



UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Engenharia Civil, Arquitetura e Urbanismo

LUCAS FILIPE DA RONCH

**ESTUDO DE VULNERABILIDADE SÍSMICA DE
OAEs TÍPICAS EM RODOVIAS FEDERAIS DO
NORDESTE DO BRASIL ATRAVÉS DA
APLICAÇÃO DE MÉTODOS EM *MACHINE
LEARNING***

CAMPINAS - SP

2024

LUCAS FILIPE DA RONCH

**ESTUDO DE VULNERABILIDADE SÍSMICA DE
OAEs TÍPICAS EM RODOVIAS FEDERAIS DO
NORDESTE DO BRASIL ATRAVÉS DA
APLICAÇÃO DE MÉTODOS EM *MACHINE
LEARNING***

Dissertação de Mestrado apresentada a Faculdade de Engenharia Civil, Arquitetura e Urbanismo da Unicamp, para obtenção do título de Mestre em Engenharia Civil na área de Estruturas e Geotécnica

Orientador: Prof. Dr. Gustavo Henrique Siqueira

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL
DA DISSERTAÇÃO DEFENDIDA PELO ALUNO LUCAS
FILIPE DA RONCH E ORIENTADA PELO PROF. DR.
GUSTAVO HENRIQUE SIQUEIRA

CAMPINAS - SP

2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

R667e Ronch, Lucas Filipe da, 1992-
Estudo de vulnerabilidade sísmica de OAEs típicas em rodovias federais do nordeste do Brasil através da aplicação de métodos em machine learning / Lucas Filipe da Ronch. – Campinas, SP : [s.n.], 2024.

Orientador(es): Gustavo Henrique Siqueira.
Dissertação (mestrado) – Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Civil, Arquitetura e Urbanismo.

1. Fragilidade. 2. Pontes de concreto. 3. Engenharia de terremotos. 4. Aprendizado de máquina. I. Siqueira, Gustavo Henrique, 1980-. II. Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Civil, Arquitetura e Urbanismo. III. Título.

Informações complementares

Título em outro idioma: Seismic risk assessment of typical bridges in Northeastern Brazil through the application of Machine Learning methods

Palavras-chave em inglês:

Fragility

Concrete bridges

Earthquake engineering

Machine learning

Área de concentração: Estruturas e Geotécnica

Titulação: Mestre em Engenharia Civil

Banca examinadora:

Gustavo Henrique Siqueira [Orientador]

Leandro Mouta Trautwein

Rogério Carrazedo

Data de defesa: 03-10-2024

Programa de Pós-Graduação: Engenharia Civil

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0009-0000-3105-1167>

- Currículo Lattes do autor: <https://lattes.cnpq.br/9609796688923974>

**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA CIVIL, ARQUITETURA E
URBANISMO**

**ESTUDO DE VULNERABILIDADE SÍSMICA DE OAEs
TÍPICAS EM RODOVIAS FEDERAIS DO NORDESTE DO
BRASIL ATRAVÉS DA APLICAÇÃO DE MÉTODOS EM
MACHINE LEARNING**

Lucas Filipe Da Ronch

Dissertação de Mestrado aprovada pela Banca Examinadora, constituída por:

**Prof. Dr. Gustavo Henrique Siqueira
Presidente e Orientador/ FECFAU/UNICAMP**

**Prof. Dr. Leandro Mouta Trautwein
FECFAU/UNICAMP**

**Prof. Dr. Rogério Carrazedo
Escola de Engenharia de São Carlos**

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 03 de outubro de 2024

Resumo

As Obras de Artes Especiais (OAEs) são as componentes mais vulneráveis em um sistema de transporte rodoviário. Logo, recomenda-se que eventos excepcionais como os terremotos sejam considerados no dimensionamento de OAEs, como por exemplo as pontes. No Brasil, nas últimas décadas, terremotos representativos, de magnitudes superiores a 5, aconteceram no Brasil e foram registrados pelo Centro de Sismologia da Universidade de São Paulo (USP). A nova norma nacional para projetos de estruturas sismo-resistentes, NBR 15421 (2023), não se aplica a OAEs. Desta forma, não é evidente se a infraestrutura existente é capaz de resistir uma demanda sísmica. Este trabalho tem como objetivo avaliar a vulnerabilidade de uma OAE (ponte) de concreto armado de geometria típica situada no Nordeste do Brasil, região conhecida pela, relativamente alta, atividade sísmica no território brasileiro. O sistema estrutural é composto por tabuleiro em seção “T” apoiado em encontros do tipo “U” ou gravitacional e em pórticos compostos por dois pilares circulares através de aparelhos de apoio elastoméricos. Os registros sísmicos recolhidos são do *Pacific Earthquake Engineering Research Center* (PEER), sendo estes compatíveis com espectros-alvo para a cidade de Natal (RN). O estudo de vulnerabilidade é conduzido através da construção de curvas de fragilidade, que fornecem a probabilidade condicional da demanda (definida por análises dinâmicas não lineares) exceder a capacidade estrutural para diferentes estados-limites de dano da estrutura. Métodos tradicionais para elaborar as curvas de fragilidade são demasiadamente demorados, principalmente devido ao relativo alto custo computacional de processar inúmeros modelos baseados no Método dos Elementos Finitos. Portanto, é apresentado um esquema computacional eficiente baseado em técnicas de Aprendizado de Máquina (*Machine Learning*) para avaliar a aplicação de ferramentas de inteligência artificial, como por exemplo o *Support Vector Machine*, *Decision Tree* e *Random Forest*, para análises de vulnerabilidade sísmica de estruturas situadas na região nordeste do Brasil. As curvas de fragilidade obtidas pelos métodos tradicionais e pelas técnicas de *Machine Learning* são similares.

Palavras-chave: Fragilidade, pontes de concreto, engenharia de terremotos, aprendizado de máquina.

Abstract

Bridges are the most vulnerable components in a transportation system. Therefore, extraordinary events such as seismic activities should also be considered in bridge design. In Brazil, since the last few decades, representative earthquakes with magnitudes above 5 have been registered by the seismological center of the University of São Paulo (USP). The newest edition of the Brazilian regulation for seismic-resistant structure design, NBR 15421 (2023), does not apply to special structures, such as bridges. It is not clear whether the existing infrastructure can resist seismic demands. This work aims to evaluate the vulnerability of a reinforced concrete bridge with typical geometry located in the northeast of Brazil, a region known for relatively high seismic activity in Brazilian territory. The structural system is composed of a "T" section deck supported by non-integral U-type or gravity abutments and by frames with two circular columns through elastomeric bearings. The seismic records were collected from the Pacific Earthquake Engineering Research Center, which is compatible with the target spectrum of the city of Natal/RN. The study of vulnerability is conducted through the assembly of fragility curves, which provide the conditional probability of the demand (obtained by nonlinear dynamic analysis) exceeding the structural capacity for different structural damage limit states. Traditional approaches to evaluating fragility curves are time-consuming, mainly because of the relatively high computational cost of running several finite element-based models. Here, we present a computationally efficient scheme based on Machine Learning techniques to evaluate the application of artificial intelligence tools, such as Support Vector Machine, Decision Tree and Random Forest, for seismic vulnerability for structures situated in the northeastern region of Brazil. The fragility curves obtained by traditional approach and Machine Learning techniques were very similar.

Keywords: Fragility, concrete bridges, earthquake engineering, machine learning.

Lista de ilustrações

Figura 1	– Sismos do Brasil de 1720 a 2016 (Fonte: Assumpção et al. (2016)).	24
Figura 2	– Catálogo uniforme, filtrado conforme detectabilidade (Fonte: adaptado de Assumpção et al. (2016)).	24
Figura 3	– Mapeamento da aceleração sísmica no Brasil (Fonte: NBR 15421 (2023)).	25
Figura 4	– Mapas de ameaça sísmica para aceleração de pico em rocha, para tempo de retorno de 475 e 2475 anos, respectivamente (Fonte: adaptado de Assumpção et al. (2016)).	26
Figura 5	– Mapas de ameaça sísmica para aceleração de pico em rocha, para tempo de retorno de 475 e 2475 anos, respectivamente (Fonte: adaptado de Petersen et al. (2018)). .	26
Figura 6	– Mapas de ameaça sísmica para acelerações com período de vibração de 0,2 e 1,0 segundo, para um tempo de retorno de 475 anos, respectivamente (Fonte: adaptado de Petersen et al. (2018)).	27
Figura 7	– Mapas de ameaça sísmica para acelerações com período de vibração de 0,2 e 1,0 segundo, para um tempo de retorno de 2475 anos, respectivamente (Fonte: adaptado de Petersen et al. (2018)).	28
Figura 8	– Fatores que compõe o risco (Fonte: autor).	28
Figura 9	– Mapa de localização das tipologias na Região Nordeste do Brasil (Fonte: adaptado de Cavalcante (2022)).	31
Figura 10	– Espectro de resposta para estruturas com um grau de liberdade com mesma taxa de amortecimento e períodos naturais diferentes solicitados por um mesmo terremoto (Fonte: adaptado de Paultre (2013)).	34
Figura 11	– Ilustração do PSDM no espaço logarítmico (Fonte: adaptado de Cavalcante (2022)).	37
Figura 12	– Curva de capacidade com os Estados Limites (LS) de Dano (Fonte: autor).	40
Figura 13	– Tipos de <i>Machine Learning</i> e alguns algoritmos comumente utilizados (Fonte: adaptado de Kong et al. (2019)).	42
Figura 14	– Exemplo de <i>SVM</i> para classificação (Fonte: adaptado de Xie et al. (2020)).	46
Figura 15	– Exemplo de <i>SVR</i> para regressão (Fonte: adaptado de Xie et al. (2020)).	47
Figura 16	– Exemplo da estrutura do algoritmo do método <i>Decision Tree</i> (Fonte: autor).	48
Figura 17	– Exemplo da estrutura do algoritmo do método <i>Random Forest</i> (Fonte: autor).	49
Figura 18	– Exemplo do gráfico <i>Response Plot</i> (Fonte: autor).	52
Figura 19	– Exemplo do gráfico <i>Response Plot</i> utilizando <i>box plot</i> (Fonte: autor).	53
Figura 20	– Exemplo do gráfico <i>Predict vs. Actual Response</i> (Fonte: autor).	53
Figura 21	– Exemplo do gráfico de resíduos (Fonte: autor).	54

Figura 22 – Exemplo da matriz de confusão para apenas 2 classes (Fonte: autor).	55
Figura 23 – Exemplo de <i>Scatter Plot</i> para apenas 3 classes (Fonte: autor).	58
Figura 24 – Exemplo de <i>Parallel Coordinates Plot</i> para apenas 3 classes (Fonte: autor).	58
Figura 25 – Exemplo de <i>Receiver Operating Characteristics Curve</i> (Fonte: autor).	60
Figura 26 – Exemplo da <i>confusion matrix</i> para 3 classes (Fonte: autor).	62
Figura 27 – Classes C1, C2 e C3 da ponte de tipologia T3 (Fonte: adaptado de Cavalcante (2022)).	64
Figura 28 – Modelo da tipologia adotada (Fonte: autor).	67
Figura 29 – Taxa de armadura dos pilares em função da seção transversal (Fonte: autor).	70
Figura 30 – Armadura adotada para as vigas travessas (Fonte: autor).	70
Figura 31 – Espectro de resposta (Fonte: adaptado de ASCE (2017)).	71
Figura 32 – Espectro de resposta para solo D considerando TR de 475 anos (Fonte: autor).	72
Figura 33 – Espectros individuais, médio e alvo para solo tipo D considerando TR de 475 anos (Fonte: adaptado de Cavalcante (2022)).	74
Figura 34 – Modelo estrutural da tipologia T3 (Fonte: adaptado de Cavalcante (2022)).	77
Figura 35 – Modelo dos elementos de concreto dividido em fibras (Fonte: adaptado de Cavalcante (2022)).	78
Figura 36 – Modelos constitutivos do concreto e do aço (Fonte: adaptado de Cavalcante (2022)).	79
Figura 37 – Modelos utilizado para a ISE para fundação profunda (Fonte: adaptado de Cavalcante (2022)).	81
Figura 38 – Modelos utilizado para os aparelhos de apoio (Fonte: adaptado de Cavalcante (2022)).	83
Figura 39 – Modelos utilizado para as juntas (<i>gaps</i>) (Fonte: adaptado de Cavalcante (2022)).	84
Figura 40 – Tipo de encontros utilizados (Fonte: adaptado de Cavalcante (2022)).	84
Figura 41 – Modelo utilizado para os encontros no sentido longitudinal (Fonte: adaptado de Cavalcante (2022)).	85
Figura 42 – Modelo utilizado para os encontros no sentido transversal (Fonte: adaptado de Cavalcante (2022)).	86
Figura 43 – Modelo utilizado para os encontros no sentido vertical (Fonte: adaptado de Cavalcante (2022)).	87
Figura 44 – Método de obtenção dos parâmetros S_c e β_c (Fonte: adaptado de Cavalcante (2022)).	89
Figura 45 – Etapas para obtenção da curvatura ϕ_y (<i>Pushover multidirecional</i>) (Fonte: adaptado de Cavalcante (2022)).	90
Figura 46 – <i>Interquartile Range Method</i> (Fonte: autor).	92
Figura 47 – Propriedades geométricas da ponte de tipologia T3 classe C1 (Fonte: adaptado de Cavalcante (2022)).	98

Figura 48 – Modos de vibração da ponte de tipologia T3 classe C1 (Fonte: adaptado de Cavalcante (2022)).	98
Figura 49 – Deslocamento longitudinal e transversal do tabuleiro ao longo do tempo (Fonte: adaptado de Cavalcante (2022)).	99
Figura 50 – Deslocamento longitudinal e transversal dos encontros ao longo do tempo (Fonte: adaptado de Cavalcante (2022)).	99
Figura 51 – Resposta histeréticas dos aparelhos de apoio (Fonte: adaptado de Cavalcante (2022)).	100
Figura 52 – Resposta histeréticas dos pilares (Fonte: adaptado de Cavalcante (2022)).	100
Figura 53 – Resultados dos PSDMs para a ponte de tipologia T3 classe C1 (Fonte: autor). . .	102
Figura 54 – Curvas de fragilidade para cada estado-limite de dano (Fonte: autor).	103
Figura 55 – Curva de fragilidade do sistema estrutural para cada estado-limite de dano (Fonte: autor).	104
Figura 56 – CC - <i>Response plot - training set</i> (Fonte: autor).	107
Figura 57 – CC - <i>Response plot - testing set</i> (Fonte: autor).	107
Figura 58 – CC - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	108
Figura 59 – CC - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	108
Figura 60 – CC - <i>Residual plot - training set</i> (Fonte: autor).	109
Figura 61 – CC - <i>Residual plot - testing set</i> (Fonte: autor).	109
Figura 62 – CC - Modelo esquemático da <i>Decision Tree</i> (Fonte: autor)	110
Figura 63 – ADA - <i>Response plot - training set</i> (Fonte: autor).	112
Figura 64 – ADA - <i>Response plot - testing set</i> (Fonte: autor).	113
Figura 65 – ADA - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	113
Figura 66 – ADA - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	114
Figura 67 – ADA - <i>Residual plot - training set</i> (Fonte: autor).	114
Figura 68 – ADA - <i>Residual plot - testing set</i> (Fonte: autor).	115
Figura 69 – ADA - Modelo esquemático da <i>Decision Tree</i>	115
Figura 70 – ADP - <i>Response plot - training set</i> (Fonte: autor).	117
Figura 71 – ADP - <i>Response plot - testing set</i> (Fonte: autor).	118
Figura 72 – ADP - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	118
Figura 73 – ADP - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	119
Figura 74 – ADP - <i>Residual plot - training set</i> (Fonte: autor).	119
Figura 75 – ADP - <i>Residual plot - testing set</i> (Fonte: autor).	120
Figura 76 – ADP - Modelo esquemático da <i>Decision Tree</i> (Fonte: autor).	120
Figura 77 – ADT - <i>Response plot - training set</i> (Fonte: autor).	123
Figura 78 – ADT - <i>Response plot - testing set</i> (Fonte: autor).	123
Figura 79 – ADT - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	124

Figura 80 – ADT - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	124
Figura 81 – ADT - <i>Residual plot - training set</i> (Fonte: autor).	125
Figura 82 – ADT - <i>Residual plot - testing set</i> (Fonte: autor).	125
Figura 83 – ADT - <i>Modelo esquemático da primeira árvore de decisão do Random Forest</i> (Fonte: autor).	126
Figura 84 – BDL - <i>Response plot - training set</i> (Fonte: autor).	128
Figura 85 – BDL - <i>Response plot - testing set</i> (Fonte: autor).	129
Figura 86 – BDL - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	129
Figura 87 – BDL - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	130
Figura 88 – BDL - <i>Residual plot - training set</i> (Fonte: autor).	130
Figura 89 – BDL - <i>Residual plot - testing set</i> (Fonte: autor).	131
Figura 90 – BDL - <i>Modelo esquemático do algoritmo Decision Tree</i> (Fonte: autor).	131
Figura 91 – BDT - <i>Response plot - training set</i> (Fonte: autor).	133
Figura 92 – BDT - <i>Response plot - testing set</i> (Fonte: autor).	134
Figura 93 – BDT - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	134
Figura 94 – BDT - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	135
Figura 95 – BDT - <i>Residual plot - training set</i> (Fonte: autor).	135
Figura 96 – BDT - <i>Residual plot - testing set</i> (Fonte: autor).	136
Figura 97 – BDT - <i>Modelo esquemático do algoritmo Decision Tree</i> (Fonte: autor).	136
Figura 98 – DU - <i>Response plot - training set</i> (Fonte: autor).	138
Figura 99 – DU - <i>Response plot - testing set</i> (Fonte: autor).	139
Figura 100 – DU - <i>Predicted vs. Actual plot - training set</i> (Fonte: autor).	139
Figura 101 – DU - <i>Predicted vs. Actual plot - testing set</i> (Fonte: autor).	140
Figura 102 – DU - <i>Residual plot - training set</i> (Fonte: autor).	140
Figura 103 – DU - <i>Residual plot - testing set</i> (Fonte: autor).	141
Figura 104 – DU - <i>Modelo esquemático do algoritmo Decision Tree</i> (Fonte: autor).	141
Figura 105 – Curvas de fragilidade para cada estado-limite de dano após modelo de regressão (Fonte: autor).	144
Figura 106 – Curva de fragilidade do sistema estrutural para cada estado-limite de dano após modelo de regressão do ML (Fonte: autor).	144
Figura 107 – ADA - <i>Modelo esquemático do algoritmo Decision Tree</i> (Fonte: autor).	147
Figura 108 – ADA - <i>Confusion Matrix</i> para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).	148
Figura 109 – ADP - <i>Modelo esquemático do algoritmo Decision Tree</i> (Fonte: autor).	150
Figura 110 – ADP - <i>Confusion Matrix</i> para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).	151

Figura 111 – BDL - Modelo esquemático do algoritmo <i>Decision Tree</i> (Fonte: autor).	153
Figura 112 – BDL - <i>Confusion Matrix</i> para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).	154
Figura 113 – BDT - <i>Confusion Matrix</i> para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).	156
Figura 114 – DU - Modelo esquemático do algoritmo <i>Decision Tree</i> (Fonte: autor).	158
Figura 115 – DU - <i>Confusion Matrix</i> para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).	159
Figura 116 – Comparação entre as curvas de fragilidade do sistema estrutural para cada estado-limite de dano - Para análise estocástica e após modelo de regressão do ML. . . .	160
Figura 117 – Curva de fragilidade do sistema estrutural para os estados-limite de dano extensivo e completo para a análise estocástica e após modelo de regressão do ML.	160
Figura 118 – Comparação entre as diferenças, em termos de probabilidades de dano, para os estados limites extensivo e completo para os modelos estocástico e regressão do ML.	161

Lista de tabelas

Tabela 1 – Terremotos, de 2000 até 2010, com maior número de mortos (Fonte: adaptado de USGS (2021)).	22
Tabela 2 – Terremotos, de 2000 até 2019, com magnitude superior a 5 (Fonte: adaptado do Centro de Sismologia da Universidade de São Paulo (USP)).	23
Tabela 3 – Divisão das zonas sísmicas no Brasil (Fonte: adaptado da NBR 15421 (2023)).	25
Tabela 4 – OAEs analisadas por estado da região Nordeste do Brasil (Fonte: adaptado de Cavalcante (2022)).	30
Tabela 5 – Exemplos de medidas de intensidades (Fonte: adaptado de Freddi et al. (2017)).	33
Tabela 6 – Valores adotados de S_C e β_C (Fonte: adaptado de Mangalathu et al. (2017)).	39
Tabela 7 – S_C e β_C obtidos através da análise <i>Pushover</i> (Fonte: adaptado de Cavalcante (2022)).	40
Tabela 8 – Parâmetros de avaliação (Fonte: adaptado de Hastie et al. (2001)).	50
Tabela 9 – Parâmetros estatísticos calculados a partir da <i>confusion matrix</i> (Fonte: autor).	57
Tabela 10 – Configuração das curvas ROC a partir de um <i>confusion matrix</i> 3x3 (Fonte: autor).	62
Tabela 11 – Características estruturais da classe adotada (Fonte: adaptado de Cavalcante (2022)).	64
Tabela 12 – Valores adotados e parâmetros das variáveis aleatórias para as pontes da classe C1 tipologia T3 (Fonte: adaptado de Cavalcante (2022)).	66
Tabela 13 – Propriedades materiais aleatórias (Fonte: adaptado de Cavalcante (2022)). .	69
Tabela 14 – Parâmetros adotados para coleta dos registros sísmicos compatíveis com o espectro de resposta (Fonte: adaptado de Cavalcante (2022)).	73
Tabela 15 – Metodologia utilizada nas ligações para a tipologia T3 (Fonte: adaptado de Cavalcante (2022)).	76
Tabela 16 – Equações utilizadas para o cálculo do coeficiente de atrito.	82
Tabela 17 – Valores adotados para divisão dos grupos (<i>TAGs</i>) (Fonte: autor).	95
Tabela 18 – Resultados do PSDM para cada Parâmetro de Engenharia (Fonte: autor). .	101
Tabela 19 – CC - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	105
Tabela 20 – CC - <i>training set</i> - <i>Decision Tree</i> (Fonte: autor).	106
Tabela 21 – CC - <i>testing set</i> - <i>Decision Tree</i> (Fonte: autor).	106

Tabela 22 – ADA - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	111
Tabela 23 – ADA - <i>training set</i> - <i>Decision Tree</i> (Fonte: autor).	111
Tabela 24 – ADA - <i>testing set</i> - <i>Decision Tree</i> (Fonte: autor).	112
Tabela 25 – ADP - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	116
Tabela 26 – ADP - <i>training set</i> - <i>Decision Tree</i> (Fonte: autor).	116
Tabela 27 – ADP - <i>testing set</i> - <i>Decision Tree</i> (Fonte: autor).	117
Tabela 28 – ADT - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	121
Tabela 29 – ADT - Deslocamento transversal dos encontros - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	122
Tabela 30 – ADT - Deslocamento transversal dos encontros - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	122
Tabela 31 – BDL - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	127
Tabela 32 – BDL - <i>training set</i> - <i>Decision Tree</i> (Fonte: autor).	127
Tabela 33 – BDL - <i>testing set</i> - <i>Decision Tree</i> (Fonte: autor).	128
Tabela 34 – BDT - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	132
Tabela 35 – BDT - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	132
Tabela 36 – BDT - Deslocamento transversal dos aparelhos de apoio - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	133
Tabela 37 – DU - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).	137
Tabela 38 – DU - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	137
Tabela 39 – DU - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	138
Tabela 40 – Resumo dos resultados obtidos - modelo de regressão (Fonte: autor).	142
Tabela 41 – Resultados do PSDM para cada Parâmetro de Engenharia após modelo de regressão (Fonte: autor).	143
Tabela 42 – ADA - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).	146
Tabela 43 – ADA - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	146
Tabela 44 – ADA - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	147
Tabela 45 – ADP - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).	149

Tabela 46 – ADP - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	149
Tabela 47 – ADP - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	150
Tabela 48 – BDL - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).	152
Tabela 49 – BDL - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	152
Tabela 50 – BDL - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	153
Tabela 51 – BDT - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).	155
Tabela 52 – BDT - <i>training set</i> - <i>Support Vector Machine</i> (Fonte: autor).	155
Tabela 53 – BDT - <i>testing set</i> - <i>Support Vector Machine</i> (Fonte: autor).	156
Tabela 54 – DU - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).	157
Tabela 55 – DU - <i>training set</i> - <i>Random Forest</i> (Fonte: autor).	157
Tabela 56 – DU - Deslocamento do tabuleiro - <i>testing set</i> - <i>Random Forest</i> (Fonte: autor).	158
Tabela 57 – Resumo dos resultados obtidos - modelo de classificação (Fonte: autor). . . .	159

Sumário

1	INTRODUÇÃO	17
1.1	Justificativa	19
1.2	Objetivo Geral	20
1.2.1	Objetivos específicos	20
1.3	Organização da dissertação	21
2	REFERENCIAL TEÓRICO	22
2.1	Situação no Brasil	22
2.2	Análise de risco sísmico	28
2.3	Tipologias estruturais	29
2.4	Curvas de fragilidade	31
2.4.1	Demanda sísmica	33
2.4.2	Capacidade estrutural	37
2.5	Machine Learning	41
2.5.1	Supervised learning	44
2.5.1.1	Regressão	49
2.5.1.2	Classificação	54
2.6	Síntese do capítulo	62
3	METODOLOGIA	63
3.1	Tipologia adotada	63
3.2	Geração do banco de dados	67
3.3	Modelagem numérica	74
3.4	Análises dinâmicas não lineares	87
3.5	Modelos de demanda e capacidade	88
3.6	Curvas de fragilidade	90
3.7	Machine Learning	90
3.7.1	Regressão	93
3.7.2	Classificação	94
3.8	Síntese do capítulo	96
4	RESULTADOS	97
4.1	Resultado estrutural	97

4.1.1	Resultado determinístico	97
4.1.2	Resultado estocástico	100
4.2	Resultado do Machine Learning	104
4.2.1	Resultados da regressão	105
4.2.1.1	Curvatura dos pilares	105
4.2.1.2	Deslocamento ativo dos encontros	110
4.2.1.3	Deslocamento passivo dos encontros	116
4.2.1.4	Deslocamento transversal dos encontros	121
4.2.1.5	Deslocamento longitudinal dos aparelhos de apoio	126
4.2.1.6	Deslocamento transversal dos aparelhos de apoio	132
4.2.1.7	Deslocamento do tabuleiro	137
4.2.1.8	Resumo dos resultados obtidos para os parâmetros de engenharia - modelo de regressão	142
4.2.1.9	Curvas de fragilidade - resultados do modelo de regressão	142
4.2.2	Resultados da classificação	145
4.2.2.1	Deslocamento ativo dos encontros	145
4.2.2.2	Deslocamento passivo dos encontros	148
4.2.2.3	Deslocamento longitudinal dos aparelhos de apoio	151
4.2.2.4	Deslocamento transversal dos aparelhos de apoio	154
4.2.2.5	Deslocamento do tabuleiro	156
4.2.2.6	Resumo dos resultados obtidos para os parâmetros de engenharia - modelo de classificação	159
4.3	Comparação da análise estocástica vs Machine Learning	159
4.4	Síntese do capítulo	162
5	CONSIDERAÇÕES FINAIS	163
5.1	Sugestão de trabalhos futuros	165
	REFERÊNCIAS	166

1 Introdução

Os eventos sísmicos são definidos por vibrações de terra derivadas de deslizamentos de falhas geológicas (i.e., intraplaca) ou pelo choque de placas (i.e., interplaca) no interior da crosta terrestre. Considerando que o risco sísmico é composto pela ameaça, exposição e vulnerabilidade, faz-se necessário a consideração dessas solicitações no dimensionamento estrutural, principalmente em obras de infraestruturas como, por exemplo, as pontes. Segundo Dukes et al. (2012) as pontes são essenciais no sistema de transporte, durante e depois dos eventos sísmicos, providenciando rotas de emergência e de evacuação. No Brasil o Departamento Nacional de Infraestrutura de Transportes (DNIT) é o órgão responsável por supervisionar os projetos, as construções, as operações, as manutenções, os reparos, as reabilitações e as substituições de mais de cinco mil obras de artes especiais situadas em rodovias federais conforme Oliveira et al. (2019). Neste estudo, serão utilizados os termos “OAE” e “ponte” de forma intercambiável, pois ambos se referem à mesma ideia central.

Considerando os aspectos não-lineares da interação entre a atividade sísmica, como a duração, energia liberada e probabilidade de ocorrência, e a estrutura de concreto armado, como o comportamento variável quando solicitado por uma ação dinâmica, o método tradicional de análise que utiliza coeficientes de majoração e minoração das ações e resistências, respectivamente, pode não garantir a segurança devida à estrutura. Diante desse cenário, modelos probabilísticos têm sido utilizados para avaliar a vulnerabilidade das estruturas situadas em locais com atividades sísmicas através das curvas de fragilidade. As curvas de fragilidade descrevem a probabilidade condicional que uma estrutura atinja ou ultrapasse um determinado nível de dano específico em função da intensidade sísmica (*intensity measure - IM*) (NIELSON, 2005). As curvas são construídas através da análise de demanda e da capacidade sísmica, sendo essas variáveis consideradas como aleatórias. Portanto, torna-se necessário gerar modelos que caracterizam uma estrutura base de interesse em função das variáveis aleatórias pré-definidas, como geometria, propriedades dos materiais, entre outras.

A demanda sísmica corresponde à resposta máxima de um componente estrutural quando solicitado por atividade sísmica. De acordo com Mangalathu e Jeon (2019) a demanda pode ser obtida através de vários tipos de análises como a *incremental dynamic analysis*, *stripe analysis* e *cloud analysis*, sendo esta última análise aplicada durante este trabalho, onde utiliza o *nonlinear time history analysis* para extrair as respostas sísmicas. Segundo Cornell et al. (2002) a relação das respostas máximas das componentes estruturais em função de uma medida de intensidade

da atividade sísmica pode ser estabelecida através de uma função exponencial, admitindo-se que segue uma distribuição lognormal. Logo, é possível efetuar uma regressão linear no espaço logarítmico, de forma a ser obtida a média e dispersão da demanda.

A capacidade sísmica das componentes é definida através dos estados limites dos elementos estruturais, que podem ser descritas como na FEMA (2003), que caracteriza os estados limites de maneira qualitativa. Ela caracteriza os estados limites em quatro classificações: danos leves (*slight*), danos moderados (*moderate*), danos extensos (*extensive*) e colapso (*complete*). De acordo com Cornell et al. (2002), admite-se que a capacidade possui uma distribuição lognormal. Os valores de média e dispersão para a capacidade dos componentes estruturais podem ser encontrados na literatura, como em Mangalathu et al. (2017). Outra maneira de obter esses valores é através de análise *Pushover*, que representa uma análise estática não-linear conduzida sob cargas gravitacionais constantes e a aplicação gradual de incrementos de cargas horizontais. Esta análise representa o comportamento da estrutura frente às ações sísmicas.

As características dos terremotos devem ser definidas como variáveis aleatórias, sendo empregado um conjunto de sinais (acelerograma). As fontes dessas incertezas são caracterizadas como aleatórias ou epistêmicas. De acordo com Nielson e DesRoches (2007) e Padgett e DesRoches (2008), a variabilidade de alguns parâmetros na modelagem e nas acelerações sísmicas na geração das curvas de fragilidade são importantes, sendo necessário uma escolha criteriosa das variáveis aleatórias e de suas respectivas distribuições estatísticas.

O processo de análise a partir de modelos probabilísticos voltados para a avaliação de vulnerabilidade estrutural frente a uma ação sísmica exige uma grande demanda computacional e, por consequência, excessivo tempo de processamento. Como alternativa a excessiva demanda de tempo e demanda computacional, surge a implementação de métodos baseados em inteligência artificial, como o *Machine Learning*, que são aplicados em diversas áreas da ciência conforme observado em Hastie et al. (2001). Estes métodos, quando aplicados como inteligência artificial, estão em constante desenvolvimento e implementam ferramentas estatísticas para proporcionar a habilidade de autoaprendizagem à linguagem computacional (KIANI et al., 2019). Ou seja, a partir de um determinado problema, deseja-se estimar respostas quantitativas e/ou qualitativas baseadas em dados de entrada, que são chamados de variáveis aleatórias. Portanto, é criada uma série de amostras que, quando analisada, gera um conjunto de respostas utilizadas para o treinamento da linguagem computacional. Com isto, é criado um modelo de previsão para novas amostras não simuladas.

Os métodos do *Machine Learning* podem ser divididos em duas partes distintas, a aprendizagem supervisionada (*supervised learning*) e a não supervisionada (*unsupervised learning*).

Basicamente, a diferença entre as duas é que a aprendizagem supervisionada utiliza respostas quantitativas e/ou qualitativas como *input* no processo e a outra não. Além do mais, dentro da aprendizagem supervisionada se destacam dois métodos, o de classificação (*classification*) e o de regressão (*regression*). Dentro do âmbito da análise de vulnerabilidade estrutural, o método de regressão pode ser empregado para estimar tanto as respostas da demanda sísmica quanto da capacidade estrutural, gerando as curvas de fragilidade a partir da aplicação de ajustes de funções nos pontos obtidos. Já o método de classificação permite que a estrutura seja classificada de acordo com os indicadores, usualmente nomeados como *tags*, do estado qualitativo estrutural previamente estabelecido.

Uma etapa fundamental para aplicação do *Machine Learning* é a análise de sensibilidade das variáveis aleatórias, que compõe os dados de entrada, uma vez que nem todas as variáveis escolhidas como aleatórias possuem importância na calibração do modelo de previsão. A aplicação de ferramentas baseadas em *Machine Learning* ainda é muito recente na área de análise de vulnerabilidade estrutural para obras de artes especiais e poucos trabalhos foram publicados, conforme Mangalathu et al. (2019). Nesse trabalho, os autores modelaram OAEs de concreto armado e aplicaram métodos de *Machine Learning* como *Naives Bayes*, *decision tree*, *random forest*, *quadratic discriminant analyses* e *K-nearest neighbors* para classificar as OAEs quanto ao risco estrutural frente a um terremoto.

1.1 Justificativa

O Brasil está situado na região central da placa sul-americana, ou seja, está menos suscetível à ocorrência de sismos comparado a regiões de fronteira entre placas conforme Assumpção et al. (2016). Entretanto, não há indicações que terremotos intraplacas são menos intensos que os interplaca. De acordo com Assumpção et al. (2016), enquanto no Brasil um terremoto de magnitude 5 ocorre, em média, a cada cinco anos, um sismo dessa magnitude ocorre, em média, duas vezes por semana em países como Chile e Peru.

Nas últimas décadas eventos sísmicos representativos, de magnitude superiores a 5, aconteceram no Brasil e foram registrados pelo Centro de Sismologia da Universidade de São Paulo (USP). A região Nordeste do Brasil quando comparada ao resto do território nacional apresenta considerável risco sísmico, uma vez que é a terceira região com maior densidade demográfica e possui uma das maiores ameaças sísmicas de acordo com Assumpção et al. (2016) e Petersen et al. (2018). De acordo com Cavalcante (2022), dentre as pontes localizadas no Nordeste do Brasil, três tipologias são predominantes. A terceira tipologia mais frequente é aquela que apresenta sistema estrutural mais complexo com encontros e pórticos intermediários

conectados por aparelhos de apoio elastoméricos, sendo esta tipologia estudada durante este trabalho.

A Associação Brasileira de Normas Técnicas (ABNT), em 2023, lançou uma nova revisão da norma específica para projeto de estruturas sismo-resistentes, a NBR 15421 (2023), no entanto esta ainda exclui sua aplicação em estruturas como as OAEs. Além disso, diferente das normas estrangeiras, como a *National Building Code of Canada* (2005) que utiliza um mapa sísmico com uma probabilidade de 2% de serem ultrapassadas em 50 anos, equivalente a um tempo de retorno (TR) de 2475 anos, a norma brasileira utiliza somente a probabilidade de 10% de serem excedidas no mesmo intervalo de tempo, ou seja, um tempo de retorno de 475 anos.

Sendo assim, a norma brasileira para projetos de estruturas sismo-resistentes apresenta valores de acelerações inferiores aos propostos recentemente por Assumpção et al. (2016) e Petersen et al. (2018). Portanto, parte das estruturas construídas podem estar inseguras quanto à probabilidade aceitável de se atingir os estados de dano, como abertura de fissuras, rupturas locais ou globais, entre outros. Logo, torna-se necessário a realização de estudos referentes à vulnerabilidade sísmica a partir de modelo probabilísticos, conforme realizado em outros países.

O processo de análise a partir de modelos probabilísticos voltados para a avaliação de vulnerabilidade estrutural frente a uma ação sísmica exige um grande custo computacional devido ao excessivo tempo de processamento. Como solução, a utilização de métodos de *machine learning* para estimar respostas quantitativas e qualitativas através de um modelo de previsão, reduzem consideravelmente os custos computacionais de simulação para os estudos de vulnerabilidade sísmica (ROKNEDDIN, 2013).

1.2 Objetivo Geral

O objetivo geral deste trabalho consiste em comparar o estudo de vulnerabilidade sísmica de uma OAE, situadas no nordeste do Brasil, através de métodos probabilísticos como a curva de fragilidade, com métodos de inteligência artificial como o *machine learning*.

1.2.1 Objetivos específicos

- Gerar um banco de dados com informações suficiente para possibilitar a implementação das ferramentas de *machine learning*;
- Obter os resultados da fragilidade sísmica para a OAE em estudo;

- Classificar as OAEs analisadas em relação aos respectivos estados de dano analisados pelo modelo probabilístico;
- Estudar os dois métodos da aprendizagem supervisionada do *machine learning*, classificação e regressão, aplicados ao estudo de vulnerabilidade estrutural de uma OAE de concreto armado;
- Comparar os resultados obtidos através do modelo probabilístico com os resultados obtidos através dos métodos do *machine learning*;
- Obter e validar uma programação generalizada para o estudo de vulnerabilidade de OAEs com sistema estrutural semelhante.

1.3 Organização da dissertação

O presente trabalho está dividido em 5 capítulos: introdução, referencial teórico, metodologia, resultados e considerações finais.

O **capítulo 1** contém a introdução do problema, a justificativa de seu estudo e os objetivos gerais e específicos do presente trabalho.

O **capítulo 2** contém a situação do Brasil, a análise de risco sísmico, as tipologias presentes no nordeste brasileiro, aspectos da geração da curva de fragilidade e conceitos teóricos do método de *machine learning*.

O **capítulo 3** traz a tipologia adotada e a metodologia empregada neste trabalho, tanto para a geração do banco de dados quanto para a modelagem numérica dos modelos de demanda e capacidade que posteriormente são utilizadas na geração das curvas de fragilidade. Além disso, este capítulo também contém a maneira como os métodos de *machine learning* foram empregados.

O **capítulo 4** apresenta os resultados obtidos através do modelo probabilístico e os resultados obtidos através dos métodos de classificação e regressão da aprendizagem supervisionada. Ademais, este capítulo também traz a comparação entre a análise numérica *vs machine learning*.

O **capítulo 5** contém as considerações finais da presente dissertação de mestrado e sugestões de trabalhos futuros.

2 Referencial Teórico

Os eventos sísmicos possuem probabilidades de ocorrência de acordo com a região em estudo, podendo ser divididos entre terremotos intraplaca e interplaca, e para quantificar uma atividade sísmica é utilizada a escala de magnitude que representa a energia liberada durante um terremoto. Os terremotos interplaca são mais frequentes e de maior magnitude quando comparado com os terremotos intraplaca, que caracterizam as atividades sísmicas no Brasil. Devido esta diferença de magnitude e probabilidade de ocorrência, a atividade sísmica intraplaca pode conduzir a interpretação de que estes eventos possuem baixo ou nenhum risco, podendo levar a situações de colapso estrutural e risco a população. De acordo com Assumpção et al. (2016) terremotos no território brasileiro já atingiram magnitude de 6,2, como por exemplo em 1955 no Mato Grosso. Em comparação com o cenário mundial, a Tabela 1 evidencia terremotos, com as respectivas magnitudes, dos anos 2000 até 2010 que tiveram maior número de mortos.

Tabela 1 – Terremotos, de 2000 até 2010, com maior número de mortos (Fonte: adaptado de USGS (2021)).

Ano	Local	Magnitude	Número de mortos
2010	Haiti	7,0	222.570
2009	Indonésia	7,5	1.787
2008	China	7,9	88.011
2007	Peru	8,0	709
2006	Indonésia	6,3	6.605
2005	Paquistão	7,6	82.364
2004	Indonésia	9,1	228.802
2003	Irã	6,6	33.819
2002	Afeganistão	6,1	1.685
2001	Índia	7,7	21.357
2000	Indonésia	7,9	231

2.1 Situação no Brasil

Terremotos com magnitudes superiores a 5 foram registrados no Brasil nas últimas décadas pelo Centro de Sismologia da Universidade de São Paulo (USP). Para demonstrar que as atividades sísmicas de origem intraplaca podem ser tão significativas quanto as de origem interplaca, a Tabela 2 apresenta terremotos, dos anos 2000 até 2019, com magnitude superior a

5. De acordo com Shearer (2009), a magnitude para ondas de corpo (mb) diz respeito a medida pela amplitude máxima da primeira chegada e de acordo com Assumpção (1983), a magnitude regional (mR) é utilizada para as atividades sísmicas captadas pelo Instituto de Astronomia, Geofísica e Ciências da USP, compatíveis com a magnitude mb para o Brasil.

Tabela 2 – Terremotos, de 2000 até 2019, com magnitude superior a 5 (Fonte: adaptado do Centro de Sismologia da Universidade de São Paulo (USP)).

Ano	Local	Magnitude
2019	Região Oeste do Brasil	6,3 mb
2018	Fronteira Peru - Brasil	7,0 mb
2017	Fronteira Peru - Brasil	5,4 mb
2016	Fronteira Peru - Brasil	6,2 mb
2015	Jordão - AC	6,2 mb
2014	Tarauaca - AC	5,2 mb
2010	Mara Rosa - GO	5,0 mb
2008	Plataforma Continental - SP	5,2 mR
2007	Plataforma Continental - AP	5,1 mb
2006	Caiena - AP	5,2 mb
2005	Pto. dos Gaúchos - MT	5,0 mb

Para ilustrar esse cenário, a Figura 1 mostra o catálogo das atividades sísmicas que foram detectados por instrumentos e as atividades definidas por relatos históricos no Brasil (ASSUMPCÃO et al., 2016). Pode-se notar a diferença entre as atividades sísmicas nas regiões Norte e Sudeste, por exemplo, onde na região Norte foram registrados terremotos mais espaçados e de maior magnitude, e na região Sudeste encontram-se sismos registrados e relatados de menor magnitude e com maior concentração. De acordo com Assumpção et al. (2016), a diferença entre as atividades sísmicas registradas pode ser explicada tanto pela existência de um rede de sismógrafos quanto o número de habitantes da região, ou seja, é provável que a região Norte tenha sofrido mais terremotos, mas por conta da falta de instrumentação e locais desabitados essas atividades sísmicas não foram sentidas. Portanto, uma melhor maneira de se avaliar quais são as regiões mais ou menos ativas é filtrar o catálogo de sismos mostrando apenas os sismos com magnitudes acima de um algum limite que dependa do tempo. A Figura 2 ilustra um mapa de epicentros com este catálogo filtrado, conhecido como catálogo uniforme, com os seguintes critérios: magnitude acima de 6.0 desde 1940, magnitude maior que 5.0 desde 1962, maior 4.5 desde 1968 e maior que 3.5 desde 1980. Os locais que apresentam maiores atividades sísmicas dentro do Brasil são a região do Pantanal, o estado do Ceará e Rio Grande do Norte e a região sul do estado de Minas Gerais (ASSUMPCÃO et al., 2016).

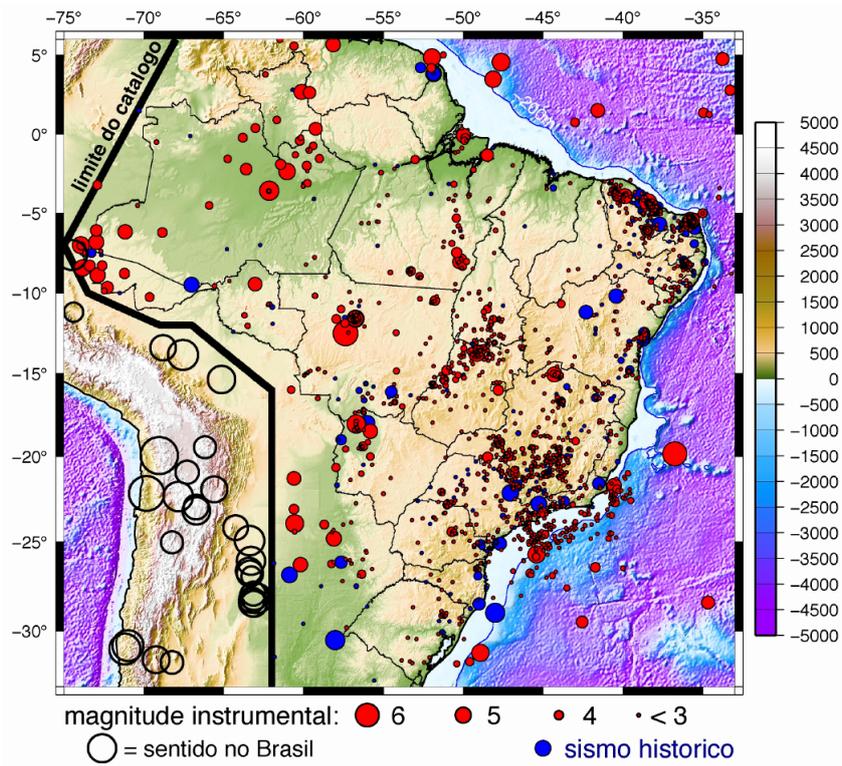


Figura 1 – Sismos do Brasil de 1720 a 2016 (Fonte: Assumpção et al. (2016)).

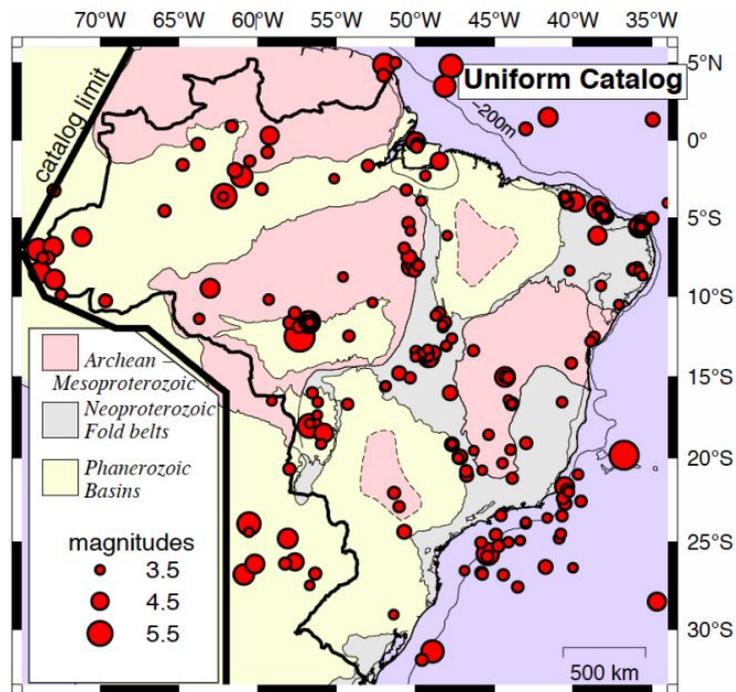


Figura 2 – Catálogo uniforme, filtrado conforme detectabilidade (Fonte: adaptado de Assumpção et al. (2016)).

Diante da necessidade de considerar as ações sísmicas no dimensionamento das estruturas brasileiras devido aos estudos realizados, em 2023, a Associação Brasileira de Normas Técnicas lançou a nova revisão da norma para projeto de estruturas sísmo-resistente, a NBR 15421 (2023). A norma traz um mapa de ameaça sísmica elaborado de acordo com os resultados obtidos por Shedlock e Tanner (1999) no projeto *Global Seismic Hazard Assessment Project (GSHAP)* e está representado na Figura 3. Este mapa sísmico da norma brasileira foi elaborado para solo classe B (rocha) e traz uma probabilidade de 10% de serem excedidas em 50 anos, ou seja, um tempo de retorno de 475 anos. De acordo com a Figura 3, podemos separar as zonas sísmicas com seus respectivos valores de aceleração, conforme apresentado na Tabela 3.

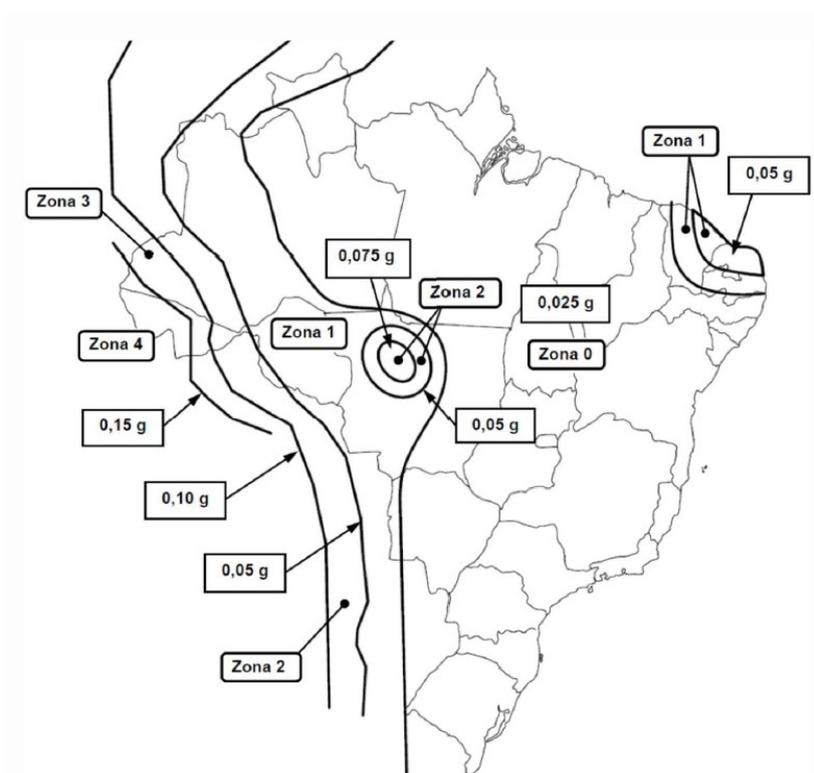
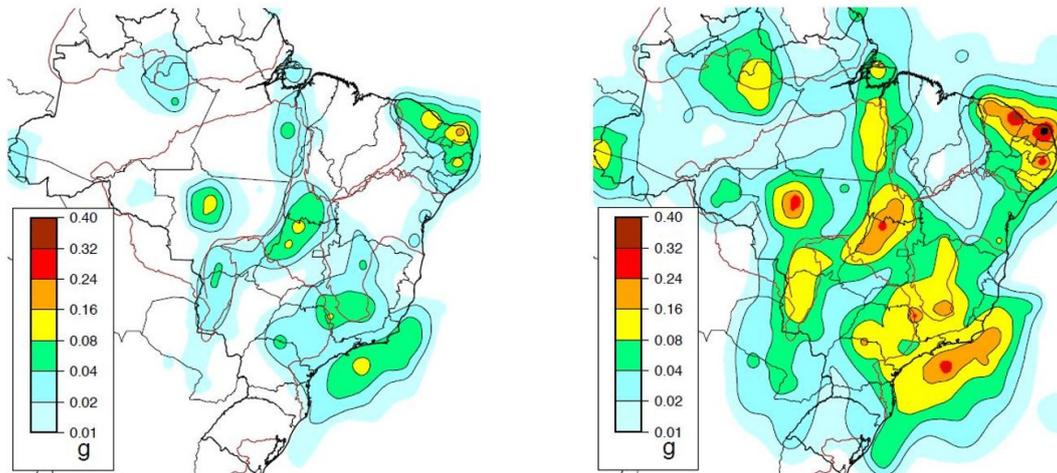


Figura 3 – Mapeamento da aceleração sísmica no Brasil (Fonte: NBR 15421 (2023)).

Tabela 3 – Divisão das zonas sísmicas no Brasil (Fonte: adaptado da NBR 15421 (2023)).

Valores da aceleração sísmica - a_g (g)	Zona sísmica
$= 0,025$	Zona 0
$0,025 \leq a_g \leq 0,05$	Zona 1
$0,05 \leq a_g \leq 0,10$	Zona 2
$0,10 \leq a_g \leq 0,15$	Zona 3
$a_g = 0,15$	Zona 4

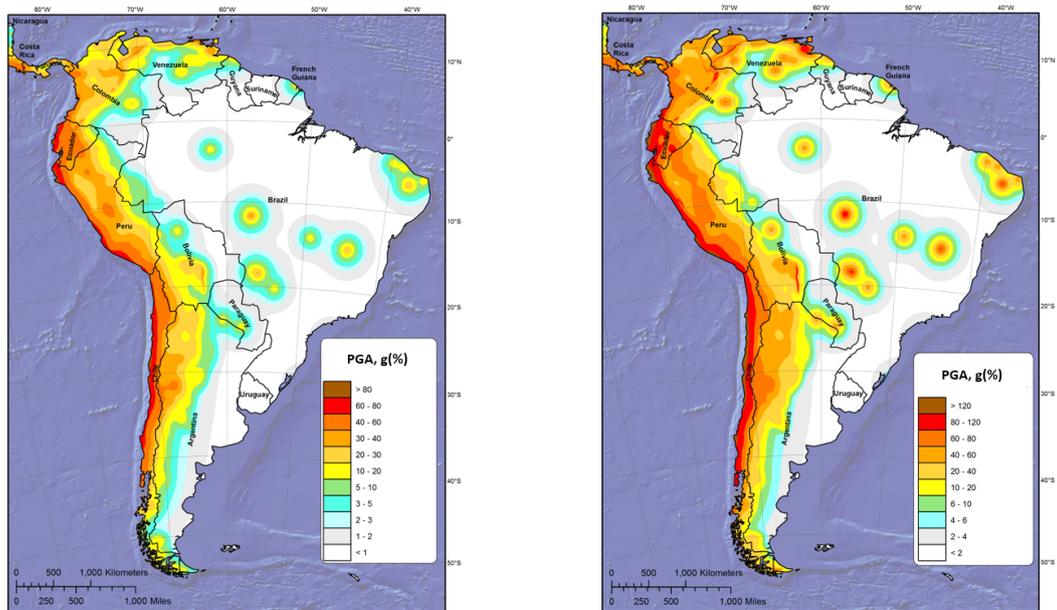
Assumpção et al. (2016) e Petersen et al. (2018) elaboraram mapas de ameaças sísmicas para o território brasileiro. A Figura 4 ilustra os mapas de ameaça sísmica para o Brasil, elaborado por Assumpção et al. (2016), com probabilidades de serem excedidas de 10% e 2% em 50 anos, ou seja, totalizando um tempo de retorno igual a 475 e 2475 anos, respectivamente. A Figura 5 ilustra os mapas de ameaça sísmica para o território brasileiro, elaborado por Petersen et al. (2018), com mesmas probabilidades de excedência.



a) Probabilidade de 10% em 50 anos.

b) Probabilidade de 2% em 50 anos.

Figura 4 – Mapas de ameaça sísmica para aceleração de pico em rocha, para tempo de retorno de 475 e 2475 anos, respectivamente (Fonte: adaptado de Assumpção et al. (2016)).

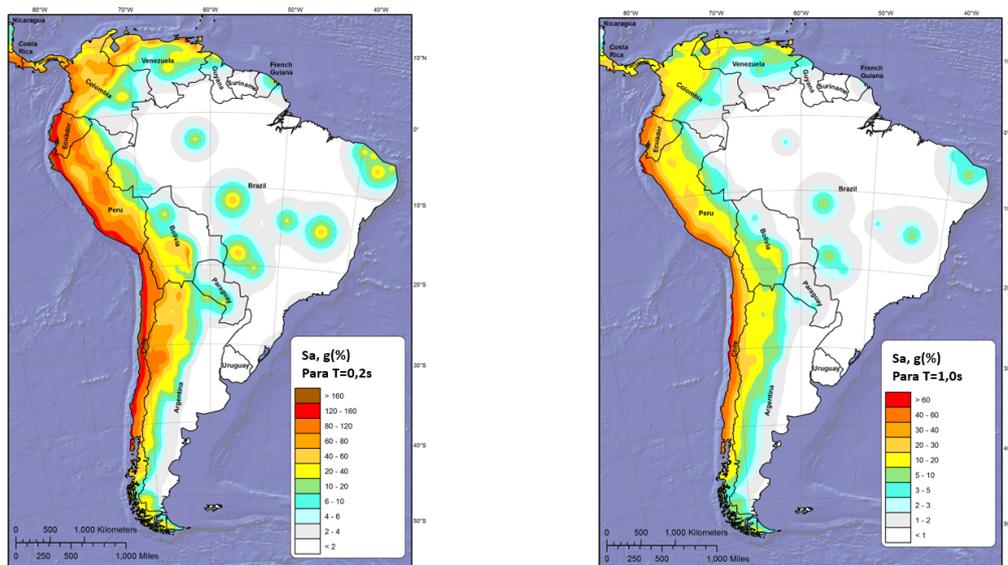


a) Probabilidade de 10% em 50 anos.

b) Probabilidade de 2% em 50 anos.

Figura 5 – Mapas de ameaça sísmica para aceleração de pico em rocha, para tempo de retorno de 475 e 2475 anos, respectivamente (Fonte: adaptado de Petersen et al. (2018)).

Além dos mapas de ameaça sísmica ilustrados na Figura 5, Petersen et al. (2018) também desenvolveu mapas de ameaça sísmica levando em consideração as diretrizes recomendadas no BSSC (2015) e no ASCE (2017) para os Estados Unidos, baseados no *United States Geological Survey (USGS)* e no *National Seismic Hazard Models (NSHMs)*. Estes mapas de ameaça sísmica consideram a probabilidade de excedência de 2% e 10% em 50 anos, como os mapas já expostos, para acelerações com períodos de vibração de 0,2 e 1,0 segundos com 5% de amortecimento e velocidade média de propagação maior que 760 m/s, velocidade esta compatível com solo classe B (rocha) de acordo com a NBR 15421 (2023). As Figuras 6 e 7 demonstram estes mapas de ameaça sísmica elaborados por Petersen et al. (2018).

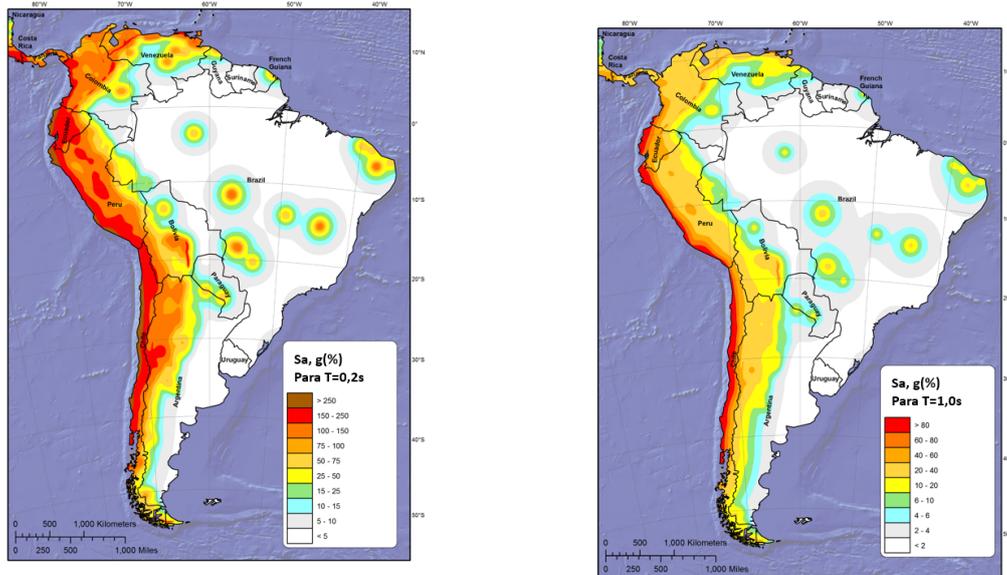


a) Probab. de 10% em 50 anos para $T = 0,2s$.

b) Probab. de 10% em 50 anos para $T = 1s$.

Figura 6 – Mapas de ameaça sísmica para acelerações com período de vibração de 0,2 e 1,0 segundo, para um tempo de retorno de 475 anos, respectivamente (Fonte: adaptado de Petersen et al. (2018)).

Portanto, a norma brasileira de projetos de estruturas sísmo-resistentes, a NBR 15421 (2023), além de excluir a sua aplicação em estruturas como as OAEs, ela também apresenta valores de acelerações inferiores aos propostos recentemente por Assumpção et al. (2016) e Petersen et al. (2018). Logo, parte das estruturas podem estar inseguras quanto à probabilidade aceitável de se atingir determinado estado de dano, tornando-se assim necessários estudos referentes à vulnerabilidade sísmica.



a) Probab. de 2% em 50 anos para $T = 0,2s$.

b) Probab. de 2% em 50 anos para $T = 1s$.

Figura 7 – Mapas de ameaça sísmica para acelerações com período de vibração de 0, 2 e 1, 0 segundo, para um tempo de retorno de 2475 anos, respectivamente (Fonte: adaptado de Petersen et al. (2018)).

2.2 Análise de risco sísmico

De maneira geral o risco pode ser definido como a combinação da probabilidade de um evento e as suas respectivas consequências negativas. Analogamente, pode-se definir risco como o potencial de prejuízos ou danos a patrimônios e o potencial de perda de vida que podem ocorrer em um período específico em um sistema ou sociedade, conforme é definido pelo *United Nations Office for Disaster Risk Reduction* (UNDRR, 2019). Desta forma, o risco pode ser determinado como uma função dos seguintes fatores: vulnerabilidade, exposição e ameaça, de acordo com UNDRR (2019), conforme ilustrado na Figura 8.

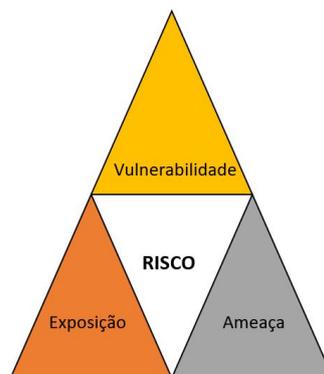


Figura 8 – Fatores que compõe o risco (Fonte: autor).

Os fatores que compõem o risco, de acordo com UNDRR (2019), podem ser definidos como: vulnerabilidade: características ou circunstâncias de uma comunidade, sistema ou patrimônio que deixam suscetível aos efeitos prejudiciais de um desastre; exposição: pessoas, propriedades, sistemas, ou outros elementos presentes em zonas de risco que estão, portanto, sujeitos a potenciais perdas; ameaça: insinuação de que algo desagradável ou perigoso vai acontecer especialmente se uma determinada ação ou ordem não for seguida.

Dentro do âmbito sísmico, McGuire (2004) define a ameaça sísmica ou perigo sísmico como as propriedades de um terremoto que podem causar perdas ou danos como, por exemplo, a frequência com que uma intensidade sísmica ocorre. A ameaça sísmica pode ser calculada pela *Probabilistic Seismic Hazard Analysis (PSHA)*, que mede a probabilidade de excedência de valores de medidas de intensidade para uma determinada localidade de acordo com McGuire (2004) e Baker (2008).

Como exposto acima, o fator da exposição está diretamente relacionado ao número de habitantes de uma região, ou seja, a probabilidade de potenciais perdas é maior em regiões com maior densidade demográfica quando comparada com regiões pouco habitadas. O fator da vulnerabilidade dentro do contexto de risco sísmico significa entender e estudar o quão suscetível uma região e as suas estruturas estão aos efeitos de um evento sísmico. Em outras palavras, o fator da vulnerabilidade procura entender a probabilidade de ser gerado um dano em uma estrutura durante um evento sísmico de determinada intensidade.

Estudos já foram realizados para avaliar a vulnerabilidade de estruturas submetidas a um terremoto no Brasil. Estes estudos fazem parte de um projeto de pesquisa conjunto desenvolvido no *Risk Engineering Laboratory (RELab)* da Universidade Estadual de Campinas (UNICAMP). Por exemplo, Cavalcante (2022) estudou a vulnerabilidade sísmica de pontes de concreto armado, enquanto Pereira (2021) e Rodrigues (2021) estudaram a vulnerabilidade de edifícios de concreto armado.

2.3 Tipologias estruturais

A região Nordeste do Brasil conta com aproximadamente 2500 pontes rodoviárias sob administração do Governo Federal. De acordo com Tavares (2012), é necessário a elaboração de um inventário destas estruturas por motivos de ineficiência computacional de se avaliar cada uma individualmente. O estudo desenvolvido por Cavalcante (2022) utilizou 250 pontes, o que representa 10% do total de pontes, para representar e caracterizar as estruturas da região Nordeste. Os dados foram levantados de maneira randômica através do Sistema de Gerenciamento de Obras de Arte (SGO) elaborado pelo Departamento Nacional de Infraestrutura e Transportes

(DNIT) (CAVALCANTE, 2022). A Tabela 4 demonstra o número de OAEs analisadas, em comparação com o número de OAEs totais, por estado, elaborado por Cavalcante (2022).

Tabela 4 – OAEs analisadas por estado da região Nordeste do Brasil (Fonte: adaptado de Cavalcante (2022)).

Estado	OAEs analisadas	OAEs totais
Alagoas	20	166
Bahia	30	434
Ceará	20	377
Maranhão	35	247
Paraíba	30	280
Pernambuco	40	573
Piauí	30	183
Rio Grande do Norte	30	203
Sergipe	15	81
Total	250	2504

Dentre as características obtidas do SGO, de acordo com Cavalcante (2022), estão: número, comprimento e largura dos vãos; número e dimensões das longarinas; ano de construção; tipo de aparelho de apoio; dimensões e tipo dos pilares e encontros etc. O estudo realizado por Cavalcante (2022) considerou somente pontes de concreto armado, moldadas *in-loco*, retas sem esconsidade e apoiadas por aparelhos de apoio elastoméricos, onde a representatividade de cada uma destas características se aproxima ou é maior que 90% das pontes analisadas.

Após coletadas, Cavalcante (2022) separou as 250 pontes em sete diferentes tipologias em função do sistema estrutural apresentado pelas pontes conforme indicado a seguir:

- Tipologia 1 (T1): vão único; possui encontros rotulados; não possui pórticos internos;
- Tipologia 2 (T2): múltiplos vãos contínuos; não possui encontros; possui pórticos internos rotulados;
- Tipologia 3 (T3): múltiplos vãos contínuos; possui encontros rotulados; possui pórticos internos rotulados;
- Tipologia 4 (T4): vão único; possui encontros engastados; não possui pórticos internos;
- Tipologia 5 (T5): múltiplos vãos contínuos; possui encontros rotulados; possui pórticos internos rotulados;

- Tipologia 6 (T6): múltiplos vãos contínuos; possui encontros rotulados; possui pórticos internos engastados;
- Tipologia 7 (T7): múltiplos vãos contínuos; não possui encontros; possui pórticos internos engastados.

As ligações caracterizadas como engastadas representam vinculações monolíticas entre as componentes estruturais e as estruturas que não possuem encontros, ou seja, que possuem vãos que estão em balanço, dispõem de cortinas para manter e conter a massa de solo. Cavalcante (2022) elaborou um mapa de localização das 250 pontes analisadas em função das tipologias estipuladas, junto ao percentual representativo de cada tipologia, na região Nordeste do Brasil ilustrado na figura 9.

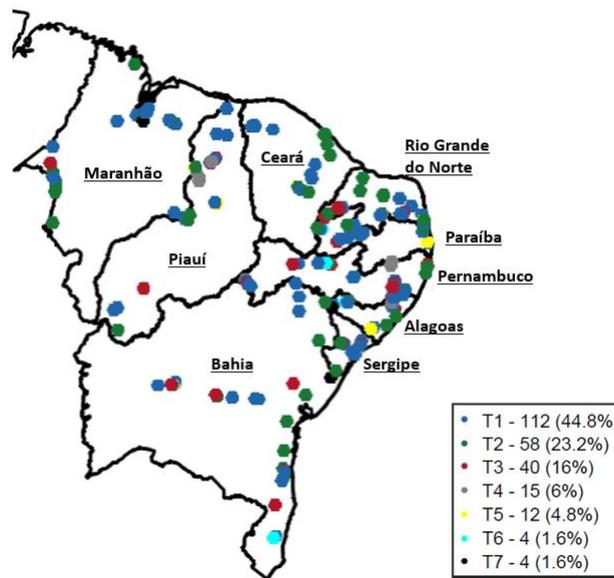


Figura 9 – Mapa de localização das tipologias na Região Nordeste do Brasil (Fonte: adaptado de Cavalcante (2022)).

Pode-se observar que as tipologias T1, T2 e T3 totalizam 210 das 250 pontes analisadas, ou seja, um percentual de 84% e, desta forma, apenas essas três tipologias foram consideradas como representativas para a região Nordeste brasileira (Cavalcante (2022)).

2.4 Curvas de fragilidade

A fragilidade de uma estrutura frente a um evento sísmico pode ser avaliada através das curvas de fragilidade, que medem a probabilidade da estrutura ou uma componente estrutural atingir ou ultrapassar um determinado estado limite de dano (os estados limites são definidos no

item 2.4.2) em função de uma medida de intensidade (*Intensity Measure - IM*) (NIELSON, 2005). A partir da resposta estrutural a um evento sísmico, referida neste trabalho como demanda estrutural, demanda sísmica ou somente demanda, e considerando a capacidade da estrutura ou de uma componente estrutural esta probabilidade pode ser determinada. A Equação 2.4.1 determina a probabilidade condicional da demanda estrutural (D) se igualar ou exceder a capacidade estrutural (C) em função de uma intensidade de medida (IM).

$$P(IM) = P[D \geq C|IM] \quad (2.4.1)$$

As curvas de fragilidade podem ser construídas através da análise de demanda e da capacidade estrutural e podem ser determinadas, quando estas são consideradas como variáveis aleatórias que seguem uma distribuição lognormal, através da Equação 2.4.2 de acordo com Nielson e DesRoches (2007). Este método, conhecido como método analítico, é realizado através de simulações computacionais voltadas a análise estrutural e pode ser observado em diversos trabalhos como, por exemplo, Siqueira et al. (2014a). Outros métodos para determinar as curvas de fragilidade podem ser realizadas, como: métodos empíricos através do reconhecimento dos danos estruturais em terremotos reais (LALLEMANT et al., 2015); através da avaliação e opinião profissional (PORTER et al., 2007); métodos híbridos que representam a utilização de todos os métodos citados a cima (KAPPOS et al., 2006).

$$P[LS|IM] = \Phi \left[\frac{\ln\left(\frac{S_D}{S_C}\right)}{\sqrt{\beta_{D|IM}^2 + \beta_C^2}} \right] \quad (2.4.2)$$

Sendo: $P[LS|IM]$ a probabilidade de se atingir um estado limite (*Limit State - (LS)*) dada uma intensidade de medida (*Intensity Measure - (IM)*) de uma atividade sísmica; Φ a função de distribuição normal acumulada (CDF); S_D a mediana da demanda estrutural; S_C a mediana da capacidade; $\beta_{D|IM}$ a dispersão logarítmica da demanda e β_C a dispersão logarítmica da capacidade.

De acordo com Silva et al. (2014), a probabilidade da demanda se igualar ou exceder um determinado estado-limite de dano pode ser calculada a partir de análises dinâmicas não-lineares através do uso de modelos numéricos computacionais. A capacidade estrutural condiciona os estados-limites de dano de uma estrutura, ou de uma componente estrutural, e são divididos em cinco níveis de dano: ausência de dano, leve, moderado, extensivo e completo. Já os valores de demanda, que representam a resposta estrutural a um terremoto, são obtidos em função de uma medida de intensidade (IM) que corresponde a uma grandeza das atividades sísmicas. A Tabela 5 a seguir demonstra alguns exemplos de medidas de intensidades utilizadas.

Tabela 5 – Exemplos de medidas de intensidades (Fonte: adaptado de Freddi et al. (2017)).

Medida de Intensidade	Definição
S_{a-T_1}	Aceleração espectral no período fundamental de vibração
$S_{a-0,2s}$	Aceleração espectral de período curto
$S_{a-1,0s}$	Aceleração espectral de período longo
S_v	Velocidade espectral no período fundamental de vibração
S_d	Deslocamento espectral no período fundamental de vibração
PGA	<i>Peak Ground Acceleration</i> - Aceleração máxima do solo
PGV	<i>Peak Ground Velocity</i> - Velocidade máxima do solo
PGD	<i>Peak Ground Displacement</i> - Deslocamento máximo do solo

Conforme exposto acima, para calcular a fragilidade da estrutura ou de uma componente estrutural torna-se necessário a definição do estado-limite. O estado limite é definido com relação ao comportamento estrutural através de medidas de interesse conhecidas como Parâmetros de Demanda de Engenharia (em inglês *Engineering Demand Parameters* - (*EDP*)). De acordo com Freddi et al. (2017), alguns exemplos destas medidas são: deformação máxima do concreto ou da armadura; curvatura máxima; esforço cortante máximo; momento fletor máximo; *Drift*; deslocamento máximo relativo. Portanto, um estado-limite é ultrapassado quando a demanda estrutural em função de uma determinada medida de intensidade (IM) supera o valor da capacidade estrutural atrelada ao respectivo estado-limite, como demonstra a Equação 2.4.1.

A seguir serão discutidas as técnicas utilizadas no cálculo de demanda e capacidade para a determinação das curvas de fragilidade através do método analítico, ou seja, através de simulações computacionais voltadas a análise estrutural. Além do mais, a seguir também será discutido a medida de intensidade (IM) utilizada neste trabalho.

2.4.1 Demanda sísmica

Segundo Bachman et al. (2004), a demanda sísmica representa a resposta estrutural de uma componente ou de um sistema frente a um terremoto e que pode ser utilizado para determinar os danos causados pela ação sísmica. A resposta de uma estrutura quando solicitada à uma ação sísmica pode ser calculada a partir de análises dinâmicas através de modelos numéricos. Desta forma, para se realizar uma análise dinâmica torna-se necessário a determinação de algumas características intrínsecas à estrutura como suas frequências naturais e modos de vibração. De acordo com Paultre (2013), a resposta dinâmica de uma estrutura pode ser determinada através de uma combinação linear de um conjunto de movimentos harmônicos simples chamados de modos naturais de vibração, onde cada modo representa uma parcela de participação na vibração

do sistema. A vibração causada em uma estrutura quando solicitada por uma carga sísmica não possui frequência determinada, uma vez que a resposta a uma ação sísmica apresenta dependência do período natural da estrutura, como pode ser observado no gráfico de espectro de resposta ilustrado na Figura 10.

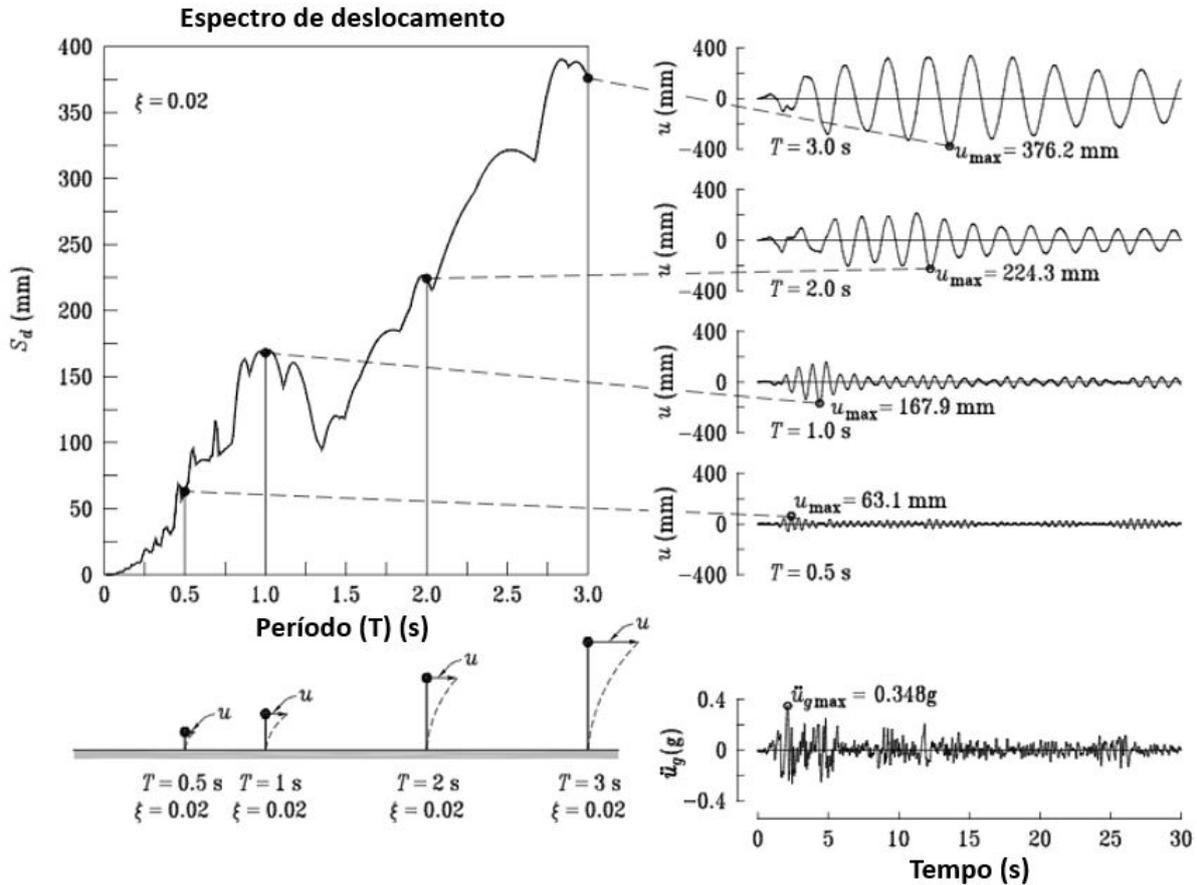


Figura 10 – Espectro de resposta para estruturas com um grau de liberdade com mesma taxa de amortecimento e períodos naturais diferentes solicitados por um mesmo terremoto (Fonte: adaptado de Paultre (2013)).

Como pode ser observado na Figura 10, a resposta dinâmica de um sistema estrutural varia de acordo com o seu respectivo período natural de vibração, portanto o espectro de resposta é um gráfico que representa o máximo valor de um parâmetro específico, seja de deslocamento, velocidade ou aceleração, em função do período natural de um sistema estrutural idealizado de um grau de liberdade.

Para determinar as respostas estruturais frente a um terremoto pode-se aplicar análises dinâmicas não-lineares em função de um histórico de acelerações que varia no tempo. A Equação 2.4.3 descreve o equilíbrio dinâmico das respostas estruturais em função de uma ação que varia

no tempo, como por exemplo, um terremoto.

$$\mathbf{p}_{ef}(t) = \mathbf{M} \cdot \ddot{\mathbf{u}}(t) + \mathbf{C} \cdot \dot{\mathbf{u}}(t) + \mathbf{K} \cdot \mathbf{u}(t) \quad (2.4.3)$$

Sendo: $\mathbf{p}_{ef}(t)$ o vetor de força efetivo; \mathbf{M} a matriz de massa; $\ddot{\mathbf{u}}(t)$ o vetor de aceleração; \mathbf{C} a matriz de amortecimento; $\dot{\mathbf{u}}(t)$ o vetor de velocidade; \mathbf{K} a matriz de rigidez e $\mathbf{u}(t)$ o vetor deslocamento.

Os registros sísmicos utilizados nas análises de fragilidade podem ser do tipo escalonado ou não escalonado, ou seja, os valores de aceleração no registro sísmico escalonado são modificados por um fator de escala enquanto o não escalonado mantém os registros conforme foram medidos. O uso do escalonamento permite a obtenção de diferentes registros para qualquer valor de medida de intensidade (IM) e, desta forma, torna-se possível obter uma distribuição da demanda estrutural para cada IM. De acordo com Vamvatsikos e Cornell (2002), o registro escalado é utilizado no método da análise dinâmica incremental (em inglês *Incremental Dynamic Analysis - IDA*) e eles a definem como uma análise paramétrica de um modelo numérico submetido a um ou mais registros sísmicos escalonados.

Por outro lado, os registros não escalonados são utilizados em métodos de análises dinâmicas não incrementais, onde os sinais são empregados sem qualquer escalonamento da medida de intensidade (IM). Um método que, por exemplo, utiliza registros não escalonado é o *Cloud Analysis*. Este método quando aplicado com uma pequena quantidade de registros sísmicos pode gerar respostas pouco representativas, pelo fato de apresentar maior dificuldade na determinação de inúmeros valores de demanda para um mesmo valor de IM. Contudo, quando aplicado com uma grande quantidade de registros sísmicos torna-se possível a determinação de vários valores de demanda onde cada resposta pode ser associada a um único registro e, portanto, uma regressão entre os registros e as respectivas demandas torna-se possível.

De acordo com Cornell et al. (2002), o método do *Cloud Analysis* pode ser utilizado para determinar a demanda em função de uma medida de intensidade (IM), ou seja, obter um valor da resposta estrutural para cada registro sísmico. Desta forma, a utilização de um número razoável de registros sísmicos resulta na geração de um grande conjunto de dados, ou “nuvem”, de pares demanda *vs* medida de intensidade (D-IM) onde um modelo de regressão pode ser aplicado. Atrelado ao modelo de regressão está a dispersão causada pela variabilidade dos modelos estruturais, conforme demonstrado na Tabela 12 e pela variabilidade dos diferentes registros sísmicos de mesma intensidade de medida utilizados.

Após realizadas as análises dinâmicas para a obtenção dos valores de demanda em função de uma IM aplica-se o modelo de regressão chamado de Modelo Probabilístico de Demanda

Sísmica (em inglês *Probabilistic Seismic Demand Model (PSDM)*). De acordo com Cornell et al. (2002), a relação entre a demanda estrutural e a intensidade de medida pode ser relacionada como uma reta no espaço logarítmico conforme a Equação 2.4.4 a seguir:

$$\ln(\hat{D}) = \ln(a) + b \cdot \ln(IM) \quad (2.4.4)$$

Sendo: $\ln(\hat{D})$ o valor médio da resposta estrutural de um determinado Parâmetro de Demanda de Engenharia (EDP) analisado no espaço logarítmico; $\ln(a)$ o coeficiente linear da reta de regressão; b o coeficiente angular desta reta.

Considerando que tanto a demanda sísmica quanto a capacidade estrutural seguem distribuições logarítmicas e utilizando os parâmetros obtidos através do PSDM, tem-se a Equação 2.4.5.

$$P[LS|IM] = \Phi \left[\frac{\ln(IM) - \theta}{\beta} \right] \quad (2.4.5)$$

$$\beta = \frac{\sqrt{\beta_{D|IM}^2 + \beta_C^2}}{b} \quad (2.4.6)$$

$$\theta = \exp \left[\frac{\ln(S_C) - \ln(a)}{b} \right] \quad (2.4.7)$$

O valor da dispersão $\beta_{D|IM}$ do modelo de regressão do PSDM pode ser obtida através da Equação 2.4.8.

$$\beta_{D|IM} = \sqrt{\frac{\sum \left[\ln(Di) - \ln \left[a \cdot (IM^b) \right] \right]^2}{n - 2}} \quad (2.4.8)$$

Sendo: Di a demanda estrutural, ou seja, a resposta máxima de um determinado Parâmetro de Demanda de Engenharia (EDP) e n o número de simulações computacionais da análise estrutural.

Segundo Padgett et al. (2008) e Freddi et al. (2017) a medida de intensidade escolhida para a análise de fragilidade deve apresentar características como: eficiência - menor dispersão $\beta_{D|IM}$ da demanda em função da medida de intensidade (IM); praticidade - relação direta entre a medida de intensidade e a demanda estrutural calculada pelo coeficiente angular b da reta do PSDM; suficiência - valores da demanda estrutural independentes das características dos eventos sísmicos (magnitude e distância do epicentro); proficiência - melhor relação entre a

dispersão e o coeficiente angular. A Figura 11 a seguir ilustra o *Probabilistic Seismic Demand Model* (PSDM).

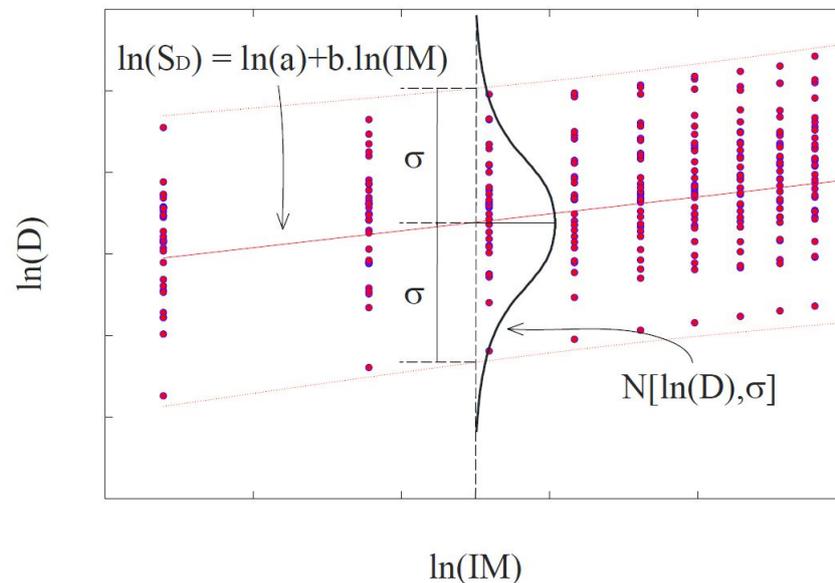


Figura 11 – Ilustração do PSDM no espaço logarítmico (Fonte: adaptado de Cavalcante (2022)).

2.4.2 Capacidade estrutural

A determinação da capacidade estrutural, ou de uma componente estrutural, é um fator preponderante no estudo de fragilidade de uma estrutura, pois é a capacidade que condiciona os estados-limites (LS) de dano de uma estrutura. Diversos estudos de fragilidade sobre pontes, como por exemplo Mangalathu et al. (2017) e Nielson e DesRoches (2006), recorrem a *Federal Emergency Management Agency* (FEMA) para definir os diferentes estados-limites de dano. De acordo com a FEMA (2010), os LS são:

- Sem Dano (*none*): não há dano ao sistema estrutural ou a algum componente estrutural;
- Dano Leve (*slight/minor damage*): definido por fissuras (aberturas de até 1 mm de espessura) e/ou leve deslocamento de concreto nos encontros, apoios, pilares ou tabuleiro. Somente reparos estéticos são necessários;
- Dano Moderado (*moderate damage*): definido por trincas (aberturas de 1 a 3 mm de espessura) de cisalhamento e deslocamento do concreto dos pilares (ainda apresentando integridade estrutural), movimentação menor que 5 cm nos encontros, fissuras em aparelhos de apoio que possuam chave de cisalhamento ou parafusos chumbados aos pórticos, ruptura do aparelho de apoio do tipo *rocker* (com pino que possibilita a rotação) ou recalque moderado da laje de acesso;

- Dano Extensivo (*extensive damage*): definido pelo processo de degradação dos pilares por ruptura de cisalhamento (integridade do pilar comprometida), deslocamento residual significativo das conexões, recalque significativo da laje de acesso, deslocamento vertical dos encontros, recalque diferencial nas conexões e ruptura das chaves de cisalhamento nos encontros;
- Dano Completo (*complete damage*): definido pelo colapso de qualquer pilar, perda de contato com os aparelhos de apoio (que provavelmente levará ao colapso do tabuleiro) e inclinação da infraestrutura devido ao colapso da fundação.

De maneira similar à demanda, também é assumido que a capacidade da estrutura ou de uma componente estrutural segue uma distribuição lognormal. Desta forma, para quantificar os respectivos estados-limites (LS) de dano expostos acima, torna-se necessário a criação de modelos que possibilitem a extração de parâmetros da distribuição lognormal da capacidade, como a mediana e dispersão logarítmica. A mediana e a dispersão podem ser definidas conforme as Equações 2.4.9 e 2.4.10 abaixo, respectivamente:

$$S_C = e^{\left[\ln(\mu_X) - \left(\frac{\sigma_Y^2}{2} \right) \right]} = e^{\mu_Y} \quad (2.4.9)$$

$$\beta_C = \sqrt{\ln \left[\left(\frac{\sigma_X}{\mu_X} \right)^2 + 1 \right]} = \sigma_Y \quad (2.4.10)$$

Sendo: S_C e β_C a mediana e a dispersão logarítmica, respectivamente; σ_X e μ_X o desvio padrão e a média de uma função de distribuição normal, respectivamente; σ_Y e μ_Y o desvio padrão e a média de uma função de distribuição lognormal, respectivamente.

No entanto, alguns autores como Nielson e DesRoches (2006), Choi et al. (2004), Tavares et al. (2012) e Mangalathu et al. (2017) apresentam valores de mediana (S_C) e dispersão logarítmica (β_C) para estruturas como as pontes. Os valores propostos por Mangalathu et al. (2017) foram adotados neste trabalho e estão demonstrados na Tabela 6.

Os valores da mediana S_C e dispersão β_C que estão representadas com “ - ” indicam que determinado estado-limite de dano para um determinado parâmetro de engenharia não é atingido por estar condicionado a outro estado-limite de dano de outro parâmetro de engenharia. Por exemplo, a possível mudança do estado-limite de dano moderado para extensivo para o deslocamento longitudinal do aparelho de apoio não ocorre por estar condicionado ao estado-limite de dano extensivo para o deslocamento do tabuleiro.

Tabela 6 – Valores adotados de S_C e β_C (Fonte: adaptado de Mangalathu et al. (2017)).

Parâmetros de Engenharia	Estado-Limite	Mediana - S_C	Dispersão - β_C
Deslocamento longitudinal do aparelho de apoio (BDL) (mm)	Leve	25	0,35
	Moderado	76	0,35
	Extensivo	-	-
	Completo	-	-
Deslocamento transversal do aparelho de apoio (BDT) (mm)	Leve	25	0,35
	Moderado	76	0,35
	Extensivo	-	-
	Completo	-	-
Deslocamento do tabuleiro (DU) (mm)	Leve	-	-
	Moderado	-	-
	Extensivo	152	0,35
	Completo	229	0,35
Deslocamento ativo do encontro (ADA) (mm)	Leve	38	0,35
	Moderado	102	0,35
	Extensivo	-	-
	Completo	-	-
Deslocamento passivo do encontro (ADP) (mm)	Leve	76	0,35
	Moderado	254	0,35
	Extensivo	-	-
	Completo	-	-
Deslocamento transversal do encontro (ADT) (mm)	Leve	25	0,35
	Moderado	102	0,35
	Extensivo	-	-
	Completo	-	-

A capacidade estrutural dos pilares é comumente obtida através de análises do tipo *Pushover*. Este tipo de análise consiste em submeter a estrutura do pilar a uma análise estática não linear, através da aplicação de cargas gravitacionais constantes e incrementos de carga horizontais aplicadas de maneira gradativa. De forma geral, o objetivo da análise *Pushover* é estabelecer o mecanismo plástico de resposta da estrutura e a distribuição do dano que é obtida através de uma curva de capacidade que representa uma função entre o deslocamento no topo da estrutura e o esforço cortante na base ou uma função do momento fletor total na base do pilar pela curvatura da seção transversal mais crítica. A Figura 12 a seguir ilustra a curva de capacidade obtida para um pórtico com dois pilares.

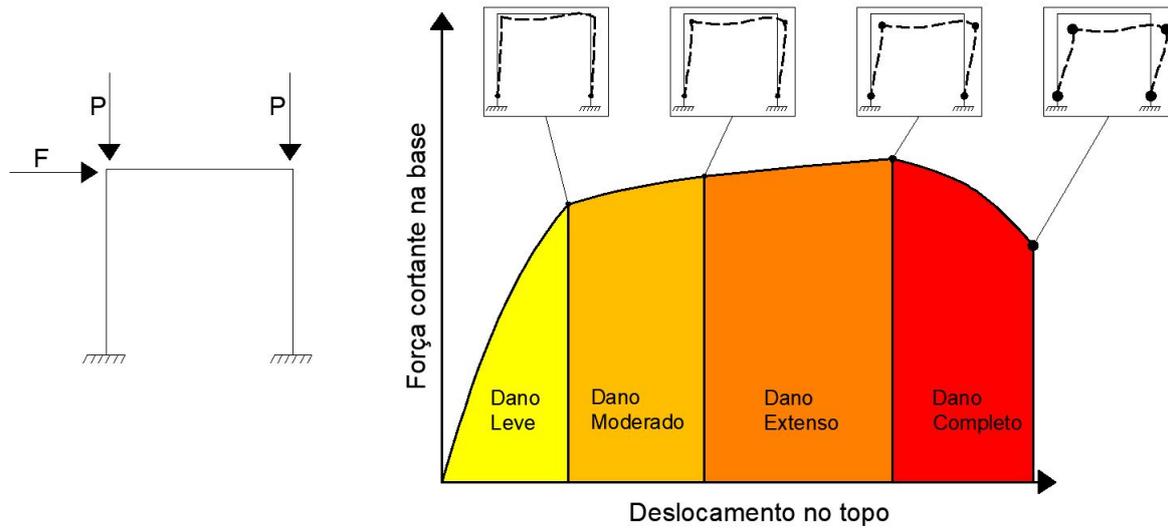


Figura 12 – Curva de capacidade com os Estados Limites (LS) de Dano (Fonte: autor).

As curvas de dano para pilares da tipologia T3 foram estabelecidas no estudo desenvolvido por Cavalcante (2022). O Parâmetro de Demanda de Engenharia (EDP) utilizado neste trabalho para os pilares é a curvatura ϕ da seção transversal. Os valores de curvatura podem ser calculados a partir dos valores de demanda de ductilidade I' , como demonstrado na Equação 2.4.11 a seguir:

$$\phi = I' \cdot \phi_y \tag{2.4.11}$$

Sendo: ϕ_y o valor da curvatura em que ocorre o primeiro escoamento das armaduras longitudinais (em ingles *Longitudinal Reinforcement Yielding Stress*) do pilar mais solicitado.

De acordo com Cavalcante (2022), os valores obtidos, da tipologia T3, através das análises de capacidade do tipo *Pushover* estão demonstrados na Tabela 7.

Tabela 7 – S_C e β_C obtidos através da análise *Pushover* (Fonte: adaptado de Cavalcante (2022)).

EDP	LS	Ductilidade - I'	Mediana - S_C	Dispersão - β_C
Curvatura da da seção transv. do pilar (CC) (1/m)	Leve	1,0	0,007	0,22
	Moderado	2,0	0,013	0,22
	Extensivo	4,0	0,027	0,22
	Completo	7,0	0,047	0,22

2.5 Machine Learning

Segundo Xie et al. (2020), os métodos de *Machine Learning* (ML), procedentes da inteligência artificial, têm evoluído rapidamente através dos anos como a promessa de alterar substancialmente e melhorar o papel da ciência dos dados (em inglês *Data Science*) em diversas disciplinas. Os avanços das técnicas de ML vêm contribuindo em diversas áreas da ciência e da engenharia como, por exemplo, a ciência dos materiais, engenharia biológica, gerenciamento de obras, engenharia de transportes e no domínio da engenharia de terremotos que compreendem as seguintes áreas: análise de risco sísmico, identificação de atividade sísmica e detecção de dano, controle estrutural para atenuação das ações sísmicas e análise da fragilidade estrutural frente a uma atividade sísmica (XIE et al., 2020). Segue uma explicação mais detalhada de cada uma das áreas que compõem o domínio da engenharia de terremotos:

- Análise de risco sísmico (*Seismic hazard analysis*): consiste nos estudos voltados para a previsão dos níveis dos tremores terrestres e suas incertezas para um determinado local. Além do mais, os métodos de ML também são utilizados para avaliar o potencial de liquefação do solo;
- Identificação de atividade sísmica e detecção de dano (*Seismic identification and damage detection*): a identificação compreende a coleta de diversos estudos que utilizaram ML para simular o sistema estrutural para realizar previsões da sua resposta sísmica determinística. A detecção de dano é, de modo geral, definida como o uso de ML para reconhecer, classificar e avaliar o dano sísmico em uma estrutura;
- Controle estrutural para atenuação das ações sísmicas (*Structural control for earthquake mitigation*): consiste em utilizar controles ativos ou semi-ativos, equipados com ML, nas estruturas para atenuar os efeitos adversos de um terremoto;
- Análise da fragilidade estrutural (*Seismic fragility assessment*): definido pelo processo de incorporação de inúmeras fontes de incertezas onde os métodos de ML são utilizados para desenvolver modelos probabilísticos de demanda sísmica (PSDM) e para parametrizar funções de fragilidade.

O último item exposto acima, análise da fragilidade estrutural, é o foco desse trabalho. Dentro do âmbito de estudo das diversas ferramentas da inteligência artificial, os métodos de *Machine Learning* podem ser divididos em dois tipos: aprendizagem supervisionada (em inglês *supervised learning*) e aprendizagem não supervisionada (em inglês *unsupervised learning*). De modo geral, a aprendizagem supervisionada usa conhecimento prévio dos dados utilizados como

entrada (*input*) para a aplicação de ML com o objetivo de aprender uma função que melhor descreva ou se aproxime da relação entre o *input* e os dados de saída (*output*), os quais também já são conhecidos. No caso da aprendizagem não supervisionada, o objetivo da aplicação de ML é explorar e obter possíveis relações de um banco de dados existente, o qual nenhum conhecimento prévio foi estabelecido e que não se sabe exatamente quais informações esses dados podem conter. O objeto de estudo deste trabalho é a utilização das ferramentas da aprendizagem supervisionada.

A aprendizagem supervisionada ainda pode ser dividida em duas seções de diferentes aplicações, dependendo das características do seu banco de dados (XIE et al., 2020). Caso os dados sejam discretos e as respostas/funções desejadas também sejam de caráter discreto, *output* qualitativo, utiliza-se o método de ML chamado de classificação (em inglês *classification*) (HASTIE et al., 2001). No entanto, caso os dados sejam contínuos e as respostas/funções desejadas também sejam de caráter contínuo, *output* quantitativo, utiliza-se o método de ML chamado de regressão (em inglês *regression*) (HASTIE et al., 2001). A Figura 13 demonstra um fluxograma esquemático da divisão entre os dois métodos do ML e algumas ferramentas (algoritmos) comumente utilizadas.

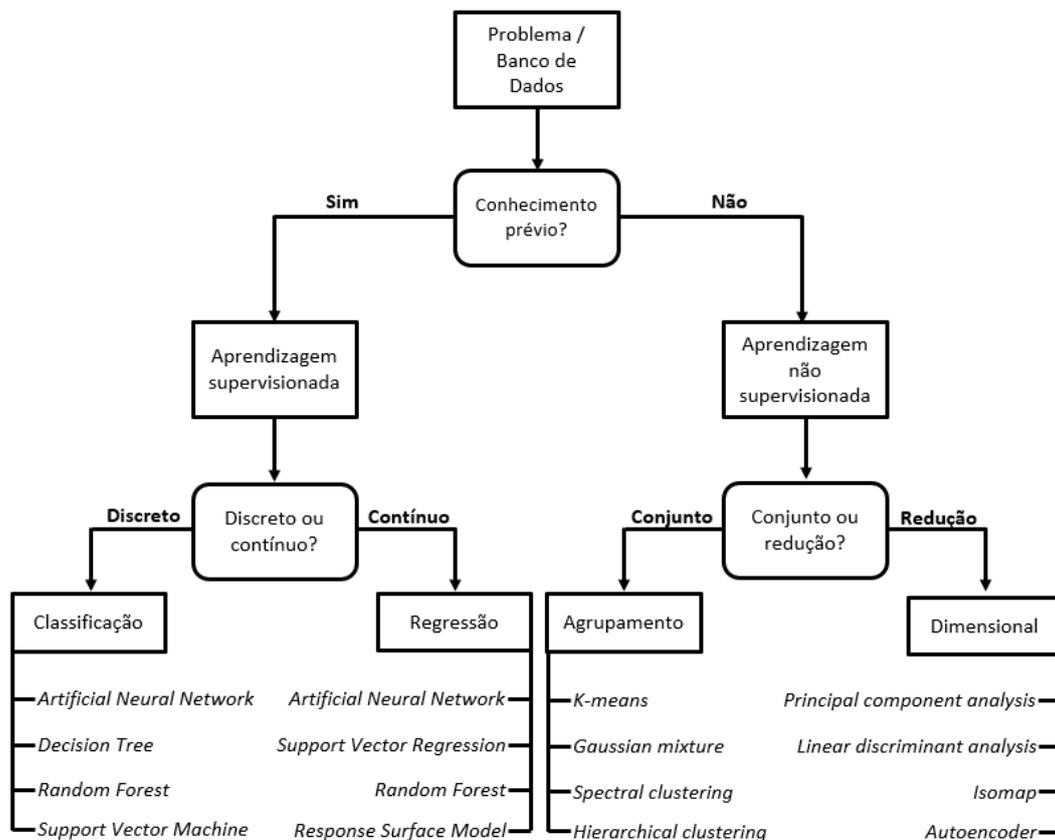


Figura 13 – Tipos de *Machine Learning* e alguns algoritmos comumente utilizados (Fonte: adaptado de Kong et al. (2019)).

Uma etapa fundamental no processo de aplicação de ML é divisão do banco de dados em um conjunto de treinamento (em inglês *training set*) e um conjunto de teste (em inglês *testing set*). Somente a partir destes dois conjuntos é possível medir a performance de um determinado algoritmo testado. De maneira geral, a divisão dos dados em 70% para o *training set* e 30% para o *testing set* é comumente utilizada. Como diz o nome, o conjunto de treinamento será utilizado para treinar um algoritmo com o objetivo de descrever uma relação entre o *input* e *output*, logo o melhor algoritmo é aquele que melhor descreve esta relação através de uma função estabelecida pelo próprio algoritmo. Não se utiliza 100% dos dados para o treinamento do algoritmo pois um dos parâmetros de sua performance é a verificação da sua precisão no processo de predição com dados não utilizados no processo de treinamento, com o objetivo de se evitar erros devido a criação de tendência. Portanto, a parcela de 30% dos dados, separados como *testing set*, servirão como confirmação da performance do algoritmo utilizado, uma vez que esses dados não foram previamente analisados pela própria ferramenta, evitando-se viés durante o processo de ML.

Consequentemente, todos os algoritmos testados durante o processo de aplicação de *machine learning* têm sua performance avaliada tanto para o *training set* quanto para o *testing set*. De maneira ideal deseja-se obter uma relação satisfatória e similar entre *input* e *output* para ambos os conjuntos, ou seja, obter um algoritmo que performe de maneira eficiente e semelhante tanto para o *training set* quanto para o *testing set*. Os algoritmos que performam de maneira muito precisa na predição dos resultados, durante o processo de treinamento, normalmente não generalizam de maneira satisfatória, ou seja, sua performance é baixa quando testada com os novos dados do *testing set*. Esse tipo de erro é chamado de *overfitting*. Já o erro denominado de *underfitting* pode ocorrer quando os algoritmos performam de maneira pouco satisfatória durante o treinamento e, por consequência, também performam de maneira insatisfatória quando avaliada com o *testing set*.

De acordo com Hastie et al. (2001), o erro *overfitting* está relacionado com a capacidade do algoritmo conseguir diferenciar entre dados relevantes e não relevantes para o processo de predição. Entende-se por dados o conjunto de valores compostos pelas variáveis aleatórias para as OAEs em estudo e as respectivas respostas de demanda no caso de uma análise de regressão ou os respectivos grupos os quais os modelos foram divididos no caso de uma análise de classificação. Portanto, uma maneira de se evitar o erro de *overfitting* é aprimorar os dados de entrada com técnicas como *feature selection*, *feature transformation* e *hyperparameter tuning*.

A técnica de *feature selection* consiste em identificar quais dados são realmente relevantes para a aplicação de ML e que, por consequência, resultarão em uma melhor performance do algoritmo testado. A técnica de *feature transformation* se resume ao processo de transformação dimensional dos dados utilizando ferramentas como *principal component analysis (PCA)*, *nonne-*

gative matrix factorization e *factor analysis*. Já a técnica de *hyperparameter tuning* consiste no processo de identificação do conjunto de parâmetros que resultam no melhor modelo preditivo, ou seja, como o algoritmo de ML ajusta o modelo para os respectivos dados. Dentre as técnicas mencionadas, o *feature selection* é a mais importante uma vez que reduzindo o número de dados utilizados como *input* o processamento computacional fica mais eficiente e seus resultados mais representativos. Seguem alguns métodos comumente utilizados na realização do *feature selection*:

- *Stepwise regression*: consiste no processo de adicionar ou remover dados até que se alcance a melhor performance;
- *Sequential feature selection*: consiste no processo de adicionar ou remover, de maneira iterativa, as variáveis e analisar o efeito que cada mudança teve na performance do algoritmo;
- *Regularization*: consiste na utilização de ferramentas de redução para remover variáveis redundantes através da redução de seus coeficientes para zero;
- *Neighborhood Component Analysis (NCA)*: definido pelo processo de investigação da relevância das variáveis para que as variáveis com menor relevância possam ser descartadas.

2.5.1 Supervised learning

Para aplicar a aprendizagem supervisionada, é necessário ter conhecimento prévio sobre as características do seu banco de dados, ou seja, o que se pretende alcançar com a aplicação de *machine learning*. Portanto, deve-se primeiro conhecer as características do seu banco de dados e quais informações deseja-se obter. Uma das formas de saber qual a melhor técnica da aprendizagem supervisionada para ser utilizada no banco de dados é responder as seguintes perguntas: “O banco de dados pode ser dividido em grupo ou ser categorizado?” ou “O banco de dados contém uma faixa variável de informações?”. Caso o banco de dados possa ser separado em grupos ou classes específicas, deve-se utilizar algoritmos de classificação para se obter uma função que melhor descreva a relação entre *input* e *output*. No entanto, caso a natureza do seu banco de dados, em especial o *output*, seja composto por números reais, como por exemplo temperatura, tempo, valores de demanda estrutural, deve-se utilizar as técnicas de regressão para melhor estabelecer a relação entre dados de entrada e saída.

A escolha de qual algoritmo apresenta a melhor performance, para técnicas de classificação ou regressão, é através do processo de tentativa e erro. Ademais, esta escolha também deve levar em considerações alguns aspectos, como: velocidade de treinamento, uso da memória computacional, precisão do processo preditivo (em inglês *Prediction Accuracy*), transparência

e o nível de interpretação das escolhas feitas pelo algoritmo, ou seja, o quão fácil se consegue entender as razões pelas quais o algoritmo realizou determinada predição.

Os parâmetros utilizados para avaliar a precisão do algoritmo durante o processo de predição é diferente para os métodos de classificação e regressão. Esses parâmetros são discutidos mais a frente nos itens 2.5.1.1 (Regressão) e 2.5.1.2 (Classificação).

Com base no que foi apresentado até este momento, pode-se estabelecer as principais etapas para a aplicação dos métodos da aprendizagem supervisionada com o objetivo de encontrar o melhor algoritmo de predição.

- Preparação do banco de dados e divisão entre *training set* e *testing set*;
- Escolher o algoritmo que apresente maior eficiência no processamento e no uso da memória computacional, na predição dos dados, tanto para o *training set* quanto para o *testing set* e maior transparência dos mecanismos utilizados durante o processo;
- Escolher o modelo com melhor performance dentro dos algoritmos escolhidos para o estudo. Lembrando que erros como *overfitting* e *underfitting* não devem ocorrer;
- Escolher um tipo de método para a avaliação da performance dos resultados gerados pelo modelo;
- Examinar a performance e repetir os passos anteriores até atingir o desempenho desejado;
- Utilizar o modelo escolhido para realizar predições com novos dados.

A exemplo do que a Figura 13 ilustrou, os algoritmos, ou técnicas, mais utilizadas no estudo da aprendizagem supervisionada são: *Artificial Neural Network* (ANN), *Decision Tree* (DT), *Random Forest* (RF), *Support Vector Machine* (SVM), *Support Vector Regression* (SVR) e *Response Surface Model* (RSM). No entanto, estes são apenas alguns dos inúmeros algoritmos existentes na execução do *Machine Learning* como o *Quadratic Discriminant Analysis* (QDA), *K-nearest Neighbor* (KNN) e *Naive Bayes*. Como o foco deste trabalho é aplicar as ferramentas de ML para estudos de vulnerabilidade sísmica, não será realizado uma análise detalhada dos métodos estatísticos utilizados. Para tal, o leitor deve se direcionar a materiais como Hastie et al. (2001). No entanto, segue uma breve descrição dos métodos mais utilizados e suas respectivas características.

De acordo com Xie et al. (2020), o *Artificial Neural Network* (ANN) consiste tipicamente na divisão entre, no mínimo, três camadas: *input layer*, *hidden layer* e *output layer*. De maneira geral, as variáveis do modelo, ou problema, são ponderadas e alimentadas na camada oculta

(*hidden layer*) que consiste em uma série de relações não-lineares como, por exemplo, funções sigmoidais, que posteriormente também são ponderadas e alimentadas a camada de saída (*output*) para estabelecer modelos de regressão ou classificação (XIE et al., 2020). As ponderações são aplicadas devido as consequências da propagação e atualização do modelo preditivo durante o processo de treinamento, com o intuito de minimizar os erros que normalmente se propagam na direção contrária das ponderações (HASTIE et al., 2001). Encontrar a melhor estrutura do ANN pode ser muito desafiador, pois esta técnica tem grande potencial de gerar erros de *overfitting*. O ANN pode ainda ser ampliado para o *Deep Neural Network (DNN)* que incorpora múltiplas camadas ocultas (LECUN et al., 2015).

O *Support Vector Machine (SVM)* é um algoritmo binário de classificação que utiliza funções de *kernel* para possibilitar uma análise dos dados no espaço de características multi-dimensionais. De acordo com Xie et al. (2020) e considerando um banco de dados que possa ser separado, o SVM estabelece um hiperplano capaz de maximizar a margem entre o hiperplano e os vetores de suporte (em inglês *supporting vectors*), o qual contém os dados que se encontram próximos ao hiperplano. A Figura 14 ilustra a aplicação do SVM para classificação. As variáveis ξ_i e ξ_j representam os dados que foram classificados de maneira incorreta. O SVM também pode ser utilizado como método de regressão, chamado de *Support Vector Regression*. O SVR possui as mesmas características do SVM, porém incorpora uma função de erro que define uma região ε em torno das respostas corretas do *training set* (VAPNIK, 1998). A Figura 15 demonstra a aplicação do SVR para regressão. Pode-se observar que os pontos internos da região ε possuem erro zero.

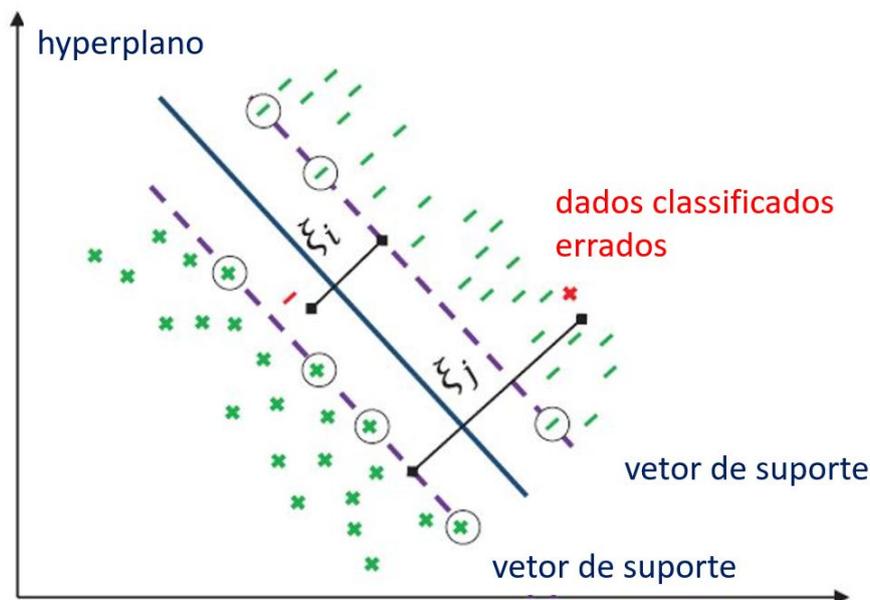


Figura 14 – Exemplo de *SVM* para classificação (Fonte: adaptado de Xie et al. (2020)).

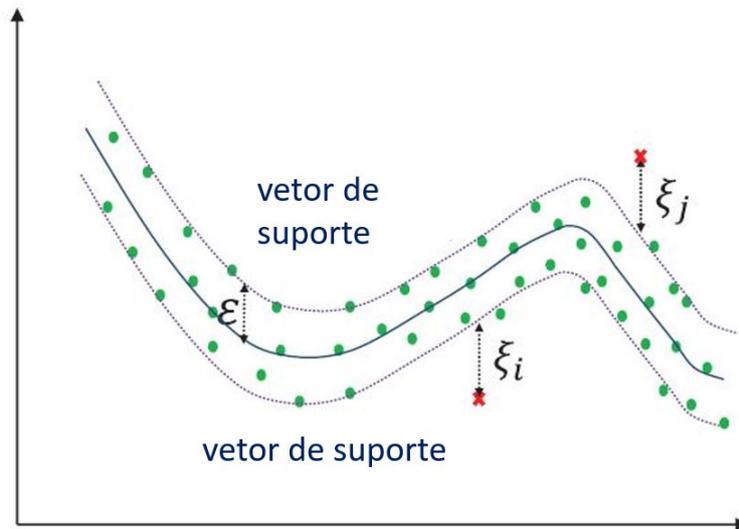


Figura 15 – Exemplo de *SVR* para regressão (Fonte: adaptado de Xie et al. (2020)).

O *Response Surface Model* (RSM) é uma técnica que vem sendo bastante utilizada em inúmeras áreas por conta da sua simplicidade, transparência e a transmissibilidade das variáveis (XIE et al., 2020), mesmo tendo sido desenvolvida como um método estatístico para a exploração de dados de um sistema e suas respectivas respostas (BOX; HUNTER, 1957). O *Response Surface Model* estima as respostas usando uma série de funções básicas. Enquanto polinômios de terceira ordem ou superiores possam ser utilizados, polinômios do primeiro e segundo grau são preferíveis para a aplicação do RSM. Assume-se que o erro no processo de predição utilizando o RSM segue uma distribuição normal com média zero e um valor de variância σ^2 . A técnica do RSM pode ser melhorada a partir da utilização de alguns métodos como o *Stepwise regression* e o *Least absolute shrinkage and selection operator (LASSO)*.

De acordo com Xie et al. (2020), o *Decision Tree* (DT) é um algoritmo não parametrizado que separa de maneira repetitiva os dados de entrada em regiões distintas que não se sobrepõem. Estas regiões podem ser agrupadas em três tipos de nós: o nó raiz (*root node*) que contém todas as informações do *input*; nós interiores (*interior nodes*); nós terminais (*terminal nodes*). O algoritmo define um modelo para cada região resultante do *input*. Um simples modelo de regressão pode ser aplicado em cada região para estudos de regressão ou uma classe pode ser atribuída em cada região para estudos de classificação. Comumente, uma função de custo, ou peso, é utilizada durante o processo com o intuito de otimizar o método de separação. A Figura 16 demonstra o processo descrito por um algoritmo de DT.

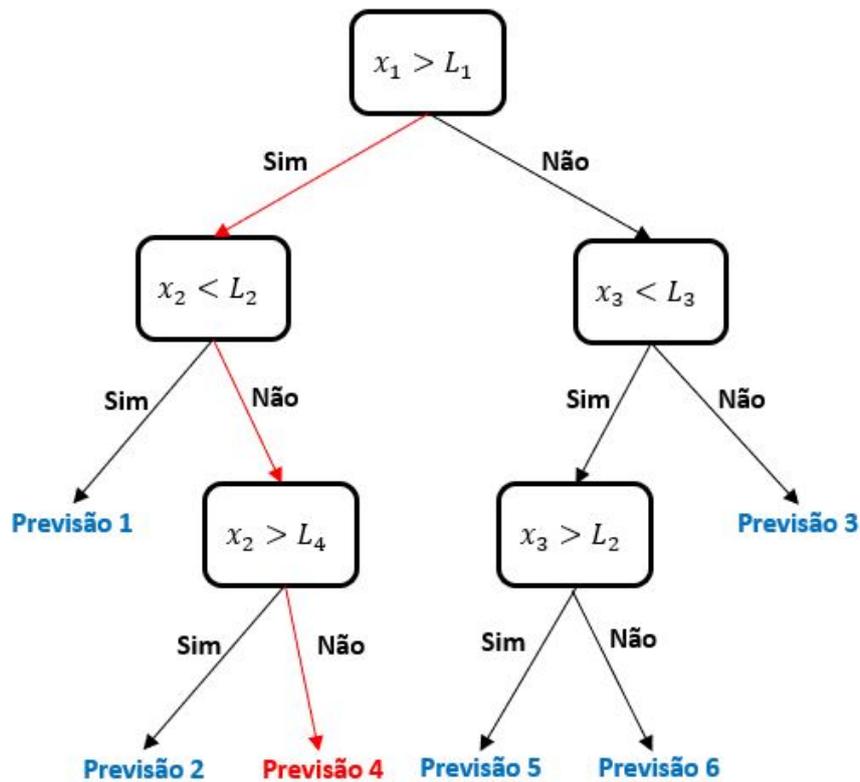


Figura 16 – Exemplo da estrutura do algoritmo do método *Decision Tree* (Fonte: autor).

Para evitar que potencias erros de *overfitting* e de instabilidade causados por modelos que contém somente um *Decision Tree*, o algoritmo do *Random Forest* (RF) é constituído por um conjunto de DTs. O RF é uma técnica de *machine learning* bastante popular e muito utilizada na área da estatística (MANGALATHU; JEON, 2019). Este método apresenta inúmeras vantagens quando comparada com outras técnicas do ML por conta da sua flexibilidade, simplicidade intuitiva e eficiência computacional (BREIMAN, 2001). Diferente de outros métodos, o RF não implica nenhuma suposição sobre as funções de distribuição dos dados de entrada e, por isso, é considerado também um método não parametrizado. De acordo com Hastie et al. (2001), por ser um método baseado em um conjunto de *Decision Trees*, o RF utiliza duas técnicas importantes do ML, conhecidas como *bagging* e *random feature selection*. A técnica de *bagging* consiste no fato que cada DT é construída individualmente com reposições randômicas (*bootstrap*) dos dados do *training set* e a média dos valores do *output* de cada DT é utilizada como previsão. A técnica de *random feature selection* faz com que o RF seja um algoritmo robusto contra o erro de *overfitting* e tenha uma boa performance quando comparada com outras técnicas como SVM e ANN. A Figura 17 demonstra o processo descrito por um algoritmo de RF.

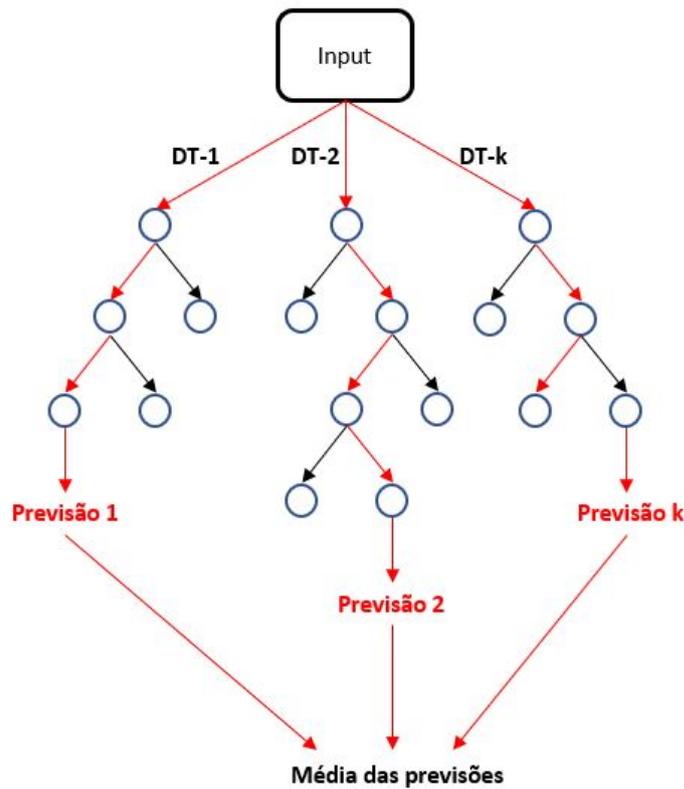


Figura 17 – Exemplo da estrutura do algoritmo do método *Random Forest* (Fonte: autor).

2.5.1.1 Regressão

De acordo com Hastie et al. (2001), o modelo de regressão é voltado para análises em que as informações do banco de dados são informações contínuas e as funções/respostas desejadas também são de caráter contínuo. Portanto, o *output* é definido como quantitativo e seus dados representam uma faixa variável de informações. Além do *prediction accuracy* outros parâmetros comumente utilizados para avaliar a precisão do algoritmo durante o processo de medição são a raiz quadrada do erro médio (*Root Mean Squared Error - RMSE*), o erro quadrático médio (*Mean Squared Error - MSE*), o erro médio absoluto (*Mean Absolute Error - MAE*) e através do coeficiente de determinação (R^2).

O RMSE representa a raiz quadrada do segundo momento das diferenças entre os valores calculados pelo modelo e os valores observados, conforme a Equação 2.5.1 a seguir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (2.5.1)$$

O erro quadrático médio (MSE) é equivalente ao quadrado do RMSE, ou seja, ele

representa o segundo momento das diferenças entre os valores calculados e os valores observados, como demonstra a Equação 2.5.2 a seguir.

$$MSE = \frac{1}{n} \sum_{j=1}^n \left(y_j - \hat{y}_j \right)^2 \quad (2.5.2)$$

O erro médio absoluto (MAE) é equivalente ao módulo do primeiro momento das diferenças entre os valores calculados e os valores observados, como demonstra a Equação 2.5.3 a seguir.

$$MAE = \frac{1}{n} \sum_{j=1}^n \left| y_j - \hat{y}_j \right| \quad (2.5.3)$$

Por sua vez, o coeficiente de determinação (R^2) é a medida de ajuste de um modelo probabilístico, variando entre 0 e 1, entre os valores calculados e os observados. Em outras palavras, o valor de R^2 representa quanto um modelo linear explica a variância da amostra. Desta forma, a Equação 2.5.4 demonstra o cálculo do coeficiente de determinação.

$$R^2 = 1 - \left(\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y}_j)^2} \right) \quad (2.5.4)$$

Sendo para as fórmulas acima: n o número total de amostras; y_j o valor observado; \hat{y}_j o valor calculado pelo algoritmo durante o processo de predição; \bar{y}_j o valor médio das observações.

A Tabela 8 a seguir demonstra as características dos parâmetros utilizados na avaliação da precisão do algoritmo e os respectivos valores idealizados.

Tabela 8 – Parâmetros de avaliação (Fonte: adaptado de Hastie et al. (2001)).

Parâmetro	Característica	Valor ideal
RMSE	Positivo; Unidade correspondente ao <i>output</i>	Menor valor possível
MSE	Positivo; Quadrado do RMSE	Menor valor possível
MAE	Positivo; Similar ao RMSE; Menos sensível a <i>outliers</i>	Menor valor possível
R^2	Sempre menor do que 1	Valor próximo de 1

De acordo com Hastie et al. (2001), em casos que o banco de dados é robusto o bastante, a avaliação da precisão do processo preditivo para o *training set* deve ser realizada através do conjunto de validação (em inglês *validation set*). De acordo com os estudos de aplicação das ferramentas de *machine learning* para a obtenção de algoritmos preditivos voltados a análise de fragilidade de pontes, como Mangalathu et al. (2019), Xie et al. (2020) e Ferrario et al. (2017),

um banco de dados é considerado robusto quando possui acima de 5 mil dados. O conjunto de validação é composto pelo próprio *training set* e representa como o algoritmo vai verificar e validar a sua própria performance. Portanto, a melhor precisão obtida por um algoritmo representa a melhor precisão do *validation set*. A seguir estão descritas as principais técnicas utilizadas para o *validation set*.

- *K-fold Cross Validation*: os dados do *training set* são divididos randomicamente em K grupos distintos para validação. Para cada grupo (K_i), o algoritmo treina um modelo preditivo usando os dados contidos neste grupo e verifica a sua performance usando os dados restantes do *validation set*. Após realizar esse procedimento para cada (K_i), o algoritmo calcula a média do erro entre todos os grupos;
- *Hold-out Validation*: uma porcentagem do *training set* é separada para ser utilizada como *validation set*. Após a separação, o algoritmo treina o modelo preditivo usando os dados do *training set* e verifica sua performance utilizando os dados do *validation set* que foram separados;
- *Resubstitution Validation*: neste caso, o algoritmo utiliza todo o *training set* para treinar um modelo preditivo e calcula o erro usando os mesmos dados, ou seja, nenhuma separação de *validation set* é utilizada.

Vale ressaltar que as técnicas de validação somente afetam a maneira em que o algoritmo calcula os parâmetros de performance, ou seja, o modelo preditivo final é sempre treinado utilizando todos os dados do *training set*. A técnica do *K-fold Cross Validation* proporciona uma excelente estimativa da precisão preditiva do modelo final treinado com todos os dados do conjunto de treinamento. A técnica do *Hold-out Validation* utiliza apenas uma porção dos dados do *training set* e, portanto, deve ser utilizada apenas para um grande conjunto de dados. No entanto, a técnica do *Resubstitution Validation* não oferece nenhuma proteção contra erros de *overfitting*, uma vez que sem a separação dos dados em um grupo de validação o algoritmo pode calcular performances utópicas para o conjunto de treinamento. Em outras palavras, é provável que o algoritmo performe de maneira muito satisfatória para o *training set* e quando testado com os dados do *testing set* sua performance seja muito inferior.

Apesar dos parâmetros demonstrados na Tabela 8 indicarem com precisão a performance do algoritmo, métodos como o gráfico de resposta (*response plot*), o gráfico de valores observados vs. valores previstos (*predicted vs. actual response plot*) e o gráfico de residuais (*residuals plot*) podem ser implementados para verificar o resultado do algoritmo testado.

Utilizando o gráfico de resposta é possível analisar os resultados do modelo de regressão estudado. Após o treinamento do modelo, o gráfico de resposta pode ser implementado para mostrar graficamente os valores previstos e os valores observados *vs.* o número de observações ou uma variável aleatória. A Figura 18 demonstra um exemplo do gráfico de resposta.

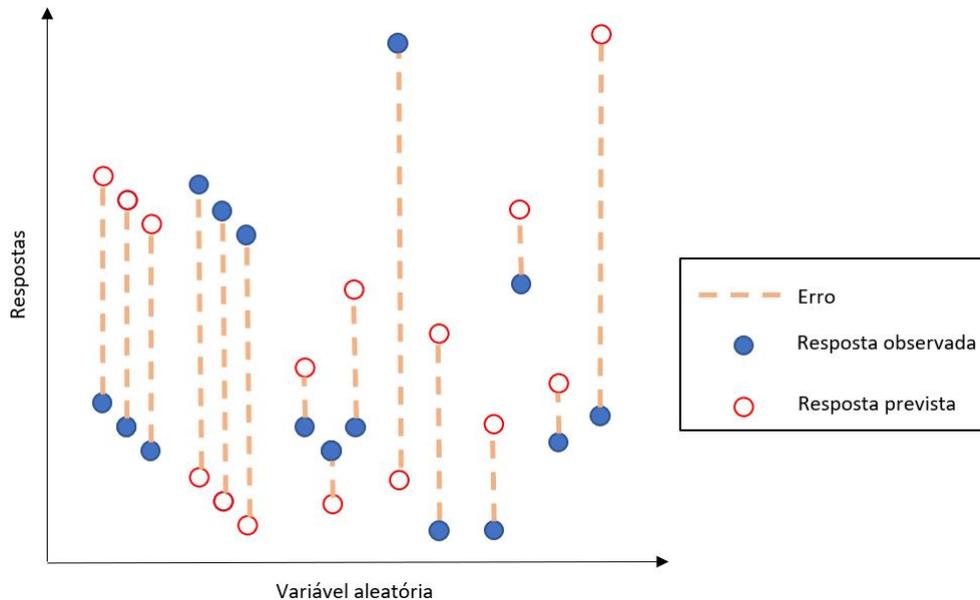


Figura 18 – Exemplo do gráfico *Response Plot* (Fonte: autor).

Quando uma variável aleatória possui uma faixa de poucos valores, pode-se utilizar *box plot* ao invés de marcadores para verificar a performance do algoritmo. O *box plot* ilustra os valores típicos das respostas e possíveis resíduos (*outliers*). A marca central indica a mediana e a marca inferior e superior indicam 25% e 75%, respectivamente. Linhas verticais, ou horizontais a depender da orientação gráfica, se estendem aos dados mais extremos do banco de dado que não são considerados *outliers*. Os resíduos por sua vez são ilustrados com marcas de “+”. A Figura 19 demonstra um exemplo do *Response Plot* utilizando *box plot*.

O gráfico de valores observados *vs.* valores previstos demonstra visualmente a performance do algoritmo, comparando os valores observados com os valores previstos. Em outras palavras, este gráfico pode ser utilizado para verificar o quão bem o modelo de regressão faz previsões para diferentes valores de resposta. Um modelo perfeito fará previsões exatas, portanto todos os pontos ilustrados no *predicted vs. actual response plot* pertencerão a linha diagonal que dividi o gráfico. A distância vertical entre a linha e qualquer ponto do gráfico representa o erro para determinado ponto. A Figura 20 demonstra um exemplo deste gráfico. De forma geral, um bom modelo de regressão tem, comumente, os pontos espalhados de maneira simétrica em torno da linha diagonal.

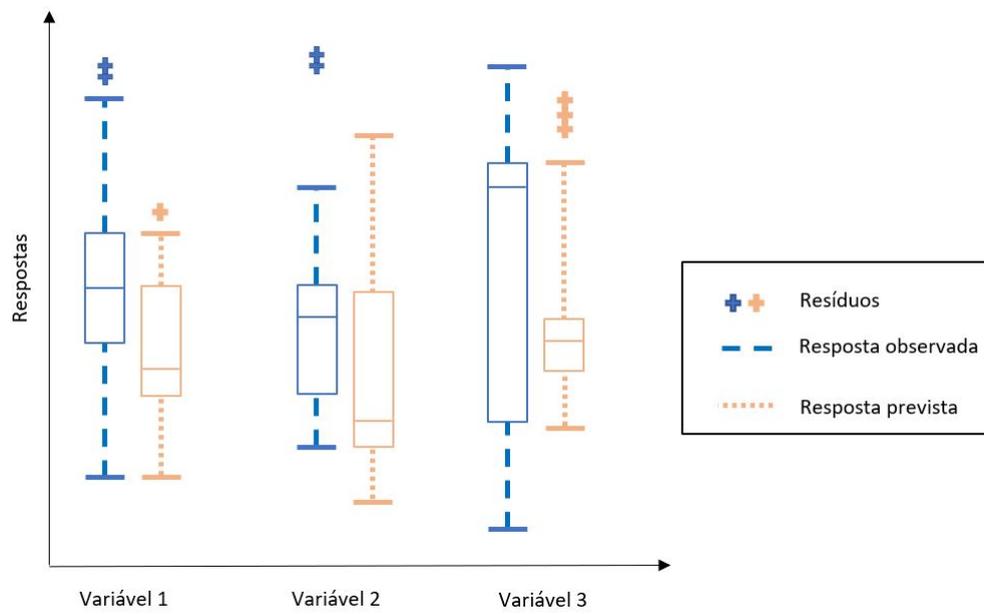


Figura 19 – Exemplo do gráfico *Response Plot* utilizando *box plot* (Fonte: autor).

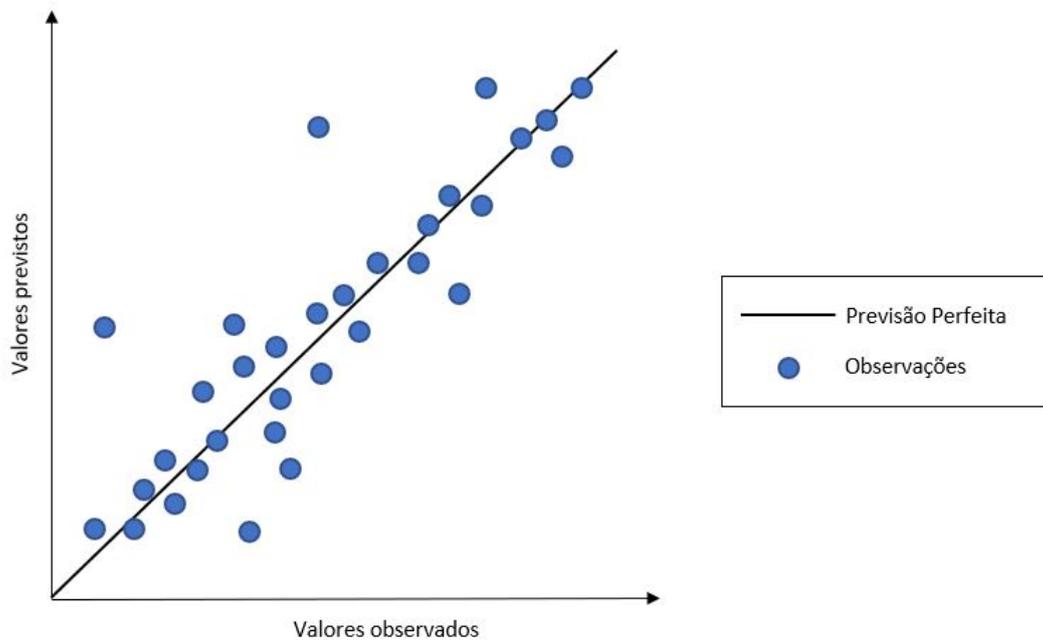


Figura 20 – Exemplo do gráfico *Predict vs. Actual Response* (Fonte: autor).

O gráfico de resíduos é uma excelente ferramenta para verificar as diferenças entre os valores previstos e os valores observados. O *residual plot* pode ser utilizado para verificar os resíduos das observações, valores previstos, número de observações ou uma variável aleatória. Uma característica que este gráfico deve conter para ser considerado adequado é que os dados plotados não apresentem nenhum tipo de padrão, ou seja, os dados devem estar distribuídos randomicamente. A Figura 21 demonstra um exemplo do *residual plot*.

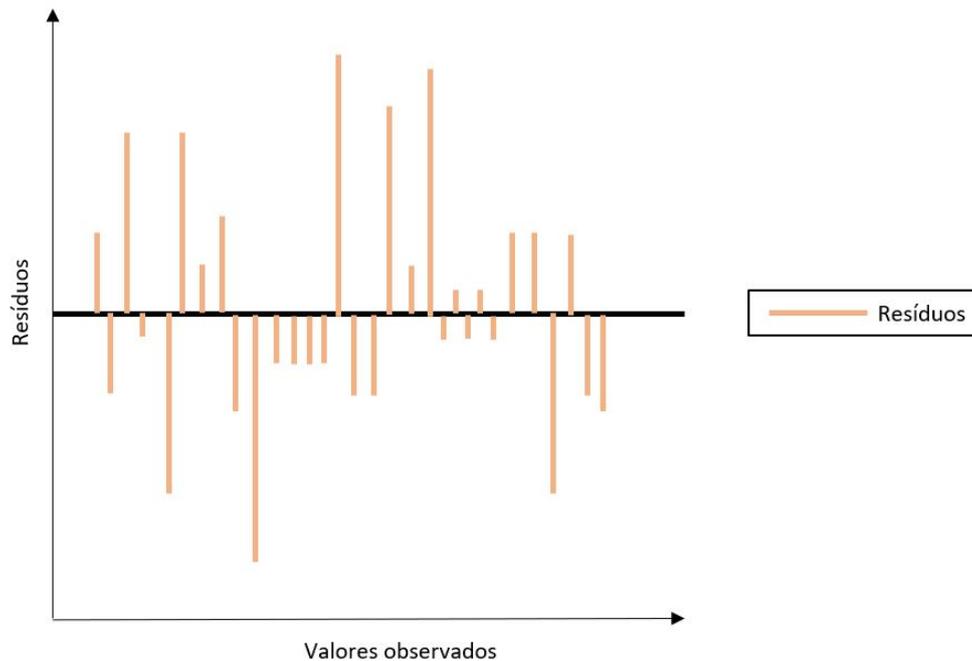


Figura 21 – Exemplo do gráfico de resíduos (Fonte: autor).

2.5.1.2 Classificação

O modelo de classificação é destinado para análises em que os dados e as respostas desejadas sejam de caráter discretos, ou seja, o *output* é considerado qualitativo e, portanto, este banco de dados pode ser dividido em grupos ou ser categorizado (HASTIE et al., 2001). Como observado no item 2.5.1.1, um parâmetro de avaliação da performance de um algoritmo de classificação é o *prediction accuracy*. No entanto, apenas avaliar o algoritmo escolhido através da porcentagem resultante da precisão do processo preditivo pode, às vezes, levar a interpretações erradas. Desta forma, utiliza-se de ferramentas gráficas para complementar o entendimento sobre a performance do algoritmo estudado.

Conforme apontado no item 2.5.1.1, a avaliação do processo preditivo para o *training set* deve ser realizada utilizando-se o *validation set* (HASTIE et al., 2001). De maneira similar aos modelos de regressão, as técnicas mais utilizadas para a construção do *validation set* para

o modelo de classificação são: *K-fold Cross Validation*, *Hold-out Validation* e *Resubstitution Validation*. Todas as características descritas anteriormente para os modelos de regressão, quanto ao *validation set*, são pertinentes para os modelos de classificação.

O valor em porcentagem da precisão de previsão, para um modelo de classificação, é comumente retirada da matriz de confusão (em inglês *confusion matrix*) elaborada a partir dos resultados de um modelo. De maneira geral, a *confusion matrix* descreve quanto o modelo acertou ou errou ao fazer as suas previsões. Em resumo, a matriz de confusão ilustra graficamente quantas vezes o algoritmo realizou a previsão de maneira correta comparada com as classes estabelecidas no modelo em estudo, e a partir desse resultado calcula parâmetros estatísticos e o *prediction accuracy*. Considerando um problema simples de classificação com apenas duas classes, onde uma das classes é considerada positiva (por exemplo a não manifestação de uma doença) e a outra negativa (por exemplo a manifestação de doença), conforme demonstrado na Figura 22, podemos obter os parâmetros estatísticos (HASTIE et al., 2001).

Classes Reais	Condição negativa	Falso Positivo	<u>Verdadeiro Negativo</u>
	Condição positiva	<u>Verdadeiro Positivo</u>	Falso Negativo
		Condição positiva prevista	Condição negativa prevista
		Classes Previstas	

Figura 22 – Exemplo da matriz de confusão para apenas 2 classes (Fonte: autor).

Os valores calculados pelo modelo de classificação podem ser divididos em 4 grupos. O primeiro grupo do **Falso Positivo (FP)** demonstra quantas vezes o algoritmo previu uma condição positiva quando na verdade, para as respectivas observações, a condição era negativa. Já o segundo grupo do **Verdadeiro Negativo (VN)** demonstra quantas vezes o algoritmo previu de maneira correta a condição negativa. O grupo do **Falso Negativo (FN)** ilustra quantas vezes o algoritmo errou ao prever, como uma condição negativa, uma condição real

positiva. Por fim, o grupo do **Verdadeiro Positivo (VP)** demonstra quantas vezes o algoritmo acertou em prever uma condição positiva. A depender das classes estipuladas, o algoritmo pode ser penalizado quanto cometer erros de previsão do tipo FP ou FN. Por exemplo, caso fosse um estudo de doenças, o erro de previsão FN é considerado o erro mais crítico, pois classifica uma observação como não contendo a doença quando na verdade ela contém, ou seja, o algoritmo irá classificar de maneira equivocada dados que deveriam ser positivos.

A partir da matriz de confusão o *prediction accuracy* pode ser calculado a partir da Equação 2.5.5 a seguir:

$$Accuracy = \frac{\sum TP + \sum TN}{\sum População} \quad (2.5.5)$$

De acordo com Hastie et al. (2001), a interpretação da performance do modelo preditivo em estudo somente a partir do *prediction accuracy* pode levar a falsas conclusões e, portanto, outros parâmetros estatísticos devem ser utilizados. A Tabela 9 a seguir demonstra os parâmetros estatísticos mais utilizados que podem ser calculados através da *confusion matrix*.

Além dos parâmetros expostos na Tabela 9, pode-se também calcular o valor F_1 (em inglês $F_1, score$) que representa a média harmônica da precisão e da sensibilidade. A Equação 2.5.6 a seguir demonstra como calcular F_1 a partir dos valores da matriz de confusão.

$$F_1 = \frac{2 \cdot PPV \cdot TPR}{PPV + TPR} \quad (2.5.6)$$

Para problemas de classificação é comum a utilização de outros métodos gráficos, além do *confusion matrix*, que ajudam a avaliar o desempenho do algoritmo testado. O gráfico de dispersão (em inglês *scatter plot*) possibilita uma análise visual dos dados utilizados no treinamento e os resultados previstos pelo modelo. Este gráfico é construído através da relação de duas variáveis aleatórias e as classes definidas para o problema em questão. A Figura 23 ilustra um exemplo do *scatter plot*.

A partir deste gráfico pode-se visualizar que as variáveis aleatórias X e Y separam de maneira satisfatória a classe 1 das demais classes, 2 e 3. Portanto, o gráfico de dispersão é uma excelente ferramenta para investigar quais variáveis aleatórias são úteis para o modelo e quais podem ser retiradas através da técnica de *feature selection*.

Da mesma forma que o gráfico de dispersão, o gráfico de coordenadas paralelas (em inglês *parallel coordinates plot*) também pode ser uma ferramenta útil na investigação de quais variáveis aleatórias ajudam o modelo a separar as classes e quais dificultam.

Tabela 9 – Parâmetros estatísticos calculados a partir da *confusion matrix* (Fonte: autor).

Parâmetros Estatísticos	Equação
Taxa de Falso Positivo (<i>False Positive Rate</i>) (FPR)	$\frac{\sum FP}{\sum \text{CondiçãoNegativa}}$
Taxa de Verdadeiro Negativo - Especificidade (<i>True Negative Rate - Specificity</i>) (TNR - SPC)	$\frac{\sum TN}{\sum \text{CondiçãoNegativa}}$
Taxa de Verdadeiro Positivo - Sensibilidade (<i>True Positive Rate - Sensitivity</i>) (TPR)	$\frac{\sum TP}{\sum \text{CondiçãoPositiva}}$
Taxa de Falso Negativo (<i>False Negative Rate</i>) (FNR)	$\frac{\sum FN}{\sum \text{CondiçãoPositiva}}$
Taxa de Descoberta Falsa (<i>False Discovery Rate</i>) (FDR)	$\frac{\sum FP}{\sum \text{CondiçãoPositivaPrevista}}$
Valor de Previsão Positivo - Precisão (<i>Positive Prediction Value - Precision</i>) (PPV)	$\frac{\sum TP}{\sum \text{CondiçãoPositivaPrevista}}$
Valor de Previsão Negativa (<i>Negative Prediction Value</i>) (NPV)	$\frac{\sum TN}{\sum \text{CondiçãoNegativaPrevista}}$
Taxa de Falsa Omissão (<i>False Omission Rate</i>) (FOR)	$\frac{\sum FN}{\sum \text{CondiçãoNegativaPrevista}}$
Taxa de Probabilidade Positiva (<i>Positive Likelihood Ratio</i>) (LR+)	$\frac{TPR}{FPR}$
Taxa de Probabilidade Negativa (<i>Negative Likelihood Ratio</i>) (LR-)	$\frac{FNR}{TNR}$
Prevalência (<i>Prevalance</i>)	$\frac{\sum \text{CondiçãoPositiva}}{\sum \text{População}}$

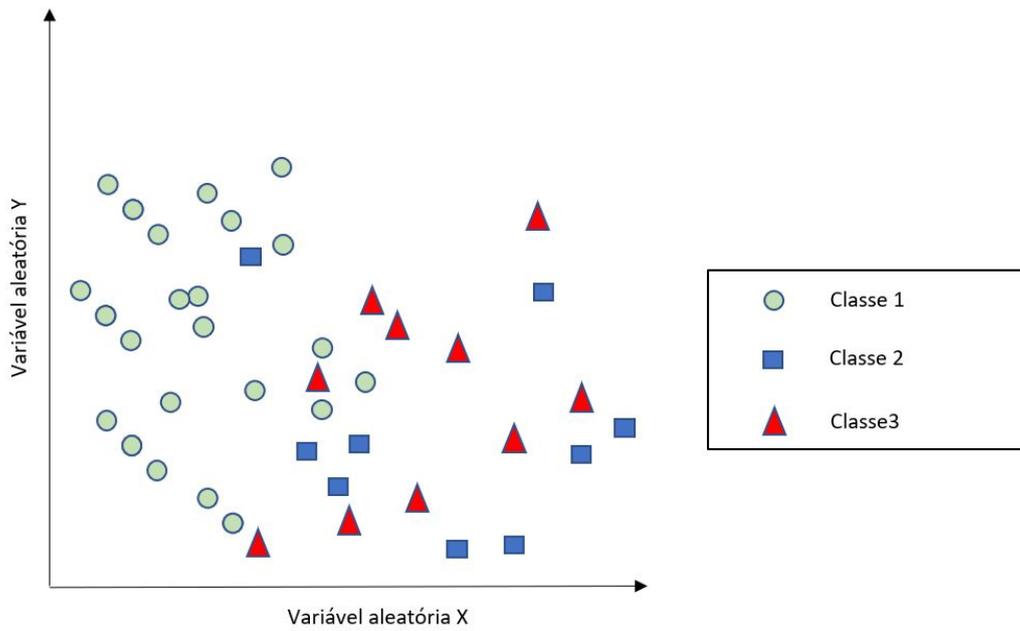


Figura 23 – Exemplo de *Scatter Plot* para apenas 3 classes (Fonte: autor).

A vantagem deste gráfico quando comparado ao *scatter plot* é a possibilidade de analisar de maneira visual em um plano 2-D dados de maiores dimensões. A Figura 24 demonstra um exemplo do *parallel coordinates plot*.

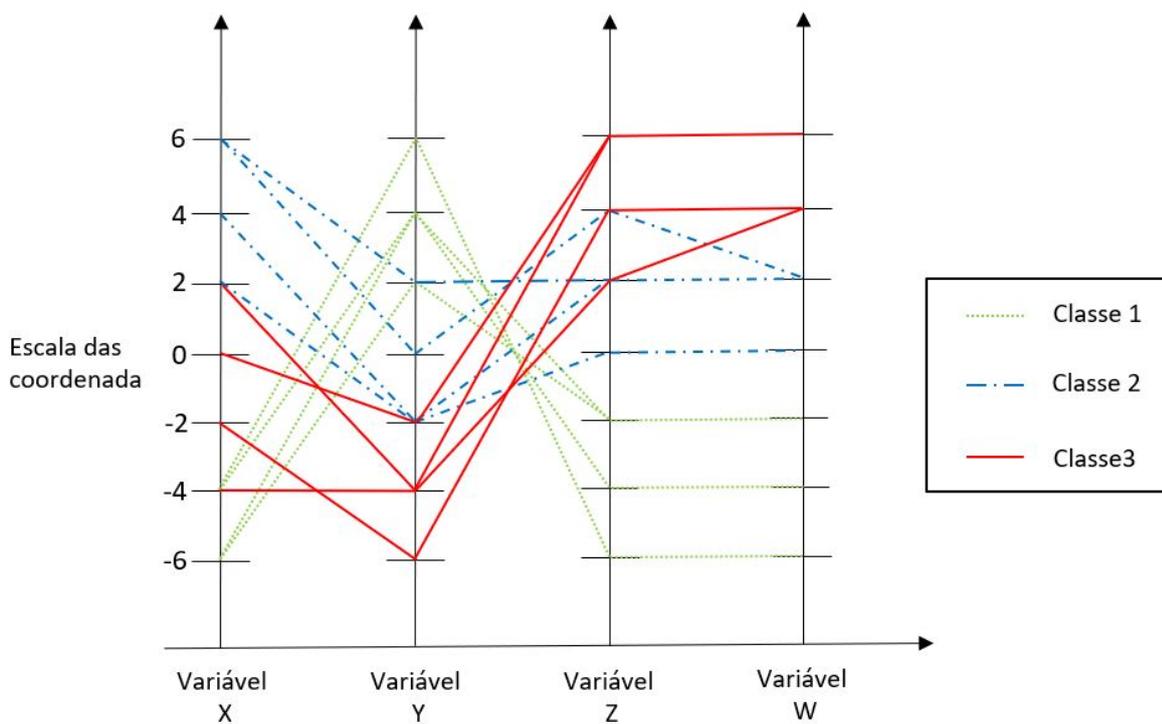


Figura 24 – Exemplo de *Parallel Coordinates Plot* para apenas 3 classes (Fonte: autor).

Conforme pode ser observado no gráfico acima, as variáveis aleatórias Z e W performam melhor que a X e Y para a separação de classes. Desta forma, as variáveis X e Y poderiam ser descartadas através do *feature selection* no avanço do estudo. O eixo das ordenadas pode ser escalonado de três formas diferentes, conforme descritas a seguir.

- *None*: os dados do banco de dados são utilizados sem escala com os mesmos valores de máximo e mínimo;
- *Range*: os dados são escalados com os valores máximos e mínimos independentes;
- *Z-Score*: os dados são escalonados de acordo com o *valor-z*. Este parâmetro, também conhecido como valor padrão, é o número de desvios padrões pelos quais o valor de um dado bruto (observação retirada do banco de dados) está abaixo ou acima do valor médio do que está sendo observado ou avaliado. Dados brutos acima do valor médio possuem valores padrões positivos, enquanto os que estão abaixo do valor médio possuem valores padrões negativos.

Por fim, uma última ferramenta gráfica comumente utilizada na avaliação da performance de um algoritmo de classificação é a curva das características operacionais do receptor (em inglês *receiver operating characteristics curve - ROC*). De acordo com Fawcett (2006), a curva ROC é muito utilizada para análises bi-dimensionais (problemas com duas classes ou categorias), no entanto, é possível elaborar curvas ROC com mais de duas classes (problemas conhecidos como multi-classes). A curva bi-dimensional será o foco deste trabalho e, portanto, para uma análise multi-dimensional o leitor deve se direcionar a trabalhos como Landgrebe e Duin (2006), Landgrebe e Duin (2007) e Fieldsend e Everson (2005).

De maneira simplificada, a curva ROC irá plotar a sensibilidade *vs.* a taxa de falso positivo, ou seja, o TPR *vs.* FPR para duas classes para diferentes valores limites (em inglês *thresholds*). O processo de variação do *threshold* é necessário para a obtenção dos pares de coordenadas que constituem a curva ROC. Desta forma, quantos mais *thresholds*, igualmente espaçados, forem utilizados, mais pares de coordenadas serão obtidos para a construção da curva. Portanto, a *receiver operating characteristics curve* ilustra a troca entre valores de TPR e FPR para diferentes valores limites. A Figura 25 demonstra um exemplo da curva ROC.

Outro parâmetro que pode ser retirado da curva ROC é a área de baixo da curva (em inglês *area under the curve - AUC*), representada pela hachura em azul. De acordo com Fawcett (2006), a AUC representa a performance do algoritmo de classificação, ou seja, uma AUC igual a 1 significa que o algoritmo classificou, para determinados *thresholds*, perfeitamente

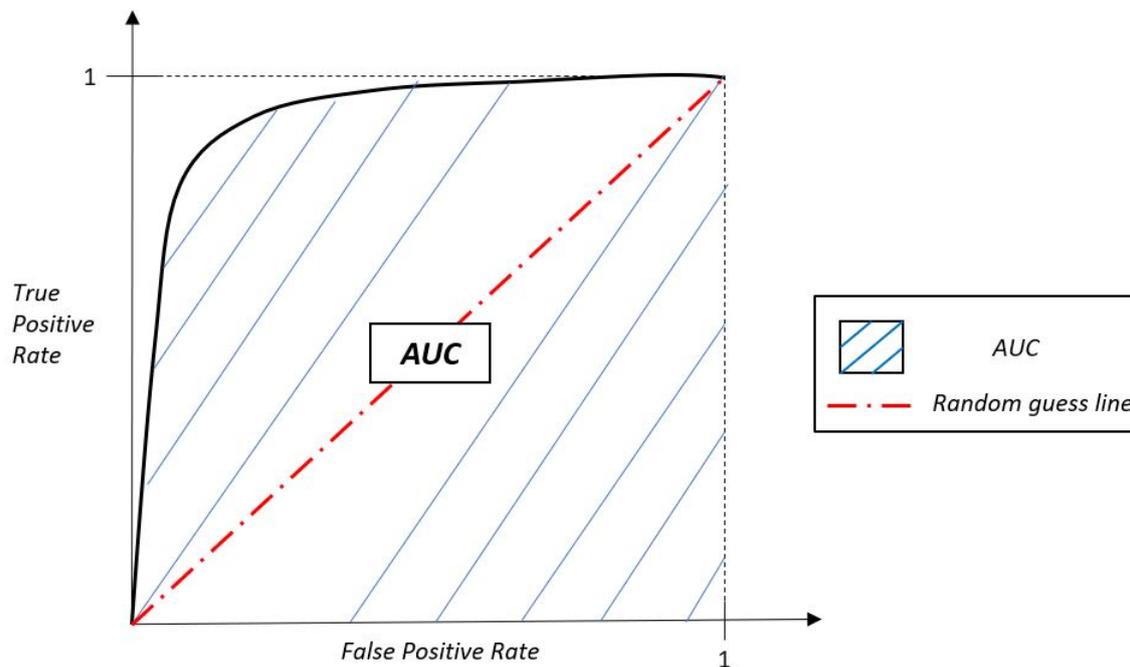


Figura 25 – Exemplo de *Receiver Operating Characteristics Curve* (Fonte: autor).

todas as informações do banco de dados. A linha vermelha em diagonal é chamada de linha de classificação aleatória (em inglês *random guess line*). Esta linha indica se o modelo em estudo performa melhor ou pior que uma classificação aleatória, em outras palavras, se a classificação realizada pelo algoritmo é melhor que uma classificação randômica. Fica evidente que se deseja obter modelos os quais a curva ROC esteja sempre acima da *random guess line* para todos os pares de coordenadas para os *thresholds* analisados.

De acordo com Fawcett (2006), alguns algoritmos de classificação produzem uma pontuação que representa a probabilidade de uma observação pertencer a uma determinada categoria ou classe. Estes valores podem ser probabilidades absolutas as quais aderem aos teoremas de probabilidade, ou valores genéricos não calibrados. Neste caso, a única propriedade que este valor contém é que quanto maior a pontuação maior a probabilidade de uma observação pertencer a uma determinada classe ou categoria. A curva ROC e a AUC são insensíveis ao fato de suas probabilidades previstas, ou seja, as pontuações calculadas, estarem ou não devidamente calibradas para realmente representar as probabilidades de determinada observação pertencer a uma certa classe ou categoria (FAWCETT, 2006). Desta forma, pode-se afirmar que a métrica da *ROC curve* desempenham unicamente para demonstrar o quão bem um algoritmo separa suas classes ou categorias, ou seja, a AUC representa quanto um modelo classificará uma observação positiva randomicamente escolhida com maior probabilidade que uma observação negativa escolhida randomicamente.

Por definição, a utilização da curva ROC e a AUC, como ferramentas para analisar o desempenho de um algoritmo estudado, é restrita para problemas bi-dimensionais, ou seja, problemas que contêm apenas duas classes ou categorias (FAWCETT, 2006). Desta forma, como este trabalho aplicou mais de duas classes/categorias para estudar as ferramentas de classificação do *machine learning*, foi utilizado um método de aproximação da curva ROC e a AUC para problemas multi-dimensionais. De acordo com Fawcett (2006), curvas ROC multi-dimensionais provocam um grande acréscimo de complexidade ao problema, uma vez que, para \mathbf{n} classes a matriz de confusão se torna uma matriz $\mathbf{n} \times \mathbf{n}$, contendo \mathbf{n} classificações corretas (valores da diagonal principal da *confusion matrix*) e $\mathbf{n}^2 - \mathbf{n}$ possíveis erros (valores fora da diagonal principal). Ao invés de gerenciar somente os prós e contras entre TP e FP, o problema multi-dimensional contém \mathbf{n} benefícios e $\mathbf{n}^2 - \mathbf{n}$ erros, ou seja, com apenas três classes o problema se torna um politopo com 6 ($\mathbf{3}^2 - \mathbf{3} = \mathbf{6}$) dimensões. Para simplificar essa análise multi-dimensional, um método que pode ser empregado é o método chamado de “um contra todos” (em inglês “one versus all”) (FAWCETT, 2006). Este método consiste em produzir \mathbf{n} curvas ROC, considerando problemas com \mathbf{n} classes, onde cada curva representa uma classe. Em outras palavras, se C é um conjunto de todas as classes ou categorias, a curva ROC i demonstrará a performance de classificação usando a classe c_i como a classe positiva, enquanto todas as outras classes são consideradas negativas. As Equações 2.5.7 e 2.5.8 a seguir resumem o método de maneira simplificada.

$$P_i = c_i \quad (2.5.7)$$

$$N_i = \bigcup_{j \neq i} c_j \in C \quad (2.5.8)$$

Sendo: P_i a classe considerada positiva (c_i) dentro do conjunto de classes formado por C e N_i a união de todas as outras classes, contidas em C , que são consideradas negativas (c_j).

Em resumo, para um problema de, por exemplo, três classes a matriz de confusão teria uma estrutura $\mathbf{3} \times \mathbf{3}$ e, através do método proposto por Fawcett (2006), seria necessário a criação de três curvas ROC. A Figura 26 e a Tabela 10 a seguir demonstram este exemplo.

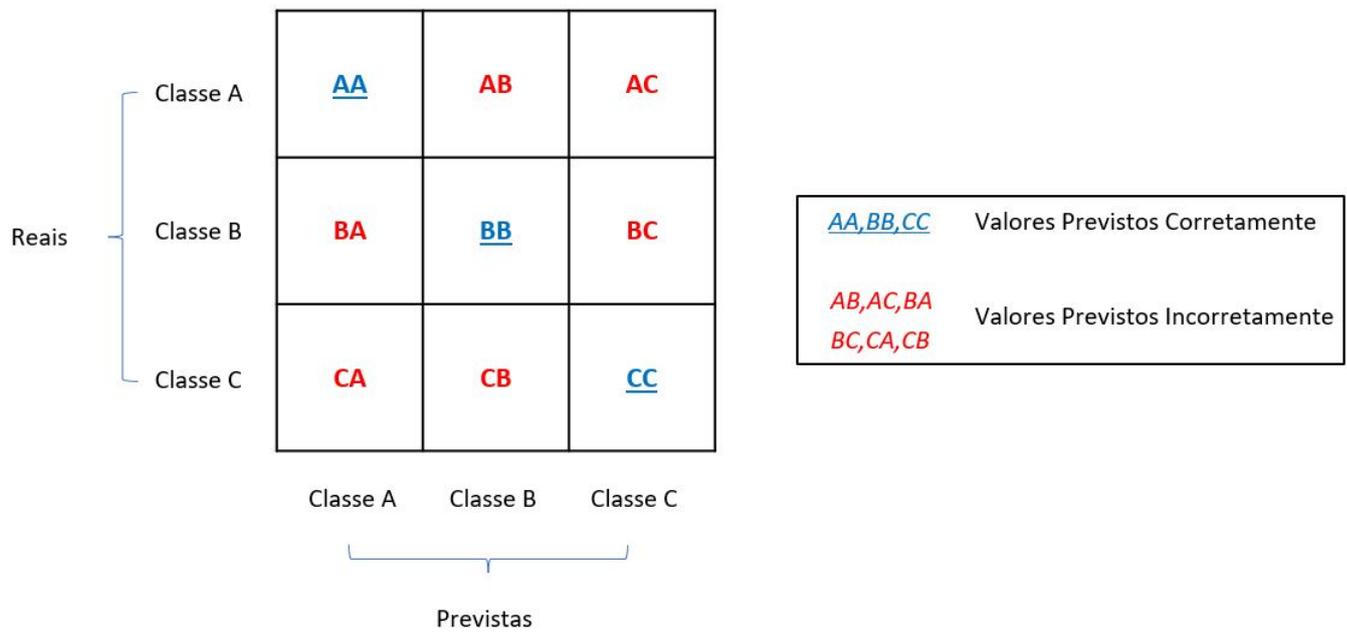


Figura 26 – Exemplo da *confusion matrix* para 3 classes (Fonte: autor).

Tabela 10 – Configuração das curvas ROC a partir de um *confusion matrix* 3x3 (Fonte: autor).

ROC	Classe Positiva - P_i	Classes negativas - N_i
Curva ROC 01	A	B e C
Curva ROC 02	B	A e C
Curva ROC 03	C	A e B

2.6 Síntese do capítulo

Neste capítulo foi discutido a situação no Brasil frente aos trabalhos sobre atividade sísmica mais recentes em comparação com a norma nacional para estruturas sismo-resistentes (NBR 15421, 2023). Ademais, foi descrita a metodologia para a classificação das OAEs da região Nordeste do Brasil e quais critérios foram adotados para a escolha da tipologia alvo deste estudo. Na sequência, foram definidos o processo de obtenção das curvas de fragilidade e a importância desta ferramenta para a análise de vulnerabilidade sísmica de estruturas. Por fim, foi discutido o uso de ferramentas de inteligência artificial, como os métodos de *Machine Learning*, e as importantes contribuições que o uso destas ferramentas vêm mostrando no âmbito da ciência dos dados.

3 Metodologia

Este capítulo irá apresentar as etapas necessárias para a realização deste trabalho. A lista a seguir demonstra cada etapa realizada.

- *Escolha da tipologia para o estudo;*
- *Geração do banco de dados;*
- *Modelagem numérica;*
- *Análises dinâmicas não lineares;*
- *Modelos de demanda e capacidade;*
- *Curvas de fragilidade;*
- *Machine Learning;*
 - Regressão;*
 - Classificação.*

3.1 Tipologia adotada

A tipologia adotada para este trabalho é a tipologia 3 (T3), que entre as tipologias consideradas representativas (T1, T2 e T3) é a que apresenta sistema estrutural mais complexo por apresentar encontros e pórticos internos. De acordo com Cavalcante (2022), esta tipologia ainda pode ser dividida em diferentes classes que apresentem comportamento similar diante de um terremoto. No caso da tipologia T3, as classes foram agrupadas em função do número de pilares por pórticos e do tipo de seção transversal do tabuleiro, sendo (Cavalcante (2022)):

- Classe 1 (C1): dois pilares por pórtico e seção transversal “T”;
- Classe 2 (C2): um pilar-parede por pórtico e seção transversal “T”;
- Classe 3 (C3): um pilar-parede por pórtico e seção caixão.

A classe C1 representa cerca de 40% do total de 40 OAEs analisadas da tipologia T3, seguidos de 25% da classe C2 e 15% da classe C3 (Cavalcante (2022)). Os 20% restantes

apresentam sistema estrutural diverso e diferente das respectivas classes da tipologia T3. A Figura 27 ilustra a divisão de classes, através de um corte transversal na região do pórtico interno, dentro da tipologia adotada para este trabalho.

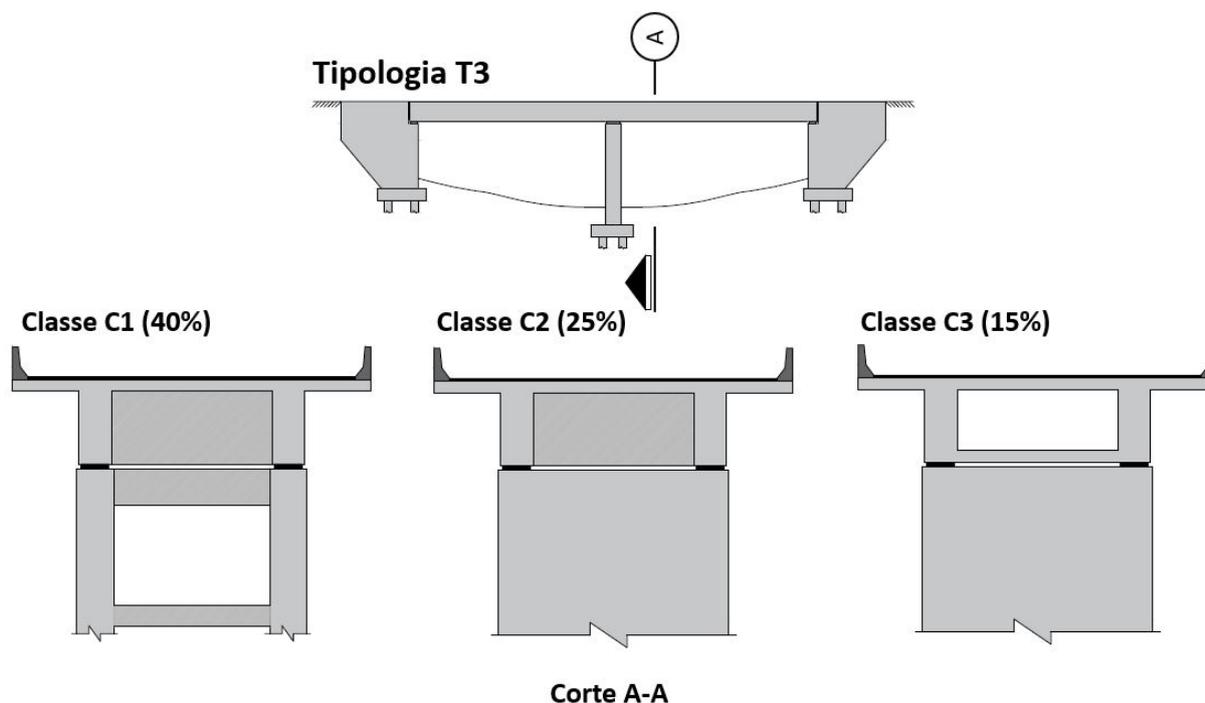


Figura 27 – Classes C1, C2 e C3 da ponte de tipologia T3 (Fonte: adaptado de Cavalcante (2022)).

Como a classe C1 representa quase metade das pontes da tipologia T3 analisadas, conforme levantado por Cavalcante (2022), as características do sistema estrutural da respectiva classe será adotada como objeto de estudo deste trabalho. A Tabela 11 demonstra a descrição final da classe adotada para a tipologia T3.

Tabela 11 – Características estruturais da classe adotada (Fonte: adaptado de Cavalcante (2022)).

Tipologia	Classe	Características
T3	C1	Múltiplos vãos contínuos Tabuleiros sem juntas intermediárias Pórticos intermediários rotulados Dois pilares por pórtico intermediário, ligado por viga travessa Seção transversal "T" Aparelhos de apoio elastoméricos Fundações profundas (adotado)

Para melhor descrever os parâmetros obtidos através DNIT, como número, comprimento e largura dos vãos, comprimento das longarinas, geometria dos pilares e vigas etc, faz-se necessário

criar funções probabilísticas que melhor representem os valores coletados (Cavalcante (2022))). São criadas funções de distribuições cumulativas (CDF) (em inglês *cumulative distribution function*) que descrevem a distribuição probabilística de uma variável aleatória de valor real X , ou seja, a CDF de uma variável X avaliada em x é a probabilidade de X assumir um valor menor ou igual a x (equação 3.1.1).

$$F_X(x) = P(X \leq x) \quad (3.1.1)$$

De acordo com Nielson (2005), pode-se gerar CDF empíricas que indicam que a probabilidade de ocorrência de um dado é igual a $1/N$, sendo N o total de dados de um determinado conjunto discreto de dados. Assumindo-se um conjunto de dados suficientemente grande e que segue uma distribuição normal, a média μ_X , o desvio padrão σ_X e a CDF podem ser definidas como:

$$\mu_X = \sum_{k=1}^N \frac{x_i}{N} \quad (3.1.2)$$

$$\sigma_X = \sqrt{\sum_{k=1}^N \frac{(x_i - \mu)^2}{N - 1}} \quad (3.1.3)$$

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sigma_X \cdot \sqrt{2 \cdot \pi}} \cdot e^{\left[-\frac{1}{2} \cdot \left(\frac{t - \mu_X}{\sigma_X}\right)^2\right]} dt \quad (3.1.4)$$

De acordo com Nowak e Collins (2012), a distribuição lognormal de uma variável aleatória pode ser adotada como $y = \ln(x)$ normalmente distribuída e, portanto, a média μ_Y , o desvio padrão σ_Y e a CDF podem ser definidas como:

$$\mu_Y = \ln(\mu_X) - \frac{\sigma_Y^2}{2} \quad (3.1.5)$$

$$\sigma_Y = \sqrt{\ln\left[\left(\frac{\sigma_X}{\mu_X}\right)^2 + 1\right]} \quad (3.1.6)$$

$$F_Y(y) = \int_{-\infty}^y \frac{1}{\sigma_Y \cdot \sqrt{2 \cdot \pi}} \cdot e^{\left[-\frac{1}{2} \cdot \left(\frac{t - \mu_Y}{\sigma_Y}\right)^2\right]} dt \quad (3.1.7)$$

Para poder comparar e validar se uma amostra segue uma distribuição de probabilidade de referência, como por exemplo a distribuição normal ou lognormal, Cavalcante (2022) aplicou

o teste *Kolmogorov-Smirnov* (teste K-S) para verificar se os dados poderiam ser ajustados a uma distribuição normal para um determinado nível de significância e aplicou o teste *Kolmogorov-Smirnov* biamostrais para o ajuste com distribuições lognormais. Todos os testes das variáveis aleatórias utilizaram nível de significância de 5%.

De acordo com Cavalcante (2022), as variáveis aleatórias para as OAEs representantes da classe C1 da tipologia T3 são as seguintes: número de vãos (n_v), comprimento dos vãos (L_v), altura das longarinas (h_{long}), altura dos pilares (h_{pil}), distância entre os eixos dos pilares (d_{pil}), diâmetro dos pilares (ϕ_{pil}), altura dos encontros (H_{enc}), comprimento das abas dos encontros (B_{aba}) e as alturas das abas dos encontros (H_{aba1}) e (H_{aba2}). Variáveis como a espessura da laje, largura das longarinas, largura e altura das vigas travessas, tipo de seção transversal dos pilares, largura do tabuleiro, largura dos encontros e número de longarinas foram mantidas constantes por representarem valores similares entre as pontes estudadas (Cavalcante (2022)). A Tabela 12 a seguir demonstra, de maneira resumida, os valores adotados fixos e os parâmetros das variáveis aleatórias para as pontes da classe C1 tipologia T3.

Tabela 12 – Valores adotados e parâmetros das variáveis aleatórias para as pontes da classe C1 tipologia T3 (Fonte: adaptado de Cavalcante (2022)).

Propriedade estrutural	Definição	Distribuição	Parâmetros
Espessura da laje	Valor fixo	-	0,4 m
Largura das longarinas	Valor fixo	-	0,4 m
Largura das travessas	Valor fixo	-	0,4 m
Altura das travessas	Valor fixo	-	1,0 m
Tipo de seção dos pilares	Valor fixo	-	Circular
Largura do tabuleiro	Valor fixo	-	10 m
Largura dos encontros	Valor fixo	-	10 m
Número de longarinas	Valor fixo	-	2 unidades
Número de vãos (n_v)	Var. aleatória	Uniforme	$n_v = 2$ (67,7%) $n_v = 3$ (26,7%) $n_v = 4$ (6,6%)
Comprimento dos vãos (L_v)	Var. aleatória	Lognormal	$\mu = 2,7m$; $\sigma = 0,37m$
Altura das longarinas (h_{long})	Var. aleatória	-	$h_{long} = 0,09 \cdot L_v + 0,07$
Altura dos pilares (h_{pil})	Var. aleatória	Normal	$\mu = 4,88m$; $\sigma = 1,57m$
Dist. entre pilares (d_{pil})	Var. aleatória	Lognormal	$\mu = 1,57m$; $\sigma = 0,04m$
Diâmetro dos pilares (ϕ_{pil})	Var. aleatória	-	$\phi_{pil} = 0,03 \cdot L_v + 0,36$
Altura dos encontros (H_{enc}) Tipo gravidade (42,9%)	Variável aleatória	Lognormal	$\mu = 1,65m$ $\sigma = 0,41m$
Altura dos encontros (H_{enc}) Tipo “U” (57,1%)	Variável aleatória	Lognormal	$\mu = 1,96m$ $\sigma = 0,13m$

Como não há um predomínio entre os tipos de encontros, foram considerados dois tipos de encontros: tipo “U” e do tipo gravidade (Cavalcante (2022)). Para mais informações referentes a obtenção dos valores expostos, o leitor é direcionado ao trabalho de Cavalcante (2022). A Figura 28 ilustra um modelo 3D da tipologia adotada para este trabalho.

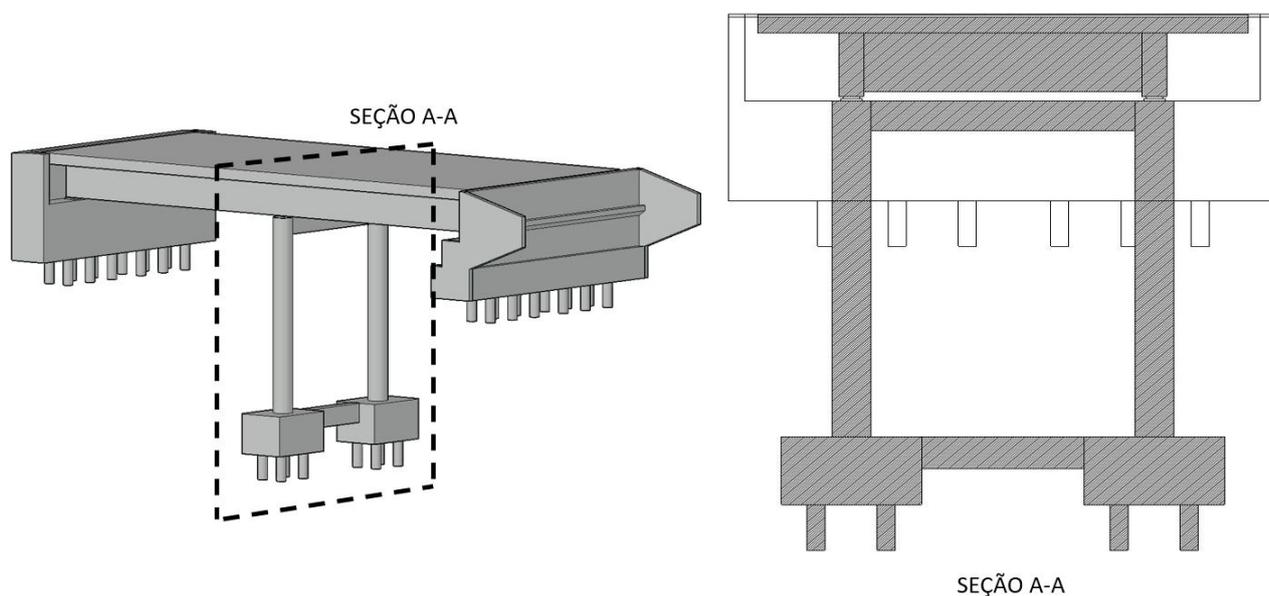


Figura 28 – Modelo da tipologia adotada (Fonte: autor).

3.2 Geração do banco de dados

Para a geração de dados, foram realizadas análises dinâmicas não lineares de 100 OAEs da tipologia T3, obedecendo a variabilidade das propriedades estruturais demonstradas pela Tabela 12, solicitadas por 100 acelerações compatíveis com a região Nordeste do Brasil, totalizando, desta forma, 10.000 (dez mil) análises.

Devido à necessidade de automatização do processo probabilístico para a aplicação das ferramentas de *machine learning*, utilizou-se uma rotina computacional no software MATLAB® com interação com a plataforma OpenSees (2019) (MCKENNA et al., 2006), uma vez que esta plataforma não possui interface gráfica. Esta rotina é denominada *Failure Engineering and Analysis of Reliability for Bridges* (FEAR-Bridge). Este programa faz parte de um projeto de pesquisa conjunto desenvolvido no Risk Engineering Laboratory (RELab) da Universidade Estadual de Campinas (UNICAMP) para análise de risco em estruturas de pontes. O RElab tem desempenhado um papel importante na análise de vulnerabilidade estrutural como mostram os trabalhos de Pereira et al. (2024), Rodrigues et al. (2024), Leitão et al. (2023) e Andrade et al. (2022).

As análises numéricas são realizadas na plataforma do OpenSees (2019), sendo este um programa computacional de elementos finitos desenvolvido para análise sísmica de estruturas e que dispõe de uma extensa biblioteca de elementos e parâmetros configuráveis (MAZZONI et al., 2006). É permitido ao usuário utilizar diversas possibilidades de elementos, que melhor se adequem ao tempo e custo computacional disponível, adotar diferentes modelos constitutivos e alterar parâmetros das análises como métodos de convergência, número de iterações e tolerâncias, dentre outros.

Desta forma, a utilização do FEAR-Bridge pode ser dividida em três processos, conforme os itens a seguir:

- *Pré-processamento*: as propriedades físicas e geométricas e os sinais de acelerações compatíveis são introduzidos para a criação dos modelos estruturais. Desta forma, os elementos são discretizados e as respectivas relações constitutivas dos materiais são criadas;
- *Processamento*: os modelos numéricos criados pelo MATLAB[®] são calculados pela plataforma OpenSees (2019);
- *Pós-processamento*: os resultados obtidos no OpenSees (2019) são importados e trabalhados pelo MATLAB[®], gerando as respostas das análises.

A rotina computacional do FEAR-bridge permite, de maneira gráfica, a introdução dos parâmetros físicos e geométricos da ponte a ser analisada. O programa possui dois módulos: determinístico e estocástico. O módulo determinístico permite a criação e o estudo, através de análises modais, estáticas não lineares e dinâmicas, de um modelo estrutural de uma ponte com suas respectivas condições de contorno e discretização dos elementos. Posteriormente a realização da análise no módulo determinístico, é possível obter: os modos de vibração e participação modal, curva de capacidade dos pilares e histórico de resposta no tempo.

Já o módulo estocástico funciona do mesmo jeito que o determinístico, porém na etapa de criação do modelo são criados n modelos em função das variáveis aleatórias das propriedades estruturais e materiais. Posteriormente a realização das análises dos modelos no módulo estocástico, é possível obter os valores de demanda em função do IM e aplicar o PSDM. Com estes valores e considerando os valores de capacidade adotados neste trabalho (ver Tabelas 6 e 7), torna-se possível a determinação das curvas de fragilidade. Os modelos estocásticos foram gerados a partir da aplicação dos métodos de Monte Carlo (MMC) e *Latin Hypercube Sampling* (LHS). O MMC é utilizado para a geração das amostras randômicas das variáveis aleatórias e o LHS é utilizado para reduzir o número de simulações necessárias para a obtenção de resultados confiáveis.

De acordo com Nowak e Collins (2012), a utilização do LHS permite uma boa distribuição das variáveis aleatórias, gerando assim uma boa representatividade para a realização das simulações. A geração das variáveis aleatórias que seguem distribuições normais e lognormais foi realizada pelo LHS, enquanto as variáveis que seguem distribuições uniformes foram geradas pelo MMC (CAVALCANTE, 2022).

De acordo com Cavalcante (2022), as propriedades materiais não foram fornecidas pelo Departamento Nacional de Infraestrutura de Transporte (DNIT). Desta forma, recorreu-se a valores fornecidos pela literatura.

As propriedades materiais consideradas aleatórias são: resistência do concreto à compressão (f_c); tensão de escoamento do aço (f_y); módulo de elasticidade longitudinal do aço (E_s); comprimento das juntas (*Gaps*); taxa de amortecimento da estrutura (ξ); módulo de elasticidade transversal dos aparelhos de apoio elastoméricos (G). A seguir, a Tabela 13 demonstra os valores da média e do desvio padrão das propriedades citadas a cima.

Tabela 13 – Propriedades materiais aleatórias (Fonte: adaptado de Cavalcante (2022)).

Variável	Distrib.	Média - μ	Desv. Padrão - σ	Referência
f_c (MPa)	Normal	26,20	4,30	Santiago e Beck (2017)
f_y (MPa)	Normal	576	63,36	Nogueira (2010)
E_s (GPa)	Normal	200	6,60	Mirza e MacGregor (1979)
<i>Gaps</i> (mm)	Normal	24,50	5	Tavares et al. (2012)
ξ (%)	Lognormal	1,956	0,176	Siqueira et al. (2014a)
G (MPa)	Lognormal	-0,203	0,141	Siqueira et al. (2014a)

Para as armaduras longitudinais dos pilares foram adotadas taxas entre 0,97% e 1,1%, uma vez que, de acordo com Padgett e DesRoches (2009), a taxa de 1% representa a taxa para os pilares de OAEs que não foram dimensionadas para eventos sísmicos. Portanto, o diâmetro e o número de barras variam em função da área da seção transversal do pilar. Os estribos foram adotados com ϕ igual a 10 mm espaçados a cada 10 cm. Para ambos pilares e vigas, o cobrimento adotado foi de 3 cm. A Figura 29 a seguir demonstra a variação de taxa de armadura dos pilares em função da seção transversal.

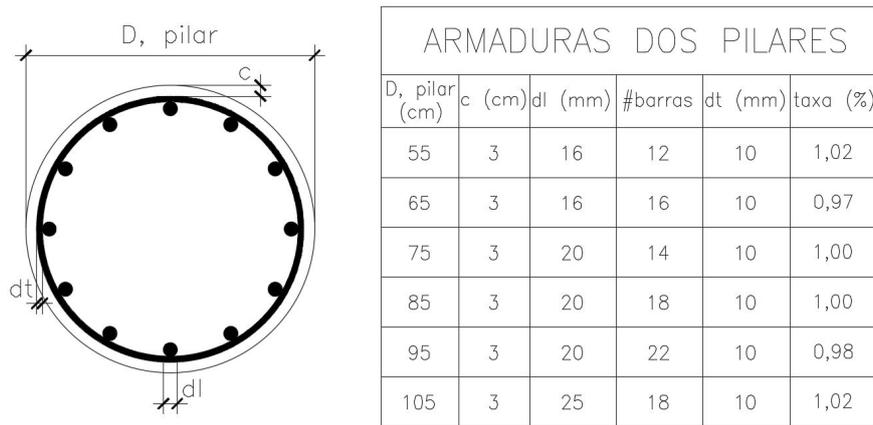


Figura 29 – Taxa de armadura dos pilares em função da seção transversal (Fonte: autor).

Como as vigas travessas para a classe C1 da tipologia T3 apresentam área da seção transversal constante, 40 cm x 100 cm, foram adotados os valores preconizados pela NBR 6118 (2024) para armadura mínima (conforme item 17.3.5 da referida norma). Sabendo que o f_c do concreto é uma variável que segue distribuição normal, o maior valor calculado para a resistência característica a compressão do concreto foi de aproximadamente 40 MPa. Desta forma, a maior taxa mínima a ser utilizada é de 0,179% (taxa mínima para concretos de 40 MPa). Portanto, como armadura de flexão foram adotadas 4 barras nas faces superiores e inferiores com ϕ igual a 16 mm (taxa igual a 0,4%), e como armadura de pele foram adotadas 5 barras laterais, por face, com ϕ igual a 10 mm (taxa igual a 0,2%), superior a taxa mínima necessária de 0,10%. Como armadura de cisalhamento foram adotadas barras de ϕ igual a 8 mm espaçadas a cada 15 cm (taxa igual a 0,18%), sendo a taxa mínima necessária de 0,14%. A Figura 30 demonstram a armadura adotada para as vigas travessas.

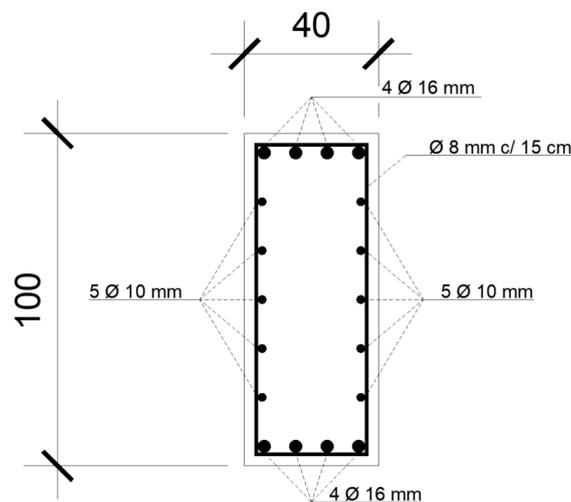


Figura 30 – Armadura adotada para as vigas travessas (Fonte: autor).

Os valores de área e espessura média dos aparelhos de apoio elastoméricos foram retirados do catálogo da empresa NEOPREX, apresentados em Cavalcante (2022). De acordo com Kelly (1993), os aparelhos de apoio foram dimensionados para tensões de compressão de serviço iguais a 7 MPa.

Os sinais de aceleração do solo utilizados neste trabalho são aqueles compatíveis com a região de estudo. Para tal, é necessário conhecer o espectro de resposta desta região. Conforme visto no item 2.4.1, a resposta dinâmica de uma estrutura varia de acordo com seu respectivo período natural de vibração, desta forma, o espectro de resposta representa o máximo valor de deslocamento, ou velocidade, ou aceleração, em função do período natural de um sistema de um grau de liberdade. Como a utilização individual de sismos é pouco prática para estudos de modelos estocásticos, utilizou-se o espectro de resposta médio definido pela norma ASCE (2017), uma vez que a norma NBR 15421 (2023) não permite a construção de espectros de resposta para períodos curtos e longos. Conforme já mencionado neste trabalho, os valores de acelerações sísmicas utilizados foram retirados de Petersen et al. (2018), considerando um tempo de retorno (TR) de 475 anos.

A ASCE (2017) utiliza um espectro de resposta com fator de amortecimento de 5% em função do tipo de solo do local de estudo, devido a alteração que solos moles ou rígidos provocam nas acelerações sísmicas. A Figura 31 a seguir demonstra o espectro de resposta de acordo com a norma ASCE (2017).

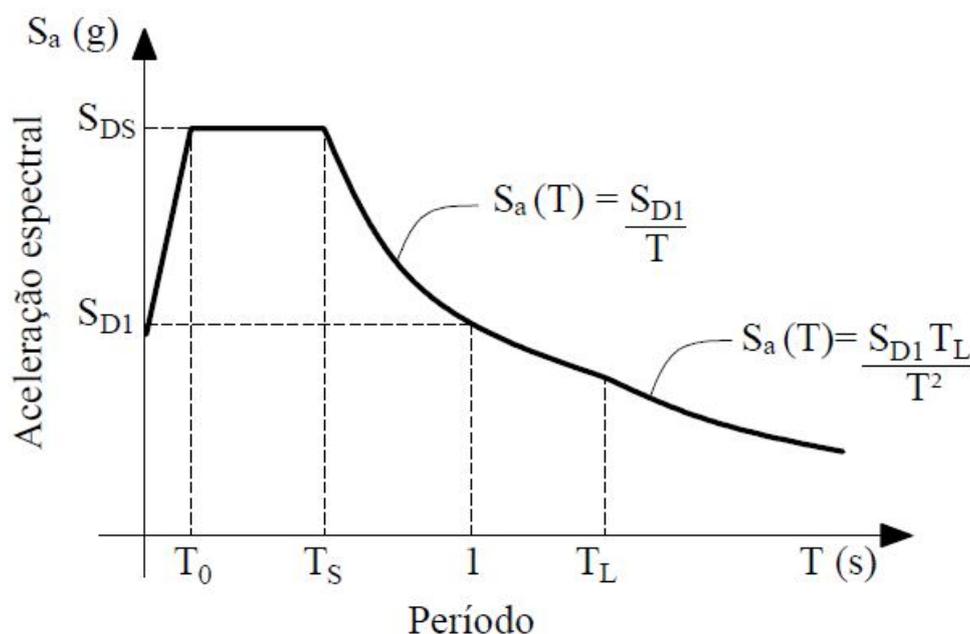


Figura 31 – Espectro de resposta (Fonte: adaptado de ASCE (2017)).

Os valores S_{DS} e S_{DI} são calculados com base no tipo de período utilizado (curto ou longo) e nos coeficientes do terreno no local de estudo. O período longo corresponde a 1 segundo e o período curto a 0,2 segundos. De acordo com Cavalcante (2022), os dados utilizados são referentes a cidade de Natal-RN, devido a localização das sondagens a percussão. Com base nas sondagens, 98,9% dos resultados apresentaram N_{spt} médios entre 15 e 50, classificando assim o solo como tipo D conforme a NBR 15421 (2023). Desta forma, utilizando a metodologia presente na norma ASCE (2017), foi construído o espectro de resposta considerando solo D para valores de acelerações espectrais para períodos curtos e longos de 0,15g e 0,03g, respectivamente. Estes valores propostos representam a média geométrica dos valores nas duas direções ortogonais, compatibilizando o espectro para análises bidirecionais. A Figura 32 a seguir demonstra o espectro de resposta para solo tipo D com TR de 475 anos.

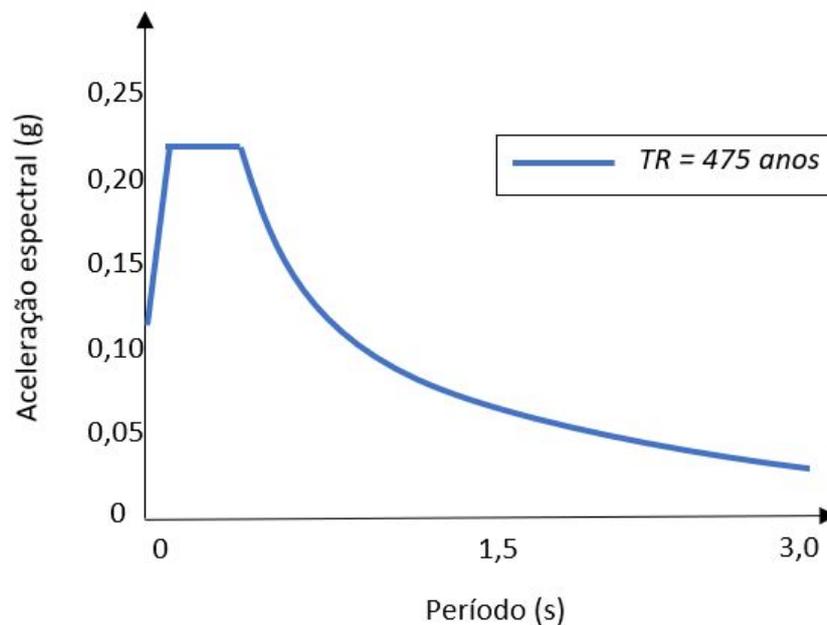


Figura 32 – Espectro de resposta para solo D considerando TR de 475 anos (Fonte: autor).

A partir do espectro de resposta adotado, é possível obter acelerogramas que sejam compatíveis. A base de dados escolhida foi a do *Pacific Earthquake Engineering Research Center* (PEER). Vale ressaltar que, como mencionado anteriormente, a compatibilidade é dada através da média geométrica dos espectros dos sinais encontrados no catálogo do PEER. Os parâmetros adotados para coleta dos registros compatíveis seguem na Tabela 14 a seguir.

Tabela 14 – Parâmetros adotados para coleta dos registros sísmicos compatíveis com o espectro de resposta (Fonte: adaptado de Cavalcante (2022)).

Parâmetro	Valor máximo	Valor mínimo
V_{s30} (m/s)	370	180
R_{rup} (km)	100	20
R_{JB} (km)	100	20
Magnitude	6,5	4,0

Sendo: V_{s30} a velocidade das ondas transversais nos 30 primeiros metros a partir da superfície terrestre; R_{rup} a menor distância do local ao plano de ruptura; R_{JB} a distância Joyner-Boore ao plano de ruptura.

As velocidades máximas e mínimas foram adotadas conforme classificação do solo (tipo D) e, de acordo com Beyer e Bommer (2007), para ambas as distâncias ao plano de ruptura não foi adotado um tipo de falha e por ser tratar de atividades sísmicas rasas os valores adotados foram os mesmos.

De acordo com Beyer e Bommer (2007), tanto a definição de qual componente é paralela ou perpendicular ao plano de ruptura quanto o ângulo de incidência das atividades sísmicas influenciam na resposta. No entanto, devido a região de estudo ser uma área muito extensa os registros sísmicos foram aplicados nas direções paralelas e perpendiculares ao tráfego de maneira aleatória e considerando que o custo computacional, em considerar o ângulo de incidência para trabalhos de análise estocástica, teria um grande acréscimo, adotou-se a orientação dos registros do PEER.

Nenhuma restrição foi imposta na duração das atividades sísmicas, pois o presente trabalho está interessado em avaliar os parâmetros de demanda através dos deslocamentos e, portanto, a duração do sismo não deve ser uma preocupação apesar de influenciar na dissipação de energia (IERVOLINO et al., 2008).

A Figura 33 a seguir demonstra os espectros individuais, o espectro médio e o espectro-alvo para um TR de 475 anos considerando uma taxa de amortecimento igual a 5% (CAVALCANTE, 2022).

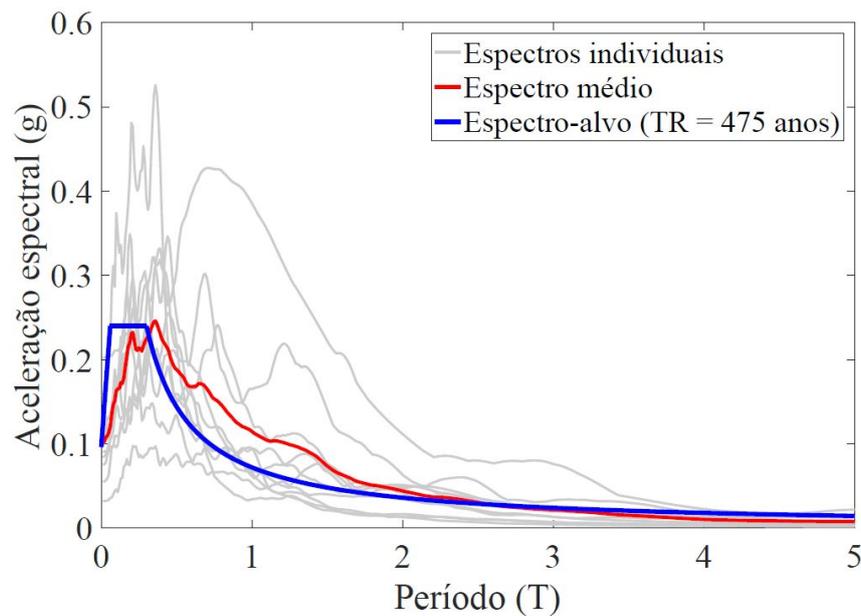


Figura 33 – Espectros individuais, médio e alvo para solo tipo D considerando TR de 475 anos (Fonte: adaptado de Cavalcante (2022)).

3.3 Modelagem numérica

Com o objetivo de representar a ponte da classe C1 da tipologia T3, estrutura em estudo (Figura 28), modelos numéricos são empregados para simular as respostas estruturais. A metodologia utilizada neste trabalho pode ser encontrada em diversos trabalhos como Tavares et al. (2012), Siqueira et al. (2014b), Pan et al. (2010) e Song et al. (2019).

De acordo com Choi e Jeon (2003), Nielson e DesRoches (2006) e Siqueira et al. (2014a), é adotado que o comportamento do tabuleiro se mantém em regime elástico linear quando solicitados por uma atividade sísmica e, portanto, estes são modelados através de elementos de pórticos elásticos lineares (em inglês *elastic beam-column element*). Os nós destes elementos estão posicionados no centro de gravidade da seção transversal e cada um possui 6 graus de liberdade (3 translacionais e 3 rotacionais). Além do peso próprio do tabuleiro, foi considerado o peso próprio da pavimentação e das defensas para a formação da matriz de massa através do método *lumped-mass*, que concentra as massas nos nós.

Já as vigas e os pilares também são modelados por elementos de pórticos, porém não lineares (em inglês *non-linear beam column element*). Estes elementos podem ser analisados com base no deslocamento com plasticidade distribuída (em inglês *displacement-based element*) ou com base no equilíbrio das forças (em inglês *force-based element*). De acordo com a plataforma OpenSees (2019), modelos baseados no equilíbrio das forças exigem maior esforço computacional pela dificuldade de convergência apresentada no processo.

Os modelos analisados com base no deslocamento utilizam equações de equilíbrio na vizinhança do nó em estudo para compatibilizar as deformações. A representação da não-linearidade do elemento é dada através do uso de deformações axiais constantes e distribuições lineares da curvatura ao longo da sua extensão.

Para modelar as ligações entre elementos são utilizadas metodologias complexas, como a utilização de sistemas de barras rígidas (em inglês *rigid-link beam*) ligadas por um elemento sem comprimento (em inglês *zero-length element*), que pode representar uma mola ou amortecedor, e a utilização de espaços (em inglês *gaps*) para simular as juntas entre os componentes estruturais. As barras rígidas são utilizadas para transferir todas as forças sem que haja deformações. Já o *zero-length element* é utilizado para introduzir o comportamento elasto-plástico das ligações.

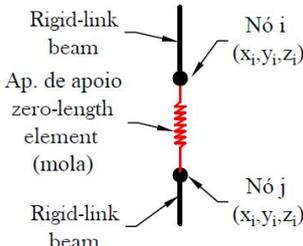
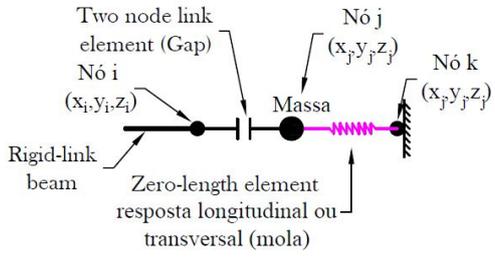
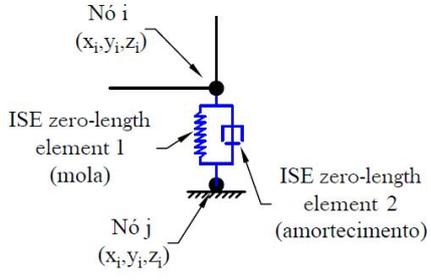
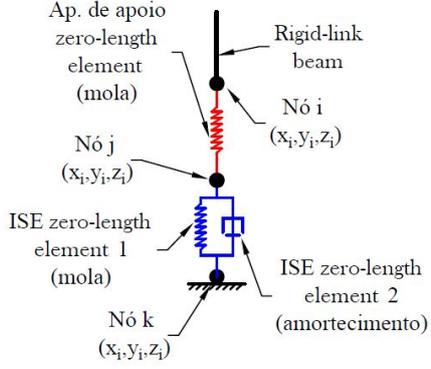
A Tabela 15 a seguir demonstra o tipo de metodologia empregada para cada tipo de ligação estrutural e o seu referente esquema utilizado. Nota-se que na ligação do encontro com o tabuleiro é utilizado uma massa para simular o aterro do encontro.

A Figura 34 a seguir demonstra o modelo matemático concebido para a tipologia T3 classe C1, com as correspondentes ligações, para as análises de fragilidade.

As vigas e os pilares modelados como *non-linear beam column element* são discretizados em fibras com plasticidade distribuída. A discretização consiste em dividir o elemento estrutural em inúmeros, porém finitos, segmentos chamados de fibras. Esta consideração é capaz de verificar a perda de rigidez do elemento finito em função do acréscimo das cargas atuantes, uma vez que quando um elemento de concreto apresenta deformações inelásticas as fissuras se distribuem, resultando em uma redistribuição da rigidez (IZADPANA; HABIBI, 2015). A seção transversal do elemento de concreto pode ser dividida em duas regiões, a região interna delimitada pela armadura transversal (estribos) e a região externa. A Figura 35 a seguir demonstra o modelo dos elementos de concreto dividido em fibras.

Enquanto a região externa apresenta comportamento descrito pelo concreto não-confinado, a região interna é definida pela relação constitutiva do concreto confinado. A região interna possui mais ductilidade e resistência do que comparada com a região externa, uma vez que a região externa não leva em consideração a contribuição das armaduras transversais. Portanto, utiliza-se o modelo definido por Chang e Mander (1994) para modelar os elementos de concreto. Já as armaduras de flexão, são concebidas como elementos pontuais definidos pela relação constitutiva do aço. A relação adotada para o modelo constitutivo do aço foi o de Giuffré-Menegetto-Pinto (GMP) com enrijecimento isotrópico (FILIPPOU et al., 1983) com 5% de enrijecimento.

Tabela 15 – Metodologia utilizada nas ligações para a tipologia T3 (Fonte: adaptado de Cavalcante (2022)).

Ligação	Metodologia	Esquema
Pórticos	<i>Rigid-link/Zero length element</i>	 <p>The diagram shows a vertical rigid-link beam. At the top is Node i with coordinates (x_i, y_i, z_i). At the bottom is Node j with coordinates (x_i, y_i, z_i). A red spring, labeled 'Ap. de apoio zero-length element (mola)', connects the two nodes.</p>
Encontros	<i>Rigid-link/Gap/Zero length element</i>	 <p>The diagram shows a horizontal rigid-link beam. Node i is at (x_i, y_i, z_i) and Node j is at (x_j, y_j, z_j). A gap is shown between Node i and a central 'Massa' (mass). A pink spring, labeled 'Zero-length element resposta longitudinal ou transversal (mola)', connects the mass to Node k at (x_j, y_j, z_j). A label 'Two node link element (Gap)' points to the gap between Node i and the mass.</p>
ISE - pórticos	<i>Rigid-link/Zero length element</i>	 <p>The diagram shows a vertical rigid-link beam. Node i is at (x_i, y_i, z_i) and Node j is at (x_i, y_i, z_i). Two blue ISE zero-length elements are shown: 'ISE zero-length element 1 (mola)' (spring) and 'ISE zero-length element 2 (amortecimento)' (damper), both connecting the nodes.</p>
ISE - encontros	<i>Rigid-link/Zero length element</i>	 <p>The diagram shows a vertical rigid-link beam. Node i is at (x_i, y_i, z_i) and Node j is at (x_i, y_i, z_i). A red spring, labeled 'Ap. de apoio zero-length element (mola)', connects Node i to Node j. Below Node j, there is a damper and spring assembly labeled 'ISE zero-length element 1 (mola)' and 'ISE zero-length element 2 (amortecimento)'. Node k is at (x_i, y_i, z_i).</p>

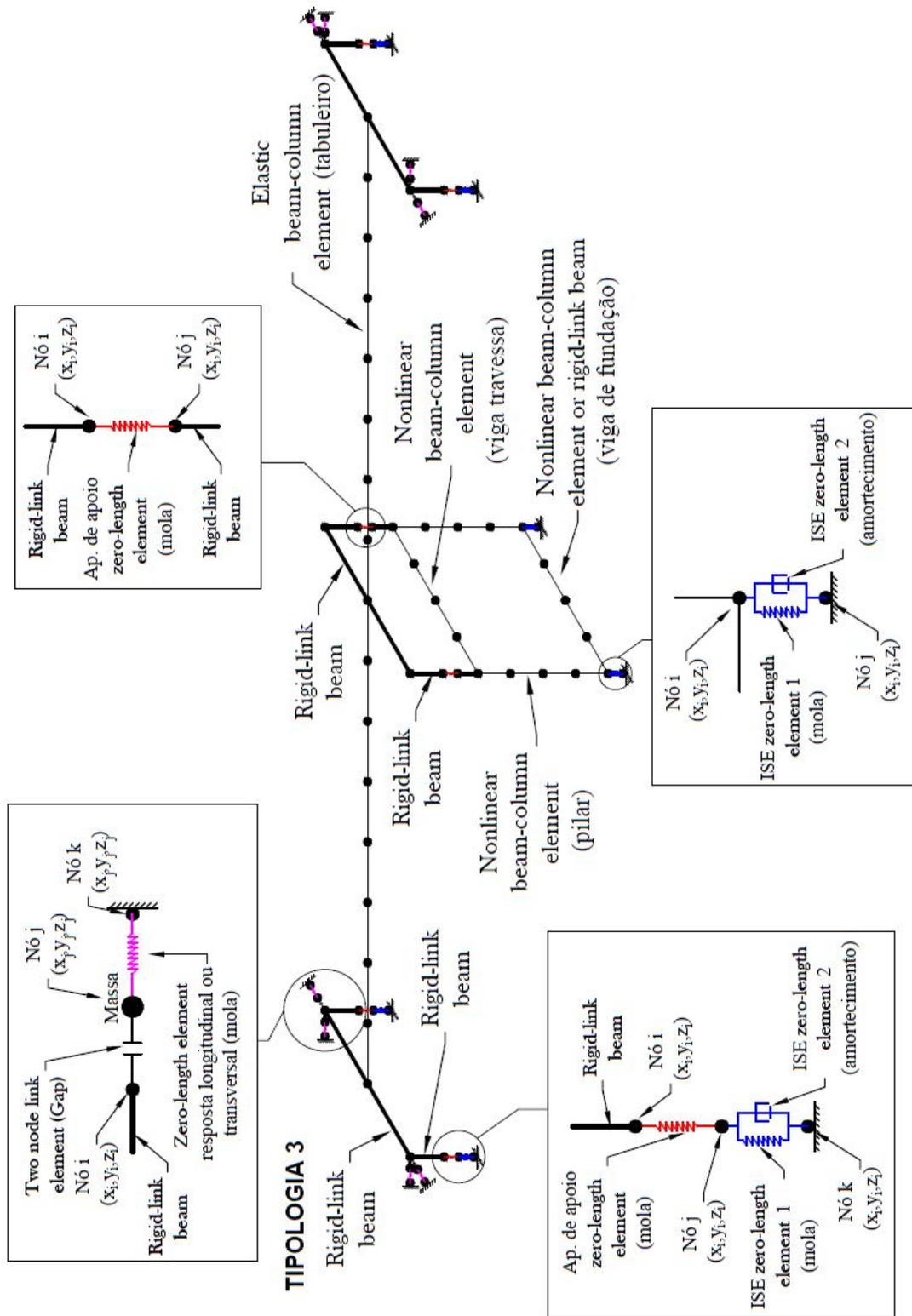


Figura 34 – Modelo estrutural da tipologia T3 (Fonte: adaptado de Cavalcante (2022)).

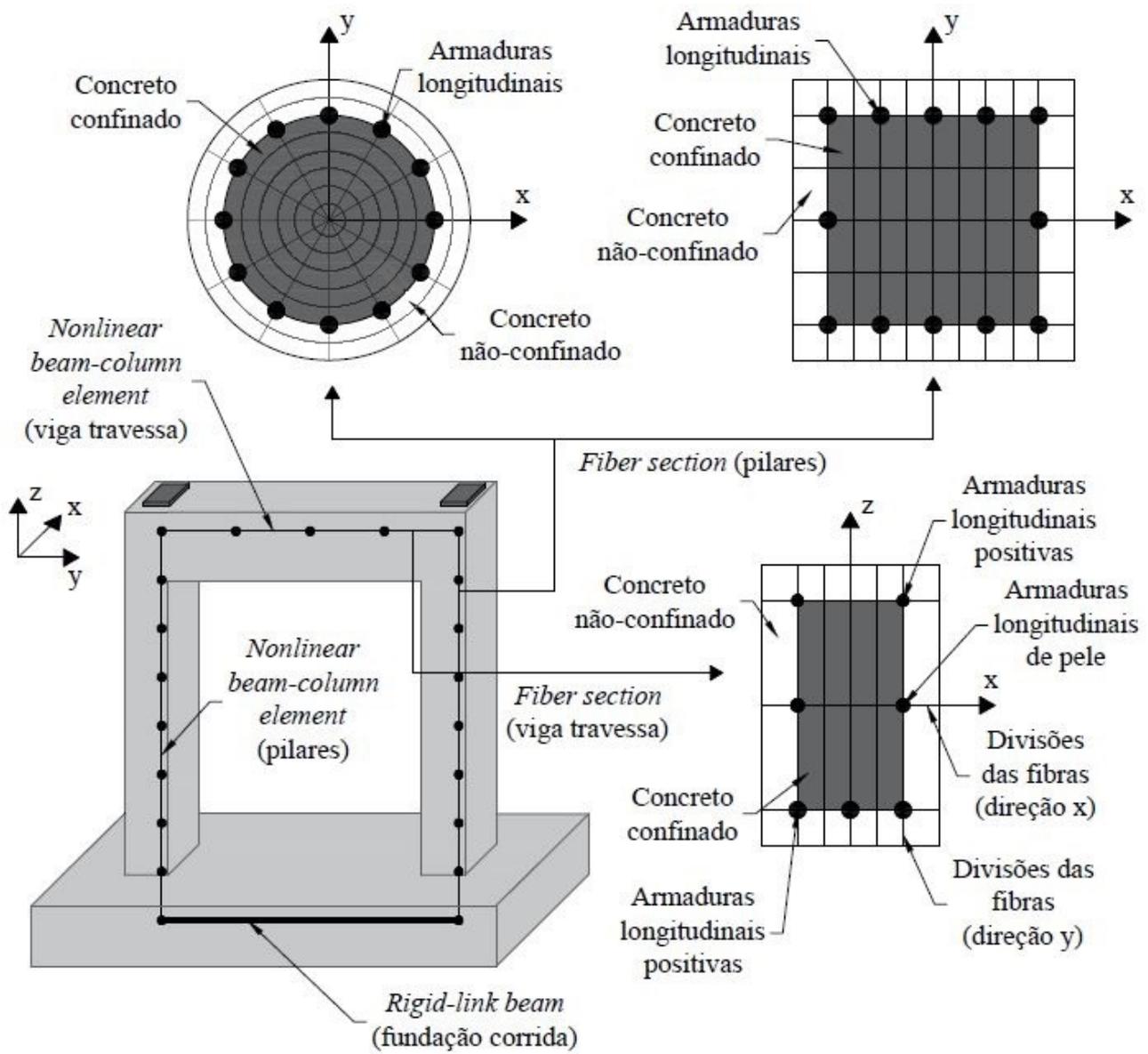


Figura 35 – Modelo dos elementos de concreto dividido em fibras (Fonte: adaptado de Cavalcante (2022)).

A Figura 36 a seguir demonstra os modelos constitutivos adotados para o concreto e para o aço, juntamente com a resposta histerética da armadura de flexão.

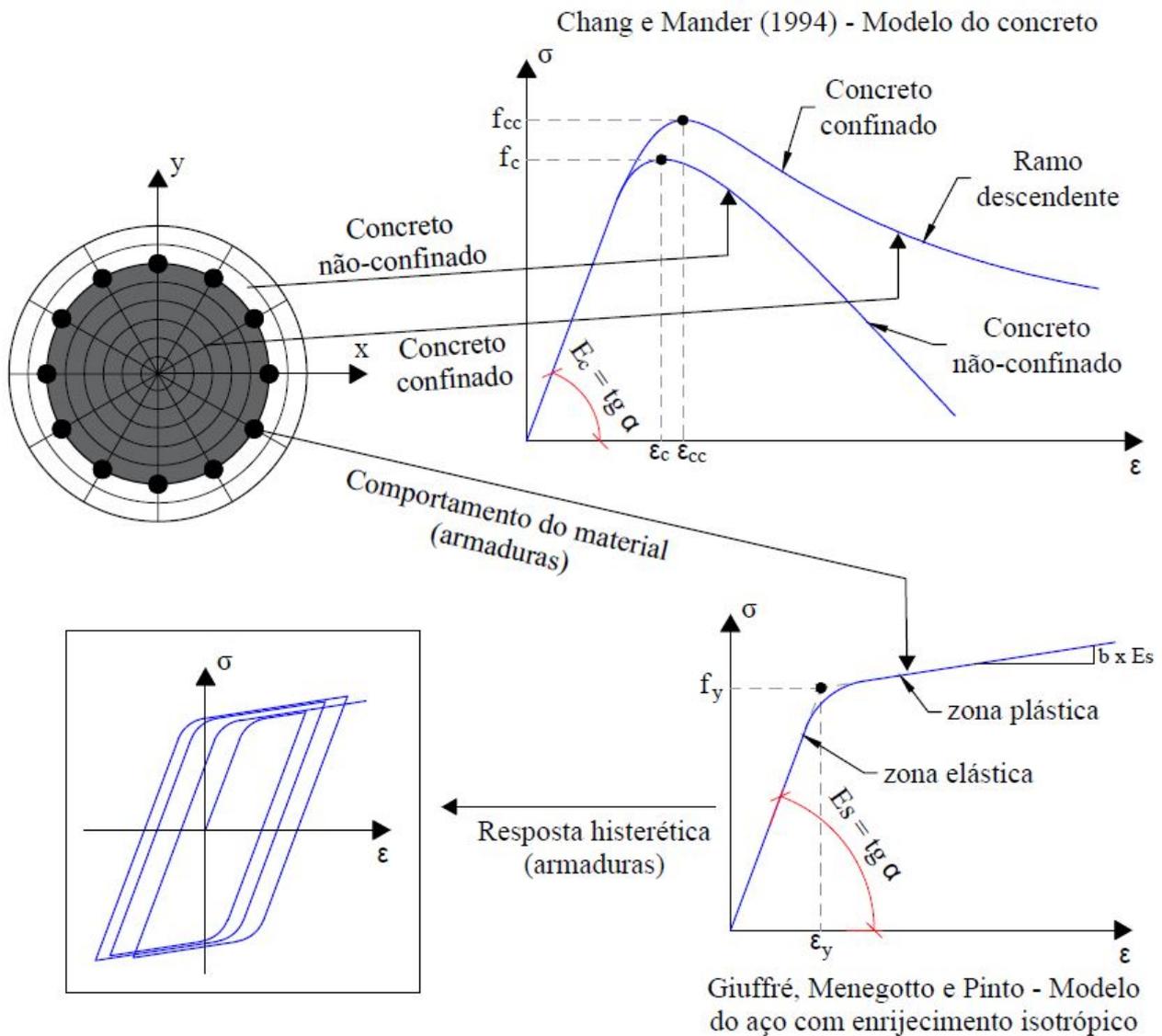


Figura 36 – Modelos constitutivos do concreto e do aço (Fonte: adaptado de Cavalcante (2022)).

O modelo de Chang e Mander (1994) foi elaborado especificamente para pilares de pontes discretizados em fibras com modelos cíclicos dinâmicos. Além do modelo permitir que as respostas históricas sejam extrapoladas para o modelo de demanda sísmica, ele é capaz de simular o fechamento e abertura de fissuras.

O modelo de GMP com enrijecimento isotrópico de 5% (FILIPPOU et al., 1983), apresenta uma mudança suave entre a zona elástica e a zona plástica. Para a calibração do modelo do aço são utilizados os resultados experimentais de Tanaka (1990), conforme exposto no *Pacific Earthquake Research Center* (PEER). Como a definição em detalhes dos modelos fogem ao

escopo deste trabalho, o leitor deve se direcionar aos materiais utilizados aqui como referência para mais informações: Chang e Mander (1994), Filippou et al. (1983) e Tanaka (1990).

A relação entre a estrutura e solo é definida pelo modelo de interação solo-estrutura (ISE). De acordo com Elnashai e Sarno (2008), as interações podem ser divididas em 3 grupos: cinemáticas - representa a transição dos esforços do solo para a estrutura ou da estrutura para o solo; inerciais - representa aumento dos períodos naturais da estrutura devido a sua flexibilização e a dissipação de energia causada pela própria interação; inelásticas - representa a não-linearidade entre o comportamento do solo e da estrutura devido a diferença de rigidez entre os dois elementos.

O modelo de ISE utilizado por Suescun (2010) foi empregado neste trabalho. Este modelo também foi utilizado em outros trabalhos como Siqueira et al. (2014a) e Tavares et al. (2013).

Como para as pontes em estudo (tipologia T3) foram adotadas fundações do tipo profunda, foi utilizado a metodologia apresentada por Novak (1974). Esta metodologia utiliza elementos finitos do tipo *zero-length* ligados em paralelo (ZLE/ZLE) ligados aos pilares e ao nó da base perfeitamente engastado.

As propriedades utilizadas para a modelagem da ISE são: peso específico do solo (ρ_{solo}); coeficiente de Poisson (ν_{solo}); velocidade de onda (V_s); módulo de elasticidade transversal (G_{solo}); número das estacas (n_{est}); comprimento das estacas (L_{est}); diâmetro das estacas (ϕ_{est}); módulo de elasticidade do concreto da estaca ($E_{sc,est}$); distância externa entre os eixos das estacas e a face do bloco nas duas direções ($S_{x,ext}$) e ($S_{y,ext}$); distância interna entre os eixos das estacas nas duas direções ($S_{x,int}$) e ($S_{y,int}$); altura do bloco (h_b).

A Equação 3.3.1 a seguir, apresenta o cálculo do módulo de elasticidade do solo a partir da velocidade de propagação de ondas.

$$G_{solo} = \rho_{solo} \cdot V_s^2 \quad (3.3.1)$$

A Figura 37 a seguir, representa a modelagem utilizada para a ISE no caso de fundação profunda.

O modelo utilizado para a interação solo-estrutura, considera os esforços de vibrações verticais quando as estacas são solicitadas por carregamentos laterais, ou seja, as estacas são modeladas conforme os graus de liberdade vertical, horizontal, rotacional e torsional (NOVAK, 1974). Para mais detalhes da modelagem da ISE utilizada, o leitor deve se direcionar aos materiais utilizados aqui como referências: Novak (1974), Prakash e Sharma (1990) e Novak e Howell (1977) e Suescun (2010).

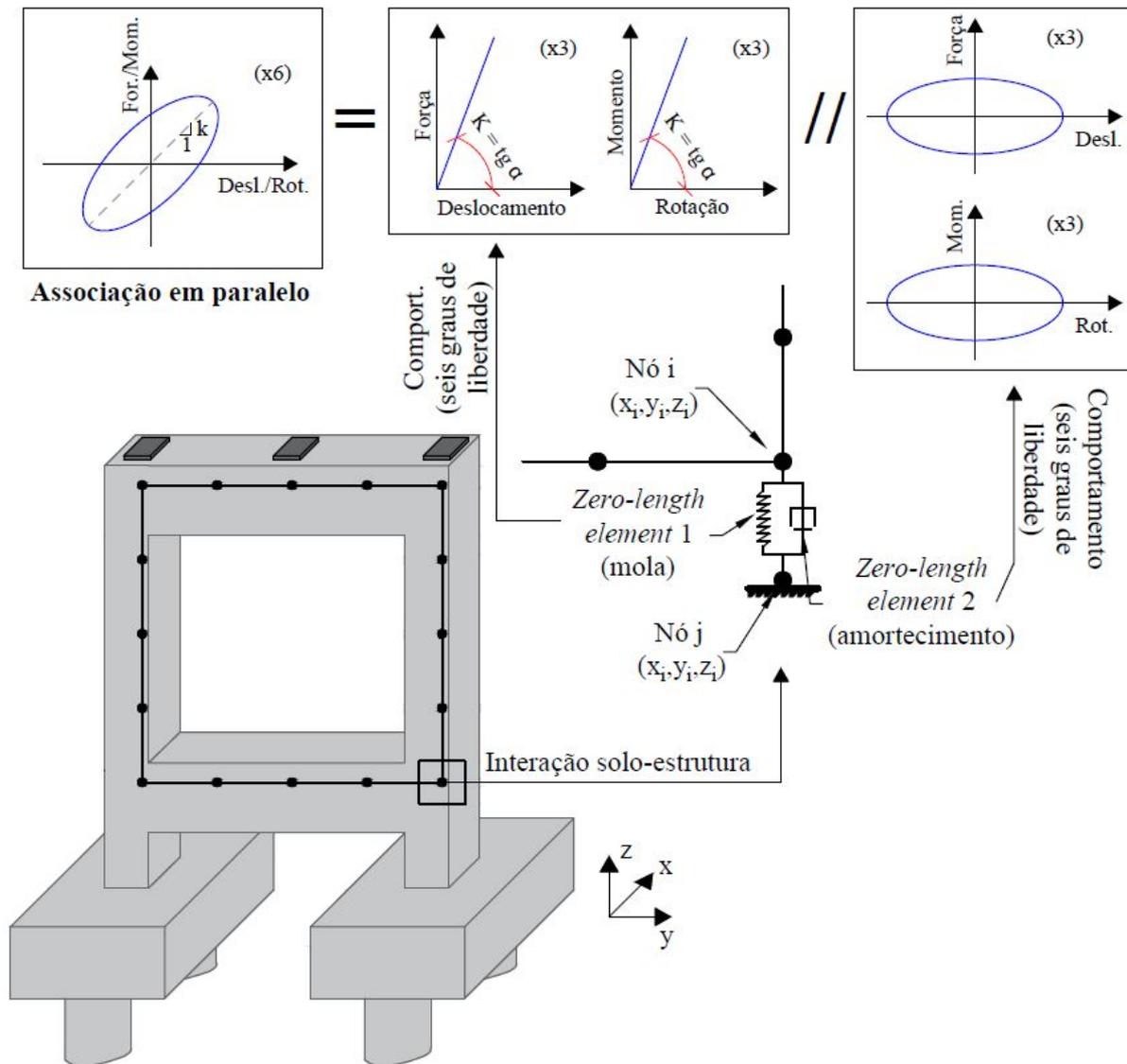


Figura 37 – Modelos utilizado para a ISE para fundação profunda (Fonte: adaptado de Cavalcante (2022)).

Os aparelhos de apoio elastoméricos apresentam comportamento elasto-plástico perfeito nas direções horizontais e são fixos na direção vertical, podendo rotacionar ou torcer livremente. Para isso, os aparelhos de apoio foram modelados por elementos do tipo *zero-length*, conforme pode ser observado em trabalhos como Padgett (2007) e Suescun (2010). A rigidez no deslocamento horizontal é calculada a partir da Equação 3.3.2 a seguir.

$$k = \frac{G \cdot A}{h} \quad (3.3.2)$$

Sendo: k a constante de rigidez dos aparelhos de apoio; G o módulo de elasticidade transversal; A a área em planta; h a altura ou espessura do aparelho de apoio.

Os aparelhos de apoio elastoméricos encontrados na região Nordeste do Brasil são quase que exclusivamente do tipo simples, ou seja, não apresentam nenhum pino de aço fixador ou camadas de chapa de aço. Logo, a força de escoamento F_y é adotada como sendo a força em que ocorre o escorregamento do tabuleiro, sendo calculada como demonstra a Equação 3.3.3 a seguir.

$$F_y = \mu \cdot N \quad (3.3.3)$$

Sendo: F_y a força de escoamento; μ o coeficiente de atrito; N a força normal no aparelho de apoio.

A Tabela 16 a seguir demonstra algumas equações utilizadas para se obter o coeficiente de atrito do aparelho de apoio. Neste trabalho foi adotado a formulação proposta por Schrage (1981).

Tabela 16 – Equações utilizadas para o cálculo do coeficiente de atrito.

Autor	Equação
Schrage (1981)	$\mu = 0,05 + \frac{0,4}{\sigma_m}$
NBR 9062 (2017)	$\mu = 0,1 + \frac{0,6}{\sigma_m}$

A tensão normal exercida no aparelho de apoio, σ_m , é calculada a partir da divisão da força normal pela área do aparelho de apoio, expressa em MPa. A Figura 38 a seguir demonstra o modelo utilizado para os aparelhos de apoio.

Os *gaps*, que são as juntas existentes entre a superestrutura e os encontros ou entre os vãos da própria superestrutura, foram modelados utilizando um modelo de impacto com amortecimento com resposta não-linear (*Hertzdamp model*), conforme proposto por Muthukumar (2003). Este modelo foi adotado para que a dissipação de energia durante o impacto fosse representada, uma vez que o modelo Hertz somente considera valores de rigidez não-lineares para representação da mola e não considera o efeito do impacto. O elemento *two node link*, conforme exposto na Tabela 15, trabalha apenas em compressão.

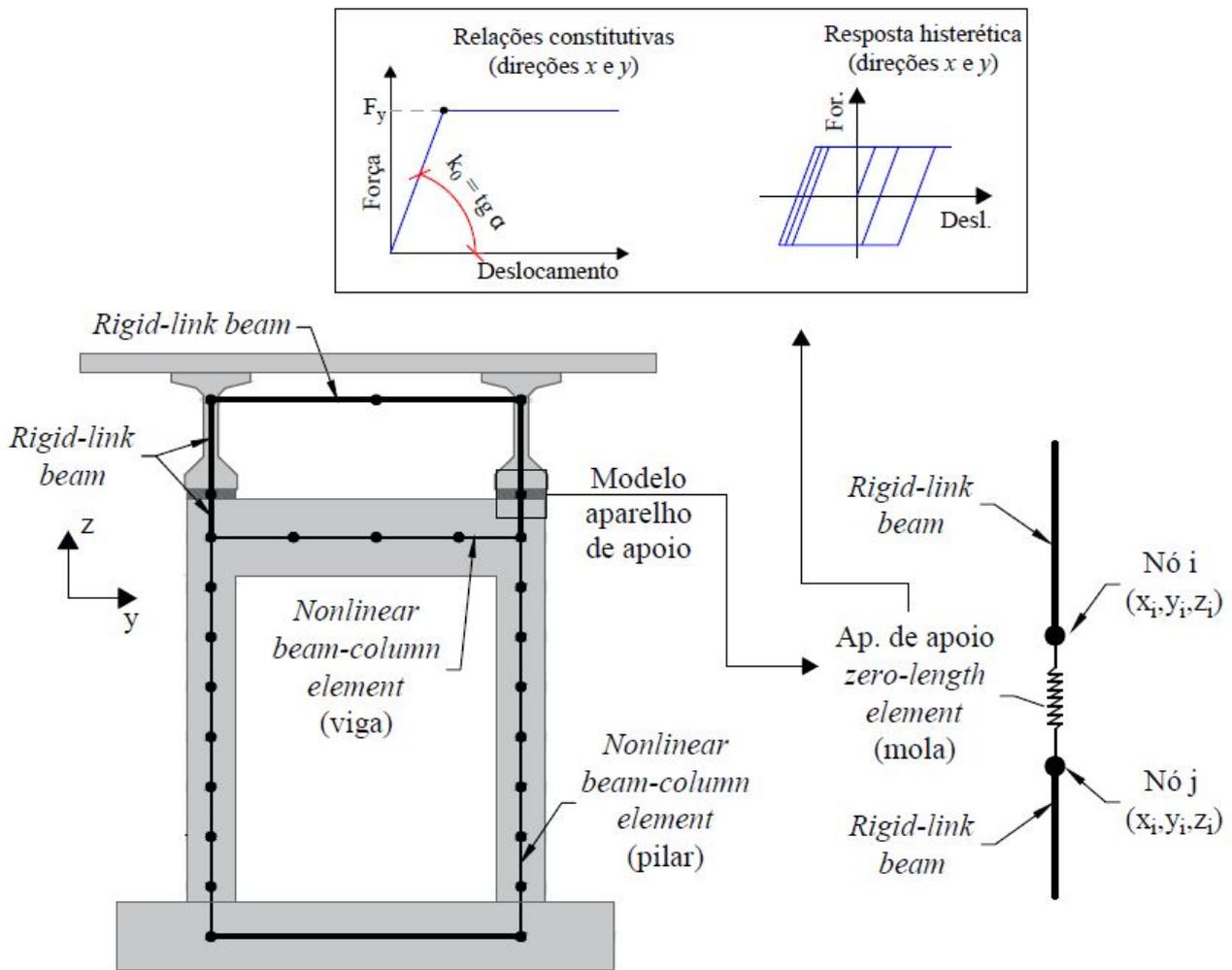


Figura 38 – Modelos utilizado para os aparelhos de apoio (Fonte: adaptado de Cavalcante (2022)).

A Figura 39 a seguir demonstra o modelo utilizado para as juntas (*gaps*) para situações de tabuleiro descontínuo, ou seja, quando há uma junta entre os vãos da estrutura. Para mais detalhes da modelagem das juntas utilizada, o leitor deve se direcionar aos materiais utilizados aqui como referências: Muthukumar (2003) e Muthukumar e DesRoches (2006).

De acordo com o item 3.1 (Tipologia Adotada), este trabalho considerou dois tipos de encontros por não haver prevalência para a classe C1 da tipologia T3. Os encontros são do tipo “U” ou gravidade. A Figura 40 a seguir demonstra com mais detalhe os dois tipos de encontros.

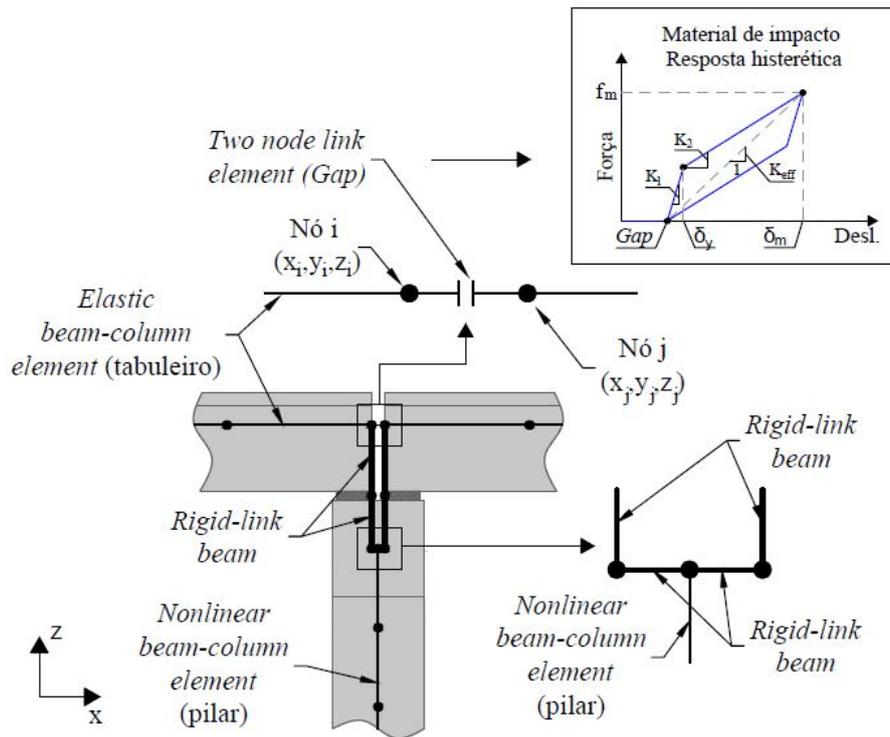


Figura 39 – Modelos utilizado para as juntas (*gaps*) (Fonte: adaptado de Cavalcante (2022)).

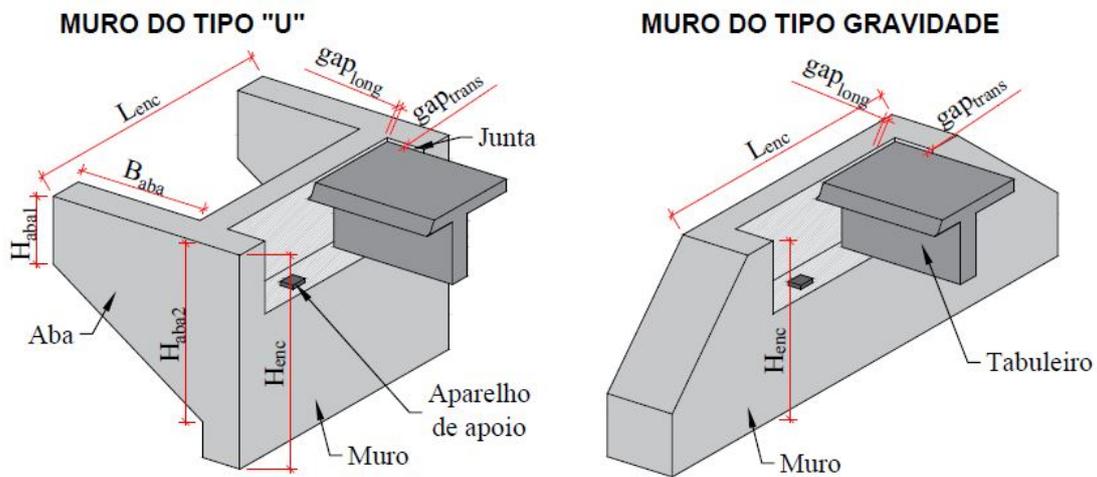


Figura 40 – Tipo de encontros utilizados (Fonte: adaptado de Cavalcante (2022)).

Dependendo da intensidade da atividade sísmica, impactos longitudinais ou transversais podem ocorrer entre os encontros e a superestrutura. Para este trabalho foram consideradas apenas encontros com fundações profundas e por este motivo adotou-se o método de Suescun (2010). Este modelo considera a contribuição das estacas, do aterro, das abas do muro tipo “U” e das juntas (*gaps*) para determinar o comportamento dos encontros. Longitudinalmente, as juntas trabalham em série com o aterro, enquanto as estacas trabalham em paralelo.

A Figura 41 a seguir demonstra o modelo utilizado para os encontros no sentido longitudinal.

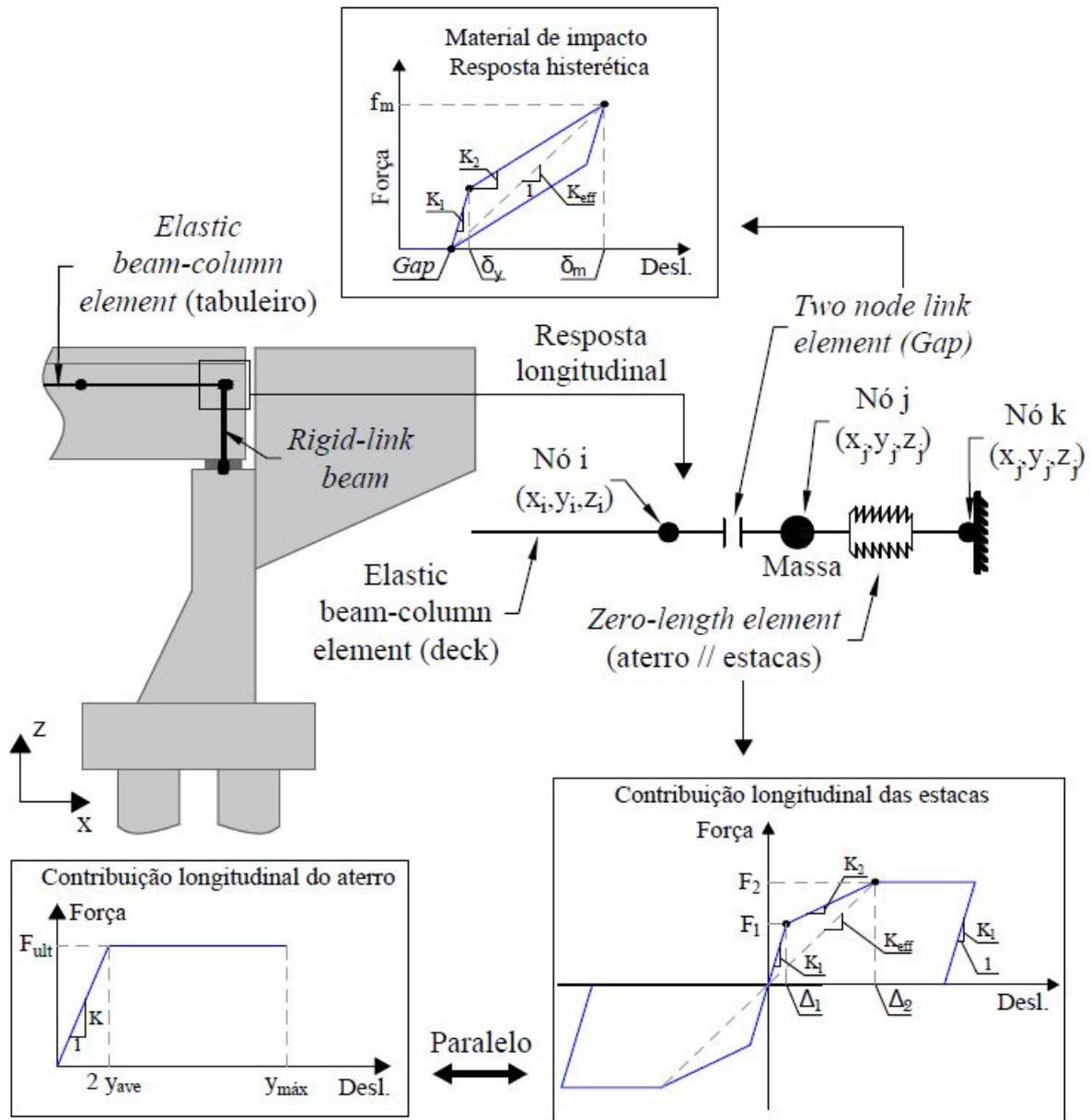


Figura 41 – Modelo utilizado para os encontros no sentido longitudinal (Fonte: adaptado de Cavalcante (2022)).

O comportamento do aterro é descrito como elastoplástico perfeito conforme proposto por Maroney (1994). Já as estacas são modeladas de acordo com o modelo proposto por Choi (2002), que descreve a contribuição das estacas nos empuxos ativos e passivos através de um modelo trilinear. Este modelo consiste em assumir que as estacas começam a plastificar a partir de um deslocamento de 2,54 cm e o que escoamento se inicia quando a força exercida na estaca

ultrapassa 30% do deslocamento último de 5,5 cm (SUESCUN, 2010). Transversalmente, as estacas trabalham em paralelo com as abas e o aterro e as juntas em série com o conjunto. A Figura 42 a seguir demonstra o modelo utilizado para os encontros no sentido transversal.

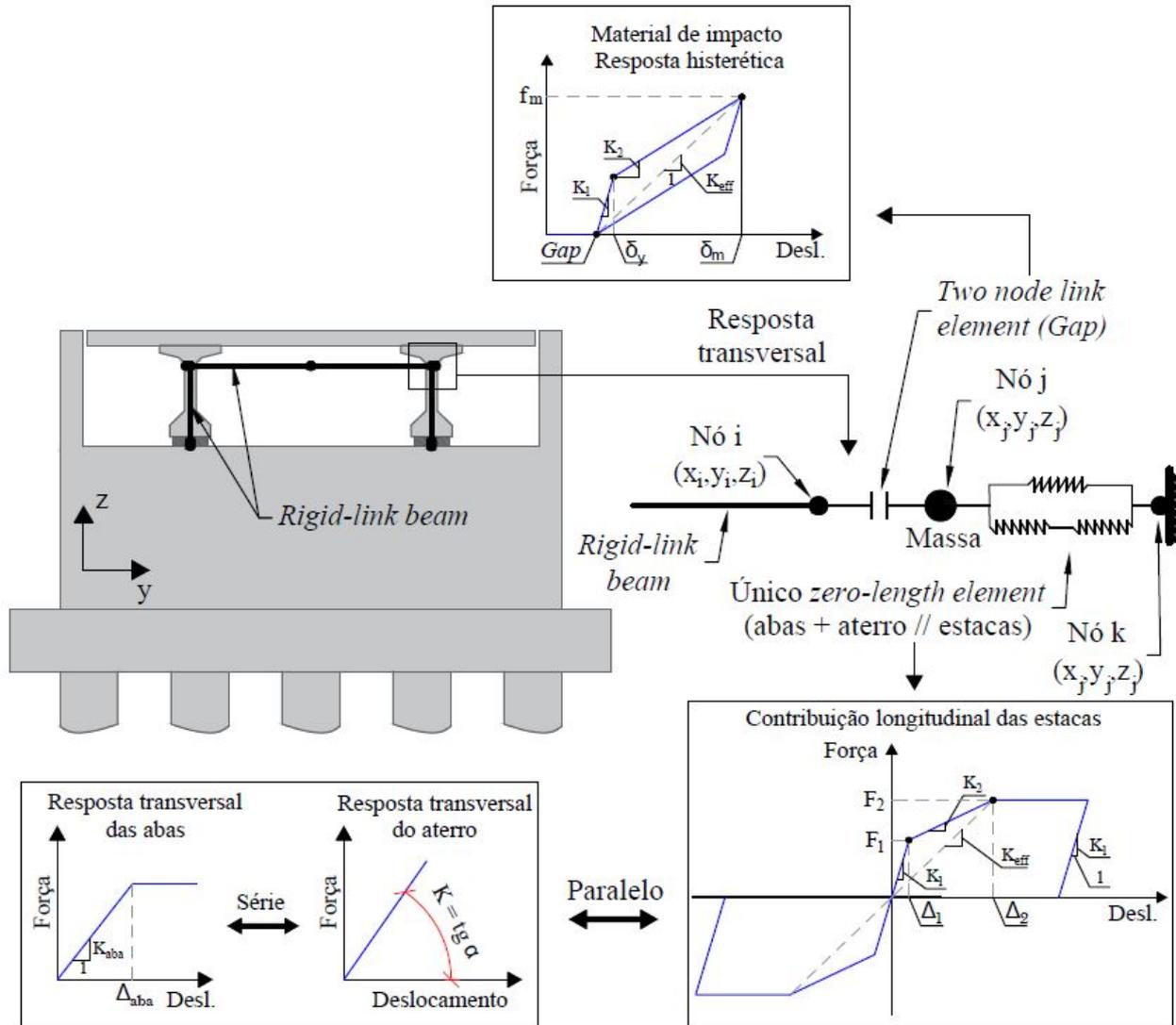


Figura 42 – Modelo utilizado para os encontros no sentido transversal (Fonte: adaptado de Cavalcante (2022)).

A rigidez transversal devida ao aterro é obtida através do modelo de Wilson e Tan (1990) e a rigidez da aba no sentido transversal é obtida a partir do modelo da Caltrans (1990). Na direção vertical, os aparelhos de apoio que apresentam comportamento elastoplástico perfeito, estão ligados aos tabuleiros e são conectados aos encontros através de dois elementos tipo *zero-length* que representam a ISE.

A Figura 43 demonstra o modelo utilizado para os encontros no sentido vertical.

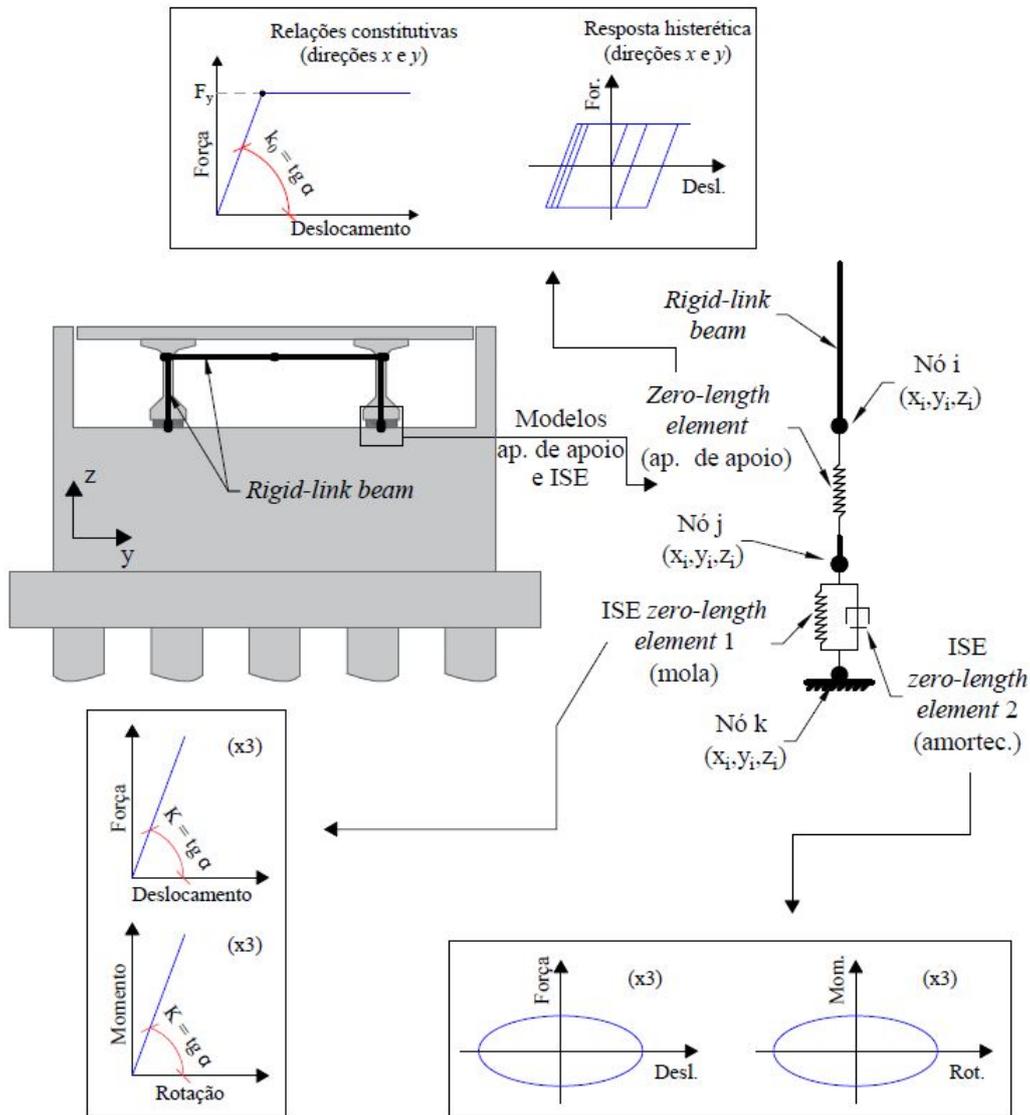


Figura 43 – Modelo utilizado para os encontros no sentido vertical (Fonte: adaptado de Cavalcante (2022)).

Para mais detalhes da modelagem dos encontros utilizadas, o leitor deve se direcionar aos materiais utilizados aqui como referências: Suescun (2010), Maroney (1994), Choi (2002), Wilson e Tan (1990) e Caltrans (1990).

3.4 Análises dinâmicas não lineares

A partir do modelo da ponte da classe C1 e tipologia T3 e dos sinais de terremotos adotados é possível conduzir as análises dinâmicas não lineares. Neste trabalho foi utilizado o método de *Cloud Analysis* que consiste em uma metodologia de análise dinâmica não incremental, onde os sinais são empregados sem qualquer tipo de escalonamento da medida de intensidade

(IM). Este método, quando aplicado com uma grande quantidade de registros sísmicos, permite a determinação de vários valores de demanda onde cada resposta pode ser associada a um único registro. Desta forma, o método de *Cloud Analysis* é utilizado para determinar a demanda em função de uma IM, ou seja, obter um valor da resposta estrutural para cada registro sísmico.

Para este trabalho foram utilizados 100 pares de acelerações sísmicas, cem no sentido longitudinal X e cem no sentido transversal Y , combinadas a partir da média geométrica, e 100 modelos de OAEs da tipologia escolhida, resultando assim em 10.000 modelos analisados, ou seja 10.000 pares 'ponte x terremoto'. Este alto número de análises segue o padrão encontrado nos trabalhos com foco na aplicação de *Machine Learning*, realizados por Mangalathu et al. (2019), Mangalathu e Jeon (2019) e Xie et al. (2020). A medida de intensidade utilizada foi a aceleração espectral S_a de período de 1 segundo (período longo) com taxa de amortecimento de 5%.

3.5 Modelos de demanda e capacidade

Posteriormente a realização das análises dinâmicas não lineares para a obtenção dos valores de demanda em função de uma IM, aplica-se o modelo de regressão referenciado neste trabalho como PSDM (Modelo Probabilístico de Demanda Sísmica). Os resultados estudados para a criação das curvas de fragilidade são: curvatura dos pilares, deslocamento longitudinal e transversal dos aparelhos de apoio, deslocamento do tabuleiro e deslocamento ativo, passivo e transversal dos encontros. Vale ressaltar que apenas os valores máximos são utilizados, de cada componente de cada modelo estrutural, para a análise de regressão. Desta forma, para cada modelo analisado são criados os Modelos Probabilísticos de Demanda Sísmica e a partir destes modelos são calculados os parâmetros θ e β necessários para a construção das curvas de fragilidade.

Conforme exposto no item 2.4.2 deste trabalho, os valores de capacidade (mediana S_c e dispersão β_c) foram retirados da bibliografia para o deslocamento longitudinal e transversal dos aparelhos de apoio, deslocamento do tabuleiro, deslocamento ativo, passivo e transversal dos encontros (MANGALATHU et al., 2017). Estes valores estão expostos na Tabela 6.

Já para os valores de capacidade dos pilares, foram utilizados os resultados obtidos por Cavalcante (2022). Estes resultados foram obtidos através de uma análise *pushover* que consiste em submeter a estrutura do pilar a uma carga estática não linear, através da aplicação de cargas gravitacionais constantes e incrementos de cargas horizontais aplicadas de maneira progressiva. Esta análise busca estabelecer o mecanismo plástico de resposta da estrutura (escoamento das armaduras) e a distribuição do dano que é obtida através da curva de capacidade que representa

uma função entre o deslocamento no topo do pilar e o esforço cortante na base ou momento fletor total. Portanto, são determinados os valores de curvatura da seção transversal a partir dos valores de demanda de ductilidade, conforme exposto por Choi et al. (2004). A partir desta análise, determinam-se os valores dos parâmetros S_c e β_c . Estes valores estão expostos na tabela 7. A Figura 44 demonstra como são obtidos os valores dos parâmetros para os limites de dano leve (DL), moderado (DM), extensivo (DE) e completo (DC).

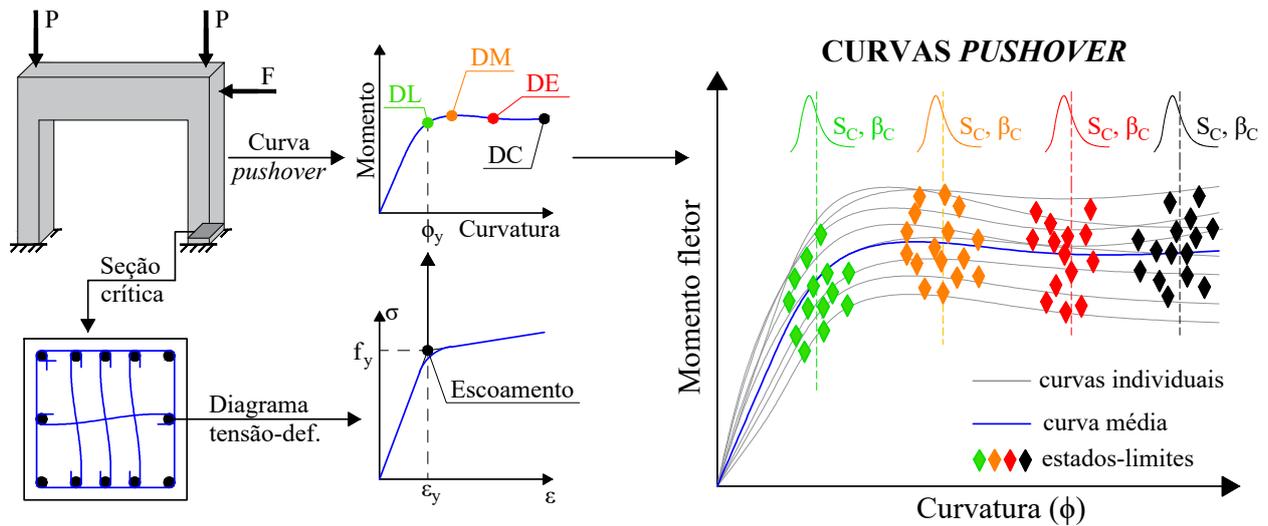


Figura 44 – Método de obtenção dos parâmetros S_c e β_c (Fonte: adaptado de Cavalcante (2022)).

Considerando que as atividades sísmicas são aplicadas nas duas direções, longitudinal (x) e transversal (y), foi utilizada metodologia proposta por Garcia et al. (2019), que consiste em aplicar os deslocamentos no topo dos pilares em diversos ângulos (*Pushover multidirecional*). Para a obtenção dos parâmetros S_c e β_c (CAVALCANTE, 2022) algumas etapas são necessárias, conforme exposto a seguir.

- *1ª etapa*: criação dos modelos numéricos e análise das forças gravitacionais;
- *2ª etapa*: substituição dos modelos globais por modelos construídos por apenas pórticos para realização de análises estáticas com as forças verticais equivalentes do modelo;
- *3ª etapa*: aplicação das forças horizontais variando o ângulo em torno do eixo vertical (z), para a obtenção dos pontos de curvatura ϕ_y onde ocorre o escoamento da armadura longitudinal;
- *4ª etapa*: criação dos gráficos de curvatura em função do ângulo de aplicação das forças horizontais;
- *5ª etapa*: aplicação dos valores de demanda de ductilidade para a determinação dos parâmetros S_c e β_c .

A Figura 45 a seguir demonstra as quatro primeiras etapas citadas acima.

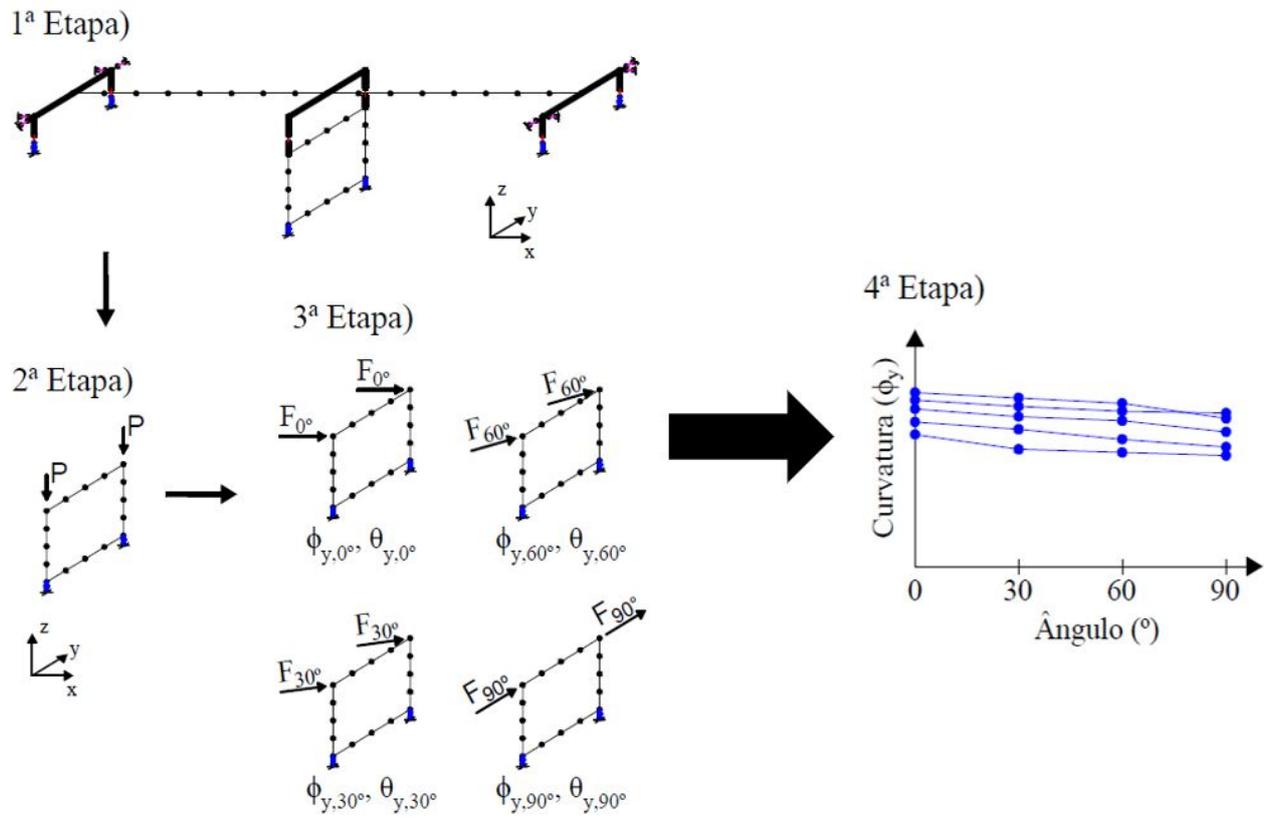


Figura 45 – Etapas para obtenção da curvatura ϕ_y (*Pushover multidirecional*) (Fonte: adaptado de Cavalcante (2022)).

3.6 Curvas de fragilidade

As curvas de fragilidade, que medem a probabilidade condicional da estrutura ou de uma componente estrutural atingir ou ultrapassar um estado limite de dano em função de uma medida de intensidade IM, são obtidas através da resposta estrutural a um evento sísmico (demanda estrutural) e da capacidade. Conforme exposto acima, os parâmetros utilizados para a capacidade estrutural foram adotadas conforme exposto por Mangalathu et al. (2017) e Cavalcante (2022), já os parâmetros de demanda são obtidos a partir da aplicação do modelo de regressão PSDM dos valores calculados através das análises dinâmicas não-lineares.

3.7 Machine Learning

Uma vez obtidos os resultados de demanda estrutural, é possível criar um banco de dados que consiste nas variáveis estocásticas referentes a tipologia T3 e as referidas respostas dinâmicas

para cada parâmetro de demanda de engenharia (EDP), totalizando 10.000 análises (100 OAES solicitadas por 100 acelerações sísmicas) para cada EDP. De acordo com Mangalathu e Jeon (2019), uma vez estabelecido o banco de dados, é possível utilizar as ferramentas de *Machine Learning*. Conforme exposto no capítulo 2, o ML pode ser dividido em duas partes, aprendizagem supervisionada, foco desta pesquisa, e não supervisionada. A aprendizagem supervisionada pode ainda ser dividida em duas frentes, classificação caso os dados trabalhados sejam discretos e regressão caso os dados trabalhados sejam contínuos (ver Figura 13).

Para aplicar as ferramentas citadas acima, o banco de dados teve de ser trabalhado uma vez que nem todos os valores de demanda sísmica (resposta estrutural) para um determinado Parâmetro de Demanda de Engenharia (EDP) representavam valores reais. Ou seja, nem todos os valores de resposta que convergiram durante as análises foram utilizados. Para encontrar e “limpar” esses valores dois métodos foram estudados, *Standard Deviation Method (Z-score)* e *Interquartile Range Method (IQR)*. Devido a natureza dos dados de demanda e a especificidade dos dois métodos, o método IQR foi adotado neste trabalho. De acordo com Hastie et al. (2001), o método IQR é mais apropriado para distribuições não-gaussianas das amostras.

O *Interquartile Range Method* consiste em encontrar os valores que possam estar acima ou abaixo do limite superior e inferior, respectivamente, e substituir pela mediana da amostragem. De maneira resumida, o método IQR pode ser aplicado da seguinte forma:

- *Cálculo da mediana*: para amostras em números pares, a mediana é média dos valores centrais. Para amostras em números ímpares, a mediana é o próprio valor central;
- *Cálculo do 25º valor*: da mesma forma que foi calculada a mediana, o valor que representa o vigésimo quinto número é a média entre o 25º e 26º valor para amostras em números pares e o próprio 25º para amostras ímpares;
- *Cálculo do 75º valor*: da mesma forma que foi calculada a mediana, o valor que representa o septuagésimo quinto número é a média entre o 75º e 76º valor para amostras em números pares e o próprio 75º para amostras ímpares;
- *Cálculo do IQR*: o valor do IQR é obtido pela subtração do 75º pelo 25º valor, ou seja $IQR = 75^{th} - 25^{th}$;
- *Cálculo do valor base de corte ou cut-off point*: de acordo com Hastie et al. (2001), o valor base de corte é normalmente calculado por uma vez e meia o valor do IQR, ou seja $cut - off = 1,5 \times IQR$;

- *Definição do limite superior e inferior*: os limites são definidos a partir do acréscimo ou decréscimo do valor base de corte ao 75º e 25º valor, respectivamente.

A Figura 46 a seguir demonstra o *Interquartile Range Method*.

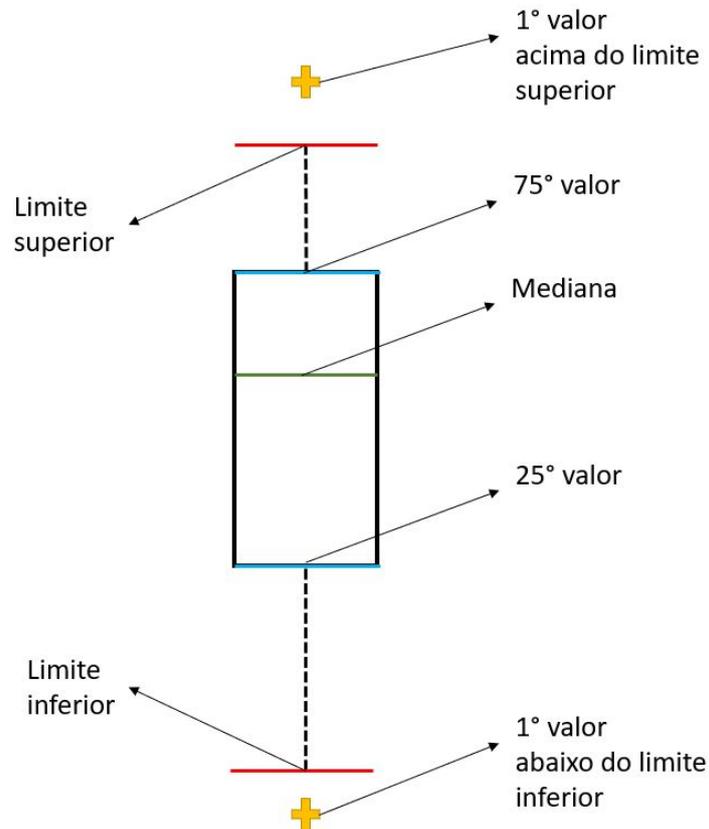


Figura 46 – *Interquartile Range Method* (Fonte: autor).

Como os valores analisados representam a demanda sísmica da estrutura quando solicitada a um terremoto, os valores que estão abaixo do limite inferior, exceto os menores que zero (negativos), não foram descartados e reajustados pela mediana, pois mesmo que inferiores ao limite esses resultados ainda são valores representativos como resposta estrutural frente a uma solicitação sísmica. Os valores negativos foram descartados pois implicam que o modelo não convergiu matematicamente.

Após os valores de demanda sísmica terem sido ajustados através do método IQR, as análises utilizando as ferramentas do *Machine Learning* foram realizadas. As variáveis estocásticas consideradas neste trabalho estão expostas nas Tabelas 12 e 13. Adicionalmente ao que está exposto nas tabelas citadas acima, a medida de intensidade (IM) também foi utilizada como variável aleatória.

3.7.1 Regressão

O modelo de regressão tem por objetivo treinar e aprender uma função que melhor descreva ou se aproxime da relação entre *input* e *output*. No caso do modelo de regressão os dados de saída são representados por valores de caráter contínuo, ou seja, os respectivos valores de demanda sísmica calculados para cada modelo de ponte x IM x parâmetro de demanda de engenharia. Levando em consideração os 7 tipos de EDP (ver tabelas 6 e 7) utilizados neste trabalho, um total de 70.000 valores foram processados nesta etapa do trabalho. Desta forma, dividiu-se o banco de dados em sete partes, uma para cada EDP, para aplicar a metodologia de regressão com o intuito de estabelecer uma função que correlacione os dados de entrada com uma resposta estrutural para cada parâmetro de demanda de engenharia. Cada um dos sub-banco de dados foi estudado de maneira separada, uma vez que a resposta sísmica para um EDP específico, por exemplo deslocamento longitudinal dos aparelhos de apoio, não deve ser utilizado como variável para outro EDP, por exemplo a curvatura dos pilares (MANGALATHU; JEON, 2019).

Com cada sub-banco separado por EDP, os dados são separados, de maneira randômica, em 70% para o *training set* e 30% para o *testing set*. Como exposto anteriormente, de acordo com Hastie et al. (2001), devido ao fato do banco de dados, ou sub-banco de dados, ser robusto o bastante, a avaliação da precisão do processo preditivo para o conjunto de treinamento deve ser realizada através do conjunto de validação (em inglês *validation set*). O conjunto de validação utilizado neste trabalho, seguindo trabalhos realizados no mesmo tema como Mangalathu e Jeon (2019) e Mangalathu et al. (2019), foi o *K-fold Cross Validation* com $K=10$. O *training set* então é utilizado para estudar qual o melhor método de regressão que estabelece a função entre *input* e *output*. Para mais informação sobre os métodos de regressão, o leitor é direcionado ao Capítulo 2 deste trabalho.

Após todos os métodos terem sido testados, é escolhido o melhor entre eles para verificar a sua performance com dados ainda não “vistos” pelo algoritmo, que compõem o *testing set*, para evitar erros devido a criação de uma tendência. A verificação da performance dos métodos testados é realizada de maneira gráfica e numérica. Os gráficos utilizados são o *Response Plot*, que demonstra em um mesmo plano todos os valores reais e os previstos pelo modelo, o *Predicted vs. Actual Response*, que demonstra a correlação entre os valores reais e previstos e por último o *Residual Plot*, que verifica os resíduos das observações. De maneira lógica, os pares de valores plotados no gráfico de resposta devem estar o mais próximos possível, para que desta forma, os valores plotados no gráfico de previstos *vs.* reais estejam sob a reta diagonal de 45 graus que demonstra a previsão perfeita. Vale ressaltar que o gráfico de resíduos não pode representar nenhum tipo de padrão, ou seja, os dados devem estar distribuídos de maneira randômica.

A verificação numérica é feita através de quatro parâmetros, sendo eles o *Root Mean Square Error* (RMSE), *Mean Square Error* (MSE), *Mean Absolute Error* (MAE) e o coeficiente de determinação R^2 . Para mais informação sobre os parâmetros de verificação numérica, o leitor é direcionado ao Capítulo 2 deste trabalho. Vale ressaltar que, exceto o R^2 , os parâmetros são sensíveis a magnitude da unidade de medida do *output*.

Depois de testado, o algoritmo com melhor performance é validado com os dados do conjunto de teste para verificar a sua performance com dados não utilizados no processo treinamento. Este passo é fundamental para investigar se o modelo escolhido produz erros de *overfitting*.

Este erro ocorre quando o algoritmo performa de maneira satisfatória durante o treinamento, mas sua performance é baixa quando testada com novos intervalos de dados. O *overfitting* está relacionado com a capacidade de o modelo conseguir diferenciar entre dados relevantes e não relevantes durante o processo de predição. Após verificado, o algoritmo que melhor performa para ambos, *training set* e *testing set*, é escolhido como modelo preditivo final.

3.7.2 Classificação

Diferente do modelo de regressão, o modelo de classificação do *machine learning* é utilizado para treinar um algoritmo que melhor descreve a relação entre os dados de entrada e saída, que possuam caráter discreto, ou seja, um *output* qualitativo. Por se tratar de uma análise qualitativa, o banco de dados é classificado, ou dividido em classes, de acordo com Hastie et al. (2001). Esta divisão em grupos leva em conta os valores de demanda sísmica e os valores de capacidade estrutural expostos no Capítulo 2.

Para este trabalho, foram utilizadas 3 classes, ou *TAGs*, distintos para classificar as respostas estruturais frente a um evento sísmico para cada parâmetro de demanda de engenharia. A Tabela 17 a seguir demonstra os valores limites dos *TAGs* adotados para este trabalho.

Tendo em vista que os valores de demanda sísmica para os parâmetros curvatura dos pilares e deslocamento transversal dos encontros não excederam os valores mínimos, estes EDPs não foram considerados para esta etapa do trabalho uma vez que todos os *TAGs* seriam *Green*.

Para os modelos de classificação, considerando que o algoritmo irá categorizar as respostas em apenas três classes, foram consideradas as respostas sísmicas para os pares “modelo de ponte x aceleração sísmica”, e sua respectiva categoria, para cada parâmetro de demanda de engenharia. Considerando apenas 5 EDPs, exclui-se aqui a curvatura dos pilares e o deslocamento transversal dos encontros, conforme mencionado acima, um total de 500 análises foram processadas, conforme exposto por Mangalathu et al. (2019).

Tabela 17 – Valores adotados para divisão dos grupos (*TAGs*) (Fonte: autor).

Parâmetro de Engenharia	<i>TAG</i>	Valores correspondentes
Desloc. long. do aparelho de apoio (mm)	<i>Green</i>	< 25
	<i>Yellow</i>	≥ 25 e < 76
	<i>Red</i>	≥ 76
Desloc. transv. do aparelho de apoio (mm)	<i>Green</i>	< 25
	<i>Yellow</i>	≥ 25 e < 76
	<i>Red</i>	≥ 76
Deslocamento do tabuleiro (mm)	<i>Green</i>	< 25
	<i>Yellow</i>	≥ 25 e < 76
	<i>Red</i>	≥ 76
Desloc. ativo dos encontros (mm)	<i>Green</i>	< 38
	<i>Yellow</i>	≥ 38 e < 102
	<i>Red</i>	≥ 102
Desloc. passivo dos encontros (mm)	<i>Green</i>	< 76
	<i>Yellow</i>	≥ 76 e < 254
	<i>Red</i>	≥ 254
Desloc. transv. dos encontros (mm)	<i>Green</i>	< 25
	<i>Yellow</i>	≥ 25 e < 102
	<i>Red</i>	≥ 102
Curvatura dos pilares (1/m)	<i>Green</i>	$< 0,007$
	<i>Yellow</i>	$\geq 0,007$ e $< 0,013$
	<i>Red</i>	$\geq 0,013$

De maneira similar ao método da regressão, cada um dos sub-banco de dados (um para cada EDP) foi estudado de maneira separada, uma vez que a resposta sísmica para um EDP específico não foi utilizada como variável para outro EDP.

Analogamente ao que foi descrito no item anterior, cada sub-banco de dados foi dividido de maneira aleatória em *training set* e *testing set*. O conjunto de validação utilizado para a o ML de classificação também foi o *K-fold Cross Validation* com $K=10$. Desta forma, o conjunto de treinamento é utilizado para estudar qual método melhor classifica as informações de *input* nos *TAGs* de *output*.

Posteriormente ao teste de todos os modelos de classificação, o modelo com a melhor performance para categorizar é escolhido. Similarmente ao método da regressão, o modelo de classificação escolhido é testado com informações não utilizadas no processo de treinamento, ou seja, o *testing set*, para sua validação final. A verificação da performance dos métodos é realizada de maneira gráfica e numérica. Ao contrário do método da regressão, a análise gráfica do método de classificação é mais importante do que a numérica, uma vez que a avaliação

numérica (em inglês *Prediction Accuracy*) pode esconder erros indesejáveis, por exemplo do tipo **Falso Positivo (FP)** ou **Falso Negativo (FN)**, conforme exposto no capítulo 2. Ou seja, o valor da acurácia total do modelo pode ser alto, por exemplo 90%, no entanto o modelo pode cometer erros de classificação como categorizar um resultado como *Green* quando na verdade a classe real é *Red*. Como este tipo de erro deve ser evitado, a análise gráfica através da *confusion matrix* é a melhor maneira de visualizar o resultado do algoritmo.

Para este trabalho, a verificação da performance dos modelos de classificação foi feita através da matriz de confusão e do valor da acurácia total do algoritmo, excluindo-se aqui verificações gráficas através do *scatter plot*, *parallel coordinates plot* e *receiver operating characteristics curve*. Similar ao método de regressão, o algoritmo escolhido é investigado para erros do tipo *overfitting* ao ser testado com os dados do conjunto de teste. Após verificado, o algoritmo com a melhor performance para ambos os conjuntos, treino e teste, é escolhido como modelo preditivo final.

3.8 Síntese do capítulo

Neste capítulo foi detalhado todo o processo de obtenção das informações necessárias para a geração do banco de dados, como quais parâmetros foram considerados fixos e quais foram considerados variáveis, e o processo de determinação dos sinais de aceleração utilizados. Na sequência, foi discutida a elaboração dos modelos numéricos utilizados para representar a ponte da tipologia escolhida.

Em seguida, foi discutido o processo utilizado para a realização das análises dinâmicas não lineares para, posteriormente, aplicar os modelos de demanda e capacidade com o objetivo de obter as informações necessárias para a construção das curvas de fragilidade. Por fim, foi detalhada toda a metodologia para a aplicação das ferramentas *Machine Learning*, como o modelo de regressão e classificação.

4 Resultados

Neste capítulo os resultados obtidos para a realização deste trabalho de mestrado serão expostos. Os resultados apresentados são referentes a tipologia T3 classe C1, conforme exposto ao longo deste trabalho. Esta tipologia possui múltiplos vãos, encontros e pórticos internos rotulados. A seção transversal é do tipo “T” e os pórticos internos são apoiados por 2 pilares circulares.

Os resultados foram divididos entre resultado estrutural, modelo determinístico e estocástico, e resultado do *machine learning*. Ademais, será exposto neste capítulo uma comparação entre a análise estrutural estocástica e a análise de inteligência artificial.

4.1 Resultado estrutural

Esta seção tem como objetivo apresentar os resultados das análises numéricas, como as análises modais e as análises dinâmicas não-lineares, e os resultados gerados pelos modelos estocásticos que foram analisados.

4.1.1 Resultado determinístico

A análise modal é uma etapa inicial para as análises dinâmicas não-lineares. Como para a tipologia T3 tanto o número de vãos quanto os respectivos comprimentos são variáveis aleatórias, será exposto, a seguir, a análise modal de um dos 100 modelos gerados. Esse modelo foi previamente testado e avaliado quanto ao comportamento dinâmico da estrutura e de cada componente como exposto por Cavalcante (2022). Para esta avaliação, o modelo escolhido foi analisado com os registros de aceleração do terremoto El Centro, aplicado na direção longitudinal e transversal, escalonado para que ocorra a plastificação das componentes estruturais ou o colapso global da estrutura.

A ponte avaliada apresenta as seguintes características (CAVALCANTE, 2022): número de vãos contínuos: 2; comprimento dos vãos: 14,9 metros; altura das longarinas: 1,4 metros; distância entre os pilares: 4,8 metros; diâmetro dos pilares: 80 centímetros; altura dos pilares: 4,8 metros; tipologia dos encontros: tipo *U*; altura dos encontros: 7,1 metros.

A Figura 47 ilustra o modelo de ponte adotado para esta avaliação.

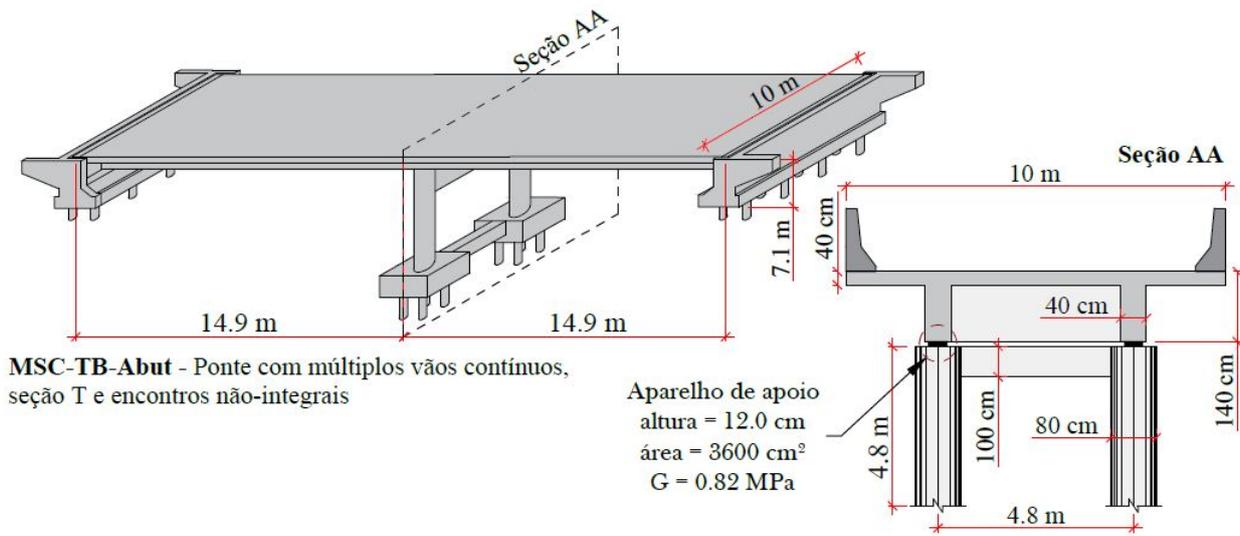


Figura 47 – Propriedades geométricas da ponte de tipologia T3 classe C1 (Fonte: adaptado de Cavalcante (2022)).

A Figura 48 a seguir ilustra os primeiros quatro modos de vibração da ponte adotada. Nota-se que os primeiros dois modos de vibração são no sentido longitudinal e transversal, respectivamente, com participação modal de 97% (CAVALCANTE, 2022). O terceiro modo de vibração é torsional, enquanto o quarto é translacional, porém ambos possuem pouca representatividade quanto sua participação modal. Além do mais, de acordo com Cavalcante (2022), a massa vertical é majoritariamente mobilizada no quinto modo de vibração.

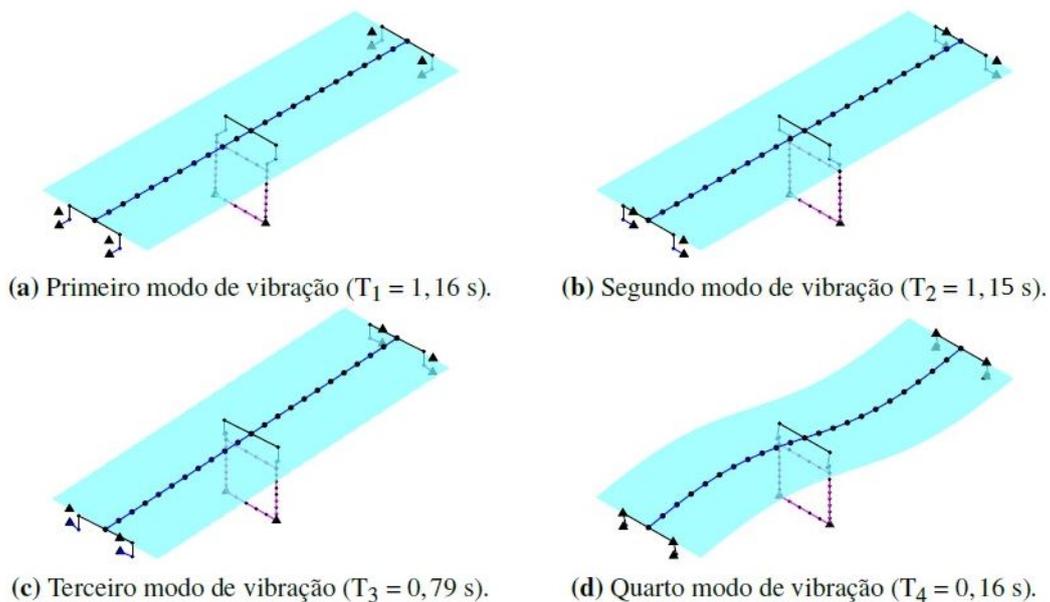


Figura 48 – Modos de vibração da ponte de tipologia T3 classe C1 (Fonte: adaptado de Cavalcante (2022)).

As respostas dinâmicas levantadas por Cavalcante (2022) para a ponte apresentada quanto solicitada ao El Centro foram ilustradas nas Figuras 49 à 52.

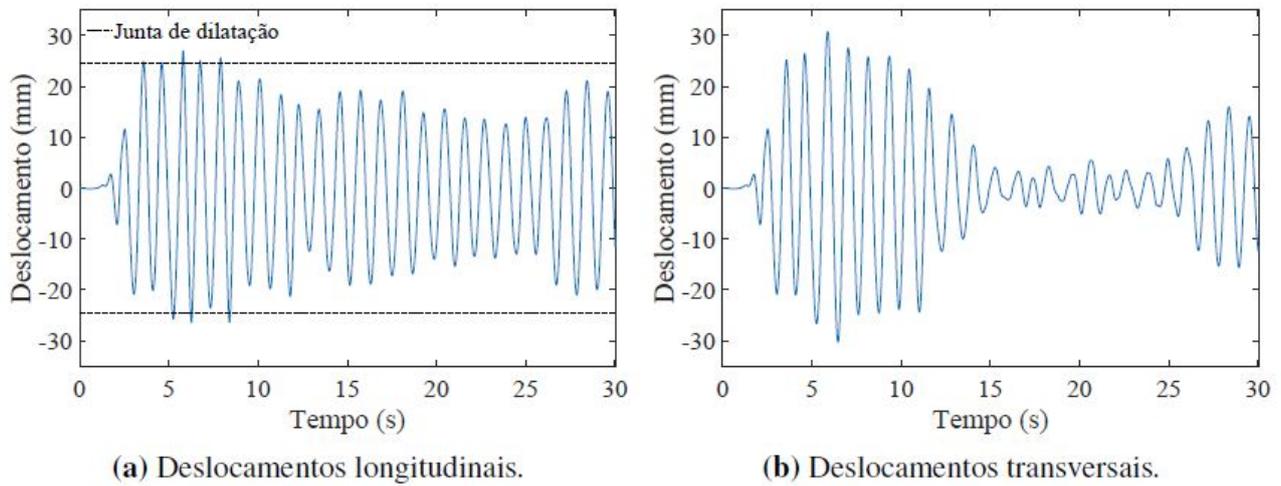


Figura 49 – Deslocamento longitudinal e transversal do tabuleiro ao longo do tempo (Fonte: adaptado de Cavalcante (2022)).

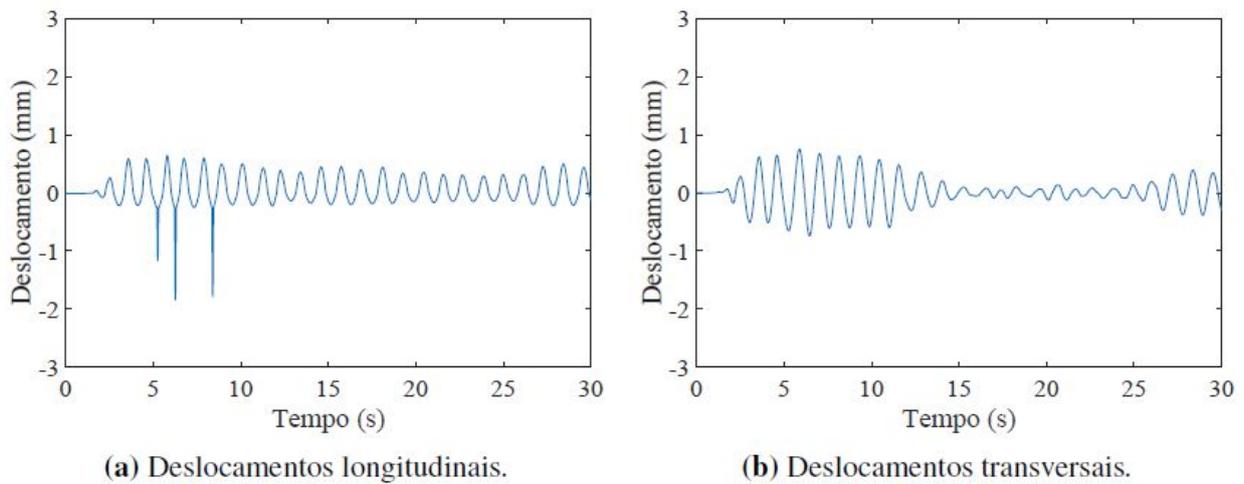


Figura 50 – Deslocamento longitudinal e transversal dos encontros ao longo do tempo (Fonte: adaptado de Cavalcante (2022)).

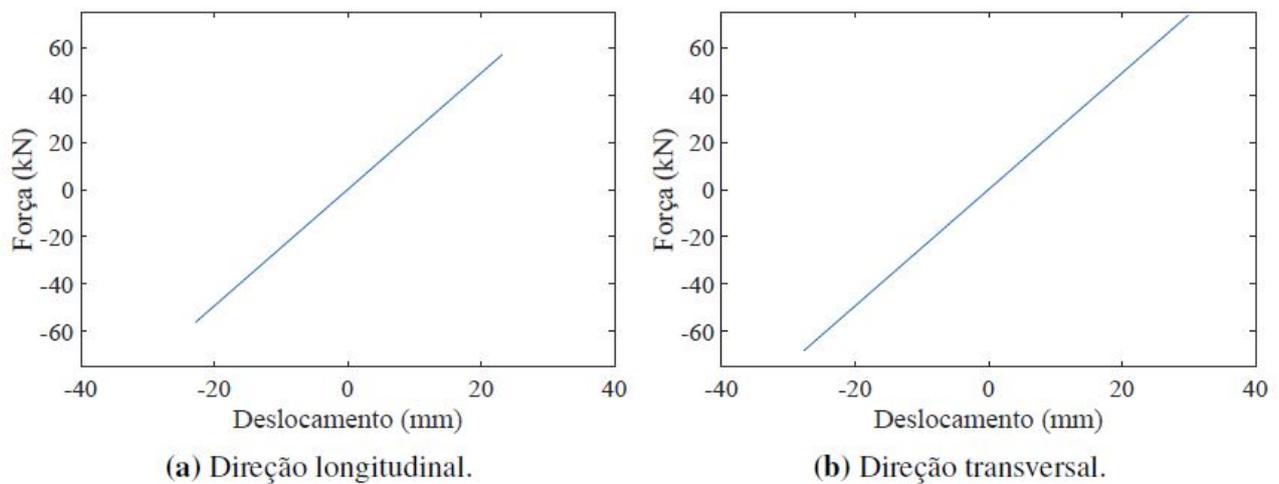


Figura 51 – Resposta histeréticas dos aparelhos de apoio (Fonte: adaptado de Cavalcante (2022)).

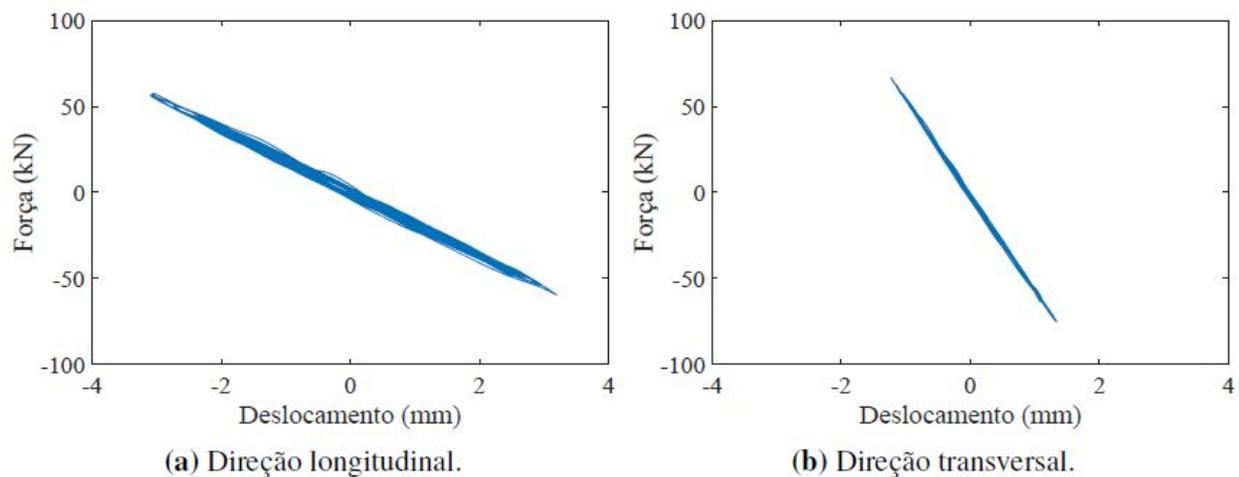


Figura 52 – Resposta histeréticas dos pilares (Fonte: adaptado de Cavalcante (2022)).

De acordo com Cavalcante (2022), os resultados apresentados mostram que as respostas estruturais estão condizentes com o esperado. Desta forma, com o modelo validado, as análises realizadas para o estudo estocástico também são aceitáveis.

4.1.2 Resultado estocástico

Esta seção tem como objetivo apresentar as curvas de fragilidade para a ponte de tipologia T3 classe C1, considerando as variáveis aleatórias expostas nos Capítulos 2 (propriedades geométricas) e 3 (propriedades físicas). Conforme exposto neste trabalho, para a construção das curvas de fragilidade é necessário obter as respostas dinâmicas, ou demanda estrutural, das componentes estruturais quando solicitadas por atividades sísmicas. As atividades sísmicas

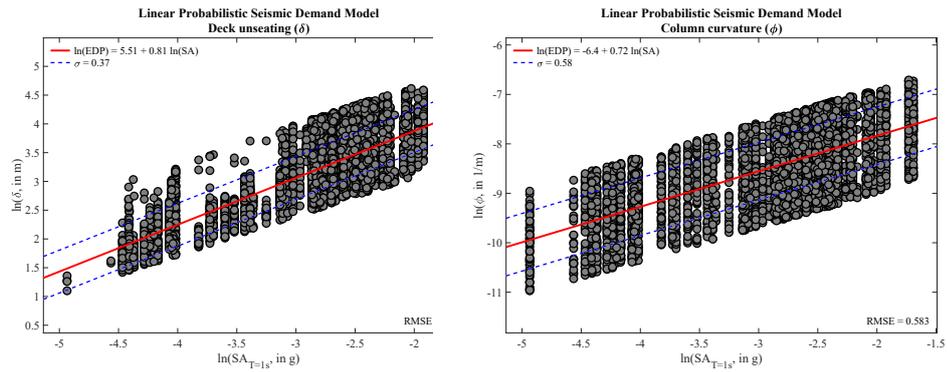
utilizadas neste trabalho são compatíveis com a região de estudo, onde utilizou-se o espectro de resposta média definido pela norma ASCE (2017) com valores retirados de Petersen et al. (2018), considerando um tempo de retorno de 475 anos, conforme exposto no Capítulo 3.

Os modelos desenvolvidos neste trabalho são resultados das análises dinâmicas não lineares. Com base no que foi exposto neste trabalho, 100 OAEs estatisticamente semelhantes foram geradas e, posteriormente, analisadas utilizando 100 pares de sinais sísmicos, totalizando 10.000 valores de demanda sísmica para cada Parâmetro de Demanda de Engenharia (EDP). Posteriormente a obtenção dos valores de demanda em função de uma intensidade de medida (IM), aplica-se o Modelo Probabilístico de Demanda Sísmica (PSDM). Conforme exposto por Cavalcante (2022), o modelo linear do PSDM é bastante representativo quando comparado com o modelo bilinear, sendo assim o modelo adotado para este trabalho. A Figura 53 a seguir, ilustra os modelos de demanda para cada um dos sete EDPs.

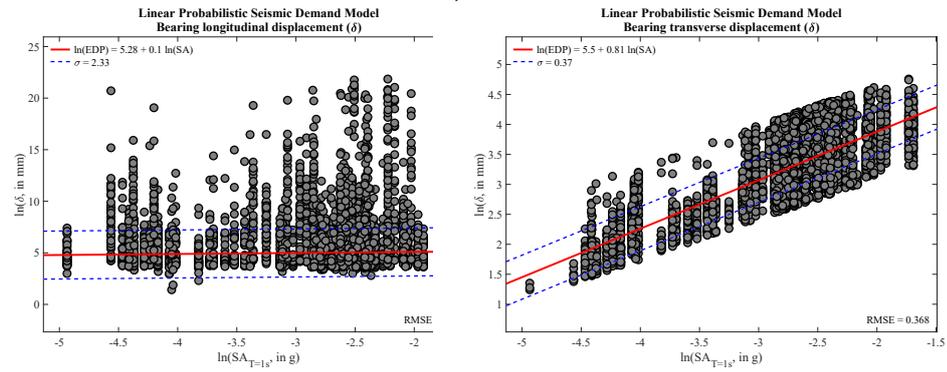
A Tabela 18 abaixo demonstra de maneira resumida os resultados dos Modelos Probabilísticos de Demanda Sísmica para cada Parâmetro de Demanda de Engenharia.

Tabela 18 – Resultados do PSDM para cada Parâmetro de Engenharia (Fonte: autor).

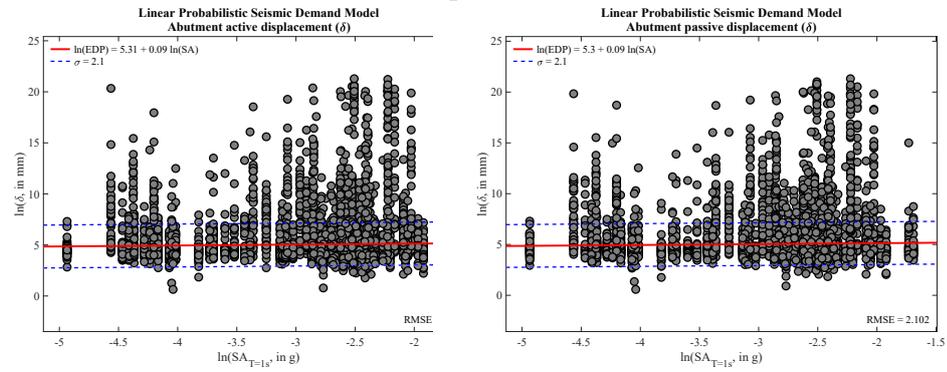
Parâmetro de Demanda	$\ln(a)$	b	Dispersão	R^2
Deslocamento do tabuleiro (mm)	5,51	0,81	0,37	0,65
Curvatura dos pilares (1/m)	-6,40	0,72	0,58	0,46
Deslocamento long. dos aparelhos de apoio (mm)	5,30	0,10	2,33	0,20
Deslocamento transv. dos aparelhos de apoio (mm)	5,50	0,81	0,37	0,66
Deslocamento ativo dos encontros (mm)	5,31	0,10	2,10	0,18
Deslocamento passivo dos encontros (mm)	5,30	0,10	2,11	0,23
Deslocamento transv. dos encontros (mm)	1,86	0,71	0,65	0,44



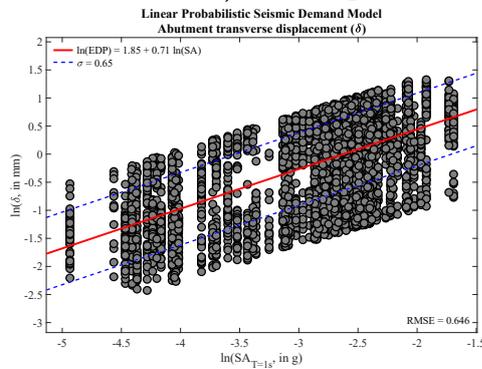
a) Deslocamento do tabuleiro. b) Curvatura dos pilares.



c) Desloc. long. dos aparel. de apoio. d) Desloc. trans. dos aparel. de apoio.



e) Desloc. ativo dos encontros. f) Desloc. passivo dos encontros.



g) Desloc. transversal dos encontros.

Figura 53 – Resultados dos PSDMs para a ponte de tipologia T3 classe C1 (Fonte: autor).

De acordo com a tabela, pode-se verificar que, para alguns parâmetros de demanda, o valor de dispersão foi alto, resultando em valores baixos de R^2 . De acordo com Nielson e DesRoches (2006), esse comportamento pode estar atribuído ao comportamento não-linear das juntas de dilação. Outro motivo que pode influenciar esse comportamento de alta dispersão é a grande variação dos parâmetros geométricos empregados neste estudo (CAVALCANTE, 2022).

Após obter os parâmetros de demanda através da regressão do PSDM, e conhecer os parâmetros de capacidade, é possível conduzir a análise de risco dinâmico através das curvas de fragilidade. Conforme exposto no Capítulo 3, as curvas de fragilidade medem a probabilidade condicional de uma estrutura ou de uma componente estrutural atingir ou ultrapassar um estado limite de dano em função de uma medida de intensidade IM. Desta forma, as curvas de fragilidade são utilizadas para mensurar a vulnerabilidade de uma estrutura que, para este estudo, se refere a uma ponte da tipologia T3 e classe C1.

A Figura 54 a seguir demonstra as curvas de fragilidade para o Parâmetros de Demanda de Engenharia para cada estado-limite de dano.

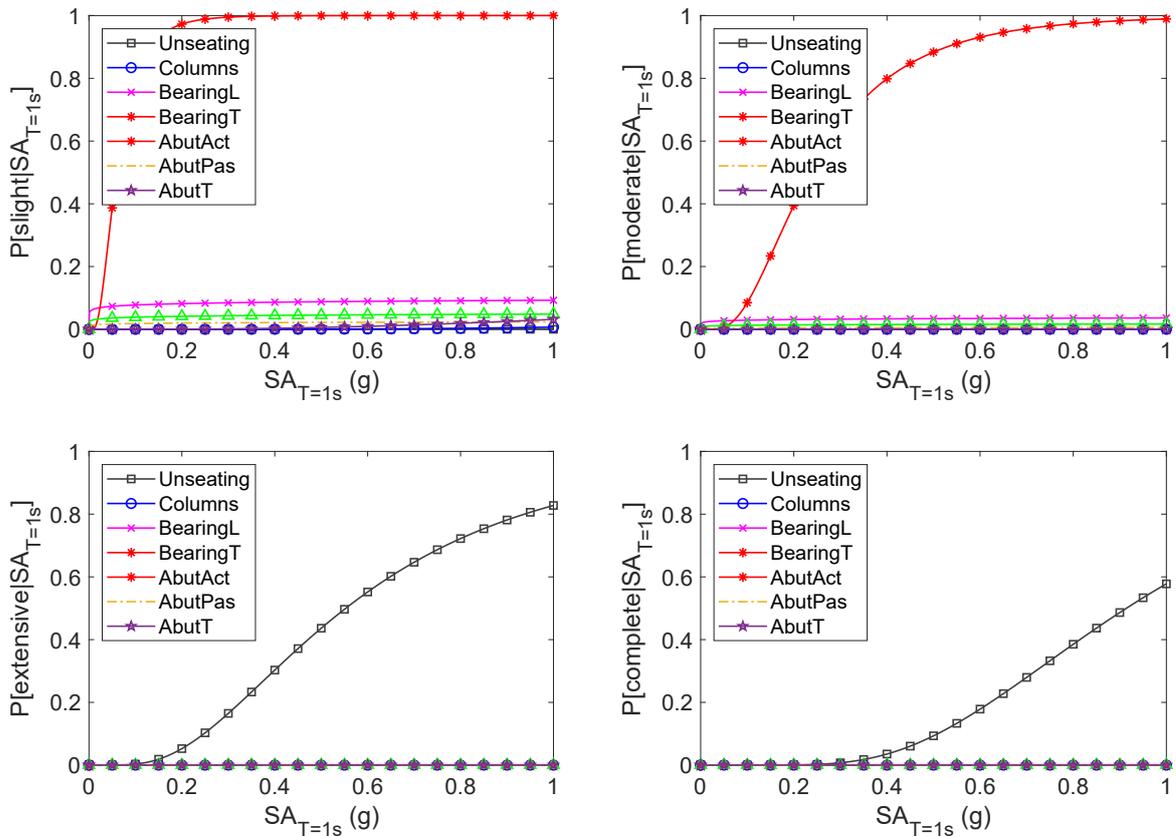


Figura 54 – Curvas de fragilidade para cada estado-limite de dano (Fonte: autor).

De acordo com a Figura 54, pode-se verificar que, para os estados-limites de dano leve e moderado, o deslocamento transversal dos aparelhos de apoio é o parâmetro mais crítico. Já para os limites de dano extensivo e completo é o deslocamento do tabuleiro que determina a fragilidade do sistema estrutural.

A Figura 55 a seguir demonstra a curva de fragilidade para o sistema estrutural para cada estado-limite de dano.

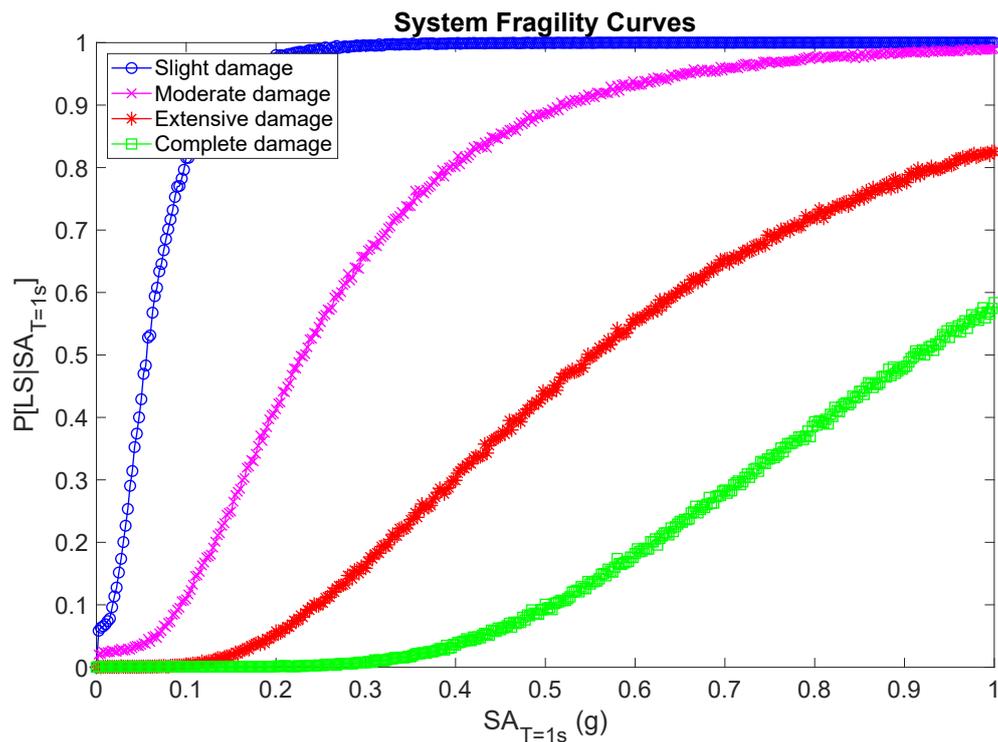


Figura 55 – Curva de fragilidade do sistema estrutural para cada estado-limite de dano (Fonte: autor).

4.2 Resultado do Machine Learning

Esta seção tem como objetivo apresentar os resultados dos processos de inteligência artificial (*machine learning*), que será dividido em duas partes: resultados dos modelos de regressão e resultados dos modelos de classificação. Para as figuras que representam um modelo esquemático do algoritmo *Decision Tree* ou *Random Forest*, entende-se que os ramos descendentes à esquerda representam a opção de menor ($<$) em relação ao valor estipulado, e os ramos descendentes à direita representam a opção de maior ou igual (\geq) ao valor estipulado.

4.2.1 Resultados da regressão

Os modelos de regressão foram treinados para cada um dos parâmetros de demanda de engenharia (EDP), contabilizando um total de setenta mil análises, conforme exposto no Capítulo 3. Para facilitar a visualização do leitor, apenas o resultado da curvatura dos pilares (CC) será exposto em sua totalidade, ou seja, as verificações gráficas irão conter 10.000 pontos. Os demais parâmetros foram filtrados para apenas 1% dos respectivos dados, totalizando 100 pontos para as verificações gráficas.

Como todos os modelos escolhidos, tanto para o *training set* quanto para o *testing set*, foram do tipo “Tree”, seja *Decision Tree* ou *Random Forest*, será ilustrado a árvore de decisão para cada EDP. Para os casos onde o modelo com melhor performance foi o *Random Forest*, será ilustrado apenas a primeira árvore de decisão. O fato dos melhores modelos de regressão serem do tipo “Tree”, condiz com os estudos sobre o mesmo tema como exposto por Mangalathu e Jeon (2019).

4.2.1.1 Curvatura dos pilares

Para o parâmetro curvatura dos pilares (CC), o modelo com melhor performance foi o *decision tree*. A Tabela 19 a seguir demonstra os resultados obtidos para a curvatura dos pilares para os modelos do conjunto de treinamento através do *Root Mean Square Error*.

Tabela 19 – CC - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE $\times 10^5$
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	18,77
	<i>Quadratic</i>	14,43
	<i>Cubic</i>	17,95
	<i>Fine Gaussian</i>	44,30
	<i>Medium Gaussian</i>	13,56
	<i>Coarse Gaussian</i>	19,33
<i>Linear Regression</i>	<i>Linear</i>	15,47
	<i>Interactions Linear</i>	15,90
	<i>Robust Linear</i>	18,42
<i>Decision Tree</i>	<i>Fine</i>	1,59
	<i>Medium</i>	2,60
	<i>Coarse</i>	4,16
<i>Random Forest</i>	<i>Bagged</i>	5,57
	<i>Boosted</i>	2,51

De acordo com a tabela de verificação da performance, o algoritmo *Decision Tree - Fine type* teve o melhor desempenho entre todos os algoritmos. A Tabela 20 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 20 – CC - *training set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	$1,59 \cdot 10^{-5}$
R^2	1,00
MSE	$2,53 \cdot 10^{-10}$
MAE	$2,71 \cdot 10^{-6}$
<i>Prediction Speed</i>	Aprox. 290000 obs/sec
<i>Training time</i>	1,07 sec
<i>Minimum leaf size</i>	4

Fica evidente que, de acordo com os dados expostos na tabela da curvatura dos pilares para o *training set*, que o modelo de regressão teve uma ótima performance, de maneira que o parâmetro R^2 ficou aproximadamente igual a 1,0. Isso significa que o modelo explica aproximadamente 100% da variância da curvatura dos pilares. No entanto, conforme exposto anteriormente e de acordo com Hastie et al. (2001), a performance do modelo também deve ser avaliada pelo conjunto de teste. Desta forma, a Tabela 21 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 21 – CC - *testing set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	$2,62 \cdot 10^{-5}$
R^2	0,98
MSE	$6,86 \cdot 10^{-10}$
MAE	$3,51 \cdot 10^{-6}$

Nota-se que o algoritmo de *Decision Tree* performou de maneira similar para o *testing set*, logo o modelo não cometeu *overfitting*.

A seguir, as Figuras 56 à 61 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

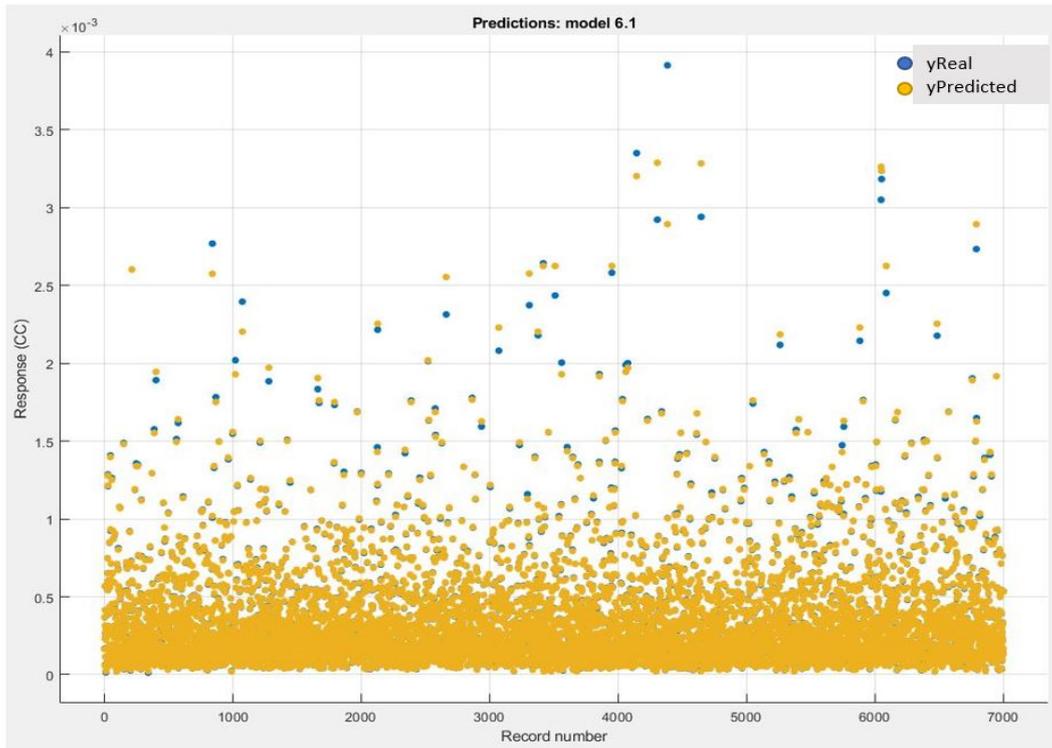


Figura 56 – CC - Response plot - training set (Fonte: autor).

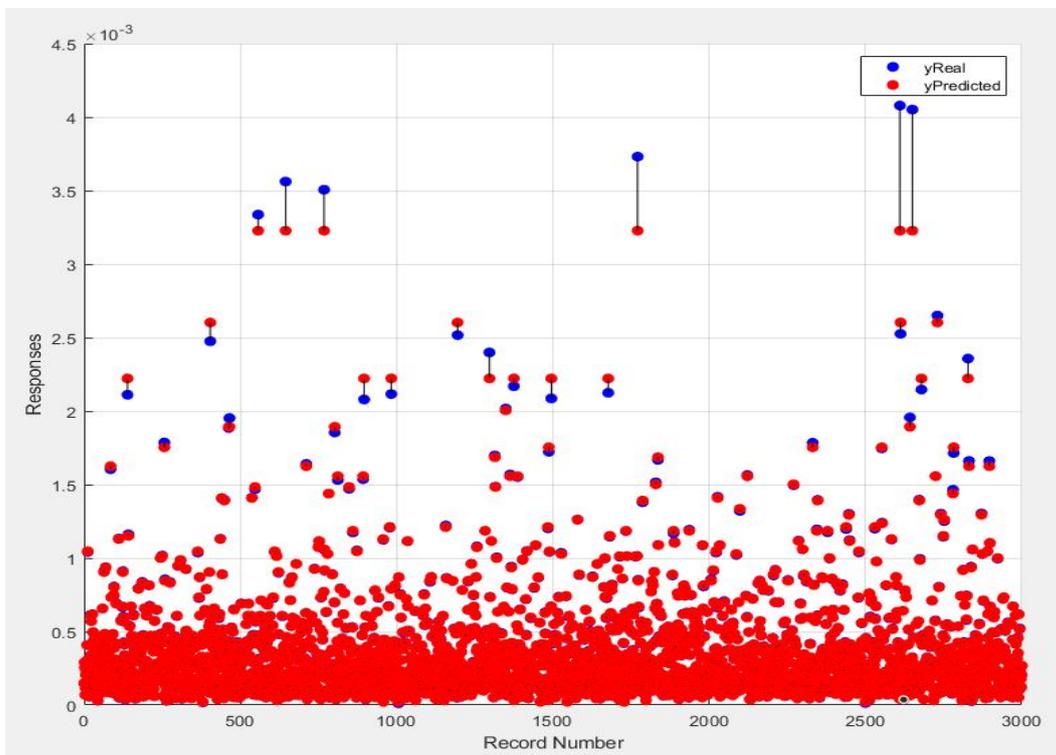


Figura 57 – CC - Response plot - testing set (Fonte: autor).

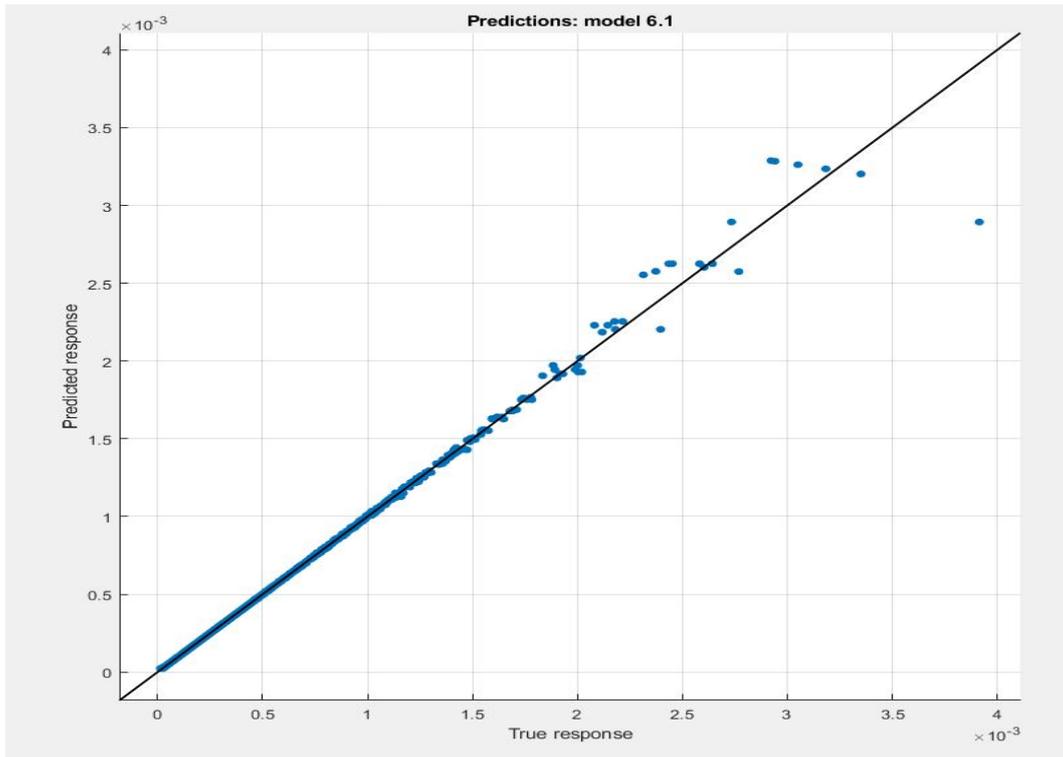


Figura 58 – CC - Predicted vs. Actual plot - training set (Fonte: autor).

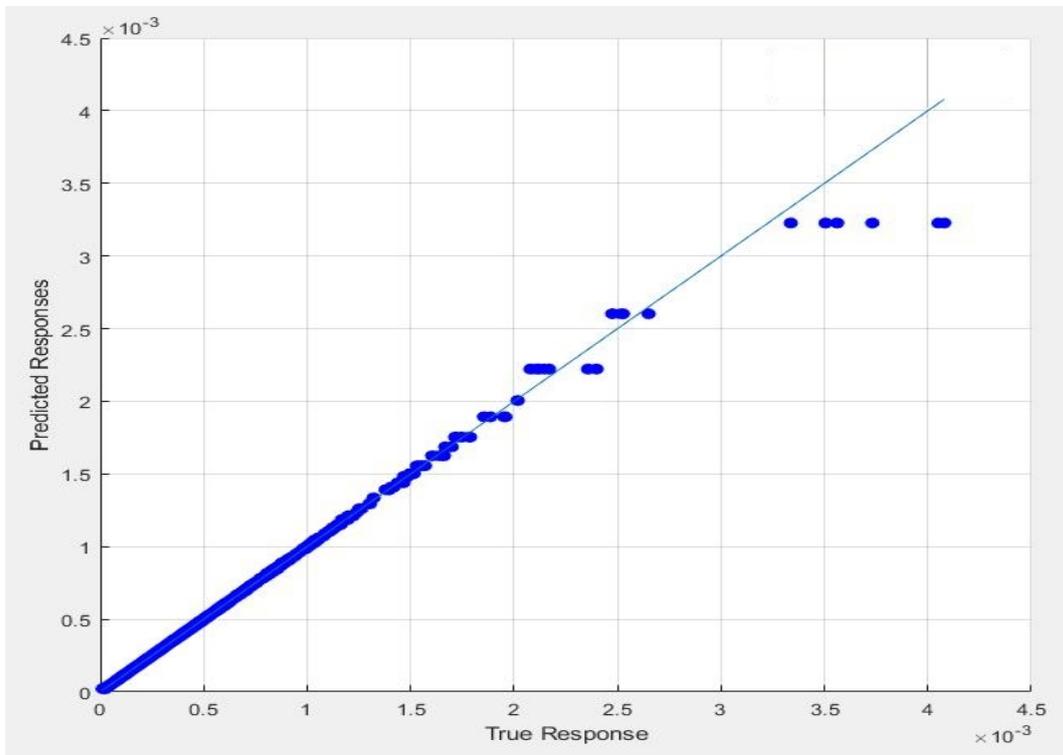


Figura 59 – CC - Predicted vs. Actual plot - testing set (Fonte: autor).

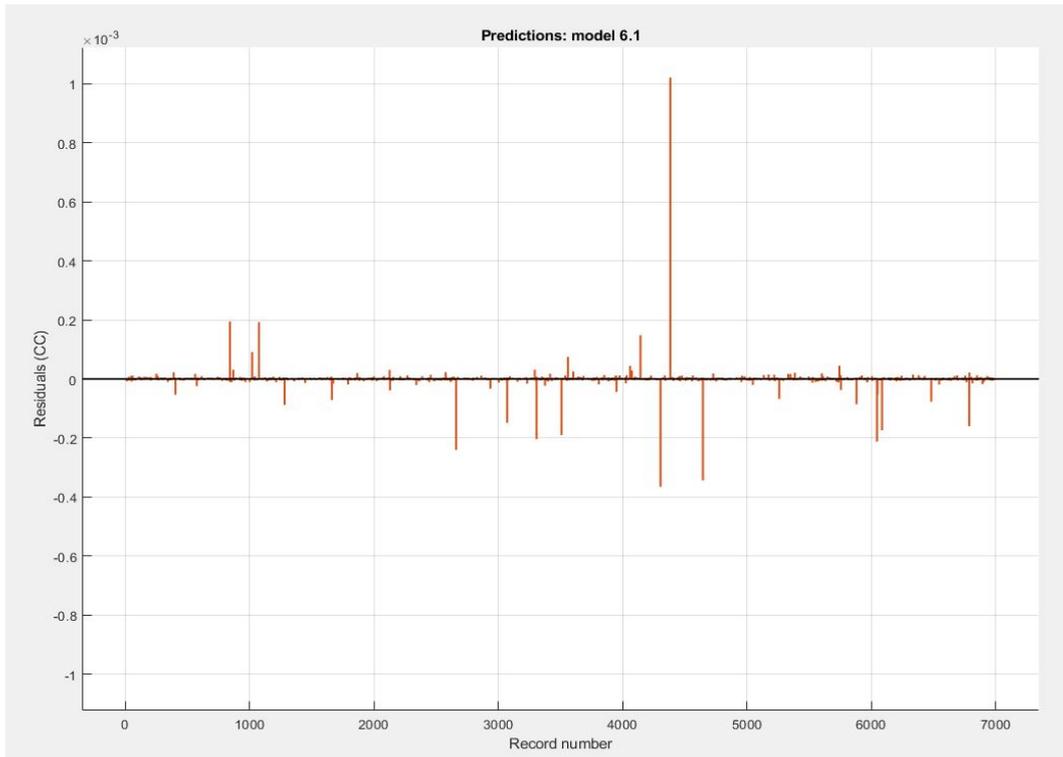


Figura 60 – CC - Residual plot - training set (Fonte: autor).

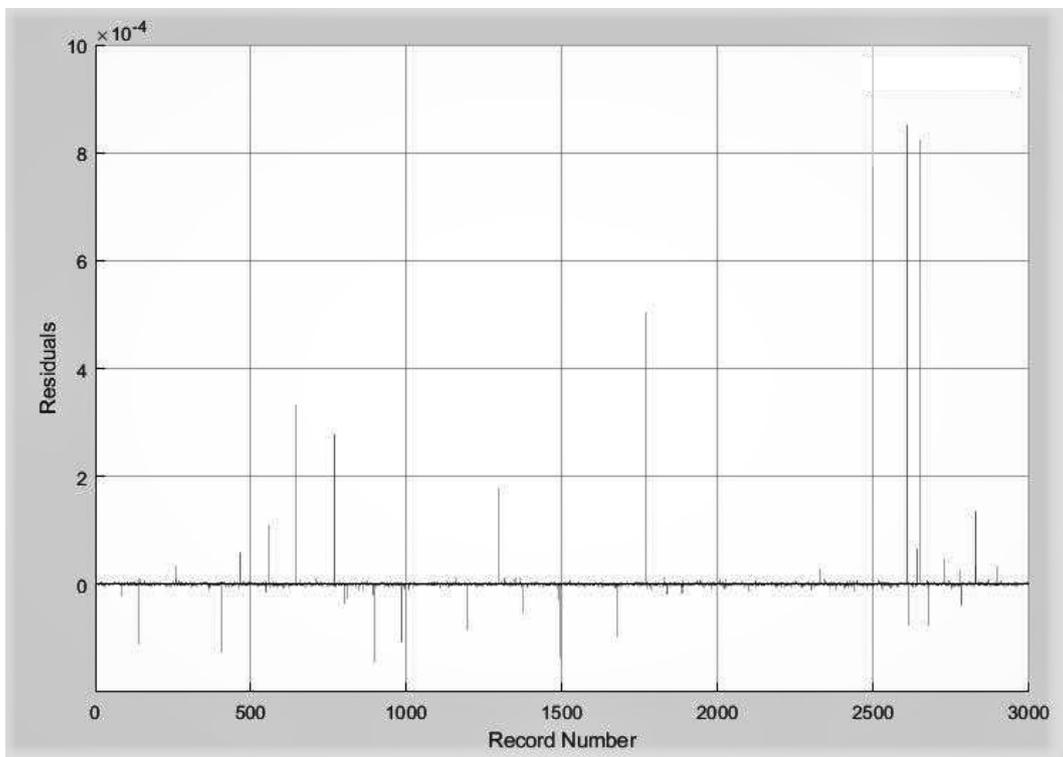


Figura 61 – CC - Residual plot - testing set (Fonte: autor).

Tabela 22 – ADA - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	0,1040
	<i>Quadratic</i>	0,1021
	<i>Cubic</i>	0,1005
	<i>Fine Gaussian</i>	0,1111
	<i>Medium Gaussian</i>	0,1052
	<i>Coarse Gaussian</i>	0,1074
<i>Linear Regression</i>	<i>Linear</i>	0,1164
	<i>Interactions Linear</i>	0,5728
	<i>Robust Linear</i>	0,1396
<i>Decision Tree</i>	<i>Fine</i>	0,0511
	<i>Medium</i>	0,0876
	<i>Coarse</i>	0,1088
<i>Random Forest</i>	<i>Bagged</i>	0,0850
	<i>Boosted</i>	0,0836

Tabela 23 – ADA - *training set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0511
R^2	0,78
MSE	0,0026
MAE	0,0256
<i>Prediction Speed</i>	Aprox. 2900 obs/sec
<i>Training time</i>	0,3923 sec
<i>Minimum leaf size</i>	4

Conforme os dados expostos na tabela do deslocamento ativo dos encontros para o *training set*, o modelo teve uma boa performance, de maneira que o parâmetro R^2 ficou aproximadamente igual a 0,8. Isso significa que o modelo explica aproximadamente 80% da variância do deslocamento ativo dos encontros. No entanto, conforme exposto para a EDP curvatura dos pilares, a performance do modelo também deve ser avaliada pelo conjunto de teste. Desta forma, a Tabela 24 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 24 – ADA - *testing set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0632
R^2	0,7750
MSE	0,0040
MAE	0,0363

Nota-se que o algoritmo de *Decision Tree* performou de maneira similar para o *testing set*, logo o modelo não cometeu *overfitting*.

A seguir, as Figuras 63 à 68 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

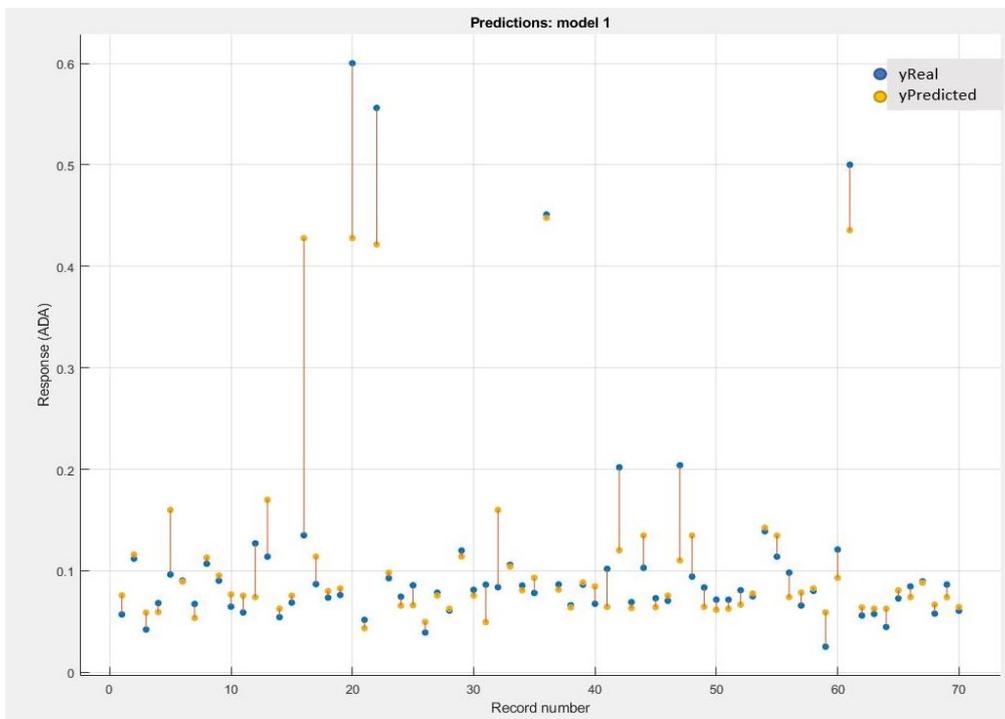


Figura 63 – ADA - *Response plot* - *training set* (Fonte: autor).

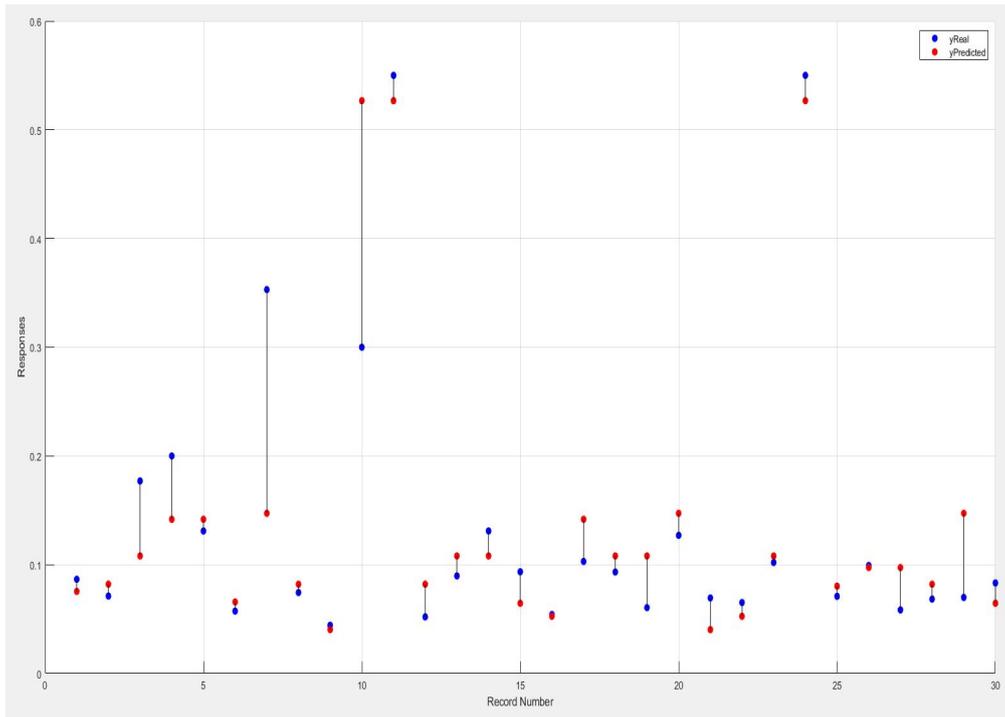


Figura 64 – ADA - *Response plot - testing set* (Fonte: autor).

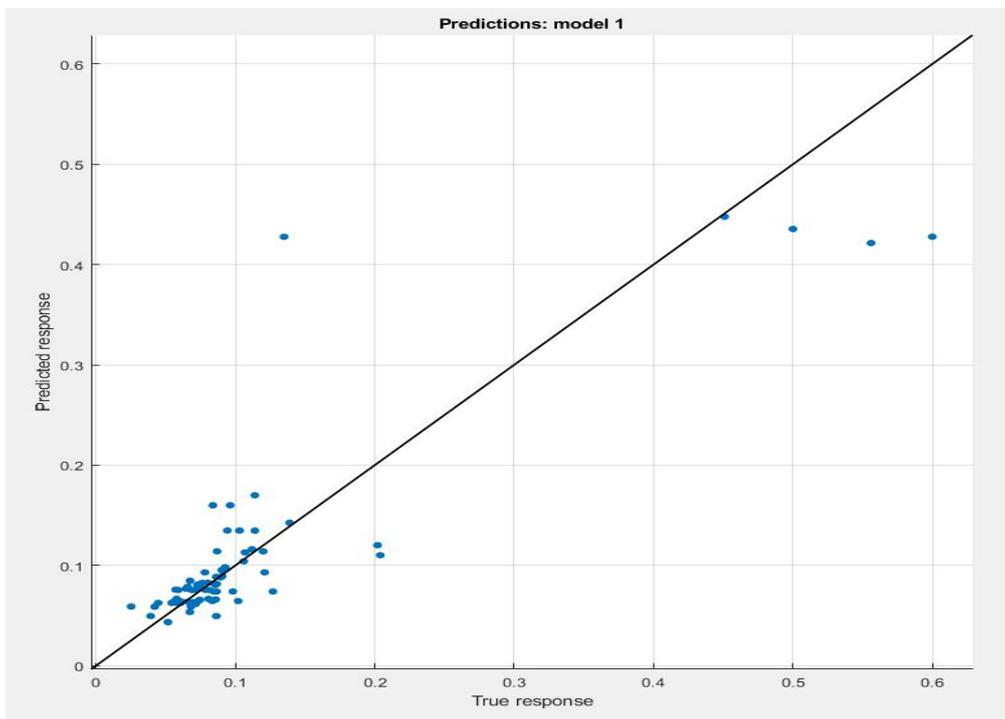


Figura 65 – ADA - *Predicted vs. Actual plot - training set* (Fonte: autor).

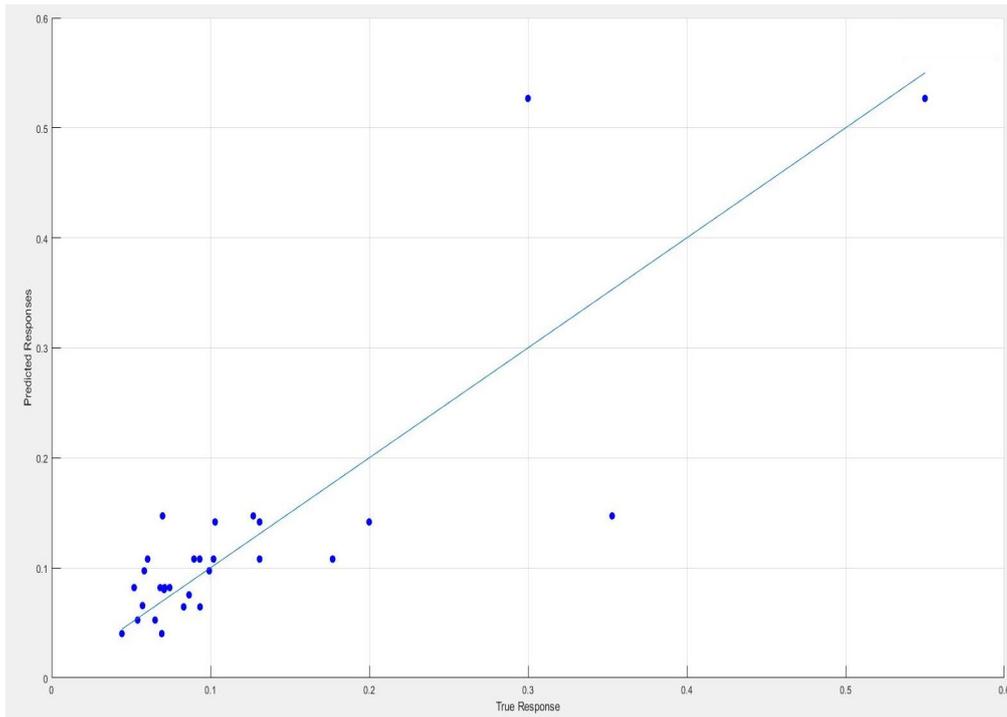


Figura 66 – ADA - Predicted vs. Actual plot - testing set (Fonte: autor).

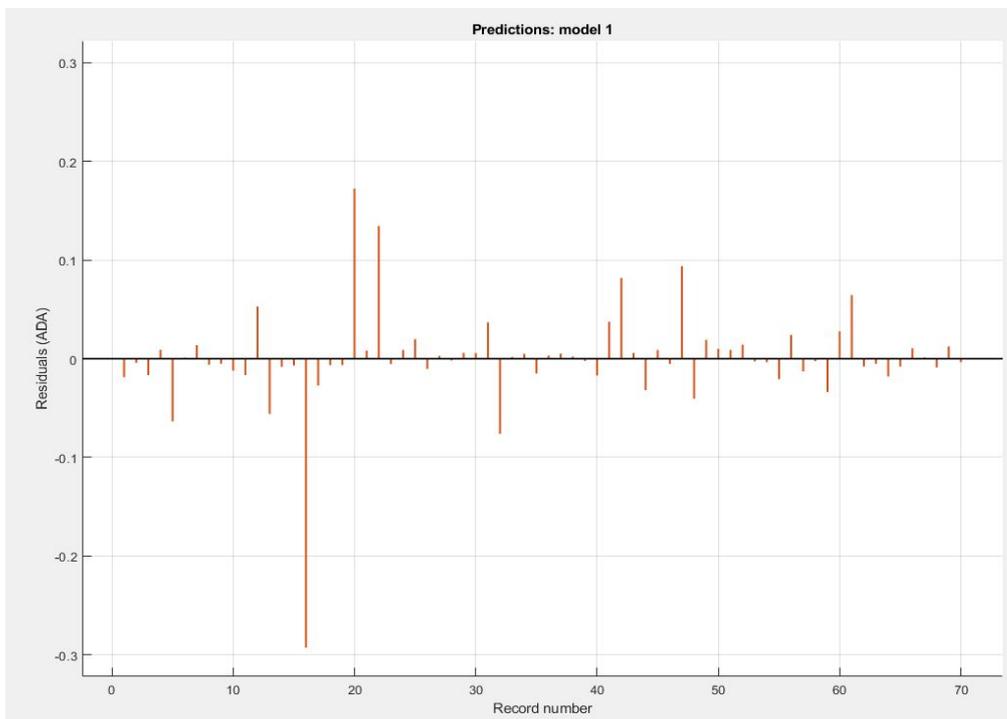


Figura 67 – ADA - Residual plot - training set (Fonte: autor).

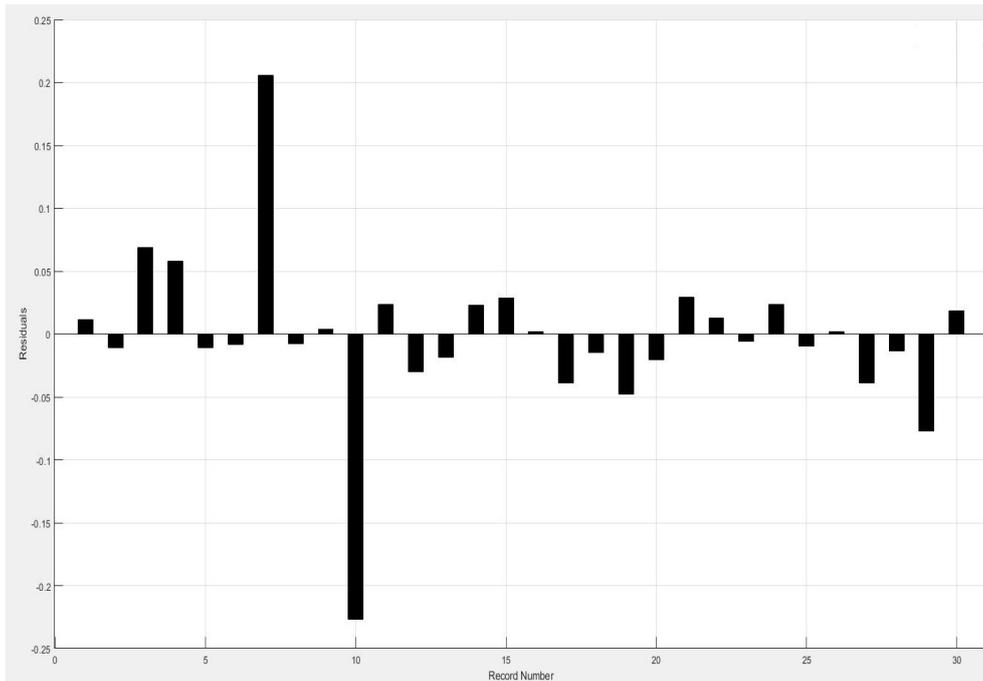


Figura 68 – ADA - Residual plot - testing set (Fonte: autor).

Pode-se observar a partir das análises gráficas que o modelo de regressão performou de maneira muito similar tanto para o *training set* quanto para o *testing set*. A Figura 69 ilustra o esquema da *Decision Tree* utilizada pelo algoritmo. É possível visualizar que a variável aleatória “*Backwall height*”, que se refere a altura dos encontros, é a mais utilizada no algoritmo.

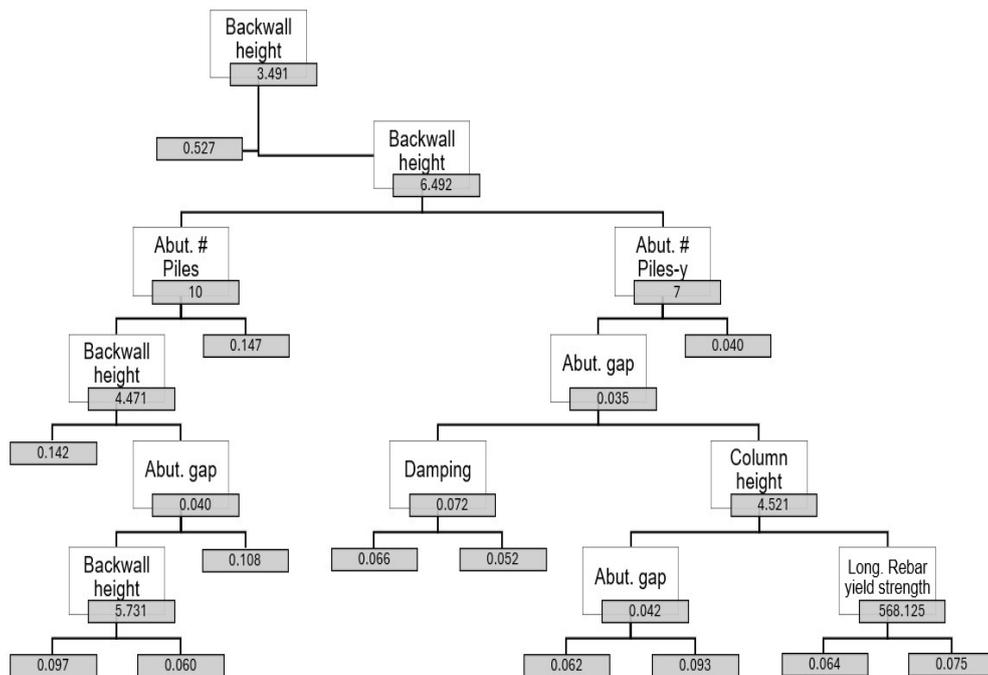


Figura 69 – ADA - Modelo esquemático da *Decision Tree*.

4.2.1.3 Deslocamento passivo dos encontros

Para o parâmetro deslocamento passivo dos encontros (ADP), o modelo com melhor performance foi, também, o *decision tree*. Diferente dos parâmetros de demanda de engenharia expostos até o momento, o deslocamento passivo dos encontros teve uma performance ruim. A Tabela 25 a seguir demonstra os resultados obtidos para o deslocamento passivo dos encontros para os modelos do conjunto de treinamento através do *Root Mean Square Error*.

Tabela 25 – ADP - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	0,0535
	<i>Quadratic</i>	0,0530
	<i>Cubic</i>	0,0500
	<i>Fine Gaussian</i>	0,0612
	<i>Medium Gaussian</i>	0,0552
	<i>Coarse Gaussian</i>	0,0571
<i>Linear Regression</i>	<i>Linear</i>	0,0566
	<i>Interactions Linear</i>	0,3045
	<i>Robust Linear</i>	0,0612
<i>Decision Tree</i>	<i>Fine</i>	0,0418
	<i>Medium</i>	0,0484
	<i>Coarse</i>	0,0600
<i>Random Forest</i>	<i>Bagged</i>	0,0483
	<i>Boosted</i>	0,0461

De acordo com a tabela de verificação da performance, o algoritmo *Decision Tree - Fine type* teve o melhor desempenho entre todos os algoritmos. A Tabela 26 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 26 – ADP - *training set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0418
R^2	0,51
MSE	0,0017
MAE	0,0257
<i>Prediction Speed</i>	Aprox. 3100 obs/sec
<i>Training time</i>	0,35 sec
<i>Minimum leaf size</i>	4

Conforme os dados expostos na tabela do deslocamento passivo dos encontros para o *training set*, o modelo teve uma performance ruim, de maneira que o parâmetro R^2 ficou aproximadamente igual a 0,5. Isso significa que o modelo explica aproximadamente 50% da variância do deslocamento passivo dos encontros. A performance do modelo também deve ser avaliada pelo conjunto de teste. Desta forma, a Tabela 27 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 27 – ADP - *testing set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0452
R^2	0,51
MSE	0,0020
MAE	0,0324

Mesmo o algoritmo não tendo uma boa performance, nota-se que o modelo performou de maneira similar para o *testing set*, logo o modelo não cometeu *overfitting*.

A seguir, as Figuras 70 à 75 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

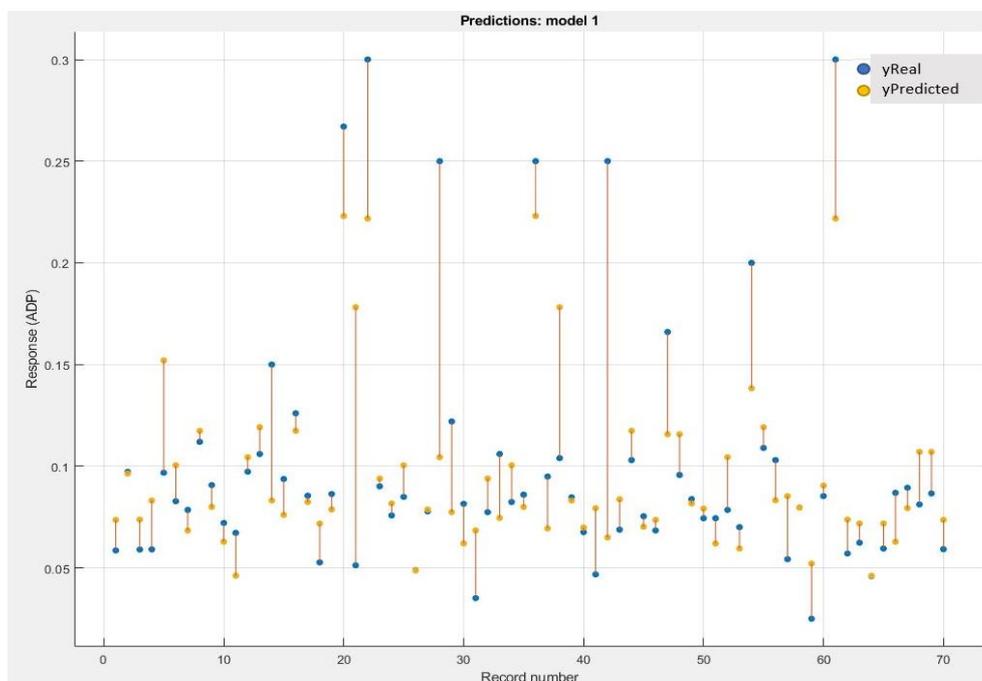


Figura 70 – ADP - *Response plot* - *training set* (Fonte: autor).

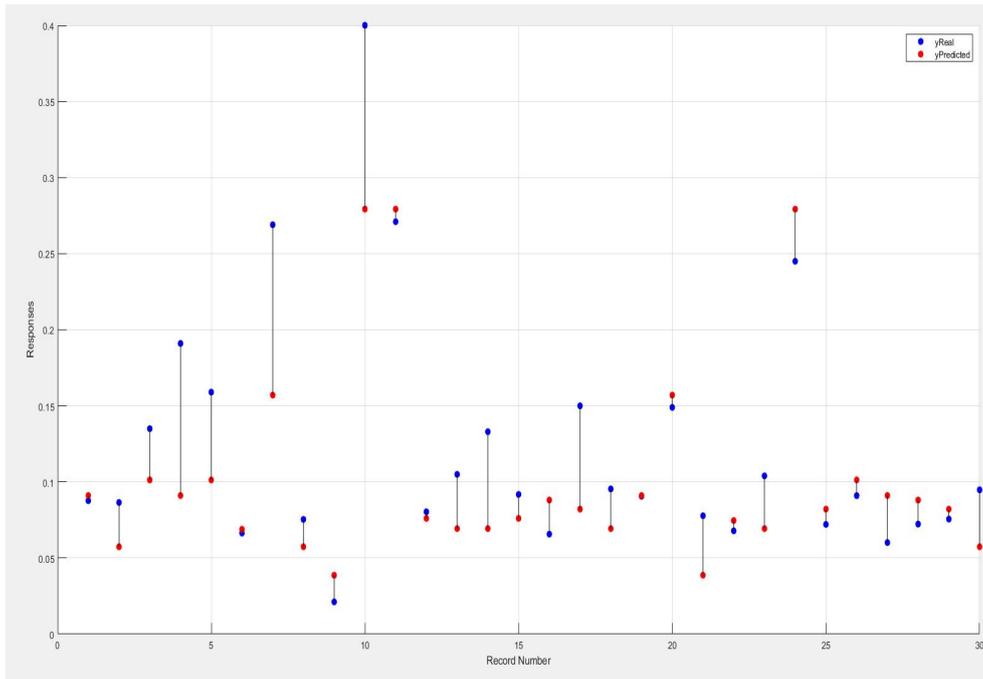


Figura 71 – ADP - Response plot - testing set (Fonte: autor).

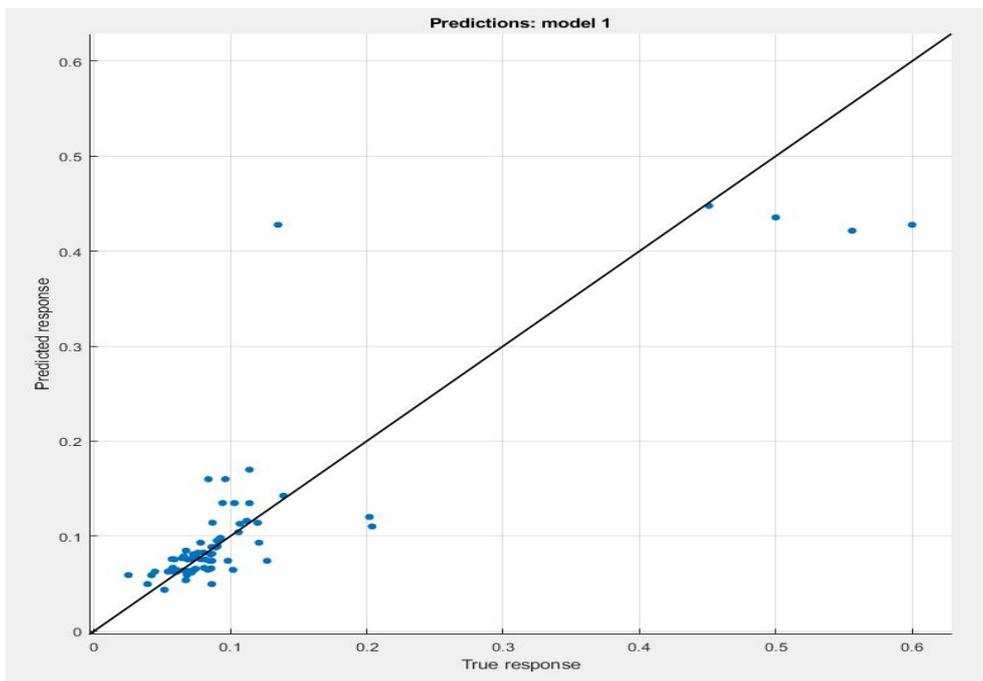


Figura 72 – ADP - Predicted vs. Actual plot - training set (Fonte: autor).

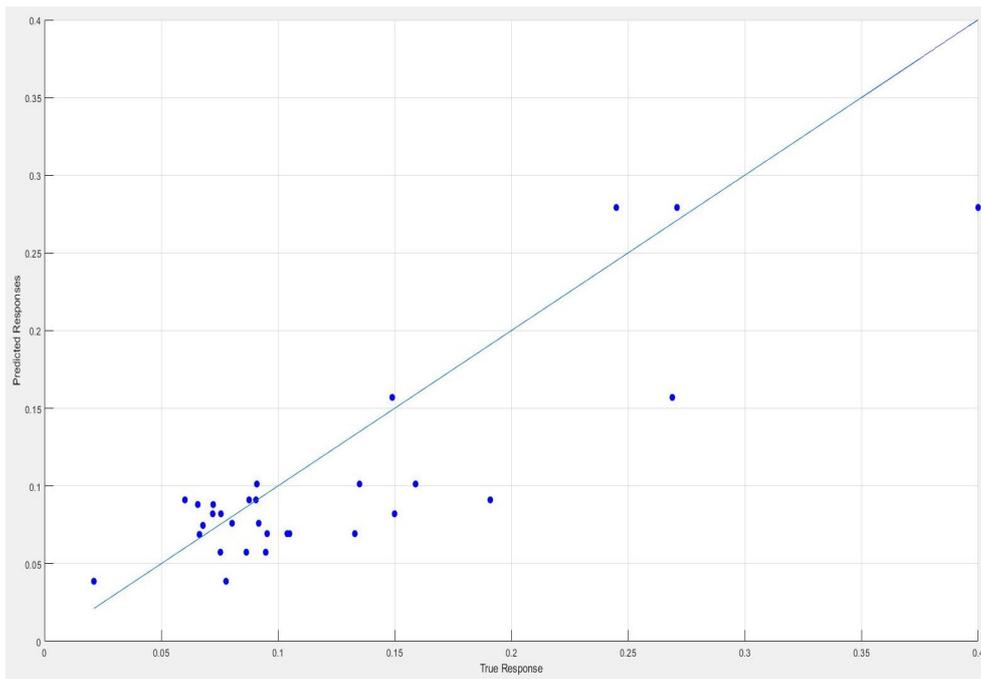


Figura 73 – ADP - Predicted vs. Actual plot - testing set (Fonte: autor).

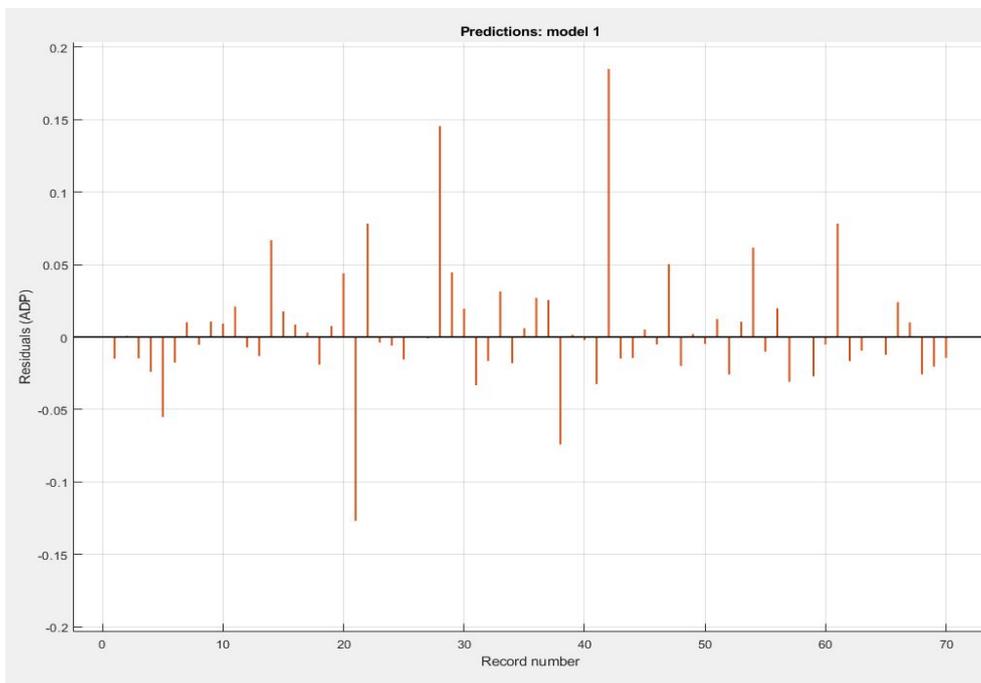


Figura 74 – ADP - Residual plot - training set (Fonte: autor).

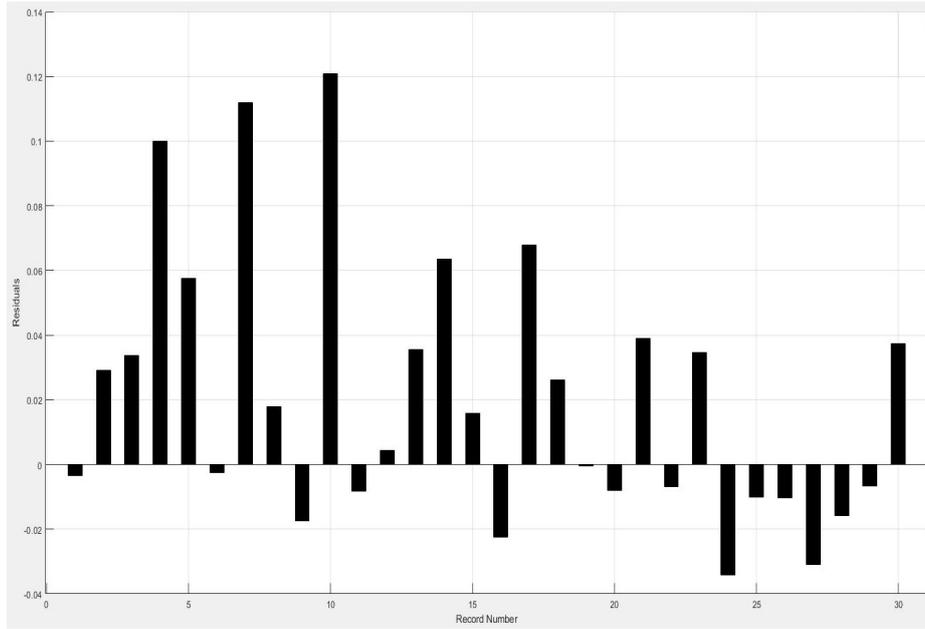


Figura 75 – ADP - Residual plot - testing set (Fonte: autor).

Pode-se observar a partir das análises gráficas que o modelo de regressão performou de maneira muito similar tanto para o *training set* quanto para o *testing set*. A Figura 76 ilustra o esquema da *Decision Tree* utilizada pelo algoritmo. É possível visualizar que a variável aleatória “*Backwall height*”, que se refere a altura dos encontros, é a mais utilizada no algoritmo.

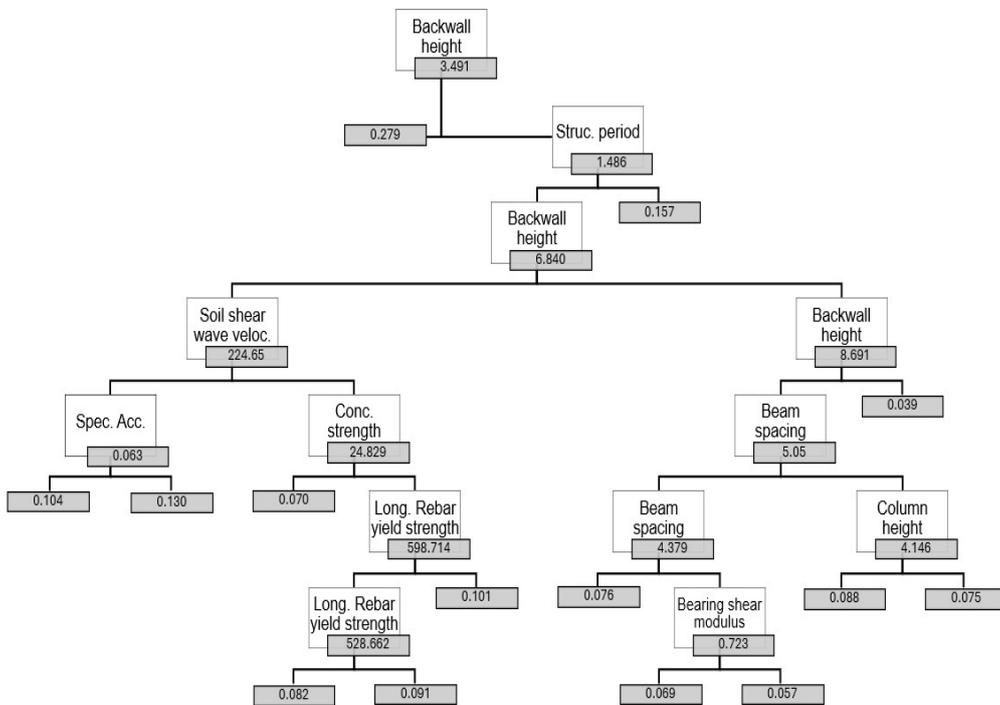


Figura 76 – ADP - Modelo esquemático da Decision Tree (Fonte: autor).

4.2.1.4 Deslocamento transversal dos encontros

Para o parâmetro deslocamento transversal dos encontros (ADT), o modelo com melhor performance foi o *Random Forest*. A Tabela 28 a seguir demonstra os resultados obtidos para o deslocamento transversal dos encontros para os modelos do conjunto de treinamento através do *Root Mean Square Error*.

Tabela 28 – ADT - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE $\times 10^5$
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	45,60
	<i>Quadratic</i>	37,11
	<i>Cubic</i>	36,61
	<i>Fine Gaussian</i>	69,35
	<i>Medium Gaussian</i>	39,92
	<i>Coarse Gaussian</i>	55,75
<i>Linear Regression</i>	<i>Linear</i>	51,37
	<i>Interactions Linear</i>	158,46
	<i>Robust Linear</i>	62,54
<i>Decision Tree</i>	<i>Fine</i>	35,72
	<i>Medium</i>	36,12
	<i>Coarse</i>	69,90
<i>Random Forest</i>	<i>Bagged</i>	37,02
	<i>Boosted</i>	30,52

De acordo com a tabela de verificação da performance, o algoritmo *Random Forest - Boosted Tree* teve o melhor desempenho entre todos os algoritmos. A Tabela 29 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 29 – ADT - Deslocamento transversal dos encontros - *training set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
RMSE	$30,52 \cdot 10^{-5}$
R^2	0,81
MSE	$9,31 \cdot 10^{-8}$
MAE	$22,87 \cdot 10^{-5}$
<i>Prediction Speed</i>	Aprox. 690 obs/sec
<i>Training time</i>	2,30 sec
<i>Minimum leaf size</i>	8
<i>Number of learners</i>	60

Conforme os dados expostos na tabela do deslocamento transversal dos encontros para o *training set*, o modelo teve uma boa performance, de maneira que o parâmetro R^2 ficou aproximadamente igual a 0,8. Isso significa que o modelo explica aproximadamente 80% da variância do deslocamento transversal dos encontros.

A performance do modelo também deve ser avaliada pelo conjunto de teste. Desta forma, a Tabela 30 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 30 – ADT - Deslocamento transversal dos encontros - *testing set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
RMSE	$28,51 \cdot 10^{-5}$
R^2	0,80
MSE	$7,98 \cdot 10^{-8}$
MAE	$20,36 \cdot 10^{-5}$

Nota-se que o modelo performou de maneira similar para o *testing set*, logo o modelo não cometeu *overfitting*.

A seguir, as Figuras 77 à 82 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

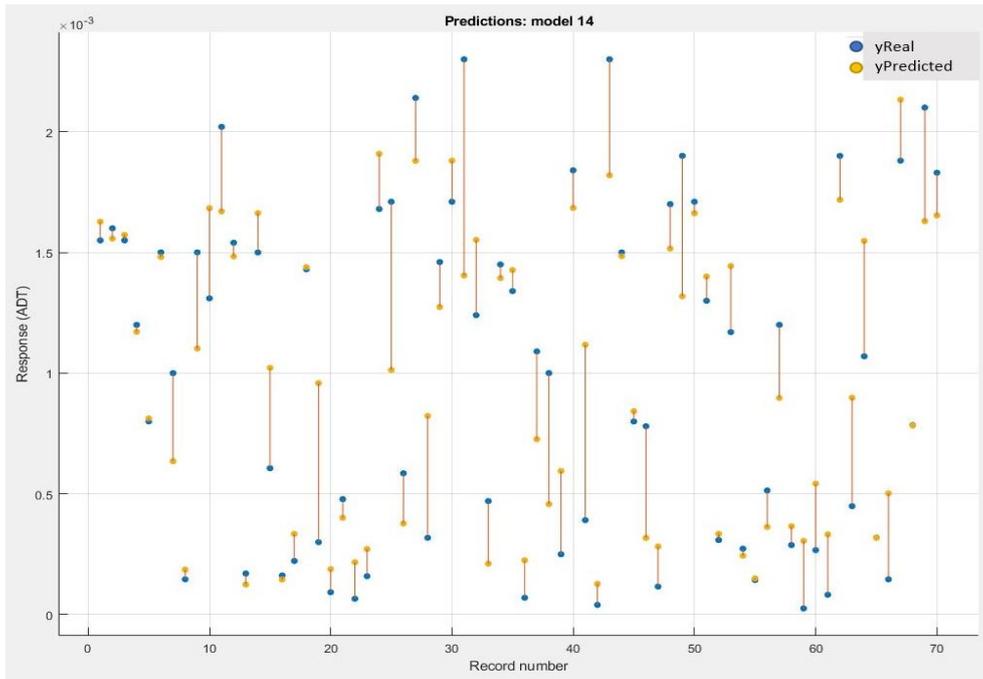


Figura 77 – ADT - Response plot - training set (Fonte: autor).

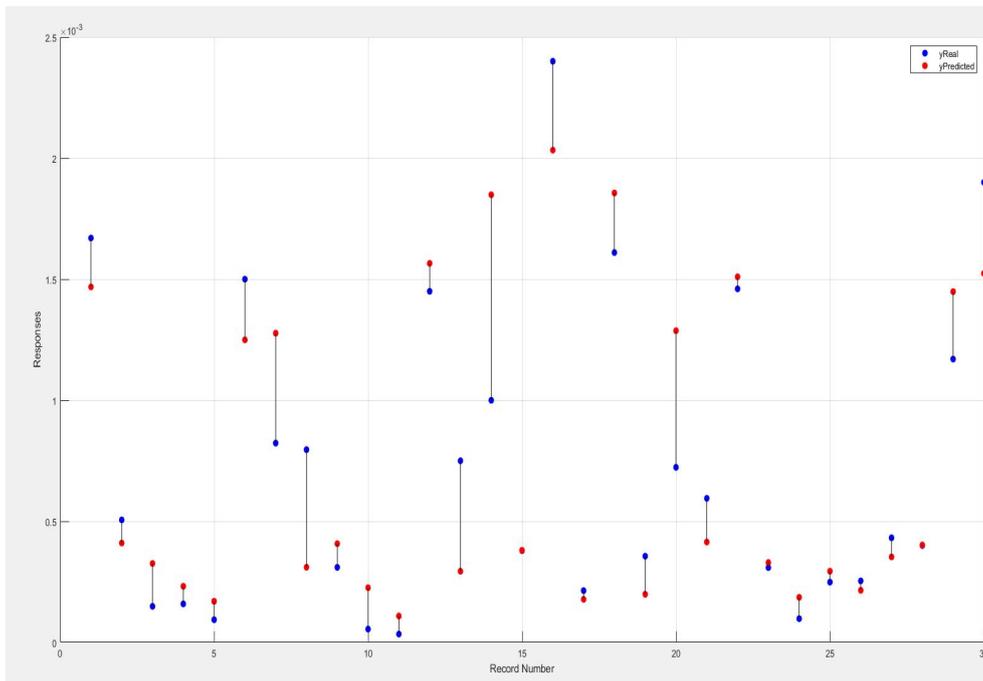


Figura 78 – ADT - Response plot - testing set (Fonte: autor).

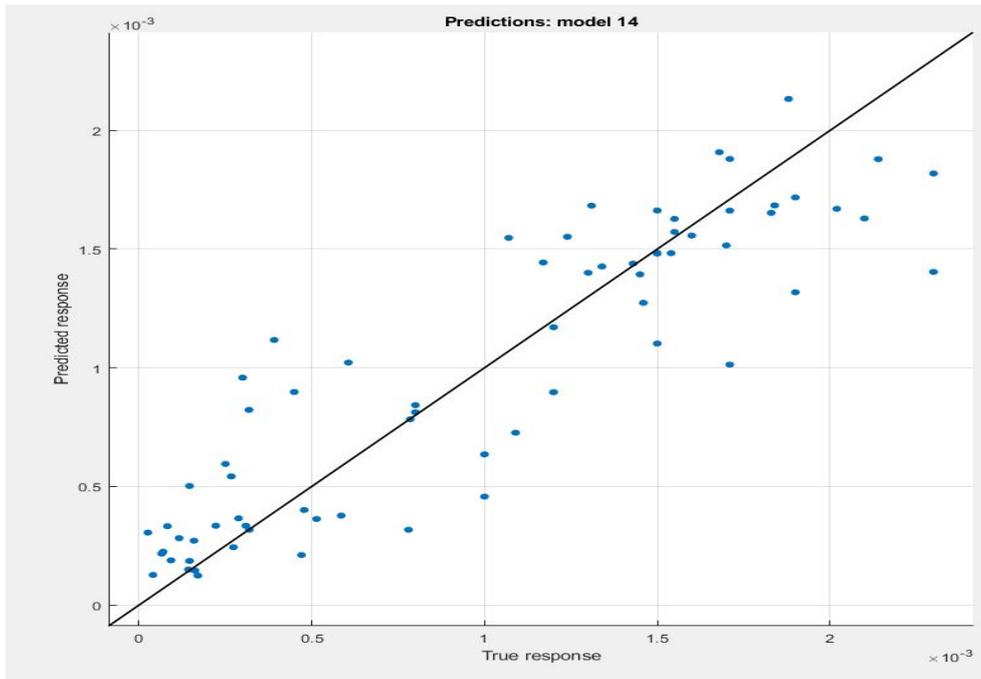


Figura 79 – ADT - Predicted vs. Actual plot - training set (Fonte: autor).

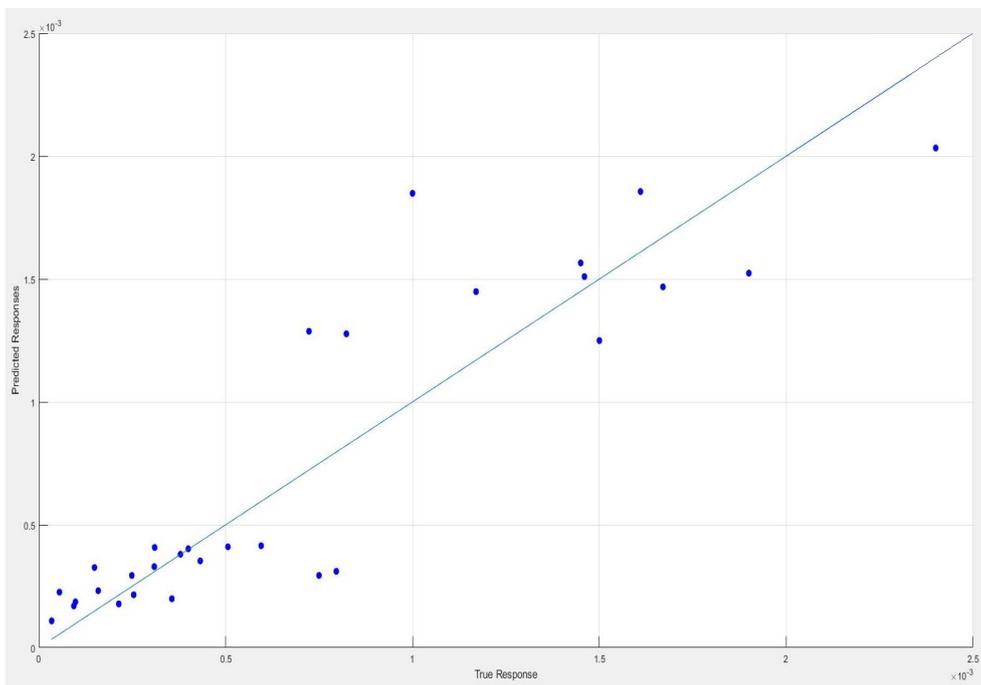


Figura 80 – ADT - Predicted vs. Actual plot - testing set (Fonte: autor).

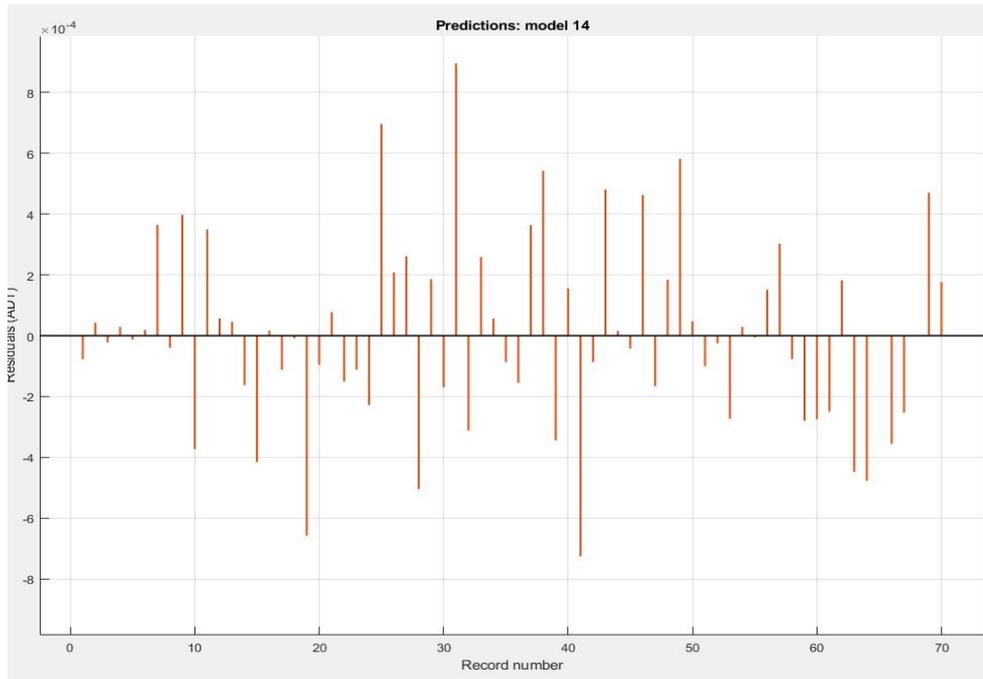


Figura 81 – ADT - Residual plot - training set (Fonte: autor).

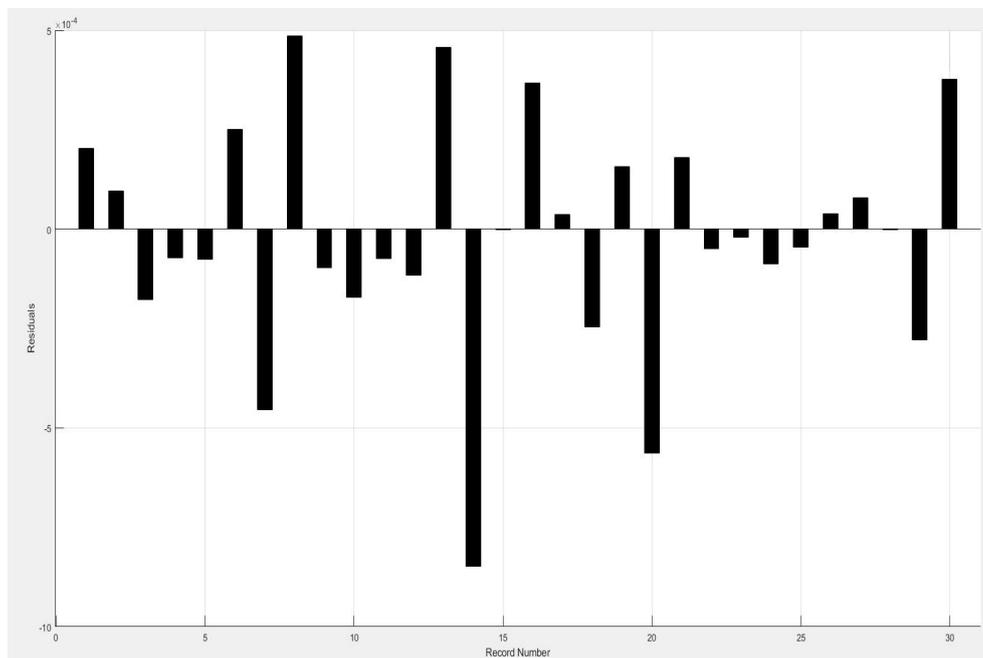


Figura 82 – ADT - Residual plot - testing set (Fonte: autor).

Pode-se observar a partir das análises gráficas que o modelo de regressão performou de maneira muito similar tanto para o *training set* quanto para o *testing set*. A Figura 83 ilustra o esquema da primeira, de um total de 60, árvore de decisão do *Random Forest* utilizada pelo

algoritmo. É possível visualizar que a variável aleatória “*Backwall height*”, que se refere a altura dos encontros, é a mais utilizada no algoritmo.

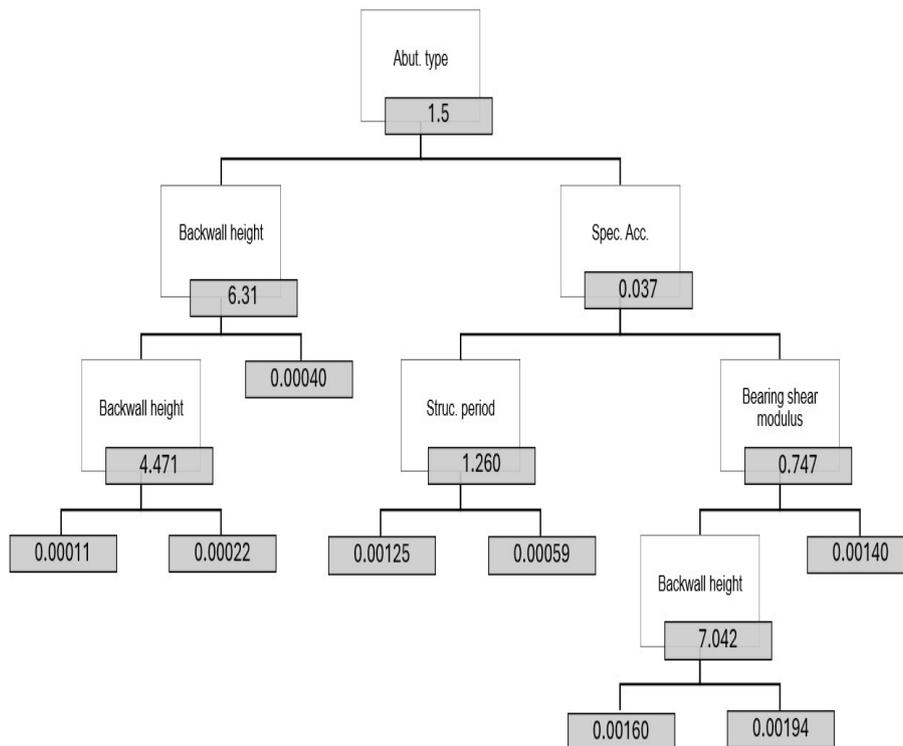


Figura 83 – ADT - Modelo esquemático da primeira árvore de decisão do Random Forest (Fonte: autor).

4.2.1.5 Deslocamento longitudinal dos aparelhos de apoio

Para o parâmetro deslocamento longitudinal dos aparelhos de apoio (BDL), o modelo com melhor performance foi o *Decision Tree*. A Tabela 31 a seguir demonstra os resultados obtidos para o deslocamento longitudinal dos aparelhos de apoio para os modelos do conjunto de treinamento através do *Root Mean Square Error*.

De acordo com a tabela de verificação da performance, o algoritmo *Decision Tree* teve o melhor desempenho entre todos os algoritmos. A Tabela 32 a seguir demonstra os parâmetros e as características do algoritmo e a velocidade de processamento.

Tabela 31 – BDL - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	0,1457
	<i>Quadratic</i>	0,1430
	<i>Cubic</i>	0,1399
	<i>Fine Gaussian</i>	0,1452
	<i>Medium Gaussian</i>	0,1470
	<i>Coarse Gaussian</i>	0,1506
<i>Linear Regression</i>	<i>Linear</i>	0,1705
	<i>Interactions Linear</i>	0,6546
	<i>Robust Linear</i>	0,1621
<i>Decision Tree</i>	<i>Fine</i>	0,0740
	<i>Medium</i>	0,1304
	<i>Coarse</i>	0,1510
<i>Random Forest</i>	<i>Bagged</i>	0,1253
	<i>Boosted</i>	0,1422

Tabela 32 – BDL - *training set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0740
R^2	0,76
MSE	0,0054
MAE	0,0428
<i>Prediction Speed</i>	Aprox. 3400 obs/sec
<i>Training time</i>	0,38 sec
<i>Minimum leaf size</i>	4

Conforme os dados expostos na tabela do deslocamento longitudinal dos aparelhos de apoio para o *training set*, o modelo teve uma performance razoável, de maneira que o parâmetro R^2 ficou aproximadamente igual a 0,75. Isso significa que o modelo explica aproximadamente 75% da variância do deslocamento longitudinal dos aparelhos de apoio.

A Tabela 33 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 33 – BDL - *testing set* - *Decision Tree* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0527
R^2	0,92
MSE	0,0028
MAE	0,0344

Nota-se que o modelo performou melhor para o *testing set*, logo o modelo não cometeu *overfitting*.

A seguir, as Figuras 84 à 89 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

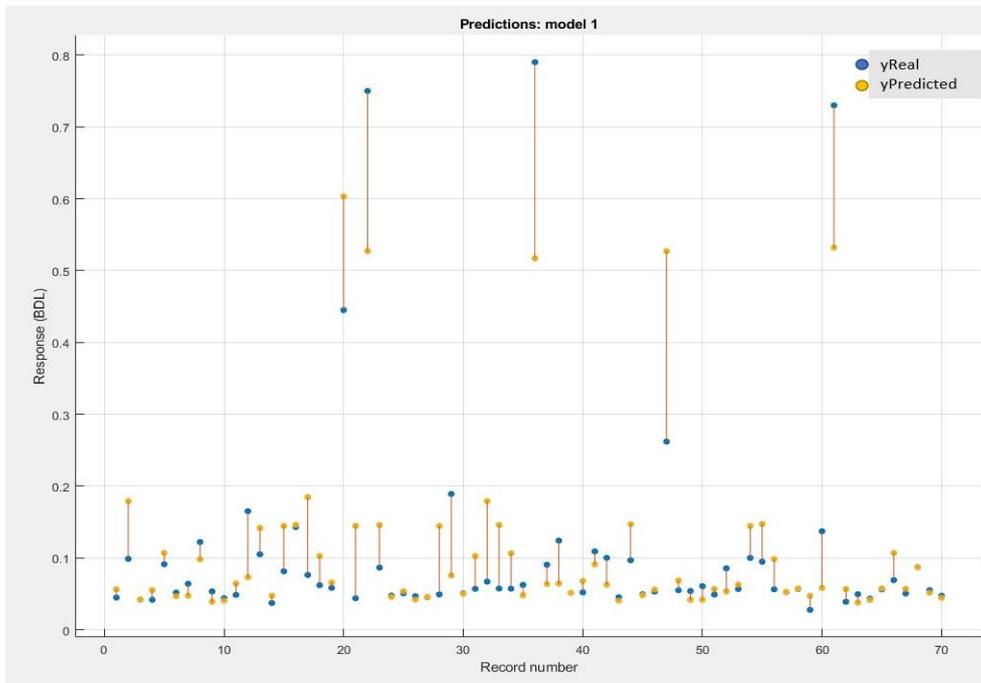


Figura 84 – BDL - *Response plot* - *training set* (Fonte: autor).

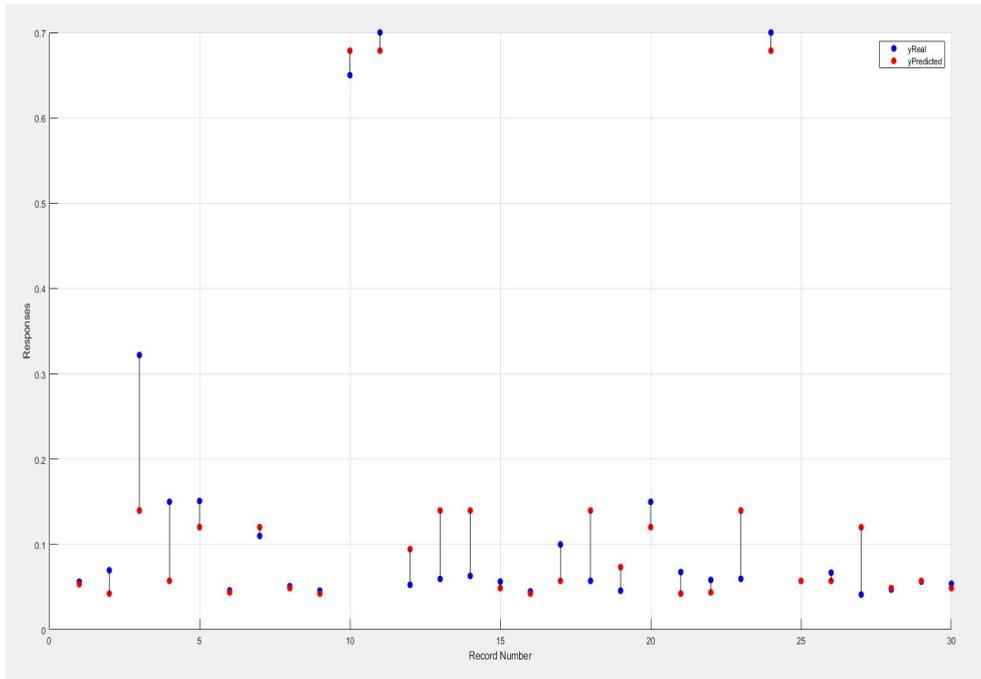


Figura 85 – BDL - Response plot - testing set (Fonte: autor).

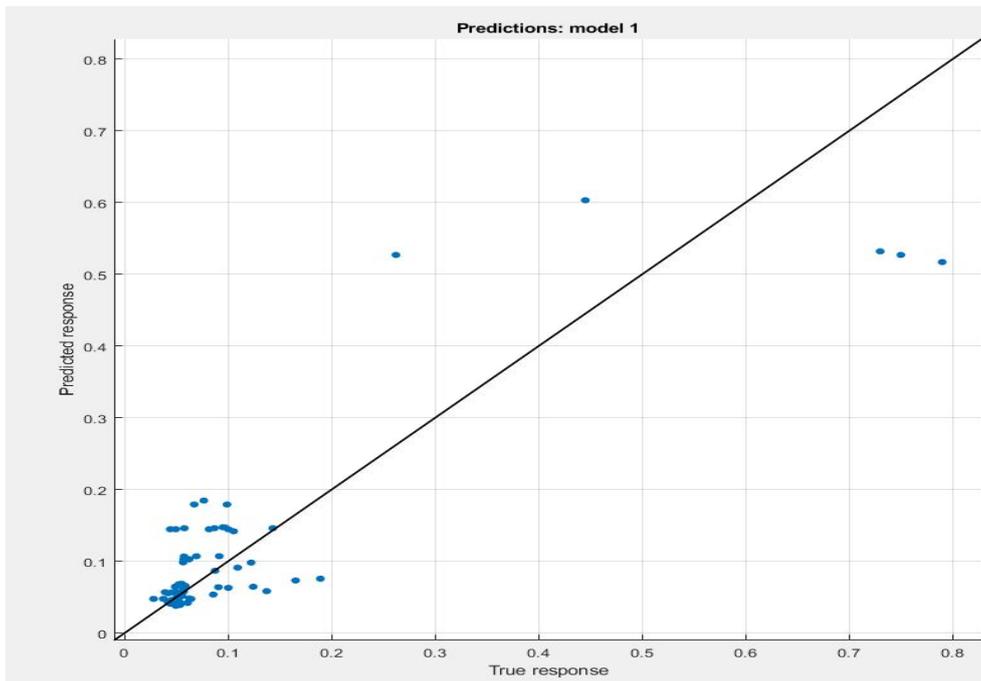


Figura 86 – BDL - Predicted vs. Actual plot - training set (Fonte: autor).

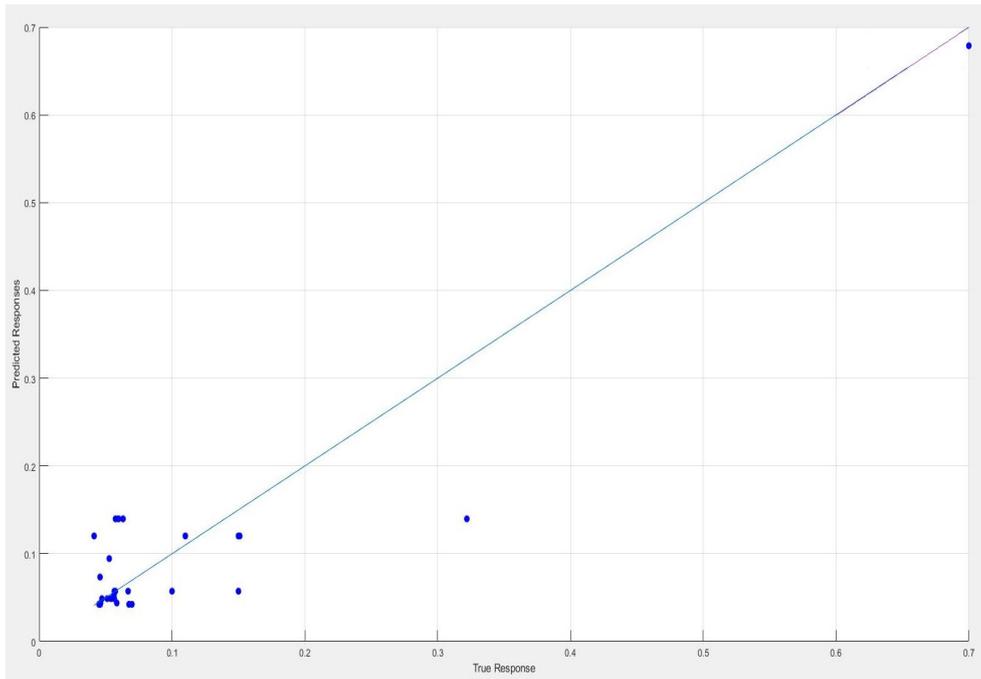


Figura 87 – BDL - Predicted vs. Actual plot - testing set (Fonte: autor).

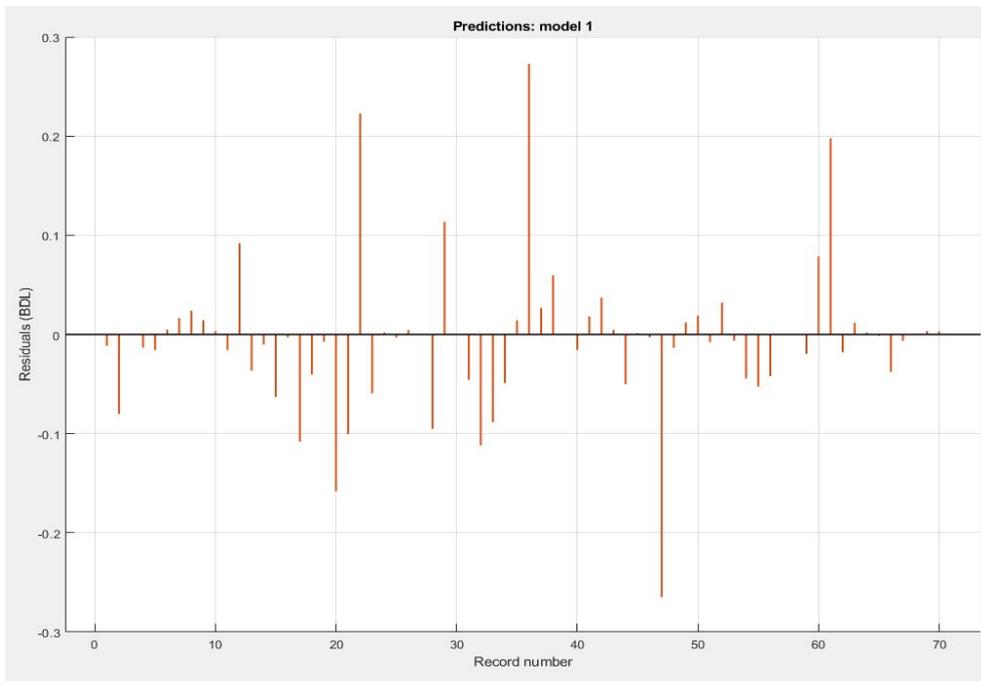


Figura 88 – BDL - Residual plot - training set (Fonte: autor).

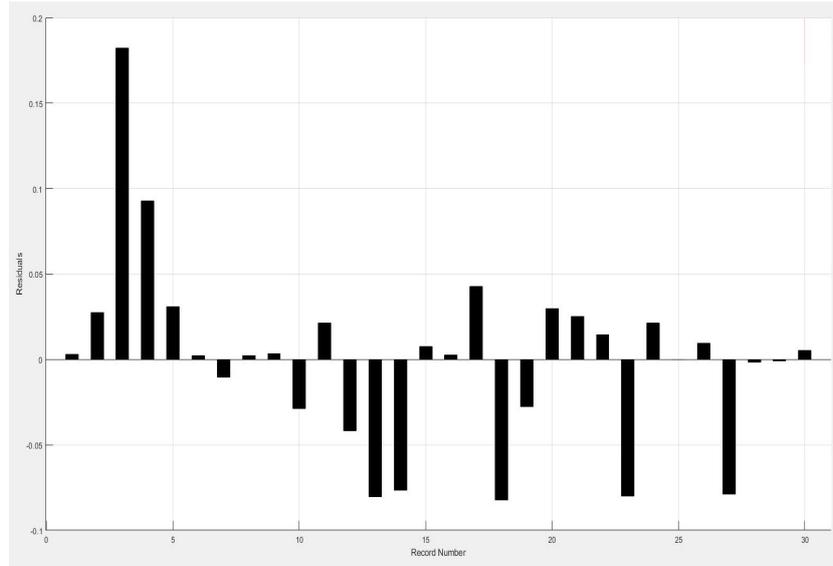


Figura 89 – BDL - Residual plot - testing set (Fonte: autor).

Pode-se observar a partir das análises gráficas que o modelo de regressão performou melhor para o *testing set* quando comparado com o *training set*. A Figura 90 ilustra o esquema da árvore de decisão utilizada pelo algoritmo. É possível visualizar que a variável aleatória “*Backwall height*”, que se refere a altura dos encontros, é a mais utilizada no algoritmo.

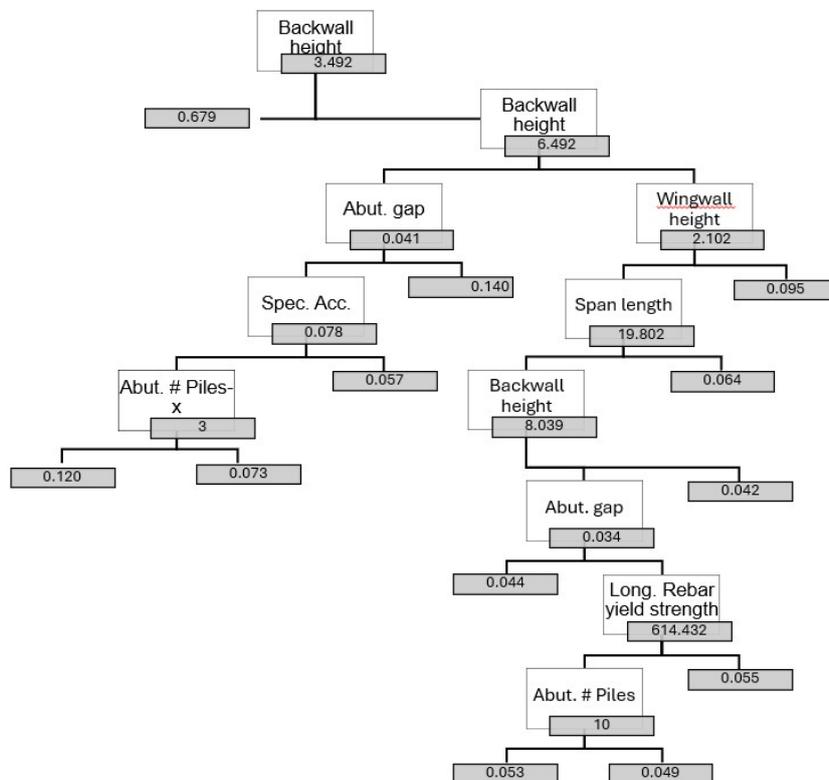


Figura 90 – BDL - Modelo esquemático do algoritmo Decision Tree (Fonte: autor).

4.2.1.6 Deslocamento transversal dos aparelhos de apoio

Para o parâmetro deslocamento transversal dos aparelhos de apoio (BDT), o modelo com melhor performance foi o *Random Forest*. A Tabela 34 a seguir demonstra os resultados obtidos para o deslocamento transversal dos aparelhos de apoio para os modelos do conjunto de treinamento através do *Root Mean Square Error*.

Tabela 34 – BDT - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	0,0150
	<i>Quadratic</i>	0,0150
	<i>Cubic</i>	0,0147
	<i>Fine Gaussian</i>	0,0190
	<i>Medium Gaussian</i>	0,0142
	<i>Coarse Gaussian</i>	0,0167
<i>Linear Regression</i>	<i>Linear</i>	0,0189
	<i>Interactions Linear</i>	0,0880
	<i>Robust Linear</i>	0,0222
<i>Decision Tree</i>	<i>Fine</i>	0,0151
	<i>Medium</i>	0,0144
	<i>Coarse</i>	0,0189
<i>Random Forest</i>	<i>Bagged</i>	0,0128
	<i>Boosted</i>	0,0130

De acordo com a tabela de verificação da performance, o algoritmo *Random Forest* - *Bagged Trees* teve o melhor desempenho entre todos os algoritmos. A Tabela 35 a seguir demonstra os parâmetros e as características do algoritmo e a velocidade de processamento.

Tabela 35 – BDT - *training set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0128
R^2	0,54
MSE	0,00016
MAE	0,0100
<i>Prediction Speed</i>	Aprox. 960 obs/sec
<i>Training time</i>	1,1043 sec
<i>Minimum leaf size</i>	8
<i>Number of learners</i>	30

Conforme os dados expostos na tabela do deslocamento transversal dos aparelhos de apoio para o *training set*, o modelo teve uma performance ruim, de maneira que o parâmetro R^2 ficou aproximadamente igual a 0,55. Isso significa que o modelo explica aproximadamente 55% da variância do deslocamento transversal dos aparelhos de apoio. A Tabela 36 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 36 – BDT - Deslocamento transversal dos aparelhos de apoio - *testing set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0091
R^2	0,32
MSE	0,0308
MAE	0,0075

Nota-se que o modelo performou pior para o *testing set* do que comparado com *training set*, logo ocorreu *overfitting*.

A seguir, as Figuras 91 à 96 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

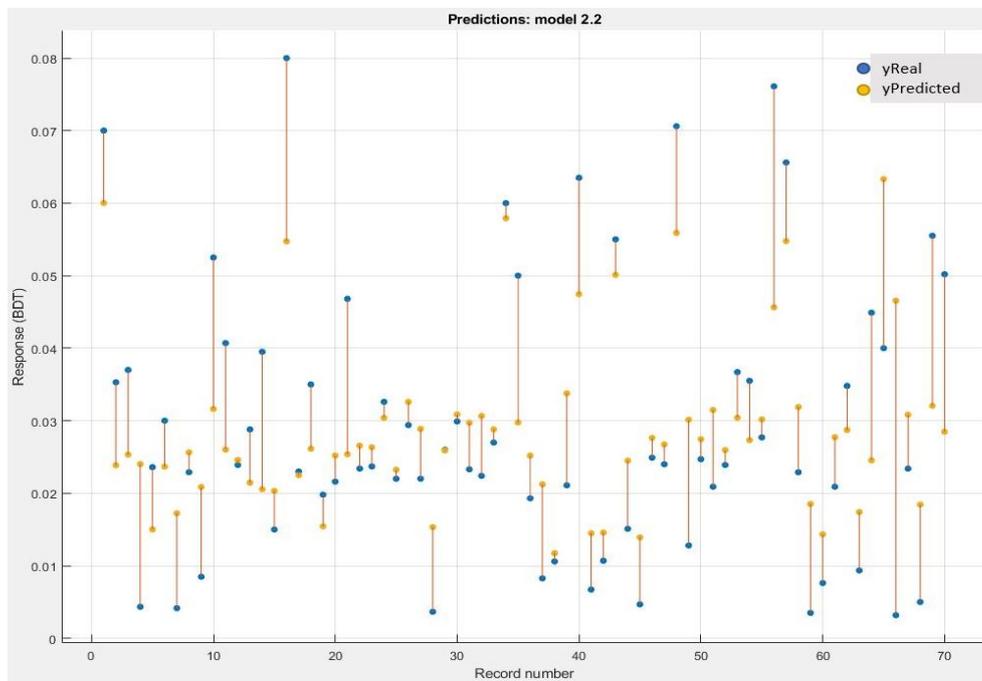


Figura 91 – BDT - *Response plot* - *training set* (Fonte: autor).

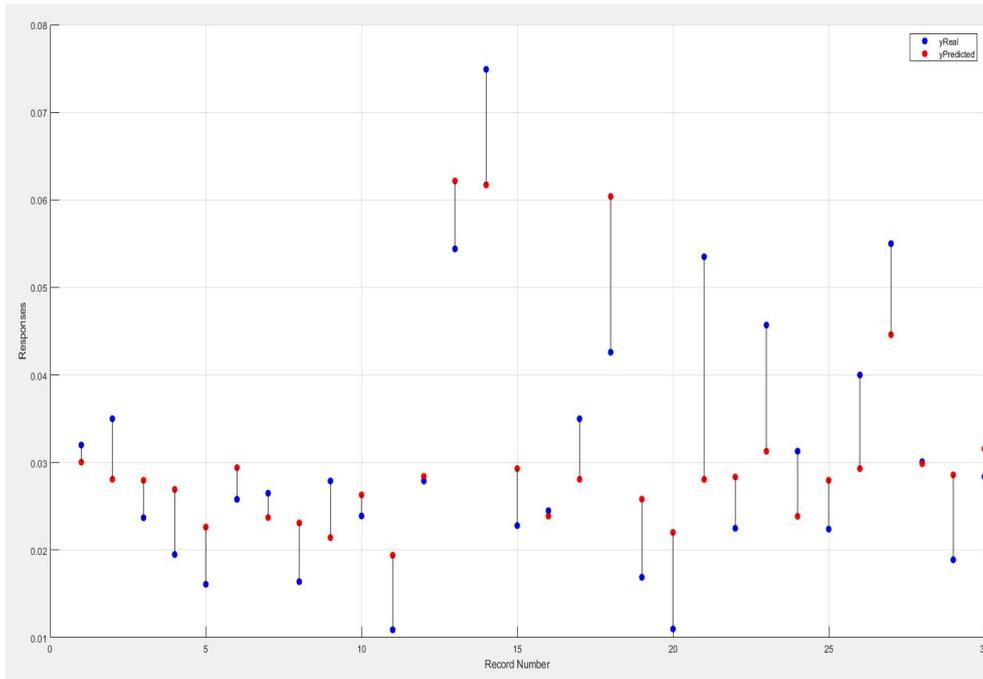


Figura 92 – BDT - Response plot - testing set (Fonte: autor).

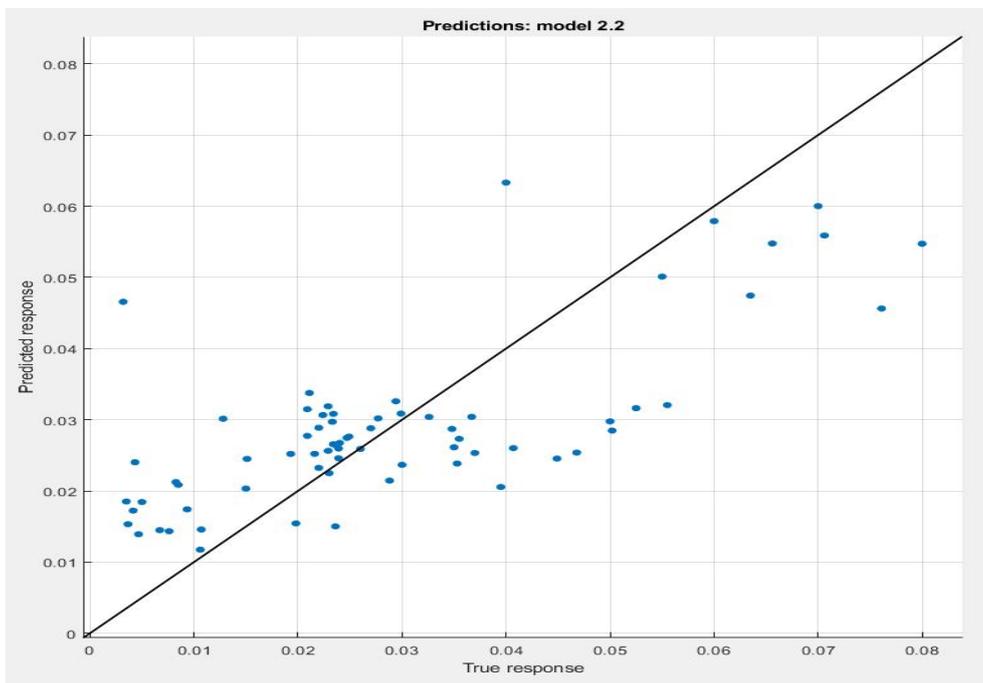


Figura 93 – BDT - Predicted vs. Actual plot - training set (Fonte: autor).

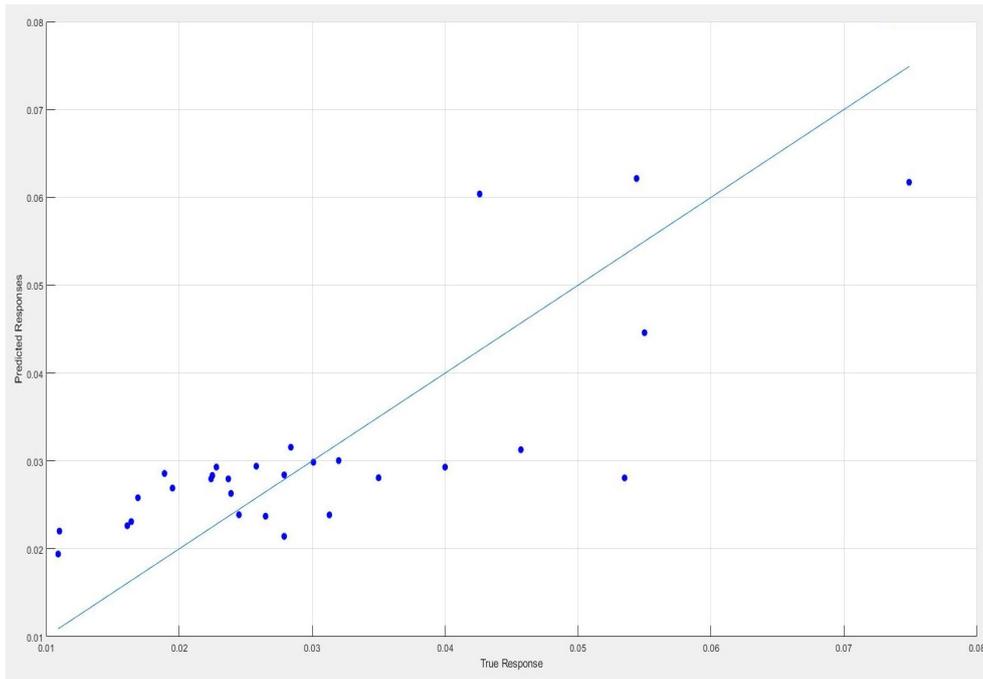


Figura 94 – BDT - *Predicted vs. Actual plot - testing set* (Fonte: autor).

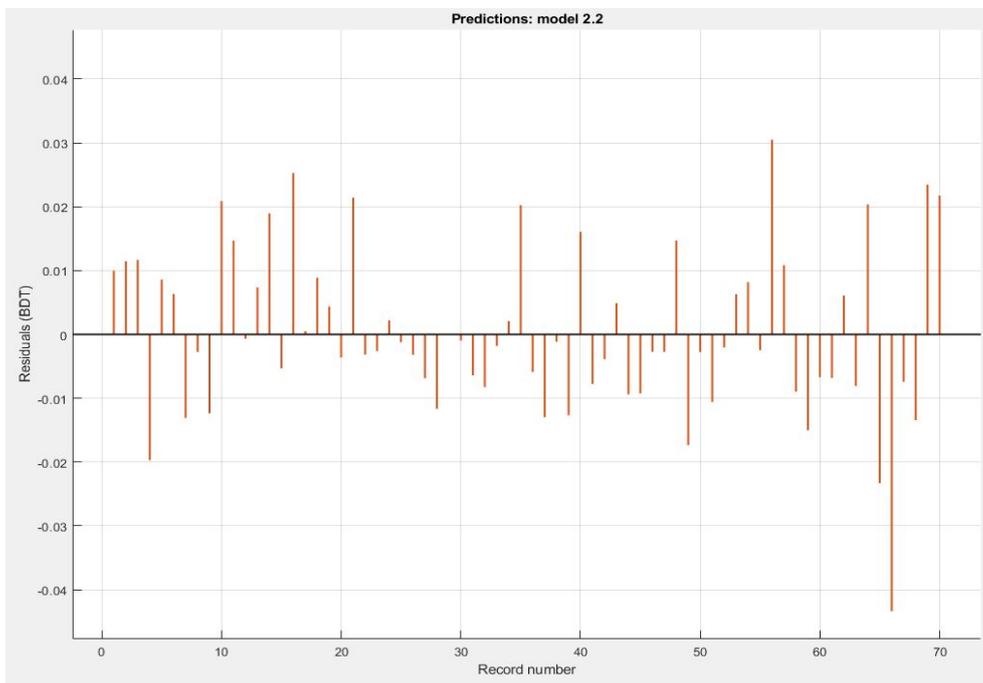


Figura 95 – BDT - *Residual plot - training set* (Fonte: autor).

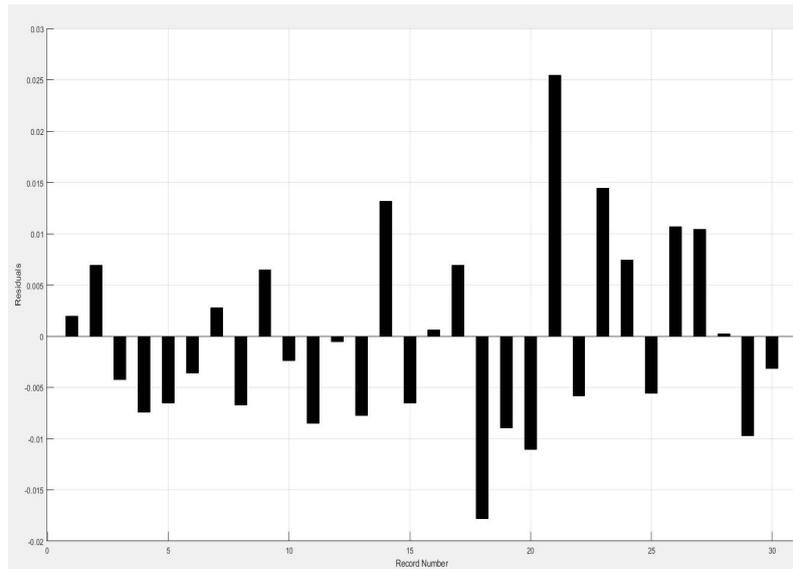


Figura 96 – BDT - Residual plot - testing set (Fonte: autor).

Pode-se observar a partir das análises gráficas que o modelo de regressão performou melhor para o *training set* quando comparado com o *testing set*. A Figura 97 ilustra a primeira árvore de decisão, entre as trinta utilizadas pelo algoritmo. É possível visualizar que as variáveis aleatória “*Lower wingwall height*” e “*Spectral acceleration*”, que se referem a altura das alas dos encontros e a aceleração espectral, respectivamente, são as mais utilizadas no algoritmo.

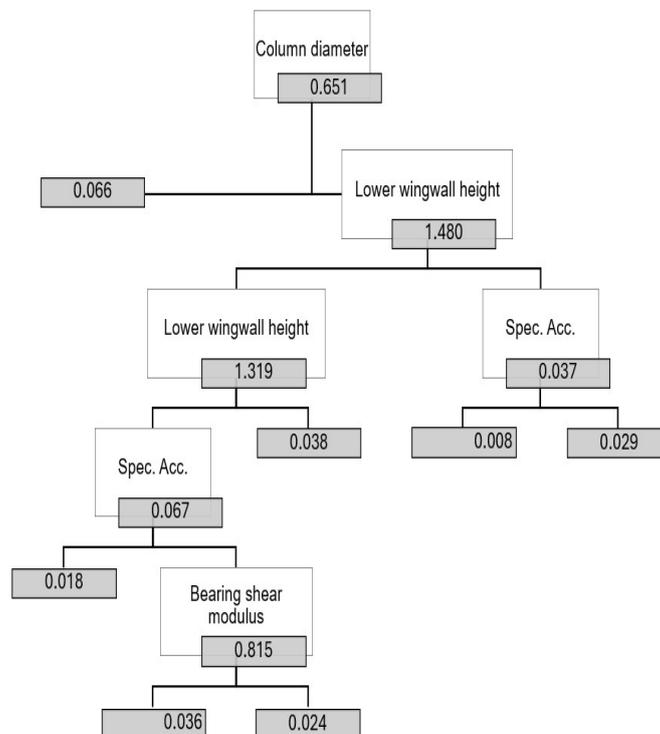


Figura 97 – BDT - Modelo esquemático do algoritmo Decision Tree (Fonte: autor).

4.2.1.7 Deslocamento do tabuleiro

Para o parâmetro deslocamento do tabuleiro (DU), o modelo com melhor performance foi o *Random Forest*. A Tabela 37 a seguir demonstra os resultados obtidos para o deslocamento do tabuleiro para os modelos do conjunto de treinamento através do *Root Mean Square Error*.

Tabela 37 – DU - Verificação da performance dos modelos de regressão através do RMSE (Fonte: autor).

Algoritmo	Tipo	RMSE
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	0,0119
	<i>Quadratic</i>	0,0110
	<i>Cubic</i>	0,0113
	<i>Fine Gaussian</i>	0,0183
	<i>Medium Gaussian</i>	0,0120
	<i>Coarse Gaussian</i>	0,0150
<i>Linear Regression</i>	<i>Linear</i>	0,0117
	<i>Interactions Linear</i>	0,0701
	<i>Robust Linear</i>	0,0120
<i>Decision Tree</i>	<i>Fine</i>	0,0118
	<i>Medium</i>	0,0121
	<i>Coarse</i>	0,0182
<i>Random Forest</i>	<i>Bagged</i>	0,0099
	<i>Boosted</i>	0,0106

De acordo com a tabela de verificação da performance, o algoritmo *Random Forest - Bagged Trees* teve o melhor desempenho entre todos os algoritmos. A Tabela 38 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 38 – DU - *training set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0099
R^2	0,70
MSE	$9,97 \cdot 10^{-5}$
MAE	0,0083
<i>Prediction Speed</i>	Aprox. 970 obs/sec
<i>Training time</i>	1,1268 sec
<i>Minimum leaf size</i>	8
<i>Number of learners</i>	30

Conforme os dados expostos na tabela do deslocamento do tabuleiro para o *training set*, o modelo teve uma performance mediana, de maneira que o parâmetro R^2 ficou aproximadamente igual a 0,70. Isso significa que o modelo explica aproximadamente 70% da variância do deslocamento do tabuleiro. A Tabela 39 a seguir demonstra os parâmetros numéricos calculados pelo modelo de regressão escolhido para o *testing set*.

Tabela 39 – DU - *testing set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
RMSE	0,0059
R^2	0,72
MSE	$3,48 \cdot 10^{-5}$
MAE	0,0051

Nota-se que o modelo performou de maneira similar para o *testing set*, logo o modelo não cometeu *overfitting*.

A seguir, as Figuras 98 à 103 ilustram os resultados gráficos para o conjunto de treinamento e para o conjunto de teste.

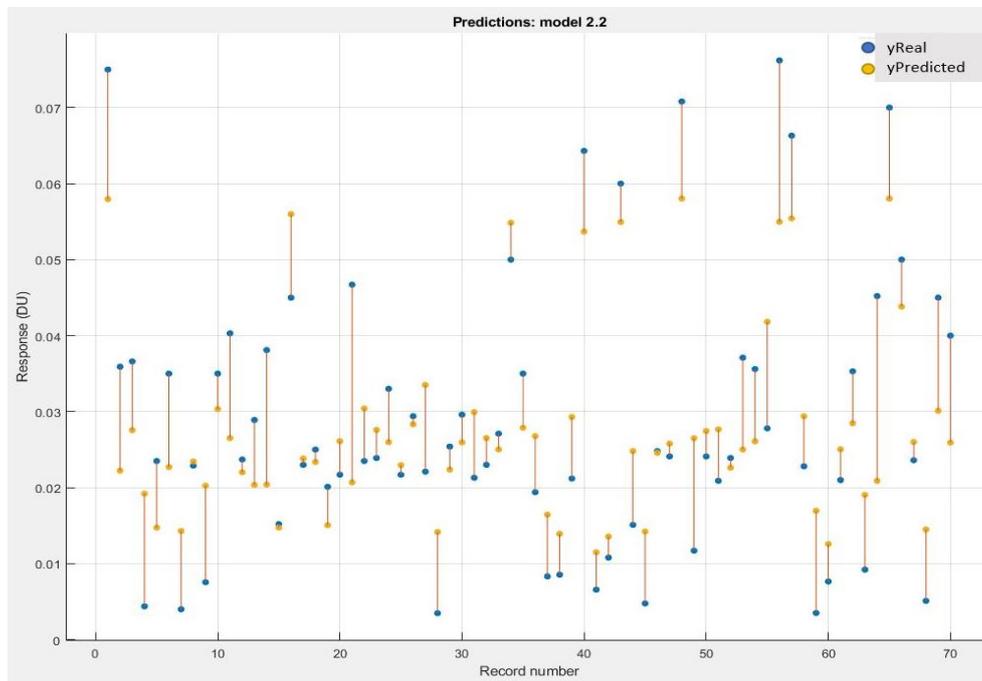


Figura 98 – DU - *Response plot* - *training set* (Fonte: autor).

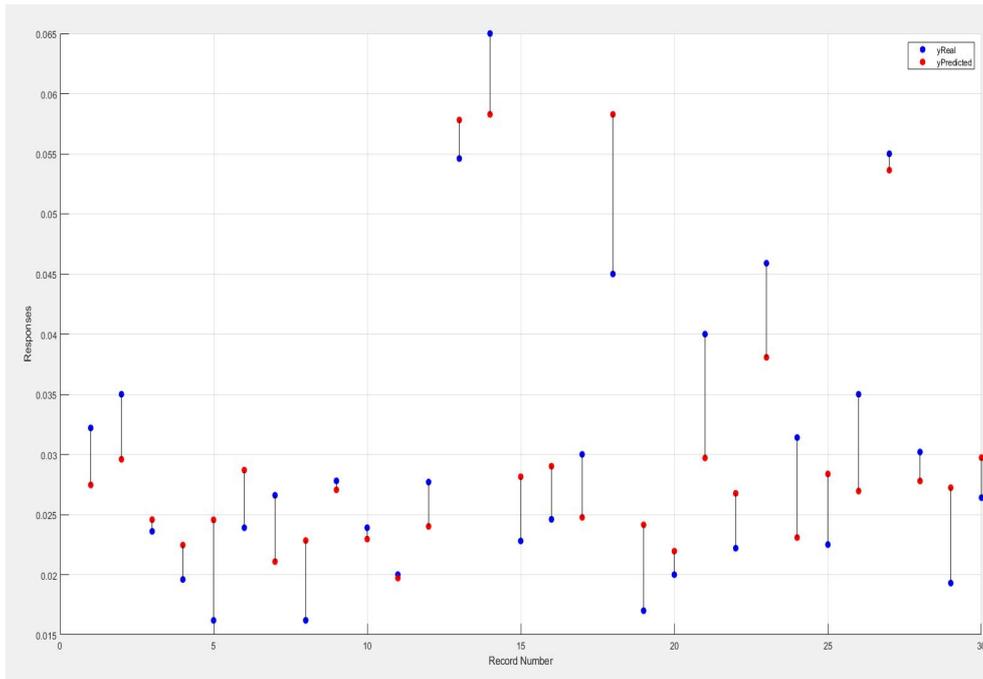


Figura 99 – DU - Response plot - testing set (Fonte: autor).

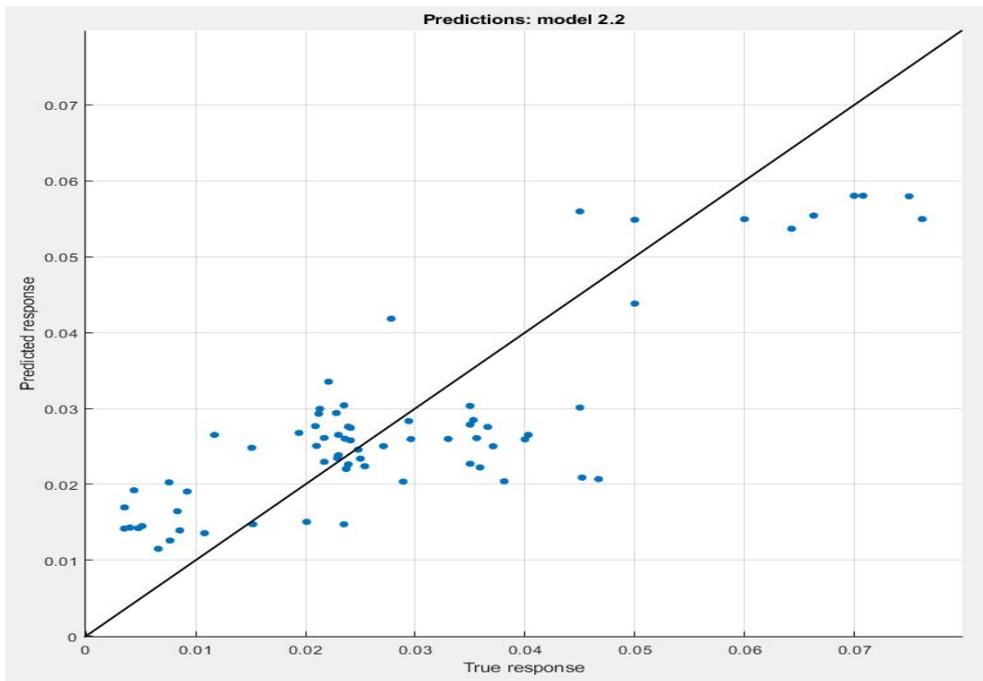


Figura 100 – DU - Predicted vs. Actual plot - training set (Fonte: autor).

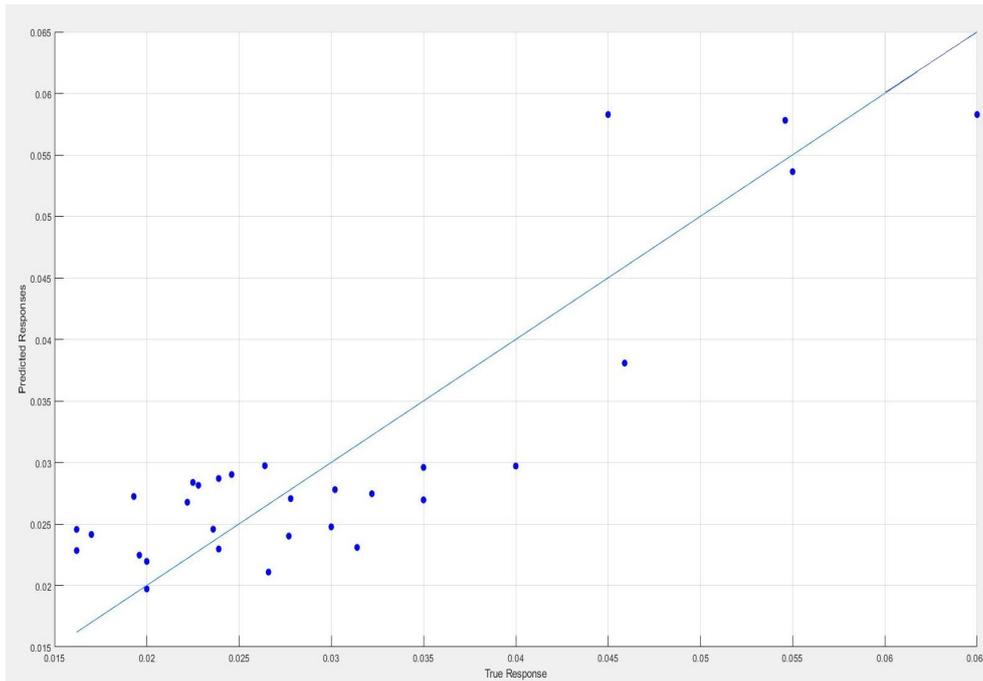


Figura 101 – DU - Predicted vs. Actual plot - testing set (Fonte: autor).

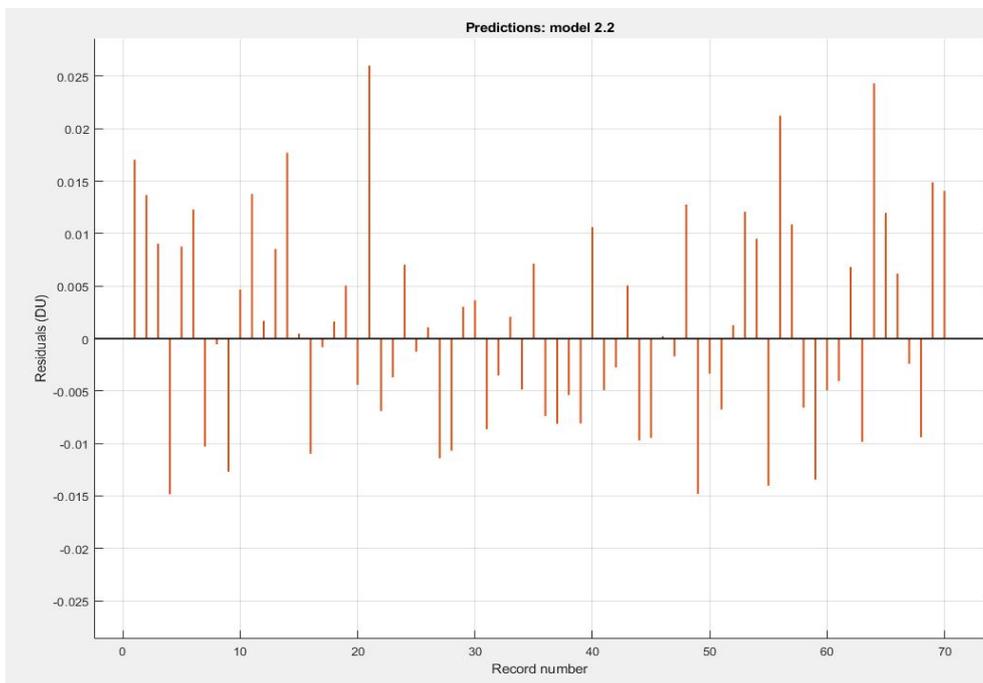


Figura 102 – DU - Residual plot - training set (Fonte: autor).

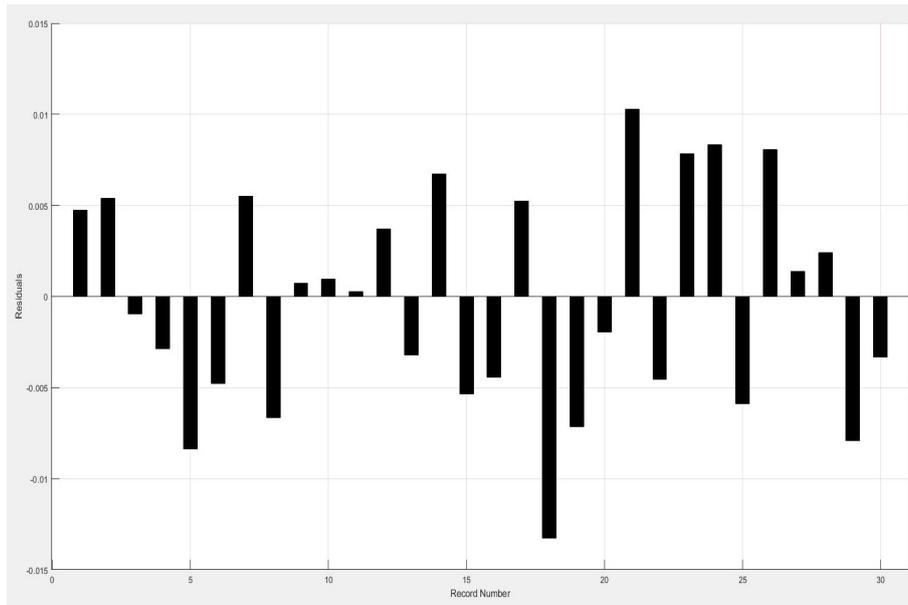


Figura 103 – DU - *Residual plot - testing set* (Fonte: autor).

Pode-se observar a partir das análises gráficas que o modelo de regressão performou de maneira similar para o *training set* e para o *testing set*. A Figura 104 ilustra a primeira árvore de decisão, entre as trinta utilizadas pelo algoritmo. É possível visualizar que as variáveis aleatórias utilizadas no algoritmo não apresentam nenhuma preferência.

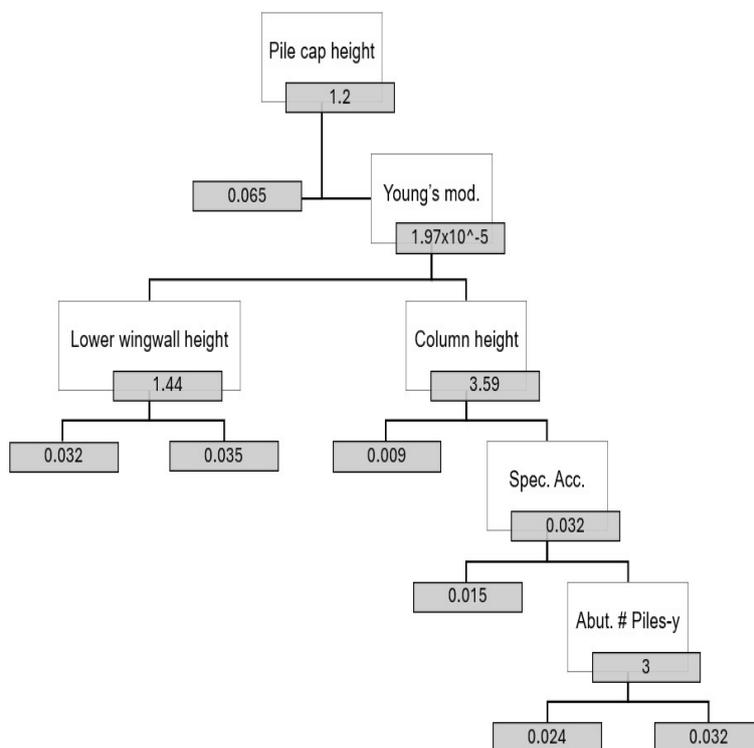


Figura 104 – DU - Modelo esquemático do algoritmo *Decision Tree* (Fonte: autor).

4.2.1.8 Resumo dos resultados obtidos para os parâmetros de engenharia - modelo de regressão

A tabela 40 a seguir demonstra o resumo dos resultados obtidos para os parâmetros de engenharia para o modelo de regressão para o conjunto de treinamento e de teste.

Tabela 40 – Resumo dos resultados obtidos - modelo de regressão (Fonte: autor).

EDP	Conjunto	Algoritmo	RMSE	R^2
CC	Treino	<i>Decision Tree</i>	$1,59 \cdot 10^{-5}$	100%
	Teste	<i>Decision Tree</i>	$2,62 \cdot 10^{-5}$	98%
ADA	Treino	<i>Decision Tree</i>	$5,11 \cdot 10^{-2}$	78%
	Teste	<i>Decision Tree</i>	$6,32 \cdot 10^{-2}$	78%
ADP	Treino	<i>Decision Tree</i>	$4,18 \cdot 10^{-2}$	51%
	Teste	<i>Decision Tree</i>	$4,52 \cdot 10^{-2}$	51%
ADT	Treino	<i>Random Forest</i>	$3,05 \cdot 10^{-4}$	81%
	Teste	<i>Random Forest</i>	$2,85 \cdot 10^{-4}$	80%
BDL	Treino	<i>Decision Tree</i>	$7,40 \cdot 10^{-2}$	76%
	Teste	<i>Decision Tree</i>	$5,27 \cdot 10^{-2}$	92%
BDT	Treino	<i>Random Forest</i>	$1,28 \cdot 10^{-2}$	54%
	Teste	<i>Random Forest</i>	$9,10 \cdot 10^{-3}$	32%
DU	Treino	<i>Random Forest</i>	$9,90 \cdot 10^{-3}$	70%
	Teste	<i>Random Forest</i>	$5,90 \cdot 10^{-3}$	72%

4.2.1.9 Curvas de fragilidade - resultados do modelo de regressão

De maneira similar ao procedimento realizado no item dos resultados estocásticos, com os resultados do modelo de regressão é possível aplicar o Modelo Probabilístico de Demanda Sísmica (PSDM). A Tabela 41 abaixo demonstra de maneira resumida os resultados dos Modelos Probabilísticos de Demanda Sísmica para cada Parâmetro de Demanda de Engenharia.

Tabela 41 – Resultados do PSDM para cada Parâmetro de Engenharia após modelo de regressão (Fonte: autor).

Parâmetros de Demanda	$\ln(a)$	b	Dispersão	R^2
Deslocamento do tabuleiro (DU) (mm)	5,92	0,95	0,46	0,45
Curvatura dos pilares (CC) (1/m)	-5,51	1,07	0,58	0,57
Deslocamento long. dos aparelhos de apoio (BDL) (mm)	4,04	0,10	1,55	0,20
Deslocamento transv. dos aparelhos de apoio (BDT) (mm)	5,90	0,94	0,46	0,65
Deslocamento ativo dos encontros (ADA) (mm)	4,43	0,10	1,80	0,15
Deslocamento passivo dos encontros (ADP) (mm)	4,46	0,10	1,82	0,24
Deslocamento transv. dos encontros (ADT) (mm)	1,09	0,47	0,6	0,59

Da mesma maneira que os resultados estocásticos, e de acordo com a tabela, pode-se verificar que, para alguns parâmetros de demanda, o valor de dispersão foi alto, resultando em valores baixos de R^2 .

Após obter os parâmetros de demanda através da regressão do PSDM para o modelo de regressão do ML, e conhecer os parâmetros de capacidade, é possível conduzir a análise de risco dinâmico através das curvas de fragilidade. A Figura 105 a seguir demonstra as curvas de fragilidade para o Parâmetros de Demanda de Engenharia para cada estado-limite de dano.

De acordo com a Figura 105, pode-se verificar que, para os estados-limites de dano leve e moderado, o deslocamento transversal dos aparelhos de apoio é o parâmetro mais crítico. Já para os limites de dano extensivo e completo é o deslocamento do tabuleiro que determina a fragilidade do sistema estrutural.

A Figura 106 a seguir demonstra a curva de fragilidade para o sistema estrutural para cada estado-limite de dano.

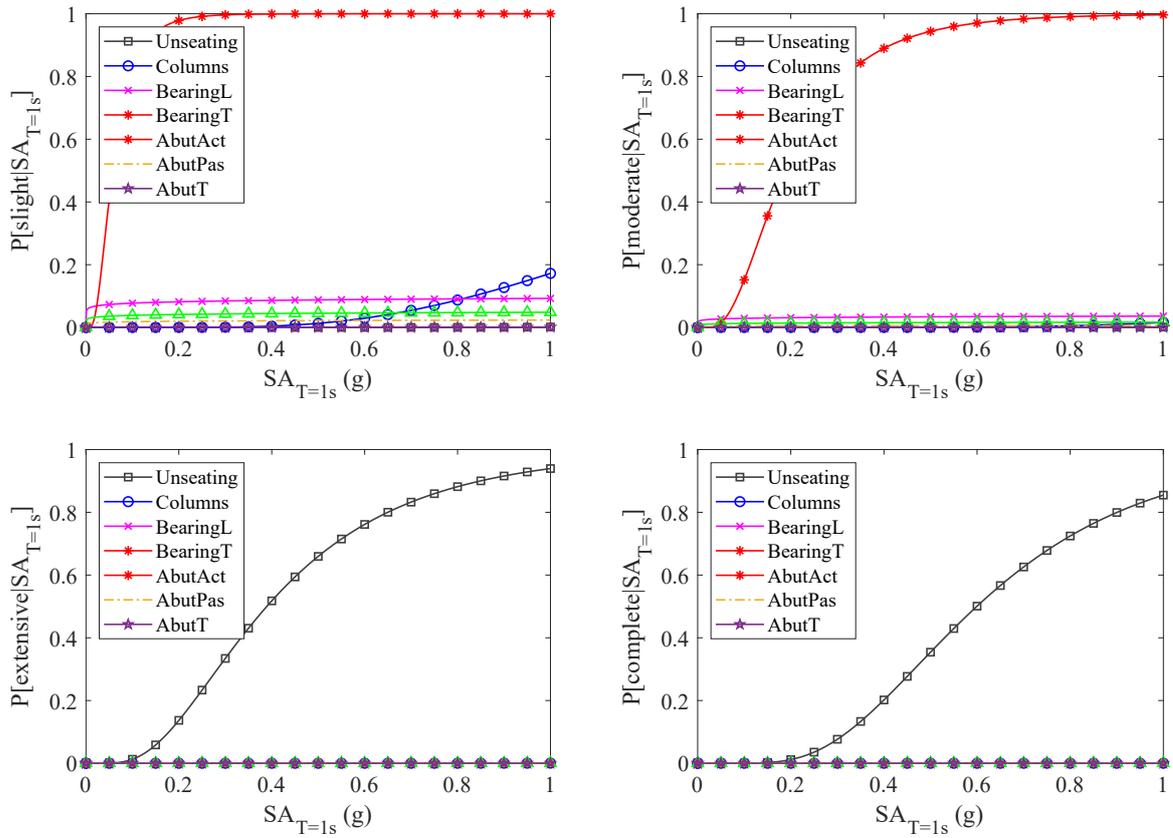


Figura 105 – Curvas de fragilidade para cada estado-limite de dano após modelo de regressão (Fonte: autor).

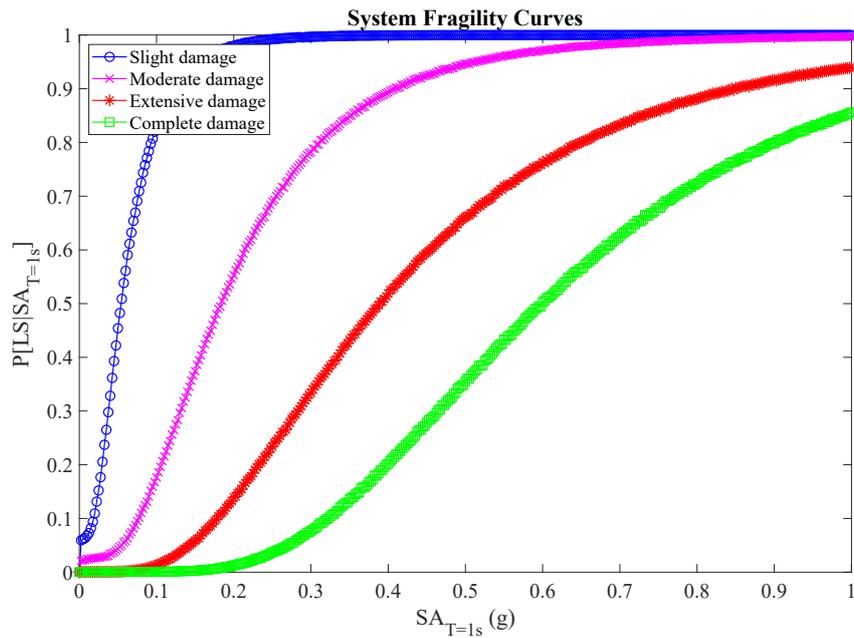


Figura 106 – Curva de fragilidade do sistema estrutural para cada estado-limite de dano após modelo de regressão do ML (Fonte: autor).

4.2.2 Resultados da classificação

Os modelos de classificação foram treinados para todos os parâmetros de demanda de engenharia, exceto a curvatura dos pilares (CC) e deslocamento transversal dos encontros (ADT), uma vez que, durante o processo de categorização das classes, nenhum valor de resposta estrutural superou o limite mínimo de 0,007 (1/m) para a curvatura e 25 (mm) para o deslocamento transversal. Para os cinco parâmetros restantes, considerando que as resposta sísmicas foram agrupadas com os modelos de OAEs, foram realizadas 500 análises.

Similar ao que foi exposto na seção dos resultados dos modelos de regressão, caso o algoritmo com melhor performance seja do tipo “*Tree*”, seja *Decision Tree* ou *Random Forest*, será ilustrada a árvore de decisão. Para os casos em que o algoritmo escolhido foi o *Random Forest*, será exibido apenas a primeira árvore de decisão.

4.2.2.1 Deslocamento ativo dos encontros

Para o parâmetro deslocamento ativo dos encontros (ADA), o modelo com melhor performance foi o *Random Forest*. A Tabela 42 a seguir demonstra os resultados obtidos para o deslocamento ativo dos encontros para os algoritmos do *training set* através do *Prediction Accuracy* (PA). Vale ressaltar que o modelo escolhido é o que possui a melhor acurácia total e a menor quantidade de erros de classificação, análise feita através das matrizes de confusão, conforme exposto no Capítulo 3.

De acordo com a tabela de verificação de performance, os algoritmos *Support Vector Machine - Linear e Medium Gaussian*, *K-nearest Neighbor - weighted* e *Random Forest - bagged* tiveram os melhores desempenhos entre todos os algoritmos. No entanto, o modelo que apresentou menos erros de classificação foi o *Random Forest*.

Tabela 42 – ADA - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).

Algoritmo	Tipo	PA
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	90,0%
	<i>Quadratic</i>	84,3%
	<i>Cubic</i>	82,9%
	<i>Fine Gaussian</i>	75,7%
	<i>Medium Gaussian</i>	90,0%
	<i>Coarse Gaussian</i>	75,7%
<i>K-nearest Neighbor</i> (KNN)	<i>Fine</i>	82,9%
	<i>Medium</i>	87,1%
	<i>Coarse</i>	75,7%
	<i>Cosine</i>	87,1%
	<i>Cubic</i>	88,6%
<i>Discriminant Analysis</i>	<i>Linear</i>	87,1%
	<i>Boosted</i>	<i>failed</i>
<i>Regression</i>	<i>Logistic</i>	87,0%
<i>Decision Tree</i>	<i>Fine</i>	85,7%
	<i>Medium</i>	85,7%
	<i>Coarse</i>	85,7%
<i>Random Forest</i>	<i>Bagged</i>	90,0%
	<i>Boosted</i>	75,7%

A Tabela 43 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 43 – ADA - *training set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
<i>Prediction Speed</i>	Aprox. 230 obs/sec
<i>Training time</i>	3,8225 sec
<i>Total Accuracy</i>	90,0%
<i>Number of learners</i>	30

Pode-se observar que, de acordo com os dados apresentados na tabela do deslocamento ativo dos encontros para o *training set*, o modelo de classificação teve uma ótima performance, classificando corretamente os *TAGs* 90% das vezes. No entanto, conforme exposto na seção dos resultados dos modelos de regressão e de acordo com Hastie et al. (2001), a performance do

modelo também deve ser avaliada pelo conjunto de teste. A Tabela 44 a seguir demonstra o valor de acurácia total para o modelo de classificação escolhido para o *testing set*.

Tabela 44 – ADA - *testing set* - *Random Forest* (Fonte: autor).

Parâmetro	Resultado
Total Accuracy	73,3%

Nota-se que o modelo *Random Forest* performou pior para o *testing set*, logo o modelo, provavelmente, cometeu *overfitting*. A Figura 107 a seguir ilustra a primeira árvore de decisão, entre as 30 utilizadas pelo algoritmo.

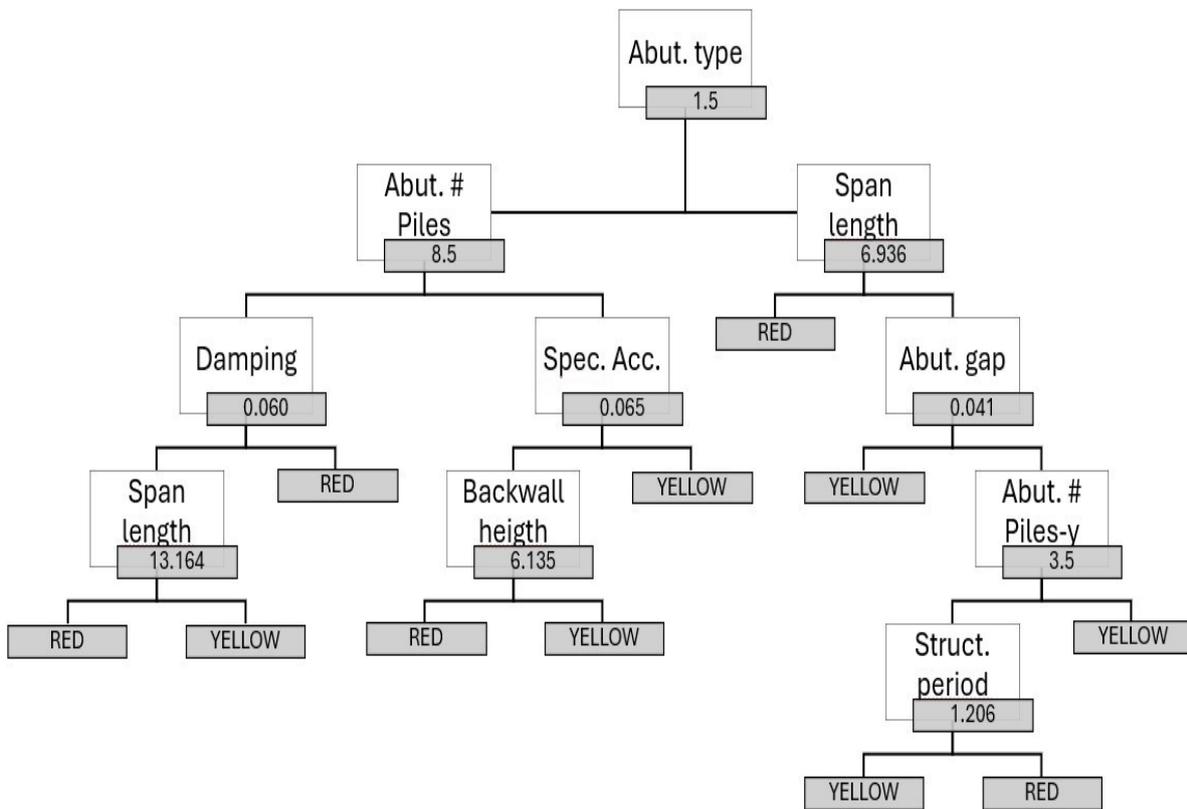


Figura 107 – ADA - Modelo esquemático do algoritmo *Decision Tree* (Fonte: autor).

Pode-se observar a partir das matrizes de confusão que os dados de resposta estrutural, tanto para o *training set* quanto para o *testing set* não apresentaram nenhum valor categorizado na classe verde. A Figura 108 a seguir demonstra a *confusion matrix* para ambos os conjuntos, treinamento e teste, respectivamente.

		Conjunto de treinamento				Conjunto de teste						
Classes reais	Red	13	4	76%	24%	Red	4	8	33%	67%		
	Yellow	3	50	94%	6%	Yellow	/		100%	/		
				TPR	FNR			TPR	FNR			
PPV		81%	93%			PPV		100%	69%			
FDR		19%	7%			FDR		/		31%		
		Red	Yellow					Red	Yellow			
		Classes previstas						Classes previstas				

Figura 108 – ADA - *Confusion Matrix* para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).

De acordo com a matriz de confusão, o modelo para o conjunto de teste cometeu mais erros de classificação, onde 67% dos dados reais categorizados como *red* foram classificados como *yellow*, ou seja, aproximadamente 70% das pontes que apresentaram deslocamento acima de 102 mm foram classificadas como tendo um deslocamento ativo máximo menor que 102 mm. Mesmo assim, o erro dos valores previstos como *yellow* é de apenas 31%. Já no caso do modelo para o conjunto de treinamento, o algoritmo cometeu mais erros para os valores previstos como *red*, contabilizando um total de 19%, quando comparado com o *testing set*, que não cometeu nenhum erro ao categorizar a mesma classe.

4.2.2.2 Deslocamento passivo dos encontros

Para o parâmetro deslocamento passivo dos encontros (ADP), o modelo com melhor performance também foi o *Random Forest*. A Tabela 45 a seguir demonstra os resultados obtidos para o deslocamento passivo dos encontros para os algoritmos do *training set* através do *Prediction Accuracy* (PA).

De acordo com a tabela de verificação de performance, o algoritmo *Random Forest - bagged* obteve o melhor desempenho entre todos os algoritmos. A Tabela 46 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 45 – ADP - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).

Algoritmo	Tipo	PA
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	75,7%
	<i>Quadratic</i>	70,0%
	<i>Cubic</i>	64,3%
	<i>Fine Gaussian</i>	55,7%
	<i>Medium Gaussian</i>	65,7%
	<i>Coarse Gaussian</i>	57,1%
<i>K-nearest Neighbor</i> (KNN)	<i>Fine</i>	62,9%
	<i>Medium</i>	54,3%
	<i>Coarse</i>	55,7%
	<i>Cosine</i>	55,7%
	<i>Cubic</i>	50,0%
<i>Discriminant</i> <i>Analysis</i>	<i>Linear</i>	72,9%
	<i>Boosted</i>	<i>failed</i>
<i>Regression</i>	<i>Logistic</i>	68,2%
<i>Decision Tree</i>	<i>Fine</i>	67,1%
	<i>Medium</i>	67,1%
	<i>Coarse</i>	71,4%
<i>Random</i> <i>Forest</i>	<i>Bagged</i>	77,1%
	<i>Boosted</i>	70,0%

Tabela 46 – ADP - *training set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
<i>Prediction Speed</i>	Aprox. 73 obs/sec
<i>Training time</i>	11,36 sec
<i>Total Accuracy</i>	77,1%
<i>Number of learners</i>	100

Pode-se observar que, de acordo com os dados apresentados na tabela do deslocamento passivo dos encontros para o *training set*, o modelo de classificação teve uma performance regular, classificando corretamente os TAGs 77% das vezes. A Tabela 47 a seguir demonstra o valor de acurácia total para o modelo de classificação escolhido para o *testing set*.

Tabela 47 – ADP - *testing set* - *Random Forest* (Fonte: autor).

Parâmetro	Resultado
<i>Total Accuracy</i>	43,3%

Nota-se que o modelo *Random Forest* teve uma péssima performance para o *testing set*, logo o modelo cometeu *overfitting*. A Figura 109 a seguir ilustra a primeira árvore de decisão, entre as cem utilizadas pelo algoritmo.

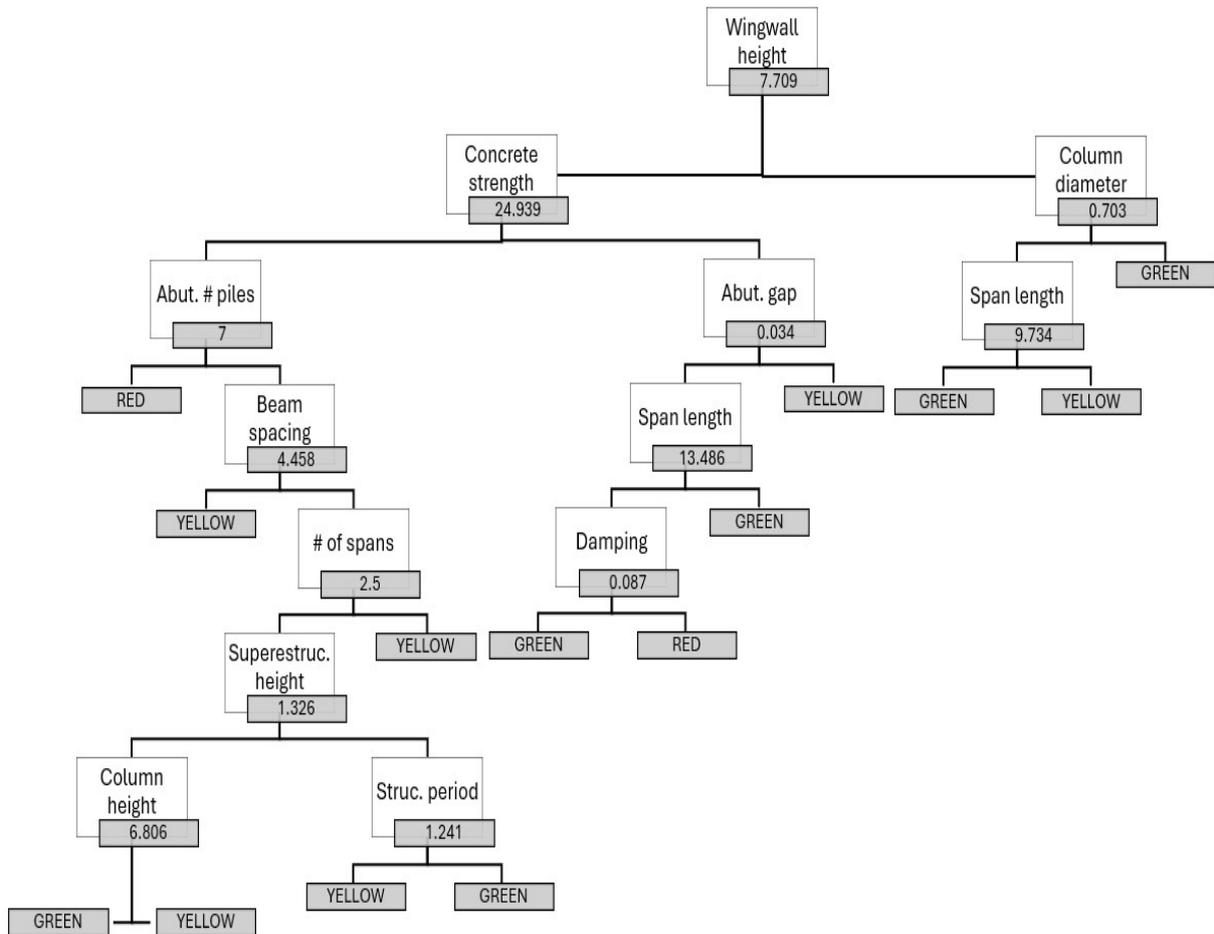


Figura 109 – ADP - Modelo esquemático do algoritmo *Decision Tree* (Fonte: autor).

Pode-se observar a partir das matrizes de confusão que os dados de resposta estrutural, tanto para o *training set* quanto para o *testing set* apresentaram todas as classes de categorização. A Figura 110 a seguir demonstra a *confusion matrix* para ambos os conjuntos, treinamento e teste, respectivamente.

		Conjunto de treinamento				Conjunto de teste						
Classes reais	Green	15		11	58%	42%	Green	9		1	90%	10%
	Red		2	3	40%	60%	Red	4	1	1	17%	83%
	Yellow	2		37	95%	5%	Yellow	11		3	21%	79%
				TPR	FNR					TPR	FNR	
PPV	88%	100%	73%			PPV	38%	100%	60%			
FDR	12%		27%			FDR	63%		40%			
		Green	Red	Yellow			Green	Red	Yellow			
		Classes Previstas					Classes Previstas					

Figura 110 – ADP - *Confusion Matrix* para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).

De acordo com a matriz de confusão, o modelo para o conjunto de teste cometeu mais erros de classificação, onde 83% dos dados reais categorizados como *red* foram classificados como *green*, ou seja, aproximadamente 83% das pontes que apresentaram deslocamento acima de 254 mm foram classificadas como tendo um deslocamento ativo máximo menor que 76 mm. Da mesma forma, 79% dos dados reais categorizados como *yellow* foram classificados como *green*, logo aproximadamente 79% das pontes que apresentaram deslocamento passivo acima de 76 mm e menor que 254 mm foram classificadas como tendo um deslocamento ativo máximo menor que 76 mm. Portanto, fica evidente que o modelo teve uma péssima performance para o *testing set*.

Já no caso do modelo para o conjunto de treinamento, o algoritmo cometeu mais erros para os valores reais categorizados como *green*, contabilizando um total de 42%, quando comparado com o *testing set*, que cometeu 10% de erro ao categorizar a mesma classe.

4.2.2.3 Deslocamento longitudinal dos aparelhos de apoio

Para o parâmetro deslocamento longitudinal dos aparelhos de apoio (BDL), o modelo com melhor performance também foi o *Random Forest*. A Tabela 48 a seguir demonstra os resultados obtidos para o deslocamento longitudinal dos aparelhos de apoio para os algoritmos do *training set* através do *Prediction Accuracy* (PA).

De acordo com a tabela de verificação de performance, o algoritmo *Random Forest - bagged* obteve o melhor desempenho entre todos os algoritmos. A Tabela 49 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 48 – BDL - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).

Algoritmo	Tipo	PA
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	91,4%
	<i>Quadratic</i>	91,4%
	<i>Cubic</i>	94,3%
	<i>Fine Gaussian</i>	77,1%
	<i>Medium Gaussian</i>	88,6%
	<i>Coarse Gaussian</i>	77,1%
<i>K-nearest Neighbor</i> (KNN)	<i>Fine</i>	91,4%
	<i>Medium</i>	85,7%
	<i>Coarse</i>	77,1%
	<i>Cosine</i>	85,7%
	<i>Cubic</i>	84,3%
<i>Discriminant</i> <i>Analysis</i>	<i>Linear</i>	92,9%
	<i>Boosted</i>	<i>failed</i>
<i>Regression</i>	<i>Logistic</i>	91,4%
<i>Decision Tree</i>	<i>Fine</i>	85,7%
	<i>Medium</i>	85,7%
	<i>Coarse</i>	85,7%
<i>Random</i> <i>Forest</i>	<i>Bagged</i>	95,0%
	<i>Boosted</i>	77,1%

Tabela 49 – BDL - *training set* - *Random Forest* (Fonte: autor).

Parâmetros	Resultados
<i>Prediction Speed</i>	Aprox. 210 obs/sec
<i>Training time</i>	4,18 sec
<i>Total Accuracy</i>	95,0%
<i>Number of learners</i>	30

Pode-se observar que, de acordo com os dados apresentados na tabela do deslocamento longitudinal dos aparelhos de apoio para o *training set*, o modelo de classificação teve uma ótima performance, classificando corretamente os *TAGs* 95% das vezes. A Tabela 50 a seguir demonstra o valor de acurácia total para o modelo de classificação escolhido para o *testing set*.

Tabela 50 – BDL - *testing set* - *Random Forest* (Fonte: autor).

Parâmetro	Resultado
<i>Total Accuracy</i>	83,3%

Nota-se que o modelo *Random Forest* teve uma excelente performance para o *testing set*, logo o modelo não cometeu *overfitting*. A Figura 111 a seguir ilustra a primeira árvore de decisão, entre as cem utilizadas pelo algoritmo.

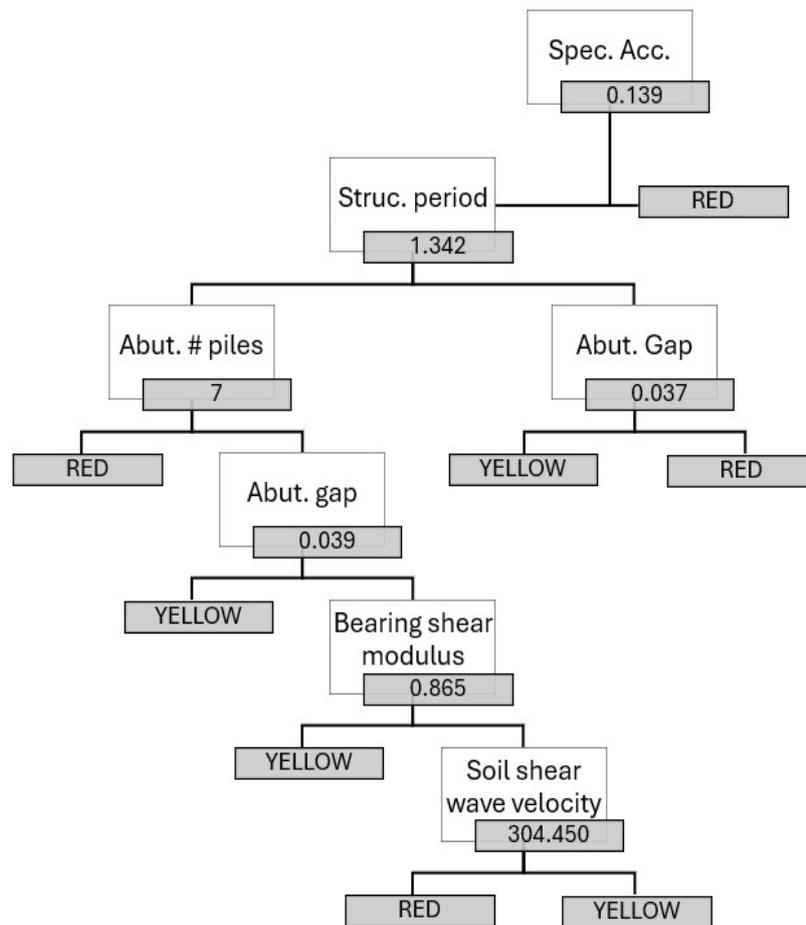


Figura 111 – BDL - Modelo esquemático do algoritmo *Decision Tree* (Fonte: autor).

Pode-se observar a partir das matrizes de confusão que os dados de resposta estrutural, tanto para o *training set* quanto para o *testing set* apresentaram apenas as classes de categorização *red* e *yellow*. A Figura 112 a seguir demonstra a *confusion matrix* para ambos os conjuntos, treinamento e teste, respectivamente.

		Conjunto de treinamento				Conjunto de teste					
Classes reais	Red	14	2	88%	13%	Red	10	2	83%	17%	
	Yellow	2	52	96%	4%	Yellow	3	15	83%	17%	
				TPR	FNR			TPR	FNR		
PPV		88%	96%			PPV		77%	88%		
FDR		13%	4%			FDR		23%	12%		
		Red	Yellow					Red	Yellow		
		Classes previstas						Classes previstas			

Figura 112 – BDL - *Confusion Matrix* para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).

De acordo com a matriz de confusão, o modelo para o conjunto de teste cometeu mais erros de classificação, onde 17% dos dados reais categorizados como *red* e *yellow* foram classificados como *yellow* e *red*, respectivamente. Mesmo assim, os erros dos valores previstos como *yellow* e *red* são de apenas 12% e 23%, respectivamente.

4.2.2.4 Deslocamento transversal dos aparelhos de apoio

Para o parâmetro deslocamento transversal dos aparelhos de apoio (BDT), o modelo com melhor performance foi o *Support Vector Machine*. A Tabela 51 a seguir demonstra os resultados obtidos para o deslocamento transversal dos aparelhos de apoio para os algoritmos do *training set* através do *Prediction Accuracy* (PA).

De acordo com a tabela de verificação de performance, o algoritmo *Support Vector Machine* obteve o melhor desempenho entre todos os algoritmos. A Tabela 52 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 51 – BDT - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).

Algoritmo	Tipo	PA
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	71,4%
	<i>Quadratic</i>	61,4%
	<i>Cubic</i>	57,1%
	<i>Fine Gaussian</i>	40,0%
	<i>Medium Gaussian</i>	61,4%
	<i>Coarse Gaussian</i>	57,1%
<i>K-nearest Neighbor</i> (KNN)	<i>Fine</i>	42,9%
	<i>Medium</i>	55,7%
	<i>Coarse</i>	42,9%
	<i>Cosine</i>	55,7%
	<i>Cubic</i>	58,6%
<i>Discriminant</i> <i>Analysis</i>	<i>Linear</i>	51,4%
	<i>Boosted</i>	48,3
<i>Regression</i>	<i>Logistic</i>	58,4%
<i>Decision Tree</i>	<i>Fine</i>	40,0%
	<i>Medium</i>	40,0%
	<i>Coarse</i>	42,9%
<i>Random</i> <i>Forest</i>	<i>Bagged</i>	55,7%
	<i>Boosted</i>	42,9%

Tabela 52 – BDT - *training set* - *Support Vector Machine* (Fonte: autor).

Parâmetros	Resultados
<i>Prediction Speed</i>	Aprox. 450 obs/sec
<i>Training time</i>	1,70 sec
<i>Total Accuracy</i>	71,4%
<i>Kernel function</i>	<i>Linear</i>

Pode-se observar que, de acordo com os dados apresentados na tabela do deslocamento transversal dos aparelhos de apoio para o *training set*, o modelo de classificação teve uma performance regular, classificando corretamente os TAGs 71,4% das vezes. A Tabela 53 a seguir demonstra o valor de acurácia total para o modelo de classificação escolhido para o *testing set*.

Tabela 53 – BDT - *testing set* - *Support Vector Machine* (Fonte: autor).

Parâmetro	Resultado
<i>Total Accuracy</i>	70,0%

Nota-se que o modelo *Support Vector Machine* teve uma performance similar para o *testing set*, logo o modelo não cometeu *overfitting*. Pode-se observar a partir das matrizes de confusão que os dados de resposta estrutural para o *training set* apresentaram todas as classes de categorização, enquanto para o *testing set* apresentaram apenas as classes *green* e *yellow*. A Figura 113 a seguir demonstra a *confusion matrix* para ambos os conjuntos, treinamento e teste, respectivamente.

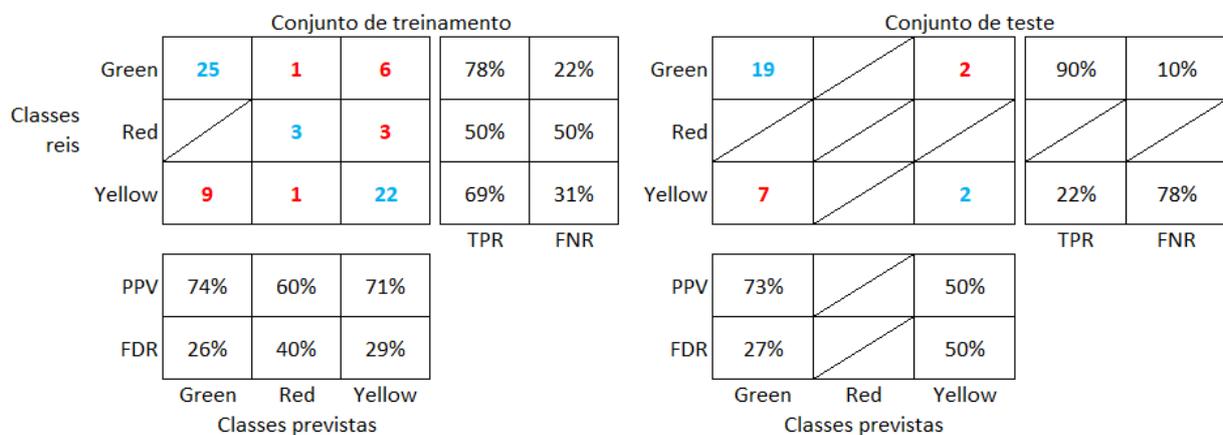


Figura 113 – BDT - *Confusion Matrix* para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).

De acordo com a matriz de confusão, o modelo para o conjunto de teste cometeu mais erros de classificação, onde 78% dos dados reais categorizados como *yellow* foram classificados como *green*, ou seja, aproximadamente 78% das pontes que apresentaram deslocamento acima de 25 *mm* foram classificadas como tendo um deslocamento ativo máximo menor que 25 *mm*. Mesmo assim, o erro dos valores previstos como *green* é de apenas 27%. Já no caso do modelo para o conjunto de treinamento, o algoritmo cometeu mais erros para os valores previstos como *red*, contabilizando um total de 50%.

4.2.2.5 Deslocamento do tabuleiro

Para o parâmetro deslocamento do tabuleiro (DU), o modelo com melhor performance foi o *Decision Tree*. A Tabela 54 a seguir demonstra os resultados obtidos para o deslocamento do tabuleiro para os algoritmos do *training set* através do *Prediction Accuracy* (PA).

Tabela 54 – DU - Verificação da performance dos modelos de classificação através do PA (Fonte: autor).

Algoritmo	Tipo	PA
<i>Support Vector Machine</i> (SVM)	<i>Linear</i>	55,7%
	<i>Quadratic</i>	54,3%
	<i>Cubic</i>	50,0%
	<i>Fine Gaussian</i>	47,1%
	<i>Medium Gaussian</i>	55,7%
	<i>Coarse Gaussian</i>	51,4%
<i>K-nearest Neighbor</i> (KNN)	<i>Fine</i>	40,0%
	<i>Medium</i>	51,4%
	<i>Coarse</i>	47,1%
	<i>Cosine</i>	54,3%
	<i>Cubic</i>	55,7%
<i>Discriminant Analysis</i>	<i>Linear</i>	51,4%
	<i>Boosted</i>	<i>failed</i>
<i>Regression</i>	<i>Logistic</i>	49,3%
<i>Decision Tree</i>	<i>Fine</i>	54,3%
	<i>Medium</i>	54,3%
	<i>Coarse</i>	60,0%
<i>Random Forest</i>	<i>Bagged</i>	52,9%
	<i>Boosted</i>	47,1%

De acordo com a tabela de verificação de performance, o algoritmo *Decision Tree* obteve o melhor desempenho entre todos os algoritmos. A Tabela 55 a seguir demonstra os parâmetros numéricos juntamente com as características do algoritmo e a velocidade de processamento.

Tabela 55 – DU - *training set* - *Random Forest* (Fonte: autor).

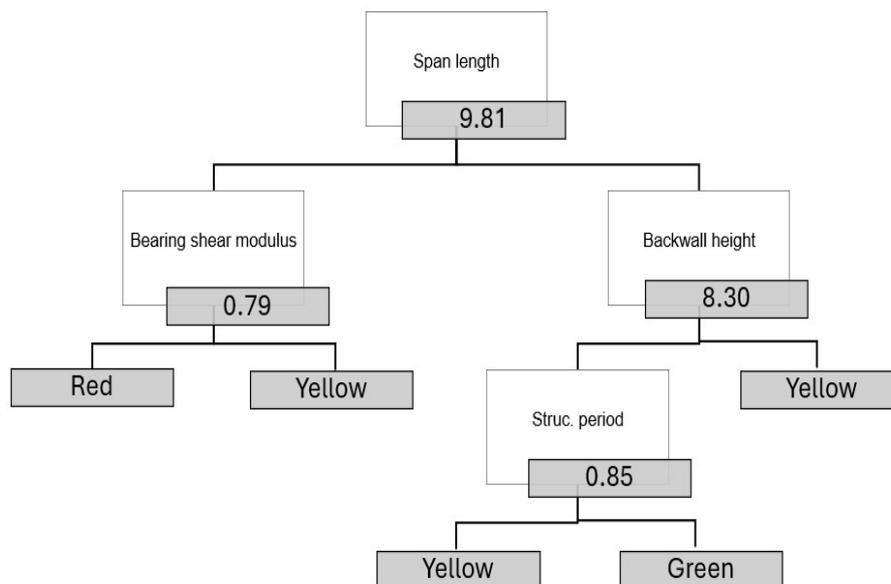
Parâmetros	Resultados
<i>Prediction Speed</i>	Aprox. 2300 obs/sec
<i>Training time</i>	0,27 sec
<i>Total Accuracy</i>	60,0%
<i>Minimum leaf size</i>	4

Pode-se observar que, de acordo com os dados apresentados na tabela do deslocamento do tabuleiro para o *training set*, o modelo de classificação teve uma performance ruim, classificando corretamente apenas os TAGs 60% das vezes. A Tabela 56 a seguir demonstra o valor de acurácia total para o modelo de classificação escolhido para o *testing set*.

Tabela 56 – DU - Deslocamento do tabuleiro - *testing set* - *Random Forest* (Fonte: autor).

Parâmetro	Resultado
<i>Total Accuracy</i>	70,0%

Nota-se que o modelo *Decision Tree* teve uma performance melhor para o *testing set*, logo o modelo não cometeu *overfitting*. A Figura 114 a seguir ilustra a árvore de decisão.

**Figura 114** – DU - Modelo esquemático do algoritmo *Decision Tree* (Fonte: autor).

Pode-se observar a partir das matrizes de confusão que os dados de resposta estrutural para o *training set* apresentaram todas as classes de categorização, enquanto para o *testing set* apresentaram apenas as classes *green* e *yellow*. A Figura 115 a seguir demonstra a *confusion matrix* para ambos os conjuntos, treinamento e teste, respectivamente.

De acordo com a matriz de confusão, o modelo para o conjunto de teste cometeu mais erros de classificação, onde 78% dos dados reais categorizados como *yellow* foram classificados como *green*, ou seja, aproximadamente 78% das pontes que apresentaram deslocamento acima de 25 *mm* foram classificadas como tendo um deslocamento ativo máximo menor que 25 *mm*. Mesmo assim, o erro dos valores previstos como *green* é baixo, de apenas 27%. Já no caso do modelo para o conjunto de treinamento, o algoritmo cometeu mais erros para os valores previstos como *red*, contabilizando um total de 50%.

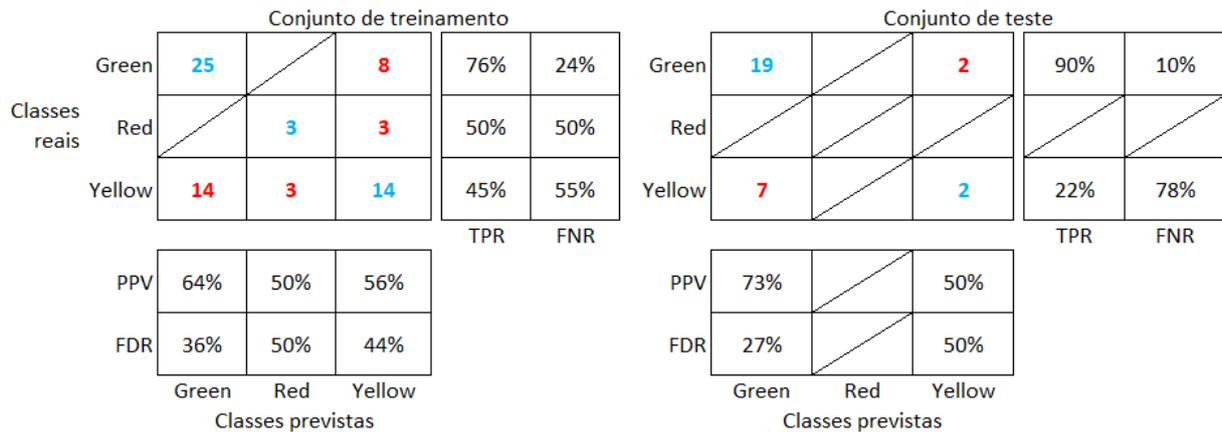


Figura 115 – DU - *Confusion Matrix* para o conjunto de treinamento e para o conjunto de teste (Fonte: autor).

4.2.2.6 Resumo dos resultados obtidos para os parâmetros de engenharia - modelo de classificação

A tabela 57 a seguir demonstra o resumo dos resultados obtidos para os parâmetros de engenharia para o modelo de classificação para o conjunto de treinamento e de teste.

Tabela 57 – Resumo dos resultados obtidos - modelo de classificação (Fonte: autor).

EDP	Conjunto	Algoritmo	<i>Prediction accuracy</i>
ADA	Treino	<i>Random Forest</i>	90%
	Teste	<i>Random Forest</i>	73%
ADP	Treino	<i>Random Forest</i>	77%
	Teste	<i>Random Forest</i>	43%
BDL	Treino	<i>Random Forest</i>	95%
	Teste	<i>Random Forest</i>	83%
BDT	Treino	<i>SVM</i>	71%
	Teste	<i>SVM</i>	70%
DU	Treino	<i>Random Forest</i>	60%
	Teste	<i>Random Forest</i>	70%

4.3 Comparação da análise estocástica vs Machine Learning

Os resultados obtidos através do *Machine Learning*, para o modelo de regressão, são similares aos resultados obtidos pela análise estocástica. Esta comparação confirma a eficácia no modelo de inteligência artificial. No entanto, também fica evidente que os erros cometidos pela análise estocástica são transferidos para o modelo de inteligência artificial, resultando em erros similares, uma vez que a base de dados utilizada pelo *Machine Learning* são os resultados da análise estocástica.

A Figura 116 a seguir demonstra as curvas de fragilidade do sistema para todos os estados-limites de dano para a análise estocástica e para o modelo de regressão do ML, respectivamente.

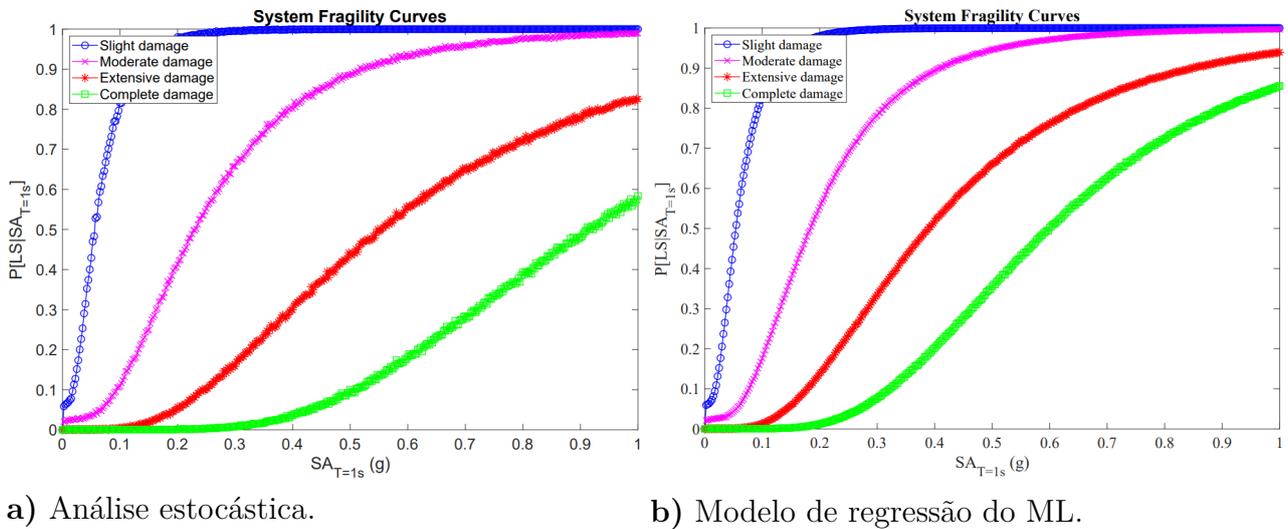


Figura 116 – Comparação entre as curvas de fragilidade do sistema estrutural para cada estado-limite de dano - Para análise estocástica e após modelo de regressão do ML.

Quando comparadas, a curva de fragilidade do sistema para o modelo do *machine learning*, para os estados-limites de dano extensivo e completo, apresentou maior probabilidade de dano para uma dada intensidade de medida (IM). Para facilitar a comparação, a Figura 117 a seguir demonstra as curvas de fragilidade do sistema para os estados-limites de dano extensivo e completo para a análise estocástica e para o modelo de regressão do ML, respectivamente.

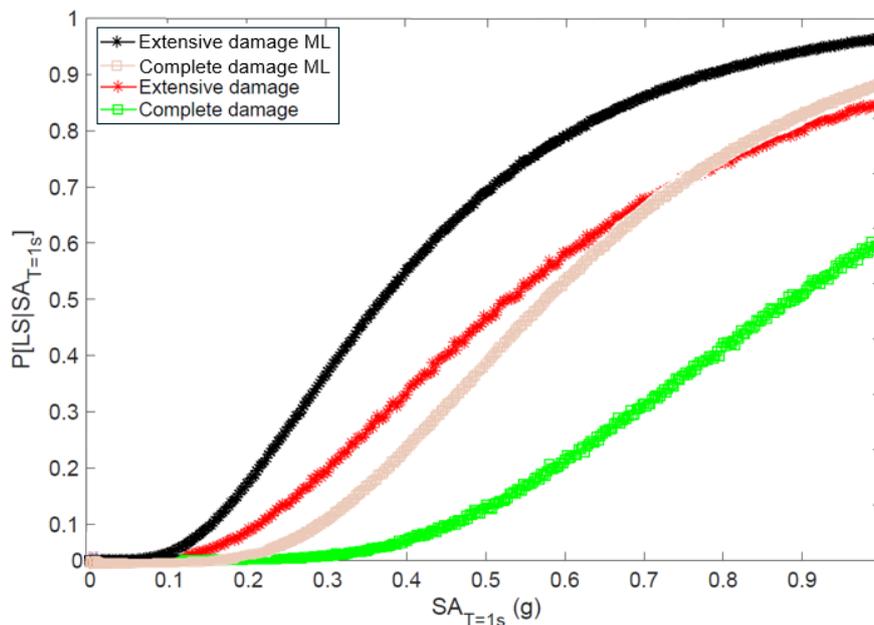


Figura 117 – Curva de fragilidade do sistema estrutural para os estados-limite de dano extensivo e completo para a análise estocástica e após modelo de regressão do ML.

Com base na figura 117, nota-se que a partir de um IM igual a 0,15(g) (para dano extensivo) e 0,25(g) (para dano completo) os modelos começam a apresentar diferentes probabilidades de dano. Esta diferença se comporta de maneira distinta em relação ao tipo de estado limite de dano em comparação. A Figura 118 a seguir demonstra as diferenças encontradas, em termos de probabilidade de dano, para os danos extensivo e completo para os modelos estocástico e ML. Percebe-se que, para o estado limite de dano extensivo, o erro entre os modelos aumenta até 0,4(g) e diminui a partir de 0,6(g). Já para o estado limite de dano completo, o erro tem um pico em 0,7(g) e logo depois diminui. Para um IM igual a 0,6(g) a probabilidade de dano extensivo é de aproximadamente 55% e 80% para o modelo estocástico e de regressão do ML, respectivamente, totalizando um erro de 25%. Já para o dano completo, para um mesmo IM, a probabilidade é de aproximadamente 20% e 50% para o modelo estocástico e de regressão do ML, respectivamente, totalizando um erro de 30%. Entende-se que este erro está a favor da segurança, uma vez que, para um certo IM, a probabilidade de dano obtida através do modelo de *machine learning* é maior do que o esperado.

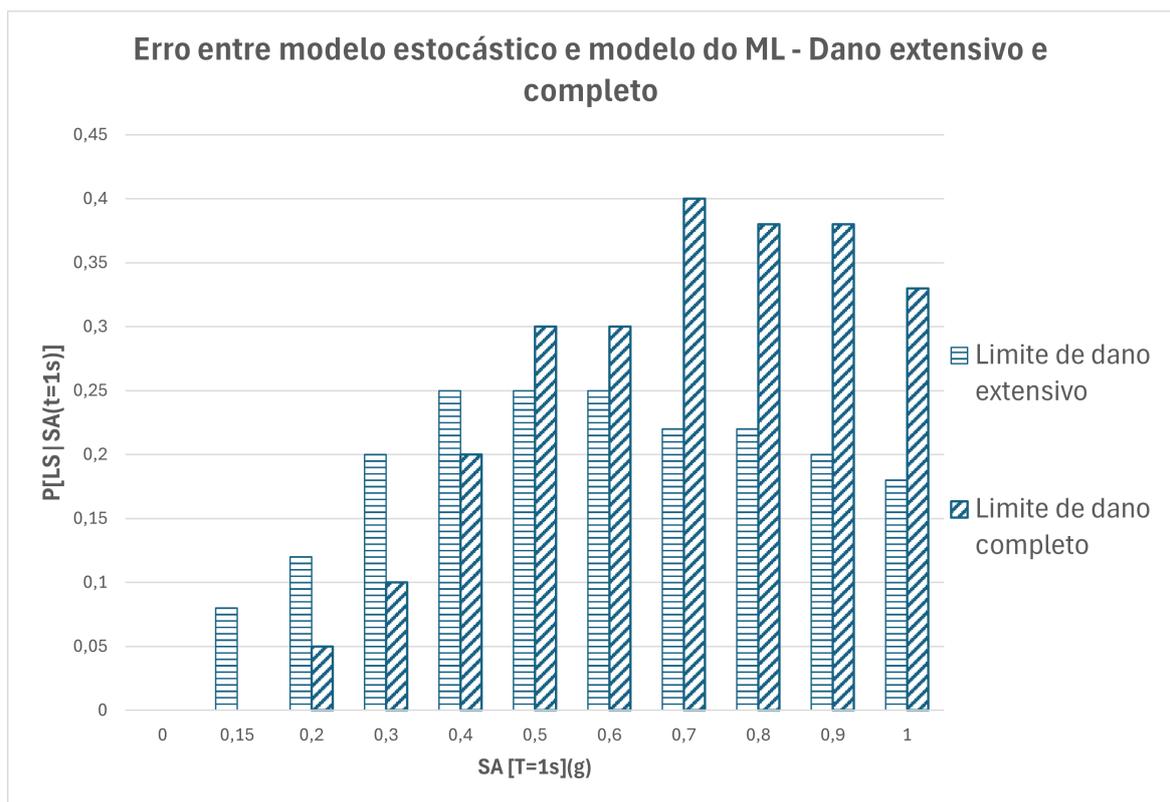


Figura 118 – Comparação entre as diferenças, em termos de probabilidades de dano, para os estados limites extensivo e completo para os modelos estocástico e regressão do ML.

4.4 Síntese do capítulo

Neste capítulo foi demonstrado os resultados obtidos através da aplicação de ML (modelos de regressão e classificação) para os conjuntos de treinamento e de teste. Os parâmetros numéricos e as características dos algoritmos escolhidos foram demonstrados para todos os parâmetros de demanda de engenharia.

Inúmeros gráficos foram utilizados com o objetivo de ilustrar os resultados obtidos para ambos os modelos. Após a análise de cada EDP, tanto para os modelos de regressão quanto para os de classificação, foram discutidos os resultados e apontados as respectivas performances. Por fim, foi discutido a comparação entre a análise estocástica e a análise de inteligência artificial.

5 Considerações finais

O principal objetivo deste trabalho é comparar o estudo de vulnerabilidade sísmica de uma OAE, situadas no nordeste do Brasil, através de métodos probabilísticos como a curva de fragilidade, com a aplicação das ferramentas de inteligência artificial como o *machine learning*.

As etapas realizadas neste trabalho foram:

- Geração do banco de dados - Foram gerados 100 modelos de OAEs estatisticamente similares. Posteriormente, estes modelos foram analisados com um conjunto de 100 pares de terremotos, para assim, obter 10.000 dados de resposta estrutural, ou demanda. As atividades sísmicas utilizadas são aquelas compatíveis com a região nordeste do Brasil.
- Modelo numérico - A modelagem é empregada com o intuito de simular as respostas estruturais. Foram adotadas para este trabalho as metodologias encontradas em Tavares et al. (2012), por exemplo;
- Análises dinâmicas não lineares - O método utilizado neste trabalho foi o *Cloud Analysis*, que consiste em uma análise dinâmica não incremental, onde os sinais são empregados sem qualquer tipo de escalonamento;
- Modelos de demanda e capacidade - Após realizadas as análises dinâmicas para a obtenção das respostas estruturais em função de um IM, aplicou-se o PSDM para calcular os parâmetros probabilísticos dos modelos. Os valores de capacidade foram retirados da bibliografia, exceto para os limites de curvatura dos pilares, que foram obtidos através da análise *pushover*;
- Curvas de fragilidade - Com os parâmetros do PSDM já calculados e os valores limites de capacidade, foi possível construir as curvas de fragilidade para cada componente estrutural e do sistema como um todo;
- Aplicação das ferramentas de inteligência artificial - Obtidos os resultados de demanda estrutural e os valores de capacidade, foi possível aplicar as ferramentas de *machine learning* para gerar os modelos de regressão e classificação;
- Construção das curvas de fragilidade pós análise de inteligência artificial - Obtidos os resultados do modelo de regressão do ML, aplicou-se o PSDM para calcular os parâmetros

probabilísticos do modelo e, conseqüentemente, a construção das curvas de fragilidade para componente estrutural e do sistema.

Considerando os resultados obtidos do modelo estocástico, verificou-se uma grande dispersão nos modelos de regressão do PSDM para o deslocamento longitudinal dos aparelhos de apoio e dos deslocamentos ativo e passivo dos encontros. Estes valores podem ser devido ao comportamento não-linear das juntas de dilatação. Essa mesma dispersão pode ser encontrada nos modelos de regressão do PSDM para os valores obtidos através do *machine learning*.

Em geral, os resultados obtidos dos modelos de inteligência artificial, tanto de regressão quanto de classificação foram satisfatórios. Verificou-se que apenas para dois parâmetros de demanda (EDP) o modelo de regressão apresentou erros consideráveis, validados pelo RMSE e pelo coeficiente de determinação R^2 , sendo estes dois parâmetros o deslocamento passivo dos encontros e o deslocamento transversal dos aparelhos de apoio. No entanto, pode-se observar que mesmo que os parâmetros do deslocamento transversal dos aparelhos de apoio não tenham atingido valores satisfatórios, o resultado da regressão probabilística (PSDM) não divergiu muito da análise estocástica. Os outros parâmetros apresentaram valores baixos de erro e dispersão com R^2 superior a 70%, tanto para o conjunto de treinamento quanto para o conjunto de teste. Conclui-se que os modelos também não apresentaram *overfitting*.

Os modelos de classificação tiveram uma performance melhor quando comparada aos modelos de regressão. O valor mais baixo de precisão calculado (*Prediction Accuracy*) foi de 60% para o *training set* para o deslocamento do tabuleiro. No entanto, quando testado no *testing set* o modelo performou de maneira satisfatória com 70% de precisão. Ao contrário do modelo de regressão, o modelo de classificação cometeu erros do tipo *overfitting* para os deslocamentos passivos dos encontros, totalizando uma precisão baixa de 43% para o conjunto de teste, contra 77% de precisão para o conjunto de treinamento.

Obtidos os valores da regressão do PSDM, tanto para a análise estocástica quanto para a análise através das ferramentas de inteligência artificial, pode-se construir as curvas de fragilidade. Exceto para os estados-limites de dano extensivo e completo, as curvas, tanto dos parâmetros de engenharia quanto do sistema, foram muito similares, e, conseqüentemente satisfatórios, exemplificando a eficácia do uso das ferramentas de *Machine Learning* no estudo de vulnerabilidade estrutural.

Entende-se que as diferenças obtidas nas curvas para os estados-limites de dano extensivo e completo estão a favor da segurança, uma vez que para determinada intensidade de medida (IM) a probabilidade de dano é maior para as curvas geradas através do ML quando comparada com as curvas da análise estocástica. Vale ressaltar, que os erros encontrados na análise estocástica

são transferidos para a análise de inteligência artificial, portanto o processo de geração dos dados deve ser controlado a fim de se evitar a propagação de erros.

5.1 Sugestão de trabalhos futuros

Para trabalhos futuros sobre a temática de análise de vulnerabilidade sísmica de OAEs, através de ferramentas de inteligência artificial, o autor sugere que a OAE em estudo tenha uma configuração geométrica mais simples. Desta forma, o modelo matemático (elementos finitos) será mais básico, reduzindo assim os erros cometidos durante o processo de análise estocástica e automaticamente os erros do processo de inteligência artificial.

Fica também como sugestão a implantação de ferramentas mais complexas de inteligência artificial, talvez no ramo da aprendizagem não supervisionada. Recomenda-se também uma investigação mais aprofundada sobre os erros cometidos pelos algoritmos no processo de predição, quantitativa ou qualitativa, para os estados limites de danos mais severos, como o extensivo e completo.

Referências

- ANDRADE, R.; PEREIRA, E.; CAVALCANTE, G.; VIEIRA, L.; SIQUEIRA, G. Seismic fragility assessment for a rc building in seismically stable brazil: A sensitivity analysis. **Journal of Building Engineering**, 2022.
- ASCE. **ASCE 7-16 - Minimum design loads and associated criteria for buildings and other structures**. [S.l.]: American Society of Civil Engineering, 2017.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 9062**: Projeto e execução de estruturas de concreto pré-moldado. Rio de Janeiro, 2017. 86 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 15421**: Projeto de estruturas resistentes a sismos. Rio de Janeiro, 2023. 27 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6118**: Projeto de estruturas de concreto - procedimento. Rio de Janeiro, 2024. 242 p.
- ASSUMPÇÃO, M. A regional magnitude scale for brazil. **Bulletin of the Seismological Society of America**, The Seismological Society of America, v. 73, n. 1, p. 237–246, 1983.
- ASSUMPÇÃO, M.; PIRCHINER, M.; DOURADO, J.; BARROS, L. Terremotos no brasil: Preparando-se para eventos raros. **Bulletin of Brazilian Society of Geophysics**. [online]. Available at: https://sbgf.org.br/home/images/Boletim_96-2016.pdf [Accessed 31 Jun. 2018]. (in Portuguese), 2016.
- BACHMAN, R.; BONOWITZ, D.; CALDWELL, P.; FILIATRAULT, A.; KENNEDY, R.; MCGAVIN, G.; MIRANDA, E. Engineering demand parameters for nonstructural components. **ATC-58 Project Task Report**, ATC, Redwood City, California, 2004.
- BAKER, J. W. An introduction to probabilistic seismic hazard analysis (psha). **White paper, version**, Citeseer, v. 1, p. 72, 2008.
- BEYER, K.; BOMMER, J. J. Selection and scaling of real accelerograms for bi-directional loading: a review of current practice and code provisions. **Journal of Earthquake Engineering**, v. 11, n. S1, p. 13–45, 2007.
- BOX, G. E.; HUNTER, J. S. Multi-factor experimental designs for exploring response surfaces. **The Annals of Mathematical Statistics**, p. 195–241, 1957.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5–32, 2001.
- BSSC. 1050 (2015) nehrp recommended seismic provisions for new buildings and other structures. **Building Seismic Safety Council, Federal Emergency Management Agency**, Washington, 2015.
- CALTRANS, S. Caltrans seismic design criteria. **California Department of Transportation**, Sacramento, CA, 1990.

- CAVALCANTE, G. **Estudo do risco sísmico de OAEs típicas em rodovias federais do nordeste do brasil**. Tese (Doutorado) — Unicamp, Campinas, Brasil, 2022.
- CHANG, G. A.; MANDER, J. B. **Seismic energy based fatigue damage analysis of bridge columns: Part 1 - Evaluation of seismic capacity**. [S.l.], 1994.
- CHOI, E. **Seismic analysis and retrofit of mid-America bridges**. Tese (Doutorado) — School of Civil and Environmental Engineering, Georgia Institute of Technology, Georgia, Estados Unidos, 2002.
- CHOI, E.; DESROCHES, R.; NIELSON, B. Seismic fragility of typical bridges in moderate seismic zones. **Engineering Structures**, v. 26, n. 2, p. 187–199, 2004.
- CHOI, E.; JEON, J.-C. Seismic fragility of typical bridges in moderate seismic zone. **KSCE Journal of Civil Engineering**, v. 7, n. 1, p. 41–51, 2003.
- CORNELL, C. A.; JALAYER, F.; HAMBURGER, R. O.; FOUTCH, D. A. Probabilistic basis for 2000 sac federal emergency management agency steel moment frame guidelines. **Journal of structural engineering**, v. 128, n. 4, p. 526–533, 2002.
- DUKES, J.; DESROCHES, R.; PADGETT, J. Sensitivity study of design parameters used to develop bridge specific fragility curves. In: **Proc. 15th World Conf. Earthquake Eng.** [S.l.: s.n.], 2012. v. 720.
- ELNASHAI, A. S.; SARNO, L. D. **Fundamentals of earthquake engineering**. [S.l.]: Wiley Online Library, 2008.
- FAWCETT, T. An introduction to roc analysis. **Pattern recognition letters**, v. 27, n. 8, p. 861–874, 2006.
- FEMA. **Hazus-MH MR5 Technical Manual. Multi-Hazard Loss Estimation Methodology Earthquake Model**. 2010.
- FEMA, H.-M. Mr3 technical manual: Multi-hazard loss estimation methodology earthquake model. **Washington DC: Federal Emergency Management Agency**, 2003.
- FERRARIO, E.; PEDRONI, N.; ZIO, E.; LOPEZ-CABALLERO, F. Bootstrapped artificial neural networks for the seismic analysis of structural systems. **Structural Safety**, v. 67, p. 70–84, 2017.
- FIELDSEND, J. E.; EVERSON, R. M. Visualisation of multi-class roc surfaces. 2005.
- FILIPPOU, F. C.; BERTERO, V. V.; POPOV, E. P. Effects of bond deterioration on hysteretic behavior of reinforced concrete joints. Earthquake Engineering Research Center, University of California Berkeley, 1983.
- FREDDI, F.; PADGETT, J. E.; DALL’ASTA, A. Probabilistic seismic demand modeling of local level response parameters of an rc frame. **Bulletin of Earthquake Engineering**, v. 15, n. 1, p. 1–23, 2017.

- GARCIA, M. S. C.; SIQUEIRA, G. H.; JUNIOR, L. C. M. V.; VIZOTTO, I. Evaluation of structural capacity of triangular and hexagonal reinforced concrete free-form shells. **Engineering Structures**, v. 188, p. 519–537, 2019.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The elements of statistical learning. springer series in statistics. In: . [S.l.]: Springer, 2001.
- IERVOLINO, I.; MADDALONI, G.; COSENZA, E. Eurocode 8 compliant real record sets for seismic analysis of structures. **Journal of Earthquake Engineering**, v. 12, n. 1, p. 54–90, 2008.
- IZADPANAH, M.; HABIBI, A. Evaluating the spread plasticity model of idarc for inelastic analysis of reinforced concrete frames. **Structural Engineering and Mechanics**, v. 56, n. 2, p. 169–188, 2015.
- KAPPOS, A. J.; PANAGOPOULOS, G.; PANAGIOTOPOULOS, C.; PENELIS, G. A hybrid method for the vulnerability assessment of r/c and urm buildings. **Bulletin of Earthquake Engineering**, v. 4, n. 4, p. 391–413, 2006.
- KELLY, J. M. Earthquake-resistant design with rubber. Springer, 1993.
- KIANI, J.; CAMP, C.; PEZESHK, S. On the application of machine learning techniques to derive seismic fragility curves. **Computers & Structures**, v. 218, p. 108–122, 2019.
- KONG, Q.; TRUGMAN, D. T.; ROSS, Z. E.; BIANCO, M. J.; MEADE, B. J.; GERSTOFT, P. Machine learning in seismology: Turning data into insights. **Seismological Research Letters**, v. 90, n. 1, p. 3–14, 2019.
- LALLEMANT, D.; KIREMIDJIAN, A.; BURTON, H. Statistical procedures for developing earthquake damage fragility curves. **Earthquake Engineering & Structural Dynamics**, v. 44, n. 9, p. 1373–1389, 2015.
- LANDGREBE, T.; DUIN, R. A simplified extension of the area under the roc to the multiclass domain. In: **Seventeenth annual symposium of the pattern recognition association of South Africa**. [S.l.: s.n.], 2006. v. 1.
- LANDGREBE, T. C.; DUIN, R. P. Approximating the multiclass roc by pairwise analysis. **Pattern recognition letters**, Elsevier, v. 28, n. 13, p. 1747–1758, 2007.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, v. 521, n. 7553, p. 436–444, 2015.
- LEITÃO, F.; ANDRADE, R.; PEREIRA, E.; CAROBENO, C.; SIQUEIRA, G. Evaluation of seismic fragility in reinforced concrete structures using machine learning techniques: A systematic literature review. **IBRACON**, 2023.
- MANGALATHU, S.; HWANG, S.-H.; CHOI, E.; JEON, J.-S. Rapid seismic damage evaluation of bridge portfolios using machine learning techniques. **Engineering Structures**, v. 201, p. 109785, 2019.

- MANGALATHU, S.; JEON, J.-S. Stripe-based fragility analysis of multispan concrete bridge classes using machine learning techniques. **Earthquake Engineering & Structural Dynamics**, v. 48, n. 11, p. 1238–1255, 2019.
- MANGALATHU, S.; SOLEIMANI, F.; JEON, J.-S. Bridge classes for regional seismic risk assessment: Improving hazus models. **Engineering Structures**, Elsevier, v. 148, p. 755–766, 2017.
- MARONEY, B. Interpretation of large scale bridge abutment test results. In: **Proc., 3rd Annual Seism. Research Workshop**. [S.l.: s.n.], 1994.
- MAZZONI, S.; MCKENNA, F.; SCOTT, M. H.; FENVES, G. L. et al. Opensees command language manual. **Pacific Earthquake Engineering Research (PEER) Center**, v. 264, 2006.
- MCGUIRE, R. K. **Seismic Hazard And Risk Analysis**. [S.l.]: Earthquake Engineering Research Institute, 2004.
- MCKENNA, F.; FENVES, G. L.; SCOTT, M. H. Opensees: Open system for earthquake engineering simulation. **Pacific Earthquake Engineering Research Center, University of California**, Berkeley, USA, 2006. Disponível em: <<http://opensees.berkeley.edu>>.
- MIRZA, S. A.; MACGREGOR, J. G. Variability of mechanical properties of reinforcing bars. **Journal of the Structural Division**, v. 105, n. ASCE 14590 Proceeding, 1979.
- MUTHUKUMAR, S. **A contact element approach with hysteresis damping for the analysis and design of pounding in bridges**. Tese (Doutorado) — Georgia Institute of Technology, Georgia, Estados Unidos, 2003.
- MUTHUKUMAR, S.; DESROCHES, R. A hertz contact model with non-linear damping for pounding simulation. **Earthquake engineering & structural dynamics**, v. 35, n. 7, p. 811–828, 2006.
- NIELSON, B. G. **Analytical fragility curves for highway bridges in moderate seismic zones**. Tese (Doutorado) — Georgia Institute of Technology, 2005.
- NIELSON, B. G.; DESROCHES, R. Influence of modeling assumptions on the seismic response of multi-span simply supported steel girder bridges in moderate seismic zones. **Engineering structures**, v. 28, n. 8, p. 1083–1092, 2006.
- NIELSON, B. G.; DESROCHES, R. Seismic fragility methodology for highway bridges using a component level approach. **Earthquake Engineering & Structural Dynamics**, Wiley Online Library, v. 36, n. 6, p. 823–839, 2007.
- NOGUEIRA, C. G. **Desenvolvimento de modelos mecânicos, de confiabilidade e de otimização para aplicação em estruturas de concreto armado**. Tese (Doutorado) — Universidade de São Paulo, São Paulo, Brasil, 2010.
- NOVAK, M. Dynamic stiffness and damping of piles. **Canadian Geotechnical Journal**, v. 11, n. 4, p. 574–598, 1974.

- NOVAK, M.; HOWELL, J. F. Torsional vibration on pile foundations. **Journal of Geotechnical and Geoenvironmental Engineering**, v. 103, n. Proc. Paper 11318 Proceeding, 1977.
- NOWAK, A. S.; COLLINS, K. R. **Reliability of structures**. [S.l.]: CRC Press, 2012.
- OLIVEIRA, C.; GRECO, M.; BITTENCOURT, T. Analysis of the brazilian federal bridge inventory. **Revista IBRACON de Estruturas e Materiais**, v. 12, n. 1, p. 1–3, 2019.
- OPENSEES. 2019. Disponível em: <https://opensees.berkeley.edu/wiki/index.php/OpenSees_User>.
- PADGETT, J.; NIELSON, B. G.; DESROCHES, R. Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios. **Earthquake Engineering & Structural Dynamics**, v. 37, n. 5, p. 711–725, 2008.
- PADGETT, J. E. **Seismic vulnerability assessment of retrofitted bridges using probabilistic methods**. Tese (Doutorado) — Georgia Institute of Technology, Georgia, Estados Unidos, 2007.
- PADGETT, J. E.; DESROCHES, R. Methodology for the development of analytical fragility curves for retrofitted bridges. **Earthquake Engineering & Structural Dynamics**, v. 37, n. 8, p. 1157–1174, 2008.
- PADGETT, J. E.; DESROCHES, R. Retrofitted bridge fragility analysis for typical classes of multispan bridges. **Earthquake Spectra**, v. 25, n. 1, p. 117–141, 2009.
- PAN, Y.; AGRAWAL, A. K.; GHOSN, M.; ALAMPALLI, S. Seismic fragility of multispan simply supported steel highway bridges in new york state. ii: Fragility analysis, fragility curves, and fragility surfaces. **Journal of Bridge Engineering**, v. 15, n. 5, p. 462–472, 2010.
- PAULTRE, P. **Dynamics of structures**. [S.l.]: John Wiley & Sons, 2013.
- PEREIRA, E. **Estudo da Fragilidade Sísmica de Pórticos de Concreto Armado com Irregularidades Estruturais**. Dissertação (Mestrado) — Unicamp, 2021.
- PEREIRA, E.; ANDRADE, R.; LEITÃO, F.; CAROBENO, C.; SIQUEIRA, G. Seismic risk evaluation of non-ductile low-rise rc buildings in brazil: Time-based and intensity-based assessments considering different performance metrics. **Journal of Building Engineering**, 2024.
- PETERSEN, M. D.; HARMSSEN, S. C.; JAISWAL, K. S.; RUKSTALES, K. S.; LUCO, N.; HALLER, K. M.; MUELLER, C. S.; SHUMWAY, A. M. Seismic hazard, risk, and design for south america. **Bulletin of the Seismological Society of America**, v. 108, n. 2, p. 781–800, 2018.
- PORTER, K.; KENNEDY, R.; BACHMAN, R. Creating fragility functions for performance-based earthquake engineering. **Earthquake Spectra**, v. 23, n. 2, p. 471–489, 2007.
- PRAKASH, S.; SHARMA, H. D. **Pile foundations in engineering practice**. [S.l.]: John Wiley & Sons, 1990.

- RODRIGUES, I. **Avaliação da Vulnerabilidade Sísmica de Edificações Regulares em Concreto Armado no Brasil através da Elaboração de Curvas de Fragilidade**. Dissertação (Mestrado) — Unicamp, São Paulo, Brasil, 2021.
- RODRIGUES, I.; CAVALCANTE, G.; PEREIRA, E.; VIEIRA, L.; LIEL, A.; SIQUEIRA, G. Seismic fragility assessment of a rc frame considering concentrated and distributed plasticity modeling. **Revista IBRACON de Estruturas e Materiais**, 2024.
- ROKNEDDIN, K. **Reliability and risk assessment of networked urban infrastructure systems under natural hazards**. Tese (Doutorado) — Rice University, Texas, Estados Unidos, 2013.
- SANTIAGO, W.; BECK, A. A new study of brazilian concrete strength conformance. **Revista IBRACON de Estruturas e Materiais**, v. 10, n. 4, p. 906–923, 2017.
- SCHRAGE, I. Anchoring of bearings by friction. In: AMERICAN CONCRETE INSTITUTE NIAGARA FALLS, NY, USA. **Joint sealing and bearing systems for concrete structures, world congress on joints and bearings**. [S.l.], 1981. v. 1.
- SHEARER, P. M. **Introduction to seismology**. [S.l.]: Cambridge University Press, 2009.
- SHEDLOCK, K. M.; TANNER, J. G. Seismic hazard map of the western hemisphere. **Annali di Geofisica**, v. 42, n. 6, p. 1199–1214, 1999.
- SILVA, V.; CROWLEY, H.; VARUM, H.; PINHO, R.; SOUSA, R. Evaluation of analytical methodologies used to derive vulnerability functions. **Earthquake Engineering & Structural Dynamics**, v. 43, n. 2, p. 181–204, 2014.
- SIQUEIRA, G. H.; SANDA, A. S.; PAULTRE, P.; PADGETT, J. E. Fragility curves for isolated bridges in eastern canada using experimental results. **Engineering Structures**, v. 74, p. 311–324, 2014.
- SIQUEIRA, G. H.; TAVARES, D. H.; PAULTRE, P.; PADGETT, J. E. Performance evaluation of natural rubber seismic isolators as a retrofit measure for typical multi-span concrete bridges in eastern canada. **Engineering structures**, v. 74, p. 300–310, 2014.
- SONG, S.; QIAN, Y.; LIU, J.; XIE, X.; WU, G. Time-variant fragility analysis of the bridge system considering time-varying dependence among typical component seismic demands. **Earthquake Engineering and Engineering Vibration**, v. 18, n. 2, p. 363–377, 2019.
- SUESCUN, J. R. **Courbes de fragilite pour les ponts au Quebec tenant compte du sol de fondation**. Tese (Doutorado) — Université de Sherbrooke, Sherbrooke, Canada, 2010.
- TANAKA, H. **Effect of lateral confining reinforcement on the ductile behaviour of reinforced concrete columns**. Tese (Doutorado) — University of Canterbury, Christchurch, New Zeland, 1990.
- TAVARES, D.; SUESCUN, J.; PAULTRE, P.; PADGETT, J. Seismic fragility of a highway bridge in quebec. **Journal of bridge engineering**, American Society of Civil Engineers, v. 18, n. 11, p. 1131–1139, 2013.

- TAVARES, D. H. **Évaluation de la vulnérabilité sismique des ponts routiers au Québec à l'aide des courbes de fragilité**. Tese (Doutorado) — Université de Sherbrooke, Sherbrooke, Canada, 2012.
- TAVARES, D. H.; PADGETT, J. E.; PAULTRE, P. Fragility curves of typical as-built highway bridges in eastern canada. **Engineering Structures**, v. 40, p. 107–118, 2012.
- UNDRR. 2019. Disponível em: <<https://www.unisdr.org>>.
- USGS. 2021. Disponível em: <<https://earthquake.usgs.gov/earthquakes>>.
- VAMVATSIKOS, D.; CORNELL, C. A. Incremental dynamic analysis. **Earthquake Engineering & Structural Dynamics**, Wiley Online Library, v. 31, n. 3, p. 491–514, 2002.
- VAPNIK, V. Statistical learning theory. 1st edwiley. **Boston, USA**, 1998.
- WILSON, J. C.; TAN, B. S. Bridge abutments: formulation of simple model for earthquake response analysis. **Journal of Engineering Mechanics**, v. 116, n. 8, p. 1828–1837, 1990.
- XIE, Y.; SICHANI, M. E.; PADGETT, J. E.; DESROCHES, R. The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. **Earthquake Spectra**, v. 36, n. 4, p. 1769–1801, 2020.