

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO

Lucas Machado Basso

**MODELO DE PREVISÃO DE DEMANDA DE ORDENS DE
MANUTENÇÃO EMERGENCIAIS BASEADA EM FATORES
CLIMÁTICOS EM UM SISTEMA DE DISTRIBUIÇÃO
DE ENERGIA ELÉTRICA**

Santa Maria, RS
2020

Lucas Machado Basso

**MODELO DE PREVISÃO DE DEMANDA DE ORDENS DE MANUTENÇÃO
EMERGENCIAIS BASEADA EM FATORES CLIMÁTICOS EM UM SISTEMA DE
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA**

Dissertação apresentado ao curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Maria (UFSM), na linha de pesquisa Inteligência Organizacional, como requisito parcial para a obtenção do título de **Mestre em Engenharia de Produção**.

Orientador: Prof. Dr. Vinícius Jaques Garcia

Santa Maria, RS
2020

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Basso, Lucas Machado

Modelo de previsão de demanda de ordens de manutenção emergenciais baseada em fatores climáticos em um sistema de distribuição de energia elétrica / Lucas Machado Basso.- 2020.

74 p.; 30 cm

Orientador: Vinícius Jaques Garcia

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Produção, RS, 2020

1. Previsão de demanda 2. Modelos preditivos 3. Manutenção emergencial 4. Mineração de dados I. Garcia, Vinícius Jaques II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

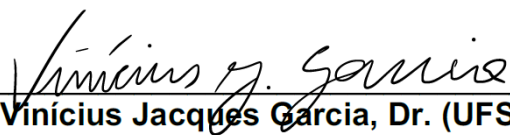
Declaro, LUCAS MACHADO BASSO, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Lucas Machado Basso

**MODELO DE PREVISÃO DE DEMANDA DE ORDENS DE MANUTENÇÃO
EMERGENCIAIS BASEADA EM FATORES CLIMÁTICOS EM UM SISTEMA
DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA**

Dissertação apresentada ao curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Maria (UFSM), na linha de pesquisa Inteligência Organizacional.

Aprovado em 26 de fevereiro de 2020:



Vinícius Jacques Garcia, Dr. (UFSM)
(Presidente/Orientador)



Carmen Brum Rosa, Dra. (UFSM)



Lynceo Falavigna Braghirolli, Dr. (UFSC)

Santa Maria, RS
2020

RESUMO

MODELO DE PREVISÃO DE DEMANDA DE ORDENS DE MANUTENÇÃO EMERGENCIAIS BASEADA EM FATORES CLIMÁTICOS EM UM SISTEMA DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

AUTOR: Lucas Machado Basso
ORIENTADOR: Prof. Dr. Vinícius Jaques Garcia

A Agência Nacional de Energia Elétrica (ANEEL) regula o mercado de distribuição de energia elétrica no Brasil e estabelece metas e indicadores de prazos de atendimento de solicitações dos consumidores às concessionárias de energia. O não cumprimento desses prazos gera multas a serem pagas pelas concessionárias. O primeiro passo para a solução deste problema é estabelecer um modelo de previsão de demanda de horas em ordens de manutenção emergenciais que sirva de base para o planejamento da capacidade necessária ao atendimento desses chamados. O objetivo deste estudo é desenvolver um modelo de previsão de demanda para ordens de manutenção emergencial em um sistema de distribuição de energia elétrica considerando fatores climáticos que possam afetar essa demanda. A metodologia utilizada no projeto segue cinco etapas: (i) revisão bibliográfica nas áreas de previsão de demanda e sistemas de distribuição de energia elétrica; (ii) estudo das características da demanda de ordens de manutenção emergenciais e sua relação com os fatores climáticos por meio de dados históricos; (iii) escolha do método de previsão de demanda mais adequado; (iv) desenvolvimento de um modelo de previsão de demanda para as ordens emergências que considere os fatores climáticos; e (v) realização da previsão dessa demanda a partir dos dados coletados. Como resultado deste projeto de pesquisa tem-se o modelo de previsão de demanda de ordens de manutenção emergenciais baseado em fatores climáticos em um sistema de distribuição de energia elétrica. A previsão da demanda de ordens de manutenção é fundamental para o planejamento dos recursos de mão-de-obra, materiais e infraestrutura necessários para atender aos chamados e garantir a disponibilidade do sistema de distribuição de energia elétrica.

Palavras-chave: Previsão de demanda. Modelos preditivos. Manutenção emergencial. Mineração de dados.

ABSTRACT

FORECASTING EMERGENCY MAINTENANCE ORDERS BASED ON CLIMATE FACTORS IN AN ELECTRICITY DISTRIBUTION SYSTEM

AUTHOR: Lucas Machado Basso
ADVISOR: Prof. Dr. Vinícius Jaques Garcia

The National Electric Energy Agency (ANEEL) regulates the electric energy distribution market in Brazil and establishes targets and deadlines for meeting requests from consumers to energy utilities. Failure to meet these deadlines generates fines to be paid by the concessionaires. The first step in solving this problem is to establish a model for forecasting the demand for hours in emergency maintenance orders that will serve as a basis for planning the capacity required to answer these calls. The objective of this study is to develop a demand forecasting model for emergency maintenance orders in an electricity distribution system considering climatic factors that may affect this demand. The methodology used in the project follows five steps: (i) bibliographic review in the areas of demand forecasting and electricity distribution systems; (ii) study of the characteristics of the demand for emergency maintenance orders and their relationship with climatic factors through historical data; (iii) choosing the most appropriate demand forecasting method; (iv) development of a demand forecast model for emergency orders that considers climatic factors; and (v) forecasting this demand from the data collected. As a result of this research project, there is a demand forecasting model for emergency maintenance orders based on climatic factors in an electricity distribution system. Forecasting the demand for maintenance orders is essential for planning the manpower, materials and infrastructure resources needed to answer calls and ensure the availability of the electricity distribution system.

Keywords: Demand forecast. Predictive models. Emergency maintenance. Data mining.

LISTA DE FIGURAS

Figura 1 – Relação entre demanda de eletricidade e temperatura	18
Figura 2 – Correlação entre todas as variáveis envolvidas no modelo	43
Figura 3 – Coeficientes do modelo de regressão múltipla	43
Figura 4 – Teste de hipótese para homocedasticidade	44
Figura 5 – Coeficientes do modelo de regressão múltipla	45
Figura 6 – Teste de hipótese para homocedasticidade	45
Figura 7 – Correlação entre variáveis modelo de regressão.....	46
Figura 8 – Valores dos coeficientes do modelo de regressão.....	47
Figura 9 – Gráfico de dispersão entre as variáveis X1 (emergência defasada) vs. Y (emergência)	47
Figura 10 – Gráfico de dispersão entre as variáveis X2 (máxima rajada dia) vs. Y (emergência).....	48
Figura 11 – Gráfico de dispersão entre as variáveis X4 (precipitação) vs. Y (emergência).....	49
Figura 12 – Gráfico de dispersão entre as variáveis X8 (Umidade) vs. Y (emergência).....	50
Figura 13 – Gráfico de dispersão entre as variáveis X9 (Velocidade média vento) vs. Y (emergência).....	51
Figura 14 – Valores dos coeficientes do modelo de regressão.....	51
Figura 15 – Teste de hipótese para homocedasticidade	52
Figura 16 – Teste de Durbin Watson	52
Figura 17 – Resumo de resultados dos resíduos.....	53
Figura 18 – Normal P Plot dos resíduos	53
Figura 19 – Teste <i>Kolmogorov-Smirnov</i>	54
Figura 20 – Valores de fator de inflação da variância	54
Figura 21 – Dados das horas de manutenção emergenciais	56
Figura 22 – Teste de hipótese <i>Augmented Dickey-Fuller</i> com tendência linear	57
Figura 23 – Teste de hipótese <i>Phillips-Perron</i> com tendência linear	58
Figura 24 – Teste de hipótese <i>Augmented Dickey-Fuller</i> com intercepto	59
Figura 25 – Teste de hipótese <i>Phillips-Perron</i> com intercepto	60
Figura 26 – Diagrama de auto correlação dos dados	61
Figura 27 – Modelo ARMA (1,1)	62
Figura 28 – Modelo ARMA (1,3)	63
Figura 29 – Modelo ARMA (3,1)	63
Figura 30 – Modelo ARMA (3,3)	64
Figura 31 – Diagrama de auto correlação modelo ARMA (1,3)	65
Figura 32 – Diagrama dos resíduos do modelo ARMA (1,3).....	66
Figura 33 – Previsão do modelo ARMA (1,3).....	67

LISTA DE TABELAS

Tabela 1 – MAPE, erro absoluto médio percentual	55
Tabela 2 – Comparativo entre os modelos ARMA	64

SUMÁRIO

1	INTRODUÇÃO	9
1.1	PROBLEMA DE PESQUISA	10
1.2	JUSTIFICATIVA.....	11
1.3	OBJETIVOS GERAL E ESPECÍFICOS	12
1.4	ESTRUTURA DO TRABALHO	12
2	REFERENCIAL TEÓRICO	14
2.1	PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA	14
2.2	SELEÇÃO DAS VARIÁVEIS	18
2.2.1	Temperatura do Ponto de Equilíbrio.....	19
2.2.2	<i>Degree Days</i>.....	19
2.3	INTRODUÇÃO A PREVISÃO DE DEMANDA	20
2.4	MODELOS DE PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA	21
2.5	CHAMADAS EMERGENCIAIS DE MANUTENÇÃO.....	23
2.6	REGRESSÃO LINEAR MÚLTIPLA.....	24
2.7	ANÁLISE DE CORRELAÇÃO	25
2.8	ESTUDO DE SÉRIES TEMPORAIS.....	26
2.8.1	Modelos para Séries Temporais.....	27
2.8.1.1	<i>Tendência e Sazonalidade</i>	28
2.8.1.2	<i>Modelos de Suavização Exponencial</i>	30
2.8.1.3	<i>Modelos ARIMA.....</i>	33
3	METODOLOGIA.....	36
3.1	CENÁRIO DA PESQUISA	36
3.2	DELINEAMENTO DA PESQUISA	36
3.3	PROCEDIMENTO DE COLETA DE DADOS	37
3.4	DESENVOLVIMENTO DA PESQUISA.....	37
4	RESULTADOS	42
4.1	ANÁLISE DOS DADOS	42
4.2	REGRESSÃO MÚLTIPLA	42
4.3	SÉRIES TEMPORAIS	55
5	CONSIDERAÇÕES FINAIS	68
5.1	LIMITAÇÕES DA PESQUISA.....	69
5.2	SUGESTÕES DE ESTUDOS FUTUROS.....	69
	REFERÊNCIAS.....	70

1 INTRODUÇÃO

As concessionárias de distribuição de energia elétrica devem oferecer serviços de fornecimento de energia que tenham qualidade, níveis de tensão apropriados e uma baixa taxa de interrupção. Para atingir esses objetivos, especificamente a qualidade, as agências reguladoras estabelecem indicadores de qualidade de energia para serviços de fornecimento, bem como metas e limites a serem alcançados pelas concessionárias de energia elétrica. Essas empresas geralmente realizam programas de manutenção preventiva e corretiva, para melhorar a confiabilidade do sistema, estabelecendo melhores condições de trabalho para prolongar a vida útil de seus equipamentos (BERTLING et al., 2005).

A previsão de demanda de energia elétrica é essencial para a gestão de energia, programação de manutenção e decisões de investimento nos mercados de energia e na flutuação dos preços dos combustíveis. A demanda por energia elétrica depende de variáveis econômicas, sociais e das condições climáticas (MIRASGEDIS et al., 2006). A previsão da demanda de eletricidade desempenha um papel fundamental no gerenciamento de sistemas de energia, especialmente para fins de operação e manutenção (GONZÁLEZ-ROMERA et al., 2006).

Políticas de manutenção são definidas como um conjunto de atividades técnicas, administrativas e gerenciais a serem aplicadas durante o ciclo de vida de uma máquina ou instalação, usadas para orientar a tomada de decisões de gerenciamento ou dedicadas a restaurar a máquina em condições (CARNERO, 2014). As decisões sobre políticas de manutenção na rede elétrica envolvem questões operacionais relacionadas a máquinas e instalações, como modos de falha, juntamente com aspectos estratégicos decorrentes da necessidade de disponibilidade, segurança e qualidade na organização, com o objetivo geral de otimizar o custo total do ciclo de vida (CARAZAS; SOUZA, 2010).

Segundo Sittithumwat et al. (2004), os clientes atendidos por uma rede de distribuição de energia elétrica podem sofrer paralisações frequentes por diversos motivos, incluindo perda de suprimento de uma subestação, falha de componente, queda de raio, acidente de carro, contato de árvore e animal e *flashover* de isolante devido ao acúmulo de sujeira e presença de umidade. O setor de serviços públicos toma várias ações preventivas para minimizar as falhas que tem impacto direto na

confiabilidade do serviço. Contudo, quando essas falhas ocorrem, é necessário o atendimento e reestabelecimento rápido do fornecimento de energia.

Para o rápido atendimento dos chamados de manutenção emergenciais, a empresa de energia deve manter capacidade de recursos humanos e materiais disponíveis para atender a demanda (FROGER et al., 2016). A quantidade de capacidade a se ter disponível depende da quantidade e complexidade dos chamados de manutenção emergenciais. Acredita-se que parte desses problemas emergenciais são decorrentes de condições climáticas como ventos e chuvas (NGUYEN et al., 2018).

Perante a essas questões, o presente trabalho tem como objetivo realizar uma previsão de demanda de ordens de manutenção emergenciais, essa previsão terá a finalidade de auxiliar a empresa alvo do estudo no planejamento de suas equipes de manutenção, tanto em termos de quantidade como alocação das mesmas.

1.1 PROBLEMA DE PESQUISA

Segundo Ahmed et al. (2018), a população em todo o mundo está aumentando continuamente, o que exige mais geração de energia elétrica para satisfazer as necessidades individuais. Além disso, com o advento de novas tecnologias e melhorias no estilo de vida, o uso médio de equipamentos elétricos também está aumentando gradualmente. Conseqüentemente, as indústrias de energia precisam empreender um planejamento efetivo de expansão e manutenção do sistema para atender à demanda de carga adicional. Para o presente trabalho, será analisada a influência de fatores climáticos na demanda por manutenção emergencial realizadas por uma empresa de distribuição de energia elétrica. Busca-se, posteriormente, usar esses fatores climáticos para realizar previsões de demanda, facilitando o planejamento dos recursos necessários à realização das manutenções emergenciais em tempo apropriado.

Os atuais sistemas de distribuição de eletricidade operam em um mercado liberalizado. Esses sistemas devem, portanto, ser capazes de fornecer eletricidade a clientes com um alto grau de confiabilidade e serem econômicos para os fornecedores (YSSAAD et al., 2014). As empresas de serviços públicos estão constantemente explorando maneiras de otimizar o uso dos recursos disponíveis para manutenção, garantindo um nível aceitável de confiabilidade (SALMAN et al., 2017).

O gerenciamento de energia é crucial para a futura prosperidade econômica e segurança ambiental. A energia está ligada à produção industrial, à produção agrícola, à saúde, ao acesso à água, à população, à educação, à qualidade de vida, etc. O gerenciamento da demanda de energia é necessário para a alocação adequada dos recursos disponíveis (SUGANTHI; SAMUEL, 2012). Para atender esse gerenciamento da demanda de energia é necessário que a manutenção dos sistemas de distribuição de energia elétrica seja eficaz. Uma previsão eficaz de ordens de manutenção emergenciais vem de encontro a resolver os problemas de quantidade de recursos humanos e alocação desses recursos para que as manutenções sejam realizadas no tempo certo e oferecendo um serviço de qualidade.

1.2 JUSTIFICATIVA

Os sistemas de distribuição de energia elétrica são fundamentais para garantir o desenvolvimento econômico e social de um país (BATISTA et al., 2013). Projeções do Plano Nacional de Energia Elétrica 2030 preveem crescimento do consumo de energia elétrica chegando a mais de 134% dos valores registrados no ano de 2005 (BRASIL, 2007).

A Agência Nacional de Energia Elétrica (ANEEL) regula o mercado de distribuição de energia elétrica no Brasil e estabelece metas e indicadores de prazos de atendimento de solicitações dos consumidores às concessionárias de energia. O não cumprimento desses prazos gera multas a serem pagas pelas concessionárias.

Neste cenário, existe um *trade-off* entre atender todas as demandas no prazo, evitar multas e ter excesso de capacidade ou reduzir o custo de pessoal e, conseqüentemente a capacidade de atendimento, e pagar eventuais multas. O primeiro passo para a solução deste problema é estabelecer uma previsão de demanda de atendimentos confiável e que possa servir de base para o planejamento da capacidade disponível.

As técnicas de previsão de demanda referem-se ao cálculo matemático e estatístico usado para transformar parâmetros numéricos, como dados históricos, em uma previsão de quantidade de demanda a ser atendida no próximo período. A previsão de demanda é sempre formada por um valor determinístico de demanda e um fator de erro da previsão. O fator de erro representa a aleatoriedade dos dados e

a incerteza existente na previsão. Quanto menor o fator de erro, mais próxima a previsão estará dos dados reais.

Desta forma, é fundamental que o modelo de previsão de demanda escolhido seja representativo das características peculiares da demanda real e seja confiável para ser utilizado no planejamento da capacidade de atendimento da concessionária.

1.3 OBJETIVOS GERAL E ESPECÍFICOS

O objetivo geral deste projeto de pesquisa é o desenvolvimento de um modelo de previsão de demanda de ordens de manutenção emergenciais baseada em fatores climáticos em um sistema de distribuição de energia elétrica. Essa previsão de demanda será baseada em fatores climáticos e dados históricos.

Como objetivos específicos têm-se:

- a) Realizar uma revisão bibliográfica sobre previsão de demanda e ordens de manutenção para compreender o funcionamento do gerenciamento de sistemas de distribuição de energia elétrica;
- b) Analisar e entender as características da demanda de ordens de manutenção emergências existentes no sistema de distribuição de energia elétrica;
- c) Determinar o(s) método(s) de previsão de demanda a ser utilizado e coletar os dados necessários;
- d) Desenvolver um modelo de previsão de demanda para as ordens emergenciais que considere as características de demanda existentes;
- e) Realizar a previsão de demanda com os dados coletados, analisar e validar o modelo proposto.

1.4 ESTRUTURA DO TRABALHO

O presente trabalho é estruturado a partir de cinco capítulos, sendo eles descritos a seguir.

No capítulo 1, a Introdução contextualiza o trabalho e introduz as ideias desenvolvidas no mesmo. O capítulo possui introdução, problema de pesquisa, justificativa e objetivos.

No capítulo 2, o Referencial Teórico busca por conhecimento na bibliografia para dar base ao presente trabalho. Estudos sobre previsão de demanda de energia elétrica, chamadas de manutenção emergencial e métodos de previsão de demanda são os principais tópicos que a atual pesquisa pretende realizar.

O capítulo 3 apresenta a Metodologia do trabalho, compreendendo como a atual pesquisa é classificada em termos de: natureza, método científico, abordagem, objetivos e procedimentos técnicos. Este capítulo descreve as etapas para a realização do estudo.

No capítulo 4, Resultados, é apresentado o desenvolvimento da pesquisa, contendo as análises dos dados e desenvolvimentos dos métodos realizados no presente trabalho.

No capítulo 5 é descrito as considerações finais que o presente trabalho conseguiu atingir e os questionamentos para futuros trabalhos.

2 REFERENCIAL TEÓRICO

O presente trabalho tem como objetivo realizar uma previsão de demanda de horas de ordens de manutenção emergencial em um sistema de distribuição de energia elétrica, durante a pesquisa bibliográfica não foi encontrado trabalhos anteriores que retratassem esse cenário, logo, a pesquisa teve o objetivo de encontrar métodos de previsão de demanda na área de energia elétrica, para usar como suporte para o atual trabalho. O método de regressão linear múltipla é citado na bibliografia como o método para utilizar variáveis climáticas para explicar a demanda de energia elétrica, o presente trabalho pretende utilizar esse método para utilizar as variáveis climáticas para explicar a demanda de ordens de manutenção emergenciais na rede elétrica. O método de séries temporais será utilizado para analisar os dados históricos de ordens de manutenção emergenciais e realizar uma previsão a partir dos mesmos.

2.1 PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA

A mudança climática global tem despertado grandes preocupações nos últimos anos devido ao seu forte impacto em muitos aspectos do meio ambiente e das atividades humanas em áreas urbanas. Uma área diretamente afetada pela mudança climática é o consumo de energia para aquecimento e resfriamento. Como a demanda de resfriamento e aquecimento geralmente segue em direções opostas, os aumentos ou reduções líquidas do consumo de energia elétrica dependem em grande parte do domínio da demanda de resfriamento ou aquecimento da região (PETRI; CALDEIRA, 2015).

No contexto atual, a mudança climática tornou-se um fenômeno inequívoco. Embora envolva principalmente a mudança de temperatura, outras variáveis climáticas, como chuva, velocidade do vento, evaporação e umidade, também podem ser afetadas como resultado da mudança climática. Abordar os impactos das mudanças climáticas na demanda de energia elétrica é essencial para prever a demanda futura. Por exemplo, os requisitos de refrigeração e aquecimento mudam significativamente em relação às mudanças climáticas que podem resultar na mudança na demanda de carga de eletricidade (AHMED et al., 2018).

Está bem documentado que fatores relacionados ao clima desempenham um papel importante no consumo de energia elétrica. Por muitos anos, empresas de

serviços públicos e a indústria de energia elétrica se interessaram pela relação entre consumo de energia elétrica e clima, e desenvolveram algoritmos empíricos de normalização de clima, visando melhorar a previsão de carga sujeita à variações no clima regional (SAILOR, 2000).

A seleção de variáveis climáticas adequadas para previsão da demanda de eletricidade é crítica, pois afeta a precisão da previsão. Diferentes variáveis climáticas podem ter diferentes impactos na demanda de energia elétrica devido as diferentes condições geográficas (VU et al., 2015). Os elementos climáticos que influenciam a demanda de eletricidade são, em ordem decrescente de importância, temperatura, umidade, vento e precipitação (MIRASGEDIS et al., 2007).

A previsão de demanda de energia elétrica tem sido uma área ativa nas pesquisas há muitos anos, bem como um tópico fundamental para operações e planejamento de sistemas de energia. A falta de precisão das previsões de demanda e carga tem impacto negativo no custo operacional das empresas de energia (HAIDA; MUTO, 1994). Com base no *lead time* da previsão, quatro diferentes classes foram definidas. As previsões de longo prazo, abrangendo um intervalo de 1 a 20 anos à frente, são fundamentais para o planejamento estratégico e construção de nova geração e capacidade de transmissão, uma vez que novos projetos nessas áreas geralmente levam anos para serem concluídos (KANDIL et al., 2002). As previsões de médio prazo, de um mês a um ano, são usadas para agendamento de manutenção e planejamento de acordos de compartilhamento de energia. A previsão de curto prazo, de uma hora adiante para algumas semanas à frente tem importância fundamental na programação da mão-de-obra, planos de compra de combustível, capacidade de segurança, manutenção de curto prazo e uso de armazenamento de curto prazo (FRIEDRICH et al., 2014). A última classe de previsão é de curtíssimo prazo a qual varia de alguns minutos a uma hora à frente e é utilizada para controle da demanda em tempo real (TAYLOR, 2008).

No novo ambiente liberalizado dos mercados de eletricidade, é importante entender e ser capaz de prever os efeitos das variáveis naturais sobre necessidade de resfriamento. Espera-se que o resfriamento contínuo nos setores de processamento de alimentos, armazenamento, industrial, comercial e doméstico e aquecimento contínuo nos processos industriais funcione o tempo todo, independentemente das variações de temperatura ao ar livre. Como o sistema de processamento e armazenamento de alimentos troca calor com os arredores,

qualquer alteração na temperatura externa pode alterar a quantidade de eletricidade necessária (HEKKENBERG et al., 2009).

Conhecer o comportamento da carga antecipadamente é crucial no planejamento, na análise e na operação de sistemas de energia, de modo a garantir um fornecimento de eletricidade ininterrupto, confiável, seguro e econômico (MIRASGEDIS et al., 2006).

O aumento da temperatura global devido às mudanças climáticas tornou-se uma grande preocupação para as estratégias sobre demanda de energia elétrica. Como o aumento da temperatura tem um impacto significativo na demanda de eletricidade, a expansão da rede elétrica e o plano de aumento relacionado das indústrias de energia serão afetados devido à mudança nas condições climáticas (AHMED et al., 2012). A mudança climática pode atribuir a maior demanda de eletricidade per capita em algumas regiões, enquanto a demanda em outras regiões pode diminuir levando a um maior resfriamento e menores requisitos de aquecimento. Portanto, é necessário saber como as mudanças climáticas desempenharão um papel no crescimento da demanda futura de eletricidade e afetarão o planejamento da expansão do sistema de energia (SALOMÃO et al., 2007).

O requisito geral de eletricidade para fins de controle de temperatura pode variar demograficamente. Uma porcentagem significativa do uso de eletricidade também depende das variáveis socioeconômicas, como população, industrialização, urbanização, tecnologia eficiente de energia, avanço tecnológico, renda e preço.

Sailor (2000) aplicou vários modelos de regressão para vários estados nos EUA e destacou que a sensibilidade da demanda de eletricidade com relação à temperatura é positiva em alguns estados, enquanto é negativa em outros estados. Estudos realizados por Mirasgedis et al. (2007) para a Grécia mostraram que a mudança climática faz com que a demanda de eletricidade aumente significativamente. Ruth e Lin (2006) implementaram modelos de regressão múltipla considerando as mudanças climáticas e atividade socioeconômicas e encontraram relações significativas entre a demanda de eletricidade e a mudança climática.

Ang et al. (1992) descobriram que o consumo residencial de eletricidade em Cingapura foi sensível a pequenas mudanças nas variáveis climáticas, particularmente a temperatura, que estava intimamente ligada à crescente difusão de aparelhos elétricos para controles ambientais, mas uma grande parte do crescimento futuro na demanda de eletricidade surge da crescente necessidade de ar

condicionado, o que levará a variações sazonais cada vez maiores no consumo de eletricidade. Atakhanova e Howie (2007) estimaram a demanda agregada do Cazaquistão por eletricidade, bem como a demanda por eletricidade nos setores industrial, de serviços e residencial, usando dados regionais. O estudo mostrou que a elasticidade de preço da demanda em todos os setores era baixa e a elasticidade de renda da demanda no agregado e em todos os modelos do setor eram inferiores a uma unidade. Finalmente, a previsão ilustrou que a demanda de eletricidade pode crescer a 3% ou 5% ao ano, dependendo das taxas de crescimento econômico e da política do governo com relação aos aumentos de preço e à eficiência promocional.

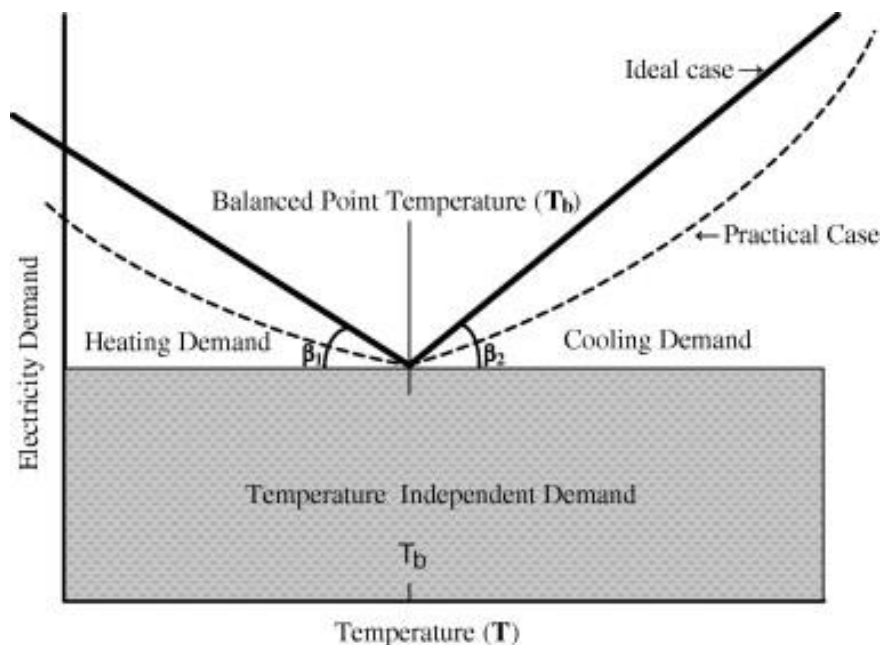
Os principais estudos indicam que a relação entre demanda de eletricidade e a temperatura é claramente não linear. Esta não linearidade refere-se ao fato de que ambos os aumentos e diminuições de temperatura, ligados a passagem de certas temperaturas “limiares”, aumentam a demanda por eletricidade. Segundo Henley e Peirson (1997), essa resposta é causada pelas diferenças entre a temperatura ambiente ou a temperatura externa e o conforto ou a temperatura interna. Quando o diferencial entre a temperatura interior e exterior aumenta, o arranque do equipamento de aquecimento ou resfriamento correspondente aumenta imediatamente a procura de energia. Naturalmente, a curva da resposta da demanda às temperaturas depende especialmente das características climáticas da área geográfica a que os dados de demanda se referem, que condicionam o tipo de equipamento instalado e as condições sob as quais ele começará a operar.

A relação entre demanda de energia elétrica e temperatura pode ser tipicamente representada por uma curva em forma de ‘U’ (FRANCO; SANSTAD, 2008). A demanda de energia aumenta com baixa temperatura devido à necessidade de aquecimento e também com a alta temperatura devido à necessidade de resfriamento. Espera-se que o resfriamento contínuo nos setores de processamento de alimentos, armazenamento, industrial, comercial e doméstico e aquecimento contínuo nos processos industriais funcione o tempo todo, independentemente das variações de temperatura ao ar livre. Como o sistema de processamento e armazenamento de alimentos troca calor com os arredores, qualquer alteração na temperatura externa pode alterar a quantidade de eletricidade necessária (HEKKENBERG et al., 2009).

Segundo Hekkenberg et al. (2009), a relação linear entre temperatura e demanda de energia pode ser representada usando graus-dias de temperatura. Os

graus-dias de aquecimento e resfriamento descrevem a partida da temperatura em um instante a partir de um nível limiar. Este ponto limiar é frequentemente definido como a temperatura do ponto de equilíbrio. Quando a temperatura é representada em termos de graus-dia, a curva de demanda de energia esperada será uma curva em forma de 'V', como podemos analisar na Figura 1. Os ângulos β_1 e β_2 da Figura 1 correspondem ao grau de resfriamento e aquecimento necessários com a mudança de temperatura. Os valores desses ângulos são específicos da localização e dependem de características socioeconômicas e geográficas. Valores mais altos de β_1 e β_2 indicam maior sensibilidade da demanda de eletricidade à mudança de temperatura.

Figura 1 – Relação entre demanda de eletricidade e temperatura



Fonte: Ahmed et al. (2012).

2.2 SELEÇÃO DAS VARIÁVEIS

Antes da análise de regressão, é essencial identificar variáveis que tenham impacto significativo na demanda de eletricidade. Como a análise de regressão baseada em clima lida com grandes conjuntos de dados, o descarte de variáveis não significativas exigirá menos processamento de dados. Entre todas as variáveis meteorológicas, a temperatura controla predominantemente a demanda de eletricidade (CHING-LAI et al., 2005).

2.2.1 Temperatura do Ponto de Equilíbrio

A temperatura do ponto de equilíbrio é a temperatura em que o consumo de eletricidade é menor. A energia necessária para fins de aquecimento e resfriamento é mínima neste momento. Esta é presumivelmente uma temperatura que não é nem muito fria nem muito quente. A temperatura de equilíbrio pode variar dependendo das características específicas do local e práticas culturais (RUTH; LIN, 2006). O ponto de equilíbrio para a análise relacionada à demanda de eletricidade pode ser considerado como 65° F (18,4° C), conforme o estudo de Ruth e Lien (2006). No entanto, na Tailândia, Parkpoom e Harrison (2008), consideram como sendo 24° C. Howden e Crimp (2001) usaram procedimentos de amostragem iterativa para determinar a temperatura do ponto de equilíbrio. Por conseguinte, é indicado que 17,5°C pode ser usado como a temperatura do ponto de equilíbrio para Sydney.

2.2.2 Degree Days

A não linearidade da resposta da demanda de eletricidade às temperaturas é causada pelo fato de que a demanda aumenta quando a temperatura diminui (temperaturas “baixas”) e quando a temperatura aumenta (temperaturas “altas”). Para incorporar esse comportamento em um modelo linear, a abordagem tradicional tem sido segmentar a variação de temperaturas em duas novas variáveis, portanto, segregando o comportamento da demanda quando as temperaturas estão “frias” e quando as temperaturas estão “quentes”.

a) Cooling Degree Days (CDD)

Dias de Grau de Refrigeração (CDD) representa a extensão do resfriamento necessário se a temperatura for superior à temperatura do ponto de equilíbrio. A expressão para o CDD é dada como:

$$CDD = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n \alpha_1 (T_{ij} - T_b) \quad (1)$$

$$\alpha_1 = \begin{cases} 1 & \text{if } T_{ij} - T_b > 0 \\ 0 & \text{if } T_{ij} - T_b < 0 \end{cases} \quad (2)$$

Onde n é o número de amostras de dados em um dia; m é o número total de dias; T_b é a temperatura do ponto de equilíbrio; e T_{ij} é a temperatura do ar da j th horas em i º dia.

b) Heating Degree Days (HDD)

Dias de grau de Aquecimento (HDD) representa a extensão do aquecimento necessário se a temperatura for inferior ao valor limite.

$$HDD = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n \alpha_2 (T_b - T_{ij}) \quad (3)$$

$$\alpha_2 = \begin{cases} 1 & \text{if } T_b - T_{ij} > 0 \\ 0 & \text{if } T_b - T_{ij} < 0 \end{cases} \quad (4)$$

Onde n é o número de amostras de dados em um dia; m é o número total de dias; T_b é a temperatura do ponto de equilíbrio; e T_{ij} é a temperatura do ar da j th horas em i º dia.

2.3 INTRODUÇÃO A PREVISÃO DE DEMANDA

De acordo com Chopra e Meindl (2016), classificamos os métodos de previsão em quatro tipos, sendo eles:

- a) Qualitativo: métodos de previsão qualitativos são essencialmente subjetivos e contam com o julgamento humano. Eles são mais apropriados quando poucos dados históricos estão disponíveis ou quando especialistas possuem inteligência de mercado que possa afetar a previsão. Esses métodos também podem ser necessários para prever a demanda por vários anos no futuro em uma indústria nova.
- b) Séries Temporais: os métodos de previsão de séries temporais utilizam a demanda histórica para fazer a previsão. Eles fundamentam-se na suposição de que o histórico da demanda passada é um bom indicador da demanda futura. Esses métodos são mais apropriados quando o padrão de demanda básico não varia significativamente de um ano para o seguinte. Esses são os métodos mais simples de implementar e podem servir como um bom ponto de partida para uma previsão de demanda.

- c) Causal: os métodos de previsão causais supõem que a previsão da demanda está altamente relacionada a certos fatores no ambiente (o estado da economia, taxas de juros etc.). Os métodos de previsão causais encontram essa correlação entre demanda e fatores ambientais e utilizam estimativas de quais serão os fatores ambientais para prever a demanda futura. Por exemplo, o preço do produto está fortemente relacionado à demanda. Assim, as empresas podem utilizar métodos causais para determinar o impacto das promoções de preço sobre demanda.
- d) Simulação: os métodos de previsão por simulação imitam as escolhas do consumidor que geram a demanda para chegar a uma previsão. Usando simulação, uma empresa pode combinar métodos de séries temporais e causais para responder a perguntas como: qual será o impacto de uma promoção? Qual será o impacto de um concorrente que abre uma loja na vizinhança?

2.4 MODELOS DE PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA

Singh et al. (2012) realizam uma revisão abrangente de métodos e proposições de modelos aplicados à previsão de demanda, incluindo técnicas como: regressão simples, regressão múltipla, suavização exponencial, mínimos quadrados ordinários, previsão de carga adaptativa, séries temporais estocásticas auto regressivas, modelo ARMA, modelo ARIMA, máquinas de vetores de suporte, modelos baseados em *soft computing*, algoritmos genéticos, lógica *fuzzy*, redes neurais e sistemas especialistas baseados em conhecimento. Podemos analisar que são vários os métodos relatados em trabalhos anteriores que podem ser utilizados para se realizar uma previsão de demanda.

Alguns trabalhos recentes nesta área de estudo apresentam a combinação de dois ou mais métodos acima mencionados. Nie et al. (2012) descreve a combinação de um modelo de Média Móvel Integrada Autoregressiva (ARIMA) para a previsão da demanda diária de energia elétrica, combinado com *Support Vector Machines* (SVM) para corrigir o desvio da tendência básica de previsão anterior. A principal desvantagem do uso de modelos ARIMA é que a relação importante com as variáveis meteorológicas não é explicada (APADULA et al., 2012). Ahmed et al. (2012) utiliza a análise de regressão linear múltipla para as variáveis climáticas históricas e não

climáticas para estabelecer uma correlação entre a demanda de eletricidade e variáveis-chave associadas. Na sequência é utilizada a análise de series temporais para prever a temperatura futura e os graus de refrigeração e aquecimento correspondente em *New South Wales*, local do estudo realizado.

Âmbar et al. (2018) comparam as capacidades e a eficiência de previsão de cinco modelos diferentes, prevendo o consumo de eletricidade de um edifício administrativo localizado em Londres, Reino Unido. Esses modelos são: Regressão Múltipla, Programação Genética, Rede Neural Artificial, Rede Neural Profunda e Máquina de vetor de Suporte. Os modelos de previsão são desenvolvidos com base em cinco parâmetros diferentes, como radiação solar, temperatura, velocidade do vento, umidade e índice de dias uteis. O consumo de eletricidade previsto pelos modelos é comparado com o consumo real, sendo assim demonstrando que a Rede Neural Artificial tem o melhor desempenho, apresentando o menor erro médio percentual absoluto (MAPE) entre os modelos.

Son e Kim (2017) realizaram um estudo para fornecer um modelo preciso para previsão de um mês da demanda de eletricidade no setor residencial. O estudo utilizou 14 variáveis meteorológicas e cinco variáveis sociais. Com base na regressão vetorial de suporte e na seleção de características *fuzzy-rough* com algoritmos de otimização de enxame de partículas, o método proposto estabeleceu um modelo com variáveis que se relacionam com a previsão ignorando variáveis que inevitavelmente levam a erros de previsão. O modelo de previsão proposto foi validado usando dados históricos da Coreia do Sul entre janeiro de 1991 e dezembro de 2012. Este modelo também foi comparado com o obtido a partir de rede neural artificial, média móvel integrada auto regressiva (ARIMA) e modelos de regressão linear múltipla, mostrando ser o melhor modelo para previsão de curto prazo.

Ao analisar os estudos encontrados na literatura utilizando os modelos citados, temos exemplos como Apadula et al. (2012), onde um modelo de regressão linear múltipla baseado em variáveis relacionadas ao calendário e ao clima é desenvolvido para estudar as relações entre as variáveis meteorológicas e a demanda de eletricidade, bem como para prever a demanda mensal de eletricidade 1 mês à frente. Ruth e Lin (2006) realizaram um estudo sobre os impactos potenciais da mudança climática no uso da eletricidade pelos setores residencial e comercial no estado de Maryland, EUA. A análise de séries temporais foi usada para quantificar relações históricas de demanda de temperatura e energia. Um modelo de computador dinâmico

usa esses relacionamentos para simular a demanda futura de energia sob uma faixa de preços de energia, temperaturas e outros fatores.

Ahmed et al. (2018) utilizaram uma abordagem de regressão múltipla baseada em eliminação regressiva para analisar a influência das variáveis climáticas na previsão de carga elétrica. Uma análise de correlação foi realizada usando o coeficiente de *Pearson* para examinar a interdependência entre diferentes variáveis climáticas no contexto de Sydney. A regressão múltipla baseada na eliminação para trás é usada para excluir as variáveis climáticas não significativas e avaliar a sensibilidade de variáveis significativas relacionadas à demanda de carga elétrica. Sailor (2000) aplicou o modelo de regressão múltipla para examinar o efeito da temperatura na demanda de carga, especialmente em vários estados dos EUA e relatou que as demandas em alguns estados têm sensibilidade positiva à temperatura, enquanto as demandas em outros estados têm sensibilidade negativa.

2.5 CHAMADAS EMERGENCIAIS DE MANUTENÇÃO

Shafiee e Sorensen (2017) definem manutenção como uma estratégia que inclui um conjunto de políticas e ações que são usadas para “reter” ou “restaurar” um equipamento, bem como o sistema de suporte à decisão no qual as atividades de manutenção são planejadas. Seus estudos indicam alguns tipos de manutenção que podem ser realizadas, entre elas:

- a) Revisão: revisão pode incluir redesenho e/ou substituição de equipamentos críticos, sendo realizadas após um longo período de tempo.
- b) Manutenção corretiva: uma ação de reparo é realizada após uma falha de sistema.
- c) Inspeções de rotina: inspeções são consideradas visitas para analisar a normalidade do sistema, essas podem ser diárias, semanais ou na frequência que se for necessário para cada tipo de sistema.
- d) Manutenção centrada em confiabilidade: é usado para otimizar as decisões de manutenção de um sistema, evitando que seu nível de confiabilidade caia abaixo de um determinado valor específico. Envolve a manutenção das funções do sistema, identificado os problemas e selecionando as tarefas de manutenção mais apropriadas. Normalmente é aplicado a componentes /

subsistemas críticos cujas falhas podem resultar em falha catastrófica do sistema ou alta perda de produção.

- e) Manutenção preditiva: inclui o uso de métodos modernos de medição e processamento de dados para prever, acompanhar e diagnosticar com precisão as condições do sistema durante as operações.
- f) Manutenção baseada em risco: visa reduzir os riscos gerais associados a falhas inesperadas do sistema. O cronograma de inspeção e manutenção é otimizado com base em riscos quantificados causados por falhas de componentes. Os componentes de alto risco são inspecionados e analisados com maior frequência, enquanto que para componentes de baixo risco o esforço é minimizado para reduzir o escopo total do programa de trabalho e custo de manutenção.
- g) Manutenção de grupo: fornece uma base para combinar atividades de manutenção e compartilhar os custos de configuração com um número de componentes do sistema. Essa estratégia de compartilhamento pode reduzir os custos ou pode resultar em custos menores em comparação ao caso em que as tarefas de manutenção são conduzidas separadamente para cada componente.

2.6 REGRESSÃO LINEAR MÚLTIPLA

Segundo Ghani e Ahmad (2010), existem seis tipos de análises de regressão disponíveis em estatísticas, sendo essas: regressão linear simples, regressões lineares múltiplas, regressão logística, regressão ordinal, regressão multinomial e análise discriminante. Regressões lineares múltiplas (MLR) é o método de estatística em regressão que analisa a relação entre a variável de resposta única (variável dependente), com duas ou mais variáveis controladas (variáveis independentes).

Segundo Oliveira et al. (2011), modelos de regressão múltipla envolvem três ou mais variáveis ou estimadores. Isto significa que uma variável é função de duas ou mais variáveis independentes (explicativas). O propósito do uso de variáveis independentes adicionais é melhorar a capacidade preditiva do modelo, em comparação com a regressão linear simples. Mesmo quando há interesse no efeito de apenas uma variável, é importante incluir variáveis que possam afetar a variável principal, realizando uma análise de regressão múltipla. Assim, pode-se reduzir o

desperdício estocástico, pois reduzir a variância residual aumenta o poder dos testes de significância. Por meio de regressões múltiplas, também podemos eliminar a tendência que poderia resultar se fosse simplesmente ignorada uma das variáveis que afeta diretamente a variável dependente.

Muitas relações teóricas também podem ser expressas com modelos de regressão. No entanto, uma restrição nos modelos de regressão é que eles são válidos apenas no intervalo de onde os dados foram extraídos. Portanto, para alcançar uma relação experimental global, é necessário ter várias séries de dados (JIA; LIU, 2009).

O objetivo geral da regressão múltipla é aprender mais sobre a relação entre várias variáveis independentes e uma variável dependente. Em geral, os procedimentos de regressão múltipla estimarão uma equação linear dada na forma a seguir.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_nX_n + \varepsilon \quad (5)$$

Onde ε é a taxa de erro do modelo. Os coeficientes de regressão ou "termos b" representam as contribuições independentes de cada variável independente para a previsão da variável dependente. A linha de regressão expressa a melhor previsão da variável dependente (Y), dadas as variáveis independentes (X). No entanto, geralmente há variação substancial dos pontos observados em torno da linha de regressão ajustada. O desvio de um ponto particular da linha de regressão (seu valor previsto) é chamado de valor residual (HADI; HOMAYOON, 2017).

2.7 ANÁLISE DE CORRELAÇÃO

A análise de correlação é uma ferramenta importante para as diferentes áreas do conhecimento, não somente como resultado final, mas como uma das etapas para a utilização de outras técnicas de análise. O método usualmente conhecido para medir a correlação entre duas variáveis é o coeficiente de correlação linear de Pearson, também conhecido como coeficiente de correlação do momento produto. Este foi o primeiro método de correlação, introduzido por Karl Pearson em 1897 (LIRA, 2004).

Este coeficiente é também conhecido como Coeficiente de Correlação do Momento Produto. Na prática, conforme apresentado em Lira (2004), o coeficiente de

correlação $\hat{\rho}$ é interpretado como um indicador que descreve a interdependência entre as variáveis X e Y. Outra forma de interpretar o coeficiente de correlação é em termos de $\hat{\rho}^2 = R^2$, denominado coeficiente de determinação ou de explicação. Quando multiplicado por 100, o $\hat{\rho}^2 = R^2$ fornece a percentagem da variação em Y (variável dependente), que pode ser explicada pela variação em X (variável independente), ou seja, o quanto de variação é comum às duas variáveis.

Cabe lembrar que o coeficiente de determinação é a relação entre a variação explicada pelo modelo linear ($X\hat{\beta} + \hat{\alpha} = Y^*$, onde $\hat{\alpha}$ e $\hat{\beta}$ são constantes) e a variação total. A significância do coeficiente de correlação estimado é verificada através de teste de hipóteses. A estatística para testar a hipótese $H_0: \rho = 0$ contra $H_1: \rho \neq 0$ tem distribuição t com $n - 2$ graus de liberdade, ou seja:

$$t = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} \sim t_{n-2} \quad (6)$$

O coeficiente de correlação linear de Pearson varia de -1 a 1. O sinal indica direção positiva ou negativa do relacionamento e o valor sugere a força da relação entre as variáveis. Uma correlação perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra. No outro oposto, uma correlação de valor zero indica que não há relação linear entre as variáveis. Todavia, como valores extremos (0 ou 1) dificilmente são encontrados na prática, é importante discutir como os pesquisadores podem interpretar a magnitude dos coeficientes. Para Cohen (1988), valores entre 0,10 e 0,29 podem ser considerados pequenos; escores entre 0,30 e 0,49 podem ser considerados como médios; e valores entre 0,50 e 1 podem ser interpretados como grandes. Dancy e Reidy (2005) apontam para uma classificação ligeiramente diferente: $r = 0,10$ até $0,30$ (fraco); $r = 0,40$ até $0,6$ (moderado); $r = 0,70$ até 1 (forte).

2.8 ESTUDO DE SÉRIES TEMPORAIS

Segundo Morettin e Tolo (2004) e Montgomery et al. (2008), os objetivos da análise de séries temporais são: (i) investigar o mecanismo gerador da série temporal; (ii) fazer previsões de valores futuros da série; descrever apenas o comportamento da série; (iii) procurar periodicidades relevantes nos dados.

De modo geral, podemos identificar uma série temporal como um vetor $Z(t)$, de ordem $r \times 1$, onde, por sua vez, t é um vetor $p \times 1$. Então, consideramos a série:

$$Z(t) = [Z_1(t), Z_2(t), \dots, Z_n(t)] \quad (7)$$

Uma série temporal estacionária representa que ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Entretanto, a maior parte das séries que encontramos na prática apresentam alguma forma de não-estacionariedade.

Para se transformar uma série temporal não-estacionária em estacionária, é necessário a aplicação de diferenças sucessivas na série original (MORETTIN; TOLOI, 2004):

A primeira diferença de $Z(t)$ é definida por:

$$\Delta Z(t) = Z(t) - Z(t - 1), \quad (8)$$

A segunda diferença é

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta [Z(t) - Z(t - 1)], \quad (9)$$

Ou seja,

$$\Delta^2 Z(t) = Z(t) - 2Z(t - 1) + Z(t - 2), \quad (10)$$

De modo geral, a n -ésima diferença de $Z(t)$ é:

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]. \quad (11)$$

2.8.1 Modelos para Séries Temporais

Os modelos utilizados para descrever séries temporais são processos estocásticos. Qualquer que seja a classificação que faça para os modelos de séries temporais, existe um número muito grande de modelos diferentes para descrever o

comportamento de uma série particular. A construção destes modelos depende de vários fatores, tais como, o comportamento do fenômeno ou o conhecimento a priori que temos de sua natureza e do objetivo da análise (BROCKWELL; DAVIS, 2002).

2.8.1.1 Tendência e Sazonalidade

Podemos considerar as observações $\{Z_t, t = 1, \dots, N\}$ de uma série temporal. Tendo a equação 12:

$$Z_t = T_t + S_t + a_t; \quad (12)$$

Onde T_t e S_t representam a tendência e a sazonalidade, respectivamente, enquanto a_t é uma componente aleatória, de média zero e variância constante.

Supondo-se que a componente sazonal S_t não esteja presente, obteremos a equação 13:

$$Z_t = T_t + a_t; \quad (13)$$

Onde a_t é ruído branco, com variância constante.

Há vários métodos para estimar a tendência T_t . Dentre eles, os mais utilizados são (MORETTIN; TOLOI, 2004):

- a) ajustar uma função no tempo, como um polinômio, uma exponencial ou outra função suave de t ;
- b) suavizar (ou filtrar) os valores da série ao redor de um ponto, para estimar a tendência naquele ponto;
- c) suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados.

Estimando-se a tendência através de \hat{T}_t , obtém-se a série ajustada para tendência ou livre de tendência, conforme a equação 14.

$$Y_t = Z_t - \hat{T}_t, \quad (14)$$

Outro procedimento utilizado para eliminar a tendência de uma série é aquele de tomar diferenças, o exemplo da primeira diferença é apresentado na equação 15:

$$\Delta Z_t = Z_t - Z_{t-1}, \quad (15)$$

Ao se tratar da componente de sazonalidade, primeiramente precisamos compreender alguns conceitos sobre a mesma, empiricamente, consideramos como sazonais os fenômenos que ocorrem regularmente de ano para ano, como um aumento de vendas de passagens aéreas no verão, aumento de vendas no comércio na época do Natal, entre outras questões.

Segundo Morettin e Tolo (2004), observa-se em series sazonais que ocorrem relações:

- a) entre observações para meses sucessivos em um ano particular;
- b) entre as observações para o mesmo mês em anos sucessivos.

Assim, a observação Z_t correspondente a dezembro do ano de 2000 é relacionada com os demais meses de 2000, bem como com os demais meses de dezembro de 1999, 2001 etc. Assim, Z_t é relacionada com Z_{t-1}, Z_{t-2}, \dots , mas também com Z_{t-s}, Z_{t-2s}, \dots . Isto implica que séries sazonais são caracterizadas por apresentarem correlação alta em “lags sazonais”, isto é, lags que são múltiplos do período s . O procedimento de ajustamento sazonal será tal que esta correlação seja removida (pelo menos em grande parte).

Considerando-se, por exemplo, o caso de dados mensais p e o número total de observações, N , é um múltiplo de 12, isto é, $N = 12p$, $p =$ número de anos. Sendo assim temos as equações 16, 17 e 18:

$$\bar{Z}_i = \frac{1}{12} \sum_{j=1}^{12} Z_{ij}, i = 1, \dots, p, \quad (16)$$

$$\bar{Z}_j = \frac{1}{p} \sum_{i=1}^p Z_{ij}, j = 1, \dots, 12, \quad (17)$$

$$\bar{Z} = \frac{1}{12p} \sum_{i=1}^p \sum_{j=1}^{12} Z_{ij} = \frac{1}{N} \sum_{t=1}^N Z_t, \quad (18)$$

A partir disto, pode-se reescrever o modelo para sazonalidade constante:

$$Z_{ij} = T_{ij} + S_{ij} + a_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, 12. \quad (19)$$

2.8.1.2 Modelos de Suavização Exponencial

A maioria dos métodos de previsão baseia-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série (MONTGOMERY et al., 2008).

Uma das classes de métodos utilizados para realizar previsões é a das suavizações. Segundo Morettin e Tolo (2004), técnicas específicas desse tipo assumem que os valores extremos da série representam a aleatoriedade e, assim, por meio da suavização desses extremos, pode-se identificar o padrão básico. A seguir serão apresentados alguns modelos para série localmente constantes.

a) Médias Móveis Simples (MMS)

A técnica de média móvel consiste em calcular a média aritmética das r observações mais recentes, é dada pelas equações 20 e 21 (ANDERSON, 1979):

$$M_t = \frac{Z_t + Z_{t-1} + \dots + Z_{t-r+1}}{r} \quad (20)$$

Ou

$$M_t = M_{t-1} + \frac{Z_t - Z_{t-r}}{r} \quad (21)$$

Assim, M_t é uma estimativa do nível μ_t que não leva em conta (ou não pondera) as observações mais antigas, o que é razoável devido ao fato do parâmetro variar suavemente com o tempo.

O nome média móvel é utilizado porque, a cada período, a observação mais antiga é substituída pela mais recente, calculando-se uma média nova.

Segundo Morettin e Tolo (2004), as principais vantagens do modelo são:

- i. simples aplicação;

- ii. é aplicável quando se tem um número pequeno de observações;
- iii. permite uma flexibilidade grande devido à variação de r de acordo com o padrão da série.

Como desvantagens têm-se:

- i. deve ser utilizado somente para prever séries estacionárias, caso contrário a precisão das previsões obtidas será muito pequena, pois os pesos atribuídos às r observações são todos iguais e nenhum peso é dado às observações anteriores a esse período;
- ii. necessidade de armazenar pelo menos $(r - 1)$ observações; e
- iii. dificuldade de determinar o valor de r .

b) Suavização Exponencial Simples (SES)

Conforme Morettin e Tolo (2004) a SES pode ser descrita matematicamente pela equação 22:

$$\bar{Z}_t = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k Z_{t-k} + (1 - \alpha)^t \bar{Z}_0, \quad t = 1, \dots, N, \quad (22)$$

Onde \bar{Z}_t é denominado valor exponencialmente suavizado e α é a constante de suavização, $0 \leq \alpha \leq 1$. Efetuando a expansão da equação 22, temos a equação 23:

$$\bar{Z}_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + \alpha(1 - \alpha)^2 Z_{t-2} + \dots \quad (23)$$

O que significa que a SES é uma média ponderada que dá pesos maiores às observações mais recentes, eliminando uma das desvantagens do método de Médias Móveis Simples.

Para a determinação da constante α , quanto menor for o valor de α mais estável serão as previsões finais, uma vez que a utilização de baixo valor α implica que pesos maiores serão dados às observações passadas e, conseqüentemente, qualquer flutuação aleatória, no presente, exercerá um peso menor no cálculo da previsão. Em geral, quanto mais aleatória for a série estudada, menores serão os valores da constante de suavização. O efeito de α grande ou pequeno é

completamente análogo (em direção oposta) ao efeito do parâmetro r no método MMS (MONTGOMERY et al., 2008).

As vantagens do método SES, segundo Montgomery e Johnson (1976):

- i. fácil entendimento;
- ii. aplicação não dispendiosa;
- iii. grande flexibilidade permitida pela variação da constante de suavização α ;
- iv. necessidade de armazenar somente \bar{Z}_t, α e Z_t ; e
- v. o valor de $\alpha = 2/(r - 1)$ fornece previsões semelhantes ao método MMS com parâmetro r .

Para os autores, a principal desvantagem é a dificuldade em determinar o valor mais apropriado da constante de suavização.

c) Suavização Exponencial de Holt (SEH)

Essa técnica se diferencia das técnicas anteriores na medida em que analisa séries temporais que apresentam tendência. Considerando, o caso de uma série temporal não sazonal, que é composta localmente da soma de nível, tendência e resíduo aleatório com média zero e variância constante (MORETTIN; TOLOI, 2004), temos a equação 24:

$$Z_t = \mu_t + T_t + a_t, \quad t = 1, \dots, N. \quad (24)$$

O método de SES quando aplicado a uma série que apresenta tendência linear positiva (ou negativa) fornece previsões que subestimam (ou superestimam) continuamente os valores reais. Para evitar esse erro sistemático, um dos métodos aplicáveis é o SEH. Esse método é similar, a diferença é que em vez de suavizar só o nível, ele usa uma nova constante de suavização para “modelar” a tendência da série.

Os valores do nível e da tendência da série, no instante t , serão estimados pelas equações 25 e 26 respectivamente:

$$\bar{Z}_t = AZ_t + (1 - A)(\bar{Z}_{t-1} + \widehat{T}_{t-1}), \quad 0 < A < 1 \text{ e } t = 2, \dots, N, \quad (25)$$

$$\widehat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\widehat{T}_{t-1}, \quad 0 < C < 1 \text{ e } t = 2, \dots, N, \quad (26)$$

Onde as variáveis A e C são denominadas constantes de suavização, as equações 25 e 26 como em todos os métodos de suavização, modificam estimativas prévias quando uma nova observação é obtida (MORETTIN; TOLOI, 2004).

d) Suavização Exponencial sazonal de Holt-Winters (HW)

Segundo Montgomery et al. (2008), esse tipo de suavização é apropriado para séries temporais que apresentam um padrão de comportamento mais complexo. Existem dois tipos de procedimentos cuja utilização depende das características da série considerada. Tais procedimentos são baseados em três equações com constantes de suavização diferentes, que são associadas a cada uma das componentes do padrão da série: nível, tendência e sazonalidade.

Ao analisar uma série com sazonalidade aditiva, têm-se a equação 27:

$$Z_t = \mu_t + T_t + F_t + a_t. \quad (27)$$

Onde o fator sazonal F_t e a tendência T_t são aditivos.

As estimativas do fator sazonal, nível e tendência da série são dadas pelas equações 28, 29 e 30:

$$\hat{F}_t = D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s}, 0 < D < 1, \quad (28)$$

$$\bar{Z}_t = A(Z_t - \bar{Z}_t) + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), 0 < A < 1, \quad (29)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1, \quad (30)$$

Onde A, C e D são constantes de suavização.

2.8.1.3 Modelos ARIMA

Nos modelos ARIMA a estratégia para construção do modelo será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados (MORETTIN; TOLOI, 2004). Os estágios do ciclo iterativo são:

a) uma classe geral de modelos é considerado para análise (especificação);

- b) há identificação de um modelo, com base na análise de auto correlações, auto correlações parciais e outros critérios;
- c) na fase de estimação, os parâmetros do modelo identificado são estimados;
- d) finalmente, na verificação ou diagnóstico do modelo ajustado, é realizada uma análise de resíduos a fim de verificar a adequação do modelo para os fins necessários.

Este modelo consiste em determinar quais os filtros que se deve utilizar na modelagem, onde a parte auto regressiva é representada por $AR(p)$, onde (p) indica a ordem do modelo. A parte que representa o número de diferenças (d) é dada por $I(d)$, onde d representa quantas diferenças são necessárias para tornar a serie estacionária. Em geral, utilizasse $d=1$ para tornar a série estacionaria em nível, e $d=2$ para estabilizar a variabilidade da série.

Modelos auto regressivos

Obtendo um modelo auto regressivo de ordem p , que denotaremos $AR(p)$:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t, \quad (31)$$

Se definirmos o operador auto regressivo estacionário de ordem p ,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad (32)$$

Então pode-se escrever:

$$\phi(B)\tilde{Z}_t = a_t, \quad (33)$$

Modelos de médias móveis

Obtendo um modelo de médias móveis de ordem q , que denotaremos $MA(q)$:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (34)$$

E sendo $\tilde{Z}_t = Z_t - \mu$, teremos:

$$\tilde{Z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t, \quad (35)$$

Onde,

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (36)$$

É o operador de médias móveis de ordem q .

3 METODOLOGIA

A condução de pesquisas científicas deve estar balizada em pressupostos metodológicos para que possa ser considerada válida e apresentar resultados coerentes (MARCONI; LAKATOS, 2010).

Dessa forma, os métodos de trabalho para atingir o objetivo proposto estão descritos neste capítulo, e abrange o delineamento da pesquisa, técnicas, etapas e instrumentos necessários para atender os objetivos propostos nesta dissertação de mestrado.

3.1 CENÁRIO DA PESQUISA

A atual pesquisa é realizada para atender as necessidades da empresa alvo do estudo. A empresa é uma distribuidora de energia localizada no estado do Rio Grande do Sul. A empresa realiza manutenções emergências na rede de distribuição de energia elétrica na região central do estado e necessita determinar a quantidade de mão de obra necessária para realizar as manutenções em tempo apropriado. O presente trabalho visa realizar uma previsão de demanda dessas ordens de manutenção emergenciais.

3.2 DELINEAMENTO DA PESQUISA

Quanto a sua natureza, esta pesquisa é classificada como pesquisa aplicada. A pesquisa de natureza aplicada visa encontrar uma solução para um problema imediato de uma sociedade ou organização empresarial. Compreende também, identificar tendências sociais, econômicas ou políticas que possam afetar determinada instituição (KOTHARI, 2004).

Quanto aos objetivos, a pesquisa é considerada exploratória, visto que tem como propósito um maior entrosamento com o tema de estudo. E assim, classificou-se também, como descritiva envolvendo técnicas padronizadas de coleta de dados, as quais, conforme Gil (2009), descreve as características de uma determinada população, fenômeno ou o estabelecimento de relação entre variáveis.

Em relação ao método de pesquisa, o trabalho é caracterizado por um estudo de modelagem e simulação. A utilização de modelos permite compreender melhor o

ambiente em questão, identificar problemas, formular estratégias e oportunidades e apoiar e sistematizar o processo de tomada de decisão (referencia).

Quanto aos seus procedimentos, esta pesquisa classifica-se como quantitativa. Segundo Neto e Pureza (2012), modelagem quantitativa são modelos abstratos descritos em linguagem matemática e computacional, que utilizam técnicas analíticas (matemáticas, estatísticas) e experimentais (simulação) para calcular valores numéricos das propriedades do sistema em questão, podendo ser usados para analisar os resultados de diferentes ações possíveis no sistema. Modelos quantitativos compreendem um conjunto de variáveis de controle que variam em um domínio específico e variáveis de desempenho que interferem a qualidade das decisões obtidas a partir de relações causais e quantitativas definidas entre essas variáveis.

3.3 PROCEDIMENTO DE COLETA DE DADOS

Para compreender com detalhamento o assunto trabalhado nesta pesquisa foi utilizada a técnica de pesquisa exploratória através da busca por referências bibliográficas de previsão de demanda de energia elétrica, ordens de manutenção e métodos causais quantitativos. A coleta de dados foi realizada a partir do acesso ao banco de dados de registro de ordens de manutenção emergencial disponibilizado pela empresa.

3.4 DESENVOLVIMENTO DA PESQUISA

A sequência de etapas que compõem o desenvolvimento desta pesquisa está delineada segundo os pressupostos referentes à previsão de demanda, demanda de energia elétrica, ordens de manutenção, regressão múltipla e series temporais. Propõe-se, para tanto, que o objetivo geral seja atingido por meio do cumprimento das cinco etapas apresentadas no quadro abaixo.

Etapa 1	Revisão bibliográfica em livros e revistas qualificadas nas áreas de previsão de demanda e sistemas de distribuição de energia elétrica.
Etapa 2	Analisar e entender as características da demanda de ordens de manutenção emergenciais existentes no sistema de distribuição de energia elétrica.
Etapa 3	Determinar o(s) método(s) de previsão de demanda a ser utilizado.
Etapa 4	Desenvolver um modelo de previsão de demanda para as ordens emergenciais que considere as características de demanda existentes.
Etapa 5	Realizar a previsão de demanda com os dados coletados.

Etapa 1 – Revisão bibliográfica: A primeira etapa teve como objetivo o melhor entendimento do cenário da pesquisa. A realização da revisão bibliográfica nas áreas de previsão de demanda e sistemas de distribuição de energia elétrica é importante para dar suporte ao desenvolvimento do presente estudo. Nesta etapa foram pesquisados os seguintes tópicos: previsão de demanda de energia elétrica, introdução a previsão de demanda, modelos de previsão de demanda de energia elétrica, regressão linear múltipla, chamadas emergências de manutenção e estudo de séries temporais.

Etapa 2 - Analisar e entender as características da demanda de ordens de manutenção emergenciais existentes no sistema de distribuição de energia elétrica: esta etapa foi realizada após o recebimento dos dados da empresa alvo do estudo. Os dados obtidos eram referentes a todas as ordens de chamados de manutenção da rede elétrica, nos períodos de 2013, 2014, 2015, 2016 e 2017, na região de Santa Maria e alguns municípios do estado do Rio Grande do Sul. Como o foco do trabalho se limita em realizar a previsão de demanda para um local específico, os dados que não são referentes ao município de Santa Maria não foram utilizados. Em função de problemas de falta de dados em alguns anos, o presente trabalho utilizou o ano de 2017, com os dados do município de Santa Maria, para o desenvolvimento desta etapa. Os dados brutos continham muitas informações, dentre elas, número da equipe designada para a realização da manutenção, tempo de deslocamento das equipes, localização, tempo de manutenção, horário em que foi realizado o serviço, causa ocorrente para ser realizada a chamada, detalhe da causa, qual componente foi danificado, previsão climática no momento da manutenção e várias outras questões relatadas sobre o serviço realizado. Após compreender o funcionamento dos processos para a realização das ordens de manutenção, iniciou-

se a mineração dos dados brutos para posteriormente serem trabalhados e simulados. Essa mineração foi realizada para eliminar: dados repetidos e os dados que não são classificados como manutenção emergencial, como por exemplo serviços de manutenção comercial na rede elétrica. Depois de analisar somente os dados que se referem ao objetivo principal do presente trabalho, os dados minerados como manutenção emergencial, foi realizado uma nova separação, classificando assim, manutenção emergencial causada por fatores climáticos e manutenção emergencial geral. Manutenção emergencial causada por fatores climáticos se refere às ordens de chamada de manutenção com o detalhe da causa sendo: inundação, descarga atmosférica, vendaval e temporal. Enquanto os dados referentes a manutenção emergencial geral, são dados que tem como detalhe da causa: falha de componente, poste com problema, defeito transitório, problemas com a vegetação (árvores), animais diversos, transformadores e outras falhas que causam chamadas emergenciais. Como o objetivo do presente trabalho é a realização de uma previsão de demanda de ordens de manutenção emergenciais, os dados trabalhados são referentes a quantidade de horas diárias que foram utilizadas para a realização dos serviços de manutenções emergenciais do ano de 2017.

Etapa 3 - Determinar o(s) método(s) de previsão de demanda a ser utilizado(s):

Após a realização da etapa 2 da metodologia, os dados para o desenvolvimento foram minerados e classificados em dois grupos, dados referentes a manutenção emergencial causadas por fatores climáticos e manutenção emergencial geral. Com base na análise dos dados coletados, determinaram-se o uso de Regressão múltipla para o grupo de dados de manutenção emergencial causadas por fatores climáticos e séries temporais para os dados de manutenção emergencial geral, para a realização da previsão de demanda.

Para o grupo de dados classificados como manutenção emergencial causadas por fatores climáticos, o método causal Regressão linear múltipla é definido para a realização da previsão. Este método foi escolhido pois irá utilizar as variáveis climáticas para explicar a demanda de ordens de manutenção emergências, dentre essas variáveis climáticas, temos: máxima rajada no dia, soma da precipitação, insolação, evaporação, umidade relativa média, velocidade média do vento, *Cooling Degree Days* e *Heating Degree Days*. Esses dados referentes as variáveis climáticas

foram obtidos a partir do (INMET) Instituto Nacional de Meteorologia (2018). Utilização de ferramentas para análises estatísticas: a primeira análise realizada foi a correlação entre a variável dependente “Y”, que representa as horas de manutenção emergencial, com as variáveis climáticas já citadas, esta correlação serve para medir a associação entre as variáveis. Alguns testes de hipótese foram realizados para definir se os dados atendem as premissas para se utilizar o método de Regressão Múltipla, dentre eles: teste ANOVA, que tem como objetivo comparar a distribuição das variáveis do problema. Teste de hipótese para analisar homocedasticidade, esse teste tem como finalidade analisar os resíduos da regressão, sabendo se estes resíduos estão correlacionados com alguma variável do modelo. Teste de hipótese de *Durbin-Watson*, para analisar a presença de auto correlação (dependência) nos resíduos do modelo de regressão múltipla. Teste de hipóteses de *Kolmogorov-Smirnov*, para medir a normalidade da distribuição. Teste de hipótese de fator de inflação da variância, que analisa a existência de multicolineariedade entre variáveis. A fim de resolver problemas com os resíduos obtidos a partir da Regressão múltipla, foi realizado a transformação logarítmica nos dados para realizar novas análises.

O outro método a ser utilizado é de séries temporais, por ser um modelo temporal, a previsão é realizada para explicar a demanda a partir dos dados históricos fornecidos pela empresa alvo do estudo. Esse método tem como objetivo utilizar os dados referentes ao grupo manutenção emergencial geral para a realização da previsão. As primeiras análises realizadas neste grupo de dados tiveram como objetivo identificar as características dos dados perante a metodologia de séries temporais, os testes de hipótese de *Augmented Dickey-Fuller* e *Phillips-Perron* servem para analisar se a série possui ou não raízes unitárias, ou seja, se a série de dados é estacionária.

Etapa 4 - Desenvolvimento de um modelo de previsão de demanda para ordens de manutenção emergenciais na rede elétrica: este modelo será realizado a partir dos métodos de previsão definidos na etapa 3: método de regressão linear múltipla e método de series temporais.

a) Método de Regressão Linear Múltipla

Este método será utilizado para desenvolver um modelo de previsão de demanda para os dados referentes a manutenção emergencial causadas por fatores

climáticos, com o objetivo de explicar a partir de variáveis climáticas o comportamento da demanda.

Os ajustes dos dados para o desenvolvimento deste método provêm da mineração dos dados e de ferramentas de análises estatísticas.

O desenvolvimento do modelo será realizado no *software STATISTICA*®.

b) Método de Séries Temporais

Este método será utilizado para desenvolver um modelo de previsão de demanda para os dados referentes a manutenção emergencial geral. Este método vai trabalhar com análises de sazonalidade, tendência, estacionariedade, normalidade e correlação.

Dentre os modelos de séries temporais, poderemos utilizar para realizar a previsão: tendência e sazonalidade, suavização exponencial, médias móveis simples, suavização exponencial de *Holt-Winters*, modelos ARIMA e ARMA.

O desenvolvimento do modelo será realizado no *software R studio*®.

Etapa 5 - Realização da previsão de demanda de ordens de manutenção emergências na rede elétrica: nessa etapa, será realizada a previsão de demanda diária de horas de manutenção emergenciais a partir dos modelos desenvolvidos na etapa 4 e dos dados obtidos com a empresa. A partir da previsão para um período de teste de 7 dias de horas de manutenção emergenciais, a validação do modelo utilizado será realizada a partir do teste MAPE, erro absoluto médio percentual, ele é calculado encontrando o erro absoluto em cada período, e então calculando a média desses erros absolutos percentuais. Fornece uma indicação de quão grandes os erros de previsão estão na comparação com os valores dos dados reais (HANKE et al., 2001).

4 RESULTADOS

4.1 ANÁLISE DOS DADOS

A primeira etapa realizada foi a separação dos dados brutos fornecidos pela empresa distribuidora de energia elétrica. Esses dados são referentes a todas as horas de atividades de manutenção emergenciais que a empresa realizou na cidade de Santa Maria no ano de 2017. Sendo uma amostra de 365 dias, e que desta amostra 7 dias serão usados para a etapa de treino e validação do modelo. Os dados foram minerados em horas de ordens de manutenção emergencial com causas relacionadas ao meio ambiente (descarga elétrica, vento, etc.) e causas gerais (desgaste de equipamentos, acidentes, etc.).

4.2 REGRESSÃO MÚLTIPLA

A primeira separação dos dados foi realizada a partir de um filtro “Meio Ambiente”, em que os dados são referentes a todos os chamados de emergências em que a empresa classificou como causa “Meio ambiente”.

Uma regressão linear múltipla foi realizada com os dados da demanda de ordens emergenciais como variável dependente “Y”, e climáticas como variáveis independentes “ x_1, x_2, \dots, x_n ”. Como variáveis climáticas utilizaram-se: máxima rajada no dia, soma da precipitação, insolação, evaporação, umidade relativa média, velocidade média do vento, *Cooling Degree Days* e *Heating Degree Days*. Para os dados de clima de Santa Maria, utilizou-se a base de dados do INMET que possui uma estação meteorológica na cidade.

Para as análises estatísticas dos dados obtidos, utilizaram-se os *softwares* STATISTICA® e R studio®.

A primeira análise realizada foi a relação das correlações entre todas as variáveis envolvidas no modelo (Figura 2). Observa-se que a correlação é realizada para medir quanto a variável dependente é explicada pela variável independente, também mensura o grau de variância comum entre as variáveis (GUJARATI; PORTER, 2011). Como podemos analisar na Figura 2, temos que a variável dependente “Y” do modelo é “emerg”, e assim podemos analisar o valor da correlação dela com as variáveis independentes.

Figura 2 – Correlação entre todas as variáveis envolvidas no modelo

Variable	Correlations (Spreadsheet1)								
	Max rajada	precipitação	Insolação	Evaporação	Umidade	vel med vento	HDD	CDD	emerg
Max rajada	1,000000	0,320192	-0,107289	0,418020	-0,231476	0,683151	-0,061600	0,043665	0,407594
precipitação	0,320192	1,000000	-0,378024	-0,090149	0,338943	0,061631	-0,239886	0,331246	0,184849
Insolação	-0,107289	-0,378024	1,000000	0,159254	-0,661201	0,000735	0,541793	-0,684710	-0,052640
Evaporação	0,418020	-0,090149	0,159254	1,000000	-0,481609	0,372643	0,399415	-0,247537	-0,016146
Umidade	-0,231476	0,338943	-0,661201	-0,481609	1,000000	-0,376206	-0,375479	0,632224	0,030977
vel med vento	0,683151	0,061631	0,000735	0,372643	-0,376206	1,000000	-0,203591	0,020035	0,099748
HDD	-0,061600	-0,239886	0,541793	0,399415	-0,375479	-0,203591	1,000000	-0,643429	-0,084346
CDD	0,043665	0,331246	-0,684710	-0,247537	0,632224	0,020035	-0,643429	1,000000	0,079204
emerg	0,407594	0,184849	-0,052640	-0,016146	0,030977	0,099748	-0,084346	0,079204	1,000000

Fonte: Elaborado pelo autor (2019) com STATISTICA®.

Na sequência, realizou-se uma regressão linear múltipla entre a variável dependente horas emergenciais e as demais variáveis climáticas. O modelo de regressão múltipla foi ajustado e os coeficientes calculados, conforme Figura 3. As variáveis em vermelho são as variáveis que resultaram significativas no modelo.

Figura 3 – Coeficientes do modelo de regressão múltipla

Regression Summary for Dependent Variable: emerg (Spreadsheet1)						
R= ,51747919 R²= ,26778471 Adjusted R²= ,25133044						
F(8,356)=16,274 p<0,0000 Std.Error of estimate: 234,81						
N=365	Beta	Std.Err. of Beta	B	Std.Err. of B	t(356)	p-level
Intercept			-224,837	198,9819	-1,12994	0,259263
Max rajada	0,744166	0,071473	55,572	5,3374	10,41187	0,000000
precipitação	-0,042383	0,054441	-0,805	1,0342	-0,77851	0,436786
Insolação	0,145155	0,075789	9,838	5,1369	1,91524	0,056262
Evaporação	-0,152349	0,064494	-16,605	7,0295	-2,36224	0,018703
Umidade	0,048611	0,084013	1,083	1,8717	0,57861	0,563220
vel med vento	-0,351336	0,072716	-113,469	23,4848	-4,83161	0,000002
HDD	-0,095930	0,073345	-11,669	8,9218	-1,30793	0,191740
CDD	0,037008	0,076002	4,522	9,2858	0,48694	0,626603

Fonte: Elaborado pelo autor (2019).

Pode-se verificar que poucas variáveis resultaram significativas para o modelo, além de um R² relativamente baixo (0,26) ou seja, a variável dependente é pouco explicada pelo modelo selecionado.

Ao analisar os resíduos do modelo de regressão múltipla, verificou-se que o modelo é inadequado, pois seus resíduos são heterocedásticos, ou seja, a variância entre eles não é constante. Essa conclusão baseou-se nos testes de hipótese de Breusch-Pagan, segundo Pesaran et al. (2008), o teste pressupõe que existe uma relação linear entre o quadrado dos resíduos padronizados e as variáveis explicativas

do modelo. Segundo Breusch e Pagan (1979), o teste de hipótese de Breusch-Pagan examina se os resíduos estimados estão correlacionados com alguma variável do modelo.

O teste considera que a hipótese nula é de homocedasticidade, tem-se um p valor $< 0,05$, logo rejeita-se a hipótese nula de que os resíduos sejam homocedásticos (Figura 4).

Figura 4 – Teste de hipótese para homocedasticidade

```

studentized Breusch-Pagan test
data: regsomeioamb
BP = 53.889, df = 8, p-value = 7.253e-09

> car::ncvTest(regsomeioamb)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 4405.891, Df = 1, p = < 2.22e-16

```

Fonte: Elaborado pelo autor (2019).

A partir disto, foi realizado uma nova seleção nos dados, sendo eles selecionados pelo detalhamento da causa, estas foram: descarga atmosférica, ventos e inundação. Outra questão foi a implementação de uma nova variável independente, representada pela variável dependente com uma defasagem. Essa defasagem é representada por uma diferença de tempo nos dados observados, sendo assim, essa variável entra no modelo para ajudar na explicação, fazendo referência se o dia anterior causa impacto no dia posterior. Este conceito segundo Mattos (2018), a defasagem de uma variável é apresentada na equação 37:

$$\Delta X_t = X_t - X_{t-1}, \quad (37)$$

Os coeficientes calculados e o modelo de regressão múltipla ajustado é apresentado na Figura 5, onde as variáveis em vermelho são consideradas significativas pelo modelo.

Figura 5 – Coeficientes do modelo de regressão múltipla

Regression Summary for Dependent Variable: emerg (Spreadshee R= ,55002896 R²= ,30253185 Adjusted R²= ,28479961 F(9,354)=17,061 p<0,0000 Std.Error of estimate: 187,62						
N=364	Beta	Std.Err. of Beta	B	Std.Err. of B	t(354)	p-level
Intercept			-175,246	159,3047	-1,10007	0,272049
Max rajada	0,764526	0,070850	46,773	4,3345	10,79080	0,000000
precipitação	-0,056038	0,053397	-0,869	0,8281	-1,04946	0,294681
Insolação	0,133442	0,074659	7,392	4,1359	1,78736	0,074735
Evaporação	-0,159230	0,063134	-14,169	5,6179	-2,52210	0,012103
Umidade	0,036491	0,082503	0,664	1,5005	0,44230	0,658542
vel med vento	-0,377287	0,071715	-99,560	18,9243	-5,26095	0,000000
HDD	-0,098858	0,071927	-9,822	7,1463	-1,37443	0,170179
CDD	0,019085	0,074375	1,904	7,4211	0,25661	0,797629
y defasado	0,175432	0,044694	0,175	0,0447	3,92519	0,000104

Fonte: Elaborado pelo autor (2019).

Novamente obtivemos um R^2 relativamente baixo, de aproximadamente 30% e podemos analisar que apenas quatro das variáveis são significantes para o modelo proposto, resultados na Figura 5. Ao analisar os resíduos do modelo de regressão múltipla, verificou-se que o modelo é inapropriado, pois seus resíduos são heterocedásticos, ou seja, a variância entre eles não é constante. Essa conclusão baseou-se novamente nos testes de hipótese de Breusch-Pagan, Figura 6.

Figura 6 – Teste de hipótese para homocedasticidade

```
studentized Breusch-Pagan test
data: regmaisdet
BP = 53.535, df = 9, p-value = 2.314e-08
> car::ncvTest(regmaisdet)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
chisquare = 4256.894, Df = 1, p = < 2.22e-16
```

Fonte: Elaborado pelo autor (2019).

A fim de resolver o atual problema de homocedasticidade dos resíduos da regressão, foi utilizado o método de transformação logarítmica. Para a transformação logarítmica foi necessário a exclusão dos dados com valores iguais a zero da nossa variável dependente. Esses dados zerados representam dias em que o sistema falhou em coletar dados e dias em que ordens emergenciais não foram registradas. Além

disso, foi realizada a exclusão dos dados *outliers*, para melhor ajuste do modelo, totalizando uma amostra de 345 dias para o desenvolvimento do modelo.

As variáveis foram reconsideradas a partir do método da defasagem, sendo elas: Y (tempo de manutenção), X1 (Y defasado), X2 (máxima rajada no dia), X3 (máxima rajada defasada), X4 (precipitação), X5 (precipitação defasada), X6 (insolação), X7 (evaporação), X8 (umidade) e X9 (velocidade média vento).

A correlação entre as variáveis do novo modelo ajustado é apresenta na Figura 7. Podemos analisar que o valor das correlações entre a variável dependente com as variáveis independentes tem valores maiores, ou seja, correlação mais forte, isso comparando com a Figura 2 que apresentava as variáveis testadas para o modelo anterior e descartado.

Figura 7 – Correlação entre variáveis modelo de regressão

Variable	Correlations (Spreadsheet2)									
	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
X1	1,000000	-0,046146	0,374960	0,038793	0,272119	-0,082473	-0,086939	0,104987	0,031240	0,395409
X2	-0,046146	1,000000	0,237313	0,169961	-0,186391	-0,106595	0,489514	-0,272839	0,747131	0,410111
X3	0,374960	0,237313	1,000000	0,003873	0,154221	-0,174689	0,266788	0,038399	0,253170	0,156676
X4	0,038793	0,169961	0,003873	1,000000	0,235020	-0,416778	-0,154833	0,586058	-0,094771	0,274325
X5	0,272119	-0,186391	0,154221	0,235020	1,000000	-0,179096	-0,581381	0,351288	-0,158991	0,093081
X6	-0,082473	-0,106595	-0,174689	-0,416778	-0,179096	1,000000	0,182161	-0,435307	-0,105757	-0,111283
X7	-0,086939	0,489514	0,266788	-0,154833	-0,581381	0,182161	1,000000	-0,458348	0,388737	0,121765
X8	0,104987	-0,272839	0,038399	0,586058	0,351288	-0,435307	-0,458348	1,000000	-0,344263	-0,073431
X9	0,031240	0,747131	0,253170	-0,094771	-0,158991	-0,105757	0,388737	-0,344263	1,000000	0,203864
Y	0,395409	0,410111	0,156676	0,274325	0,093081	-0,111283	0,121765	-0,073431	0,203864	1,000000

Fonte: Elaborado pelo autor (2019).

Na Figura 8 temos os valores da regressão múltipla para o modelo ajustado. As variáveis em vermelho (p valor <0,05) foram consideradas significativas pelo modelo de regressão.

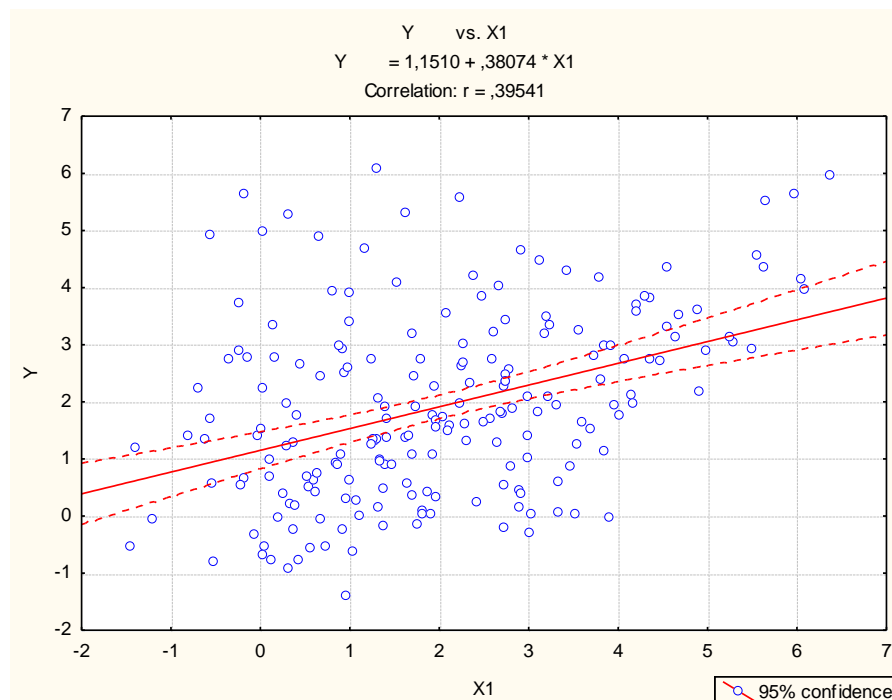
Figura 8 – Valores dos coeficientes do modelo de regressão

Regression Summary for Dependent Variable: Y (Spreadsheet)						
R= ,65789960 R²= ,43283188 Adjusted R²= ,40624587						
F(9,192)=16,280 p<0,0000 Std.Error of estimate: 1,2431						
N=202	Beta	Std.Err. of Beta	B	Std.Err. of B	t(192)	p-level
Intercept			7,46790	3,623685	2,06086	0,040665
X1	0,454561	0,060879	0,43770	0,058622	7,46659	0,000000
X2	0,528335	0,098584	2,48847	0,464333	5,35924	0,000000
X3	-0,079473	0,069061	-0,36777	0,319585	-1,15077	0,251258
X4	0,260686	0,079940	0,06462	0,019815	3,26104	0,001313
X5	0,048726	0,074935	0,01186	0,018237	0,65025	0,516306
X6	-0,046103	0,065701	-0,01588	0,022626	-0,70170	0,483715
X7	-0,024905	0,087502	-0,04613	0,162076	-0,28463	0,776238
X8	-0,256586	0,084332	-2,31177	0,759812	-3,04256	0,002674
X9	-0,236026	0,089759	-1,02301	0,389045	-2,62954	0,009243

Fonte: Elaborado pelo autor (2019).

Afim de analisar as variáveis que obtiveram um p-valor significativo para o atual modelo desenvolvido, foi realizado os gráficos de dispersão entre estas variáveis independentes significativas com a variável depende do modelo de regressão. As Figuras 9, 10, 11, 12 e 13 apresentam os gráficos de dispersão da variável dependente Y com as variáveis significativas X1, X2, X4, X8 e X9, respectivamente.

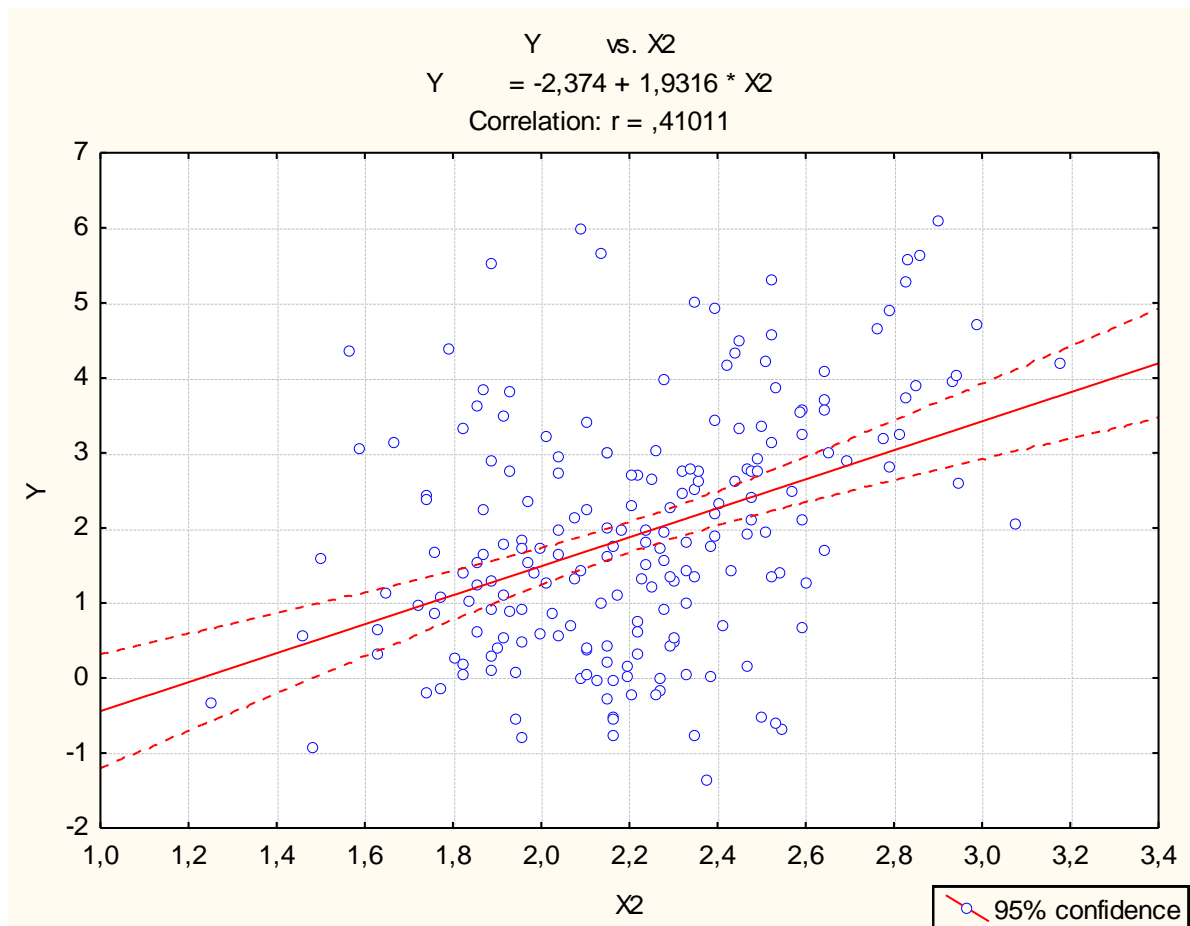
Figura 9 – Gráfico de dispersão entre as variáveis X1 (emergência defasada) vs. Y (emergência)



Fonte: Elaborado pelo autor (2019).

A Figura 9 representa o resultado gerado pela combinação da variável dependente Y (emergência) que é nossa série histórica de horas de manutenção emergencial por dia vs. X1 (emergência defasada). Temos um coeficiente de correlação $r = 0,39$, ou seja, uma correlação positiva, mas de média intensidade, isso representa a grande dispersão dos pontos representados no gráfico.

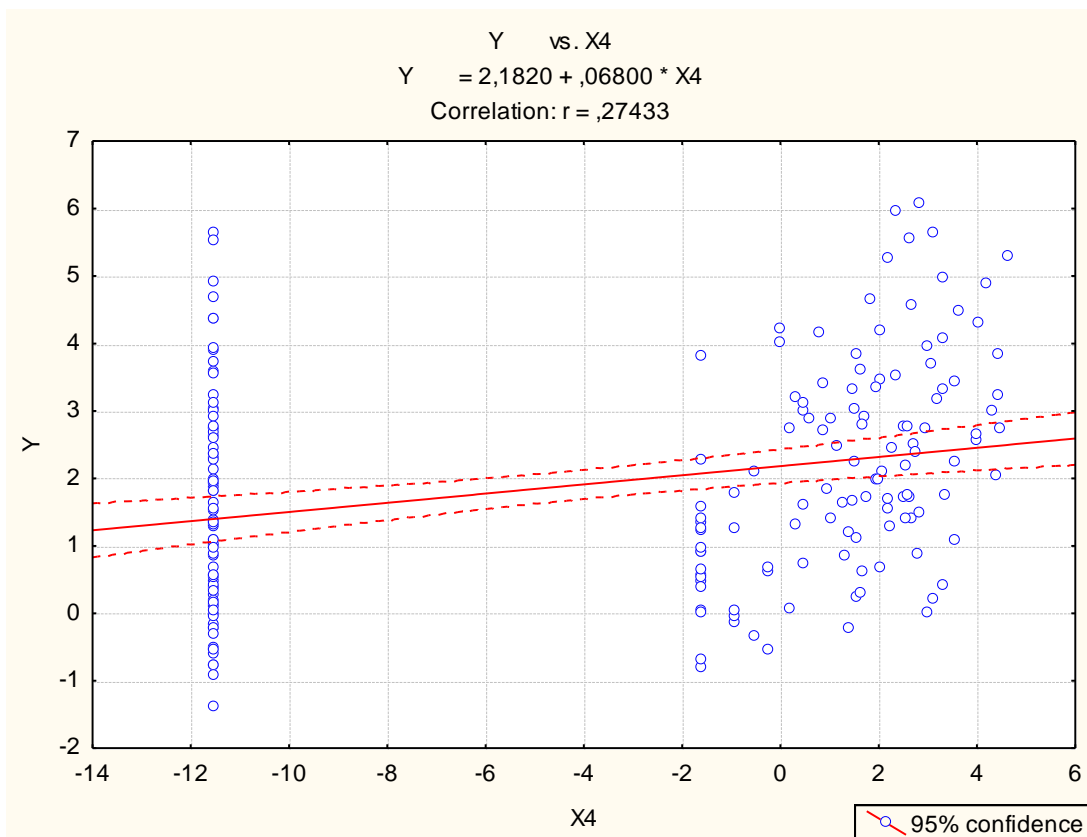
Figura 10 – Gráfico de dispersão entre as variáveis X2 (máxima rajada dia) vs. Y (emergência)



Fonte: Elaborado pelo autor (2019).

A Figura 10 representa o resultado gerado pela combinação da variável dependente Y (emergência) vs. X2 (máxima rajada dia). Temos um coeficiente de correlação $r = 0,41$, ou seja, uma correlação positiva e de média intensidade, isso representa a grande dispersão dos pontos representados no gráfico.

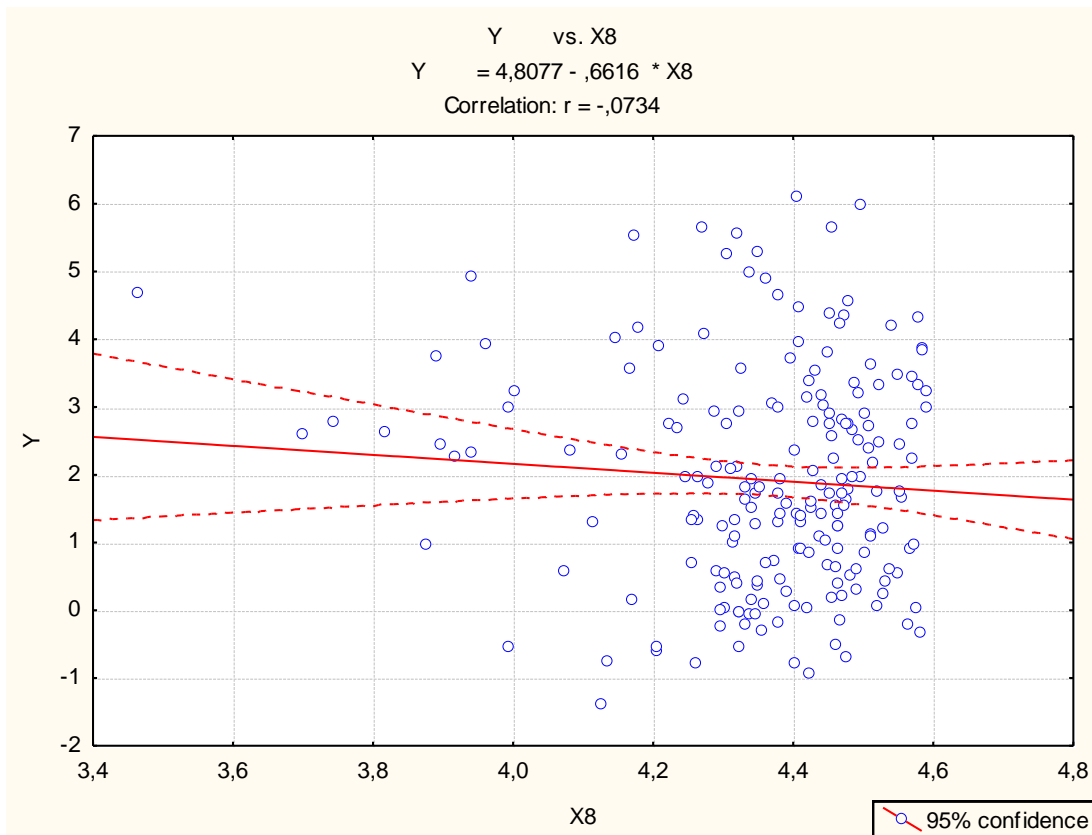
Figura 11 – Gráfico de dispersão entre as variáveis X4 (precipitação) vs. Y (emergência)



Fonte: Elaborado pelo autor (2019).

A Figura 11 representa o resultado gerado pela combinação da variável dependente Y (emergência) vs. X4 (precipitação). Temos um coeficiente de correlação $r = 0,27$, ou seja, uma correlação positiva e de baixa intensidade, a explicação para que exista uma linha de pontos separada da nuvem de pontos clássicos de um gráfico de dispersão é pelo fato dos dias em que não houve registro de valor (precipitação).

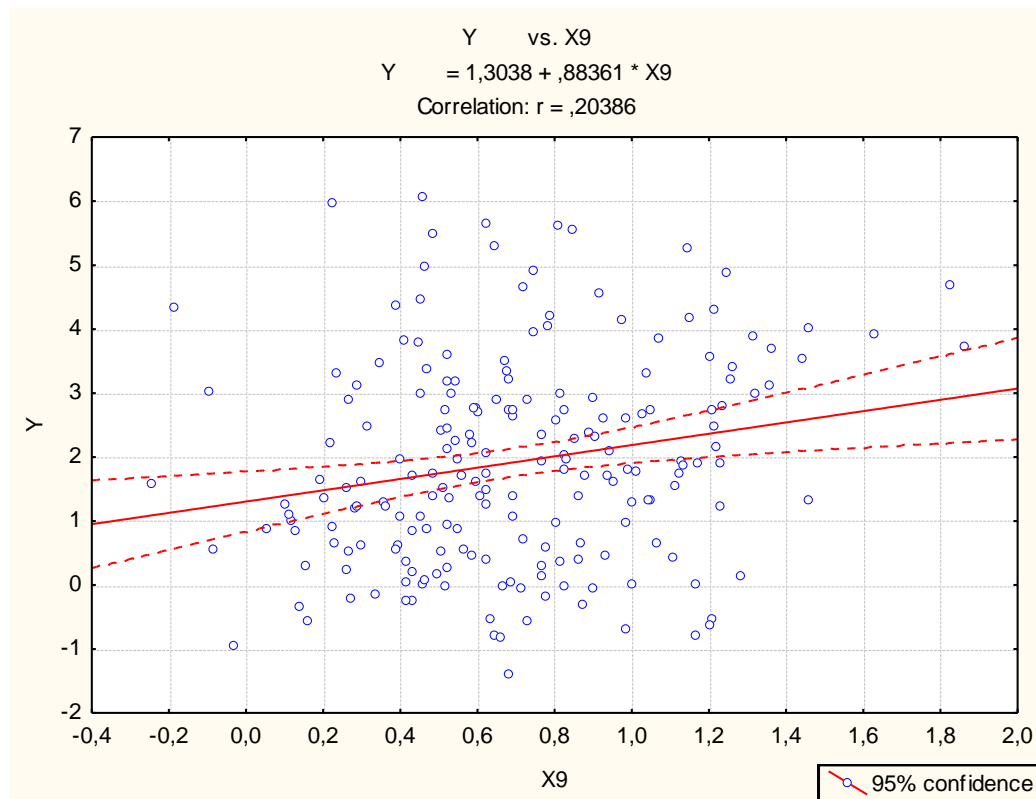
Figura 12 – Gráfico de dispersão entre as variáveis X8 (Umidade) vs. Y (emergência)



Fonte: Elaborado pelo autor (2019).

A Figura 12 representa o resultado gerado pela combinação da variável dependente Y (emergência) vs. X8 (umidade). Temos um coeficiente de correlação $r = -0,07$, ou seja, uma correlação negativa e de baixa intensidade, o que ajuda a entender a grande dispersão da nuvem de pontos identificadas no gráfico.

Figura 13 – Gráfico de dispersão entre as variáveis X9 (Velocidade média vento) vs. Y (emergência)



Fonte: Elaborado pelo autor (2019).

Para o melhor ajuste do modelo, as variáveis não significativas (p valor $>0,05$) foram excluídas. Esse procedimento foi realizado com a exclusão de uma variável por vez. O modelo de regressão em que todas as variáveis são significativas é apresentado na Figura 14.

Figura 14 – Valores dos coeficientes do modelo de regressão

Regression Summary for Dependent Variable: Y (Spreadsheet)						
R= ,65080506 R ² = ,42354722 Adjusted R ² = ,40884179						
F(5,196)=28,802 p<0,0000 Std.Error of estimate: 1,2404						
N=202	Beta	Std.Err. of Beta	B	Std.Err. of B	t(196)	p-level
Intercept			6,21611	3,416536	1,81942	0,070374
X1	0,438289	0,054967	0,42204	0,052929	7,97365	0,000000
X2	0,485211	0,092894	2,28536	0,437533	5,22328	0,000000
X4	0,291586	0,076597	0,07228	0,018986	3,80674	0,000188
X8	-0,235701	0,075952	-2,12360	0,684311	-3,10327	0,002197
X9	-0,225853	0,087368	-0,97892	0,378683	-2,58506	0,010462

Fonte: Elaborado pelo autor (2019).

As estimativas dos valores estatísticos do modelo foram de R^2 múltiplo igual a 0,423 e R^2 ajustado de 0,408. O coeficiente de determinação do modelo ajustado é de 0,4235, ou seja, 42,35% da variável dependente Y (tempo de manutenção) é explicada pelas variáveis independentes do modelo de regressão.

Os testes de homocedasticidade foram satisfatórios, sendo apresentado na Figura 15.

Figura 15 – Teste de hipótese para homocedasticidade

```
> lmtest::bptest(regteste2) # Breusch-Pagan test
      studentized Breusch-Pagan test
data:  regteste2
BP = 9.6202, df = 5, p-value = 0.08674

> car::ncvTest(regteste2)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 2.380327, Df = 1, p = 0.12287
```

Fonte: Elaborado pelo autor (2019).

Ao interpretar o teste de hipótese para homocedasticidade, temos um p valor = 0,08674, ou seja, p valor > 0,05, sendo assim, não foi possível rejeitar a hipótese nula, ou seja, não existe forte evidência que a variação dos resíduos não é constante, portanto, assume-se que eles sejam homocedásticos.

Foi realizado o teste de *Durbin-Watson*, para analisar a presença de auto correlação (dependência) nos resíduos do modelo de regressão múltipla (JOHNSTON, 1977).

O resultado do teste é apresentado na Figura 16, o p valor = 0,82, ou seja, p valor > 0,05, sendo assim, não foi possível rejeitar a hipótese nula, portanto, assume-se que os resíduos são independentes.

Figura 16 – Teste de Durbin Watson

```
> durbinwatsonTest(regteste2)
lag Autocorrelation D-w Statistic p-value
1 -0.02181643 2.037078 0.82
Alternative hypothesis: rho != 0
```

Fonte: Elaborado pelo autor (2019).

A Figura 17 apresenta um resumo de resultados dos resíduos da regressão. Podemos analisar dois fatores importantes: i) a coluna do *Standard Residual*, apresenta valores entre -3 e 3, o que significa que não existem mais outliers no modelo de regressão; ii). A coluna *Cook's Distance* apresenta valores entre -1 e 1, o que representa que não temos valores de outliers, que possam ser influentes no modelo de regressão. Segundo Kim et al. (2001), a distância de Cook serve para detectar observações influentes, que podem levar a exclusão de uma dada observação.

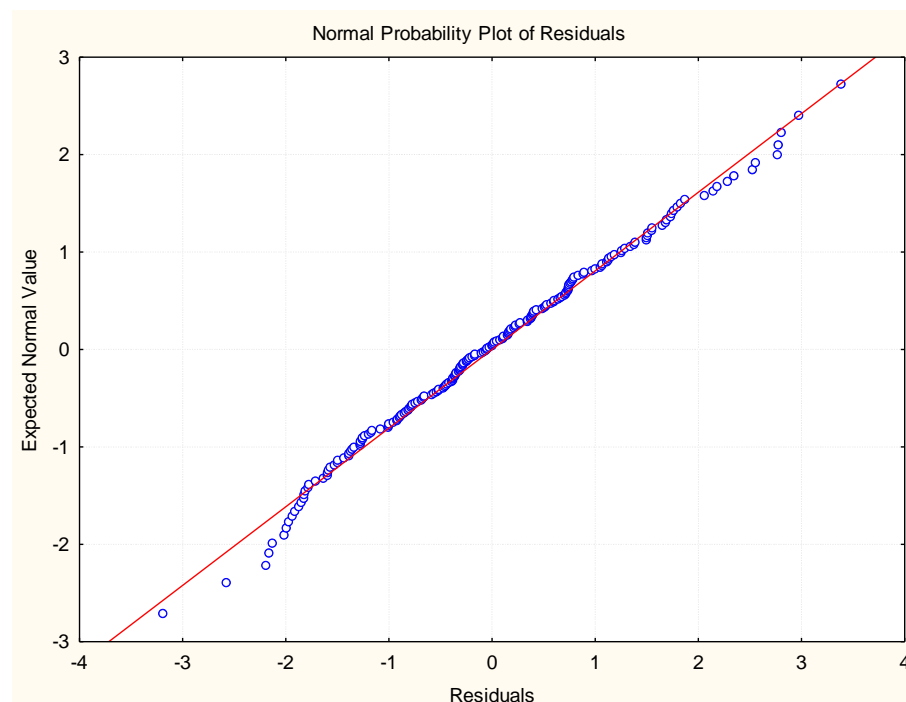
Figura 17 – Resumo de resultados dos resíduos

Predicted & Residual Values (Spreadsheet2)									
Dependent variable: Y									
Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
Minimum	-1,38741	-0,854331	-3,18651	-2,64280	-2,56899	0,127060	1,11411	-3,25792	0,000000
Maximum	6,08757	4,544397	3,38703	2,49929	2,73065	0,507123	32,60326	3,54338	0,060014
Mean	1,92037	1,920372	-0,00000	-0,00000	0,00000	0,206571	4,97525	0,00577	0,005648
Median	1,74586	1,818038	-0,04623	-0,09747	-0,03727	0,195965	4,02198	-0,04719	0,001994

Fonte: Elaborado pelo autor (2019).

A Figura 18 ajuda a analisar visualmente se os resíduos estão distribuídos normalmente. Para confirmar tal característica foi realizado o teste de hipóteses de *Kolmogorov-Smirnov*, apresentado na Figura 14.

Figura 18 – Normal P Plot dos resíduos



Fonte: Elaborado pelo autor (2019).

Figura 19 – Teste *Kolmogorov-Smirnov*

```
> ks.test(residregteste2, "pnorm")

one-sample Kolmogorov-Smirnov test

data: residregteste2
D = 0.079946, p-value = 0.1512
alternative hypothesis: two-sided
```

Fonte: Elaborado pelo autor (2019).

Como resultado do teste temos um p valor = 0,1512, ou seja, p valor > 0,05, sendo assim, não é possível rejeitar a hipótese nula, portanto, assume-se que os resíduos seguem uma distribuição normal padrão.

A Figura 20 apresenta o valor do fator de inflação da variância, que representa a existência de multicolineariedade entre as variáveis.

Figura 20 – Valores de fator de inflação da variância

Current Status of Sweep Matrix; DV: Y (Spreadsheet2)						
	X1	X2	X4	X8	X9	Y
X1	-1,02731	-0,16546	0,01508	0,12249	0,19931	0,438289
X2	-0,16546	-2,93405	1,13448	-0,74257	2,04916	0,485211
X4	0,01508	1,13448	-1,99489	1,27061	-0,59971	0,291586
X8	0,12249	-0,74257	1,27061	-1,96144	-0,00386	-0,235701
X9	0,19931	2,04916	-0,59971	-0,00386	-2,59538	-0,225853
Y	0,43829	0,48521	0,29159	-0,23570	-0,22585	0,576453

Fonte: Elaborado pelo autor (2019).

Ao analisar a Figura 20, observa-se que estes valores em módulo são menores que 4, o que confirma que não existe multicolineariedade entre as variáveis.

Com as verificações a partir de ferramentas estatísticas e de mineração de dados concluímos que todos os testes foram satisfatórios para confirmar que o modelo desenvolvido segue os padrões necessários para se realizar a previsão com os dados utilizados, sendo as variáveis independentes e a variável dependente.

O modelo de regressão linear múltipla pode ser descrito a partir de sua equação 38:

$$Y = 6,216 + 0,422 X1 + 2,285 X2 + 0,072 X4 - 2,123 X8 - 0,978 X9 \quad (38)$$

Para a realização da previsão a partir do modelo de regressão linear foi utilizado um período de treino de 7 dias. Para analisar os erros gerados pela previsão utilizou-se o teste MAPE, erro absoluto médio percentual, apresentado na Tabela 1, a equação 39 apresenta o conceito para se calcular o MAPE.

$$\frac{|Real-Previsão|}{|Real|} * 100 \quad (39)$$

Tabela 1 – MAPE, erro absoluto médio percentual

Previsão	Real	MAPE
2,67 horas	3,42 horas	22%
2,93 horas	3,90 horas	25%
3,12 horas	4,01 horas	22%
3,77 horas	3,13 horas	20%
2,10 horas	3,25 horas	35%
3,71 horas	2,63 horas	41%
2,09 horas	3,10 horas	33%
Total 20,42 horas	Total 23,47 horas	Total 28%

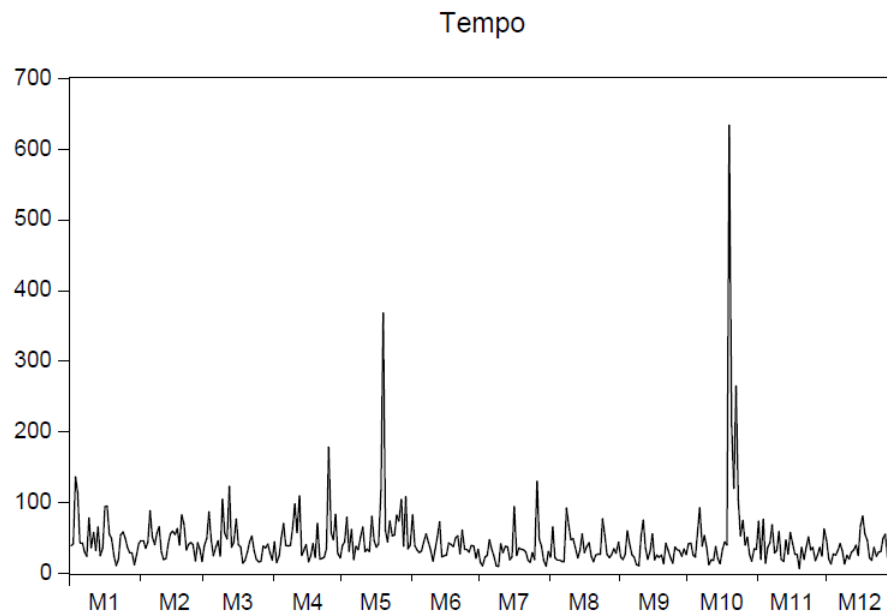
Fonte: Elaborado pelo autor (2019).

Considerando o resultado de 28% na comparação entre os erros dos dados previstos com os dados reais coletados, significa que o modelo pode ser utilizado para a previsão da demanda de ordens de manutenção com um erro de 28%.

4.3 SÉRIES TEMPORAIS

A segunda parte da previsão, não relacionada a fatores climáticos, foi realizada a partir da metodologia de séries temporais. Foram utilizados dados de horas em ordens de manutenção da rede elétrica da cidade de Santa Maria referentes ao ano de 2017. A Figura 21 apresenta os dados originais da série temporal do presente estudo. Este gráfico foi desenvolvido a partir do software EViews®.

Figura 21 – Dados das horas de manutenção emergenciais



Fonte: Elaborado pelo autor (2019).

Para iniciar a modelagem da série temporal devemos analisar alguns aspectos das mesmas para saber se atendem os requisitos para o desenvolvimento da previsão. A primeira questão a ser analisada sobre a série é a estacionariedade. Para isso foram realizados dois testes de hipóteses que consistem em analisar se a série possui raízes unitárias. Em caso positivo, a série é não estacionária, em caso negativo, a série é estacionária. Os testes realizados foram: *Augmented Dickey-Fuller* e *Phillips-Perron*. Para Mattos (2018) o desenvolvimento do teste DF foi motivado pela necessidade de verificar se uma série precisa ser diferenciada para se tornar estacionária, se o gráfico da série temporal indicar que ela é não estacionaria, então devemos diferenciá-la até apresentar um padrão estacionário.

As Figuras 22, 23, 24 e 25 apresentam os resultados dos testes. Cada teste foi feito com dois parâmetros diferentes, analisando se a série apresenta uma tendência linear e se a série não apresenta média zero.

Figura 22 – Teste de hipótese *Augmented Dickey-Fuller* com tendência linear

Augmented Dickey-Fuller Unit Root Test on TEMPO

Null Hypothesis: TEMPO has a unit root Exogenous: Constant, Linear Trend Lag Length: 2 (Automatic - based on SIC, maxlag=16)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-7.710661	0.0000
Test critical values:	1% level		-3.983612	
	5% level		-3.422286	
	10% level		-3.133995	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(TEMPO) Method: Least Squares Date: 12/10/18 Time: 16:52 Sample (adjusted): 1/04/2017 12/31/2017 Included observations: 362 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
TEMPO(-1)	-0.542904	0.070410	-7.710661	0.0000
D(TEMPO(-1))	-0.198663	0.064085	-3.099992	0.0021
D(TEMPO(-2))	-0.172267	0.051832	-3.323586	0.0010
C	25.75750	5.653699	4.555866	0.0000
@TREND("1/01/2017")	-0.009699	0.021322	-0.454886	0.6495
R-squared	0.379267	Mean dependent var		-0.315272
Adjusted R-squared	0.372312	S.D. dependent var		53.37369
S.E. of regression	42.28625	Akaike info criterion		10.34052
Sum squared resid	638361.2	Schwarz criterion		10.39427
Log likelihood	-1866.634	Hannan-Quinn criter.		10.36189
F-statistic	54.53161	Durbin-Watson stat		1.975717
Prob(F-statistic)	0.000000			

Fonte: Elaborado pelo autor (2019).

Figura 23 – Teste de hipótese *Phillips-Perron* com tendência linear

Phillips-Perron Unit Root Test on TEMPO

Null Hypothesis: TEMPO has a unit root Exogenous: Constant, Linear Trend Bandwidth: 6 (Newey-West automatic) using Bartlett kernel				
			Adj. t-Stat	Prob.*
Phillips-Perron test statistic			-14.50935	0.0000
Test critical values:	1% level		-3.983471	
	5% level		-3.422218	
	10% level		-3.133955	
*MacKinnon (1996) one-sided p-values.				
Residual variance (no correction)				1840.046
HAC corrected variance (Bartlett kernel)				2202.260
Phillips-Perron Test Equation Dependent Variable: D(TEMPO) Method: Least Squares Date: 12/10/18 Time: 16:56 Sample (adjusted): 1/02/2017 12/31/2017 Included observations: 364 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
TEMPO(-1)	-0.708915	0.050362	-14.07648	0.0000
C	34.76723	5.148649	6.752690	0.0000
@TREND("1/01/2017")	-0.016883	0.021514	-0.784749	0.4331
R-squared	0.354378	Mean dependent var		-0.046804
Adjusted R-squared	0.350801	S.D. dependent var		53.45922
S.E. of regression	43.07363	Akaike info criterion		10.37191
Sum squared resid	669776.9	Schwarz criterion		10.40403
Log likelihood	-1884.687	Hannan-Quinn criter.		10.38467
F-statistic	99.07531	Durbin-Watson stat		2.041691
Prob(F-statistic)	0.000000			

Fonte: Elaborado pelo autor (2019).

Como mostra a Figura 22 do teste de hipótese de *Augmented Dickey-Fuller* com tendência linear, obteve-se como resultado um valor *t-Statistic* = -7.71, e um *p* valor = 0,00001, sendo esse valor significativo, ou seja, rejeitamos a hipótese nula de que a série possui raízes unitárias, com isso, a série é estacionária. Porém, ao analisar-se a variável de tendência, tem-se um valor *t-Statistic* = -0,45, e um *p* valor = 0,6495, ou seja, com um *p* valor não significativo, nossa série não apresenta tendência linear. A Figura 23 do teste de hipótese *Phillips-Perron* ajuda a confirmar os mesmos

pressupostos citados no teste anterior. No teste *Phillips-Perron* obtivemos como resultado um valor t-Statistic = -14,5, e um p valor = 0,00001, confirmando que a série não possui raiz unitária, ou seja, é estacionária. A variável tendência obteve como resultado o valor de t-Statistic = -0,078, e um p valor = 0,4331, não significativo, confirmando que a série não apresenta tendência linear.

Figura 24 – Teste de hipótese *Augmented Dickey-Fuller* com intercepto

Augmented Dickey-Fuller Unit Root Test on TEMPO

Null Hypothesis: TEMPO has a unit root				
Exogenous: Constant				
Lag Length: 2 (Automatic - based on SIC, maxlag=16)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-7.706300	0.0000
Test critical values:				
	1% level		-3.448211	
	5% level		-2.869307	
	10% level		-2.570975	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(TEMPO)				
Method: Least Squares				
Date: 12/10/18 Time: 16:53				
Sample (adjusted): 1/04/2017 12/31/2017				
Included observations: 362 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
TEMPO(-1)	-0.540641	0.070156	-7.706300	0.0000
D(TEMPO(-1))	-0.200031	0.063943	-3.128252	0.0019
D(TEMPO(-2))	-0.172998	0.051749	-3.343015	0.0009
C	23.87628	3.850758	6.200410	0.0000
R-squared	0.378907	Mean dependent var		-0.315272
Adjusted R-squared	0.373702	S.D. dependent var		53.37369
S.E. of regression	42.23938	Akaike info criterion		10.33557
Sum squared resid	638731.2	Schwarz criterion		10.37857
Log likelihood	-1866.738	Hannan-Quinn criter.		10.35267
F-statistic	72.80111	Durbin-Watson stat		1.976231
Prob(F-statistic)	0.000000			

Fonte: Elaborado pelo autor (2019).

Figura 25 – Teste de hipótese *Phillips-Perron* com intercepto

Phillips-Perron Unit Root Test on TEMPO

Null Hypothesis: TEMPO has a unit root				
Exogenous: Constant				
Bandwidth: 6 (Newey-West automatic) using Bartlett kernel				
			Adj. t-Stat	Prob.*
Phillips-Perron test statistic			-14.49783	0.0000
Test critical values:				
	1% level		-3.448111	
	5% level		-2.869263	
	10% level		-2.570952	
*MacKinnon (1996) one-sided p-values.				
Residual variance (no correction)				1843.185
HAC corrected variance (Bartlett kernel)				2207.913
Phillips-Perron Test Equation				
Dependent Variable: D(TEMPO)				
Method: Least Squares				
Date: 12/10/18 Time: 16:57				
Sample (adjusted): 1/02/2017 12/31/2017				
Included observations: 364 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
TEMPO(-1)	-0.706877	0.050268	-14.06217	0.0000
C	31.59476	3.186646	9.914739	0.0000
R-squared	0.353277	Mean dependent var		-0.046804
Adjusted R-squared	0.351490	S.D. dependent var		53.45922
S.E. of regression	43.05077	Akaike info criterion		10.36812
Sum squared resid	670919.5	Schwarz criterion		10.38953
Log likelihood	-1884.997	Hannan-Quinn criter.		10.37663
F-statistic	197.7446	Durbin-Watson stat		2.042584
Prob(F-statistic)	0.000000			

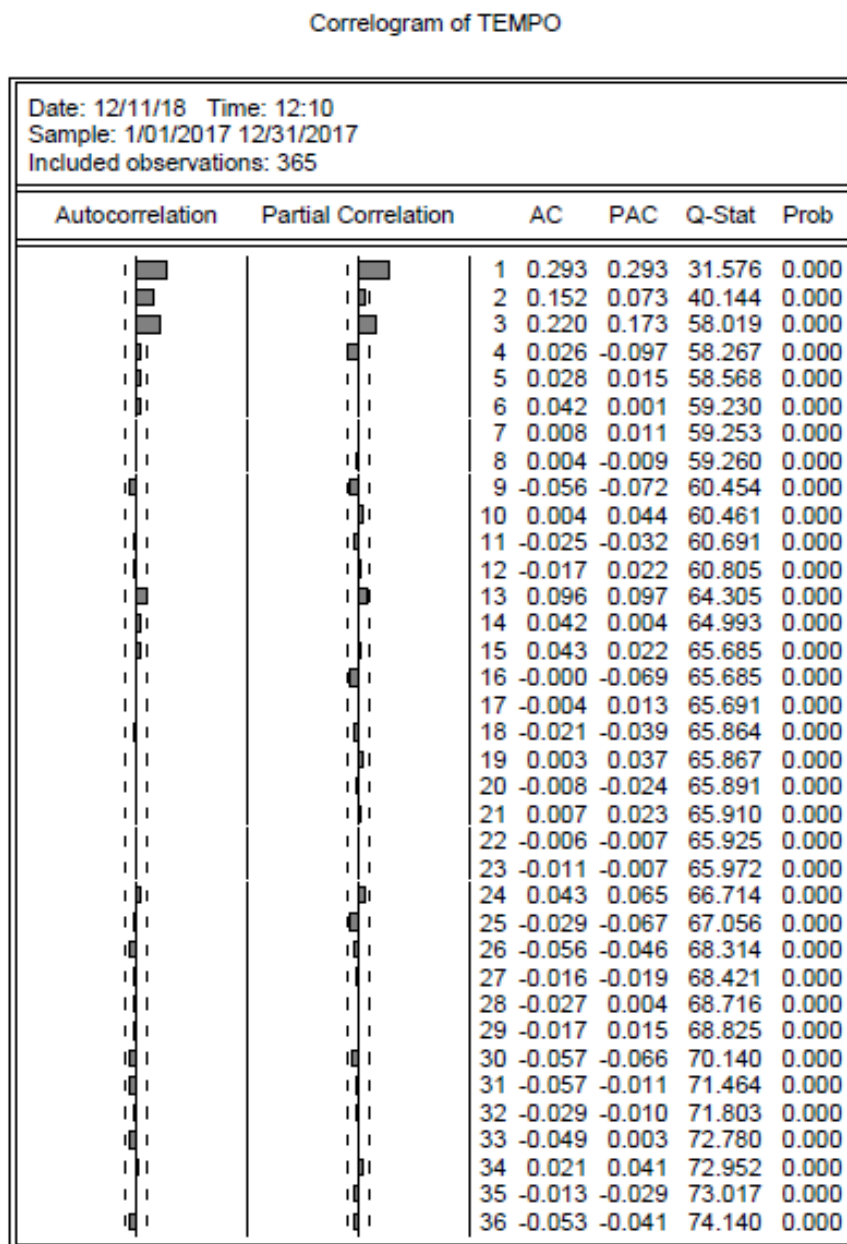
Fonte: Elaborado pelo autor (2019).

O resultado teste de hipótese *Augmented Dickey-Fuller* com intercepto é apresentado na Figura 24. Observa-se um valor t-Statistic = -7,70, com p valor = 0,0001, ou seja, a série não possui nenhuma raiz unitária sendo ela estacionária. Ao analisar os coeficientes das variáveis, segundo o teste, todas são significativas. A Figura 25 apresenta os resultados obtidos a partir do teste de *Phillips-Perron*, o que ajuda a confirmar que a série temporal é estacionária e as variáveis são significativas.

O próximo é analisar as funções de auto correlação e auto correlação parcial dos dados de tempo de ordens de manutenção emergenciais na rede elétrica. Para analisar essas duas funções, fez-se o diagrama de auto correlação da série temporal, como é apresentado no Figura 26. O diagrama mostra que na função de

autocorrelação os lags 1, 2 e 3 apresentaram valores maiores que os limites de significância. E na função de correlação parcial os lags 1 e 3 apresentaram valores além dos limites de significância. Esses lags servem para analisar tanto o valor AR (auto regressivo) como MA (medias moveis). Como já foi concluído que nossa serie temporal é estacionária, vamos utilizar o modelo ARMA. Pois o modelo ARIMA é representado com o $I(d)$, que significa quantas diferenças foram necessárias para transformar a série original em estacionária.

Figura 26 – Diagrama de auto correlação dos dados



Fonte: Elaborado pelo autor (2019).

Ajustes de modelos:

Como a Figura 26 apresentou lags 1 e 3 como fora dos limites de significância, será testado 4 tipos de modelo, ARMA (1, 1), ARMA (1,3), ARMA (3,1) E ARMA (3,3), estes modelos são apresentados nas Figuras 27, 28, 29 e 30.

ARMA (1,1)

Figura 27 – Modelo ARMA (1,1)

Dependent Variable: TEMPO
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 12/11/18 Time: 13:31
 Sample: 1/01/2017 12/31/2017
 Included observations: 365
 Convergence achieved after 33 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.64987	7.601155	5.874091	0.0000
AR(1)	0.662109	0.074395	8.899916	0.0000
MA(1)	-0.415400	0.107476	-3.865034	0.0001
SIGMASQ	1812.999	34.71973	52.21813	0.0000
R-squared	0.098402	Mean dependent var		44.70087
Adjusted R-squared	0.090910	S.D. dependent var		44.90432
S.E. of regression	42.81458	Akaike info criterion		10.36287
Sum squared resid	661744.7	Schwarz criterion		10.40561
Log likelihood	-1887.224	Hannan-Quinn criter.		10.37985
F-statistic	13.13340	Durbin-Watson stat		1.981091
Prob(F-statistic)	0.000000			
Inverted AR Roots	.66			
Inverted MA Roots	.42			

Fonte: Elaborado pelo autor (2019).

ARMA (1,3):

Figura 28 – Modelo ARMA (1,3)

Dependent Variable: TEMPO
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/11/18 Time: 13:35
Sample: 1/01/2017 12/31/2017
Included observations: 365
Convergence achieved after 21 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.65150	7.060253	6.324349	0.0000
AR(1)	0.281972	0.055072	5.120080	0.0000
MA(3)	0.192001	0.034189	5.615889	0.0000
SIGMASQ	1765.147	32.08548	55.01390	0.0000

R-squared	0.122199	Mean dependent var	44.70087
Adjusted R-squared	0.114904	S.D. dependent var	44.90432
S.E. of regression	42.24577	Akaike info criterion	10.33634
Sum squared resid	644278.7	Schwarz criterion	10.37908
Log likelihood	-1882.383	Hannan-Quinn criter.	10.35333
F-statistic	16.75161	Durbin-Watson stat	2.010281
Prob(F-statistic)	0.000000		

Inverted AR Roots	.28		
Inverted MA Roots	.29+.50i	.29-.50i	-.58

Fonte: Elaborado pelo autor (2019).

ARMA (3,1):

Figura 29 – Modelo ARMA (3,1)

Dependent Variable: TEMPO
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/11/18 Time: 13:38
Sample: 1/01/2017 12/31/2017
Included observations: 365
Convergence achieved after 22 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.68722	6.607121	6.763493	0.0000
AR(3)	0.207910	0.037270	5.578419	0.0000
MA(1)	0.255146	0.055212	4.621172	0.0000
SIGMASQ	1775.924	32.05440	55.40343	0.0000

R-squared	0.116840	Mean dependent var	44.70087
Adjusted R-squared	0.109500	S.D. dependent var	44.90432
S.E. of regression	42.37454	Akaike info criterion	10.34244
Sum squared resid	648212.1	Schwarz criterion	10.38518
Log likelihood	-1883.495	Hannan-Quinn criter.	10.35942
F-statistic	15.91976	Durbin-Watson stat	1.952334
Prob(F-statistic)	0.000000		

Inverted AR Roots	.59	-.30-.51i	-.30+.51i
Inverted MA Roots	-.26		

Fonte: Elaborado pelo autor (2019).

ARMA (3,3):

Figura 30 – Modelo ARMA (3,3)

Dependent Variable: TEMPO
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 12/11/18 Time: 13:40
 Sample: 1/01/2017 12/31/2017
 Included observations: 365
 Convergence achieved after 23 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.70857	5.301346	8.433438	0.0000
AR(3)	0.201265	0.462800	0.434885	0.6639
MA(3)	0.020856	0.462476	0.045097	0.9641
SIGMASQ	1912.303	36.14230	52.91040	0.0000
R-squared	0.049018	Mean dependent var		44.70087
Adjusted R-squared	0.041116	S.D. dependent var		44.90432
S.E. of regression	43.97150	Akaike info criterion		10.41627
Sum squared resid	697990.8	Schwarz criterion		10.45901
Log likelihood	-1896.969	Hannan-Quinn criter.		10.43326
F-statistic	6.202593	Durbin-Watson stat		1.436748
Prob(F-statistic)	0.000406			
Inverted AR Roots	.59	-.29-.51i	-.29+.51i	
Inverted MA Roots	.14+.24i	.14-.24i	-.28	

Fonte: Elaborado pelo autor (2019).

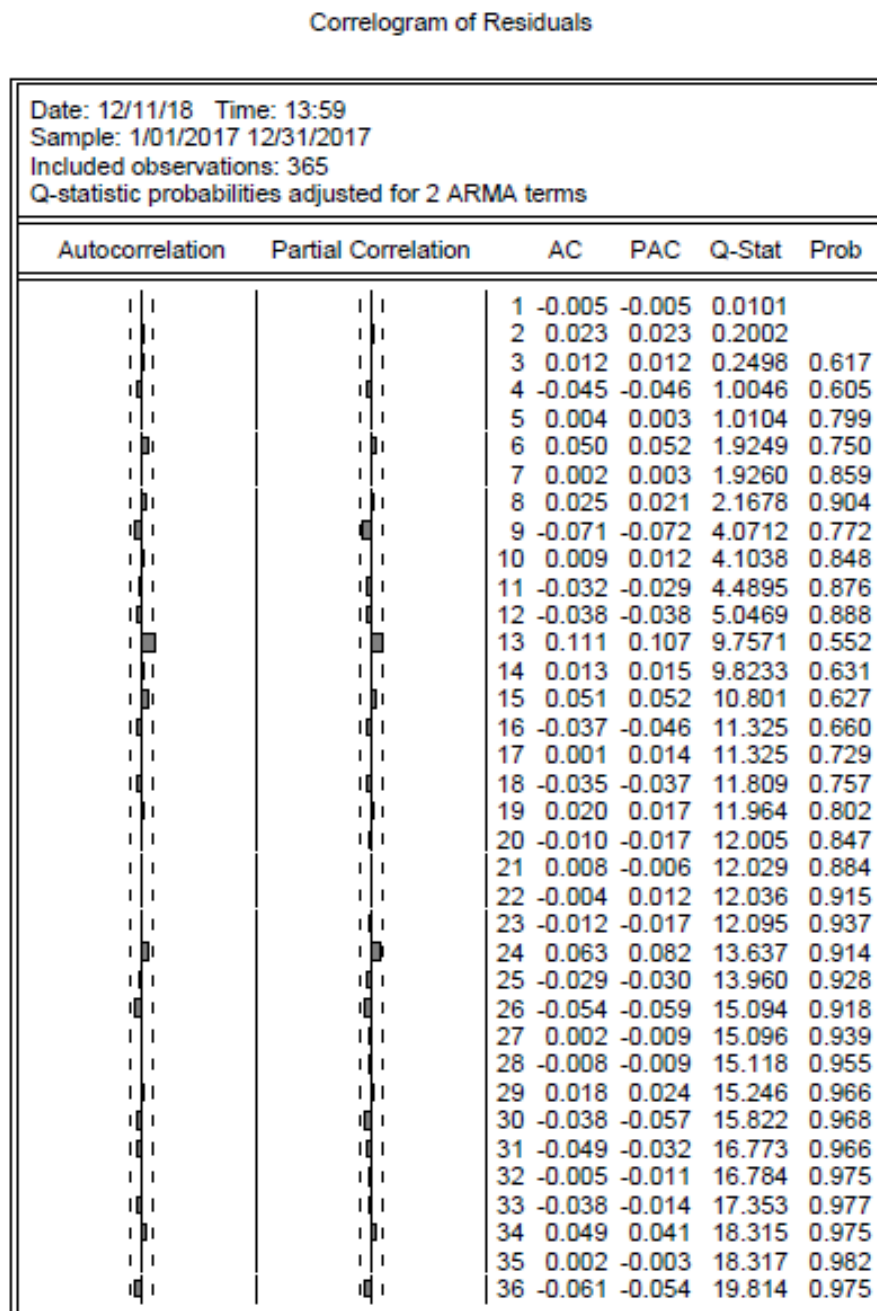
O comparativo entre os modelos é apresentado na Tabela 2, o objetivo para classificação destes é: Menores coeficientes de AIC (*Akaike info criterion*), BIC (*Schwarz criterion*) e SIGMASQ, e o maior *R-squared*.

Tabela 2 – Comparativo entre os modelos ARMA

Modelo	AIC	BIC	SIGMASQ	R-squared
ARMA (1,1)	10.362	10.405	1812	0.098
ARMA (1,3)	10.336	10.379	1765	0.122
ARMA (3,1)	10.342	10.385	1775	0.116
ARMA (3,3)	10.416	10.459	1912	0.041

Na Figura 31 é apresentado o diagrama de auto correlação dos resíduos do modelo ARMA (1,3), definido como o melhor modelo a partir dos resultados obtidos:

Figura 31 – Diagrama de auto correlação modelo ARMA (1,3)



Fonte: Elaborado pelo autor (2019).

Todos os “lags” se encontram dentro do intervalo de confiança, ou seja, todas as informações da série temporal foram capturadas pelo modelo ajustado.

O diagrama dos resíduos ao quadrado é apresentado na Figura 32:

Figura 32 – Diagrama dos resíduos do modelo ARMA (1,3)

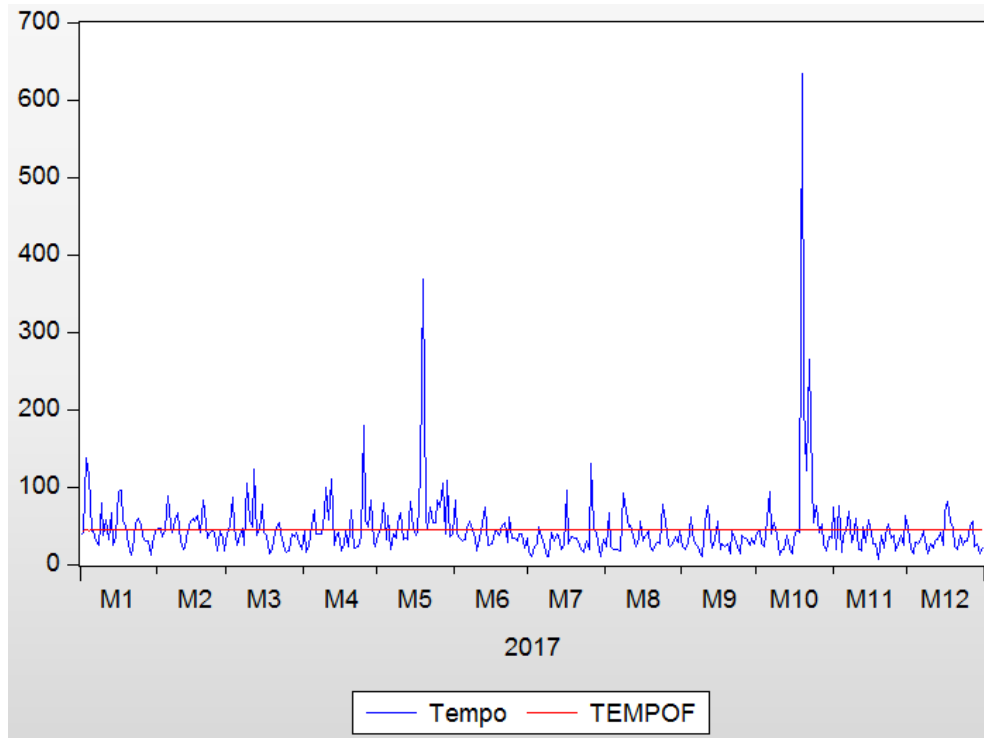
Date: 12/11/18 Time: 14:23
 Sample: 1/01/2017 12/31/2017
 Included observations: 365

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.000	-0.000	2.E-06	0.999
		2 -0.005	-0.005	0.0097	0.995
		3 0.012	0.012	0.0673	0.995
		4 -0.006	-0.006	0.0822	0.999
		5 -0.005	-0.005	0.0918	1.000
		6 -0.006	-0.006	0.1060	1.000
		7 -0.007	-0.007	0.1223	1.000
		8 -0.005	-0.005	0.1311	1.000
		9 -0.004	-0.004	0.1360	1.000
		10 -0.003	-0.003	0.1388	1.000
		11 -0.007	-0.007	0.1551	1.000
		12 -0.007	-0.007	0.1732	1.000
		13 0.004	0.004	0.1785	1.000
		14 -0.004	-0.004	0.1847	1.000
		15 -0.002	-0.002	0.1870	1.000
		16 -0.001	-0.001	0.1871	1.000
		17 -0.008	-0.008	0.2117	1.000
		18 -0.008	-0.008	0.2360	1.000
		19 -0.004	-0.004	0.2429	1.000
		20 -0.006	-0.006	0.2580	1.000
		21 -0.007	-0.007	0.2781	1.000
		22 -0.007	-0.008	0.3000	1.000
		23 -0.005	-0.006	0.3113	1.000
		24 0.007	0.006	0.3286	1.000

Fonte: Elaborado pelo autor (2019).

Previsão ARMA (1,3):

Figura 33 – Previsão do modelo ARMA (1,3)



Fonte: Elaborado pelo autor (2019).

Ao analisar a Figura 28 podemos notar que mesmo o modelo ARMA (1,3) ter tido os melhores coeficientes para realizar a previsão, o modelo não consegue identificar claramente o real fenômeno que descreve a série temporal.

5 CONSIDERAÇÕES FINAIS

A presente pesquisa teve como objetivo geral desenvolver um modelo de previsão de ordens de manutenção emergenciais para um sistema de distribuição de energia elétrica, baseado em fatores climáticos. O ponto de partida da pesquisa foi realizar um levantamento bibliográfico para identificar as técnicas de previsão de demanda realizadas nas áreas de energia elétrica e manutenção, bem como modelos preditivos.

O estudo na bibliografia identificou a utilização de variáveis climáticas para o desenvolvimento de modelos preditivos de consumo, demanda e manutenção em sistemas de distribuição de energia elétrica, sendo assim, o presente trabalho identificou quais variáveis climáticas poderiam ser utilizadas para o desenvolvimento do modelo de previsão de ordens de manutenção emergenciais.

Para solucionar o questionamento proposto e executar os objetivos do trabalho, foram realizadas análises para entender as características dos dados de horas de manutenção emergenciais, e com isso determinar o método de regressão linear múltipla para elaboração do modelo de previsão de demanda. A mineração dos dados foi realizada a fim de classificar quais variáveis climáticas foram utilizadas no modelo, utilizando testes de hipótese para ajustar os dados perante os pressupostos necessários.

Por meio da análise dos resultados obtidos a partir do modelo desenvolvido de regressão linear múltipla foi possível identificar a relação de variáveis climáticas como máxima rajada do dia, quantidade de chuva, umidade e velocidade média do vento com as horas de manutenção emergências coletadas no projeto, sendo este modelo com um coeficiente de determinação (R^2) de 0,4235, ou seja, 42,35% dos dados coletados são explicados pelas variáveis. A previsão de demanda de ordens de manutenção emergenciais baseada em fatores climáticos foi realizada e a validação do modelo foi feita a partir do teste estatístico MAPE (erro absoluto médio percentual), que apresentou um resultado de 28%, comparando os erros gerados dos dados previstos com os dados reais coletados.

Com contribuição, pode-se destacar a utilização de modelos de regressão linear múltipla para a elaboração de modelos de previsão de demanda no setor de energia elétrica, esta prática pode ajudar as empresas deste segmento a otimizarem a utilização de seus recursos. Também acredita-se ser uma contribuição a descoberta

dos fatores climáticos que de fato influenciam na previsão da demanda, podendo ser o ponto de partida para estudos futuros em empresas distribuidoras de energia elétrica.

5.1 LIMITAÇÕES DA PESQUISA

Durante o desenvolvimento dessa dissertação, ocorreram eventos impeditivos a obtenção de alguns resultados. Um deles foi a coleta dos dados reais para serem utilizados como base para o desenvolvimento de modelos preditivos. Com a falta de dados referentes aos anos de 2018, 2019 e 2020. Dos dados obtidos para o presente trabalho, apenas os dados referentes ao ano de 2017 estavam aptos e foram utilizados, sendo estes uma amostra de 365 dias, dos quais 345 foram minerados para o modelo desenvolvido e 7 para a validação da previsão.

5.2 SUGESTÕES DE ESTUDOS FUTUROS

Para pesquisas futuras a determinação de outros métodos de previsão de demanda a ser utilizados pode gerar outras análises referentes aos dados coletados e as variáveis escolhidas, sendo métodos de previsão de demanda como: Redes neurais e Árvores de decisão.

REFERÊNCIAS

- AHMED, T.; MUTTAQI, K. M.; AGALGAONKAR, A. P. Climate change impacts on electricity demand in the State of New Wales, Australia. **Applied Energy**, v. 98, p. 376-383, 2012.
- AHMED, T.; VU, D. H.; MUTTAQI, K. M.; AGALGAONKAR, A. P. Load forecasting under changing climatic conditions for the city of Sydney, Australia. **Energy**, v. 142, p. 911-919, 2018.
- ANDERSON, O. D. **Time Series Analysis and Forecasting** – The Box and Jenkins Approach. London and Boston: Butterworths, 1976.
- ANG, B. W.; GOH, T. N.; LIU, X. Q. Demanda residencial de eletricidade em Cingapura. **Energy**, v. 17, n. 1, p. 37-46, 1992.
- APADULA, F.; BASSINI, A.; ELLI, A.; SCAPIN, S. Relationships between meteorological variables and monthly electricity demand. **Applied Energy**, v. 98, p. 346-356, 2012.
- ATAKHANOVA, Z., HOWIE, P. Demanda de eletricidade no Cazaquistão. **Energy Policy**, v. 35, n. 7, p. 3729-3743, 2007.
- BATISTA, O. E. et al. A Brazilian Experience in Energy Management: Low-cost Actions as Strategy to Reduce Electricity Costs. **IFAC Proceedings**, v. 46, n. 24, p. 40-47, sept. 2013.
- BERTLING, L.; ALLAN, R.; ERIKSSON, R. A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. **IEEE Power System**, v. 20, n. 1, 2005.
- BRASIL. Ministério de Minas e Energia. **Plano Nacional de Energia 2030**. Brasília: Ministério de Minas e Energia; Empresa de Pesquisa Energética, 2007. 12v.
- BREUSCH, T. S.; PAGAN, A. R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. **Econometrica, Journal of the Econometric Society**, v. 47, n. 5, p. 1287-1294, 1979.
- BROCKWELL, D. J., DAVIS, R. A. **Introduction to Time Series and forecasting**. 2. ed. New York: Springer, 2002.
- BROWN, D. E. Introduction to Data Mining for Medical Informatics. **Clinics in Laboratory Medicine**, v. 28, n. 1, p. 9-35, 2008.
- CARAZAS, F. G.; SOUZA, G. F. M. Risk-based decision making method for maintenance policy selection of thermal power plant equipment. **Energy**, v. 35, n. 2, p. 964-975, 2010.
- CARNERO, M. C. **MCDAs Techniques in Maintenance Policy Selection**. Encyclopedia of Business Analytics and Optimization. 2014.

- CHING-LAI, H.; SIMON, J. W.; MAJITHIA, S. Analisando o impacto das variáveis meteorológicas na demanda mensal de eletricidade. **IEEE Trans Power Syst**, v. 20, n. 4, 2005.
- CHOPRA, S.; MEINDL, P. **Gestão da cadeia de suprimentos, estratégia, planejamento e operações**. 6. ed. São Paulo: Pearson education do Brasil, 2016.
- FRANCO, G.; SANSTAD, A. Mudança climática e demanda de eletricidade na Califórnia. **Clim Change**, v. 87, p. 139-151, 2008.
- FRIEDRIH, L.; ARMSTRONG, P.; AFSHARI, A. Mid-term forecasting of urban electricity load to isolate air-conditioning impact. **Energy and Buildings**, v. 80, p. 72-80, 2014.
- FROGER, A.; GENDREAU, M.; MENDOZA, J. E.; PINSON, E.; ROUSSEAU, L. M. Maintenance scheduling in the electricity industry: A literature Review. **European Journal of Operational Research**, v. 251, n. 3, p. 695-706, 2016.
- GHANI, I. M.; AHMAD, S. Stepwise Multiple Regression Method to Forecast Fish Landing. **Procedia - Social and Behavioral Sciences**, v. 8, p. 549-554, 2010.
- GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 4. ed. São Paulo: [s.n.], 2009.
- GONZÁLEZ-ROMERA, E.; JARAMILLO-MORÁN, M. A.; CARMONA-FERNÁNDEZ, D. Monthly electric energy demand forecasting based on trend extraction. **IEEE Transactions on Power Systems**, v. 21, n. 4, p. 1946-1953, 2006.
- GORDAN, M.; ISMAIL, Z.; RAZAK, H. A.; GHAEDI, K.; IBRAHIM, Z.; TAN, Z. X.; GHAYEB, H. H. Data mining-based damage identification of a slab-on-girder bridge using inverse analysis. **Measurement**, v. 151, 2020.
- GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5. ed. São Paulo: McGraw Hill Brasil, 2011.
- HADI, F.; HOMAYOON, K. New empirical model to evaluate groundwater flow into circular tunnel using multiple regression analysis. **International Journal of Mining Science and Technology**, v. 27, n. 3, p. 415-421, 2017.
- HAIDA, T.; MUTO, S. Regression based peak load forecasting using a transformation technique, Power Systems. **IEEE Transactions**, v. 9, n. 4, p. 1788-1794, 1994.
- HANKE, J. E.; WICHERN, D. W.; REITSCH, A. G. **Business forecasting**. 7. ed. New York: Prentice Hall, 2001.
- HARRISON, J. H. Introduction to the Mining of Clinical Data. **Clinics in Laboratory Medicine**, v. 28, n. 1, p. 1-7, 2008.
- HEKKENBERG, M.; MOLL, H. C. Dynamic temperature dependence patterns in future energy demand models in the context of climate change. **Energy**, v. 34, n. 11, p. 1797-1806, 2009.

HENLEY, A., PEIRSON, J. Não linearidades na demanda de eletricidade e temperatura: Métodos paramétricos versus não paramétricos. **Boletim Oxford de Economia e Estatística**, v. 59, n. 1, p. 149-162, 1997.

HOWDEN, S.M., CRIMP, S. Efeito do clima e das mudanças climáticas na demanda de eletricidade na Austrália. *Ecosistema Sustentável CSIRO*, 2001.

INMET. **Instituto Nacional de Meteorologia**. Brasil, 2018.

JIA, T. G.; LIU, J. Stability of mine ventilation system based on multiple regression analysis. **Mining Science and Technology**, China, v. 19, n. 4, p. 463-466, 2009.

JOHNSTON, J. **Métodos econométricos**. São Paulo: Atlas, 1977.

KANDIL, M.; EL-DEBEIKY, S. M.; HASANIEN, N. Long-term load forecasting for fast developing utility using a knowledge-based expert system, *Power Systems*. **IEEE Transactions**, v. 17, n. 2, p. 491-496, 2002.

KIM, C.; YONJOO, L.; PARK, B. U. Cook's distance in local polynomial regression. **Statistics & Probability Letters**, v. 54, n. 1, p. 33-40, 2001.

KOTHARI, C. R. **Research methodology: methods and techniques**. [S.l.]: New Age International, 2004.

LAKATOS, E. M.; MARCONI, M. A. **Metodologia Científica**. 7. ed. São Paulo: Atlas, 2010.

LIRA, S. A. **Análise de correlação: abordagem teórica e de construção dos coeficientes com aplicações**. 2004. 196 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia dos Setores de Ciências Exatas e de Tecnologia) – Universidade Federal do Paraná, Curitiba, 2004.

LIU, J.; KONG, X.; ZHOU, X.; WANG, L.; ZHANG, D.; LEE, I.; XU, B.; XIA, F. Data Mining and Information Retrieval in the 21st century: A bibliographic review. **Computer Science Review**, v. 34, 2019.

MATTOS, R. S. **Tendências e Raízes Unitárias**. Juiz de Fora: Universidade Federal de Juiz de Fora, 2018.

MENNIS, J.; GUO, D. Spatial data mining and geographic knowledge Discovery – An Introductio. **Computers, Environment and Urban Systems**, v. 33, n. 6, p. 403-408, 2009.

MIRASGEDIS, S.; SARAFIDIS, Y.; GEORGOPOULOU, E.; KOTRONI, V.; LAGOVARDOS, K. Modeling framework for estimating impacts of climate change on electricity demand at regional level: Case of Greece. **Energy Convers Manage**, v. 48, n. 5, p. 1737-1750, 2007.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to Time Series Analysis and Forecasting**. New Jersey: [s.n.], 2008.

- MONTGOMERY, D. C.; JOHNSON, L. A. **Forecasting and Time Series Analysis**. New York: [s.n.], 1976.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo: [s.n.], 2004.
- NETO, R. M.; PUREZA, V. Modelagem e Simulação. In: Miguel, P. A. C. (Org.). **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. Rio de Janeiro: Elsevier. ABEPRO, 2012.
- NGUYEN, T. A. T.; CHOU, S. Y. Maintenance strategy selection for improving cost-effectiveness of offshore Wind systems. **Energy Conversion and Management**, v. 157, p. 86-95, 2018.
- OLIVEIRA, M. O.; MARZEC, D. P.; BORDIN, G.; BRETAS, A. S.; BERNARDON, D. Climate Change Effect on Very Short-Term Electric Load Forecasting. In: IEEE TRONDHEIM POWERTECH, 2011. **Proceeding...** Trondheim, Norway: IEEE, 2011.
- OLKIN, I.; SAMPSON, A. R. **Multivariate Analysis: Overview**. Oxford: International Encyclopedia of the Social & Behavioral Sciences, 2001.
- PANG-NING, T.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**, Pearson Addison-Wesley. Boston, 2006.
- PARKPOOM, S.; HARRISON, G. P. Analisando o impacto das mudanças climáticas na demanda futura de eletricidade na Tailândia. **IEEE Trans Power Syst**, v. 23, n. 3, p. 1441-1448, 2008.
- PESARAN, M. H.; ULLAH, A.; YAMAGATA, T. A Bias-adjusted LM test of error cross-section Independence. **The Econometrics Journal**, v. 11, n. 1, p. 105-127, 2008.
- PETRI, Y.; CALDEIRA, K. Impactos do aquecimento global nos graus-dias de aquecimento e resfriamento residencial nos Estados Unidos. **Scientific Reports**, v. 5, 2015.
- RUTH, M.; LIN, D. C. Regional energy demand and adaptations to climate change: Methodology and application to the state of Maryland, USA. **Energy Policy**, v. 34, n. 17, p. 2820-2833, 2006.
- SAILOR, D. J. Relacionando as cargas de eletricidade do setor residencial e comercial ao clima, avaliando as vulnerabilidades e vulnerabilidades do estado. **Energy**, v. 26, n. 7, p. 645-657, 2000.
- SALMAN, A. M.; LI, Y.; BASTIDAS-ARTEAGA, E. Maintenance optimization for power distribution systems subjected to hurricane hazard, timber decay and climate change. **Reliability Engineering & System Safety**, v. 168, p. 136-149, 2017.
- SALOMÃO S. et al. **Mudança climática 2007: a base da ciência física**. Grupo de Trabalho I ao Quarto Relatório de Avaliação do Painel Intergovernamental sobre Mudança do Clima. Cambridge, UK, Nova Iorque, EUA: Cambridge University Press, 2007.

SHAFIEE, M.; SORENSEN, J. D. Maintenance optimization and inspection planning of Wind energy assets: Models, methods and strategies. **Reliability Engineering & System Safety**, v. 192, 2017.

SING, A. K.; IBRAHEEM, I.; KHATOON, S.; MUAZZAM, M.; CHATUVERVEDI, D. Load forecasting techniques and methodologies: A review. In: INTERNATIONAL CONFERENCE ON POWER, CONTROL AND EMBEDDED SYSTEMS, 2., 2012. **Proceeding...** Allahabad, India: IEEE, 2012.

SITTITHUMWAT, A.; SOUDI, F.; TOMSOVIC, K. Optimal allocation of distribution maintenance resources with limited information. **Electric Power Systems Research**, v. 68, n. 3, p. 208-220, 2004.

SON, H.; KIM, C. Short-term forecasting of electricity demand for the residential sector using weather and social variables. **Resources, Conservation and Recycling**, v. 123, p. 200-207, 2017.

SUGANTHI, L.; SAMUEL, A. A. Energy models for demand forecasting – A review. **Renewable and Sustainable Energy Reviews**, v. 16, n. 2, p. 1223-1240, 2012.

TAN, P.-N. **Introduction Data Mining**. Omstructor's Solution Manual Pearson Education India, 2007.

TAYLOR, J. W. An evaluation of methods for very short-term load forecastig using minute-by-minute British data. **International Journal of Forecasting**, v. 24, n. 4, p. 645-658, 2008.

VU, D. H.; MUTTAQI, K. M.; AGALGAONKAR, A. P. A variance inflation fator and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. **Applied Energy**, v. 140, p. 385-394, 2015.

YSSAAD, B.; KHIAT, M.; CHAKER, A. Reliability centered maintenance optimization for power distribution systems. **International Journal of Electrical Power & Energy Systems**, v. 55, p.108-115, 2014.