



Estimação do valor do preço de mercado da energia elétrica através de modelos não-lineares

Felipe Claudio da Silva Santos

Projeto de Graduação apresentado ao Curso
de Engenharia Eletrônica e de Computação
da Escola Politécnica, Universidade Federal
do Rio de Janeiro, como parte dos requisitos
necessários à obtenção do título de Enge-
nheiro.

Orientador: Luiz Pereira Calôba

Coorientador: Natanael Nunes de Moura Ju-
nior

Rio de Janeiro
Outubro de 2008

Título

Felipe Claudio da Silva Santos

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO
DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO DA ESCOLA PO-
LITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE ENGENHEIRO ELETRÔNICO E DE COMPUTAÇÃO

Autor:

Felipe Claudio da Silva Santos

Orientador:

Prof. Luis Pereira Calôba, Dr.Ing

Co-orientador:

Natanael Nunes de Moura Junior, Dr.

Examinador:

Flávio Luis de Mello

Examinador:

Carlous Jourdan Gadelha Vieira

Rio de Janeiro

Outubro de 2008

Declaração de Autoria e de Direitos

Eu, *Felipe Claudio da Silva Santos* CPF 427.279.208-30, autor da monografia *Estimação do valor do preço de mercado da energia elétrica através de modelos não-lineares*, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na disciplina de Projeto de Graduação da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.
2. Excetuam-se do item 1. eventuais transcrições de texto, figuras, tabelas, conceitos e idéias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
4. O autor pode, excepcionalmente, encaminhar à Comissão de Projeto de Graduação, a não divulgação do material, por um prazo máximo de 01 (um) ano, improrrogável, a contar da data de defesa, desde que o pedido seja justificado, e solicitado antecipadamente, por escrito, à Congregação da Escola Politécnica.
5. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
6. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
7. Por ser verdade, firmo a presente declaração.

Felipe Claudio da Silva Santos

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Escola Politécnica - Departamento de Eletrônica e de Computação
Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária
Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfiltrar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

AGRADECIMENTO

Meus agradecimentos vão, em primeiro lugar para os meus pais Georgea e Ricardo e minha irmã Júlia, por todo o apoio dado durante a minha jornada acadêmica. Sem os mesmos não conseguiria trilhar metade do caminho que trilhei.

Gostaria também de agradecer meus orientadores Luiz Calôba e Natanael Moura por todo esforço, tempo investido, paciência e boa vontade mostrada. Além disso, agradeço o voto de confiança dado e espero ter retribuído ao nível merecido. Esses aprendizado proporcionado me tornaram um profissional melhor do que quando comecei o trabalho.

Gostaria de agradecer ao Gabriel Torres e ao Luiz Évora pela colaboração direta no trabalho. Hora trazendo um pouco da experiência, hora revisando o conteúdo do texto.

Gostaria de agradecer outros familiares que estiveram comigo nessa jornada assim como minha avó Sônia e minhas tias Silvia, Silvane e Maria

Em último lugar, mas não menos importante, gostaria de agradecer ao meus amigos que são parte chave na minha vida. Em especial, Rafael Chimienski que sempre esteve comigo nos momentos mais complicados da minha vida.

RESUMO

Os contratos de fornecimento de energia são feitos com base em quantidades fixas de MWhs que devem ser entregues ao consumidor mensalmente. Geralmente o valor produzido não é exatamente igual ao valor combinado. A diferença entre o que foi contratado e o que foi efetivamente fornecido é liquidada no mercado de curto prazo ao valor do PLD - Preço da Liquidação das Diferenças, o qual é calculado semanalmente pela CCEE.

Este trabalho traz uma abordagem que utiliza decomposição em séries temporais, processamento de sinais e redes neurais para estimar o valor do PLD médio mensal. Ao final do trabalho são feitas estimativas para um período de até 10 meses, trazendo o comportamento da rede conforme o número de meses a frente.

Palavras-Chave: PLD, séries temporais, redes neurais, MLP, aprendizado de máquina, setor elétrico brasileiro.

ABSTRACT

The electrical power supply contracts use fixed amount of MWHs that must be provided to the consumer monthly. The generated power usually is not equal the value agreed upon contract. The difference between the supplied and the contracted amount is liquidated using PLD - Differences Liquidation Price. This values is calculated by CCEE.

This project uses times series decomposition, signal processing and neural networks to estimate PLD's monthly value. In the last part of this work, there are estimates for a 10 months period, showing the neural network behavior for each number of months ahead.

Key-words: PLD, time series, neural networks, MLP, machine learning, Brazilian's Electrical Sector.

SIGLAS

UFRJ - Universidade Federal do Rio de Janeiro

EPE - Empresa de Pesquisa Energética

SIN - Sistema Interligado Nacional

SEB - Sistema Elétrico Brasileiro

RE-SEB - Reestruturação do Sistema Elétrico Brasileiro

ANEEL - Agência Nacional de Energia Elétrica

MME - Ministério de Minas e Energias

CCEE - Câmara de Comercialização de Energia Elétrica

EPE - Empresa de Energia Elétrica

ONS - Operador Nacional do Sistema Elétrico

MAE - Mercado Atacadista de Energia

CMSE - Comitê de Monitoramento do Setor Elétrico

ACR - Ambiente de Contratação Regulada

ACL - Ambiente de Contratação Livre

PLD - Preço de Liquidação das Diferenças

MCP - Mercado de Curto Prazo

CMO - Custo Marginal de Operação

CVU - Custo Variável Unitário

UTE - Usina Térmica

ENA - Energia Natural Armazenada

EAR - Energia Armazenada

CCEAR - Contrato de Comercialização de Energia Elétrica no Ambiente Regulado

PMO - Planejamento Mensal de Operação

CFURH - Compensação Financeira pelo Uso dos Recursos Hídricos

MLP - Multilayer Perceptron

FFT - Fast Fourier Transform

IIR - Infinite Impulse Response

FIR - Finite Impulse Response

AR - Autoregressive

ARIMA - Autoregressive Integrated Moving Average

GARCH - Generalized Auto-Regressive Conditional Heteroscedasticity

ACF - Autocorrelation Function

PACF - Partial Autocorrelation Function

AIC - Akaike Info Criterion

PSF - Partial Sequence-based Forecasting

RBFN - Radial Basis Function Network

CSV - Comma Separated Values

MSE - Mean Squared Error

RMSE - Root Mean Squared Error

ACC - Accuracy

LSTM - Long Short-Term Memory

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Tema | 1 |
| 1.2 | Delimitação | 1 |
| 1.3 | Justificativa | 2 |
| 1.4 | Objetivos | 2 |
| 1.5 | Metodologia | 2 |
| 1.6 | Descrição | 3 |
| 2 | O Sistema Elétrico Brasileiro | 4 |
| 2.1 | Histórico do Setor Elétrico Brasileiro | 4 |
| 2.2 | O Cálculo do Preço da Liquidação das Diferenças | 9 |
| 2.3 | Revisão bibliográfica | 15 |
| 3 | Séries Temporais e Aprendizado de Máquina | 20 |
| 3.1 | Séries Temporais | 20 |
| 3.1.1 | Tendência | 22 |
| 3.1.2 | Sazonalidade | 24 |
| 3.1.3 | Ciclos Senoidais | 26 |
| 3.1.4 | Componente Residual | 26 |
| 3.2 | Processamento de Sinais | 28 |
| 3.3 | Redes Neurais Artificiais | 29 |
| 3.3.1 | Backpropagation | 33 |
| 3.3.2 | Rede perceptron multicamadas | 34 |
| 3.3.3 | Treinamento | 35 |

| | |
|---|-----------|
| 4 Método | 38 |
| 4.1 Coleta de dados | 38 |
| 4.2 Pré-processamento | 39 |
| 4.2.1 Tendência | 40 |
| 4.2.2 Sazonalidade | 41 |
| 4.2.3 Ciclos Senoidais e Resíduo | 41 |
| 4.3 Seleção dos sinais de entrada | 42 |
| 4.3.1 Seleção dos atrasos do sinal de saída | 42 |
| 4.4 Rede Neural | 42 |
| 4.4.1 Treinamento das redes neurais | 42 |
| 5 Resultados | 45 |
| 5.1 Aquisição dos dados | 45 |
| 5.2 Pré-Processamento | 46 |
| 5.2.1 Seleção dos sinais de entrada | 46 |
| 5.2.2 Seleção dos atrasos do sinal de saída | 47 |
| 5.2.3 Tendência | 47 |
| 5.2.4 Sazonalidade | 49 |
| 5.2.5 Ciclos Senoidais e Resíduo | 50 |
| 5.3 Treinamento das redes neurais | 51 |
| 5.3.1 Treinamento da rede para obter a saída no mês atual | 52 |
| 5.3.2 Treinamento da rede para obter a saída um mês a frente | 54 |
| 5.3.3 Treinamento da rede para obter a saída para vários meses a frente | 60 |
| 6 Conclusões | 63 |
| Bibliografia | 65 |
| A Outros resultados | 71 |
| A.1 Previsão para 2 meses a frente | 71 |
| A.2 Previsão para 3 meses a frente | 74 |
| A.3 Previsão para 4 meses a frente | 76 |
| A.4 Previsão para 5 meses a frente | 79 |

| | | |
|-----|---|----|
| A.5 | Previsão para 6 meses a frente | 81 |
| A.6 | Previsão para 7 meses a frente | 84 |
| A.7 | Previsão para 8 meses a frente | 86 |
| A.8 | Previsão para 9 meses a frente | 89 |
| A.9 | Previsão para 10 meses a frente | 91 |

Listas de Figuras

| | | |
|------|---|----|
| 2.1 | Instituições do setor elétrico brasileiro. | 6 |
| 2.2 | Mapa do Sistema Elétrico Brasileiro no horizonte de 2017 | 7 |
| 2.3 | Mapa do Sistema Elétrico Brasileiro no horizonte de 2017 | 9 |
| 2.4 | Comparação entre a matriz elétrica mundial e a brasileira | 10 |
| 2.5 | Instituições do setor elétrico brasileiro. | 11 |
| 2.6 | Previsão indireta do PLD | 15 |
| 2.7 | Previsão direta do PLD | 15 |
| 2.8 | Previsão do PLD utilizando o modelo híbrido | 16 |
| 2.9 | Previsão do PLD utilizando o modelo com árvore de classificação | 16 |
| 2.10 | Instituições do setor elétrico brasileiro. | 18 |
| 3.1 | Exemplo de decomposição de série temporal. | 21 |
| 3.2 | Exemplo de extração de tendência com filtro de média. | 24 |
| 3.3 | Extração de sazonalidade com periodicidade anual ($P = 12$) para o dataset de nascimentos em Nova York. | 25 |
| 3.4 | Autocorrelação para o sinal residual do dataset de nascimentos em Nova York. | 26 |
| 3.5 | Resíduo sem remoção de ciclos senoidais para o dataset de nascimentos em Nova York. | 27 |
| 3.6 | Resíduo com remoção de ciclos senoidais para o dataset de nascimentos em Nova York. | 27 |
| 3.7 | Sinal residual no tempo para o dataset de nascimentos em Nova York [1]. | 28 |
| 3.8 | Filtro Notch IIR digital. | 29 |
| 3.9 | Neurônio biológico | 30 |
| 3.10 | Neurônio artificial | 30 |
| 3.11 | Função logística. | 32 |

| | | |
|------|--|----|
| 3.12 | Tangente hiperbólica. | 32 |
| 3.13 | ReLU. | 33 |
| 3.14 | Tangente hiperbólica. | 34 |
| 3.15 | Rede Perceptron Multicamadas. | 35 |
| 3.16 | Influência da taxa de aprendizado. | 36 |
| 3.17 | Validação Cruzada. | 37 |
| 4.1 | Diagrama do Método. | 38 |
| 5.1 | Correlação dos dados de entrada. | 46 |
| 5.2 | Autocorrelação do PLD. | 47 |
| 5.3 | Análise do mse pelo tamanho da janela na extração da tendência. | 48 |
| 5.4 | Histograma e FFT da tendência obtida. | 48 |
| 5.5 | Análise do mse pelo tamanho da janela na extração da sazonalidade. | 49 |
| 5.6 | Histograma e FFT da sazonalidade obtida. | 49 |
| 5.7 | Análise do mse pelo tamanho da janela na extração da sazonalidade. | 50 |
| 5.8 | Distribuição e FFT após extração da componente de maior energia. | 50 |
| 5.9 | Resultado Final com as componentes separadas utilizando o método descrito acima. | 51 |
| 5.10 | RMSE pelo número de neurônios no conjunto de validação. | 52 |
| 5.11 | RMSE pelo número de neurônios no conjunto de teste. | 52 |
| 5.12 | Erro pelo número de época e por fold. | 53 |
| 5.13 | Sinal completo. | 54 |
| 5.14 | RMSE pelo número de neurônios no conjunto de validação. | 55 |
| 5.15 | RMSE pelo número de neurônios no conjunto de teste. | 55 |
| 5.16 | Erro pelo número de épocas e por fold. | 56 |
| 5.17 | Sinal completo. | 57 |
| 5.18 | RMSE pelo número de neurônios no conjunto de validação. | 57 |
| 5.19 | RMSE pelo número de neurônios no conjunto de teste. | 58 |
| 5.20 | Erro pelo número de épocas e por fold. | 59 |
| 5.21 | Sinal completo. | 59 |
| 5.22 | Histograma de previsão sem e com a correção de erro. | 60 |
| 5.23 | Erro no sinal residual pelo número de passos no conjunto de dados de teste. | 61 |

| | |
|--|----|
| 5.24 Erro no sinal residual pelo número de passos no conjunto de dados completo. | 61 |
| A.1 RMSE pelo número de neurônios no conjunto de validação. | 71 |
| A.2 RMSE pelo número de neurônios no conjunto de teste. | 72 |
| A.3 Erro pelo número de épocas e por fold. | 73 |
| A.4 Sinal completo. | 73 |
| A.5 RMSE pelo número de neurônios no conjunto de validação. | 74 |
| A.6 RMSE pelo número de neurônios no conjunto de teste. | 74 |
| A.7 Erro pelo número de épocas e por fold. | 75 |
| A.8 Sinal completo. | 76 |
| A.9 RMSE pelo número de neurônios no conjunto de validação. | 76 |
| A.10 RMSE pelo número de neurônios no conjunto de teste. | 77 |
| A.11 Erro pelo número de épocas e por fold. | 78 |
| A.12 Sinal completo. | 78 |
| A.13 RMSE pelo número de neurônios no conjunto de validação. | 79 |
| A.14 RMSE pelo número de neurônios no conjunto de teste. | 79 |
| A.15 Erro pelo número de épocas e por fold. | 80 |
| A.16 Sinal completo. | 81 |
| A.17 RMSE pelo número de neurônios no conjunto de validação. | 81 |
| A.18 RMSE pelo número de neurônios no conjunto de teste. | 82 |
| A.19 Erro pelo número de épocas e por fold. | 83 |
| A.20 Sinal completo. | 83 |
| A.21 RMSE pelo número de neurônios no conjunto de validação. | 84 |
| A.22 RMSE pelo número de neurônios no conjunto de teste. | 84 |
| A.23 Erro pelo número de épocas e por fold. | 85 |
| A.24 Sinal completo. | 86 |
| A.25 RMSE pelo número de neurônios no conjunto de validação. | 86 |
| A.26 RMSE pelo número de neurônios no conjunto de teste. | 87 |
| A.27 Erro pelo número de épocas e por fold. | 88 |
| A.28 Sinal completo. | 88 |
| A.29 RMSE pelo número de neurônios no conjunto de validação. | 89 |
| A.30 RMSE pelo número de neurônios no conjunto de teste. | 89 |
| A.31 Erro pelo número de épocas e por fold. | 90 |

| | |
|--|----|
| A.32 Sinal completo. | 91 |
| A.33 RMSE pelo número de neurônios no conjunto de validação. | 91 |
| A.34 RMSE pelo número de neurônios no conjunto de teste. | 92 |
| A.35 Erro pelo número de épocas e por fold. | 93 |
| A.36 Sinal completo. | 93 |

Lista de Tabelas

| | | |
|-----|--|----|
| 2.1 | Critérios vigentes para se tornar consumidor Livre. | 8 |
| 2.2 | Variáveis utilizadas no cálculo do PLD semanal. | 12 |
| 2.3 | Preço mínimo e máximo do PLD. | 13 |
| 2.4 | Arquivos de entrada do DECOMP. | 14 |
| 4.1 | Dados utilizado no trabalho. | 39 |
| 5.1 | Resultados obtidos com as redes no dataset de validação. | 53 |
| 5.2 | Resultados obtidos com as redes no dataset de validação. | 56 |
| 5.3 | Resultados obtidos com as redes no dataset de validação. | 58 |
| A.1 | Resultados obtidos com as redes no dataset de validação. | 72 |
| A.2 | Resultados obtidos com as redes no dataset de validação. | 75 |
| A.3 | Resultados obtidos com as redes no dataset de validação. | 77 |
| A.4 | Resultados obtidos com as redes no dataset de validação. | 80 |
| A.5 | Resultados obtidos com as redes no dataset de validação. | 82 |
| A.6 | Resultados obtidos com as redes no dataset de validação. | 85 |
| A.7 | Resultados obtidos com as redes no dataset de validação. | 87 |
| A.8 | Resultados obtidos com as redes no dataset de validação. | 90 |
| A.9 | Resultados obtidos com as redes no dataset de validação. | 92 |

Capítulo 1

Introdução

Neste capítulo, será introduzido o tema do projeto, além de mostrar a relevância do mesmo e quais métodos serão utilizados para alcançar o determinado objetivo. Ao final exibida a estrutura organizacional do texto.

1.1 Tema

O sistema elétrico brasileiro é funcional de maneira bem diferente dos demais sistemas elétricos pelo mundo, pois há grande abundância de energia barata gerada a partir de fontes hídricas. Essas fontes são responsáveis por aproximadamente 65,2% da energia consumida no país [2].

A volatilidade desses recursos torna bastante complexa a atividade de prever e planejar a utilização de forma a diminuir o custo da energia elétrica. A diferença entre a quantidade de energia negociada e efetivamente gerada é liquidada utilizando como base o PLD - Preço da Liquidação das Diferenças, o qual deseja-se determinar através de modelos matemáticos.

1.2 Delimitação

O trabalho tem como foco a previsão do PLD mensal e tem aplicação para os seguintes setores: elétrico, bancário, planejamento e pesquisa. Para isso serão utilizados dados públicos presentes em [3] e [4] para treinar redes multicamadas Perceptron - *MLP* para prever a parte residual do PLD médio mensal.

1.3 Justificativa

Os resultados da pesquisa são de interesse do setor elétrico, pois permitem, por exemplo, que comercializadoras de energia elétrica consigam obter lucro com a compra e venda de energia.

Para o setor financeiro, o trabalho pode ser utilizado como ferramenta para análise de risco de crédito para empresas do setor elétrico. No ramo de planejamento e pesquisa, os resultados podem ser utilizados para entender melhor a dinâmica do preço e realizar planejamentos mais precisos quanto à utilização de recursos.

1.4 Objetivos

O objetivo deste trabalho é obter e avaliar um modelo de predição de séries temporais aplicado ao problema de previsão do PLD mensal. Para isso busca-se (1) definir a forma como o pré-processamento será realizado, buscando obter o sinal residual necessário para treinar a rede neural; (2) Definir a arquitetura utilizada no treinamento; (3) Avaliar o resultados obtidos de forma a conseguir determinar o erro esperado na previsão para meses a frente.

1.5 Metodologia

Serão utilizados como entrada os dados de geração de energia para diversas fontes, além da carga de energia, energia natural afluente e informações sobre as vazões nos rios e bacias que compõem a região sudeste e centro-oeste, conforme a divisão feita no sistema elétrico brasileiro. O PLD médio mensal será utilizado como base para a saída da rede neural.

O PLD será decomposto em tendência, sazonalidade, ciclos senoidais e resíduo, de forma a fornecer como entrada da rede somente a parte não-determinística do sinal (resíduo). Além disso, será feito uma análise para determinar quais sinais serão utilizados na entrada e quantos atrasos dos mesmos serão necessários.

Finalmente serão treinadas redes neurais com somente uma camada oculta. Em um primeiro estágio será treinada uma rede que consiga obter o resíduo para o mesmo mês da entrada fornecida. Nos estágios seguintes serão treinadas redes para prever o resultado residual.

1.6 Descrição

O capítulo 2 traz um resumo sobre as características e o funcionamento do setor elétrico brasileiro, além de fazer uma breve revisão bibliográfica. O capítulo 3 apresenta as técnicas que serão utilizadas no trabalho e descreve matematicamente o funcionamento das mesmas. No capítulo 4 são descritos os métodos utilizados no trabalho, conectando os conhecimentos mostrados nos capítulos 2 e 3. No capítulo 5 são exibidos os resultados para cada um dos passos descritos no capítulo anterior. O capítulo 6 traz a conclusão e trabalhos futuros que podem ser realizados.

Capítulo 2

O Sistema Elétrico Brasileiro

Este capítulo tem como objetivo introduzir o leitor à estrutura do setor elétrico brasileiro, assim como expor os maiores desafios enfrentados. Também será apresentado e detalhado o problema a ser resolvido nesse projeto.

2.1 Histórico do Setor Elétrico Brasileiro

No começo da década de 90 a estrutura do setor elétrico brasileiro era verticalizada, composta por grandes estatais que tinham o controle sobre toda a cadeia produtiva (geração, transmissão e distribuição). Houve então uma reestruturação do setor (RE-SEB), baseado no princípio de que ”a eficiência no setor elétrico será assegurada através da competição, onde possível, e da regulamentação, onde necessária” [5]. Sendo assim, o mercado foi aberto para o setor privado. O SEB foi então dividido em empresas de geração, transmissão, distribuição e comercialização de energia. Muitas empresas públicas foram leiloadas, dentre elas a a Light (21/05/1996) e a CERJ(20/11//1996) [6].

Assim como mencionado, um dos pilares do RE-SEB foi a regulamentação. Então em 1996 foi criada a Agência Nacional de Energia Elétrica - ANEEL, a qual tinha como função regular e fiscalizar a produção, transmissão, distribuição e comercialização de energia elétrica segundo as diretrizes definidas pelo governo federal.

Foram aprovadas medidas que incentivavam a realização de contratos de longo prazo entre distribuidores e geradores [6]. Além disso, foi dado livre acesso aos pe-

queños produtores (potenciais hidráulicos menores que 1MW e usinas termoelétricas com potencial inferior a 5MW) ao sistemas de transmissão e distribuição das concessionárias e permissionárias do serviço público. Isto facilitou o aparecimento de novas fontes de energias e novos produtores no SEB.

Em 2001 houve uma grande crise energética no país que foi fruto de um período seco. Então, mudanças estruturais foram implantadas no SEB, buscando assim melhorar a gestão, evitar incidentes, garantir a qualidade e incentivar a expansão do mesmo. Com isso, a organização da estrutura foi modificada, sendo que as instituições responsáveis pelo pleno funcionamento deste modelo são descritas a seguir [4]:

- **Ministério de Minas e Energias - MME:** Responsável pela formulação e implantação de políticas relacionadas ao setor elétrico.
- **Comitê de Monitoramento do Setor Elétrico - CMSE:** Acompanha e avalia a continuidade e segurança do suprimento elétrico em todo o território nacional
- **Conselho Nacional de Política Energética - CNPE:** Formula políticas e diretrizes de energia que garantam o suprimento de energia elétrica para todo o país
- **Agência Nacional de Energia Elétrica - ANEEL:** Regula e fiscaliza a produção, transmissão, distribuição e comercialização de energia elétrica segundo as diretrizes definidas pelo governo federal.
- **Câmara de Comercialização de Energia Elétrica - CCEE:** Atua desde a medição de energia gerada e efetivamente consumida até a liquidação financeira dos contratos de compra e venda no curto prazo [4]. Anteriormente se chamava Mercado Atacadista de Energia - MAE.
- **Operador Nacional do Sistema Elétrico - ONS:** Tem como objetivo a coordenação e controle da instalações de geração e transmissão de energia elétrica no Brasil, buscando garantir a todos os agentes o livre acesso à rede

de transmissão, o menor custo de expansão do sistema e as melhores condições operacionais futuras [3]

- **Empresa de Pesquisa Energética - EPE:** Realiza pesquisas em áreas relacionadas com energia elétrica, petróleo, gás biocombustíveis e planejamento energético [2].

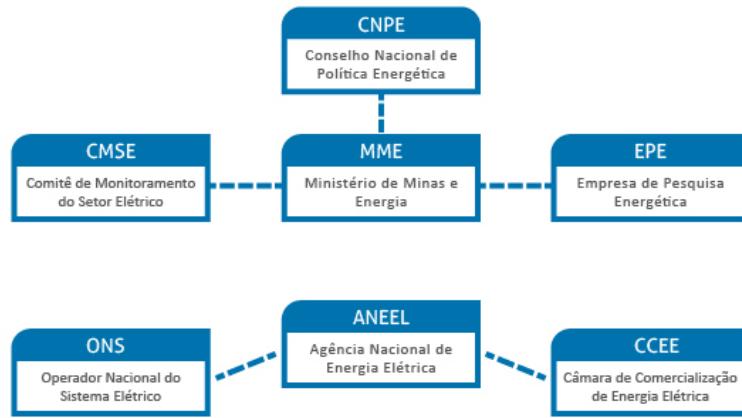


Figura 2.1: Instituições do setor elétrico brasileiro. Fonte: EPE [4].

Outra medida tomada para facilitar o gerenciamento de toda a rede de geração, transmissão, e consumo de energia elétrica foi dividir o Sistema Interligado Nacional - SIN [3] em quatro subsistemas sendo estes: Norte, Nordeste, Sudeste/Centro-Oeste e Sul. Vale ressaltar que existe conexão entre as partes, de forma que um subsistema pode auxiliar o suprimento da demanda de energia de outro em períodos de seca. Participam desse sistema diversos agentes [4], sendo eles:

- **Agentes geradores:** são agentes autorizados a gerar energia elétrica para o sistema.
- **Agentes de transmissão:** são agentes que detêm o meios de transmissão de energia.
- **Agentes de distribuição:** são agentes autorizados a distribuir energia contratada dos serviços de transmissão em uma área definida pela concessão
- **Consumidores livres:** são consumidores que têm a opção de escolher o distribuidor de energia, diferente do que acontece no consumo doméstico.

- **Agentes comercializadores:** são agentes que negociam energia por meio de contratos bilaterais realizados no ambiente de contratação livre. Esse tipo de agente pode negociar a energia com outras comercializadoras, geradoras, consumidores livres e distribuidoras por meio de leilões de ajuste no ambiente de contratação regulada. Além desse, existem outros tipos de consumidores, assim como visto em [4].

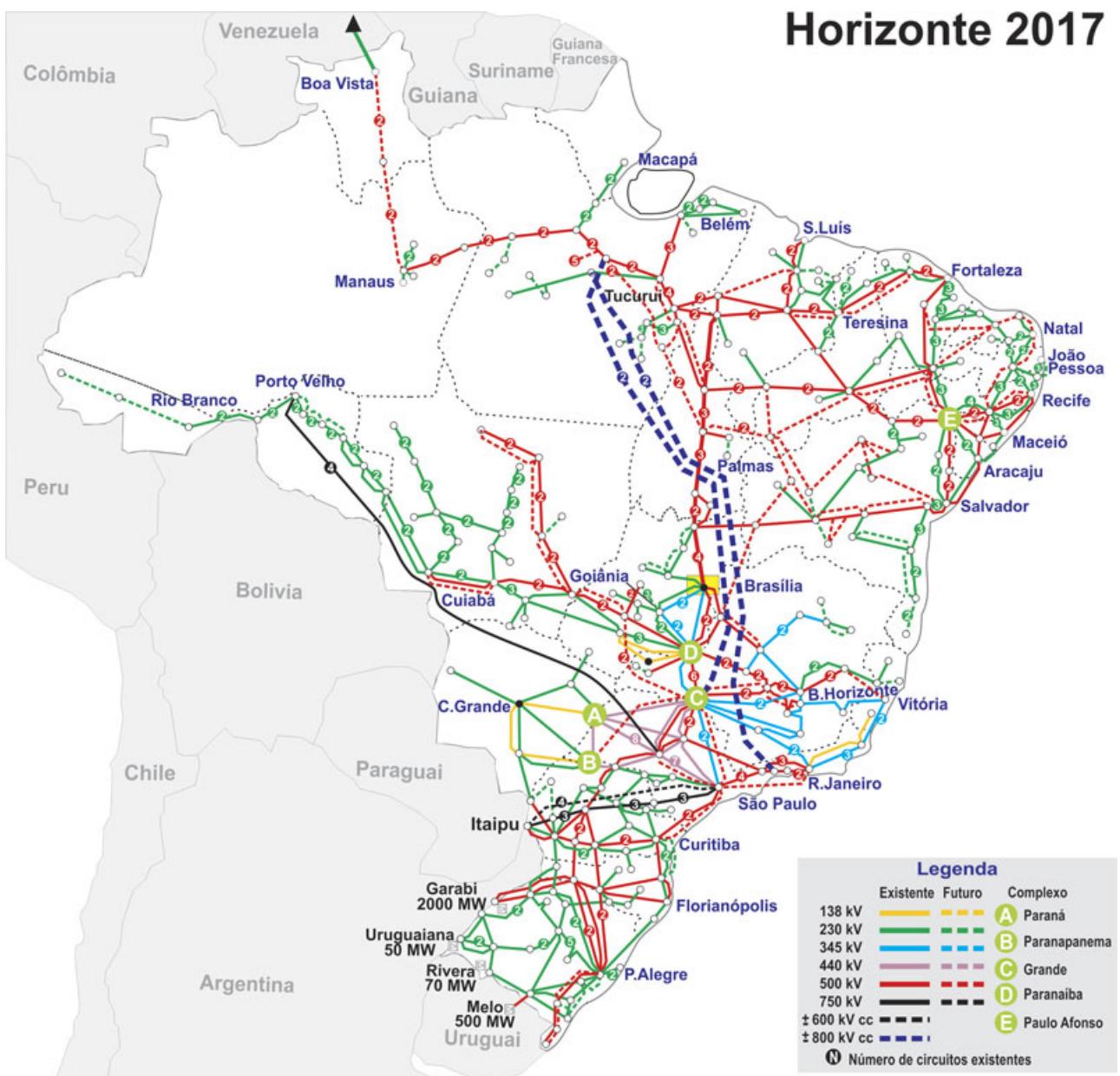


Figura 2.2: Mapa do Sistema Elétrico Brasileiro no horizonte de 2017[3]

A reestruturação também dividiu o mercado de energia elétrica brasileiro em dois ambientes de comercialização: Ambiente de Contratação Regulada - ACR e Ambiente de Contratação Livre - ACL. O primeiro ambiente é focado na contratação de energia para consumidores residenciais e indústrias de baixo consumo em geral. As negociações são feitas através de leilões onde geralmente os contratos de fornecimento de energia tem foco no médio e longo prazo (duração maior que 5 anos). As distribuidoras são obrigadas a comprar 100% da energia utilizada através desse meio. O segundo ambiente é focado em contratações bilaterais, de forma a trazer benefícios para grande consumidores, conforme mostrado abaixo, pois podem negociar o preço da energia direto com os geradores [7]. Os agentes que podem realizar compra por esse tipo de mercado são os agentes livres, conforme mostrados na tabela a seguir:

Tabela 2.1: Critérios vigentes para se tornar consumidor Livre.

| Demandá Mínima | Tensão mínima de fornecimento | Data da ligação do consumidor |
|-----------------------|--------------------------------------|--------------------------------------|
| 3MW | Qualquer Tensão | Após 07/07/1995 |
| 3MW | 69KV | Antes 07/07/1995 |

Dado o dinamismo do consumo, dificilmente a energia utilizada será igual a contratada. Uma das funções da CCEE é realizar medições tanto do que é produzido quanto o que é consumido. Esta instituição concentra informações sobre o montante de energia negociado em cada contrato do SEB. A diferença entre o que foi negociado e o que foi consumido é liquidado utilizando como multiplicador o Preço de Liquidação das Diferença - PLD, o qual será explicado mais a frente. O saldo dos agentes no mercado de curto prazo se dá pelas seguintes equações[7].

$$Receita_{Gerador} = (Energia_{Gerada} - Energia_{Contratada}) * PLD \quad (2.1)$$

$$Despesa_{Consumidor} = (Energia_{Consumida} - Energia_{Contratada}) * PLD \quad (2.2)$$

Esse processo de liquidação das diferenças é realizado entre as partes envolvidas na negociação. A CCEE somente realiza o cálculo de quanto cada parte envolvida na negociação deve pagar ou receber. A importância do PLD não se resume somente ao MCP, pois muitos dos contratos do mercado livre são baseados nesse valor[8]. Isto permite que os agentes tomem estratégias diferentes para aumentar o lucro ou diminuir o prejuízo. Um exemplo disso seria um agente consumidor, que ao prever o aumento de preços nos próximos meses, pode buscar realizar contratos para abastecimento futuro a um preço menor do que seria pago. Outro exemplo seria um distribuidor, que ao prever um baixo PLD no MCP do mês atual, decide vender o excedente de energia contratada no mercado livre ao invés de receber o valor proporcional ao PLD.

2.2 O Cálculo do Preço da Liquidação das Diferenças

O grande desafio do setor é utilizar a energia de maneira eficiente, reduzindo o custo da mesma e o risco de deficit tanto no mês atual quanto nos meses seguintes. Isto demanda a criação e utilização de novas tecnologias auxiliares ao planejamento.

Conforme observado na figura abaixo, o preço depende diretamente da quantidade de água nos reservatórios no momento atual e nos meses futuros.

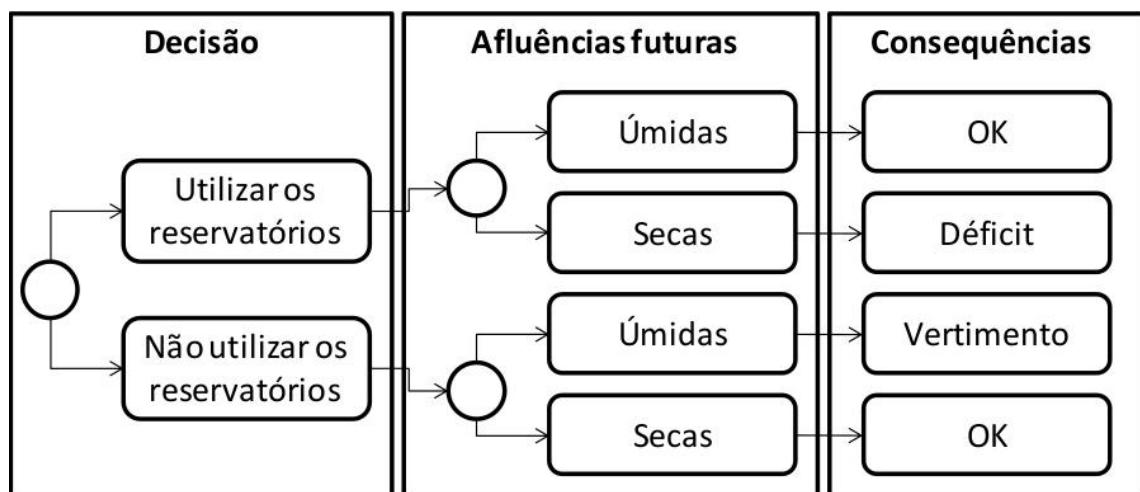
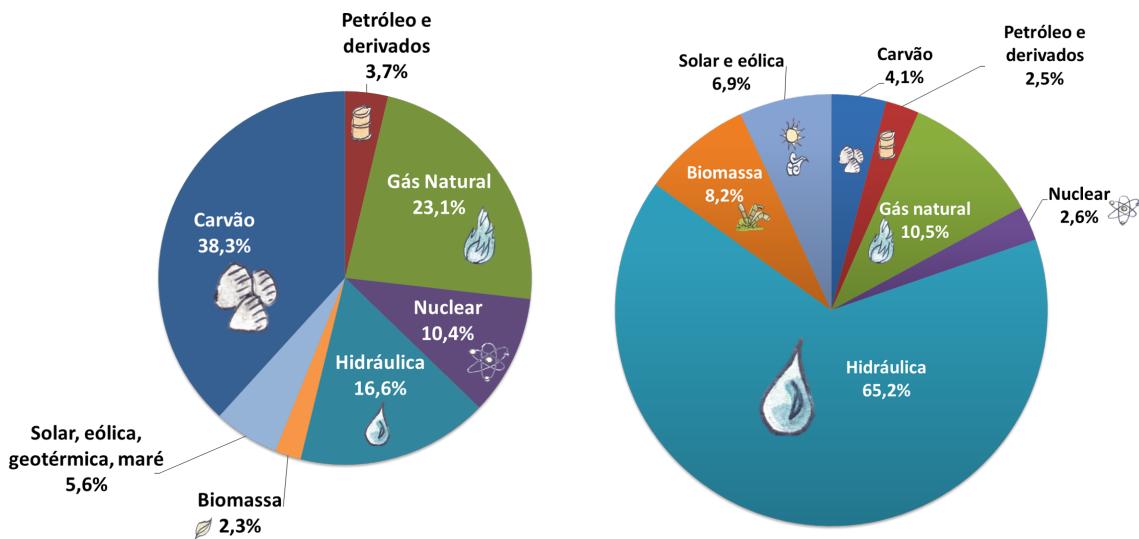


Figura 2.3: Planejamento da produção de energia do sistema hidrotérmico[9]

Sendo assim, utilizar as usinas hidrelétricas - UHEs na capacidade máxima gerará o menor custo de energia possível no mês atual, porém, se no futuro for necessário utilizar outras fontes para suprir a demanda energética, o custo pode ser muito maior do que em uma situação onde a produção energética da UHE foi distribuída durante os meses consecutivos. Por um outro lado, caso a quantidade de água armazenada em um reservatório supere o limiar máximo do mesmo, torna-se necessário liberar o excedente para o rio, situação chamada de "vertimento". Isto é indesejável, pois resulta em um desperdício de energia barata. Portanto, um planejamento equivocado pode causar altas no PLD, o que afeta diretamente os agentes que participam do mercado de energia elétrico brasileiro.

O Brasil atualmente possui uma matriz elétrica diferente do que em geral é observado no mundo. Em média 65,2% da energia utilizada no país vem de fontes hidráulicas, diferente do que ocorre em média no resto do mundo (aproximadamente 16,6%), assim como visto em [2]. Essa fonte além de ser renovável, tem impacto menor no meio ambiente e custos menores quando comparado com fontes baseadas em combustíveis fósseis. O grande ponto negativo é a característica volátil desse recurso, pois depende baste de condições climáticas, o que dificulta o planejamento para períodos longos.



(a) Matriz Elétrica Mundial. Fonte: EPE [2]. (b) Matriz Elétrica Brasileira. Fonte: EPE [2].

Figura 2.4: Comparação entre a matriz elétrica mundial e a brasileira

Uma tendência mundial no setor é a transição da rede elétrica atual para os smart grids. Este termo engloba uma série de tecnologias, tendo como foco o sensoriamento dos componentes que compõem a rede. Isto permite ter informações sobre o consumo em um período de tempo menor do que nas redes antigas, além de conseguir averiguar falhas na rede de maneira remota, mudando a postura das distribuidoras de reativa para ativa, pois torna-se possível resolver os erros antes que os clientes liguem para a concessionária.

Assim como visto em [10], no ano de 2012 a ANEEL aprovou a resolução normativa N° 482 [11], que "Estabelece as condições gerais para o acesso de microgeração e minigeração distribuída aos sistemas de distribuição de energia elétrica, o sistema de compensação de energia elétrica, e dá outras providências". Desde março de 2016 é permitido que qualquer fonte de energia renovável participe do sistema de geração distribuído, não limitando somente à solar e eólica. Sendo assim a tendência é que parte da produção de energia seja proveniente de pequenos geradores, diminuindo parcialmente a dependência de grandes produtores de energia, o que pode acarretar na redução do risco de deficit de energia.

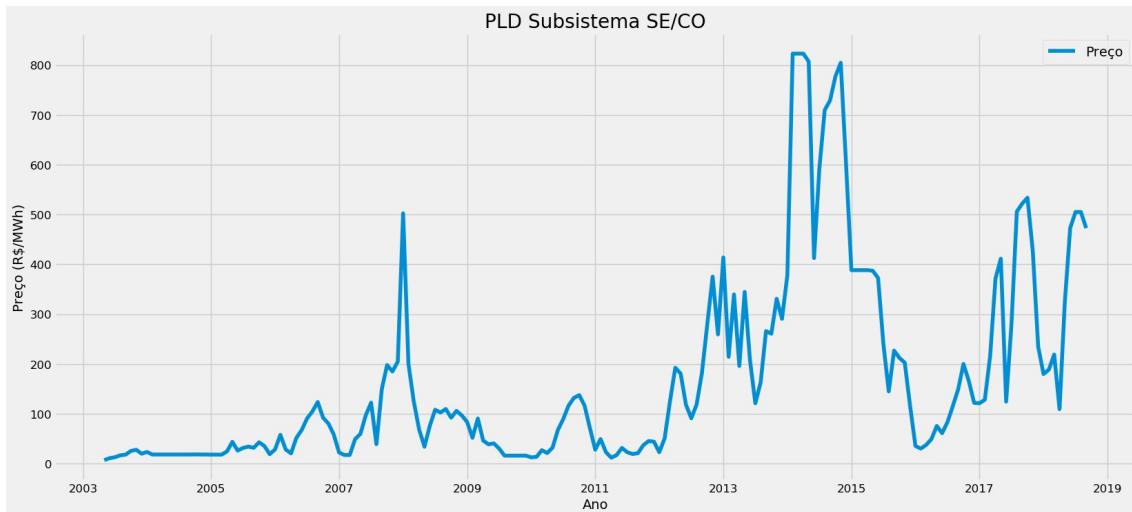


Figura 2.5: PLD mensal. Fonte: CCEE [4].

No gráfico acima observa-se que existem pontos onde o SIN foi afetado por crises, fazendo com que o PLD aumentasse abruptamente. No ano de 2007 ocorreu a primeira crise mostrada no gráfico. O SEB tinha passado por uma recente reestruturação da organização, a qual não foi suficiente para evitar o acontecimento.

Durante o período de racionamento (2001/06 - 2002/02) a demanda foi menor do que a oferta, porém com o passar dos anos ela voltou a se aproximar da curva de oferta. Após abril de 2005, houve um período sem investimento na capacidade energética sistema. Junto com esse fato, ocorreu um período de baixa afluência nas usinas hidrelétricas.

Outra crise notável ocorreu no ano de 2014, onde segundo [12], houve a diminuição das tarifas impostas pelo governo em um momento de demanda crescente, o atraso em obras de geração e transmissão, além da escassez de chuvas.

O PLD é calculado semanalmente e disponibilizado sexta-feira pela CCEE conforme a fórmula 2.3, com vigência na semana seguinte. Para cada um dos submercados do SIN são divulgados valores para os três patamares de carga: leve, médio e pesado. Esses estão relacionados com a hora do dia e o consumo, sendo que o patamar é pesado em horários de alto consumo e leve nos períodos de baixo consumo energético. O PLD mensal por subsistema é calculado também pela CCEE através da média dos preços ponderados pelo tempo em cada patamar.

$$PLD_{s,r,\omega} = \min(\max(CMO_{s,r,\omega}, PLD_{min_{ano}}), PLD_{max_{ano}}) \quad (2.3)$$

Onde:

Tabela 2.2: Variáveis utilizadas no cálculo do PLD semanal.

| Variável | Significado |
|-------------------|--|
| s | Subsistema ("N", "NE", "S", "SE\CO") |
| r | Patamar de carga ("leve", "médio", "pesado") |
| w | Semana do mês |
| CMO | Custo para produção de um MWh adicional no SIN |
| $PLD_{min_{ano}}$ | PLD mínimo estabelecido anualmente pela ONS |
| $PLD_{max_{ano}}$ | PLD máximo estabelecido anualmente pela ONS |

O PLD máximo é calculado com base no custo variável unitário - CVU mais elevado de uma Usina térmica - UTE em operação comercial, a gás natural, contratada

por meio Contrato de Comercialização de Energia Elétrica no Ambiente Regulado - CCEAR , definido no planejamento mensal de operação - PMO de dezembro e será aplicado entre a primeira e última semana operativa do ano subsequente, para todos os submercados.

O PLD mínimo é calculado com base no maior valor entre: i) o calculado com base na Receita Anual de Geração - RAG das usinas hidrelétricas em regime de cotas, nos termos da Lei nº 12.783/2013, excluídos os valores relacionados à remuneração e reintegração de investimentos, e adicionada a estimativa de Compensação Financeira pelo Uso dos Recursos Hídricos - CFURH; e ii) as estimativas dos custos de geração da usina de Itaipu para o ano seguinte, fornecidas pela Itaipu Binacional para fins de reajustes e/ou revisões tarifárias.

O valores de definidos para os últimos 5 anos foram:

Tabela 2.3: Preço mínimo e máximo do PLD.

| Ano | PLD Mínimo | PLD Máximo |
|------|------------|------------|
| 2015 | 30,26 | 388,48 |
| 2016 | 30,25 | 422,56 |
| 2017 | 33,68 | 533,82 |
| 2018 | 40,16 | 505,18 |
| 2019 | 42,35 | 513,89 |

O planejamento e o cálculo do PLD é feito com base em programas desenvolvidos especificamente para o SIN. Alguns deles são [13]:

- NEWAVE: Utilizado no planejamento de médio prazo (até 5 anos), com espaçamento mensal entre as previsões
- DECOMP: Utilizado no planejamento de curto prazo (até 12) com base semanal

cálculo do PLD é feito com base no CMO obtido do programa DECOMP, o qual é disponibilizado somente para agentes cadastrados no CCEE e tem como entrada os seguintes dados:

Tabela 2.4: Arquivos de entrada do DECOMP.

| Arquivo | Conteúdo |
|----------------|---|
| CASO.DAT | Contém o nome do Arquivo índice |
| Arquivo Índice | Contém a lista de arquivos de entrada sob gerenciamento do usuário |
| DADGER.DAT | Dados gerais de planejamento |
| VAZOES.XXX | Vazões incrementais afluentes aos aproveitamentos da configuração considerada |
| HIDR.DAT | Cadastros de dados das usinas |
| MLT.DAT | Médias mensais de longo termo |
| LOSS.DAT | Perdas no sistema |
| DADGNL.XXX | dados de usinas térmicas |

O Termo MLT - Média de Longo Termo refere-se à média com período anual, utilizando uma série histórica dos valores de ENA - Energia Natural Afluente para cada um dos meses desde 1931. A ENA, por sua vez é a energia que pode ser gerada através da vazão da água em uma determinada bacia.

Os arquivos HIDR.DAT, MLT.DAT e VAZOES.DAT vêm em formato binário e para serem lidos pelo operador, precisam de programas auxiliares para a conversão, sendo eles Hydroedit para o primeiro e Vazedit para os dois últimos. Ambos são fornecidos pela ONS. Já o arquivo de dados gerais tem 43 parâmetros, sendo alguns deles contratos de importações/exportação de energia, custo de déficit e a função de custo futuro. Esta última determinada pelo NEWAVE.

Obter um modelo matemático baseado no DECOMP que possa ser utilizado na previsão do PLD é de interesse não só dos agentes que participam do mercado elétrico brasileiro, como também de empresas indiretamente relacionadas ao setor financeiro e o meio acadêmico, pois, o entendimento da dinâmica do preço pode ajudar na formulação de novos modelos de previsão e alteração do já existente, buscando melhorias nas atividades de planejamento feitas pelos órgãos.

2.3 Revisão bibliográfica

Em [14] foi utilizada uma abordagem onde são previstos os valores de energia afluente, energia proveniente de hidroelétricas, energia proveniente de usinas térmicas e carga dos subsistemas, a fim de obter o valor do PLD indiretamente. A outra forma proposta foi utilizar a própria informação sobre o PLD na previsão do mesmo. Em ambos os casos, foram utilizadas redes neurais artificiais autorregressiva não-linear - NARNET e os valores previstos são para até 3 semanas à frente.

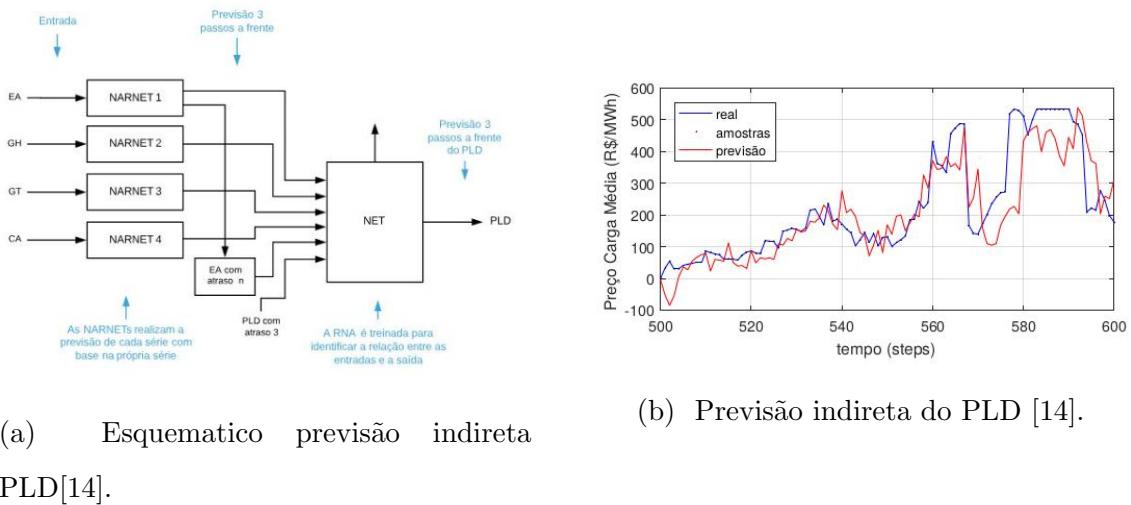


Figura 2.6: Previsão indireta do PLD

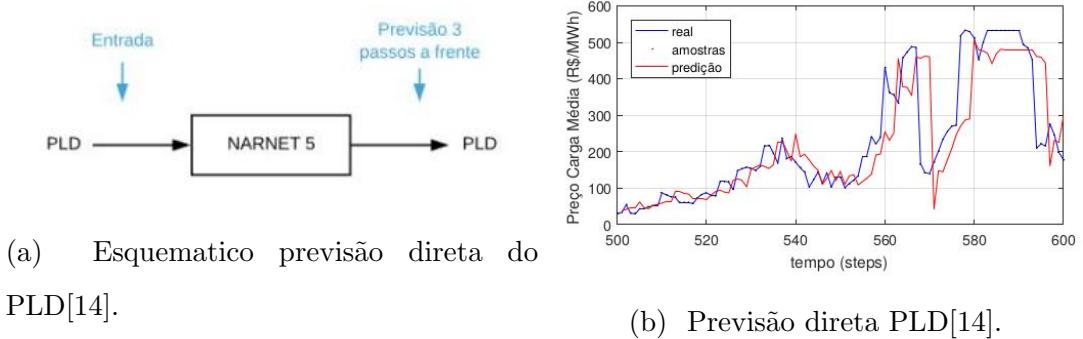
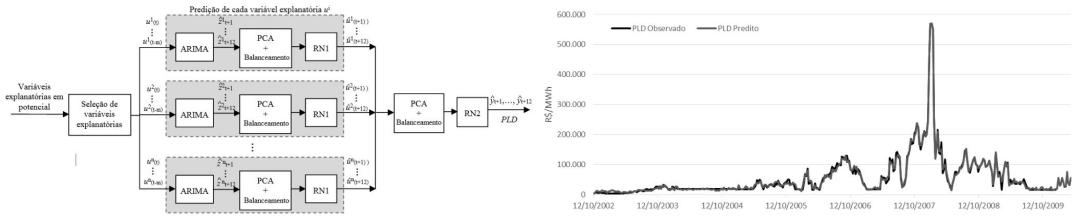


Figura 2.7: Previsão direta do PLD

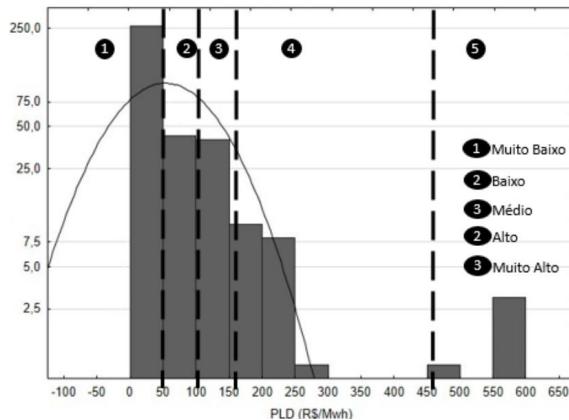
Outras abordagens são vistas em [9], onde é proposto um modelo híbrido com a utilização de *Autoregressive Integrated Moving Average* - ARIMA, análise de componentes principais e redes neurais, conforme a figura a seguir:



(a) Esquematico modelo híbrido de previsão do PLD. [9]. (b) Resultado da previsão do modelo híbrido. [9].

Figura 2.8: Previsão do PLD utilizando o modelo híbrido

Na segunda abordagem proposta pelo mesmo autor utilizou-se uma combinação do algoritmo C5.0 e árvore de classificação e regressão - CART para determinar se o preço do PLD será estará dentro de um patamar num instante futuro.



(a) Classes de saída do algorítmo [9].

| | Muito Baixo | Baixo | Médio | Alto | Muito Alto |
|--------------------|--------------------|--------------|--------------|-------------|-------------------|
| Muito Baixo | 342 | 0 | 0 | 2 | 0 |
| Baixo | 1 | 50 | 0 | 0 | 0 |
| Médio | 1 | 1 | 76 | 5 | 0 |
| Alto | 1 | 0 | 0 | 95 | 0 |
| Muito alto | 0 | 0 | 0 | 0 | 7 |

(b) Matriz de confusão [9].

Figura 2.9: Previsão do PLD utilizando o modelo com árvore de classificação

Em ambos os casos acima o autor realiza previsões para um período de 12 semanas à frente.

No trabalho visto em [15], têm se a análise das variáveis utilizadas na entrada de cada submercado através da correlação. Além disso, é feita a comparação entre os resultados obtidos pelo modelo ARIMA, rede neurais recorrentes, rede neural direta sem seleção de lags e com a seleção de lags. Já em [16] foram utilizadas como variáveis de entrada a geração das hidroelétricas, térmicas e eólica, além da Energia Natural Afluente - ENA, ENA máxima, EAR - Energia Armazenada e o PLD para cada um dos patamares nas 4 últimas semanas, utilizando uma rede neural com 3 camadas ocultas. Como métrica de avaliação dos resultados, faz-se um plot do alvo pelo resultado previsto pela rede e então é feito uma ajuste linear . No caso ideal a equação $y = a \cdot x + b$ se torna $y = x$ (previsão perfeita) . Os preços são previstos em um horizonte de até 6 semanas. Em [17] utiliza-se uma rede neural recorrente para obter previsões em um horizonte de até seis meses a frente.

Outra abordagem diferente das observadas acima é a vista em [18], onde o autor utiliza modelos *Generalized Auto-Regressive Conditional Heteroscedasticity*- GARCH para a previsão do PLD. Para a estimação dos parâmetros foram utilizadas análises com base na função de autocorrelações - ACF, autocorrelações parciais - PACF, função de verossimilhança (LLF) e testes estatísticos como o *Akaike info criterion* - AIC.

A previsão do preço da energia elétrica é um interesse comum de vários agentes envolvidos com o mercado de energia mundialmente. Em [19] o autor utiliza o algoritmo *K-means* para clusterizar os dados de entrada passados e o algoritmo *Partial Sequence-based Forecasting* - PSF para prever o resultado com base na média dos resultados obtidos no *clusters* anteriores. O modelo obtido foi utilizado para prever preço e demanda de energia para o mercado espanhol, australiano e nova-iorquino e obteve resultados relevantes.

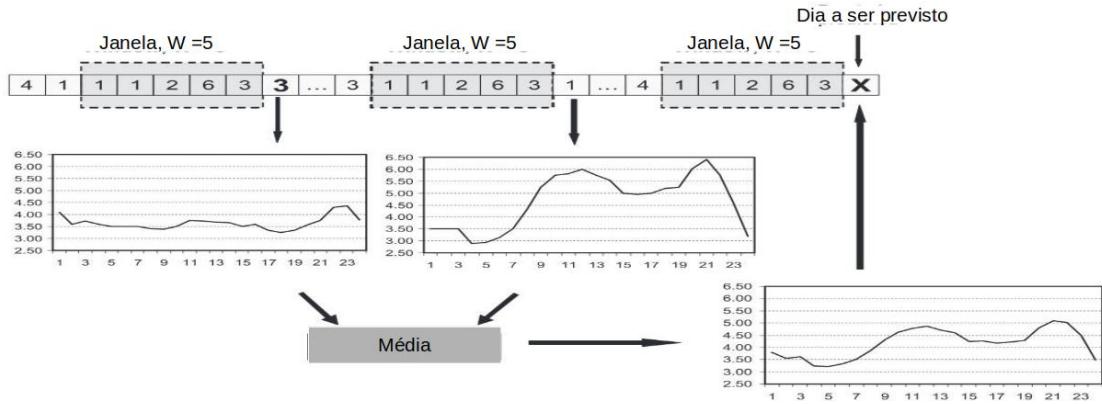


Figura 2.10: Utilização do algoritmo PSF para previsão de séries temporais [19].

Em [20] são mostradas algumas abordagens simples para o problema de previsão do preço *spot* (PLD é um exemplo de preço *spot*) no mercado de energia de Ontário. Foi utilizada uma rede neural com uma camada escondida, uma rede neural misturada com lógica fuzzy e otimização com o algoritmo IMO. A previsão foi feita hora a hora para um período de 24 horas. Concluiu-se que a rede neural com lógica fuzzy obteve o melhor resultado, além disso a relevância das entradas para a previsão foi analisada e viu-se que as entradas mais relevantes eram a demanda e a interrupção da geração.

A revista [21] traz a revisão de alguns resultados obtidos na previsão dos preços de mercados de energia elétrica como o da Noruega, Espanha e Califórnia. São utilizadas algumas abordagem lineares como AR, ARIMA e GARCH. Além disso são exibidos resultados de abordagens não lineares utilizando redes neurais. Alguns variantes mostrados são: RNA com lógica fuzzy, com filtro de Kalman extendido e com transformada de Fourier e Hartley como filtro para os dados de preço utilizados como entrada da rede. O paper [22] mostra os resultados para a competição global de previsão de energia, a qual ocorreu em 2014. São mostrados resultados tanto para previsão de preços da energia elétrica, carga de energia, energia eólica e solar.

Em [23] foi utilizada uma abordagem híbrida, onde o modelo wavelet-ARIMA seria responsável por prever a parte linear e a parte composta pelo modelo *Radial Basis Function Network* - RBFN seria responsável por corrigir o erro residual proveniente

da primeira parte.

Capítulo 3

Series Temporais e Aprendizado de Máquina

O objetivo deste capítulo é trazer a fundamentação teórica das técnicas utilizadas neste trabalho. Em primeiro lugar será abordada a teoria, aplicações e dificuldades relacionadas às séries temporais. Logo após, serão apresentados os modelos de aprendizados de máquina utilizados como redes neurais MLP.

3.1 Séries Temporais

Uma série temporal é composta por uma coleção de observações feitas de forma sequencial e dependente [24]. Essa ordem da sequência é dada pelo tempo, o qual pode ser contínuo ou discreto. No primeiro caso, $T = t : t_1 < t < t_2$ e a série temporal é definida como $\{X(t) : t \in T\}$, já no segundo caso, $T = \{t_1, t_2, \dots, t_n\}$ e a série temporal é definida como $\{X_t : t \in T\}$, onde X é a variável observada. Geralmente define-se o T para o caso discreto como $T = \{1, 2, \dots, n\}$ por questões de simplicidade. Neste trabalho só serão abordados os casos discretos, pois todos os sinais utilizados no mesmo são dessa natureza.

Assim como o tempo, os valores da variável X_t podem ser contínuos ou discretos de acordo com o fenômeno que se observa. Alguns exemplos de fenômenos temporais com valores contínuos são a temperatura em um determinada região, volume de água em uma bacia hidrográfica e o peso de um indivíduo. Já como exemplo de

fenômenos temporais com valores discretos podem ser citados o número viagens de avião, quantidade de nascimentos, de carros produzidos por uma montadora etc. Todos esse casos estão relacionados com um período de observação próprio, podendo ser uma janela de meses, anos, até mesmo décadas de observações de uma determinada variável.

A análise de séries temporais pode ser feita com diferentes intuições, sendo os mais comuns a predição de valores futuros com base no histórico já conhecido, o controle de um processo, a explicação e descrição de fenômenos [24].

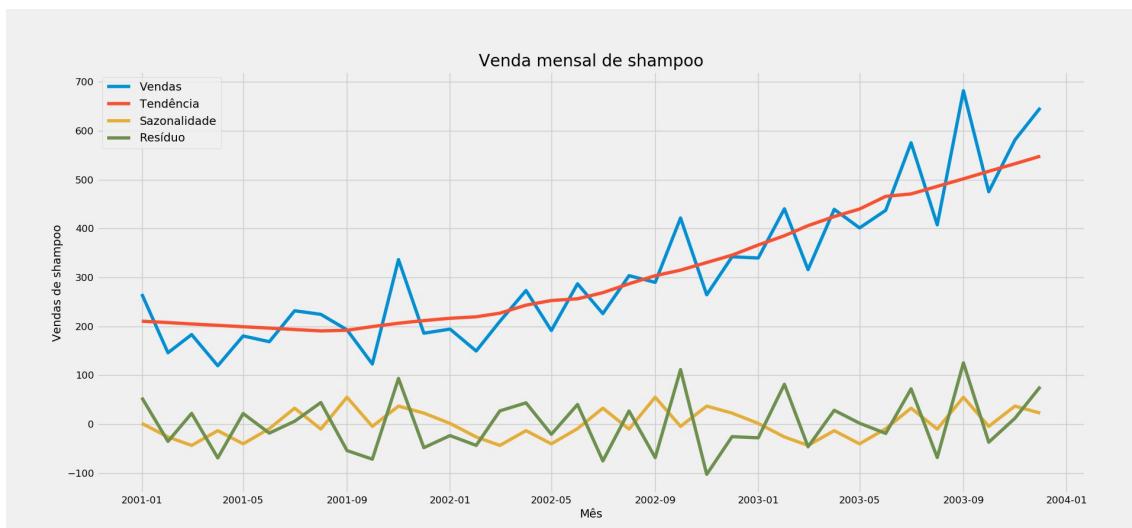


Figura 3.1: Exemplo de decomposição de série temporal. Fonte: Makridakis, Wheelwright and Hyndman (1998) [25]

Séries temporais geralmente são decompostas em outras mais simples. A decomposição utilizada será a seguinte, conforme visto em [26]:

$$X_t = \text{tend}_t + \text{sz}_t + \text{cs}_t + \text{rest}_t$$

Onde tend_t é a tendência, sz_t é a sazonalidade, cs_t é o ciclo senoidal e rest_t é o resíduo. Tanto a tendência, quanto a sazonalidade e os ciclos senoidais são determinísticos. Conforme [27], boa parte das séries temporais são não estacionárias, sendo que as componentes de tendência e sazonalidade são as maiores responsáveis por esse efeito. Para que o modelo com redes neurais faça boas previsões, que é o que

se busca nesse trabalho, é necessário utilizar séries estacionárias, portanto somente a parte residual será usada na entrada das redes neurais.

3.1.1 Tendência

Segundo [24], a tendência pode ser vista como "uma mudança de longo prazo no nível médio da série" e a forma mais simples de modelar pode ser vista pela equação a seguir.

$$tend_t = \alpha + \beta t + \epsilon_t \quad (3.1)$$

Onde α e β são constantes a serem estimadas e ϵ_t denota um erro aleatório com média zero. Geralmente chama-se o termo $m_t = \alpha + \beta t$ de termo de tendência, mas alguns autores chamam o termo β de tendência, já que $\beta = m_t - m_{t-1}$. Essa variável indica a inclinação da função durante o tempo.

A função utilizada na aproximação da tendência pode ser escolhida de acordo com a série que está sendo analisada. Uma forma bastante comum é a utilização de uma função polinomial na extração de tendência.

Para séries monotonicamente crescente ou decrescente, utilizar $p = 1$ (função linear) ou $p = 2$ (função quadrática) geralmente é suficiente para a extração da tendência, porém caso a série seja mais complexa, pode ser necessário utilizar funções de ordem mais altas.

$$tend_t = \epsilon_t + \sum_{n=0}^p \beta_n t^n \quad (3.2)$$

Alguns métodos de filtragem podem ser utilizados também na extração de tendência. É comum utilizar filtros lineares nessa tarefa. Esses são definidos pela seguinte equação:

$$y_t = \sum_{j=-q}^s a_j x_{t+j} \quad (3.3)$$

Onde a_j são os pesos que multiplicam o sinal x_{t+j} . Para o filtro de médias móveis geralmente utiliza-se $q = s$ e $a_{-r} = a_r$, garantindo a simetria do filtro. Além disso faz-se que $\sum_{j=-q}^s a_j = 1$, de modo que $\min\{x_t\} \leq y_t \leq \max\{x_t\}$. O caso mais simples de média móvel é aquele onde todos os pesos tem o mesmo valor:

$$y_t = \frac{1}{2q+1} \sum_{j=-q}^q x_{t+j} \quad (3.4)$$

O resultado do filtro acima é não-causal, o que impede que o processamento seja utilizado para a previsão de séries. Sendo assim, uma outra abordagem possível é fazer um deslocamento no filtro para que sejam utilizadas somente amostras do passado, conforme a equação a seguir:

$$y_t = \frac{1}{2q+1} \sum_{j=-2q}^0 x_{t+j} \quad (3.5)$$

Outro problema observado nas abordagens de média móvel descritas acima é que somente se obtém a tendência para $N - 2q$ pontos. Caso seja necessário obter a tendência para todos os pontos da série, pode-se aplicar métodos de extração sobre o resultado obtido.

Uma terceira abordagem para extração de tendência é utilizar um filtro com pesos que decaem geometricamente, com j , priorizando assim, as amostras mais recentes da série temporal:

$$y_t = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j} \quad (3.6)$$

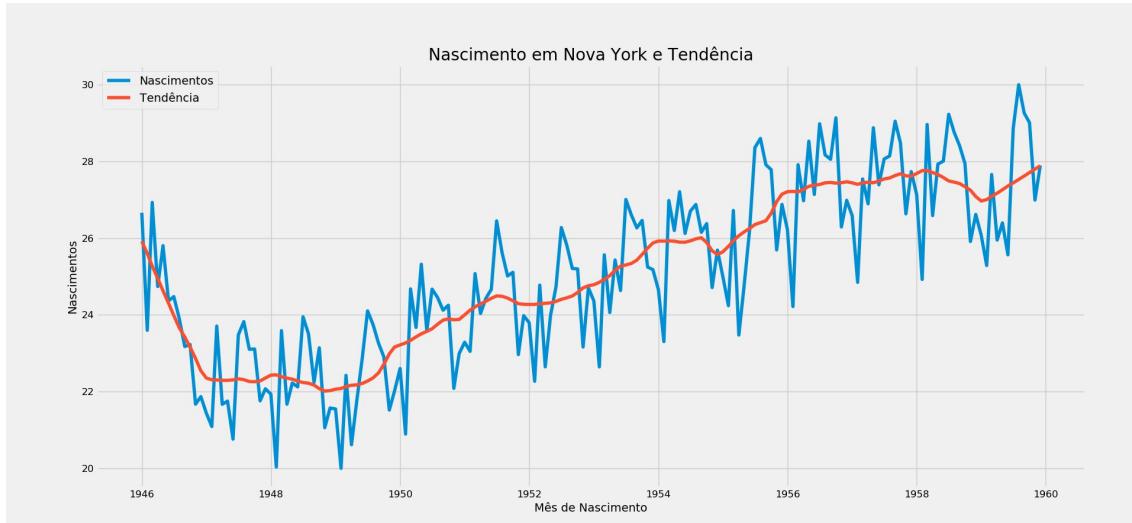


Figura 3.2: Exemplo de extração de tendência com filtro de média. Fonte: Newton (1998). [1]

Outra forma de extrair a tendência é através da diferenciação. Para dados não sazonais, a primeira diferença costuma ser suficiente para garantir a estacionariedade aproximada da série restante [24]:

$$y_t = x_t - x_{t-1} = \nabla x_t \quad (3.7)$$

3.1.2 Sazonalidade

É comum encontrar nas séries observadas alguns padrões que se repetem periodicamente. Esse efeito é denominado sazonalidade e deve ser removido para que se obtenha ao final do processamento uma série residual estacionária [26].

Segundo [26], a sazonalidade pode ser determinada pela seguinte fórmula:

$$sz_t = \frac{1}{Int(N/P)} \sum_{k=0}^{Int(N/P)} s_i(i + kP) \quad i = 1, \dots, P \quad (3.8)$$

Onde N é o número de amostras, P é o período sazonal, $Int(N/P)$ é o resultado inteiro da divisão de N/P e s_i é o sinal com a tendência previamente removida. Sendo assim, é feito uma média dos pontos da série temporal espaçados pelo período

P . A sazonalidade se repete durante a série temporal, então, caso seja desejado obter a sazonalidade em um tempo $0 < t < N$ utiliza-se a seguinte fórmula:

$$sz_t = sz[Resto(t/P)] \quad (3.9)$$

A periodicidade do fenômeno sazonal pode ser obtida através do conhecimento prévio da série que está sendo analisada ex: Espera-se que a venda de protetores solares seja maior no período de verão, pois é quando as pessoas costumam ir mais às praias. O número de pessoas que frequentam o metrô deve diminuir durante o fim de semana, pois a maioria trabalha durante a semana etc.

Outra forma de se obter o período é fazendo uma inspeção visual sobre o gráfico da série após a remoção da tendência. Em alguns casos serão visíveis os padrões periódicos.

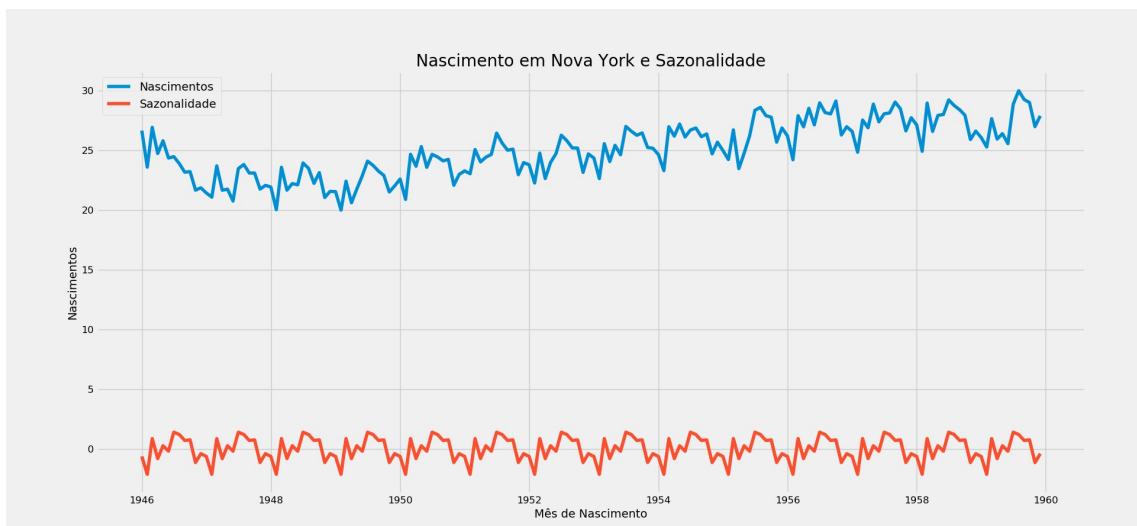


Figura 3.3: Extração de sazonalidade com periodicidade anual ($P = 12$) para o dataset de nascimentos em Nova York. Fonte: Newton (1988). [1]

Também pode-se obter a informação sobre o período através da observação do gráfico de autocorrelação, onde picos de magnitude que seguem um padrão de espaçamento podem indicar a periodicidade da sazonalidade.

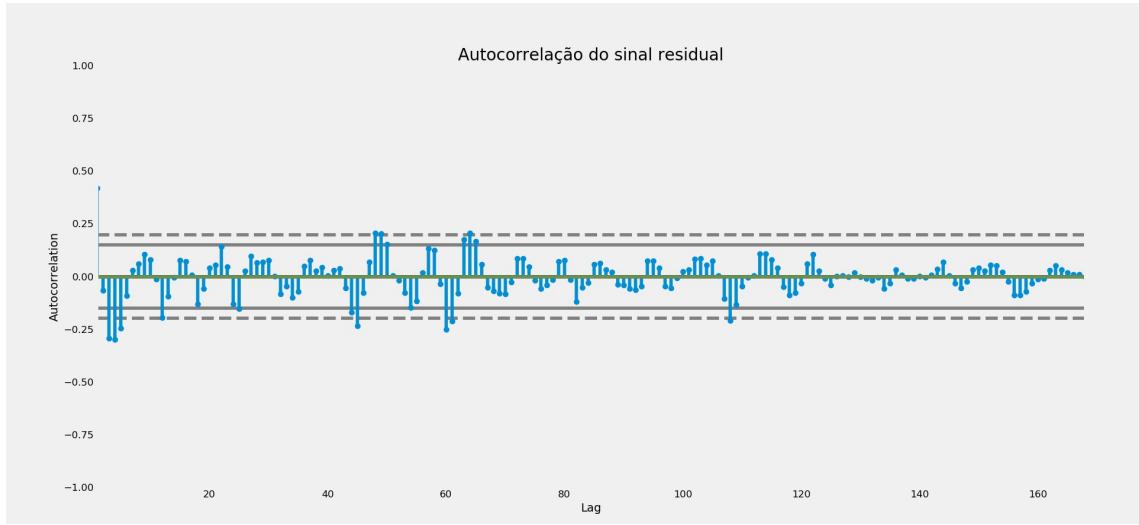


Figura 3.4: Autocorrelação para o sinal residual do dataset de nascimentos em Nova York. Fonte: Newton (1988). [1]

3.1.3 Ciclos Senoidais

Os ciclos senoidais representam um caso bem específico de sazonalidade, sendo representados por senoides de período P . Essa senoide é extraída através da análise do espectrograma dado pela FFT-*Fast Fourier Transform*. Os parâmetros de saída são os termos a e b , conforme vistos nas equações abaixo[28]:

$$cs_t = a \cdot \cos(2\pi ft) + b \cdot \sin(2\pi ft) \quad (3.10)$$

Os pontos onde a magnitude do sinal ($\sqrt{a^2 + b^2}$) são muito maiores que os outros indicam provável ciclos senoidais que devem ser removidos.

3.1.4 Componente Residual

Caso o processo de extração de componentes descrito nos tópicos acima seja realizado com sucesso, será obtida uma componente residual estacionária. Seguindo esta abordagem, este sinal residual é a única parte aleatória do modelo. Sendo assim, o problema se reduz a prever o comportamento da parte residual da série temporal.

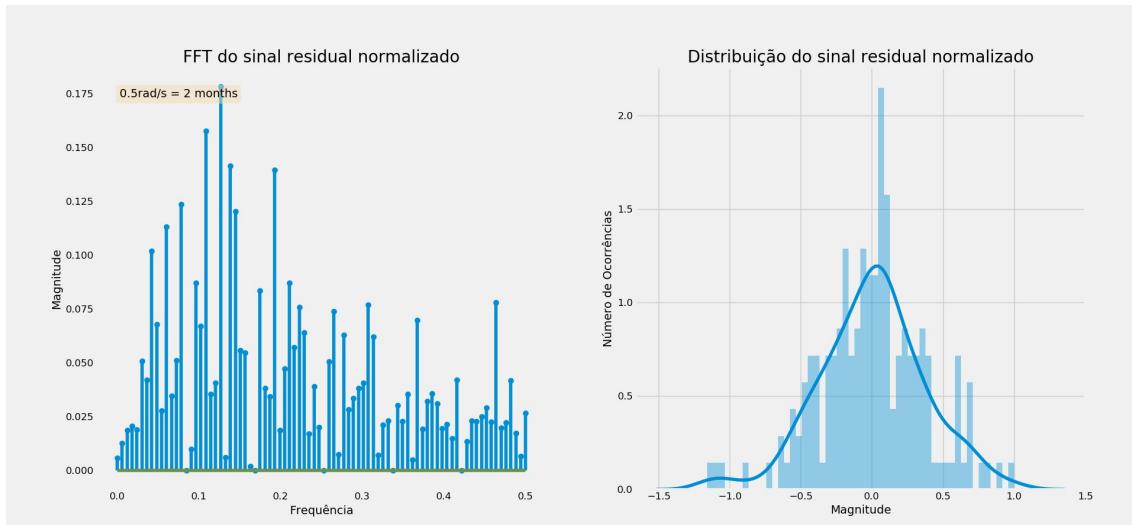


Figura 3.5: Resíduo sem remoção de ciclos senoidais para o dataset de nascimentos em Nova York. Fonte: Newton (1988). [1]

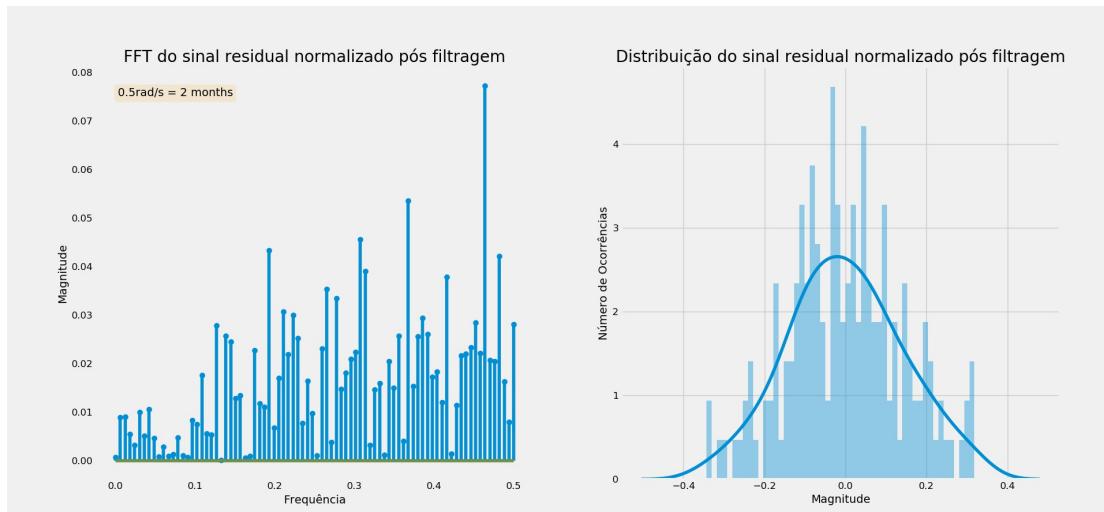


Figura 3.6: Resíduo com remoção de ciclos senoidais para o dataset de nascimentos em Nova York. [1]

Como pode-se observar, a discrepância no sinal residual nas frequências próximas de $\omega_0 = 0.135\text{rad}\backslash\text{amostra}$ diminui. Além disso, a distribuição se tornou menos espaçada (O desvio padrão mudou de aproximadamente $\sigma = 0.37$ para $\sigma = 0.14$).

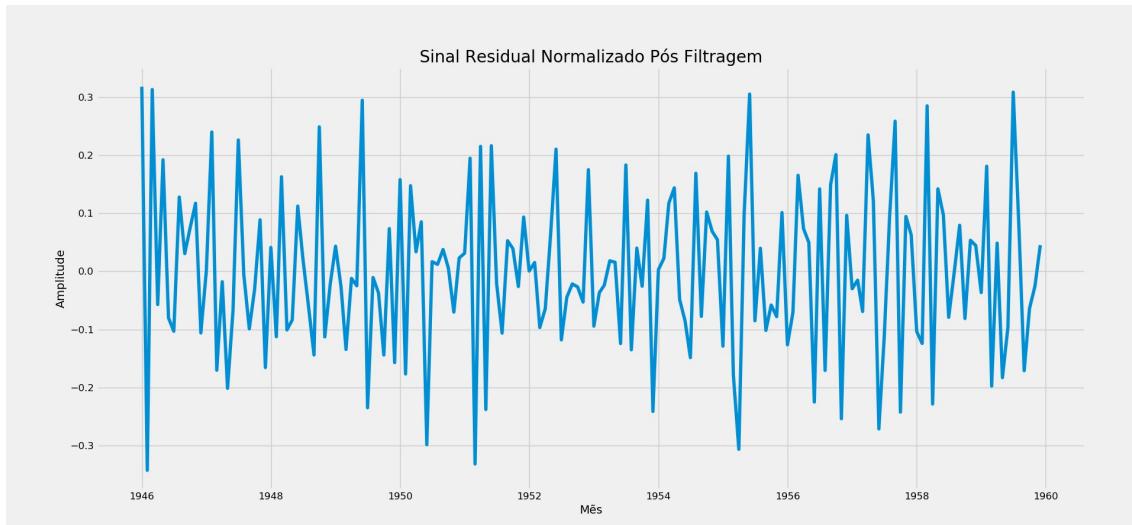


Figura 3.7: Sinal residual no tempo para o dataset de nascimentos em Nova York Fonte: Newton (1988). [1]

3.2 Processamento de Sinais

Para remover os ciclos senoidais do espectrograma visto em 3.5 é necessário realizar algum tipo de filtragem sobre sinal. Uma das formas de se classificar um filtro é pela sua resposta em frequência, sendo as mais comuns: passa-baixa, passa-alta, passa-banda e rejeita-banda.

No problema de remoção de ciclos senoidais, busca-se um filtro que remova somente a frequência com maior magnitude, sem afetar muito as magnitudes das outras frequências presentes no espectrograma. Para isso procura-se um filtro que seja rejeita-banda com a banda de rejeição bem estreita e banda de passagem aproximadamente plana.

Um filtro bastante conhecido na literatura que atende a esse critério é o Notch. A escolha pela versão IIRs se dá pela possibilidade de obter atenuações maiores e banda de rejeição mais estreita para um mesma ordem N quando comparado com os filtros FIRs. O custo dessa escolha é a existência de defasagem não linear, a qual não é problema para esta aplicação. A função de transferência do filtro Notch de segunda ordem se dá pela equação a seguir: [28]:

$$H(z) = b \cdot \frac{1 - 2 \cos \omega_0 z^{-1} + z^{-2}}{1 - 2b \cos \omega_0 z^{-1} + (2b - 1)z^{-2}} \quad (3.11)$$

e

$$b = \frac{1}{1 + \beta} = \frac{1}{1 + \frac{\sqrt{1-G_b^2}}{G_b} \tan(\frac{\Delta\omega}{2})} \quad (3.12)$$

Onde ω_0 é a frequência que se deseja rejeitar, $\Delta\omega$ é a banda de rejeição, G_b é a atenuação na frequência de corte. Geralmente utiliza-se $G_b = 3dB$. O parâmetro Q citado na seção 3.1.4 pode ser definido também como $Q = \frac{\omega_0}{bw}$. bw por sua vez é a banda de rejeição do filtro Notch.

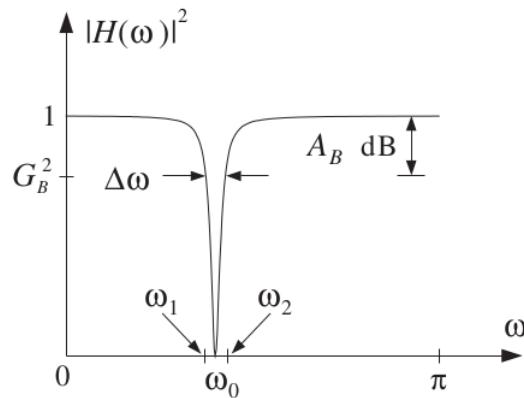


Figura 3.8: Filtro Notch IIR digital Fonte: Introduction to Signal Processing. [28]

3.3 Redes Neurais Artificiais

Redes Neurais artificiais são modelos computacionais que tentam reproduzir o comportamento observado na estrutura cerebral dos seres vivos. O neurônio pode ser considerado a célula básica de processamento do cérebro humano. Sua estrutura é dividida em três partes principais [26] [29] [30]:

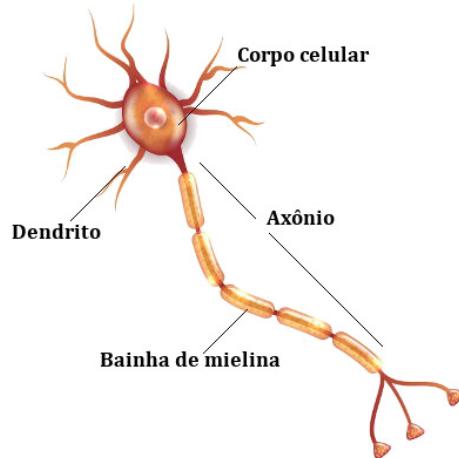


Figura 3.9: Neurônio biológico. Fonte: [31]

- **Dendritos:** São responsáveis por receber estímulos elétricos de outros neurônios
- **Corpo celular:** Processa as informações recebidas pelos dendritos e determina se será disparado um impulso elétrico
- **Axônio:** Transmite o impulso elétrico, e, através das sinapses, envia a informação para outros neurônios. Isto ocorre sem contato entre os mesmos.

A representação matemática desse modelo é dada pela seguinte estrutura:

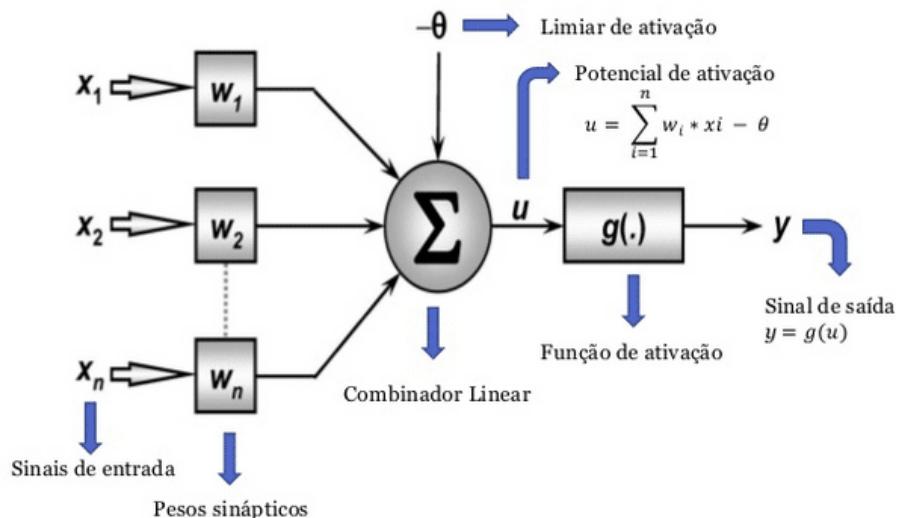


Figura 3.10: Neurônio Artificial. Fonte: [32]

A entrada do neurônio é um vetor $X = [x_1, x_2, \dots, x_n]$ análogo ao sinais elétricos transmitidos no cérebro humano. Essa entrada é ponderada por um conjunto de pesos $W = [w_1, w_2, \dots, w_N]$ e somada em um combinador linear junto com um limiar de ativação θ . O somatório das entradas gera um potencial de ativação u , o qual passa por uma função de ativação e gera um sinal de saída que poderá ser propagado para outros neurônios [29]. As informações descritas acima se resumem nas seguintes equações [33]:

$$u_j = \sum_{i=1}^N w_{ji} \cdot x_i - \theta \quad (3.13)$$

$$y = g(u) \quad (3.14)$$

Sendo que se considerar $x_0 = 1$ e $w_0 = -\theta$, pode-se definir a equação 3.13 como:

$$u = \sum_{i=0}^n w_i \cdot x_i \quad (3.15)$$

A função de ativação pode ter diferentes formatos. Caso seja identidade, obtém se um regressor linear [34]. Este tipo de abordagem traz uma grande desvantagem, pois a saída do sistema sempre será linear. Isto vem do fato de que uma composição de transformações lineares é também uma transformação linear. Sendo assim, nas redes neurais são utilizadas funções não-lineares. Alguns exemplos são:

- **Função Logística:**

$$g(u) = \frac{1}{1 + e^{-\beta u}} \quad (3.16)$$

Onde β é uma constante real que modifica a inclinação da reta.

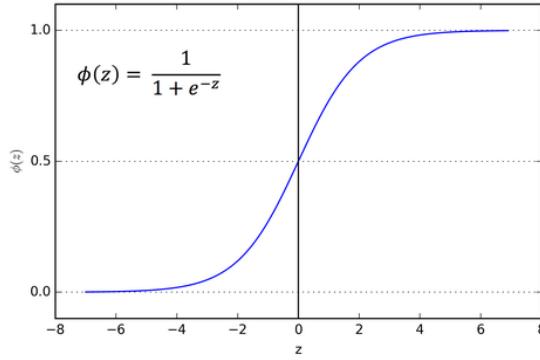


Figura 3.11: Neurônio Artificial. Fonte: [35]

- **Tangente Hiperbólica:**

$$g(u) = \frac{1 - e^{-\beta u}}{1 + e^{-\beta u}} \quad (3.17)$$

Onde $-1 \leq g(u) \leq 1$ para qualquer u e assim como em 3.16, β também modifica a inclinação da reta.

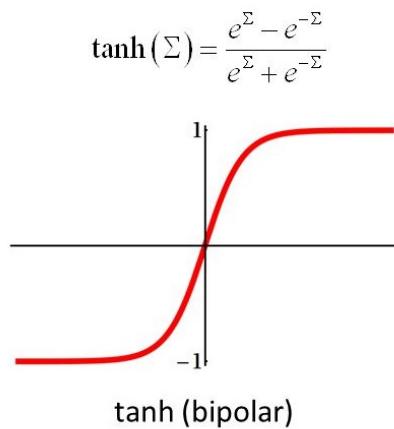


Figura 3.12: Tangente Hiperbólica. Fonte: [36]

- **Unidade Linear Retificada - ReLU [37]:**

$$g(u) = \max(0, u) \quad (3.18)$$

Esta função é linear na parte positiva e zero na parte negativa.

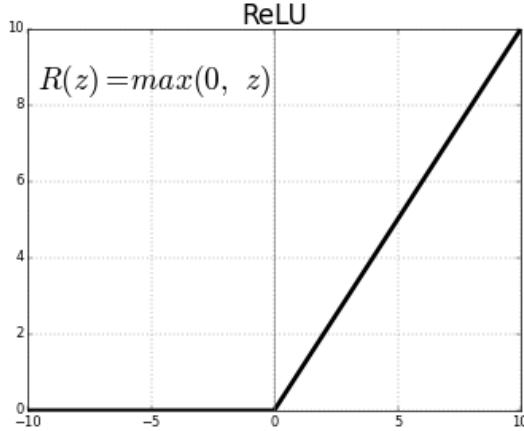


Figura 3.13: ReLU. Fonte: [38]

As funções de ativação 3.16 e 3.17 são deriváveis em todos os pontos e a 3.18 só não é derivável no ponto zero, porém contorna-se essa limitação fazendo $g'(0) = 0$. A ReLU tem sido essencial para o estado da arte de redes neurais [38] [39] [40] [41]. A derivada da função de ativação é utilizada pelos algoritmos de treinamento baseados no gradiente do erro assim como será visto mais à frente.

3.3.1 Backpropagation

O algoritmo de backpropagation é bastante utilizado no treinamento de redes neurais e utiliza o gradiente do erro como base dos cálculos, assim como mencionado anteriormente [34] [42]. Busca-se mover o vetor dos pesos na direção do mínimo global. A expressão de atualização dos pesos é da seguinte forma:

$$w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E_n w^{(\tau)} \quad (3.19)$$

O qual deve ser repetido até que o erro se torne suficientemente pequeno. O gradiente do erro nessa fórmula é dado por:

$$\nabla E^{(\tau)} = \frac{\partial E}{\partial W_{ji}^{(\tau)}} = \frac{\partial E}{\partial Y_j^{(\tau)}} \cdot \frac{\partial Y_j^{(\tau)}}{\partial u_j^{(\tau)}} \cdot \frac{\partial u_j^{(\tau)}}{\partial W_{ji}^{(\tau)}} \quad (3.20)$$

Têm-se como ideia principal do mesmo avaliar quanto que um determinado peso em uma camada influencia no erro da saída e assim, modificá-lo de forma a tornar

esse erro menor. Um ponto importante para o sucesso do algoritmo é a normalização da entrada, visto que diminui o tempo de convergência.

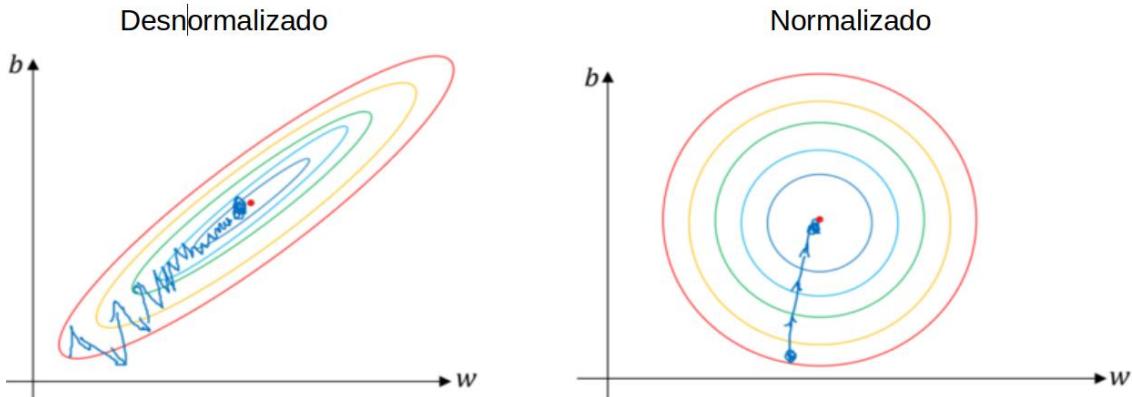


Figura 3.14: Gradiente Descendente desnormalizado e normalizado. Fonte: [38]

3.3.2 Rede perceptron multicamadas

Assim como no cérebro, os neurônios artificiais podem ser agrupados em estrutura mais complexas. Para uma camada inicial com N entradas têm se na J -ésima saída:

$$Y_j^{(1)} = g\left(\sum_{i=0}^N W_{ji}^{(1)} \cdot X_i\right) \quad (3.21)$$

Nas camadas seguintes utiliza-se a saída da camada anterior (com M neurônios) como entrada na camada atual. Na fórmula busca-se obter a saída para o P -ésimo neurônio da camada H .

$$Y_P^{(H)} = g\left(\sum_{i=0}^M W_{pi}^{(H)} \cdot Y_i^{(H-1)}\right) \quad (3.22)$$

As redes MLP tem sido utilizadas em diferentes classes de problemas como classificação de elementos e previsão de séries temporais [26] [29] [43]. Com o grande crescimento do número de dados disponível para utilização e o desenvolvimento das tecnologias computacionais, têm-se atualmente redes com muitas camadas e neurônios. Essas redes tem como objetivo aumentar a capacidade de separação, previsão, além de poder obter informações relevantes sobre neurônios intermediários da rede [44] [45] [46].

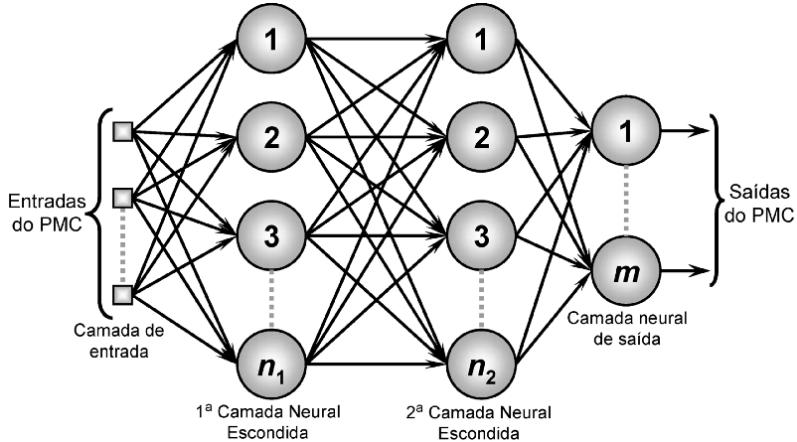


Figura 3.15: Rede Perceptron Multicamadas. Fonte: [47]

3.3.3 Treinamento

Para que a rede consiga de fato ”aprender” com os dados é necessário realizar o treinamento. Para que o mesmo seja bem sucedido deve-se atentar para alguns fatores como:

- **Inicialização dos pesos:** Os pesos não devem ser iniciados com o mesmo valor, pois isto faria com que cada neurônio interprete a entrada da mesma forma, gerando então uma estrutura simétrica [38] [48] [49].
- **Função custo:** Há necessidade de definir qual função de custo será utilizada na avaliação dos resultados da rede. A função mais comum é o erro médio quadrático - *MSE*, porém dependendo da análise que se deseja fazer e do problema a ser resolvido, outras funções podem ser utilizadas como o erro médio quadrático - *RMSE*, erro absoluto - *MAE* e acurácia - *ACC* [27].
- **Curva de aprendizado:** É comum também utilizar um gráfico do erro na saída em função da época de treinamento para o conjunto de treinamento e validação. Através do mesmo é possível observar características como overfitting e underfitting e selecionar o conjunto de pesos que tem o melhor compromisso [50].
- **Quantidade de dados X complexidade da rede:** Outro fator importante a ser observado é a quantidade de dados disponível para treinamento, visto que

quanto maior a complexidade estrutural da rede, maior a capacidade de gerar funções mais complexas, portanto torna-se necessário uma maior quantidade de dados para que a mesma seja treinada sem o efeito de overfitting.

- **Taxa de aprendizado:** Caso o fator η da fórmula 3.19 seja um valor muito grande, o algoritmo não conseguirá convergir para um mínimo, porém se η for um número muito grande, o treinamento pode levar muita épocas até convergir. Cabe então a quem especifica os parâmetros da rede neural escolher um η adequado de forma com que a convergência ocorra e não demore demais.

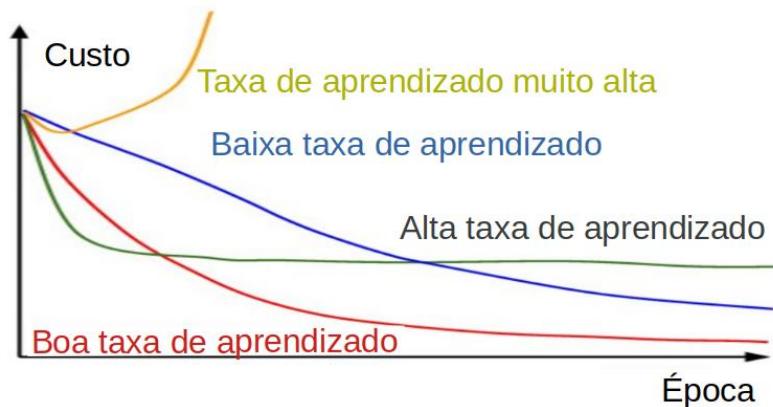


Figura 3.16: Influência da taxa de aprendizado. Fonte: [51]

- **Divisão dos dados de entrada:** Uma prática comum para se obter resultados consistentes é dividir os dados em um conjunto de testes e outro de validação, de forma que o conjunto de testes não seja usado no treinamento, sendo assim usado para verificar o quanto bom são os resultados da rede para dados desconhecidos. Outra prática comum é dividir o conjunto de treinamento em treino e validação e utilizar a validação cruzada, de forma a realizar vários treinos com a mesma arquitetura, permitindo obter melhores resultados no treinamento [52].

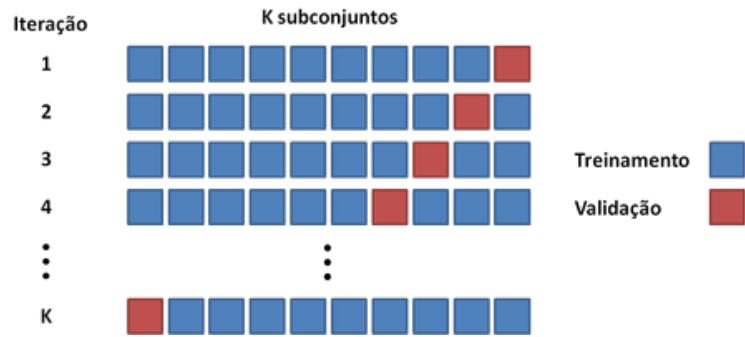


Figura 3.17: Validação cruzada. Fonte: [53]

Capítulo 4

Método

O objetivo deste capítulo é descrever o método utilizado no desenvolvimento do trabalho. Será descrito o fluxo de dados utilizado na previsão do PLD. Isso inclui o pré-processamento dos dados e o treinamento das redes neurais.

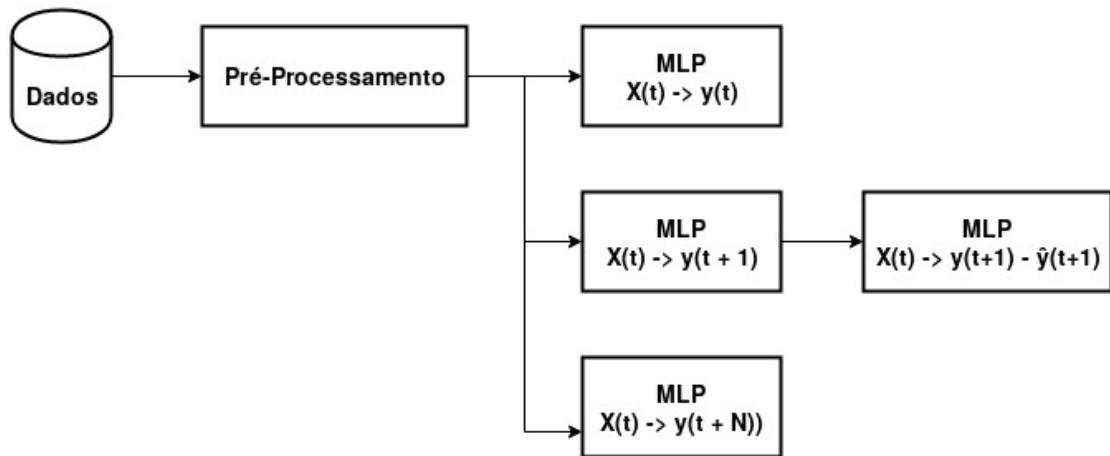


Figura 4.1: Diagrama do Método.

4.1 Coleta de dados

Serão utilizados nos modelos os seguintes dados do setor elétrico:

Todos os dados acima estão disponíveis publicamente e são séries temporais com intervalo de amostragem de 1 mês.

Tabela 4.1: Dados utilizado no trabalho.

| Dado | Tipo | Fonte |
|--|---------|----------|
| PLD | Saída | CCEE [4] |
| Energia Total Armazenada | Entrada | ONS [3] |
| Energia Gerada pelas Usinas Hidroelétricas | Entrada | ONS [3] |
| Energia Gerada pelas Usinas Térmicas | Entrada | ONS [3] |
| Energia Gerada pelas Usinas Solares | Entrada | ONS [3] |
| Energia Gerada pelas Usinas Eólicas | Entrada | ONS [3] |
| Energia de Carga | Entrada | ONS [3] |
| Energia Natural Afluente | Entrada | ONS [3] |
| Soma das Vazões Afluentes | Entrada | CCEE [4] |
| Valor Útil da Soma das Vazões Afluentes | Entrada | CCEE [4] |

Para a obtenção da soma das vazões, utilizou-se o arquivo VAZOES.DAT fornecido pela CCEE. Este arquivo é binário e para ser lido precisa da utilização do executável vazedit também fornecido pela CCEE. Após o processamento deste programa, obtém-se um arquivo texto com as vazões de cada posto fluviométrico para cada mês em cada ano desde de 1931. Esses valores foram somados para os postos que pertencem exclusivamente à região Sudeste.

O ”valor útil da soma das vazões afluentes” é um cálculo semelhante ao descrito no parágrafo anterior, porém removendo em cada posto a vazão mínima necessária para a produção de energia. Para obter essa informação, utiliza-se o arquivo HIDR.DAT, também fornecido pela CCEE, o qual possui o cadastro de todas as usinas consideradas nos cálculos do PLD. Assim como no caso do arquivo VAZOES.DAT, o HIDR.DAT também é binário. Para processar o mesmo é necessário um executável também fornecido pela CCEE chamado hidroedit.

4.2 Pré-processamento

A etapa de pré-processamento descrita nessa seção foi aplicada tanto nos sinais de entrada quanto na saída. Busca-se decompor o sinal conforme mencionado em

3.1. Ao final do processamento, o sinal X_t poderá ser obtido através das soma dos sinais extraídos, assim como visto na formula a seguir: $X_t = tend_t + sz_t + cs_t + rest_t$. O primeiro passo a ser realizado é a extração da tendência.

4.2.1 Tendência

Dado o tamanho reduzido da série temporal, não foi possível aplicar a média móvel diretamente pois isso implicaria na perda de alguns pontos da série. Sendo assim, utilizou-se uma abordagem híbrida na extração da tendência onde as primeiras amostras são obtidas através de uma regressão linear e a média móvel é usada no restante da série. A transição entre os dois métodos costuma ser bastante brusca, gerando altas frequências na transformada de Fourier. A solução para esse novo problema foi implementar um transição gradual entre os métodos de extração. O processamento realizado na obtenção da tendência pode ser visto na fórmula abaixo:

$$\hat{y} = \begin{cases} X_i, & \text{se } i \leq 2 \\ LR(1, X_i), & \text{se } 2 < i < W \\ \alpha LR(1, X_W) + (1 - \alpha) MA(W), & \text{se } W \leq i < W + K \\ MA(W), & \text{se } i \geq W + K \end{cases} \quad (4.1)$$

Para um filtro de média de comprimento W com K passos de transição e com $0 \leq \alpha \leq 1$ proporcional ao número de passos de transição, define-se o filtro acima para resolver o problema da falta de pontos para realizar a média móvel. LR se refere à regressão linear e MA refere-se à média móvel e a qual é definida pela seguinte equação:

$$MA(M) = \frac{1}{W+1} \sum_{j=-W}^0 x_{t+j} \quad (4.2)$$

A outra proposta de extração de tendência utilizada foi a seguinte:

$$\hat{y} = \begin{cases} X_i, & \text{se } i \leq W \\ MA(W), & \text{se } i > W \end{cases} \quad (4.3)$$

As duas fórmulas acima dependem do parâmetro (W) o qual varia conforme o problema a ser resolvido. A escolha do W foi feita a partir da análise do MSE pelo

tamanho da janela. Como o objetivo é obter uma função que tenha a capacidade de prever a tendência um passo à frente com o mínimo erro possível, busca-se um valor de W que não seja pequeno demais (pouca generalização) nem muito grande (pouco sensível às variações).

4.2.2 Sazonalidade

Após a extração da tendência, o sinal restante é descrito pela seguinte fórmula: $s_{1t} = sz_t + cs_t + res_t$. O método de extração é similar ao descrito acima, porém, no caso da sazonalidade a série é decimada conforme um parâmetro T que indica o tamanho do lag temporal utilizado.

As fórmulas utilizadas na extração da sazonalidade foram as seguintes:

$$\hat{y} = \begin{cases} X_{RESTO(i/T)}, & \text{se } i \leq 2T \\ LR(1, X^*), & \text{se } i > 2T \end{cases} \quad (4.4)$$

$$\hat{y} = \begin{cases} X_{RESTO(i/T)}, & \text{se } i \leq 2T \\ MA(1, X^*), & \text{se } i > 2T \end{cases} \quad (4.5)$$

Onde X^* é o conjunto composto por todas as amostras em um determinado instante onde $RESTO(i/T) = 0$.

Em 4.4 utiliza-se a regressão linear dos pontos decimados para tentar prever a sazonalidade, já em 4.5 utiliza-se a média móvel. Em ambos os casos utilizam-se todos os pontos fornecidos ao filtro como base para a previsão, eliminando o problema de selecionar o parâmetro W . Todavia, ainda existe o problema de selecionar o valor correto do parâmetro T para que se obtenha o menor erro possível na extração da componente sazonal.

4.2.3 Ciclos Senoidais e Resíduo

O ciclo senoidal foi obtido através da seleção do maior pico na transformada de Fourier do sinal. Para a extração do mesmo foi utilizado o filtro Notch. O resultado dessa operação é um sinal composto somente pela componente residual. Esta é a

parte não determinística da série. Finalmente o sinal é normalizado utilizando a técnica *MinMaxScaler* [54].

4.3 Seleção dos sinais de entrada

A seleção dos sinais de entrada foi feita através do gráfico de correlação dos mesmos. Buscou-se sinais com alta correlação, para que pudesse ser removido um dos pares, visto que isso é um indicativo de que há uma redundância na informação fornecida no modelo. A remoção dos sinais com alta correlação diminui o tempo de treinamento e também pode diminuir o ruído na entrada, fazendo assim com que o treinamento tenha uma precisão maior.

4.3.1 Seleção dos atrasos do sinal de saída

A entrada do modelo é composta por sinais no tempo atual e sinais defasados em períodos de tempo específicos. A seleção das defasagens foi feita através do cálculo da autocorrelação do sinal de saída. Todos os valores com autocorrelação maior ou próximo do limite do intervalo de 95% de confiança foram selecionados. Os valores de defasagem foram replicados para todos os sinais utilizados na entrada do modelo. A função de ativação ser

4.4 Rede Neural

Foi definido que a rede neural utilizada será um MLP de uma camada escondida somente, com ReLU 3.13 como função de ativação. A saída conterá somente um neurônio com função de ativação linear, de modo a conseguir obter o sinal utilizado na previsão do resíduo do PLD.

4.4.1 Treinamento das redes neurais

O treinamento será realizado modificando o número de neurônios na camada intermediária, de forma a avaliar qual arquitetura fornece o menor erro na saída. Para a avaliação dos resultados será utilizado como métrica o RMSE no conjunto de validação pelo número de neurônios. Outro critério para a avaliação utilizado também

foi o gráfico de pontos do dado previsto pelo original de forma similar ao visto em 2.3.

Dado que as entradas foram selecionadas e previamente processadas, restou definir a arquitetura da rede neural utilizada. Devido à simplicidade e a boa capacidade de solução de problemas, foi utilizada a rede Perceptron Multicamadas com somente uma camada intermediária. Foi definido para a camada de entrada e intermediária que a função de ativação utilizada seria a ReLU, dado a velocidade de processamento e os bons resultados obtidos em pesquisas recentes 3.3.

Para a camada de saída, a função de ativação escolhida foi a linear, para que se possa construir a série temporal. O algoritmo de aprendizado utilizado foi o AdaDelta. O α , parâmetro que regula a taxa de aprendizado, foi escolhido de forma a fazer com que o gráfico de erro no dataset de treino pelo número de neurônios convergisse para todos os folds utilizados na validação cruzada.

O modelo foi treinado utilizando validação cruzada. Em cada fold foram feitas algumas inicializações aleatórias utilizando o algoritmo visto em [37], de forma a reduzir o impacto de uma inicialização ruim do modelo.

Para facilitar a seleção dos modelos, foi feito uma tabela contendo informações sobre o erro médio, desvio padrão, coeficiente angular e offset médios para o dataset de validação. Criou-se a métrica para medir o desvio do coeficiente angular (a) e do offset(b) de cada modelo, fazendo com que um primeiro erro fosse obtido através da fórmula $\epsilon_1 = ABS(1 - a)$ e um segundo fosse obtido por $\epsilon_2 = ABS(b)$, onde ABS é o valor absoluto do número. Além disso, para fins de ranqueamento, criou-se uma terceira métrica descrita como $\epsilon_3 = NORM(\epsilon_1) + NORM(\epsilon_2)$, onde $NORM$ é a normalização feita dividindo todos os valores pela norma infinita do vetor. O modelo a ser utilizado é então o que tiver o menor erro ϵ_3 .

Na primeira fase do treinamento, fez-se uma rede que tinha como entrada $x(t)$ e obtém-se como saída $y(t)$, de modo a verificar se o processo utilizado no treinamento e seleção da rede estava adequado, além de servir como caso base para os modelos de

previsão. Dado que as previsões são somente variações desse modelo básico, torna-se de extrema importância a boa adesão do modelo ao problema.

A saída utilizada no processamento foi o sinal residual do PLD, dado que a tendência, sazonalidade e ciclos sazonais são determinísticos. Sendo assim, para obter valor final da PLD é necessário combinar a parte prevista pela rede com a parte determinística.

Da segunda fase em diante foram feitas previsões do resíduo do PLD mensal de fato. Na segunda fase em específico, previu-se o resíduo PLD para o mês seguinte. Além disso, propôs-se uma abordagem com o treinamento de uma segunda rede neural para corrigir o erro entre a previsão feita pela primeira rede e o sinal original.

Nos passos seguintes somente foi feita a comparação entre o resíduo original e o previsto para N meses a frente. Ao final pode-se obter um gráfico com o erro e desvio padrão pela quantidade de meses a frente.

Capítulo 5

Resultados

Neste capítulo, são mostrados os resultados obtidos através da aplicação do método descrito no capítulo anterior. Serão detalhados as características dos sinais utilizados assim como os valores dos parâmetros aplicados em cada algoritmo. Alguns resultados obtidos serão expostos no apêndice para facilitar a leitura.

5.1 Aquisição dos dados

Os dados utilizados e suas respectivas fontes podem ser observados na tabela 4.1. No início do trabalho definiu-se que o foco seria a região sudeste, sendo assim, todos os dados utilizados foram filtrados para a região SE/CO. Os dados provenientes da ONS foram baixados direto do site no formato *Comma Separated Values - CSV*.

Para facilitar o processamento, foram retiradas todas as colunas que não fossem a data no formato ”mês (por extenso) ano”. Essas foram então renomeadas para ”month” e ”value” e o valor do dado referido (ex: MWh produzidos por usinas hidroelétricas no mês), facilitando o desenvolvimento do código.

Dado que as regras para o cálculo do PLD variam durante os anos, utilizou-se somente os dados no período entre 01/2015 e 12/2018 de modo a tentar reduzir ao máximo esse efeito. Sendo assim, o *dataset* possui 48 somente pontos.

5.2 Pré-Processamento

Antes de poder utilizar os sinais obtidos no treinamento da rede, fez-se necessário selecionar quais sinais seriam realmente fornecidos como entrada para o modelo e quantos atrasos temporais seriam necessários para determinar a saída. Além disso há a necessidade de extrair a componente residual do PLD e normalizar os dados. Cada um desses tópicos será detalhado nos parágrafos a seguir.

5.2.1 Seleção dos sinais de entrada

Os sinais de entrada foram selecionados conforme a seguinte matriz de autocorrelação abaixo:

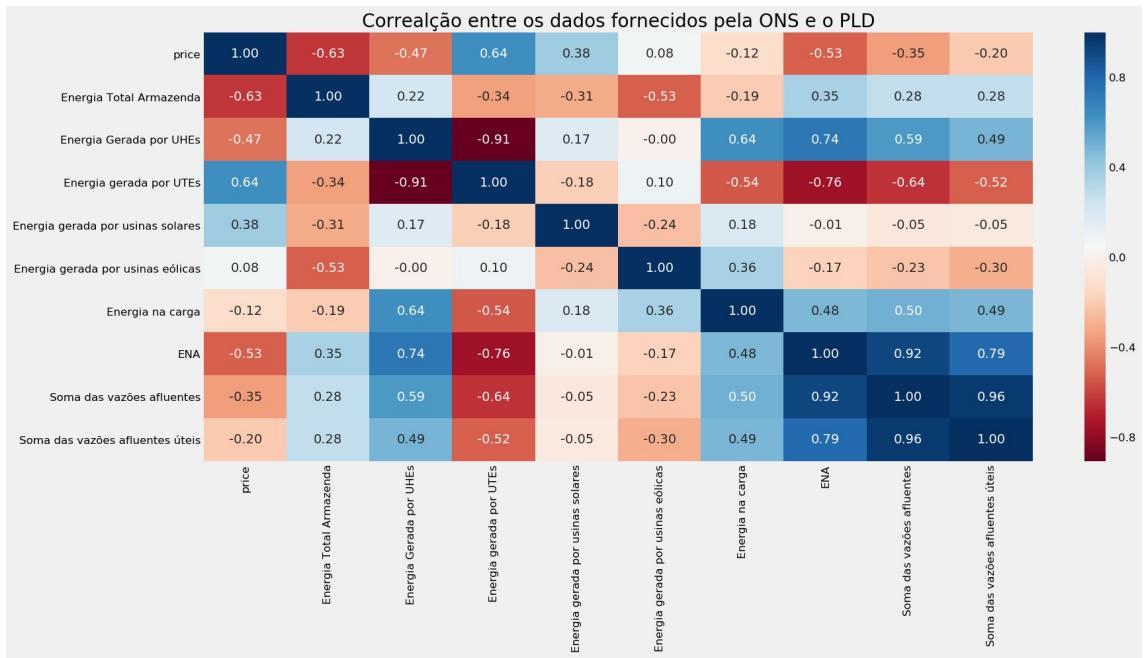


Figura 5.1: Correlação dos dados de entrada.

Observou-se alta correlação negativa entre a energia produzida pelas hidroelétricas e térmicas. Isso ocorre pelo fato da energia gerada pelas termoelétricas ser a principal substituta para a energia proveniente de usinas térmicas. Sendo assim, decidiu-se então remover a série da energia gerada pelas UTEs de forma a remover a redundância dos dados fornecidos.

5.2.2 Seleção dos atrasos do sinal de saída

O gráfico de autocorrelação expressa de forma numérica uma relação linear entre o valor atual e os atrasos. Com isso, têm-se os atrasos com maior importância na formação do sinal. Assim como descrito no capítulo anterior, foram selecionados os pontos que tinham correlação maior ou próximo do limite do intervalo de 95% de confiança.

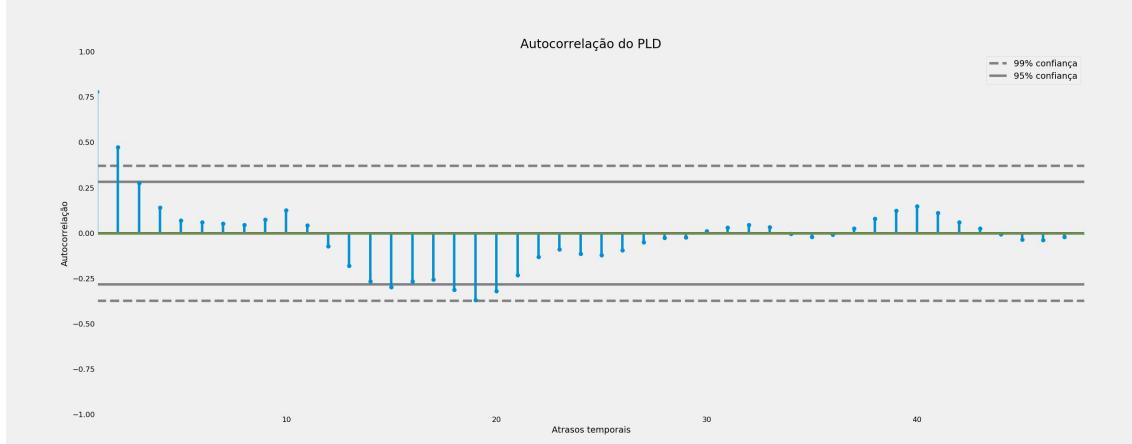


Figura 5.2: Autocorrelação do PLD.

Os atrasos obtidos através dessas análises foram replicados para os sinais da entrada, podendo assim definir quantos atrasos seriam necessários para obter informação suficiente para reconstruir a saída. Como conclui-se pelo gráfico que até 20 atrasos temporais são necessários para representar o sinal, o *dataset* de 48 pontos foi reduzido a 28 pontos.

5.2.3 Tendência

Para a definição do parâmetro W , responsável pelo tamanho da janela utilizada no processamento, fez-se um gráfico do MSE pelo tamanho da janela conforme visto na figura a seguir:

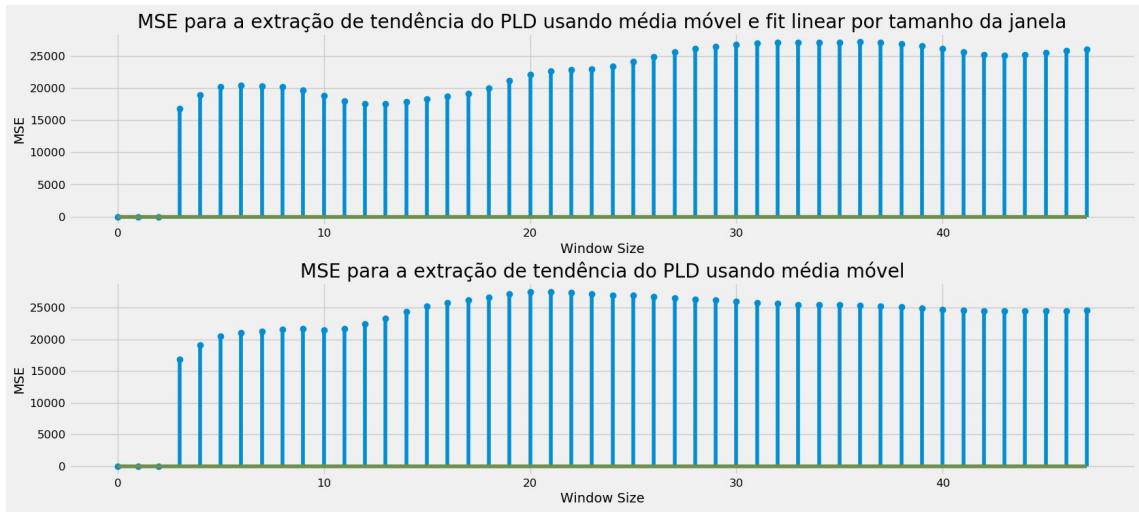


Figura 5.3: Análise do mse pelo tamanho da janela na extração da tendência.

Tendo como base o gráfico acima, observou-se que a extração proposta em 4.1 obteve os melhores resultados e selecionou-se $W = 12$. O parâmetro K foi definido como $K = 5$ de forma empírica, de forma a obter uma transição suave entre a regressão linear e a média móvel. Aplicando esse filtro obteve-se o seguinte resultado para a extração da tendência:

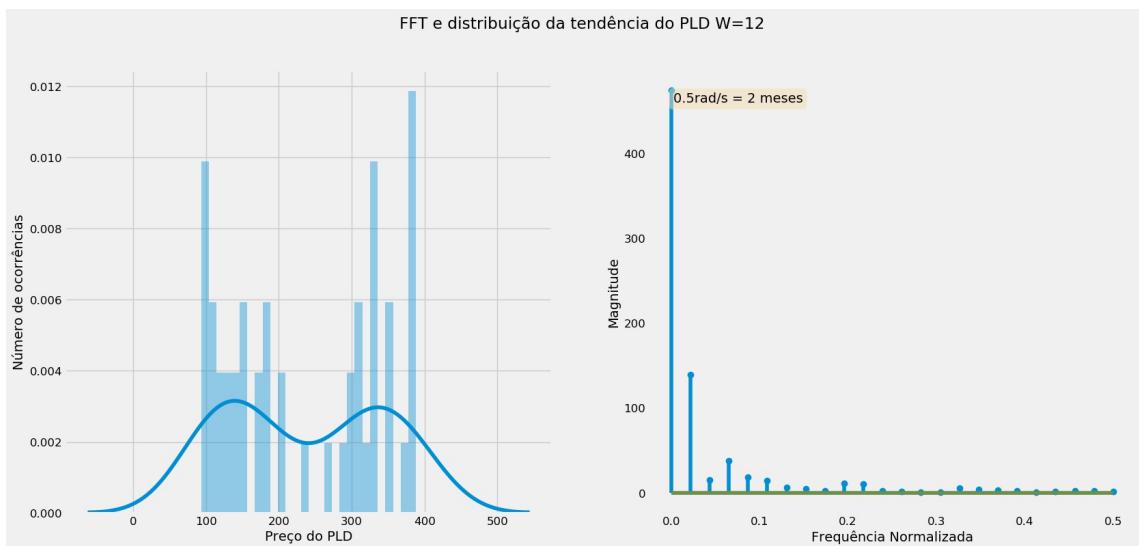


Figura 5.4: Histograma e FFT da tendência obtida.

5.2.4 Sazonalidade

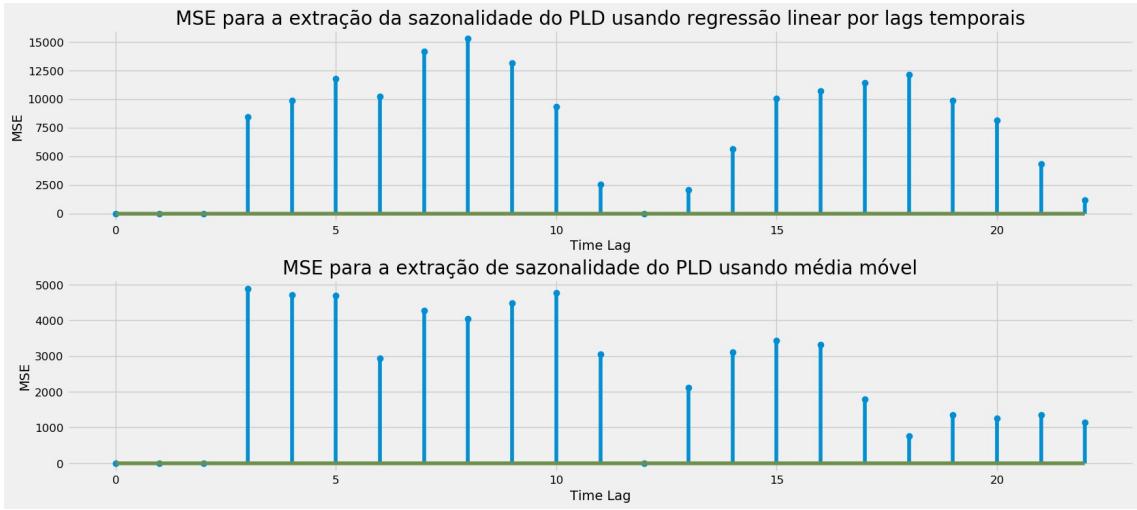


Figura 5.5: Análise do mse pelo tamanho da janela na extração da sazonalidade.

Nesse caso observou-se que existe um mínimo local em $T = 6$ e um mínimo global em $T = 12$. Escolheu se o $T = 6$ por ser um múltiplo comum dos dois. Entre as duas análises, a que trouxe melhores resultados foi a que utilizou somente a média móvel. Finalmente obteve-se como resultado da extração os seguintes gráficos:

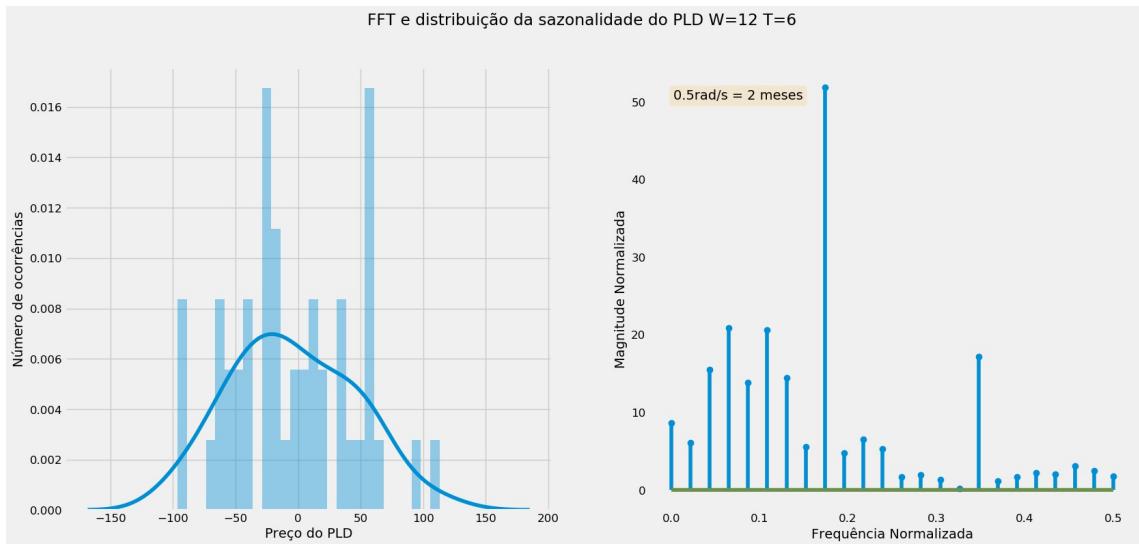


Figura 5.6: Histograma e FFT da sazonalidade obtida.

5.2.5 Ciclos Senoidais e Resíduo

Após o processamento descrito acima, obtém sinal $s_{2t} = cs_t + res_t$, o qual tem a seguinte distribuição e FFT.

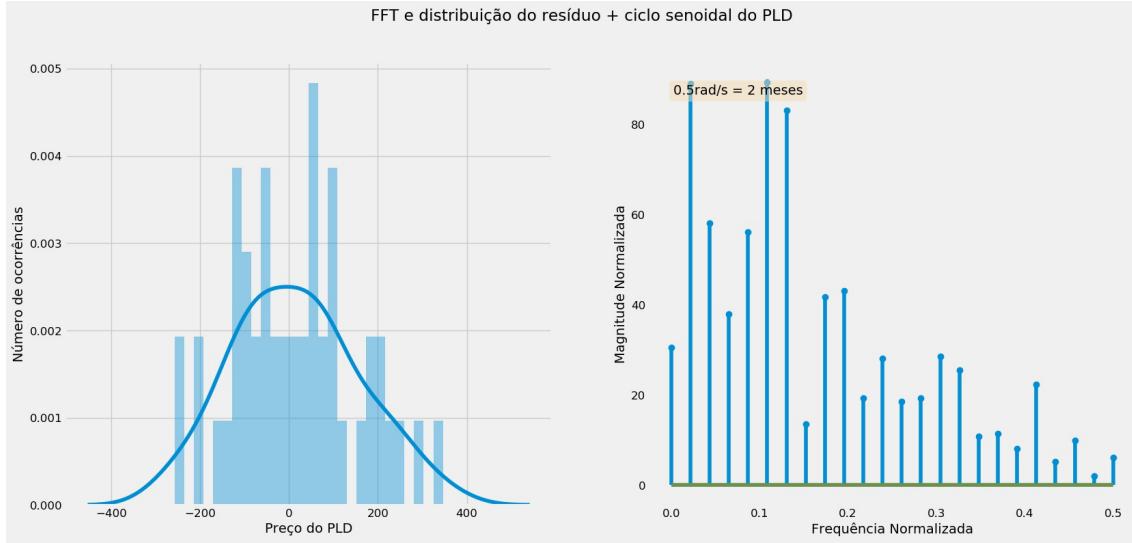


Figura 5.7: Análise do mse pelo tamanho da janela na extração da sazonalidade.

Para a extração do resíduo de fato utilizou-se um filtro notch com frequência $w = 0.10869$ rad/amostra e $Q = 0.01$ removendo então a frequência com maior energia na transformada de Fourier. Obtendo a seguinte distribuição e FFT:

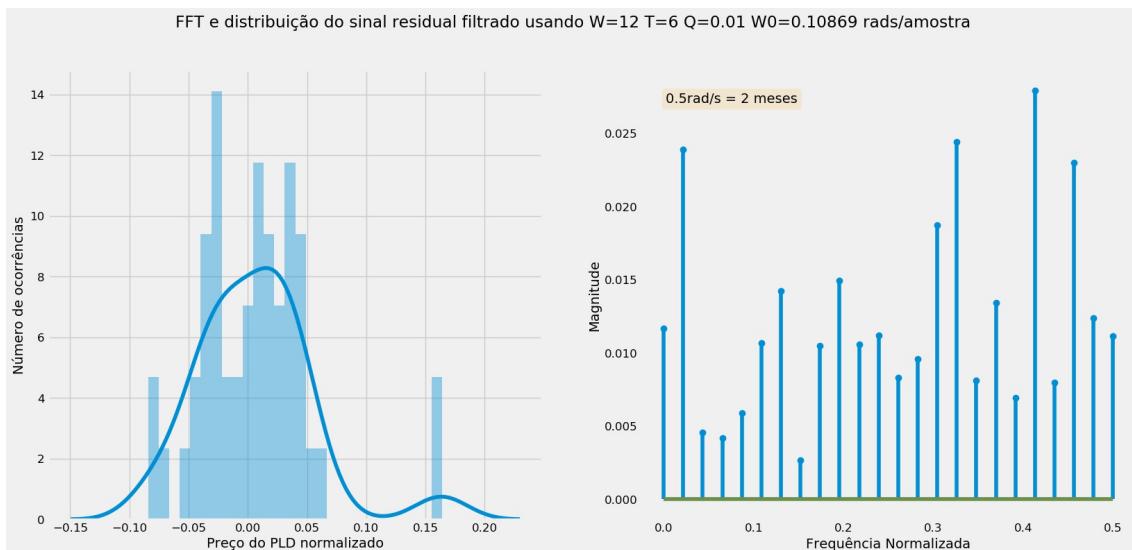


Figura 5.8: Distribuição e FFT após extração da componente de maior energia.

Ao final do processo, os seguintes resultados foram obtidos:

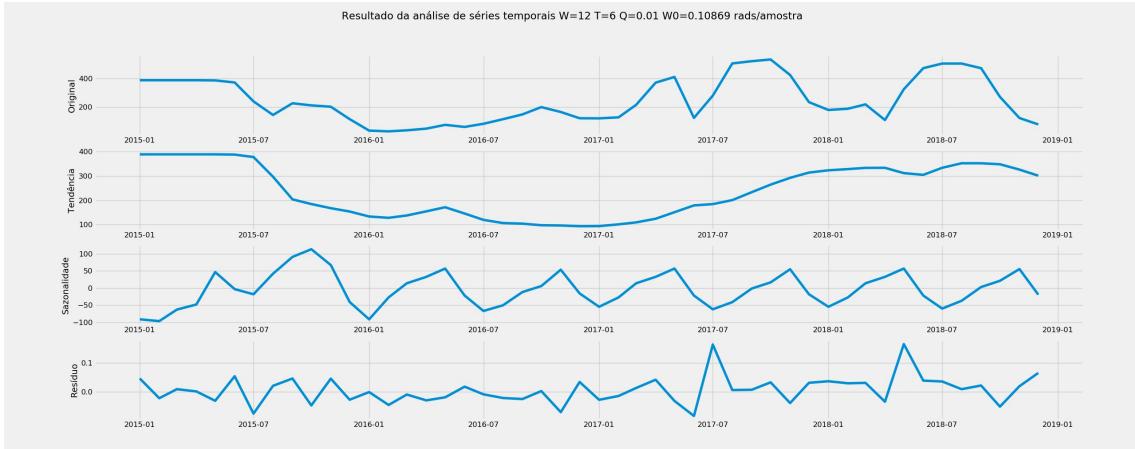


Figura 5.9: Resultado Final com as componentes separadas utilizando o método descrito acima.

5.3 Treinamento das redes neurais

O treinamento foi feito utilizando validação cruzada com 8 folds no dataset de treinamento. Foram deixados 3 pontos da série temporal para realizar o teste. Para cada fold foram feitas 3 inicializações aleatórias utilizando o método chamado *he uniform*, o qual está descrito em [37]. Os 3 últimos foram separados para teste e os outros 25 foram utilizados na validação cruzada. Essa será feita com valores entre 3 e 8 folds. Após a extração residual, os dados são normalizados para facilitar a convergência do treinamento, assim como mostrado em 3.3.3.

O algoritmo de aprendizado utilizado foi o adadelta com learning rate de $\alpha = 0.01$ [55] por conta dos bons resultados obtidos no treinamento e rápida convergência para o estado final. Foi utilizado o critério de parada antecipada *early stop* para que caso a rede não melhorasse o erro obtido em 25 épocas, o processamento seria então interrompido.

Apesar das definições acima, ainda há necessidade de definir a quantidade de neurônios na camada intermediária. Foi feito então um treinamento variando o número de neurônios entre 1 e 90 e, a partir disso, gerou-se um gráfico com o RMSE pelo número de neurônios no conjunto de validação e de teste.

5.3.1 Treinamento da rede para obter a saída no mês atual

Para o treinamento rede para a saída atual, obteve-se o seguinte gráfico de rmse pelo número de neurônios no conjunto de treinamento:

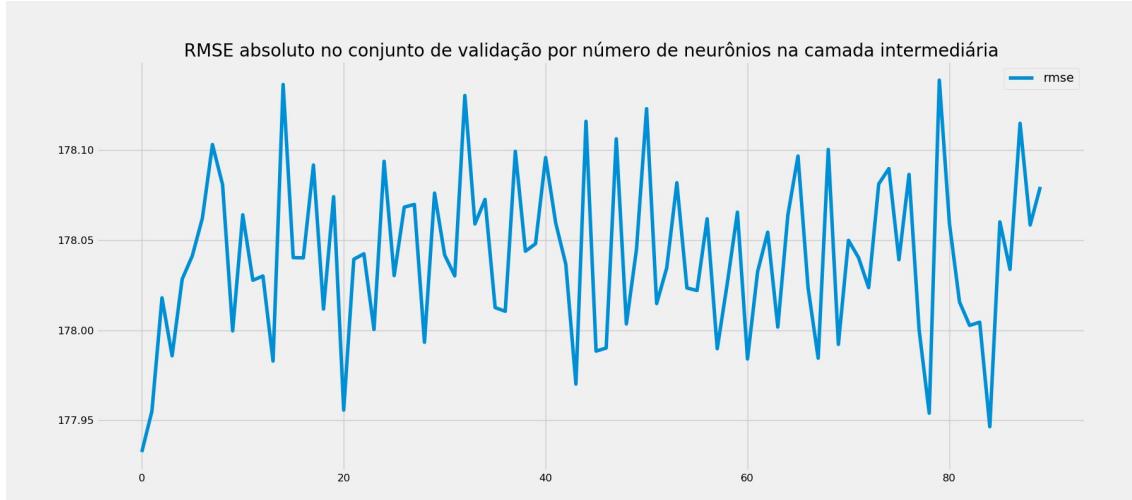


Figura 5.10: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de testes o resultado foi o seguinte:

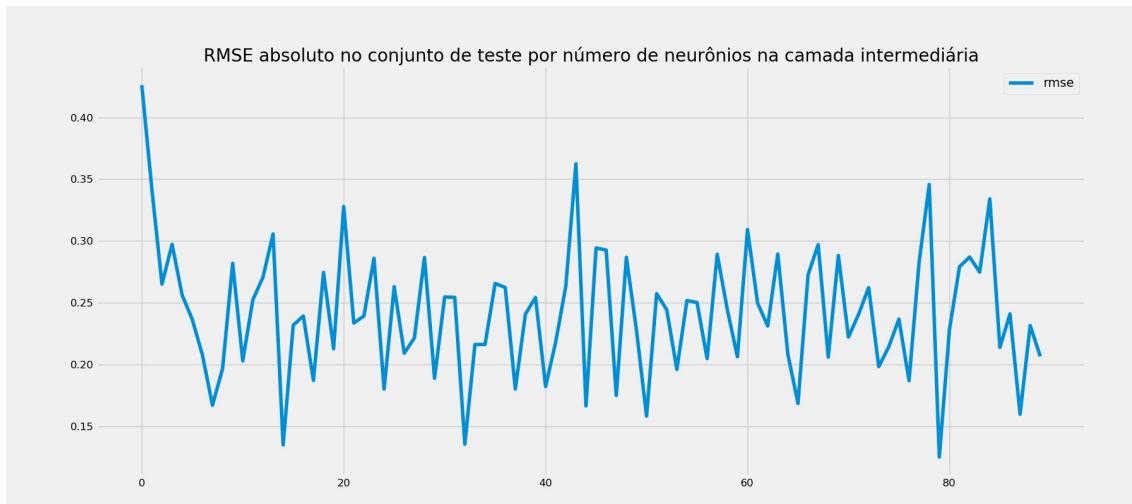


Figura 5.11: RMSE pelo número de neurônios no conjunto de teste.

As 5 melhores estruturas de redes ranqueadas pelo critério ϵ_3 definido anteriormente são as seguintes:

Tabela 5.1: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 61 | 64,435 | 61,180 | 0,998 | 0,008 | -0,007 | 0,175 | 0,184 |
| 38 | 66,896 | 63,894 | 1,005 | 0,019 | -0,003 | 0,175 | 0,194 |
| 97 | 63,509 | 61,957 | 1,000 | 0,001 | 0,046 | 0,224 | 0,225 |
| 77 | 56,788 | 54,954 | 1,003 | 0,010 | 0,044 | 0,224 | 0,234 |
| 92 | 55,370 | 53,135 | 0,983 | 0,063 | 0,009 | 0,224 | 0,287 |

E com isso o modelo com 61 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas e por folds:

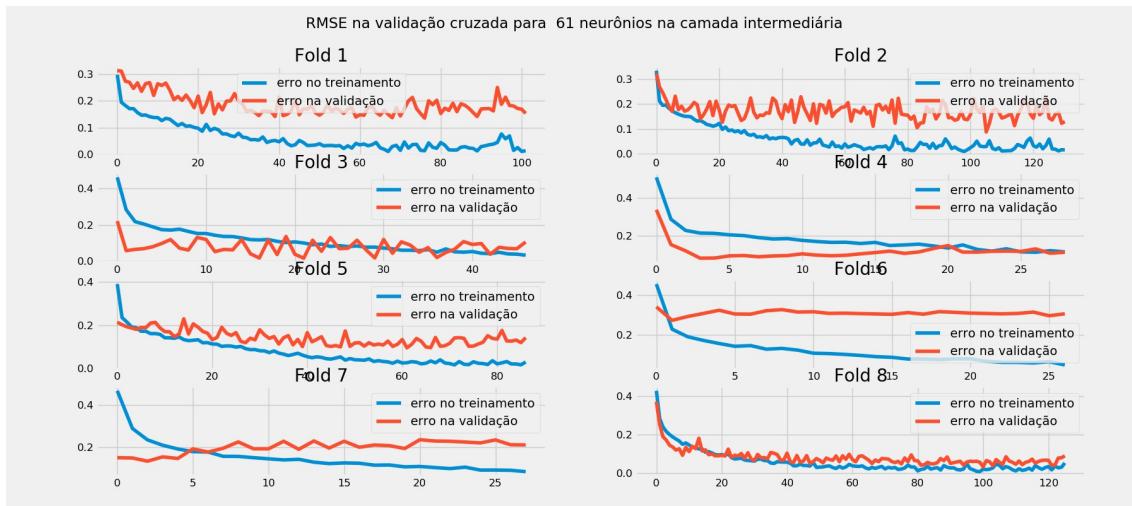


Figura 5.12: Erro pelo número de época e por fold.

E os seguintes resultados foram obtidos:

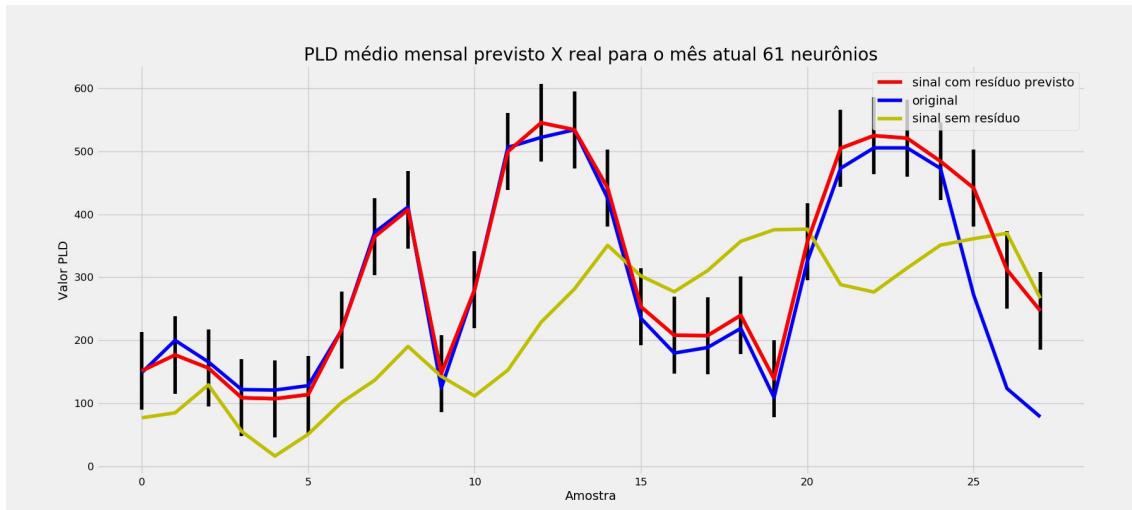


Figura 5.13: Sinal completo.

Os 3 últimos pontos são do conjunto de dados de teste. Observa-se que a rede não conseguiu prevê-los tão bem quanto os pontos do conjunto de treinamento, assim como esperado.

Dado a pouca quantidade de dados, a rede não obteve o comportamento esperado nos treinamentos, porém o resultado obtido mostrou que a mesma conseguiu mapear bem a entrada na saída. Sendo assim, viu-se que a rede está sendo treinada de forma coerente para obter $y(t)$ a partir de $x(t)$. Esse caso também será utilizado como base para as previsões, pois espera-se que o erro para N passos a frente seja sempre maior ou igual ao caso base.

5.3.2 Treinamento da rede para obter a saída um mês a frente

Seguindo o procedimento previamente determinado, obteve-se os seguintes resultados para a previsão do PLD para 1 mês a frente:

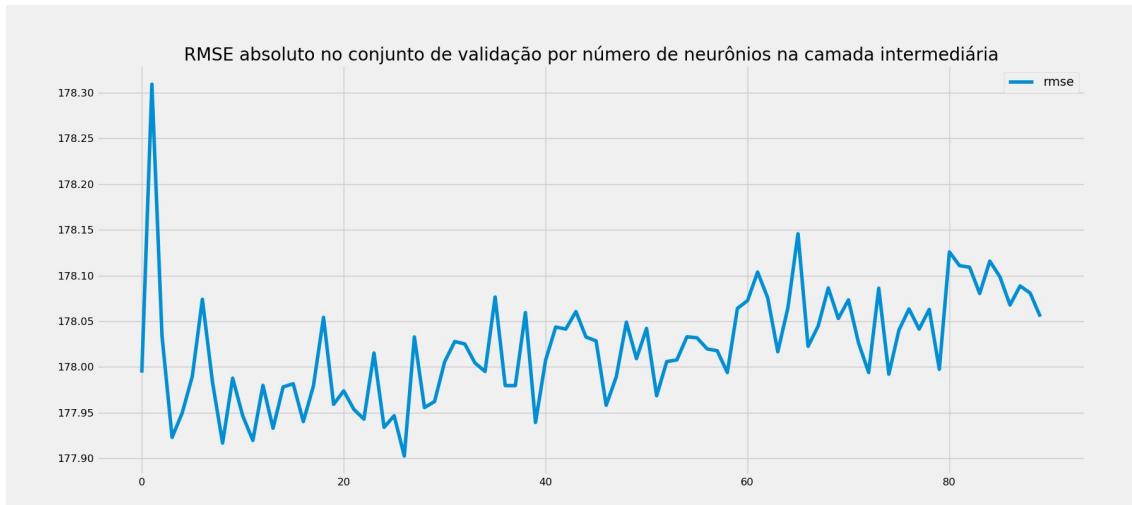


Figura 5.14: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

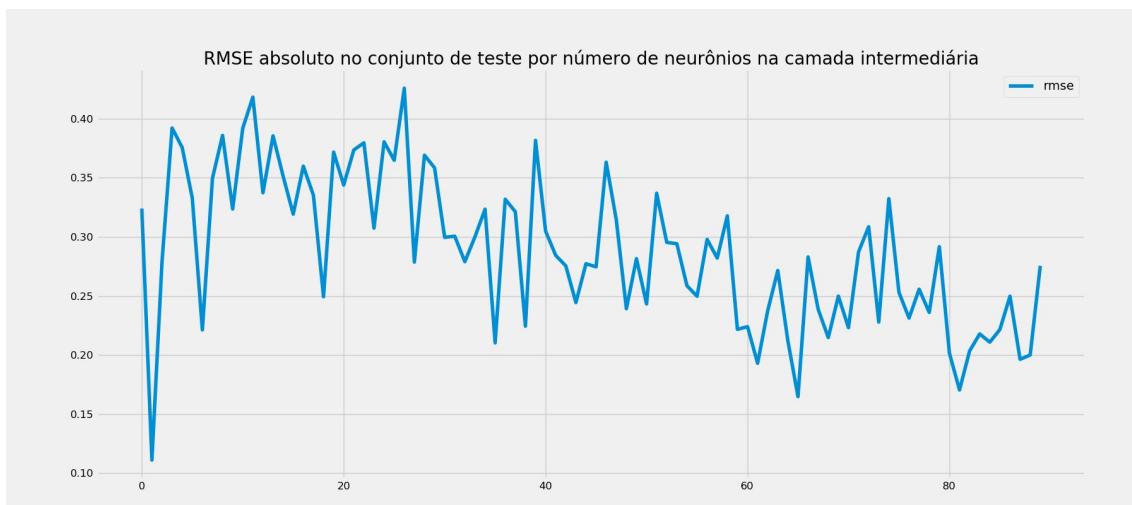


Figura 5.15: RMSE pelo número de neurônios no conjunto de teste.

Tabela 5.2: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|---------|--------|-------------|--------------|-------------|--------------|--------------|
| 63 | 82,911 | 69,521 | 0,993 | 0,002 | 0,005 | 0,003 | 0,004 |
| 71 | 88,327 | 74,076 | 0,985 | 0,004 | -0,027 | 0,014 | 0,018 |
| 25 | 93,041 | 73,840 | 0,966 | 0,008 | 0,020 | 0,010 | 0,018 |
| 50 | 103,137 | 93,552 | 0,927 | 0,018 | 0,032 | 0,016 | 0,034 |
| 51 | 94,937 | 81,015 | 0,860 | 0,034 | 0,000 | 0,000 | 0,034 |

E com isso o modelo com 63 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

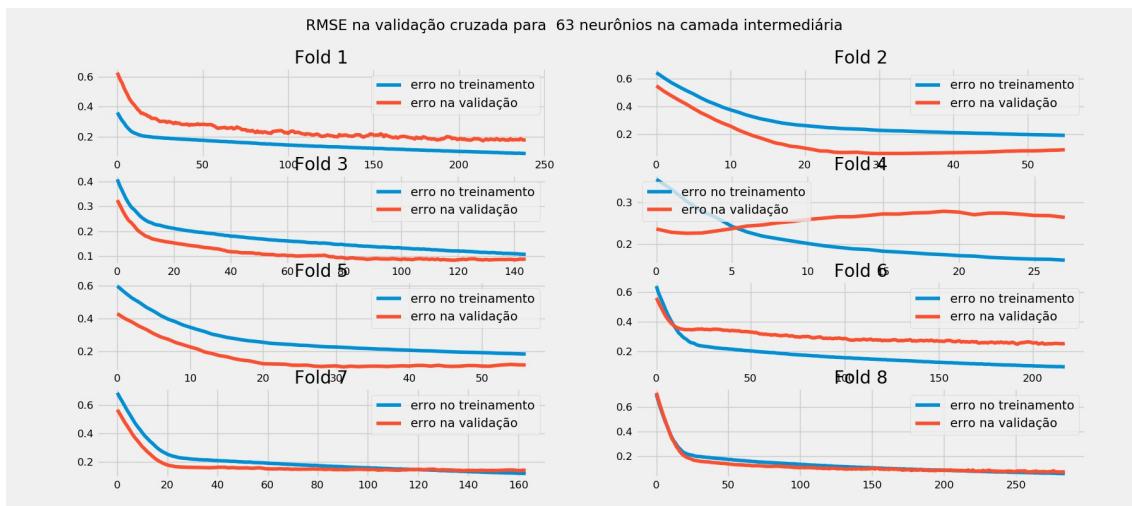


Figura 5.16: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

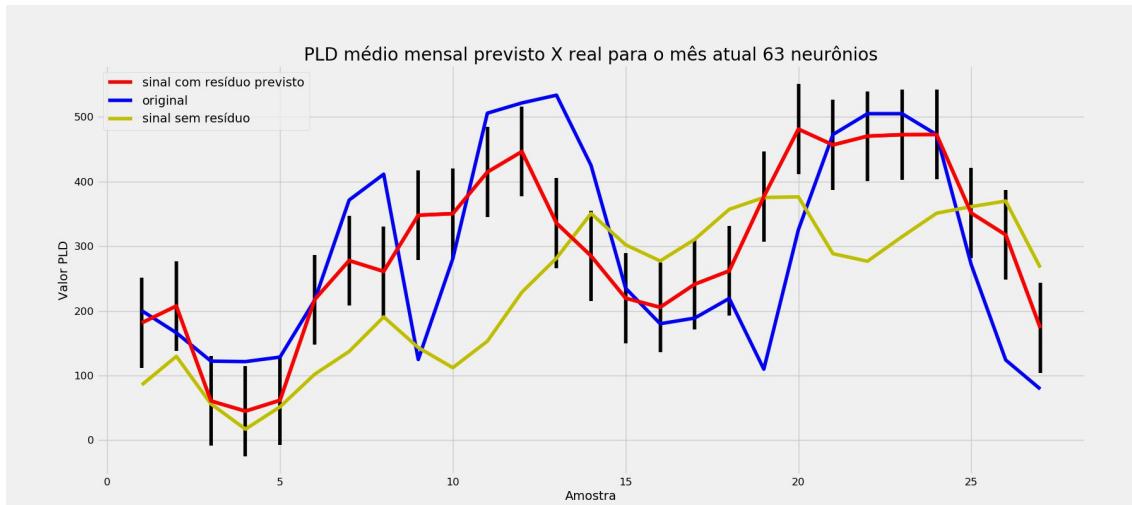


Figura 5.17: Sinal completo.

Para essa previsão, foi treinada uma rede para reduzir o erro de estimativa usando como entrada os dados já mencionados anteriormente. Os erros de estimativa entre o sinal reconstruído e o original foram utilizados como alvos da rede. Obteve-se então os seguintes resultados:

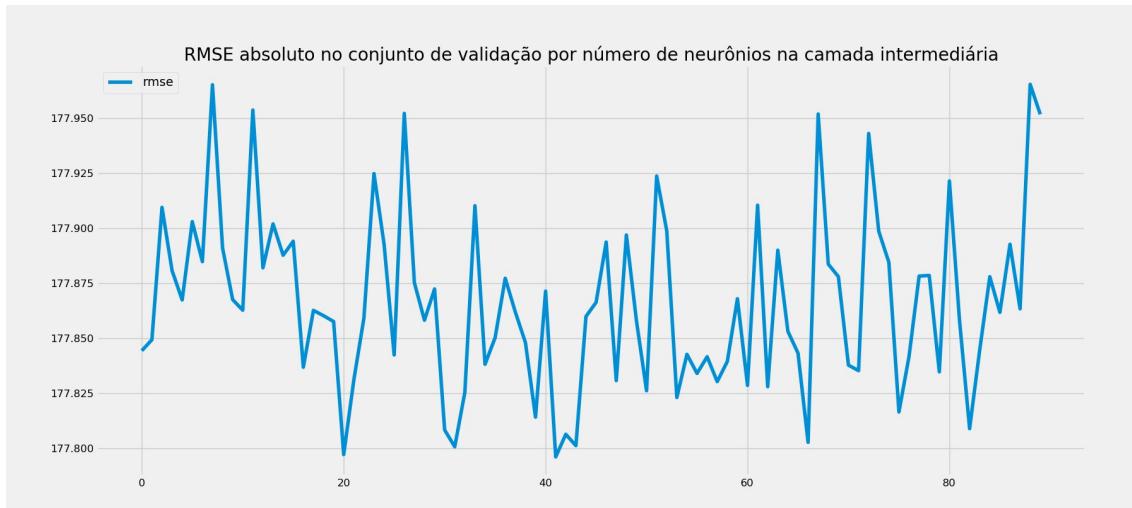


Figura 5.18: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

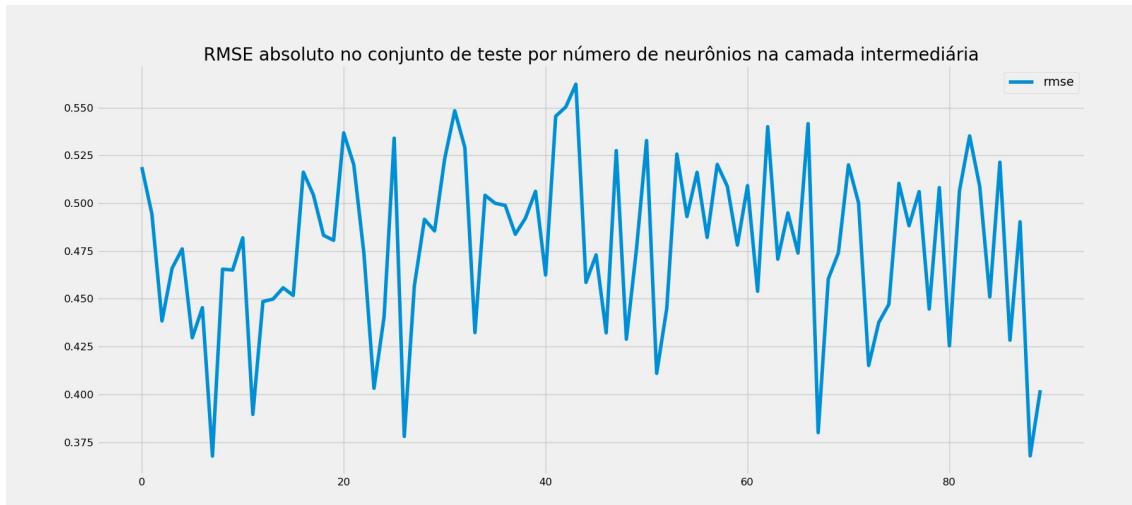


Figura 5.19: RMSE pelo número de neurônios no conjunto de teste.

Tabela 5.3: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|---------|---------|-------------|--------------|-------------|--------------|--------------|
| 90 | 126,678 | 101,233 | 1,029 | 0,001 | -0,001 | 0,000 | 0,001 |
| 14 | 130,352 | 98,738 | 0,973 | 0,001 | -0,002 | 0,000 | 0,001 |
| 46 | 137,971 | 99,784 | 1,017 | 0,001 | -0,017 | 0,001 | 0,002 |
| 76 | 136,014 | 105,032 | 0,982 | 0,001 | 0,057 | 0,003 | 0,004 |
| 60 | 123,420 | 98,844 | 1,070 | 0,002 | -0,031 | 0,002 | 0,004 |

E com isso o modelo com 90 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

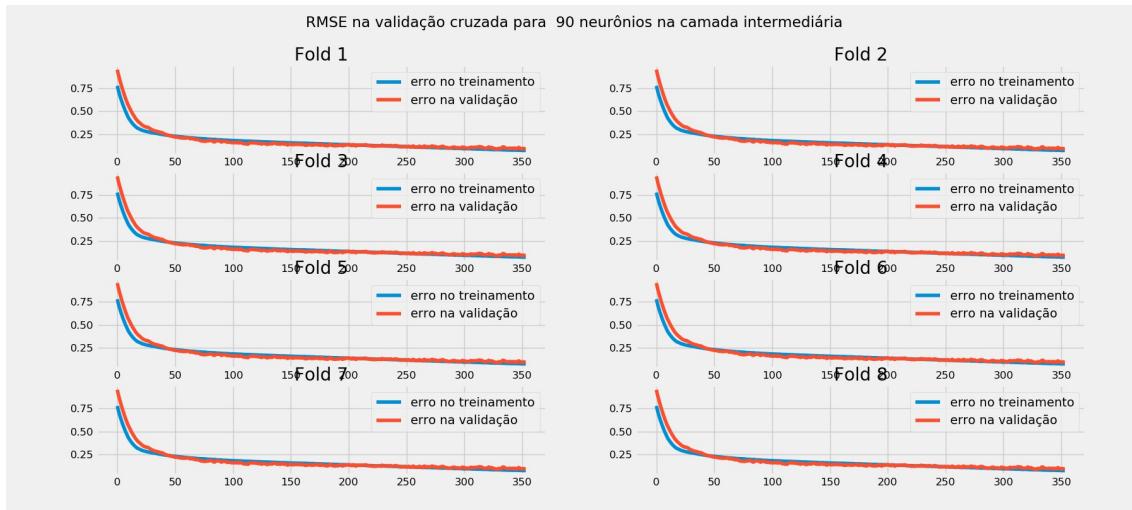


Figura 5.20: Erro pelo número de épocas e por fold.

E finalmente obteve-se os seguintes resultados:

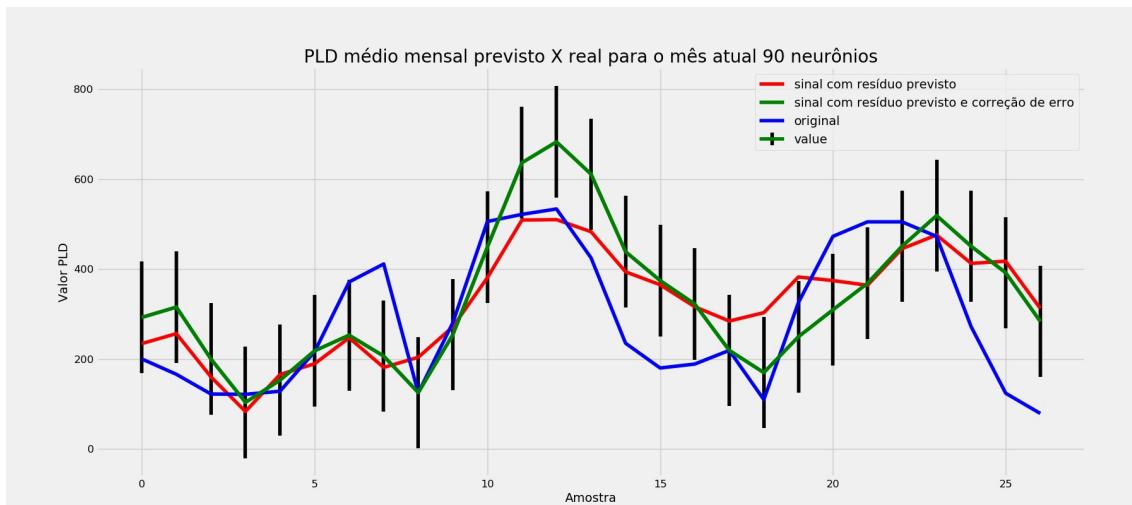


Figura 5.21: Sinal completo.

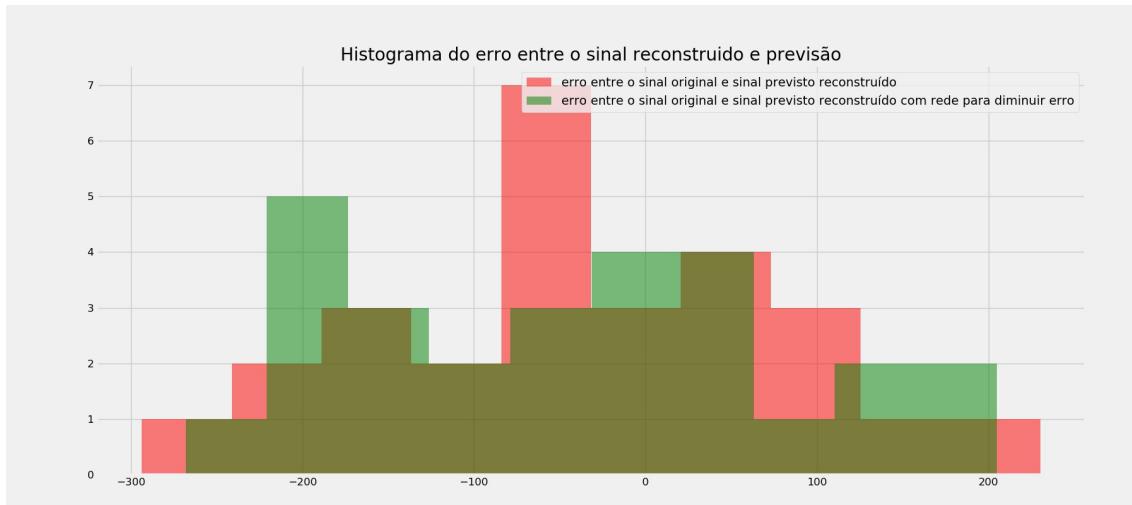


Figura 5.22: Histograma de previsão sem e com a correção de erro.

Pelo histograma é possível ver que a rede de correção do erro de estimativa diminuiu o erro máximo e mínimo, porém a estrutura sem a correção do erro tem maior concentração de pontos próximos do zero. Sendo assim, espera-se que a saída com a correção de erro erre mais, porém com discrepância pequena, enquanto espera-se que sem a correção de erro o resultado seja melhor na maioria das vezes, podendo correr o risco de errar muito em alguns pontos.

Cabe a quem for utilizar analisar se compensa mais errar menos na média, porém ter uma chance de errar muito em casos extremos ou errar um pouco mais na média e ter o erro dentro de uma faixa menor.

5.3.3 Treinamento da rede para obter a saída para vários meses a frente

A rede para vários meses à frente foi treinada de forma similar ao observado anteriormente, porém nesse caso observou-se somente a parte residual e não foi treinada uma rede secundária para corrigir o erro de estimativa em cada caso.

A partir dos resultados, gerou-se o seguinte gráfico que mostra o erro da rede em relação ao sinal original para diferentes números de passos a frente:

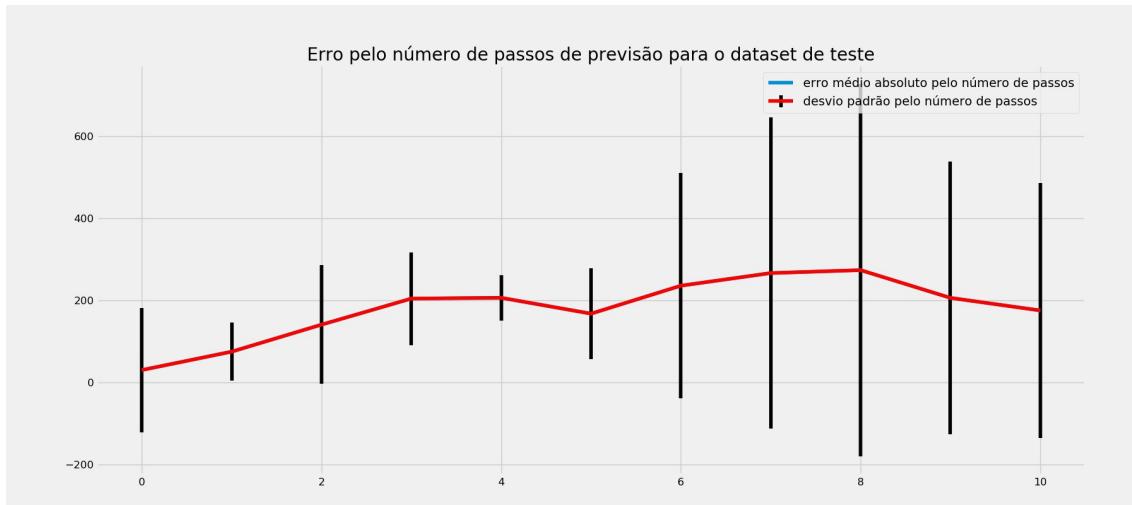


Figura 5.23: Erro no sinal residual pelo número de passos no conjunto de dados de teste.

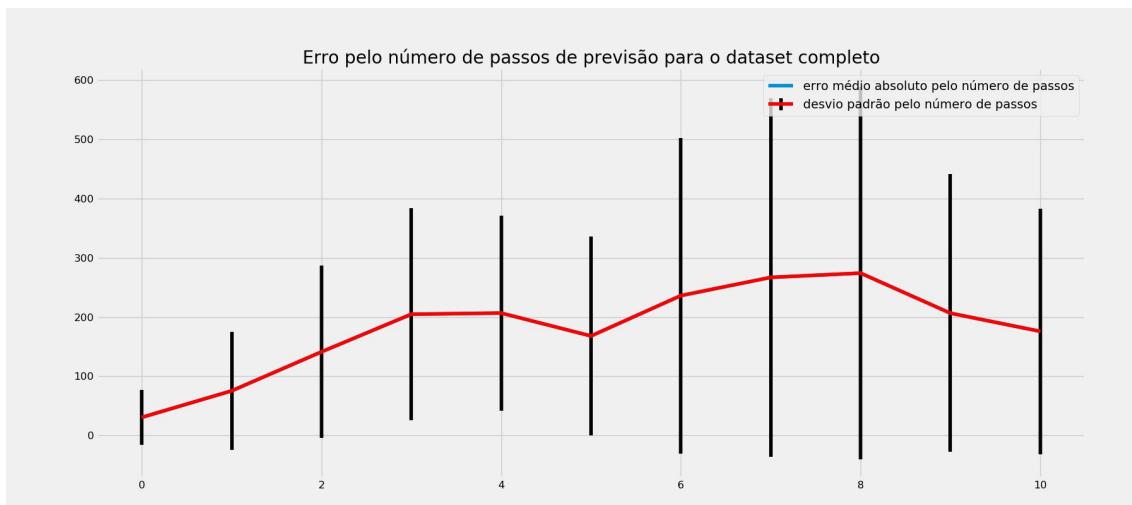


Figura 5.24: Erro no sinal residual pelo número de passos no conjunto de dados de completo.

A tabela com os dados utilizados na seleção de cada modelo, o gráfico de convergência e os resultados obtidos estão presente nos apêndice A.

Assim como o esperado, tanto o erro quanto o desvio padrão aumentam com o número de passos a frente, salvo algumas exceções. Sendo assim, a previsão para muitos passo a frente traz um risco maior para a análise. Cabe a quem for usar mensurar quanto de risco pode ser tolerado baseado nesses gráficos.

Ainda assim cabe ressaltar que a análise foi feita somente para a parte residual. Provavelmente ao reconstruir o PLD, os erros ainda devem aumentar, pois serão somados os erros de estimação da tendência, sazonalidade e ciclos senoidais. A estratégia da rede para a correção dos erros pode ser aplicada para cada um dos casos buscando deixar a distribuição do erro mais concentrada em torno do zero.

Capítulo 6

Conclusões

Neste trabalho foi proposto um modelo baseado em redes neurais, cujo o objetivo era prever o valor médio do PLD para alguns meses a frente. Para isso, foi necessária a decomposição do sinal, deixando para o treinamento da rede somente a parte não-determinística. Foram testados diferentes métodos de extração das componentes até chegar na que foi utilizada no capítulo 5. Além disso, as MLPs foram treinadas para quantidades variadas de neurônios na camada intermediária, buscando a melhor arquitetura que solucionasse o problema proposto.

Diante dos resultados obtidos com o trabalho, observou-se a dificuldade em utilizar redes neurais MLP para conjuntos de dados pequenos. Apesar disso, atingiu-se os objetivos mencionados no começo do projeto.

O método utilizado para o pré-processamento dos dados pode ser replicado para outros tipos de séries temporais, assim como o método para a seleção da melhor rede pode ser explorado em outros trabalhos.

Quanto às conclusões obtidas através do experimento, observou-se que o erro tem uma tendência crescente conforme o número de meses a frente na previsão, assim como esperado, tornando cada vez mais complexa a utilização do modelo para cenário muito distantes. A aplicação da rede para correção dos erros trouxe para o a previsão um compromisso entre erro máximo e erro médio, conforme a arquitetura escolhida.

Para trabalhos futuros, seria interessante comparar os resultados obtidos com modeos diferentes, como por exemplo, SARIMA e GARCH e ADALASSO. Além disso, podem ser exploradas outras formas de extração das componentes da série temporal, de forma a obter um sinal residual com menos energia dos que os que foram encontrados.

Outra abordagem pode ser a de considerar os dados para os anos antes de 2015 e analisar quanto isso afeta a previsão do PLD médio mensal. Nesse caso, cabe também utilizar estruturas de redes neurais mais complexas, como por exemplo LSTM - *Long Short-Term Memory*.

Referências Bibliográficas

- [1] (1988), N., “Monthly New York City births: unknown scale. Jan 1946 ? Dec 1959”, <https://datamarket.com/data/set/22nv/monthly-new-york-city-births-unknown-scale-jan-1946-dec-1959#!ds=22nv&display=line>, 1988, (Acesso em 9 de Abril 2019).
- [2] EPE, “Empresa de Pesquisa Energética”, <http://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>, 2018, (Acesso em 26 Janeiro 2019).
- [3] ONS, “Operador Nacional do Sistema Elétrico”, <http://ons.org.br/>, 2019, (Acesso em 27 Janeiro 2019).
- [4] CCEE, “Câmara de Comercialização de Energia Elétrica”, <https://www.ccee.org.br/>, 2019, (Acesso em 31 Janeiro 2019).
- [5] SILVA, E. F. D., *Principais Condicionantes das alterações no modelo de comercialização*. M.Sc. dissertation, Escola Politécnica da Universidade de São Paulo, 2008.
- [6] SILVA, B. G. D., *Evolução do setor elétrico brasileiro no contexto econômico nacional: Uma análise histórica e econométrica de longo prazo*. Ph.D. dissertation, Universidade de São Paulo - Programa de pós-graduação em energia EP-FEA-IEE-IF, 2011.
- [7] CASTRO, N. J. D., LEITE, A. L. D. S., “Preço spot de eletricidade: teoria e evidências do caso brasileiro”, *IV Encontro de Enconomia Catarinense, 2010, Criciúma, ,* 2010.

- [8] CAVALIERE, M. A., “Previsão de preços futuros de energia eletrônica no ambiente de contratação livre - uma abordagem de equilíbrio de mercado sob incertezas”, 2017.
- [9] FILHO, J. C. R., *Previsão multi-passo a frente do preço de energia elétrica de curto prazo no mercado brasileiro*. Ph.D. dissertation, Universidade Federal do Pará - Instituto de Tecnologia, programa de pós-graduação em engenharia elétrica, 2014.
- [10] BARRETO, E. P. B. M., “Smart grid: Eficiência energética e a geração distribuída a partir das redes inteligentes”, 2018.
- [11] ANEEL, “Resolução normativa nº 482, de 17 de abrid de 2012”, <http://www2.aneel.gov.br/cedoc/ren2012482.pdf>, 2019, (Acesso em 26 Fevereiro 2019).
- [12] UFJF, E. I., “Especial: A crise energética brasileira”, <http://energiainteligenteufjf.com/especial/especial-a-crise-energetica-brasileira/>, 2019, (Acesso em 26 Fevereiro 2019).
- [13] CEPEL, “Documentação Técnica das metodologias e modelos de otimização energética do CEPEL”, <http://www.cepel.br/sala-de-imprensa/noticias/menu/noticias/documentacao-tecnica-das-metodologias-e-modelos-de-otimizacao-energetica-do-cepel.htm>, 2019, (Acesso em 26 Fevereiro 2019).
- [14] SILVA, A. P. S., “Previsão do preço de liquidação das diferenças por meio de redes neurais artificiais”, 2018.
- [15] JUNIOR, F. R. P., *Redes neurais diretas e recorrentes na previsão do preço de energia elétrica de curto prazo no mercado brasileiro*. M.Sc. dissertation, Universidade Federal do Pará, 2016.
- [16] RODRIGUES, A. L., *Redes neurais artificiais aplicadas na previsão de preços do mercado spot de energia elétrica*. M.Sc. dissertation, Universidade de São Paulo, 2009.

- [17] NASCIMENTO, F. B. D., “Redes Neurais Artificiais Aplicadas à Predição do Preço De Liquidação das Diferenças no Mercado De Energia”, 2017.
- [18] SOUSA, A. J. S. D., *Análise e Previsão da Volatilidade do Preço de Liquidação das Diferenças no Mercado Brasileiro Utilizando o Modelo GARCH*. M.Sc. dissertation, Universidade Federal da Bahia, 2013.
- [19] Martinez Alvarez, F., Troncoso, A., Riquelme, J. C., *et al.*, “Energy Time Series Forecasting Based on Pattern Sequence Similarity”, *IEEE Transactions on Knowledge and Data Engineering*, v. 23, n. 8, pp. 1230–1243, Aug 2011.
- [20] Rodriguez, C. P., Anders, G. J., “Energy price forecasting in the Ontario competitive power system market”, *IEEE Transactions on Power Systems*, v. 19, n. 1, pp. 366–374, Feb 2004.
- [21] Amjadi, N., Hemmati, M., “Energy price forecasting - problems and proposals for such predictions”, *IEEE Power and Energy Magazine*, v. 4, n. 2, pp. 20–29, March 2006.
- [22] HONG, T., PINSON, P., FAN, S., *et al.*, “Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond”, *International Journal of Forecasting*, v. 32, n. 3, pp. 896 – 913, 2016.
- [23] SHAFIE-KHAH, M., MOGHADDAM, M. P., SHEIKH-EL-ESLAMI, M., “Price forecasting of day-ahead electricity markets using a hybrid forecast method”, *Energy Conversion and Management*, v. 52, n. 5, pp. 2165 – 2169, 2011.
- [24] EHLERS, R. S., *Análise de Séries Temporais*, 2009.
- [25] MAKRIDAKIS, W., HYNDMAN, “Sales of shampoo over a three year period”, <https://datamarket.com/data/set/22r0/sales-of-shampoo-over-a-three-year-period#!ds=22r0&display=line>, 1998, (Acesso em 9 de Abril 2019).
- [26] CALÔBA, L. P., *Introdução ao Uso de Redes Neurais na Modelagem de Sistemas Dinâmicos e Séries Temporais.*, 2002.

- [27] PEREIRA, D. F. R., “Aprendizado de máquina e aprendizado profundo para apoio à decisão no mercado financeiro”, 2018.
- [28] ORFANIDIS, S. J., *Introduction to Signal Processing*, 2002.
- [29] SILVA, I. N. D., SPATTI, D. H., FLAUZINO, R. A., *Redes neurais artificiais para engenharia e ciências aplicadas*. Artliber Editora, 2010.
- [30] GURNEY, K., *An Introduction to Neural Networks*. Bristol, PA, USA, Taylor & Francis, Inc., 1997.
- [31] ESCOLA, B., “O que é neurônio?”, <https://brasilescola.uol.com.br/o-que-e/biologia/o-que-e-neuronio.htm>, 2019.
- [32] VINICIUS, A., “Redes Neurais Artificiais”, <https://medium.com/@avinicius.adorno/redes-neurais-artificiais-418a34ea1a39>, 2017.
- [33] HILDEBRANDT, T. H., DINGLE, A., “Improving C++ Performance Using Temporaries”, *Computer*, v. 29, n. 03, pp. 31–41, mar 1998.
- [34] BISHOP, C. M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg, Springer-Verlag, 2006.
- [35] SHARMA, S., “Redes Neurais Artificiais”, <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>, 2017.
- [36] BHATTARAI, S., “tanh-activation-function”, <https://saugatbhattarai.com.np/what-is-activation-functions-in-neural-network-nn/logistic-sigmoid-unipolar-tanh-bipolar/>, 2018.
- [37] HE, K., ZHANG, X., REN, S., *et al.*, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”, *CoRR*, v. abs/1502.01852, 2015.
- [38] SARKAR, K., “ReLU : Not a Differentiable Function: Why used in Gradient Based Optimization? and Other Generalizations of ReLU.”,

<https://medium.com/@kanchansarkar/relu-not-a-differentiable-function-why-used-in-gradient-based-optimization-7fef3a4cece>, 2018.

- [39] AGARAP, A. F., “Deep Learning using Rectified Linear Units (ReLU)”, *CoRR*, v. abs/1803.08375, 2018.
- [40] CHENG, H.-T., KOC, L., HARMSEN, J., *et al.*, “Wide & Deep Learning for Recommender Systems”. In: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, DLRS 2016, pp. 7–10, New York, NY, USA, 2016.
- [41] TANG, Y., “Deep Learning using Support Vector Machines”, *CoRR*, v. abs/1306.0239, 2013.
- [42] Hecht-Nielsen, “Theory of the backpropagation neural network”. In: *International 1989 Joint Conference on Neural Networks*, pp. 593–605 vol.1, 1989.
- [43] GARDNER, M., DORLING, S., “Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences”, *Atmospheric Environment*, v. 32, n. 14-15, pp. 2627–2636, 1998.
- [44] SCHMIDHUBER, J., “Deep learning in neural networks: An overview”, *Neural Networks*, v. 61, pp. 85 – 117, 2015.
- [45] DENG, L., YU, D., “Deep Learning: Methods and Applications”, *Foundations and Trends® in Signal Processing*, v. 7, n. 3?4, pp. 197–387, 2014.
- [46] HINTON, G., DENG, L., YU, D., *et al.*, “Deep Neural Networks for Acoustic Modeling in Speech Recognition”, *IEEE Signal Processing Magazine*, v. 29, pp. 82–97, November 2012.
- [47] MOREIRA, S., “Rede Neural Perceptron Multicamadas”, <https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9>, 2018.
- [48] YAM, J. Y., CHOW, T. W., “A weight initialization method for improving training speed in feedforward neural network”, *Neurocomputing*, v. 30, n. 1, pp. 219 – 232, 2000.

- [49] HSIAO, T.-C. R., LIN, C.-W., CHIANG, H. K., “Partial least-squares algorithm for weights initialization of backpropagation network”, *Neurocomputing*, v. 50, pp. 237 – 247, 2003.
- [50] Russo, D., Zou, J., “How much does your data exploration overfit? Controlling bias via information usage”, *arXiv e-prints*, p. arXiv:1511.05219, Nov 2015.
- [51] DABBURA, I., “Gradient Descent Algorithm and Its Variants”, <https://towardsdatascience.com/gradient-descent-algorithm-and-its-variants-10f652806a3>, 2017.
- [52] Arlot, S., Celisse, A., “A survey of cross-validation procedures for model selection”, *arXiv e-prints*, p. arXiv:0907.4728, Jul 2009.
- [53] COUTO, E., “Bias vs. Variância (Parte 2)”, <https://ericcoutho.wordpress.com/2013/07/18/bias-vs-variancia-parte-2/>, 2013.
- [54] LEARN, S., “sklearn.preprocessing.MinMaxScaler”, <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>, 2019.
- [55] ZEILER, M. D., “ADADELTA: An Adaptive Learning Rate Method”, *CoRR*, v. abs/1212.5701, 2012.

Apêndice A

Outros resultados

A.1 Previsão para 2 meses a frente

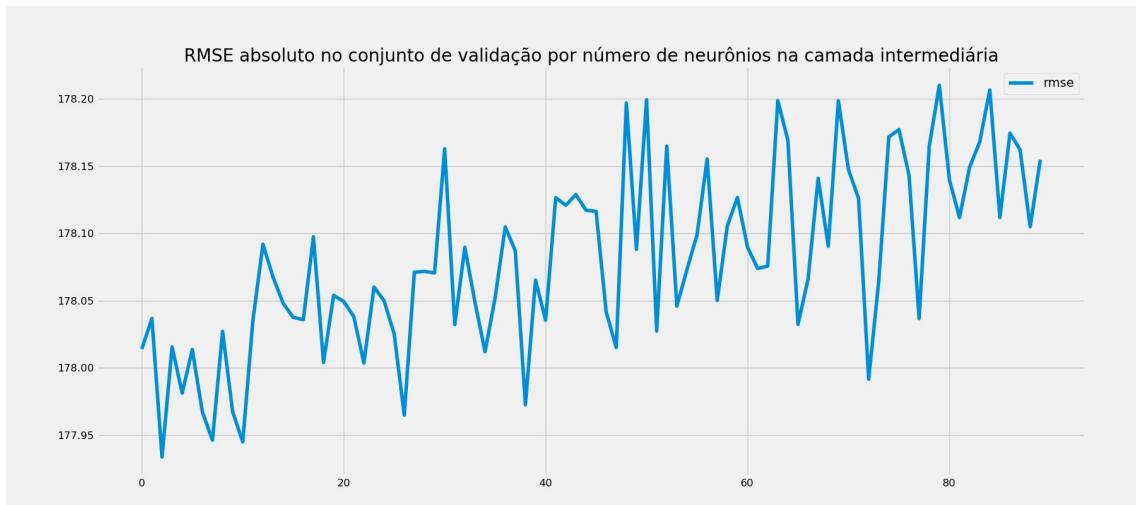


Figura A.1: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

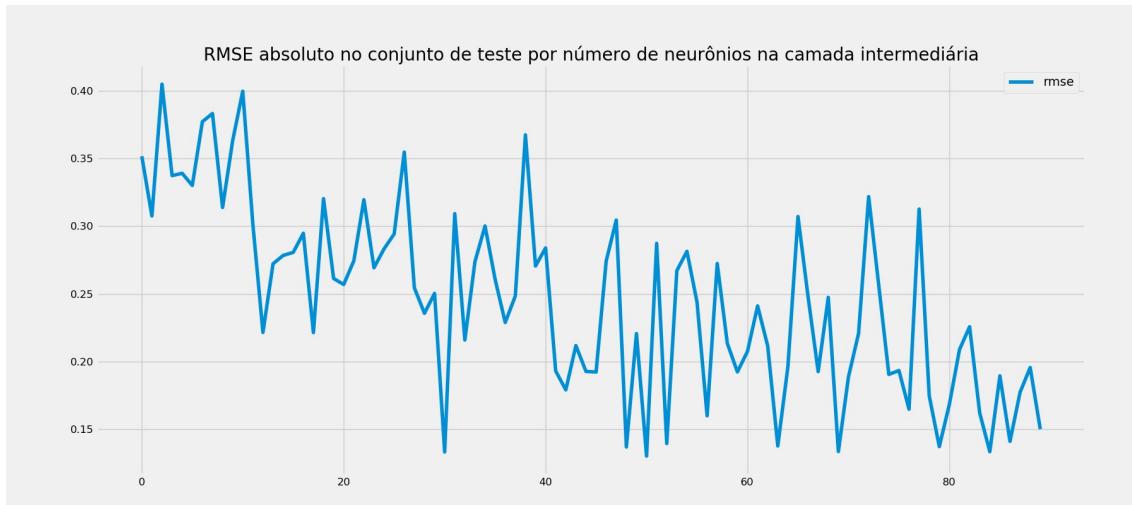


Figura A.2: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.1: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|---------|--------|-------------|--------------|-------------|--------------|--------------|
| 69 | 97,141 | 76,580 | 1,029 | 0,001 | 0,011 | 0,000 | 0,001 |
| 79 | 96,834 | 84,359 | 0,945 | 0,001 | 0,022 | 0,001 | 0,002 |
| 42 | 100,878 | 80,880 | 1,056 | 0,001 | -0,022 | 0,001 | 0,002 |
| 86 | 96,153 | 77,890 | 0,945 | 0,001 | -0,032 | 0,001 | 0,002 |
| 58 | 108,011 | 83,478 | 1,181 | 0,003 | 0,001 | 0,000 | 0,003 |

E com isso o modelo com 69 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

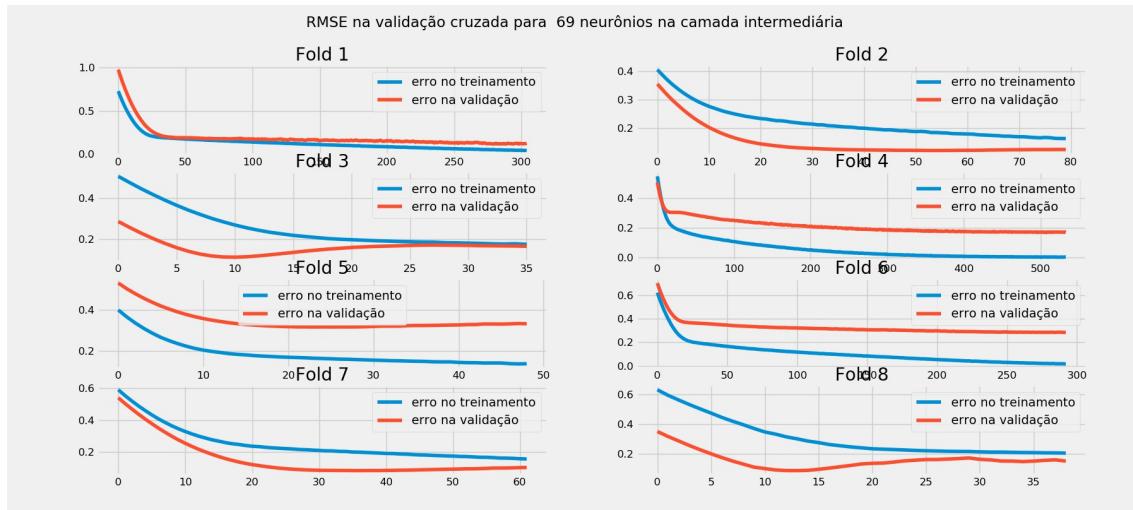


Figura A.3: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

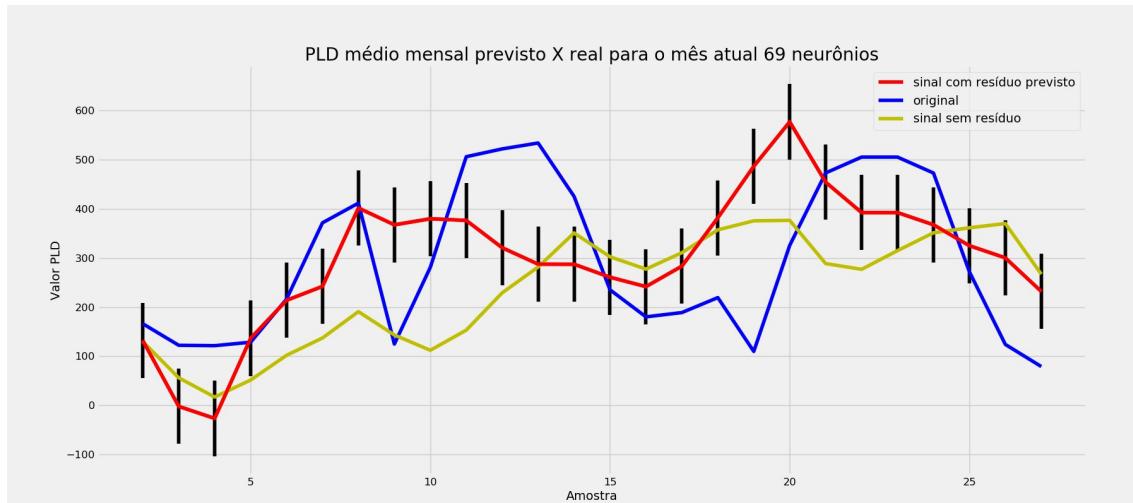


Figura A.4: Sinal completo.

A.2 Previsão para 3 meses a frente

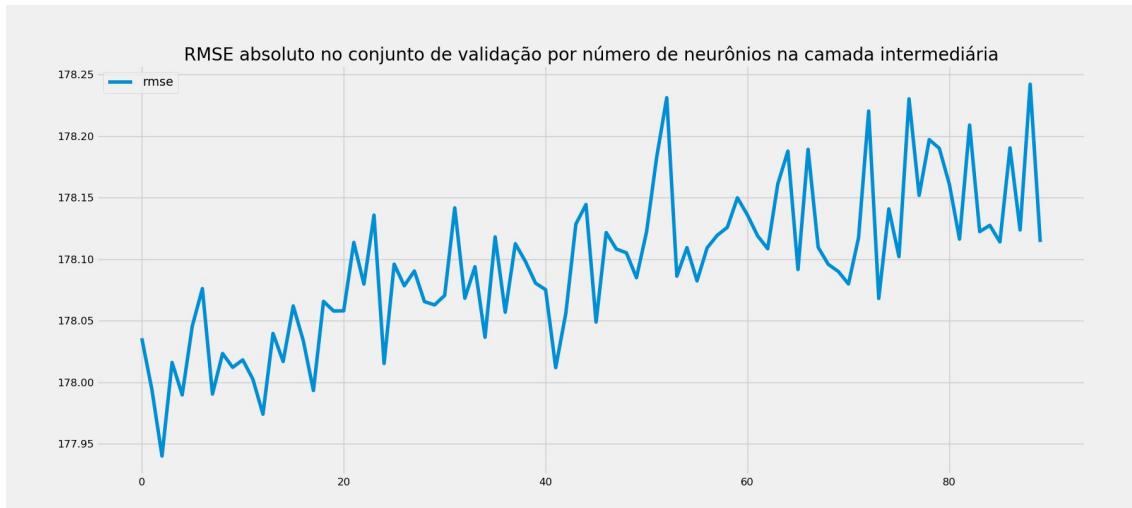


Figura A.5: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

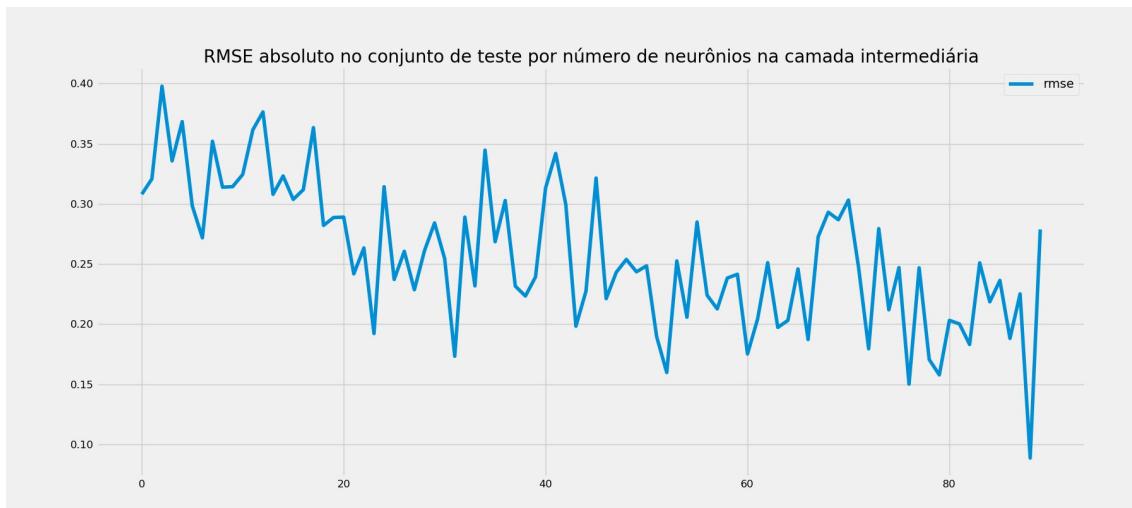


Figura A.6: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.2: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 88 | 75,835 | 69,489 | 1,003 | 0,000 | 0,019 | 0,004 | 0,004 |
| 39 | 79,729 | 69,457 | 0,906 | 0,006 | -0,008 | 0,002 | 0,008 |
| 62 | 78,011 | 70,262 | 0,954 | 0,003 | 0,052 | 0,011 | 0,014 |
| 61 | 88,784 | 75,252 | 0,845 | 0,010 | 0,023 | 0,005 | 0,015 |
| 27 | 72,325 | 64,862 | 0,906 | 0,006 | 0,049 | 0,010 | 0,016 |

E com isso o modelo com 88 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

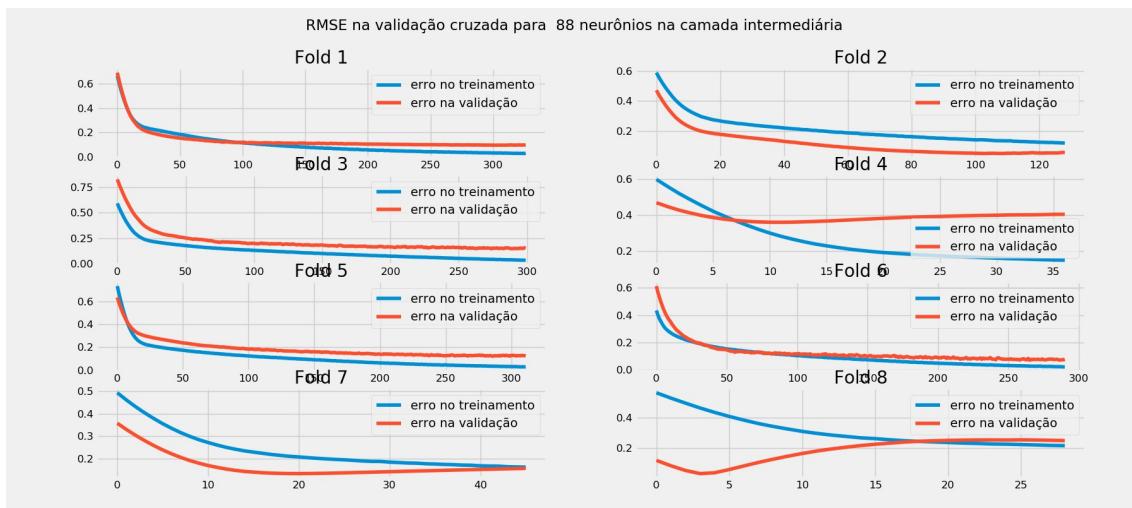


Figura A.7: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

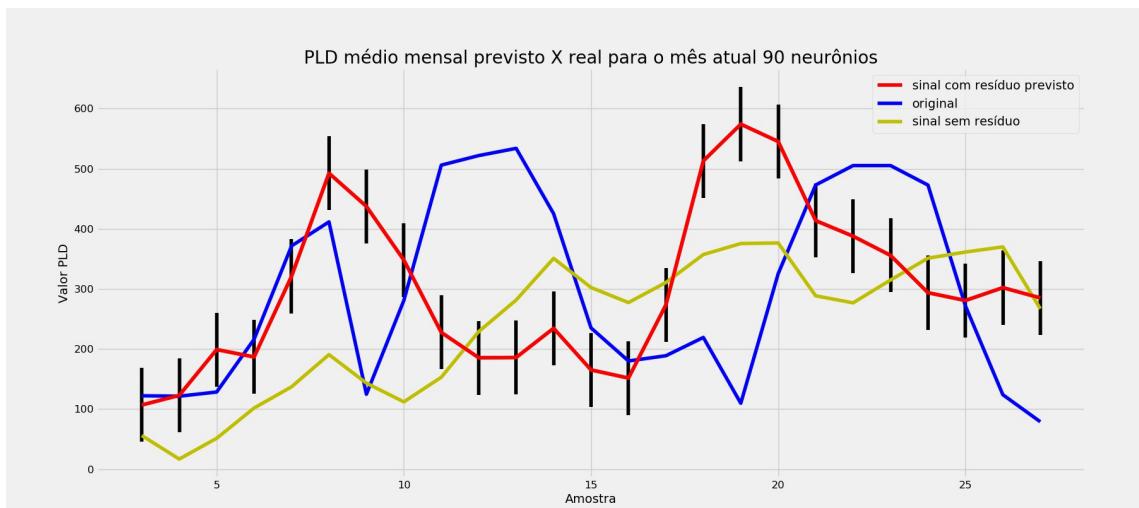


Figura A.8: Sinal completo.

A.3 Previsão para 4 meses a frente

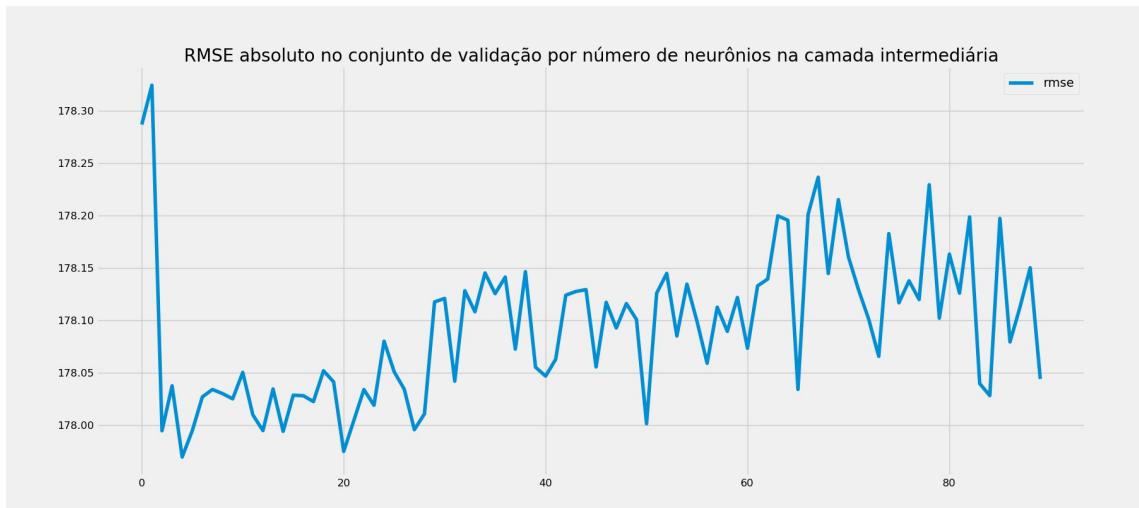


Figura A.9: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

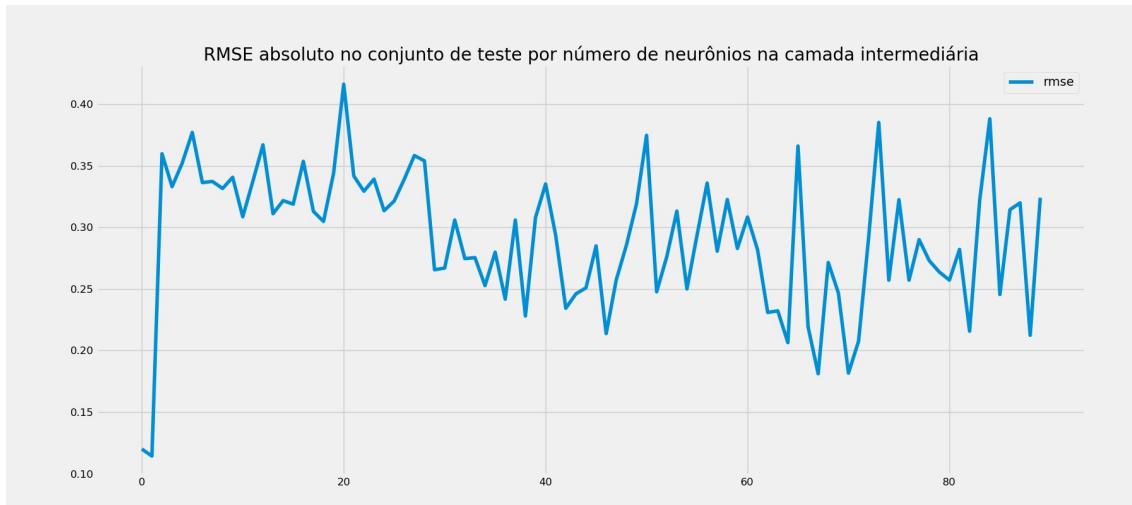


Figura A.10: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.3: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 87 | 66,068 | 49,039 | 1,093 | 0,002 | 0,029 | 0,001 | 0,003 |
| 14 | 84,320 | 61,843 | 0,939 | 0,001 | 0,107 | 0,004 | 0,005 |
| 81 | 92,701 | 75,397 | 1,148 | 0,003 | 0,069 | 0,002 | 0,005 |
| 37 | 79,235 | 56,037 | 0,916 | 0,002 | 0,119 | 0,004 | 0,006 |
| 45 | 93,947 | 62,609 | 1,074 | 0,001 | 0,127 | 0,004 | 0,006 |

E com isso o modelo com 88 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

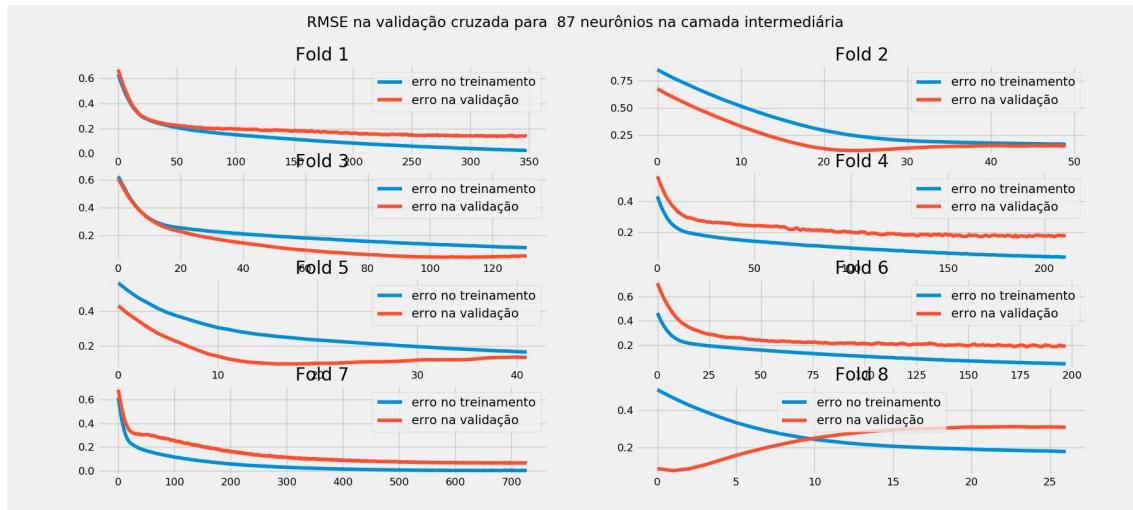


Figura A.11: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

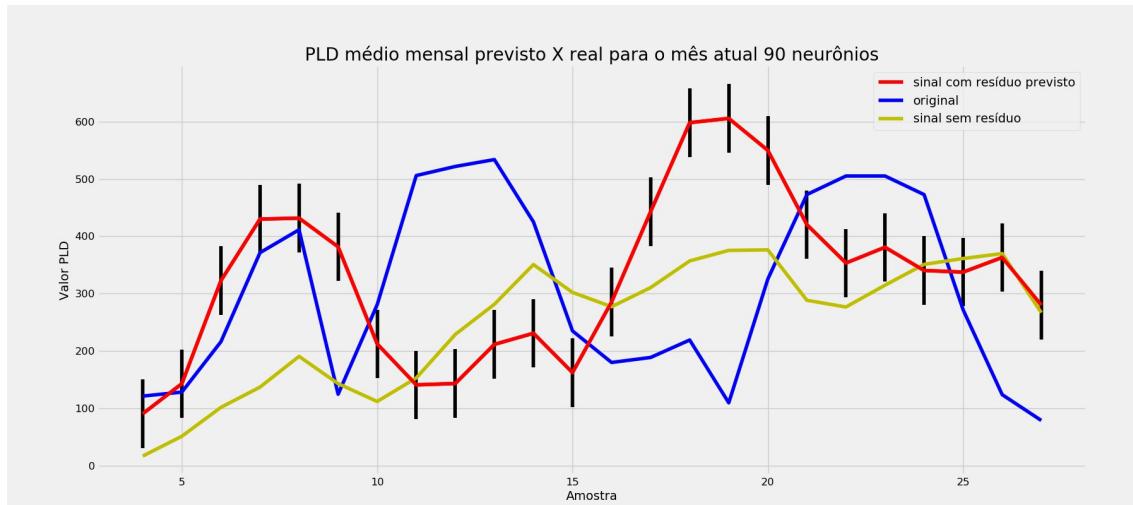


Figura A.12: Sinal completo.

A.4 Previsão para 5 meses a frente

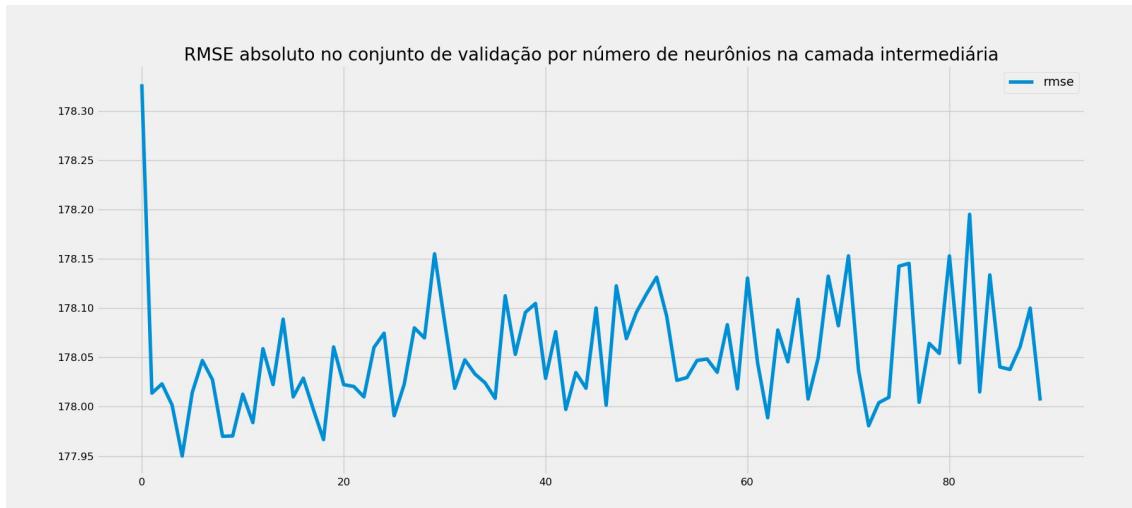


Figura A.13: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

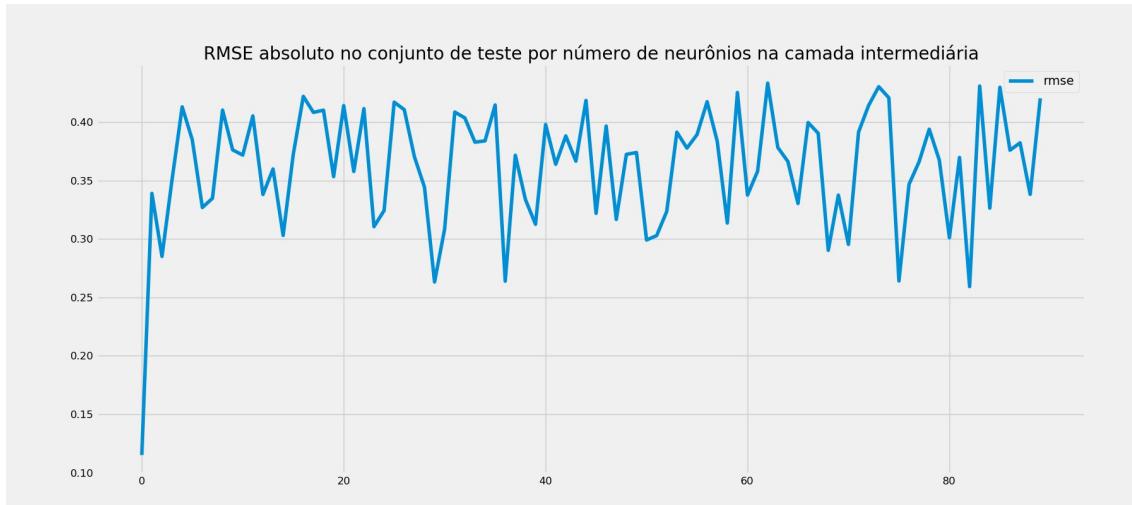


Figura A.14: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.4: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 34 | 63,022 | 54,479 | 0,956 | 0,000 | -0,014 | 0,000 | 0,001 |
| 90 | 72,624 | 68,375 | 1,067 | 0,001 | -0,005 | 0,000 | 0,001 |
| 53 | 62,794 | 56,172 | 0,916 | 0,001 | -0,002 | 0,000 | 0,001 |
| 86 | 62,563 | 57,758 | 1,033 | 0,000 | -0,038 | 0,001 | 0,001 |
| 41 | 66,194 | 62,095 | 0,970 | 0,000 | -0,047 | 0,001 | 0,001 |

E com isso o modelo com 34 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

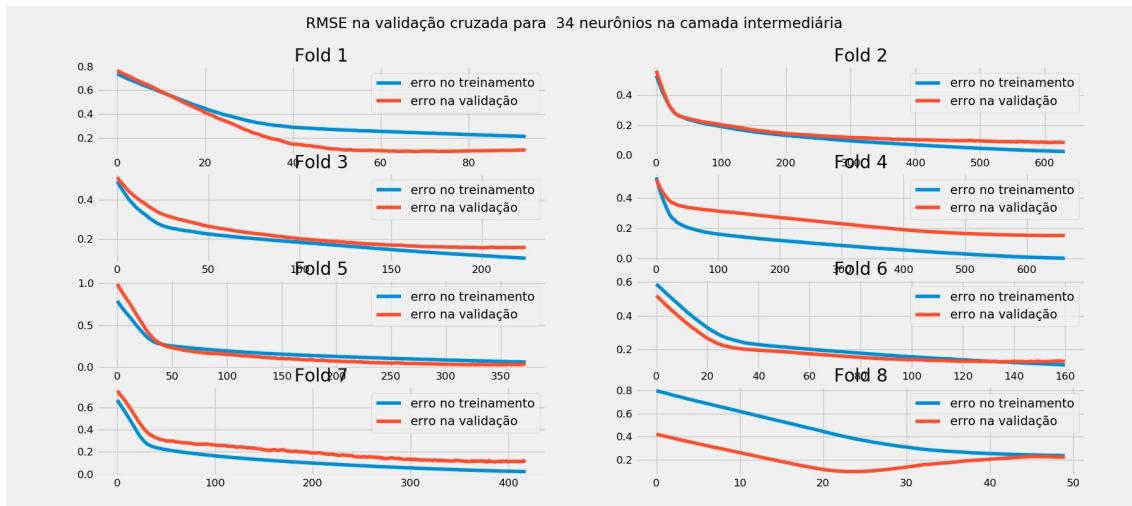


Figura A.15: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

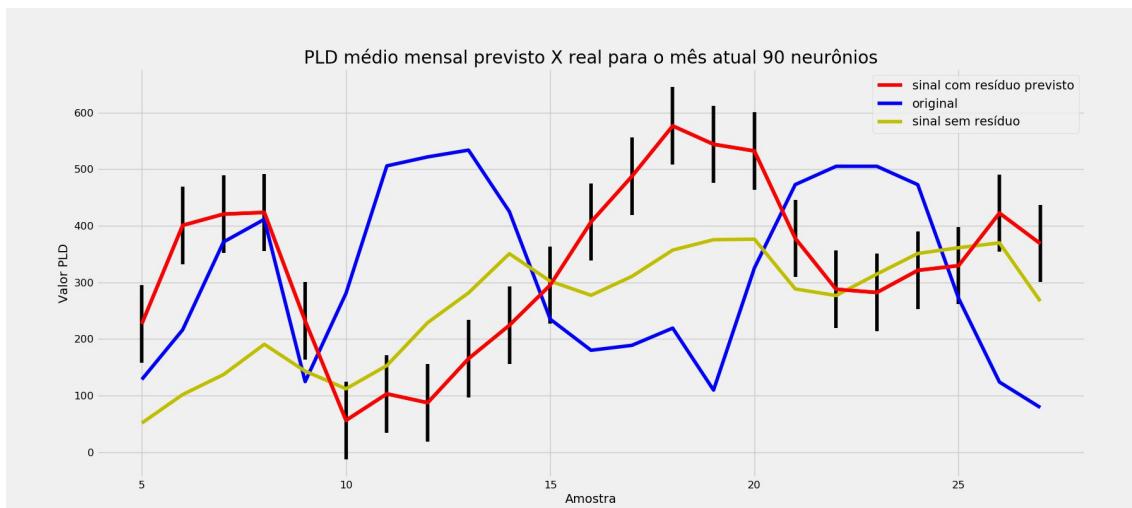


Figura A.16: Sinal completo.

A.5 Previsão para 6 meses a frente

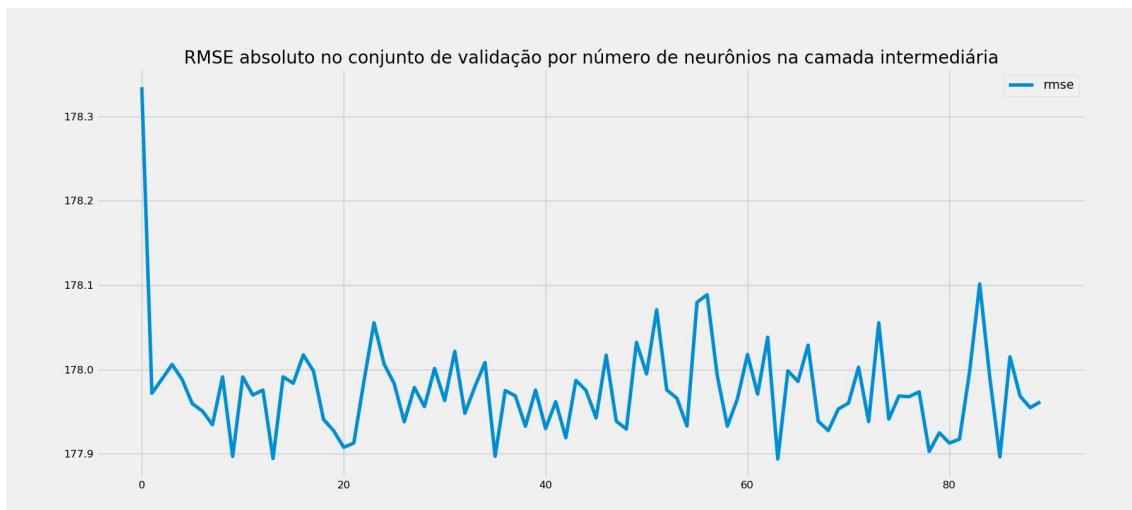


Figura A.17: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

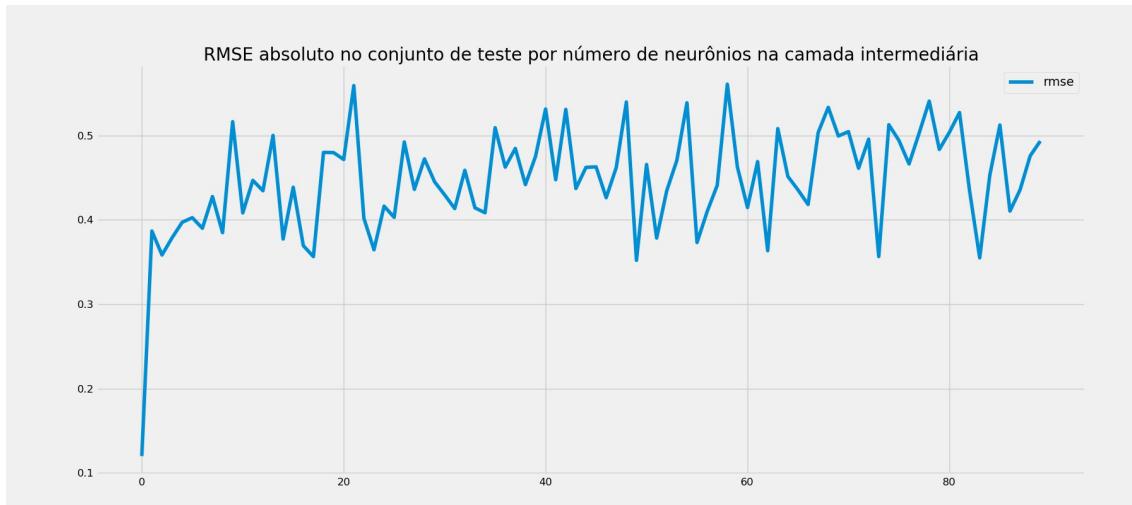


Figura A.18: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.5: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 57 | 88,770 | 79,111 | 0,975 | 0,000 | -0,008 | 0,000 | 0,001 |
| 20 | 95,084 | 88,687 | 0,953 | 0,001 | -0,007 | 0,000 | 0,001 |
| 71 | 81,320 | 75,164 | 1,074 | 0,001 | 0,001 | 0,000 | 0,001 |
| 85 | 73,433 | 70,922 | 0,978 | 0,000 | 0,032 | 0,001 | 0,002 |
| 17 | 72,576 | 66,944 | 1,089 | 0,002 | -0,028 | 0,001 | 0,003 |

E com isso o modelo com 57 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

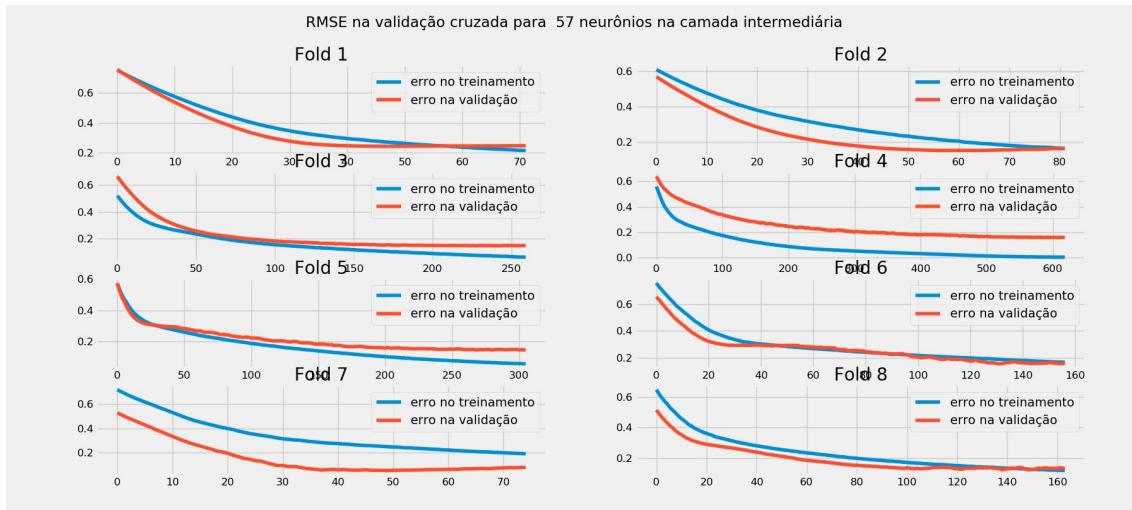


Figura A.19: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

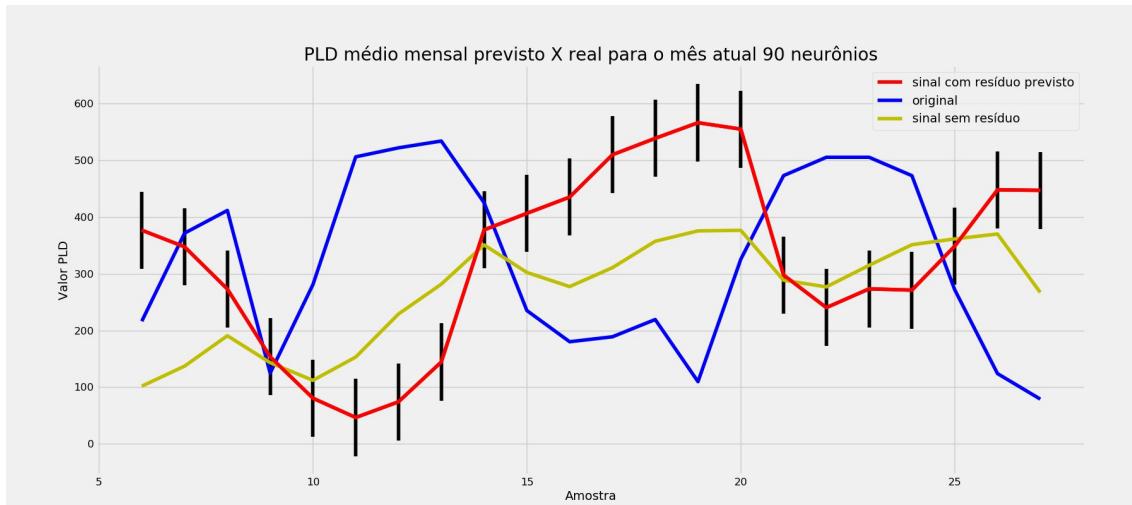


Figura A.20: Sinal completo.

A.6 Previsão para 7 meses a frente

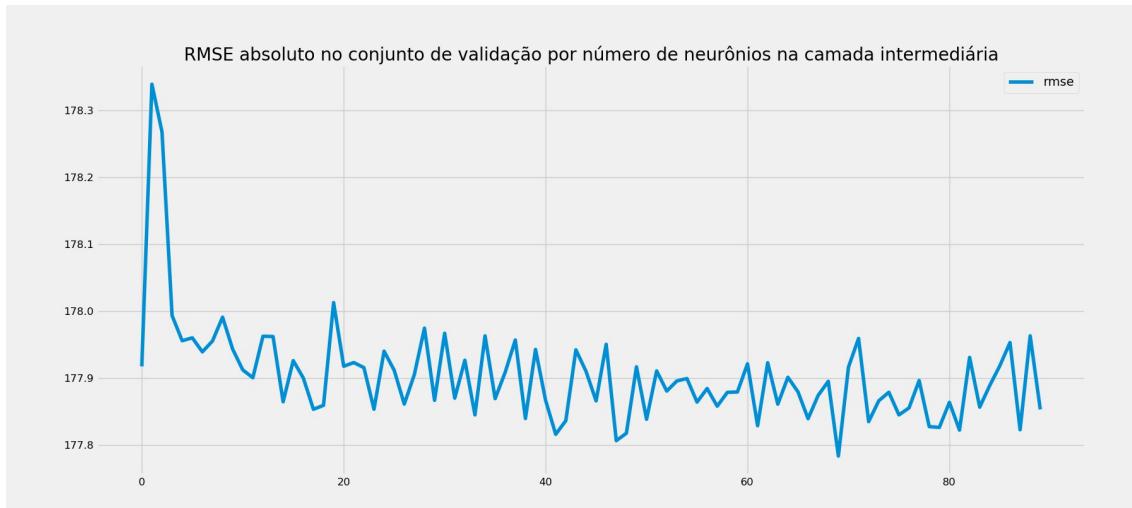


Figura A.21: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

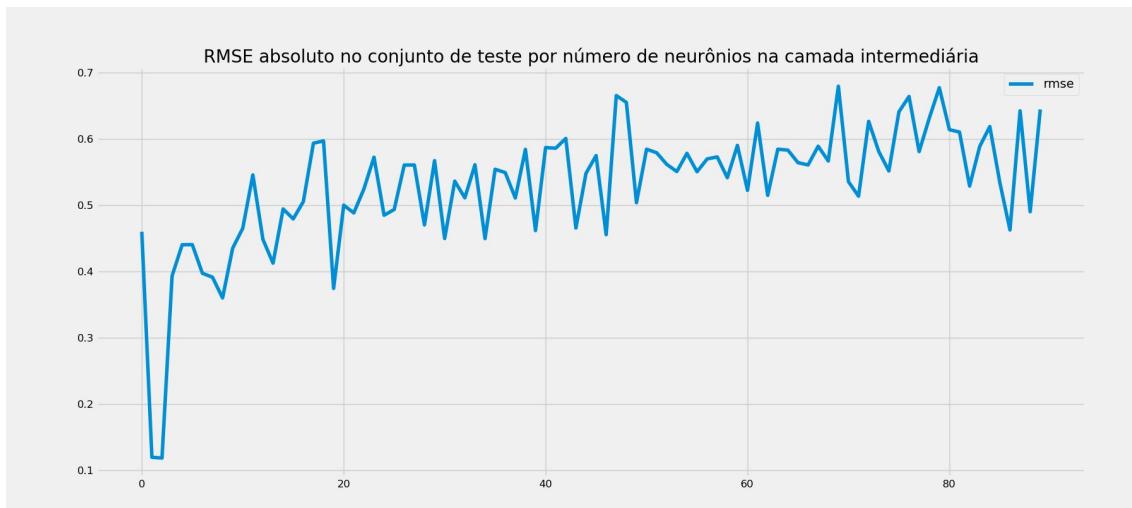


Figura A.22: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.6: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 85 | 81,374 | 74,638 | 1,073 | 0,001 | -0,015 | 0,000 | 0,001 |
| 59 | 87,220 | 79,156 | 0,963 | 0,000 | 0,034 | 0,001 | 0,001 |
| 55 | 94,054 | 84,463 | 0,964 | 0,000 | -0,041 | 0,001 | 0,001 |
| 88 | 88,838 | 80,536 | 1,105 | 0,001 | -0,038 | 0,001 | 0,002 |
| 45 | 82,782 | 78,053 | 0,891 | 0,001 | -0,048 | 0,001 | 0,002 |

E com isso o modelo com 85 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

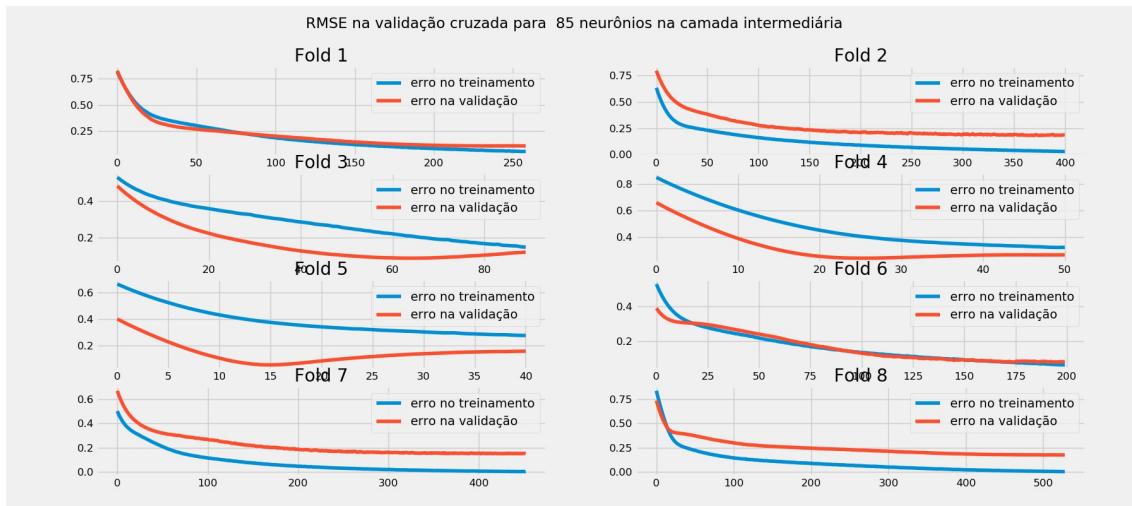


Figura A.23: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

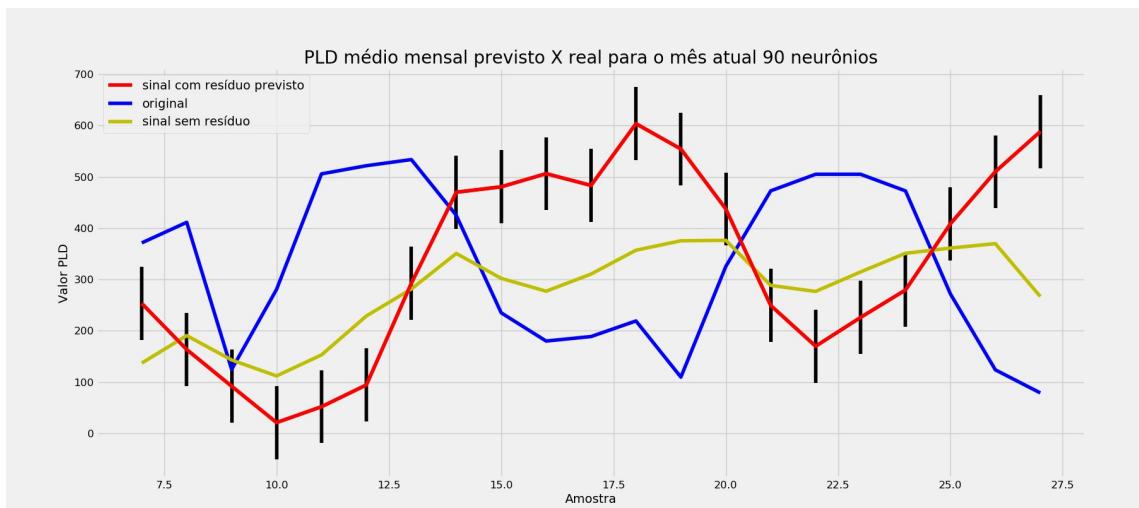


Figura A.24: Sinal completo.

A.7 Previsão para 8 meses a frente

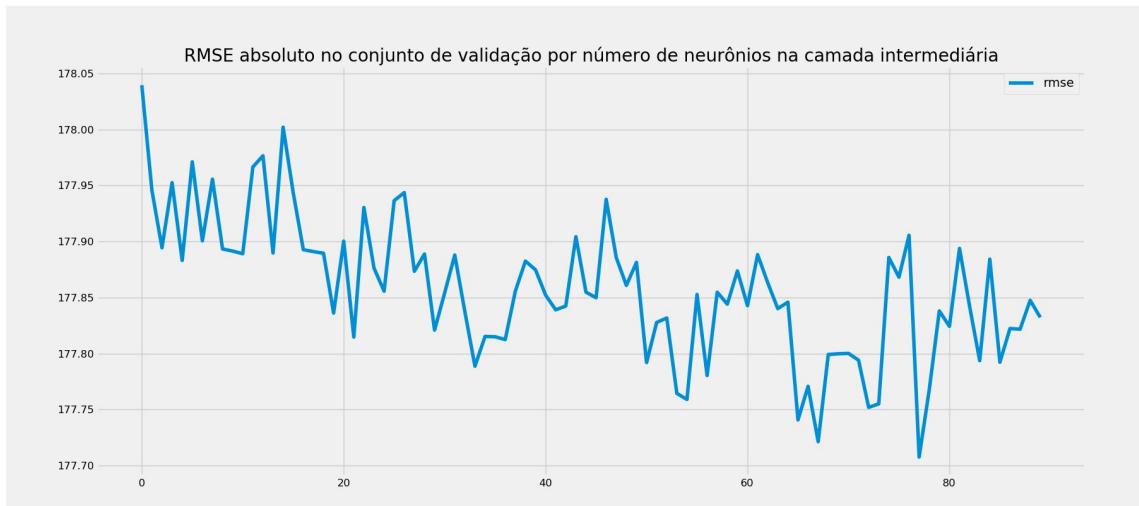


Figura A.25: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

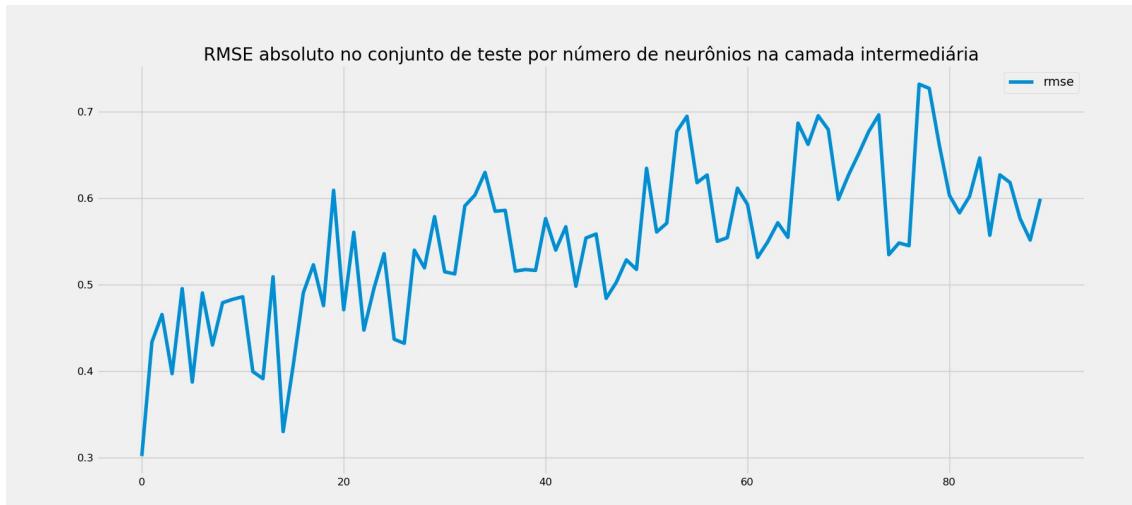


Figura A.26: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.7: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 29 | 93,498 | 61,547 | 1,060 | 0,001 | -0,002 | 0,000 | 0,001 |
| 24 | 94,052 | 69,505 | 1,055 | 0,001 | -0,035 | 0,001 | 0,002 |
| 9 | 83,198 | 57,893 | 1,089 | 0,001 | -0,029 | 0,001 | 0,002 |
| 66 | 86,705 | 61,198 | 1,068 | 0,001 | -0,042 | 0,001 | 0,002 |
| 51 | 92,674 | 62,083 | 1,036 | 0,001 | 0,060 | 0,002 | 0,003 |

E com isso o modelo com 29 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

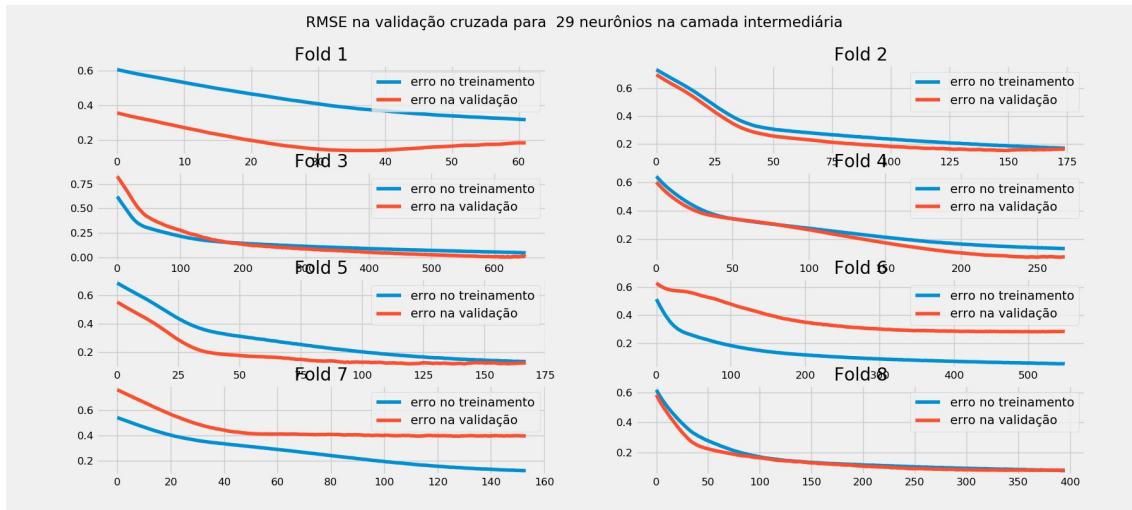


Figura A.27: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

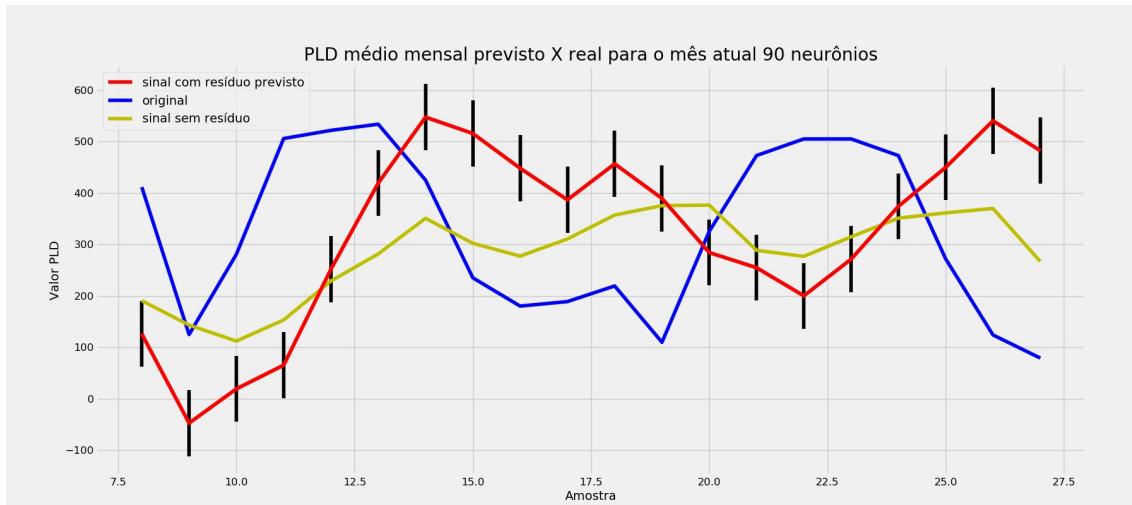


Figura A.28: Sinal completo.

A.8 Previsão para 9 meses a frente

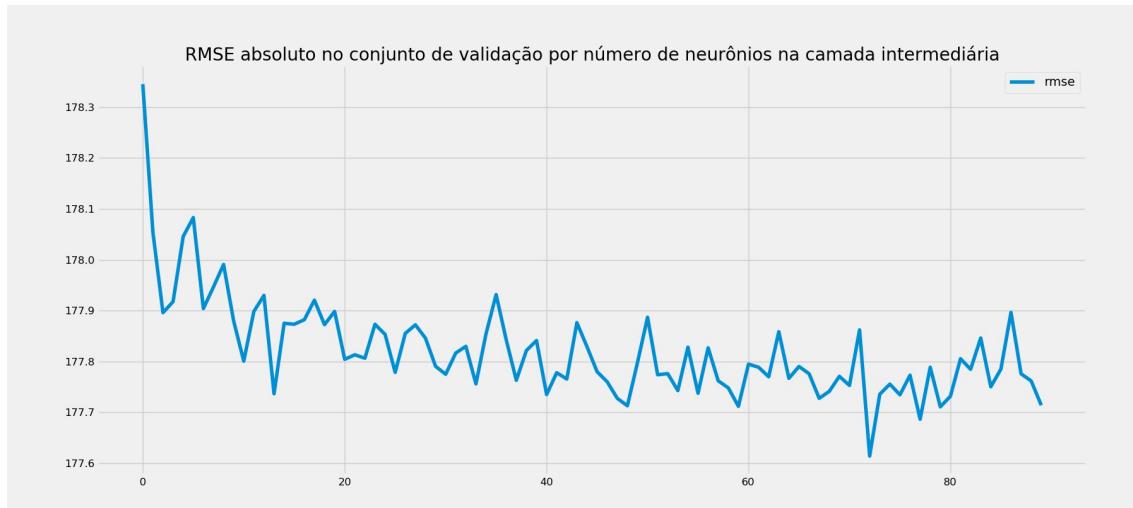


Figura A.29: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

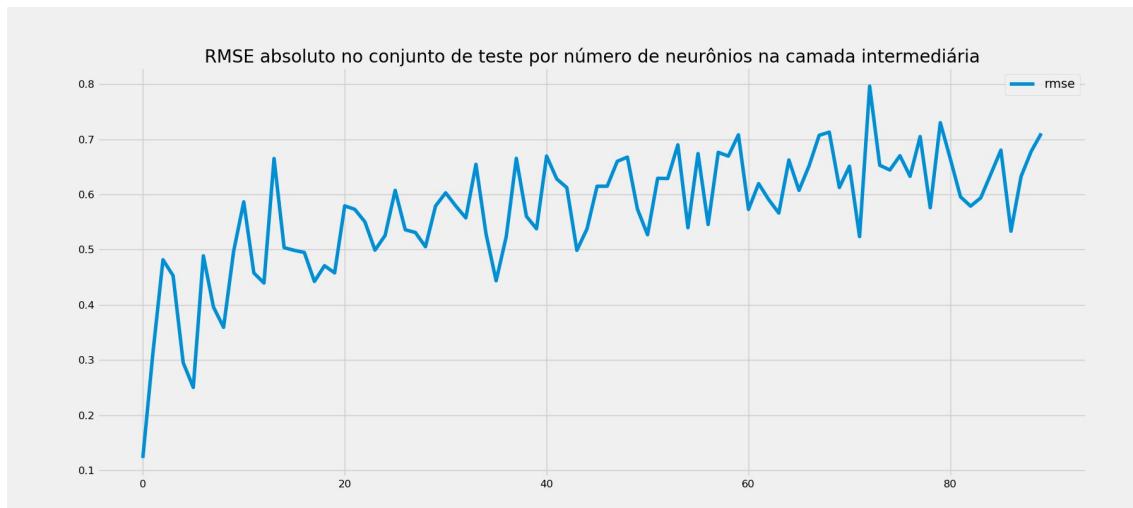


Figura A.30: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.8: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 66 | 53,146 | 35,811 | 0,968 | 0,000 | -0,132 | 0,001 | 0,002 |
| 30 | 81,717 | 66,631 | 1,147 | 0,001 | 0,050 | 0,000 | 0,002 |
| 39 | 74,664 | 53,134 | 0,996 | 0,000 | -0,220 | 0,002 | 0,002 |
| 74 | 63,687 | 45,221 | 0,708 | 0,002 | -0,017 | 0,000 | 0,002 |
| 90 | 93,000 | 79,354 | 0,944 | 0,000 | 0,220 | 0,002 | 0,003 |

E com isso o modelo com 66 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

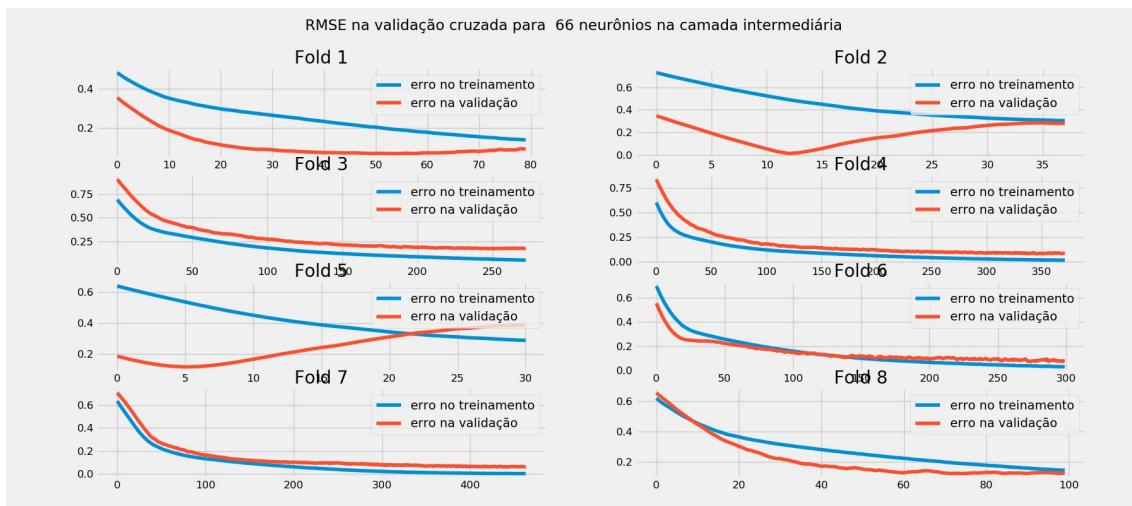


Figura A.31: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

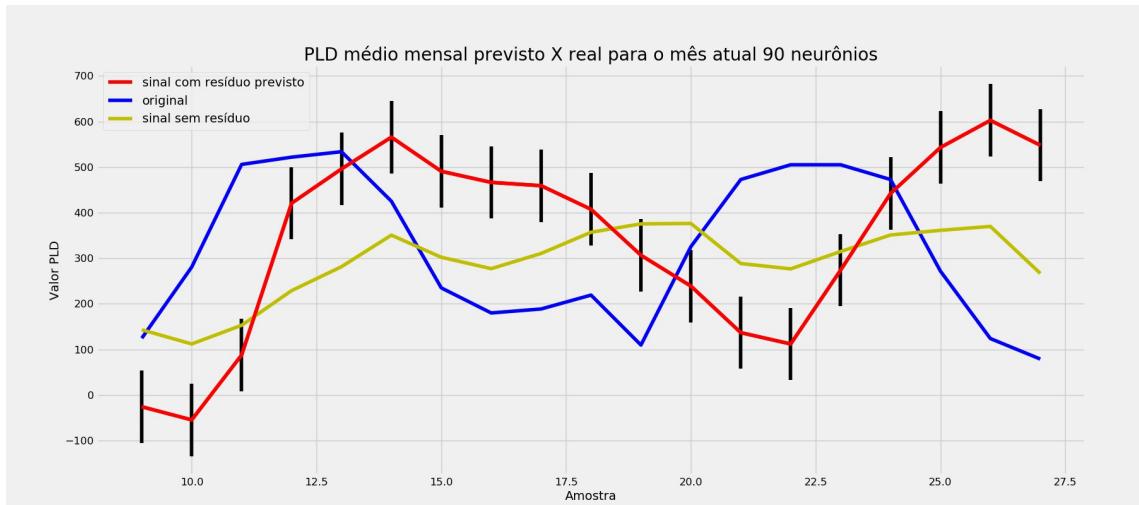


Figura A.32: Sinal completo.

A.9 Previsão para 10 meses a frente

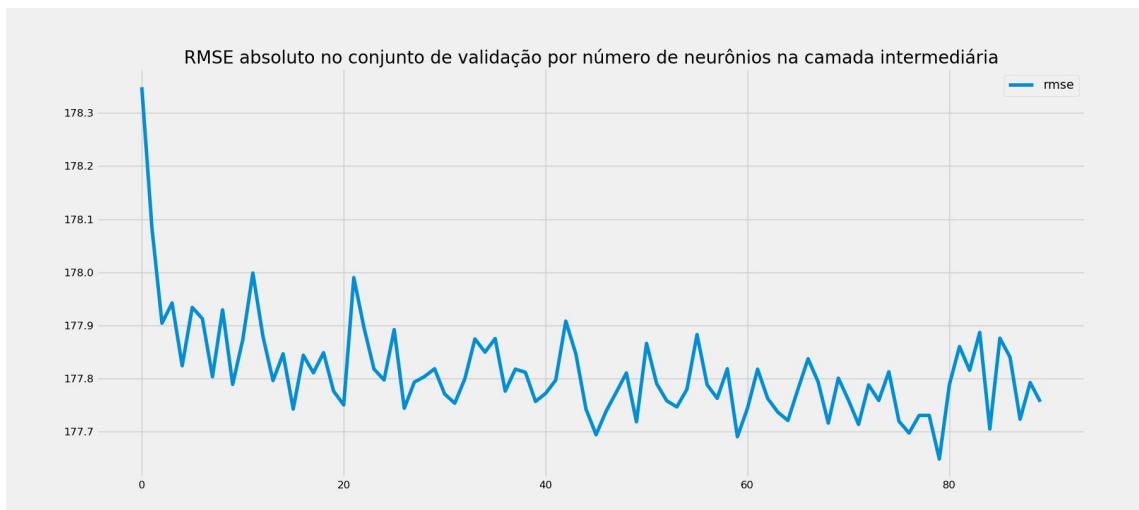


Figura A.33: RMSE pelo número de neurônios no conjunto de validação.

No conjunto de teste o resultado foi o seguinte:

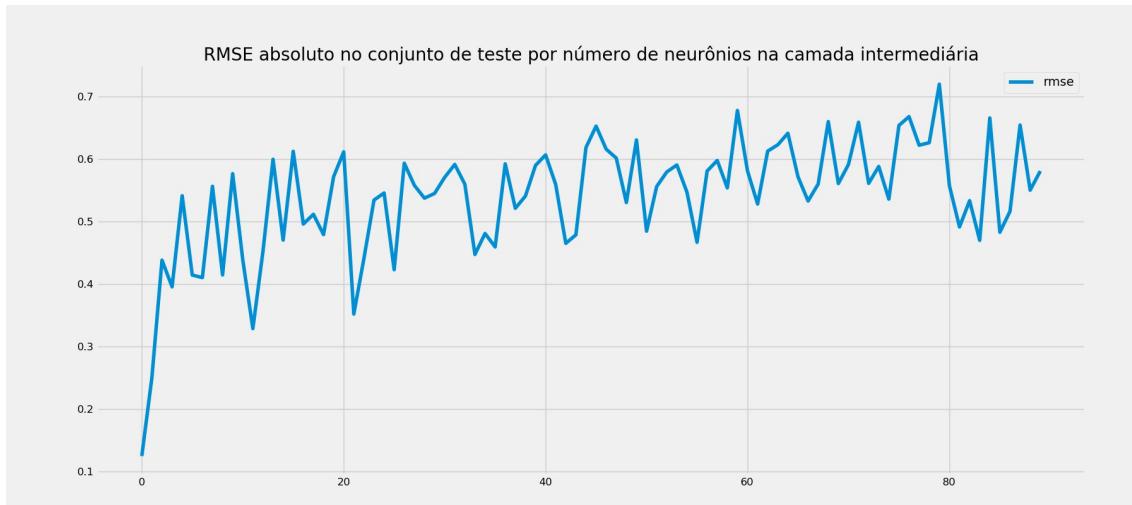


Figura A.34: RMSE pelo número de neurônios no conjunto de teste.

Tabela A.9: Resultados obtidos com as redes no dataset de validação.

| #neurônios | RMSE | STD | a (média) | ϵ_1 | b (média) | ϵ_2 | ϵ_3 |
|------------|--------|--------|-------------|--------------|-------------|--------------|--------------|
| 74 | 47,276 | 24,496 | 0,979 | 0,000 | 0,092 | 0,002 | 0,002 |
| 82 | 50,150 | 38,221 | 0,794 | 0,002 | -0,065 | 0,002 | 0,004 |
| 86 | 71,629 | 49,838 | 0,744 | 0,003 | -0,061 | 0,001 | 0,004 |
| 55 | 47,440 | 34,009 | 1,408 | 0,005 | 0,011 | 0,000 | 0,005 |
| 32 | 41,108 | 27,709 | 1,073 | 0,001 | 0,181 | 0,004 | 0,005 |

E com isso o modelo com 74 neurônios foi escolhido. O treinamento dessa estrutura obteve os seguintes erros pelo número de épocas:

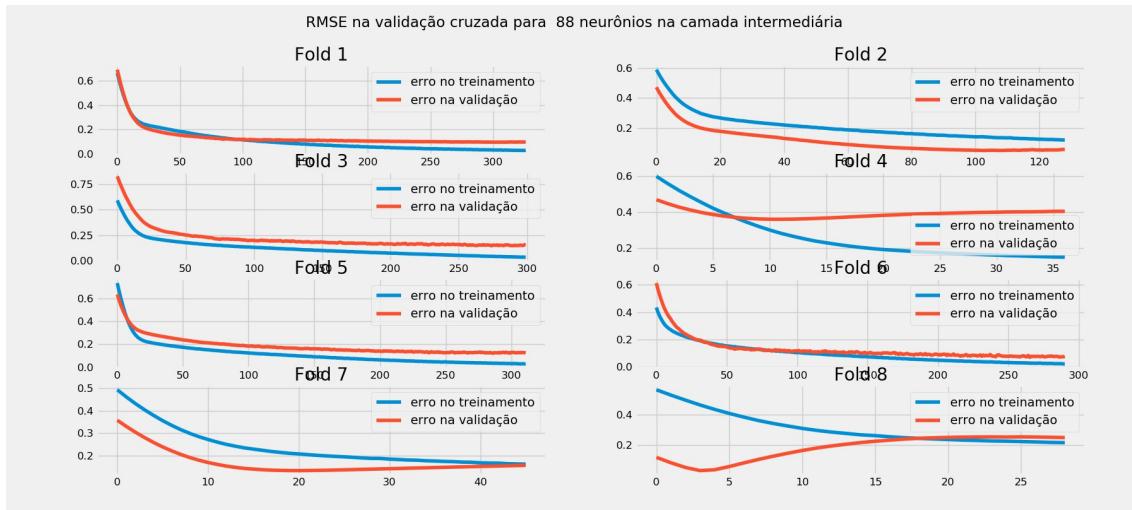


Figura A.35: Erro pelo número de épocas e por fold.

E obteve-se os seguintes resultados:

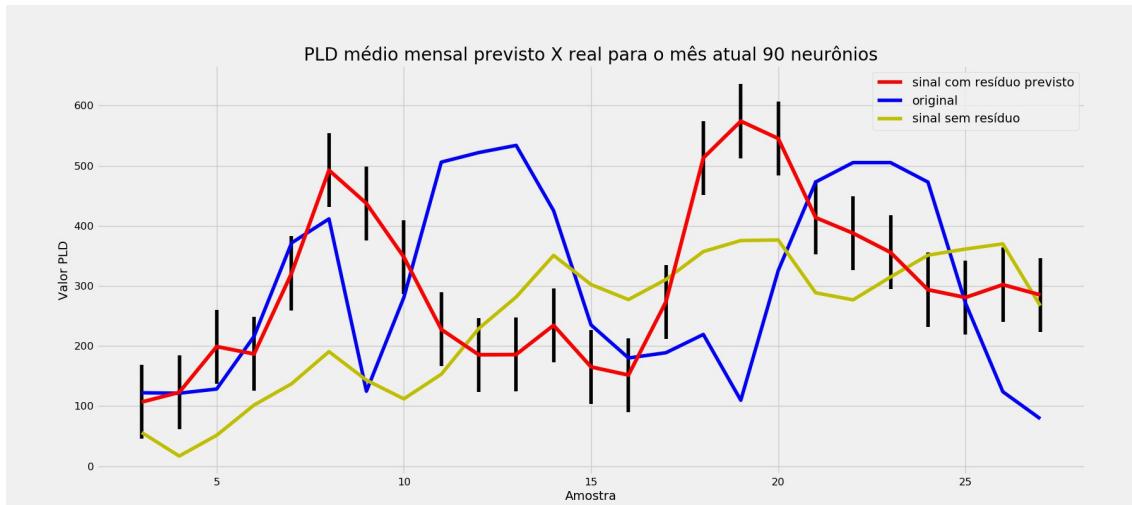


Figura A.36: Sinal completo.