FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Monitorização da integridade de estruturas com recurso a data analytics

João Diogo Alves Morais

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Vera Miguéis

31 de Julho de 2020

Resumo

O processo de implementação de uma estratégia de identificação de dano numa estrutura é referido como *structural health monitoring* (SHM). O dano é definido como as alterações nas propriedades físicas e/ou geométricas de uma estrutura.

SHM é uma área de grande interesse tecnológico e científico. Os elevados custos de reparação das estruturas, resultantes do envelhecimento e de acontecimentos naturais ou não naturais, tornaram o SHM uma área cada vez mais importante.

Com o constante crescimento e amadurecimento dos sistemas de informação, torna-se cada vez mais viável a utilização de técnicas de *Data Mining*, capazes de extrair correlações, padrões comportamentais e tendências através de uma grande quantidade de dados, recorrendo a tecnologias de reconhecimento de padrões e técnicas estatísticas e matemáticas. O trabalho desta dissertação foca-se no desenvolvimento de modelos de deteção de anomalias nas estruturas. Para isto, modelou-se o comportamento regular de uma estrutura através de ferramentas de data mining, nomeadamente algoritmos de regressão como redes neuronais, *support vector machines* e *random forests*. Para além disto, recorreu-se a ferramentas de deteção de anomalias, designadamente a carta de controlo *Hotteling* T^2, a fim de detetar desvios significativos entre o comportamento regular previsto e o comportamento monitorizado.

A implementação de vários modelos usando este conjunto de técnicas permitiu concluir que a *Random Forest* é a técnica mais adequada para realizar as previsões das variáveis estruturais da estrutura em análise. Ambas as abordagens propostas revelaram-se capazes de detetar dano apenas quando ele realmente existe.

Palavras Chave: *Data Mining*/ *Random Forest*/ Redes Neuronais Artificiais/ Regressão/ *Structure Health Monitoring*

Abstract

The process of implementing a damage identification strategy is refered as Stuctural Health Monitoring (SHM). Demage is defined as changes to the material and/or geometric properties of these structures.

SHM is a area of great technological and scientific interest. The high costs of reparairing a structure, resulting from aging and natural and/or unnatural causes, have made SHM an increasingly important area.

With the constant growth and maturation of information systems, it becomes in increasingly feasible to use Data Mining tecniques, able to extract correlations, behavioral patterns and trends through a large amount of data, using pattern recognition tools and mathematical and statistical techniques. The work of this dissertation focuses on the development of anomaly detection models in structures. For this, the regular behavior of a structure was modeled using Data Mining tools, namely regression algorithms as neural networks, support vector machines and random forests. Beyond this, anomaly detection tools were used, as Hotteling T^2 control chart, in order to detect significant deviations between the expected regular behavior and the monitored behavior.

The implementation of several models using this set of techniques allowed us to conclude that the Random Forest is the most adequate technique to make the predictions of the structural variables of the structure under analysis. Both proposed approaches proved to be able to detect damage only when it really exists.

Agradecimentos

À minha família por todo o apoio psicológico
Aos meus colegas que me ajudaram em tudo o que puderam
À minha orientadora Vera Miguéis, pela paciência

João Morais

"I failed over and over again... That's why I succeed"

Michael Jordan

Conteúdo

Lista de Figuras

Lista de Tabelas

Abreviaturas e Símbolos

SHM	Strutural Health Monitoring
DM	Data Mining
KDD	Knowledge Discovery in Databases
AANN	Auto Associative Neural Network
SVM	Support Vector Machine
PCs	Principal Components
SVR	Support Vector Regression
RNA	Redes Neuronais Artificiais
RF	Random Forest

Capítulo 1

Introdução

1.1 Contexto

Para uma gestão eficiente do património construído, nomeadamente tendo em vista o extenso parque de obras de arte existente, a monitorização da integridade estrutural apresenta valiosas potencialidades. Ao longo do tempo de vida útil restante das estruturas, poderá ser acompanhado o seu comportamento e poderá ser avaliada a evolução do seu desempenho, permitindo projetar eficientemente, em conformidade com a efetiva condição estrutural, as intervenções de manutenção que se revelem necessárias.

Em termos gerais, uma danificação pode ser definida como uma mudança introduzida num sistema que afeta negativamente a sua atual ou futura performance. O conceito dano, nasce através da comparação entre dois estados: o primeiro é denominado o estado inicial, geralmente caracterizado como estado íntegro; o segundo é o estado atual da estrutura.

A danificação em estruturas começa pelo defeito no material, tendo todos os materiais diferentes percentagens de degeneração. Devido a vários tipos de acontecimentos, os defeitos vão crescendo, causando danos em componentes, e consequentemente, danificando a estrutura. À medida que o dano cresce, poderá chegar a um ponto onde é comprometido o estado operacional da estrutura. Este ponto é referido como falha. Os danos acumulam com o passar do tempo, associados também à fadiga e à corrosão do próprio material. A curto-prazo, os danos poderão surgir pela má utilização por parte do ser humano, ou por efeitos naturais, como por exemplo um terramoto.

O processo de implementar um sistema capaz monitorizar a condição de uma estrutura é referido como structural health monitoring (SMH). Este processo envolve a observação da estrutura, recorrendo a medições periódicas, a extração de relações *demage-sensitive* através das medições e de análises estatísticas, por forma a determinar o estado atual do sistema. Para um sistema SHM a longo termo, deverão ser realizadas inspeções periódicas, resguardando a sua boa funcionalidade, tendo em conta o inevitável envelhecimento e a acumulação de danos resultante das condições ambientais e de utilização. Num caso extremo, tal como o acontecimento de um terramoto, o sistema SHM é usado para verificar rapidamente o estado atual da estrutura [5].

1.2 Motivação

Atualmente, quase todas as organizações privadas e públicas pretendem detetar possíveis degenerações nas suas infraestruturas o mais rápido possível. Como tal, a importância dos sistemas SHM tem vindo a crescer consideravelmente, sendo motivada também pelos possíveis impactos económicos e pela possibilidade de garantir a segurança dos utilizadores das estruturas. Como por exemplo, companhias aeroespaciais juntamente com o governo estão a investigar a tecnologia SHM para identificação de danos em naves espaciais escondidas pelos protetores térmicos. Finalmente, muitas partes das infraestruturas técnicas estão a exceder o seu tempo de vida. Devido a problemas económicos, muitas dessas estruturas estão a ser utilizadas apesar do seu mau estado, realçando ainda mais a relevância da habilidade de monitorizar estruturas.

Muitas das manutenções (estruturais e mecânicas) de um sistema são feitas com base no tempo. SHM é a tecnologia que permitirá evoluir para uma filosofia de manutenção baseada na condição da estrutura. Este tipo de manutenção, consiste num sistema de sensores que monitoriza a resposta do sistema e que notifica o operador caso detete uma anomalia. Os benefícios de segurança de vida e económicos associados a esta filosofia, apenas serão concretizados se o sistema conseguir avisar antes que o dano se torne numa falha da estrutura. Este tipo de implementações requer um *hardware* de monitorização e um procedimento de análise de dados mais sofisticados.

1.3 Objetivos

Este documento foca-se no desenvolvimento de um sistema capaz de detetar e classificar danos em estruturas. Este processo é chamado de *Structural Health Monitoring* (SHM).

O desenvolvimento de um sistema SHM envolve 3 etapas fundamentais. A primeira é a normalização de dados, onde estes são tratados na tentativa de remover os efeitos ambientais, chamados efeitos dinâmicos, ou seja, remover os efeitos que vêm com a temperatura e a humidade. Posteriormente, pode haver a necessidade de isolar outros efeitos, tal como os ruídos, melhorando a qualidade do futuro modelo preditivo.

A segunda etapa consiste na modelação de um modelo preditivo capaz de prever com precisão as respetivas variáveis estruturais do sistema físico em analise. Naturalmente, que associada às previsões realizadas, estará associado um desvio, determinado através da diferença entre o valor previsto e o observado, levando assim à terceira e ultima etapa. Esta etapa consiste na deteção de danos através da informação produzida pela analise residual. A analise anterior permitirá a deteção de padrões comportamentais anormais baseados na amostras recolhidas. Técnicas como *Hotelling Control Charts e Cluster Analysis* irão ser usadas para deteção de anomalias.

1.4 Estrutura da Dissertação

Para além da introdução, esta dissertação contém mais 5 capítulos. No capítulo 2, é descrito o estado da arte e são apresentados trabalhos relacionados. No capítulo 3 é apresentada a me-

todologia utilizada durante o trabalho realizado. No capítulo 4 são apresentados os resultados obtidos. Finalmente, no capitulo 5 são apresentadas conclusões sobre o trabalho realizado, e são apresentadas algumas implementações que poderão ser realizadas no futuro.

Capítulo 2

Revisão Bibliográfica

2.1 Data Mining

A sociedade de hoje em dia cada vez mais vive rodeada de infraestruturas que são usadas regularmente, portanto é de extrema importância a monitorização da integridade estrutural das mesmas. Com os avanços da sensorização, aquisição de dados, computação, comunicação, e gestão de dados e informação fez com que seja possível a criação de sistemas capazes de acompanhar o comportamento e avaliar o desempenho estrutural, permitindo projetar de forma eficiente as intervenções de manutenções que se revelem necessárias fazer [6]. Durante a analise dos valores obtidos pela sensorização é recorrente termos uma vasta quantidade de dados, inclusive dados que não são relevantes para o pretendido, deste modo, é necessário recorrer a técnicas *data mining* (DM). DM é uma dos passos chaves no *Knowledge Discovery in Databases*. Este processo tem a capacidade de descobrir correlações, padrões e tendências através de uma grande quantidade de dados, recorrendo a tecnologias de reconhecimento de padrões e técnicas estatísticas e matemáticas [7].

Normalmente, DM é constituído por dois tipos de modelação que são aplicados na fase de modelação: um modelo descritivo que trata do reconhecimento de padrões e de relações no dados recolhidos pela monitorização; um outro modelo, intitulado de preditivo que é responsável pela previsões de dados. Algumas das técnicas mais conhecidas são:

- *Clustering*, trata-se de um modelo descritivo e consiste em dividir as amostras em grupos correlacionados;
- Previsão, este método preditivo determina padrões, regras ou modelos para prever um determinado valor que pode ser usado para outras funções;
- *Association Rule Discovery*, método descritivo que consiste em detetar relação entre as regras de modo a detetar uma ocorrência com base noutras occrências;
- Classificação, esta técnica é usada quando é necessário determinar a que subconjunto é que determinadas novas observações pertencem. A classificação, através do conjunto de dados de treino cria um padrão e usa esse mesmo padrão para classificar os novos dados [8].

2.2 Classificação e regressão baseada em *Data Mining* para a detecção de danos em estruturas

Observou-se nos últimos anos um grande avanço na investigação da temática da mineração de dados. Um exemplo desta temática diz respeito às técnicas de construção de modelos preditivos. Foram realizados, até ao momento, inúmeros trabalhos e estudos baseados em *Data Mining* para previsão de dados, mas quando chega a altura de escolher uma técnica de previsão, o leque de opções é reduzido. Os algoritmos baseados em *data mining* podem ser divididos em técnicas de aprendizagem supervisionadas (tais como redes neuronais artificiais, regressão linear, classificação, *support vector machine, random forest, redes neuronais artificiais, etc.*) ou não supervisionadas (tais como *k-means clustering, principal component analysis, fuzzy logic, etc.*). A grande diferença entre as abordagens, é que na supervisionada os valores a prever são conhecidos, contrariamente às não supervisionadas. Esta distinção ajuda a sumariar a categorização entre as diversas técnicas *data mining* e as suas aplicações em SHM [7].

Tendo em conta que o objetivo deste trabalho é desenvolver um modelo de regressão nomeadamente, um algoritmo que permita prever valores de determinadas variáveis relacionadas com a condição das estruturas, são apresentados mais abaixo, de forma pormenorizada, alguns algoritmos que o permitem fazer.

2.2.1 Redes Neuronais Artificiais

O cérebro humano é estudado há dezenas de anos. Assim, e com os avanços na tecnologia, é natural que houvesse a tentativa de replicar o mesmo processo de pensamento na sua forma inteligente. Como tal as redes neuronais foram criadas.

Redes neuronais são sistemas adaptativos criados com a finalidade de simular a forma como os neurónios humanos estão ligados. Esta característica é de facto distintiva, onde "aprender com exemplos"substitui a tradicional "programação"para resolver problemas. Esta característica tornou o modelo RNA bastante apelativo no domínio das regressões, pois nas situações em que o problema não é totalmente compreensível, a rede permite resolver o problema apenas com os dados de treino [9].

O modelo computacional RNA é capaz de resolver funções através de reconhecimento de padrões. Geralmente, os modelos RNA são utilizados para reconstruir uma aprendizagem de relação não linear através do treino da mesma.

Genericamente, as redes neuronais são constituídas por duas partes: unidade de processamento (neurónios), localizados na camada de rede, e conexão entre os elementos. Por sua vez, as camadas de rede são dividas em: camada de entrada (*input layer*), camada escondida (*hidden layer*) e camada de saída (*output layer*), tal como é ilustrado na figura 2.1 [10].

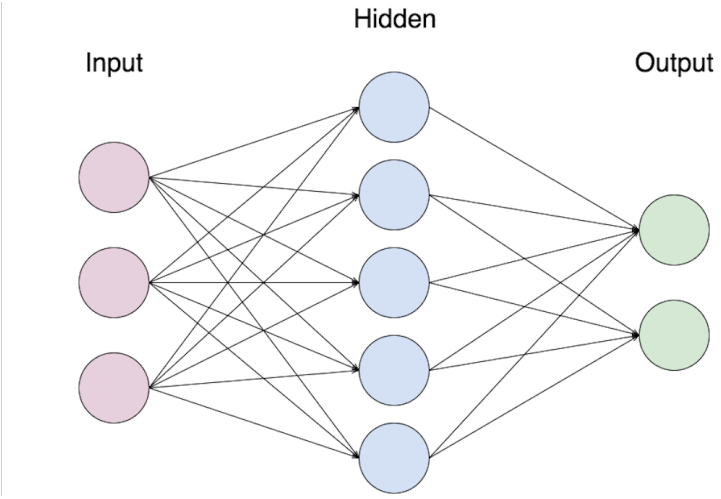

Figura 2.1: Exemplo de uma rede neuronal

Há uma diversidade enorme de tipos de redes neuronais. O tipo de modelo define-se pela estrutura interna e pela maneira como os neurónios estão conectados entre si. De seguida irão ser apresentados os tipos de redes neuronais mais conhecidos:

- *Feedforward neural network*, este tipo de redes, tal como o próprio nome indica, consiste numa propagação para a frente da informação, sendo o fluxo unidirecional [11]. Geralmente, são empregadas em problemas de classificação de padrões;

- Recorrentes, este tipo de redes foram desenvolvidas para lidar com informação sequencial. As redes neuronais recorrentes incorporam variáveis de estado que guardam a informação passada, que juntamente com as entradas atuais determinam as saídas do sistema. Em alguns casos os valores de ativação passam por um processo de relaxamento até atingirem um estado estável. Alguns exemplos deste tipo de redes são as redes neuronais auto associativas [12].

Uma aplicação prática das redes neuronais foi apresentada no artigo [12](2018). Neste estudo, foram implementadas duas redes neuronais: a primeira para detetar a presença de danos e a segunda para localizar os mesmos. Para o primeiro objetivo, um conjunto de frequências obtidas a partir de um estado saudável da estrutura foram usadas como valores de entrada. A rede neuronal auto associativa (AANN) é depois implementada. Uma vez treinada, esta rede retorna valores semelhantes aos valores da entrada da rede, desde que a entrada tenha sido obtida através de um sistema integro. Assim se frequências de um sistema danificado forem introduzidas na rede, esta nunca poderá retornar valores idênticos na saída, já que a realidade não está em conformidade com as suposições assumidas. Para ter em consideração a temperatura ambiente, a temperatura foi usada como input para a rede, juntamente com o conjunto de frequências obtidas a partir de um estado saudável da estrutura. Para a localização de danos, foi implementada uma segunda rede neuronal. A partir de experimentações numéricas, identificou-se que, ao contrário do impacto dos danos locais em variáveis globais, exemplo das frequências modais, o impacto na previsão

do erro da AANN é preponderante e consistente. Foi observado também que, cada cenário de dano corresponde a um único padrão na previsão de erro da AANN, fazendo com que seja possível identificar o dano seguindo apenas o padrão correspondente. No caso prático deste estudo, as variáveis dependentes consideradas foram as propriedades modais, que foram medidas para diferentes condições de temperatura, já a variável independente considerada foi a temperatura.

2.2.2 *Support Vector Machine*

Com o decorrer dos anos, cada vez mais se tem usado esta técnica baseada em aprendizagem estatística. Tal como as redes neuronais, esta passa também por uma fase de treino em que o modelo é alimentado com conjuntos de valores de entrada (variáveis independentes) e valores alvo de saída (variável dependente) [1]. Este modelo tem duas ideias chave. A primeira é um classificador de margem ideal, ou seja, é um classificador linear que constrói um hiperplano de separação, onde a distância entre o positivo e o negativo é maximizado, que no caso de uma estrutura seria por exemplo estrutura danificada no caso negativo e estrutura não danificada no caso positivo. A segunda ideia chave consiste no uso de funções *kernel*. A função *kernel* faz o produto escalar entre dois vetores, que aplicando um mapeamento não linear adequado do *kernel* aos dados iniciais os torna linearmente separáveis no espaço de recurso de alta dimensão, apesar de não o serem no espaço da entrada original (ver figura 2.2) [1] [13].

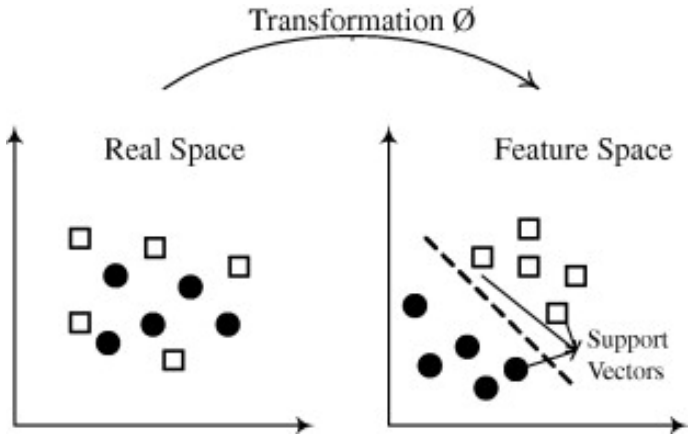

Figura 2.2: Exemplo de uma transformação SVM (adaptada de [1]

No caso de estudo analisado [6] é utilizada a técnica *principal component analysis* juntamente com o método *support vector regression* (SVR), e teve como principal objetivo modelar a variabilidade frequência-temperatura das frequências modais para estruturas que utilizam sistemas de monitorização. O estudo realizado consistia em comparar a performance de dois modelos SVR. Um modelo SVR foi desenvolvido utilizando dados comprimidos através de PCA e a sua performance foi comparada com o modelo treinado diretamente através dos dados originalmente medidos em termos de precisão de modelos e custos computacionais. No caso prático, foi utilizada a temperatura como variável independente e as frequências medidas através dos vários sensores (por exemplo, acelerómetros, transdutores, etc) foram consideradas como variáveis dependentes.

Para o modelo SVR foi treinado utilizando dados comprimidos por PCA. Em primeiro lugar foram extraídos os PCs da temperatura, que juntamente com os dados da frequência medida foram inseridos num algoritmo vetor de suporte para formular o modelo SVR. Uma vez que a performance de um modelo SVR depende da escolha os hiperparâmetros, usou-se um método de pesquisa em grelha com validação cruzada, seguido de um método heurístico, enquanto que os PCs eram selecionados por tentativa erro. Através de um ciclo, o modelo SVR foi construído utilizando dados de treino e os erros de previsão foram calculados comparando o valor previsto e o valor alvo para os dados novos de validação seguindo um esquema de validação cruzada K-fold, evitando assim o problema de *overfitting*. O conjunto de dados medidos são divididos equitativamente. Cada sub-conjunto é testado alimentando o conjunto de dados de treino separadamente para cada combinação de hiperparâmetros. O processo foi repetido até que todos os sub-conjuntos fossem testados. Seguidamente, calculou-se o erro quadrático médio (MSE) para cada sub-conjunto testado. O hiperparâmetro do SVR a ser utilizado foi assim determinado, correspondendo ao valor mínimo do MSE.

2.2.3 *Random Forest*

RF é umas técnicas de mineração de dados mais utilizada, para a sua popularidade contribui a simplicidade do método, e a fácil compreensão dos resultados obtidos.

Uma árvore de decisão mapeia os resultados possíveis para uma série de decisões tomadas. Esta apresenta os seus resultados numa estrutura regressiva.

Geralmente uma árvore de decisão começa com um único nó, que, consoante o atributo, se divide em vários resultados. Cada atributo representa um teste a ser feito dando origem a um resultado. Para cada possível resultado é originada uma nova sub árvore, e assim sucessivamente, sendo que cada resultado é independente.

Após implementada a árvore, a previsão de um determinado item é feita através da navegação pela árvore, começando na raiz, até à ultima folha, sendo que o resultado apresentado na última folha representa a previsão desse mesmo item [14].

Na figura 2.3 podemos observar um exemplo bastante simples de uma árvore de decisão. A decisão seria se "naquele dia devo ir jogar futebol ou não"mediante os três atributos desse dia em questão. Como podemos analisar através da imagem a decisão é sim ou não, constituindo assim as classes possíveis, os atributos, como será fácil de compreender são, se aparenta fazer sol naquele dia, se aparenta que o tempo estará nublado, ou se aquele dia tem aspeto de que choverá. Um conjunto de atributos sucede-se a cada decisão tomada. Por exemplo, o dia aparenta que virá chuva, com a chuva virá vento forte? Virá vento fraco? Se a probabilidade de estar vento forte for elevado, a decisão a tomar será não ir jogar futebol. Pelo contrário, se estiver vento fraco a decisão será ir jogar futebol. Neste acontecimento, os atributos Temperatura e Humidade não são considerados, pois são desnecessário para classificar este exemplo.

De salientar, que o modelo apresentado é bastante simples e não modela um problema real, as árvores podem, de facto, atingir valores elevados de complexidade, consoante o número de variáveis e classes. De realçar, que o *layouts* apresentado para árvores de decisão nem sempre são

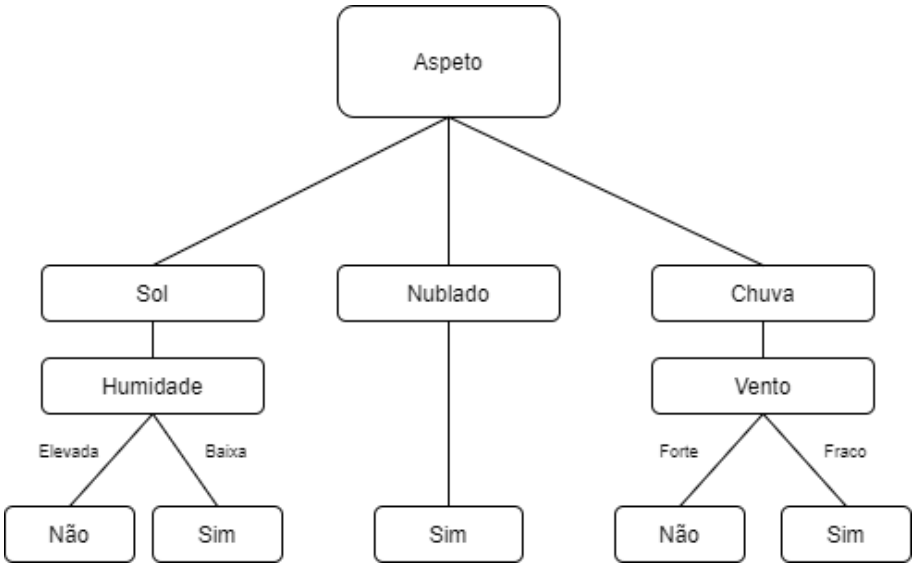

Figura 2.3: Exemplo simples de um modelo de uma Árvore de decisão

apresentados como no exemplo, é frequente apresentar o valor relativo de cada classe, ou seja, a probabilidade de ocorrência.

A construção de um modelo de árvore de decisão é feita através de dados de treino, ou seja, como já foi referido anteriormente, é um método de aprendizagem supervisionada. A construção é feita de forma iterativa ao dividir o conjunto de dados. A árvore forma-se dependendo dos testes realizados para cada variável. O processo de partição está completo quando todas as observações de um determinado nó pertencem à mesma classe (tem o mesmo valor para a variável alvo), ou quando o processo de divisão já não origina um aumento de valor nas previsões.

Contudo, as árvores de decisão carregam um alto risco de *overfitting*, visto que não conseguem voltar atrás depois de uma divisão de um certo nó. Como solução, foram criadas as *Random Forest*, que têm a capacidade de combinar várias árvores de decisão num modelo único, como se pode observar pela figura 2.4.

Random Forest é uma técnica que combina os métodos de aprendizagem regressão e classificação.

As árvores na RF correm em paralelo e não existe interação entre cada uma das árvores enquanto estão a ser construídas.

A construção de um modelo RF, tal como as árvores de decisão, começa pelo treino do modelo, construindo assim uma série de árvores. A saída do modelo será a média (por vezes ponderada) das previsões de cada árvore.

As *Random Forest*, como já foi referido anteriormente, combina o resultado de múltiplas previsões, agregando várias árvores de decisão. Deste modo, o modelo possui uma série de características, entre elas:

- o número de partições de cada nó é limitado a uma certa percentagem de um total, esse total é conhecido como hiperparâmetro, assegurando assim que o modelo não dependa de

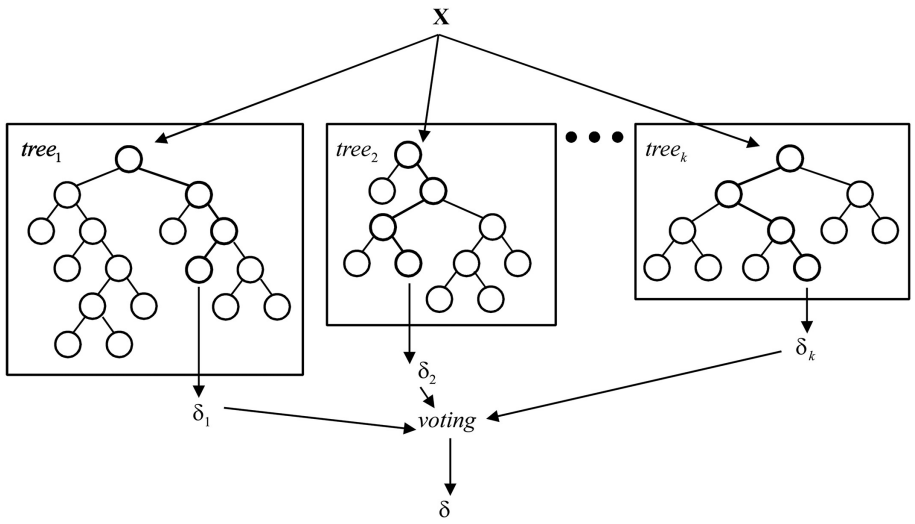

Figura 2.4: Estrutura de um modelo *random forest* (retirada de [2])

nenhum atributo, distribuindo assim a importância de todos o parâmetros potencialmente preditivos;

- cada árvore, gera aleatoriamente uma amostra ao conjunto de dados original quando é gerada uma divisão de um certo nó, adicionando assim um elemento de aleatoriedade reduzindo a possibilidade de *overfitting*.

No caso prático estudado [2] usou-se o método RF como ferramenta de regressão. Treinou-se e aplicou-se a RF para prever os deslocamentos concretos de uma barragem e estabelecer a monitorização da deformação da mesma. No artigo analisado usou-se como entradas do modelo os coeficientes relacionados com a componente hidráulica (por exemplo, pressão), com a componente térmica e com a componente de envelhecimento. O estudo realizado tem como objetivo mostrar a evolução de um modelo estatístico para para um modelo otimizado *random forest regression* (RFR). Para desenvolver um modelo otimizado RFR, o modelo estatístico é usado para estabelecer as variáveis input, para selecionar os parâmetros apropriados do Mtry e Ntree e para extrair fortes variáveis explicativas. O modelo RF consegue medir a importância das variáveis para a previsão da deformação. A vantagem do metodo RFR é que consegue extrair fatores representativos que influenciam o comportamento da estrutura baseado na importância da variável.

Primeiro dividiu-se os dados da monitorização em conjunto de teste e de treino, de seguida calculou-se a correlação de cada variável e multiplicou-se os coeficientes da regressão e os correspondentes fatores influenciáveis, que viriam a servir de inputs para o RFR. Selecionou-se os parâmetros RFR iniciais, de acordo com a dimensão do vetor de entrada. A previsão da deformação da barragem foi definida como a média de cada árvore de decisão. De seguida calculou-se a importância de cada variável de entrada, retirou-se as variáveis menos importantes e repetiu-se o passo de calculo da previsão, obtendo assim a previsão otimizada.

2.2.4 *Principal Component Analysis*

Esta não é uma técnica de regressão ou classificação, é um método de analise de dados, que é usado como ferramenta de redução de dimensionalidade dos dados da amostra. O conceito principal é transformar uma grande dimensão de variáveis correlacionadas numa pequena dimensão de variáveis não correlacionadas, através de uma projeção ortogonal, estas variáveis são conhecidas como *principal components* (figura 2.5) [10].

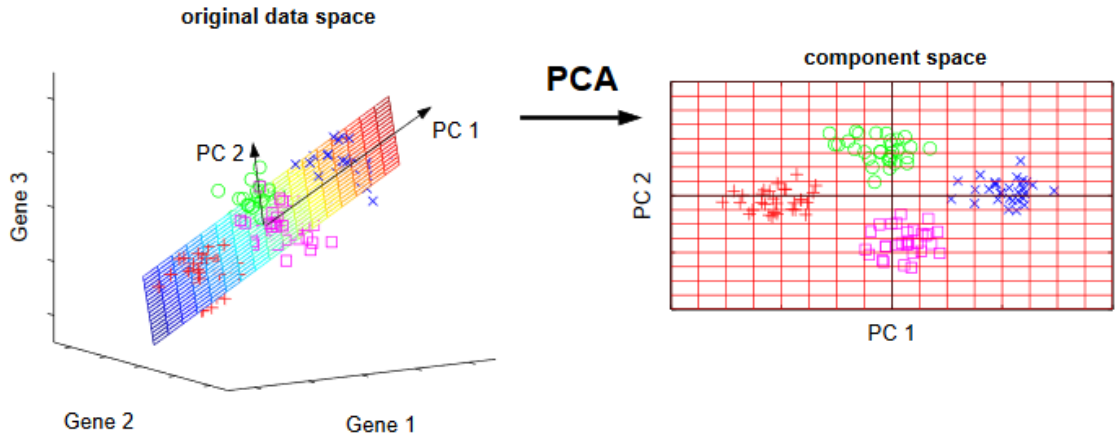

Figura 2.5: Exemplo de uma transformação PCA (retirada de [3])

Em diversos casos de estudo foi usada esta técnica, sempre com o objetivo de reduzir os dados da amostra. No artigo [15] após normalizar os valores, foi adotada a técnica PCA.

Consegue-se também redefinir cada propriedade e obter novos valores através do PCA. Estas varáveis são independentes entre elas, tornando mais preciso o treino do modelo. Neste estudo, usou-se o PCA para reduzir a quantidade de variáveis para 200 para, posteriormente, os introduzir na rede neuronal.

2.2.5 *Fuzzy Logic*

Consiste em 4 componentes importantes: *fuzzyfication, fuzzy rule-base, fuzzy inference e defuzzification*. *Fuzzification* consiste em mappear as entradas em conjunto de *memberships*. *Fuzzy rule-base* consiste em definir um conjunto de regras às variáveis *fuzzy* e descrever como funções *membership*. *Fuzzy inference* é o mecanismo de decisão propriamente dito. O *defuzzifier* altera as consequências do fuzzy de regras para valores (figura 2.6). Este método é regularmente usado juntamente com redes neuronais artificiais [10].

[16] propuseram uma abordagem que combina algoritmos de treino não supervisionados de redes neuronais e *fuzzy* para detetar e classificar os danos por níveis.

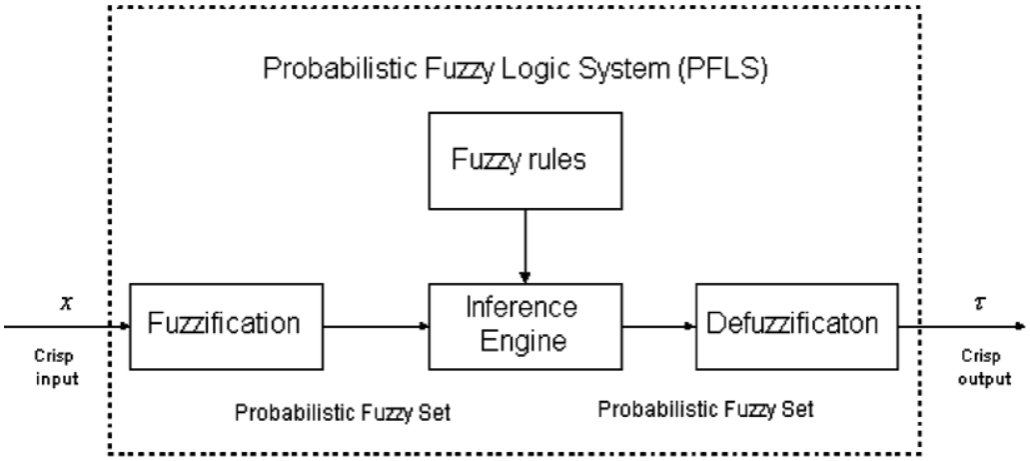

Figura 2.6: Exemplo de um sistema *fuzzy* (retirada de [4])

2.2.6 *Bayesian Analysis*

Método estatístico aplicado ao reconhecimento de padrões. É um método de classificação baseado no teorema de Bayes's. É usado para endereçar incertamente dados e modelar erros numa estrutura de deteção de danos.

No artigo referido anteriormente, [16], foi usada a técnica de *Bayesian* juntamente com *fuzzy*, ou seja, na partição dos potenciais niveis de dano da estrutura em conjuntos *fuzzy* usou-se a técnica de *Baye's*.

Capítulo 3

Metodologia

Neste estudo, usou-se uma estrutura real, para dar suporte ao desenvolvimento de modelos SHM. Este capítulo tem como objetivo descrever a estrutura utilizada e detalhar os modelos utilizados.

3.1 Estrutura

Tal como referido anteriormente, foi utilizada uma estrutura real para dar suporte na recolha de dados. Este modelo físico comporta-se tal como uma torre, ou um pilar de uma ponte com a secção a variar consoante a altura. O modelo é composto por duas barras de ferro, unidas por chapas soldadas, sendo a barra inferior mais fina que a barra superior. A estrutura foi equipada com uma série de sensores, tais como acelerómetros, extensómetros, inclinómetros, transdutores e termómetros. Na imagem 3.1 é ilustrada a disposição dos sensores. O posicionamento de cada sensor foi estrategicamente definido de acordo com a grandeza a ser medida.

A estrutura foi então dividida em 3 secções: a primeira denominada por secção inferior do modelo, localizada na base da estrutura, zona mais propícia a recolha de dados de extensões e rotações; a segunda secção foi definida no centro geométrico do modelo, e foi denominada como secção intermédia, zona propícia para a recolha do deslocamento horizontal, acelerações, rotações e extensões; a terceira secção é dada no nível superior do modelo, onde são de interesse a recolha do deslocamento horizontal, acelerações e rotações. Cada sensor foi responsável por captar medições de 15 em 15 minutos, perfazendo um total de 96 medições a cada 24h. O período de recolha teve a duração de 1 ano, começando em Abril de 2018 e acabando em Abril de 2019.

Durante o período experimental, foi necessário que o modelo estivesse submetido a ações ambientais. Além das ações ambientais, a estrutura foi exposta a dois tipos de cenários diferentes: um primeiro cenário onde a estrutura foi apenas influenciada pelas suas próprias propriedades estáticas e dinâmicas que mais tarde seriam usadas como referência (entre 27/04/2018 e 10/09/2018); um segundo cenário onde seria induzido um certo grau de dano (de 10/09/2018 até ao final do mês de Abril de 2019). A dano foi induzido através da ligação de um cabo de ferro à estrutura.

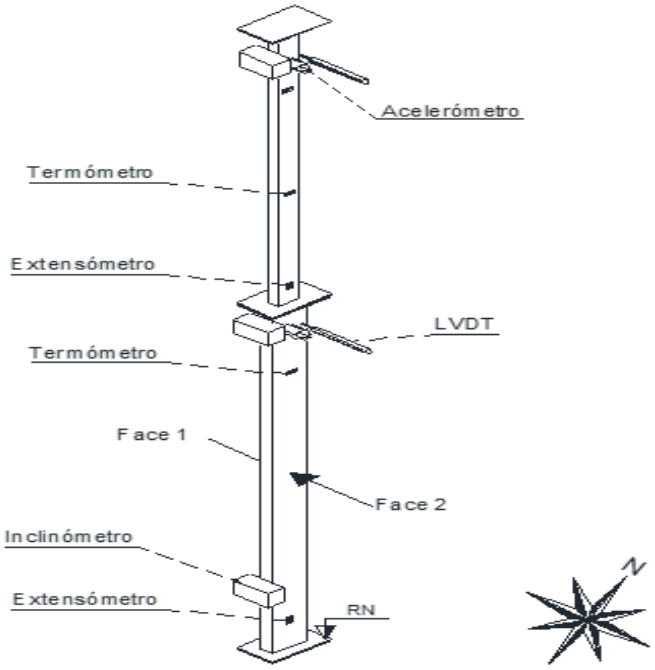

Figura 3.1: Estrutura física do modelo

3.2 Dados

Posteriormente, são apresentados os dados comparativos entre as grandezas medidas pelos sensores. Como configuração geral para a realização das previsões, temos as variáveis de saída e as variáveis preditivas, apresentadas a seguir.

- variáveis de saída, compreendem as grandezas medidas por 9 sensores, tais como rotações, extensões e deslocamentos. As variáveis de saída estão apresentadas na tabela 3.1.

Inclinómetro inferior
Inclinómetro intermédio
Inclinómetro superior
LVDT intermédio
LVDT superior
Strain face1 inferior
Strain face2 inferior
Strain face1 intermédio
Strain face2 intermédio

Tabela 3.1: Variáveis de saída

- variáveis preditivas, são as variáveis relacionadas com a temperatura. A temperatura medida no interior da caixa de escadas foi retirada por ter um comportamento dependente dos utilizadores do edifício, alterando a componente natural necessária As variáveis preditivas estão apresentadas na tabela 3.2.

Temperatura inferior face 2
Temperatura intermédia face 1
Temperatura intermédia face 2
Temperatura superior face 1
Temperatura superior face 2
Temperatura superior face 2 desprotegida
Temperatura ambiente

Tabela 3.2: Variáveis preditivas

Primeiramente, tal como referido anteriormente, os conjuntos de dados recolhidos foram divididos em dois sub conjuntos. O primeiro sub conjunto que contem os dados pré dano. O segundo sub conjunto que contem os dados pós dano. Este tipo de estratégia foi adotada por forma a simular um caso o mais real possível, onde a rede é treinada com dados livres de dano, de tal modo que futuramente consiga detetar valores anormais, pois são valores com os quais a rede não foi treinada. É expectável que a rede preveja valores diferentes dos observados durante o período pós dano.

3.3 Modelos preditivos

Depois de perceber e interpretar os dados recolhidos, os mesmos foram usados para dar suporte ao sistema de deteção de dano. Para o desenvolvimento do modelo preditivo, foram testadas três diferentes técnicas de previsão, as *random forest*, as redes neuronais e as *suport vector machines*.

Utilizando os valores medidos pelos sensores durante o período pré dano, tanto como variáveis estruturais como os valores medidos pelos termómetros, foi possível desenvolver um modelo preditivo de uma estrutura saudável que seria considerado como referência. Comparando os valores previstos com os esperados, é possível determinar qualquer comportamento anormal na estrutura.

Em primeiro lugar, para o desenvolvimento de um modelo preditivo, os diferentes tipos variáveis precisam de ser definidos. Como referido anteriormente, definiu-se a variáveis estruturais como variáveis de saída e as variáveis relacionadas com a temperatura como variáveis preditivas. Isto é, para cada variável de saída será criado um modelo específico, resultando assim num total de 9 modelos.

As variáveis relacionadas com a temperatura foram definidas como preditivas, uma vez que afetam diretamente as variáveis estruturais e fazem parte das propriedades dinâmicas da mesma.

Uma vez definidas as variáveis, temos a capacidade de prever os valores esperados de uma estrutura íntegra, podendo assim comparar com os valores observados. A diferença entre os valor esperado e o valor observado é designado de desvio.

Depois de calculado o desvio, o mesmo poderá ser utilizado como entrada para a carta de controlo *Hotelling* T^2. Consequentemente, podemos determinar se existe qualquer tipo de anomalia

na estrutura ou não [17]. Logicamente, o modelo prevê valores diferentes dos observados durante o período de dano. Adicionalmente, para o período pré dano, será expectável que os valores previstos sejam semelhantes aos observados.

O sistema proposto deve detetar anomalias, contudo deve evitar alertar com base em falsos positivos, ou seja, situações em que o modelo classifica como anormal e efetivamente não há dano. Tendo em mente este requisito, a capacidade do modelo preditivo foi testada no período de teste, representado na figura 3.2. Apenas os dados referentes ao período que antecede o período de teste foram usados para desenvolver o modelo preditivo. Deste modo, haverá uma fase de treino e validação do modelo, e depois haverá uma fase de teste dos modelos preditivos propostos. Na figura 3.2 é ilustrada, de forma mais detalhada, a divisão dos respetivos dados.

Os modelos preditivos foram desenvolvidos em 3 etapas. Na primeira etapa, foi desenvolvido um modelo considerando todas a variáveis preditivas. Dividiu-se os dados referentes ao período pré dano em treino, validação e teste, e apenas foram testados diferentes valores possíveis para os parâmetros das diferentes técnicas de aprendizagem. Após determinados os parâmetros, efetuou-se o treino com os dados de treino e validação em conjunto, utilizando os parâmetros selecionados, para depois realizar o teste para o período pré dano. A segunda fase passou por dividir aleatoriamente os dados do período pré dano, exceptuando os dados de teste, em 10 amostras diferentes. A técnica consiste em treinar o modelo com cada amostra, determinando ao mesmo tempo quais seriam os melhores parâmetros para cada amostra. Esta técnica tem como principal objetivo aumentar a precisão e evitar o *overfitting*. A ultima etapa passou pela tentativa de melhorar os resultados, tornando o modelo ainda mais preciso. Foi então implementada uma técnica chamada *forward selection*. O objetivo principal deste método de seleção é identificar e selecionar as variáveis independentes mais importantes e não redundantes do grande conjunto de potenciais variáveis [18]. A metodologia adotada foi bastante simples: primeiro era testada cada variável preditora independentemente, guardando o respetivo MAPE, de seguida era selecionada a variável com o melhor resultado. Depois de determinada a primeira variável, o processo para determinar a segunda variável seria exatamente o mesmo, e assim sucessivamente, até não ser encontrada uma variável cujo o MAPE seja melhor do que o anterior [19].

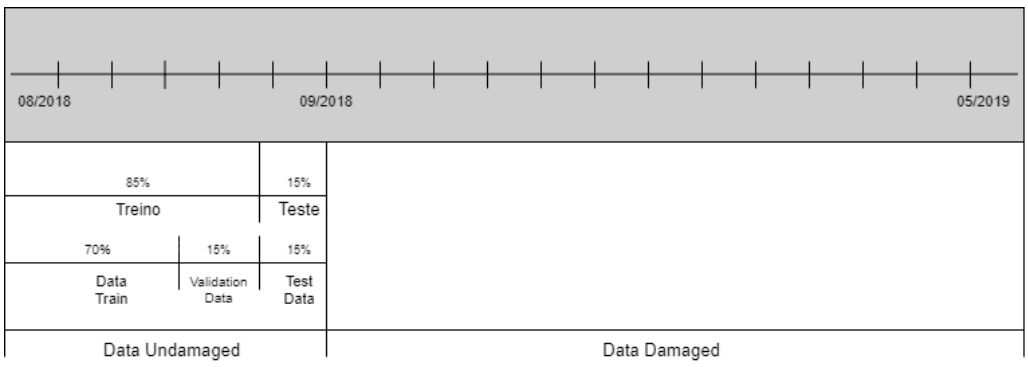

Figura 3.2: Divisão temporal dos dados observados

Em primeiro lugar, começou-se por desenvolver a técnica *random forest* utilizando 1000 árvores de decisão. Este valor deve ser tanto maior quanto maior for o conjunto de dados, sendo que o número de árvores de decisão estabiliza o erro associado ao modelo, contudo não deverá ser muito elevado, já que um valor elevado de árvores de decisão levará ao *overfitting*. O modelo RF foi desenvolvido com o parâmetro *Mtry* afinado. O valor *Mtry* define o número de variáveis selecionadas aleatoriamente a serem considerados em cada divisão, e naturalmente, este valor está limitado ao número máximo de variáveis preditivas. Em segundo lugar alimentou-se o modelo das redes neuronais com os dados e com um número de camadas escondidas e o peso do decaimento afinados. Em relação às camadas escondidas, a sua funcão é permitir estabelecer uma relação complexa e não linear entre as variáveis preditivas e a variável target. Quanto maior o número de camadas escondidas maior é eficácia na resolução de problemas complicados, porém, deverá haver um meio termo, porque um número elevado de camadas escondida poderá levar ao *overfitting*. Quanto ao peso de decaimento é um termo adicional na regra de atualização de pesos que faz com que os pesos se deteriorem exponencialmente a zero, se nenhuma outra atualização estiver agendada.

O processo de escolha dos parâmetros foi feito em conjunto com o treino, ou seja, a respetiva rede foi treinada utilizando todos os diferentes valores dos parâmetros, guardando o respetivo MAPE obtido nos dados de validação. Deste modo, os parâmetros escolhidos, para as respetivas técnicas de aprendizagem, seriam aqueles cujo o MAPE fosse o valor mais reduzido. O próximo passo seria testar o modelo num conjunto de dados ainda não observado. Deste modo, testou-se o modelo no conjunto de dados de teste durante o período pré dano. Tendo testado e considerado um bom modelo usou-se para prever para os valores pós dano. Os resultados obtidos devem ser precisos para o período pré dano, contrariamente ao que deverá acontecer para o período pós dano.

As métricas escolhidas para verificar a performance dos modelos preditivos foram o MAPE (percentagem de erro absoluto médio) e o MSE (erro quadrático médio).

Quanto às cartas de controlo, tal como foi referido anteriormente, usou-se a carta de controlo *Hotelling* T^2. A nova fase consiste no procedimento multivariado mais usado para controlar as alterações num vetor uniforme de características de qualidade correlacionadas [20]. Devido ao facto de se tratar de uma carta multivariada é possível acompanhar simultaneamente a evolução das variáveis de saída e determinar se a estrutura está danificada ou não.

Em resumo, a carta de controlo *Hotelling* T^2 considera a média de cada resíduo individual e uma matriz de covariância entre cada par de resíduos. Isto significa que qualquer mudança (como a que se espera vir a detetar no período pós dano) que resulte numa alteração da média e da variância dos desvios, deverá resultar numa subida no gráfico *Hotelling* T^2. Então um limite de controlo superior é criado e qualquer valor que ultrapasse esse limite será considerado como uma situação anormal em comparação ao resto dos dados.

A elaboração da carta de controlo divide-se em duas fases distintas. A primeira fase consiste em utilizar as amostras preliminares, ou seja, onde o processo estava sob controlo e utilizá-las como referência contra a qual observações futuras do processo serão comparadas, calculando limites. Na segunda fase, os parâmetros estimados anteriormente são comparados com as novas

observações.

No caso prático em estudo, o período correspondente à validação do modelo coincide com a primeira fase, enquanto que o período correspondente ao teste durante o período pré dano e ao teste durante período pós dano coincide com a segunda fase.

Depois de desenvolvida a carta de controlo, foi elaborado modelo, que consiste em determinar quanto tempo uma anomalia demora a ser detetada.

Capítulo 4

Resultados

4.1 Modelos preditivos

De modo a validar o uso dos modelos preditivos para detetar anomalias, é importante determinar a performance de cada técnica proposta. Naturalmente, cada técnica dará origem a resultados diferentes, que devem ser comparados.

Random Forest									
Variável dependente	Incl inf	Incl int	Incl sup	LVDT int	LVDT sup	Strain face1 inf	Strain face2 inf	Strain face1 int	Strain face2 int
Variáveis a prever	6	7	7	7	7	7	5	7	7
Mape(%)	0.57	0.62	0.65	116	12.71	2.90	12.28	2.63	3.57
MSE	68.82	89.36	115.97	0.0063	0.027	50.02	82.78	49.08	58.58

Tabela 4.1: Resultados obtidos para o modelo *Random Forest*

Redes Neuronais									
Variável dependente	Incl inf	Incl int	Incl sup	LVDT int	LVDT sup	Strain face1 inf	Strain face2 inf	Strain face1 int	Strain face2 int
Decay	0.046	0.01	0.215	0.01	0.464	0.1	1	0.001	1
Size	80	70	40	30	20	80	60	40	40
Mape(%)	0.55	0.70	1.10	88.49	14.26	2.88	13.14	2.74	4.02
MSE	69.25	156.8583	302.53	0.008	0.06	50.27	98.64	58.31	76.49

Tabela 4.2: Resultados obtidos no modelo Redes Neuronais

As tabelas apresentam as combinações de parâmetros que permitiram obter melhores resultados nos dados de validação, sendo que as métricas de desempenho respeitam ao período de teste. Como se pode verificar, nenhum dos modelos conseguiu prever com exatidão os valores da característica estrutural LVDT intermédio, consequentemente foi desconsiderada para efeitos de deteção de anomalias. Através os resultados obtidos (ilustrados nas tabelas 4.1 e 4.2) pode-se concluir que os modelos suportados pela técnica *Random Forests* obtiveram, em média, resultados ligeiramente melhores relativamente aos resultados obtidos através dos modelos sustentados pela técnica Redes Neuronais Artificiais, produzindo um MAPE médio de 4.5%. Contudo as duas técnicas produziram resultados muito semelhantes, obtendo um MAPE médio de 4.9% nas Redes Neuronais. Com base nos resultados obtidos, optou-se por utilizar as duas técnicas para explorar o potencial incremento de desempenho perante uma seleção de variáveis independentes a incluir nos modelos. Pode-se concluir também que as variáveis, para as quais foram obtidas previsões

mais precisas, foram as que se referem aos inclinómetros, seguido dos extensómetros. Por último as variáveis para as quais os modelos apresentaram com menor precisão foram referentes aos transdutores.

Depois do desenvolvimento dos modelos considerando todas as potenciais variáveis preditivas, foi implementa a técnica *forward selection*, que como já foi referido anteriormente, selecionaria as variáveis preditivas com maior potencial preditivo. Os resultados obtidos estão ilustrados nas tabelas das figuras 4.1 e 4.2.

Técnica	**Random Forests**								
Variável de saída	Inclinómetro Inf	Inclinómetro Int	Inclinómetro Sup	LVDT Int	LVDT sup	Strain_f1_inf	Strain_f2_inf	Strain_f1_int	Strain_f2_int
MAPE	0.56	0.70	0.75	129	12.86	2.96	12.92	2.69	3.59
MSE	70.2	124.8	163.1	0.007	0.03	51.3	87.9	52.1	60.6
Mtry	2	3	3	2	3	2	1	3	3
Variáveis Preditoras Selecionadas	Temp_int_f2 / Temp_sup_f1/ Temp_inf_f2/ Temp_sup_f2/ Temp_amb	Temp_inf_f2/ Temp_sup_f1/ Temp_amb	Temp_inf_f2/ Temp_sup_f1/ Temp_amb/ Temp_sup_f2	Temp_int_f2/ Temp_int_f1/ Temp_supr_f1	Temp_sup_f2_despr / Temp_inf_f2/ Temp_amb/ Temp_sup_f1	Temp_int_f1/ Temp_int_f2/ Temp_sup_f1/ Temp_inf_f2/ Temp_amb	Temp_inf_f2/ Temp_sup_f2_despr / Temp_int_f2	Temp_int_f1/ Temp_int_f2/ Temp_amb/ Temp_inf_f2	Temp_int_f2/ Temp_amb/ Temp_inf_f2/ Temp_sup_f1

Figura 4.1: Parâmetros de entrada, paramétricas de erro e as variáveis preditoras obtidas usando o método *Forward Selection* para cada variável dependente utilizando o modelo preditivo *Random Forests*

Técnica	**Neural Networks**								
Variável de saída	Inclinómetro Inf	Inclinómetro Int	Inclinómetro Sup	LVDT Int	LVDT sup	Strain_f1_infr	Strain_f2_inf	Strain_f1_int	Strain_f2_int
MAPE (%)	0.58	0.75	0.75	78.8	13.6	3.2	12.9	3.93	4.07
MSE	79.1	193.1	179.9	0.0075	0.055	61.1	92.1	234	74.8
size	80	70	30	40	10	80	40	60	60
decay	0.04641589	0.01	0.01	0.2154435	0.2154435	1	1	0.1	1
Variáveis Preditoras Selecionadas	Temp_int_f2/ Temp_sup_f1/ Temp_amb/ Temp_int_f1/ Temp_sup_f2_desp	Temp_int_f2/ Temp_int_f1/ Temp_sup_f2_desp/ Temp_amb/ Temp_sup_f1	Temp_inf_f2/ Temp_sup_f2/ Temp_sup_f1/ Temp_int_f1/ Temp_int_f2/ Temp_amb	Temp_inf_f2/ Temp_amb/ Temp_int_f1/ Temp_int_f2/ Temp_sup_f1/ Temp_sup_f2_desp	Temp_inf_f2/ Temp_sup_f1/ Temp_sup_f2_desp/ Temp_int_f1/ Temp_int_f2	Temp_int_f1/ Temp_int_f2/ Temp_inf_f2/ Temp_amb	Temp_int_f1/ Temp_sup_f2_desp/ Temp_inf_f2/ Temp_amb	Temp_int_f1/ Temp_int_f2/ Temp_amb/ Temp_inf_f2	Temp_int_f1/ Temp_int_f2/ Temp_amb/ Temp_sup_f1

Figura 4.2: Parâmetros de entrada, paramétricas de erro e as variáveis preditoras obtidas usando o método *Forward Selection* para cada variável dependente utilizando o modelo preditivo Redes Neuronais Artificiais

As tabelas apresentam as combinações de parâmetros e as variáveis preditivas selecionadas que permitiram obter melhores resultados nos dados de validação, sendo que as métricas de desempenho respeitam ao período de teste. Como se pode observar, os resultados não diferem dos anteriormente obtidos. Havendo este cenário, deve-se ao facto da técnica *forward selection* ser "gulosa". O método consiste na seleção de variáveis passo a passo. Este processo é iniciado com um modelo vazio e vai adicionando variáveis uma a uma. Em cada passo é adicionada a variável que mais potencia o modelo [19]. Esta foi a abordagem adotada, pois apresenta a melhor relação entre tempo e desempenho. Através dos resultados alcançados, será possível concluir que em geral, a técnica de aprendizagem *Random Forest* resulta num modelo ligeiramente mais preciso.

Como referido anteriormente, nenhum dos modelos se evidenciou no resultados obtidos através das métricas de erro. Assim foram desenvolvidos dois modelos, um que recorre ao algoritmo *Random Forest*, outro onde foi aplicado o método redes neuronais. Neste capitulo apresentam-se os resultados obtidos no algoritmo Random Forest para uma variável exemplo. Focando na variável Inclinómetro Inferior, o modelo atingiu uma previsão bastante alta. Como se pode verificar no

gráfico da figura, referente ao período de validação, 4.3 quando se compara os valores observados, a cor preta, e os valores previstos, representados pela cor azul.

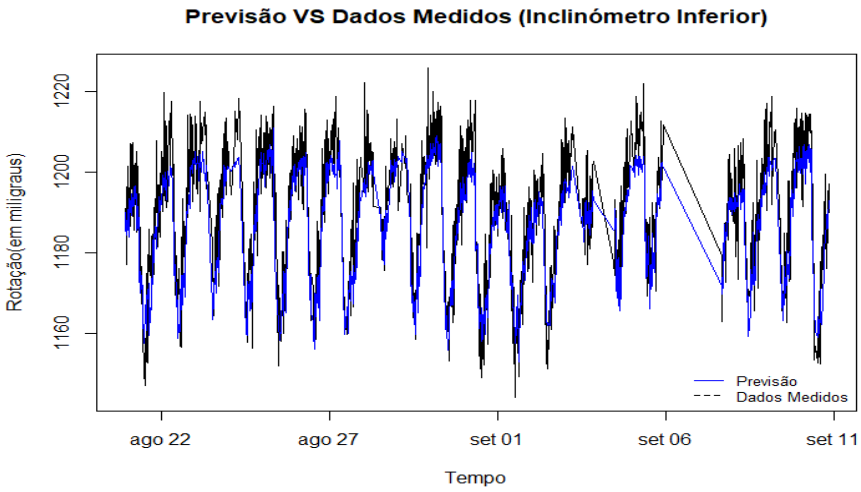

Figura 4.3: Comparação entre os valores observados e os valores previstos durante a fase de teste (período pré dano) utilizando o modelo preditivo *Random Forests*

Para o mesmo período de observação, pode-se observar que no gráfico da figura 4.5 o desvio entre os valores observados e os previstos é bastante reduzido, o que realça ainda mais a qualidade das previsões realizadas. Depois de validada a performance do modelo, é de extrema importância verificar se o algoritmo realmente é capaz de detetar anomalias ou não. Como se pode apurar através da análise do gráfico da figura 4.4, durante o período pré dano, as previsões estão alinhadas com os valores observados. Contudo, após o dano, as previsões, representadas a vermelho na imagem 4.4, não estão alinhadas com os valores observados. O mesmo se pode verificar na imagem 4.5, após a linha vertical representada a vermelho, período em que o dano é induzido. Pode-se constatar que o desvio entre os valores observados e os previstos sobe consideravelmente. Com o conjunto de resultados obtidos poderá ser desenvolvido um sistema que seja capaz de detetar anomalias.

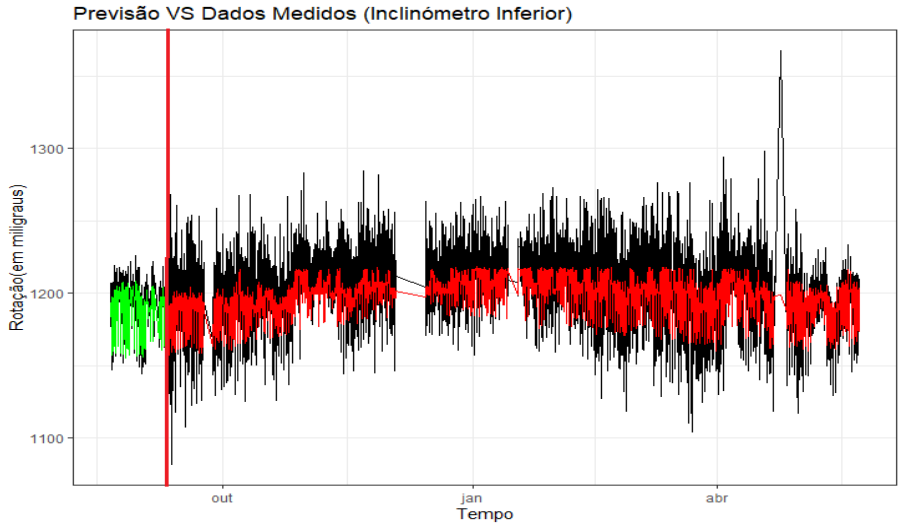

Figura 4.4: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

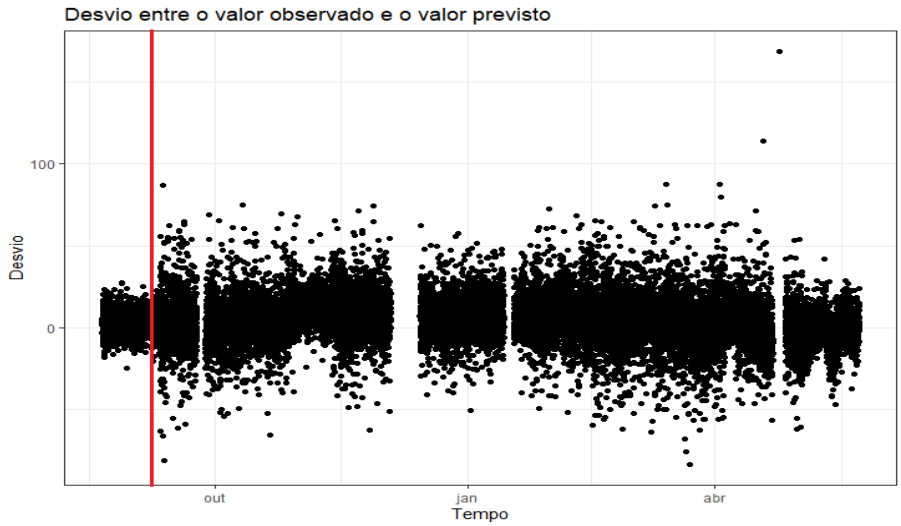

Figura 4.5: Desvio entre os valores observados (Inclinómetro Inferior)

Por forma a conceber um sistema que considere todos os desvios entre os valores previstos e observados, foi desenvolvido um sistema suportado pelas cartas de controlo *Hotelling* T^2. Ao examinar a figura4.6, pode-se verificar que durante o período de teste (pré dano), apenas alguns pontos ultrapassam os limites de controlo, precisamente 2.37% dos pontos.

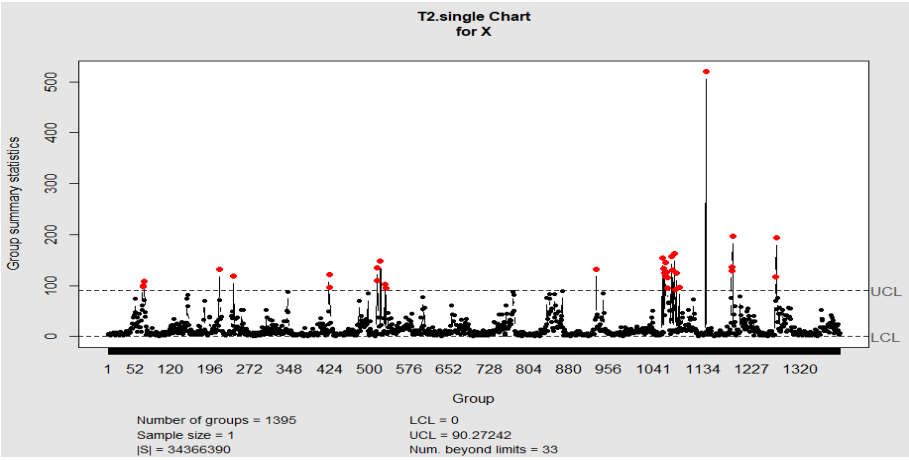

Figura 4.6: Carta de controlo *Hotelling* T^2 para o período de teste

A carta de controlo foi aplicada ao período de teste, bem como ao período pós dano. Como se pode verificar através da figura 4.7, o número de pontos acima do limite superior aumentou consideravelmente, nomeadamente 18.1% do pontos, o que evidência a eficácia da carta de controlo utilizada.

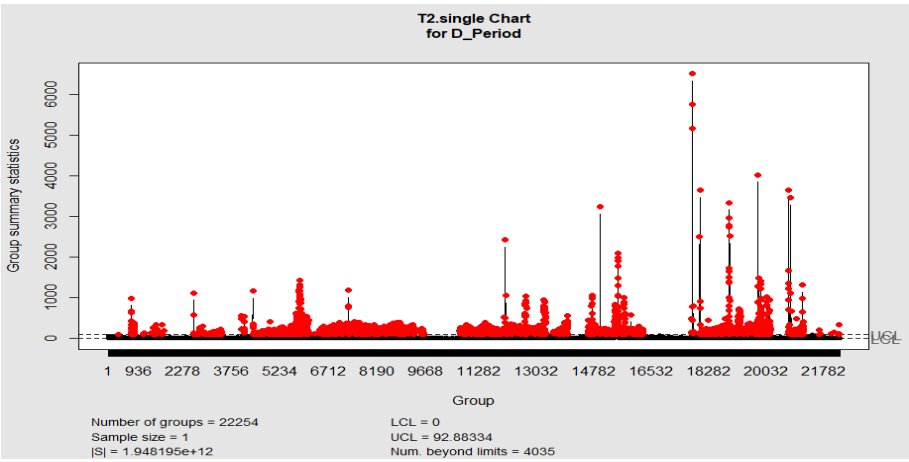

Figura 4.7: Carta de controlo *Hotelling* T^2 para o período pós dano

Do mesmo modo que foi desenvolvida a cartas de controlo para o modelo *Random Forest*, foi elaborada a carta de controlo *Hotelling* T^2 para o modelo computacional Redes Neuronais Artificiais. Como se pode verificar nas imagens apresentadas, os resultados entre os dois modelos foram bastante semelhantes. Durante o período de teste pré dano (figura 4.8, o modelo deteta que 6% dos pontos estão acima do limite superior, sinalizando assim como anomalia. Quanto ao resultado apresentado durante o período pré dano, pode-se concluir que as previsões realizadas pelo

método *Random Forest* são mais precisas que as realizadas pela técnica RNA, visto que ao longo do período pré dano, a percentagem de pontos acima do limite superior deveria ser aproximadamente zero. O mesmo pode ser concluído a partir da análise do gráfico da carta *Hotelling* da figura 4.9, pois ao detetar cerca de 14.4% de pontos acima do limite superior, continua a ser inferior à percentagem obtida pelas *Random Forest*.

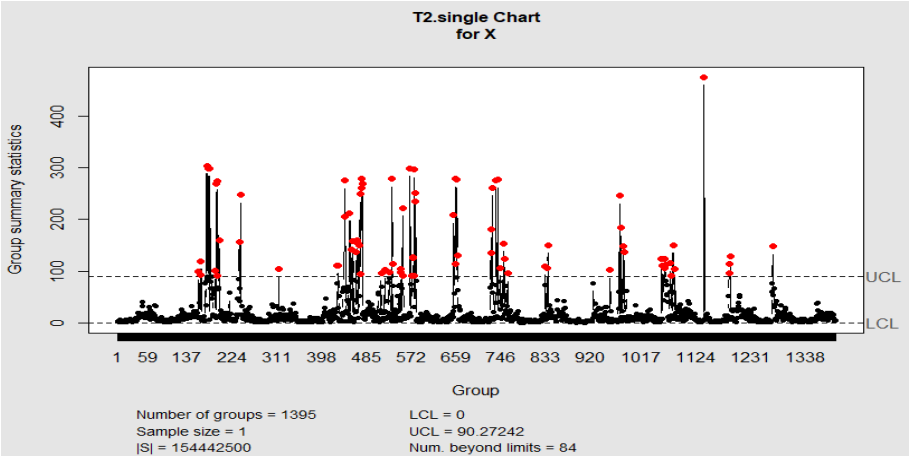

Figura 4.8: Carta de controlo *Hotelling* T^2 para o período pré dano utilizando a técnica Redes Neuronais

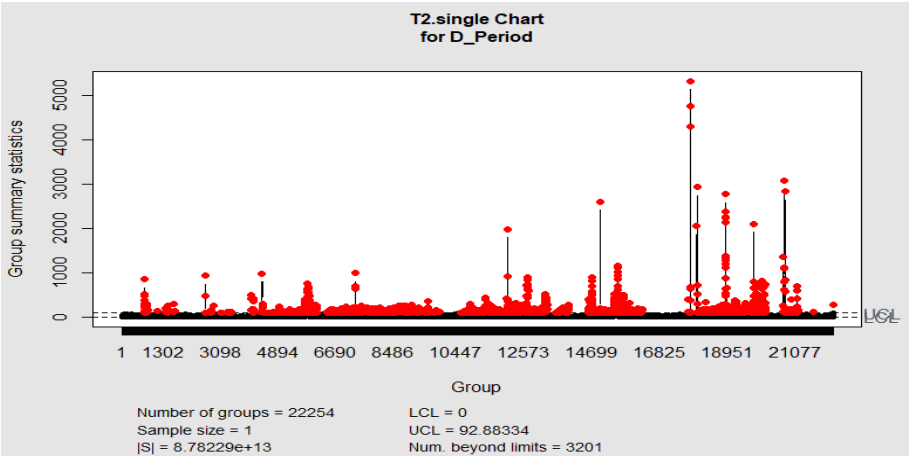

Figura 4.9: Carta de controlo *Hotelling* T^2 para o período pós dano utilizando a técnica Redes Neuronais

Depois de construídas as cartas de controlo para cada técnica, é importante calcular quanto tempo demora uma anomalia a ser detetada, ou seja, o primeiro ponto acima dos limtes de controlo. Para o modelo RNA o tempo de deteção foi 7 dias 6 horas e 15 min, quanto o modelo RF o tempo de deteção foi de 3 dias 6 horas e 45 min, evidenciando assim as qualidades de previsão do modelo RF.

Capítulo 5

Conclusão e trabalho futuro

Ao longo do capítulo 5 serão apresentadas as conclusões, assim como possíveis melhorias a serem concretizadas no futuro, de forma a completar todo o trabalho realizado ao longo da dissertação e responder às limitações do modelo atual.

5.1 Conclusão

Sendo a monitorização da integridade de estruturas, um tema cada mais importante na sociedade atual é de extrema importância o desenvolvimento de técnicas capazes de detetar danificação, antes que se tornem irreparáveis. Este facto contribui para a crescente popularidade, não só em trabalhos académicos como no próprio meio empresarial. O sistema deve ser robusto e deve evitar alertar sobre falsos positivos.

O modelo foi desenvolvido a partir de um contexto da vida real, utilizando como suporte uma estrutura semelhante a uma ponte. A estrutura foi submetida a um período livre de dano e posteriormente a um período onde foi induzido um certo grau de dano. De salientar, que a estrutura demonstrou um comportamento satisfatório durante as diferentes fases de monitorização, atendendo à expectativa de representação de uma estrutura real.

As técnicas de deteção de dano apresentadas demonstraram um desempenho satisfatório. Os modelos desenvolvidos mostraram-se capazes de identificar mudanças de padrão comportamentais da estrutura. Tendo em conta as técnicas de aprendizagem apresentadas, a técnica *Random Forest* demonstrou mais qualidade nas previsões realizadas do que a técnica Redes Neuronais. A implementação do método *forward selection* revelou resultados satisfatórios, sendo a abordagem com a melhor relação qualidade/tempo.

Finalmente, foram implementadas cartas de controlo. Estas demonstraram um desempenho satisfatório. Primeiro foram implementadas para o período pré dano, onde apenas se detetaram alguns picos para ambas as técnicas, posteriormente foram implementadas para o período pós dano, permitindo sinalizar uma mudança considerável no comportamento da estrutura.

5.2 Trabalho futuro

Ao longo da dissertação foram concretizados diversos avanços ao nível da modelação do comportamento regular de uma estrutura e ao nível da deteção de anomalias. No entanto, no futuro poder-se-ão implementar algumas melhorias.

No futuro poder-se-á investir em aplicar algumas técnicas de tratamento de dados de forma a permitir uma maior qualidade na previsão. Além disso, poder-se-á considerar a utilização de outros algoritmos preditivos, nomeadamente redes neuronais recorrentes. Admite-se ainda que este trabalho poderá ser estendido, dando lugar a um sistema de alarmes que possa ajudar os gestores da manutenção de estruturas físicas.

Anexo A

Anexos

A.1 Resultados do modelos preditivos

A.1.1 Random Forest

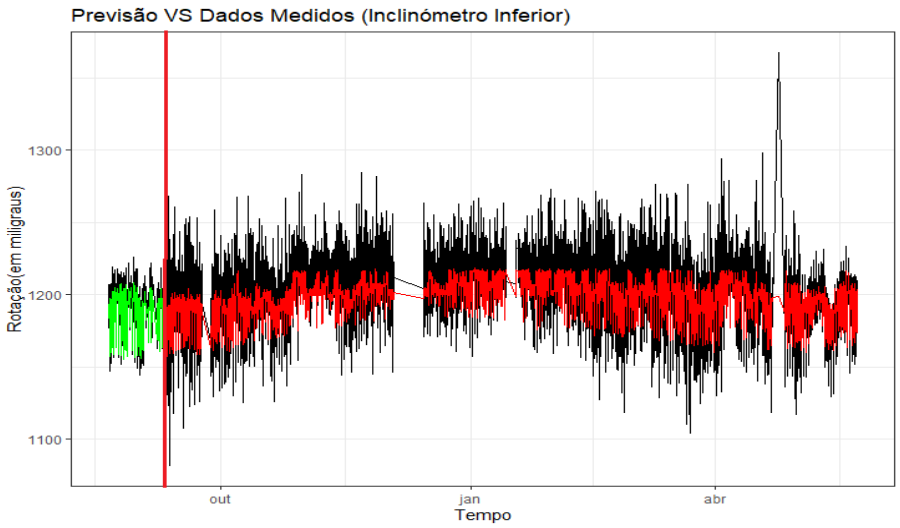

Figura A.1: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

A.1.2 Redes Neuronais Artificiais

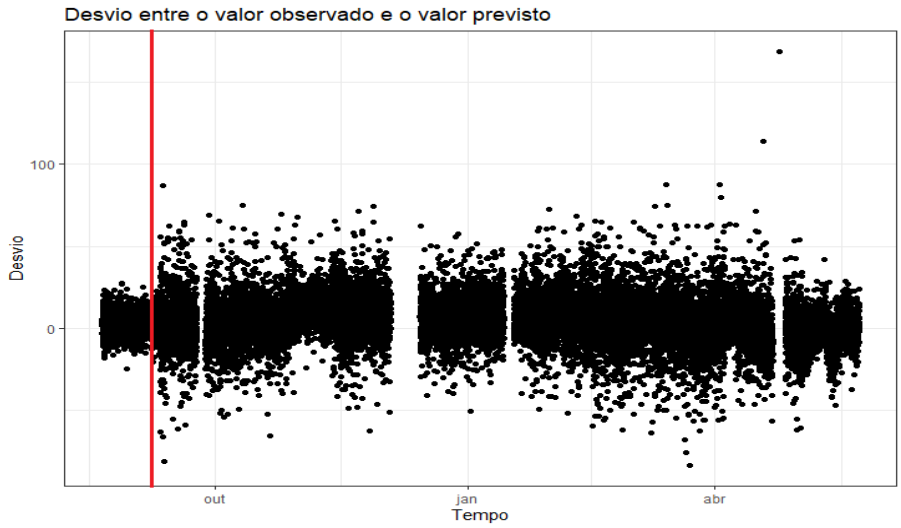

Figura A.2: Desvio entre os valores observados (Inclinómetro Inferior)

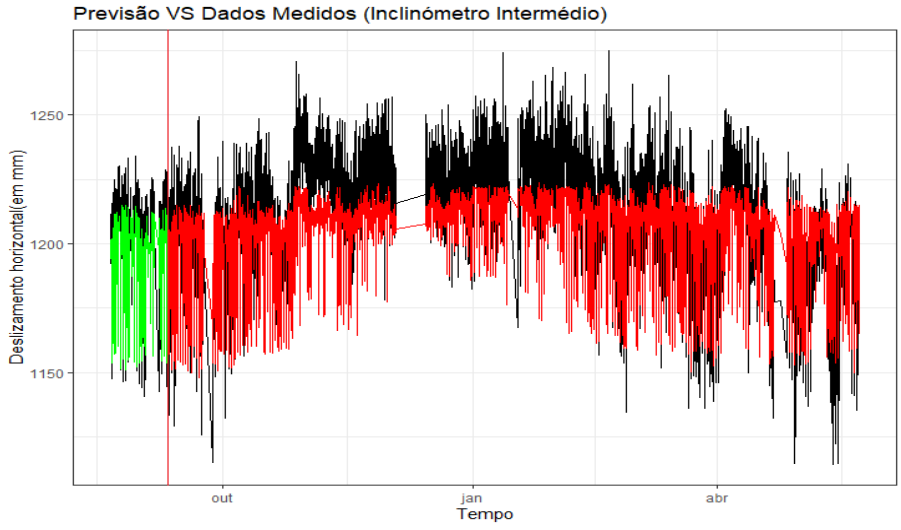

Figura A.3: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

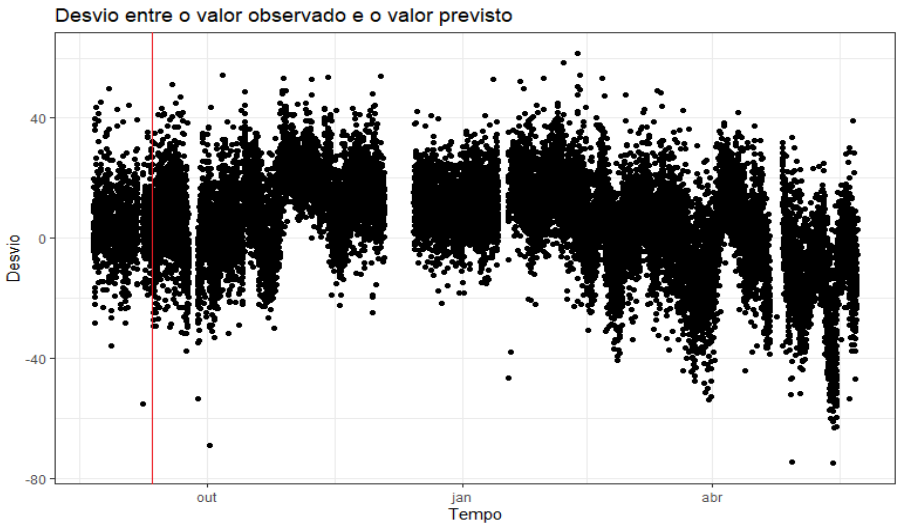

Figura A.4: Desvio entre os valores observados (Inclinómetro Intermédio)

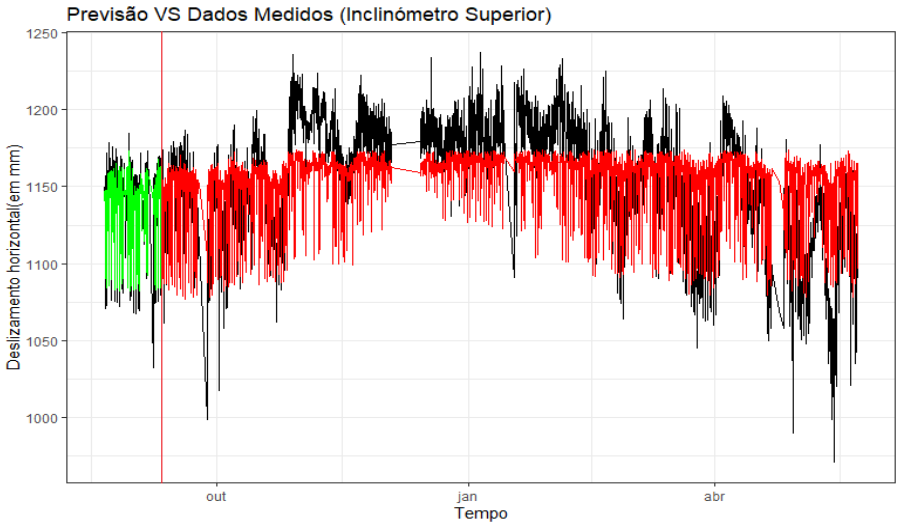

Figura A.5: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

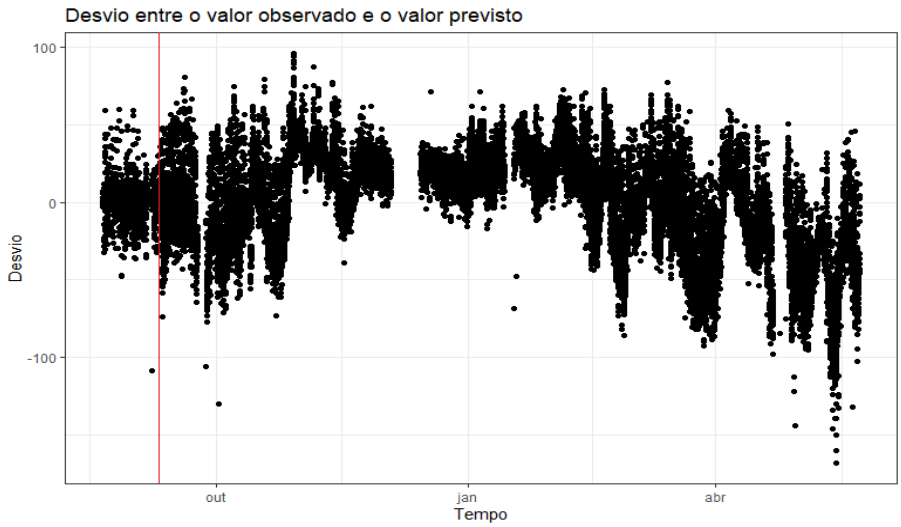

Figura A.6: Desvio entre os valores observados (Inclinómetro Superior)

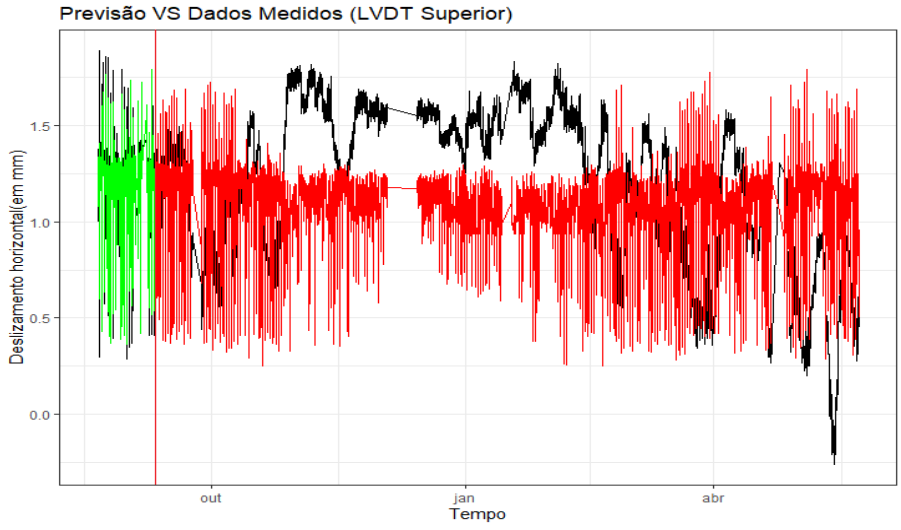

Figura A.7: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

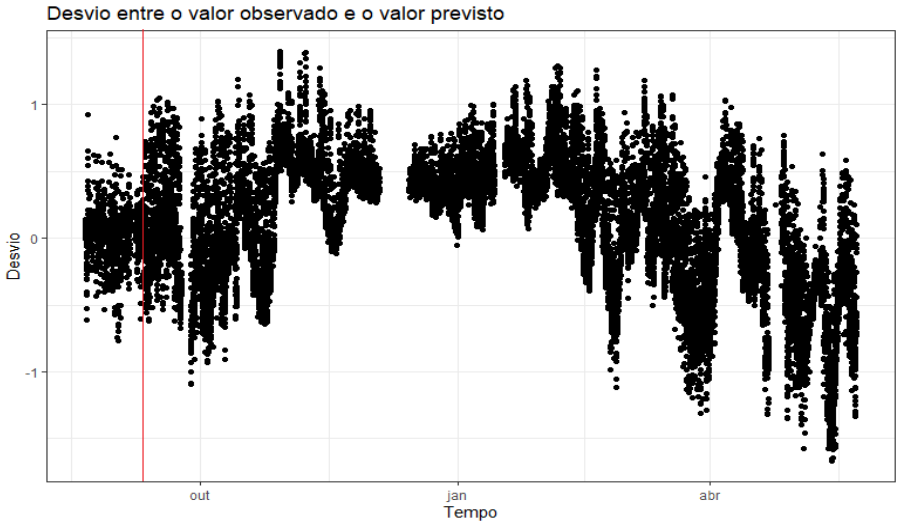

Figura A.8: Desvio entre os valores observados (LVDT Superior)

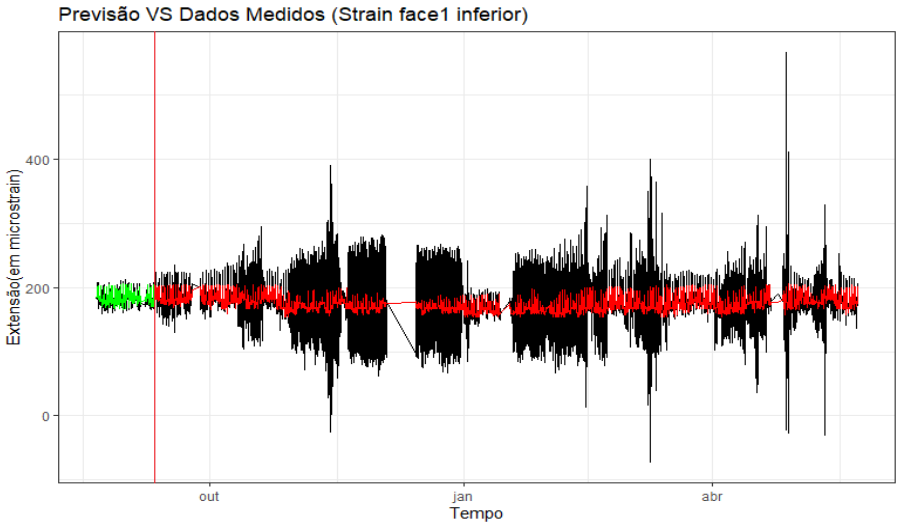

Figura A.9: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

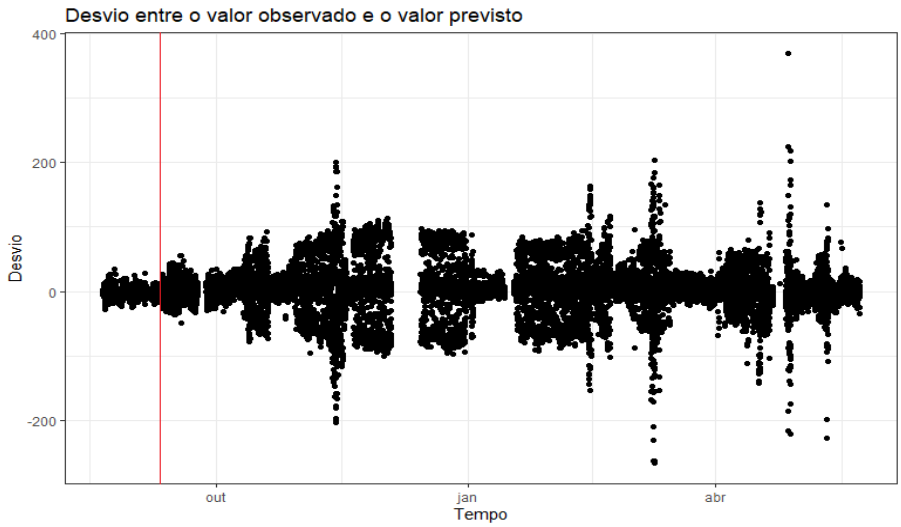

Figura A.10: Desvio entre os valores observados (Strain face1 inferior)

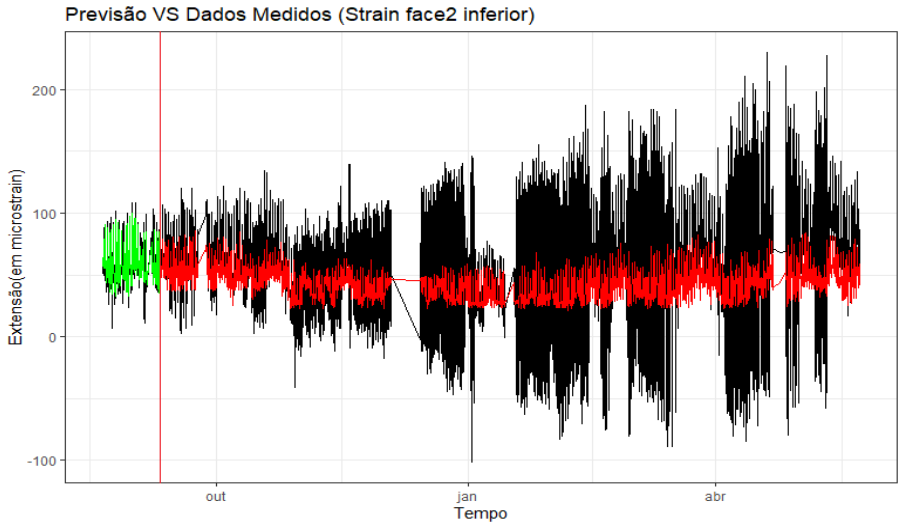

Figura A.11: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

Figura A.12: Desvio entre os valores observados (Strain face2 inferior)

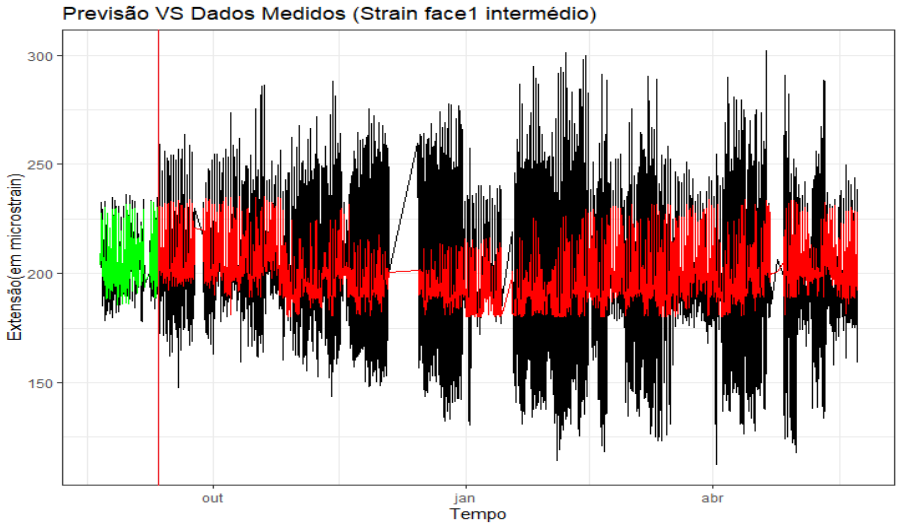

Figura A.13: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

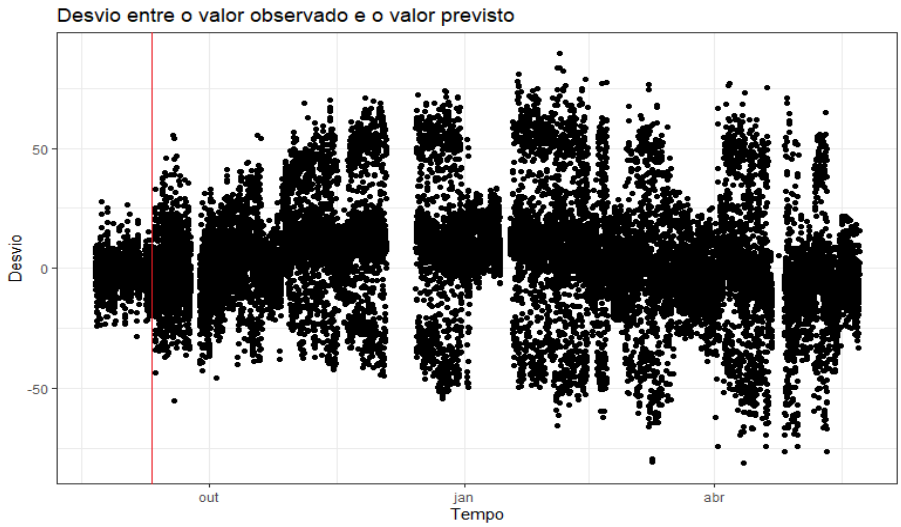

Figura A.14: Desvio entre os valores observados (Strain face1 intermédio)

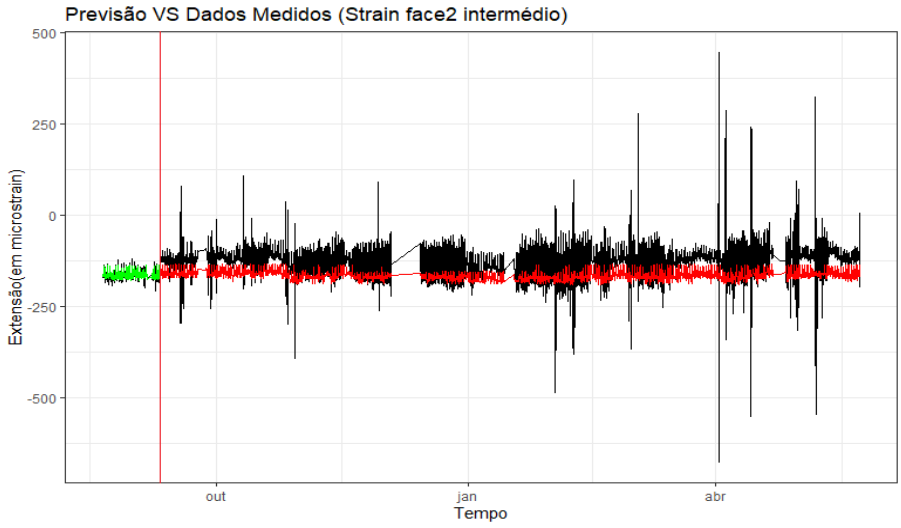

Figura A.15: Comparação entre os valores observados e previstos do modelo *Random Forests*. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

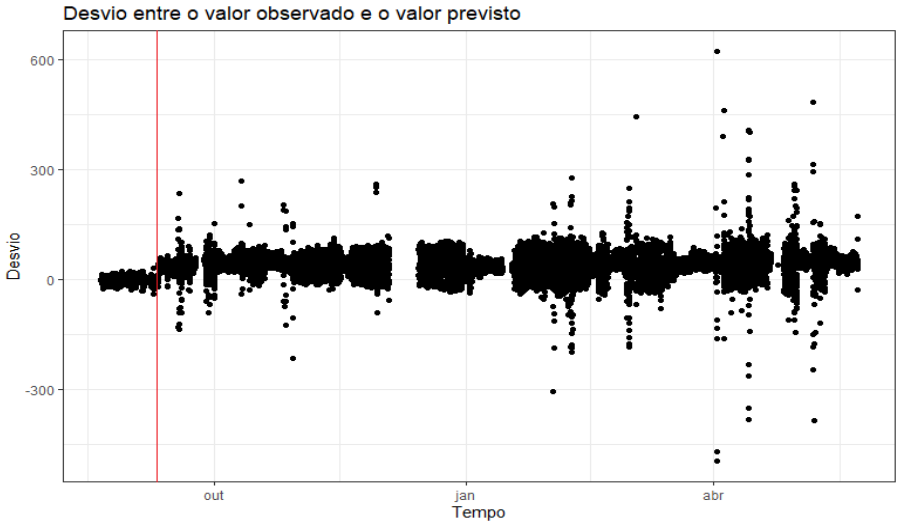

Figura A.16: Desvio entre os valores observados (Strain face2 intermédio)

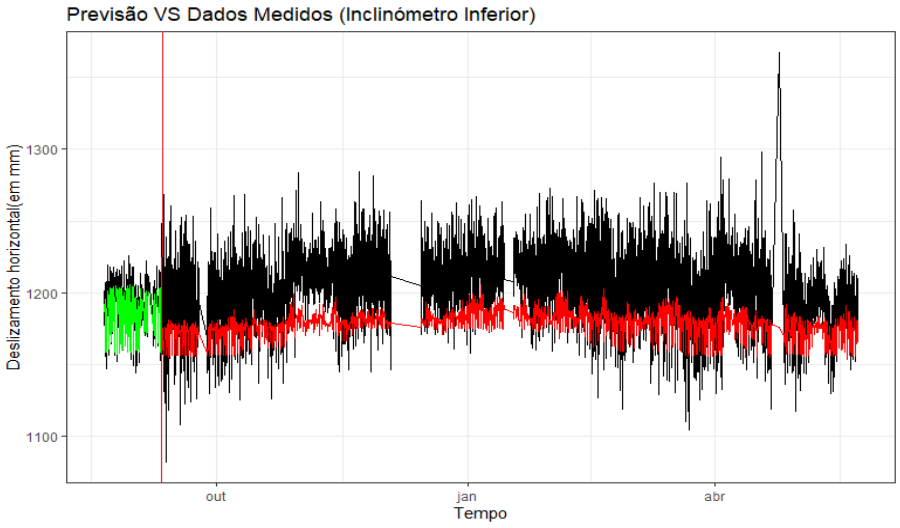

Figura A.17: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

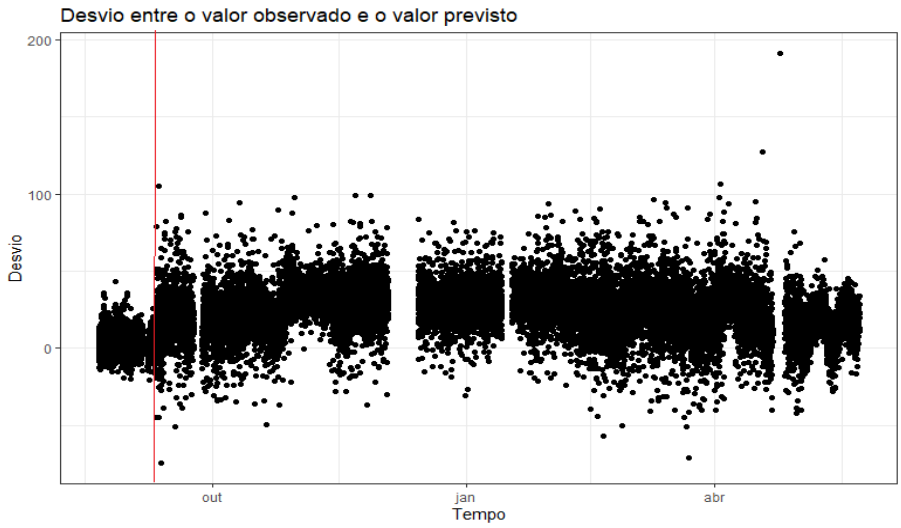

Figura A.18: Desvio entre os valores observados (Inclinómetro inferior)

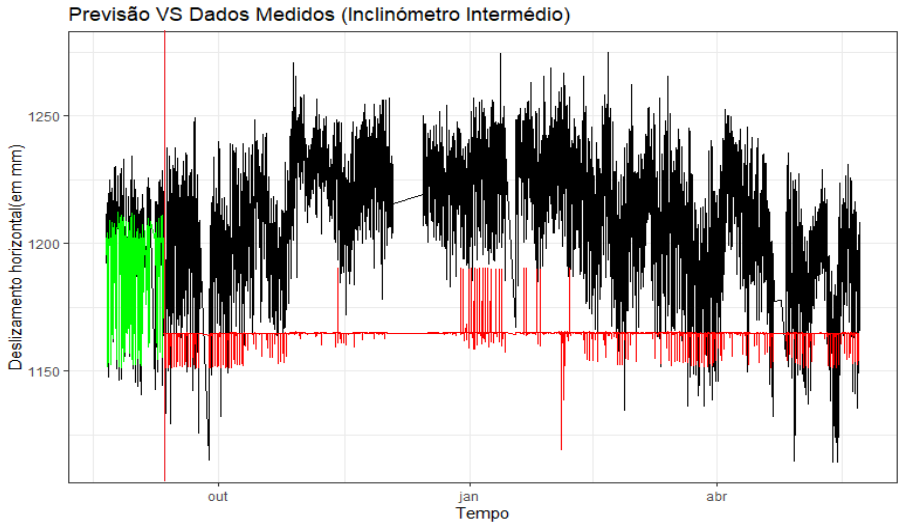

Figura A.19: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

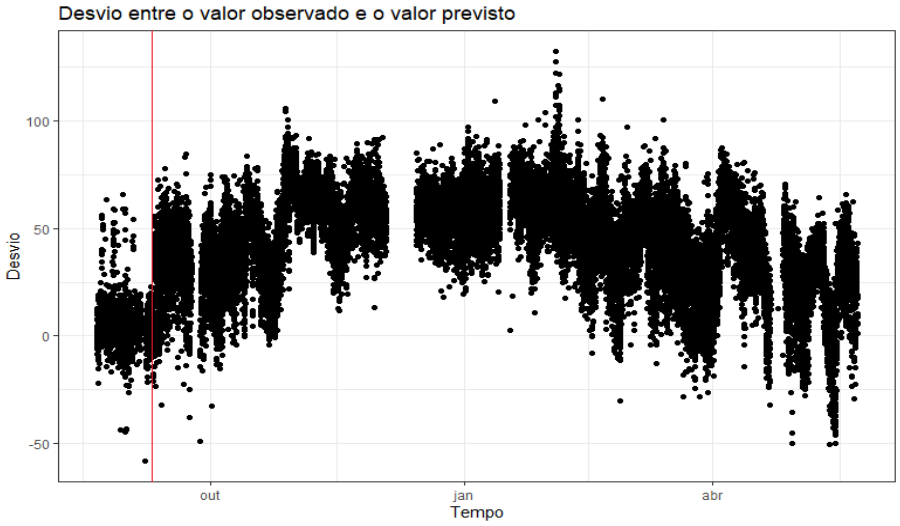

Figura A.20: Desvio entre os valores observados (Inclinómetro intermédio)

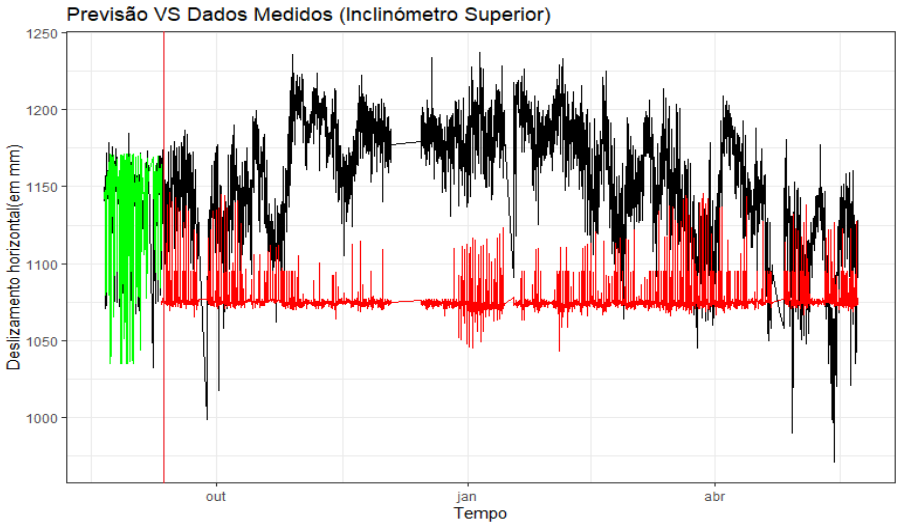

Figura A.21: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

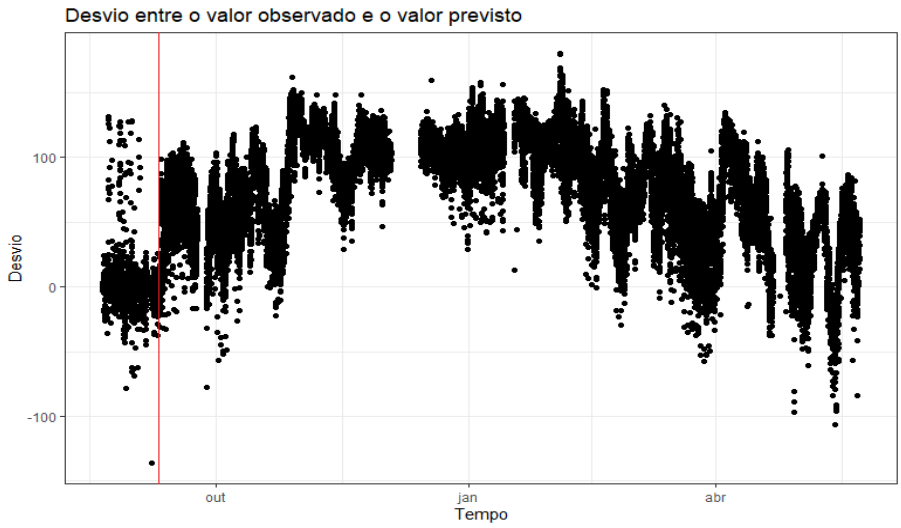

Figura A.22: Desvio entre os valores observados (Incinómetro Superior)

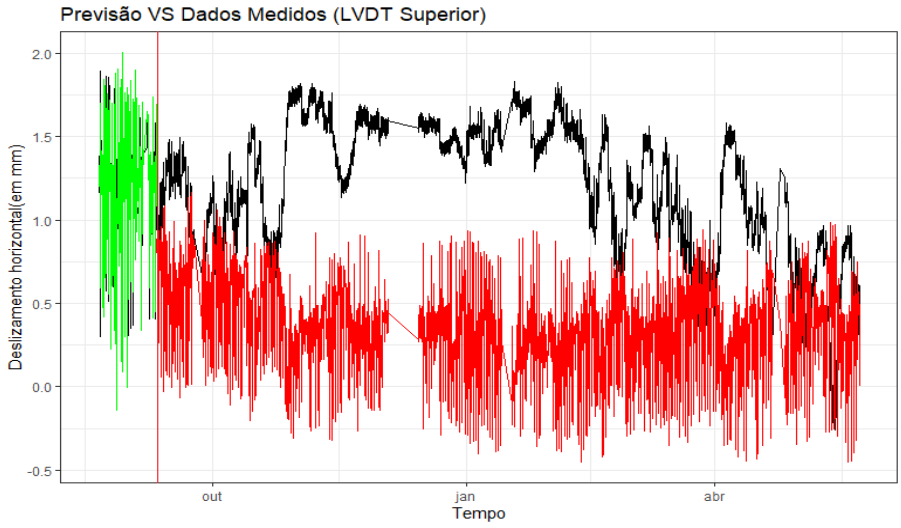

Figura A.23: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

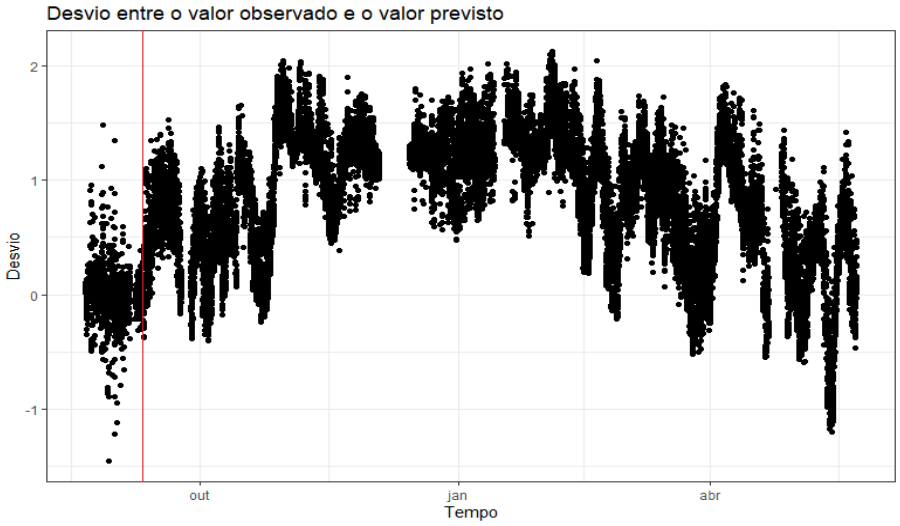

Figura A.24: Desvio entre os valores observados (LVDT superior)

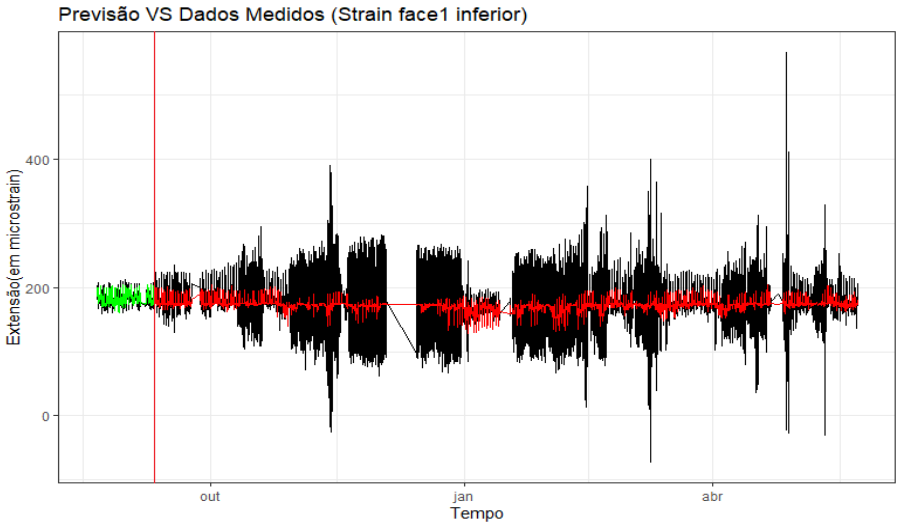

Figura A.25: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

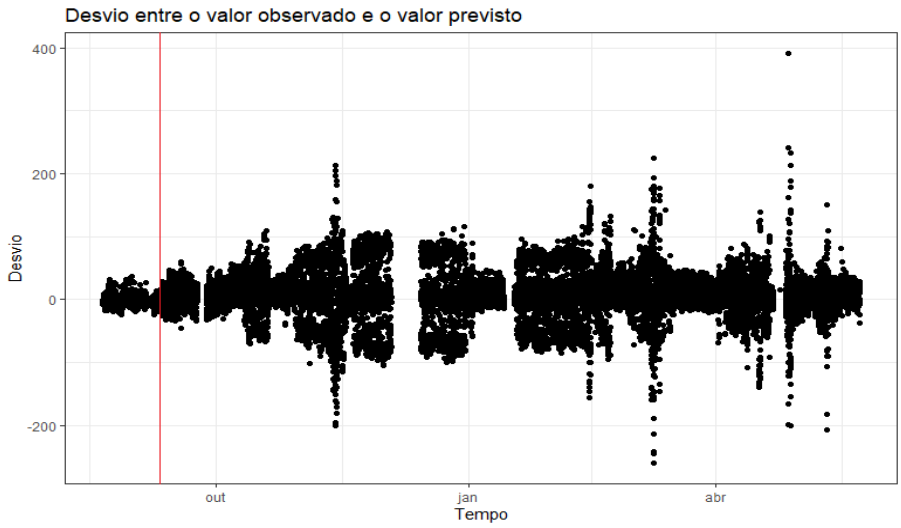

Figura A.26: Desvio entre os valores observados (Strain face1 inferior)

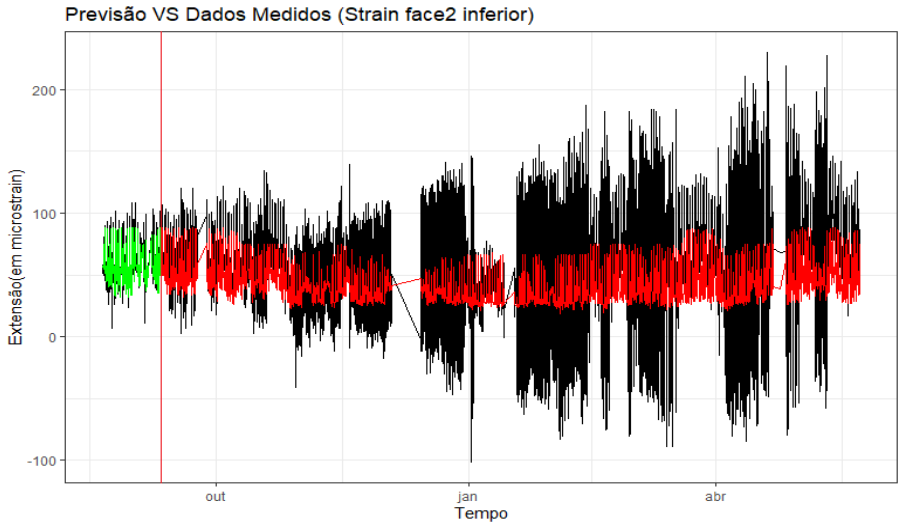

Figura A.27: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

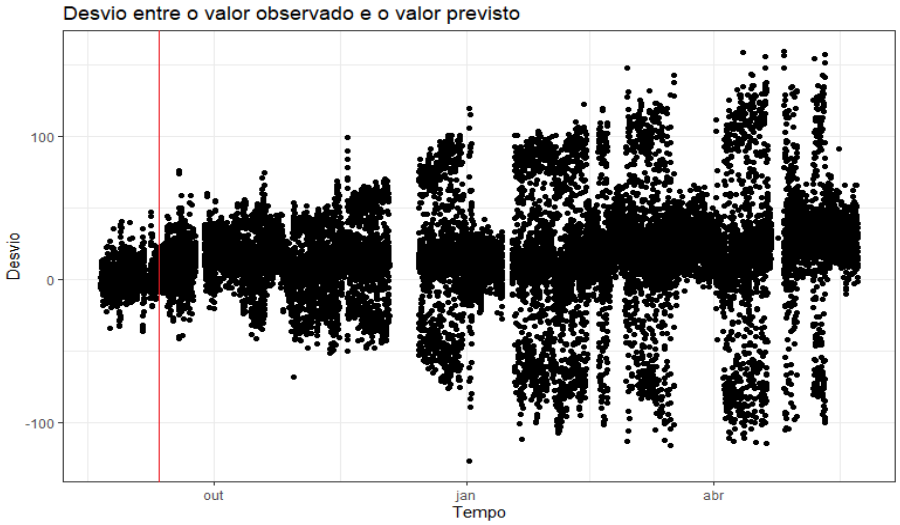

Figura A.28: Desvio entre os valores observados (Strain face2 inferior)

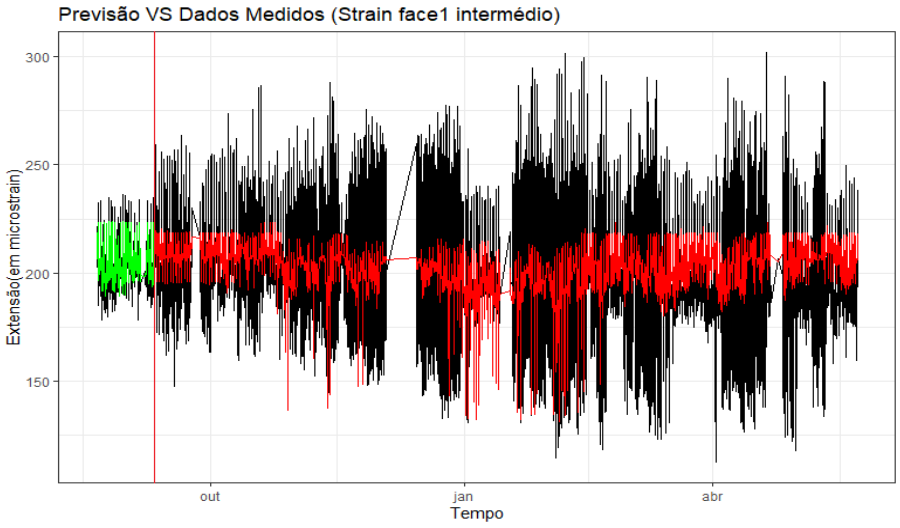

Figura A.29: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

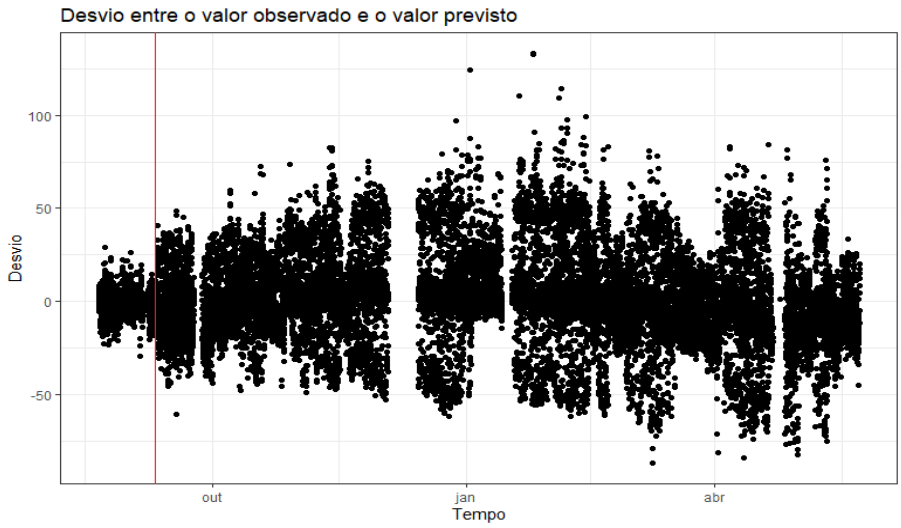

Figura A.30: Desvio entre os valores observados (Strain face1 intermédio)

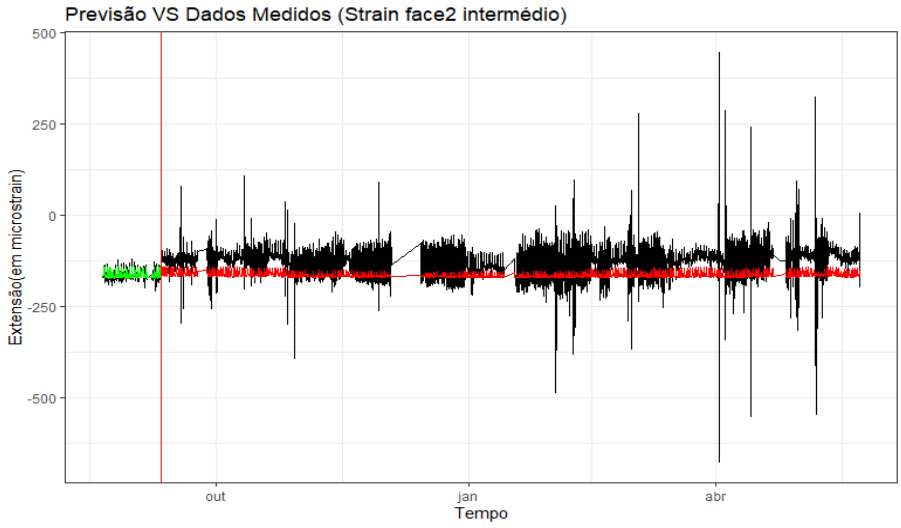

Figura A.31: Comparação entre os valores observados e previstos pelo modelo RNA. A preto temos os valores observados durante o período pré dano e o período pós dano. A verde estão as previsões realizadas para o período de teste pré dano. A vermelho estão as previsões realizadas para o período pós dano

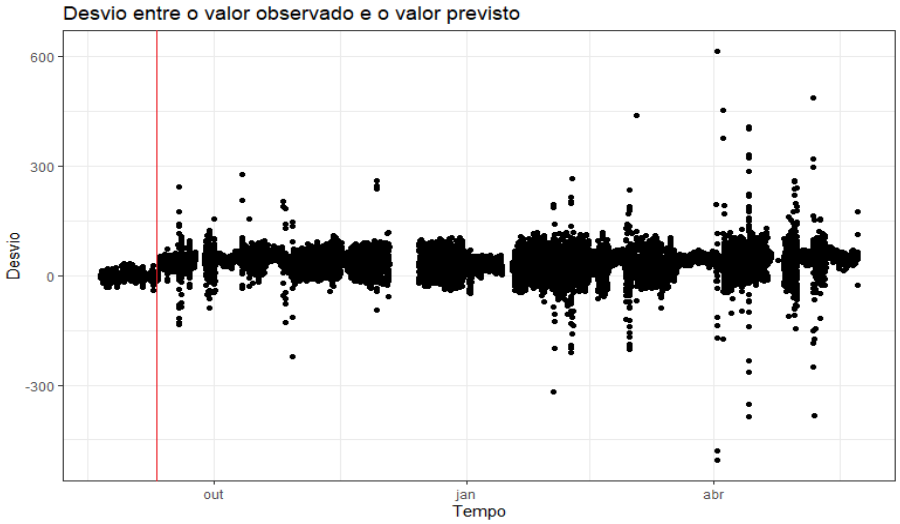

Figura A.32: Desvio entre os valores observados (Strain face2 intermédio)

Referências

[1] Anthony TC Goh e SH Goh. Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. *Computers and Geotechnics*, 34(5):410–421, 2007.

[2] Bo Dai, Chongshi Gu, Erfeng Zhao, e Xiangnan Qin. Statistical model optimized random forest regression model for concrete dam deformation monitoring. *Structural Control and Health Monitoring*, 25(6):e2170, 2018.

[3] S Matthias. Approaches to analyse and interpret biological profile data. *Potsdam University*, 2006.

[4] Zhi Liu e Han-Xiong Li. A probabilistic fuzzy logic system for modeling and control. *IEEE Transactions on Fuzzy Systems*, 13(6):848–859, 2005.

[5] Charles R Farrar e Keith Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):303–315, 2007.

[6] XG Hua, YQ Ni, JM Ko, e KY Wong. Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. *Journal of Computing in Civil Engineering*, 21(2):122–135, 2007.

[7] Daniel T Larose e Chantal D Larose. *Discovering knowledge in data: an introduction to data mining*, volume 4. John Wiley & Sons, 2014.

[8] Fadhilah Ahmad, Nur Hafieza Ismail, e Azwa Abdul Aziz. The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129):6415–6426, 2015.

[9] Mohamad H Hassoun et al. *Fundamentals of artificial neural networks*. MIT press, 1995.

[10] Meisam Gordan, Hashim Abdul Razak, Zubaidah Ismail, e Khaled Ghaedi. Recent developments in damage identification of structures using data mining. *Latin American Journal of Solids and Structures*, 14(13):2373–2401, 2017.

[11] David J Montana e Lawrence Davis. Training feedforward neural networks using genetic algorithms. Em *IJCAI*, volume 89, páginas 762–767, 1989.

[12] Smriti Sharma e Subhamoy Sen. Damage detection in presence of varying temperature through residual error modelling approach with dual neural network.

[13] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[14] S Chehreh Chelgani, SS Matin, e James C Hower. Explaining relationships between coke quality index and coal properties by random forest method. *Fuel*, 182:754–760, 2016.

[15] Aiping Guo, Ajuan Jiang, Jie Lin, e Xiaoxiao Li. Data mining algorithms for bridge health monitoring: Kohonen clustering and lstm prediction approaches. *The Journal of Supercomputing*, páginas 1–16, 2019.

[16] MM Reda Taha e J Lucero. Damage identification for structural health monitoring using fuzzy pattern recognition. *Engineering Structures*, 27(12):1774–1783, 2005.

[17] Emanuel Sousa Tomé, Mário Pimentel, e Joaquim Figueiras. Damage detection under environmental and operational effects using cointegration analysis–application to experimental data from a cable-stayed bridge. *Mechanical Systems and Signal Processing*, 135:106386, 2020.

[18] F Guillaume Blanchet, Pierre Legendre, e Daniel Borcard. Forward selection of explanatory variables. *Ecology*, 89(9):2623–2632, 2008.

[19] Kezhi Z Mao. Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):629–634, 2004.

[20] Francisco Aparisi e César L Haro. Hotelling's t2 control chart with variable sampling intervals. *International Journal of Production Research*, 39(14):3127–3140, 2001.