

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO DE JOINVILLE
CURSO DE ENGENHARIA AEROESPACIAL

GABRIELA CAROLINE LIVON LOPES

APRENDIZADO DE MÁQUINA PARA DESENVOLVIMENTO DE MODELO
SEMI-EMPÍRICO DE PREDIÇÃO DE RUÍDO DE MOTOR AERONÁUTICO

Joinville
2021

GABRIELA CAROLINE LIVON LOPES

APRENDIZADO DE MÁQUINA PARA DESENVOLVIMENTO DE MODELO
SEMI-EMPÍRICO DE PREDIÇÃO DE RUÍDO DE MOTOR AERONÁUTICO

Trabalho apresentado como requisito
para obtenção do título de bacharel
em Engenharia Aeroespacial do
Centro Tecnológico de Joinville da
Universidade Federal de Santa Catarina.

Orientador: Dr Rafael Gigena Cuenca

Joinville
2021

GABRIELA CAROLINE LIVON LOPES

APRENDIZADO DE MÁQUINA PARA DESENVOLVIMENTO DE MODELO
SEMI-EMPÍRICO DE PREDIÇÃO DE RUÍDO DE MOTOR AERONÁUTICO

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de bacharel em Engenharia Aeroespacial, na Universidade Federal de Santa Catarina, Centro Tecnológico de Joinville.

Joinville (SC), 10 de Setembro de 2021.

Banca Examinadora:

Prof. Dr. Rafael Gigena Cuenca
Orientador
Presidente

Prof. Dr. Filipe Dutra da Silva
Membro(a)
Universidade Federal de Santa Catarina

Prof. Dr. Pablo Andretta Jaskowiak
Membro(a)
Universidade Federal de Santa Catarina

AGRADECIMENTOS

Agradeço à UFSC pelas inúmeras oportunidades de crescimento profissional e pessoal que me proporcionou.

Agradeço à minha família pelo apoio, em especial ao meu pai por ter me conferido o privilégio de cursar a graduação que tanto desejava.

Agradeço ao Professor Dr. Rafael Gigena Cuenca que me deu a oportunidade de realizar este trabalho, me orientando diligentemente e contribuindo decisivamente para o direcionamento da minha carreira para uma área pela qual tenho tanto apreço.

E por fim, agradeço ao engenheiro aeroacústico Maicon Secchi, meu mentor de estágio, pelos muitos ensinamentos que serão utilizados durante toda a minha carreira, sendo que muitos deles foram aplicados neste trabalho.

RESUMO

O aumento do tráfego aéreo eleva o nível de ruído existente no entornos dos aeroportos, impactando a qualidade de vida da população que habita tais espaços. Nesse contexto, a predição de ruído é uma importante ferramenta de projeto preliminar de aeronaves orientado à atenuação de ruído, visando ao conforto acústico e conformidade com as normas existentes em cada país, que tem se tornado cada vez mais rígidas. Uma das principais fontes de ruído em uma aeronave durante a decolagem é o seu sistema propulsivo, sendo o motor turbo fan predominante na aviação. Com o objetivo de obter ferramentas de predição de ruído adequadas, foi proposta uma modificação da regressão utilizada por Heidmann (1979) para a predição de ruído de banda larga de conjunto rotor estator característico de turbofan. A metodologia utilizada para obter modelo de predição consistiu na aplicação de três técnicas de aprendizado de máquina (*random florest*, *k nearest neighbors* e Multi Layer Perceptron) em dados de ensaio de ruído existentes. Para a seleção de modelo utilizou-se validação cruzada e análise das métricas de erro nos modelos obtidos. Os modelos selecionados como melhores foram os seguintes: *random florest* com erro médio de 0.676, *k nearest neighbors* com erro médio de 0.56 e um modelo que utiliza Multi Layer Perceptron com erro médio de 1.286. Apesar do modelo utilizando Multi Layer Perceptron possuir um erro médio maior e ser mais caro computacionalmente que o modelo *random florest* e KNN, o modelo Multi Layer Perceptron foi considerado mais adequado para a predição devido à suavidade apresentada na predição de nível de pressão sonora total.

Palavras-chave: Aeroacústica. Aprendizado de máquina. Predição de ruído de banda larga. Predição de ruído de motor aeronáutico.

ABSTRACT

The increase in air traffic raises the noise level around airports, impacting the quality of life of the population that inhabits these spaces. In this context, noise prediction is an important tool for aircraft preliminary design oriented to noise attenuation, in a manner that acoustic comfort and compliance with existing legal standards are accounted for, which is important because those standards have become increasingly restricted. One of the main sources of noise in an aircraft during takeoff is the propulsive system, being the turbofan engine the most used. To obtain adequate noise prediction tools, a modification of the regression used by Heidmann (1979) for broadband noise prediction of a turbofan rotor-stator assembly was proposed. The methodology used to obtain a new prediction model consists of the application of three machine learning techniques (random forest, k nearest neighbors and Multi Layer Perceptron) in existing engine noise data. Cross-validation and analysis of error were applied in the models obtained for model selection. The best models selected were the following: random forest with an average error of 0.676, k nearest neighbours with an average error of 0.56 and a model that uses a Multi Layer Perceptron with an average error of 1.286. Although the highest error is found on the Multi Layer Perceptron model and it has a greater computational cost than the other models, the Multi Layer Perceptron was considered more suitable for prediction due to the smoothness presented in the prediction of overall sound pressure level.

Keywords: Aeroacoustics. Machine Learning. Broadband Noise Prediction. Aero-Engine Noise Prediction.

LISTA DE FIGURAS

Figura 1 – Gráfico do nível de pressão sonora indicando a componente de ruído de banda-larga e tonal.	15
Figura 2 – Representação da instalações de ensaio acústico.	16
Figura 3 – Representação esquemática do funcionamento da Random forest. .	19
Figura 4 – Representação esquemática do funcionamento da Multi Layer Perceptron com uma camada escondida.	20
Figura 5 – Representação esquemática da validação cruzada aplicada (10 fold).	21
Figura 6 – Nível de pressão sonora disponibilizado no NASA TM X-71763. . . .	24
Figura 7 – Comparação entre NPS com e sem remoção de NPS harmônico. . .	27
Figura 8 – Fluxograma da metodologia de seleção de modelo.	30
Figura 9 – Aplicação de validação cruzada para diferentes configurações de modelo.	31
Figura 10 – Histograma do coeficiente de determinação dos modelos selecionados quando o treinamento foi realizado utilizando o NPS com e sem remoção de tonal. (a) Modelos treinados com NPS sem remoção de tonal. (b) Modelos treinados com NPS de banda larga.	34
Figura 11 – Predição de NPS sem remoção de tonal (dados total) e NPS de banda larga (dados banda larga) utilizando o modelo de predição Floresta Randômica para o fan B.	35
Figura 12 – Dados para 20 repetições de 10 fold por configuração de modelo de predição. (a) Média aritmética de R^2 , (b) Média aritmética do Erro médio.	36
Figura 13 – Histograma da validação cruzada do coeficiente de determinação dos modelos RF, KNN, GP e MLP selecionados.	37
Figura 14 – Comparação entre predições de NPS realizada com modelos selecionados, dados de ensaio e modelo de predição NASA TM X-71763.	38
Figura 15 – Comparação entre o modelo MLP e RF para o Fan A, com a diretividade de 10 graus.	39
Figura 16 – Nível de pressão sonora global, predito por 4 modelos selecionados, para o fan A.	40
Figura 17 – Nível de pressão sonora global predito por modelos MLP e RF selecionados, em diferentes velocidades percentuais, para o fan A.	41

LISTA DE TABELAS

Tabela 1 – Informações dos conjunto fan/estator.	25
Tabela 2 – Número de árvores utilizados nas diversas configurações de modelo utilizando a técnica <i>random forest</i>	28
Tabela 3 – Número de vizinhos e funções de ponderação utilizados nas diversas configurações de modelo utilizando a técnica <i>k nearest neighbors</i>	28
Tabela 4 – Número de neurônios e de camadas utilizados nas diversas configurações de modelo utilizando a técnica <i>Multi Layer Perceptron</i>	29
Tabela 5 – Métricas de erro obtidas na validação cruzada de cada configuração e tempo de processamento da validação cruzada.	33

LISTA DE SIGLAS

ANNOP Aircraft noise Prediction Program

FAA Federal Aviation Administration

ICAO Organização da Aviação Civil Internacional

KNN k nearest neighbors

MLP Multi Layer Perceptron

NPS Nível de Pressão Sonora - Sound Pressure Level (SPL)

OASPL Nível de Pressão Sonora Global

RF Random Forest

rps Rotação por segundo

RSS Distância rotor/estator

LISTA DE SÍMBOLOS

ΔT	Aumento de temperatura total no fan/compressor
ΔT_0	Valor de referência do aumento de temperatura total no fan/compressor (0.555 K)
f_b	Frequência fundamental
f_i	Predição no ponto i
γ	Coeficiente de expansão adiabática do ar (1.4)
\dot{m}	Fluxo mássico através do fan/compressor
\dot{m}_0	Valor de referência do fluxo mássico através do fan/compressor (0.453 kg/s)
M_{TR}	Número de Mach de ponta de pá relativo na entrada do rotor
$(M_{TR})_D$	Número de Mach de ponta de pá relativo na entrada do rotor no ponto de design
n_{blade}	Número de pás
n	Número de pontos do conjunto de dados.
rps	Rotação por segundo
R^2	Coeficiente de determinação
n_{harm}	Número do harmônico
T_{amb}	Temperatura ambiente
θ	Diretividade
VEL_{perc}	Velocidade percentual
yt_i	Ponto i do conjunto de dados de teste
\bar{y}	Média dos dados avaliados
y_i	Ponto i do conjunto de dados
P_T	Pressão total

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivo Geral	13
1.1.1	Objetivos Específicos	13
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Espectro de ruído de um motor turbofan	14
2.2	Ensaio e modelo de predição exposto em Heidmann (1979)	16
2.3	Aprendizado de máquina	17
2.3.1	Pré Processamento	18
2.3.2	Regressão	18
2.3.3	<i>Random Forest</i>	18
2.3.4	<i>k nearest neighbors</i>	19
2.3.5	Multi Layer Perceptron	19
2.4	Validação cruzada	20
2.5	Métricas de erro	22
2.5.1	Coeficiente R^2	22
2.5.2	Bias	22
3	MÉTODO	23
3.1	Metodologia de extração de dados e pré processamento	23
3.2	Remoção do ruído harmônico	26
3.3	Treinamento dos modelos	26
3.3.1	<i>Random Forest</i>	27
3.3.2	<i>k nearest neighbors</i>	27
3.3.3	Multi layer Perceptron	28
3.4	Seleção de modelo	29
4	RESULTADOS	32
4.1	Modelos de predição treinados com NPS com e sem remoção de harmônico	32
4.2	Modelos de predição treinados com NPS de banda larga (com remoção de tonal).	35
5	CONCLUSÕES	42
5.1	Trabalhos futuros	42
	REFERÊNCIAS	44

1 INTRODUÇÃO

De acordo com Wolfgang (2011), o ruído excessivo pode reduzir significativamente a qualidade de vida da população, impactando significativamente a saúde humana, levando ao aumento do nível de estresse e até gerando maior risco de doenças cardiovasculares. Tais impactos são mais significativos na vida da população vizinha de aeroportos. Por volta dos anos 70, os critérios de certificação de aeronaves tornaram-se mais rígidos, passando a limitar o ruído externo de aeronaves. É preciso considerar que, historicamente, as cidades se expandiram para as proximidades de aeroportos, ocupando espaços até então pouco habitados. Além disso, devido ao aumento do tráfego de aeronaves e da carga paga transportada, a ICAO (padrão para as principais agências de aviação civil), utilizando o chamado *Balanced Approach*, vem gradativamente restringindo os níveis de ruído aceitos para as aeronaves (ICAO, 2019).

As principais fontes que contribuem para o Ruído no motor aeronáutico são o difusor, a câmara de combustão, o Fan, o compressor, a turbina e o bocal de saída, sendo que, para turbo fans de alto bypass, o conjunto rotor/estator é a maior fonte de ruído. Por isso, as técnicas de predição de ruído para esse conjunto são relevantes como ferramentas de projeto (DOBRZYNSKI, 2010).

De acordo com (WARD, 2011), atualmente os motores mais utilizados em aeronaves comerciais são turbofans, pois possuem maior eficiência energética e são mais silenciosos em comparação aos motores turbojato. No motor turbofan a energia gerada pela combustão é utilizada para rodar as pás do rotor que são circundadas por uma nacela que ocasiona o jato frio, principal responsável pela propulsão. A existência da nacela permite a utilização de liners para atenuação do ruído gerado pelo conjunto rotor/estator, reduzindo o ruído externo.

Para a estimativa de nível sonoro gerado por máquinas e equipamentos, de acordo com Bistafa (2011), podem ser utilizados métodos semi-empíricos, semi-analíticos e computacionais. O método de predição semi-empírico, como o método de Heidmann (1979), utiliza dados de ensaios realizados com diferentes geometrias para realizar predições, possuindo custo computacional muito baixo, limitação de complexidade geométrica, e por ser baseado em medições experimentais, os resultados têm confiabilidade satisfatória.

De acordo com Moreau e Enghardt (2008) e Ventres, Theobald e Mark (1982), os modelos semi-analíticos são uma modelagem teórica, utilizadas para prever as fontes e propagação do ruído. Dahl (2012) cita alguns exemplos de códigos semi-analíticos utilizados pela NASA para previsão de ruído de fan, sendo eles: V072, TFanNS, RSI, BFanNS e Superpose. Além disso, Dahl escreve que a quantidade de

informação necessária para usar tais técnicas, assim como o custo computacional são muito maiores que nas técnicas semi-empíricas. Tanigawa, Yamasaki e Ooishi (2009) e Brochine (2013) utilizam simulações de CFD com formulação URANS, para modelar a turbulência, obtendo o escoamento em um conjunto rotor/estator e utiliza a técnica semi-analítica para resolver as fontes acústicas e compor o espectro dos motores.

Já a predição de ruído aeroacústico por técnicas computacionais, resolve simulações fluido dinâmicas, para computar o ruído utilizando, utilizando por exemplo, métodos Lattice Boltzman para a predição de ruído e a formulação de Farassat, ou analogia de Ffowck-Williams and Hawkings, para computar a propagação do som, como no estudo de Keating et al. (2015), onde o programa PowerFLOW é utilizado para a predição do ruído de trem de pouso. Para tais simulações, foi necessário alto poder computacional e informações detalhadas da geometria, consumindo um tempo de 21 horas de simulação para as malhas grosseiras e 195 horas para as malhas mais refinadas. Entre os métodos apresentados, os métodos semi-empíricos, como o de Heidmann, podem ser melhorados com aplicação de técnicas mais modernas, como demonstrado por Guimarães (2017), em sua modernização de predição, realizada a partir do princípio de similaridade e ampliação de base de dados utilizada. De acordo com o mesmo autor, tal princípio implica que fans similares possuem espectros de ruído similares e é aplicado para reduzir o erro em casos mal representados pelo modelo original. Isso motiva a experimentação de técnicas mais modernas no método de Heidmann, por exemplo, o uso de inteligência artificial (IA).

Inteligência artificial é amplamente utilizada para predição e classificação em diversas áreas, como no trabalho de Libbrecht (2015) que apresenta aplicações de IA em genética, entre outros. Essa diversidade de aplicações ocorre devido ao relativo baixo custo computacional e versatilidade de aplicação dos algoritmos. Como exemplo de técnica de IA utilizada para predição de ruído, temos a regressão Kriging, aplicada por Cuenca (2017) para previsão de ruído de um modelo de um rotor/estator em escala. As técnicas de regressão utilizadas neste trabalho são: rede neural, árvore de decisão, *Random Forest* e *k Nearest Neighbors*. Para a obtenção de métodos utilizando tais técnicas foram utilizadas funções em python do pacote scikit-learn, que, como apontado por Pedregosa et al. (2014), permite a utilização de aprendizado de máquina para fins comerciais e científicos por não especialistas. Inspirado por tais trabalhos, este estudo propõem uma metodologia que consiste na extração dos dados obtidos por Heidmann, disponíveis no seu trabalho (1979), aplicando neles diferentes métodos de aprendizado de máquina. Sendo que proposta de utilização de mais de uma técnica de regressão tem como propósito identificar a que melhor se adapta à aplicação desejada. E por fim, a comparação dos modelos utilizando aprendizado de máquina com o modelo obtido por Heidmann.

1.1 OBJETIVO GERAL

Utilizar dados de ensaio de ruído publicados por Heidmann (1979) para obtenção de novos modelos de predição de ruído de banda-larga de motores turbofan, aplicando diferentes técnicas de regressão utilizadas em aprendizado de máquina, buscando diminuir as limitações do método proposto por Heidmann (1979).

1.1.1 Objetivos Específicos

- Utilização de diferentes técnicas de aprendizado de máquina;
- Obtenção de ferramenta de predição de ruído de motores turbofan para projeto conceitual;
- Verificar se a aplicação é vantajosa relacionada ao modelo tradicional.

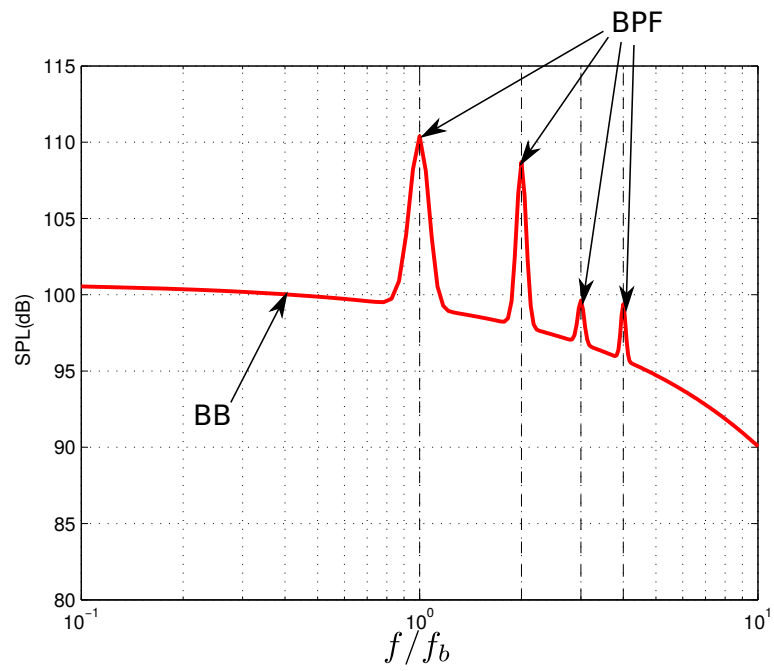
2 FUNDAMENTAÇÃO TEÓRICA

2.1 ESPECTRO DE RUÍDO DE UM MOTOR TURBOFAN

O ruído gerado por rotores é composto pela componente de NPS (Nível de Pressão Sonora) tonal e de banda larga. A componente tonal do ruído consiste em picos de NPS resultantes do deslocamento do ar causado pela rotação do *fan*, sendo que no espectro pode-se observar picos de NPS nos números do harmônico (n_{harm}). O número do harmônico é obtido pela razão entre a frequência (f) e a frequência fundamental (f_b), sendo que a frequência fundamental é representada pela Equação 1 como o produto do número de pás do rotor (n_{blade}) com a rotação por segundo do rotor (rps). Na Figura 1, pode-se observar os picos de NPS (ou em inglês Sound Pressure Level (SPL)) descritos, estando destacados como *BPF* (frequência fundamental, ou em inglês *Blade Passing Frequency*). Para simplificar, o eixo da frequência é normalizado pela f_b ($\frac{f}{f_b}$), dessa forma, os inteiros da frequência normalizada coincidem com os tons harmônicos da BPF. A Figura 1 também permite observar a outra componente de NPS citada: NPS de banda larga. Esta componente de NPS está indicada como BB (banda larga, ou em inglês *broadband*), sendo que pode-se observar a forma do espectro do BB resultante da característica de distribuição aleatória da intensidade do ruído. É importante ressaltar que este trabalho tem o objetivo de prever apenas a componente de NPS de banda-larga.

$$f_b = n_{blade} \cdot rps \quad (1)$$

Figura 1 – Gráfico do nível de pressão sonora indicando a componente de ruído de banda-larga e tonal.

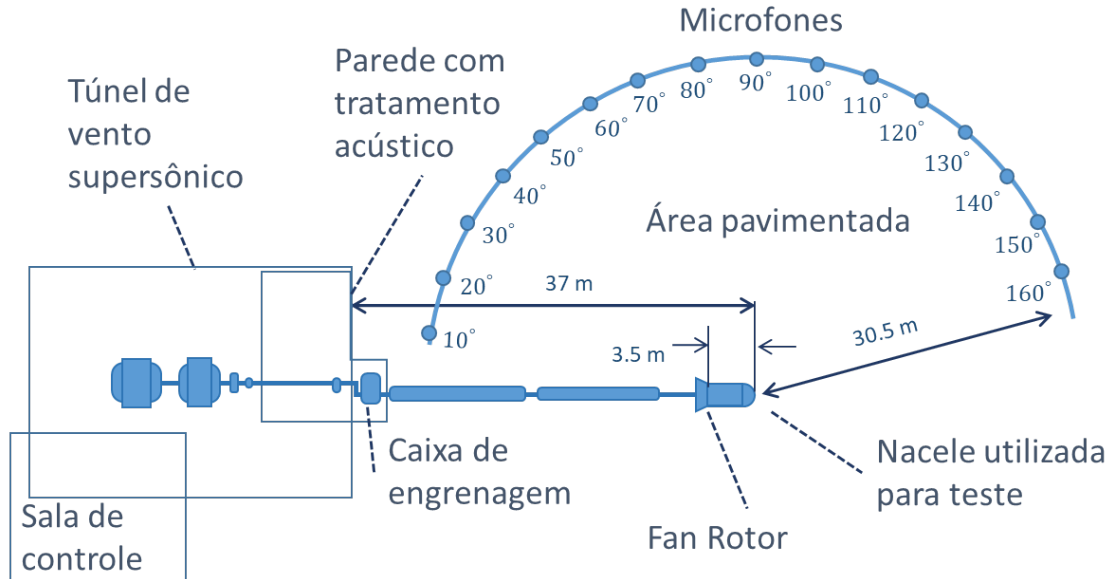


Fonte: Cuenca (2017)

2.2 ENSAIO E MODELO DE PREDIÇÃO EXPOSTO EM HEIDMANN (1979)

O memorando técnico NASA TM X-71763 (Heidmann (1979)) se baseia em medições experimentais, disponibilizadas em gráficos de nível de pressão sonora por frequência, de modelos em escala 1:1 de conjuntos Rotor/Estator. As medições são feitas por microfones à 100 pés de distância do fan em 16 ângulos de diretividade distintos, como mostra a Figura 2. Para tal, os modelos de conjunto Rotor/estator operaram em 4 diferentes velocidades de rotação. Na Figura 2, também pode-se observar o isolamento acústico do prédio onde o motor responsável pela movimentação do rotor estava instalado. De acordo com Heidmann e Feiler (1973) as paredes das instalações eram as únicas superfícies refletoras, além do solo, que poderiam afetar as medições, com uma distância de cerca de 120 pés em relação ao rotor, e tais paredes foram cobertas com 6 polegadas de espuma de poliuretano para evitar as reflexões sonoras. Os dados para cada condição foram medidos 3 vezes para permitir o cálculo da média como resultado final, em banda de um terço de oitava para a faixa de frequência entre 50 e 20000 hz. Os dados medidos foram corrigidos para temperatura e umidade padrão de 59° Farenheit e 70% de umidade relativa.

Figura 2 – Representação da instalações de ensaio acústico.



Fonte: Adaptado de Heidmann e Feiler (1973)

No começo da década de 70, Dunn e Peart (1973) desenvolveram um modelo de predição de ruído baseados em dados coletados na mesma época e publicados por Heidmann e Feiler (1973). No estudo de Dunn e Peart, os modelos são desenvolvidos para vários tipos de aeronaves: turbojato, turboprop, turbofan, V/STOL e helicópteros. Mais para o final da década, Heidmann (1979) utilizou a base de informações

complementada por novos estudos (HEIDMANN, 1979) sobre o ruído de motores como base de seu estudo, porque, além de ser um estudo extensivamente documentado, utilizou dados de motores em escala real.

No trabalho de Heidmann (1979), os dados das medições foram utilizados para definir uma regressão de ruído de Broadband baseada em um ajuste de curvas considerado adequado. A regressão proposta para o NPS está apresentada na equação 2 como composta por dois termos F_4 e L_c . A Equação 3 apresenta a função (F_4) que define o formato do espectro de NPS no domínio da frequência, pela equação, podemos concluir que o NPS terá um formato logarítmico que utiliza as variáveis frequência (f) e frequência fundamental (f_b). Já L_c não afeta o formato da curva, apenas seu valor absoluto, sendo tal função representada pela Equação 4, que é composta por 5 termos. Os dois primeiros termos, dependem de: ΔT , o aumento de temperatura total no fan/compressor, ΔT_o , valor de referência de ΔT (igual à 0.555 K), \dot{m} é vazão mássica através do fan/compressor e seu valor de referência, $\dot{m}_o = 0.453kg/s$. Para obter detalhes sobre as funções F_1 , F_2 e F_3 é necessário consultar a referência (HEIDMANN, 1979), mas é relevante observar quais são os parâmetros utilizados em tais funções. Uma das variáveis é M_{TR} , o número de Mach de ponta de pá (a razão entre a velocidade de ponta de pá e a velocidade do som) relativo de entrada do rotor, o $(M_{TR})_D$, que é o M_{TR} no ponto de design e, por fim, θ , que é o ângulo de diretividade.

$$NPS(f) = L_c + F_4 \left(\frac{f}{f_b} \right) \quad (2)$$

$$F_4 = 10 \log_{10} \left(e^{-\frac{1}{2}} \left[\frac{\ln \left(\frac{f}{2.5f_b} \right)}{\ln 2.2} \right] \right) \quad (3)$$

$$L_c = 20 \log_{10} \left(\frac{\Delta T}{\Delta T_o} \right) + 10 \log_{10} \left(\frac{\dot{m}}{\dot{m}_o} \right) + F_1 [(M_{TR}), (M_{TR})_D] + F_2(RSS) + F_3(\theta) \quad (4)$$

2.3 APRENDIZADO DE MÁQUINA

Métodos de regressão, de acordo com Harrington (2012), são amplamente utilizados para a previsão de dados. É particularmente útil quando o objeto de estudo não tem forma analiticamente descrita e se trata de um grande banco de dados. Para problemas onde existe um banco de dados definido e se deseja obter um resultado à partir de determinadas variáveis, deve-se utilizar técnicas de regressão de aprendizado de máquina supervisionado. Antes da aplicação da(s) técnica(s) de regressão, os dados devem ser pré processados.

2.3.1 Pré Processamento

O Pré Processamento deve ser realizado antes da regressão, e sua importância depende do banco de dados e regressão à serem utilizados. Uma etapa do pré processamento é a determinação das variáveis relevantes e pode ser realizada com base na teoria do fenômeno estudado ou, como detalhado por Kuhn (2019), empiricamente por meio de um teste de sensibilidade. Neste teste, o modelo é treinado com todas as variáveis que podem ser utilizadas, e compara-se o erro obtido por este modelo, e por um treinado após a retirada de uma ou mais variáveis do grupo de treinamento. Com tal comparação, pode-se observar quais variáveis são mais relevantes para a predição adequada. A última etapa do pré processamento é uma normalização, onde as variáveis são normalizadas para não haver uma diferença muito significativa de ordem de grandeza. De acordo com Géron (2017), essa normalização é essencial para a boa performance da maioria dos modelos de aprendizado de máquina, e deve ser baseada apenas nos dados do conjunto de treinamento, mas aplicada também no conjunto de teste.

2.3.2 Regressão

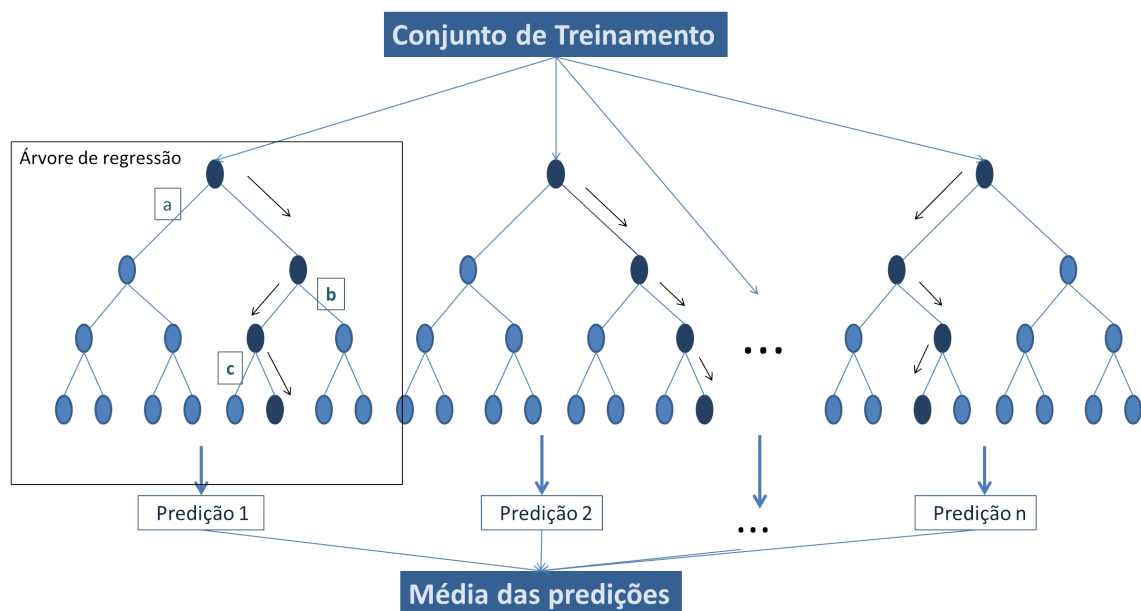
Existem diversas técnicas de regressão, sendo que para cada técnica, o desempenho depende do tipo do fenômeno estudado. Um exemplo claro, é que técnicas que realizam boas predições em casos lineares, não necessariamente realizam boas predições em casos não lineares. Por esse motivo, de acordo com Ahmed et al. (2010) é interessante aplicar mais de uma técnica de modelagem a um determinado problema, com a finalidade de utilizar métricas de erro para treinar o modelo mais adequado para a predição do fenômeno estudado. Para definir o modelo mais adequado, é importante observar tanto as métricas de erro para o grupo de treinamento, quanto se não ocorre *overfitting* (resultado com ótima predição do grupo de treinamento, porém péssima para o grupo de teste). Cada uma das técnicas de modelagem à seguir, apresentam diferentes performances de acordo com as configurações utilizadas para o treinamento. Sendo que as configurações foram definidas com o intuito de observar a influência do aumento da complexidade dos modelos, ou da variação de parâmetros para os modelos obtidos.

2.3.3 *Random Forest*

A técnica de regressão *random forest* é uma técnica de *ensemble* que combina mais de uma árvore de regressão, como está representado no esquema da Figura 3, e utiliza a média das predições de cada árvore. A representação de árvores de regressão utilizadas no *random forest* também pode ser observada na Figura 3. Dentro do retângulo em preto, observa-se nós com ramificações. Cada um dos nós indicados

pela letra 'a', 'b' e 'c', representa a aplicação de critérios que irão reduzir os possíveis resultados à serem obtidos pela regressão. O treinamento consiste no teste de possíveis critérios e verificação de sua adequação utilizando métricas de erro. A árvore de regressão, de acordo com Harrington (2012), possui a vantagem de exigir baixo custo computacional e poder ser utilizada quando há dados ausentes ou irrelevantes no conjunto de dados. Por outro lado, pode existir *overfitting*. *Random forest* possui a vantagem de diminuir a ocorrência de *overfitting* e aumentar a acurácia em relação à árvore de regressão, de acordo com Pedregosa et al. (2011).

Figura 3 – Representação esquemática do funcionamento da Random forest.



Fonte: Autora (2021)

2.3.4 *k nearest neighbors*

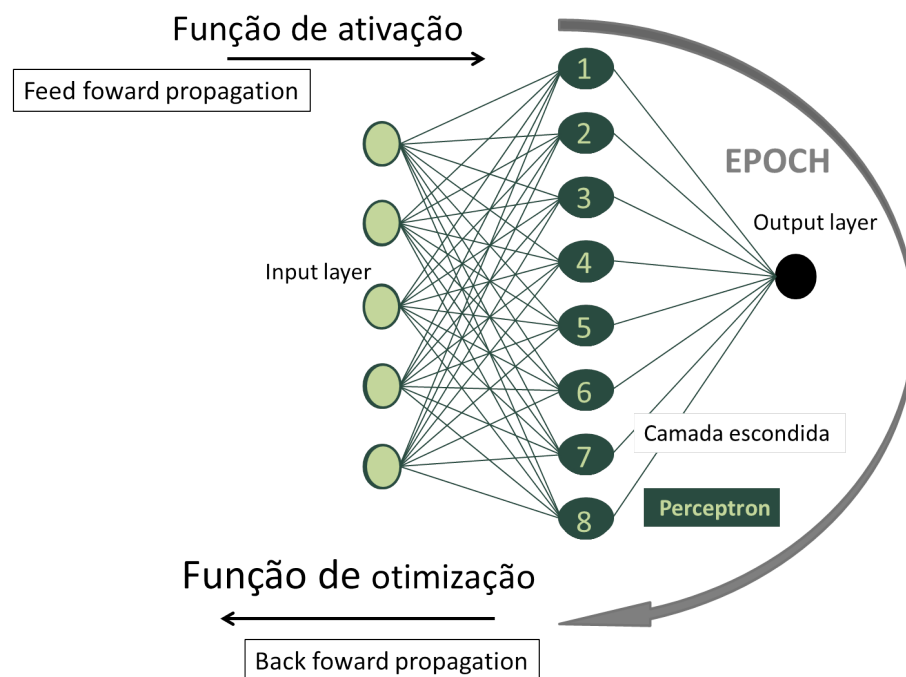
Essa regressão utiliza um número definidos de pontos próximos e uma função de peso, como peso uniforme para cada ponto vizinho ou o inverso da distância do ponto vizinho para realizar o treinamento. De acordo com Timbers, Campbell e Lee (2021) esse modelo representa bem comportamentos de função não lineares, mas pode não representar bem problemas com muitas variáveis e valores fora da faixa de valores do conjunto de dados de treinamento.

2.3.5 Multi Layer Perceptron

A Figura 4 contém a representação esquemática de uma *Multi Layer Perceptron* (MLP) com uma camada escondida e 8 perceptrons, sendo que, segundo Galushkin (2007), um perceptron recebe inputs, aplica a estes uma função de ativação, e obtém

um *output*. Caso exista mais de uma camada escondida, os *outputs* da camada anterior se tornam *inputs* da camada a seguir. Quando se obtém o *output* da última camada escondida, terminou o *feed foward propagation*. Então, é calculado o erro e utilizada uma função de otimização para definir quais serão as alterações dos pesos atribuídos. O processo que utiliza a função de otimização de camada para camada escondida é chamado *back foward propagation*. O processo de *feed* e *back foward* é chamado *epoch* e o número de *epochs* pode ser definido pelo usuário, ou pela estagnação do erro obtido.

Figura 4 – Representação esquemática do funcionamento da Multi Layer Perceptron com uma camada escondida.



Fonte: Autora (2021)

2.4 VALIDAÇÃO CRUZADA

A seleção do modelo mais adequado, idealmente, é realizada após a implementação de vários modelos e seu cálculo de erro, utilizando validação cruzada. Sendo a validação cruzada útil para analisar se está ocorrendo *overfitting* e *underfitting*. De acordo com Harrington (2012), *overfitting* pode ser definido como um treinamento muito adequado para uma fração do conjunto de pontos, porém, não representa adequadamente o fenômeno. Uma forma de reduzir esse problema, é limitando o número de variáveis utilizadas. Também de acordo com Harrington (2012), *underfitting*, ao contrário de *overfitting*, a predição fica muito distante dos valores originais, também não representando de forma adequada o fenômeno. o autor sugere trocar o modelo de

regressão utilizado quando existir *overfitting*. Para detectar esses problemas, de acordo com Picard e Cook (1984), pode-se usar como validação cruzada, as estimativas de erro *Leave one out*, *kfold* e Montecarlo.

Kfold é o modelo de validação cruzada no qual o conjunto de pontos é dividido em grupo de treinamento e grupo de testes, como apresentado na figura 5. Os pontos designados para cada grupo são permutados em um determinado número de combinações e é obtido o erro de cada conjunto. Picard e Cook (1984) também concluem que utilizar 10 pastas resulta em uma boa análise sem custo computacional muito elevado. Dessa forma, neste trabalho, será utilizado o 10 fold.

Figura 5 – Representação esquemática da validação cruzada aplicada (10 fold).



Fonte: Autora (2021)

2.5 MÉTRICAS DE ERRO

De acordo com Montgomery e Runger (2009), usualmente, dois valores são utilizados para caracterizar uma distribuição de probabilidades: a média e a variância. Sendo que a variância de um conjunto de dados pode ser representada pelo coeficiente R^2 e a média pelo bias. A importância de analisar a variância e o bias, na avaliação de modelos de aprendizagem de máquina, é exposta por Harrington (2012), relatando que há um ponto ótimo em que a complexidade diminui o bias e a variância. Após o ponto ótimo, o aumento da complexidade do modelo pode continuar diminuindo o erro, porém, ao mesmo tempo, aumentando a variância.

2.5.1 Coeficiente R^2

O coeficiente de determinação expressa a variância do erro que varia de $-inf$ a 1, sendo 1 o valor representando pontos iguais para o conjunto de dados e para o resultado da predição. Uma característica deste parâmetro, de acordo com Heij et al. (2004), é que o valor do coeficiente independe da escala dos dados e quanto menor, maior o erro.

$$\begin{aligned}
 R^2 &= 1 - \frac{SS_{res}}{SS_{tot}} & (5) \\
 SS_{tot} &= \sum_i (yt_i - \bar{y})^2 \\
 SS_{res} &= \sum_i (yt_i - f_i)^2
 \end{aligned}$$

Sendo que yt_i é o ponto i do conjunto de dados de teste, \bar{y} é a média dos dados avaliados e f_i é a predição no ponto i .

2.5.2 Bias

De acordo com Ahmed et al. (2010) o bias é a diferença entre o erro esperado e o erro medido, permitindo observar se há *underfitting*. Em um modelo de predição de ruído, essa métrica permite estimar o erro obtido na predição de NPS.

$$Bias_{medio} = \frac{1}{n} \cdot \sum_i (y_i - f_i) \quad (6)$$

Sendo que y_i é o ponto i do conjunto de dados e n é o número de pontos do conjunto de dados.

3 MÉTODO

A metodologia consiste na obtenção, normalização dos dados, treinamento e aplicação de validação cruzada para diferentes tipos de regressão. Também foi realizada análise dos resultados em comparação com o modelo estabelecido em NASA TM X-71763, e por fim, proposição de modelo para predição de ruído de banda larga.

3.1 METODOLOGIA DE EXTRAÇÃO DE DADOS E PRÉ PROCESSAMENTO

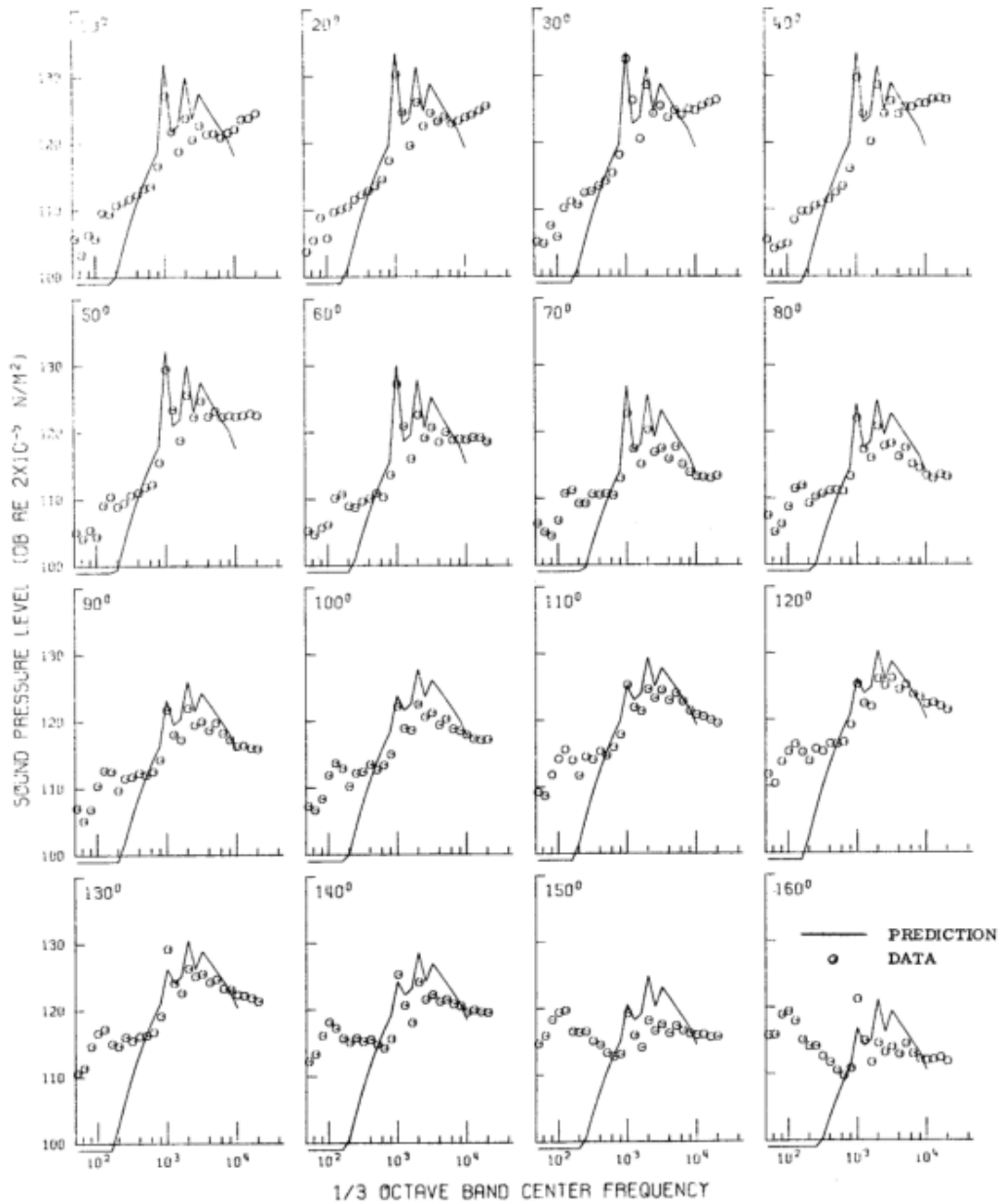
Neste trabalho, tais dados foram organizados na forma de um banco de dados tabelado para a utilização em novos ajustes de curvas, não mais definindo a forma das Equações 2 e 4, mas sim utilizando algoritmos de aprendizado de máquina para definir a forma que melhor se ajusta aos dados. Os dados dos experimentos realizados como parte do programa ANOPP foram disponibilizados em gráficos por Heidmann (1979), como exibido na Figura 6, para frequência de 50 à 20000 Hz em 1/3 de banda de oitava. Como os dados estavam disponibilizados publicamente apenas nos gráficos, utilizou-se a ferramenta online desenvolvida por Digitalizer (10 de ago. de 2019.) para extrair dados das figuras. As principais informações das condições dos experimentos estão disponibilizadas na Tabela 1. A primeira coluna apresenta os nomes atribuídos a cada um dos 8 *fans* utilizados, e a última coluna apresenta as velocidades percentuais utilizados nos experimentos para cada *fan*. Para cada *fan*, em cada velocidade percentual, foram realizadas medições em 16 ângulos de diretividade (Figura 2) e cada medição resultou em 27 valores NPS (correspondentes às frequências 50 à 20000 Hz em 1/3 de banda de oitava). Sendo assim, o número de objetos adquiridos para a predição é 13824. Cada objeto tem 5 *features* (variáveis) para a predição do NPS: θ (ângulo de diretividade), f (frequência), M_{TR} (número de Mach de ponta de pá relativo na entrada do rotor), $(M_{TR})_D$ (M_{TR} no ponto de design) e RSS (distância rotor/estator).

Heidmann (1979) afirma que a normalização reduz a complexidade das correções que serão aplicadas, ao reduzir o número de variáveis do problema, ao mesmo tempo que não afeta a precisão do método. De acordo com o autor, o critério de definição dos parâmetros fluxo mássico (\dot{m}) e variação de temperatura (ΔT), foi a conveniência dessas variáveis já serem frequentemente utilizadas no estudo de desempenho de motores turbofan. Dessa forma, aplica-se a normalização:

$$NPS_{normalizado} = NPS - 20 \cdot \log \left(\frac{\Delta T}{\Delta T_o} \right) - 10 \cdot \log \left(\frac{\dot{m}}{\dot{m}_o} \right) \quad (7)$$

onde ΔT é a variação de temperatura, obtida pela equação:

Figura 6 – Nível de pressão sonora disponibilizado no NASA TM X-71763.



(f) Fan B, configuration 108, 70-percent speed.

Fonte: Heidmann (1979)

$$\Delta T = \left(\frac{n_{blade}}{f_b} \cdot P_T^{\frac{\gamma-1}{\gamma}} \cdot \frac{rps}{VEL_{perc}} \right) \cdot T_{amb}$$

onde P_T é a pressão total, \dot{m} é o fluxo mássico de ar que passa pelo rotor, f_b é a frequência fundamental do rotor, rps é a rotação nominal do eixo, em revoluções por segundo e VEL_{perc} é a velocidade percentual do rotor. Os valores desses parâmetros, para cada motor, estão apresentados na Tabela 1. O valor para a temperatura ambiente (T_{amb}) é 288.15 K e o coeficiente de expansão adiabática do ar (γ) é 1.4.

Tabela 1 – Informações dos conjunto fan/estator.

Fan	Pressão total	Vazão mássica, [kg/s]	Mach tip	Mach tip design	Rotação por segundo, [rps]	Número de pás	RSS - Distância rotor/estator, [%]	Razão de Solidez	Diâmetro do fan, [m]	Velocidade, [%]
A	1.5	430	1.04	1.2	61	40	200	1.45	1.86	0.6 0.7 0.8 0.9
B	1.5	430	1.04	1.2	60	26	200	1.3	1.86	0.6 0.7 0.8 0.9
C	1.6	415	1.39	1.52	87	26	200	1.4	1.74	0.6 0.7 0.8 0.9
QF-1	1.5	396	0.99	1.12	59	53	367	1.34	1.82	0.6 0.7 0.8 0.9
QF-3	1.4	396	0.99	1.12	59	53	367	1.34	1.82	0.6 0.7 0.8 0.9
QF-5	1.6	385	0.98	1.14	61	36	227	1.38	1.75	0.6 0.7 0.8 0.85
QF-6	1.2	396	0.67	0.88	40	42	400	1.19	1.82	0.6 0.7 0.8 0.9
QF-9	1.2	403	0.63	0.87	37	15	200	0.89	1.84	0.6 0.7 0.86 0.93

3.2 REMOÇÃO DO RUÍDO HARMÔNICO

Como este trabalho tem por objetivo propor um novo modelo para o ruído de banda larga, objetiva-se avaliar se é necessário filtrar ou retirar os harmônicos dos espectros similares aos apresentados na seção 2.1. Tal separação pode ser realizada de diversas formas, por exemplo, Sree e Stephens (2016), aplicam no sinal do ruído uma correlação para alinhar as componentes tonais em cada período, permitindo filtrar o ruído aleatório de banda larga do ruído tonal. Como no presente trabalho os dados de nível de pressão sonora foram obtidos apenas no domínio da frequência, não é possível aplicar o método proposto por Sree e Stephens (2016). Então, a abordagem de separação dos tonais foi realizada no domínio da frequência.

Seguindo a Equação 1, foram encontrados os valores de frequência correspondente aos 11 harmônicos (n_{harm}) iniciais, sendo que em tais bandas de frequência o ruído tonal está presente. Por isso, foi verificado em qual banda de 1/3 de oitava cada harmônico pertencia e foi realizada a média do valor de NPS da banda seguinte e da banda anterior à banda que continha o valor de NPS harmônico, como indicado na Equação 8. A Figura 7 apresenta a diferença entre os dados total e banda larga obtidos desta forma.

$$NPS(i) = \frac{NPS(i-1) + NPS(i+1)}{2} \quad (8)$$

$$n_{harm} = 1, 2, 3, \dots, 11.$$

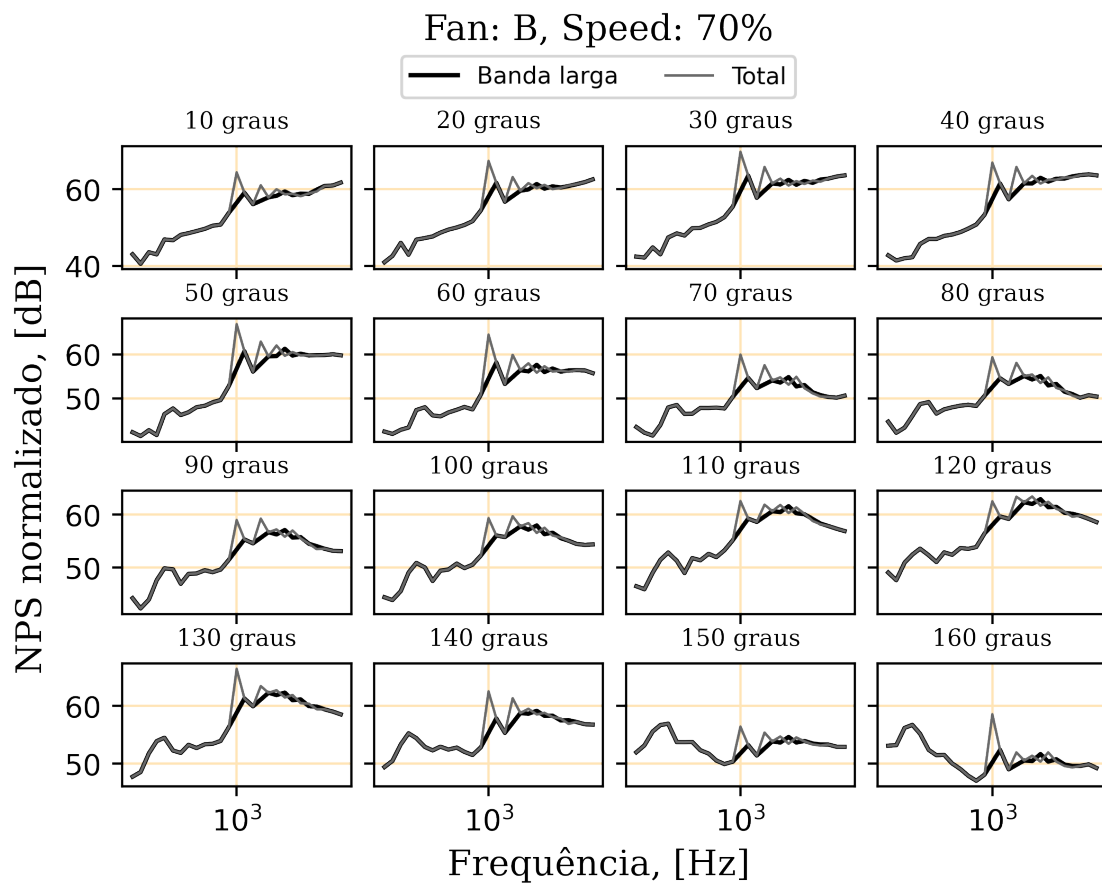
$$\forall i = f_b \cdot n_{harm}$$

A metodologia de extração não se mostrou satisfatória apenas para 10 espectros de um fan em uma velocidade percentual, então, para tais condições, os valores dos picos observados foram substituídos por valores referentes aos pontos vizinhos. Após a separação das componentes de ruído, foram treinados modelos com os conjuntos de dados, com e sem separação, com o intuito de verificar se as diferentes abordagens iriam apresentar diferença significativa na representação dos tonais.

3.3 TREINAMENTO DOS MODELOS

São aplicadas quatro técnicas de regressão pra treinamento: *random forest* (RF), *k nearest neighbors* (KNN) e Multi Layer Perceptron (MLP). Tais modelos são treinados utilizando 20 repetições de 10 fold cada, sendo uma das divisões o conjunto de dados utilizado pra teste e os outros conjuntos de dados utilizados para treinamento. Estas técnicas de regressão foram aplicadas utilizando diferentes configurações, que estão detalhadas abaixo.

Figura 7 – Comparação entre NPS com e sem remoção de NPS harmônico.



Fonte: Autora (2021)

3.3.1 *Random Forest*

As configurações de *random forest* foram utilizadas variando o número de árvores de regressão. Esta variação tem como objetivo observar as métricas de erro com o aumento da complexidade do modelo (quanto maior o número de árvores, mais complexo). É esperado que os modelos mais complexos tenham melhor desempenho. O número de árvores utilizado variou de 2 a 80, como pode ser observado na Tabela 2.

3.3.2 *k nearest neighbors*

As configurações de KNN, detalhadas na Tabela 3, variam entre 3 e 6 para o número de vizinhos utilizados no treinamento. Foram utilizadas funções de ponderação uniforme e distância para cada intervalo de número vizinhos. É esperado que este modelo apresente resultados com boas métricas de erro, sem utilizar muito custo computacional, e que seja possível observar diferença nas métricas de erro, entre as funções que utilizaram a função de ponderação uniforme e distância. Isso porque, as frequências estão em banda de 1/3 de oitava, aumentando consideravelmente as

Tabela 2 – Número de árvores utilizados nas diversas configurações de modelo utilizando a técnica *random forest*.

Configuração	Número de árvores
1	2
2	5
3	7
4	10
5	15
6	20
7	30
8	40
9	50
10	60
11	70
12	80

distâncias para as frequências mais altas, de forma que há diferença considerável dessa ponderação para a ponderação uniforme.

Tabela 3 – Número de vizinhos e funções de ponderação utilizados nas diversas configurações de modelo utilizando a técnica *k nearest neighbors*.

Configuração	Ponderação	Número de vizinhos
13	distância	3
14	distância	4
15	distância	5
16	distância	6
17	uniforme	3
18	uniforme	4
19	uniforme	5
20	uniforme	6

3.3.3 Multi layer Perceptron

O Multi layer Perceptron (MLP) é utilizado com a função de ativação tanh e função de otimização lbfgs. Tais funções, foram selecionadas após verificar o coeficiente R^2 de testes realizados com algumas funções de ativação e de otimização. Tal verificação foi realizada utilizando validação cruzada, e não está documentada neste trabalho. Na implementação utilizada o critério de parada é o *Early stopping*, quando o erro obtido no treinamento aumenta duas *epochs* seguidas, o treinamento é encerrado, retornando para os parâmetros antes do erro começar a aumentar. A Tabela 4 apresenta o número de perceptrons e camadas escondidas de cada configuração de MLP analisada. É esperado que esse modelo apresente bons resultados, desde que a

configuração do modelo tenha a complexidade adequada. Também é esperado que o custo computacional seja alto.

Tabela 4 – Número de neurônios e de camadas utilizados nas diversas configurações de modelo utilizando a técnica *Multi Layer Perceptron*.

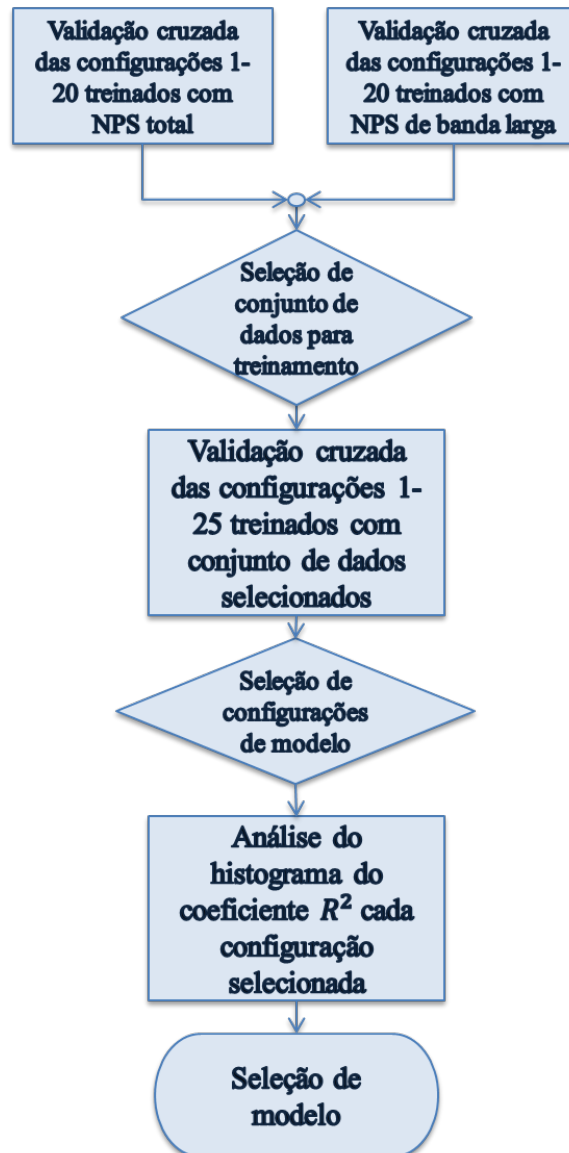
Configuração	Número de Perceptrons	Número de camadas ocultas
21	8	1
22	10	1
23	12	1
24	8	2
25	40	1

3.4 SELEÇÃO DE MODELO

A metodologia de seleção de modelo está representada pelo fluxograma na figura 8. Primeiramente, a validação cruzada é realizada com diferentes conjuntos de dados de treinamento: NPS total (sem remoção de harmônico) e NPS de banda larga (com remoção de harmônico). A validação cruzada está esquematizada na figura 9, utiliza o 10 fold e permite obter a média do Bias e R^2 dos 200 modelos. Tal validação é realizada para diversas configurações da técnica kNN e RF (1-20), pois estas são baratas, computacionalmente, se comparadas com configurações da técnica MLP. A comparação dos resultados obtidos não necessariamente representa o comportamento do MLP. No entanto, o melhor desempenho em 2 das 3 regressões é considerado suficiente para selecionar qual conjunto de dados será utilizado para a validação cruzada de todas as configurações (1-25). Para cada configuração selecionada é obtido um histograma do coeficiente R^2 , observando o seu desvio padrão, e entre os 200 modelos obtidos para a configuração, é selecionado um destes modelos treinados para realizar predição de ruído e comparar com os dados originais.

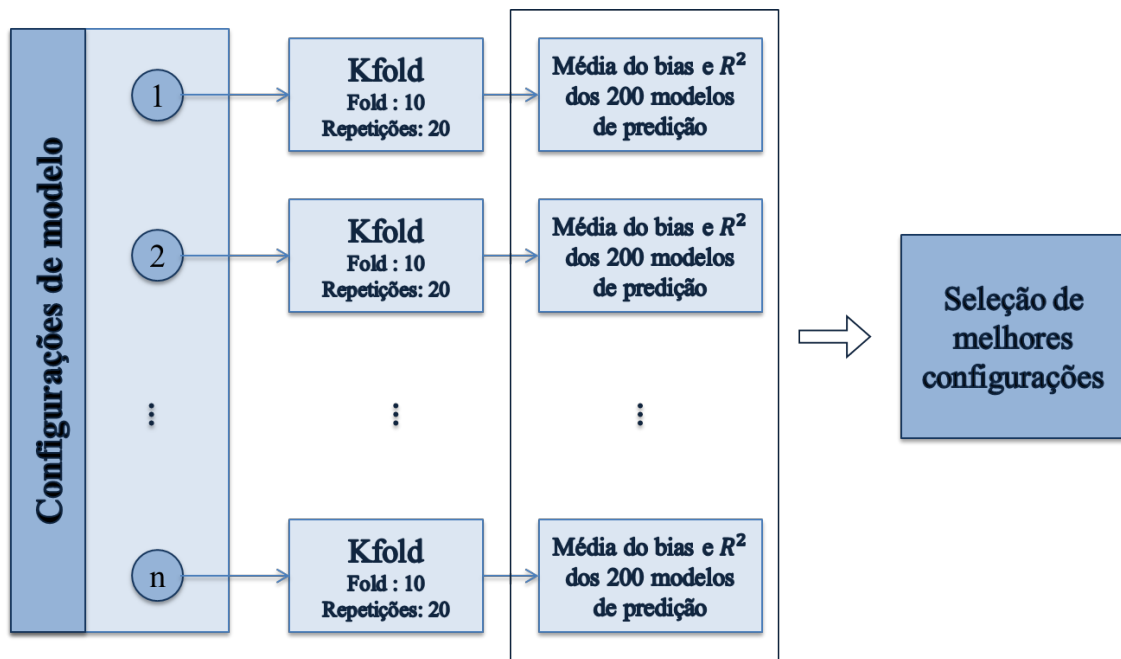
Para comparar as diferentes configurações de cada técnica de regressão, detalhadas nas Tabelas 2, 3 e 4, foi aplicado o 10 fold para cada configuração de regressão como esquematizado na figura 9. Como também pode ser observado em tal figura, a média do Bias obtido em cada um dos 200 modelos, é utilizada para comparação da performance dos modelos.

Figura 8 – Fluxograma da metodologia de seleção de modelo.



Fonte: Autora (2021)

Figura 9 – Aplicação de validação cruzada para diferentes configurações de modelo.



Fonte: Autora (2021)

4 RESULTADOS

Os resultados obtidos consistem na comparação dos modelos treinados utilizando o NPS normalizado total (sem remoção do harmônico) e NPS normalizado de banda larga; comparação de diversos modelos treinados utilizando NPS de banda larga; seleção de configuração e comparação com o modelo de predição de ruído de banda larga NASA TM X-71763.

Para o desenvolvimento deste estudo, utilizou-se um computador modelo LAPTOP-4FS2BLIT, com um processador Intel(R) i7-8565U de 8 núcleos com velocidade de 1.99 GHz e 8GB de RAM instalada. A obtenção dos métodos foi feita em linguagem Python, utilizando a biblioteca SciKit-Learn (PEDREGOSA et al., 2014) utilizado no sistema operacional Windows 10 Home. Com o intuito de indicar a diferença de custo computacional entre as configurações, foi disponibilizado na Tabela 5 o tempo aproximado para realizar os treinamentos da validação cruzada.

4.1 MODELOS DE PREDIÇÃO TREINADOS COM NPS COM E SEM REMOÇÃO DE HARMÔNICO

A figura 12 apresenta o gráfico da média do coeficiente R^2 e do erro médio para os 200 treinamentos de cada configuração de cada modelo de predição. Os parâmetros selecionados para cada configuração estão detalhados na seção 3.3, e na Tabela 5 observa-se valores médios de métricas de erro para os treinamentos realizados com NPS de banda larga para tais configurações. Nessa tabela, o desvio padrão e média do erro máximo também é disponibilizado. Este erro máximo é o maior valor resultante da subtração do ponto de ensaio e predito. Tais valores foram considerados aceitáveis, e relevantes de serem apresentados apenas para indicar quais seriam os maiores erros que poderiam ser obtidos na predição do NPS.

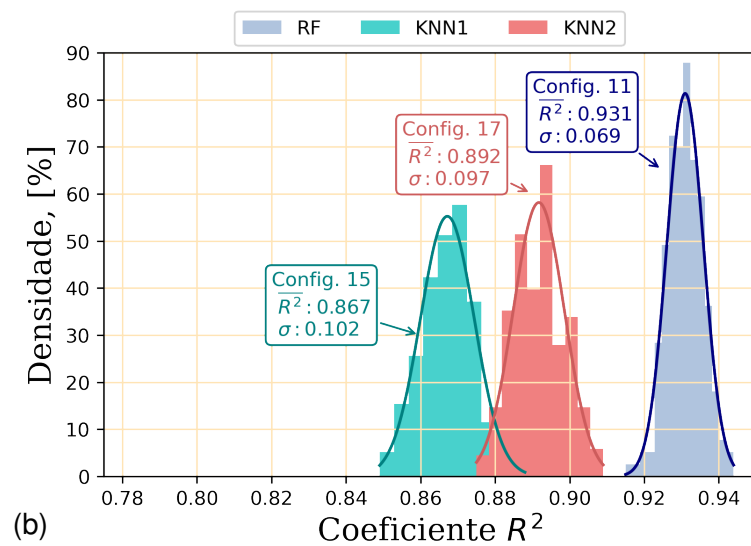
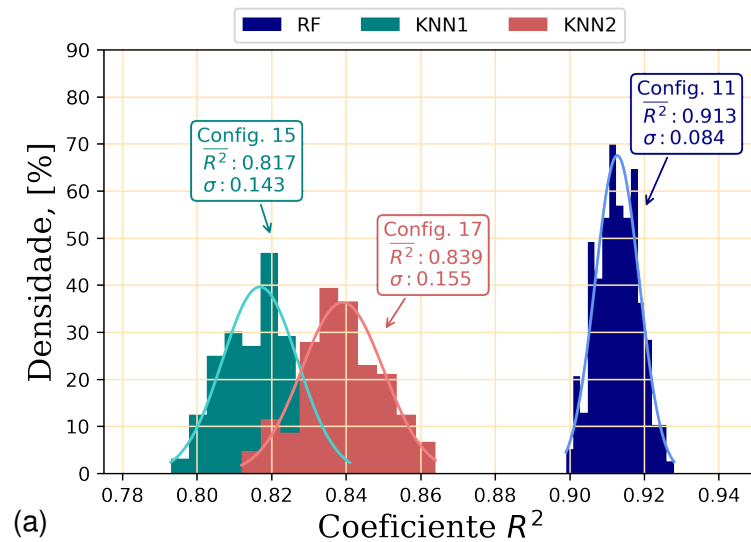
Tabela 5 – Métricas de erro obtidas na validação cruzada de cada configuração e tempo de processamento da validação cruzada.

Configuração	Regressão	$\overline{R^2}$	Desvio Padrão R_2	\overline{Bias}	Desvio Padrão Bias	$\overline{Erro\ máximo}$	Desvio Padrão Erro máximo	minutos total
1	RF	0.91	0.091	0.80	0.095	12.00	21.19	7
2	RF	0.92	0.068	0.74	0.053	10.79	21.64	7
3	RF	0.92	0.074	0.73	0.045	10.42	21.53	7
4	RF	0.93	0.063	0.71	0.045	10.24	20.35	8
5	RF	0.93	0.069	0.70	0.039	10.14	22.66	9
6	RF	0.93	0.067	0.69	0.035	9.96	20.55	9
7	RF	0.93	0.061	0.69	0.031	9.88	19.91	10
8	RF	0.93	0.066	0.68	0.030	9.92	21.51	12
9	RF	0.93	0.067	0.68	0.030	9.83	21.60	13
10	RF	0.93	0.069	0.68	0.030	9.87	22.18	15
11	RF	0.93	0.069	0.68	0.028	9.68	22.63	15
12	RF	0.93	0.066	0.68	0.029	9.83	21.77	20
13	KNN	0.872	0.105	0.56	0.046	11.26	25.77	6
14	KNN	0.869	0.112	0.56	0.046	11.14	26.35	5
15	KNN	0.867	0.102	0.56	0.053	11.28	24.46	5
16	KNN	0.864	0.110	0.56	0.056	11.23	25.36	5
17	KNN	0.892	0.097	1.06	0.068	9.64	17.87	6
18	KNN	0.888	0.097	1.21	0.074	9.68	13.00	6
19	KNN	0.886	0.099	1.30	0.079	10.03	14.33	6
20	KNN	0.875	0.103	1.36	0.112	10.34	16.91	6
21	MLP	0.737	0.555	2.57	2.976	12.88	14.82	196
22	MLP	0.782	0.567	2.32	3.189	12.26	14.38	344
23	MLP	0.818	0.524	2.10	3.167	11.92	13.43	499
24	MLP	0.776	3.597	2.16	14.77	12.19	34.29	1800
25	MLP	0.923	0.104	1.30	0.833	12.64	12.50	5040

As configurações foram selecionadas considerando maiores coeficientes R^2 e menores bias. Tais informações podem ser visualizadas na Figura 12, onde observa-se, ainda, que a média de R^2 e bias indicam maior erro para conjuntos treinados com NPS tonal, indicando que os tonais não são tão bem representados pelos modelos, quanto o NPS de banda larga.

As configurações selecionadas para análise estão destacadas em cinza na Tabela 5. Os 200 modelos de cada configuração selecionada estão representados na figura 10 a, que apresenta o histograma do coeficiente de determinação treinados com o conjunto de dados de NPS harmônico + banda larga enquanto a figura 10 b apresenta o histograma para mesmas configurações, porém utilizando NPS de banda larga para treinamento. Comparando os dois gráficos, é possível observar que os valores de coeficiente de determinação, são maiores para os treinamentos utilizando o NPS de banda larga, e que os maiores valores de R^2 são para a configuração de floresta randômica.

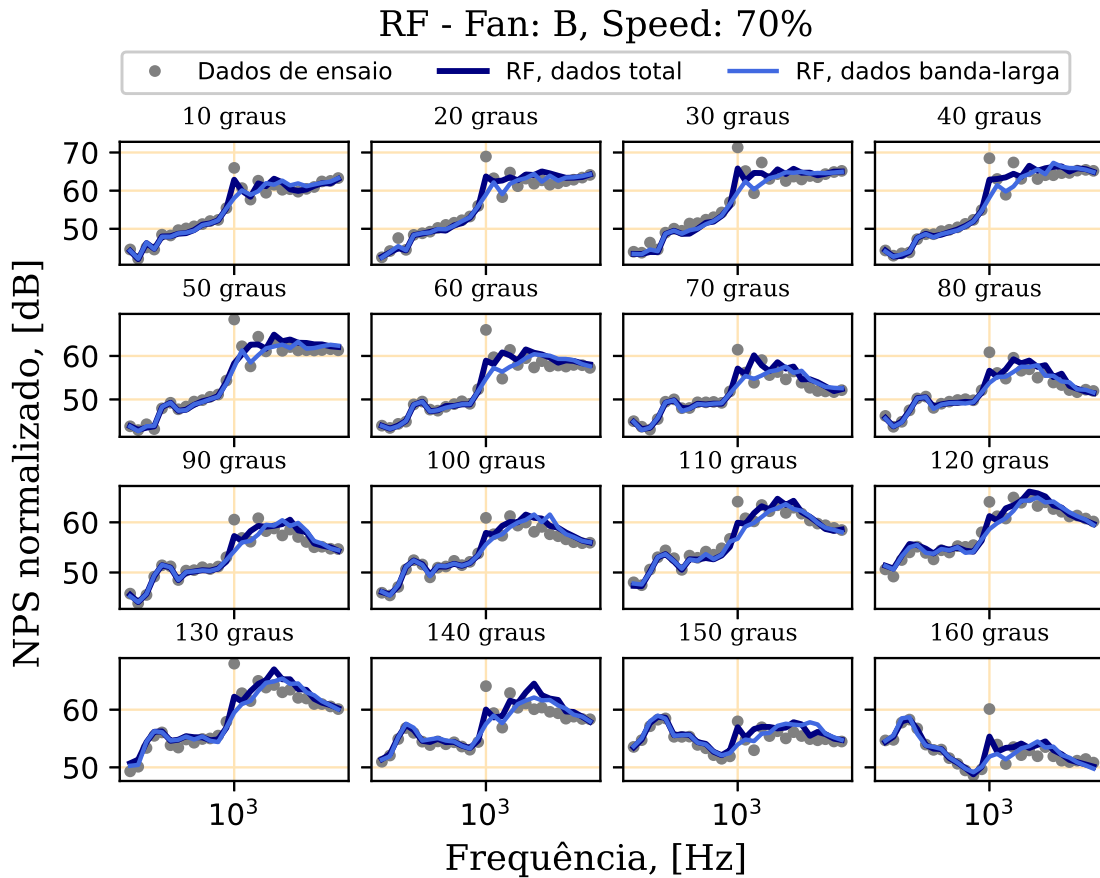
Figura 10 – Histograma do coeficiente de determinação dos modelos selecionados quando o treinamento foi realizado utilizando o NPS com e sem remoção de tonal. (a) Modelos treinados com NPS sem remoção de tonal. (b) Modelos treinados com NPS de banda larga.



Fonte: Autora (2021)

Utilizando o mesmo critério (R^2 mais alto) a regressão com maior R^2 da configuração 11 (floresta randômica) foi selecionada para análise. As regressões obtidas com os diferentes treinamentos são comparadas com os dados de ensaio (NPS normalizado sem separação de componentes) na figura 11. Pode-se observar que no treinamento realizado com NPS normalizado sem remoção da componente harmônica, a predição reproduz parcialmente os picos dos harmônicos, assim, caso a remoção dos tonais não tivesse sido realizada, o treinamento ou a seleção do modelo, deveria ser capaz de prever o ruído de broadband sem representar os tons que estão presentes na

Figura 11 – Predição de NPS sem remoção de tonal (dados total) e NPS de banda larga (dados banda larga) utilizando o modelo de predição Floresta Randômica para o fan B.

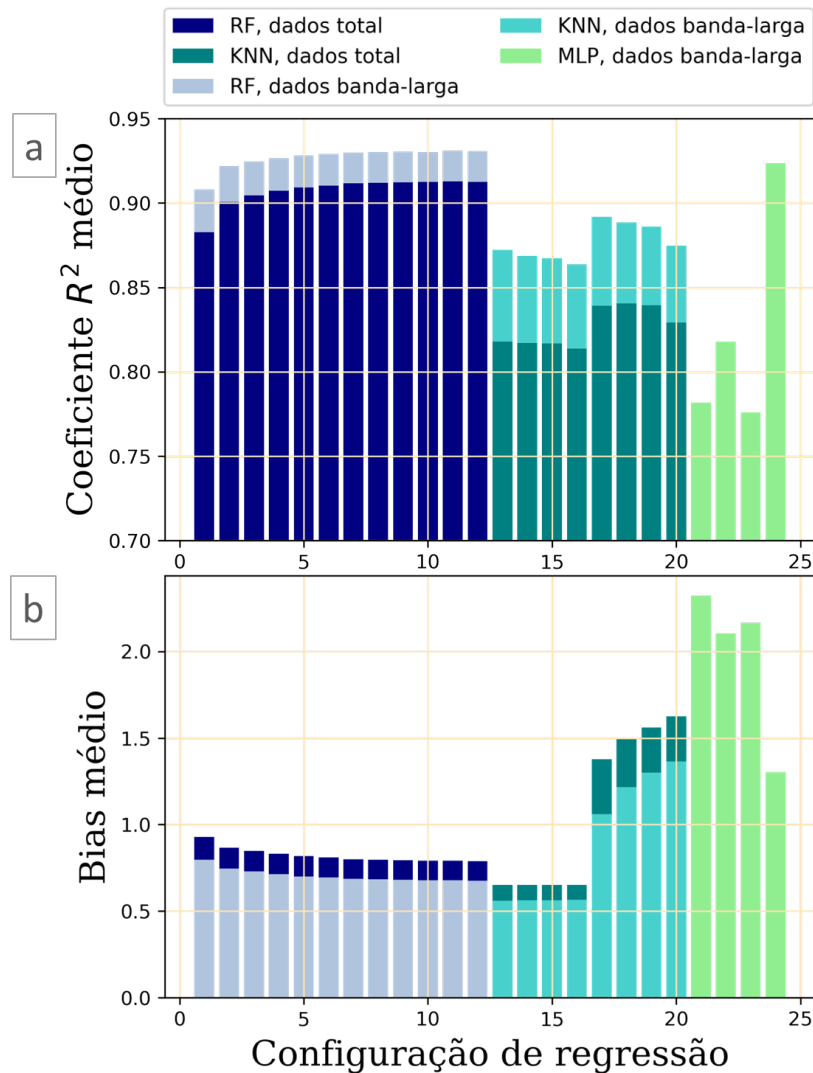


base de dados, visto que é de se esperar que a regressão tente acompanhar a forma dos tonais. Tais observações resultam na seleção de NPS de banda larga para todas as demais análises..

4.2 MODELOS DE PREDIÇÃO TREINADOS COM NPS DE BANDA LARGA (COM REMOÇÃO DE TONAL).

Com os dados da figura 12 e Tabela 5, são selecionadas as melhores configurações, considerando o maior R^2 para a técnica de regressão RF, MLP, KNN com função de ponderação distância e KNN com função de ponderação uniforme. Para cada configuração foi selecionado o modelo com maior R^2 . No caso da técnica RF, a configuração selecionada possui um R^2 médio de 0.931 e o bias médio desta é 0.68, enquanto o modelo selecionado desta configuração possui R^2 igual a 0.944 e bias de 0.68. Para a técnica KNN foram selecionados dois modelos, um dos modelos com a configuração 15 (para a função de ponderação distância), com R^2 médio de 0.867 e bias médio de 0.56. O modelo selecionado para a configuração 15 possui R^2 de 0.887 e

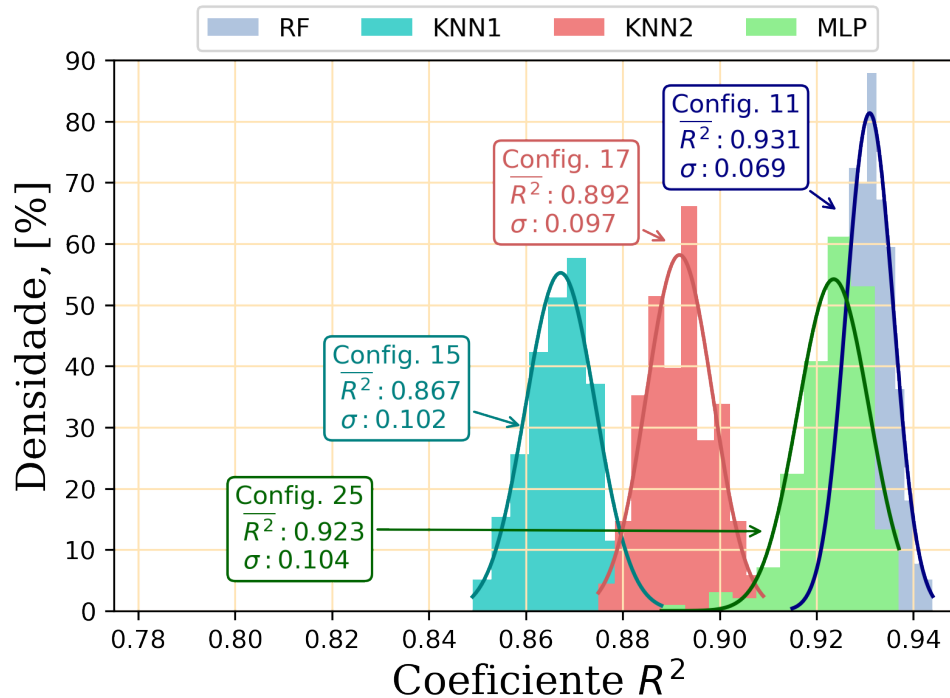
Figura 12 – Dados para 20 repetições de 10 fold por configuração de modelo de predição. (a) Média aritmética de R^2 , (b) Média aritmética do Erro médio.



Fonte: Autora (2021)

bias de 0.557. O outro modelo da técnica KNN (para a função de ponderação uniforme) foi selecionado com a configuração 17 e possui R^2 médio de 0.892 e bias médio de 1.06. Enquanto o modelo selecionado para esta configuração possui R^2 de 0.909 e bias de 1.054. Por fim, para o modelo selecionado utilizando a técnica MLP, o valor de R^2 médio é 0.923 e bias médio é 1.3, e o modelo selecionado possui R^2 0.937 e bias de 1.233. A figura 13 apresenta um histograma do coeficiente de determinação dos 200 modelos treinados com o NPS de banda larga para cada configuração selecionada.

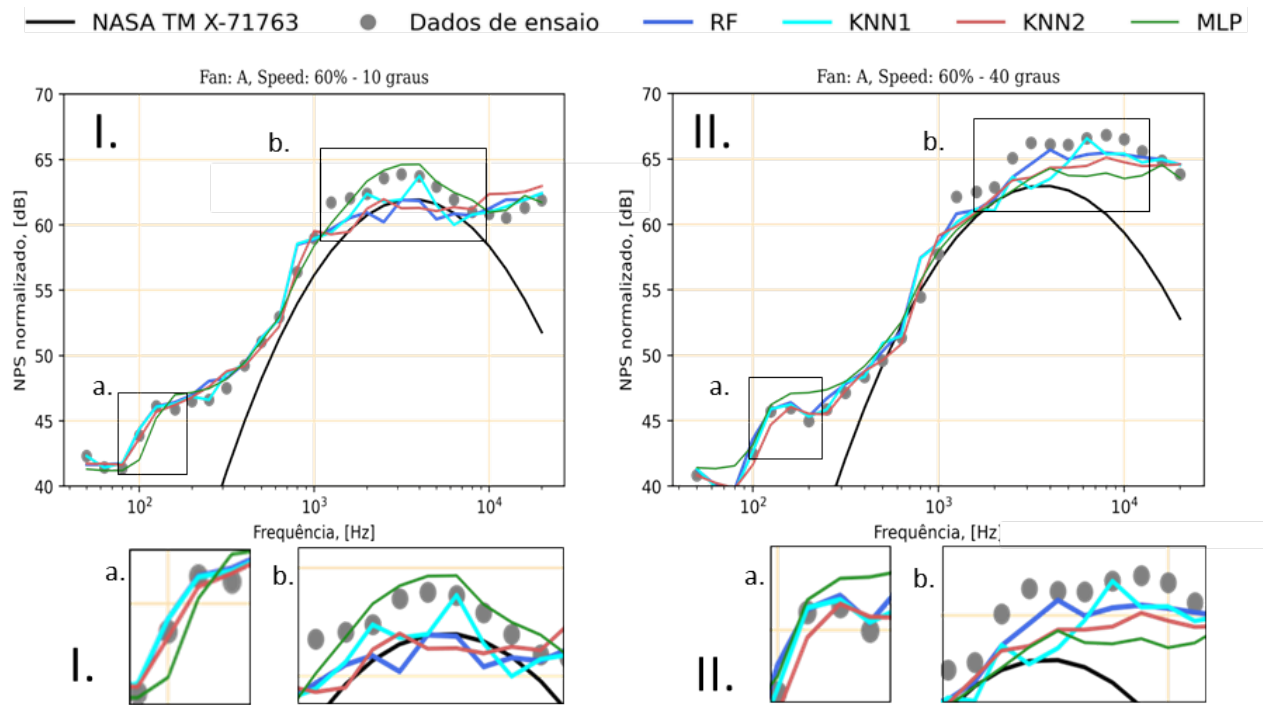
Figura 13 – Histograma da validação cruzada do coeficiente de determinação dos modelos RF, KNN, GP e MLP selecionados.



Fonte: Autora (2021)

Realizou-se a comparação dos espectros obtidos na figura 14. Observa-se na figura que todos os modelos propostos são mais capazes em reproduzir a forma dos espectros, se comparados ao modelo NASA TM X-71763, pois o formato característico de tal modelo semi-empírico não consegue representar a forma do ruído medido nas altas e baixas frequências. O espectro representado na Figura 14 I apresenta uma predição suave e próxima aos níveis desejados, utilizando o modelo MLP. Pode-se observar, comparando I e II, que a predição se aproxima mais do NPS de ensaio em I em relação a II. Comparando as mesmas imagens, observa-se que para o modelo RF ocorre o oposto em II, a solução se aproxima mais dos níveis obtidos em ensaio, sendo que a solução obtida não é tão suave.

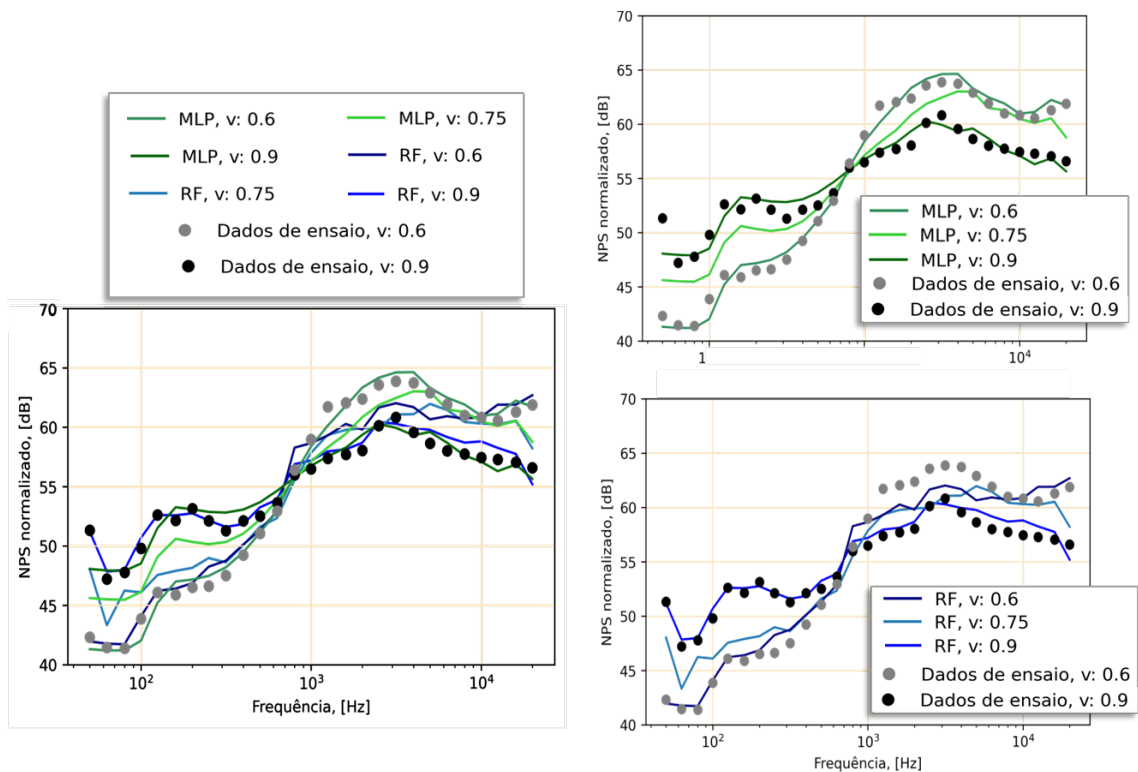
Figura 14 – Comparação entre previsões de NPS realizada com modelos seleccionados, dados de ensaio e modelo de predição NASA TM X-71763.



Fonte: Autora (2021)

A comparação espectral também é realizada para predição em uma velocidade percentual que não está presente no banco de dados (0.75). É exibido na figura 15 os espectros medidos para o Fan A, diretividade de 10° em duas velocidades diferentes, 0.6 e 0.9. Na mesma figura, 3 predições são mostradas para as velocidades 0.6 e 0.9, para observar a predição nas velocidades medidas e na velocidade 0.75 que não foi medida experimentalmente, ou seja, não existe na base de dados nenhuma informação para essa velocidade, em nenhum fan. Pode se observar que para as predições nas velocidades 0.6 e 0.9, a tendência das curvas acompanham os dados medidos. Observa-se também que a velocidade 0.75 (intermediária às duas velocidades medidas) apresenta uma tendência intermediária às das velocidades 0.6 e 0.9.

Figura 15 – Comparação entre o modelo MLP e RF para o Fan A, com a diretividade de 10 graus.

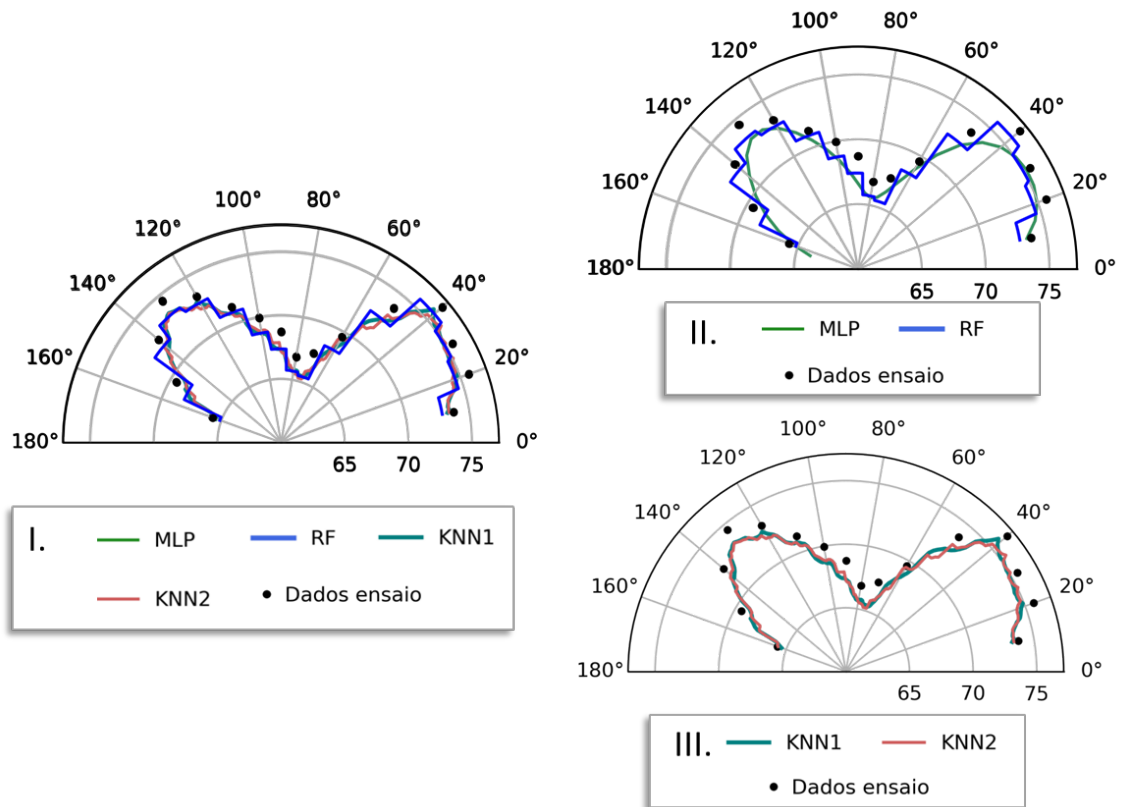


Fonte: Autora (2021)

Além da comparação espectral, também foi realizada a comparação do nível de pressão sonora global (OASPL). Na figura 16, observa-se o OASPL obtido com NPS medido no ensaio e predito com os modelos selecionados para o fan A, com a velocidade percentual de 60%. A predição foi realizada para ângulos de 10 à 160 graus, com intervalo de 1 grau de diretividade. Observando o OASPL, observa-se que existem platôs observáveis no *random forest*, ou seja, a solução é menos suave que a solução dos modelos que utilizam KNN, que, por sua vez, é menos suave que a solução do modelo MLP. É importante ressaltar que o OASPL representa o NPS global à 30.5 metros do conjunto rotor estator, para os ângulos de diretividade já apresentados na figura

2.

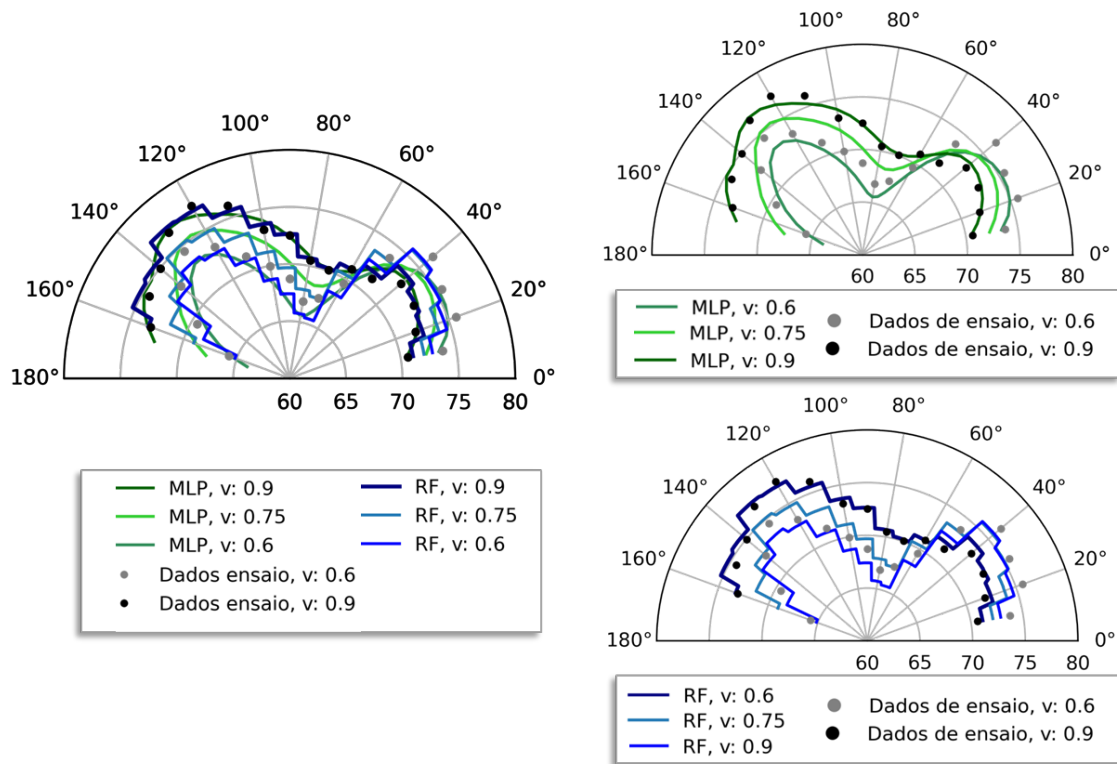
Figura 16 – Nível de pressão sonora global, predito por 4 modelos seleccionados, para o fan A.



Fonte: Autora (2021)

A Figura 17 apresenta a comparação do nível de pressão sonora global do modelo com o resultado mais suave (MLP) e o modelo com melhores métricas de erro (RF). A comparação é realizada para o fan A, com velocidade percentual de 60%, 75% e 90%. Como observado na figura 15, os modelos conseguem representar o comportamento do ruído para as velocidades utilizadas no treinamento, e a tendência para a velocidade 0.75 segue uma tendência intermediária às velocidades 0.6 e 0.9. No entanto, observa-se no NPS global, na predição utilizando o modelo RF (Figuras 16 e 17), um comportamento em platô, o que não ocorre no fenômeno físico em questão. Portanto, o modelo de predição utilizando o RF mostra-se menos adequado do que o MLP. Por outro lado, pode-se observar que a predição realizada pelo modelo utilizando MLP, além de representar adequadamente a variação de velocidade, possui um comportamento mais suave do que o RF. Por esses motivos, dá-se prioridade à MLP na seleção do modelo para a predição de ruído apresentada neste trabalho.

Figura 17 – Nível de pressão sonora global predito por modelos MLP e RF seleccionados, em diferentes velocidades percentuais, para o fan A.



Fonte: Autora (2021)

5 CONCLUSÕES

O presente trabalho dedicou-se ao estudo da predição de ruído de banda larga de turbofan, com o intuito de desenvolver uma ferramenta de predição de ruído, objetivando a predição de ruído com maior acurácia na etapa de projeto conceitual de aeronaves, para, assim, contribuir para a elaboração de projetos de aeronaves mais silenciosas.

A metodologia para desenvolver tal ferramenta incluiu modelagem utilizando aprendizado de máquina, métricas de erro para analisar o desempenho de modelos com diferentes configurações, permitindo à autora, além de obter conhecimentos relativos à aeroacústica e predição de ruído externo, também desenvolver habilidades com modelagem e tratamento de dados, que poderão ser aplicados em diferentes problemas de engenharia, que envolvam, ou não, predição de ruído.

Os resultados evidenciam que a regressão retirando os harmônicos resultou em modelos com maior suavidade, mostrando que é interessante melhorar a técnica de separação dos harmônicos e aplicar a regressão apenas no espectro de banda larga. Com isso, implementou-se diversos modelos de aprendizado de máquina utilizando *random forest*, *k nearest neighbors* e rede neural (*Multi Layer Perceptron*). Com a comparação dos modelos obtidos com os treinamentos, dados de ensaio e o modelo de Heidmann (1979), observou-se que a aplicação de aprendizado de máquina modelou melhor o ruído em baixas e altas frequências, melhor se ajustando aos dados experimentais, principalmente para os motores cujos espectros de banda larga não se comportam como a proposta de Heidmann (1979) espera.

A observação do NPS global com a diretividade evidenciou que o *random forest* não tem o comportamento suave esperado para o fenômeno físico. No entanto, como observado na Figura 13, os histogramas de erro dos métodos RF e MLP possuem uma grande intersecção, o que pode justificar que não existe diferença estatística entre os dois métodos. Este modelo utiliza a técnica MLP com 1 camada, 40 neurônios, função de ativação 'tanh' e a função de otimização 'lbfgs', sendo que o coeficiente de determinação do modelo escolhido é 0.934 e erro médio 1.286.

5.1 TRABALHOS FUTUROS

Toda experiência obtida neste trabalho, motiva algumas melhorias que estão listadas abaixo:

- Obter métrica adequada para verificar suavidade do modelo;
- Separação do grupo de dados com maior amostra de teste, para verificar se permite detectar overfitting de forma mais adequada;

- Separação de ruído tonal e harmônico utilizando *ensemble average* para maior precisão;
- Modelo de predição do nível de pressão sonora do tonal;
- Implementação de modelo para diferentes bancos de dados.

REFERÊNCIAS

- AHMED, N. K. et al. An empirical comparison of machine learning models for time series forecasting. **Econometric Reviews**, Taylor Francis, v. 29, n. 5-6, p. 594–621, 2010. Disponível em: <https://doi.org/10.1080/07474938.2010.481556>.
- BISTAFA, S. R. **Acústica aplicada ao controle do ruído**. 2. ed. São Paulo: Blucher, 2011.
- BROCHINE, C. Metodologia semi-analítica para predição de ruído de banda larga causado pela interação entre a esteira turbulenta do rotor com as aletas da estatora em motores turbofans. Dissertação (Mestrado em Engenharia) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.
- CUENCA, R. G. Aeroacústica de Motores Aeronáuticos: Uma Abordagem por Meta-Modelo. Tese (Doutorado em Ciências) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2017.
- DAHL, M. D. **Assessment of NASA's Aircraft Noise Prediction Capability**. [S.l.], 2012.
- DIGITALIZER, W. P. **Web Plot Digitalizer**. 10 de ago. de 2019. Disponível em: <https://apps.automeris.io/wpd/>.
- DOBRZYNSKI, W. **Almost 40 Years of Airframe Noise Research: What Did We Achieve?** DLR, German Aerospace Center, 38108 Braunschweig, Germany: JOURNAL OF AIRCRAFT Vol. 47, No. 2, 2010.
- DUNN, D. G.; PEART, N. A. **AIRCRAFT NOISE SOURCE AND CONTOUR ESTIMATION**. Washington, DC: National Aeronautics and Space Administration, 1973.
- GALUSHKIN, A. I. **Neural Network Theory**. 1. ed. [S.l.]: Springer, 2007. ISBN 3540481249,9783540481249.
- GUIMARÃES, R. C. C. Desenvolvimento de modelo semi-empírico de predição de ruído de banda larga de fan com adaptação de correlações a partir do princípio de similaridade. Dissertação (Mestrado em Engenharia) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2017.
- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 1. ed. [S.l.]: O'Reilly Media, 2017. ISBN 1491962291,9781491962299.
- HARRINGTON, P. **Machine Learning in Action**. Shelter Island: MANNING, 2012.
- HEIDMANN, M. F. Interim prediction method for fan and compressor source noise. **Scientific and Technical Information Branch**, 1979.
- HEIDMANN, M. F.; FEILER, C. E. **NOISE COMPARISONS FROM FULL-SCALE FAN TESTS AT NASA LEWIS RESEARCH CENTER**. NASA Lewis Research Center Cleveland, Ohio: AIAA, 1973.

- HEIJ, C. et al. **Econometric methods with applications in business and economics**. 1. ed. [S.l.]: Oxford University Press, 2004. ISBN 0199268010,9780199268016,0191533238,9780191533235.
- ICAO, S. 50 years of annex 16 – the special meeting on aircraft noise in the vicinity of airports. https://www.icao.int/environmental-protection/Documents/EnvironmentalReports/2019/ENVReport2019_pg39-43.pdf, 2019.
- KEATING, A. et al. Computational aeroacoustics validation and analysis of a nose landing gear. Exa Corporation, Burlington, MA, 01803, USA, 2015.
- MAX KUHN, K. J. **Feature Engineering and Selection: A Practical Approach for Predictive Models**. 1. ed. [S.l.]: Chapman and Hall/CRC, 2019. (Chapman Hall/CRC Data Science Series). ISBN 1138079227,9781138079229.
- LIBBRECHT, M. Machine learning applications in genetics and genomics. *Nat Rev Genet* 16, 321–332 <https://doi.org/10.1038/nrg3920>, 2015.
- MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 4. ed. [S.l.]: LTC, 2009.
- MOREAU, A.; ENGHARDT, L. Improvements of a parametric model for fan broadband and tonal noise. **The Journal of the Acoustical Society of America**, v. 123, p. 3540, 06 2008.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12, Parietal, INRIA Saclay Neurospin, Bât 145, CEA Saclay 91191 Gif sur Yvette – France, 2014.
- PICARD, R.; COOK, R. D. Cross-validation of regression models. **Journal of the American Statistical Association**, v. 79, n. 387, 1984.
- SREE, D.; STEPHENS, D. B. Improved separation of tone and broadband noise components from open rotor acoustic data. **Aerospace**, 2016.
- TANIGAWA, K.; YAMASAKI, N.; OOISHI, T. Improved hybrid prediction of fan noise. In: **Proceeding of 15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference)**. [S.l.: s.n.], 2009.
- TIMBERS, T.-A.; CAMPBELL, T.; LEE, M. **Data Science: A First Introduction**. 1. ed. [S.l.]: Bookdown, 2021.
- VENTRES, C.; THEOBALD, M.; MARK, W. **Turbofan Noise Generation Volume 1: Analysis**. [S.l.], 1982.
- WARD, T. A. **Aerospace propulsion systems**. 2. ed. Singapore: Wiley, 2011.
- WOLFGANG, B. e. a. **Burden of disease from environmental noise:Quantification of healthy life years lost in Europe**. DK-2100 Copenhagen Ø, Denmark: World Health Organization - WHO Regional Office for Europe, 2011.