

ANÁLISE E CARACTERIZAÇÃO DO CONSUMO DE ÁGUA DE MUNICÍPIOS DO RIO GRANDE DO NORTE NA REGIÃO SEMIÁRIDA

Suzana Alves Cândido *
Gustavo Marques Calazans Duarte**

RESUMO: A disponibilidade dos recursos hídricos tem sido pauta relevante mundialmente nos últimos anos e na região semiárida não é diferente. Por apresentar frequentes registros de escassez hídrica, a gestão dos recursos hídricos no semiárido e consequentemente na Região Seridó, necessitam constantemente tomar decisões e elaborar estratégias que melhor atendam as demandas da população. Com base nisso, o consumo *per capita* de água (QPC) enquadra-se como uma componente importante nesse processo de planejamento, pois o QPC sofre significativa interferência de diversos fatores. Diante disso, o presente trabalho, foi realizado com a finalidade de analisar quais os fatores mais interferem no consumo *per capita* de água de 23 municípios atendidos pela Companhia de Águas e Esgotos do Rio Grande do Norte (CAERN), na regional Seridó. Para alcançar tal objetivo, realizou-se uma coleta de dados quantitativos referentes aos serviços de água e às condições socioeconômicas, climáticas e geográficas de cada município; assim como, foram realizadas análises por meio de métodos estáticos como a PCA e RF, para melhor entender e delimitar quais fatores melhor explicavam o QPC da região. Por meio disso, com base nas análises realizadas, constatou-se que os fatores que explicam melhor o QPC dos municípios analisados são: o Salário médio mensal dos trabalhadores formais, o Pessoal ocupado, a Pluviometria quantis anual, o Índice de atendimento urbano de água e a Extensão da rede de água por ligação. Obteve-se ainda um modelo matemático que, por meio dessas variáveis, é possível estimar o QPC para os municípios atendidos pela regional Seridó.

Palavras-chave: Consumo *per capita* de água; recursos hídricos; fatores intervenientes; PCA.

ABSTRACT: The availability of water resources has been a relevant topic worldwide in recent years and the semi-arid region is no different. Due to frequent records of water scarcity, the management of water resources in the semi-arid region and consequently in the Seridó Region, constantly needs to make decisions and develop strategies that best meet the demands of the population. Based on this, per capita water consumption (QPC) is an important component in this planning process, as the QPC suffers significant interference from several factors. Therefore, the present work was carried out with the purpose of analyzing which factors most interfere in the per capita water consumption of 23 municipalities served by the large northern river Water and Sewage Company (CAERN), in the Seridó region. To achieve this objective, quantitative data was collected regarding water services and the socioeconomic, climatic and geographic conditions of each municipality; as well as, analyzes were carried out using static methods such as PCA and RF, to better understand and define which factors best explained the QPC in the region. Therefore, based on the analyzes carried out, it was found that the factors that best explain the QPC of the municipalities analyzed are: the average monthly salary of formal workers, the employed personnel, the annual quantile rainfall, the urban water service index and Extension of the water network by connection. A mathematical model was also obtained which, through these variables, makes it possible to estimate the QPC for the municipalities served by the Seridó region.

Key words: *Per capita* water consumption; water resources; intervening factors; PCA.

*Bacharela em Ciência e Tecnologia e graduanda em Engenharia Civil – UFERSA

E-mail: suzanaalvescandido@gmail.com

**Engenheiro Civil, Mestre em Engenharia Sanitária (Professor Magistério Superior – UFERSA)

E-mail: gustavomcalazans@live.com

1 INTRODUÇÃO

No Brasil, a gestão dos recursos hídricos tem sido uma componente fundamental para o monitoramento e desenvolvimento dos planos de sustentabilidade para o país. O crescimento populacional, industrial e agrícola acelerado e concentrado em determinadas regiões do país, proporciona o aumento da demanda de água, o qual por vezes, os sistemas de abastecimento existentes associados aos fatores climáticos de cada região não comportam capacidade suficiente para suprir tais demandas de consumo, resultando no frequente problema de escassez hídrica; ressaltando com isso, a necessidade de aprimorar as estratégias para uma gestão eficiente dos recursos hídricos (PNSH, 2019, p. 13-17).

Mesmo sendo um recurso abundante no planeta, a disponibilidade hídrica em qualidade e quantidade adequada para toda a população brasileira ainda é um desafio, assunto este, que tem sido pauta incessante nos debates dos gestores para a elaboração de políticas públicas e para a concepção e ampliação de sistemas de abastecimento de água (SAA) (PNSH, 2019, p. 13-17). Na região semiárida do Rio Grande do Norte (RN), a situação da disponibilidade hídrica é pauta relevante diante das condições climáticas da região, caracterizada por apresentar de 7 a 8 meses por ano de estiagem. Além disso, apresenta relevante densidade populacional, de forma que a realidade regional exige constantes análises, para viabilizar e garantir a disponibilidade de água para as populações (MACEDO; TROLEIS; DE ALMEIDA, 2021).

Diante disso, uma componente importante na concepção, ampliação e monitoramento dos SAA e nas análises de disponibilidade hídrica, é a demanda do consumo de água a ser atendido. O consumo de água é subdividido em consumo doméstico, comercial, público e industrial. Vale destacar que alguns outros consumos devem ser considerados, como o consumo para o funcionamento do próprio SAA, as perdas, entre outros (HELLER; DE PÁDUA, 2010, p. 108). Toda a demanda do consumo de água a ser atendido, é o ponto de partida para se obter o consumo *per capita* de água (qpc); o qual, conforme Heller e De Pádua (2010, p. 108) é obtido por meio da divisão do total de demanda que foi disponibilizada pelo sistema e a população abastecida, dado em L/hab.dia.

$$qpc \text{ (L/hab. dia)} = \frac{\text{média diária do volume anual consumido por uma dada população (m}^3\text{)} \times 1000}{\text{população abastecida (hab)}} \text{ (Eq. 01)}$$

Azevedo Netto e Fernández (2018, p. 408) descrevem que “a demanda e o consumo são função de uma série de fatores inerentes à localidade a ser abastecida, varia de cidade para cidade, assim como pode variar de um setor de distribuição para outro, numa mesma cidade”. Os autores destacaram o clima, os hábitos, os costumes, o padrão de vida da população, a qualidade da água fornecida e a estrutura tarifária do consumo, entre outros fatores, como principais influenciadores do consumo de água.

Heller e De Pádua (2010, p. 133-138) abordam também em sua publicação os diversos fatores que influenciam no consumo *per capita* de água, tais como o nível socioeconômico da população, o clima, o porte, características e topografia da cidade e a administração do sistema de abastecimento de água.

Para bom entendimento do consumo *per capita* de água, torna-se primordial o conhecimento de algumas variáveis e conceitos, como a macromedição, a micromedição, os sistemas de controle de perdas, entre outros. De acordo com o Sistema Nacional de Informações Sobre Saneamento - SNIS (2022, p. 7) a macromedição consiste no “conjunto de medições realizadas no sistema público de abastecimento de água, desde a captação de água bruta até as extremidades de jusante da rede de distribuição”; ainda de acordo com o autor, a micromedição consiste na “medição do consumo realizada no ponto de abastecimento de um determinado usuário, independentemente de sua categoria ou faixa de consumo”, ou seja, é o volume de água consumido registrado por meio dos hidrômetros.

As perdas são ocorrências comuns aos SAA, que, devem ser cuidadosamente monitoradas e sanadas. As perdas são divididas em duas classificações, sendo as perdas reais e as perdas aparentes. As perdas reais ou físicas “se referem a toda água disponibilizada para distribuição que não chega aos consumidores; já as perdas aparentes, estão relacionadas ao volume de água que foi efetivamente consumido pelo usuário, mas que, por algum motivo, não foi medido” (SNIS, 2022, p. 25-26). Embora, as perdas recebam esta nomenclatura, é um fator diretamente relacionado ao consumo *per capita* de água pois, apesar da possibilidade em alguns casos desta água não ter sido efetivamente consumida, ela foi disponibilizada e quantificada pelo sistema, com um fim nobre que é atender os consumidores e as demandas de operação, enquadrando-se como volume de água consumido.

Diversos estudos realizados no Brasil, tem demonstrado a importância do desenvolvimento de pesquisas referente ao consumo *per capita* de água, tendo em vista que se trata de uma variável fundamental para o bom desempenho e estimativa da garantia da disponibilidade hídrica para as populações, como também para a sustentabilidade das regiões (FERNANDES NETO, 2003, p.17-23). Como por exemplo o trabalho de Silva et al (2020), que em sua pesquisa “Análise *per capita* do abastecimento de água no município de Soledade (Rio Grande do Sul)”, observaram que o município de Soledade possui um consumo per capita abaixo da média nacional e estadual, como também observaram que uma densidade maior de residências do que indústrias no município pode ter total influência nos resultados obtidos.

Para o desenvolvimento de estudos que necessitam da análise de uma variabilidade de dados, ao longo dos tempos foram desenvolvidos diversos métodos com o intuito de auxiliar no tratamento e na interpretação de dados e resultados, como por exemplo, o método da Análise de Componentes Principais (PCA) e método *random forest* (RF). Criado por Karl Pearson (1901), o método PCA ao longo dos tempos recebeu diversas contribuições e aprimoramentos, como por exemplo, a descrição de métodos computacionais mais práticos com a finalidade de analisar estruturas de correlação (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2016).

De acordo com Luciano (2021), o método estatístico PCA tem o objetivo de transformar dados multivariados em dados de menor dimensão, sem que a relação entre eles seja prejudicada; o autor destaca que dessa forma, torna-se possível “descobrir, visualizar e interpretar diferenças existentes entre variáveis e examinar as relações que podem existir entre as amostras, podendo detectar amostras com componentes atípicos”. O método *random forest*, consiste em um algoritmo proposto por Breiman (2001), com o qual são criadas diversas árvores de decisões, utilizando um conjunto de dados, que por fim apresenta como resultado final a média das decisões de cada árvore, possibilitando uma tomada de decisão menos complexa (AMARATUNGA; CABRERA; LEE, 2008 e DA CONCEIÇÃO, 2022).

O SNIS, é o setor de domínio público responsável por coletar dados e fornecer informações de caráter operacional, gerencial, financeiro e de qualidade, sobre a prestação de serviços de água e esgotos e sobre os serviços de manejo de resíduos sólidos urbanos do Brasil. O SNIS fornece informações de todo o país por meio de dados fornecidos pelas companhias prestadoras de serviços de saneamento (SNIS, 2022 e PORTELA; COSTA; CORREIA, 2023). Por meio desses dados, torna-se possível o desenvolvimento de estudos mais específicos para cada região brasileira, favorecendo o desenvolvimento de pesquisas que servem de subsídio para a tomada de decisões que visam uma gestão dos recursos hídricos mais eficiente e sustentável.

Com esse propósito, a presente pesquisa teve como objetivo geral, analisar o consumo de água dos municípios monitorados pela Regional Seridó, geridos pela Companhia de Águas e Esgotos do Rio Grande do Norte (CAERN) e com base em ciência de dados compreender os fatores que influenciam esse consumo. Como objetivos específicos, o presente trabalho realizou: a caracterização do consumo de água dos 23 municípios monitorados pela Regional Seridó; analisou os fatores que influenciam o índice de consumo de água desses municípios;

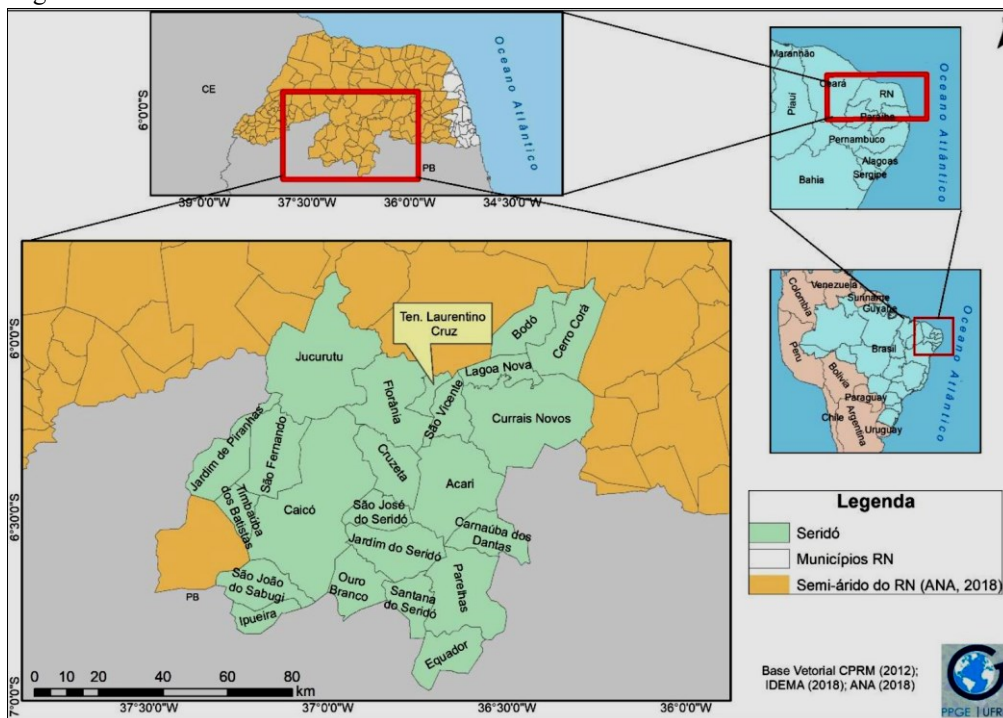
comparou os resultados obtidos na pesquisa com as referências estabelecidas na literatura e por fim, observou a importância da compreensão das variáveis diretamente relacionadas ao consumo de água, para o bom desempenho da gestão dos recursos hídricos e da sustentabilidade dos municípios.

2 DESENVOLVIMENTO

2.1 Caracterização da área de estudo

A região Seridó é uma das 7 principais localizações do estado do RN, monitorada pela CAERN. Possuindo a denominação de Regional Seridó (RSE), com sua sede no município de Caicó, a Regional é responsável pela prestação de serviços em 23 municípios, são eles: Acari, Bodó, Caicó, Carnaúba dos Dantas, Cerro Corá, Cruzeta, Currais Novos, Equador, Florânia, Ipueira, Jardim de Piranhas, Jardim do Seridó, Jucurutu, Lagoa Nova, Ouro Branco, Parelhas, Santana do Seridó, São Fernando, São João do Sabugi, São José do Seridó, São Vicente, Tenente Laurentino e Timbaúba dos Batistas, conforme apresentado no mapa de localização da Figura 1 (ACS CAERN, 2020 e MACEDO; TROLEIS; DE ALMEIDA, 2021).

Figura 1 – Mapa de localização nacional e estadual, com destaque dos municípios atendidos pela Regional Seridó.



Fonte: MACEDO; TROLEIS; DE ALMEIDA (2021).

Após análises bibliográficas, a região de estudo do presente trabalho, foi escolhida por compor a região semiárida brasileira, que possui uma característica climática natural de apresentar maiores períodos de estiagem e menores períodos de chuva durante o ano (CORREIA, 2011); como também por possuir registros de déficits hídricos não somente devido ao clima, mas também devido às características geomorfológicas da região, a qual apresenta um embasamento de rochas cristalinas ígneas e metamórficas, dificultando assim o acesso à aquíferos subterrâneos (MACEDO; TROLEIS; DE ALMEIDA, 2021). Com base nisso, a área de estudo foi escolhida, com o objetivo de analisar o consumo *per capita* de água da região e os fatores que podem interferir nesse consumo.

2.2 Metodologia

Para o desenvolvimento do presente trabalho, inicialmente realizou-se uma análise bibliográfica para a fundamentação e definição teórica dos assuntos abordados, assim como para a escolha e delimitação da área de estudo. Posteriormente, efetuou-se um levantamento de dados quantitativos no Sistema Nacional de Informações Sobre Saneamento (SNIS), referente aos serviços de água, os quais são apresentados como “Indicadores Operacionais de Água” (Quadro 1), da qual obteve-se a planilha de dados mais atual, disponível no período do levantamento de dados, com ano de referência de 2020. Os dados obtidos no SNIS são de abrangência nacional, dessa forma, realizou-se posteriormente a seleção de dados referente aos 23 municípios, objeto de estudo do presente trabalho; é importante destacar que o IN057 - Índice de fluoretação de água (%), não foi considerado nas análises, pois nas 23 cidades estudadas, nenhuma apresentou este indicador.

Quadro 1 – Código, descrição e unidade de medida, dos indicadores operacionais de água – SNIS.

Indicadores Operacionais de Água	
IN055	Índice de atendimento total de água (%)
IN023	Índice de atendimento urbano de água (%)
IN001	Densidade de economias de água por ligação (econ./lig.)
IN043	Participação das economias residenciais de água no total das economias de água (%)
IN011	Índice de macromedicação (%)
IN009	Índice de hidromedicação (%)
IN010	Índice de micromedicação relativo ao volume disponibilizado (%)
IN044	Índice de micromedicação relativo ao consumo (%)
IN052	Índice de consumo de água (%)
IN025	Volume de água disponibilizado por economia (m ³ /mês/econ)
IN053	Consumo médio de água por economia (m ³ /mês/econ)
IN014	Consumo micromedido por economia (m ³ /mês/econ)
IN017	Consumo de água faturado por economia (m ³ /mês/econ)
IN022	Consumo médio per Capita de água (l/hab.dia)
IN058	Índice de consumo de energia elétrica em sistemas de abastecimento de água (kWh/m ³)
IN020	Extensão da rede de água por ligação (m/lig.)
IN028	Índice de faturamento de água (%)
IN013	Índice de perdas faturamento (%)
IN049	Índice de perdas na distribuição (%)
IN050	Índice bruto de perdas lineares (m ³ /dia/km)
IN051	Índice de perdas por ligação (l/dia/lig.)

Fonte: Autor (2023).

Em seguida, em continuidade ao levantamento de dados quantitativos, realizou-se uma sondagem por meio do banco de dados do Instituto Brasileiro de Geografia e Estatística (IBGE) e do banco de dados da Empresa de Pesquisa Agropecuária do Rio Grande do Norte (EMPARN), para a obtenção de amostras de diversas variáveis para cada um dos 23 municípios estudados. As informações quantitativas coletadas foram referentes as condições socioeconômicas e climáticas. Por fim, com o auxílio da ferramenta computacional *Google Earth*, aferiu-se de forma aproximada a distância em metros, de um ponto do centro urbano até um corpo hídrico importante de cada cidade. O Quadro 2 apresenta a lista de variáveis coletadas nessas bases de informações, para subsidiar a aplicação do método.

Quadro 2 – Descrição e unidade de medida das demais variáveis utilizadas.

Variáveis do IBGE, EMPARN e Google Earth
População (hab.)
Salário médio mensal dos trabalhadores formais (salários mínimos)
Pessoal ocupado (pessoas)
PIB per capita (R\$)
Índice de Desenvolvimento Humano Municipal (IDHM)
Pluviometria quantis anual (mm)
Temperatura média (°C)
Distância do centro urbano até o corpo hídrico mais importante (m)

Fonte: Autor (2023).

Tendo finalizado a coleta e seleção de dados, elaborou-se uma planilha no Excel® com todos os dados para cada um dos 23 municípios, para a partir desta, realizar as análises estatística dos dados. Optou-se pela criação de siglas para cada variável, com o intuito de uma melhor estética e visualização dos dados nos gráficos e tabelas (Quadro 3).

Quadro 3 – Siglas, descrição e unidade de medida das variáveis utilizadas para as análises.

Siglas, descrição e unidades das variáveis	
IAT	Índice de atendimento total de água (%)
IAU	Índice de atendimento urbano de água (%)
DEAL	Densidade de economias de água por ligação (econ./lig.)
PERATEA	Participação das economias residenciais de água no total das economias de água (%)
I.Macro.	Índice de macromedicação (%)
I.Hidro.	Índice de hidromedicação (%)
I.Micro.RVD	Índice de micromedicação relativo ao volume disponibilizado (%)
IMicro.RC	Índice de micromedicação relativo ao consumo (%)
ICA	Índice de consumo de água (%)
IVADE	Volume de água disponibilizado por economia (m³/mês/econ)
CMed.AE	Consumo médio de água por economia (m³/mês/econ)
CMicro.E	Consumo micromedido por economia (m³/mês/econ)
CAFE	Consumo de água faturado por economia (m³/mês/econ)
QPC	Consumo médio <i>per Capita</i> de água (l/hab.dia)
ICEESAA	Índice de consumo de energia elétrica em sistemas de abastecimento de água (kWh/m³)
ERAL	Extensão da rede de água por ligação (m/lig.)
IFA	Índice de faturamento de água (%)
IPF	Índice de perdas faturamento (%)
IPD	Índice de perdas na distribuição (%)
IBPL	Índice bruto de perdas lineares (m³/dia/km)
IPL	Índice de perdas por ligação (l/dia/lig.)
P	População (hab.)
SMT	SMT - Salário médio mensal dos trabalhadores formais (salários mínimos)
PO	Pessoal ocupado (pessoas)
PIB	PIB per capita (R\$)
IDHM	Índice de Desenvolvimento Humano Municipal (IDHM)
Pluv.	Pluviometria quantis anual (mm)
T	Temperatura média (°C)
DCUCH	Distância do centro urbano até o corpo hídrico mais importante (m)

Fonte: Autor (2023).

Os dados coletados inicialmente foram analisados por meio do método estatístico Análise de Componentes Principais (PCA), utilizado com o objetivo de identificar as variáveis ou fatores importantes, que mais explicassem variações no consumo per capita de água dos 23 municípios. É importante destacar, que no presente trabalho, entende-se por variáveis cada

informação coletada nos bancos de dados como mencionado nos parágrafos anteriores do presente subtópico (2.2).

Para a implementação do PCA e respectiva elaboração dos gráficos, utilizou-se o software R Studio versão 4.3.1 (R *Beagle Scouts*, 2023), com o script:

```
{#1 Pacotes library(tidyverse) library(FactoMineR) library(factoextra) #2 Dados
dados_agua <- readxl::read_xlsx("G:/Meu Drive/UFERSA/2023/2023.1/TCC/Suzana Alves
Candido/dados_agua_PCA.xlsx") #3 Modl factominer res_pca <- PCA(dados_agua, quali.sup
= 1, graph = TRUE, scale=TRUE) summary(res_pca)
sweep(res_pca$var$coord,2,sqrt(res_pca$eig[1:ncol(res_pca$var$coord),1]), FUN="/")}
```

Diante da grande variabilidade de dados, muitos apresentam escala numérica e unidades diferentes, portanto se faz necessário realizar a normalização ou o auto escalonamento dos dados, etapa esta, obtida de forma automática no software R Studio, por meio da função *scale*, resultando assim em variáveis adimensionais e que exercem influências igualitárias.

Em seguida utilizando também o software R Studio, realizou-se uma nova análise, utilizando o método *random forest* (RF), com o objetivo de analisar e identificar a importância de cada variável por meio deste método, para descrever e explicar o consumo *per capita* de água.

Por fim, realizou-se no mesmo software, uma análise para os dados resultantes do método PCA e RF, por meio do método de regressão linear, logarítmica e polinomial, com a finalidade de se obter um modelo matemático para a predição do consumo per capita de água (QPC) para os municípios estudados.

Todas as etapas realizadas foram fundamentais, para o alcance dos objetivos propostos no presente trabalho. Com isso, a classificação do presente estudo configura-se como uma pesquisa bibliográfica, de natureza descritiva com uma abordagem quantitativa, no qual são quantificados, analisados e descritos algumas características socioeconômicas e dos fatores que influenciam o consumo *per capita* de água de cada município atendido pela RSE, apresentando também uma abordagem explicativa com a finalidade de enfatizar a importância da compreensão das variáveis diretamente relacionadas ao consumo de água, para o bom desempenho da gestão dos recursos hídricos e da sustentabilidade dos municípios.

3 RESULTADOS E DISCUSSÕES

3.1 Método Análise de Componentes Principais (PCA)

De acordo com os dados coletados do SNIS, do IBGE, da EMPARN e do *Google Earth*, foram obtidas 29 variáveis relacionadas aos indicadores de água, à questões socioeconômicas, climáticas e geográficas, para cada um dos 23 municípios atendidos pela Regional Seridó (Apêndice). No Apêndice é possível observar as 29 variáveis de cada município, com suas respectivas descrições, valores e siglas.

Durante a análise dos dados, observou-se que o município de Lagoa Nova apresentou dados muito discrepante dos demais, o que levou a decisão de desconsiderar este município nas análises, devido à sensibilidade dos métodos utilizados a valores com alta variabilidade. Lagoa Nova apresentou consumo *per capita* de 260 L/hab.dia, ao passo que a média da região foi de 96 L/hab.dia. Esse consumo maior que a média pode ser explicado pela intensa atividade de mineração que ocorre na região ou a algum equívoco na inserção dos dados na plataforma SNIS.

Das 29 variáveis analisadas, foram avaliadas as variáveis que não explicavam bem o consumo *per capita*, e para auxílio na redução da dimensionalidade do problema e diminuição da confusão do modelo, foram removidas as variáveis que eram obtidas a partir do QPC ou que em nada se relacionava com essa. Tais foram: o IAT - Índice de atendimento total de água, o I.Macro. - Índice de macromedição, o I.Micro.RVD - Índice de micromedição relativo ao volume disponibilizado, o IMicro.RC - Índice de micromedição relativo ao consumo, o ICA -

Índice de consumo de água, o IVADE - Volume de água disponibilizado por economia, o CMed.AE - Consumo médio de água por economia, o CMicro.E - Consumo micromedido por economia, o CAFÉ – Consumo de água faturado por economia, o ICEESAA - Índice de consumo de energia elétrica em sistemas de abastecimento de água, o IFA - Índice de faturamento de água, o IPF - Índice de perdas faturamento, o IPD - Índice de perdas na distribuição, o IBPL - Índice bruto de perdas lineares e IPL - Índice de perdas por ligação. Restaram 14 variáveis. O Quadro 4 apresenta a lista das 14 variáveis remanescentes, pós avaliação.

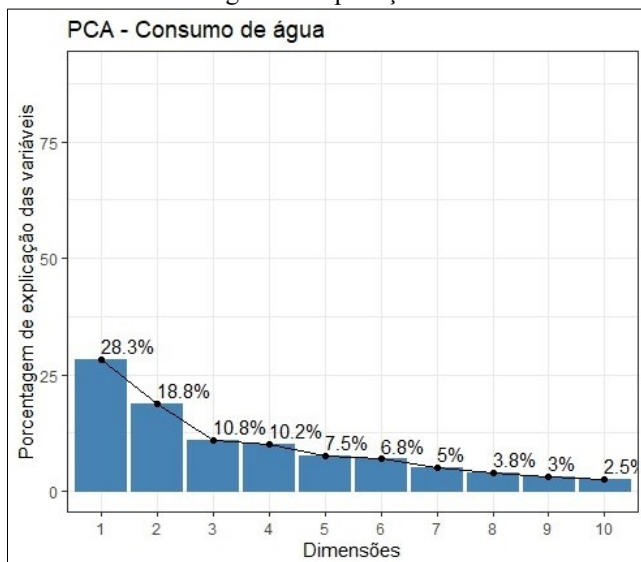
Quadro 4 – Variáveis remanescentes, pós avaliação.

Siglas e variáveis pós avaliação	
IAU	Índice de atendimento urbano de água (%)
DEAL	Densidade de economias de água por ligação (econ./lig.)
PERATEA	Participação das economias residenciais de água no total das economias de água (%)
I.Hidro.	Índice de hidrometração (%)
QPC	Consumo médio per Capita de água (l/hab.dia)
ERAL	Extensão da rede de água por ligação (m/lig.)
P	População (hab.)
SMT	Salário médio mensal dos trabalhadores formais (salários mínimos)
PO	Pessoal ocupado (pessoas)
PIB	PIB per capita (R\$)
IDHM	Índice de Desenvolvimento Humano Municipal (IDHM)
Pluv.	Pluviometria quantis anual (mm)
T	Temperatura média (°C)
DCUCH	Distância do centro urbano até o corpo hídrico mais importante (m)

Fonte: Autor (2023).

Essas variáveis explicam, por meio das cinco primeiras Componentes Principais (PCs) ou dimensões, 75,6% dos dados analisados relacionados ao consumo *per capita* de água, como pode-se observar no Gráfico 1. Jolliffe e Cadima (2016) destacam em sua publicação, que se utiliza 70% da variabilidade total como um ponto de corte comum, ou seja, das variáveis analisadas, as PCs que individualmente ou em conjunto, apresentarem de 70% acima, da percentagem de variância total contabilizada, nestas estão as variáveis que mais explicam determinada análise de dados.

Gráfico 1 – Porcentagem de explicação das 14 variáveis.



Fonte: Autor (2023).

Dessas 14 variáveis analisadas, após a análise PCA tornou-se possível reduzir a dimensionalidade do problema para 6 (seis) variáveis, sendo escolhidas as quais representavam melhor o conjunto de dados nas primeiras 5 dimensões, foram elas: o SMT - Salário médio mensal dos trabalhadores formais, o PO - Pessoal ocupado, a Pluv. - Pluviometria quantis anual, o IAU - Índice de atendimento urbano de água, a ERAL - Extensão da rede de água por ligação, e o QPC - Consumo médio per capita de água (Tabela 1). É importante destacar que a variável I.Hidro. - Índice de hidrometração na dimensão 3, apresentou em módulo um *escore* maior que a Pluv. - Pluviometria quantis anual, porém ao analisar conceitualmente essas duas variáveis, observou-se que a variável Pluv. apresentava uma melhor explicação dos dados analisados.

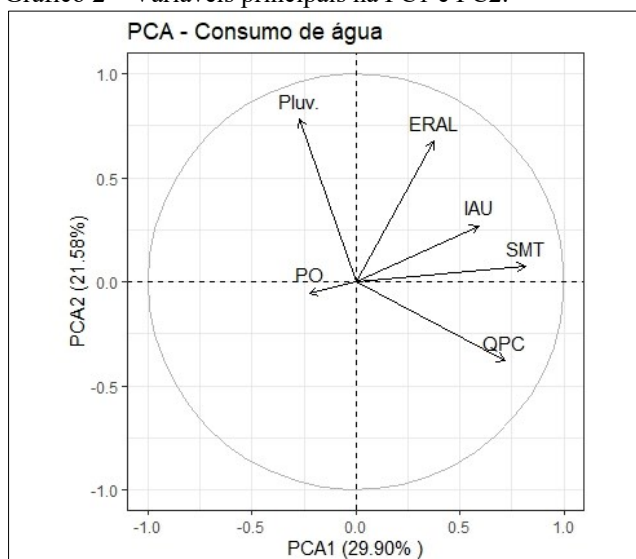
Tabela 1 – Valores dos *escores* das cinco variáveis principais nas 5 primeiras dimensões.

Variá.	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
P	0.91055498	0.19633041	-0.18214782	0.0293175412	0.05686445
SMT	-0.06209170	0.89022554	0.05627157	-0.0756943223	0.07492960
PO	0.94603021	0.14695740	-0.10457228	0.0494739270	0.10164718
PIB	-0.06001606	0.37547060	0.50707970	-0.0809773517	0.19899098
IDHM	0.65263440	-0.47925155	0.25137113	-0.2069857917	-0.01976612
Pluv.	0.12135250	-0.10719663	-0.57807698	0.6028178892	0.11784859
T	0.11777010	-0.76288394	0.11699292	-0.0820533506	0.40630303
DCUCH	0.68029572	0.15858307	0.15476590	0.0185623139	-0.38772595
IAU	-0.13514506	0.25696312	0.29955203	0.7026790969	-0.43701911
DEAL	0.75130719	0.33630207	-0.31243648	-0.0004918915	0.09791622
PERATEA	-0.73643837	-0.10812823	-0.31143655	-0.1595883016	-0.23391773
I.Hidro.	0.03559582	0.07732116	0.65567491	0.3746078268	0.34138362
QPC	-0.15416671	0.71147393	-0.07193374	-0.4773651917	0.10579709
ERAL	-0.40940171	0.30013328	-0.23737005	0.3253843476	0.53141236

Fonte: Autor (2023).

Após a redução da dimensionalidade dos dados, realizou-se uma nova análise PCA, apenas com as seis variáveis principais (SMT, PO, Pluv., IAU, ERAL, QPC), resultando em uma explicação de 51,5%, dos dados avaliados relacionados ao consumo *per capita* de água, por meio das componentes principais 1 e 2 (Gráfico 2). Por meio do Gráfico 2, é possível observar ainda uma possibilidade de maior relação entre algumas variáveis e o consumo per capita de água, possibilitando analisar entre as 6 variáveis, qual pode melhor explicar o QPC. Devido à quantidade elevada de variáveis estudadas, o Gráfico 2 não deixa clara as relações entre elas, por isso, realizou-se uma categorização das variáveis.

Gráfico 2 – Variáveis principais na PC1 e PC2.



Fonte: Autor (2023).

Essa categorização das variáveis (Quadro 5) teve como objetivo obter melhor classificação e descrição de cada variável, para a posterior análise da variável QPC - consumo *per capita* de água, objeto principal de análise do presente estudo, com as demais variáveis de cada categoria.

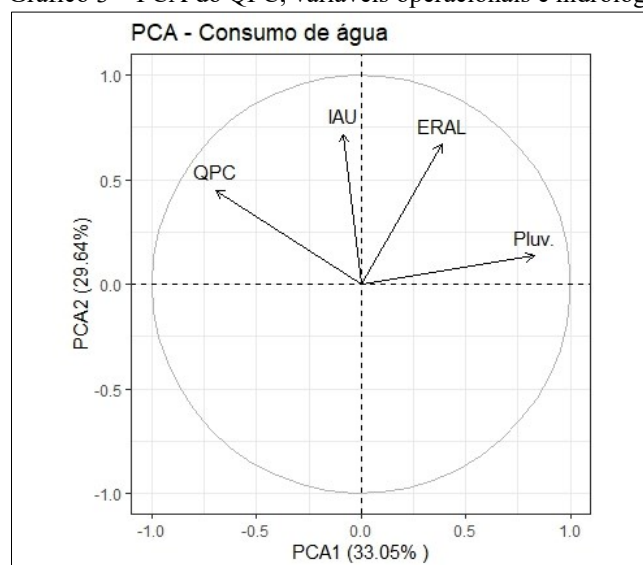
Quadro 5 - Categorização das variáveis principais da PCA.

Categorização das Variáveis		
Variável	Categoria	Sigla
SMT - Salário médio mensal dos trabalhadores formais	Econômica	ECO
PO - Pessoal ocupado	Econômica	ECO
Pluv. - Pluviometria quantis anual	Hidrológica	HIDRO
IAU - Índice de atendimento urbano de água	Operacional	OP
ERAL - Extensão da rede de água por ligação	Operacional	OP

Fonte: Autor (2023).

Tendo finalizado a categorização das variáveis, realizou-se uma nova análise PCA, do consumo *per capita* de água com as variáveis da categoria operacional e da categoria hidrológica. As duas primeiras PCs (PCA1 e PCA2) demonstraram uma explicação de 62,69% dos dados, como apresentado no Gráfico 3. É possível observar ainda por meio do Gráfico 3 que entre as variáveis operacionais e a variável hidrológica, a explicação das variáveis operacionais apresenta uma relação positiva com o QPC, ou seja, as variáveis operacionais IAU e ERAL explicam positivamente o crescimento do consumo per capita de água, à medida que essas duas variáveis crescem, o QPC tende a aumentar; por outro lado, como esperado, observa-se, que a variável hidrológica apresenta-se de forma oposta, demonstrando uma relação negativa com o QPC, representando assim, que quanto maior os índices pluviométricos o consumo *per capita* de água tende a diminuir.

Gráfico 3 – PCA do QPC, variáveis operacionais e hidrológica.

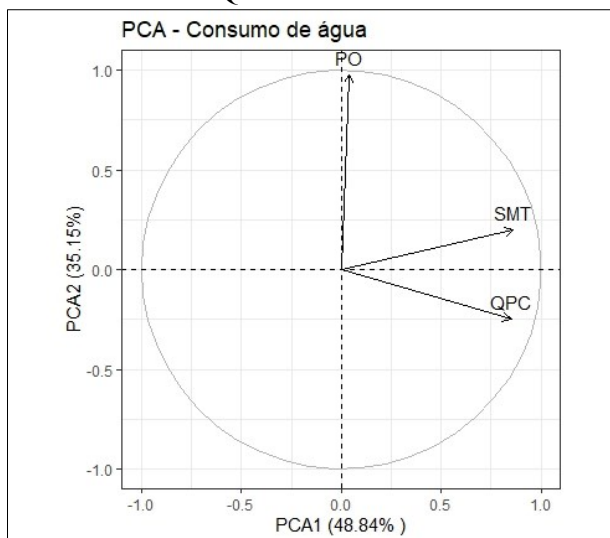


Fonte: Autor (2023).

Em seguida, uma outra análise PCA, desta vez para o QPC e as variáveis da categoria econômica, as duas primeiras PCs (PCA1 e PCA2) demonstraram uma explicação de 83,99% dos dados, como apresentado no Gráfico 4. Isso indica que as variáveis econômicas explicam melhor o consumo *per capita* desse conjunto de dados. Pode-se observar, que entre o pessoal ocupado (PO) e o salário médio mensal dos trabalhadores formais (SMT), o SMT possui uma

influência positiva no consumo *per capita* de água, ou seja, quanto maior o poder aquisitivo da população, conseqüentemente maior será o seu consumo *per capita* de água.

Gráfico 4 - PCA do QPC e variáveis econômicas.

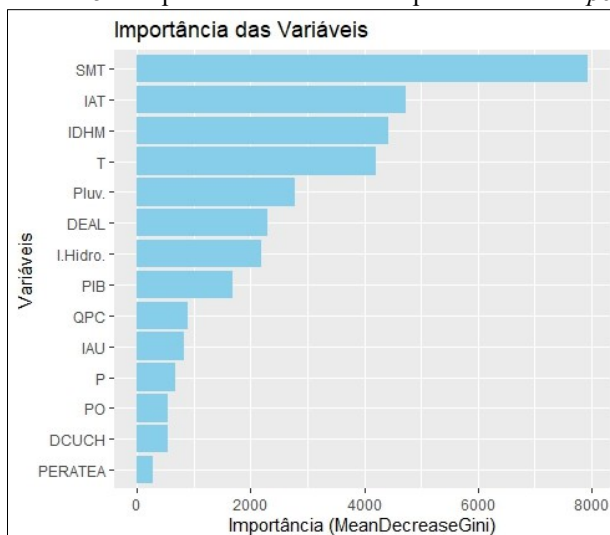


Fonte: Autor (2023).

3.2 Método *Random Forest* (RF)

Posteriormente, com a análise das variáveis por meio do método *random forest* (RF), obtiveram-se 5 variáveis com maiores graus de importância para descrever o conjunto de dados, foram elas: o SMT - Salário médio mensal dos trabalhadores formais, o IAT - Índice de atendimento total de água, o IDHM - Índice de desenvolvimento humano municipal, a T - Temperatura média e a Pluv. - Pluviometria quantis anual, como apresentado no Gráfico 5. É possível observar, que algumas das variáveis importantes apresentadas pelo método RF, coincidem com as apresentadas pelo método PCA, no entanto, chamando uma maior atenção para a variável SMT que demonstrou significativo grau de influência para o consumo *per capita* de água. A delimitação do destaque de apenas 5 variáveis com maiores graus de importância, foi realizada de forma arbitrária, com o intuito de equiparar com a quantidade de variáveis resultantes do método PCA.

Gráfico 5 – Importância das variáveis para o consumo *per capita* de água, a partir do método RF



Fonte: Autor (2023).

3.3 Método Análises de Regressão Linear, Logarítmica e Polinomial

Com todos os testes finalizados, foram realizadas as análises de regressão linear, logarítmica e polinomial para os dados resultantes dos dois métodos PCA e RF, com o objetivo de se obter um modelo matemático para a determinação estimada do QPC dos municípios estudados, dados as variáveis SMT, PO, PLUV, IAL, ERAL, IAT, IDHM, T.

Com os dados do método PCA obteve-se os seguintes modelos matemáticos:

- Equação da Regressão Linear:

$$QPC = 22,5014 + (85,5299 \times SMT) + (-0,0017 \times PO) + (-0,0261 \times Pluv.) + (-0,3489 \times IAU) + (-0,216 \times ERAL) \quad (\text{Eq. 02})$$

- Equação da Regressão Logarítmica:

$$QPC = 382,6353 + [275,9347 \times \log(SMT)] + [-12,3188 \log(PO)] + [-29,8166 \log(Pluv.) + [-105,6354 \times \log(IAU)] + [-2,6246 \times \log(ERAL)] \quad (\text{Eq. 03})$$

- Equação da Regressão Polinomial:

$$QPC = 96,92 - (1095,8451 \times SMT) + (1420,6965 \times SMT^2) - (965,802 \times SMT^3) + (729,2425 \times SMT^4) - (480,603 \times SMT^5) - (270,297 \times PO) + (504,6504 \times PO^2) + (420,7333 \times PO^3) + (373,2073 \times PO^4) - (25,172 \times Pluv.) - (9,0284 \times Pluv.^2) + (80,4912 \times Pluv.^3) + (1522,6334 \times IAU) - (1761,8786 \times IAU^2) + (245,0703 \times IAU^3) + (7,935 \times ERAL) + (61,6101 \times ERAL^2) + (2,6937 \times ERAL^3) - (54,5554 \times ERAL^4) - (35,6223 \times ERAL^5) \quad (\text{Eq. 04})$$

Nas quais:

QPC é o Consumo médio *per capita* de água (l/hab.dia); SMT é o Salário médio mensal dos trabalhadores formais (salários mínimos); PO é o Pessoal ocupado (pessoas); Pluv. é a Pluviometria quantis anual (mm); IAU é o Índice de atendimento urbano de água (%) e ERAL é a Extensão da rede de água por ligação (m/lig.).

Em seguida, utilizou-se dois modelos estatísticos de análise de erro, o modelo do Erro Quadrático Médio (MSE - da sigla em inglês *Mean Squared Error*) e o modelo Coeficientes R_squared (R^2 - R quadrado), com o objetivo de verificar o grau de ajuste do conjunto de dados a cada um dos três modelos matemáticos obtidos com os dados da PCA, como apresentado na Tabela 2. Vale destacar que na análise do erro por meio do método R^2 para a equação polinomial, os expoentes destas equações foram ajustados para que o R^2 fosse maior que 0,9.

Tabela 2 - Análise da margem de erro dos modelos matemáticos – PCA.

MSE para cada modelo:	Coefficientes e R^2 para cada modelo
Linear: 490,1503	Linear: 22,5014 e 0,3003
Logarítmico: 507,6566	Logarítmico: 85,5299 e 0,2753
Polinomial: 0,3109	Polinomial: - 0,0017 e 0,9995

Fonte: Autor (2023).

Diante dos resultados obtidos das análises de margem de erro, para garantir um grau de confiabilidade satisfatório, é importante que os modelos apresentem a menor margem de erro possível, tratando-se especificamente do modelo do erro quadrático médio (MSE), ou seja, quanto maior o desvio/valor resultante do MSE, menos preciso é o modelo matemático em questão, dessa forma, observa-se que a margem de erro menos significativa foi a da equação da regressão polinomial, com um erro de 0,3109. Tratando-se do modelo R^2 para garantir um nível de confiabilidade de 90% as equações necessitariam apresentar uma margem de erro aproximadamente de 0,90, pois, no método do R^2 quanto mais próximo de 1 o valor, maior o seu grau de confiabilidade, com isso, é possível observar que equação da regressão polinomial se sobressaiu apresentando um valor de 0,9995 para a análise do R^2 .

Na Tabela 3 é possível observar o cálculo do consumo *per capita* de água (QPC) com cada um dos modelos matemáticos obtidos, utilizando os dados do Salário médio mensal dos trabalhadores formais, do Pessoal ocupado, da Pluviometria quantis anual, do Índice de atendimento urbano de água e da Extensão da rede de água por ligação, de cada um dos 22 municípios analisados, lembrando que o município de Lagoa Nova foi desconsiderado nas análises, como explicado no início do subtópico (3.1).

Tabela 3 – Cálculo do consumo *per capita* de água dos 22 municípios com os modelos matemáticos da PCA.

Município	SMT	PO	Pluv.	IAU	ERAL	QPC Real	QPC Calc. Linear	Erro Abs. Linear	Erro Perc. Linear	QPC Calc. logaritmico	Erro Abs. logaritmico	Erro Perc. logaritmico	QPC Calc. Polinomial	Erro Abs. Polinomial	Erro Perc. Polinomial
Acari	1,50	1291,00	601,40	100,00	14,20	83,72	94,95	11,23	13,41	91,80	8,08	9,65	83,16	0,56	0,67
Bodo	1,90	421,00	619,80	100,00	17,05	122,19	129,53	7,34	6,01	125,25	3,06	2,51	122,19	0,00	0,00
Caicó	1,70	12748,00	866,10	100,00	12,76	99,80	86,14	13,66	13,69	90,11	9,69	9,71	99,80	0,00	0,00
Carnaúba dos Dantas	1,30	1399,00	855,50	100,00	11,89	94,03	71,53	22,50	23,93	70,13	23,90	25,42	94,03	0,00	0,00
Cerro Cora	1,80	943,00	909,00	100,00	24,78	90,44	110,88	20,44	22,60	108,52	18,08	19,99	90,54	0,10	0,11
Cruzeta	1,50	1060,00	612,90	100,00	9,17	85,39	96,13	10,74	12,58	93,76	8,37	9,80	86,15	0,76	0,89
Currais Novos	1,70	6557,00	711,50	84,83	10,50	81,28	106,40	25,12	30,90	104,27	22,99	28,29	81,28	0,00	0,00
Equador	1,40	405,00	445,00	98,62	9,25	73,34	93,53	20,19	27,53	95,40	22,06	30,08	73,34	0,00	0,00
Florânia	1,50	896,00	871,90	100,00	28,93	79,20	85,37	6,17	7,79	87,08	7,88	9,95	79,72	0,52	0,65
Ipueira	1,70	284,00	240,20	100,00	15,72	113,27	122,86	9,59	8,46	126,52	13,25	11,70	113,26	0,01	0,00
J. de Piranhas	1,50	1448,00	1096,80	100,00	11,29	85,25	82,38	2,87	3,37	84,01	1,24	1,46	85,21	0,04	0,05
J. do Seridó	1,50	1915,00	682,00	100,00	12,22	88,81	92,22	3,41	3,84	88,46	0,35	0,40	88,17	0,64	0,72
Jucurutu	1,60	2508,00	823,80	100,00	9,18	88,71	96,73	8,02	9,04	93,05	4,34	4,90	89,10	0,39	0,44
Ouro Branco	1,50	564,00	866,40	100,00	8,83	102,66	90,42	12,24	11,92	92,75	9,91	9,65	101,67	0,99	0,97
Parelhas	1,50	4226,00	750,90	100,00	11,22	83,20	86,74	3,54	4,26	83,20	0,00	0,00	83,17	0,03	0,03
Santana do Seridó	1,60	513,00	757,70	100,00	26,79	80,80	98,01	17,21	21,30	99,82	19,02	23,53	79,37	1,43	1,77
S. Fernando	1,70	587,00	911,90	100,00	53,53	94,42	96,63	2,21	2,34	102,14	7,72	8,18	94,42	0,00	0,00
S. J. do Sabugi	1,70	468,00	1036,60	100,00	11,21	99,86	102,73	2,87	2,87	105,80	5,94	5,95	99,87	0,01	0,01
S. J. do Seridó	1,20	1240,00	759,10	72,64	15,08	89,56	74,62	14,94	16,69	76,77	12,79	14,28	89,56	0,00	0,00
S. Vicente	1,60	780,00	663,60	100,00	12,79	88,57	103,05	14,48	16,34	101,23	12,66	14,30	89,61	1,04	1,18
Tenente L. Cruz	1,80	648,00	522,60	100,00	17,98	207,02	122,93	84,09	40,62	118,54	88,48	42,74	206,92	0,10	0,05
T. dos Bastistas	1,50	326,00	825,20	100,00	24,60	100,72	88,48	12,24	12,15	93,63	7,09	7,04	101,70	0,98	0,98
Média						96,92	96,92	14,78	14,17	96,92	13,95	13,16	96,92	0,35	0,39

Fonte: Autor (2023).

Observando a Tabela 3, é possível perceber que o modelo matemático mais preciso para os dados da PCA, ou seja, o que apresentou uma menor margem de erro, é o modelo polinomial, com uma média de erro absoluto de 0,35 e uma média de erro percentual de 0,39%; por outro lado, é importante destacar que apesar do modelo ter se mostrado satisfatório, este apresenta considerável dificuldade de utilização por conter muitos termos na equação.

Quanto a análise da regressão linear, logarítmica e polinomial para os dados resultantes do método RF, obteve-se os seguintes modelos matemáticos:

- Equação da Regressão Linear:

$$QPC = -12,3839 + (83,8533 \times SMT) + (-99,4717 \times IDHM) + (-0,0355 \times Pluv.) + (3,0717 \times T) + (-0,251 \times IAT) \quad (\text{Eq. 05})$$
- Equação da Regressão Logarítmica:

$$QPC = -38,3978 + [290,3361 \times \log(SMT)] + [-141,9028 \times \log(IDHM)] + [-49,9904 \times \log(Pluv.)] + [222,0355 \times \log(T)] + [-29,3871 \times \log(IAT)] \quad (\text{Eq. 06})$$
- Equação da Regressão Polinomial:

$$QPC = 104,0317 + [217,8161 \times (SMT)] - [112,0023 \times (SMT)^2] - [361,8785 \times (SMT)^3] - [125,7557 \times (SMT)^4] - [168,4735 \times (SMT)^5] - [29,2054 \times (IDHM)] - [14,0202 \times (Pluv.)] - [19,8063 \times (T)] - [199,3326 \times (IAT)] - [29,2054 \times (IDHM)] - [14,0202 \times (Pluv.)] -$$

$$[19,8063 \times (T)] - [199,3326 \times (IAT)] + 320,7778 \times (IAT)^2 - [346,9151 \times (IAT)^3] + [293,0078 \times (IAT)^4] - [123,9035 \times (IAT)^5] + [320,7778 \times (IAT)^2] - [346,9151 \times (IAT)^3] + [293,0078 \times (IAT)^4] - [123,9035 \times (IAT)^5] \quad (\text{Eq. 07})$$

Nas quais:

QPC é o Consumo médio *per capita* de água (l/hab.dia); SMT é o Salário médio mensal dos trabalhadores formais (salários mínimos); IAT é o Índice de atendimento total de água (%), IDHM é o Índice de desenvolvimento humano municipal (IDHM); T é a Temperatura média (°C) e Pluv. é a Pluviometria quantis anual (mm).

Em seguida, realizou-se também a análise da margem de erro, com o modelo do Erro Quadrático Médio (MSE) e o modelo Coeficientes R-quadrado (R^2), com o objetivo de verificar o grau de ajuste dos três modelos matemáticos, obtidos com os dados do método RF, como apresentado na Tabela 4.

Tabela 4 - Análise da margem de erro dos modelos matemáticos – RF.

MSE para cada modelo:	R^2 para cada modelo
Linear: 468,7395	Linear: 0,3309
Logarítmico: 478,3224	Logarítmico: 0,3172
Polinomial: 43,9353	Polinomial: 0,9373

Fonte: Autor (2023).

Ante os resultados obtidos das análises de margem de erro para os modelos do método RF, como mencionado anteriormente, com o objetivo de garantir um grau de confiabilidade satisfatório, é importante que os modelos apresentem a menor margem de erro possível tratando-se especificamente do modelo do erro quadrático médio (MSE), ou seja, quanto maior o desvio/valor resultante do MSE, menos preciso é o modelo matemático em questão, para o método RF, observa-se que a margem de erro menos significativa foi também a da equação da regressão polinomial, com um erro de 43,9353. Tratando-se do modelo R^2 é possível observar que equação da regressão polinomial se sobressaiu novamente apresentando um valor de 0,9373 para a análise do R^2 , ou seja, demonstrando um considerável grau de confiabilidade, pois quanto mais próximo de 1 o valor, maior o seu grau de confiabilidade.

Na Tabela 5 é possível observar o cálculo do consumo *per capita* de água (QPC) com cada um dos modelos matemáticos obtidos, utilizando os dados do Salário médio mensal dos trabalhadores formais, do Índice de desenvolvimento humano municipal, da Pluviometria quantis anual, da Temperatura média e do Índice de atendimento total de água de cada um dos 22 municípios analisados.

Observando ainda a Tabela 5, é possível perceber que o modelo matemático mais preciso para os dados do método RF, ou seja, o que apresentou uma menor margem de erro, é também o modelo polinomial, com uma média de erro absoluto de 5,58 e uma média de erro percentual de 6,08%, por outro lado, apresentando também significativa dificuldade de utilização por conter muitos termos na equação.

Tabela 5 - Cálculo do consumo *per capita* de água dos 22 municípios com os modelos matemáticos do método RF.

Município	SMT	IDHM	Pluv.	T	IAT	QPC Real	QPC Calc. Linear	Erro Abs. Linear	Erro Perc. Linear	QPC Calc. logaritmico	Erro Abs. logaritmico	Erro Perc. logaritmico	QPC Calc. Polinomial	Erro Abs. Polinomial	Erro Perc. Polinomial
Acari	1,50	0,68	601,40	27,50	90,98	83,72	86,15	2,43	2,90	84,68	0,96	1,15	89,40	5,68	6,78
Bodo	1,90	0,63	619,80	26,60	71,14	122,19	126,23	4,04	3,30	122,57	0,38	0,31	122,56	0,37	0,31
Caicó	1,70	0,71	866,10	27,66	100,00	99,80	88,67	11,13	11,15	87,58	12,22	12,25	93,09	6,71	6,72
Carnaúba dos Dantas	1,30	0,66	855,50	28,40	83,15	94,03	67,09	26,94	28,65	66,58	27,45	29,19	95,86	1,83	1,95
Cerro Cora	1,80	0,61	909,00	24,72	66,58	90,44	105,15	14,71	16,26	104,51	14,07	15,56	93,86	3,42	3,78
Cruzeta	1,50	0,65	612,90	31,10	92,83	85,39	98,82	13,43	15,73	97,85	12,46	14,60	81,76	3,63	4,25
Currais Novos	1,70	0,69	711,50	31,64	75,99	81,28	114,30	33,02	40,63	114,55	33,27	40,93	92,42	11,14	13,71
Equador	1,40	0,62	445,00	28,10	82,77	73,34	92,78	19,44	26,51	92,69	19,35	26,38	67,45	5,89	8,03
Florânia	1,50	0,64	871,90	28,22	88,14	79,20	83,16	3,96	5,00	83,50	4,30	5,42	88,99	9,79	12,36
Igueira	1,70	0,68	240,20	28,11	100,00	113,27	115,33	2,06	1,82	119,73	6,46	5,70	119,26	5,99	5,29
J. de Piranhas	1,50	0,60	1096,80	28,52	83,11	85,25	81,25	4,00	4,69	85,12	0,13	0,15	93,81	8,56	10,04
J. do Seridó	1,50	0,66	682,00	33,07	100,00	88,81	99,72	10,91	12,29	98,43	9,62	10,83	93,73	4,92	5,54
Jucurutu	1,60	0,60	823,80	30,06	83,68	88,71	104,10	15,39	17,35	104,55	15,84	17,86	90,93	2,22	2,50
Ouro Branco	1,50	0,65	866,40	29,93	86,81	102,66	88,64	14,02	13,65	89,47	13,19	12,85	87,88	14,78	14,40
Parelhas	1,50	0,68	750,90	28,44	98,28	83,20	82,20	1,00	1,21	81,11	2,09	2,51	85,17	1,97	2,37
Santana do Seridó	1,60	0,64	757,70	28,34	71,36	80,80	100,18	19,38	23,98	101,30	20,50	25,37	71,95	8,85	10,95
S. Fernando	1,70	0,61	911,90	30,20	99,97	94,42	105,00	10,58	11,21	104,50	10,08	10,67	99,90	5,48	5,80
S. J. do Sabugi	1,70	0,66	1036,60	29,33	100,00	99,86	93,23	6,63	6,64	94,30	5,56	5,57	91,01	8,85	8,87
S. J. do Seridó	1,20	0,69	759,10	33,13	57,32	89,56	79,65	9,91	11,06	81,68	7,88	8,80	89,32	0,24	0,27
S. Vicente	1,60	0,64	663,60	29,90	88,05	88,57	104,11	15,54	17,55	103,17	14,60	16,48	84,10	4,47	5,05
Tenente L. Cruz	1,80	0,62	522,60	29,78	77,07	207,02	130,16	76,86	37,12	128,58	78,44	37,89	201,12	5,90	2,85
T. dos Batistas	1,50	0,64	825,20	29,61	100,00	100,72	86,31	14,41	14,31	85,81	14,91	14,80	98,67	2,05	2,03
Média						96,92	96,92	14,99	14,68	96,92	14,72	14,33	96,92	5,58	6,08

Fonte: Autor (2023).

Diante de todos os resultados obtidos, o modelo e o método que apresentou uma maior precisão para o conjunto de dados analisados, para os 22 municípios estudados, foi o método PCA e o modelo matemático por meio da regressão linear polinomial.

Dessa forma, as variáveis do conjunto de dados em questão que melhor explicam o consumo *per capita* de água desses municípios consistem em variáveis enquadradas nos contextos econômicos, hidro climáticos e operacionais, reforçando com isso diversos estudos realizados e apresentados por autores como Azevedo Netto e Fernández (2018, p. 408) e Heller e De Pádua (2010, p. 133-138).

Com isso, o presente estudo torna-se uma forma de subsídio para a estimativa do consumo *per capita* de água com base nas variáveis principais apresentadas, possibilitando assim análises, comparações e estudos, para questões importantes como a gestão dos recursos hídricos, planejamento e dimensionamento para a disponibilização de água de forma emergencial para determinada localidade da região estudada.

Silva *et al.* (2008), destaca que a gestão de recursos hídricos consiste na forma com a qual pretende-se mensurar e resolver questões relacionadas a escassez dos recursos hídricos, assim como incentivar e fazer o uso adequado deste recurso tão essencial para a vida, objetivando uma otimização em benefício de toda a sociedade.

4 CONSIDERAÇÕES FINAIS

Mediante o exposto no desenvolvimento do presente trabalho, tornou-se possível observar características do consumo *per capita* de água (QPC) dos 23 municípios atendidos pela regional Seridó, destacando-se a considerável variação do QPC do município de Lagoa Nova, o qual apresentou-se bem acima da média da região, suscitando a possibilidade do desenvolvimento de estudos pontuais para analisar o QPC do município.

Com o auxílio de análises de ciência dos dados, verificou-se que para a região estudada e fatores considerados na análise, os fatores que mais explicaram o consumo *per capita* de água,

ou seja, aqueles que mais interferem no QPC são os fatores considerados de características econômicas, hidrológicas e operacionais, convergindo de forma satisfatória com diversos estudos registrados na literatura.

Para o estudo em questão os fatores/variáveis que apresentaram maior significância foram: o SMT - Salário médio mensal dos trabalhadores formais, o PO - Pessoal ocupado, a Pluv. - Pluviometria quantis anual, o IAU - Índice de atendimento urbano de água, a ERAL - Extensão da rede de água por ligação e o QPC - Consumo médio per capita de água, sendo as 5 primeiras consideradas variáveis independentes, e a última (QPC) considerada variável dependente.

Ante as análises por meio dos métodos estatísticos PCA e RF e com o auxílio do software R Studio, além de analisar quais fatores mais influenciam no QPC da região, foram obtidos modelos matemáticos que possibilitaram a estimativa do QPC. O modelo que apresentou o maior grau de confiabilidade foi a equação da regressão polinomial, obtida com os resultados da análise PCA, a qual depende dos seguintes parâmetros: SMT, PO, Pluv., IAU e ERAL. A equação polinomial, apesar de dispor de um significativo grau de confiabilidade, por apresentar muitos termos, consequentemente possui uma certa dificuldade de utilização.

Contudo, pode-se perceber que estudos mais detalhados sobre os comportamentos e características das populações, que impactam diretamente no consumo *per capita* de água, são relevantes, servindo de subsídio para as análises e tomadas de decisões quanto aos planejamentos e estratégias da gestão dos recursos hídricos e da sustentabilidade de uma determinada região.

REFERÊNCIAS

- ACS CAERN. A Companhia. **Companhia de Águas e Esgotos do Rio Grande do Norte**, 2020. Disponível em: <<http://adcon.rn.gov.br/ACERVO/caern/Conteudo.asp?TRAN=ITEM&TARG=496&ACT=&PAGE=0&PARM=&LBL=A+Caern>>. Acesso em: 16 ago. 2023.
- AMARATUNGA Dhammika; CABRERA, Javier; LEE, Yung-Seop, Florestas aleatórias enriquecidas. **Bioinformática**, V. 24, Ed. 18, setembro de 2008, p. 2010–2014. Disponível em: <https://doi.org/10.1093/bioinformatics/btn356>. Acesso em: 27 set. 2023.
- CORREIA, Rebert Coelho *et al.* A região semiárida brasileira. In: VOLTOLINI, T. V. (Ed.). **Produção de caprinos e ovinos no Semiárido**. Petrolina: Embrapa Semiárido, 2011. cap. 1, p. 21-48. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/54762/1/01-A-regiao-semiarida-brasileira.pdf-18-12-2011.pdf>>. Acesso em: 16 ago. 2023.
- DA CONCEIÇÃO, Lia Souto Manhães. **A utilização do random forest para prever a inflação**. 2022. p. 38. Monografia de final de curso - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2022. Disponível em: https://www.econ.puc-rio.br/uploads/adm/trabalhos/files/Lia_Souto_Manhaes_da_Conceicao_Mono_22.2.pdf. Acesso em: 27 set. 2023.
- EMPRESA DE PESQUISA AGROPECUÁRIA DO RN – EMPARN. **Gráfico pluviométrico (Quantis Anual) – 2021a**. Disponível em: <<http://meteorologia.emparn.rn.gov.br/graficos>>. Acesso em: 16 ago. 2023.
- EMPARN. **Relatório de variáveis meteorológicas – 2021b**. Disponível em: <<https://meteorologia.emparn.rn.gov.br/relatorios/relatorio-variaveis>>. Acesso em: 16 ago. 2023.
- FERNANDES NETO, Maria de Lourdes. **Avaliação de parâmetros intervenientes no consumo per capita de água**: estudo para 96 municípios do estado de minas gerais. 146 f. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia da Universidade Federal de Minas Gerais, Belo Horizonte, 2003. Disponível em: <<https://www.smarh.eng.ufmg.br/defesas/68M.PDF>>. Acesso em: 09 ago. 2023.
- GOOGLE. Google Earth. 2019.
- HELLER, Léo; DE PÁDUA, Valter Lúcio. **Abastecimento de água para consumo humano**. 2. ed. rev. e atual.- Belo Horizonte: Editora UFMG, 2010. ISBN: 978-85-7041-841-8 (v. 1).
- HONGYU, Kuang; SANDANIELO, Vera Lúcia Martins; OLIVEIRA JUNIOR, Gilmar Jorge de. **Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação**. E&S Engineering and Science, [S. l.], v. 5, n. 1, p. 83–90, 2016. DOI: 10.18607/ES201653398. Disponível em: <https://periodicoscientificos.ufmt.br/ojs/index.php/eng/article/view/3398>. Acesso em: 25 set. 2023.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Portal cidades@**. Rio de Janeiro: IBGE, 2022. Disponível em: <<https://cidades.ibge.gov.br/>>. Acesso em: 16 ago. 2023.
- JOLLIFFE, Ian; CADIMA, Jorge. **Análise de componentes principais: uma revisão e desenvolvimentos recentes**. Philos Trans A Math Phys Eng Sci. 13 de abril de 2016; 374(2065):20150202. Disponível em: <<http://doi.org/10.1098/rsta.2015.0202>>. Acesso em: 24 set. 2023.
- LUCIANO, Thiago Lehmkuhl. **Aplicação de ferramentas quimiométricas para avaliação da concentração de metais em resíduos da indústria de óleo e gás**. 2021. p. 51. Monografia (Graduação em Química) - Universidade Federal de Santa Catarina, Florianópolis, 2021. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/220001/TCC.pdf?sequence=1>. Acesso em: 14 set. 2023.
- MACEDO, Yuri Marques; TROLEIS, Adriano Lima; DE ALMEIDA, Lutiane Queiroz. Risco de Desabastecimento Hídrico na Região Seridó, Rio Grande do Norte, Brasil. **Revista Brasileira de Geografia Física**, v. 14, n. 06, p. 3711-3735, 2021. Disponível em: <https://encurtador.com.br/wzHIP>. Acesso em: 17 jul. 2023.
- AZEVEDO NETTO, José Martiniano de; FERNÁNDEZ, Miguel Fernández Y. **Manual de hidráulica**. 9. ed. São Paulo: Blucher, 2018. 632 p.; PDF. E-book. ISBN 978-85-212-0889-1. Disponível em: <<https://app.minhabiblioteca.com.br/#/books/9788521208891/>>. Acesso em: 08 ago. 2023.

PEARSON, Karl. (1901). «On Lines and Planes of Closest Fit to Systems of Points in Space» (PDF). *Philosophical Magazine*. 2 (6): 559–572.

PNSH - PLANO NACIONAL DE SEGURANÇA HÍDRICA. Brasília: **Agência Nacional de Águas**. ANA, 2019. ISSN ISBN: 978-85-8210-059-2. Disponível em: <https://encurtador.com.br/EGJN8>. Acesso em: 5 jul. 2023.

PORTELA, Raphael Ricardo de Jesus; COSTA, Reginaldo B. da; CORREIA, Jonas Benevides. Análise das perdas nos sistemas de abastecimento de água nos municípios do Mato Grosso do Sul. **Geofronter**, [S. l.], v. 9, n. 1, 2023. Disponível em: <https://periodicosonline.uems.br/index.php/GEOF/article/view/7570>. Acesso em: 11 jul. 2023.

R BEAGLE SCOUTS. R: The R project for statistical somputing. **R version 4.3.1**, 16-06-2023. Disponível em: <https://www.r-project.org/>. Acesso em: 24 set. 2023.

SILVA, Gustavo Maciel da *et al.* Análise per capita do abastecimento de água no município de Soledade (Rio Grande do Sul). **Revista Estudo & Debate**, [S.l.], v. 27, n. 2, jul. 2020. ISSN 1983-036X. Disponível em: <<http://www.univates.br/revistas/index.php/estudoedebate/article/view/2541>>. Acesso em: 09 ago. 2023. doi:<http://dx.doi.org/10.22410/issn.1983-036X.v27i2a2020.2541>.

SILVA, Welitom Ttatom Pereira da *et al.* Quota per capita de água, fatores intervenientes e modelagem: estudo de caso para classes socioeconômicas de Cuiabá-MT. **Sociedade & Natureza**, v. 20, n. 2, p. 219-230, 2008. Disponível em: <https://www.scielo.br/j/sn/a/yDf5LtjmgJg9mFFvp8ZscRN/abstract/?lang=pt#>. Acesso em: 27 set. 2023.

SNIS - Sistema Nacional de Informações sobre Saneamento. **Diagnóstico temático serviços de água e esgoto**. Brasília, 2022. Disponível em: <https://www.gov.br/mdr/pt-br/assuntos/saneamento/snis/produtos-do-snis/diagnosticos>. Acesso em: 5 jul. 2023.

APÊNDICE

Variáveis, siglas e dados coletados para análise, dos 23 municípios estudados.

INDICADORES OPERACIONAIS - ÁGUA										
Município	Índice de atendimento total de água (%)	Índice de atendimento urbano de água (%)	Densidade de economias de água por ligação (econ./lig.)	Participação das economias residenciais de água no total das economias de água (%)	Índice de macromedicação (%)	Índice de hidromedicação (%)	Índice de micromedicação relativo ao volume disponibilizado (%)	Índice de micromedicação relativo ao consumo (%)	Índice de consumo de água (%)	Volume de água disponibilizado por economia (m³/mês/econ)
Sigla >	IAT	IAU	DEAL	PERATEA	I.Macro.	I.Hidro.	I.Micro.RVD	IMicro.RC	ICA	MADE
Acari	90,98	100,00	1,01	95,47	100,00	99,33	66,90	99,10	67,50	12,65
Bodó	71,14	100,00	1,00	94,05	60,77	99,03	69,31	99,07	69,96	17,43
Caicó	100,00	100,00	1,10	91,38	99,68	99,63	57,64	99,47	57,95	15,94
Carnaúba dos Dantas	83,15	100,00	1,01	97,57	94,69	95,23	60,50	94,48	64,04	15,14
Cerro Corá	66,58	100,00	1,01	96,73	66,12	91,93	65,79	91,57	71,84	13,74
Cruzeta	92,83	100,00	1,02	94,98	83,93	87,23	56,01	83,87	66,78	12,63
Currais Novos	75,99	84,83	1,00	95,10	77,96	96,49	50,07	95,64	52,35	15,36
Equador	82,77	98,62	1,01	95,84	100,00	99,15	72,65	99,82	72,78	10,41
Florânia	88,14	100,00	1,01	95,02	78,56	98,10	66,11	97,57	67,76	11,88
Ipueira	100,00	100,00	1,00	95,96	100,00	100,00	78,34	99,97	78,37	12,54
Jardim de Piranhas	83,11	100,00	1,02	94,44	86,82	96,88	40,09	96,33	41,61	21,74
Jardim do Seridó	100,00	100,00	1,03	93,35	93,24	99,85	62,65	99,82	62,76	13,27
Jucurutu	83,68	100,00	1,02	94,97	84,70	97,56	67,94	97,27	69,85	12,34
Lagoa Nova	48,68	88,96	1,05	96,06	30,12	93,97	67,05	96,86	69,22	41,73
Ouro Branco	86,81	100,00	1,00	95,43	90,40	100,00	54,20	100,00	54,20	18,60
Parelhas	98,28	100,00	1,02	93,86	100,00	95,13	51,80	93,76	55,25	15,15
Santana do Seridó	71,36	100,00	1,00	96,81	91,43	97,78	51,92	97,80	53,09	16,37
São Fernando	99,97	100,00	1,01	96,68	100,00	99,91	68,11	99,80	68,25	14,45
São João do Sabugi	100,00	100,00	1,01	96,59	80,82	95,37	57,99	94,99	61,05	15,75
São José do Seridó	57,32	72,64	1,02	95,59	98,06	94,46	80,34	93,16	86,24	10,29
São Vicente	88,05	100,00	1,00	95,82	65,18	99,94	66,65	99,93	66,69	12,87
Tenente Laurentino Cruz	77,07	100,00	1,00	95,44	56,56	99,96	67,50	100,00	67,50	31,17
Timbaúba dos Batistas	100,00	100,00	1,00	93,71	100,00	100,00	78,35	100,00	78,35	12,52

INDICADORES OPERACIONAIS - ÁGUA											
Município	Consumo médio de água por economia (m³/mês/econ)	Consumo micromedido por economia (m³/mês/econ)	Consumo de água faturado por economia (m³/mês/econ)	Consumo médio per Capita de água (l/hab.dia)	Índice de consumo de energia elétrica em sistemas de abastecimento de água (kWh/m³)	Extensão da rede de água por ligação (m/lig.)	Índice de faturamento de água (%)	Índice de perdas faturamento (%)	Índice de perdas na distribuição (%)	Índice bruto de perdas lineares (m³/dia/km)	Índice de perdas por ligação (l/dia/lig.)
Sigla >	CMed.AE	CMicro.E	CAFE	QPC	ICEESAA	ERAL	IFA	IPF	IPD	IBPL	IPL
Acari	8,13	8,13	11,18	83,72	0,49	14,20	92,77	7,23	32,50	6,73	130,28
Bodó	12,20	12,20	16,57	122,19	18,72	17,05	95,05	4,95	30,04	7,61	172,15
Caicó	9,19	9,28	12,65	99,80	0,57	12,76	79,76	20,24	42,05	15,97	241,34
Carnaúba dos Dantas	9,21	9,14	12,00	94,03	1,62	11,89	83,43	16,57	35,96	11,77	171,44
Cerro Corá	9,50	9,46	13,29	90,44	0,99	24,78	100,48	-0,48	28,16	4,25	123,86
Cruzeta	8,02	7,74	11,27	85,39	0,44	9,17	93,87	6,13	33,22	11,92	133,22
Currais Novos	8,01	7,94	11,62	81,28	0,73	10,50	75,93	24,07	47,65	16,73	239,94
Equador	7,19	7,25	11,09	73,34	0,45	9,25	112,18	-12,18	27,22	7,54	89,26
Florânia	8,05	8,03	11,73	79,20	4,96	28,93	98,74	1,26	32,24	3,55	127,66
Ipueira	9,34	9,34	12,48	113,27	0,29	15,72	104,77	-4,77	21,63	5,00	84,84
Jardim de Piranhas	8,57	8,55	11,97	85,25	1,97	11,29	58,12	41,88	58,39	29,51	402,74
Jardim do Seridó	7,91	7,95	11,55	88,81	0,09	12,22	91,60	8,40	37,24	11,34	159,51
Jucurutu	8,59	8,61	11,93	88,71	0,40	9,18	96,99	3,01	30,15	11,38	124,38
Lagoa Nova	28,89	30,77	32,48	260,49	6,05	23,96	77,84	22,16	30,78	12,15	443,43
Ouro Branco	10,08	10,08	13,97	102,66	1,37	8,83	75,13	24,87	45,80	21,49	281,30
Parelhas	7,95	7,88	11,57	83,20	0,46	11,22	80,44	19,56	44,75	15,46	216,80
Santana do Seridó	8,26	8,26	11,75	80,80	2,57	26,79	75,55	24,45	46,91	6,10	240,20
São Fernando	9,42	9,41	12,60	94,42	1,56	53,53	91,31	8,69	31,75	2,54	144,95
São João do Sabugi	9,18	9,15	12,32	99,86	0,71	11,21	81,92	18,08	38,95	15,36	193,58
São José do Seridó	8,87	8,75	12,40	89,56	1,03	15,08	120,54	-20,54	13,76	1,66	47,23
São Vicente	8,58	8,59	12,82	88,57	4,47	12,79	99,64	0,36	33,31	9,68	141,07
Tenente Laurentino Cruz	21,04	21,05	25,38	207,02	1,58	17,98	81,41	18,59	32,50	13,79	333,86
Timbaúba dos Batistas	9,40	9,40	12,68	100,72	2,60	24,60	105,63	-5,63	21,65	3,22	85,77

DADOS SOCIOAMBIENTAIS E CLIMÁTICOS								
Município	Dados do Partal Cidades - IBGE					Dados EMPARN		Dados <i>Google Earth</i>
	População (hab.)	Salário médio mensal dos trabalhadores formais (salários mínimos)	Pessoal ocupado (pessoas)	PIB per capita (R\$)	Índice de Desenvolvimento Humano Municipal (IDHM)	Pluviometria quantis anual (mm)	Temperatura média (°C)	Distância do centro urbano até o corpo hídrico mais importante (m)
Sigla >	P	SMT	PO	PIB	IDHM	Pluv.	T(°C)	DCUCH
Acari	10597,00	1,50	1291,00	11857,64	0,679	601,40	27,50	5062,18
Bodó	2306,00	1,90	421,00	119380,48	0,63	619,80	26,60	1655,70
Caicó	61146,00	1,70	12748,00	18883,59	0,71	866,10	27,66	5624,73
Carnaúba dos Dantas	7990,00	1,30	1399,00	13425,32	0,66	855,50	28,40	1774,00
Cerro Corá	11000,00	1,80	943,00	11101,65	0,61	909,00	24,72	886,64
Cruzeta	8005,00	1,50	1060,00	14208,18	0,65	612,90	31,10	946,51
Currais Novos	41318,00	1,70	6557,00	16670,54	0,69	711,50	31,64	3240,11
Equador	5360,00	1,40	405,00	11572,03	0,62	445,00	28,10	2967,80
Florânia	10196,00	1,50	896,00	11186,21	0,64	871,90	28,22	492,73
Ipueira	2035,00	1,70	284,00	13322,16	0,679	240,20	28,11	684,45
Jardim de Piranhas	13977,00	1,50	1448,00	12314,12	0,60	1096,80	28,52	1063,03
Jardim do Seridó	11655,00	1,50	1915,00	15171,02	0,66	682,00	33,07	1024,25
Jucurutu	17793,00	1,60	2508,00	18522,99	0,60	823,80	30,06	787,41
Lagoa Nova	15573,00	2,00	1498,00	15608,51	0,59	604,00	24,14	1710,07
Ouro Branco	4913,00	1,50	564,00	11326,53	0,65	866,40	29,93	1068,89
Parelhas	21499,00	1,50	4226,00	15862,52	0,676	750,90	28,44	4396,93
Santana do Seridó	2696,00	1,60	513,00	13002,92	0,64	757,70	28,34	893,95
São Fernando	3492,00	1,70	587,00	15395,51	0,61	911,90	30,20	608,86
São João do Sabugi	5956,00	1,70	468,00	12536,16	0,66	1036,60	29,33	4600,20
São José do Seridó	4558,00	1,20	1240,00	16905,22	0,69	759,10	33,13	777,90
São Vicente	6310,00	1,60	780,00	11052,24	0,64	663,60	29,90	908,32
Tenente Laurentino Cruz	5891,00	1,80	648,00	17090,47	0,62	522,60	29,78	2745,92
Timbaúba dos Batistas	2348,00	1,50	326,00	14273,90	0,64	825,20	29,61	2229,09

Fonte: Autor (2023).