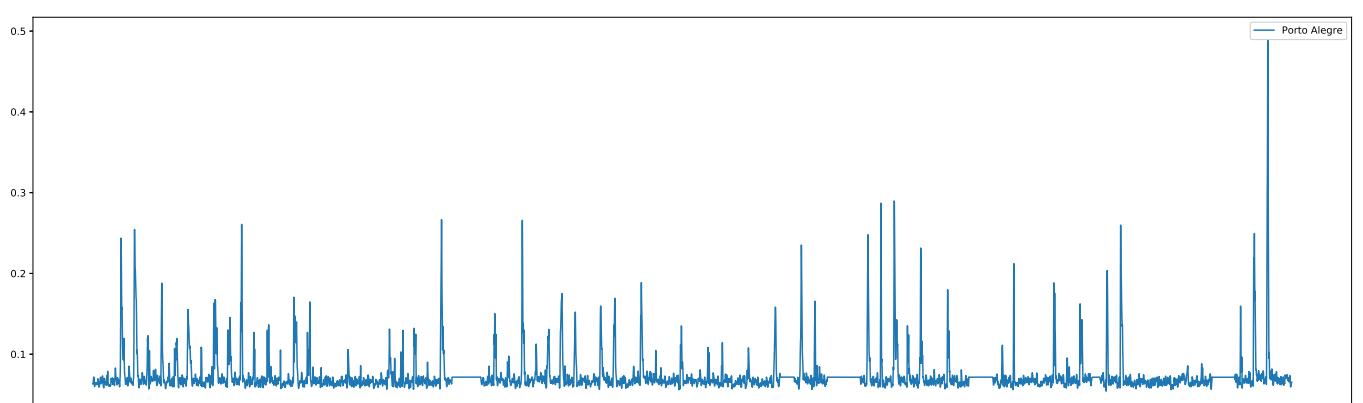
### 3.7 Seleção fina dos dados de S4

Após o pré-processamento dos dados de S4 é realizada uma seleção mais fina das estações que serão utilizadas neste trabalho, para tal foi empregado o notebook 04\_-reanalyze\_data.ipynb. O primeiro conjunto de estações descartas o foi, pois apresentava poucos pontos ao longo do período de janeiro de 2013 à dezembro de 2014 que levaram a curvas interpoladas que não são condizentes com a variável observada, o que foi evidenciado pela plotagem das séries temporais. Os dados foram amostrados em um período menor, assim, foi realizado um recorte na série temporal de S4, de maneira a ambos terem o mesmo período de amostragem. Assim, o segundo conjunto de estações descartadas são as que não mediram no período de tempo selecionado. A tabela 3.7 apresenta o grupo final de estações selecionados para o trabalho. As figuras 3.7, 3.7 e 3.7 exibem a série temporal S4 para este grupo de estações, enquanto a figura 3.7 apresenta um mapa com estas.

Cidade	Est.	Cód. de Id.	Alt.	Lat.	Lon.	Lat. Mag.	Lon. Mag.
Belo Horizonte	MG	bhz	858.000	-19.868500	-43.954200	-25.426147	24.786619
Brasília	DF	bsa	1050.000	-15.764200	-47.869400	-24.348659	22.352744
Cachoeira Paulista	SP	cpa	580.000	-22.410000	-45.000000	-24.456556	22.960540
Campo Grande	MS	32	Nan	-20.497000	-54.615000	-21.417704	14.873907
Cuiabá	MT	cub	278.000	-15.555200	-56.069800	-14.336068	14.530440
Dourados	MS	dou	756.120	-22.110000	-54.550000	-23.627266	14.698554
Fortaleza	CE	24	Nan	-3.742000	-38.539000	Nan	Nan
Guaratinguetá	SP	33	Nan	-22.789000	-45.220000	-24.188879	22.620120
Ilhéus	BA	ios	0.000	-14.470000	-39.100000	-13.470248	30.548727
Inconfidentes	MG	25	Nan	-22.318000	-46.329000	-26.299459	22.004117
Macaé	RJ	11	Nan	-22.823000	-41.785700	-20.542047	25.191448
Natal	RN	nta	0.000	-5.836162	-35.121000	Nan	Nan
Palmas	RO	3	Nan	-10.200000	-48.312000	-12.264838	23.425112
Pirassununga	SP	30	Nan	-21.989000	-47.334000	-23.990783	21.003125
Porto Alegre	RS	4	Nan	-30.071000	-51.119000	-22.954879	15.550843
Presidente Prudente	SP	6	Nan	-22.120000	-51.407000	-21.640946	17.249042
Rio de Janeiro	RJ	34	Nan	-22.823000	-43.238000	-20.105803	23.888647
Salvador	BA	26	Nan	-13.001000	-38.508000	-12.123350	31.680944
Santa Maria	RS	sta	110.100	-29.712591	-53.717206	-22.659740	13.628064
São José dos Campos	SP	sj2	593.440	-23.207000	-45.859000	-24.835610	22.002028
Tefé	AM	tfe	0.057	-3.180000	-64.440000	6.385157	9.314963
Cidade	Alt. da Cidade	Lat. da Cidade	Lon. da Cidade				
Belo Horizonte	767.0	-19.81570	-43.9542				
Brasília	1130.0	-15.78010	-47.9292				
Cachoeira Paulista	545.0	-22.67370	-44.9973				
Campo Grande	612.0	-20.44350	-54.6478				
Cuiabá	180.0	-15.59890	-56.0949				
Dourados	448.0	-22.22180	-54.8064				
Fortaleza	14.0	-3.71839	-38.5434				
Guaratinguetá	526.0	-22.81620	-45.1935				
Ilhéus	9.0	-14.79730	-39.0355				
Inconfidentes	864.0	-22.31710	-46.3284				
Macaé	7.0	-22.37170	-41.7857				
Natal	38.0	-5.79448	-35.2110				
Palmas	260.0	-10.16890	-48.3317				
Pirassununga	625.0	-21.99600	-47.4268				
Porto Alegre	22.0	-30.02770	-51.2287				
Presidente Prudente	471.0	-22.12760	-51.3856				
Rio de Janeiro	20.0	-22.90350	-43.2096				
Salvador	12.0	-12.97040	-38.5124				
Santa Maria	139.0	-29.69140	-53.8008				
São José dos Campos	593.0	-23.17910	-45.8872				
Tefé	28.0	-3.32073	-64.7236				







### 3.8 Visualizando S4 e VTEC

Feita uma seleção mais fina das estações que fornecerão as séries temporais de S4, assim, como um recorte adequado no tempo, é adequado plotar a série temporal do VTEC, juntamente com a do S4. Esta etapa é realizada no notebook 05\_visualize\_vtec\_s4\_data. Os gráficos são feitos em dois grupos distintos, o primeiro apresenta uma amostra das séries temporais, fornecendo um resolução visual melhor, enquanto o segundo fornece a série completa. Algumas observações pode ser feitas analisando visualmente ambos os conjuntos. Nota-se, por exemplo:

- uma periodicidade em ambos os dados;
- os valores de S4 sobem conforme o do VTEC diminui;
- os valores de pico de S4 aparecem em quedas e mínimos do VTEC;
- os dados de S4 apresentam maior ruído e menor disponibilidade;
- flutuações mais intensas do S4 aparecem no máximo do VTEC.

### 3.9 Análises: S4×VTEC em São José dos Campos

Existem várias formas de se buscar por padrões e correlações em um conjunto de dados, tal como a visualização por meio da plotagem de um gráfico representativo de uma série temporal. Assim, a seção 3.8 forneceu um indicativo da correlação em sua conclusão. Observado tal fato o próximo passo é buscar por um modelo que faça um mapeamento entre VTEC e S4. Este pode ter como conjunto de entrada o valor de VTEC e o de saída o S4, uma vez que ambas as variáveis são contínuas se busca realizar uma regressão.

Poder-se-ia buscar um mapeamento tal que a cada valor de VTEC seja associado um valor de S4 por estação, entretanto, usando os padrões observados juntamente das referências (), derivou-se um conjunto adicional de variáveis, que são a derivada temporal primeira e segunda do VTEC, a diferença do VTEC entre São José dos Campos e Pirassununga, a diferença do VTEC entre São José dos Campos e Brasília, assim como as derivadas temporais primeira de ambas as diferenças.

As derivadas temporais são interessantes, pois os valores de S4 aumentam conforme o valor de VTEC diminui, portanto existe uma variação no tempo que pode ser melhor extraída pela derivada temporal. As bolhas ionosféricas se deformam e propagam ao

longo de um meridiano magnético, as estações de Brasília, Pirassununga e São José dos Campos se encontram aproximadamente sobre o mesmo meridiano magnético, usando ambos os fatos considere que exista uma bolha em Brasília, e não em São José dos Campos, a diferença de VTEC terá um valor positivo, enquanto que se estiverem em ambas as cidades se tem um valor aproximadamente nulo, e com a bolha apenas em São José dos Campos um valor de diferença negativa, isto exibe claramente a propriedade da diferença espacial do VTEC em mapear a localização da bolha. A derivada temporal das diferenças espaciais fornece um indicativo da dinâmica da bolha.

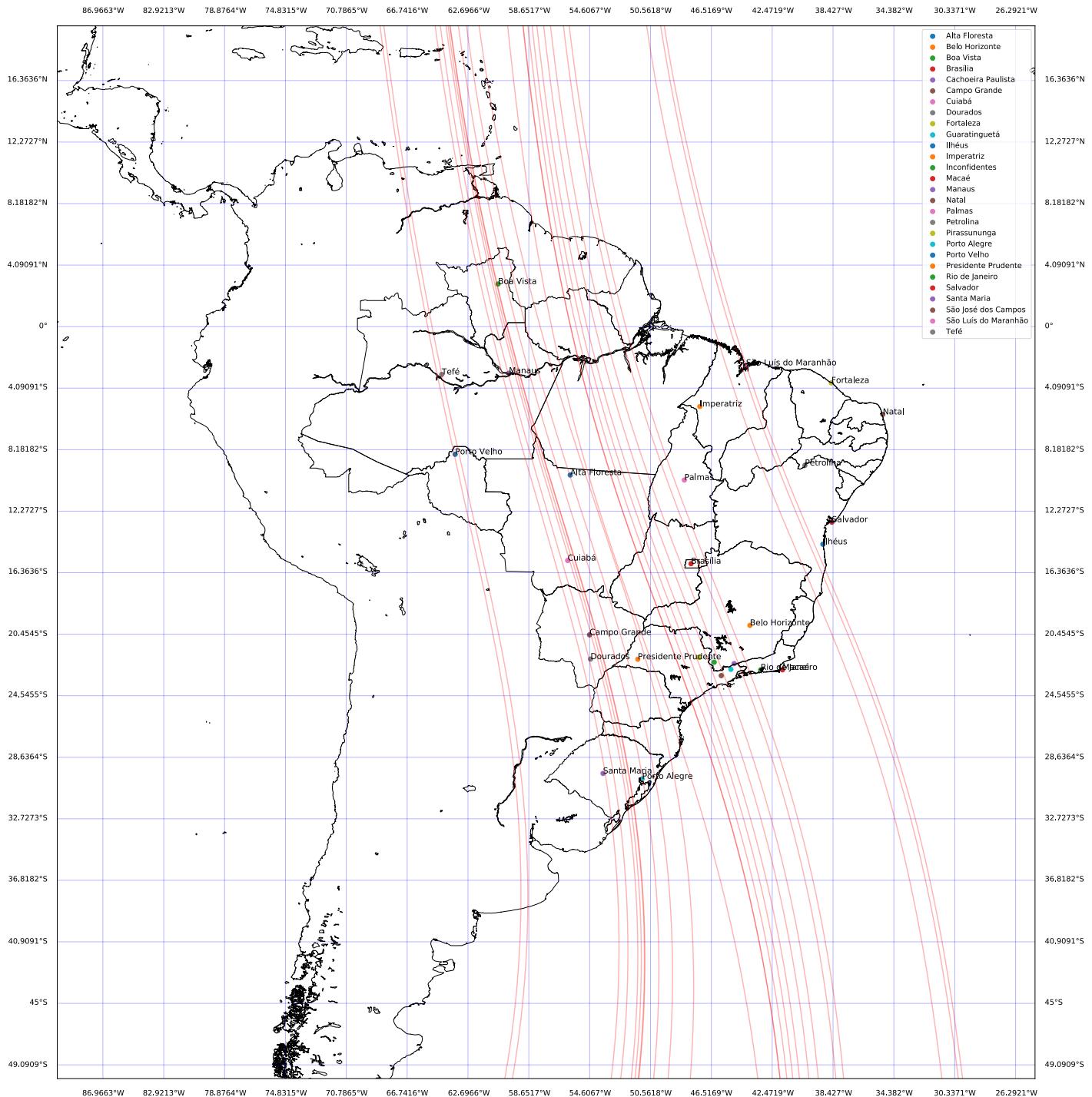
Sumarizando, ficou-se com o seguinte conjunto de variáveis/atributos, mais uma variável alvo S4:

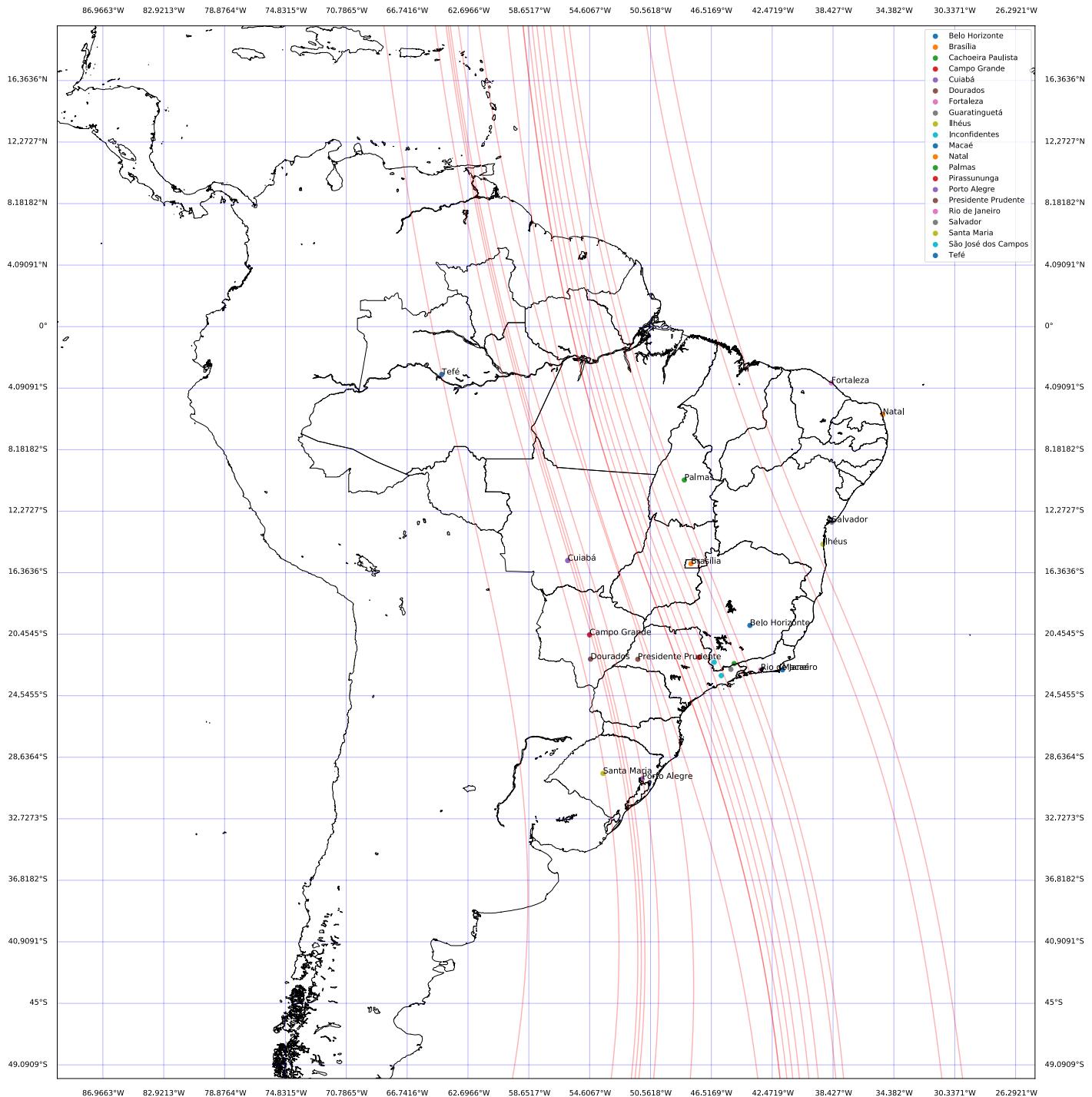
- **vtec** - conteúdo eletrônico total vertical;
- **vtec\_dt** - diferença finita de primeira ordem no tempo do VTEC, calculada por  $vtec_i - vtec_{i-1}$ ;
- **vtec\_dt2** - diferença finita de segunda ordem no tempo do VTEC, calculada por  $vtec_{i+1} - 2vtec_i + vtec_{i-1}$ ;
- **gvtec1** - diferença entre o VTEC de São José dos Campos e Pirassununga;
- **gvtec1\_dt** - diferença finita de primeira ordem no tempo do gvtec1, calculada por  $gvtec1_i - gvtec1_{i-1}$ ;
- **gvtec2** - diferença entre o VTEC de São José dos Campos e Brasília;
- **gvtec2\_dt** - diferença finita de primeira ordem no tempo do gvtec1, calculada por  $gvtec2_i - gvtec2_{i-1}$ ;
- **S4** - índice de cintilação ionosférico em São José dos Campos;

O papel do notebook 06\_analise\_sj2.ipynb é o de construir as variáveis definidas acima, assim como o de concatenar tais variáveis em uma tabela que possa ser utilizada em algoritmos de aprendizagem de máquina para desenvolvimento de um modelo.

Em posse desse conjunto de variáveis utilizou-se de um método de normalização para restringir o valor destas ao intervalo  $[0, 1]$  seguindo do treinamento de uma árvore de regressão, uma floresta randômica de regressão e uma máquina de suporte

vetorial também de regressão. Esses modelos foram desenvolvidos respectivamente nos notebooks 07\_analise\_sj2\_tree.ipynb, 07\_analise\_sj2\_random\_forest.ipynb e 07\_analise\_sj2\_svm.ipynb. Além disso, para todos os modelos seguiu-se com uma análise de sensibilidade aos atributos, removendo uma de cada vez e reavaliando as métricas do modelo gerado.







## REFERÊNCIAS BIBLIOGRÁFICAS

- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN 0123814790, 9780123814791. 3
- KIRCHHOFF, V. W. J. H. **Introdução à geofísica espacial**. [S.l.: s.n.], 1991. 152 p. ISBN 9788572330015. 8
- LAUNDAL, K. M.; RICHMOND, A. D. Magnetic coordinate systems. **Space Science Reviews**, v. 206, n. 1, p. 27–59, Mar 2017. ISSN 1572-9672. Disponível em: <<https://doi.org/10.1007/s11214-016-0275-y>>. 2
- LIMA, G. R. T. d.; STEPHANY, S.; PAULA, E. R. d.; BATISTA, I. S.; ABDU, M. A.; REZENDE, L. F. C.; AQUINO, M. G. S.; DUTRA, A. P. S. Correlation analysis between the occurrence of ionospheric scintillation at the magnetic equator and at the southern peak of the equatorial ionization anomaly. **Space Weather**, v. 12, n. 6, p. 406–416, June 2014. ISSN 1542-7390. 5
- LIMA, G. R. T. d.; STEPHANY, S.; PAULA, E. R. d.; BATISTA, I. S.; ABDU, M. A. Prediction of the level of ionospheric scintillation at equatorial latitudes in brazil using a neural network. **Space Weather**, v. 13, n. 8, p. 446–457, Aug 2015. ISSN 1542-7390. 5
- NEGRETI, P. M. S. **Estudo do conteúdo eletrônico total na região brasileira em períodos magneticamente perturbados**. 323 p. Sid.inpe.br/mtc-m19/2012/05.10.21.43-TDI. Tese (Doutorado em Geofísica Espacial/Ciências do Ambiente Solar-Terrestre) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19/2012/05.10.21.43>>. 7
- RAGHAVARAO, R.; NAGESWARARAO, M.; SASTRI, J. H.; VYAS, G. D.; SRIRAMARAO, M. Role of equatorial ionization anomaly in the initiation of equatorial spread f. **Journal of Geophysical Research: Space Physics**, v. 93, n. A6, p. 5959–5964, 1998. Disponível em: <<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JA093iA06p05959>>. 5
- RAGHUNATH SWAPNA ; VENKATA RATNAM, D. Detection of ionospheric spatial and temporal gradients for ground based augmentation system applications. **Indian Journal of Radio Space Physics (IJRSP)**, v. 45, n. 1, p. 11–19, 2016. 5

RAY, S.; PAUL, A.; DASGUPTA, A. Equatorial scintillations in relation to the development of ionization anomaly. **Annales Geophysicae**, European Geosciences Union, v. 24, n. 5, p. 1429–1442, jul. 2006. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00318073>>. 5

REZENDE, L. F. **Mineração de dados aplicada à análise e predição de cintilação ionosférica**. 176 p. (INPE-16080-TDI/1537). Dissertação (Mestrado em Compuatação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2009. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/06.22.15.52>>. 5

Rishbeth, H.; Garriott, O. K. **Introduction to ionospheric physics**. [S.l.: s.n.], 1969. 7

STAFFORD, J. V. Implementing precision agriculture in the 21st century. **Journal of Agricultural Engineering Research**, v. 76, n. 3, p. 267 – 275, 2000. ISSN 0021-8634. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0021863400905778>>. 2

TAKAHASHI, H.; WRASSE, C. M.; DENARDINI, C. M.; PÁDUA, M. B.; PAULA, E. R.; COSTA, S. M. A.; OTSUKA, Y.; SHIOKAWA, K.; MONICO, J. F. G.; IVO, A.; SANT'ANNA, N. Ionospheric tec weather map over south america. **Space Weather**, v. 14, n. 11, p. 937–949, 2016. Disponível em: <<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016SW001474>>. 2, 9

ZHU, X.; GOLDBERG, A. B. **Introduction to Semi-Supervised Learning: Morgan & Claypool**. 2009. 5