

## I-1141- MODELO DE PREVISÃO PARA VOLUME DE ÁGUA TRATADA DISPONIBILIZADO EM REDES DE DISTRIBUIÇÃO UTILIZANDO SERIES TEMPORAIS

**Tobias Ribeiro Barreto<sup>(1)</sup>**

Engenheira Civil, Especialista em Engenharia Sanitária pela PUC-BH, Mestre em Sistemas Aplicados a Engenharia e Gestão pelo Instituto Federal Fluminense (IFF). Especialista de dados do Grupo Águas do Brasil.

**Milton Erthal Júnior<sup>(2)</sup>**

Engenheiro Agrônomo pela Universidade Federal de Viçosa, Mestre e Doutor em Produção Vegetal pela Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF). Professor do Instituto Federal de Educação, Ciência e Tecnologia Fluminense, campus Campos-Guarus, onde atua nos cursos técnicos em Meio Ambiente, graduação em Engenharia Ambiental e Mestrado Profissional em Sistemas Aplicados a Engenharia e Gestão.

**Endereço<sup>(1)</sup>:** Av. Marquês do Paraná, 110 - Centro, Niterói - RJ, 24030-211. E-mail: [tobiasr.barreto@gmail.com](mailto:tobiasr.barreto@gmail.com)

### RESUMO

A previsão de volume de água tratada a ser distribuída é de grande importância no planejamento estratégico das empresas responsáveis de saneamento básico. O objetivo dessa pesquisa é propor um modelo matemático para prever o volume de água a ser distribuído no curto e no longo prazo, buscando a otimização do processo de abastecimento urbano de água. Um estudo de caso foi conduzido em uma empresa sediada no município de Campos dos Goytacazes, RJ, a partir de uma coleta de dados de 2 anos. O modelo estatístico ARIMA (Auto-Regressive Integrated Moving Average) foi empregado para prever demandas diárias de produção de água tratada. Desta forma, o modelo proposto possui o objetivo de melhorar as métricas para estimar a quantidade de água a ser distribuída no futuro, reduzindo custos econômicos e ambientais para a empresa e a região, sem afetar o serviço de distribuição de água para a população. Conclui-se que a metodologia proposta possibilita um planejamento estratégico mais apurado e alinhado com a realidade do município estudado, além de existir replicabilidade em outros modelos de negócio que possuem uma necessidade semelhante.

**PALAVRAS-CHAVE:** Saneamento, Volume Disponibilizado, Rede de Distribuição, Series Temporais, ARIMA.

### INTRODUÇÃO

A previsão do volume disponibilizado de água tem um alto significado financeiro e ambiental. Inúmeros componentes podem influenciar direta ou indiretamente a demanda de água. Incorporar precipitação, temperatura, demografia, chegar utilizar, estimar e controlar. As condições climáticas têm sido amplamente utilizadas como insumos de modelos multivariados mensuráveis para modelagem e previsão de consumo de água. A estimativa das futuras necessidades comuns de água é essencial para a organização de um sistema territorial de abastecimento de água.

Quando se trata de criar um modelo de previsão para o volume de água tratada disponibilizado para uma cidade, não existe um único modelo, existem diversas variáveis que irão influenciar a curva da série temporal em estudo. Informações como quantidade de habitantes, quantidade de ligações de água ativas, economias ativas, perfil do consumidor (residencial, comercial ou industrial), bem como a cultura de uso da água na região são exemplos de variáveis que criam essa diferença entre cada série temporal estudada. Essas especificidades criam aumentam o grau de dificuldade de criar um modelo previsão aceitável.

Esse tipo de modelo se torna uma ferramenta altamente estratégica para a empresa, uma vez que, com um modelo devidamente ajustado, é possível prever uma série de custos ao longo do ano, como, por exemplo, custos com produto químico, otimizando o estoque, custos relacionados a energia gasta na captação e tratamento da água, bem como custos envolvendo pessoas. Além de poder ser utilizado na gestão de perdas de água, tema de alta relevância no saneamento.

Campos dos Goytacazes é um município brasileiro no interior do estado do Rio de Janeiro, Região Sudeste do país. Está localizado na região Norte Fluminense do estado. De acordo como o Instituto Brasileiro de Geografia e Estatística (IBGE), possui uma população estimada de 514 643 habitantes, é a mais populosa cidade do interior

do estado, a sétima mais populosa do estado e o município com a maior extensão territorial do estado, ocupando uma área de 4 032,5 quilômetros quadrados.

Segundo o instituto Trata Brasil, realizado em 2022, Campos dos Goytacazes possui 97,9% de sua população atendida com água, sendo 2,1 cerca de 11 mil habitantes sem acesso à água potável.

Geralmente, as empresas de saneamento têm um setor de cadastro técnico responsável pela atualização e manutenção das plantas cadastrais, a partir de informações que vêm de diversos outros setores como planejamento, projetos, obras, manutenção e operação. Esta atualização é imprescindível para as ações cotidianas na área da operação e manutenção. Um cadastro desatualizado torna-se um problema cada dia maior, uma vez que a perda da confiabilidade gera um ciclo vicioso, onde quem consulta passa a não confiar e para utilizar quem executa, sabendo que o dado não será utilizado, não dá à atenção ao cadastro, que se torna cada vez mais desatualizado.

O cadastro técnico é um conjunto de informações que apresenta as características do sistema de abastecimento de água como redes, equipamentos e acessórios, resultando em um conjunto de informações para profissionais que tenham necessidade de acesso às informações técnicas do sistema.

Procura-se reunir informações da situação atual do sistema, bem como projetos futuros e implantações previstas. Geralmente, o acervo técnico reúne um conjunto de plantas e relatórios em mapotecas, com o avanço tecnológico, toda esta documentação deve ser substituída pelo meio digital, agilizando tanto a consulta quanto eventuais atualizações.

O Objetivo desse estudo é executar uma previsão a longo e curto prazo, do volume de água disponibilizado para a rede de distribuição de água do município de Campos dos Goytacazes, RJ. Com essa previsão espera-se ser possível realizar uma melhor análise de orçamento dos gastos da empresa ao longo do tempo, melhorando seus indicadores e controle de volume de água disponibilizado ao longo dos anos.

Para realização desse estudo, foram utilizados da Concessionária responsável pelo serviço ligado ao saneamento básico no município.

## METODOLOGIA

Com auxílio do cadastro técnico e de operadores diretamente ligados ao sistema de abastecimento de água (SAA). A partir desses dados, foi possível realizar todos os estudos desse trabalho. Os dados foram previamente analisados pela empresa responsável pelos serviços no município.

Foram coletados 1095 registros de volume de água disponibilizado (VD) diariamente pela concessionária em estudo. Totalizando uma coleta de 2 anos, de janeiro de 2019 a dezembro de 2021.

Uma metodologia bastante utilizada na análise de uma série temporal é Box e Jenkins. Conforme (Tápia, 2000), a utilização de séries temporais pelo método Box e Jenkins é representada pelo conjunto de processos estocásticos ARIMA (do inglês Autoregressive Integrated Moving Average) representado pelas letras (p, d, q), em que p indica o número de parâmetros autorregressivo, d representa o número de diferenças efetuadas na série para que se possa tornar estacionária e q indica o número de parâmetros de médias móveis.

Assim uma série temporal é estacionária e sua média, variância e covariância permanecem constantes, independente do período que foram analisadas, ou seja, elas não variam no decorrer do tempo. Um processo de grande importância das séries temporais estocásticas discretas é a presença de ruído branco.

Conforme Ediger e Akar (2007), o ARIMA é um dos mais populares modelos para análise da previsão de demanda para séries temporais. Os modelos ARIMA são modelos que utilizam apenas dados históricos de séries temporais com o intuito de expressar como as séries reagem de acordo com a variação estocástica anterior (BABAI; ALI; BOYLAN, 2013).

Os modelos ARIMA geram previsões através de informações contidas na própria série cronológica. Estes modelos baseiam-se em equações contendo termos estocásticos, em uma classe de equações lineares conveniente para a modelagem de previsões de séries temporais. Conforme destacado por Werner e Ribeiro (2003), o emprego da metodologia se baseia em três parâmetros para a construção dos modelos ARIMA, são eles: autorregressivo (p) – número de termos do modelo que descreve a dependência entre observações sucessivas; diferenças (d) – avaliar o aspecto estacionário da série e em caso negativo realiza processos de diferenciação até alcançar a estacionariedade dos dados; e, médias móveis (q) – média aritmética que se baseia no impacto dos dados mais recentes da série.

Assim que identificado esses componentes, os mesmos são descritos matematicamente como ARIMA (p, d, q). As técnicas de previsão utilizadas pelo modelo ARIMA, fundamentam-se na ideia de transformar as séries temporais em estacionárias pelo processo de diferenciação.

Conforme Almeida (2018), a estacionariedade é uma propriedade que indica se os atributos estatísticos (média, variância e função de autocorrelação) permanecem constantes ao longo do tempo. Desse modo, para validar os

requisitos de estacionariedade, normalmente são utilizados testes de hipóteses, como o teste de Dickey-Fuller aumentado, com a finalidade de verificar se a função que representa a série contém raiz unitária nos operadores de retardo (ALMEIDA, 2018), bem como são utilizadas funções de Autocorrelação Amostral (ACF ou FAC) e Autocorrelação Amostral Parcial (PACF ou FACP) para se analisar a influência das observações passadas no comportamento futuro da série (MARTIN et al., 2016).

Conforme Zhang (2003), os modelos ARIMA são bastante flexíveis na medida em que podem representar vários tipos diferentes de séries temporais, ou seja, autorregressiva pura (AR), média móvel pura (MA) e séries AR e MA combinadas (ARMA). Segundo Cowpertwait e Metcalfe (2009), o processo que gera a série temporal em um modelo ARMA, apresenta-se da seguinte forma na Equação 1:

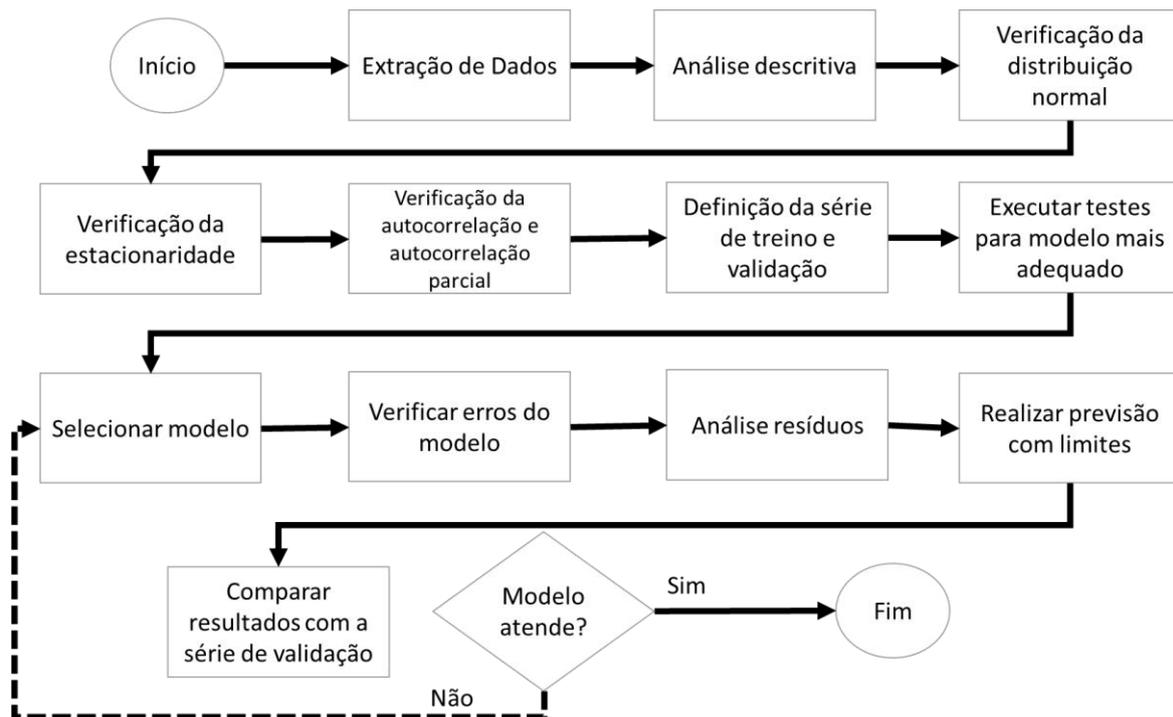
$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Equação 1

Onde  $y_t$  e  $\epsilon_t$  são o valor real da série e o erro aleatório no tempo  $t$ , respectivamente. Os erros aleatórios ( $\epsilon_t$ ) são assumidos como ruído branco, isto é, se comportam como independentes e identicamente distribuídos com média zero e variância constante (ALMEIDA, 2018). Já os termos  $\phi_i$  ( $i = 1, 2, 3, \dots, p$ ) e  $\theta_j$  ( $j = 1, 2, 3, \dots, q$ ) são parâmetros do modelo, no qual  $p$  e  $q$  são números inteiros associados às ordens dos processos AR e MA, respectivamente (ALMEIDA, 2018).

Para essa pesquisa, foi utilizado um fluxo de análise baseada na metodologia de Box e Jenkins, ilustrada na figura a seguir:

**Figura 1 – Fluxo de análise da pesquisa**



Fonte: Elaborado pelo autor com base no modelo de Box e Jenkins

A construção do modelo usando séries temporais segue um ciclo iterativo da metodologia de Box e Jenkins, é composto de algumas etapas. Para o fluxo adaptado algumas análises e verificações importante para o modelo ARIMA foi adicionado para tentar chegar a um modelo mais apurado.

A análise descritiva é um estudo simplificado e exploratório para entender um pouco mais do fenômeno e encontrar outlier, caso existam. Após análise descritiva, é feito um teste de normalidade, nesse trabalho usa-se o teste de Shapiro-Wilk e posteriormente o teste de estacionariedade (ADF e KPSS) da curva, para garantir que não existe tendencia no modelo. Isso é realizado pois os modelos de previsão autorregressivos baseiam-se nas

características de linearidade da série temporal e estacionariedade. De acordo com Gujarati (2000), uma série temporal é estacionária se suas médias e variância forem constantes ao longo do tempo e o valor da covariância entre dois períodos depender da defasagem entre dois períodos e não do período de tempo efetivo. Do contrário, a série é definida como sendo uma série temporal não-estacionária. A estacionariedade de uma série temporal deve ser comprovada através de testes. No presente estudo utilizou-se o teste da raiz unitária desenvolvido por Kwiatkowski (1992). O teste elaborado por Kwiatkowski (1992), conhecido por teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin) testa a hipótese nula de estacionariedade contra a hipótese alternativa, da seguinte forma:  $H_0$  quando a série é estacionária e  $H_1$  para quando série é não estacionária. Os valores encontrados a partir do teste de KPSS devem ser confrontados com os valores críticos para comprovação das hipóteses. Em seu estudo, Kwiatkowski (1992) apresentam uma tabela com os valores críticos. Estes valores normalmente são apresentados pelos softwares de análise juntamente com o resultado do cálculo do indicador KPSS, o que permite agilidade na interpretação dos resultados.

Posteriormente foi necessário identificar os valores apropriados para os parâmetros. Para determinar suas ordens e valores, a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) auxiliam nessa tarefa. Com todas as análises realizadas, a série original foi segmentada em duas séries menores. 2/3 da série será usada para treinar o algoritmo e criar um modelo adequado de previsão (utilizando o algoritmo em R presente no apêndice A). enquanto que o restante dos dados, 1/3 do total da série original, será usada como validação para comparar os valores de previsão do modelo final com dados reais da própria série e analisar se a previsão atende ou não às necessidades do estudo.

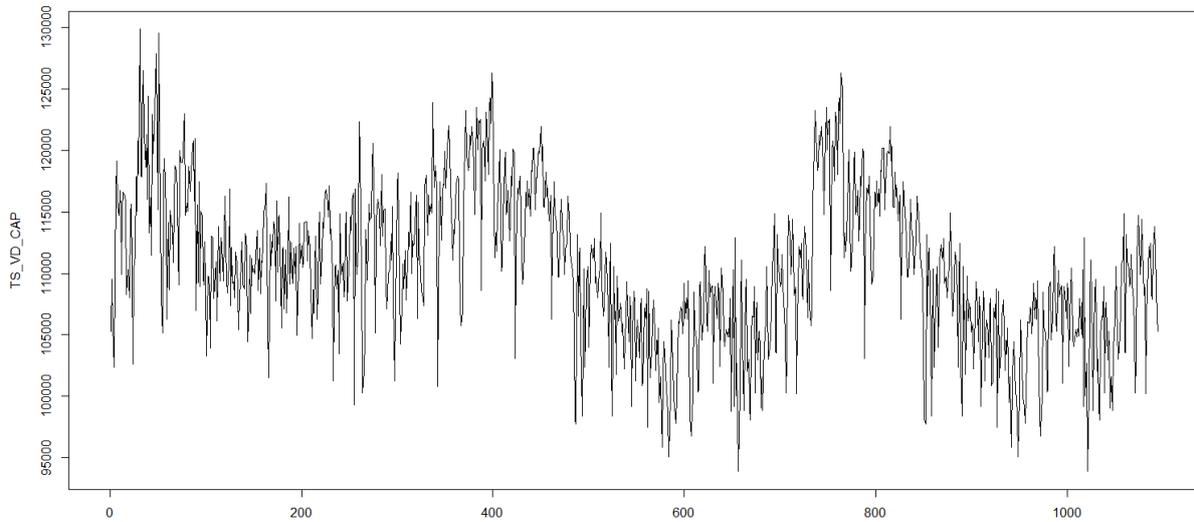
Para gerar o modelo mais adequado, foi aplicada a função `auto.arima` no algoritmo R, e seu resultado é apresentado no apêndice B. O critério de seleção adotado foi escolher os menores valores AIC e BIC (Akaike Information Criterion e Bayesian Information Criterion, respectivamente), tendo em vista que, estes critérios têm por objetivo indicar o modelo com menor número de parâmetros, uma vez que são construídos com base na variância estimada e no tamanho da amostra. O modelo que apresentar o menor valor de AIC e BIC será o que melhor se ajustará aos dados.

## RESULTADOS

Com o objetivo de analisar o comportamento da série de dados de volume disponibilizado de água tratada no município de Campos dos Goytacazes. Utilizou-se os dados diários do período de janeiro de 2020 a dezembro de 2021, formando uma série de 2 anos, com 660 observações. Para uma análise previa desenvolveu-se, inicialmente, um gráfico da série original. Essa primeira observação é importante para identificar possíveis discrepâncias ou erros que podem existir na extração dos dados, bem como entender, superficialmente, como é o comportamento da curva em relação ao tempo estudado.

Observa-se que a série assume um comportamento estacionário, pois não apresenta uma tendência bem definida e picos muito elevados levando em conta a série original. Porém, uma análise visual não é comprobatória, sendo necessário realizar testes que validem essas afirmações, norteados a construção de um modelo viável.

**Figura 2 – Série Temporal – Volume Disponibilizado**



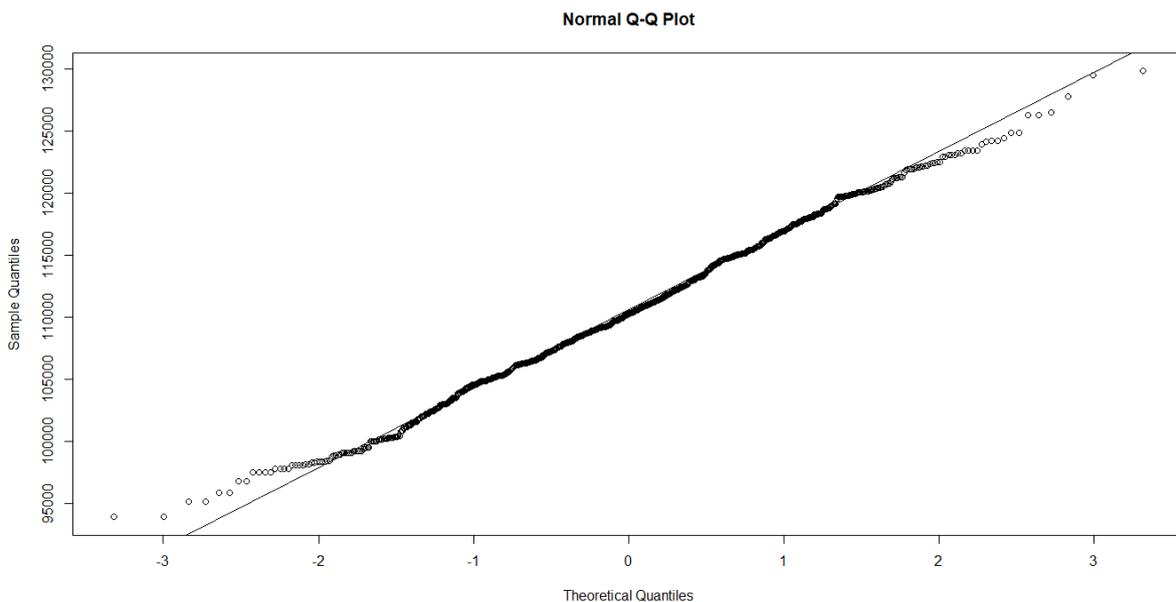
**Fonte: Elaborado pelo autor.**

Na segunda série, o objetivo foi o estudo de polímeros naturais (amidos de batata, mandioca e araruta), como No eixo y, são os valores de volume disponibilizado (VD) em m<sup>3</sup> e no eixo x, temos os valores em dias, sendo o dia 1 (01/01/2019) e o dia 1095 (31/12/2021).

Antes de iniciarmos os ajustes necessários para criação do modelo, é interessante realizarmos uma análise descritiva da série, sendo possível entender melhor seu comportamento. A série possui 1095 observações, com um valor mínimo de 93876 m<sup>3</sup>, máximo de 129844 m<sup>3</sup>, média aritmética de 110464 m<sup>3</sup> e mediana de 110298 m<sup>3</sup>.

Os dados descritivos podem ser utilizados como futuros validadores ao decorrer do ajuste do modelo. A série não apresenta grandes picos e outliers grandes o suficiente para justificar algum tratamento. Dessa forma, o próximo passo é analisar se a série possui uma distribuição gaussiana.

**Figura 3 – Distribuição dos dados da série**



**Fonte: Elaborado pelo autor**

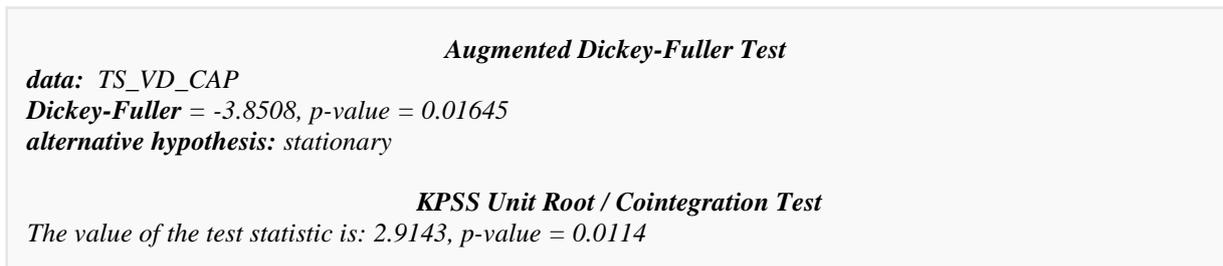
Analisando a figura acima, podemos inferir que a série possui um comportamento normal, para garantir essa afirmativa, aplica-se o teste de Shapiro.

Para o teste de Shapiro-Wilk, aplicado na série temporal original, foi encontrado o valor da estatística W de 0,996 e uma significância (p-value) de 0,167, maior que 0,05, portanto não difere significativamente de uma distribuição normal.

Visualmente identifica-se que a série é estacionária, portanto, para confirmarmos esta estacionariedade fizemos o teste de Kwiatkowski, Phillips, Schmidt and Shin (KPSS), e o Teste da raiz unitária Augmented Dickey-Fuller (ADF), com base nas funções de autocorrelação da série original, e verificados os testes de significância para embasar a conclusão de que a série não possui uma tendência definida.

Para o ADF, é considerado H0 para afirmar a não estacionariedade da série e H1 para a sua estacionariedade., usando uma significância estatística de 5% (p-value < 0,05), com auxílio do R:

**Figura 4 – Resultado do Teste de Estacionariedade**

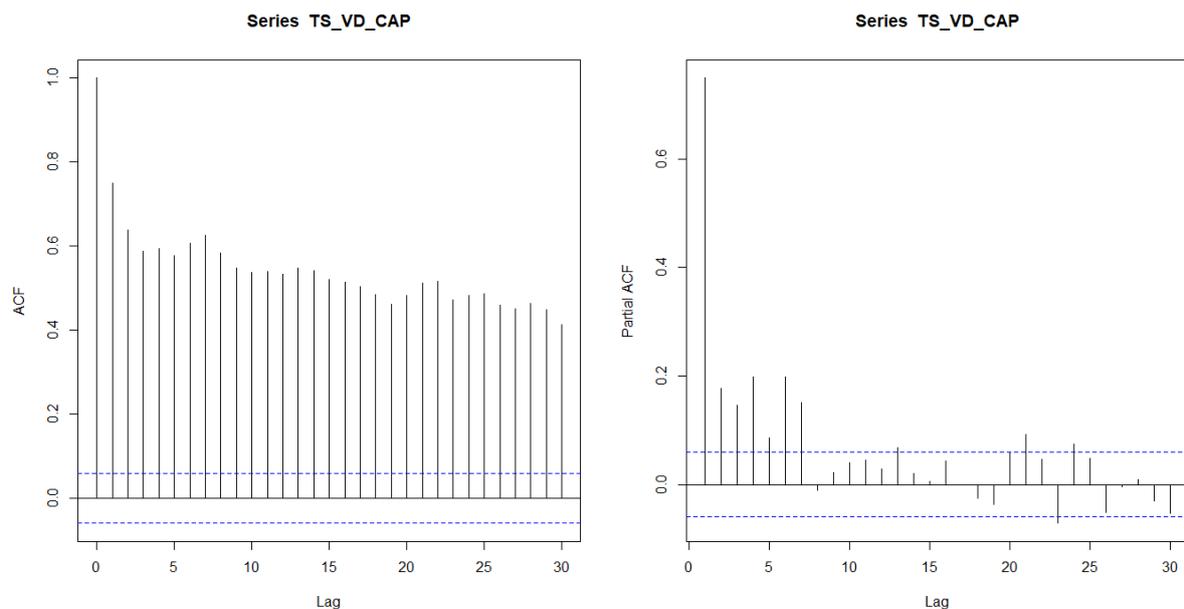


**Fonte: Elaborado pelo autor com auxílio do Rstudio**

Dessa forma podemos concluir que a série estudada é estacionária, conforme evidenciado pelo teste ADF e KPSS, na figura acima.

Além das análises de variância, estacionariedade e normalidade, foram feitas as autocorrelações e autocorrelações parciais e verificar se a série está dentro das bandas máximas e mínimas estipuladas pelos testes (FAC e FACP), conforme resultado apresentado a seguir:

**Figura 5 – Autocorrelação e Autocorrelação Parcial da série original**



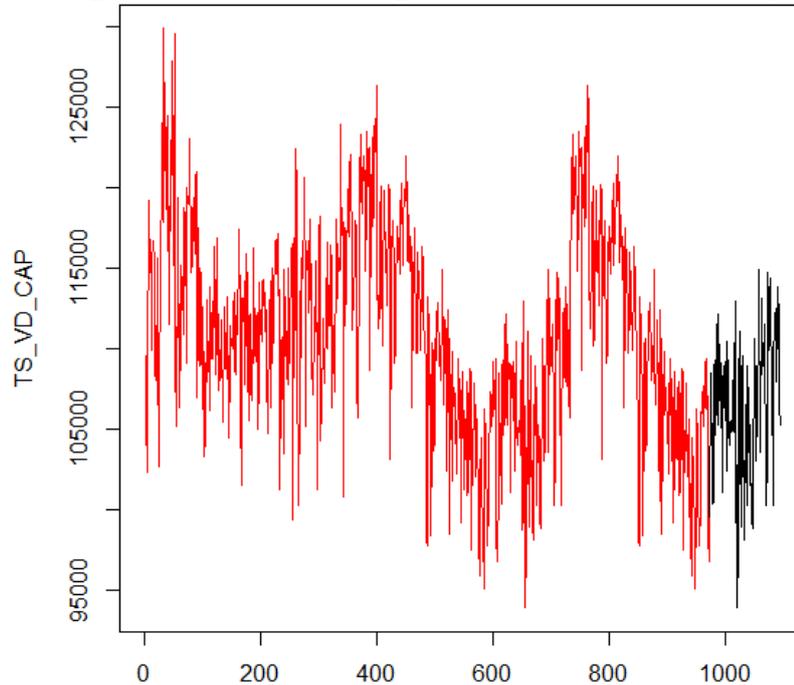
**Fonte: Elaborado pelo autor.**

Para ambos os gráficos, temos observações que ultrapassam as bandas superiores e inferiores, isso mostra a necessidade de realização de ajuste na série.

Para isso, a série original foi segmentada em 2 partes, uma série de treino, contendo 2/3 das observações e uma série de validação, contando 1/3 das observações.

Dessa forma, a série original foi dividida da seguinte forma:

**Figura 6 – Divisão da série original em série treino e série validação**



**Fonte: Elaborado pelo autor.**

Em vermelho, temos a série de treino e em preto a série que será usada para validação no fim desse estudo. Analisada a série, o próximo passo é identificar quais os modelos significativos que apresentam um comportamento semelhante a série original, para isso, foram testados uma série de modelos encontrados por meio da variação dos parâmetros do modelo do tipo ARIMA. Como regra de decisão para selecionar o modelo mais adequado, usou-se os critérios AIC e BIC, por meio dos menores valores encontrados para essas duas variáveis. Os teste e resultados obtidos encontram-se na tabela abaixo. O algoritmo aplicado no RStudio, seleciona o modelo que apresenta o menor indicador AIC.

**Figura 7 – Teste de modelos de previsão, variando os parâmetros ARIMA**

*Fitting models using approximations to speed things up...*

ARIMA (2,1,2) with drift	: 18824.98
ARIMA (0,1,0) with drift	: 19071.57
ARIMA (1,1,0) with drift	: 18994.73
ARIMA (0,1,1) with drift	: 18895.92
ARIMA (0,1,0)	: 19069.56
ARIMA(1,1,2) with drift	: 18816.29
ARIMA(0,1,2) with drift	: 18827.07
ARIMA(1,1,1) with drift	: 18815.16
ARIMA(2,1,1) with drift	: 18826.32
ARIMA(2,1,0) with drift	: 18947.1
ARIMA(1,1,1)	: 18813.38
ARIMA(0,1,1)	: 18893.93
ARIMA(1,1,0)	: 18992.72
ARIMA(2,1,1)	: 18824.35
ARIMA(1,1,2)	: 18814.46
ARIMA(0,1,2)	: 18825.1
ARIMA(2,1,0)	: 18945.09
ARIMA(2,1,2)	: 18823.07

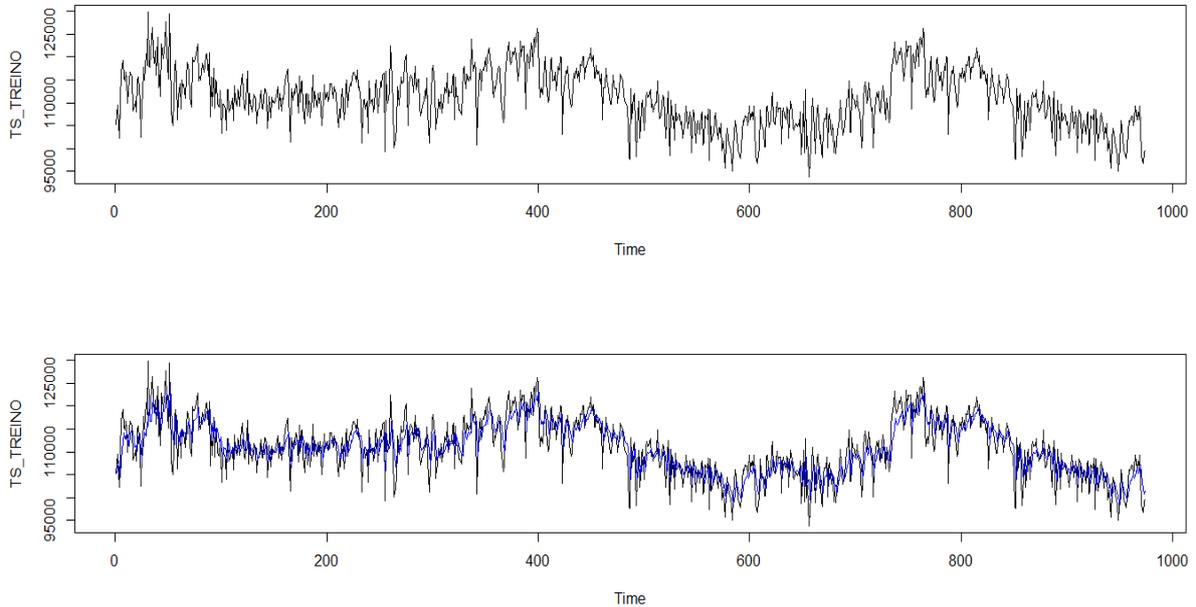
*Now re-fitting the best model(s) without approximations*  
*Best model: ARIMA(1,1,1)*  
*Series: TS\_TREINO*  
*ARIMA(1,1,1)*

**Fonte: Elaborado pelo autor**

Dessa forma, encontramos que o modelo mais adequado é o ARIMA (1,1,1), com isso, podemos criar um modelo baseado na série treino e analisar seus erros em relação ao treino.

Na figura a seguir, temos respectivamente, a série em preto demonstrando a série treino e posteriormente o resultado as séries de previsão usando ARIMA (1,1,1). A seguir podemos analisar a acurácia do modelo a partir de seus erros:

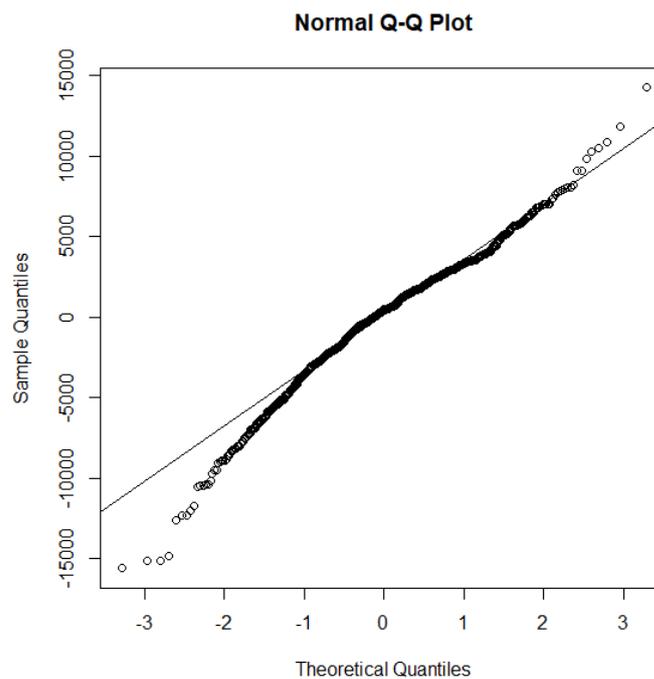
**Figura 8 – Comparativo entre a série de treino e a série de modelo para previsão**



**Fonte: Elaborado pelo autor.**

Para o modelo ARIMA (1,1,1) foi calculado um MAPE de 0,018, em relação ao modelo de treino, sendo esse, considerado um valor de erro médio percentual absoluto aceitável para o objetivo desse estudo. Encontrados o modelo adequado, o próximo passo é verificar as funções de autocorrelação (FAC) e autocorrelação Parcial (FACP), verificando se os resíduos não estão auto correlacionados assim como a normalidade dos resíduos:

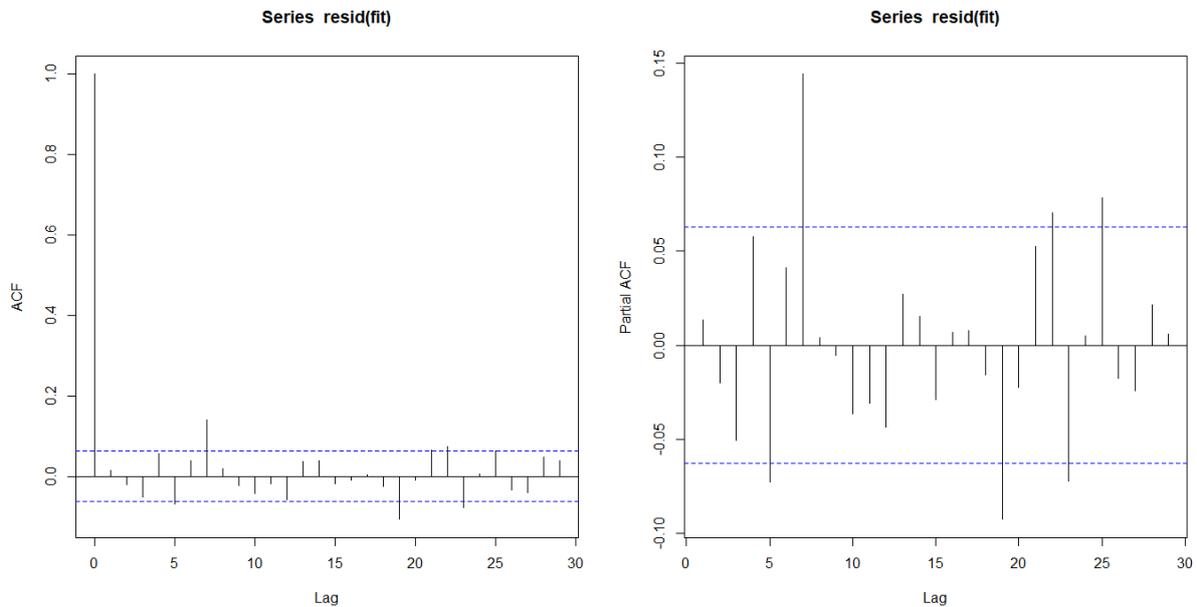
**Figura 9 – Teste de normalidade para os resíduos**



**Fonte: Elaborado pelo autor**

Analisando os resíduos gerados com os resultados no modelo e do treino é possível inferir visivelmente que possuem uma dispersão próxima à normalidade. Já para analisar a autocorrelação e autocorrelação parcial dos resíduos é usado o FAC e FACP:

**Figura 10 – FAC e FACP para os resíduos**



**Fonte: Elaborado pelo autor**

Com a série ajustada, temos os resíduos da previsão dentro dos limites máximos e mínimos estipulados pelo teste, com poucos pontos saindo desses limites, em comparação a série original. Mostrando um melhor ajuste do modelo ARIMA (1,1,1).

Com isso, podemos realizar uma previsão para os próximos 10 dias a uma probabilidade de 95%, e comparar com a série de validação, conforme tabela a seguir:

**Tabela 1 - Resultados da previsão para o modelo ARIMA**

<b>Dia</b>	<b>Validação</b>	<b>Limite Inferior</b>	<b>Limite Superior</b>	<b>Range</b>
975	108086	93893.38	108990.9	Atende
976	108882	93781.45	110590.8	Atende
977	113217	93774.98	114164.3	Atende
978	114980	93724.63	115430.8	Atende
979	107789	93647.20	111590.5	Atende
980	110326	93558.49	111710.6	Atende
981	111697	93465.84	111815.2	Atende
982	114588	93372.23	115913.4	Atende
983	113461	93278.83	112008.5	Atende

**Fonte: elaborado pelo autor com auxílio do Rstudio. Os dados são apresentados em m<sup>3</sup>**

Na tabela, podemos analisar que todos os dados para os próximos 10 dias estão dentro do limite inferior e superior determinado pelo modelo ARIMA, dessa forma o status para o range de 95% de probabilidade é classificado como “Atende”. No apêndice é apresentado uma previsão de 200 dias com a mesma análise.

## DISCUSSÕES

A estimativa de longo prazo é geralmente necessária para a organização e projetos de saneamento; enquanto a estimativa de curto prazo é apropriada na operação e administração.

A maioria dos métodos de previsão de séries temporais se baseia na suposição de que as observações passadas contêm todas as informações sobre o padrão de comportamento da série temporal e esse padrão é recorrente no tempo. O propósito dos métodos de previsão consiste em distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever os valores futuros da série temporal. Assim, pela identificação desse componente, a previsão para períodos subsequentes ao observado pode ser desenvolvida.

Existem inúmeras metodologias e modelos para estudar as informações de séries temporais, encontrar um modelo de previsão adequado para uma previsão de volume de água a ser disponibilizado diariamente em uma cidade não é uma tarefa simples. A previsão da demanda da água na rede de distribuição ao longo do tempo, utilizando séries temporais é uma zona de pesquisa dinâmica que atraiu a convergência do grupo de analistas nas últimas duas décadas. O ponto principal da exibição de séries temporais é reunir com precisão e considerar as estimativas passadas de uma série temporal para construir um modelo adequado.

As primeiras metodologias propostas para prever a demanda de água eram baseadas em modelos estatísticos normais (Jain et al., 2001). Zhou et al. (2002), foram os pioneiros em propor o modelo ARIMA criaram modelos de séries temporais para utilização diária de água em Melbourne, Austrália. Esses níveis de demanda podem ser comunicados como um elemento de fatores atmosféricos (por exemplo, temperatura do ar, volume e evento de precipitação) e demanda anterior de água.

Duas estratégias diferentes recomendadas por (Chen e Zhang 2006), em que a demanda de água por hora é antecipada utilizando mínimos quadrados bayesianos e não bayesianos.

Para o estudo realizado no município de Campos dos Goytacazes, o uso do modelo ARIMA se mostrou eficaz, apresentando um erro inferior ao erro baixo e dados de validação dentro do limite estipulado pelo próprio modelo, assim como demonstrado no estudo de Zhou, 2002.

Autores como Razali et al. (2018) realizaram um estudo para prever o consumo de água na Universidade Tun Hussein Onn Malaysia (UTHM). Para esse estudo os modelos propostos Holt Winter's ARIMA, fazendo um comparativo com o presente trabalho, dado o perfil da curva, modelos que levam em consideração a característica sazonal não apresentam bons resultados quando aplicados a uma série de 2 anos no município. A sazonalidade pode existir, porém ela não se faz significativa nesse período de dados coletados dessa forma o modelo SARIMA, Holt Winter's, entre outros, não seriam eficientes para esse caso específico.

Com isso, podemos apresentar um algoritmo em R, para realizar todos os passos desse trabalho de maneira automática, podendo ser recalibrado a cada novo dado incluído na base de dados onde o algoritmo for aplicado. Esse algoritmo é apresentado no Apêndice A, com todas as bibliotecas e dependências técnicas já incluídas e automatizadas para futuras aplicações.

## CONCLUSÃO

Estimar o volume disponibilizado de água tratada, no curto ou longo prazo, em uma cidade, é uma tarefa de alto valor tanto para o negócio, que por sua vez poderá planejar melhor seus projetos, estudar possíveis reduções de custo em produto químico, energia, lodo e pessoas, Tanto para o meio ambiente, onde abre a possibilidade de realização de diversos estudos nos mananciais de captação de água bruta, entendendo sua dinâmica hídrica prevendo possíveis momentos em que a demanda de água tratada da cidade ultrapassa a capacidade de disponibilidade de água no corpo hídrico. Esse estudo tem como objetivo iniciar uma discussão apresentando uma possibilidade de aplicação de séries temporais em problemas de saneamento básico.

Uma limitação dessa pesquisa é o período de dados que foi utilizado, para períodos mais longos é esperado um comportamento sazonal do consumo de água, como consequência o volume de água tratada disponibilizado também teria uma componente sazonal. É possível que em series com período maiores essa característica se apresente com grande relevância e o modelo ARIMA não seja o ideal a ser aplicado.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. ALMEIDA, D. M. Modelos híbridos de séries temporais aplicados ao sistema automotivo On Board Diagnostics. 2018, 75 f. Dissertação. (Mestrado em Ciência da Computação) - Universidade Federal de Pernambuco, Recife, 2018. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/32682/1/DISSERTA%20de%20Almeida.pdf>. Acesso em 09 de Jan. 2020.
2. BABAI, M. Z.; ALI, M.; BOYLAN, J. Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: Theory and empirical analysis. *International Journal Of Production Economics*, v. 143, n. 2, p.463-471, 2013.
3. CHEN, F.; YANG, B.; ZHANG, W.; MA, J.; LV, J.; YANG, Y. Enhanced recycling network for spent e-bicycle batteries: a case study in Xuzhou, China. *Waste Manag*, v. 60, p. 660–665, 2017
4. COWPERTWAIT, P. S. P.; METCALFE, A. V. *Introductory Time Series with R*. 1. ed. New York: Springer, 2009.
5. EDIGER, V. Ş.; AKAR, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, v. 35, n. 3, p.1701-1708, mar. 2007.
6. GUJARATI, D. N. *Econometria Básica*. 3. ed. São Paulo: Makron Books, 2000.
7. MARTIN, A. C.; HENNING, E.; WALTER, O. M. F. C.; KONRATH, A. C. Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville. *Revista Espacios*, v. 37, n. 6, 2016.
8. S N A M Razali and M S Rusiman and N I Zawawi and N Arbin, Forecasting of Water Consumptions Expenditure Using Holt-Winter's and ARIMA, *Journal of Physics: Conference Series*. 10.1088/1742-6596/995/1/012041
9. S.L Zhou, T.A McMahon, Q.J Wang, Frequency analysis of water consumption for metropolitan area of Melbourne. *Journal of Hydrology*, Volume 247, Issues 1–2, 2001, Pg 72-84, ISSN 0022-1694
10. TÁPIA, Milena, *Redes Neurais Artificiais: Uma Aplicação na Previsão de Preços de Ovos*, Dissertação de mestrado, UFSC, 2000.
11. WERNER, L; RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos Box Jenkins na área de assistência técnica de computadores pessoais. *G&P – Gestão & Produção*, vol. 10, n. 1, p. 47-67, abr. 2003.
12. ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural model. *Neurocomputing*. [s.l], n. 50, p. 159-175, 2003.
13. ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, v. 50, n.1, p. 159-175, 2003.