



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE ENGENHARIA
CÂMPUS DE ILHA SOLTEIRA

ANGELO ANTONIO DE CARVALHO BONVICINE

PREVISÃO DE CARGA PARA A PROGRAMAÇÃO DIÁRIA ELETROENERGÉTICA DO
SISTEMA INTERLIGADO NACIONAL POR REDES NEURAIAS ARTIFICIAIS

Ilha Solteira
2021



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

ANGELO ANTONIO DE CARVALHO BONVICINE

PREVISÃO DE CARGA PARA A PROGRAMAÇÃO DIÁRIA ELETROENERGÉTICA DO
SISTEMA INTERLIGADO NACIONAL POR REDES NEURAIAS ARTIFICIAIS

Trabalho de conclusão de curso apresentado à
Faculdade de Engenharia de Ilha Solteira –
Unesp como parte dos requisitos para
obtenção do título de Engenheiro Eletricista .

Prof^a Dr. Anna Diva Plasencia Lotufo
Orientadora

Ilha Solteira
2021

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

B723p Bonvicine, Angelo Antonio de Carvalho.
Previsão de carga para a programação diária eletroenergética do Sistema Interligado Nacional por Redes Neurais Artificiais / Angelo Antonio de Carvalho Bonvicine. -- Ilha Solteira: [s.n.], 2021
77 f. : il.

Trabalho de conclusão de curso (Graduação em Engenharia Elétrica) -
Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira, 2021

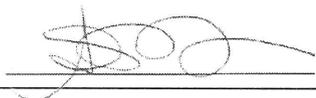
Orientadora: Anna Diva Plasencia Lotufo
Inclui bibliografia

1. Previsão de carga. 2. Planejamento da operação eletroenergética. 3.
Modelo DESSEM. 4. Séries temporais. 5. Redes Neurais Recorrentes. 6. LSTM.


Raiane da Silva Santos

ATA DE DEFESA DE TRABALHO DE GRADUAÇÃO

Aos trinta dias do mês de junho do ano de dois mil e vinte e um, o discente *Angelo Antonio de Carvalho Bonvicine*, matriculado sob o nº 142054909, tendo como banca examinadora a sua orientadora, a *Profª. Drª. Anna Diva Plasencia Lotufo*, o *Doutorando Lucas do Carmo Yamaguti* e o *Doutorando Leonardo Brain Garcia Fernandez*, apresentou o Trabalho de Graduação intitulado "PREVISÃO DE CARGA PARA A PROGRAMAÇÃO DIÁRIA ELETROENERGÉTICA DO SISTEMA INTERLIGADO NACIONAL POR REDES NEURAIS ARTIFICIAIS" obtendo a nota 9,0 (nove vírgula zero) e conceito APROVADO.



Profª. Drª. Anna Diva Plasencia Lotufo

- orientador -



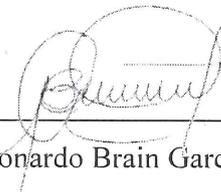
Angelo Antonio de Carvalho Bonvicine

- discente -



Doutorando Lucas do Carmo Yamaguti

- Membro da Banca -



Doutorando Leonardo Brain Garcia Fernandez

- Membro da Banca -



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

DEDICATÓRIA

Ao meu amor.

AGRADECIMENTOS

Agradeço, inicialmente, a minha família, em especial meu pai, pela oportunidade de ingressar nessa jornada, e por tanto apoiar quando mais precisei.

À minha namorada Thayná e sua mãe Lucilene, minha segunda família.

Aos meus amigos mais próximos pelo carinho e parceria de longa data.

Aos meus colegas de república, por dividirem suas vidas e rotinas comigo, e aos quais tenho grande admiração.

Ao meu colega de profissão e amigo, Bruno Morelli, pelos conselhos e ajuda na escolha do tema deste trabalho.

Aos demais amigos e colegas da UNESP, por compartilharem momentos incríveis e experiências nessa jornada.

À minha orientadora, professora Anna Diva, pela confiança em meu trabalho e pela pronto suporte na orientação e desenvolvimento deste trabalho de graduação.

RESUMO

Em janeiro de 2021 houve a inclusão de um novo modelo na cadeia de modelos de planejamento eletroenergético desenvolvida pelo Centro de Pesquisas de Energia Elétrica, o CEPEL. Assim como os já consolidados modelos de médio e curto prazo desta cadeia, o modelo DESSEM requer uma gama de dados de entrada para seu funcionamento, dentre eles, a inclusão de uma curva de carga prevista para o horizonte do estudo de planejamento a ser executado, a qual o Operador Nacional do Sistema Elétrico, o ONS, constrói diariamente utilizando de suas metodologias. O objetivo deste trabalho é trazer uma contribuição utilizando conceitos de Redes Neurais Recorrentes com estruturas LSTM, para elaboração de um modelo de previsão de carga, que possa ser utilizado alternativamente ao modelo ONS para elaboração da curva de carga associada a programação diária eletroenergética do Sistema Interligado Nacional, o SIN.

Palavras-Chave: Previsão de Carga, Planejamento da operação eletroenergética, Modelo DESSEM, Séries Temporais, Redes Neurais Recorrentes, LSTM.

ABSTRACT

In January 2021, a new model was included in the chain of electro-energy planning models developed by the Electric Energy Research Center, CEPEL. As well as the consolidated medium and short-term models of this chain, the DESSEM model requires a range of input data for its operation, as for example, the inclusion of a forecasted load curve for the horizon of the planning study to be executed, where the National Electric System Operator, ONS, builds daily using its methodologies. The objective of this work is to contribute using concepts of Recurrent Neural Networks with LSTM structures, for the elaboration of a load forecast model, which can be used alternatively to the ONS model for building the load curve associated with the daily electrical energy schedule of the Interconnected System National, the SIN.

Keywords: Electrical Load forecast, Planning of electrical energy operation, DESSEM model, Time Series, Recurrent Neural Networks, LSTM.

LISTA DE FIGURAS

| | | |
|-------------|---|----|
| Figura 1 - | Agentes institucionais do setor elétrico Brasileiro | 19 |
| Figura 2 - | Problema de decisão da operação | 20 |
| Figura 3 - | Funções de custo imediato e custo Futuro | 21 |
| Figura 4 - | Estrutura de REE's e Subsistemas de energia | 23 |
| Figura 5 - | Cadeia de modelos de otimização desenvolvido pelo CEPEL | 25 |
| Figura 6 - | Representação do sistema hidrotérmico pelo modelo DESSEM | 27 |
| Figura 7 - | Representação dos horizontes de estudo do modelo DESSEM. | 28 |
| Figura 8 - | Representação em patamares cronológicos do modelo DESSEM | 29 |
| Figura 9 - | Visão geral do Processo de Consolidação da previsão de Carga Diária | 32 |
| Figura 10 - | Estrutura da rede ANNSTLF. | 33 |
| Figura 11 - | Modelo de neurônio artificial. | 37 |
| Figura 12 - | Função de ativação Limiar (Heaviside) | 38 |
| Figura 13 - | Função linear por partes | 39 |
| Figura 14 - | Função sigmóide | 40 |
| Figura 15 - | Redes neurais de camada única e multicamada | 41 |
| Figura 16 - | Rede recorrente com neurônios ocultos | 42 |
| Figura 17 - | Versão desdobrada do grafo computacional de uma RNN. | 44 |
| Figura 18 - | O problema do desaparecimento do gradiente. | 49 |
| Figura 19 - | Estrutura do Bloco de memória LSTM | 50 |
| Figura 20 - | Preservação do gradiente pela adição de LSTM | 51 |
| Figura 21 - | Séries temporais de Carga [MW médio] e Temperatura [°C] Sudeste/Centro-Oeste | 56 |
| Figura 22 - | Histograma 2D - Carga versus Temperatura | 57 |

| | |
|---|----|
| Figura 23 - Matriz de correlação de Pearson - Carga versus Temperatura. | 58 |
| Figura 24 - Análise de FFT para séries de temperatura e Carga. | 59 |
| Figura 25 - Séries temporais sintéticas de frequência diária. | 60 |
| Figura 26 - Estimador de Linha de Base (repetição). | 64 |
| Figura 27 - Funções de custo de treino e validação. | 67 |
| Figura 28 - Performance no conjunto de dados de Teste. | 68 |
| Figura 29 - Análise da previsão semanal do Modelo LSTM | 69 |
| Figura 30 - Previsão de carga diária para o sábado - 22 de maio de 2021 | 70 |

LISTA DE TABELAS

| | | |
|------------|---|----|
| Tabela 1 - | Proporção de segmentação dos conjuntos de dados de Treino, Validação e Teste. | 61 |
| Tabela 2 - | Datasets de treinamento, validação e teste do modelo de Rede Neural LSTM. | 61 |
| Tabela 3 - | Sumário do modelo de rede neural LSTM. | 63 |
| Tabela 4 - | Sumário do modelo de Rede Neural Recorrente sem estruturas LSTM. | 64 |
| Tabela 5 - | Valores de Hiperparâmetros da etapa de <i>Backpropagation</i> . | 66 |

LISTA DE QUADROS

Quadro 1 - Estrutura de arquivos de entrada do modelo PrevCargaDessem

54

LISTA DE ABREVIACES E SIGLAS

| | |
|---------|--|
| ADAM | <i>Adaptive Moment Estimation</i> |
| ANEEL | Agncia Nacional de Energia Eltrica |
| ANNSTLF | <i>Artificial Neural Network Short-term Load Forecaster</i> |
| B(R)LF | <i>Base or regular load forecaster</i> |
| BPTT | <i>Backpropagation Through Time</i> |
| C(D)LF | <i>Change or Delta Load Forecaster</i> |
| CCEE | Cmara de Comercializao de Energia Eltrica |
| CEPEL | Centro de Pesquisas de Energia Eltrica |
| CMSE | Comit de Monitoramento do Setor Eltrico |
| CNPE | Conselho Nacional de Poltica Energtica |
| DECOMP | Modelo de Planejamento da Operao de Sistemas Hidrotrmicos Interligados de Curto Prazo |
| DESSEM | Modelo de Despacho Hidrotrmico de Curto Prazo |
| EPE | Empresa de Pesquisa Energtica |
| EPRI | <i>Electric Power Research Institute</i> |
| FCF | Funo de Custo Futuro |
| FCI | Funo de Custo Imediato |
| FFT | <i>Fast Fourier Transform</i> |
| GPU | <i>Graphic Processing Units</i> |
| LSTM | <i>Long-Short Term Memory</i> |
| MAPE | <i>Mean Absolute Percentage Error</i> |
| MME | Ministrio de Minas e Energia |
| MSE | <i>Mean Squared Error</i> |
| NEWAVE | Modelo de Planejamento da Operao de Sistemas Hidrotrmicos Interligados de Longo e Mdio Prazo |
| ONS | Operador Nacional do Sistema |
| PDDE | Programao Dinmica Dual Estocstica |

| | |
|------|--|
| PDE | Programação diária Eletroenergética |
| REE | Reservatório Equivalente de Energia |
| RFFT | <i>Real Fast Fourier Transform</i> |
| RLS | <i>Recursive Least Squares</i> |
| RNN | Redes neurais Recorrentes |
| RTRL | <i>Real Time Recurrent Learning</i> |
| SEB | Sistema Elétrico Brasileiro |
| SIGA | Sistema de Informações de Geração da Aneel |
| SIN | Sistema Interligado Nacional |
| SVR | <i>Support Vector Regression</i> |

LISTA DE SÍMBOLOS

| | |
|---------------------|--|
| \overline{x}_i | i -ésimo elemento do vetor de Entrada da Rede Neural. |
| \overline{w}_{hi} | i -ésimo peso sináptico da camada oculta h . |
| u_{hk} | Somatório das entradas do neurônio k da camada h . |
| v_{hk} | Campo local induzido do neurônio k na camada h . |
| y_i | i -ésimo elemento do vetor de Saída da rede Neural |
| $\varphi(v)$ | Função de ativação. |
| $L(y, \hat{y})$ | Função de custo (<i>Loss Function</i>). |
| δ_{hk} | Delta de erro do neurônio k da camada h |
| f_i^t | Ativação do <i>Forget Gate</i> da célula LSTM i no estágio t . |
| g_i^t | Ativação do <i>Input Gate</i> da célula LSTM i no estágio t . |
| q_i^t | Ativação do <i>Output Gate</i> da célula LSTM i no estágio t . |
| s_i^t | Estado culto da célula LSTM i no estágio t . |

SUMÁRIO

| | |
|--|-----------|
| 1 - INTRODUÇÃO | 17 |
| 1.2 - Resumo da estrutura Setor Elétrico Brasileiro | 18 |
| 1.3 - Planejamento da operação do sistema elétrico | 20 |
| 2 - MODELOS DE OTIMIZAÇÃO ENERGÉTICA | 25 |
| 2.1 - Modelo NEWAVE | 26 |
| 2.2 - Modelo DECOMP | 26 |
| 2.3 - Modelo DESSEM | 27 |
| 3 - PREVISÃO DE CARGA PARA O MODELO DESSEM | 30 |
| 3.1 - Definição de Carga pelo ONS | 30 |
| 3.2 - Previsão de carga para programação diária da operação eletroenergética (PDE) | 31 |
| 3.2 - Modelos de Previsão de Carga | 32 |
| 4 - REDES NEURAIS ARTIFICIAIS | 35 |
| 4.1 - A estrutura do neurônio artificial. | 36 |
| 4.2 - Funções de ativação | 38 |
| 4.3 - Redes neurais de multicamada | 41 |
| 4.4 - Redes Neurais Recorrentes | 42 |
| 5 - TREINAMENTO DE REDES NEURAIS | 45 |
| 5.1 - Função de Custo (Loss Function) | 45 |
| 5.2 - Algoritmo de Retropropagação (backpropagation) | 46 |
| 6 - LONG SHORT TERM MEMORY (LSTM) | 49 |
| 6.1 - Estrutura geral de redes LSTM | 50 |
| 6.2 - Equações das estruturas LSTM | 52 |
| 6.3 - Cálculo do Gradiente | 53 |
| 7 - DADOS DE ENTRADA DO MODELO DE REDE NEURAL LSTM | 54 |
| 7.1 - Principais dados de entrada | 54 |
| 7.2 - Análise dos dados históricos de Carga e Temperatura. | 56 |
| 7.3 - Preparação dos dados de Treinamento, Validação e Teste | 60 |
| 7.4 - Ajuste de escala dos dados de entrada | 62 |
| 8 - ELABORAÇÃO E TREINAMENTO DO MODELO | 63 |
| 8.1 - Estimadores de Referência | 63 |

| | |
|--|-----------|
| 8.2 - Ambiente de construção dos modelos | 65 |
| 8.3 - Treinamento do modelo LSTM | 65 |
| 9 - RESULTADOS E DISCUSSÕES | 67 |
| 10 - CONCLUSÕES | 71 |
| 10.1 - Sugestões para trabalhos futuros | 72 |
| REFERÊNCIAS | 73 |

1 - INTRODUÇÃO

O Brasil é o país com uma das mais complexas estruturas elétricas do mundo. Operar um sistema elétrico de proporções continentais exige um planejamento eficiente e adequado, que possibilite a exploração inteligente de seus recursos.

Planejamento e investimento, são pilares que estão vinculados à saúde do sistema elétrico como um todo, e episódios históricos de relacionamentos mostraram a importância desses pilares, que em sua ausência trouxeram consequências sociais severas.

Um planejamento otimizado tem como base fornecer uma política de evolução para o sistema elétrico como um todo, garantindo alguns preceitos fundamentais como:

- Aumento da energia disponível através do uso inteligente de seus recursos disponíveis.
- Garantia de maior confiabilidade dos consumidores de energia, assegurando suprimento contínuo, mesmo nos períodos de maior demanda.
- Minimizar os custos de operação do sistema, sempre levando em consideração as restrições sociais, ambientais, políticas, geográficas-temporais e financeiras.

Dentre as entidades do setor elétrico, a responsabilidade de coordenar e gerenciar esse sistema pertence, atualmente, ao operador nacional do sistema ONS. Esse planejamento executado pelo operador abrange diferentes horizontes, longo, médio e curto prazo.

Essa peculiaridade do planejamento se deve a natureza da matriz energética do Sistema Elétrico Brasileiro (SEB), sendo majoritariamente composta por usinas hidrelétricas, e que comporta volumosos intercâmbios de energia, através do Sistema interligado nacional, o SIN.

Para executar este planejamento uma série de modelos computacionais, desenvolvidos pelo Centro de Pesquisas de Energia Elétrica, o CEPREL, em parceria com o ONS, foram desenvolvidos. Esses modelos têm como objetivo representar da melhor forma possível a situação atual do SIN e promover uma operação otimizada, para cada um dos horizontes de planejamento. Inicialmente, os modelos NEWAVE e DECOMP executavam este papel de dar

apoio ao ONS nessa difícil tarefa, concebendo planejamentos com horizontes de médio e curto prazo, em bases mensal e semanal respectivamente.

Com a crescente busca por melhorias de todo este processo, em janeiro de 2021 foi inserido oficialmente um terceiro modelo a essa cadeia, o modelo DESSEM, que por sua vez é responsável por fornecer o planejamento da programação diária da operação, em base horária ou semi-horária.

Assim como os demais modelos, o DESSEM necessita de insumos de entrada para sua execução. A exemplo, é necessário fornecer uma curva de carga prevista, para ser utilizada no horizonte de planejamento do modelo, para que assim o modelo consiga determinar a melhor estratégia de uso dos recursos energéticos disponíveis para atender esta demanda.

O ONS possui metodologias para elaboração dessa curva, as quais devem ser constantemente revisadas e melhoradas. O objetivo deste trabalho é fornecer uma contribuição neste sentido, utilizando uma abordagem com redes neurais artificiais para executar a etapa de previsão de carga, que poderá ser utilizada para alimentar o modelo DESSEM.

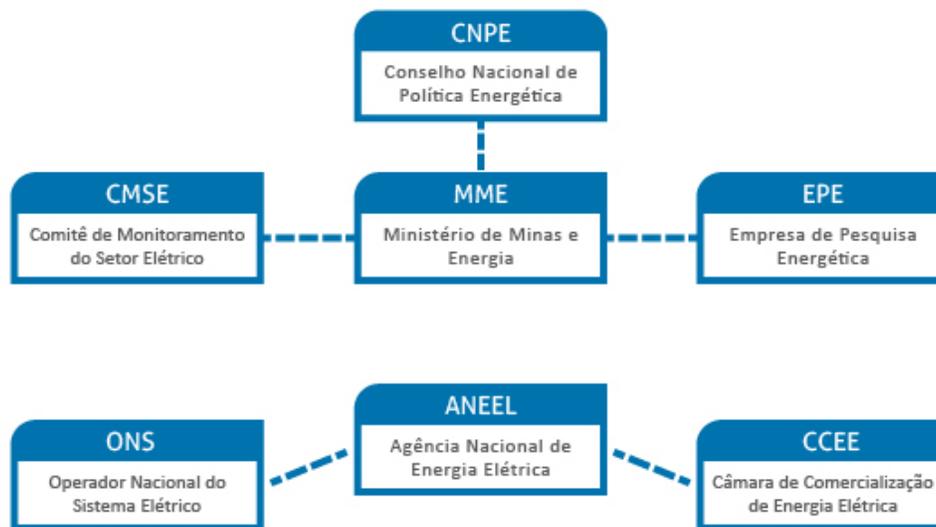
1.2 - Resumo da estrutura Setor Elétrico Brasileiro

Sob o novo modelo descrito por Tolmasquim (2015), o governo federal definiu um conjunto de agentes institucionais com competências atribuições bem-definidas, objetivando garantir o bom funcionamento setorial e alcançar os três objetivos almejados, a saber, modicidade tarifária, segurança no suprimento e universalização do acesso. Segundo o autor, os agentes podem ser classificados em 3 níveis, seguindo a natureza jurídica do ente e suas competências jurídicas.

Em breve resumo, as atividades do governo são exercidas pelo Conselho Nacional de Política Energética (CNPE), pelo Ministério de Minas e Energia (MME) e pelo CMSE. As atividades de regulação são exercidas pela ANEEL e, adicionalmente, entes operacionais executam atividades técnicas, como planejamento da expansão do setor (EPE), planejamento e

programação da operação (ONS) e a viabilização das atividades de comercialização (CCEE) (Tolmasquim, 2015).

Figura 1 - Agentes institucionais do setor elétrico Brasileiro



fonte: CCEE (2021)

O Sistema Elétrico Brasileiro (SEB) apresenta características que fazem com que a operação e o planejamento do sistema sejam uma atividade complexa e de difícil implementação. Um atributo técnico fundamental do setor é o equilíbrio físico, o qual requer a coordenação do sistema, isso porque suas partes operam com forte interdependência.

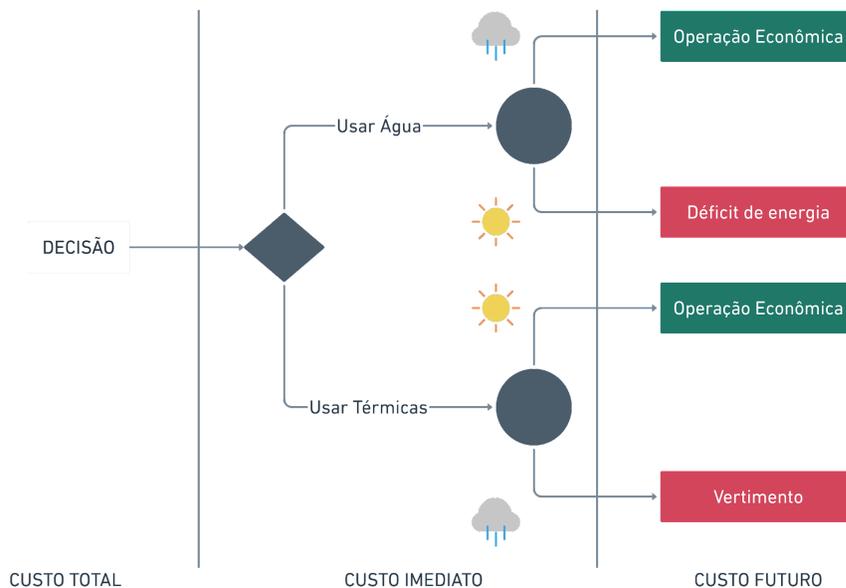
A geração de energia pode ser obtida de diversas fontes, com diferentes custos e impactos socioambientais. No caso das hidrelétricas, por exemplo, o insumo é um fluxo aleatório baseado no regime de chuvas, o que implica grande complexidade no processo de planejamento de curto e médio prazos, tarefa executada pelo ONS (Souza, 2014).

1.3 - Planejamento da operação do sistema elétrico

Segundo dados do SIGA (2021), o sistema de geração de energia elétrica do Brasil conta com uma capacidade instalada de aproximadamente 175.700 MW de potência instalada, sendo 109.325 MW de natureza hidrelétrica. Outra grande parcela dessa geração é complementada basicamente por usinas termolétricas, juntamente com outras parcelas de fontes intermitentes.

Nesse contexto, Souza (2014) destaca que em um sistema hidrotérmico em que tem-se; de um lado a disponibilidade futura de geração hídrica incerta, a um custo de geração considerado zero, de outro, a geração térmica futura supostamente conhecida, a um custo de geração elevado, fica evidente a necessidade de decidir a cada instante o nível de utilização térmica e hidráulica, onde a decisão tomada no presente, tem consequências diretas ao sistema, conforme exemplificado na figura 2.

Figura 2 - Problema de decisão da operação



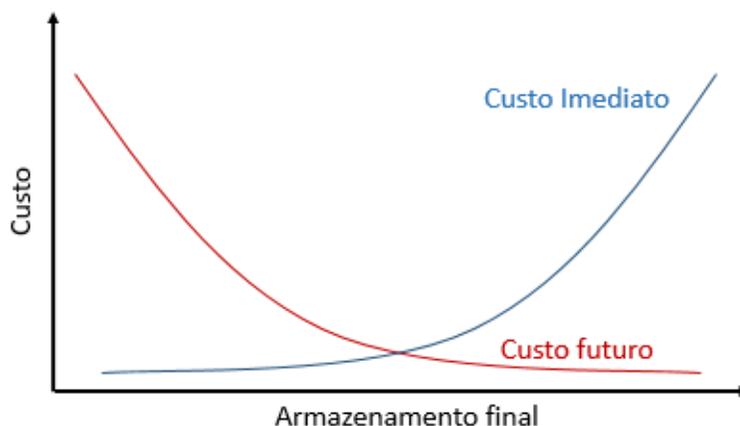
fonte: Adaptado de Souza (2014)

A exemplo, caso ocorra um despacho hidrelétrico expressivo antes de um período seco, corre-se o risco de despachar térmicas com custo muito mais elevado no futuro. Por outro lado, um despacho térmico anterior a um período úmido poderá acarretar em vertimentos, incorrendo em desperdício de energia. Este problema é conhecido como “dilema do operador”, e é um exercício que mostra com clareza o acoplamento temporal entre as decisões tomadas na elaboração do planejamento da operação (Souza, 2014).

Dessa forma uma operação equilibrada, como definido por Kelman (2001), seria aquela onde é possível gerar energia suficiente para atender a carga, com um pequeno risco de falha, havendo um compromisso mútuo entre deplecionar (usar água) ou não deplecionar (usar térmicas) os reservatórios. A variável de decisão é o volume de água armazenado ao final do período de operação.

Tal decisão se associa a um custo imediato, vinculado as térmicas que deverão ser despachadas no horizonte de planejamento, representado pela função de custo imediato FCI , e a custos futuros associados à expectativa de despacho térmico, indicado pela função de custo futuro FCF , conforme descrito por Pereira et al. (1998), conforme a figura 3.

Figura 3 - Funções de custo imediato e custo Futuro



Fonte: Adaptado de Pereira et al. (1998)

Na figura 3, observa-se o crescimento da função de custo imediato à medida que o armazenamento esperado ao final do horizonte de planejamento aumenta, o que pressupõe-se um gasto mais elevado com térmicas para preservar estes reservatórios. Todavia, a função de custo futuro tem comportamento contrário a função de custo imediato, pois quanto maior o estoque final de energia ao final do planejamento, menores serão os gastos com combustíveis no futuro. (Souza 2014).

Assim, a coordenação do planejamento da operação de um sistema hidrotérmico, como o SIN, pode ser desenhada como um problema de otimização de grande porte, com o acoplamento temporal e espacial da operação, dinâmico, estocástico, interconectado e não-linear.

O SEB adota uma cadeia de modelos para a solução deste problema, que considera diferentes horizontes de planejamento, e que são acoplados através da função de custo futuro *FCF*, cujos horizontes adotados são: longo prazo, médio prazo, curto prazo e programação diária.

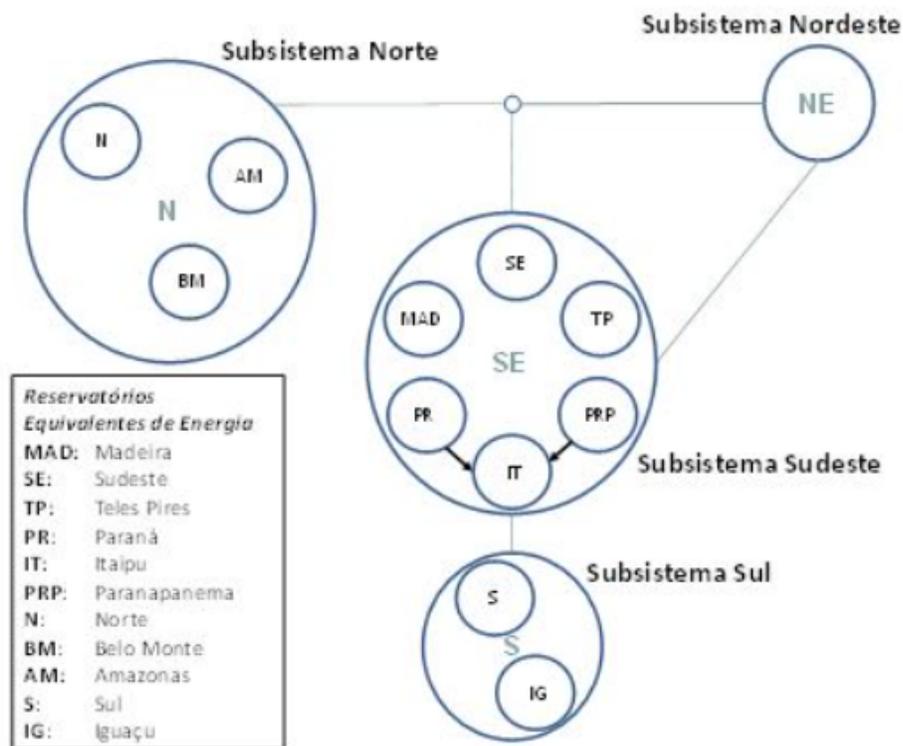
Mais especificamente, definem-se como planejamento da operação, os esforços para desenhar o comportamento do sistema em um horizonte de até cinco anos, promovendo um aproveitamento racional dos recursos. A programação da operação tem por objetivo estabelecer a operação de curto prazo do sistema hidrotérmico, fornecendo sugestões de decisões operativas do sistema de geração que sejam factíveis dado um sistema de transmissão, e que respeitem as metas estabelecidas no planejamento da operação (Souza 2014).

A solução destes modelos é efetuada através de algoritmos de Programação Dinâmica Dual Estocástica (PDDE) (Pereira e Pinto, 1991). Essa metodologia faz uso da decomposição de Benders (Benders, 1962), buscando encontrar estratégias ótimas para a operação dos subsistemas interligados. O regime de aflúncias é tratado por um modelo auto-regressivo periódico (McLeod, 1994). Esta estratégia realiza uma amostragem aleatória de diversos cenários gerados através do modelo auto-regressivo periódico, e realiza uma série de iterações, calculando as funções de custo futuro através da aproximação linear por partes de N retas tangentes, sendo N o número de iterações completas do modelo.

Objetivando reduzir o esforço computacional exigido pelos modelos, para o planejamento da operação a médio prazo, executada pelo modelo NEWAVE, é utilizada a agregação das usinas em reservatórios equivalentes de energia (REE) e submercados (ou subsistemas).

Cada subsistema do SIN contém um conjunto de REE's, e atualmente a modelagem do planejamento conta com 12 REE's divididos em 4 submercados de energia (Sudeste e Centro-Oeste, Sul, Nordeste e Norte) conforme mostra a figura 4.

Figura 4 - Estrutura de REE's e Subsistemas de energia



fonte: CEPEL , 2018b

Porém, para verificar se a política da operação obtida pelo modelo de decisão estratégica será viável, é necessário desagregar a solução obtida para os reservatórios equivalentes em usinas individualizadas, ou seja, averiguar se as usinas que compõem o sistema equivalente serão

capazes de atender ao montante de geração hidráulica determinado para o sistema pelo modelo de decisão da programação de curto prazo, dada pelo modelo DECOMP. Em seguida, ajusta-se a proposta de despacho semanal ao modelo diário, pelo uso do modelo DESSEM (Souza 2014).

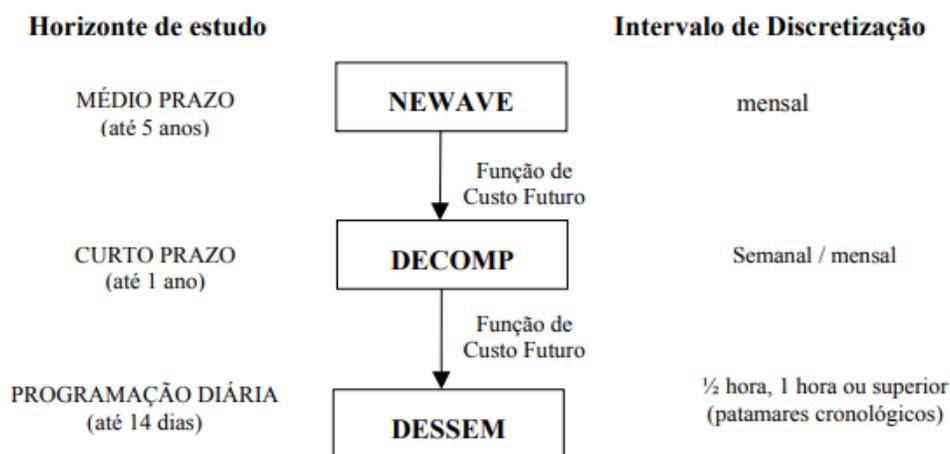
2 - MODELOS DE OTIMIZAÇÃO ENERGÉTICA

O modelo NEWAVE, como já citado, tem foco em otimizar o uso da água, dependendo fortemente da estocasticidade, a qual é fundamentalmente representada por séries sintéticas geradas a partir de modelos de séries temporais autoregressivos, inseridos no módulo de energias afluentes (Souza, 2014).

Para horizontes menores, é importante inserir outras variáveis de importância neste problema, a exemplo restrições de equipamentos de geração e transmissão. Assim, cabe ao modelo DECOMP minimizar o valor do custo total do sistema, dadas essas restrições.

Neste horizonte menor, tanto DECOMP quanto DESSEM representam as usinas do sistema de forma individualizada, expressando suas características e restrições operativas, conforme descrito em Souza (2014).

Figura 5 - Cadeia de modelos de otimização desenvolvido pelo CEPEL



fonte: CEPEL, 2011

Destaca-se que é importante observar a redução de incertezas com relação às afluências conforme o nível de detalhamento do sistema avança, justificando a estratégia de acoplar os modelos através da função de custo futuro, *FCF*. (Souza, 2014)

2.1 - Modelo NEWAVE

O NEWAVE é um modelo de otimização para planejamento de médio e longo prazo, com discretização mensal e representação por sistemas equivalentes. Seu objetivo é minimizar o valor esperado do custo de operação do sistema, para todo período de planejamento, sob alguns parâmetros de confiabilidade. (Souza, 2014)

Um dos principais resultados deste modelo, são as funções de custo futuro, com o objetivo de mensurar o impacto no uso dos reservatórios. Neste modelo, é possível representar os limites das interligações entre subsistemas do SIN, e tanto a carga quanto o custo de *déficit* são representados em patamares. Esses patamares são associados a valores de energia semelhantes em determinados períodos, que representam diferentes etapas de utilização do sistema, onde temos demandas de energia altas, médias e baixas (EPE, 2009).

2.2 - Modelo DECOMP

Assim como o NEWAVE o modelo DECOMP também busca obter uma operação ótima para o sistema Hidrotérmico, mas em um horizonte de tempo mais curto. O DECOMP recebe a função de custo obtida pelo NEWAVE juntamente com dados de carga, vazões, disponibilidades e limites de transmissão entre submercados. Assim, o DECOMP obtém o resultado otimizado para o planejamento do primeiro mês em base semanal (Souza, 2014).

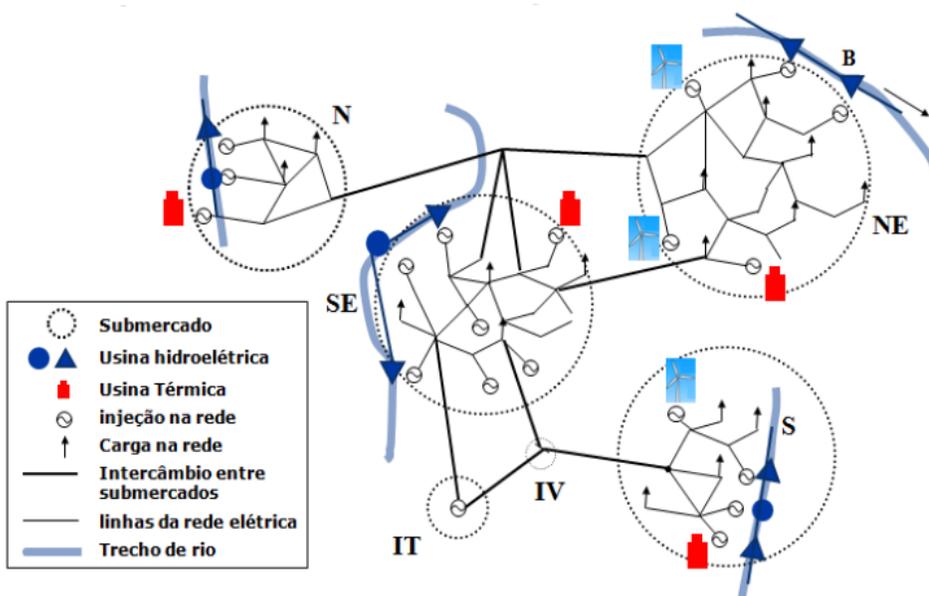
É importante observar que as vazões previstas e a aleatoriedade das vazões do restante período são obtidas por um árvore de possibilidades e um parque gerador individualizado.

Conforme observado, o DECOMP é utilizado para operação de curto prazo do SEB, e assim como o NEWAVE depende fortemente de modelos estocásticos. (CEPEL, 2018a)

2.3 - Modelo DESSEM

O terceiro modelo da cadeia, o modelo DESSEM, é desenvolvido pelo CEPEL desde 1998, e entrou em operação oficial a partir de janeiro de 2021. Este modelo se acopla com o modelo DECOMP através da função de custo futuro ao final de seu horizonte de planejamento, e tem foco na elaboração da programação diária da operação de sistemas hidrotérmicos, onde o despacho é efetuado com horizonte semi-horário, assim como a obtenção do preço horário de energia para o dia seguinte.

Figura 6 - Representação do sistema hidrotérmico pelo modelo DESSEM

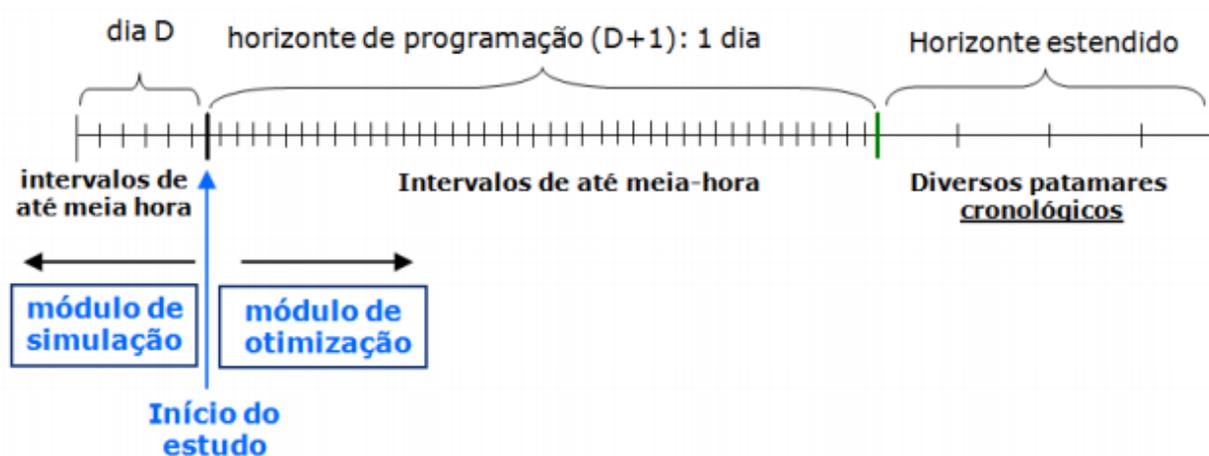


fonte: CEPEL, 2019

O modelo DESSEM representa as usinas a nível de suas unidades geradoras e considera uma modelagem detalhada da rede elétrica, através de uma modelagem DC com ou sem perdas, além da inclusão de possíveis restrições. Também são representadas as restrições de *unit commitment* das usinas termoeletricas e a operação das usinas de ciclo combinado, definindo suas rampas de tomada e alívio de carga. Em termos hidrológicos, o modelo representa de forma precisa o balanço dos reservatórios, através da inclusão dos tempos fixos de viagem ou de curvas de propagação das vazões (CEPEL, 2019).

Apesar do foco do modelo ser a programação diária, este dispõe de um máximo horizonte de estudo de até 14 dias, com a possibilidade de representar os intervalos de tempo em agregações: semi-horária, horária ou em termos de patamares cronológicos de duração variável.

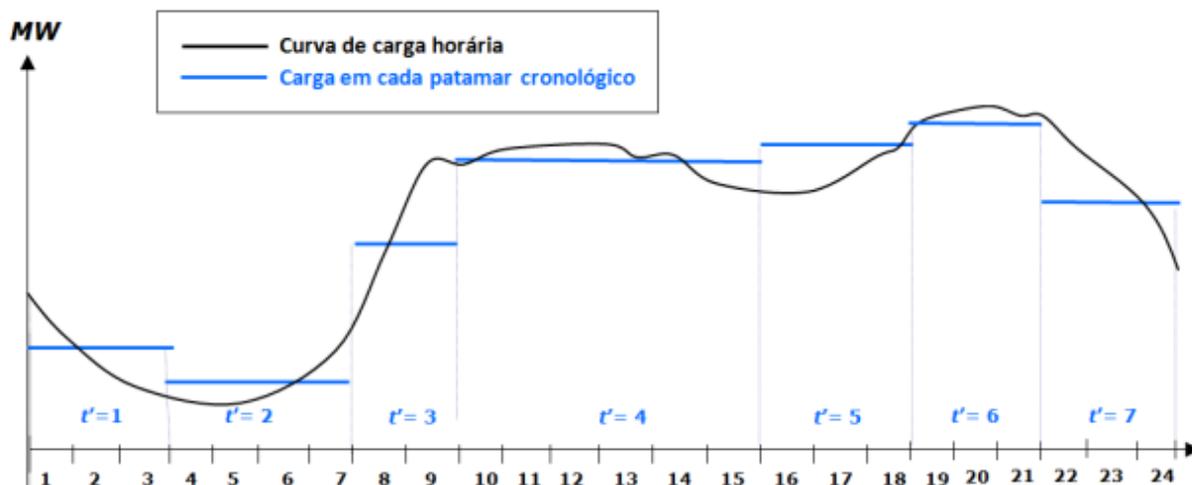
Figura 7 - Representação dos horizontes de estudo do modelo DESSEM.



Fonte: CEPEL, 2019

Esses patamares cronológicos seriam compostos pela média dos valores associados às horas que compõem cada patamar. Para o caso da carga, essa curva seria agregada em intervalos superiores a uma hora (CEPEL, 2019).

Figura 8 - Representação em patamares cronológicos do modelo DESSEM



fonte: CEPEL, 2019

É importante ressaltar, que esta representação em patamares está fortemente guiada pela curva de carga fornecida pelo modelo, porém não estão associadas, em termos de modelagem, aos patamares de carga adotados pelos modelos NEWAVE e DECOMP (CEPEL, 2019).

Neste contexto, a curva de carga fornecida pelo ONS ao modelo DESSEM é obtida através de processos de previsão de carga, associados especificamente à programação diária da operação eletroenergética.

3 - PREVISÃO DE CARGA PARA O MODELO DESSEM

A etapa de previsão de carga é uma peça fundamental na programação diária da operação. Estimar de forma precisa a demanda de energia ao longo do dia garante um despacho otimizado de usinas, de forma a atender essa demanda a determinados critérios e margens de segurança.

Dentre os modelos de otimização, o foco deste trabalho está em abordar a etapa de previsão de carga para o modelo DESSEM, embora existam etapas semelhantes para os demais modelos, NEWAVE e DECOMP. Mais adiante, será discutida a metodologia proposta para elaboração desta curva de carga.

De antemão, é necessário conhecer a metodologia padrão utilizada pelo ONS para elaboração desta curva, um processo detalhado e que adota diversos modelos para estimar a curva de carga para o dia seguinte, a chamada previsão de carga para a programação diária da operação (ONS, 2019).

3.1 - Definição de Carga pelo ONS

O objetivo do sistema elétrico como um todo é atender a demanda dos consumidores conectados à rede. Todo o trabalho de otimização e planejamento da operação visa garantir o fornecimento contínuo de energia, ao menor custo, seguindo determinados critérios de segurança.

Os consumidores conectados à rede geram uma demanda de potência, que deve ser atendida pelos diferentes recursos de geração disponíveis. Entretanto, deve-se considerar as perdas técnicas e não-técnicas associadas ao sistema de transmissão e distribuição. Dessa forma a Carga é definida como a demanda de potência mais as perdas do sistema de transmissão e distribuição (ONS, 2019)

$$Carga = Demanda + Perdas \quad (3.1)$$

3.2 - Previsão de carga para programação diária da operação eletroenergética (PDE)

Conforme ONS (2019), as previsões são elaboradas um dia antes da data da operação, em intervalos de 30 minutos, com um horizonte que varia entre 2 a 7 dias à frente. Essas previsões normalmente são para dias úteis ou finais de semana precedidos ou seguidos de feriados e dias de pontes. Previsões para datas especiais, como jogos do Brasil em copas do mundo e eleições, são elaboradas com maior antecipação porque consideram comportamentos de consumo de energia elétrica fora dos padrões normalmente observados, exigindo medidas operativas especiais. No ONS, o processo responsável pelas previsões de carga para PDE é chamado de Consolidação da previsão da carga diária e se divide em:

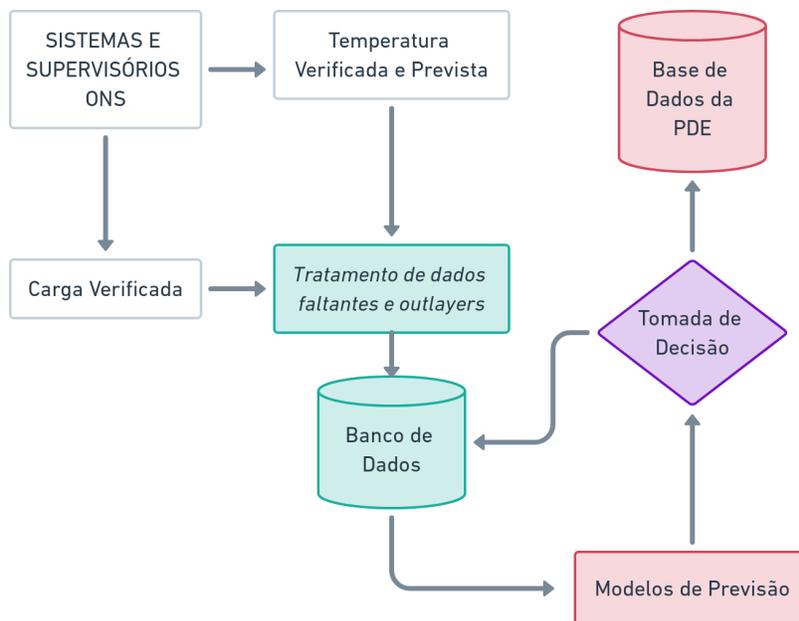
- a) Leitura das séries temporais de cargas verificadas por áreas e subsistemas em intervalos no dia corrente até o momento da previsão.
- b) Leituras das séries de temperaturas verificadas e previstas, em intervalos horários para o dia da operação. As previsões de temperatura são obtidas através dos modelos de previsão do CPTEC para as capitais do país.
- c) Geração de cenários de carga através dos modelos de previsão de carga de curto prazo considerando os principais cenários meteorológicos.
- d) Tomada de decisão.
- e) Envio das previsões para o processo da PDE.

A etapa mais crítica é a tomada de decisão, pois ela consiste na elaboração de uma curva de carga diária para cada área e subsistema para o dia da operação. O objetivo desse processo é reduzir os desvios entre a previsão de carga nas áreas e subsistemas monitorados pelo operador a fim de garantir a segurança eletroenergética do sistema e a modicidade tarifária. (ONS 2019).

Os dados faltantes e outliers são tratados utilizando modelos estatísticos, desenvolvidos pela UFRJ, CEPEL e ONS. A geração de cenários é feita através dos modelos de previsão de

curto prazo, desenvolvidos pelo CEPEL e ONS, além do ANNSTLF (*Artificial Neural Network Short-term Load Forecaster*) (ONS 2019).

Figura 9 – Visão geral do Processo de Consolidação da previsão de Carga Diária



Fonte: Adaptado de ONS, 2019

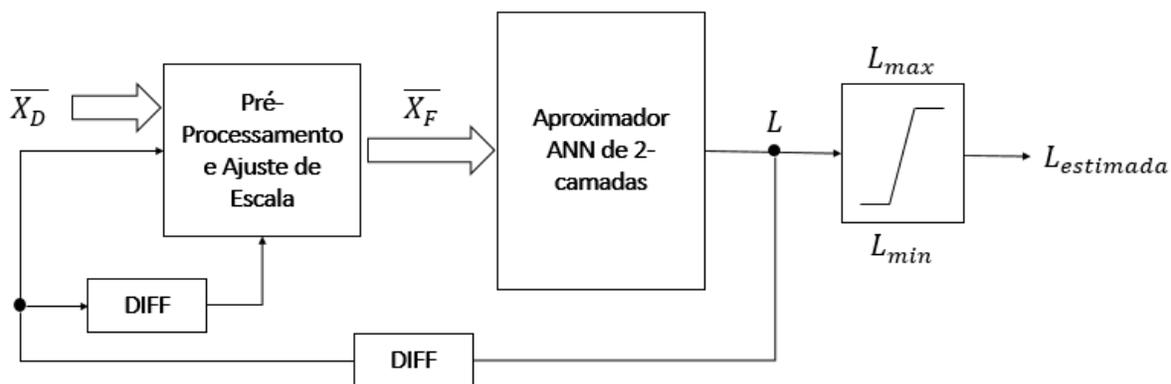
Por fim, todas as previsões geradas pelos modelos que serviram de base para a tomada de decisão, bem como as previsões finais para as áreas e subsistemas que foram enviadas para o processo da PDE são armazenadas na base de dados para fins de análise de desvios, aprimoramento dos modelos e auditoria (ONS 2019).

3.2 - Modelos de Previsão de Carga

Os atuais modelos de previsão de carga utilizados pelo ONS (2019), para o processo de consolidação de carga diária são: ANNSTLF, CEPEL e ONS.

O ANNSTLF é um modelo desenvolvido pelo EPRI (*Electric Power Research Institute*) baseado em redes neurais do tipo *feedforward*. Este modelo é composto de duas redes neurais: B(R)LF (*Base or regular load forecaster*) e C(D)LF (*Change or Delta Load Forecaster*) além do modulo RLS (*Recursive Least Squares*), responsável pela combinação linear das saídas das redes anteriores e pelo resultado final do modelo. Este modelo encontra-se embarcado num software fechado e licenciado para uso do ONS. (ONS 2019).

Figura 10 - Estrutura da rede ANNSTLF



\bar{X}_D : Vetor Descritivo [variáveis de entrada]

\bar{X}_F : Vetor de Previsão [\bar{X}_D , DIFF(L), DIFF(DIFF(L))]

fonte: Adaptado de SHARAF et. al., 1993

O modelo do CEPEL, é baseado em SVR (*Support Vector Regression*) e foi desenvolvido em linguagem R (R é uma linguagem de programação e um ambiente de software livre para computação estatística e gráficos e é apoiado pela R *foundation for Statistical Computing*). Os dados do modelo são inseridos via “*decks*” de entrada e são lidos no formato de *dataframe*, i.e, um tipo de dados do R (ONS 2019).

O modelo do ONS é um combinação de modelos, dentre eles: Regressão Dinâmica, Redes Neurais, SVR, Regressão Quantílica e Similaridade de Temperatura, todos desenvolvidos em linguagem R.

A tomada de decisão é efetuada através de avaliação de analistas especialistas para compor a curva de previsão de carga diária.

A consolidação da carga diária, tal como vem sendo feito no ONS, tem garantido a eficiência e a eficácia dos processos da carga junto às áreas clientes, têm reduzido riscos elétricos e energéticos na operação em tempo real e tem contribuído de forma efetiva para a modicidade tarifária.

Contudo, a dinâmica do mercado de energia elétrica exigirá mudanças constantes nos processos da carga, sobretudo na aquisição de dados e no aprimoramento dos métodos analíticos. O escopo deste trabalho, envolve propor uma metodologia, que possa ser utilizada de forma isolada, ou combinada com os demais modelos do ONS para efetuar estudos de previsão de carga.

Neste contexto, a implementação deste modelo se dará através do uso de redes neurais recorrentes, projetadas para efetuar regressão da série temporal de carga.

4 - REDES NEURAIAS ARTIFICIAIS

As redes neurais artificiais, ou simplesmente redes neurais, são modelos de algoritmo de aprendizado de máquina, baseados no conceito de inteligência artificial. Estes modelos utilizam de grandes quantidades de dados, onde são treinados para estabelecer as correlações e interações entre as diferentes variáveis de determinado problema.

Segundo Haykin (1999), a motivação para estudo e implementação de redes neurais se baseiam no reconhecimento de que nosso cérebro humano processa informações de uma forma inteiramente diferente que um computador tradicional. Complexo, não-linear e de processamento paralelo, nosso cérebro é capaz de realizar tarefas (como perceber padrões e realizar controle motor) com extrema agilidade, muito mais rápido que os computadores digitais existentes, e para isso utiliza de uma quantidade muito grande de células para efetuar este procedimento, os chamados neurônios.

As redes neurais, de forma análoga, possuem os chamados neurônios artificiais, que são as unidades de processamento associadas, os chamados neurônios artificiais, e se assemelham ao cérebro convencional em dois aspectos:

- O conhecimento é adquirido pela rede a partir de seu ambiente por um processo de aprendizagem.
- Forças de conexão entre os neurônios, chamados de pesos sinápticos, são utilizados para armazenar o conhecimento adquirido.

Esse processo de aprendizagem se dá através de um algoritmo de aprendizagem, ao qual modifica os pesos da rede de forma ordenada para alcançar o objetivo desejado do projeto. (Haykin, 1999).

A escolha do modelo de redes neurais para executar a metodologia de previsão de carga, se baseia nos benefícios de adotar esta estratégia. Dentre as mais relevantes para esta tarefa em questão destaca-se:

Não-Linearidade. As conexões entre os neurônios da rede podem ser lineares, mas também podem ser adotadas relações não-lineares. Essas propriedades são importantes para correlacionar variáveis não linearmente relacionadas. (Haykin, 1999).

Adaptabilidade. As redes neurais são capazes de adaptar seus pesos a modificações no ambiente. Em particular, uma rede treinada para operar em um ambiente específico pode ser retreinada para lidar com pequenas modificações nas condições operativas no ambiente (Haykin, 1999).

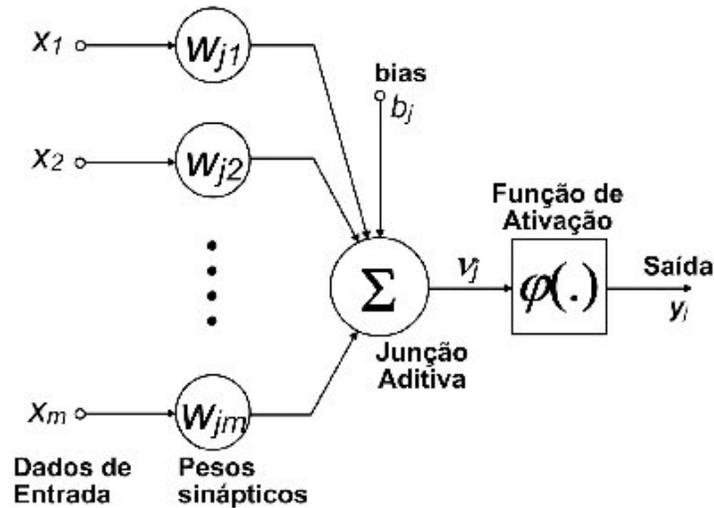
Esses parâmetros são intrínsecos ao processo de previsão de carga, como as relações não-lineares entre suas principais variáveis e a constante evolução do SEB, exigindo certa adaptabilidade por parte do modelo utilizado.

4.1 - A estrutura do neurônio artificial.

A estrutura principal de uma rede neural são seus neurônios. Eles são responsáveis pelo processamento das informações que chegam ao modelo, e são essenciais para a operação da rede neural.

Dentre elementos básicos da rede, descritos segundo Haykin (1999), temos o conjunto dos m sinais de entrada, representados pelo vetor $\bar{x} = [x_1, x_2, \dots, x_m]$. Esses sinais se relacionam com o neurônio k através de seu conjunto de pesos sinápticos, representados pela matriz de pesos $\bar{w} = [w_{k1}, w_{k2}, \dots, w_{km}]$. Dessa forma o peso sináptico w_{k1} relaciona o elemento x_1 do vetor de sinais de entrada com o neurônio k .

Figura 11 - Modelo de neurônio artificial.



fonte: Adaptado de Haykin, 1999

Estes pesos podem assumir valores positivos ou negativos, e no neurônio os sinais de entrada são combinados através de um somador, ponderados por seus respectivos pesos. Até este ponto, trata-se de uma simples transformação linear. Por fim, uma função de ativação $\varphi(\cdot)$ irá restringir a amplitude de saída de um neurônio a um valor finito. O bias b_k pode ser inserido para aumentar ou diminuir a entrada líquida da função de ativação (Haykin, 1999).

O sinal de bias, agregado ao valor do somatório de sinais de entrada v_k , pode ser representado pela equação 4.1, onde é chamado de campo local induzido (Haykin, 1999).

$$v_k = u_k + b \tag{4.1}$$

Em termos matemáticos, podemos expressar o funcionamento de um neurônio artificial através do par de equações 4.2 e 4.3, sendo y_k o sinal de saída do neurônio (Haykin, 1999):

$$u_k = \sum_{j=1}^m w_{kj} x_j \tag{4.2}$$

$$y_k = \varphi(v_k) \tag{4.3}$$

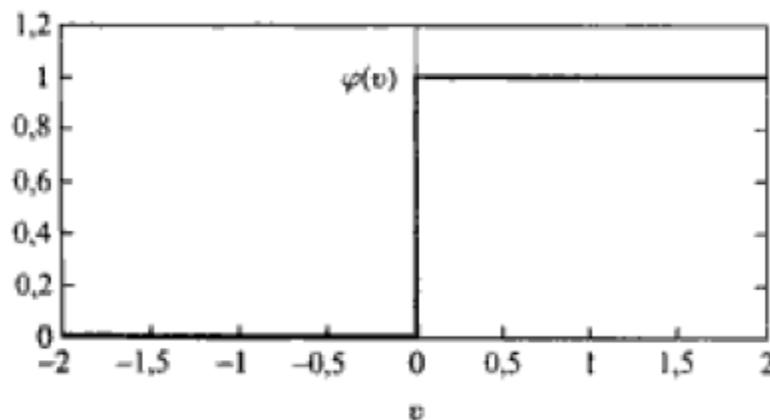
4.2 - Funções de ativação

A função de ativação $\varphi(u)$ define a saída do neurônio em função do campo local induzido v_k . Haykin (1999) cita alguns exemplos de funções de ativação, como funções de limiar, linear por partes e Sigmóide.

A Função de limiar (*Heaviside*), a exemplo, pode ser definida pela seguinte expressão:

$$\begin{aligned} \varphi(v) &= 1; \text{ se } v \geq 0 \\ \varphi(v) &= -1; \text{ se } v < 0 \end{aligned} \tag{4.4}$$

Figura 12 - Função de ativação Limiar (*Heaviside*)



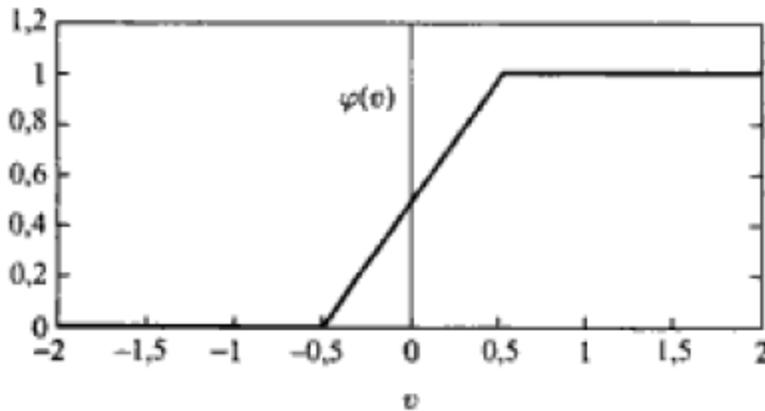
fonte: Haykin, 1999

Este modelo é referido na literatura como o modelo de McCulloch-Pitts, em reconhecimento aos trabalhos de McCulloch e Pitts (1943). Neste modelo, caso a saída do campo local induzido seja negativa, a ativação é automaticamente nula

A Função linear por partes, é descrita por uma equação de forma:

$$\begin{aligned} \varphi(v) &= 1; \text{ se } v \geq 1/2 \\ \varphi(v) &= v; \text{ se } v > -1/2 \\ \varphi(v) &= 0; \text{ se } v < -1/2 \end{aligned} \tag{4.5}$$

Figura 13 - Função linear por partes



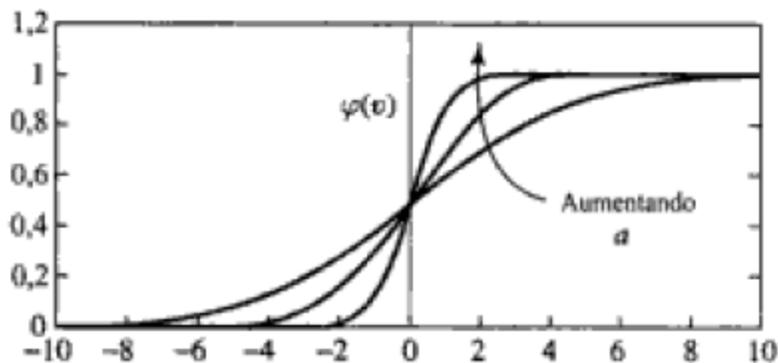
Fonte: Haykin, 1999.

Esta forma de função de ativação pode ser vista como uma aproximação de um amplificador não-linear, com duas regiões de saturação.

A função sigmóide possui um gráfico em forma de 's', e é uma das mais comuns funções de ativação utilizadas para construção de redes neurais. Trata-se de uma função estritamente crescente, e um exemplo de de função sigmóide é a função logística.

$$\varphi(v) = 1/(1 + e^{-av}) \tag{4.6}$$

Figura 14- Função sigmóide



Fonte: Haykin, 1999.

O parâmetro a ajusta a inclinação da função, cuja variação produz curvas com diferentes inclinações. No limite, com o parâmetro de inclinação tendendo a infinito, a função sigmóide se torna a função de limiar.

As funções mencionadas anteriormente possuem intervalo de ativação entre 0 e +1. Em certos casos é desejável, ou mesmo necessário, que a ativação forneça também valores negativos. Neste caso, pode-se utilizar a função tangente hiperbólica.

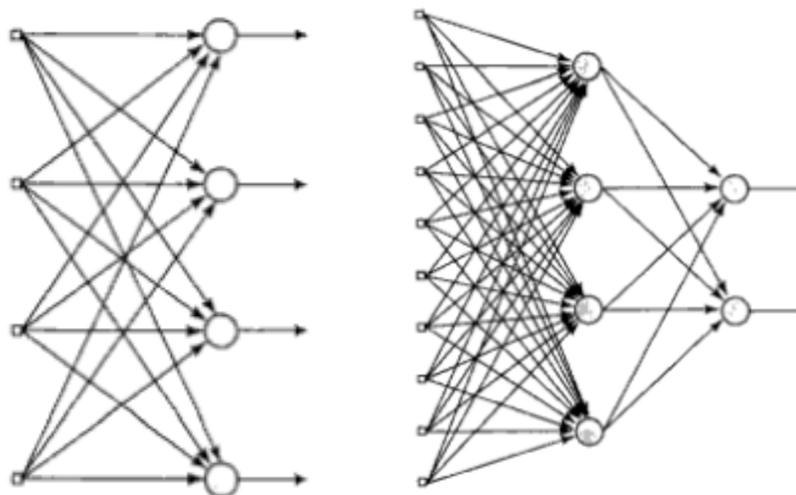
$$\varphi(v) = \tanh(v) \tag{4.7}$$

A função tangente hiperbólica possui forma correspondente a de uma função sigmóide, entretanto prescreve um saída anti-simétrica entre -1 e +1. Esta função traz diversos benefícios analíticos em certos casos, e em projetos de redes neurais recorrentes, esta função se faz muito eficaz.

4.3 - Redes neurais de multicamada

Em uma topologia de rede neural, os neurônios são distribuídos em forma de camadas. Em redes neurais mais simples, há apenas uma única camada de processamento, dessa forma existe uma camada de entrada e uma camada de saída. Uma segunda topologia, seria a adição de mais camadas de processamento, chamadas de camadas ocultas (*hidden layers*), formadas por neurônios ocultos (*hidden units*).

Figura 15 - Redes neurais de camada única e multicamada



Fonte: Haykin, 1999.

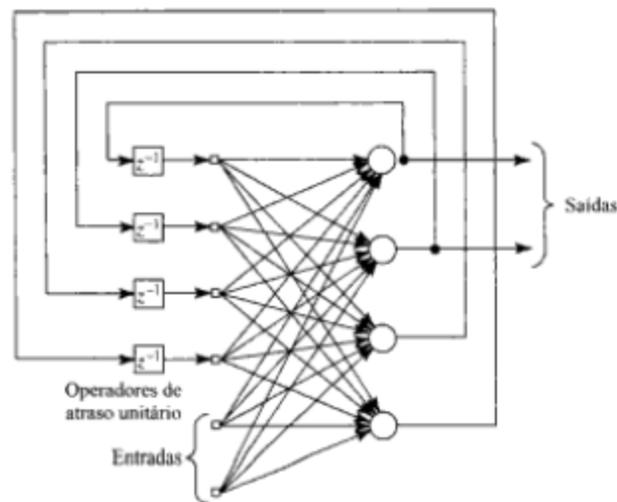
Essa técnica permite extrair estatísticas de ordem mais elevada dos dados de entrada. Em geral, é uma característica muito valiosa quando se tratam de dados com um volume elevado de sinais de entrada (Churchland e Sejnowski, 1992).

Em termos de fluxo da informação, dizemos que se tratam de modelos de rede de alimentação adiante (*feedforward*), uma vez que não há nenhum laço de realimentação nestes modelos.

4.4 - Redes Neurais Recorrentes

Redes neurais recorrentes se diferenciam das redes neurais de realimentação adiante normais pela existência de pelo menos um laço de realimentação.

Figura 16 - Rede recorrente com neurônios ocultos



fonte: Haykin, 1999

Embora as diferenças entre uma rede neural padrão e uma rede recorrente possam parecer triviais, as implicações do aprendizado por sequência são vastas. Uma rede neural multicamada consegue mapear somente as interações de entrada e saída, em adição, uma rede recorrente pode em princípio mapear sobre o todo o histórico de entradas anteriores para cada saída. O ponto chave é que os laços de realimentação, também chamados de conexões recorrentes, adicionam

certo grau de memória sobre as entradas, permitindo que sejam mantidas nos estados internos da rede (internal states) e influenciando nas saídas (Graves, 2012).

A etapa de alimentação adiante é semelhante a de uma rede multicamadas, com a exceção de que a ativação irá conter elementos dos intervalos de tempo anteriores. Considerando uma sequência de entrada \bar{x} de tamanho I e duração $0 \leq t \leq T$ em um neurônio k , onde x_i^t é o valor de entrada i no intervalo de tempo t . Para esta rede temos uma camada oculta de H unidades ocultas, e uma camada de K unidades de saída. Como apresentado por Graves (2012), para a camada oculta, o campo local induzido pela entrada na unidade oculta h , v_h^t pode ser definido por:

$$v_h^t = \sum_{i=1}^I w_{ih} x_i^t + \sum_{h'=1}^H w_{hh'} b_{h'}^{t-1} \quad (4.8)$$

Nesta equação, podemos observar o termo de interação com o histórico de ativações, sendo representado pelo segundo somatório da equação 4.8, onde temos os conjuntos de pesos. Em termos de ativação, podemos representá-la em termos do campo local da unidade h , onde $\varphi_h(v)$, é a função de ativação da unidade oculta h :

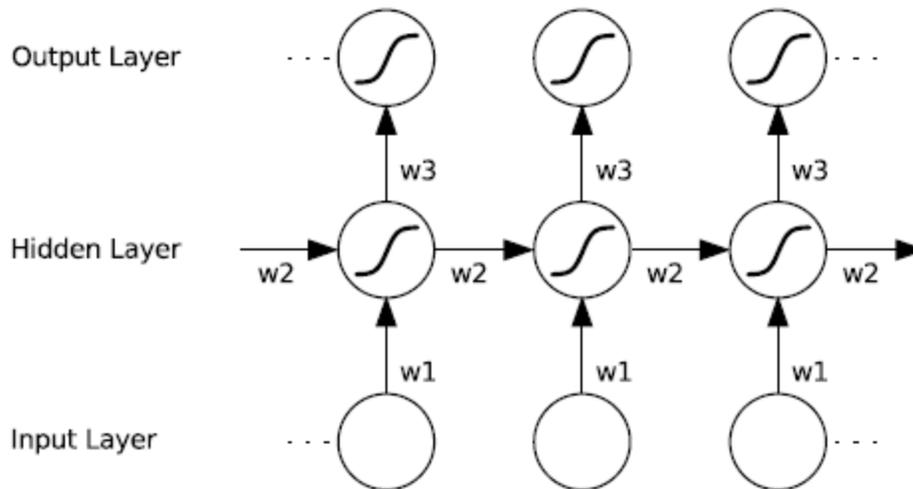
$$b_h^t = \varphi_h(v_h^t) \quad (4.9)$$

A sequência completa de ativações ocultas pode ser calculada recursivamente a partir de $t = 0$, incrementando t em cada intervalo de tempo. As saídas $\bar{y} = [y_1, y_2, \dots, y_K]$ podem ser computadas, através do conjunto de pesos $\bar{w}_k = [w_{1k}, w_{2k}, \dots, w_{Hk}]$ da unidade de saída k :

$$y_k^t = \sum_{h=1}^H w_{hk} b_h^t \tag{4.10}$$

Uma forma mais interessante para representar uma rede recorrente, é através de seu grafo remodelado, contendo uma versão ‘desdobrada’ da rede juntamente com suas sequências de entrada. Essa visualização permite observar melhor o repasse de informações ao longo dos intervalos de tempo, e generalizar melhor redes com dependências mais complexas (Graves, 2012).

Figura 17 - Versão desdobrada do grafo computacional de uma RNN.



Fonte: Graves, 2012.

5 - TREINAMENTO DE REDES NEURAIAS

A rotina de aprendizado da rede neural é executada através de um algoritmo de treinamento correspondente. Este processo pode ser entendido como um processo de otimização, onde os pesos da rede são ajustados a uma melhor combinação, que minimiza uma determinada função de custo. Para esta tarefa, é amplamente utilizado o algoritmo de retropropagação (Werbos, 1974).

5.1 - Função de Custo (Loss Function)

A função de custo, ou função objetivo, é a métrica principal do processo de treinamento de uma rede neural. O processo de ajuste dos pesos é executado através do erro obtido na saída da rede, em relação ao valor de referência utilizado para treinamento. Os valores de erro obtidos ao longo do processo de treinamento compõem a função de custo.

O objetivo do processo, é minimizar os erros, obtendo uma resposta mais próxima possível do valor esperado na saída da rede. Intrinsecamente, minimizar os erros envolve minimizar a função de custo.

A definição matemática da função de custo depende da tarefa a ser realizada pela rede neural. Se tratando de um problema de classificação, a função de custo pode ser expressa, por exemplo, em função das probabilidades de cada classe associada ao problema (Bishop, 1995).

Entretanto, neste trabalho, o processo de obtenção da curva de carga prevista envolve um processo de regressão. Para tal é necessário medir a diferença entre os valores previstos e os valores esperados no processo. Dessa forma, métricas estatísticas como MAPE (Mean Absolute

Percentage Error) e MSE (Mean Squared Error) podem ser utilizadas, pois ambas geram funções de custo diferenciáveis.

As funções de custo $L(y, \hat{y})$, sendo y o valor esperado e \hat{y} o valor previsto, podem ser expressas por:

$$L(y, \hat{y})_{MAPE} = \frac{100}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|) \quad (5.1)$$

$$L(y, \hat{y})_{MSE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2 \quad (5.2)$$

Ambos oferecem boas formas para representar o erro de saída da rede, entretanto a métrica de MSE é mais sensibilizada pela escala dos dados, penalizando muito os erros em valores grandes não normalizados, enquanto a MAPE não oferece bons resultados em valores de escala pequena, resultando em erros excessivamente grandes com poucas variações.

5.2 - Algoritmo de Retropropagação (*backpropagation*)

Após computar as entradas pela etapa de alimentação adiante, a rede neural realiza a etapa de retropropagação, reajustando os pesos seguindo o algoritmo de retropropagação (Werbos, 1974).

O algoritmo é simplesmente a aplicação da regra da cadeia para derivadas parciais. O primeiro passo é calcular as derivadas da função de custo $L(y, \hat{y})$ em respeito às unidades de saída. Após esta etapa, calcula-se as demais derivadas parciais utilizando a regra da cadeia. A equação 5.4 mostra por exemplo obtenção do delta de erro na camada de saída δ_y .

$$\delta_y = dL(y, \hat{y})/dy \quad (5.4)$$

Através desse método é possível propagar os erros a todos os elementos, e conseqüentemente ajustar os pesos da rede. O ajuste de pesos é executado pela regra delta (Widrow e Hoff, 1960), calculando o valor do ajuste de peso Δw_{kj} em função dos diferenciais de erro obtidos na etapa de retropropagação:

$$\Delta w_{ij} = \eta \delta_{ij} x_i \quad (5.5)$$

Na equação 5.5 temos os termos de diferenciais de erro retropropagado δ_{ij} obtidos para os i pesos das j camadas da rede, multiplicados por sua respectiva entrada e uma constante positiva η denominada de taxa de aprendizagem (*learning rate*) (Haykin, 2005). Ao calcular os valores de ajuste, o respectivo peso é atualizado, e preparado para a próxima iteração:

$$w_{ij(n+1)} = w_{ij(n)} - \Delta w_{ij(n)} \quad (5.6)$$

Este processo se repete até que a função de custo assuma um valor de erro inferior a uma tolerância arbitrária, ou quando o algoritmo atinge um número máximo de iterações.

$$L(y, \hat{y}) \leq Tol_{min} \quad (5.7)$$

$$n_{iterações} \leq N_{máx} \quad (5.8)$$

Se tratando de redes recorrentes, a modelagem não se difere ao extremo, apenas devemos abstrair do fato de que assim como as ativações da rede recorrente dependem de entradas passadas na etapa adiante, os deltas sofrem influência das iterações ao longo dos períodos de tempo (Graves, 2012).

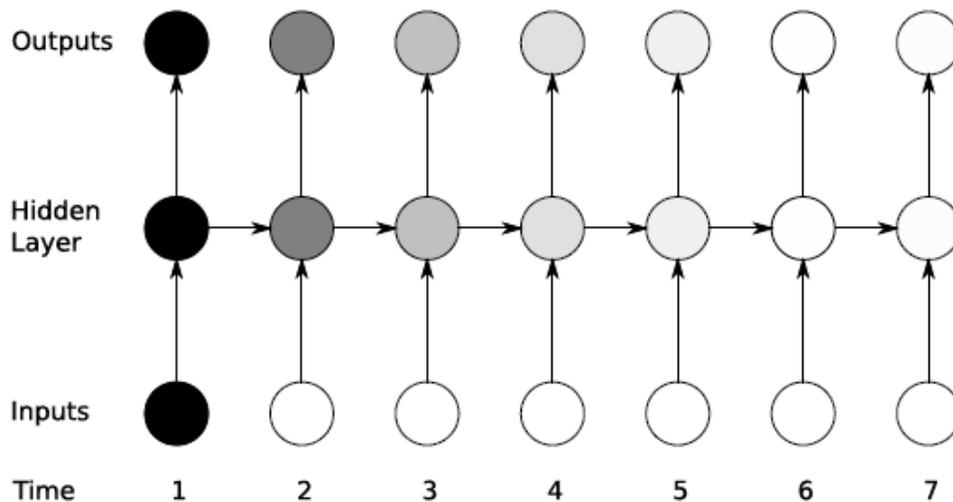
Assim, a estratégia mais eficiente é ‘desdobrar’ a rede recorrente e aplicar etapa de retropropagação através dos períodos de tempo da rede de forma inversa, iniciando em $t = T_e$ decrementado os valores. Essa técnica é conhecida como BPTT - *backpropagation through time*. (Williams e Zipser, 1995).

6 - LONG SHORT TERM MEMORY (LSTM)

O benefício de utilizar redes recorrentes, como já destacado, se baseia em sua capacidade de aprender características provenientes do histórico de entradas e saídas da rede. Infelizmente, para redes neurais recorrentes comuns, essa capacidade se mostra limitada. Isso se deve ao fato que o contexto fornecido por uma entrada pode ser dissolvido no repasse a camadas ocultas, decaindo exponencialmente (ou explodindo em alguns casos) à medida que circula pelas conexões recorrentes. Este problema é referido na literatura como problema do desaparecimento do gradiente (vanishing gradient problem) (Bengio et al., 1994).

Na figura 17, há uma representação deste problema, onde o contexto tende a ‘sumir’ (quanto mais escuro maior a influência do contexto) com o passar dos intervalos de tempo.

Figura 18- O problema do desaparecimento do gradiente



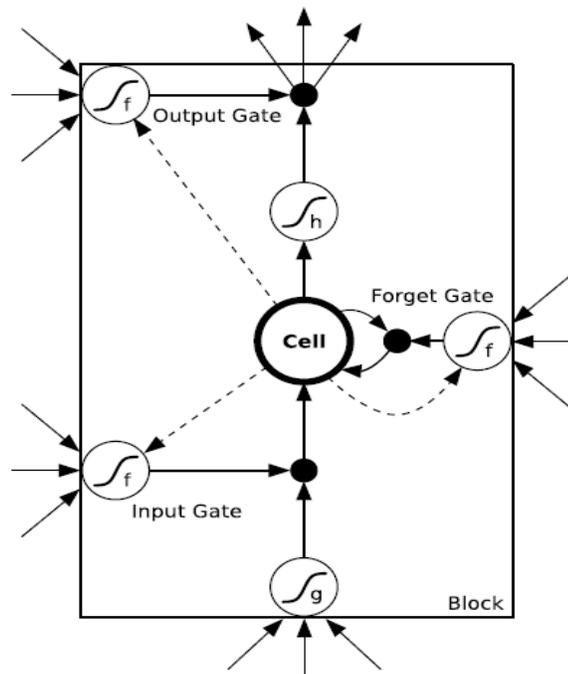
Fonte: Graves, 2012.

No problema de séries temporais, como no caso da previsão de carga, é desejável manter as informações retidas por mais tempo nas unidades da rede. Para isso, será adotada a estratégia de utilizar a arquitetura de memórias LSTM (*long short-term memory*) (Hochreiter e Schmidhuber, 1997).

6.1 - Estrutura geral de redes LSTM

A estrutura LSTM consiste em um conjunto de sub-redes recorrentemente conectadas, conhecidas como blocos de memória. Estes blocos podem ser vistos como uma versão diferenciável de chips de memória em computadores digitais (Graves, 2012).

Figura 19 - Estrutura do Bloco de memória LSTM

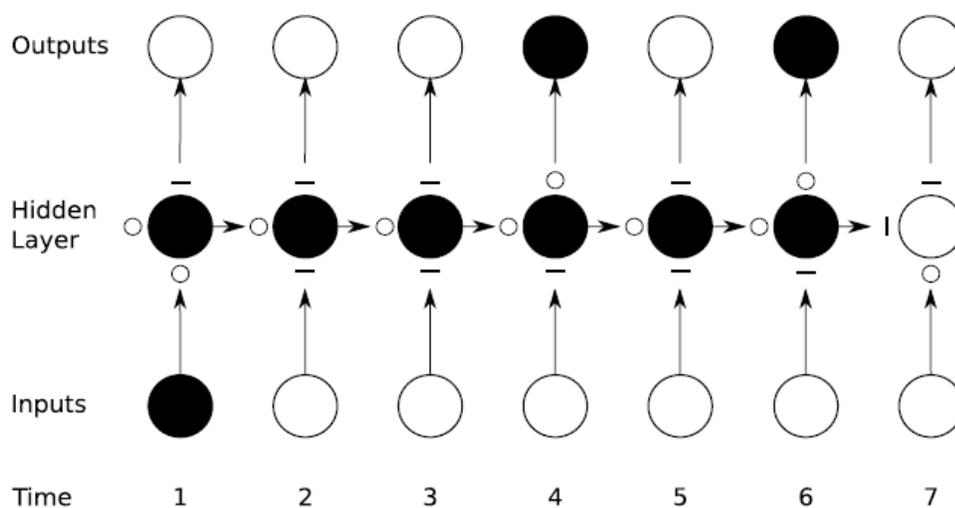


Fonte: Graves, 2012.

Cada bloco contém uma ou mais células de memória auto-conectadas, e três multiplicadores de entrada - input gate (g), output gate (q) e forget gate (f). Esses multiplicadores fazem o papel análogo às funções de leitura, escrita e reset nos blocos (Graves, 2012).

A função de ativação das portas (representada pela letra f na figura 18) é geralmente uma sigmóide logística, dado que as ativações das mesmas se dá entre 0 (porta fechada) e 1 (porta aberta), ao passo que as funções de ativação da entrada e saída da célula (representadas na figura 18 pelas letras g e h) podem variar, desde as funções tanh e sigmóide (Graves, 2012).

Figura 20 - Preservação do gradiente pela adição de LSTM



fonte: Graves, 2012

O sombreamento na figura 19 mostra a sensibilidade das entradas, os nós escuros representam alta sensibilidade e os nós brancos sensibilidade nula. Os estados das portas de entrada, saída e esquecimento são mostrados pelos símbolos abaixo, acima e à esquerda dos nós, respectivamente. Para simplicidade, as portas possuem apenas estados aberto (O) e fechado (-).

A célula de memória retém informações da primeira entrada enquanto a porta de esquecimento permanece aberta e a porta de entrada estiver fechada, permitindo sua propagação

ao longo dos períodos de tempo. A sensibilidade da camada de saída pode ser alterada pela porta de saída, sem influenciar no estado da célula (Graves, 2012).

6.2 - Equações das estruturas LSTM

Um importante componente de uma rede LSTM é sua unidade de estado interno s_i^t . Esta unidade é controlada pela porta de esquecimento f_i^t , para um intervalo de tempo t em uma célula i , que ajusta este peso a um valor entre 0 e 1, dado sua ativação sigmóide $\sigma(\cdot)$. Conforme definido em Goodfellow et al. (2016), a porta de descrita pelo equacionamento:

$$f_i^t = \sigma(b_i^f + \sum_j U_{ij}^f x_j^t + \sum_j W_{ij}^f h_j^{(t-1)}) \quad (6.1)$$

Na equação 6.1, x^t é o vetor de entrada atual e h^t o vetor da camada oculta atual, contendo a saída de todas as células LSTM, e b^f, U^f, W^f são respectivamente, o *bias*, os pesos da entrada e os pesos recorrentes para a porta de esquecimento. O estado interno é atualizado conforme a equação 6.2.

$$s_i^t = f_i^t s_i^{t-1} + g_i^t \sigma(b_i + \sum_j U_{ij} x_j^t + \sum_j W_{ij} h_j^{(t-1)}) \quad (6.2)$$

Onde b, U, W respectivamente denotam o *bias*, os pesos de entrada e pesos recorrentes da célula LSTM. A porta de entrada g_i^t é computada de forma similar a porta de esquecimento, mas contendo seus próprios parâmetros (Goodfellow et al., 2016).

$$g_i^t = \sigma(b_i^g + \sum_j U_{ij}^g x_j^t + \sum_j W_{ij}^g h_j^{(t-1)}) \quad (6.3)$$

A saída h_i^t da célula LSTM pode ser desligada, via porta de saída q_i^t , que também usa ativação sigmóide, e possui seus próprios parâmetros b^o, U^o, W^o (Goodfellow et al., 2016).

$$h_i^t = \tanh(s_i^t) q_i^t \quad (6.4)$$

$$q_i^t = \sigma(b_i^o + \sum_j U_{ij}^o x_j^t + \sum_j W_{ij}^o h_j^{(t-1)}) \quad (6.5)$$

6.3 - Cálculo do Gradiente

Assim como observado nas topologias anteriores, LSTM é uma é uma aproximação de uma função diferenciável, e pode ser treinada utilizando descida do gradiente. O algoritmo original de treinamento para LSTM (Hochreiter e Schmidhuber, 1997) utiliza um erro de gradiente aproximado com uma combinação de RTRL - *Real Time Recurrent Learning* (Robinson e Fallside, 1987) e BPTT - *Backpropagation Through Time* (William e Zipster, 1995).

A BPTT é truncada após um período de tempo, devido a percepção que as dependências de longo prazo podem ser tratadas com os blocos de memória, e não via fluxo de ativações em torno das conexões recorrentes. Esta é uma propriedade importante para tarefas de controle contínuo e previsão de séries temporais (Graves, 2012).

7 - DADOS DE ENTRADA DO MODELO DE REDE NEURAL LSTM

Para elaboração do modelo LSTM foi necessário coletar dados de entrada para serem utilizados no treinamento do modelo de redes neurais. Os conjuntos de dados são basicamente *decks* de entrada utilizados pelo modelo ONS de previsão de carga para o modelo DESSEM, o *prevCargaDessem*, e são disponibilizados diariamente pelo próprio ONS, através do portal SINtegre (ONS, 2021).

7.1 - Principais dados de entrada

Os dados fornecidos são históricos de variáveis importantes para previsão da curva de carga do modelo LSTM. Os arquivos são fornecidos em formato csv (*comma-separated values*) para facilitar a leitura pelo programa, compilados em um arquivo compactado (.zip), e separados em pastas por subsistema energético (Sudeste/Centro-Oeste, Sul, Nordeste e Norte).

Quadro 1 - Estrutura de arquivos de entrada do modelo *PrevCargaDessem*

| Nome do Arquivo | Conteúdo do Arquivo |
|-----------------------------|--|
| XX_AAAA-MM-DD_CARGAHIST.CSV | Dados históricos de Carga verificada. |
| XX_AAAA-MM-DD_COMBINA.CSV | Pesos para cálculo de previsão combinada. |
| XX_AAAA-MM-DD_FERIADOS.CSV | Data de feriados. |
| XX_AAAA-MM-DD_HORAVERAO.CSV | Dados de horário de verão. |
| XX_AAAA-MM-DD_HORIZONTE.CSV | Horizonte de previsão. |
| XX_AAAA-MM-DD_PATAMARES.CSV | Patamares de carga e sua duração. |
| XX_AAAA-MM-DD_SEPARADOR.CSV | Separador decimal padrão. |
| XX_AAAA-MM-DD_TEMP HIST.CSV | Dados históricos de temperatura. |
| XX_AAAA-MM-DD_TEMP PREV.CSV | Previsões de temperatura para o horizonte. |

fonte: Adaptado de CEPEL, 2020

Dentre esses arquivos podemos destacar os mais importantes para a modelagem do modelo LSTM de rede neural.

1. Arquivo CARGHIST. Histórico de carga global horária, verificada e ajustada e para o respectivo subsistema. Este histórico é previamente tratado pelo ONS para remoção de efeitos adversos (contingências), prevalecendo apenas fatores de temperatura e atividade de consumidores.. O histórico disponibilizado tem início no ano de 2016, e horizonte até a data mais recente da consulta (CEPEL, 2020).
2. Arquivo TEMPHIST. Histórico de temperatura equivalente por subsistema com base horária. Esta temperatura é uma variável *proxy* modelada pela melhor combinação linear de temperaturas medidas e pontos estratégicos do respectivo subsistema. Os pesos atribuídos aos diferentes pontos de medição de temperatura (aeródromos) são calculados utilizando o melhor ajuste de correlação, calculado pelo ONS utilizando técnicas de algoritmos bio-inspirados (CEPEL, 2020).

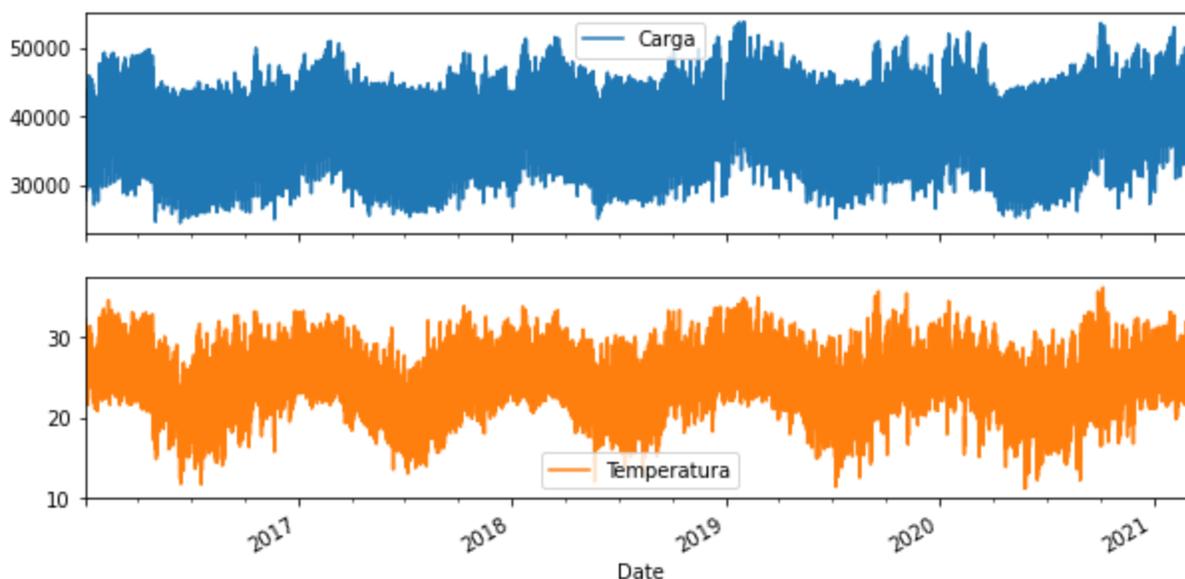
Para compor os dados de entrada, os arquivos são armazenados em ambiente de cloud (Google Drive), para que possam ser consumidos pelo algoritmo, que será rodado com auxílio da ferramenta Google Collaboratory, um ambiente de notebooks python gratuito, rodando diretamente em computadores na nuvem.

Os arquivos são lidos separadamente utilizando a biblioteca de processamento de dados Pandas, e salvos em uma estrutura de dados *dataframe*. Assim são obtidas variáveis com os históricos de carga e temperatura.

7.2 - Análise dos dados históricos de Carga e Temperatura.

Para que o modelo tenha um melhor desempenho, é necessário analisar os dados de entrada e verificar seus parâmetros gerais. Embora os dados de entrada sejam previamente tratados pelo ONS, pequenos ajustes são necessários, dado que a base de dados utilizada é preparada inicialmente para ser utilizada em outra metodologia.

Figura 21 - Séries temporais de Carga [MW médio] e Temperatura [°C] | Sudeste/Centro-Oeste



fonte: próprio autor

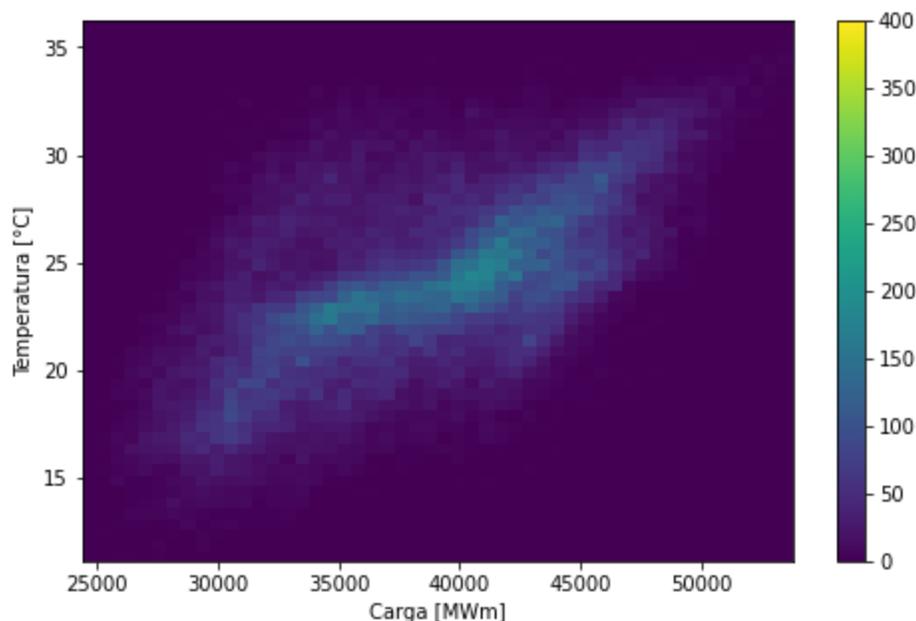
A curva de temperatura utilizada é uma série equivalente de temperatura, obtida através da contribuição de temperatura de diversas capitais e cidades de maior relevância comercial (ONS,2019).

Ambas as séries apresentam um efeito sazonal, onde observa-se a elevação de temperatura nos meses de verão e uma redução considerável da média de temperatura em períodos de meses de inverno.

Observando as séries nota-se o efeito de correlação entre carga e a temperatura. Para investigar mais profundamente este efeito, é possível plotar ambas as séries em um histograma de duas dimensões, assim é observável a interação entre ambas as séries. Pela figura 22, nota-se a correlação positiva entre a série temporal de carga e a série de temperatura.

Este efeito se associa a intensificação do uso de dispositivos e sistemas de refrigeração em períodos de maiores temperaturas, tais como ar condicionados, ventiladores e freezers.

Figura 22 - Histograma 2D - Carga *versus* Temperatura



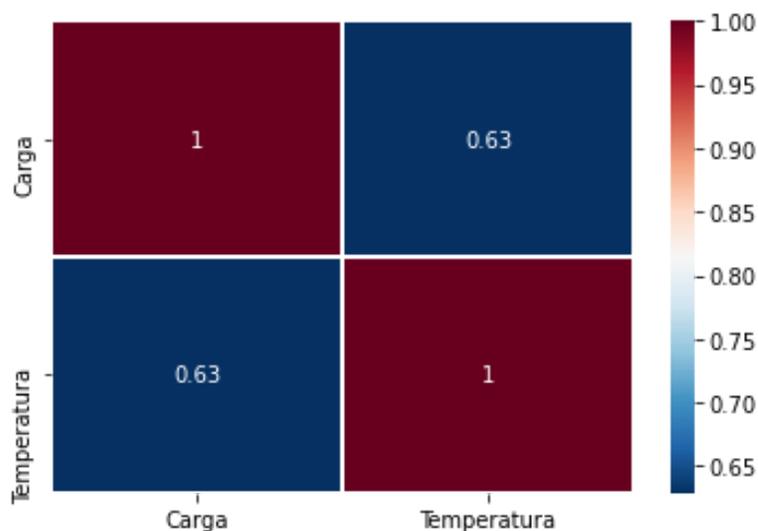
fonte: Próprio autor

Em termos numéricos, esse efeito de correlação pode ser obtido através do cálculo da matriz de correlação de Pearson, onde calcula-se o grau de correlação ρ entre dois vetores X e Y de tamanho N , em termos de sua covariância e suas variâncias individuais em relação a suas respectivas médias:

$$\rho(X, Y) = cov(X, Y) / (var(X) \cdot var(Y)) \quad (7.1)$$

Este coeficiente pode ser calculado para a combinação de variáveis de temperatura e carga para obtenção da matriz. A diagonal principal representa os valores de autocorreção, e os demais valores contém a correlação entre ambas as variáveis. O coeficiente de valor maior que zero indica correlação positiva entre os vetores analisados, e menor que zero em caso de correlação negativa, com seu módulo variando entre 0 e 1, sendo unitário em caso de acoplamento perfeito entre as variáveis.

Figura 23 - Matriz de correlação de Pearson - Carga *versus* Temperatura



fonte: Próprio Autor

O coeficiente de correlação de Pearson calculado foi de 0.63. Este valor é suficiente para demonstrar numericamente o efeito de correlação positiva entre as variáveis.

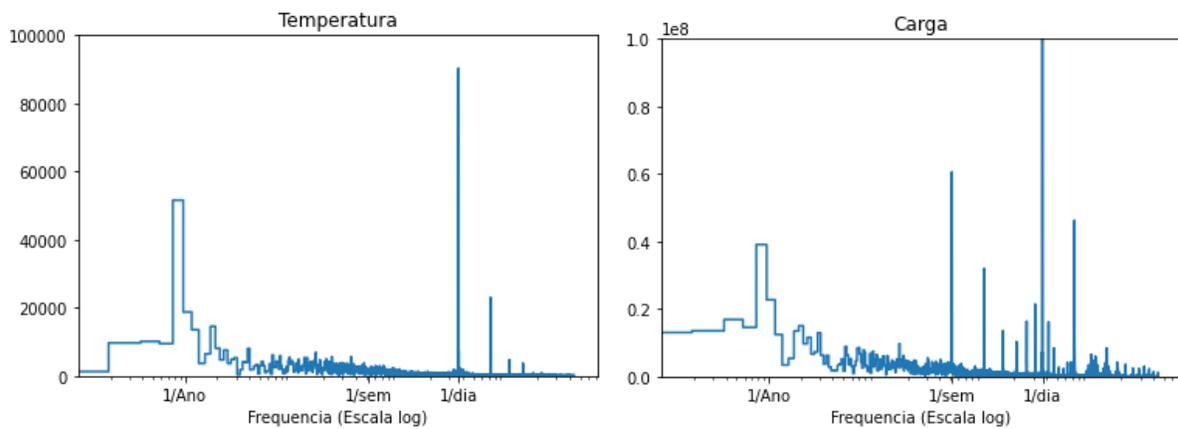
Outro fator de influência nas duas séries, é a presença de efeitos de sazonalidade. Com o auxílio da transformada rápida de Fourier (FFT), é possível decompor o espectro de frequências de ambas as séries.

A transformada foi implementada em linguagem Python, com suporte do módulo de processamento de sinais da biblioteca TensorFlow (TensorFlow 2021), para toda a amostra de

sinal disponível, calculada utilizando o algoritmo de RFFT. As frequências de interesse são marcadas pelos períodos de 1 ano, 1 semana e 1 dia, respectivamente transformados em hertz (Hz).

Para a modelagem, dado o escopo do problema, que envolve uma previsão de horizonte diário, a componente diária apresenta maior influência em ambas as séries, e é desejável que o modelo LSTM aprenda este comportamento. Para tal, a estratégia se baseia em gerar uma série temporal que possa ser introduzida no modelo, e carregue as informações de sazonalidade diária.

Figura 24 - Análise de FFT para séries de temperatura e Carga



fonte: próprio autor

Dessa forma, utilizando as marcações de tempo disponíveis, foram geradas duas séries temporais, de seno e cosseno, contendo a frequência diária desejada.

$$S_{sin}(t) = \sin\left(2\pi\frac{1}{86400}t\right) \quad (7.1)$$

$$S_{cos}(t) = \cos\left(2\pi\frac{1}{86400}t\right) \quad (7.2)$$

As séries resultantes (figura 25) quando combinadas fornecem as características sazonais desejadas, e permitem que sejam inseridas como dados de entrada. Os valores de 86400 equivalem ao número de segundos no período de um dia.

Este processo de rearranjo dos dados foi compilado em funções de apoio, permitindo rápida manipulação dos dados em caso de reexecução. Assim obteve-se o *Dataset* apropriado para execução do treinamento, contendo 45552 entradas de dados, contendo as séries sazonais de seno e cosseno, temperatura e carga verificadas, para o subsistema Sudeste/Centro-Oeste.

Figura 25 - Séries temporais sintéticas de frequência diária



fonte: próprio autor.

7.3 - Preparação dos dados de Treinamento, Validação e Teste

Para realização das etapas de treinamento, validação e teste, o *Dataset* foi segmentado em três conjuntos, cada qual designado para uma dessas etapas. Essa repartição tem como objetivo treinar o modelo LSTM com a maior combinação de casos possíveis, validar estes dados durante o treinamento permitindo assim ajustes dos parâmetros de treino, e por fim mensurar seu desempenho final.

Tabela 1 - Proporção de segmentação dos conjuntos de dados de Treino, Validação e Teste

| Dataset | Proporção (do total) | Tamanho |
|-----------|----------------------|---------|
| Treino | 80% | 36442 |
| Validação | 10% | 4555 |
| Teste | 10% | 4555 |

fonte: próprio autor

O processo de previsão consiste em obter a carga prevista para o próximo dia do horário de planejamento, mais precisamente, para as próximas 24 horas. Para tal será fornecido ao modelo LSTM o histórico das últimas 168 horas, equivalente aos 7 dias anteriores ao início do período de previsão, para as 4 séries de entrada.

Tabela 2 - *Datasets* de treinamento, validação e teste do modelo de Rede Neural LSTM

| Dataset | X (entrada) | Y (saída) |
|----------------------------------|---------------|--------------|
| $[X_{treino}, Y_{treino}]$ | [36442,168,4] | [36442,24,1] |
| $[X_{validação}, Y_{validação}]$ | [4555,168,4] | [4555,24,1] |
| $[X_{teste}, Y_{teste}]$ | [4555,168,4] | [4555,24,1] |

fonte: próprio autor

Para isso os conjuntos passam por um processo de janelamento, dividindo as entradas do modelo em n janelas de entrada e saída, sendo as entradas x_n uma matriz com dimensões $dim(x_n) = [168, 4]$, e as saídas y_n de dimensões $dim(y_n) = [24, 1]$, contendo os valores

reais de carga. Os *datasets* finais $X = [x_1, x_2, \dots, x_n]$ e $Y = [y_1, y_2, \dots, y_n]$ possuem 3 dimensões, sendo elas o número de exemplos do conjunto, o número de intervalos de tempo.

7.4 - Ajuste de escala dos dados de entrada

Para aumentar a performance do modelo LSTM, uma estratégia é padronizar os dados de entrada. Para isso subtrai-se a média e divide-se os dados pelo seu desvio padrão. Esta transformação é efetuada nas séries temporais de carga e temperatura, produzindo séries temporais de média nula e variância unitária.

Um processo alternativo para ajuste de escala envolveria normalizar os dados, em termos dos valores máximos e mínimos. Entretanto, o processo de padronizar os dados em torno da média e desvio padrão reduz a influência de *outlayers* durante o processo de treinamento (Géron, 2019).

Este ajuste é necessário apenas para as séries de carga e temperatura, uma vez que as séries sintéticas sazonais de seno e cosseno, já possuem média nula e variância unitária.

8 - ELABORAÇÃO E TREINAMENTO DO MODELO

O modelo LSTM desenvolvido contém uma camada de entrada com 64 neurônios, e uma camada de saída de 24 neurônios. Para este modelo utiliza-se uma arquitetura de sequencia-sequência, onde o modelo retorna uma sequência de 24 valores a cada intervalo de tempo.

Este modelo se opõe a versão mais simples de sequencia-vetor, onde os 24 períodos de tempo são previstos apenas na última iteração. Essa estratégia permite que mais gradientes de erro fluam através das camadas, auxiliando no processo de aprendizagem (Géron, 2019).

Tabela 3 - Sumário do modelo de rede neural LSTM

| Camada (tipo) | Unidades | Nº de Parâmetros treináveis |
|---------------|-----------|-----------------------------|
| Camada LSTM | 64 | 17664 |
| Camada DENSA | 24 | 1560 |
| TOTAL | 88 | 19224 |

fonte: próprio autor

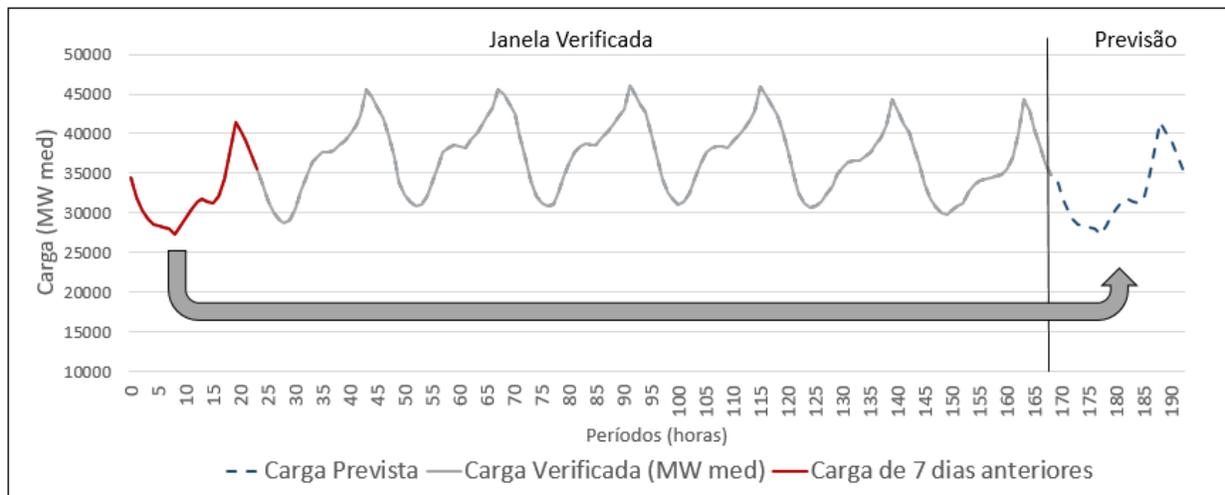
8.1 - Estimadores de Referência

Para verificar o desempenho, foram formulados modelos de linha de base (*baseline*), para referência da acurácia do modelo LSTM nas previsões. Um dos modelos de linha de base utilizados é um estimador por repetição (TensorFlow, 2021).

Este estimador tem fácil implementação, e sua função é replicar a entrada referente ao sétimo dia anterior à previsão. A exemplo, caso a previsão esteja sendo efetuada para uma quarta-feira da semana corrente, o estimador replica a entrada referente a última quarta-feira

verificada dos dados de entrada. Esta aproximação pode ser eficiente dado que os dias da semana possuem perfis de carga bem definidos.

Figura 26 - Estimador de Linha de Base (repetição)



fonte:Próprio autor

Outro possível estimador que pode ser utilizado para avaliar o desempenho do modelo LSTM é utilizar uma rede neural recorrente simples, sem conexões LSTM.

Tabela 4 - Sumário do modelo de Rede Neural Recorrente sem estruturas LSTM

| Camada (tipo) | Unidades | Nº de Parâmetros treináveis |
|---------------|-----------|-----------------------------|
| Camada LSTM | 64 | 4416 |
| Camada DENSA | 24 | 1560 |
| TOTAL | 88 | 5976 |

fonte: próprio autor

Este modelo de referência utiliza a mesma topologia do modelo LSTM, em termos de número de unidades e camadas, porém sem a inclusão das unidades de memória. Dessa forma, pode-se observar um número muito menor de parâmetros de treinamento na camada recorrente, uma vez que as estruturas LSTM acrescentam diversos parâmetros treináveis ao modelo.

Por se tratar de um modelo mais complexo, é esperado que o modelo LSTM obtenha performances superiores a ambos os estimadores de referência adotados.

8.2 - Ambiente de construção dos modelos

Para auxiliar na criação e treinamento do modelo LSTM e dos classificadores de base, foi utilizado o suporte de computação na nuvem, através da ferramenta Google Colaboratory (Google, 2021). O Colab permite a execução de códigos em notebooks Python, construídos com base da tecnologia de Jupyter Notebooks (Jupyter, 2021), e fornece acesso a máquinas remotas de alto processamento com GPU (*Graphic Processing Units*), permitindo treinar e ajustar os modelos (LSTM e demais estimadores de referência) com mais facilidade e rapidez.

Para produção dos modelos, foi utilizado o suporte da biblioteca de código aberto Python TensorFlow (TensorFlow, 2021), especializada em tarefas de aprendizado de máquina e criação de redes neurais.

8.3 - Treinamento do modelo LSTM

O modelo LSTM é então compilado seguindo os dados de topologia da Tabela 3. Durante o processo serão utilizados os dados padronizados, assim, a métrica de função de custo utilizada foi a MSE, bem como a métrica para para avaliação de seu desempenho.

O número máximo permitido de iterações foi de 100 épocas, dividindo o conjunto de dados de entrada em 8 lotes (*batches*) de treinamento. A taxa de aprendizagem inicial adotada é de $\eta = 5 \cdot 10^{-6}$. A atualização de pesos se deu através algoritmo de ADAM (*Adaptive Moment*

Estimation) (Kingma e Ba, 2015), com fatores de momento $\beta_1 = 0.900$ e $\beta_2 = 0.999$. Este algoritmo tem como funcionalidade melhorar o processo de convergência do modelo, adicionando momento ao processo de atualização de pesos, reduzindo ou aumentando sua intensidade quando necessário.

A tabela 5 contém os valores agrupados para os hiperparâmetros utilizados no treinamento.

Tabela 5 - Valores de Hiperparâmetros da etapa de *Backpropagation*

| Hiperparâmetro | Valor |
|---|-------------|
| Máximo de Épocas (N) | 100 |
| Tamanho dos lotes (<i>Batch Size</i>) | 8 |
| Taxa de Aprendizagem (η) | 5.10^{-6} |
| Momento (β_1) | 0.900 |
| Momento (β_2) | 0.999 |

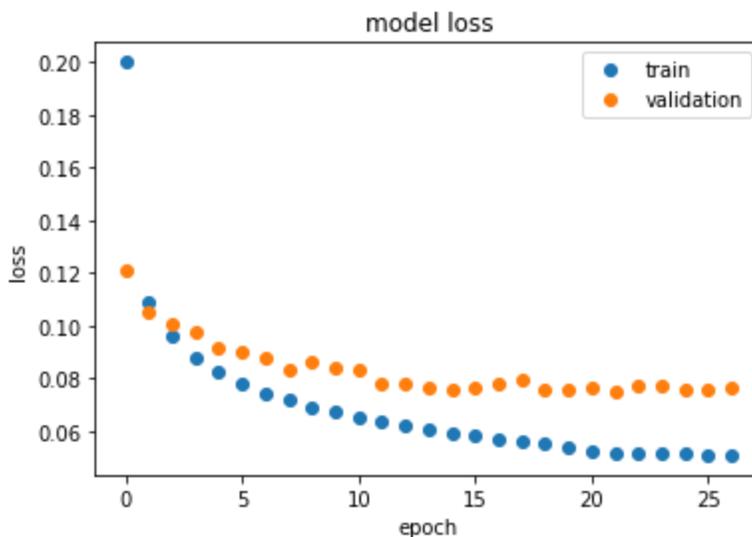
fonte: Próprio autor

Estes parâmetros foram determinados com base em diversas rodadas de treinamento, onde chegaram-se aos valores que produziam os melhores resultados. Ambos os modelos recorrentes, foram treinados com a mesma configuração de *BackPropagation*.

9 - RESULTADOS E DISCUSSÕES

O modelo LSTM então foi treinado por cerca de 20 épocas, onde foi interrompido pela rotina de *Early Stopping*, pois não houve melhoras na função de custo na ordem de 10^{-9} por 3 épocas consecutivas nos dados de validação. Na figura 27, é possível observar uma queda acentuada na função de custo de treino logo nas primeiras iterações. Na curva de validação também observa-se declínio ao longo das épocas, que embora seja de maneira mais suave.

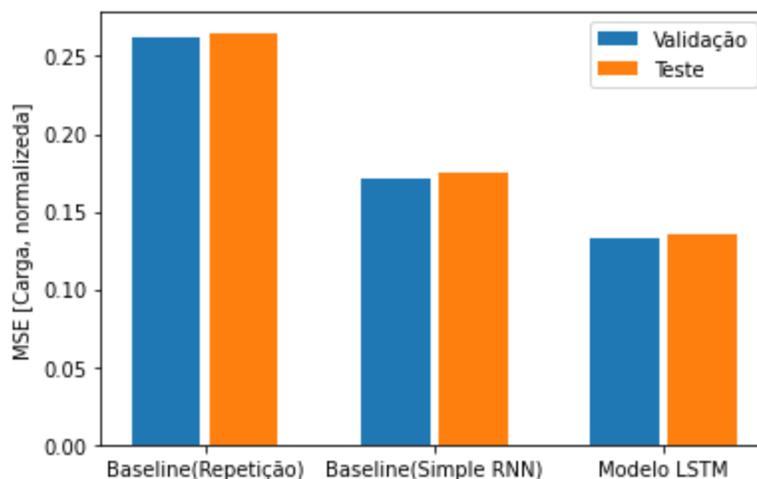
Figura 27 - Funções de custo de treino e validação



fonte: próprio autor

Os modelos de referência (Repetição e RNN simples) também foram treinados, sendo a RNN simples com especificações semelhantes a sua versão LSTM. Após o treinamento, os modelos foram avaliados utilizando os dados de teste, onde é possível observar um desempenho superior do modelo LSTM (Figura 28). Na Figura 29 pode-se observar a previsão elaborada para uma semana completa.

Figura 28 - Performance no conjunto de dados de Teste.



fonte: próprio autor

A performance média obtida no conjunto de dados de validação e teste são semelhantes uma vez que os conjuntos possuem tamanhos equivalentes, e ambos apresentam dados inéditos ao modelo, ou seja, que não foram submetidos durante o processo de treinamento.

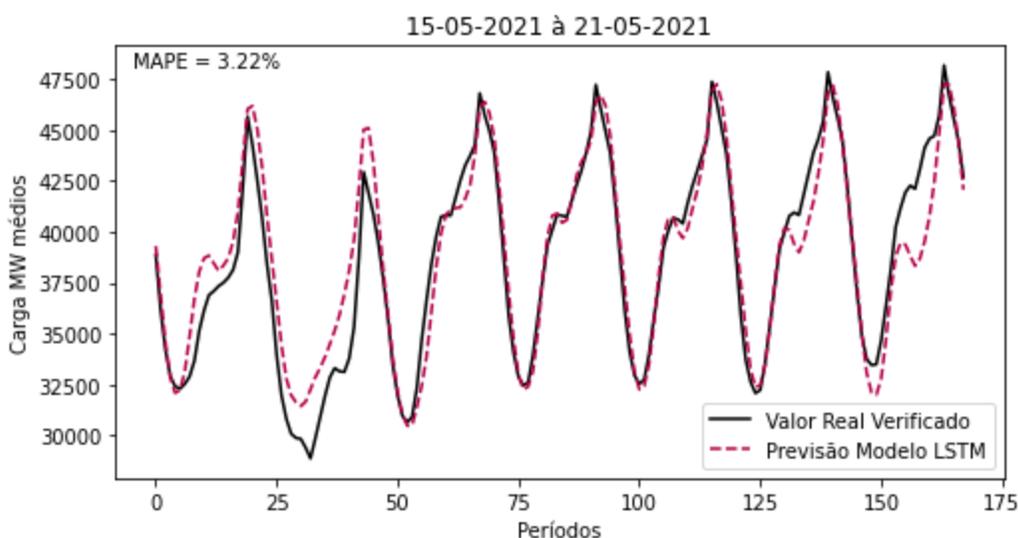
O equilíbrio de performance entre os dois conjuntos mostra que a generalização do modelo é boa, portanto não houve sobreajuste do modelo aos dados de treinamento, conforme já indicado pela curva de validação do modelo (figura 27).

O estimador de repetição apresentou performance inferior, conforme esperado, pois não há uma modulação eficiente entre a curva de carga sem conhecimento do perfil de temperatura, uma vez que entre as semanas ocorrem variações de temperaturas máximas e mínimas, além da temperatura média do dia. Estas variações podem alterar os picos de carga máxima e mínima, deslocando a carga média e assim dificultando a previsão.

Dessa forma os modelos baseados em RNN possuem performance superior, devido ao seu melhor ajuste da carga devido às variações da temperatura.

Em termos de MAPE, a previsão da rede neural LSTM apresentou um desvio de cerca de 3.22% em relação aos valores verificados para uma semana completa de previsão. Verificou-se desvios mais elevados em determinados períodos, principalmente no sábado e domingo da semana avaliada.

Figura 29 - Análise da previsão semanal do Modelo LSTM

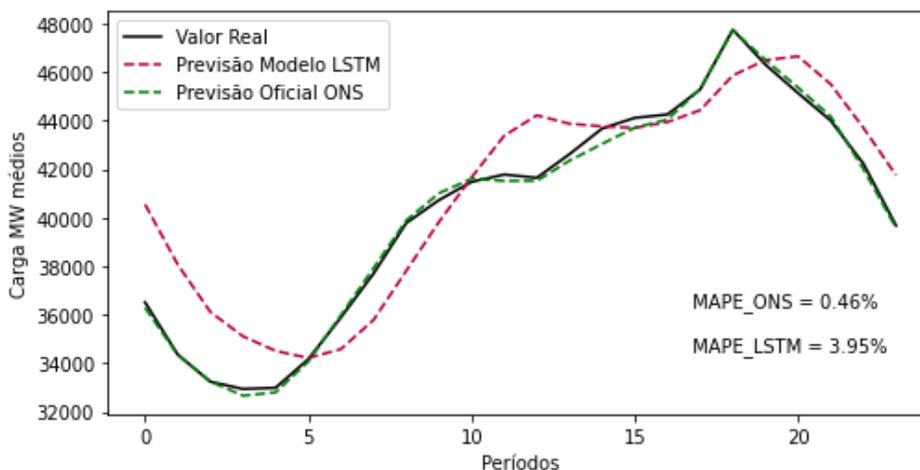


fonte: próprio autor

Para as previsões diárias, o resultado é semelhante. A figura 30 mostra a previsão feita para o dia 22 de maio de 2021, um sábado, juntamente com o valor verificado e a previsão oficial do modelo ONS.

Neste ponto, é possível notar mais adequadamente a dificuldade do modelo em se ajustar adequadamente a dias de carga mais específica, o que evidencia uma necessidade de melhor tratamento de dados desses dias especiais, sendo eles: feriados, eventos especiais (eleições, jogos da copa, etc.), sábados e domingos.

Figura 30 - Previsão de carga diária para o sábado - 22 de maio de 2021



fonte: Próprio autor.

A metodologia do ONS oferece melhor ajuste de previsão, pois envolve a melhor combinação estatística de diversos modelos (ANNSTLF, SVM, regressão quantílica, dentre outros) e ajustes pontuais feitos por especialistas. O modelo LSTM desenvolvido obteve um desvio menor que 5% para o dia avaliado, sendo considerado um resultado satisfatório, dado uma previsão de um único modelo sem os ajustes posteriores.

10 - CONCLUSÕES

A utilização de redes neurais no processo de previsão de carga abre uma possibilidade de novas metodologias que podem ser utilizadas para melhorar o processo e exatidão dos insumos necessários para execução da programação diária da operação eletroenergética.

Por meio do uso de uma topologia de redes neurais recorrentes e fazendo uso de um modelo com estruturas de memória LSTM foi possível obter estimativas da curva de carga para ser utilizada como entrada para o DESSEM, e conseqüentemente no cálculo do planejamento da programação diária da operação.

O modelo LSTM não apresentou sobreajuste dos dados, graças à boa generalização obtida treinando o modelo em um número otimizado de épocas, o suficiente para reconhecimento dos padrões, e a conseqüente interrupção do treinamento pela rotina de *early stopping*.

A previsão apresentou um MAPE da ordem de 3.22% para previsões executadas ao longo da semana verificada, sendo o modelo LSTM visivelmente superior à sua versão sem o uso de estruturas LSTM e ao estimador de repetição, ambos utilizados como modelos de referência. Apesar de possuir uma performance inferior à previsão oficial do modelo ONS, o que indica a necessidade de novas melhorias, o resultado é considerado satisfatório, dado a complexidade do problema.

De forma geral o modelo LSTM apresentou bons resultados na determinação da carga máxima os dias, sendo um fator importante na alocação de recursos em horários de maior demanda do sistema.

Os desvios encontrados em patamares de carga mais baixa mostram que há a necessidade de melhorar a entrada de dados horários, bem como uma melhor inserção de informações de dias especiais e finais de semana, com a finalidade de dar maior aderência a previsões nestes patamares.

Por fim, ainda existe a possibilidade de combinar o modelo com outras metodologias, melhorando a performance da previsão, e assim aproximando o resultado final da acurácia da metodologia oficial do ONS.

10.1 - Sugestões para trabalhos futuros

A título de sugestão para trabalhos futuros, as seguintes melhorias podem ser estudadas e testadas, a fim de melhorar a qualidade das previsões do modelo de previsão de carga desenvolvido.

- Adição de estruturas convolucionais nas camadas superiores da rede neural.
- Combinação do modelo com outras metodologias de previsão.
- Classificação e inserção de datas de feriados e dias especiais no modelo LSTM.
- Abertura da previsão para padrão semi-horário (30 em 30 minutos).
- Previsão semanal e mensal.
- Melhorias no tratamento e constituição dos dados de entrada.
- Testes com topologias de rede GNU ou Echo State.

REFERÊNCIAS

BENDERS, J. F.. Partitioning procedures for solving mixed variables programming problems. **Numerische Mathematik**, São Paulo, v. 4, n. 1, p. 238-252, jan./1962.

BENGIO, Yoshua; PATRICE, Simard; FRASCONI, Paolo. **Learning Long-Term dependencies is difficult**. 1. ed. Canadá: Università di Firenze, 1994.

BISHOP, C. M.. **Neural Networks for pattern recognition**. 1. ed. Oxford: Oxford University Press, 1995.

CCEE. **Portal da Câmara de comercialização de energia elétrica**. Disponível em: <https://www.ccee.org.br/portal>. Acesso em: 12 abr. 2021.

CEPEL. **DECOMP - Modelo de planejamento da operação de sistemas hidrotérmicos interligados de curto prazo, 2018a**. Disponível em: http://www.cepel.br/pt_br/produtos/decomp-modelo-de-planejamento-da-operacao-de-sistemas-hidrotermicos-interligados-de-curto-prazo.htm. Acesso em: 6 jan. 2021.

CEPEL. **Modelo DESSEM v16.7**: Programação diária da operação com representação detalhada das Unidades geradoras, Considerando rede elétrica e Restrições de Segurança. 16. ed. São Paulo: Centro de Pesquisas de Energia Elétrica, 2019.

CEPEL. **NEWAVE - Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo, 2018b**. Disponível em: http://www.cepel.br/pt_br/produtos/newave-modelo-de-planejamento-da-operacao-de-sistemas-hidrotermicos-interligados-de-longo-e-medio-prazo.htm. Acesso em: 6 jan. 2021.

CEPEL. **Relatório de projeto - DEA - 2461/ 2020 Cliente - ONS**: Manual de metodologia e uso do PrevCargaDessem. 1. ed. Rio de Janeiro: Centro de pesquisas de Energia Elétrica, 2020. p. 1.

CHURCHLAND, K. P. S; SEJNOWSKI, Terrence J.. **The Computational Brain, Computational Neuroscience Series**. 1. ed. Cambridge, MA: MIT Press, 1992.

EPE. **Atualização do valor para patamar único de custo de déficit**. 1. ed. Rio de Janeiro: Empresa de Pesquisa Energética , 2009.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1. ed. MIT: MIT Press, 2016.

GOOGLE. **Google Colab**. Disponível em: <https://colab.research.google.com/>. Acesso em: 2 fev. 2021.

GRAVES, Alex. **Supervised Sequence Labelling with Recurrent Neural Networks**. 1. ed. Berlin: Springer Books, 2012.

GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow**: Concepts, tools, and Techniques to Build Intelligent Systems. 2. ed. Canadá: O'REILLY, 2019.

HAYKIN, Simon. **Redes Neurais** : Princípios e prática. 2. ed. Porto Alegre: Bookman, 1999.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. **Long Short-Term Memory**: Neural Computation . 9. ed: Massachusetts Institute of Technology, 1997. p. 1735-1780.

JUPYTER. **Jupyter Project**. Disponível em: <https://jupyter.org/>. Acesso em: 26 abr. 2021.

KELMAN, Jerson. **Relatório da Comissão de Análise do Sistema Hidrotérmico de Energia Elétrica** . 1. ed., Brasília, 2001.

KINGMA, Diederik P.; BA, Jimmy. **Adam**: A Method for Stochastic Optimization. 1. ed, Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, USA, v. 5, n. 1, p. 115-133, dez./1943.

MCLEOD, A. I.. Diagnostic checking of periodic autoregression model with application. **Journal of Time Series Analysis**, São paulo, v. 15, n. 2, p. 221-233, jan./1994.

ONS. - **SINTEGRE: Portal de Relacionamento com os agentes**. Disponível em: <https://sintegre.ons.org.br/>. Acesso em: 13 mai. 2021.

ONS. **NT 096 - 2019** : Previsão de Carga para a programação diária eletroenergética. 1. ed. Rio de Janeiro, 2019.

PEREIRA, M. V. F; PINTO, L. M. V. G. Multi-stage optimization applied to energy planning . **Mathematical Programming** , São Paulo, v. 52, n. 3, p. 359-375, dez./1991.

PEREIRA, Mario; CAMPODÓNICO, Nora; KELMAN, Rafael. **Long-term hydro scheduling based on stochastic models**. 1. ed. Rio de Janeiro: EPSOM'98, 1998.

ROBINSON, A. J.; FALLSIDE, Frank. **The utility driven dynamic error propagation network.**: Technical Report CUED/F-INFENG/TR.1. 1. ed. Cambridge: Cambridge University Engineering Department., 1987.

SHARAF, A. M.. **A Neural Network based short-term load forecast model.** 1. ed. 1993: IEEE, 1993.

SIGA. **Sistema de informações de geração da ANEEL.** Disponível em: <https://app.powerbi.com/view?r=eyJrIjoiNjc4OGYyYjQtYWM2ZC00YjllLWJlYmEtYzdkNTQ1MTc1NjM2liwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBIMSIsImMiOjR9>. Acesso em: 15 mai. 2021.

SOUZA, Reinaldo C.; OLIVEIRA, Fernando Cyrino; FERREIRA, Pedro Costa. **Planejamento da operação de sistemas hidrotérmicos no Brasil:** Geração de Cenários e otimização. 2. ed. Rio de Janeiro, RJ: PUC RIO, 2014.

TENSORFLOW. **Tutorials :Time series forecasting.** Disponível em: <https://www.tensorflow.org>. Acesso em: 2 fev. 2021.

TOLMASQUIM, Maurício. **Novo modelo do Setor Elétrico Brasileiro** . 2. ed. Rio de Janeiro: Syngenta, EPE, 2015.

WERBOS, Paul. **Beyond regression : new tools for prediction and analysis in the behavioral sciences:** Dissertação (mestrado). 1. ed. Harvard: Harvard University, 1974.

WIDROW, B.; HOFF, M. E.. **Adaptive Switching Circuits.** 1. ed. Institute of Radio Engineers: Western Electronic Show and Convention, 1960.

WILLIAM, Ronald J.; ZIPSTER, David. **Gradient-Based Learning Algorithms for Recurrent Networks and Their Computational Complexity**. 1. ed. Stanford: Stanford Press, 1995.