MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

RECONSTRUÇÃO DO CONTEÚDO ELETRÔNICO TOTAL DA IONOSFERA UTILIZANDO ANÁLISE ESPECTRAL DE DADOS HISTÓRICOS

João Vítor Bernardi Rohr

Relatório de Iniciação Científica do programa PIBITI, orientada pelo Dr. Adriano Petry.

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

INPE
São José dos Campos
2022

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

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

RECONSTRUÇÃO DO CONTEÚDO ELETRÔNICO TOTAL DA IONOSFERA UTILIZANDO ANÁLISE ESPECTRAL DE DADOS HISTÓRICOS

João Vítor Bernardi Rohr

Relatório de Iniciação Científica do programa PIBITI, orientada pelo Dr. Adriano Petry.

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

INPE
São José dos Campos
2022

Dados Internacionais de Catalogação na Publicação (CIP)

Sobrenome, Nomes.

Cutter Reconstrução do Conteúdo Eletrônico Total da Ionosfera utilizando Análise Espectral de Dados Históricos / Nome Completo do Autor1; Nome Completo do Autor2. – São José dos Campos : INPE, 2022.

xv + 34 p. ; ()

Dissertação ou Tese (Mestrado ou Doutorado em Nome do Curso) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, AAAA.

Orientador : José da Silva.

1. Palavra chave. 2. Palavra chave 3. Palavra chave. 4. Palavra chave. 5. Palavra chave I. Título.

CDU 000.000

ATENÇÃO! A FOLHA DE APROVAÇÃO SERÁ INCLUIDA POSTERIORMENTE.
Mestrado ou Doutorado em Nome do Curso

RESUMO

A descrição analítica de modelos para o Conteúdo Eletrônico Total (TEC) da ionosfera apresenta alta complexidade e elevado custo computacional. Sendo assim, com a grande quantidade de dados já existentes é de interesse a utilização de métodos de aprendizado de máquina e ciência de dados para a mais rápida predição do TEC da ionosfera através de indicadores de atividade solar. Dando continuidade ao trabalho anterior, neste também se buscou a modelagem da dinâmica da ionosfera durante longos períodos (1 a 3 anos) fazendo uso da análise espectral através da Transformada Discreta do Cosseno (DCT). As variáveis solares analisadas como features do modelo foram o número relativo de manchas solares (RSN), índice F10.7 e 39 bandas de fluxo fotônico (PF) de extremo ultravioleta (EUV) obtidos através do modelo empírico Solar2000. A fim de condensar as 39 bandas de fluxo fotônico em apenas uma variável foi proposta uma nova feature baseada na média ponderada pelos coeficientes de correlação de Pearson de cada uma das bandas, nomeada por simplicidade como PF combinado. Para o presente trabalho foram considerados apenas os modelos de regressão linear e máquina de vetores de suporte (SVM). Além disso, o conjunto de dados foi aumentado para 19 anos (2003-2021) tanto em valores de TEC como de dados solares o que permitiu a análise de desempenho com o aumento dos anos de teste e maior tempo de treino. Por fim, desenvolveu-se um código geral em que é possível se subdividir a simulação em diferentes modelos individuais separados por um período de dias do ano, sazonal por exemplo, pelos solstícios e equinócios. Analisando o desempenho de modelos treinados com certas combinações de features ficou notável que aqueles com F10.7 e algumas bandas separadas de PF desempenharam muito abaixo do que aqueles treinados somente com RSN e/ou PF combinado, os quais atingiram erros de 2.8 TECu, quando comparados com dados de TEC fornecidos pelo International GNSS Service (IGS). De outro modo, com variáveis de F10.7, RSN e PF das três primeiras bandas, mas agora com divisão sazonal, o valor de RMSE ficou em torno de 2 TECu para todo o período de teste. Todavia, observou-se que a utilização apenas de RSN e/ou PF combinado não obtém benefícios significativos da divisão sazonal, pois com estas features o comportamento periódico sazonal é suprimido.

Palavras-chave: Conteúdo Eletrônico Total. Previsão. Aprendizado de Máquina. Modelagem Sazonal. Ionosfera.

RECONSTRUCTION OF THE TOTAL ELECTRON CONTENT OF THE IONOSPHERE USING SPECTRAL ANALYSIS OF HISTORICAL DATA

ABSTRACT

The analytical description of models for the ionospheric Total Electron Content (TEC) shows high complexity and computational cost. Therefore, with the large amount od data already available it is of interest the application of machine learning and data science methods to forecast the ionospheric TEC more rapidly through solar activity proxies. In that way, it was done the dynamical modeling of the ionosphere during long periods of time (1 to 3 years) using spectral analysis by means of the Discrete Cosine Transform (DCT). The variables analyzed as the model's features were the Relative Sunspot Number (RSN), the F10.7 index and 39 bandwidths of Photon Fluxes (PF) in the Extreme Ultra-Violet (EUV) spectrum, all of those obtained from the empirical model Solar2000. Aiming to condensate the 39 bandwidths of photon flux into only one variable it was proposed a new feature based on the weighted average where the weights are the Pearson coefficient with respect to the TEC for each bandwidth, this variable was named combined PF. The models considered were linear regression and Support Vector Machine (SVM), based on previous results. Moreover, the data base on this analysis constituted 19 years (2003-2021) of TEC and solar data, for which performance over long periods of time changing training and testing. Based on previous observed evidence of error seasonal variation, it was tested the seasonal subdivision of the model to lower the Root Mean Square Error (RMSE) during solstices. Evaluating the RMSE error of models trained and tested with different sets of features it was evident that those with F10.7 e some bands of PF performed well lower than those trained only with RSN and/or combined PF, who reached errors of 2.8 TECu, when compared with TEC data obtained from the International GNSS Service (IGS). Otherwise, considering F10.7, RSN and the three first bandwidths of PF, but now with seasonal division, the value of RMSE stayed in the range of 2 TECu for the whole test period. Nevertheless, using only RSN and/or combined PF it was not observed significant benefits from seasonal division, the reason for that being the suppression of the error seasonal behavior when applying those features.

Keywords: Total Electron Content. Forecasting. Machine Learning. Seasonal Modeling. Ionosphere.

LISTA DE FIGURAS

Pág.

LISTA DE TABELAS

Pág.

LISTA DE ABREVIATURAS E SIGLAS

TEC	–	Conteúdo Eletrônico Total
SC	–	Ciclo Solar
SSN	–	Número de Manchas Solares
SIDC	–	Centro de Análises de Dados de Influência Solar
SI	–	Sistema Internacional de Unidades
SFU	–	Unidade de Fluxo Solar
PF	–	Fluxo Fotônico

SUMÁRIO

Pág.

1 INTRODUÇÃO

A Ionosfera contém apenas 1% de toda a massa da atmosfera, entretanto, tem uma importância muito maior para as telecomunicações, (HANSLMEIER, 2006). Na ionosfera existem partículas carregadas (íons) as quais influenciarão ondas de rádio passantes. A ionização destas partículas depende de radiação solar, no caso as radiações com energia suficiente para isto estão no espectro do extremo ultra violeta e dos raios X. Por conta destes íons, esta camada da atmosfera acaba sendo um condutor de eletricidade e ondas de rádio.

A ionosfera pode ser dividida em três sub-regiões. A primeira, chamada D, vai de 50-90 km e tem a menor ionização. A segunda, chamada E, vai de 90-150 km e tem como principais íons o O_2^+ e NO^+. A última camada é a F, que pode ainda ser dividida em F1 e F2. A camada F1 vai de 150-400 km e a camada F2 vai de 400-1000 km, entretanto não há uma definição exata entre o plasma da ionosfera e os limites do campo magnético da Terra, (HANSLMEIER, 2006).

Na camada F2 se tem a maior concentração de elétrons, sendo assim, a mais importante. Os picos de concentração de elétrons ocorrem nos polos para altas altitudes e no equador magnético para baixas latitudes. As duas principais causas de variação da concentração de elétrons na ionosfera se deve às fontes de ionização que variam, no caso o Sol e as auroras, e mudanças na parte neutra da termosfera, (HANSLMEIER, 2006).

Em sistemas de comunicação HF (*High Frequency*), que se baseiam na reflexão de ondas de rádio nas camadas da ionosfera, as frequências máximas e mínimas utilizáveis são definidas pela densidade de elétrons presentes na ionosfera. Um dos sistemas mais afetados por estes íons são os de GPS (*Global Positioning System*), que devem se comunicar com satélites em órbitas de 22000 km de altitude, portanto, é necessário aplicar uma correção nos valores de modo a obter a posição com menor erro.

A medida da quantidade de elétrons na ionosfera se dá através do Conteúdo Eletrônico Total (sigla em inglês TEC). O TEC, é uma medida da quantidade de elétrons integrados entre dois pontos do espaço, portanto o TEC depende da direção em que é medido. O TEC mais utilizado é o diretamente vertical, TECv que será utilizado neste trabalho. O TEC é diretamente proporcional ao efeito de delay de propagação da ionosfera sobre ondas de rádio, o que o torna uma ferramenta útil para a correção destes efeitos.

1.1 Objetivos

Tendo em vista a importância do TEC, é de objetivo deste trabalho prever seu valor utilizando dados de atividade solar em modelos de aprendizado de máquina. Buscou-se então modelar a dinâmica do TEC ionosférico através de técnicas de análise espectral. Tendo em vista o que foi desenvolvido anteriormente na pesquisa foi proposto os seguintes objetivos:

- Aumentar a escala das análises o máximo possível, ou seja, aumentar o número de dados;
- Tendo mais dados, é possível estudar o desempenho dos modelos com mudanças nos conjuntos de treino e teste;
- Avaliar qualitativamente as *features* do modelo;
- Explorar a dependência sazonal do erro através da divisão em submodelos.

Proposição: Utilização de dados de TEC do IGS, com espaçamento temporal de 2 horas e espacial de 2,5°x5,0° em latitude e longitude respectivamente; Aplicação da Transformada Discreta do Cosseno (DCT) para a exclusão do domínio temporal e a criação de 12 frequências; Aplicação de modelos de aprendizado de máquina relacionando cada frequência de cada ponto espacial separadamente com features de atividade solar. Métrica de performance escolhida RMSE (Root Mean-Square Error).

1.2 Definições e Visão Geral

Como dados de referência para o treino e teste dos modelos utilizou-se os valores de TEC disponíveis pelo IGS (*International GNSS Service*). Estes valores estão espaçados espacialmente em 2,5° x 5,0° em latitude e longitude respectivamente, formando um grid global. Além disso, os valores tem dimensão temporal também, espaçada de 2 horas formando 13 conjuntos de dados espaciais de TEC por dia. Um mapa TEC gerado com dados do IGS pode ser visto na Figura 1.1.

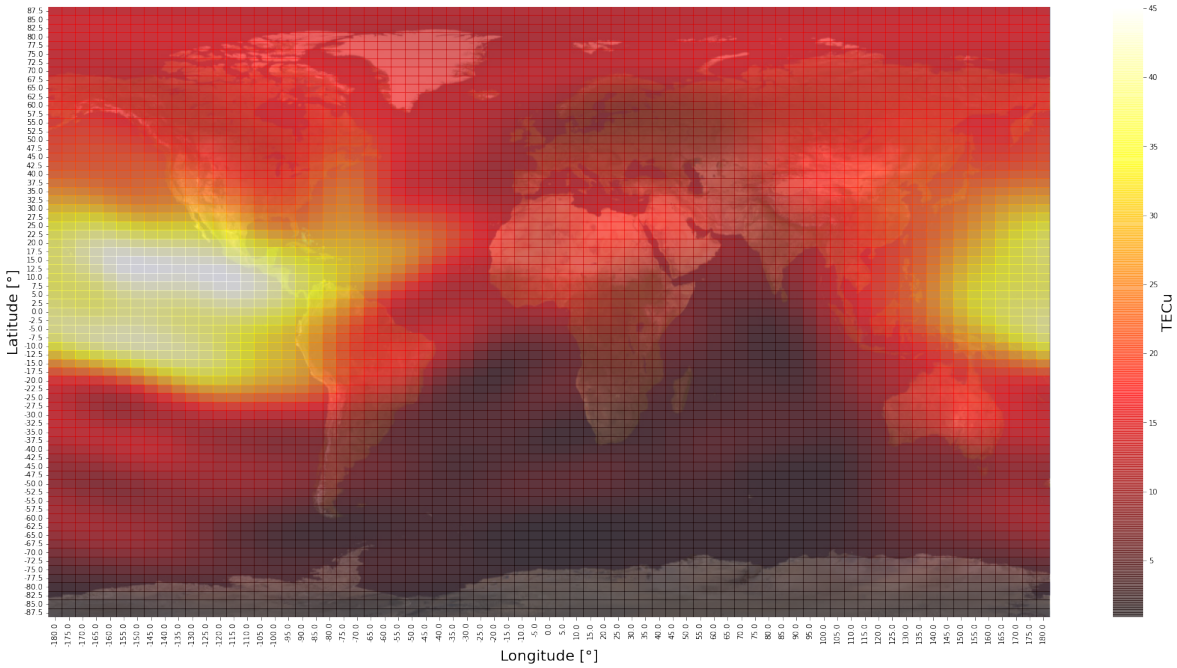

Figura 1.1 - Mapa de TEC para o dia 200 de 2009 às 08:00 UTC.

As features presentes são o número relativo de manchas solares (RSN), o índice F10.7 de ondas de rádio e 39 bandas de fluxo fotônico no espectro do extremo ultra violeta. As features foram obtidas do modelo semi-empírico Solar2000, o qual dá como saída seus valores diários para todo o planeta. O conjunto de anos de análise foi estendido para este trabalho, indo dos anteriores 5 anos (2015-2019) para 19 anos (2003-2021), sendo um limite imposto pela qualidade dos dados apresentados pelo IGS, pois anteriormente a 2003 o padrão de medida temporal muda e existem valores impossíveis.

O procedimento para modelagem começa com a aplicação da Transformada Discreta do Cosseno (DCT) para a exclusão do domínio temporal e a criação de 12 frequências. Em seguida, uma vez que não há mais a dimensão temporal e os dados de TEC podem ser correlacionados com os dados solares para cada frequência de cada ponto no grid espacial, é aplicado um modelo de aprendizado de máquina que pode ser variado. Este modelo terá diferentes valores de treino e teste. A métrica escolhida para medir o desempenho de cada modelo foi o RMSE (Root Mean-Square Error).

2 DESENVOLVIMENTO

2.1 Features

Todas as features aqui presentes foram obtidas do modelo empírico de irradiância solar SOLAR2000 (TOBISKAA et al., 2000).

2.1.1 Número de Manchas Solares

As manchas solares são fenômenos temporários que ocorrem na fotosfera sendo regiões relativamente mais escuras que apresentam fortes campos magnéticos (0.4T) e temperaturas mais baixas que a temperatura média da superfície do Sol. As manchas solares são formadas por regiões nomeadas umbra e penumbra, sendo a primeira a região escura central e a segunda regiões adjacentes em formato de filamento menos escuras (HANSLMEIER, 2006). Não há uma explicação completa e detalhada sobre a formação das manchas solares, entretanto, é amplamente aceito que estas manchas são resultado de tubos de fluxo magnético na zona de convecção do Sol. A diminuição da temperatura nessas zonas se deve à inibição de convecção na fotosfera de modo que a temperatura na superfície acabe sendo menor que nas redondezas não afetadas pelo campo magnético (CHOUDHURI, 2015).

A ocorrência de manchas solares pode se dar de forma única ou em grupos e se concentra em latitudes baixas (5°-30°), perto do Equador Solar. O tempo de existência destas manchas, entretanto, é curto indo de alguns dias a algumas rotações solares (CANDER, 2019). A atividade das manchas solares segue um ciclo oscilatório de 11 anos chamado de ciclo de manchas solares ou Ciclo Solar (SC). Além disso, a mensuração das manchas se dá através do número relativo de manchas solares (RSN), descrito primeiramente por Wolf (1851).

Este número relativo de manchas solares é dado então por,

$$RSN = k(10g + f) \tag{2.1}$$

onde k é um fator que depende na atmosfera local média, parâmetros ambientais e instrumentais do observatório que realiza a medida, este fator é determinado pelo Centro de Análises de Dados de Influência Solar (SIDC) (ROYAL OBSERVATORY OF BELGIUM, 2022), g é o número de grupos e f o número de manchas individuais.

Os RSN foram obtidos para o período de 2003-2021 espaçados diariamente. Este período de tempo compreende a segunda metade do SC 23, todo o SC 24 e breve

começo do SC 25. Uma visão geral dos dados pode ser vista na Figura 2.1. Percebe-se claramente a relação cíclica.

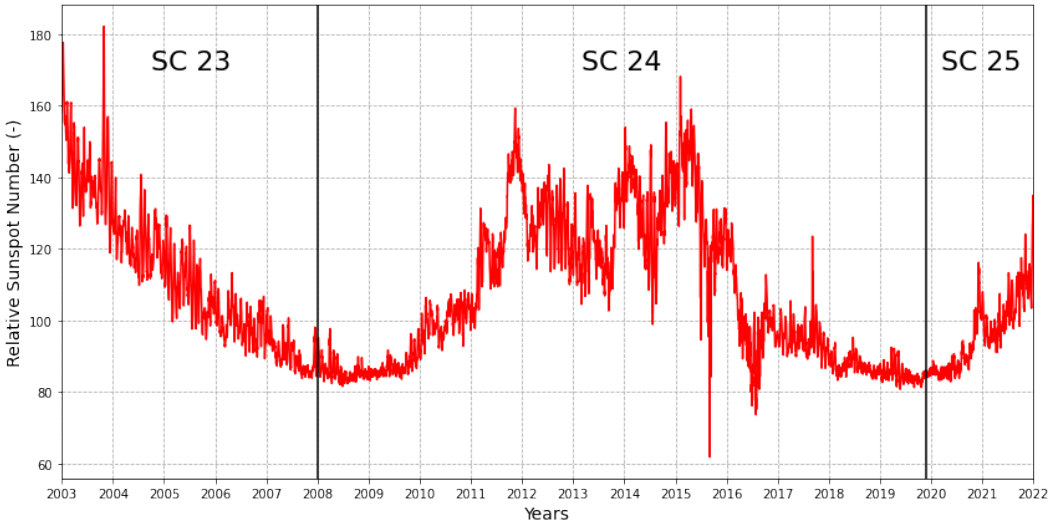

Figura 2.1 - Conjunto de dados de RSN obtidos para análise.

2.1.2 F10.7

A feature F10.7 é amplamente utilizada na modelagem ionosférica uma vez que tem papel de representante das emissões ultravioletas extremas do Sol (EUV) (SCHONFELD et al., 2015). Estas EUVs aquecem e ionizam a atmosfera em altitudes superiores a 150 km mas não conseguem chegar até o solo, sendo assim, é necessário utilizar o índice F10.7 por este ser medido em micro-ondas que podem passar livre pela atmosfera terrestre. Estas micro-ondas com comprimento de onda na ordem de centímetros são bastante sensíveis às condições na alta cromosfera e na base da coroa solar, fazendo delas bons indicadores de atividade solar (TAPPING, 2013).

Além das EUVs o F10.7 também é um bom indicador de atividade magnética solar, irradiância solar e emissões de raios X (HUANG et al., 2009). Fisicamente esta feature representa uma medição do total de emissões no comprimento de onda de 10,7 cm (2,8 GHz) de todas as fontes presentes no disco solar, feito durante um período de 1 hora centrado na época dada pelo valor, esta quantidade na verdade é uma densidade de fluxo e não um fluxo propriamente dito (TAPPING, 2013). Seus valores são medidos em SFU (*Solar Flux Units*) cuja equivalência no S.I. é 1 $SFU = 10^{22}\ W/m^2/Hz$.

Este dado é altamente correlacionado com o número de manchas solares como pode

ser visto em Okoh e Okoro (2020). Entretanto, recentemente Svalgaard e Hudson (2010) apontou que este pode não ser mais o caso para os atuais ciclos solares. Esta mudança de correlação pode estar ligada à mudança temporal da magnitude dos campos magnéticos na umbra de manchas solares (HENNEY et al., 2012).

Assim como os RSN, os índices F10.7 foram obtidos para o período de 2003-2021 espaçados diariamente. Uma visão geral dos dados pode ser vista na Figura 2.2. Percebe-se claramente certa correlação com o RSN porém uma curva com mais ruído nos picos o que tende a achatar o gráfico.

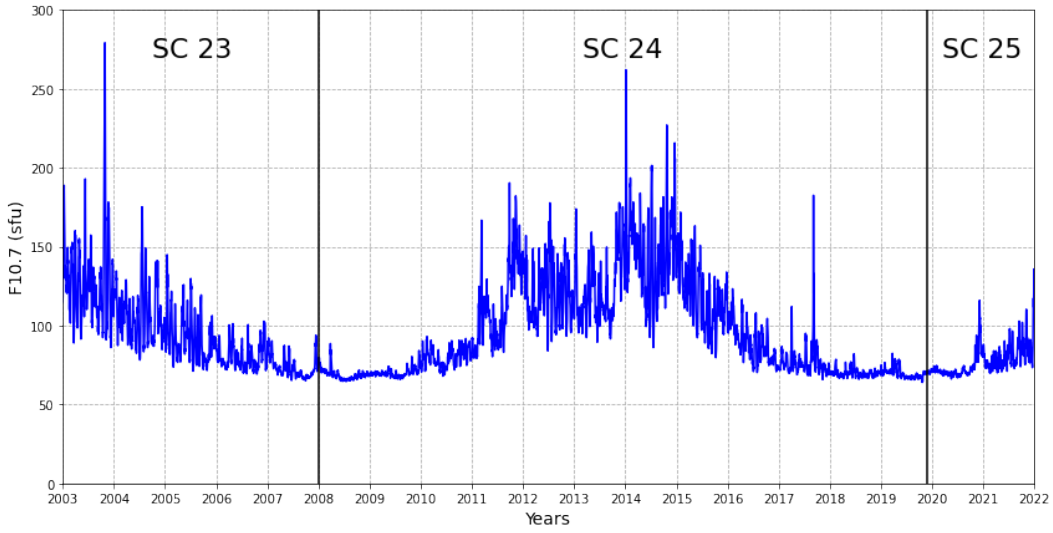

Figura 2.2 - Conjunto de dados de F10.7 obtidos para análise.

2.1.3 Fluxos Fotônicos

Os fluxos fotônicos (PF) representam a quantidade de fótons atingindo a atmosfera terrestre por unidade de área e tempo. Estes fótons serão responsáveis pela criação de elétrons na ionosfera e através de suas diferentes bandas pode se ter uma estimativa da energia depositada para cada faixa de comprimentos de onda.

Obteve-se para análise valores de fluxo fotônico para 39 diferentes bandas que podem ser vistas na Figura 2.3. O período obtido também foi de 2003-2021, sendo que pode se ver os valores de fluxo normalizado para todo o período das bandas inicial (1,86-2,95 nm), média (55,44-59,96 nm) e final (100,10-105,00 nm) do conjunto na Figura 2.4.

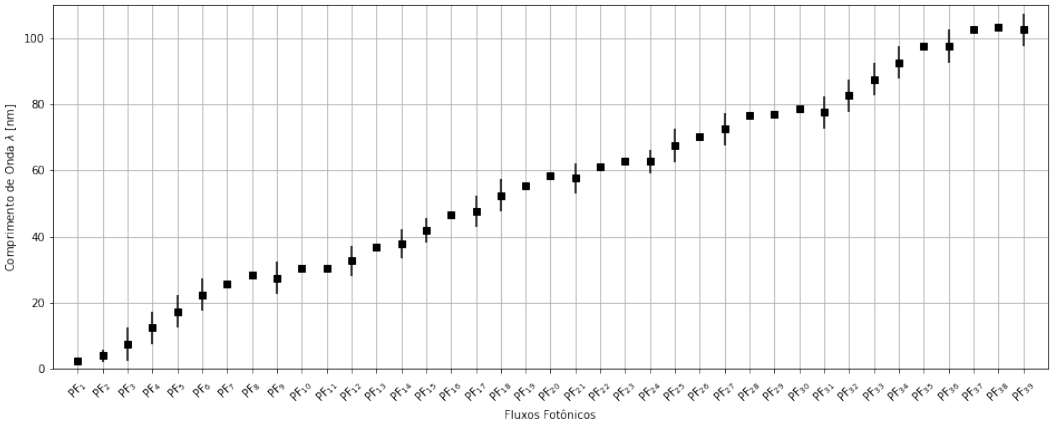

Figura 2.3 - Conjunto de bandas disponíveis para o período de 2003-2021.

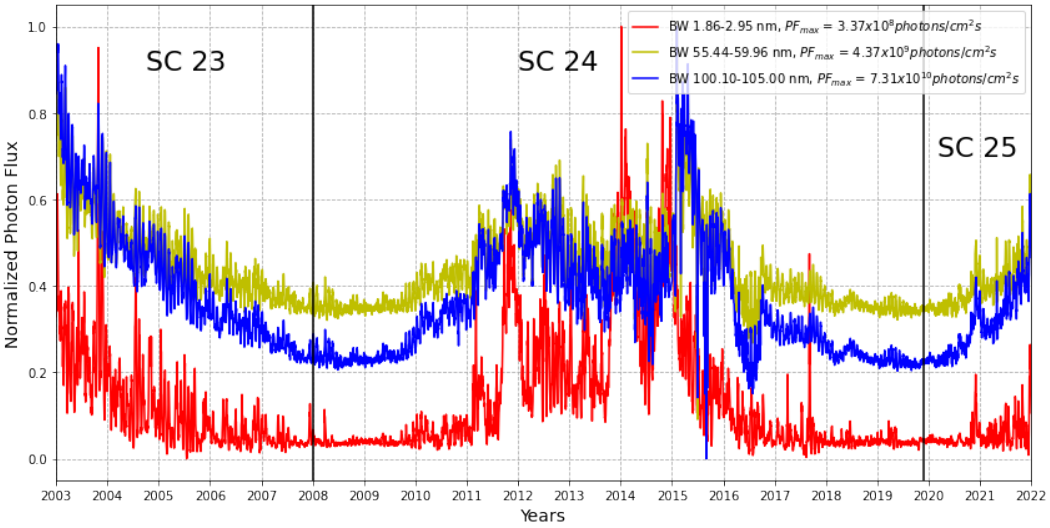

Figura 2.4 - Dados de PF normalizado em três bandas representativas para o período.

2.2 Correlação

A correlação é uma medida da relação entre duas variáveis. Por exemplo, se o aumento de uma variável não causa aumento significativo em outra variável, ou se um pequeno aumento causa grande mudança, as variáveis serão ditas com pouca correlação. No caso em que as variáveis são perfeitamente correlacionadas, um aumento de 1 em uma delas corresponderá em um aumento equivalente (na grandeza da variável dependente) de 1 também.

Como mostrado pela pesquisa desenvolvida no ano anterior, a correlação entre as variáveis é altamente linear, de modo que para medir a correlação (coeficiente an-

gular) foi utilizado neste trabalho o Coeficiente de Correlação de Pearson, Boslaugh (2012). A correlação que se deseja medir é entre o TEC e as features anteriormente apresentadas. Como o TEC é dado para cada ponto do grid global e a cada 2 horas, é necessário fazer uma média diária, criando assim um $TEC_{méd}$.

Utilizando este TEC médio e cada uma das features é possível, com a Equação 2.2, calcular o coeficiente de correlação onde teremos x sendo cada feature e y sendo o $TEC_{méd}$.

$$\rho = \frac{cov(x,y)}{\sqrt{var(x) \cdot var(y)}} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{2.2}$$

Observa-se que o coeficiente de correlação será dado pela razão entre a covariância das duas variáveis e a raiz do produto das variâncias de cada variável. Onde N é o número de dias computados e a barra superior indica a média aritmética.

O coeficiente de correlação calculado para cada feature mudando o tempo da amostra pode ser visto na Figura 2.5. Fez-se esta mudança no conjunto de tempo dos dados para ver se a correlação mudaria quais são as features mais correlatas ao TEC médio.

Fica perceptível que existe uma diminuição geral da correlação conforme se utiliza menos tempo de dados, o que é compreensível uma vez que o menor número de pontos acaba por deixar o efeito do desvio dos valores maior. Porém se for comparado a mudança de correlação de todo o período para apenas de 2021 percebe-se que a feature F10.7 tem uma queda mais acentuada que as outras passando a ser a terceira pior correlação. Isto mostra que para anos recentes, com pouca atividade solar, a feature F10.7 se mostra como uma baixa correlação.

Dentre os fluxos fotônicos a maior correlação ocorreu para a 9^a banda que é comprimento de onda de 25,11-29,95 nm.

Vale lembrar que uma alta correlação não indica necessariamente que uma feature será melhor que outra para um modelo, mas ajuda a filtrar aquelas que não apresentam mesmo nível de associação que as demais. Exemplo disso, é que se treinássemos o modelo apenas com a banda de fluxo fotônico com máxima correlação o erro dos resultados preditos seria alto uma vez que estamos avaliando apenas uma parte específica do espectro do EUV, o que causa a exclusão de todas as outras bandas significativas fisicamente.

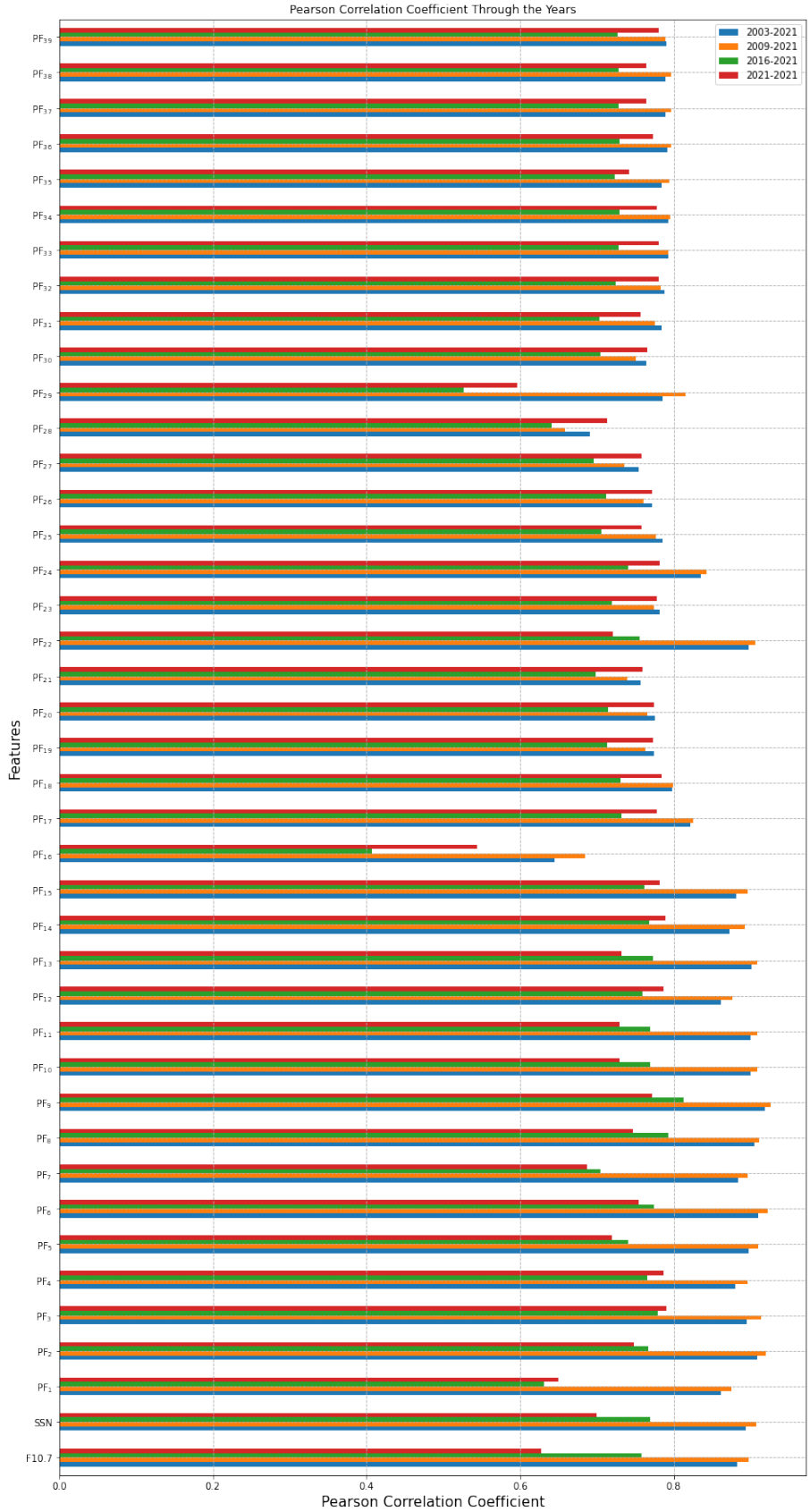

Figura 2.5 - Correlação entre o TEC médio diário e todas as features para diferentes períodos de tempo.

2.3 Combinação de Fluxos Fotônicos

De modo a condensar as 39 bandas de fluxo fotônico em apenas uma variável, visando que o número de entradas do modelo fosse reduzido e tentando manter o máximo de informação para o modelo possível, foi proposta a combinação das bandas em apenas uma. A combinação foi feita utilizando a média ponderada pelo coeficiente de correlação de Pearson com relação ao TEC médio diário, como pode ser visto na Equação 2.3.

$$PF_{comb} = \frac{\sum_{i=1}^{39} \rho(i) PF(i)}{\sum_{i=1}^{39} \rho(i)} \tag{2.3}$$

Este método permitiu dar mais importância àquelas bandas de fluxo fotônico que apresentaram maior correlação, de modo que a feature final (nomeada PF combinado) tem coeficiente de correlação mais alto do que se fosse uma simples média aritmética. Pode ser visto na Figura 2.6 as correlações das diferentes features utilizadas.

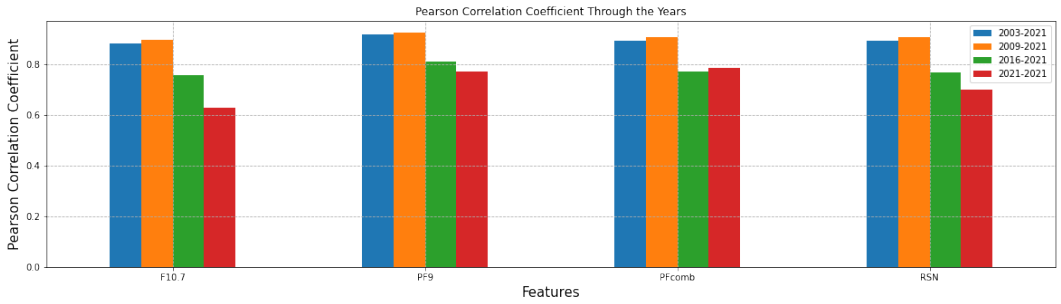

Figura 2.6 - Comparação entre coeficientes de correlação para diferentes combinações de fluxos fotônicos.

É perceptível que a correlação da banda de fluxo fotônico com maior correlação (PF_9) seja praticamente a mesma do PF_{comb}. Entretanto, quando é avaliado o RMSE gerado para um mesmo período de treino e mesmo período de teste com estas diferentes features, percebe-se que o PF_9 acaba gerando um erro anual maior para todos os anos de teste. O erro gerado é em torno de 0,5 TECu a mais para os anos de teste de 2019-2021, como pode ser visto na Tabela 2.1.

Ano	RMSE [TECu]	
	PF_9	PF_{comb}
2019	3,526	3,055
2020	3,545	3,082
2021	3,560	3,098

Tabela 2.1 - RMSE anual para as previsões dos modelos com PF_9 e PF_{comb}.

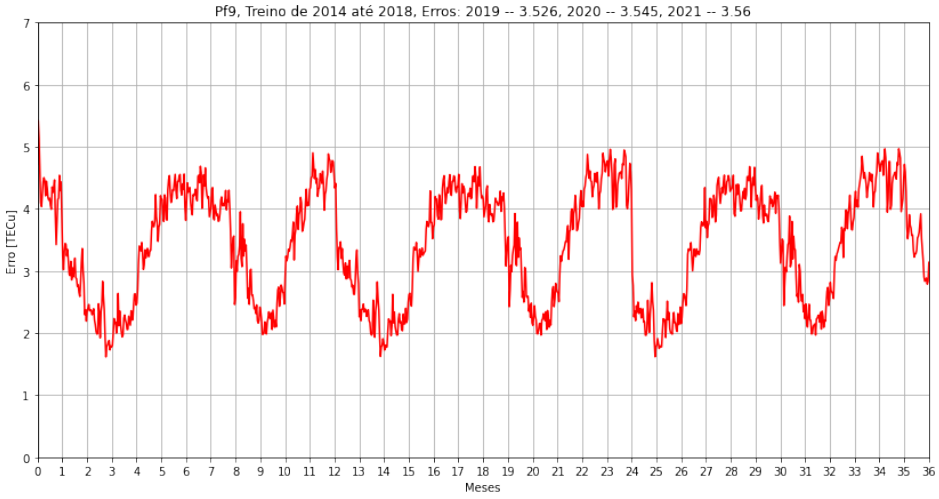

Figura 2.7 - RMSE para modelo linear treinado com dados de PF_9 de 2014-2018 e testado de 2019-2021.

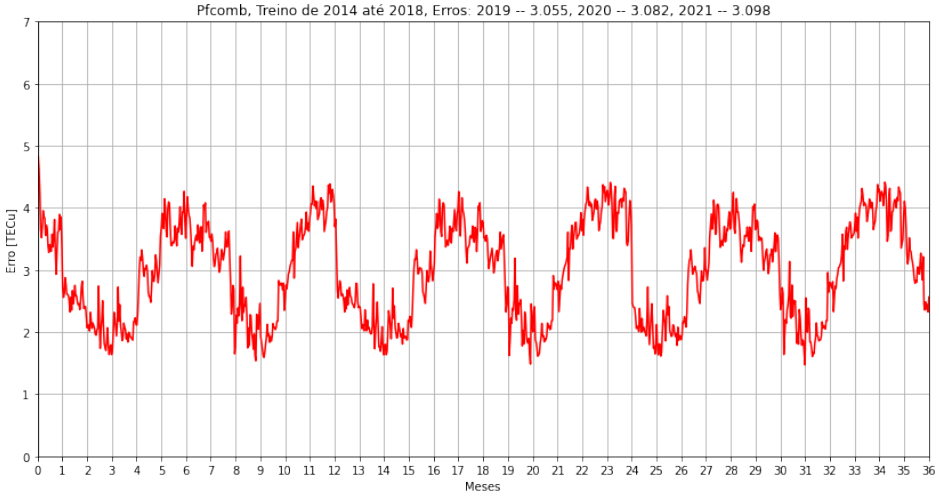

Figura 2.8 - RMSE para modelo linear treinado com dados de PF_{comb} de 2014-2018 e testado de 2019-2021.

2.4 Modelagem

Como comentado na Seção 1.2 os valores de TEC da ionosfera foram discretizados em um grid de retângulos 2,5°x5,0º em latitude e longitude respectivamente. Cada ponto deste grid terá 13 valores de TEC para cada dia do ano. Sendo assim, primeiramente fez-se a transformação dos dados indo do domínio do tempo para o domínio da frequência, por meio da transformada discreta do cosseno (DCT).

A transformada discreta do cosseno tem uma compactação maior da informação nas primeiras frequências do que a transformada discreta de Fourier (DFT), por conta disso, seria possível, no futuro, otimizar o modelo pela exclusão das frequências mais altas e manter mesmo assim um erro baixo. Neste trabalho esta exclusão não foi feita, então a transformação dos dados pode ser descrita pela Equação 2.4.

$$f(i) = C_i \sqrt{\frac{2}{N}} \sum_{j=0}^{N-1} TEC(j) \cdot cos\left[\frac{\pi i}{N}\left(j + \frac{1}{2}\right)\right] \tag{2.4}$$

Onde $f(i)$ são as frequências que vão de i = 0, 2, ..., N-1. N são o número de dados de TEC por dia, neste caso 13. O coeficiente C_i é 0 para j = 0 e $\sqrt{1/2}$ para qualquer outro valor de j.

Este procedimento é feito para cada posição no grid espacial de modo que ao final existirão 13 frequências para cada posição espacial. Isto faz com que todos os dados possam ser agrupados em relações diretas entre cada frequência e uma ou mais entradas (features). A Figura 2.9 mostra a distribuição de todos os pontos (2003-2021) da primeira frequência para diferentes features.

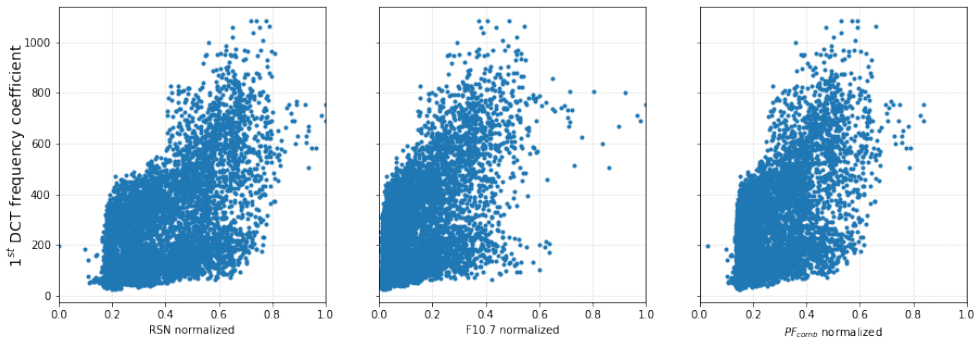

Figura 2.9 - Primeiro coeficiente de frequência após aplicação da DCT.

Com isto é possível utilizar um modelo de aprendizado de máquina para prever o comportamento de cada frequência com relação à uma entrada de atividade solar. Uma vez com os dados previstos de frequência utiliza-se a transformada discreta inversa do cosseno (IDCT) dada pela Equação 2.5.

$$TEC(j) = \sqrt{\frac{2}{N}} \sum_{i=0}^{N-1} C_i \cdot f(i) \cdot cos\left(\frac{(2j+1)i\pi}{2N}\right) \tag{2.5}$$

Um ponto importante para o bom funcionamento do modelo é a necessidade de normalizar os dados pois se tem fluxos fotônicos na ordem dos bilhões e F10.7 e RSN na ordem das centenas. De modo a explorar o efeito dos diferentes tipos de normalização foram testados três modos de normalizar para diferentes combinações de features e anos de treino.

O primeiro método é a normalização, a qual é dada pela expressão 2.6 para cada dado de índice i. A normalização tem por objetivo deixar as features dentro de uma escala, no caso, de [0, 1].

$$x_{norm}(i) = \frac{x(i) - min(x)}{max(x) - min(x)} \tag{2.6}$$

Outra forma de se normalizar os dados é utilizar a padronização, a qual tem por objetivo deixar os dados com uma distribuição normal, ou seja, com média 0 e desvio padrão de 1. A expressão matemática para a padronização é dada pela Equação 2.7

$$x_{pad}(i) = \frac{x(i) - \mu(x)}{\sigma(x)} \tag{2.7}$$

Onde $\mu(x)$ é a média dos dados x e $\sigma(x)$ é o desvio padrão dos dados x.

Os resultados do RMSE anual para modelos treinados pelo mesmo período de tempo (começando em 2018 e indo cada vez mais ao passado) e mesmas features (RSN, F10.7, PF1-3) pode ser visto na Figura 2.10, os anos de teste foram todos 2019. Percebe-se que o tipo de método de escala não tem influência no modelo linear, entretanto apresenta efeito no modelo de Máquinas de Vetores de Suporte (SVM). Isto se explica pela insensibilidade do modelo, pois depende apenas de uma transformação linear e assim os coeficientes são ajustados para dar um mesmo resultado independente do método de escala dos dados. Já como o SVM é um algoritmo ba-

seado na distância entre os pontos, a normalização acaba tendo um efeito benéfico quando comparada com a padronização, pois a última acaba assumindo a mesma importância para features diferentes.

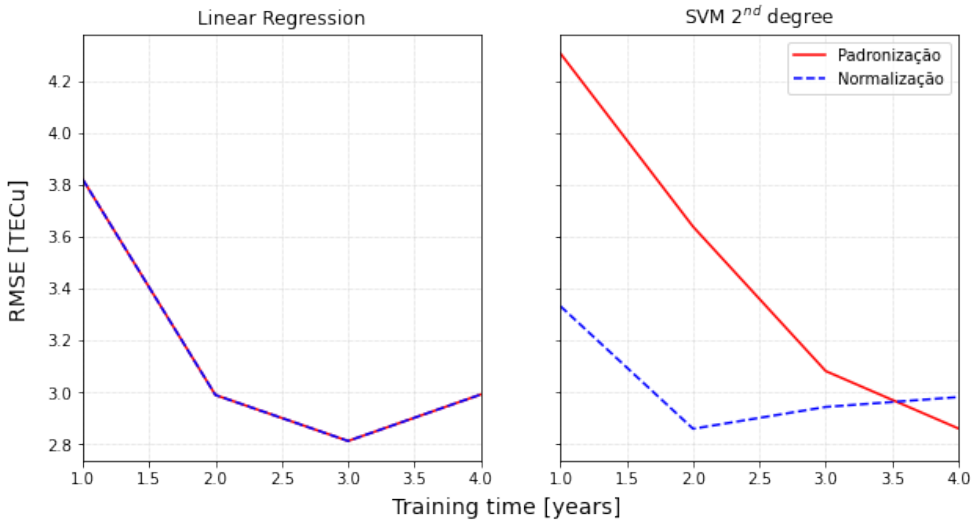

Figura 2.10 - Comparação do efeito da normalização e padronização no RMSE anual.

Por conta disso, utilizou-se apenas a normalização como método de escala dos dados para este trabalho. Como visto na pesquisa anterior, os métodos que apresentaram melhores resultados foram a regressão linear e o SVM.

2.4.1 Regressão Linear

O modelo de previsão por regressão linear é dado pela Equação 2.8 (GOODFELLOW et al., 2016).

$$\bar{Y}(\bar{x}) = \mathbf{w}\bar{x} + \bar{b} \tag{2.8}$$

Onde $\mathbf{w}$ é a matriz de pesos, $\bar{x}$ é o vetor com as features normalizadas de teste, $\bar{b}$ é o vetor de termos independentes e $\bar{Y}(x)$ são os valores do coeficiente de uma frequência específica prevista para as features de entrada. O processo deve ser feito para as 13 frequências. A matriz de pesos $\mathbf{w}$ será de formato MxK, onde M é o número de dados e K o número de features, $\bar{b}$ e $\bar{Y}(\bar{x})$ são de dimensão Mx1 e $\bar{x}$ de dimensão Kx1.

Os pesos para o modelo são obtidos através da minimização de uma função de custo. O método mais comum é o de mínimos quadrados que resulta na Equação 2.9, (MURPHY, 2012).

$$\mathbf{w} = (\hat{x}^T \hat{x})^{-1} \hat{x}^T \hat{Y} \tag{2.9}$$

Onde $\hat{x}$ são as features de treino e $\hat{Y}$ os coeficientes das frequências de treino.

2.4.2 Regularização

Para modelos lineares, há um método para evitar sobre ajustamento dos dados através da regularização. Este método se baseia na limitação dos pesos da matriz de pesos **w**. Esta regularização pode ser dada pelo método Ridge, Lasso ou ainda uma combinação dos dois, chamada Elastic Net, (HASTIE et al., 2017).

2.4.3 Regressão Ridge

A Regressão Ridge (também chamada de regressão L_2 por conta da norma utilizada) é a versão da regressão linear em que um termo linear α é posto na função de custo, proporcional ao quadrado dos pesos da função peso. A soma deste termo força com que o algoritmo não apenas se ajuste aos dados mas também mantenha os coeficientes o mais baixo possível, (GéRON, 2019).

O termo que controla o quanto de regularização é imposta sobre o modelo é chamado de hiperparâmetro α. Para o caso limite de $\alpha = 0$, o modelo se torna uma regressão linear, para α muito alto os pesos vão para zero e o modelo não tem valor para previsão. A Equação 2.10 apresenta a função de custo para este método de regularização.

$$J(\mathbf{w}) = MSE(\mathbf{w}) + \alpha \frac{1}{2} \sum_{i=1}^{n} w_i^2 \tag{2.10}$$

Onde n é o número de pesos (features) e $MSE(\mathbf{w})$ é a função de custo do modelo sem regularização (Mean Square Error), no caso linear, dada por

$$MSE(\mathbf{w}) = MSE(X,\ h_{\mathbf{w}}) = \frac{1}{m} \sum_{i=1}^{m} (\mathbf{w}^T x^{(i)} - y^{(i)})^2 \tag{2.11}$$

2.4.4 Regressão Lasso

A Regressão Lasso (do inglês *Least Absolute Shrinkage and Selection Operator Regression*) também chamada de L_1 por conta de utilizar a norma ℓ_1 ao invés de metade da norma ℓ_2, (GéRON, 2019). A Equação 2.12 apresenta a função de custo para a um modelo com regularização L_1.

$$J(\mathbf{w}) = MSE(\mathbf{w}) + \alpha \sum_{i=1}^{n} |w_i| \tag{2.12}$$

A Regressão com regularização Lasso tende a eliminar completamente os pesos das features de pouca importância, diferentemente da Ridge que não tem capacidade de zerar completamente um coeficiente de peso, (HASTIE et al., 2017).

2.4.5 Elastic Net

A regularização utilizando Elastic Net faz uma ponderação das regularizações Lasso e Ridge. O termo que controla a taxa de mistura dos dois métodos é o r, sendo α o hiperparâmetro, como já comentado. Quando $r = 0$, o Elastic Net é equivalente ao Ridge e quando $r = 1$, equivalente ao Lasso. A Equação 2.13 representa a função de custo para um modelo com regularização Elastic Net.

$$J(\mathbf{w}) = MSE(\mathbf{w}) + r\alpha \sum_{i=1}^{n} |w_i| + \alpha \left(\frac{1-r}{2} \right) \sum_{i=1}^{n} w_i^2 \tag{2.13}$$

2.4.6 Máquina de Vetores de Suporte

As Máquinas de Vetores de Suporte (SVM) foram criadas a partir da teoria de aprendizado estatístico feita por Vapnik-Chervonenkis, (UNPINGCO, 2016). O modelo se baseia na separação de grupos de dados através de uma linha com uma certa zona de *outliers*, que pode ser reta, quando for um SVM linear ou de qualquer outra ordem quando se tem um SVM de grau superior. Os SVMs foram originalmente feitos para classificação binária, porém é possível adotar o método para regressão, (MURPHY, 2012).

Dentro dos SVMs existem dois hiperparâmetros que limitarão o tamanho da zona de *outliers* e a quantidade de *outliers* que poderão existir nesta área, (GéRON, 2019). O hiperparâmetro ϵ controla o tamanho da margem de *outliers*, onde um valor grande de ϵ cria uma solução com maiores erros e um valor pequeno penaliza cada erro

e eleva o custo computacional. Para controlar a regularização do SVM se utiliza o hiperparâmetro C, que dita o nível de regularização do tipo L_2. Valores menores de C adicionam mais regularização, fazendo com que se reduza o sobre ajustamento porém também aumentando custo computacional.

2.5 Dependência Sazonal

Observou-se no projeto de pesquisa anterior (BENOIT; PETRY, 2021) que o RMSE médio diário, entre o TEC previsto e o observado, varia durante o ano. Esta variação está relacionada com os períodos de equinócios e solstícios solares. Pode ser visto na Figura 2.11 a variação deste erro para um modelo linear treinado de 2014 a 2018 e testado em 2019, com as features de RSN, F10.7 e $PF_1 - PF_3$ (as mesmas utilizadas no trabalho anterior).

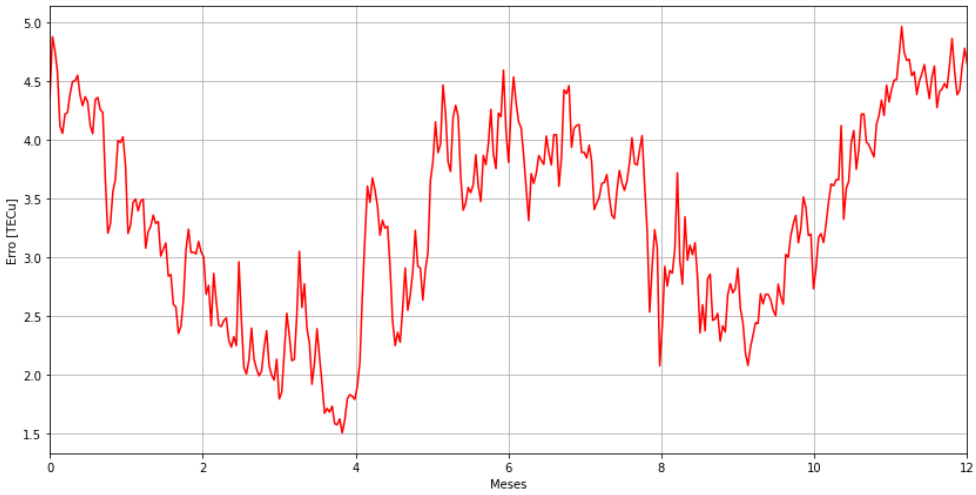

Figura 2.11 - Variação do RMSE diário para o mesmo modelo do trabalho anterior.

Durante o ano a posição relativa do Sol com o eixo de rotação da Terra muda, ou seja, um hemisfério recebe mais radiação solar que o outro durante um período (verão) e menos durante outro (inverno). O momento em que o Sol atinge a maior inclinação positiva com relação ao eixo de rotação da Terra é chamado de solstício de verão, já o contrário acontece para o solstício de inverno. A transição entre estes períodos (duas anualmente) são a primavera e o outono, onde o Sol energiza de forma mais uniforme a ionosfera. O período em que o Sol está perfeitamente alinhado com o Equador terrestre é chamado de equinócio.

O modelo treinado sem uma divisão dos dados com relação ao período do ano, gera um erro menor para o período em que lhe é apresentado mais dados, ou seja, os equinócios. Desta forma explica-se o comportamento senoidal do erro com a passagem dos meses do ano.

Visando a utilização deste comportamento para a obtenção de um erro anual menor, foi proposto a subdivisão do modelo em três. Para a divisão dos dados se passou o dia central e foi feita repartição dos dias buscando deixar um espaçamento igual entre os períodos.

O primeiro modelo foi treinado apenas com dados centrados no solstício de verão do hemisfério norte (20 de junho, dia 171 do ano). O segundo modelo foi treinado apenas com o período de dias cujo dia central é o solstício de inverno do hemisfério norte (21 de dezembro, dia 355 do ano). Por fim, o terceiro modelo fica com os dias que tem os equinócios como centro (20 de março e 22 de setembro, respectivamente dias 79 e 265 do ano).

3 RESULTADOS

3.0.1 Seleção de Features

Para a seleção de features, além de toda a análise de correlação e combinação dos fluxos fotônicos, foram feitas análises da evolução do RMSE em um ano de teste específico, variando o conjunto de *features* do modelo e também os anos de treino. A Figura 3.1 mostra os resultados de RMSE médio anual para o ano de teste de 2019 sendo os períodos de treino começando todos em 2018 e indo cada vez mais para trás no tempo.

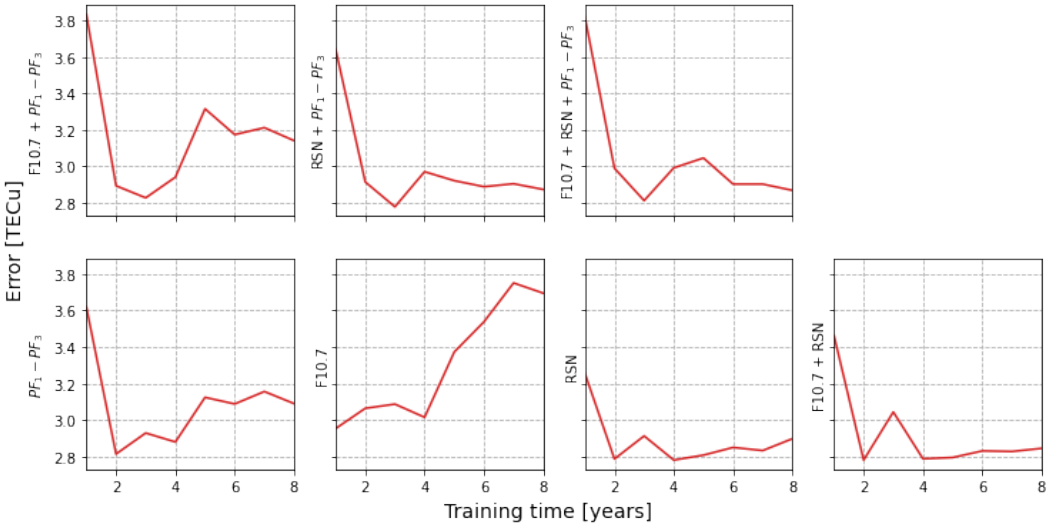

Figura 3.1 - Variação do RMSE conforme se aumenta o número de anos de treino começando em 2018, sendo o treino sempre em 2019.

Existem algumas conclusões que podem ser tiradas dos gráficos presentes em 3.1. A primeira delas é que, os modelos com a *feature* RSN tendem a ter um desempenho melhor que outros modelos. Em segundo lugar, modelos treinados apenas com F10.7 tendem a apresentar resultados muito piores que os demais conforme se aumenta os anos de treino. De um modo geral, com apenas 2 ou 3 anos de treino, os erros encontrados já são próximos, se não os menores observados, indicando que longos períodos de treino não são proveitosos para a diminuição do RMSE.

Com base nesses resultados a *feature* F10.7 foi retirada de análises futuras por conta de seu desempenho pobre, sendo seu papel rediscutido ao final do trabalho.

3.0.2 Sintonização de Hiper-parâmetros

Como apresentado na Seção 2.4.2 para controlar os coeficientes dos modelos de aprendizado de máquina pode se utilizar técnicas de regularização. Neste trabalho foi utilizado o método de gridsearch para encontrar os valores dos hiper-parâmetros α e r do *Elastic Net* que minimizassem o RMSE médio anual no modelo de regressão linear. O período de treino foi sempre o mesmo de 2018 a 2014 e os anos de teste foram 2019 a 2021. Os resultados encontrados para os anos de teste não variaram mais do que $0,06\ TECu$, portanto, apenas os resultados de 2021 que foram os com maiores erros serão apresentados. O processo de *gridsearch* foi feito para *features* de apenas RSN e RSN e PF_{comb}.

As Figuras 3.2 e 3.3 apresentam os resultados para os modelos treinados apenas com RSN e com RSN e PF_{comb} respectivamente. Percebe-se que há uma diferença desprezível entre os resultados encontrados para os dois tipos de *features*.

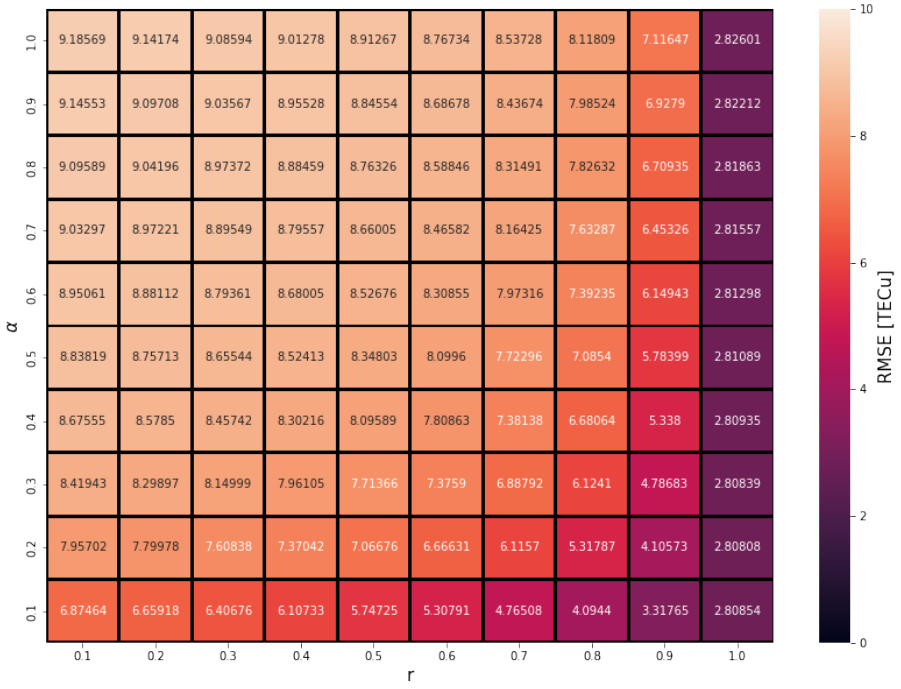

Figura 3.2 - Mapa de calor dos coeficientes de regularização do modelo linear para RSN, testado em 2021.

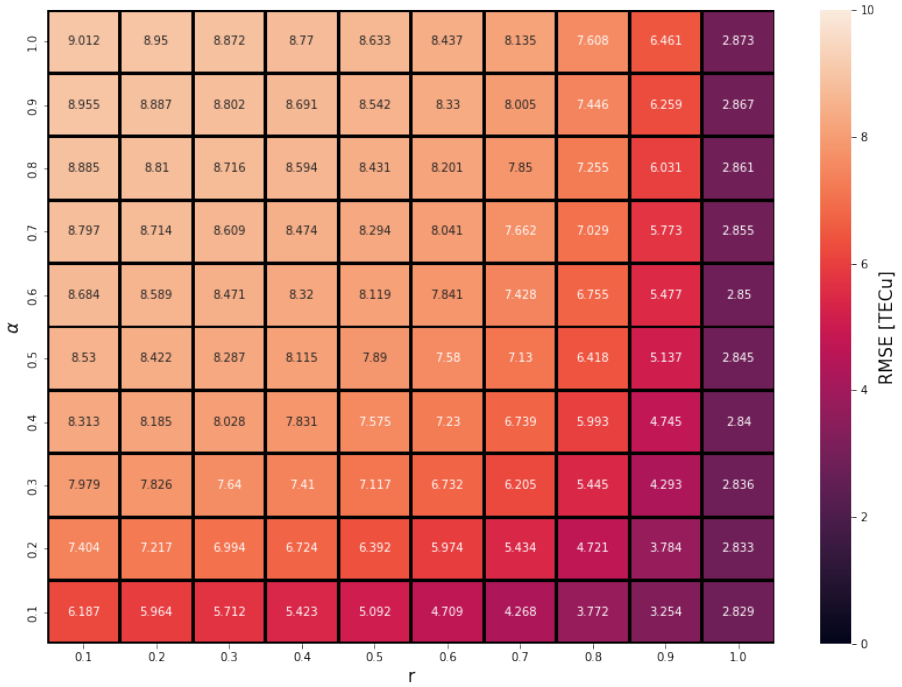

Figura 3.3 - Mapa de calor dos coeficientes de regularização do modelo linear para RSN e PF_{comb}, testado em 2021.

Observou-se que um r de 1 foi o que apresentou no geral menor RMSE médio anual, o que indica que uma regularização Lasso traz resultados melhores que uma Ridge. O valor do hiper-parâmetro α com melhores resultados foi o menor analisado, de 0,1. Isto indica que uma diminuição dos coeficientes do modelo linear trazem menor RMSE médio anual, pois o efeito de α próximo de 0 é diminuí-los. Os menores valores de RMSE encontrados são iguais àqueles encontrados para modelos sem nenhuma regularização, o que indica que a seleção de *features* como RSN e RSN + PF_{comb} não necessita de regularização para a obtenção de resultados ótimos, uma hipótese que pode ser levantada também é que os modelos treinados com RSN já teriam normalmente coeficientes baixos e valores previstos que não variam muito com o tempo devido às características do próprio RSN.

3.0.3 Comparação entre Modelos Normal e Sazonal

Como comentado na Seção 2.5 a variação do RMSE diário varia com o tempo, seguindo um período proporcional às estações do ano. Este comportamento sazonal pode ser visto na Figura 3.4 onde se tem a variação do RMSE do TEC previsto para o ano de 2019, tendo os mesmos anos de treino e *features* do trabalho anterior.

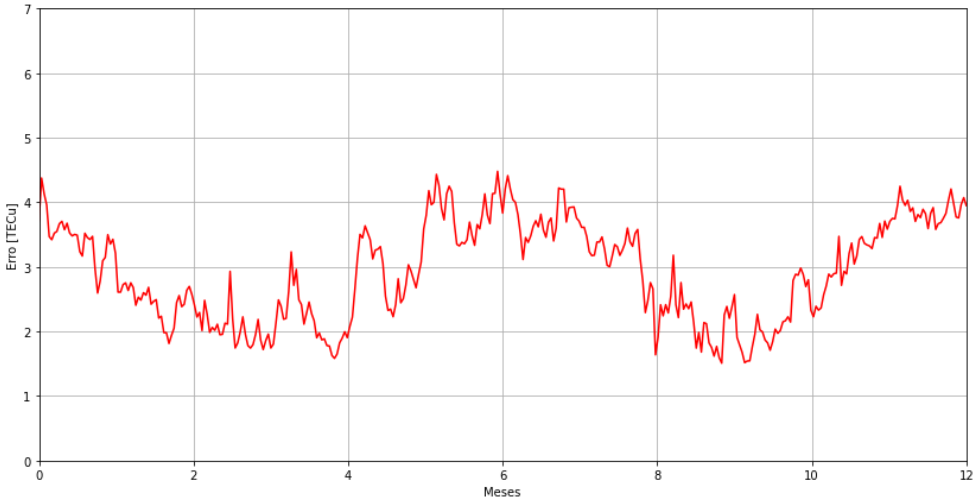

Figura 3.4 - Variação durante o ano do RMSE para modelo treino de 2014-2018 com RSN, F10.7 e $PF_1 - PF_3$.

Com este comportamento surge a oportunidade de subdividir o modelo em submodelos com dados de treino em apenas uma posição solar (verão, inverno e outono e primavera juntos). Com isso, os três modelos foram treinados com apenas estes dados e seus valores previstos podem ser vistos da Figura 3.5 a 3.7. Percebe-se que para o período em que receberam dados de treino, os erros obtidos são muito menores, como esperado.

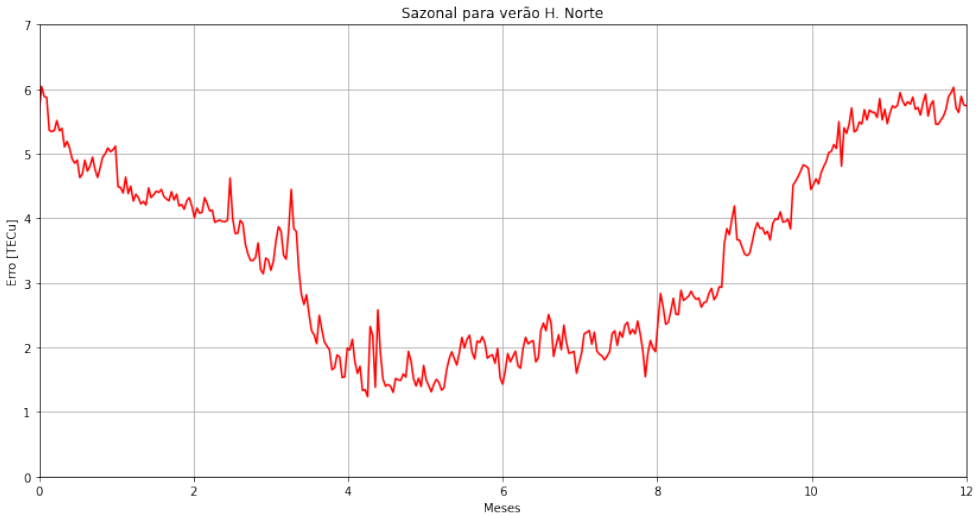

Figura 3.5 - Variação do RMSE para modelo treinado com dados centrados no solstício de verão no Hemisfério Norte do dia 171 ao 264.

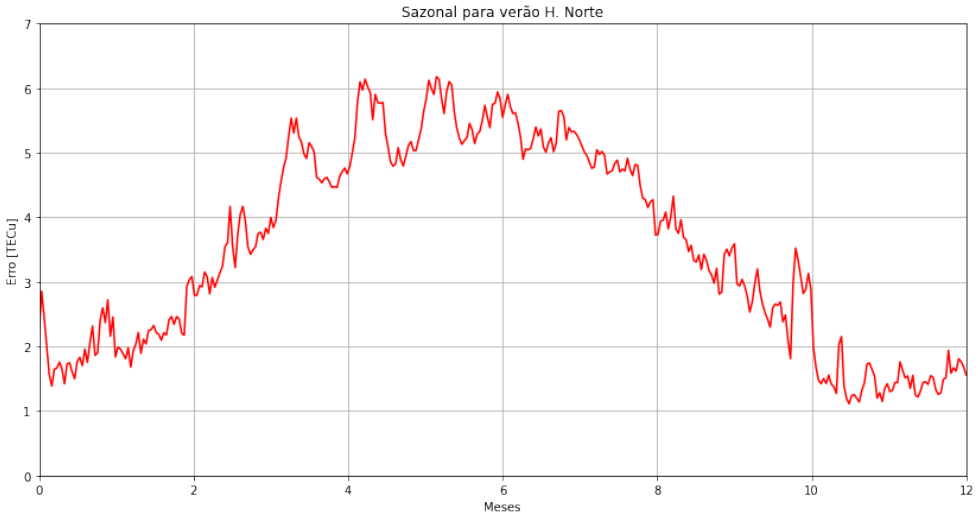

Figura 3.6 - Variação do RMSE para modelo treinado com dados centrados no solstício de verão no Hemisfério Sul do dia 354 ao 78.

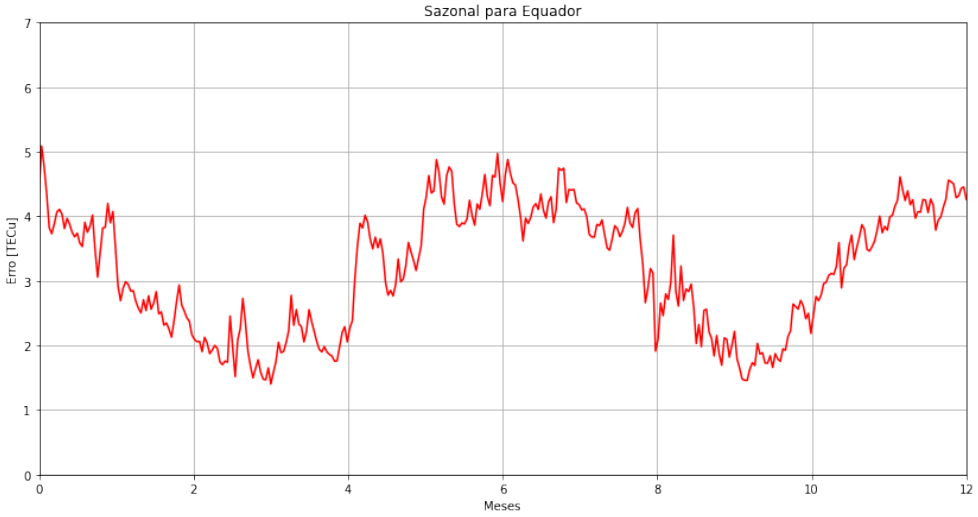

Figura 3.7 - Variação do RMSE para modelo treinado com dados centrados nos equinócios (períodos aproximados de primavera e outono) dos dias 78 ao 171 e 264 ao 358.

Para que o RMSE geral seja menor que o obtido sem a divisão sazonal, é necessário resgatar apenas os valores previstos no período em que o modelo é treinado para ser bom. Fazendo isso e agrupando em um só grupo de valores previstos, se tem a previsão para um ano inteiro baseado nas previsões feitas por cada submodelo. O RMSE gerado por essa subdivisão sazonal pode ser vista na Figura 3.8 onde

percebe-se claramente um RMSE menor.

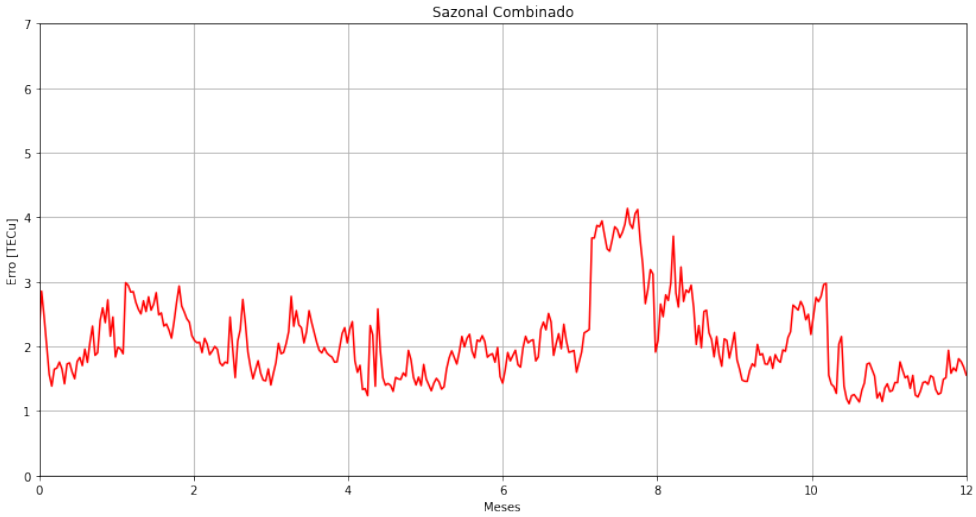

Figura 3.8 - Variação durante o ano de 2019 do RMSE para treino de 2014-2018 com RSN, F10.7 e $PF_1 - PF_3$ e divisão sazonal.

Se tem ainda um pico por volta de Agosto onde o segundo período do modelo dos equinócios começa, correspondendo à primavera no Hemisfério Sul. Este pico poderia ser apagado caso se aumentasse o período do inverno no Hemisfério Sul, o qual tem um erro menor para este período como pode ser visto na Figura 3.5.

Esta mesma divisão sazonal (utilizando os mesmos dias de divisão), para um modelo linear treinado com PF_{comb} gera os resultados mostrados na Figura 3.9. Como para o modelo do trabalho anterior quando na divisão sazonal, existem picos de erro em certas regiões do ano, as quais poderiam ser ajustadas com a escolha dos dias para a divisão.

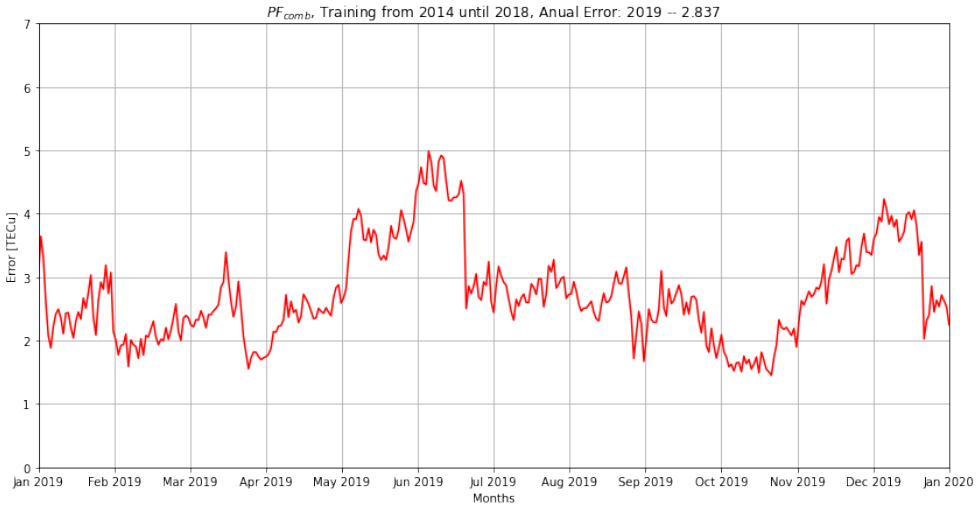

Figura 3.9 - Variação durante o ano de 2019 do RMSE para regressão linear, treino de 2014-2018 com PF_{comb} e divisão sazonal.

Por fim, para um modelo linear utilizando RSN e PF_{comb} os resultados de RMSE para os anos de teste de 2019 a 2021 podem ser vistos na Figura 3.10. Os picos também estão presentes porém o erro médio anual é 0,3 TECu menor que o menor erro anual medido para o modelo sem sazonalização com estas mesmas *features*. Com uma seleção melhor dos dias para divisão e uma avaliação das *features* levando em conta a divisão sazonal, é possível que os resultados encontrados sejam ainda melhores.

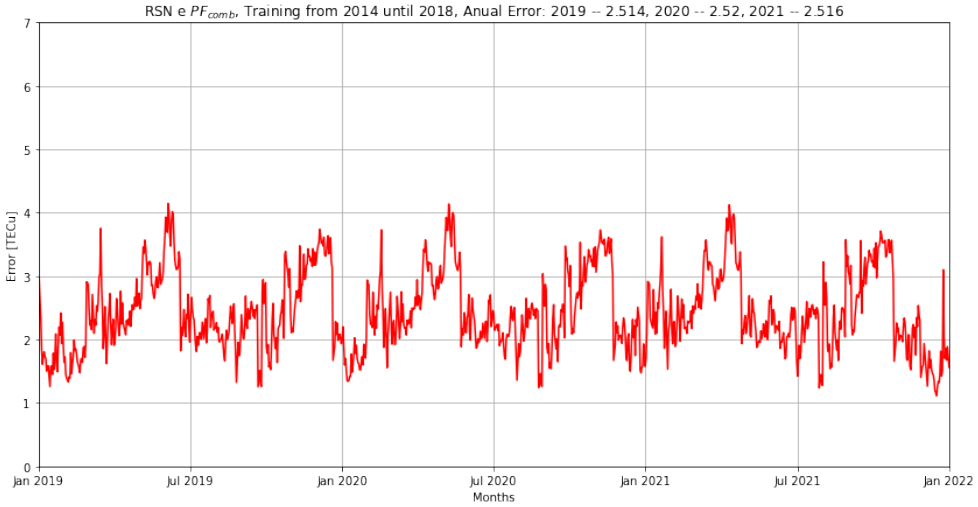

Figura 3.10 - Variação durante o ano de 2019, 2020 e 2021 do RMSE para regressão linear, treino de 2014-2018 com PF_{comb} e RSN e divisão sazonal.

3.0.4 Desempenho ao Longo do Tempo

As *features* selecionadas para analisar o desempenho do modelo sem divisão sazonal com o tempo foram RSN, PF_{comb} e PF_9. Foram alterados os anos de treino, sempre começando em 2018 e pondo o ano de treino inicial cada vez mais no passado indo até 2010. Os anos de teste foram 2019, 2020 e 2021, sendo que os RMSEs encontrados aumentam com o tempo, como seria esperado pois os dados de treino se tornam mais antigos e menos correlatos com a atualidade. Os resultados são apresentados na Figura 3.11.

Primeiramente, fica claro que o modelo treinado apenas com PF_9 apresenta resultados bem piores que os demais conforme se aumenta os anos de treino para além de 4 anos, sendo que para pouco período de treino os resultados são relativamente bons. De um modo geral a partir de 3 anos de treino os resultados de RMSE não variam mais tanto, com exceção dos modelos apenas com PF_9 e PF_{comb} que se beneficiam de mais dados de treino.

Dentre as *features* analisadas, aqueles modelos treinados com RSN, seja com outro fluxo fotônico ou não, apresentaram RMSE menores num geral, sendo que a combinação RSN + PF_{comb} obteve os menores valores, que foram logo abaixo do 2,8 TECu.

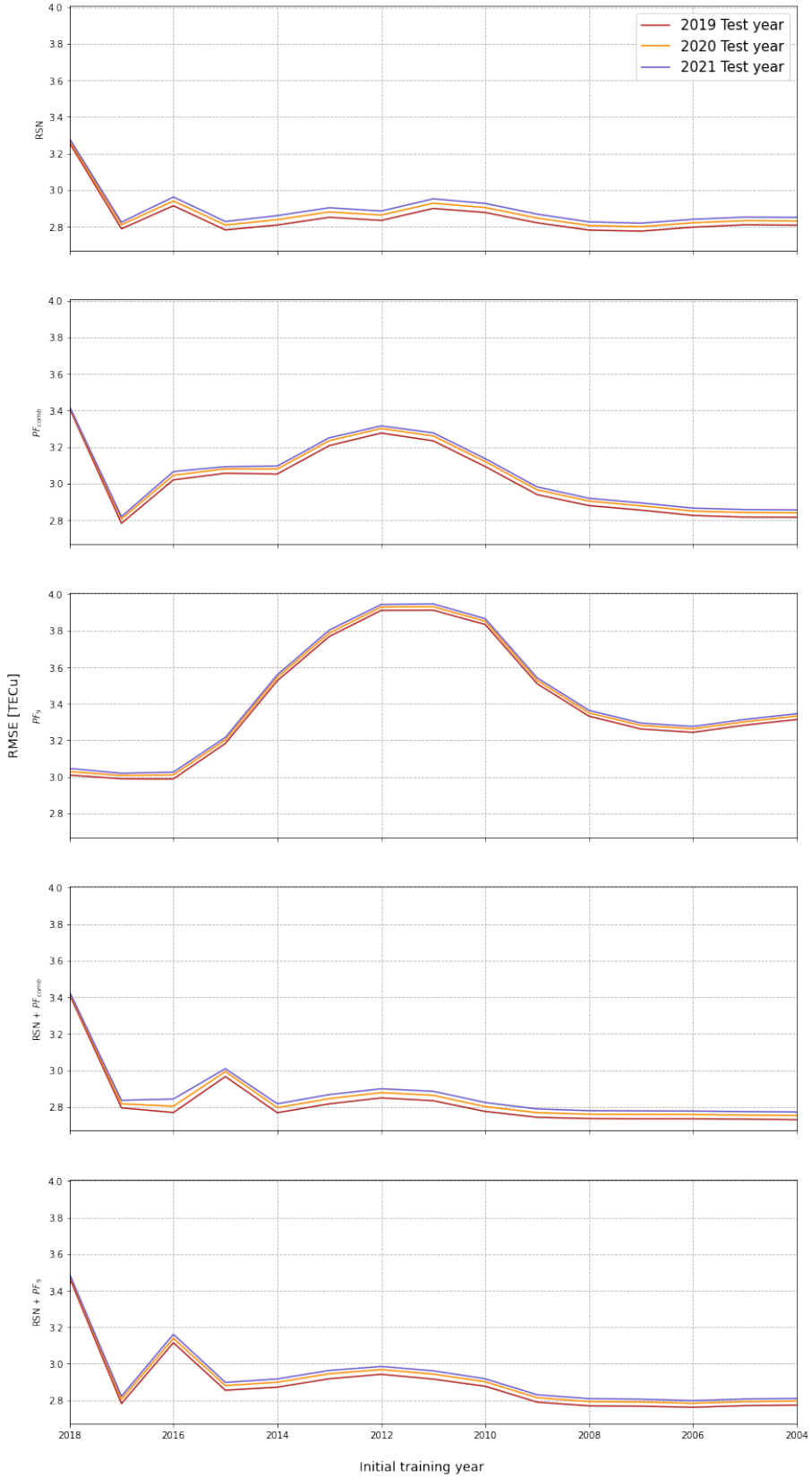

Figura 3.11 - Comparação entre RMSE para diferentes anos de teste, treino e conjuntos de features.

4 CONCLUSÕES

Com relação aos resultados obtidos durante a pesquisa pode se apontar as seguintes conclusões:

- Foi possível diminuir em mais da metade o tempo para uma mesma simulação pela otimização do código;
- Conseguiu-se aumentar o banco de dados ao máximo dos dados disponíveis de TEC e os respectivos dados de atividade solar para o mesmo período;
- Pode concluir-se que o índice F10.7 sozinho não é uma boa feature para estes modelos. Entretanto, quando em conjunto com outras features causa o efeito de sazonalidade no erro, que pode ser utilizado para obter um erro global menor através da subdivisão explorada;
- Apenas pela seleção correta de features é possível obter erros equivalentes àqueles atingidos com regularização no trabalho anterior;
- O aumento demasiado dos anos de treino (mais do que 5 anos, para o modelo sem divisão sazonal) causa, de modo geral, aumento de RMSE. Esta conclusão pode não ser válida para os casos de subdivisão sazonal, pois a quantidade de dados é menor.

Como recomendações de trabalhos futuros há a possibilidade de seleção dos períodos de divisão sazonal que geram os melhores resultados, assim como a seleção de *features* baseado também nos resultados de divisão sazonal. Para isso é necessário levar em conta se o gráfico de RMSE gerado tem um comportamento senoidal estável, sem muito ruído o que acaba por gerar um RMSE sazonal menor, estas características não estão presentes no RSN, por exemplo, mesmo este apresentando menor RMSE global sem sazonalização.

REFERÊNCIAS BIBLIOGRÁFICAS

BENOIT, A.; PETRY, A. Evaluation of f10.7, sunspot number and photon flux data for ionosphere tec modeling and prediction using machine learning techniques. **Atmosphere**, 2021. Disponível em: <https://dx.doi.org/10.3390/atmos1010000>. 18

BOSLAUGH, S. **Statistics in a Nutshell**: A desktop quick reference. 2^{nd} edition. Sebastopol, CA, USA: O'Reilley, 2012. 9

CANDER, L. R. **Ionospheric Space Weather**. 1^{st} edition. Cham, Switzerland: Springer, 2019. (Springer Geophysics). Astrophysics and Space Science Library. 5

CHOUDHURI, A. R. **Nature's Third Cycle**: A story of sunspots. 1^{st} edition. Oxford, United Kingdom: Oxford University Press, 2015. 5

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1^{st} edition. Cambridge, MA, USA: MIT Press, 2016. 15

GéRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow**: Concepts, tools and techniques to build intelligent systems. 2^{nd} edition. Sebastopol, CA, USA: O'Reilley, 2019. 16, 17

HANSLMEIER, A. **The Sun and Space Weather**. 2^{nd} edition. Dordrecht, The Netherlands: Springer, 2006. (Astrophysics and Space Science Library). 1, 5

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**: Data mining, inference, and prediction. 2^{nd} edition. [S.l.]: Springer, 2017. (Springer Series in Statistics). 16, 17

HENNEY, C. J.; TOUSSAINT, W. A.; WHITE, S. M.; ARGE, C. N. Forecasting f10.7 with solar magnetic flux transport modeling. **Space Weather**, v. 10, n. 2, 2012. Disponível em: <https://doi.org/10.1029/2011SW000748>. 7

HUANG, C.; LIU, D.-D.; WANG, J.-S. Forecast daily indices of solar activity, f10.7, using support vector regression method. **Research in Astronomy and Astrophysics**, IOP Publishing, v. 9, n. 6, p. 694–702, may 2009. Disponível em: <https://doi.org/10.1088/1674-4527/9/6/008>. 6

MURPHY, K. P. **Machine Learning**: A probabilistic perspective. 1^{st} edition. Cambridge, MA, USA: MIT Press, 2012. 16, 17

OKOH, D.; OKORO, E. On the relationships between sunspot number and solar radio flux at 10.7 centimeters. **Solar Physics**, v. 295, n. 1, 2020. Disponível em: <https://doi.org/10.1007/s11207-019-1566-8>. 7

ROYAL OBSERVATORY OF BELGIUM. **Solar Influences Analysis Center**. 2022. Disponível em: <https://www.sidc.be/>. Acesso em: 20 de junho 2022. 5

SCHONFELD, S. J.; WHITE, S. M.; HENNEY, C. J.; ARGE, C. N.; MCATEER, R. T. J. CORONAL SOURCES OF THE SOLAR fsub10.7/subRADIO FLUX. **The Astrophysical Journal**, American Astronomical Society, v. 808, n. 1, p. 29, jul 2015. Disponível em: <https://doi.org/10.1088/0004-637x/808/1/29>. 6

SVALGAARD, L.; HUDSON, H. S. The solar microwave fluxand the sunspot number, insoho-23: Understanding a peculiar solar minimum. **ASP Conf. Ser.**, Astron. Soc. of the Pac., v. 428, p. 325, mar 2010. Disponível em: <https://doi.org/10.48550/arXiv.1003.4281>. 7

TAPPING, K. F. The 10.7 cm solar radio flux (f10.7). **Space Weather**, v. 11, p. 394–406, 2013. Disponível em: <https://doi.org/10.1002/swe.20064>. 6

TOBISKAA, W. K.; WOODSB, T.; EPARVIERB, F.; VIERECKC, R.; FLOYDD, L.; BOUWERE, D.; ROTTMANB, G.; WHITEF, O. R. The solar2000 empirical solar irradiance model and forecast tool. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 62, p. 1233–1250, 2000. 5

UNPINGCO, J. **Python for Probability, Statistics, and Machine Learning**. 1^{st} edition. San Diego, CA, USA: Springer, 2016. 17

WOLF, J. R. Sonnenflecken-beobachtungen in der zweiten hälfte des jahres 1850. **Mittheilungen der Naturforschenden Gesellschaft in Bern**, p. 89–95, 1851. 5