



Predição da ocorrência de cintilação ionosférica de amplitude na região do Maranhão

L. P. De Lemos Júnior¹, Otávio Carvalho², Alex Oliveira Barradas Filho³

¹ Universidade Federal do Maranhão, São Luís, MA, Brasil
Graduando em Engenharia Elétrica
Centro de Ciências Exatas e Tecnologia - CCET

² Universidade Federal do Maranhão, São Luís, MA, Brasil
Mestrando do programa de Pós-graduação em Engenharia Aeroespacial,
Rede Nordeste Aeroespacial, polo Maranhão

³ Universidade Federal do Maranhão, São Luís, MA, Brasil
Programa de Pós-graduação em Engenharia Aeroespacial,
Rede Nordeste Aeroespacial, polo Maranhão

luiz.lemos@discente.ufma.br

Resumo. Neste trabalho implementou-se a predição de Cintilação Ionosférica (CI) de amplitude, fenômeno responsável por erros de localização em sistemas de geoposicionamento por satélites. A predição é feita com o uso de técnicas de Aprendizado de Máquina largamente empregadas pela comunidade científica para automatizar a predição de CI. Seguindo essa tendência, implementou-se uma predição de Cintilação Ionosférica de amplitude com 5 minutos de antecedência com resultados de acurácia em torno de 90% por meio dos algoritmos Floresta Aleatória, Rede Neural Perceptron de Multicamadas e do banco de dados dos Projetos CALIBRA/CIGALA. Dentre esses dois modelos, avaliou-se qual obteve melhor desempenho a partir dos dados utilizados, ao realizar-se a comparação de ambos através de métricas de desempenho. Ao final do trabalho são sugeridas algumas ações que visem melhorar a abordagem aqui empregada.

Palavras-chave: Cintilação Ionosférica; Sistema GPS; Aprendizado de Máquina; Floresta Aleatória; Rede Neural Perceptron Multicamadas.

1. Introdução

A cintilação de uma onda eletromagnética é percebida pela flutuação de sua amplitude e fase à medida que esta se propaga pela atmosfera (WHEELON, 2004). O termo Cintilação Ionosférica, CI, é empregado quando esse fenômeno ocorre na ionosfera e afeta sinais de rádio, como os sinais do Sistema de Navegação Global por Satélite, GNSS (*Global Navigation Satellite System*). Irregularidades na densidade dos íons presentes na ionosfera são as responsáveis por esse fenômeno (CRANE, 1977). A cintilação de amplitude é mais comum em baixas latitudes do que em altas latitudes (AKALA et al., 2012). No Maranhão esse fenômeno é recorrente já que o estado está situado dentro da região equatorial.



A modelagem da Cintilação Ionosférica é uma tarefa complexa devido ao grande número de fatores que influenciam esse fenômeno, dos quais é possível citar as atividades solar e geomagnética, a latitude, o clima local e o tempo (PRIYADARSHI, 2020). Por essa razão, tem-se observado um número crescente de trabalhos científicos que fazem uso da aprendizagem de máquina como principal abordagem no estudo da Cintilação Ionosférica, por exemplo, (LINTY et al., 2019; JIAO; HALL; MORTON, 2017; LIU; MORTON; JIAO, 2018; MCGRANAGHAN et al., 2018; LIMA et al., 2015; CARVALHO; FILHO, 2020).

A Aprendizagem de Máquina, AM, é um subcampo da Inteligência Artificial e se baseia na utilização de algoritmos que identificam padrões e regras dentro de um conjunto de dados. Os algoritmos diferenciam-se quanto a forma como o modelo de aprendizado é criado. Existem, em termos gerais, três tipos de aprendizado: supervisionado, não supervisionado e aprendizado por reforço (CAMPESATO, 2020). A aprendizagem supervisionada refere-se a qualquer processo de aprendizagem de máquina que gere uma função por meio da compreensão de exemplos definidos por valores de entrada e saída. Os dois tipos mais comuns de aprendizagem supervisionada são classificação e regressão, cuja as variáveis de saída serão, respectivamente, categóricas e numéricas (SAMMUT; WEBB, 2011).

No trabalho de (CARVALHO; FILHO, 2020), analisa-se a viabilidade do uso do algoritmo Máquina de Vetores de Suporte para predição de CI de amplitude e fase. Os seus resultados são contextualizados com os de (LINTY et al., 2019), para concluir que a infraestrutura aeroespacial já instalada no Brasil possibilita o uso de Aprendizado de Máquina para realizar este tipo de predição.

O objetivo deste trabalho é processar os dados oriundos dos sinais enviados da constelação GPS para estação receptora localizada em São Luís e submetê-los a detecção de padrões através de dois algoritmos de AM: Floresta Aleatória (FA) e Rede Neural Perceptron Multicamadas (RN-MLP). Esta análise resultará numa predição de CI com antecedência de 5 minutos. Em seguida, cada modelo será avaliado e comparado através de métricas de avaliação de desempenho. Ao final, serão propostas abordagens futuras que complementem o trabalho atual.

2. Fundamentação Teórica

Nesta seção aborda-se inicialmente a definição matemática do parâmetro de CI chamado de índice S_4 , por meio de sua equação. Posteriormente, formula-se a teoria da FA, da RNMLP e das métricas de desempenho.

2.1. Índice S_4

As técnicas tradicionais de medição de cintilação ionosférica de amplitude baseiam-se no valor do índice S_4 (JIAO; HALL; MORTON, 2017) definido matematicamente como o desvio padrão da potência do sinal recebido normalizado pela média da potência do sinal recebido (BRIGGS; PARKIN, 1963).

$$S_4 = \frac{1}{R^2} \sqrt{(R^2 - \overline{R^2})}, \quad (1)$$

onde R é a amplitude do sinal, R^2 é a potência do sinal e $\overline{R^2}$ é a média da potência do sinal e $\sqrt{(R^2 - \overline{R^2})}$ é o desvio padrão da potência do sinal.

Os trabalhos que utilizam o índice S_4 como medida de classificação de cintilação ionosférica se diferenciam entre si a partir do limiar adotado para considerar a presença ou não do fenômeno



de cintilação. Este trabalho se baseia na abordagem tradicional e que considera a cintilação presente para valores acima de 0,12, como em (JIAO et al., 2013).

2.2. Rede Neural Perceptron Multicamadas

A Rede Neural Perceptron Multicamadas é um algoritmo de aprendizado supervisionado que foi proposto para gerar, através das informações fornecidas por um bando de dados, uma função não-linear (PEDREGOSA et al., 2011):

$$f(\cdot) : R_m \rightarrow R_o, \quad (2)$$

onde m é o número de dimensões da entrada, o é o número de dimensões da saída.

Dado um conjunto de treinamento $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)$ onde $\mathbf{x}_i \in R^n$ uma Rede Neural de Multicamadas Perceptron aprende a função, $f(x)$, onde $\mathbf{w}_1 \in R^m$ e $\mathbf{w}_2, b_1, b_2 \in R$ são os parâmetros do modelo:

$$f(x) = \mathbf{w}_2 g(\mathbf{w}_1^T x + b_1) + b_2, \quad (3)$$

$\mathbf{w}_1, \mathbf{w}_2$ representam, respectivamente, os pesos da camada de entrada e da camada oculta. b_1, b_2 representam, nessa ordem, o viés adicionado à camada oculta e à camada de saída;

Dado $g(\cdot) : R \rightarrow R$, a função de ativação de classificação binária é:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

A função de custo para classificação é a Entropia Cruzada, no caso de classificação binária, é definida da seguinte forma (HO; WOOKEY, 2020):

$$J = -\frac{1}{M} \sum_{m=1}^M [\mathbf{w} \times \mathbf{y}_m \times \log(h_{\theta}(\mathbf{x}_m)) + (1 - \mathbf{y}_m) \times \log(1 - h_{\theta}(\mathbf{x}_m))], \quad (5)$$

onde M é o número de exemplos treinados, w é o peso, y_m alvo para o exemplo de treinamento m é a regularização L2 que penaliza modelos complexos; $\alpha > 0$ controla a magnitude da penalização L2 (KINGMA; BA, 2015). A Rede Neural MLP inicia utilizando pesos aleatórios \mathbf{w} , e minimiza os erros da Função Custo por meio de atualizações recorrentes desses pesos. Após o cálculo dos erros, que neste caso é a Função Custo, o algoritmo envia esse erro da camada de saída para a camada anterior e fornece valores de pesos atualizados que minimizam o erro dessa camada. Esse processo se repete até que a camada inicial seja alcançada.

O gradiente descendente é representado matematicamente por:

$$\mathbf{w}^{i+1} = \mathbf{w}^i - \epsilon \nabla J_{\mathbf{w}}^i, \quad (6)$$

onde i é o passo de iteração e ϵ é a taxa de aprendizado, valor maior que 0. O algoritmo para quando um número pré definido de iterações é alcançado ou quando o erro alcançado é suficientemente pequeno.



2.3. Floresta Aleatória

A Floresta Aleatória é um classificador formado por um conjunto de Árvores de Decisão estruturadas $\{h(\mathbf{x}, \Theta_k), k = 1, \dots\}$, onde Θ_k é um vetor aleatório identicamente distribuído, criado de forma independente dos outros vetores em k . Cada Árvore de Decisão vota apenas uma vez para a escolha da classe mais popular da entrada \mathbf{x} . Detalha-se que para a k –ésima Árvore de Decisão o vetor Θ_k é gerado de forma independente dos vetores já produzidos $\Theta_1, \dots, \Theta_{k-1}$, porém esses vetores possuem algo em comum, suas distribuições. Então, uma Árvore de Decisão é gerada com base no banco de dados de treinamento e no vetor Θ_k . O resultado disso é um classificador representado matematicamente pela função $h(\mathbf{x}, \Theta_k)$, onde \mathbf{x} é um vetor de entrada. Dois métodos de aleatorização são utilizados na Floresta Aleatória: o *bagging* e a seleção por divisão aleatória. Estes dois métodos de aleatorização são utilizadas para diminuir a variância do estimador, já que as Árvores de Decisão naturalmente possuem alta variância e tendem a sobreajuste. Após a criação de um número grande de Árvores de Decisão, estas votam para a escolha da classe mais popular (BREIMAN, 2001). Este conjunto de procedimentos descrevem o funcionamento do que se denomina Floresta Aleatória. Neste trabalho a implementação de Floresta Aleatória usada combina classificadores selecionados a partir de suas probabilidades de predição corretas (PEDREGOSA et al., 2011), em contraste com a publicação original que escolhe a classe mais votada por cada classificador (BREIMAN, 2001).

2.4. Métricas de Desempenho

As métricas de desempenho foram obtidas a partir da criação da matriz de confusão. Destaca-se que a matriz de confusão em si não é uma métrica, porém ela fornece a base de entendimento para as métricas existentes. No caso de uma classificação binária, a matriz de confusão divide as predições em 4 tipos: verdadeiro positivo (VP): prediz a classe positiva como positiva; verdadeiros (VN): prediz a classe negativa como negativa; falso positivo (FP): prediz a classe negativa como positiva; falso negativo (FN): prediz a classe positiva como negativa (HOSSIN; SULAIMAN, 2015).

Em seguida utilizou-se os dados da matriz de confusão para o cálculo das seguintes métricas:

A acurácia revela quantas predições estão corretas.

$$ACC = \frac{VP + VN}{VP + VN + FN + FP} \quad (7)$$

A sensibilidade avalia a quantidade de valores verdadeiros corretamente classificados.

$$SENS = \frac{VP}{VP + FN} \quad (8)$$

A Especificidade avalia a capacidade do modelo em classificar com sucesso os resultados negativos.

$$ESP = \frac{VN}{VN + FP} \quad (9)$$

A taxa de falso positivo TFP, é a taxa global de FP

$$TFP = \frac{FP}{VN + FP} \quad (10)$$



A taxa de falso negativo TFN, é a taxa global de FN

$$TFN = \frac{FN}{VN + FN} \quad (11)$$

A Medida F1 representa a média harmônica entre TFP e TFN.

$$MF1 = \frac{2 \times TFP \times TFN}{TFP + TFN} \quad (12)$$

O Valor de AUC representa a área sobre a Curva ROC, quanto mais próximo de 1, melhor. A Curva ROC avalia o modelo de forma geral, valores acima da linha vermelha representa que o modelo está aprendendo e vice-versa. Defini-se S_p como a soma de todos os exemplos classificados como positivos, n_p e n_n denotam o número de exemplos positivos e negativos, respectivamente:

$$AUC = \frac{S_p - n_p(n_n + 1)/2}{n_p \times n_n} \quad (13)$$

3. Metodologia

Os códigos foram escritos utilizando a linguagem de programação Python e a biblioteca Scikit-Learn (PEDREGOSA et al., 2011). Os dados foram adquiridos com uso da ferramenta web ISMR Query Tool do projeto CALIBRA/CIGALA (VANI et al., 2014).

3.1. Aquisição e tratamento dos dados

Como foi proposto o estudo da Cintilação Ionosférica na região do Maranhão, os dados foram colhidos da estação SLMA localizada em São Luís - MA. Foram coletados dados dos meses de setembro, outubro, novembro e dezembro do ano de 2014. Esse período foi selecionado devido ao registro de uma maior incidência de cintilação. A base de dados original continha informações referentes a satélites de diversas constelações, no entanto, para esse trabalho considerou-se apenas aqueles da constelação GPS devido ao seu uso generalizado em diversas atividades. É importante ressaltar que os satélites de GPS geram sinais de baixa intensidade e, portanto, sinais mais suscetíveis à perda de informação por conta da cintilação ionosférica (KINTNER; LEDVINA; PAULA, 2007). Além disso, foram usados apenas os dados referentes à frequência L1 (1575,42 MHz). Amostras com valores ausentes foram descartadas da base de dados. Amostras cujo ângulo de elevação do satélite correspondente fosse menor ou igual a 30° foram descartadas para evitar o efeito do multicaminho do sinal que simula a ocorrência de cintilação por interferir no índice S_4 (DUBEY et al., 2005).

Uma vez selecionados os dados, criou-se um *dataframe* com uso das funções da biblioteca *pandas*, cujas colunas estão descritas na **Tabela 1**. A coluna S_4 possui valores adiantados de 5 minutos em relação às demais colunas. Isso significa, que em vez de uma linha do *dataframe* estar relacionada ao valor atual do índice S_4 ela está relacionada com o valor futuro do índice S_4 , pra ser mais específico, 5 minutos no futuro. Esta coluna, teve seus valores numéricos categorizados da seguinte forma: $S_4 < 0,12 = 0$ (não-cintilação, NC); $S_4 > 0,12 = 1$ (cintilação, C).

Após esta ação, verificou-se desbalanceamento de classes, a base de dados possuía mais amostras da classe NC do que da C. Essa questão foi contornada subamostrando a classe mais numerosa até que a quantidade de amostras das duas classes fosse igual, resultando em uma base de dados com **368854** amostras, com as duas classes representadas de forma equivalente. Em seguida, criou-se dois *dataframes* que foram divididos em dados de teste e treinamento.



Tabela 1. Parâmetros de Entrada dos Modelos

Parâmetro	Descrição
S_4	Índice de CI
azim	Ângulo Azimute
elev	Ângulo de Elevação
avg_cn0_l1	Média L1 C/N0
avgccd_l1	Média da divergência da onda portadora
sigmaccd_l1	Desvio padrão da divergência da onda portadora
tec_45	Conteúdo total de elétrons medido em um intervalo de 45 segundos
dtec_6045	TEC diferencial medido em um intervalo de 45 e 60 segundos
tec_30	Conteúdo total de elétrons medido em um intervalo de 30 segundos
dte_4530	TEC diferencial medido em um intervalo de 45 e 30 segundos
tec_15	Conteúdo total de elétrons medido em um intervalo de 15 segundos
dtec_3015	TEC diferencial medido em um intervalo de 30 e 15 segundos
tec	Conteúdo total de elétrons
dtec_15tow	TEC diferencial medido em um intervalo de 15 segundos
f2nd_tec_cn0	Média C/N0 da frequência secundária usada para calculo do TEC

3.2. Construção dos modelos preditivos e avaliação de desempenho

Ambos os algoritmos, FA e RNMLP, possuem cada um seu próprio método de aprendizado. As suas funções de aprendizagem foram obtidas pelas informações do conjunto de treinamento e avaliadas com o conjunto de testes. Por fim, verifica-se o desempenho dos modelos ao calcular suas métricas de desempenho obtidas pela comparação entre valores preditos e valores reais **Tabela 4**.

A seguir, são fornecidos os parâmetros de entrada dos algoritmos FA e RNMLP nas **Tabelas 2**, e **3**, respectivamente.

Tabela 2. Parâmetros de entrada utilizados para Floresta Aleatória

n_estimators=50	criterion='gini'	max_depth=None
min_samples_split=2	min_samples_leaf=1	min_weight_fraction_leaf=0.0
max_features='auto'	max_leaf_nodes=None	min_impurity_decrease=0.0
min_impurity_split=None	bootstrap=True	oob_score=False

Tabela 3. Parâmetros de entrada utilizados para Rede Neural MLP

hidden_layer_sizes=100	activation='relu'	solver='adam'
alpha=0.001	batch_size='auto'	learning_rate='constant',
learning_rate_init=0.001	power_t=0.5	max_iter=200
shuffle=True	random_state=None	tol=0.0001
verbose=False	warm_start=False	momentum=0.9
nesterovs_momentum=True	early_stopping=False	validation_fraction=0.1
beta_1=0.9	beta_2=0.999	epsilon=1e-08
n_iter_no_change=10	max_fun=15000	



4. Resultados e Discussão

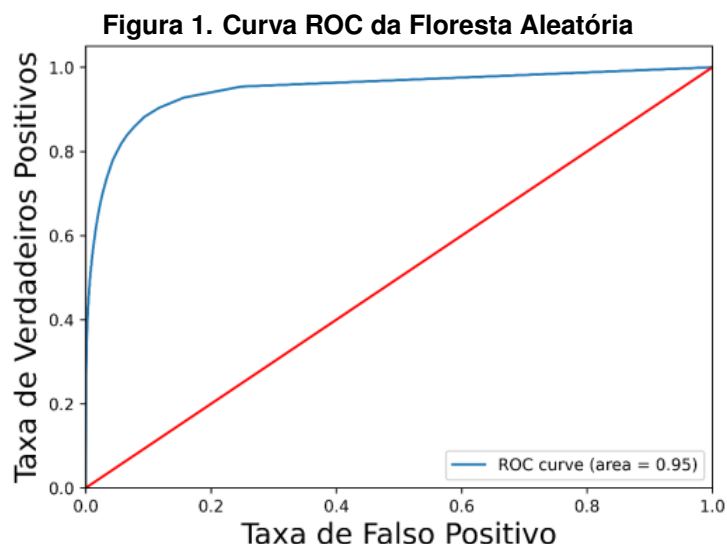
Na Tabela 3 são disponibilizadas as métricas de desempenho calculadas para os dois modelos. A acurácia, sensibilidade, especificidade, medida F1 e AUC de treino e teste são interpretadas da seguinte forma: quanto maior seu valor, melhor o desempenho. A TFP e TFN são avaliadas de modo oposto, quanto menor seus valores, melhor o desempenho. A acurácia, sensibilidade e especificidade da Floresta Aleatória foram, em torno de 6% a 7%, maiores do que as da Rede Neural MLP. A TFP, taxa de falso positivo, e a TFN, taxa de falso negativo, apresentaram valores menores para Floresta Aleatória, respectivamente, 5% e 7% menores, o que demonstra a superioridade da FA, também nestas duas métricas. A Medida F1 foi 7% maior para a Floresta Aleatória. Os AUC de treino e de teste da FA, foram, respectivamente 8% e 4% maiores do que as da Rede Neural MLP.

Tabela 4. Métricas de Avaliação dos Algoritmos

Algoritmo	ACC	SENS	ESP	TFP	TFN	Medida F1	AUC de treino	AUC de teste
Floresta Aleatória	0,9060	0,9267	0,8801	0,1198	0,0733	0,9060	0,9999	0,9626
Rede Neural MLP	0,8403	0,8514	0,8293	0,1707	0,1486	0,8417	0,9213	0,9218

4.1. Floresta Aleatória

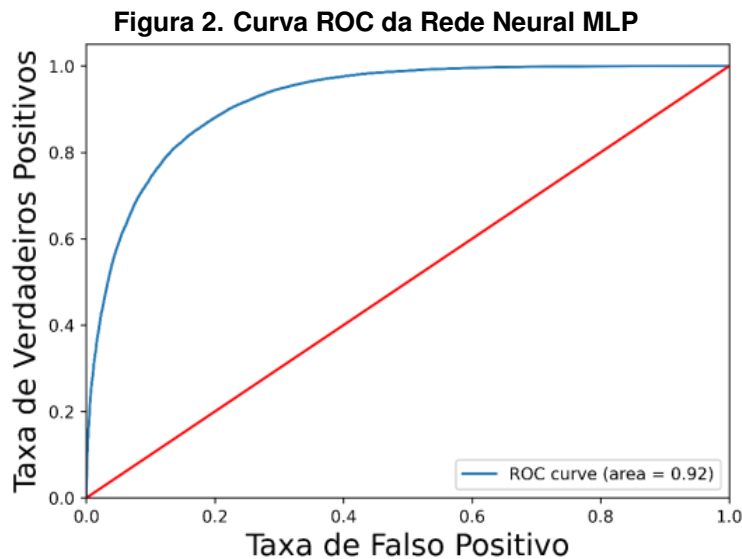
A Floresta Aleatória performou melhor para um número de estimadores igual a **50** e com os demais parâmetros inalterados. A FA apresentou, em todas as métricas, desempenho melhor do que da Rede Neural MLP. Como visto na Figura 1, sua área da curva ROC foi de **0,95**, muito próxima do valor ideal que é 1. Um AUC alto significa que a qualidade do valor de acurácia é satisfatória, conclusão que também pode ser obtida da análise da especificidade que é de **88%**. Ressalta-se que o modelo apresenta apenas uma taxa de **11%** de falsos alarmes.



4.2. Rede Neural Perceptron de Multicamadas

A rede neural performou melhor para um $\alpha = 0.001$, valor próximo ao recomendado em (KINGMA; BA, 2015). Os demais parâmetros não foram alterados. A sua Curva ROC é fornecida na Figura 2, e apresenta um performance 3% menor do que Curva ROC da FA.

Em (JIAO; HALL; MORTON, 2017), constata-se que os dados de cintilação ionosférica são linearmente separáveis apenas em um espaço multidimensional. Neste caso, não foi necessário



submeter os dados à transformações dimensionais. Dessa forma, é possível deduzir que a RN-MLP adaptou-se às não linearidades dos dados de cintilação.

Em relação as métricas, seus valores de acurácia, sensibilidade, especificidade, medida F1 e AUC foram, em média 7% menores do que da floresta aleatória. Já seu TFP e TFN foram maiores em torno de 7% em comparação a Floresta Aleatória.

O resultado obtido com a Rede Neural MLP demonstra que, apesar de sua menor performance em comparação a Floresta Aleatória, sua aplicação pode apresentar um resultados promissores desde que sejam feitos alguns ajustes no banco de dados com acréscimo de mais dados de entrada. Além disso, a complexidade da Rede Neural MLP implementada não foi muito alta, isso pode ser constatado por meio da análise dos seus parâmetros de entrada (Tabela 2). Ainda assim, sua acurácia de predição foi de 84%, 6% menor do que da FA.

5. Conclusão

Constantou-se por meio de análise das métricas de desempenho que algoritmo de melhor performance utilizado para lidar com estes dados foi a Floresta Aleatória. A robustês da natureza das Árvores de Decisões que compõem a FA, revela sua capacidade em lidar com dados que não sofreram tantos pré-processamentos ou que possuem lacunas de informações decorrentes da necessidade de escolha dos satélites que possuem ângulos de elevação maiores que 30°.

Acrescenta-se, por fim, a importância da qualidade do pré-tratamento dos dados. Algumas informações são perdidas ao se tentar evitar os efeitos do multicaminho. A escolha dos satélites que estejam a um ângulo de elevação maior do que 30° gera perda de informações que podem ser importantes para a melhora da performance dos algoritmos de Aprendizado de Máquina. Cerca de 35% a 45% dos dados são descartados por meio dessa prática, segundo (SPOGLI et al., 2014). Essa decisão faz parte do método tradicional de filtragem de interferência do multicaminho. Portanto, para trabalhos futuros se propõe a aplicação de uma filtragem dos dados de multicaminho que descarte o mínimo de informação sobre os efeitos de cintilação ionosférica nos sinais de rádio.

Agradecimentos: Os autores agradecem a Universidade Federal do Maranhão, ao financiamento do programa CAPES/PIBIC, ao INPE, ao PPG em Engenharia Aeroespacial, Rede



Nordeste Aeroespacial, polo Maranhão. Além disso, os autores agradecem pelo fornecimento dos dados provenientes da rede CIGALA/CALIBRA, projeto financiado pela Comissão Europeia (CE) no âmbito do FP7-GALILEO-2009-GSA e FP7-GALILEO-2011-GSA-1a, respectivamente, e Projeto FAPESP Número 06 / 04008-2.

Referências

- AKALA, A. O. et al. Impacts of ionospheric scintillations on gps receivers intended for equatorial aviation applications. *Radio Science*, v. 47, 2012. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2012RS004995>).
- BREIMAN, L. Random forests. *Machine Learning*, v. 45, p. 5–32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>).
- BRIGGS, B. H.; PARKIN, I. A. On the variation of radio star and satellite scintillations with zenith angle. *Journal of Atmospheric and Terrestrial Physics*, v. 25, p. 339–365, 1963. Disponível em: <https://ui.adsabs.harvard.edu/abs/1963JATP...25..339B/abstract>).
- CAMPESATO, O. *Artificial Intelligence, Machine Learning and Deep Learning*. 1. ed. [S.l.]: Mercury Learning and Information, 2020. ISBN 13: 978-1683924678.
- CARVALHO, O.; FILHO, A. B. O. Aplicação de svm para a previsão de ocorrência de cintilação ionosférica. *11° Workshop de Engenharia e Tecnologia Espaciais*, 2020. Disponível em: <http://mtc-m21c.sid.inpe.br/rep/8JMKD3MGP3W34R/435F4LB>).
- CRANE, R. K. Ionospheric scintillations. *IEEE, Proceedings, Army-supported research*, v. 65, p. 180–199, 1977. Disponível em: <https://ui.adsabs.harvard.edu/abs/1977IEEEP..65..180C/abstract>).
- DUBEY, S. et al. Study of amplitude and phase scintillation at gps frequency. *Indian Journal of Radio & Space Physics*, v. 34, p. 402–407, 2005. Disponível em: <http://nopr.niscair.res.in/bitstream/123456789/25675/1/IJRSP%2034%286%29%20402-407.pdf>).
- HO, Y.; WOOKEY, S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access*, v. 8, 2020. Disponível em: <https://ieeexplore.ieee.org/document/8943952>).
- HOSSIN, M.; SULAIMAN, M. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining Knowledge Management Process (IJDKP)*, v. 5, 2015. Disponível em: <https://www.airconline.com/ijdkp/V5N2/5215ijdkp01.pdf>).
- JIAO, Y.; HALL, J. J.; MORTON, Y. T. Automatic equatorial gps amplitude scintillation detection using a machine learning algorithm. *IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS*, v. 53, p. 405 – 418, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7812651>).
- JIAO, Y. et al. Automatic equatorial gps amplitude scintillation detection using a machine learning algorithm. *Radio Science*, v. 48, p. 698–708, 2013. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2013RS005259>).
- KINGMA, D. P.; BA, J. *Published as a conference paper at the 3rd International Conference for Learning Representations*, 2015. Disponível em: <https://arxiv.org/abs/1412.6980>).
- KINTNER, P. M.; LEDVINA, B. M.; PAULA, E. R. de. Gps and ionospheric scintillations. *Space Weather*, v. 5, 2007. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2006SW000260>).
- LIMA, G. R. T. de et al. Prediction of the level of ionospheric scintillation at equatorial latitudes in brazil using a neural network. *Space Weather*, v. 13, p. 446–457, 2015. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018SW002018>).
- LINTY, N. et al. Detection of gnss ionospheric scintillations based on machine learning decision tree. *IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS*, v. 55, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8398529>).



LIU, Y.; MORTON, Y. J.; JIAO, Y. Application of machine learning to the characterization of gps II ionospheric amplitude scintillation. *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2018. ISSN 17823216. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8373500>.

MCGRANAGHAN, R. M. et al. New capabilities for prediction of high-latitude ionospheric scintillation: A novel approach with machine learning. *Space Weather*, v. 16, p. 1817–1846, 2018. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018SW002018>.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PRIYADARSHI, S. Ionospheric scintillation modeling needs and tricks. *IntechOpen*, 2020. Disponível em: <https://www.intechopen.com/chapters/69333>.

SAMMUT, C.; WEBB, G. Supervised learning. *Encyclopedia of Machine Learning*. Springer, Boston, MA, 2011. Disponível em: https://doi.org/10.1007/978-0-387-30164-8_803.

SPOGLI, L. et al. A filtering method developed to improve gnss receiver data quality in the calibra project, mitigation of ionospheric threats to gnss: an appraisal of the scientific and technological outputs of the transmit project. *IntechOpen*, 2014. Disponível em: <https://www.intechopen.com/chapters/47138>.

VANI, B. et al. Visualizing the quality of gnss multivariate data. *Proceedings of the Brazilian Symposium on GeoInformatics*, p. 95–106, 2014. Disponível em: <http://hdl.handle.net/11449/171743>.

WHEELON, A. D. *Electromagnetic Scintillation 1. Geometrical Optics*. 1. ed. [S.l.]: Cambridge University Press, 2004. ISBN 0-511-03246-3.