



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I - CAMPINA GRANDE  
PRÓ-REITORA DE PÓS-GRADUAÇÃO E PESQUISA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA AMBIENTAL  
MESTRADO EM CIÊNCIA E TECNOLOGIA AMBIENTAL**

**AUGUSTO PONCIANO BARBOSA DA SILVA**

**ANÁLISE EXPLORATÓRIA DE DADOS DE MONITORAMENTO DE  
EFLUENTES URBANOS**

**CAMPINA GRANDE - PB**

**2021**

**AUGUSTO PONCIANO BARBOSA DA SILVA**

**ANÁLISE EXPLORATÓRIA DE DADOS DE MONITORAMENTO DE EFLUENTES  
URBANOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia Ambiental - PPGCTA, em cumprimento às exigências finais para obtenção do título de Mestre em Ciência e Tecnologia Ambiental.

**Área de concentração:** Qualidade de sistemas ambientais

**Orientador:** Prof. Dr. Rui de Oliveira

**CAMPINA GRANDE - PB**

**2021**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586a Silva, Augusto Ponciano Barbosa da.  
Análise exploratória de dados de monitoramento de efluentes urbanos [manuscrito] / Augusto Ponciano Barbosa da Silva. - 2021.  
63 p. : il. colorido.

Digitado.  
Dissertação (Mestrado em Ciência e Tecnologia Ambiental) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2021.  
"Orientação : Prof. Dr. Rui de Oliveira, Coordenação do Curso de Engenharia Sanitária e Ambiental - CCT."

1. Mineração de dados. 2. Gestão de recursos hídricos. 3. Séries temporais. I. Título

21. ed. CDD 519.232

**AUGUSTO PONCIANO BARBOSA DA SILVA**

**ANÁLISE EXPLORATÓRIA DE DADOS DE MONITORAMENTO DE  
EFLUENTES URBANOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia Ambiental - PPGCTA, em cumprimento às exigências finais para obtenção do título de Mestre em Ciência e Tecnologia Ambiental.

**Área de concentração:** Qualidade de sistemas ambientais

Aprovada em: 26/08/2021.

**BANCA EXAMINADORA**



Prof. Dr. Rui de Oliveira (Orientador)  
Universidade Estadual da Paraíba (UEPB)



Profa. Dra. Ruth Silveira do Nascimento  
Universidade Estadual da Paraíba (UEPB)

**Juscelino Alves**

Assinado de forma digital por Juscelino  
Alves Henriques:06751456433

**Henriques:06751456433** Dados: 2021.08.31 19:53:35 -03'00'

Prof. Dr. Juscelino Alves Henriques  
Instituto Federal do Sertão Pernambucano (IFSertãoPE)

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus por ter proporcionado cada momento em minha vida e mostrar que tudo na vida é um ensinamento.

. À minha noiva Cinthia por sempre estar ao meu lado, mesmo nos momentos difíceis, sempre buscando me incentivar e me encorajar a dar passos cada vez mais largos mostrando dedicação e boa vontade

Ao professor doutor Rui de Oliveira que teve contribuição grandiosa nesse trabalho, que sempre se preocupou com a minha compreensão sobre o assunto e sempre se mostrou disposto a contribuir.

Aos professores do Mestrado em Ciência e Tecnologia Ambiental da UEPB, que contribuíram ao longo do período do mestrado, por meio das disciplinas e debates, para o desenvolvimento desta pesquisa.

A todos que passaram pela minha vida e de alguma forma deixaram um pouco de seus ensinamentos e experiências que me ajudaram a crescer como pessoa e como profissional.

## RESUMO

Com o objetivo de aprofundar o conhecimento sobre a variação temporal da qualidade de águas residuárias urbanas, este trabalho foi baseado na análise de dados de qualidade de esgoto bruto obtidos de amostras coletadas no Interceptor Leste do sistema de esgoto da cidade de Campina Grande, Paraíba, Nordeste do Brasil. Dados sobre os indicadores demanda bioquímica de oxigênio ( $DBO_5$ ), nitrogênio amoniacal, coliformes termotolerantes, alcalinidade total e sólidos suspensos, obtidos no período 1977-1989, na Estação Experimental de Tratamentos Biológicos de Esgotos Sanitários (EXTRABES), então estação de pesquisa da Universidade Federal da Paraíba, foram analisados através de Análise de Agrupamento e Métodos Mann-Kendall e Box-Jenkins. A análise exploratória buscou estabelecer padrões sazonais de relações entre a qualidade do esgoto e fatores ambientais, modelar tendências de variação temporal dos indicadores de qualidade do esgoto urbano e avaliar o desempenho dos modelos que interpretam essa variação. A análise de séries temporais de dados de qualidade de águas residuárias foi capaz de demonstrar tendências significativas ao longo do período estudado, mas não apresentou o mesmo desempenho para os dados de precipitação e temperatura. A aplicação do método Box-Jenkins permitiu modelar o comportamento tanto das variáveis de qualidade do esgoto quanto das variáveis hidrológicas, sendo os modelos considerados satisfatórios com base na avaliação de erros e resíduos.

**PALAVRAS-CHAVE:** Mineração de dados. Gestão de recursos hídricos. Séries temporais.

## **ABSTRACT**

Aiming the development of knowledge on the temporal variation of urban wastewater quality, this work was based on the analysis of raw sewage quality data from samples collected from the East Interceptor of the sewerage system of Campina Grande City, Paraíba state, northern Brazil. Data on the indicators biochemical oxygen demand ( $BOD_5$ ), ammonia nitrogen, thermotolerant coliforms, total alkalinity and suspended solids, obtained in the period 1977-1989, at the Experimental Station of Biological Treatment of Sewage (EXTRABES), then a research station at the Federal University of Paraíba, were analysed through Cluster Analysis and Mann-Kendall and Box-Jenkins methods. Exploratory analysis sought to establish seasonal patterns of relationships between sewage quality and environmental factors, to model temporal variation trends of urban sewage quality indicators and to evaluate the performance of models interpreting this variation. Time series analysis for data on wastewater quality was able to demonstrate significant trends over the studied period, but did not show the same performance for data on both precipitation and temperature. The application of the Box-Jenkins method allowed modelling the behaviour of both the sewage quality variables and the hydrological variables, with the models being considered satisfactory based on the evaluation of errors and residues.

**KEYWORDS:** Data mining. Management of water resources. Time series.

## LISTA DE FIGURAS

Figura 1 - Etapas do processo de extração de conhecimento (KDD).....	15
Figura 2 - Modelo esquemático de um dendrograma.....	22
Figura 3 – Bacia Urbana do Prado do município de Campina Grande - PB.....	23
Figura 4 – Mapa das bacias drenadas pelo sistema de esgotamento sanitário do município de Campina Grande.....	25
Figura 5 - Dendrograma referente aos anos de 1977 e 1984: (a) precipitação (b) temperatura média ..	29
Figura 6 - Gráfico da precipitação e da temperatura média mensal entre os anos 1977 e 1984. ....	29
Figura 7 - Gráfico da análise de tendência: (a) precipitação; (b) temperatura .....	31
Figura 9 - Gráficos do FAC e FACP para os dados de: (a) precipitação (b) temperatura.....	33
Figura 10 – Análise dos resíduos do modelo proposto para os dados de: .....	34
(a) precipitação (b) temperatura .....	34
Figura 11 – Previsão com o modelo proposto para os dados de: (a) precipitação (b) temperatura.....	34
Figura 12 - Dendrograma da DBO referente ao período de 1977 a 1989. ....	35
Figura 13 - Dendrograma das médias mensais da DBO no período de 1977 a 1989.....	36
Figura 14 - Gráfico das médias mensais de DBO ao longo do período de 1977 a 1989.....	36
Figura 15 - Gráfico da análise de tendência da DBO.....	37
Figura 16 - Gráficos da decomposição para os dados de DBO.....	37
Figura 17- Gráficos do FAC e FACP para os dados de DBO.....	38
Figura 18 – Análise dos resíduos do modelo proposto para os dados de DBO.....	39
Figura 19 – Previsão com o modelo proposto para os dados de DBO.....	39
Figura 20 - Dendrograma em anos, referente aos valores do nitrogênio amoniacal .....	40
Figura 21 - Dendrograma referente à concentração média mensal de nitrogênio amoniacal.....	40
Figura 22 - Variação da concentração média mensal de nitrogênio amoniacal entre 1977 e 1989.....	41
Figura 23 - Gráfico da análise de tendência da amônia. ....	42
Figura 24 - Gráficos da decomposição para os dados de amônia.....	42
Figura 25 - Gráficos do FAC e FACP para os dados de nitrogênio amoniacal.....	43
Figura 26 – Análise dos resíduos do modelo proposto para os dados de nitrogênio amoniacal. ....	43
Figura 27 – Previsão com o modelo proposto para os dados de nitrogênio amoniacal.....	44
Figura 28 - Dendrograma dos dados referentes aos coliformes termotolerantes ao longo dos anos de 1977 a 1989.....	45
Figura 29 - Dendrograma referente às médias mensais de coliformes termotolerantes ao longo do período estudado. ....	45
Figura 30 - Variação das médias mensais de unidades formadoras de colônia ao longo do período estudado.....	46
Figura 31 - Gráfico da análise de tendência dos coliformes. ....	46
Figura 32 - Gráficos da decomposição para os dados de coliformes. ....	47
Figura 33 - Gráficos do FAC e FACP para os dados de coliformes. ....	47
Figura 34 – Análise dos resíduos do modelo proposto para os dados de coliformes. ....	48
Figura 35 – Previsão com o modelo proposto para os dados de coliformes. ....	48



Figura 36 - Dendrograma de agrupamento dos dados de alcalinidade total entre os anos de 1977 e 1989.....	50
Figura 37 - Dendrograma referente à alcalinidade média mensal, entre os anos de 1977 e 1989.....	50
Figura 38 - Distribuição da alcalinidade total média ao longo dos meses. ....	50
Figura 39 - Gráfico da análise de tendência da alcalinidade. ....	51
Figura 40 - Gráficos da decomposição para os dados de alcalinidade. ....	51
Figura 41 - Gráficos do FAC e FACP para os dados de alcalinidade. ....	52
Figura 42 – Análise dos resíduos do modelo proposto para os dados de alcalinidade.....	53
Figura 43 – Previsão com o modelo proposto para os dados de alcalinidade. ....	53
Figura 44 - Dendrograma dos sólidos suspensos totais referentes aos anos de 1977 a 1989.....	54
Figura 45 - Dendrograma dos sólidos suspensos totais referentes aos meses, ao longo de 1977 a 1989 .....	54
Figura 46 - Gráfico dos dados médio mensais de sólidos suspensos totais, ao longo de 1977 a 1989. .	55
Figura 47 - Gráfico da análise de tendência dos sólidos suspensos totais.....	55
Figura 48 - Gráficos da decomposição para os dados de sólidos suspensos totais. ....	56
Figura 49 - Gráficos do FAC e FACP para os dados de sólidos suspensos totais. ....	56
Figura 50 – Análise dos resíduos do modelo proposto para os dados de sólidos suspensos totais. ....	57
Figura 51 – Previsão com o modelo proposto para os dados de sólidos suspensos totais.....	57

## LISTA DE TABELAS

Tabela 1 - Áreas atendidas pelo projeto da Tecnosan em 1985. ....	25
Tabela 2 - Resumo descritivo dos dados da precipitação e temperatura média mensal. ....	28
Tabela 3 - Resultado do teste de Mann-Kendall para a precipitação e temperatura ....	30
Tabela 4 - Resultados da comparação entre modelos para dados de precipitação. ....	33
Tabela 5 - Resultados da comparação entre modelos para dados de temperatura. ....	33
Tabela 6 - Resumo estatístico descritivo dos dados de DBO. ....	35
Tabela 7 - Resultado do teste Mann-Kendall para DBO média. ....	37
Tabela 8 - Resultado do modelo para dados de DBO. ....	38
Tabela 9 - Resumo estatístico descritivo para os dados de nitrogênio amoniacal. ....	40
Tabela 10 - Resultado do teste de Mann-Kendall para a amônia. ....	41
Tabela 11- Resultado dos modelos propostos para dados de nitrogênio amoniacal. ....	43
Tabela 12 - Resumo estatístico descritivo dos dados dos coliformes termotolerantes. ....	44
Tabela 13 - Resultado do teste de Mann-Kendall para coliformes termotolerantes. ....	46
Tabela 14- Resultado dos modelos propostos para dados de coliformes. ....	48
Tabela 15 - Resumo estatístico descritivo dos dados da alcalinidade. ....	49
Tabela 16 - Resultado do teste de Mann-Kendall para os dados de alcalinidade. ....	51
Tabela 17- Resultado dos modelos propostos para dados de alcalinidade. ....	52
Tabela 18 - Resumo estatístico descritivo dos dados de sólidos suspensos totais. ....	53
Tabela 19 - Resultado do teste de Mann-Kendall para os dados de sólidos suspensos totais. ....	55
Tabela 20 - Resultado dos modelos propostos para dados de sólidos suspensos totais. ....	56

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>10</b>
<b>2</b>	<b>OBJETIVOS</b> .....	<b>12</b>
2.1	Objetivo geral .....	12
2.2	Objetivos específicos .....	12
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>13</b>
3.1	Monitoramento de águas do sistema de esgotamento sanitário .....	13
3.1.1	<i>Tipos de monitoramento da água</i> .....	13
3.1.2	<i>Efeitos alcançados pelo monitoramento</i> .....	14
3.2	Mineração de dados .....	14
3.2.1	<i>Séries temporais</i> .....	16
3.2.1.1	<i>Metodologia Box-Jenkins</i> .....	17
3.2.1.2	<i>Avaliação do modelo de previsão</i> .....	19
3.2.1.3	<i>Tendência e sazonalidade</i> .....	20
3.2.1.4	<i>Análise multivariada</i> .....	21
3.2.1.4.1	Análise de agrupamento .....	21
<b>4</b>	<b>MATERIAL E MÉTODOS</b> .....	<b>23</b>
4.1	Caracterização da área de estudo .....	23
4.2	Estudo histórico da área .....	24
4.2.1	<i>Evolução do sistema de esgotamento sanitário de Campina Grande</i> .....	24
4.3	Dados obtidos para a análise da qualidade das águas residuárias .....	26
4.4	Processo de análise do comportamento dos dados .....	26
4.4.1	<i>Pré-tratamento dos dados</i> .....	26
4.4.2	<i>Estatística descritiva</i> .....	26
4.4.3	<i>Análises de agrupamento e tendência</i> .....	27
4.4.4	<i>Metodologia das séries temporais</i> .....	27
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>28</b>
5.1	Processo de análise do comportamento dos dados .....	28
5.1.1.	<i>Comportamento da precipitação e temperatura</i> .....	28
5.1.1	<i>Comportamento da DBO</i> .....	35
5.1.2	<i>Comportamento do nitrogênio amoniacal</i> .....	39
5.1.3	<i>Comportamento dos coliformes termotolerantes</i> .....	44
5.1.4	<i>Comportamento da alcalinidade total</i> .....	49
5.1.5	<i>Comportamento dos Sólidos Suspensos Totais</i> .....	53
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>59</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>60</b>

## 1 INTRODUÇÃO

Devido ao processo deficiente da gestão de recursos hídricos existente na sociedade atual, ainda não é bem estabelecido uma metodologia para o entendimento ou associação de indicadores de qualidade dos recursos hídricos com fatores populacionais, industriais e comerciais, surgindo, então, a necessidade de um conhecimento mais aprofundado e particular para cada bacia de drenagem, a fim de obter um diagnóstico sobre a qualidade dos seus recursos hídricos, tornando esse conhecimento uma ferramenta de extrema importância, uma vez que, a partir dessas informações é possível inferir sobre as condições da bacia hidrográfica como um todo (QUEIROZ *et al.*, 2010).

A avaliação da qualidade da água inclui monitoramento, análise dos dados, elaboração de relatórios e disseminação de informações sobre as condições do ambiente aquático (BOYACIOGLU, 2008). O monitoramento e a avaliação das águas são procedimentos que permitem a caracterização e a análise de tendências em bacias hidrográficas, sendo essenciais para várias atividades, tais como planejamento, outorga, cobrança e enquadramento dos recursos hídricos (ANA, 2019).

As águas residuárias urbanas, cuja composição, segundo Campos (1999), compreende 99,9% de água e 0,1% de sólidos, constituem importante fonte de recursos hídricos, particularmente em regiões áridas e semiáridas, sujeitas aos eventos críticos das estiagens. Além dos contaminantes químicos e microbiológicos, a carga de nutrientes é outro aspecto relevante a ser considerado no processo de gestão de recursos hídricos, principalmente quando se considera o elevado potencial de reuso dessas águas.

As características das águas residuárias variam sazonalmente e, também, de acordo com os hábitos e situações socioeconômicas da população de cada região, requerendo a aplicação de métodos específicos de análise da variação de medidas de indicadores de qualidade em relação a fatores ambientais, econômicos e sociais, bem como fatores operacionais dos sistemas de esgotamento sanitário.

A monitoração da qualidade dessas águas residuárias, por longos períodos, produz grandes bancos de dados cuja complexidade também tem relação com o caráter estocástico de variações ao longo de séries de dados históricos. A análise dessas séries de dados pode requerer abordagens exploratórias com vistas à extração de informações específicas, compreensíveis e de utilidade para a gestão da qualidade dessas águas.

Métodos estatísticos multivariados têm sido citados como uma excelente ferramenta exploratória para a interpretação desses complexos conjuntos de informações. Trindade (2013)

afirma que, normalmente, são comuns trabalhos empregando algumas dessas técnicas (análise de cluster, análise discriminante e análise de componentes principais ou análise fatorial) para alcançar diferentes objetivos: extrair informações sobre as similaridades e as diferenças entre períodos e pontos de amostragem da qualidade da água, reconhecer os fatores responsáveis pelas variações e tendências espaciais e temporais dessa qualidade, identificar as fontes de poluição de diferentes pontos de um corpo d'água, identificar e analisar as relações entre química da água, uso e ocupação do solo e geologia.

Para estudar séries temporais que sofrem influência das condições climáticas, como precipitação, temperatura e umidade, Ferrari (2013) cita o teste de Mann-Kendall que, normalmente, é utilizado para verificar a ocorrência de tendências estatisticamente significativas em séries temporais. Recomendado pela Organização Meteorológica Mundial (OMM), o teste de Mann-Kendall requer que os dados sejam independentes e aleatórios, sendo o teste usado principalmente em séries climatológicas, associando mudanças climáticas com séries temporais, obtendo-se, assim noção do início, fim e comportamento de cada tendência.

Uma das técnicas quantitativas mais difundidas para o estudo de séries temporais é a metodologia de Box-Jenkins, descrita por esses autores na década de 1970. Os modelos de Box-Jenkins partem da ideia de que cada valor da série temporal pode ser explicado por valores prévios, a partir do uso da estrutura de correlação temporal que geralmente há entre os valores da série.

Portanto, o estudo sobre a variação temporal da qualidade das águas residuárias, neste trabalho, representa uma contribuição científica relevante no âmbito da aplicação de técnicas multivariadas associadas a séries temporais para avaliar e desenvolver conhecimento sobre o comportamento das águas residuárias ao longo de uma série de treze anos, no município de Campina Grande, fornecendo assim, subsídios técnicos consistentes que podem servir de orientação para estudos futuros.

## **2 OBJETIVOS**

### **2.1 Objetivo geral**

Desenvolver conhecimento sobre a variação temporal da qualidade de águas residuárias urbanas pela análise de series temporais com a metodologia Box Jenkins e o auxílio da Análise de agrupamento e teste de Mann-Kendall.

### **2.2 Objetivos específicos**

- Levantar os padrões de variação de dados sobre a qualidade de águas residuárias urbanas com base em análise exploratória de dados históricos;
- Identificar os principais fatores hidrológicos relacionados a variação do conjunto amostral;
- Modelar as tendências de variação temporal de indicadores de qualidade de águas residuárias urbanas;
- Avaliar o desempenho de modelos de interpretação da variação da qualidade de águas residuárias urbanas.

### 3 FUNDAMENTAÇÃO TEÓRICA

#### 3.1 Monitoramento de águas do sistema de esgotamento sanitário

De acordo com Von Sperling (2014), o aproveitamento adequado dos dados levantados em trabalhos experimentais ou observacionais, como programas de monitoramento, controle operacional ou pesquisa é de grande importância na Engenharia Sanitária e Ambiental.

Os dados coletados guardam em si conhecimento que uma análise superficial não permite adquirir, como por exemplo, a verificação de correlações entre parâmetros, mas com base no processo de monitoramento e trabalho estatístico sobre esses dados pode-se extrair informações poderosas para o gerenciamento desses efluentes.

Normalmente, em programa de monitoramento, a etapa de tratamento dos dados visa um maior conhecimento sobre suas características e objetiva obter inferências que podem ser confrontadas com as hipóteses ou modelos formulados.

O conhecimento sobre as características dos dados e sobre o comportamento do conjunto amostral normalmente precede a etapa mais aprofundada de comparação dos dados com as hipóteses ou modelos. Esta primeira etapa é coberta por procedimentos ligados à estatística descritiva visando descrever e resumir os dados.

##### *3.1.1 Tipos de monitoramento da água*

O processo de monitoramento das águas apresenta algumas finalidades distintas, tais como, o conhecimento sobre como os indicadores de qualidade se comportam com a influência de fatores internos e externos como precipitação, temperatura, insolação ou oxigenação da água, ou ainda, a determinação das condições e usos permitidos para os diversos fins.

De acordo com ANA (2019) é possível identificar quatro diferentes tipos de avaliações a respeito da qualidade das águas:

- i. O monitoramento básico que é feito em pontos estratégicos para o acompanhamento da evolução da qualidade das águas e identificação de tendências, em que os resultados obtidos permitem a identificação de locais onde é necessário um maior detalhamento e a frequência do monitoramento básico normalmente acompanha os ciclos hidrológicos;
- ii. Monitoramento de inventários, que compreende observações associadas à avaliação intensiva de um espectro mais ou menos amplo de indicadores com o objetivo de

- estabelecer um diagnóstico da qualidade das águas de um trecho específico de curso d'água;
- iii. Monitoramento de vigilância, nesse tipo de monitoramento incluem-se as observações efetuadas em locais onde a qualidade das águas é de fundamental importância para um determinado uso, especialmente para consumo humano, sendo necessário um monitoramento praticamente em tempo real;
  - iv. Monitoramento de conformidade, no qual busca-se a verificação do atendimento a requisitos, mesmo através de automonitoramento.

### ***3.1.2 Efeitos alcançados pelo monitoramento***

O monitoramento das águas em si não se comporta apenas como uma ferramenta de estudo, mas como fomentador de ideias e medidas para controlar e limitar os diversos usos das águas. Silva *et al.* (2017), realizaram o monitoramento da qualidade das águas em uma seção do córrego Limoeiro do município de Presidente Prudente e em uma seção do rio Pirapozinho no município de Pirapozinho, no período de 2006 a 2015.

Os autores constataram que com o passar dos anos, após a implantação das estações de tratamento de esgoto, em cada um dos municípios, ocorreu gradualmente a melhora dos indicadores de qualidade dessas águas, atestando assim, a melhoria na qualidade dos corpos aquáticos, com a diminuição da matéria orgânica e aumento da concentração de oxigênio dissolvido no meio aquático. O trabalho mostrou-se bastante útil para os gestores dos recursos hídricos, contribuindo para o planejamento das demandas e ofertas das águas superficiais.

Trindade *et al.* (2017) confirmam que a análise dos dados referentes a indicadores de qualidade das águas da sub-bacia do Rio das Velhas, em Minas Gerais, pode trazer informações como as similaridades e as diferenças entre os períodos do ano e os pontos de amostragem analisados, o reconhecimento dos indicadores responsáveis pela variação das tendências espaciais e temporais e a identificação dos pontos de poluição de um determinado fluxo de água. Além disso, o monitoramento desses indicadores pode ser usado como ferramenta auxiliadora na gestão da qualidade das águas em bacias hidrográficas.

## **3.2 Mineração de dados**

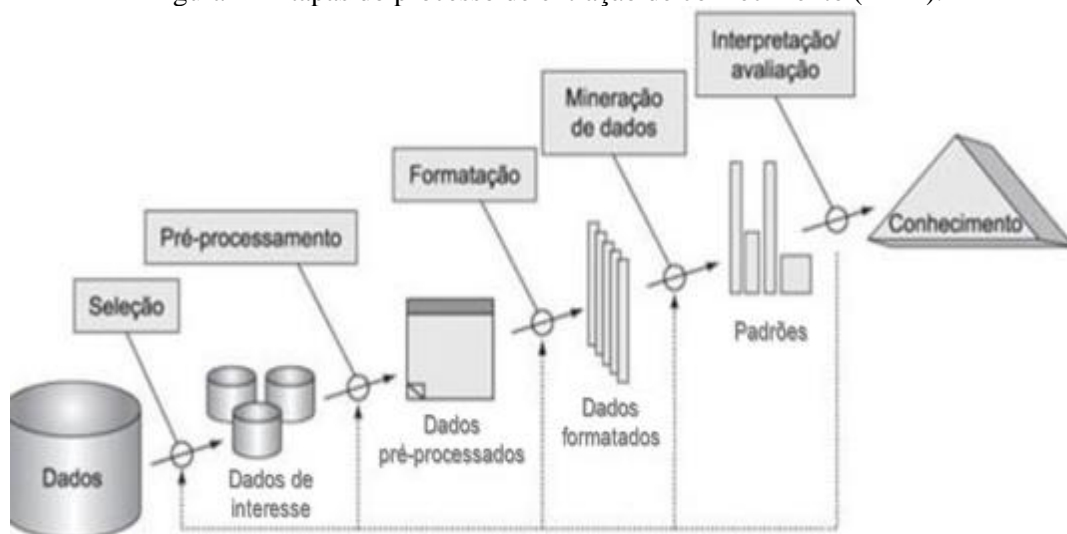
Segundo SAS (2019), a mineração de dados é o processo que se resume em encontrar anomalias, padrões e correlações em grandes conjuntos de dados, por meio de algoritmos



computacionais, para prever resultados. Com o levantamento de informações no processo de mineração de dados pode-se obter conhecimentos nos setores da economia, saúde e tecnologia, podendo assim, descobrir novas formas de lidar com problemas associados aos respectivos setores.

O processo de extração de conhecimento (*Knowledge Discovery in Databases - KDD*), representado esquematicamente na Figura 1, é definido por Fayyad *et al.* (1996) como “o processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”.

Figura 1 - Etapas do processo de extração de conhecimento (KDD).



Fonte: FAYYAD *et al.*, 1996.

Esse processo divide-se em cinco etapas:

- i. Seleção - etapa que visa selecionar um conjunto de dados, ou focar num subconjunto, onde a descoberta deve ser realizada;
- ii. Pré-processamento - fase em que ocorre a limpeza dos dados, que engloba operações básicas como a remoção de ruídos, quando necessária, a coleta da informação indispensável para modelar ou estimar ruído, a escolha de estratégias para lidar com os campos de dados ausentes, a formatação de dados de forma a adequá-los à ferramenta de mineração;
- iii. Formatação - etapa de transformação dos dados, na qual localizam-se características úteis para representar os dados dependendo do objetivo da tarefa, visando a redução do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados, bem como o enriquecimento das informações;
- iv. Mineração de dados - etapa na qual são selecionados os métodos a serem utilizados

para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação ou conjunto de representações e da busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão;

- v. Interpretação ou avaliação - etapa em que ocorre a implantação do conhecimento descoberto no desempenho do sistema, ou a documentação ou relato às partes interessadas.

Dentro desse processo podem ser aplicados diversos métodos estatísticos e algoritmos computacionais que auxiliam na identificação das correlações, podendo assim, gerar uma gama de conhecimentos que podem ser aplicados em qualquer área de estudo. Um exemplo da aplicação desse processo é o trabalho realizado por Gerolin (2018) com dados de águas subterrâneas, fornecidos pela Companhia Ambiental do Estado de São Paulo (CETESB), tendo sido utilizada a mineração de dados através de técnicas de classificação e associação para obter informações importantes a respeito das amostras coletadas, a fim de descobrir correlações entre parâmetros até então não conhecidas.

Todo conhecimento gerado pelo processo de mineração de dados traz vantagens concretas que podem ser utilizadas no processo de criação e adoção de medidas públicas ambientais e sociais, onde comunidades de baixa renda, por exemplo, podem se beneficiar com a disponibilidade de águas de reuso, próprias para irrigação ou dessedentação de animais, gerando emprego e renda.

Além disso, comunidades que recebem parcialmente ou quase não recebem ações de saneamento básico e que acabam contribuindo com o lançamento irregular de esgoto nos córregos e rios, serão mais notadas, uma vez identificadas como geradores de descargas ilegais e mostrando como esses lançamentos afetam a biodiversidade dos corpos hídricos, servindo como alerta para as autoridades responsáveis.

A mineração de dados voltada para o saneamento é um viés de mão dupla, pois ajuda os gestores a se organizarem e se planejarem, mas também serve para comprovar que medidas corretivas têm efeito a médio e a longo prazos, como a implantação de uma Estação de Tratamento de Esgoto (ETE) ao longo de um rio que, com o passar do tempo, acaba melhorando a qualidade das águas a jusante.

### ***3.2.1 Séries temporais***

Uma série temporal é um conjunto de observações ordenadas no tempo, nem sempre igualmente espaçadas, mas apresentando dependência entre intervalos de tempo. Pode ser enquadrada nesta categoria uma grande quantidade de fenômenos de natureza física, biológica, econômica, dentre outras. A maneira tradicional de analisar uma série temporal é através da sua decomposição (MORETTIN, 1987).

As séries temporais são construídas através de componentes como: ciclo, que é o movimento ondulatório que se repete ao longo do tempo; sazonalidade, que é o estudo do movimento ondulatório ao longo do tempo; tendência, análise do deslocamento da série como um todo de forma crescente ou decrescente, e aleatoriedade, caracterizadas por fatos que ocorrem ao longo da série sem uma sequência lógica no comportamento. (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).

As previsões alcançadas através da análise de séries temporais partem da ideia de que os valores atuais servirão para prever os valores futuros, utilizando abordagem estatística de previsão, buscando caracterizar padrões de comportamento para auxiliar no processo.

É importante destacar que os métodos de previsão não são exatos, sempre haverá erros, além de fatores internos e externos que influenciarão nas análises. A previsão em si não almeja um fim, mas sim, uma forma de viabilizar informações para auxiliar uma tomada de decisões de forma mais assertiva (TUBINO, 2017).

Partindo do princípio de que toda previsão não é exata, é importante em séries temporais, avaliar o comportamento dos valores da série antes de seguir com a previsão. Segundo Farrel e Stewart (2006) os testes estatísticos mais utilizados para avaliar as características dessas séries são:

- i. Normalidade – Teste de Kolmogorov-Smirnov, Anderson-Darling e Shapiro-Wilk.
- ii. Sazonalidade – Kruskal-Wallis e Friedman.
- iii. Tendência – Mann-Kendall, Wald-Wolfowitz, Cox-Stuart e Cuzik.
- iv. Correlação – Pearson, Spearman e Durbin-Watson.
- v. Estacionariedade – Dickey-Fuller aumentado, Phillips-Perron e KPSS.

### 3.2.1.1 Metodologia Box-Jenkins

Para a criação de um modelo Box-Jenkins deve ser levada em conta a estrutura de correlação temporal dos valores presentes da série para explicar o seu comportamento (WERNER; RIBEIRO, 2003). A metodologia Box-Jenkins já foi utilizada por autores como:

Sabino, Laje e Noronha (2017) para estudar as variações sazonal e temporal da qualidade das águas do córrego Gameleira em Uberaba - MG; Barbosa *et al.* (2015) que utilizaram um método para previsão de temperatura média mensal na cidade de Bauru - SP e Albuquerque *et al.* (2020) que utilizaram o método para dados de precipitação no sertão de Pernambuco.

Os modelos de Box-Jenkins, também conhecidos como modelos ARIMA traduzidos como modelos autorregressivos integrados de médias móveis, são modelos matemáticos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores dispostos na série temporal e, com base no comportamento dessas correlações, realizar previsões.

Segundo Fava (2000), os modelos ARIMA resultam da combinação de três componentes: o componente autorregressivo (AR) que se resume pela comparação do dado atual com valores do passado, levando em conta a correlação de todos os valores significativos de um período, o filtro de integração (I) - utilizado em séries não estacionárias, e o componente de médias móveis (MA) que também é a comparação do valor atual com o valor passado, mas levando em conta apenas a correlação do valor mais significativo do período.

. Uma série pode ser modelada pelos três filtros ou apenas por parte deles, resultando em modelos para:

- **Séries estacionárias:** modelo autorregressivo (AR), modelo de médias móveis (MA) e modelo autorregressivo de médias móveis (ARMA); e
- **Séries não estacionárias:** modelo autorregressivos integrado de médias móveis (ARIMA) e modelo sazonal autorregressivo integrado de médias móveis (SARIMA)

Esse último descrito como SARIMA (p, d, q) (P, D, Q)<sub>s</sub>, onde p e P são a ordem da componente autorregressiva, d e D números de diferenciações feitas na série, q e Q a ordem da componente média móvel, e S o ciclo sazonal estudado da série.

De acordo com Morretin e Tolo (1987), a metodologia Box-Jenkins se divide em um ciclo de etapas, no qual a escolha do modelo é feita com base na análise fornecida dos erros e resíduos, sendo elas, segundo Box e Jenkins (1976), a identificação, que consiste em descobrir qual dentre os vários modelos de Box-Jenkins, sejam eles sazonais ou não, descreve o comportamento da série, a partir da interpretação das funções de autocorrelações (ACF) e das funções de autocorrelações parciais (PACF), onde os componentes são estimados de acordo com o nível de significância, a estimativa dos parâmetros p e P do componente autorregressivo e os parâmetros q e Q do componente de médias móveis, e por fim, a verificação que avalia se o modelo estimado é adequado para descrever o comportamento dos dados.

Caso o modelo não seja adequado, o ciclo é repetido. Uma abordagem muito utilizada

é a não utilização de um único modelo, mas sim, alguns para serem avaliados. Quando se obtém um modelo satisfatório, chega a fase de realizar previsões, considerada a última etapa do método.

### 3.2.1.2 Avaliação do modelo de previsão

Todo modelo de previsão possui um certo grau de incerteza que precisa ser avaliado para se ter um bom desempenho, para avaliar esses modelos existem indicadores que auxiliam na avaliação do desempenho e na validação do modelo além de ajudarem no comparativo entre eles. Os valores previstos normalmente são constituídos de duas partes, uma que mede o valor esperado e outra aleatória que se distancia desse valor, onde normalmente essa parte aleatória se manifesta na forma de erro de previsão (RITZMAN; MALHORTA; KAJEWSKI, 2008).

De forma inicial, para se gerar medidas de erros de previsão de um período determinado ( $E_t$ ) se observa a diferença ou desvio entre o valor real ( $D_t$ ) do valor previsto ( $F_t$ ), conforme equação 1 (SOUZA; SAMOHYL; MIRANDA, 2008).

$$E_t = D_t - F_t \quad (\text{eq. 1})$$

Os indicadores mais utilizados na avaliação da precisão de um modelo são o Erro Médio Absoluto (MAE), que mede a dispersão dos erros da parte aleatória, quanto menor o MAE mais próximo do valor real a previsão se encontra, e o Erro Absoluto Médio Percentual (MAPE), que atua como um indicador da acurácia da previsão como uma porcentagem.

Uma outra forma de comparar modelos é a utilização do Critério de Informação de Akaike (AIC), de acordo com Sobral e Barreto (2016). Este critério é baseado na Divergência de Kullback-Leibler, que é uma medida da distância entre o modelo identificado e um teórico modelo real. Como o modelo real não é conhecido, Akaike (1973) desenvolveu uma forma de estimar esta distância através dos dados utilizados na modelagem, usando a função de verossimilhança e a ordem do modelo, como apresentado na Equação 2.

$$AIC = -2 \log(L(\hat{\theta} | \text{dados})) + 2p \quad (\text{eq. 2})$$

A medida que a verossimilhança aumenta, o termo  $-2 \log(L(\hat{\theta} | \text{dados}))$  decresce, enquanto o termo  $2p$  cresce sempre que a ordem do modelo for maior. Dessa forma, o critério de Akaike pondera entre a adequação aos dados e a complexidade do modelo, e quanto menor

o valor de AIC mais bem ajustado é o modelo.

### 3.2.1.3 *Tendência e sazonalidade*

A tendência de uma série temporal pode ser entendida como um fenômeno que reflete na elevação, no declínio ou na estabilidade, quando não há uma tendência, do valor médio da série ao longo do tempo, sem levar em conta mudanças periódicas, ou não significativas (YEVJEVICH, 1972; CHATFIELD, 1996).

Segundo Farrel e Stewart (2006) um dos testes mais utilizados para verificar a existência de tendência é o de Mann-Kendall que, de acordo com Luiz, Cardoso e Ribeiro (2012) tem sido um excelente teste para análise de tendências em séries de indicadores meteorológicos.

O teste de Mann-Kendall (MANN, 1945; KENDALL, 1975) é um método estatístico, sequencial e não-paramétrico, ou seja, não exige a distribuição normal dos dados, utilizado para verificar se uma série de dados apresenta uma tendência temporal, que seja estatisticamente significativa, sendo uma de suas principais vantagens o fato de ser pouco influenciado por mudanças severas ou séries que apresentem características não homogêneas (SALVIANO *et al.*, 2016).

Com a evolução dos anos foi proposto o teste de Kendall sazonal, uma adaptação feita para eventos que levam em conta tendências que ocorrem em determinadas etapas do ano, como por exemplo, as estações climáticas que influenciam drasticamente o clima. Então, todas as categorias que se encaixam no perfil “verão” são analisadas ao longo dos anos da série temporal, diferenciando a série por estação climática, não analisando-a como um todo.

Segundo Lamounier (2007) outro componente das séries temporais que está relacionado com o movimento acima e abaixo do valor médio é a sazonalidade. A diferença da sazonalidade e do ciclo está na preocupação nas observações em um ciclo de 12 meses, observando nesse intervalo períodos regulares de variações sazonais, com menor intervalo de tempo, as observações normalmente são dispostas em semanas, meses, trimestres, para que, assim, fique mais fácil de observar e modelar a sazonalidade.

Alguns métodos podem ser empregados para a análise do componente sazonal em uma série temporal, desde análise visual do gráfico da série, que mesmo sendo uma forma rudimentar de análise da série ainda é muito útil na prática, passando pelos modelos de regressão linear com variáveis independentes binárias (Variáveis Dummy), até a utilização de métodos mais sofisticados como proporcionado pelos modelos de análise espectral e SARIMA.

### 3.2.1.4 Análise multivariada

De acordo com Loesh (2012), a análise multivariada compreende um conjunto de técnicas estatísticas, envolvendo mais de duas variáveis, utilizadas para a análise de dados de uma amostra de observações.

Moita Neto (2004) afirma que cada método multivariado possui uma função, seja ela mostrar, prever ou otimizar relações entre variáveis. Dentre esses métodos, os mais utilizados são a Análise por Agrupamento Hierárquico (HCA) e a Análise de Componentes Principais (ACP), aplicados quando o interesse é verificar como as amostras se relacionam, a regressão linear múltipla e as redes neurais, cuja finalidade é fazer previsões e o planejamento fatorial, necessário quando se quer otimizar um processo.

#### 3.2.1.4.1 Análise de agrupamento

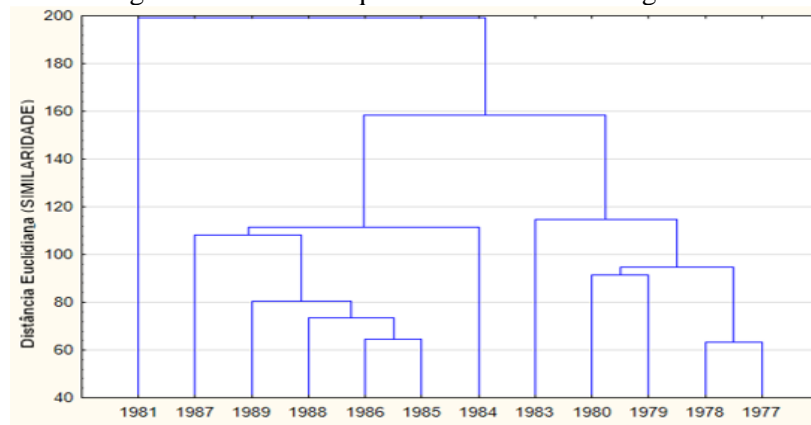
A análise de Cluster, também chamada de análise de agrupamento, é um método não supervisionado, ou seja, busca analisar padrões e correlações dentro de um banco de dados, dividindo uma grande quantidade de dados em pequenos conjuntos que são agrupados de acordo com as características que possuem. Esses conjuntos devem apresentar homogeneidade em suas características internas e também devem apresentar interrelações externas entre grupos (ZHANG *et al.*, 2011).

Wunderlin *et al.* (2001), afirmam que a análise de agrupamento se divide em hierárquica e não hierárquica. Nos métodos hierárquicos, o agrupamento é feito de forma sequencial, começando com o par mais semelhante entre os dados, formando sequencialmente outros grupos com níveis cada vez mais altos de correlação.

A similaridade dos dados para formar os grupos é dada pela distância entre dois pontos, também conhecida como distância euclidiana. Já para os métodos não hierárquicos é feita a avaliação geral da distribuição de pares de dados e depois há a classificação desses pares em um determinado número de grupos; nesse caso, o número de grupos (clusters) é determinado.

Tanto nos agrupamentos hierárquicos como nos não hierárquicos, o processo de formação e interligação entre os grupos é repetido até que um único grupo, contendo todas as amostras, seja obtido, sendo o resultado representado em um dendrograma conforme ilustrado na Figura 2.

Figura 2 - Modelo esquemático de um dendrograma.



Onde, no eixo das abscissas são apresentadas as categorias analisadas do mesmo parâmetro e no eixo das ordenadas a distância euclidiana, ou seja, a distância entre dois pontos, ou ainda a proximidade média dos dados de duas categorias.

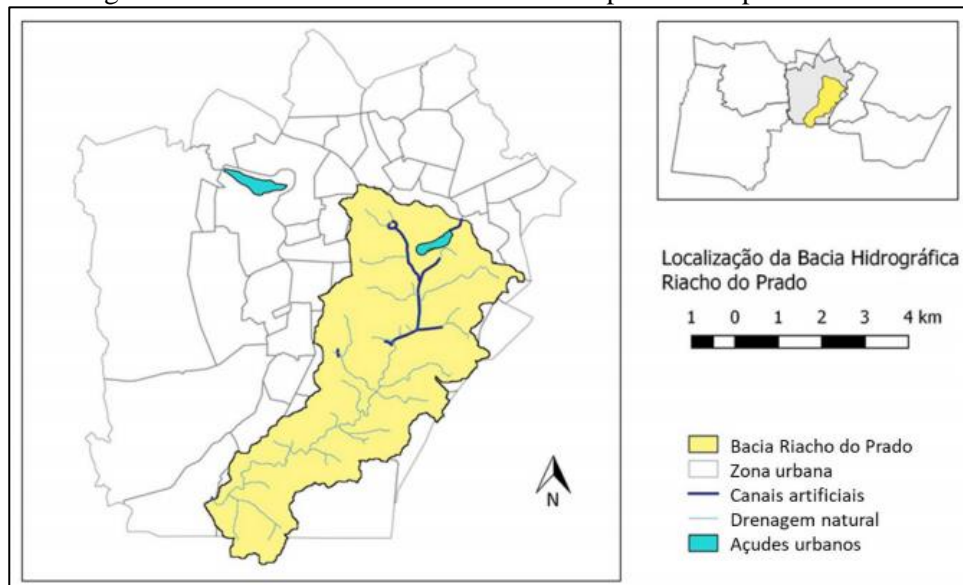


## 4 MATERIAL E MÉTODOS

### 4.1 Caracterização da área de estudo

A área de estudo analisada será a porção do sistema de esgotamento sanitário da cidade de Campina Grande, estado da Paraíba, situada na bacia urbana do Prado, ilustrada na Figura 3 que possui uma área de 37,15 km<sup>2</sup> e um perímetro de 43,78 km, sendo composta por 9 sub-bacias, abrangendo, parcial ou totalmente, 22 bairros. As águas residuárias dessa bacia são transportadas pelo Interceptor Leste (Interceptor da Depuradora), do sistema de esgotamento sanitário de Campina Grande.

Figura 3 – Bacia Urbana do Prado do município de Campina Grande - PB.



Fonte: CAMELO *et al.*, 2020.

As amostras que deram origem aos dados da série temporal de 13 anos, que foi analisada neste trabalho, foram oriundas de um projeto de monitoramento realizado entre 1977 e 1989, e tiveram como ponto de coleta um poço de visita do referido interceptor, no interior da EXTRABES, localizada no município de Campina Grande – PB.

Estão dispostas nesse banco de dados 607 observações, em média, para cada um dos seguintes indicadores de qualidade de águas residuárias, DBO, nitrogênio amoniacal, coliformes fecais, alcalinidade total e sólidos suspensos totais. As amostras de águas residuárias eram coletadas às 8 horas da manhã, com frequência variável no início do programa de monitoração, após cerca de um ano do início do programa essa frequência se tornou mais regular.

## 4.2 Estudo histórico da área

O estudo hidrológico histórico da área se deu por meio da análise de dados fornecidos pelo Instituto Nacional de Meteorologia (INMET), a exemplo dos dados de precipitação e temperatura, tendo sido obtido um banco de dados com 225 observações registradas pela estação 82795 situada na cidade de Campina Grande, no bairro Centenário, com coordenadas 7° 13' 48.00 de latitude sul e 35° 53' 60.00 de longitude oeste, com data de observação inicial em 28 de fevereiro de 1963 até 31 de dezembro de 1984.

Os dados operacionais do sistema de esgotamento sanitário foram levantados na Companhia de Água e Esgotos da Paraíba (CAGEPA), sendo de interesse deste trabalho o levantamento dos dados históricos a respeito de como novas áreas de drenagem foi sendo incorporada ao, então, novo sistema de esgotamento sanitário de Campina Grande, e quais possíveis formas de alteração no sistema poderiam alterar o comportamento dos indicadores de qualidade.

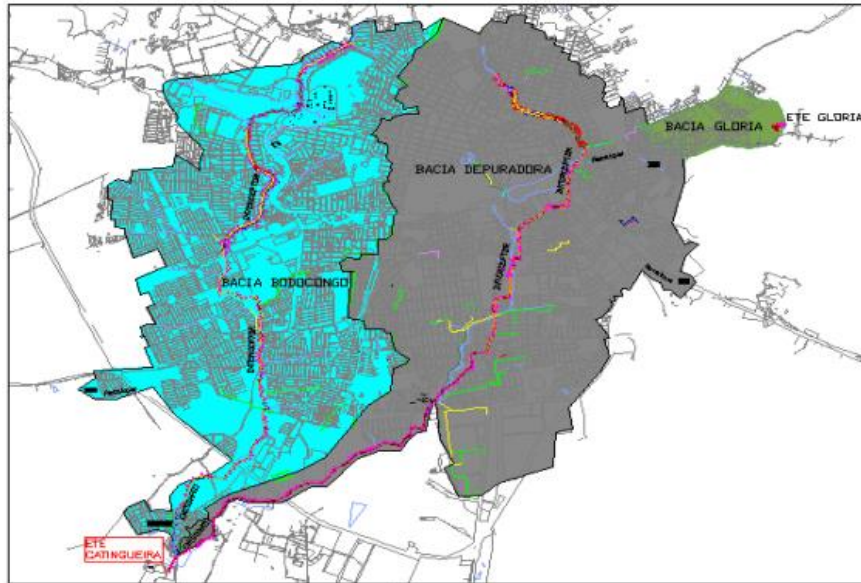
### *4.2.1 Evolução do sistema de esgotamento sanitário de Campina Grande*

De acordo com a CAGEPA (1964, 1985) o primeiro projeto de esgotamento sanitário de Campina Grande data de 1935, elaborado pelo Escritório Saturnino de Brito, o qual era um sistema de separação absoluta e escoamento por gravidade, atendendo as regiões dos bairros da Prata, São José e Centro. Além do sistema de esgotamento, esse escritório apresentou o projeto de uma estação de tratamento secundário de esgoto, denominada Estação Depuradora, já que nas imediações de Campina Grande não existe rio de elevada vazão capaz de suportar o lançamento direto dos despejos.

Atualmente, a antiga Estação Depuradora encontra-se desativada, funcionando no local, a EXTRABES que é utilizada para fins de pesquisa pela Universidade Estadual da Paraíba (UEPB) e Universidade Federal de Campina Grande (UFCG).

Entre os anos de 1969 e 1972, a CAGEPA contratou a empresa Planidro – Consultores de Engenharia Hidráulica e Sanitária Ltda. para a elaboração de um novo projeto compreendendo a ampliação da rede de esgotamento, o reforço do sistema já existente, a construção dos interceptores da Depuradora e Bodocongó e a construção de uma estação de tratamento através de duas lagoas aeradas em série, sendo previstos doze aeradores em cada lagoa. No entanto, na primeira lagoa foram instalados dez aeradores e na segunda dois aeradores (CAGEPA, 1985). A partir da Figura 4 é possível identificar as bacias drenadas pelo sistema de esgotamento sanitário de Campina Grande após a conclusão do projeto.

Figura 4 – Mapa das bacias drenadas pelo sistema de esgotamento sanitário do município de Campina Grande.



FONTE: GOMES, 2013.

Segundo Almeida (2007) e Gomes (2013), o projeto da Planidro foi dividido em duas etapas, a primeira teve início em 1973 e fim em 1974, sendo concluídos parcela da rede coletora, o Interceptor Leste ou da Depuradora, o emissário e as duas lagoas aeradas. Na segunda etapa, realizada em 1984, foram iniciadas as obras do Interceptor Oeste ou do Bodocongô.

Em 1985 a empresa Tecnosan – Engenharia S/A foi contratada pela CAGEPA para aprimorar a rede de esgotamento de Campina Grande, para atender, prioritariamente, 10 áreas até então não cobertas pelo sistema de esgotamento sanitário, assim como elencadas na Tabela 1.

Tabela 1- Áreas atendidas pelo projeto da Tecnosan em 1985.

ÁREA	LOCAL	EXTENSÃO (ha)	BACIA DE DESTINO
Área nº 1	Campus Universitário	263	Bacia de Bodocongô
Área nº 2	BR 230	52	Bacia de Bodocongô
Área nº 3	Conjunto Habitacional Caixa Econômica	11	Bacia de Bodocongô
Área nº 4	Juscelino Kubitschek	62	Bacia da Depuradora
Área nº 5	Colégio Palmeira	77	Bacia da Depuradora
Área nº 6	Alto Branco	39	Bacia da Depuradora
Área nº 7	Parque das Nações	97	Bacia da Depuradora
Área nº 8	Av. Manoel Tavares	232	Bacia da Depuradora
Área nº 9	Mirante	31	Bacia da Depuradora
Área nº 10	Estádio Ernani Sátiro	180	Bacia da Depuradora

Fonte: CAGEPA (1985).

O limite a oeste da área nº 2 é a faixa de servidão da CHESF, ao sul o projeto da área urbana e a o limite da área de projeto da Planidro. A área nº 3 limita-se ao norte com o conjunto habitacional Severino Cabral, a oeste com o conjunto habitacional bodocongó e a leste com o limite da área de projeto da Planidro. Já a área nº 5 situa-se ao final da rua 15 de novembro e à direita desta, no sentido centro periferia. A área nº 7 possui sua artéria principal a Avenida das Nações. As áreas nº 7, 8, 9 e 10 requereram estações elevatórias para fazer chegar seus efluentes até o Interceptor da Depuradora.

### **4.3 Dados obtidos para a análise da qualidade das águas residuárias**

Os arquivos de dados de qualidade das águas residuárias urbanas, sob a guarda dos atuais administradores da EXTRABES (professores da Universidade Estadual da Paraíba pertencentes ao grupo de pesquisa Saúde Ambiental) foram cedidos para análise exploratória neste trabalho.

Foram analisados o comportamento dos dados de precipitação e temperatura, assim como dos indicadores de qualidade das águas residuárias (DBO, Coliformes termotolerantes, Sólidos Suspensos Totais, Alcalinidade e Nitrogênio amoniacal). Esses são dados históricos do comportamento dos indicadores de qualidade do esgoto coletado na EXTRABES, por um período de treze anos, os quais servirão como base para o desenvolvimento de uma metodologia exploratória para o estudo do comportamento do esgoto em centros urbanos.

## **4.4 Processo de análise do comportamento dos dados**

### ***4.4.1 Pré-tratamento dos dados***

Em uma abordagem inicial, os dados foram encontrados de forma desordenada no tempo e, para se ter uma homogeneidade nas amostras e facilidade nas análises os mesmos foram distribuídos em médias mensais e os intervalos de tempo que não apresentavam análise alguma foram retirados da série, dando, assim, continuidade à metodologia com dados em médias mensais e com os intervalos de dados presentes.

### ***4.4.2 Estatística descritiva***

A primeira intervenção no sentido interpretativo dos dados foi a estatística descritiva que colaborou na compreensão do comportamento, da tendência central e a dispersão dos dados. Esta etapa foi realizada com o auxílio do *software* Action versão estudantil.

#### ***4.4.3 Análises de agrupamento e tendência***

Na análise da série como um todo e de como as variáveis podiam estar relacionadas com os registros da evolução histórica foi realizada a análise de agrupamento para a verificação de comportamentos semelhantes de variáveis diferentes e de uma mesma variável em espaços de tempos distintos.

Simultaneamente a essa análise foi realizado o levantamento da tendência da série a fim de corroborar o entendimento do comportamento e a formulação de teorias a respeito do procederl da série. Os dois métodos foram aplicados com o auxílio do *software* STATISTICA 12.

#### ***4.4.4 Metodologia das séries temporais***

Ao adotar o método Box-Jenkins para analisar as séries temporais, é verificado inicialmente se a série é estacionaria ou não através do teste de Dickey-Fuller aumentado, se a mesma não for, será necessário diferenciar os valores da mesma até atingir tal condição, o número de diferenciações feitas será o número do filtro de integração (I), com a série estacionaria é plotado então, o gráfico de autocorrelação (FAC) que indica a ordem da componente autorregressivo (AR) através do *lag* inicial mais significativo do gráfico, e o de autocorrelação parcial (FACP) que aponta a ordem da componente média móvel (MA) também através do *lag* inicial mais significativo do gráfico,

Caso os gráficos FAC e FACP apresentem sazonalidade significativa, ou seja, comportamento periódico significativo dos *lags*, as mesmas componentes de média móvel e autorregressiva podem ser encontradas para a parte sazonal da série, tornando então um modelo SARIMA (p, d, q) (P, D, Q)<sub>s</sub>.

Após o levantamento dos modelos possíveis através da análise de significância dos *lags*, os modelos foram avaliados de acordo com o menor valor de AIC, MAE e MAPE, em seguida foi realizado uma análise dos resíduos para verificar a interferência dos mesmos na modelagem da série e, por fim, é plotado o gráfico das previsões com o modelo escolhido e verificado se o comportamento das previsões se assemelha ao comportamento da série. Toda a análise de séries temporais foi realizada com o auxílio do *software* R com os pacotes forecast (HYNDMAN, 2021), fpp2 (HYNDMAN, 2020), readxl (WICKHAM H., 2019), urca (PFAFF B., 2016), tseries (TRAPLETTI A., 2020) e FitAR (MCLEOD A., 2013).

## 5 RESULTADOS E DISCUSSÃO

Neste capítulo serão descritos os resultados obtidos na análise estatística descritiva, na análise de agrupamento, no teste de Mann-Kendall e da metodologia de séries temporais para as variáveis precipitação, temperatura, DBO, nitrogênio amoniacal, coliformes termotolerantes, alcalinidade total e Sólidos Suspensos Totais.

### 5.1 Processo de análise do comportamento dos dados

#### 5.1.1. Comportamento da precipitação e temperatura

Os dados de precipitação medidos pela estação OMM:82795 e de temperatura média mensal, para o período 1977 a 1984, foram fornecidos pelo INMET e estão resumidos através de parâmetros estatísticos descritivos como: média, mediana, desvio padrão e tamanho da amostra, conforme a Tabela 2.

Tabela 2 - Resumo descritivo dos dados da precipitação e temperatura média mensal.

	Precipitação	Temperatura média
1º quartil	14 mm	22,6°C
Média	57,3 mm	23,7 °C
Mediana	41,8 mm	24,0 °C
3º quartil	85,4 mm	24,9 °C
Desvio Padrão	53,6mm	1,4 °C
Tamanho da Amostra	94	94

Verifica-se que entre o 1º e 3º quartil, o que representa 75% do total dos dados amostrais, os dados de precipitação estão entre 14 e 85,4 mm, com temperatura variando entre 22,6 e 24,9°C. Além disso, foi possível observar um desvio padrão de 53,6mm para a precipitação, indicando que os valores amostrais estão mais dispersos em torno da média. Por outro lado, o desvio padrão para temperatura foi de 1,4 °C e nesse caso, os dados estão mais próximos da média. Com base nos valores médios mensais foram construídos os dendrogramas de similaridade para os meses do período estudado e os gráficos da precipitação e temperatura, ilustrados respectivamente pelas Figuras 5 e 6.

O eixo horizontal do dendrograma representa os meses do ano e o eixo vertical, o quadrado da distância euclidiana, quanto menor o valor do quadrado da distância euclidiana,

maior será a similaridade entre o conjunto de meses dentro de um grupo e menor será a semelhança entre os grupos formados.

Figura 5 - Dendrograma referente aos anos de 1977 e 1984: (a) precipitação (b) temperatura média

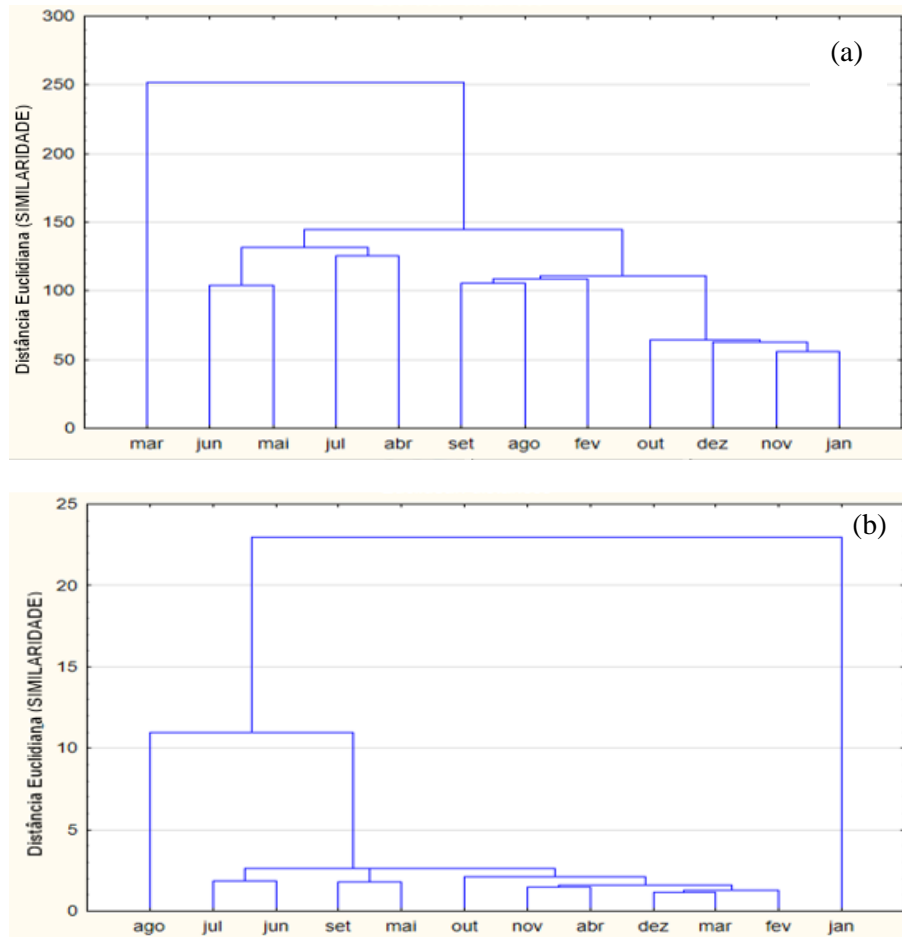
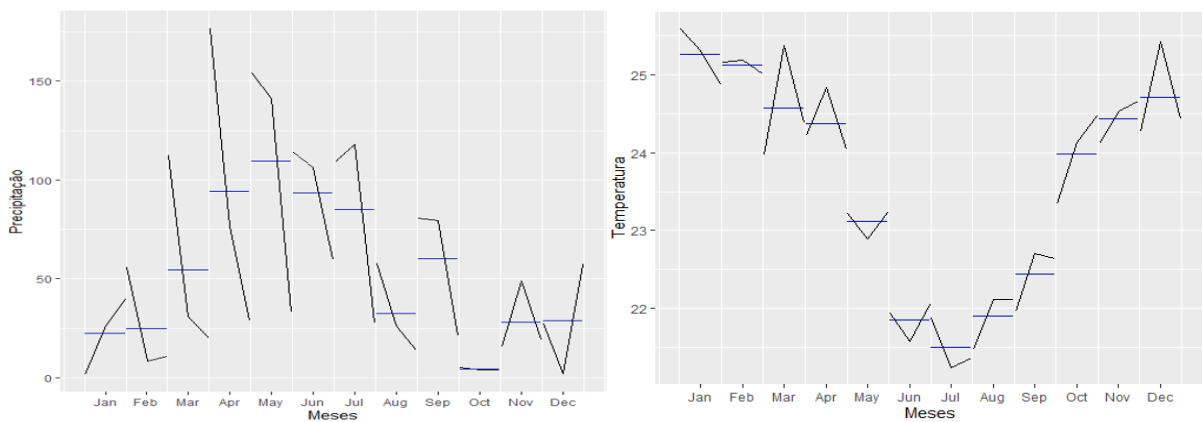


Figura 6 - Gráfico da precipitação e da temperatura média mensal entre os anos 1977 e 1984.



Analisando o dendrograma da precipitação, Figura 5(a) é possível observar a formação de três grupos distintos bem formados. O primeiro grupo corresponde aos meses que vão de

abril a julho, em que de acordo com o gráfico da precipitação média mensal são caracterizados por um período mais chuvoso, o segundo grupo representa os meses setembro, agosto e fevereiro, que apresenta características com índices de precipitação de nível médio e um terceiro grupo formado pelos meses entre outubro e janeiro que são os meses com menores índices de precipitação

O dendrograma da Figura 5(b) não possibilitou a caracterização de uma divisão muito clara das estações do ano contudo, o gráfico da temperatura média ao longo dos meses para o período de anos estudado indica uma baixa na temperatura entre os meses de junho a setembro que compreende o período de inverno da região, seguido por um período em que a temperatura permanece constante ao longo dos meses do ano.

Corroborando com os resultados encontrados Cavalcante, Lucena e Moura (2019) descrevem o clima do agreste Paraibano, com base na classificação de Köppen, como clima semiárido, apresentando elevadas temperaturas entre os meses de dezembro a maio e temperaturas mais amenas nos meses de junho, julho e agosto.

Diniz (2013) realizou o estudo da variabilidade da precipitação no município de Campina Grande, por análise dos coeficientes de correlação, e observou que as maiores médias mensais de precipitação são verificadas entre os meses de maio e julho, o que possivelmente explicaria os meses de menores temperaturas na região.

Na aplicação do teste de Mann-Kendall, exposto na Tabela 3, a presença de uma tendência estatisticamente significativa é avaliada utilizando o valor do índice ZMK. Essa estatística é usada para testar a hipótese nula, ou seja, que nenhuma tendência existe. Um valor positivo de ZMK indica um aumento de tendência, enquanto que um valor para ZMK negativo indica uma tendência decrescente.

Tabela 3 - Resultado do teste de Mann-Kendall para a precipitação e temperatura

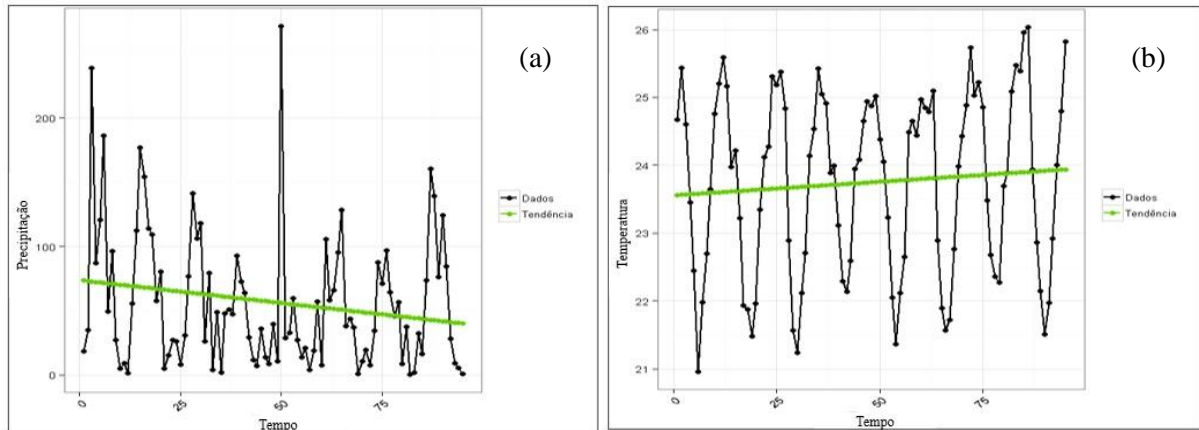
	<b>Precipitação</b>	<b>Temperatura</b>
ZMK	-0,109543011	0,06024636
P-Valor	0,116652489	0,388883352
Tamanho da Amostra	94	94
Hipótese Nula	Não há tendência	Não há tendência
Hipótese Alternativa	Há tendência	Há tendência

Com a aplicação do teste de Mann-Kendall ao analisar o p-valor do teste, é possível afirmar que em ambos os casos o p-valor calculado foi maior que o nível de significância ( $\alpha = 0,05$ ) portanto, a hipótese nula é aceita, indicando que não há tendência significativa na série, ou seja, não há mudança significativa no comportamento dos dados ao longo dos anos. Além



disso, é possível verificar que o teste indica uma tendência decrescente para a precipitação, de acordo com o valor do ZMK negativo, e uma tendência crescente para a temperatura. Podendo ser confirmado a partir dos gráficos expostos na Figura 7.

Figura 7 - Gráfico da análise de tendência: (a) precipitação; (b) temperatura



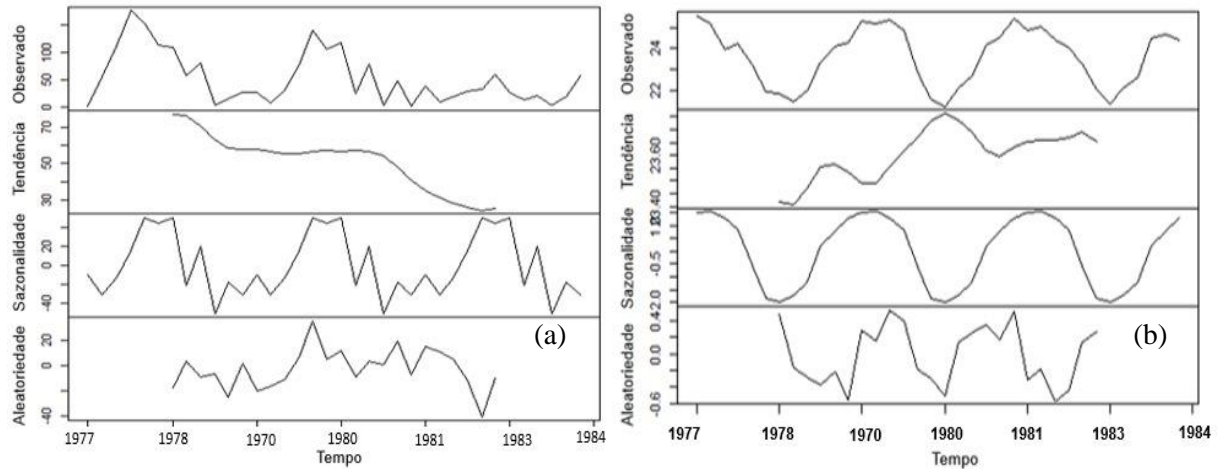
Observa-se, por meio da curva de regressão linear da Figura 7(a) que os dados apresentaram tendência de diminuição da precipitação com o passar dos anos, bem como um aumento da temperatura, Figura 7(b), confirmando o teste Mann-Kendall.

Esse resultado pode ser explicado, possivelmente, em função de eventos meteorológicos comumente presentes na região analisada, além das modificações no uso do solo e as altas taxas de crescimento urbano experimentadas pelas cidades brasileiras de médio a grande porte causando alterações nos índices de precipitação.

Dantas, Santos e Olinda (2015), afirmam em seu trabalho que as temperaturas mínimas e máximas têm aumentado gradativamente ao longo das últimas décadas em Campina Grande, fato possivelmente relacionado à intensa expansão urbana na região associada ao aumento do número de ruas com pavimentação asfáltica e redução de áreas com cobertura vegetal.

Como o teste de Mann-Kendall é um teste não paramétrico e que não é influenciado por pequenas alterações nos dados, analisaram-se os gráficos da decomposição, conforme demonstrado na Figura 8.

Figura 8 - Gráficos da decomposição: (a) dados de precipitação (b) dados de temperatura



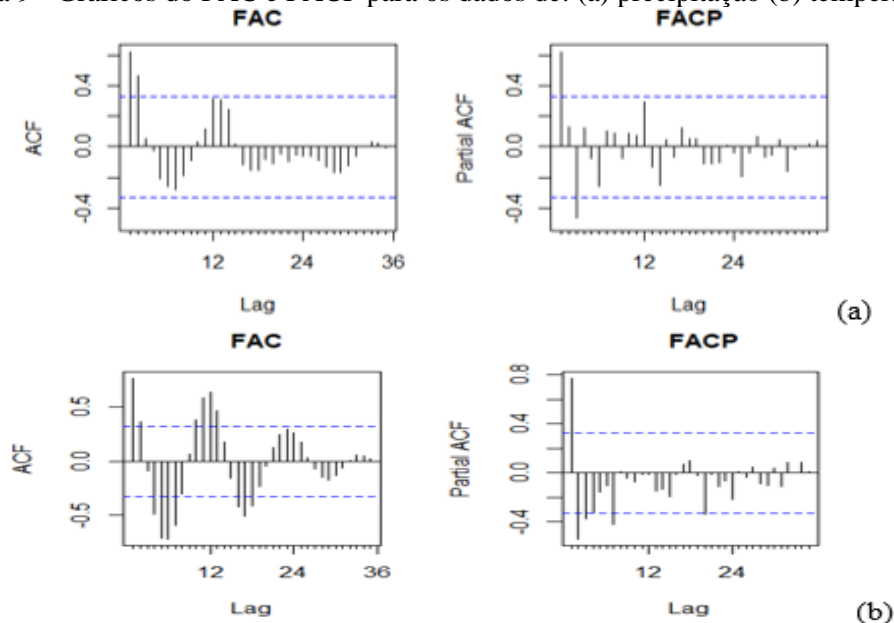
Ao decompor a série temporal, observa-se o efeito da tendência decrescente para precipitação (Figura 8a) e uma tendência com comportamento crescente para a temperatura ao longo do tempo, mas com certa estabilidade durante o ano de 1979 indicando pouca variabilidade dos dados e o efeito sazonal característico da série. O efeito sazonal da temperatura apresenta uma diminuição nos períodos de maior precipitação mensal indicando possível correlação entre as variáveis.

Dantas, Santos e Olinda (2015) realizaram uma pesquisa envolvendo as tendências anuais e sazonais nos extremos de temperatura do ar e precipitação na cidade de Campina Grande - PB entre os anos de 1975 e 2011 e detectaram que a precipitação não apresentava tendência estatisticamente significativa para o resultado encontrado. Além disso, verificaram que a ausência de tendências com valor significativo para os índices anuais de precipitação pode estar relacionada à irregularidade das chuvas sobre a região estudada, apresentando alta variabilidade espaço-temporal.

Observando os gráficos FAC e FACP da série temporal do indicador precipitação (Figura 9a), e sabendo pelo teste Dickey-Fuller aumentado que a série só se torna estacionária após a primeira diferenciação com p-valor igual a 0,03, podemos chegar a três modelos: SARIMA (1,1,1) (0,1,0)<sub>12</sub> e SARIMA (3,1,1) (0,1,0)<sub>12</sub> e SARIMA (0,0,0) (0,1,0)<sub>12</sub>, sendo os dois primeiros obtidos a partir de interpretações gráficas e último sugerido pela ferramenta auto.arima do *software* R.

Por outro lado, os gráficos FAC e FACP da série temporal para temperatura (Figura 9b) indicam que a série já é estacionária em p-valor igual a 0,01 portanto, podemos chegar ao modelo de SARIMA (1,0,1) (0,1,1)<sub>12</sub>, obtidos a partir de interpretações gráficas e ao modelo SARIMA (1,0,0) (1,1,0)<sub>12</sub>, sugerido pela ferramenta auto.arima do *software* R.

Figura 9 - Gráficos do FAC e FACP para os dados de: (a) precipitação (b) temperatura.



A partir dos dados expostos na Tabela 4 e 5 é possível identificar os resultados da comparação realizada entre os modelos obtidos para precipitação e temperatura, respectivamente.

Tabela 4 - Resultados da comparação entre modelos para dados de precipitação.

MODELO	AIC	MAE	MAPE
$(0,0,0)(0,1,0)_{12}$	250,72	22,42	100,14
$(1,1,1)(0,1,0)_{12}$	240,91	20,56	149,27
$(3,1,1)(0,1,0)_{12}$	240,48	18,99	138,72

Comparando os resultados do AIC, MAE e MAPE de cada modelo, chegamos a um melhor ajuste no modelo SARIMA  $(3,1,1)(0,1,0)_{12}$ , foi preciso diferenciar uma vez o modelo para torná-lo estacionário, tanto analisando como um todo como analisando sua parte sazonal de 12 meses.

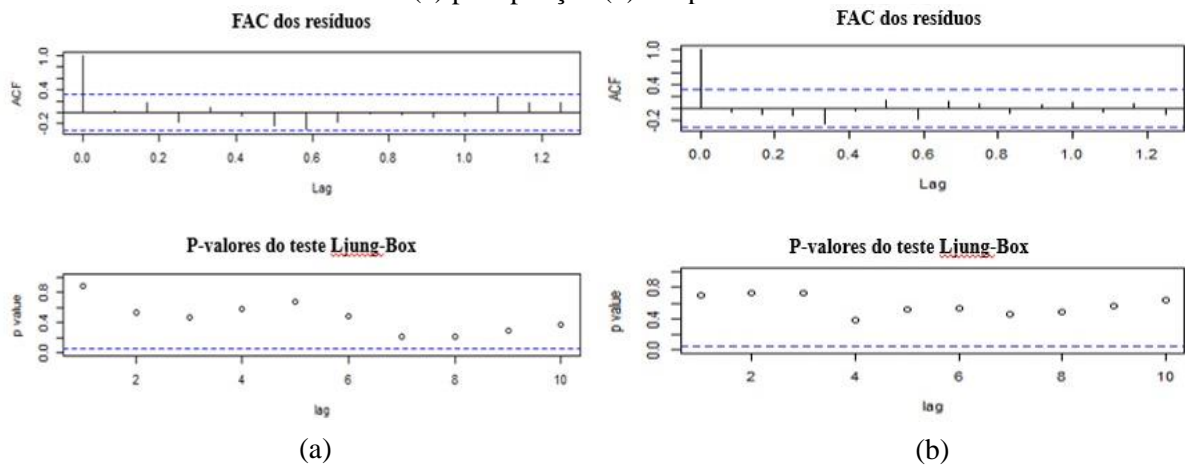
De forma similar, o comparativo dos resultados do AIC, MAE e MAPE para os dados da temperatura (Tabela 5), indicam uma leve diferença tanto no MAE quanto no MAPE, dessa forma, utilizou-se o AIC como critério. Portanto, o modelo mais ajustado aos dados foi o SARIMA  $(1,0,0)(1,1,0)_{12}$ .

Tabela 5 - Resultados da comparação entre modelos para dados de temperatura.

MODELO	AIC	MAE	MAPE
$(1,0,1)(0,1,1)_{12}$	45,58	0,22	0,95
$(1,0,0)(1,1,0)_{12}$	42,81	0,25	1,05

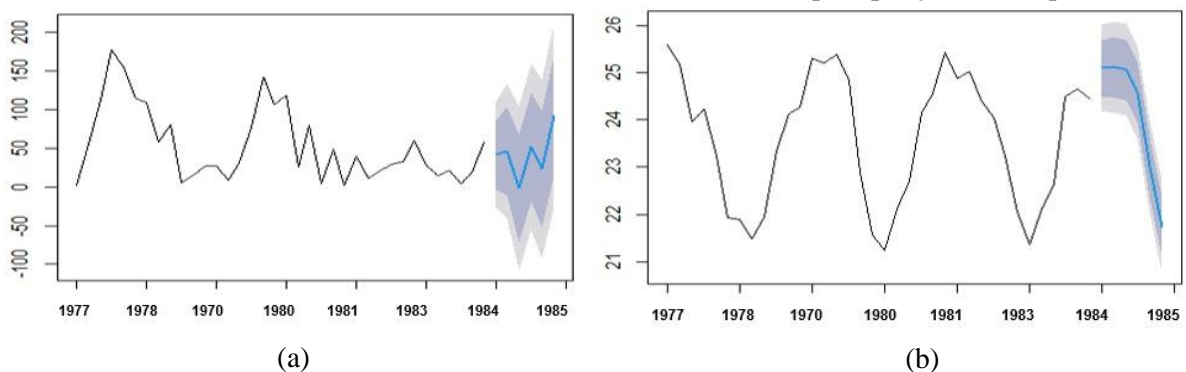
Por fim, foi realizado a análise dos resíduos do modelo SARIMA (3,1,1) (0,1,0)<sub>12</sub> e SARIMA (1,0,0)(1,1,0)<sub>12</sub>, bem como a previsão com o modelo proposto para a precipitação e temperatura, dos seis meses seguintes, demonstradas nas Figuras 10 e 11, respectivamente.

Figura 10 – Análise dos resíduos do modelo proposto para os dados de:  
(a) precipitação (b) temperatura



Analisando o gráfico 10a e 10b é possível verificar que os resíduos do modelo não evidenciam nenhum padrão, ou seja, não apresentam autocorrelação, reforçado pelo teste estatístico de Ljung-Box. Conclui-se que os resíduos são não autocorrelacionados, confirmando a boa adequação do modelo à série em questão. Logo, ambos os modelos podem ser utilizados para fazer previsões futuras.

Figura 11 – Previsão com o modelo proposto para os dados de: (a) precipitação (b) temperatura.



Na Figura 11a temos a série temporal do indicador precipitação e em destaque, as previsões para os 6 meses seguintes, seguindo o modelo SARIMA (3,1,1) (0,1,0)<sub>12</sub>. Por outro lado, na Figura 11b está apresentado a série temporal do indicador temperatura em que pode-se perceber um decaimento de temperatura média mensal nos seis meses seguintes.

O período de seis meses foi adotado tomando como base o planejamento para ações a curto prazo, lembrando que quando maior o período de previsão mais distorcido o valor pode

ser do valor real. É possível verificar que as previsões geradas pelo modelo fornecem comportamento similar ao comportamento passado da serie.

### 5.1.1 Comportamento da DBO

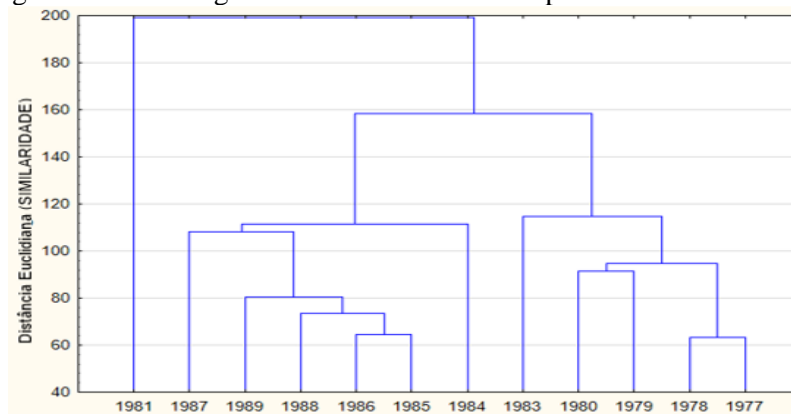
Os parâmetros estatísticos descritivos de média, mediana, desvio padrão e tamanho da amostra para os valores dos dados de DBO do período compreendido entre o período de 1977 a 1989 estão descritos na Tabela 6.

Tabela 6 - Resumo estatístico descritivo dos dados de DBO.

1º quartil	146,5
Média	223,8
Mediana	221,5
3º quartil	297,3
Desvio Padrão	90,4
Tamanho da Amostra	127

Verifica-se que entre o 1º e 3º quartil, os dados de DBO apresentam-se entre 146,5 e 297,3 mg.L<sup>-1</sup> com um desvio padrão de 90,4 mg.L<sup>-1</sup>, indicando a alta variação dos valores amostrais em torno da média. Com base nos valores anuais foi construído o dendrograma de similaridade para período estudado, como pode ser verificado na Figura 12.

Figura 12 - Dendrograma da DBO referente ao período de 1977 a 1989.



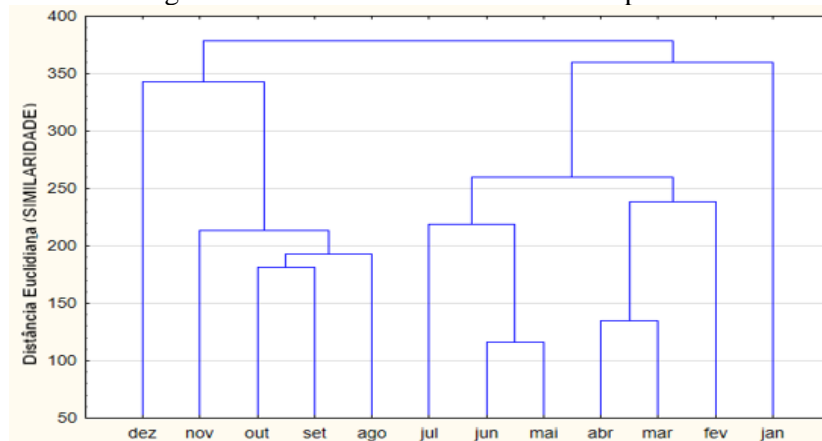
A análise de agrupamentos (clusterização) para os dados do indicador de qualidade DBO, no período de 1977 a 1989, permite a observação de dois grupos distintos, em que o primeiro grupo compreende os anos de 1984 a 1989 e o segundo grupo os anos de 1977 a 1980 e o ano de 1983, indicando que as aplicações feitas no sistema de esgotamento sanitário nos anos de 1984 e 1985, estão influenciando o comportamento dos dados, mostrando possível

diluição do efluente ou autodepuração do mesmo.

É importante notar que dentro dos grupos ainda existem particularidades que deixam um ano ligado ao outro de forma mais próxima, indicando que eles apresentam características semelhantes entre si. Por seu turno, a distinção entre alguns grupos de anos pode ser atribuída a alterações na malha de transporte desses efluentes, a exemplo de expansões na rede, com a integração de novas contribuições.

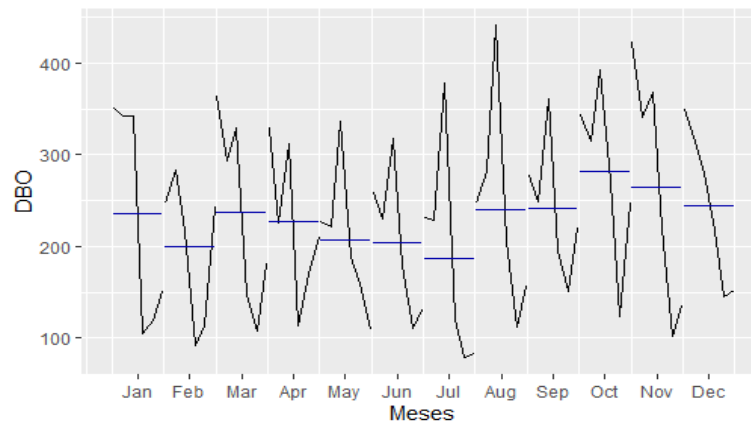
De forma semelhante, foi realizado o dendrograma das médias mensais da DBO para o mesmo período analisado, bem como seu gráfico, podendo ser observado nas Figuras 13 e 14, respectivamente.

Figura 13 - Dendrograma das médias mensais da DBO no período de 1977 a 1989.



Ao se analisar o comportamento médio mensal da DBO, verifica-se uma leve queda nos períodos de maior precipitação, como nos meses de maio, junho e julho, ocorrendo, como previsto, a diminuição da DBO devido à diluição dos esgotos por água pluvial que tem acesso à rede coletora.

Figura 14 - Gráfico das médias mensais de DBO ao longo do período de 1977 a 1989.



Com a aplicação do teste de Mann-Kendall (Tabela 7), ao analisar o p-valor ( $\alpha = 0,05$ ) do teste, aceita-se a hipótese alternativa, em que há existência de tendência significativa na

série, ou seja, há mudança significativa no comportamento dos dados ao longo dos anos além disso, é possível verificar que o teste indica uma tendência decrescente de acordo com o valor do ZMK negativo.

Tabela 7 - Resultado do teste Mann-Kendall para DBO média.

ZMK	-0,426427186
P-Valor	9,54E-13
Tamanho da Amostra	128
Hipótese Nula	Não há tendência
Hipótese Alternativa	Há tendência

Sabendo que o teste de Mann-Kendall não é influenciado por pequenas alterações nos dados, foi realizado a análise do gráfico da tendência da DBO, bem como os gráficos da decomposição, assim como pode ser observado, respectivamente, nas Figuras 15 e 16.

Figura 15 - Gráfico da análise de tendência da DBO

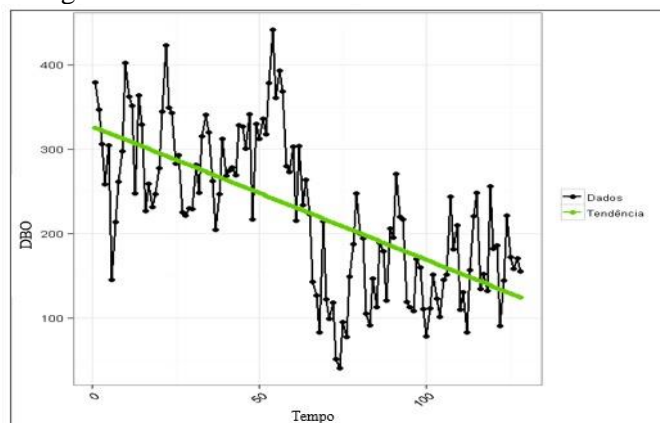
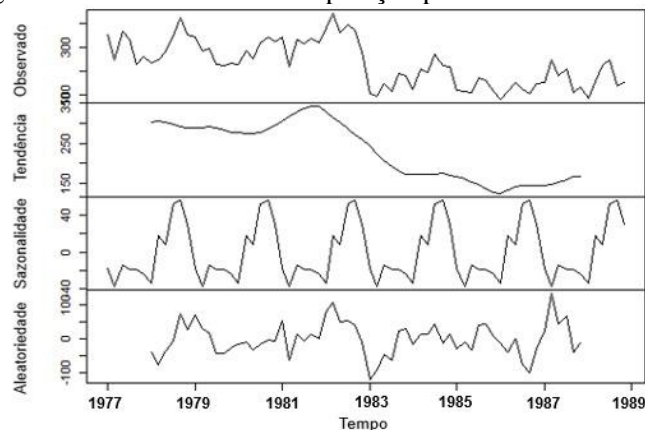


Figura 16 - Gráficos da decomposição para os dados de DBO.



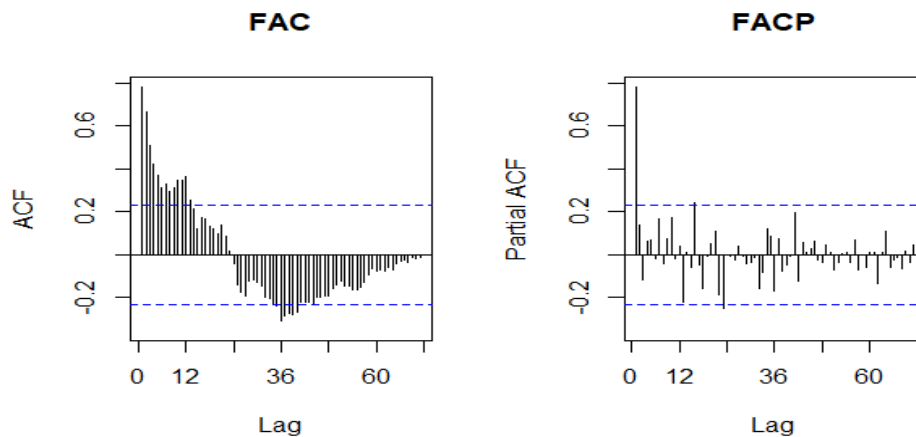
A partir da análise do gráfico de tendência foi possível verificar que há uma queda no valor da DBO com o passar dos anos, confirmando o teste de Mann-Kendall. Com relação a decomposição da série temporal do indicador DBO identifica-se um comportamento de queda

da tendência iniciada no segundo semestre de 1981 indicando possível diluição, por ação antrópica ou modificação do sistema, e um comportamento sazonal característico de elevação dos valores da DBO iniciando-se no meio de cada ano e finalizando no início do ano seguinte.

A diminuição da DBO ao longo dos anos pode ser explicada a partir do aumento do sistema de esgotamento sanitário, iniciado em 1984, segundo Gomes (2013), causando a autodepuração ou a diluição do efluente a partir de novas colaborações ao sistema.

Analisando os gráficos FAC e FACP da série temporal do indicador DBO, apresentados na Figura 17, e sabendo pelo teste Dickey-Fuller aumentado que a série só se torna estacionária após a sua primeira diferenciação (p-valor = 0,01), pode-se chegar ao modelo SARIMA  $(1,1,1)(1,0,0)_{12}$  sugerido pela ferramenta *auto.arima* do *software R*.

Figura 17- Gráficos do FAC e FACP para os dados de DBO.



A partir dos dados expostos na Tabela 8 é possível verificar os resultados do AIC, MAE e MAPE para o modelo sugerido.

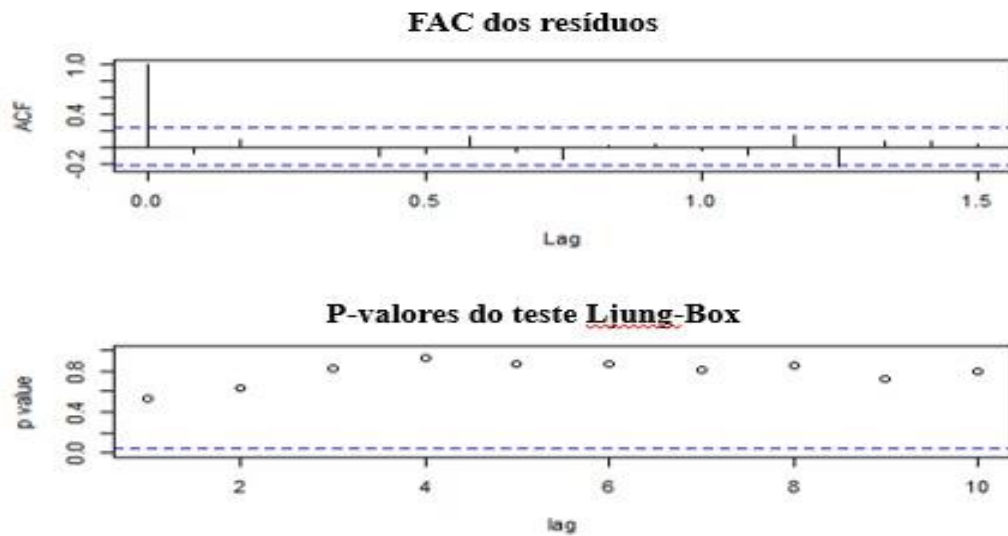
Tabela 8 - Resultado do modelo para dados de DBO.

MODELO	AIC	MAE	MAPE
$(1,1,1)(1,0,0)_{12}$	776	41,23	22,73

levantamento dos resíduos do modelo SARIMA  $(1,1,1)(1,0,0)_{12}$  foi realizado utilizando o *software R* como auxiliador, bem como a previsão com o modelo proposto para a DBO dos seis meses seguintes, como pode ser verificado, respectivamente, nas Figuras 18 e 19.



Figura 18 – Análise dos resíduos do modelo proposto para os dados de DBO.



A análise do gráfico da Figura 18 indica que não há autocorrelação significativa, e os p-valores do teste Ljung-Box não são significativos, com isso definimos que os resíduos não interferem na modelagem da série.

Figura 19 – Previsão com o modelo proposto para os dados de DBO.



A previsão com o modelo proposto do indicador DBO indica que os 6 meses seguintes fornecem estimativas razoáveis de acordo com o comportamento já mostrado da série.

### ***5.1.2 Comportamento do nitrogênio amoniacal***

A partir do resumo estatístico, demonstrado na Tabela 9, é possível verificar que 75% dos dados de nitrogênio amoniacal estão entre 31,6 e 49,2 mgN.L<sup>-1</sup> apresentando um desvio padrão de 10,1 mgN.L<sup>-1</sup>, o que indica a alta dispersão dos dados em torno da média amostral.

Tabela 9 - Resumo estatístico descritivo para os dados de nitrogênio amoniacal.

1° quartil	31,6
Média	39,9
Mediana	40
3° quartil	49,2
Desvio Padrão	10,1
Tamanho da Amostra	123

Oliveira (2006) realizou a avaliação do comportamento de espécies nutrientes (N e P) em uma lagoa anaeróbia profunda, que trata esgotos domésticos e está localizado no Bairro do Roger na cidade de Joao Pessoa. Verificou-se que na ETE do Roger que o Nitrogênio amoniacal apresentou média de  $37,9 \text{ mgN.L}^{-1}$  e um desvio padrão de  $11,3 \text{ mgN.L}^{-1}$  para o esgoto bruto. Dados similares aos verificados nessa pesquisa.

Foram construídos os dendrogramas de similaridade para os anos compreendidos entre 1977 a 1989 (Figura 20), bem como para a concentração média mensal de nitrogênio amoniacal (Figura 21).

Figura 20 - Dendrograma em anos, referente aos valores do nitrogênio amoniacal

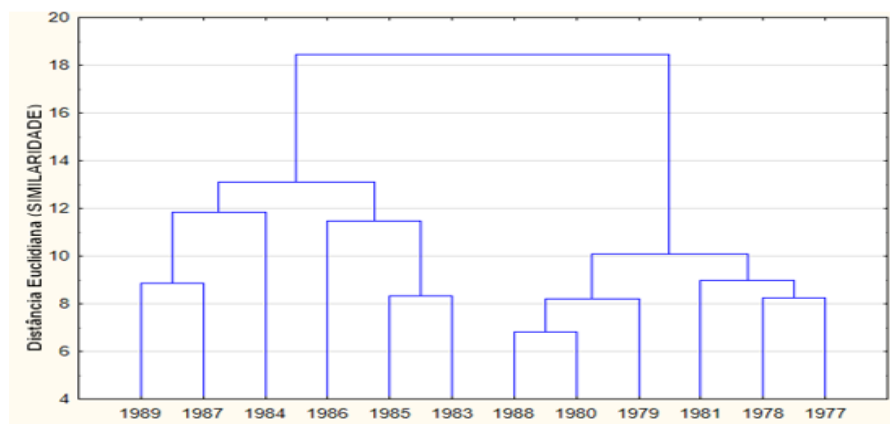
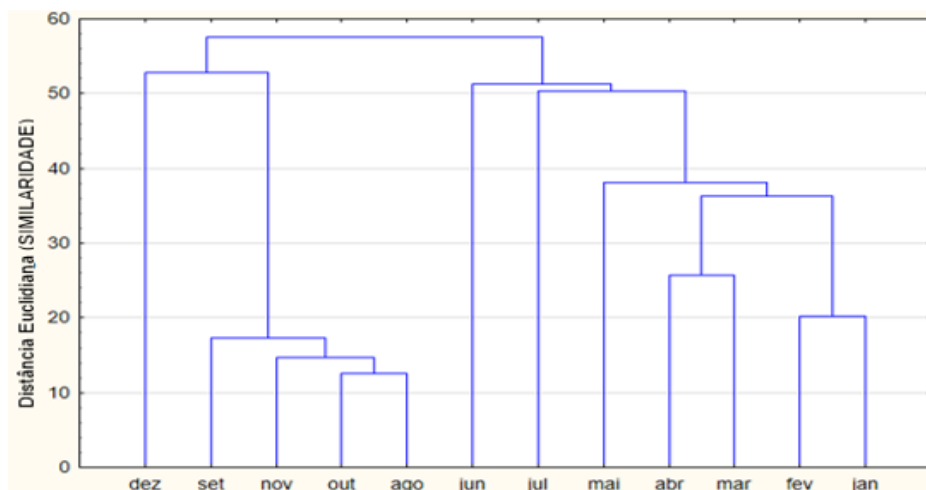


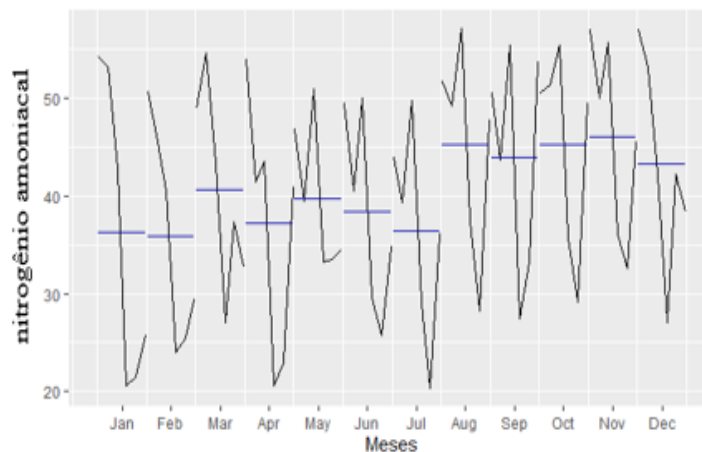
Figura 21 - Dendrograma referente à concentração média mensal de nitrogênio amoniacal.



A partir da clusterização dos dados do indicador nitrogênio amoniacal foram observados dois grupos distintos. O primeiro grupo é formado pelos anos de 1977 a 1981. Já no segundo grupo estão os anos de 1983 a 1989. A distinção entre os grupos pode ser atribuída às expansões da rede.

A análise do comportamento médio da amônia ao longo dos meses permite verificar o aumento de concentração entre os meses de agosto a dezembro, fato, possivelmente, relacionado aos menores índices de precipitação durante o período. O gráfico da variação da concentração média mensal de nitrogênio amoniacal entre os anos de 1977 e 1989, apresentado na Figura 22, confirma o comportamento médio da amônia ao longo dos meses.

Figura 22 - Variação da concentração média mensal de nitrogênio amoniacal entre 1977 e 1989.

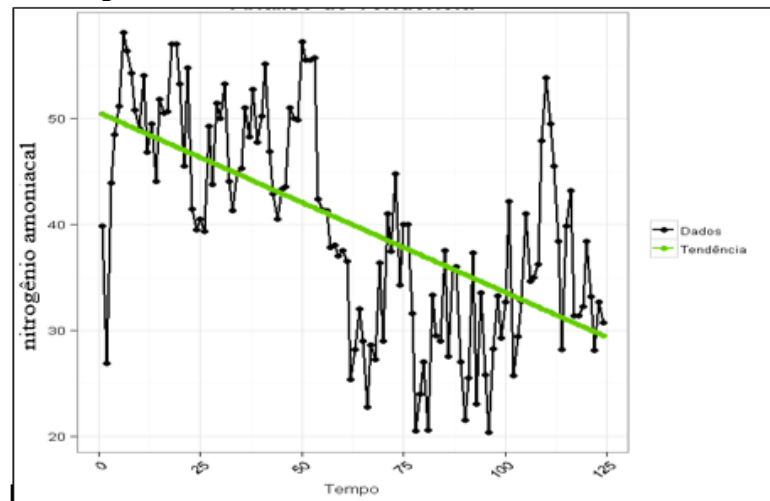


A aplicação do teste de Mann-Kendall (Tabela 10) analisa o p-valor e nos permite verificar que o p-valor calculado é menor que o nível de significância ( $\alpha = 0,05$ ), dessa forma, a hipótese alternativa é aceita. Além disso, o valor negativo para ZMK indica a existência de tendência significância decrescente da concentração de nitrogênio amoniacal, o que pode ser comprovado pela curva de regressão linear da Figura 23.

Tabela 10 - Resultado do teste de Mann-Kendall para a amônia.

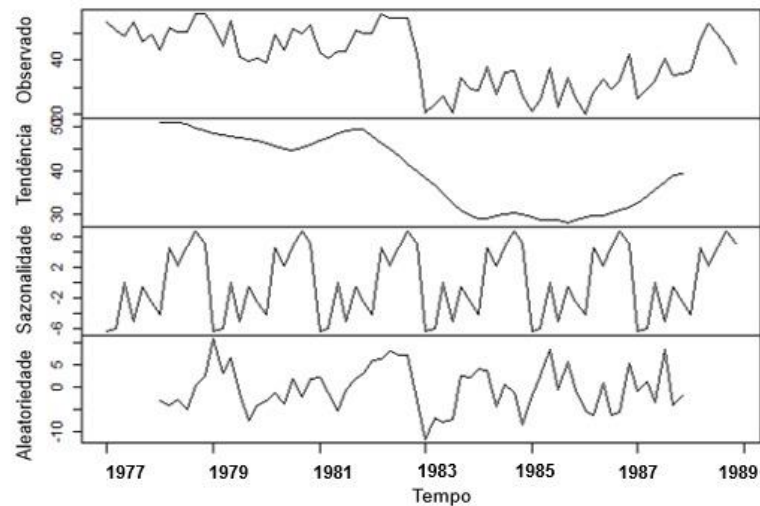
ZMK	-0,425674558
P-Valor	2,52E-12
Tamanho da Amostra	124
Hipótese Nula	Não há tendência
Hipótese Alternativa	Há tendência

Figura 23 - Gráfico da análise de tendência da amônia.



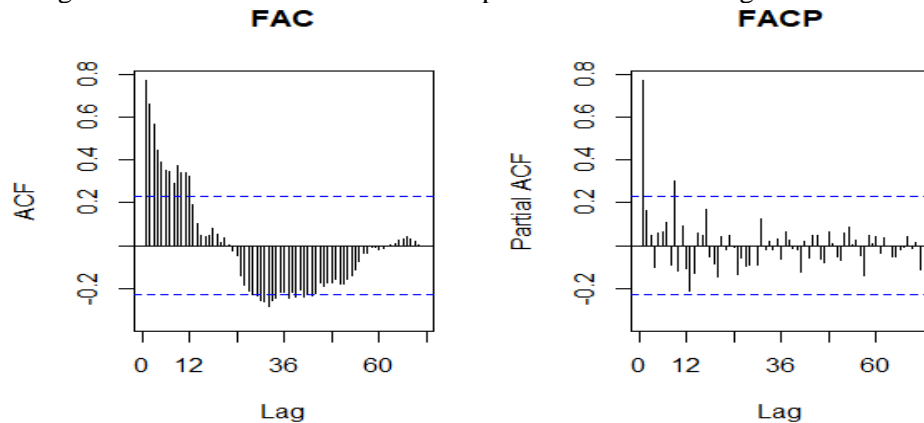
Ao analisarmos os gráficos da decomposição da série temporal do indicador nitrogênio amoniacoal na Figura 24, é possível notarmos um comportamento similar ao da DBO, tanto com relação a tendência quanto como a sazonalidade, apresentando uma diminuição dos valores no segundo semestre de 1981 e uma elevação no segundo semestre de cada ano, indicando a possível relação entre elas e a diluição do efluente iniciada em 1981.

Figura 24 - Gráficos da decomposição para os dados de amônia.



Dando sequência a observação dos gráficos FAC e FACP da série temporal do indicador nitrogênio amoniacoal, ilustrado na Figura 25, e sabendo pelo teste Dickey-Fuller aumentado que a série só se torna estacionária após a sua primeira diferenciação para o p-valor igual a 0,01, podemos chegar aos modelos SARIMA (1,1,1)(1,1,1)<sub>12</sub>, e SARIMA (0,1,1)(1,0,0)<sub>12</sub>, sendo esse último sugerido pela ferramenta *auto.arima* do *software* R.

Figura 25 - Gráficos do FAC e FACP para os dados de nitrogênio amoniacoal.



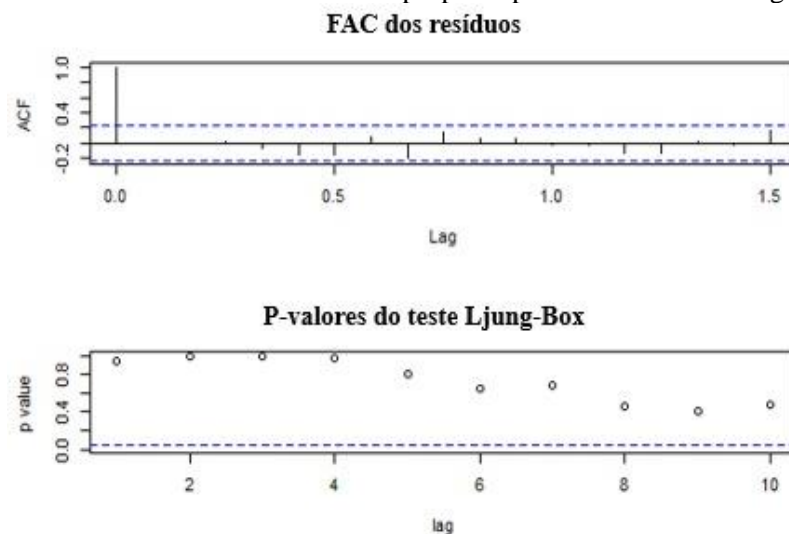
A partir dos dados expostos na Tabela 11 é possível identificar os resultados da comparação realizada entre os modelos obtidos para nitrogênio amoniacoal.

Tabela 11- Resultado dos modelos propostos para dados de nitrogênio amoniacoal.

MODELO	AIC	MAE	MAPE
$(0,1,1)(1,0,0)_{12}$	475	4,89	13,77
$(1,1,1)(1,1,1)_{12}$	405	3,59	10,49

Comparando os resultados do AIC, MAE e MAPE para cada modelo chegamos a um melhor ajuste no modelo SARIMA  $(1,1,1)(1,1,1)_{12}$ . Por fim, foi realizado um levantamento dos resíduos do modelo escolhido, podendo ser verificado na Figura 26, bem como a previsão com o modelo proposto para o nitrogênio amoniacoal dos seis meses seguintes (Figura 27).

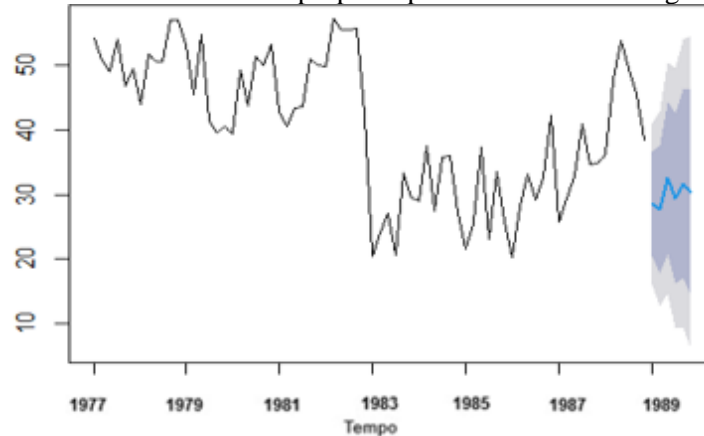
Figura 26 – Análise dos resíduos do modelo proposto para os dados de nitrogênio amoniacoal.



É notório que o mesmo não apresenta autocorrelação significativa, além dos p-valores

do teste Ljung Box não serem significativos. Com isso, conclui-se que os resíduos não interferem na modelagem da série. Na previsão para os seis meses seguintes utilizando o modelo proposto para o indicador nitrogênio amoniacal é possível verificar estimativas razoáveis.

Figura 27 – Previsão com o modelo proposto para os dados de nitrogênio amoniacal



Chen *et al.* (2018) analisaram a carga poluidora da demanda química de oxigênio (DQO), do nitrogênio amoniacal ( $\text{NH}_3\text{-N}$ ) e do FT sob diferentes condições de chuva e verificaram que o valor máximo dos poluentes apareceu nos estágios inicial e médio da chuva, enquanto diminuiu gradualmente com o aumento das chuvas.

### 5.1.3 Comportamento dos coliformes termotolerantes

A partir do resumo estatístico descritivo de coliformes termotolerantes, apresentado na Tabela 12, verifica-se que 75% dos dados estão entre  $1,19 \times 10^7$  e  $3,64 \times 10^7$  CF/100mL e apresentam desvio padrão igual a  $1,63 \times 10^7$ , indicando que os valores amostrais estão distantes da média.

Tabela 12 - Resumo estatístico descritivo dos dados dos coliformes termotolerantes.

1º quartil	$1,19 \times 10^7$
Média	$2,60 \times 10^7$
Mediana	$2,55 \times 10^7$
3º quartil	$3,64 \times 10^7$
Desvio Padrão	$1,63 \times 10^7$
Tamanho da Amostra	123

O dendrograma dos dados referente aos CT ao longo dos anos de 1977 a 1989, bem como das médias mensais de coliformes termotolerantes pode ser verificado a partir das Figuras

28 e 29, respectivamente.

Figura 28 - Dendrograma dos dados referentes aos coliformes termotolerantes ao longo dos anos de 1977 a 1989.

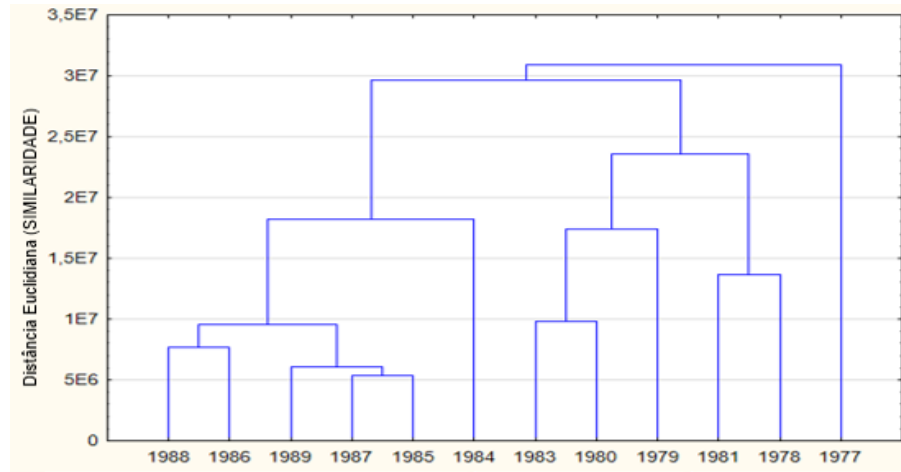
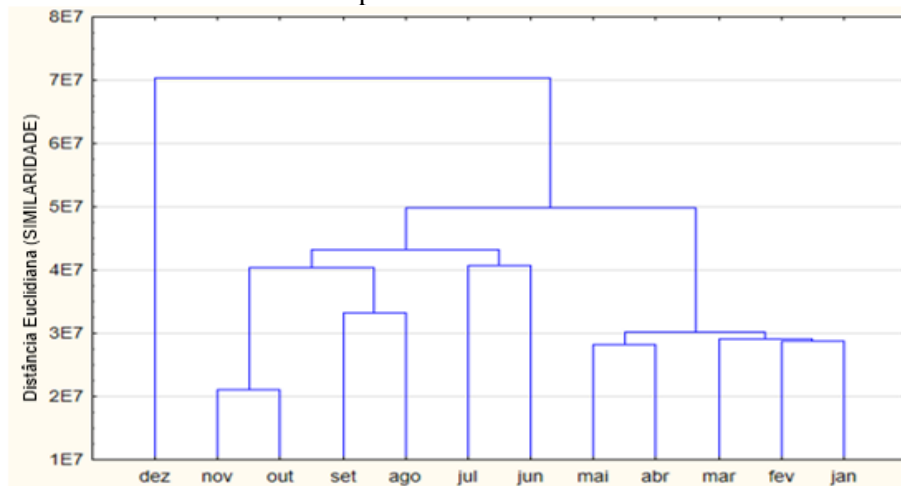


Figura 29 - Dendrograma referente às médias mensais de coliformes termotolerantes ao longo do período estudado.

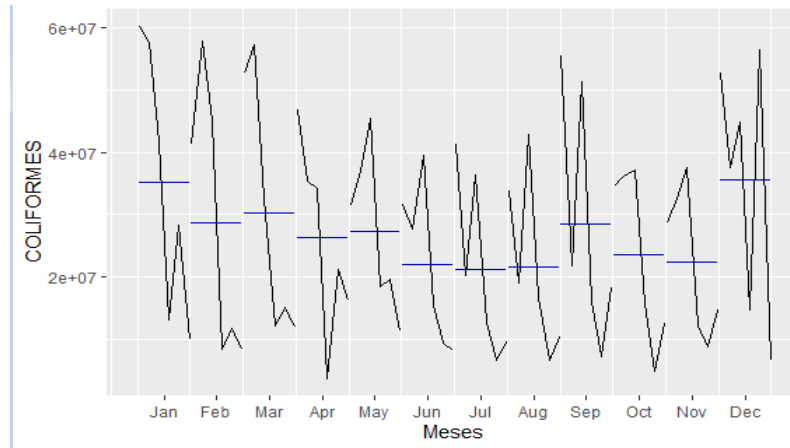


Observa-se a formação de dois grupos distintos relacionados ao comportamento de coliformes termotolerantes ao longo dos anos da série temporal estudada. A distinção entre o primeiro grupo que varia entre os anos de 1978 a 1983 e o segundo que engloba os anos entre 1984 e 1989 pode ser devida à expansão da rede coletora ao longo do período, uma vez que novas contribuições são acrescentadas ao sistema de esgotamento sanitário justamente entre os anos de 1984 e 1985.

O dendrograma referente às médias mensais de coliformes termotolerantes, indica a tendência de maior concentração em período de estiagem, de modo análogo ao que acontece com a DBO e o nitrogênio amoniacal. Na Figura 30 está ilustrado a variação mensal da concentração de Unidades Formadoras de Colônia (UFC) de coliformes termotolerantes com

base nos dados da série temporal estudada, comprovando o que foi verificado no dendrograma anterior.

Figura 30 - Variação das médias mensais de unidades formadoras de colônia ao longo do período estudado.

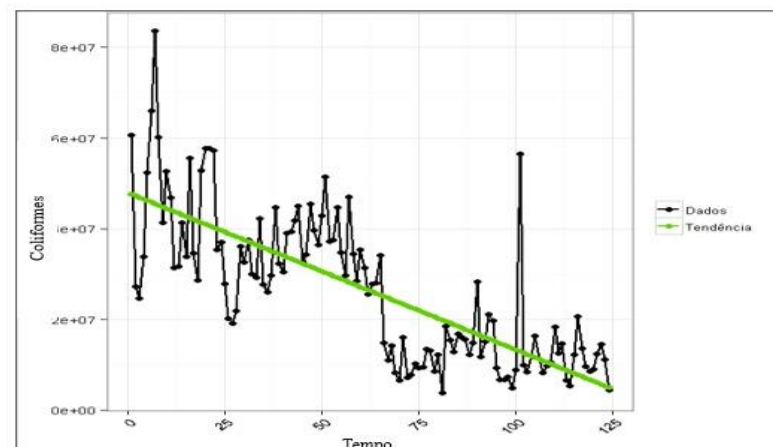


De acordo com o teste de Mann-Kendall (Tabela 13), ao analisarmos que o p-valor calculado é menor que o nível de significância ( $\alpha = 0,05$ ), a hipótese alternativa é aceita. Além disso, uma vez que o valor do ZMK é negativo verifica-se a existência de tendência significância decrescente da concentração de coliformes termotolerantes, o que pode ser constatado a partir do gráfico apresentado na Figura 31

Tabela 13 - Resultado do teste de Mann-Kendall para coliformes termotolerantes.

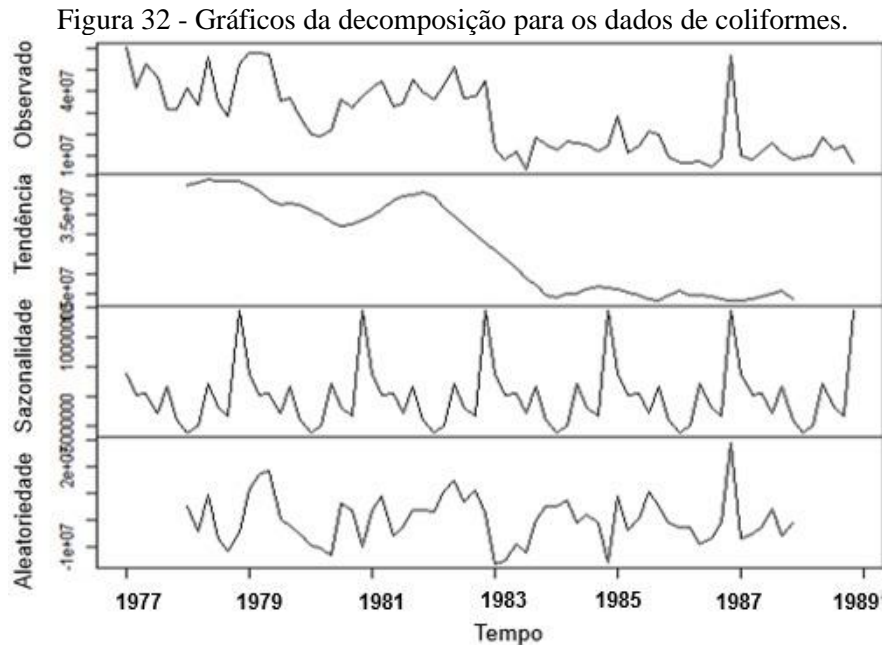
Estatística	-0,531383216
P-Valor	2,19E-18
Tamanho da Amostra	124
Hipótese Nula	Não há tendência
Hipótese Alternativa	Há tendência

Figura 31 - Gráfico da análise de tendência dos coliformes.



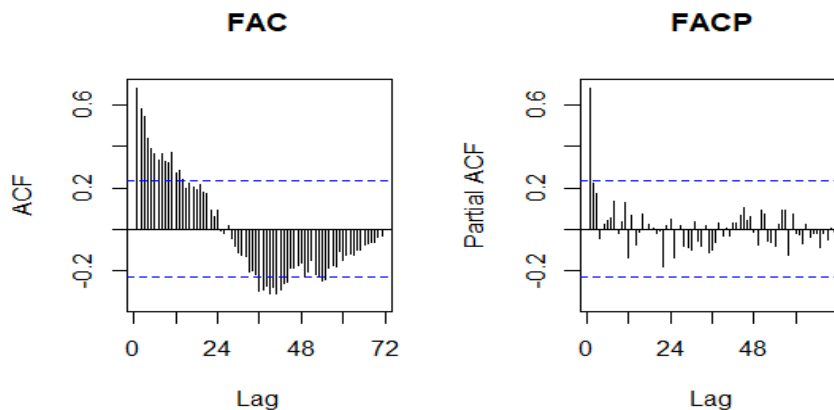


É notório que ao se observar a decomposição da série (Figura 32) a tendência também decai a partir do segundo semestre de 1981, assim como os indicadores de qualidade anteriormente citados. A sazonalidade indica um pico no aumento das UFC no quarto trimestre do ano.



Observando os gráficos FAC e FACP da série temporal do indicador coliformes termotolerantes, ilustrado na Figura 33, e sabendo pelo teste Dickey-Fuller aumentado que a série já é estacionária, com p-valor igual a 0,04, podemos chegar a dois modelos: SARIMA (1,0,1)(0,1,2)<sub>12</sub>, e o modelo ARIMA (1,1,1)<sub>12</sub>, sendo esse último sugerido pela ferramenta *auto.arima* do *software* R.

Figura 33 - Gráficos do FAC e FACP para os dados de coliformes.



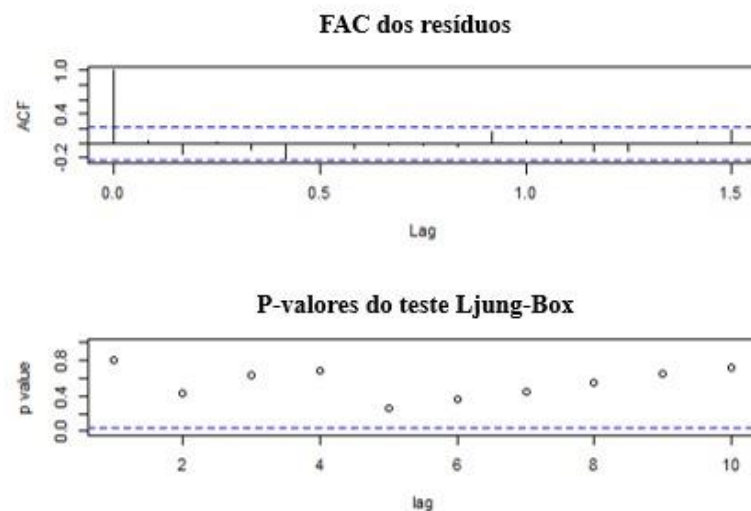
A partir dos dados expostos na Tabela 14 é possível identificar os resultados da comparação realizada entre os modelos obtidos os dados dos coliformes termotolerantes.

Tabela 14- Resultado dos modelos propostos para dados de coliformes.

MODELO	AIC	MAE	MAPE
$(1,1,1)_{12}$	2510,42	7880069	45,82
$(1,0,1)(0,1,2)_{12}$	2142,13	6944029	42,07

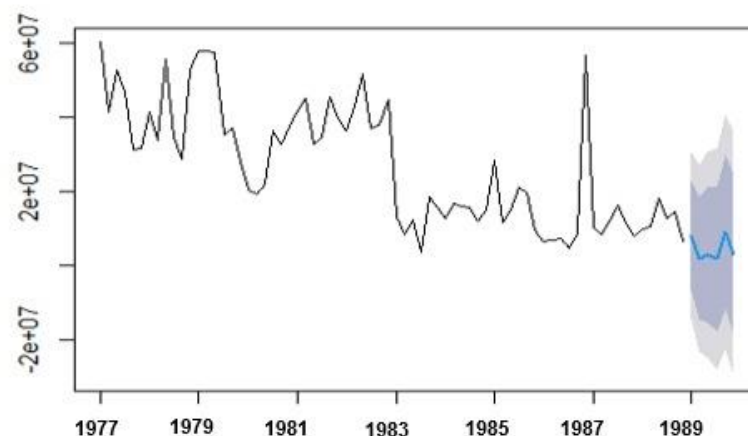
Comparando os resultados do AIC, MAE e MAPE de cada modelo, verifica-se que há um melhor ajuste no modelo SARIMA  $(1,0,1)(0,1,2)_{12}$ . Por fim, foi realizado a análise dos resíduos do modelo selecionado, representado pela Figura 34, bem como a previsão com o modelo dos dados de coliformes para os seis meses seguintes, podendo ser observado pelo gráfico exposto na Figura 35.

Figura 34 – Análise dos resíduos do modelo proposto para os dados de coliformes.



Observando os resíduos do modelo escolhido é notório que o mesmo não apresenta autocorrelação significativa, reforçado pelo teste estatístico de Ljung Box, com isso, definimos que os resíduos não interferem na modelagem da série.

Figura 35 – Previsão com o modelo proposto para os dados de coliformes.



É possível verificar que as previsões geradas para os seis meses seguintes utilizando o modelo SARIMA (1,0,1)(0,1,2)<sub>12</sub> fornecem estimativas razoáveis, apresentando decaimento dos dados.

#### ***5.1.4 Comportamento da alcalinidade total***

Pelo resumo estatístico descritivo da alcalinidade total, exposto na Tabela 15, pode ser verificado que 75% dos dados estão entre 260,6 e 333,2mg CaCO<sub>3</sub>.L<sup>-1</sup>, e apresentam desvio padrão de 50,1 mg CaCO<sub>3</sub>.L<sup>-1</sup>, indicando que os valores amostrais estão mais dispersos em torno da média.

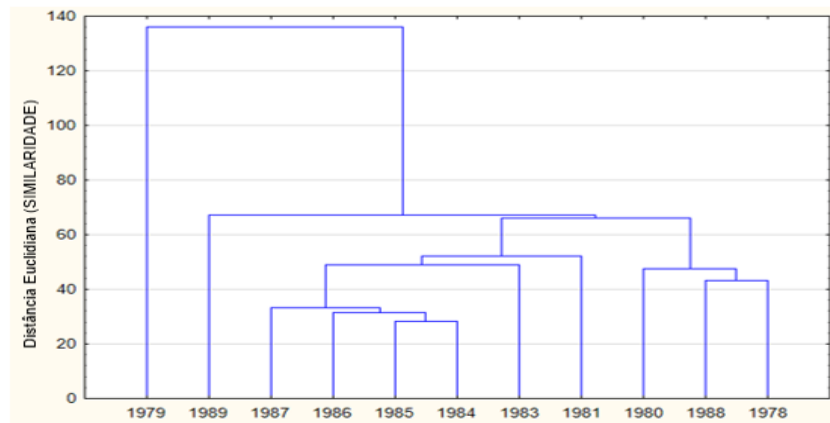
Tabela 15 - Resumo estatístico descritivo dos dados da alcalinidade.

1° quartil	260,6
Média	296,2
Mediana	286
3° quartil	333,2
Desvio Padrão	50,1
Tamanho da Amostra	112

O dendrograma ilustrado na Figura 36 é o resultado da análise de agrupamento aplicada aos dados de alcalinidade total e mostra a formação de dois grupos de anos que, mesmo sendo bem próximos, ainda se comportam de forma diferente.

As diferenças entre o primeiro grupo formado pelo intervalo entre 1981 a 1987 e o segundo (1980, 1988, 1978) podem ser atribuídas a motivos distintos ao aumento do sistema de esgotamento sanitário já que a formação de grupos não coincide com os anos de alteração no sistema.

Figura 36 - Dendrograma de agrupamento dos dados de alcalinidade total entre os anos de 1977 e 1989.



Nas Figuras 37 e 38 estão apresentadas a variação da alcalinidade total ao longo dos meses, podendo ser verificado o aumento dessa propriedade agregada entre agosto e dezembro, período de menor precipitação pluviométrica.

Figura 37 - Dendrograma referente à alcalinidade média mensal, entre os anos de 1977 e 1989.

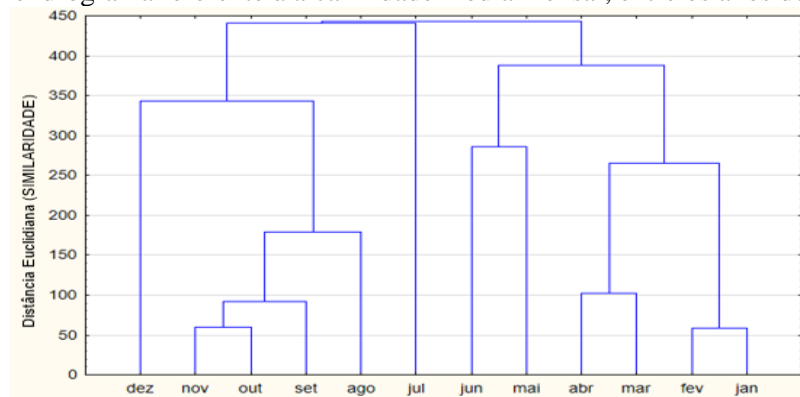
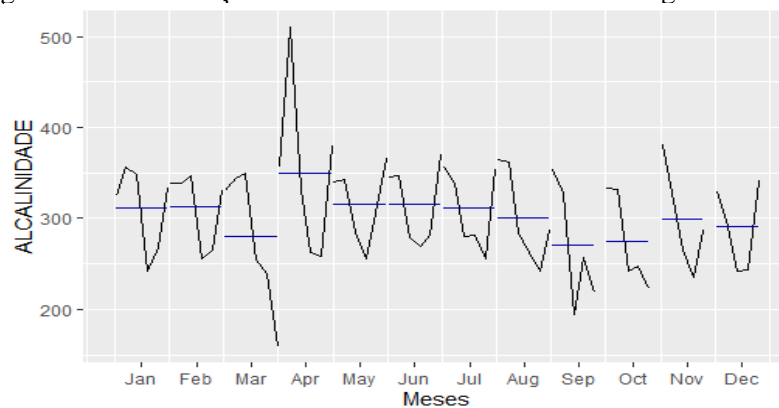


Figura 38 - Distribuição da alcalinidade total média ao longo dos meses.



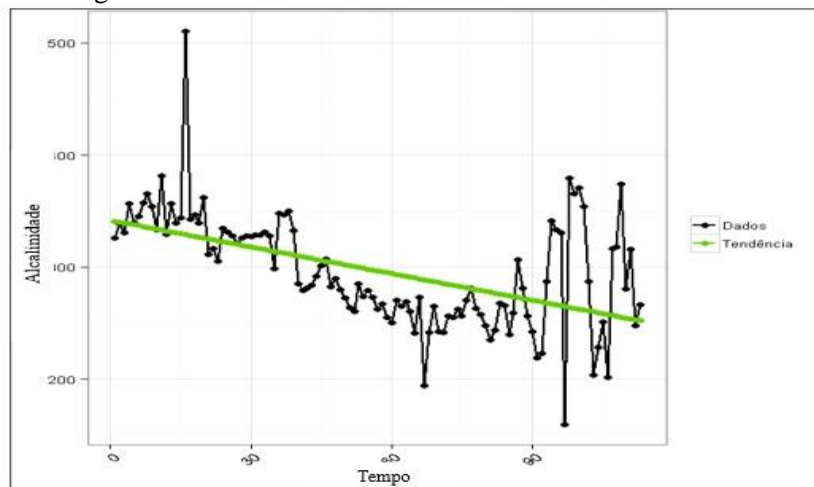
No teste de Mann-Kendall (Tabela 16) ao analisar o p-valor ( $\alpha = 0,05$ ) aceita-se a hipótese alternativa, em que permite verificar a existência de tendência significância decrescente da concentração da alcalinidade total, como pode ser verificado a partir da curva

de regressão linear apresentada na Figura 39.

Tabela 16 - Resultado do teste de Mann-Kendall para os dados de alcalinidade.

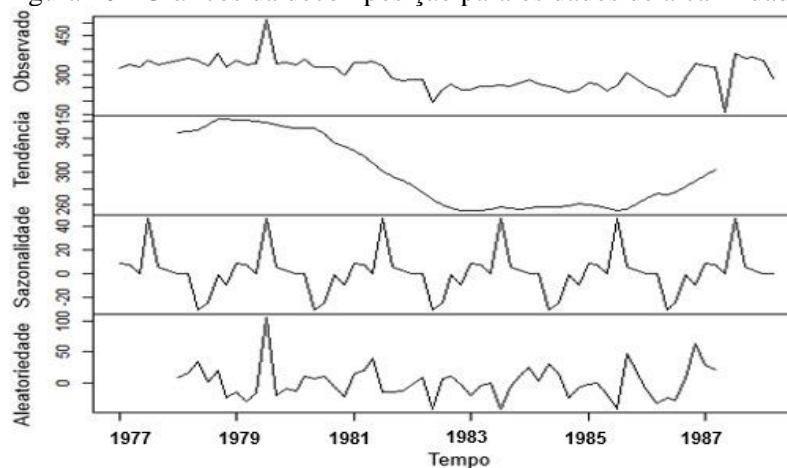
Estatística	-0,43361336
P-Valor	1,07E-11
Tamanho da Amostra	113
Hipótese Nula	Não há tendência
Hipótese Alternativa	Há tendência

Figura 39 - Gráfico da análise de tendência da alcalinidade.



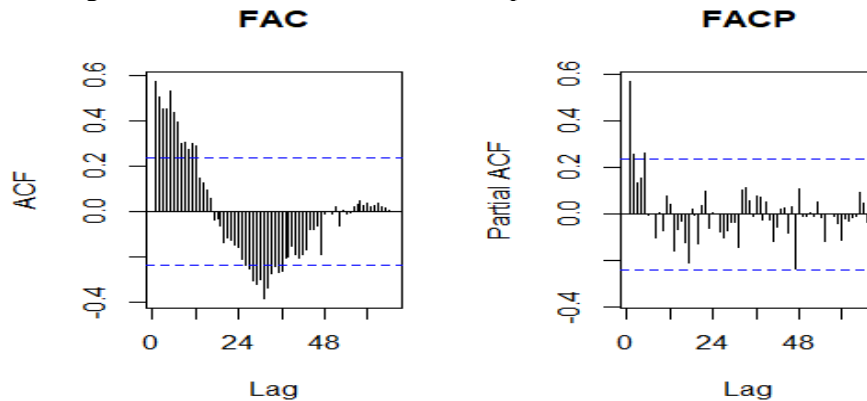
Verificando a decomposição da série temporal do indicador alcalinidade (Figura 40), observamos que, de forma mais suave ao longo dos anos, há um declínio na tendência, contudo, diferente do que ocorreu nos demais indicadores, a partir do ano de 1979 iniciou-se um decaimento dos dados da variável analisada, indicando possível diluição da água ou elevação da acidez do efluente, já a sazonalidade apresenta picos de alcalinidade no primeiro semestre do ano.

Figura 40 - Gráficos da decomposição para os dados de alcalinidade.



Observando os gráficos FAC e FACP da série temporal do indicador alcalinidade, ilustrado na Figura 41, e sabendo pelo teste Dickey-Fuller aumentado que a série se torna estacionária após a primeira diferenciação (p-valor = 0,01), podemos chegar a dois modelos: SARIMA (1,1,1)(0,1,0)<sub>12</sub>, e o modelo sugerido pela ferramenta auto.arima do *software* R, SARIMA(0,1,1)(1,0,0)<sub>12</sub>.

Figura 41 - Gráficos do FAC e FACP para os dados de alcalinidade.



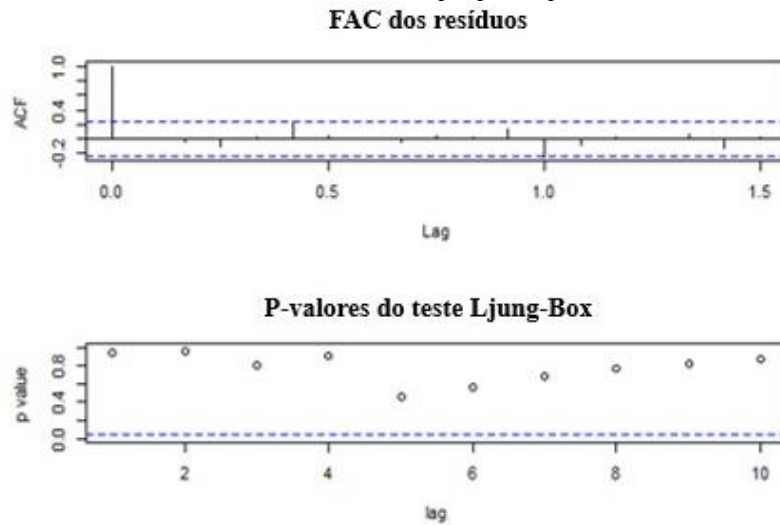
Comparando os resultados do AIC, MAE e MAPE de cada modelo, apresentado na Tabela 17, pode-se verificar que mesmo com MAE e MAPE um pouco mais elevado escolheu-se o modelo SARIMA (1,1,1)(0,1,0)<sub>12</sub> como modelo que melhor se ajusta ao valor real uma vez que o mesmo apresentou menos AIC.

Tabela 17- Resultado dos modelos propostos para dados de alcalinidade.

MODELO	AIC	MAE	MAPE
(0,1,1)(1,0,0) <sub>12</sub>	699.17	28.63	9.9
(1,1,1)(0,1,0) <sub>12</sub>	595.79	29.06	10.17

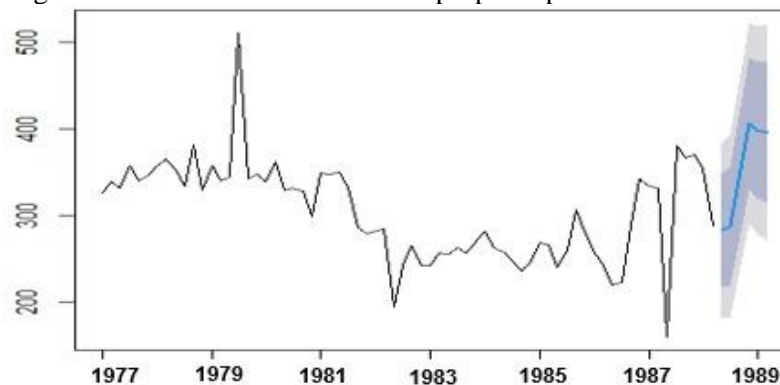
Observando os resíduos do modelo SARIMA (1,1,1)(0,1,0)<sub>12</sub> (Figura 42), é notório que o mesmo não apresenta autocorrelação significativa, reforçado pelo teste estatístico de Ljung-Box. Com isso definimos que os resíduos não interferem na modelagem da série.

Figura 42 – Análise dos resíduos do modelo proposto para os dados de alcalinidade.



Na Figura 43 temos a série temporal do indicador alcalinidade, e em destaque, as previsões para os 6 meses seguintes, seguindo o modelo SARIMA (1,1,1)(0,1,0)<sub>12</sub>, indicando o aumento dos dados.

Figura 43 – Previsão com o modelo proposto para os dados de alcalinidade.



### 5.1.5 Comportamento dos Sólidos Suspensos Totais

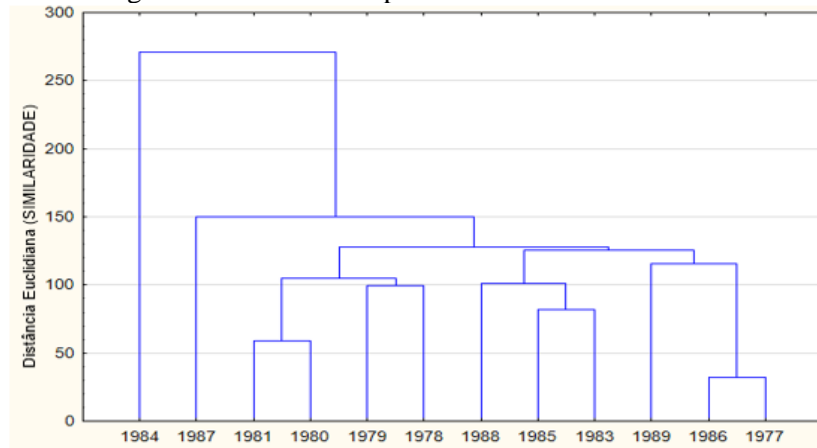
Com o resumo estatístico descritivo dos Sólidos Suspensos Totais, descrito na Tabela 18, verifica-se que 75% dos dados estão entre 240,2 e 372,3 mg.L<sup>-1</sup>, apresentando um desvio padrão de 99,2 mg.L<sup>-1</sup>, demonstrando uma alta dispersão dos dados em torno da média.

Tabela 18 - Resumo estatístico descritivo dos dados de sólidos suspensos totais.

1° quartil	240,2
Média	302,3
Mediana	307,8
3° quartil	372,3
Desvio Padrão	99,2
Tamanho da Amostra	125

Na clusterização dos dados para indicador Sólidos Suspensos Totais, no período de 1977 a 1989, é possível perceber a formação de 3 grupos distintos, como pode ser verificado pelo dendrograma apresentado na Figura 44.

Figura 44 - Dendrograma dos sólidos suspensos totais referentes aos anos de 1977 a 1989



O primeiro grupo apresenta os anos de 1978 a 1981, o segundo grupo os anos de 1983, 1985 e 1988 e o terceiro grupo é formado pelos anos 1977, 1986 e 1989. A distinção entre os grupos de anos não pode ser atribuída a alterações na malha de transporte desses efluentes já que a formação de grupos não coincide com os anos de alteração no sistema.

Analisando o comportamento médio dos sólidos suspensos totais ao longo dos meses, apresentados nas Figuras 45 e 46, verifica-se uma elevação dos valores no período do inverno que na região varia de julho a setembro, ocorrendo, como previsto, o aumento dos sólidos suspensos devido ao aumento das chuvas e o carreamento de materiais para a rede coletora.

Figura 45 - Dendrograma dos sólidos suspensos totais referentes aos meses, ao longo de 1977 a 1989

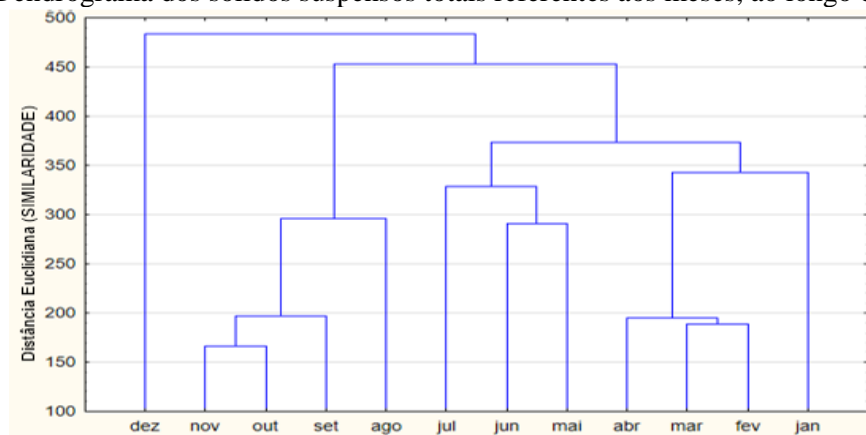
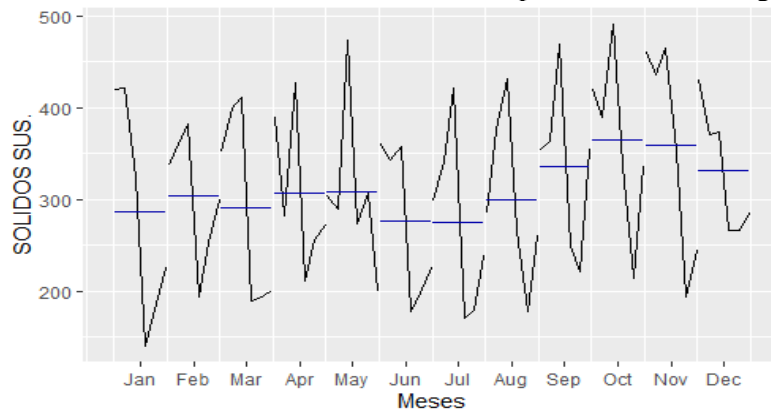




Figura 46 - Gráfico dos dados médio mensais de sólidos suspensos totais, ao longo de 1977 a 1989.



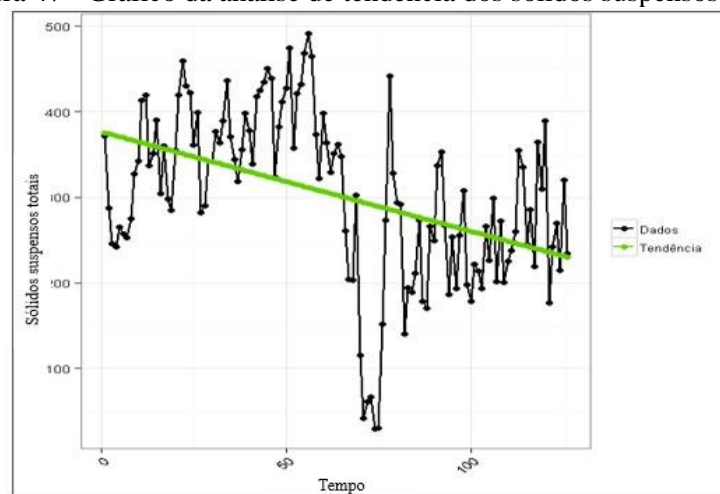
No teste de Mann-Kendall (Tabela 19) verifica-se que o p-valor ( $\alpha = 0,05$ ) é favorável a hipótese alternativa, em que considera a existência de tendência significativa na série.

Tabela 19 - Resultado do teste de Mann-Kendall para os dados de sólidos suspensos totais.

ZMK	-0,268969476
P-Valor	8,03E-06
Tamanho da Amostra	126
Hipótese Nula	Não há tendência
Hipótese Alternativa	Há tendência

A análise do ZMK negativo, assim como a curva de regressão linear apresentada na Figura 47 indicam uma tendência de diminuição dos Sólidos Suspensos Totais com o passar dos anos.

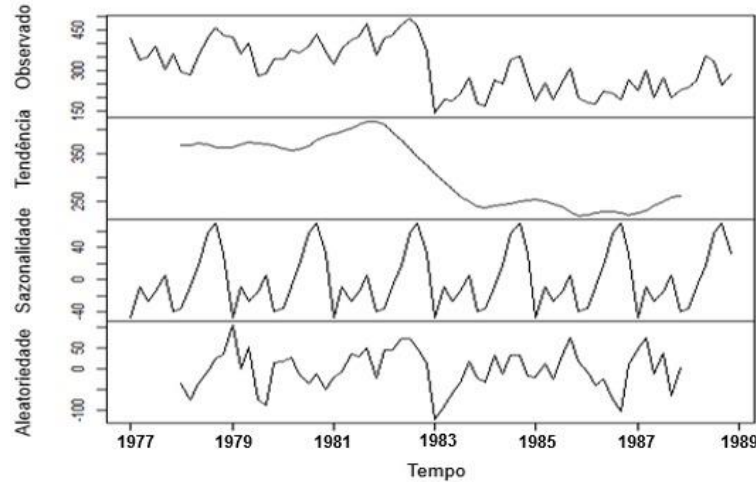
Figura 47 - Gráfico da análise de tendência dos sólidos suspensos totais.



A análise da decomposição da série temporal do indicador sólidos suspensos totais, ilustrada na Figura 48, mostra uma tendência que permanece constante até o segundo semestre

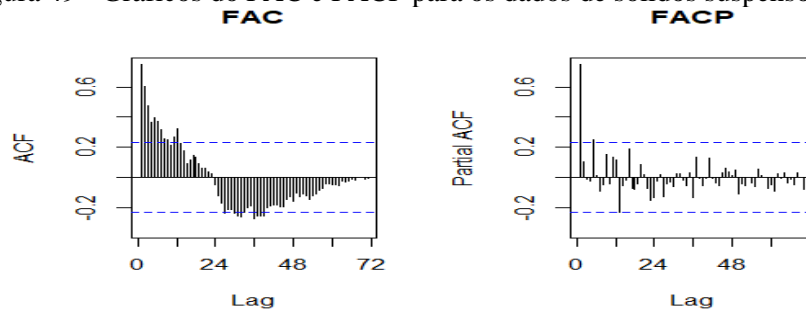
de 1980 e em seguida decai até o segundo semestre de 1981 onde volta a ficar constante novamente, indicando possível diluição do efluente nesse período específico, já a sazonalidade apresenta picos de concentração entre o terceiro e o quarto trimestre de cada ano.

Figura 48 - Gráficos da decomposição para os dados de sólidos suspensos totais.



Verificando os gráficos FAC e FACP da série temporal do indicador sólidos suspensos totais (Figura 49), e sabendo pelo teste Dickey-Fuller aumentado que a série se torna estacionária após a primeira diferenciação ( $p$ -valor = 0,01), podemos chegar a dois modelos: SARIMA (1,1,1)(0,1,0)<sub>12</sub>, e o modelo SARIMA(1,1,1)(1,0,0)<sub>12</sub> sugerido pela ferramenta auto.arima do *software* R.

Figura 49 - Gráficos do FAC e FACP para os dados de sólidos suspensos totais.



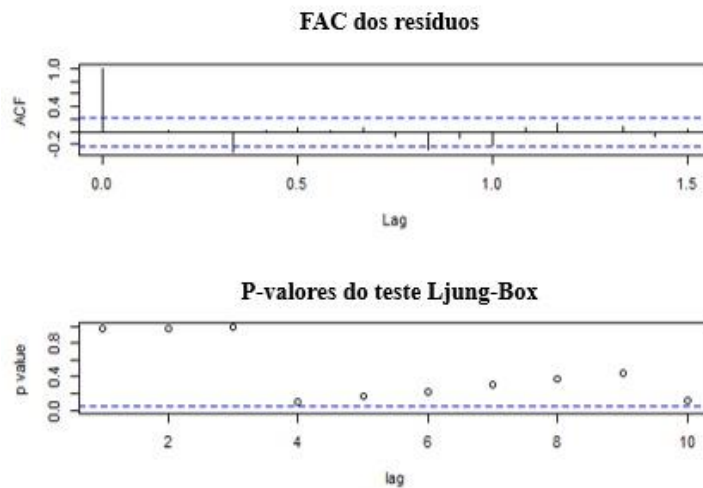
Comparando os resultados do AIC, MAE e MAPE de cada modelo, presentes na Tabela 20, verifica-se que modelo SARIMA (1,1,1)(0,1,0)<sub>12</sub> é o que melhor se ajusta aos dados.

Tabela 20 - Resultado dos modelos propostos para dados de sólidos suspensos totais.

MODELO	AIC	MAE	MAPE
(1,1,1)(0,1,0) <sub>12</sub>	676,6	45,38	17,81
(1,1,1)(1,0,0) <sub>12</sub>	782,7	44,49	17,09

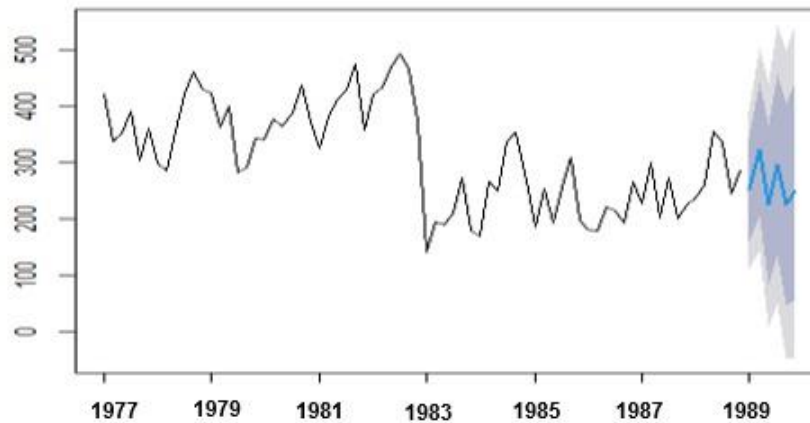
Observando os resíduos do modelo escolhido (Figura 50), é notório que o mesmo não apresenta autocorrelação significativa, além dos p-valores do teste Ljung Box não serem significativos, com isso definimos que os resíduos não interferem na modelagem da série.

Figura 50 – Análise dos resíduos do modelo proposto para os dados de sólidos suspensos totais.



Na Figura 51 temos a série temporal do indicador sólidos suspensos totais, e em destaque, as previsões para os próximos 6 meses seguindo o modelo SARIMA  $(1,1,1)(0,1,0)_{12}$ .

Figura 51 – Previsão com o modelo proposto para os dados de sólidos suspensos totais.





## 6 CONSIDERAÇÕES FINAIS

No decorrer deste trabalho foram observadas relações sazonais de indicadores de qualidade de águas residuárias como: DBO, nitrogênio amoniacal, coliformes termotolerantes, alcalinidade total e sólidos suspensos totais, relacionados com os fatores hidrológicos precipitação e temperatura.

Verificou-se que a análise da série temporal para as variáveis de qualidade de águas residuárias foi capaz de demonstrar tendências significativas ao longo do período estudado, já para as variáveis precipitação e temperatura, não foi possível obter o mesmo resultado, no entanto, foi possível identificar uma relação inversa entre precipitação e temperatura, possivelmente ocasionada devido a diminuição da temperatura em períodos de maiores índices de precipitação.

Já com relação a análise do comportamento da tendência dos indicadores de qualidade foi possível observar padrões de mudanças relacionados, tanto a diluição das águas residuárias durante o período chuvoso, quanto a expansão do sistema de esgotamento sanitário para integração de novas e diversificadas contribuições entre os anos de 1984 e 1985 que influenciou diretamente o comportamento dos indicadores que apresentam tratamento biológico, como DBO, nitrogênio amoniacal e coliformes termotolerantes.

Por fim, a análise de séries temporais pelo método Box-Jenkins nos permitiu modelar o comportamento dos dados de precipitação e temperatura, além dos indicadores de qualidade das águas residuárias, com modelos ditos satisfatórios após avaliação de erros e resíduos, também descrevendo efeito sazonal em todas essas variáveis, assim, dando ideia do comportamento futuro dessas variáveis.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. **Information theory and an extension of the maximum likelihood principle.** In B. N. Petrov & B. F. Csaki (Eds.), *Second International Symposium on Information Theory*, (pp. 267–281). Academiai Kiado: Budapest. 1973.
- ALBUQUERQUE R. C., OLINDA R. A., CUNHA A. L. X., TAVARES R. G., ARRUDA V. C. M., FILHO M. C. **Aplicação da metodologia Box & Jenkins em dados de precipitação na região do sertão do estado de Pernambuco – Brasil.** Revista GEAMA, Scientific Journal of Environmental Sciences and Biotechnology, 6 (1): 25-30, 2020.
- ALMEIDA, S. A. B. **Contribuição à aplicação de coeficientes de consumo em projetos de abastecimento de água e esgotamento sanitário em comunidades urbanas de baixa renda no nordeste do Brasil – Estudo de caso.** Dissertação (Mestrado em Engenharia Civil e Ambiental). CTRN, UFCG, 2007.
- ANA - **Agência Nacional das Águas.** Portal da Qualidade das Águas. Disponível em: [http://pnqa.ana.gov.br/rede/rede\\_avaliacao.aspx](http://pnqa.ana.gov.br/rede/rede_avaliacao.aspx). Acesso em: 01 de Abril de 2019.
- BARBOSA, E. C.; SÁFADI, T.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; SILVA, C. H. O.; MANULI, R. C. **Metodologia Box & Jenkins para previsão de temperatura média mensal da cidade de Bauru (SP).** Revista Bras. Bion, São Paulo, v. 33, n-1, p. 104 – 117, 2015.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis forecasting and control.** San Francisco: HoldenDay, 1976.
- BOYACIOGLU, H. **Investigation of Temporal Trends in Hydrochemical Quality of Surface Water in Western Turkey.** Bulletin of Environmental Contamination and Toxicology, n. 80, p. 469–474, 2008.
- CAGEPA - Companhia de Água e Esgotos da Paraíba. **Projeto Técnico Complementar do Sistema de Esgotamento Sanitário de Campina Grande-PB** (Relatório Parcial)- TECNOSAN ENGENHARIA S/A. Campina Grande-PB, Janeiro 1985.
- CAGEPA - Companhia de Água e Esgotos da Paraíba. **Projeto Técnico Complementar do Sistema de Esgotamento Sanitário de Campina Grande-PB Volume I – Memorial Descritivo-** TECNOSAN ENGENHARIA S/A. Campina Grande-PB, Julho 1985.
- CAGEPA - Companhia de Água e Esgotos da Paraíba. **SANESA - Saneamento de Campina Grande S/A Esgotos** (Relatório e Orçamentos). Escritório Saturnino de Brito, 1962. Campina Grande, PB, 1964.
- CAMELO, S. M; Coura, M. A. C; Rodrigues, A. C. L; Oliveira, R; Filho, F. C. C; Vidal, I. C. A. **Modelagem da qualidade da água em sistemas de macrodrenagem de bacias urbanas.** Revista Engenharia Sanitária e Ambiental. V. 25, n.6 p. 873-885. 2020.
- CAMPOS, J. R. **Tratamento de esgotos sanitários por processo anaeróbio e disposição controlada no solo.** Rio de Janeiro: ABES, 464p., 1999.
- CAVALCANTE, G. P.; LUCENA, D. B.; MOURA, M. O. **As chuvas na região agreste do nordeste brasileiro: variabilidade interanual.** Revista Equador, v. 8, n. 2, p.290-308. 2019.

CHATFIELD, C. **The analysis of time series - an introduction**. 5. ed. London: Chapman & Hall, 1996.

CHEN, X.; SHE, J.; LAI, C.; CHEN, L.; WANG, Y.; ZHONG, K.; CHEN, J.; WANG, Z. Analysis on Water Quality Characteristics of Typical Black and Stinking River in Chengdu City by SWMM. *Hydrology*, v. 6, n. 4, p. 100-106, 2018.

DANTAS, L. G.; SANTOS, C. A. C.; OLINDA, R. A. **Tendências anuais e sazonais nos extremos de temperatura do ar e precipitação em Campina Grande – PB**. *Revista Brasileira de Meteorologia*, v. 30, n. 4, 423 - 434, 2015.

DINIZ, J. M. T. Variabilidade da precipitação e do número de dias com chuvas em duas cidades distintas da Paraíba. *Holos*, v. 3, p. 171-180, 2013.

FARREL, P. J. STEWART, K. R. **Comprehensive Study Of Tests For Normality And Symmetry: Extending The Spiegelhalter Test**. *Journal of Statistical Computation and Simulation*. v. 76, n. 9, pp. 803–816, 2006.

FAVA, V. L. **Manual de econometria**. In: VASCONCELOS, M. A. S.; ALVES, D. São Paulo: Editora Atlas, 2000.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**, AAAI Press/The MIT Press, England, p.1-34. 1996.

FERRARI, A. L.; COLABONE, R. O.; LUCAS, L. A. ; LIMA, C. G.; VECCHIA, F. A. S. **Análise estatística comparativa das condições meteorológicas de nevoeiro, temperatura e umidade relativa do ar em Pirassununga - SP**. *Revista da UNIFA (Impresso)*, v. 26, p. 59-64, 2013.

GEROLIN, P. H. **MINERAÇÃO DE DADOS NA GESTÃO DE RECURSOS HÍDRICOS SUBTERRÂNEOS: Estudo de caso** (Monografia) Especialização Lato Sensu em Gerenciamento de Banco de Dados. Universidade Nove de Julho, São Paulo, 2018.

GOMES E. F. **Perdas de vazão e seus efeitos na operação do sistema de esgotamento sanitário de Campina Grande-PB**. 2013. 84f. (Dissertação) Mestrado em Engenharia Civil e Ambiental, Programa de Pós-Graduação em Engenharia Civil e Ambiental, Universidade Federal de Campina Grande - Campina Grande - Paraíba - Brasil, 2013.

HYNDMAN. (2020). **fpp2: Data for "Forecasting: Principles and Practice" (2nd Edition)**. R package version 2.4. Disponível em: <<https://cran.rproject.org/web/packages/fpp2/>>. Acesso em: 30 ago. 2021.

HYNDMAN. (2021). **forecast: Forecasting Functions for Time Series and Linear Models**. R package version 9.15. Disponível em: <<https://cran.r-project.org/web/packages/forecast/>>. Acesso em: 30 ago. 2021.

KENDALL, M.G. **Rank Correlation Methods**, 4th ed., Charles Griffin: London. 1975.

LAMOUNIER, W. M. **Tendência, ciclos e sazonalidade nos preços spot do café brasileiro na NYBOT**. *Gest. Prod.*, São Carlos, v. 14, n. 1, p. 13-23, jan.-abr. 2007

LINDEN, R. **Técnicas de Agrupamento**. *Revista de Sistemas de Informação da FSMA*, N.

4, p.18-36. 2009.

LOESCH, C. **Métodos estatísticos multivariados**, São Paulo; Saraiva, 288p. 2012.

LUIZ, G. C.; CARDOSO, H. C.; RIBEIRO, L. L. **Aplicação do teste sazonal de mann kendall na análise de tendência da temperatura e umidade relativa do ar goiâniago: série histórica 1961 a 2008**. Revista Geonorte, Edição Especial 2, V.1, N.5, p.414 – 427, 2012.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. **Forecasting: Methods and Applications**, 3rd ed.; JohnWiley & Sons: New York, NY, USA, 1998.

MANN, H.B. **Non-parametric test against trend**. *Econométrica*, vol. 13, n. 3, p. 245-259, 1945.

MCLEOD A. (2013). **FitAR: Subset AR Model Fitting**. R package version 1.94. Disponível em: <<https://cran.r-project.org/web/packages/FitAR/>>. Acesso em: 30 ago. 2021.

MOITA NETO, J. M. **Estatística multivariada – Uma visão didática-metodológica**. Crítica. Revista de Filosofia e Ensino, v.1, n.1, p. 1-1, 2004.

MORETTIN P. A.; TOLOI C. M. C. **Análise de séries temporais**. 2ª ed. São Paulo: Egard Blucher, 2006.

MORETTIN, P. A.; TOLOI, C. M. C. **Previsão de series temporais**, São Paulo, Atual. 436 p, 1987.

OLIVEIRA, J. B. **Avaliação das concentrações de nutrientes (N e P) na lagoa anaeróbia profunda da pedra n° 7 do polo de tratamento de esgotos do baixo Paraíba**. (Dissertação), Programa de Pós-Graduação em Engenharia Civil e Ambiental, Universidade Federal de Campina Grande - Paraíba, 2006.

PFAFF B. (2016). **URCA: Unit Root and Cointegration Tests for Time Series Data**. R package version 1.3.0. Disponível em: <<https://cran.rproject.org/web/packages/urca/>>. Acesso em: 30 ago. 2021.

QUEIROZ, M. M. F.; IOST, C.; GOMES, S. D.; VILAS BOAS, M. A. **Influência do uso do solo na qualidade da água de uma microbacia hidrográfica rural**. Revista Verde de Agroecologia e Desenvolvimento Sustentável, Mossoró, v. 5, n. 4, p. 200-210, 2010.

RITZMAN, L. P.; MALHORTA, M.; KAJEWSKI, L. **Administração de Produção e Operações**. 8 ed. São Paulo: Prentice Hall, 2008.

SABINO, C.V.S.; LAGE, L.V.; NORONHA, C.V. **Variação sazonal e temporal da qualidade das águas em um ponto do Córrego Gameleiras usando técnicas quimiométricas robustas**. Revista Engenharia Sanitária e Ambiental, v.22, N.5, p.969-983. 2017.

SALVIANO. M. F.; GROppo. J. D.; PELLEGRINO. G. Q. **Análise de Tendências em Dados de Precipitação e Temperatura no Brasil**. Revista Brasileira de Meteorologia, v.31, n. 1, 64-73, 2016.



SAS – **Analytic Software e Solutions**. Disponível em:

[https://www.sas.com/pt\\_br/insights/analytics/mineracao-de-dados.html#dmusers](https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html#dmusers). Acesso em: 07 de maio de 2019.

SILVA, M. A.; ARAÚJO, R. R. **Análise temporal da qualidade da água no córrego Limoeiro e no rio Pirapozinho no estado de São Paulo**. Revista Formatação, v. 1; n. 24, p. 182-203. 2017.

SOBRAL T. E. L.; BARRETO G. **Utilização dos critérios de informação na seleção de modelos de regressão linear**. Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, v. 4, N. 1, 2016

SOUZA, G. P.; SAMOHYL, R. W.; MIRANDA, R. G. **Métodos simplificados de previsão empresarial**. Rio de Janeiro: Ciência Moderna, 2008.

TRAPLETTI A. (2020). **tseries: Time Series Analysis and Computational Finance**. R package version 0.10-48. Disponível em: <<https://cran.rproject.org/web/packages/tseries/>>. Acesso em: 30 ago. 2021.

TRINDADE, A. L. **Aplicação de técnicas estatísticas para avaliação de dados de monitoramento de qualidade das águas superficiais da porção mineira da Bacia do Rio São Francisco**. Dissertação de mestrado da Universidade Federal de Minas Gerais, Belo Horizonte MG, 2013.

TRINDADE, A. L. C.; ALMEIDA, K. C. B.; BARBOSA, P. E.; OLIVEIRA, S. M. A. C. **Tendências temporais e espaciais da qualidade das águas superficiais da sub-bacia do Rio das Velhas, estado de Minas Gerais**. Revista Engenharia Sanitária e Ambiental. v. 22, n. 1, p. 13-24. 2017.

TUBINO, D. F. **Planejamento e Controle da Produção teoria e prática**. 3 ed. São Paulo: Atlas, 2017.

VON SPERLING, M. **Introdução à qualidade das águas e ao tratamento de esgotos**. 4. Ed. – Belo Horizonte: Editora UFMG, 472p. 2014.

WERNER, N.; RIBEIRO, J. L. D. **Previsão de Demanda: uma aplicação dos modelos Box-Jenkins**. Revista Gestão & Produção, v. 10, n.1, p 47-67, abr. 2003.

WICKHAM H. (2019). **readxl: Read Excel Files**. R package version 1.3.1. Disponível em: <<https://cran.r-project.org/web/packages/readxl/>>. Acesso em: 30 ago. 2021.

WUNDERLIN, D. A; DÍAZ, M. D. P; AMÉ, M. V; PESCE, S. F; HUED, A. C; BISTONI, M. L. A.. **Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquíá River Basin Córdoba-Argentina**. Water. Research., v. 35, n. 12, p. 2881-2894, 2001.

YEVJEVICH, V. **Probability and statistics in hydrology**. Fort Collins: Water Resources Publication, 1972.

ZHANG, X; WANG, Q; LIU, Y; WU, J; YU, M. **Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong**. Environmental Monit Assess, n. 173, p. 17-27, 2011.