

**Universidade de Brasília – UnB
Instituto de Química – IQ
Curso de Bacharelado em Química Tecnológica**

**AVALIAÇÃO ESTATÍSTICA DE TRÊS ESTAÇÕES DE TRATAMENTO DE ESGOTO DA
COMPANHIA DE SANEAMENTO AMBIENTAL DO DISTRITO FEDERAL (CAESB) POR
ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)**

**Estêvan Barbosa Colmanetti
Orientador: Prof. Dr. Jez Willian Batista Braga**

**Brasília, DF
2020**



Estêvan Barbosa Colmanetti

**AVALIAÇÃO ESTATÍSTICA DE TRÊS ESTAÇÕES DE TRATAMENTO DE ESGOTO DA
COMPANHIA DE SANEAMENTO AMBIENTAL DO DISTRITO FEDERAL (CAESB) POR
ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)**

Trabalho de conclusão de curso submetido ao curso de graduação em Química Tecnológica da Universidade de Brasília como requisito parcial para obtenção do Título de Bacharel em Química Tecnológica.

Orientador: Dr. Jez Willian Batista Braga

**Brasília, DF
2020**

Colmanetti, Estêvan Barbosa.

Avaliação Estatística De Três Estações De Tratamento De Esgoto Da CAESB Por Análise De Componentes Principais / Colmanetti, Estêvan Barbosa. Brasília: UnB, 2020. 45 p. : il. ; 29,5 cm.

Monografia (Graduação) – Universidade de Brasília, Instituto de Química, Brasília, 2020. Orientação: Dr. Jez Willian Batista Braga

1. Análise de componentes principais. 2. R. 3. Estações de tratamento de esgoto. 4. CAESB

I. Braga, Jez Willian Batista.

CDU Classificação

Estêvan Barbosa Colmanetti

Monografia submetida como requisito parcial para obtenção do Título de Bacharel em Química Tecnológica do Instituto de Química da Universidade de Brasília, em 14/12/2020 apresentada e aprovada pela banca examinadora abaixo:

Prof. Dr. Jez Willian Batista Braga, IQ/UnB
Presidente

Prof. Dr. Floriano Pastore Júnior, IQ/UnB
Membro Titular

Profa. Dra. Sarah Silva Brum, IQ/UnB
Membro Titular

Brasília, DF
2020

“De uma forma suave, você pode sacudir o mundo.”

Mahatma Gandhi

AGRADECIMENTOS

Agradeço inicialmente à minha família pelo apoio incondicional ao longo da minha graduação. Em especial à minha mãe, Estefânia, pelo crescente incentivo e por todo esforço.

Ao meu padrasto, Ricardo, por todas os válidos conselhos e por todas as advertências.

Ao meu pai, Vanilcio, por todos os frequentes retornos e por, principalmente, me incentivar dentro da minha área de atuação.

Ao meu amigo, Jorge Júnior, pelo crescente incentivo e amizade ao longo dos anos.

Aos meus avós, Gessi e José Eduardo, por existirem na minha vida e serem tão incríveis.

Ao meu grande amigo, Alexandre Costa, pelos ensinamentos em linguagens de programação e por toda a ajuda aferida.

Ao meu orientador, Jez, por toda a paciência e por toda a disponibilidade ao longo desse período de pandemia, me incentivando frequentemente e ajudando na construção deste trabalho.

Aos meus colegas de profissão e grandes amigos Matheus, Carolina, Giorgiane, Igor, Luiz, Saulo e Isabela pelas noites mal dormidas e principalmente pela longa caminhada juntos em busca dessa árdua profissão que é ser um Químico.

Todos vocês são ou foram em algum momento partes integrantes da minha vida e, por esse motivo, serei eternamente grato.

RESUMO

A Companhia de Saneamento Ambiental do Distrito Federal possui dezessete estações de tratamento de esgoto espalhadas por todo o Distrito Federal. Entre essas, três tiveram seus dados coletados durante o período de 1 de janeiro de 2017 à 31 de julho de 2020 e foram avaliadas neste trabalho pelo método de análise das componentes principais (PCA). Esse método realiza uma decomposição de matrizes de dados em chamadas componentes principais, que são combinações lineares das variáveis originais. De forma que as informações que possuem variâncias positivas entre si serão correlacionadas e as que não possuem, não serão correlacionadas. Foi possível identificar valores anômalos por meio do método PCA e suas respectivas causas, além de notar a diferença entre o afluente que chega à estação e o efluente que sai da estação. Foi possível observar também uma repetitividade nos resultados da CAESB, demonstrando a qualidade do tratamento. Além disso, algumas variáveis se mostraram altamente correlacionadas como Fósforo Total (PT), Sólidos Suspensos (SS) e Demanda Química de Oxigênio (DQO) pelo método. O PCA, contudo, não se mostrou eficiente em análises de variáveis com dados faltantes, mesmo utilizando de interpolação linear.

Palavras-Chave: PCA, R, DQO, CAESB, Tratamento de Esgoto.

ABSTRACT

The Environmental Sanitation Company of the Distrito Federal (CAESB) has seventeen sewage treatment plants throughout Distrito Federal. Among these, three had their data collected during the period from January 1st, 2017 to July 31, 2020 and were evaluated in this study by Principal Component Analysis (PCA). This method performs a decomposition of arrays in the so-called principal components, which are linear combinations of the original variables. The PCA model enable to identify anomalous values and its respective causes. Besides a clear difference between the affluent that arrives at the station and the effluent that leaves the station was observed. It was also possible to observe a similarity of the effluent treated by the CAESB, demonstrating the quality of the treatment. Some variables presented highly correlated values, such as Total Phosphorus (TP), Suspended Solids (SS) and Chemical Oxygen Demand (COD) by the method. However, PCA was unable to model efficiently variables presenting missing values, even using linear interpolation.

Keywords: PCA, R, COD, CAESB, Sewer Treatment.

LISTA DE ABREVIATURAS

PCA: Principal Component Analysis

PC-1: Componente principal 1

PC-2: Componente principal 2

PC-n: Componente principal n

ETE: Estação de Tratamento de Esgoto

CAESB: Companhia de Saneamento Ambiental do Distrito Federal

DQO: Demanda Química de Oxigênio

DBO: Demanda Bioquímica de Oxigênio

SS: Sólidos Suspensos

NT: Nitrogênio Total

PT: Fósforo Total

CT: Coliformes Totais ou Termotolerantes

DF: Distrito Federal

LISTA DE FIGURAS

<i>Figura 1- Representação de dois afluentes desaguardo em um curso de água principal.....</i>	<i>3</i>
<i>Figura 2- Representação de um efluente sendo jogado em um corpo de água (rio, lago, etc).....</i>	<i>3</i>
<i>Figura 3- Representação de afluente e efluente dentro da estação de tratamento de esgoto (ETE).....</i>	<i>4</i>
<i>Figura 4- Matriz de covariâncias de um conjunto de dados com três variáveis.....</i>	<i>5</i>
<i>Figura 5- Etapa de Gradeamento. / Foto: Departamento de Engenharia Hidráulica e Ambiental- DEHA/UFC.....</i>	<i>7</i>
<i>Figura 6- Decantadores em Estação de Tratamento de Esgoto de Sobradinho.....</i>	<i>7</i>
<i>Figura 7- Esquema simplificado da Estação de Tratamento de Esgoto de Sobradinho.....</i>	<i>9</i>
<i>Figura 8 – Porcentagem de variâncias explicadas para os afluentes das estações.....</i>	<i>14</i>
<i>Figura 9- Gráfico de pesos dos afluentes das estações.....</i>	<i>15</i>
<i>Figura 10 – Gráfico de escores dos afluentes das estações: ETE Sul (S •), ETE Norte (N •) e ETE Gama (G •).....</i>	<i>17</i>
<i>Figura 11- Gráfico da porcentagem de variâncias explicadas para os efluentes das estações com os valores anômalos.....</i>	<i>18</i>
<i>Figura 12- Gráfico de pesos dos efluentes das estações com os valores anômalos.....</i>	<i>18</i>
<i>Figura 13- Gráfico de escores dos efluentes das estações. (Amostras com comportamento anômalo identificadas por setas vermelhas).....</i>	<i>19</i>
<i>Figura 14 - Gráfico da porcentagem de variâncias explicadas para os efluentes das estações sem os valores anômalos.....</i>	<i>20</i>
<i>Figura 15 - Gráfico de pesos dos efluentes das estações sem os valores anômalos.....</i>	<i>21</i>
<i>Figura 16 - Gráfico de escores dos efluentes das estações sem os valores anômalos.....</i>	<i>22</i>
<i>Figura 17 - Gráfico da porcentagem de variâncias explicadas para a ETE Brasília Norte.....</i>	<i>23</i>
<i>Figura 18 - Gráfico da porcentagem de variâncias explicadas para a ETE Brasília Sul.....</i>	<i>23</i>
<i>Figura 19 - Gráfico da porcentagem de variâncias explicadas para a ETE Gama.....</i>	<i>24</i>
<i>Figura 20 - Gráficos de pesos para as três estações de tratamento de esgoto. Resultados, da esquerda para a direita, respectivamente, da ETE Norte*, ETE Sul e ETE Gama.....</i>	<i>24</i>
<i>Figura 21– Gráfico de escores da ETE Brasília Norte. Afluente (•) e Efluente (•).....</i>	<i>25</i>
<i>Figura 22– Gráfico de escores da ETE Brasília Sul. Afluente (•) e Efluente (•).....</i>	<i>25</i>
<i>Figura 23- Gráfico de escores da ETE Brasília Norte. Afluente (•) e Efluente (•).....</i>	<i>26</i>
<i>Figura 24 - Gráfico da porcentagem de variâncias explicadas dos afluentes das estações acrescentando mais variáveis.....</i>	<i>27</i>
<i>Figura 25 - Gráfico de pesos dos afluentes das estações acrescentando mais variáveis.....</i>	<i>27</i>
<i>Figura 26 - Gráfico de escores dos afluentes das estações, acrescentando mais variáveis.....</i>	<i>28</i>
<i>Figura 27- Gráfico da porcentagem de variâncias explicadas dos efluentes das estações acrescentando mais variáveis.....</i>	<i>29</i>
<i>Figura 28- Gráfico de pesos dos efluentes das estações acrescentando variáveis de menor periodicidade.....</i>	<i>29</i>
<i>Figura 29 - Gráfico de escores dos efluentes das estações acrescentando mais variáveis.....</i>	<i>31</i>

LISTA DE TABELAS

<i>Tabela 1 - Matriz de correlação dos afluentes da estação</i>	<i>15</i>
<i>Tabela 2 – Matriz de correlação dos efluentes da estação com os valores anômalos.....</i>	<i>19</i>
<i>Tabela 3 – Matriz de correlação dos efluentes sem os valores anômalos.....</i>	<i>21</i>
<i>Tabela 4 – Matriz de correlação em análise de afluentes com mais variáveis.....</i>	<i>27</i>
<i>Tabela 5 – Matriz de correlação em análise de efluentes com mais variáveis.....</i>	<i>29</i>
<i>Tabela 6 - Remoção em porcentagem de cada parâmetro nas estações de tratamento.....</i>	<i>30</i>

SUMÁRIO

LISTA DE FIGURAS	XI
LISTA DE TABELAS.....	XII
1. INTRODUÇÃO	1
2. OBJETIVOS	2
2.1 OBJETIVO GERAL	2
2.2 OBJETIVOS ESPECÍFICOS	2
3. REVISÃO BIBLIOGRÁFICA	2
3.1 AFLUENTE E EFLUENTE.....	2
3.1 PCA.....	4
3.3. Estações de Tratamento de Esgoto	6
4. MATERIAIS E MÉTODOS.....	9
4.1 DISPONIBILIZAÇÃO E ORGANIZAÇÃO DOS DADOS	9
4.2 SOFTWARE R E PCA	12
5. RESULTADOS E DISCUSSÃO.....	13
5.1 AFLUENTES	13
5.2 EFLUENTES.....	17
5.3 PCA INDIVIDUAL.....	22
5.4 PCA ACRESCENTANDO VARIÁVEIS (2020)	25
6. CONCLUSÃO.....	30
7. PERSPECTIVAS.....	31
8. BIBLIOGRAFIA	32

1. INTRODUÇÃO

A importância do tratamento do esgoto e do saneamento básico se dá desde os primórdios do que se entende por sociedade. As primeiras políticas públicas de saneamento e recursos hídricos no Brasil datam da era do colonialismo português inicialmente por meio do sistema de sesmarias, que atribuía a particulares a posse de terras e águas e a função de produzir, sendo reservadas aos governantes as moendas e engenhos. (MURTHA, 2015).

Inicialmente, o objetivo era simplesmente desviar o esgoto doméstico de estabelecimentos urbanos e ao longo dos séculos houve a preocupação em se reutilizar essa água residuária, havendo indícios de civilizações pré-históricas (hindus, mesopotâmios e minoan) reutilizando águas de esgoto para irrigação. Aplicações de águas residuárias em terras de produção são práticas antigas e passaram por diferentes níveis de desenvolvimento. (ANGELAKIS, 2015)

Como a água é um bem necessário e essencial, a preocupação com sua qualidade para consumo data dos primórdios da humanidade e a sobrevivência de populações sempre dependeu da disponibilidade desse bem natural (KONZEN, 2015). Nos dias atuais, entretanto, o uso excessivo e o desperdício de água têm causado preocupação e a água tem aparecido, portanto, cada vez mais como um dos eixos centrais de discussões ambientais, econômicas e sociais. Dessa forma, se fez necessária a implementação de parâmetros de qualidade e a criação de agências regulamentadoras que pudessem controlar e mensurar a qualidade do tratamento disponibilizado para a água.

A CAESB (Companhia de Saneamento Ambiental do Distrito Federal), empresa fundada em 1969 juntamente com a construção de Brasília, é a responsável por tratar as águas do Distrito Federal e entorno (DIRETORIA COLEGIADA, 2014). A ADASA (Agência Reguladora de águas, Energia e Saneamento do Distrito Federal), por sua vez, é a agência regulamentadora e reguladora, que avalia periodicamente a qualidade do tratamento da água de forma que este seja padronizado e responda aos critérios ambientais e resoluções do CONAMA (Conselho Nacional do Meio Ambiente).

O tratamento efetuado pela CAESB passa por um alto critério de avaliação e responsabilização e por este motivo deve responder com resultados que serão reavaliados e validados. A validação destes dados se dá por padronização de parâmetros e análises além de rigorosas determinações estatísticas. Sendo assim os resultados das análises permite a avaliação externa para eventual identificação de possíveis erros e melhorias.

Tratando-se mais especificamente do tratamento de águas residuárias, os parâmetros medidos nas estações de tratamento de esgoto deste trabalho são diferenciados entre ‘afluente à estação’ e ‘efluente final’ e são os mesmos para as três estações, sendo que nove das dezessete estações de tratamento da CAESB possuem tratamento a nível terciário, incluindo-se as estações avaliadas neste trabalho; enquanto as demais possuem tratamento secundário (CAESB, 2015).

2. OBJETIVOS

2.1 Objetivo geral

O trabalho teve como objetivo identificar e agrupar informações relevantes a respeito da qualidade da água das estações de tratamento de esgoto: ETE Brasília Sul, ETE Brasília Norte e ETE Gama; explicitando a qualidade do mesmo por método estatístico de análise dos componentes principais (PCA).

2.2 Objetivos específicos

- Identificação da eficiência do tratamento por repetitividade de resultados;
- Comparação e correlação de dados entre afluente e efluente;
- Identificar diferenças entre dados de análises das três ETEs.

3. REVISÃO BIBLIOGRÁFICA

3.1 Afluente e Efluente

Afluentes são cursos de água que desaguam em rios principais. No caso, rios e cursos de água menores. Sua vazão contribui diretamente para o aumento de outro corpo d’água. Para que um curso d’água possa ser chamado de afluente este precisa necessariamente estar ligados a um rio principal, mesmo que esse rio deságue em mares/oceanos. (RIBEIRO, 2020)



Figura 1 – Representação de dois afluentes desaguando em um curso de água principal.

Efluente, por sua vez, é o escoamento de água de uma estação de tratamento de águas residuais para um corpo natural de água (FOGLER, 2005).



Figura 2 – Representação de um efluente sendo jogado em um corpo de água (rio, lago, etc).

Nas estações de tratamento de esgoto, é chamado de afluente todo o fluxo de água que entra na estação, no caso, todo o material que será tratado. Ao passar pela estação e pelas etapas de tratamento, o fluxo passa a ser chamado de efluente, tendo etapas intermediárias até chegar ao efluente final que é o efluente tratado pronto para ser despejado.

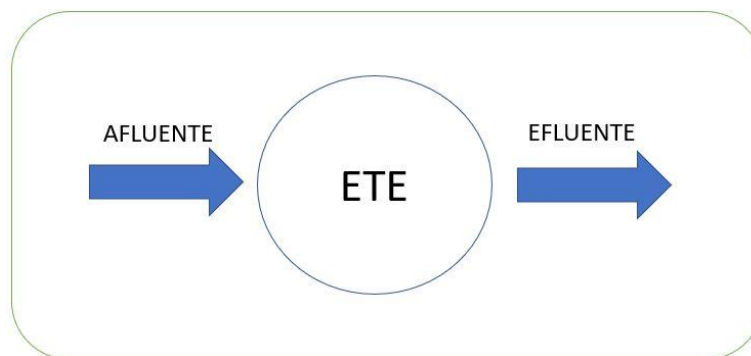


Figura 3 – Representação de afluente e efluente dentro da estação de tratamento de esgoto (ETE).

3.2 Principal Component Analysis (PCA)

A análise de componentes principais (*Principal Component Analysis*) permite agrupar informações relevantes e diminuir a dimensionalidade dos dados em um novo sistema de eixos denominados componentes principais. Esses novos eixos na realidade são, matematicamente, a decomposição de uma matriz **X** em um produto de duas matrizes denominadas de escores (**T**) e pesos (**P**) mais uma matriz de erros (**E**).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

Equação 1 – Decomposição da matriz **X** em um produto de duas novas matrizes (**T** e **P**).

As amostras podem ser vistas como pontos localizados em um espaço n-dimensional, sendo n igual ao número de variáveis. Com a redução da dimensionalidade proporcionada pela PCA, as amostras passam a ser pontos localizados num espaço de dimensão reduzida, definido pelas PCs, por exemplo, bi ou tridimensionais (SOUZA, 2010).

Para explicar o método de componentes principais é necessário compreender que a intenção principal é reduzir o conjunto de dados com um grande número de variáveis a um outro com menos variáveis sem perder a informação relevante contida nos dados. Isso porque grandes quantidades de dados tendem a ter muitas informações redundantes. Para isso, são efetuados uma sequência de passos que validam o método.

Inicialmente, caso a unidade das variáveis não for a mesma, é necessário efetuar a padronização ou escalamento dos dados. A ideia é que as variáveis que estejam em unidades de medida diferentes contribuam igualmente para as componentes principais. Sendo z o valor normalizado, a equação utilizada para padronização dos valores é explicitada abaixo (HOLLAND, 2008).

$$z = \frac{(\text{valor} - \text{média})}{\text{desvio padrão}}$$

Equação 2 – Equação de autoescalamento dos valores do conjunto de dados a ser analisado.

Então, com os valores normalizados, o próximo passo é compreender como as variáveis do conjunto de dados se distanciam da média e identificar se há alguma relação entre elas. Para isso, computa-se as matrizes de covariância. De forma a identificar as amostras que são altamente correlacionadas evitando dados redundantes. Uma matriz de covariância é uma matriz simétrica no tamanho $n \times n$ em que n é o número de variáveis. Portanto, uma representação de uma matriz de covariância com três variáveis seria como demonstrada abaixo.

$$\begin{bmatrix} Cov(x, x) & Cov(x, y) & Cov(x, z) \\ Cov(y, x) & Cov(y, y) & Cov(y, z) \\ Cov(z, x) & Cov(z, y) & Cov(z, z) \end{bmatrix}$$

Figura 4 – Matriz de covariâncias de um conjunto de dados com três variáveis.

Uma matriz de covariância com três variáveis geraria uma matriz simétrica com nove termos (3×3) como demonstrado pela Figura 4. Como a covariância de uma variável por ela mesma (*e.g.* $Cov(x,x)$) é a própria variância ($Var(x)$) e a covariância é comutativa (*e.g.* $Cov(x,y)=Cov(y,x)$), os elementos da matriz de covariância dependem única e exclusivamente da diagonal principal. E o que determina a relação entre possíveis variáveis é o sinal da matriz de covariância. Caso este seja positivo, as variáveis crescem e decrescem juntas, ou seja, são correlacionadas; se o inverso, elas não são correlacionadas (HOLLAND, 2008).

Como as componentes principais são novas variáveis moldadas a partir de combinações lineares das variáveis iniciais, a ideia é que à medida em que as matrizes de covariância forem sendo calculadas, elas serão comprimidas a valores correlacionados das primeiras componentes principais (SMITH, 2002).

Ou seja, dessa forma, é possível identificar quais variáveis possuem a maior quantidade de contribuição para explicar o conjunto de dados, reduzindo assim a dimensionalidade. Geometricamente falando, os pesos das componentes principais representam as “direções” dos dados que explicam a maior quantidade de variância, ou seja, capturam a maior quantidade de informação. (HOLLAND, 2008). Assim, a componente principal 1 (PC-1) terá a maior variância possível, seguida da componente principal 2 e assim por diante. A componente

principal 2, entretanto, é calculada a partir da não correlação com a primeira componente principal, tendo assim a segunda maior variância.

Por fim, como os autovetores de uma matriz de covariância representam as direções dos eixos onde há maior variância (consequentemente, maior informação), é lógico dizer que as componentes principais são os próprios autovetores da matriz de covariância do conjunto de dados. Ou seja, um conjunto de dados com três variáveis, terá uma matriz (3 x 3) com três respectivos autovetores e três autovalores atrelados a esses autovetores, sendo que os autovalores nada mais são do que a quantidade de variância atrelada a cada componente principal (SMITH, 2002).

3.3. Estações de Tratamento de Esgoto

As estações de tratamento de esgoto possuem função essencial no tratamento de águas residuárias. Elas têm a função de remover impurezas da água antes de a mesma retornar aos corpos d'água como rios, lagos e oceanos. Como não é encontrada água pura na natureza, qualquer distinção entre água poluída e água limpa depende inteiramente da concentração das impurezas. Tecnicamente falando, uma água se diz impura ou poluída quando ela é inadequada para determinado uso como banho, consumo, pesca, etc. (NATHANSON; AMBULKAR, 2020). Dentro de uma estação de tratamento, os processos podem ser divididos e caracterizados de acordo com suas respectivas funções e complexidades. Estes processos são usualmente divididos em pré-tratamento ou tratamento preliminar, processos primários, processos secundários e processos terciários (SAAE, 2020).

O tratamento preliminar os mecanismos predominantes são físicos e o objetivo é eliminar os sólidos indevidamente lançados no esgoto, sólidos grosseiros e que possam causar algum tipo de dano para os equipamentos. Nessa etapa são utilizadas grades, caixas de areia ou peneiras.



Figura 5 – Etapa de Gradeamento. / Foto: Departamento de Engenharia Hidráulica e Ambiental – DEHA/UFC.

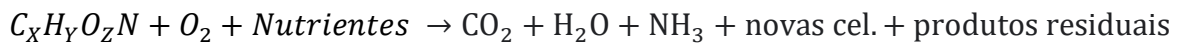
O tratamento primário, por sua vez, possui mecanismos físicos e físico-químicos e tem como objetivo a remoção dos sólidos sedimentáveis ou em suspensão, sendo estes colocados em grandes decantadores e depois removidos como lodo primário. Os processos mais comuns são o de sedimentação, coagulação/floculação. Ao fim dessa etapa, o efluente deve estar quase que livre de sólidos visíveis.



Figura 6 – Decantadores em Estação de Tratamento de Esgoto de Sobradinho.

Em seguida, ocorre o tratamento secundário. Este consiste de mecanismos essencialmente biológicos como bactérias aeróbias e anaeróbias. O objetivo é a remoção da

matéria orgânica biodegradável. Em processos aeróbios, os microrganismos converterão a matéria orgânica essencialmente em gás carbônico e água.



Equação 3 – Equação química da degradação aeróbica do águas residuárias.

Em processos anaeróbios, os microrganismos converterão a matéria orgânica em gás metano, gás carbônico, amônia e produtos residuais, dependendo da complexidade das cadeias orgânicas biodegradadas (SABESP, 2020). Essa consiste na última etapa do tratamento em algumas das estações de tratamento de esgoto da CAESB.



Equação 4 – Equação química da degradação anaeróbica de águas residuárias.

Em ETEs que possuem tratamento terciário, o objetivo é essencialmente a remoção dos micronutrientes, mais especificamente Fósforo e Nitrogênio, que são mais dificilmente biodegradados. Estes são removidos com o chamado polimento final que consiste em processos desde ultrafiltração, oxidação química/biológica ou remoção desses micronutrientes por meio de clarificadores (TCHOBANOGLIOUS et al, 2003). Importante ressaltar que as três ETEs analisadas neste trabalho possuem tratamento terciário. Um exemplo de esquema simplificado da ETE Sobradinho abaixo.

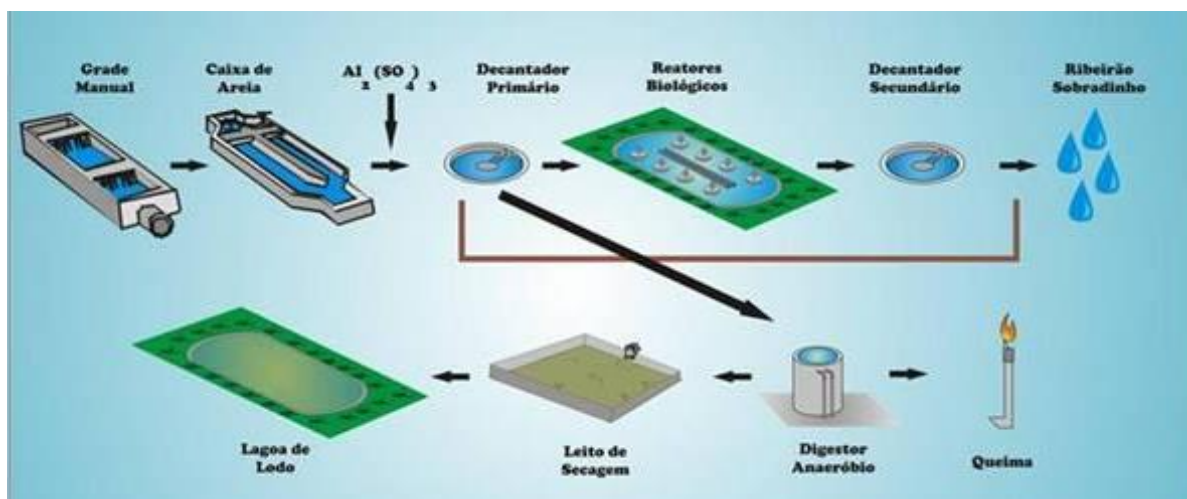


Figura 7 – Esquema simplificado da Estação de Tratamento de Esgoto de Sobradinho. Imagem retirada de Urbanizadora Paranoazinho (2012).

Importante ressaltar que na Figura 8 é especificado o agente Sulfato de Alumínio (Al_2SO_4) após a etapa de utilização da caixa de areia. Este é comumente utilizado em estações

de tratamento de água e esgoto na etapa de tratamento primário para coagulação/floculação. O agente químico é acrescentado às águas residuárias de forma a desestabilizar os agentes coloidais e facilitar a formação de flocos pela agregação dos mesmos (TCHOBANOGLOUS et al, 2003). Geralmente, pás de metal ficam girando em uma velocidade pré-determinada para aumentar a eficiência desse processo. Os flocos formados então, são decantados ou flotados dependendo da estação.

4. MATERIAIS E MÉTODOS

4.1 Disponibilização e Organização dos Dados

Os dados analisados neste trabalho foram disponibilizados pela CAESB e consistem nas análises realizadas desde 1 de janeiro de 2017 a 31 de julho de 2020 em três estações de tratamento de esgoto: ETE Brasília Sul, ETE Brasília Norte e ETE Gama.

Cada parâmetro medido possui um método e uma periodicidade pré determinada. Os resultados disponibilizados foram referentes às seguintes análises:

- Análise de nutrientes: Fósforo Total (PT) (análise diária), Nitrogênio Total (NT) (análise diária), Nitrogênio Amoniacal (N-NH₃) (análise realizada apenas em efluentes e semanal).
- Análise de matéria orgânica biodegradável e não biodegradável: Demanda Química de Oxigênio (DQO) (análise diária) e Demanda Bioquímica de Oxigênio (DBO) (análise semanal).
- Análise de sólidos em suspensão (SS). (análise diária).
- Análises microbiológicas: Coliformes Totais ou Termotolerantes (CT)* e *E. coli* (análise mensal).

**É importante ressaltar que a análise de Coliformes Totais (CT) foi interrompida no ano de 2018 nas ETEs deste trabalho, inviabilizando a análise dessa variável por falta de preenchimento.*

Os métodos aplicados para cada variável analisada neste trabalho seguem padronizações internacionais, descritas e utilizadas em outras estações de tratamento. Essas análises são todas realizadas regularmente no laboratório das próprias estações de tratamento de esgoto e no laboratório central da CAESB.

4.1.1 Nutrientes (NT, PT e N-NH₃)

Os nutrientes provenientes de material residuário como urina e fezes são ricos em substâncias como nitrogênio e fósforo. Essas substâncias são altamente eutrofizantes para

corpos receptores (MENDONÇA, 2012). Isso porque essas substâncias servem de nutrientes para as algas. Ou seja, em altas concentrações desses materiais, as algas crescem de maneira desenfreada, consumindo todo o oxigênio dissolvido existente naquele corpo hídrico e causando, conseqüentemente, a morte de organismos ali existentes. Além disso, por mais que os métodos de remoção dessas substâncias sejam plenamente conhecidos, são métodos caros e geralmente fazem parte de etapas terciárias de tratamento. (HALLS, 2002). Os métodos padronizados de medição de fósforo total (PT), nitrogênio total (NT) e nitrogênio amoniacal (N-NH₃) são métodos colorimétricos, comuns em análises da CAESB. Esses métodos consistem na digestão das amostras de água residuária em forno microondas em temperatura controlada por um período fixo de tempo. Após a digestão, são adicionados agentes complexantes às amostras. A ideia é formar um complexo colorido que tenha em sua estrutura o material de interesse e, então, identificar a concentração desse material por curva de calibração em espectrofotômetro Uv-Vis. O método completo e os procedimentos detalhados estão descritos em APHA (2005).

4.1.2 Demanda Bioquímica de Oxigênio (DBO)

A DBO é uma medida indireta da quantidade de matéria orgânica biodegradável presente numa amostra e é necessária em águas residuárias para detectar a capacidade do corpo receptor de decompor essa matéria orgânica que ali será despejada após tratamento. Sua detecção se faz pela quantidade de oxigênio consumido na respiração aeróbia de microrganismos presentes.

As bactérias presentes consomem o oxigênio dissolvido e degradam a matéria orgânica existente enquanto realizam o processo de respiração. É um teste de duração de 5 dias com temperatura constante de 20°C. Para um melhor desempenho dos organismos presentes é necessário apresentar um ambiente favorável, com nutrientes necessários para a proliferação do meio consumido, como descrito em BERTOLAMI (2017).

4.1.3 Demanda Química de Oxigênio (DQO)

A medida de DQO é feita por procedimento padrão descrito em APHA (2005) com adição de dicromato de potássio em refluxo fechado. O dicromato se comporta como agente oxidante, de forma que o cromo reduz de seu estado de oxidação heptavalente (Cr₂O₇²⁻) ao seu estado trivalente (Cr³⁺) e as matérias orgânica e inorgânica existentes são oxidadas em presença de ácido sulfúrico altamente concentrado. (VALENTE, PADILHA, & SILVA, 1997). O

método utilizado para identificação da concentração é também colorimétrico, similar ao método dos nutrientes, descrito em 4.1.1.

4.1.4 Sólidos Suspensos (SS)

A análise de sólidos suspensos é a análise que se refere aos sólidos de tamanho característico que não foram dissolvidos em águas residuárias. Na maior parte dos casos, são estruturas coloidais, eletricamente estabilizadas. Em contato com corpos hídricos, estes sólidos aumentam a turbidez da água, podendo impedir a passagem de luz, atrapalhando no aquecimento dos cursos de água e podendo diminuir a quantidade de oxigênio dissolvido (LOUGON et al, 2011). A sua medição em águas de esgoto se dá por método gravimétrico, no caso, pela diferença entre sólidos totais e sólidos voláteis e o procedimento completo adotado pela CAESB está descrito também em APHA (2005).

4.1.5 Coliformes Totais ou Termotolerantes (CT) e *E. coli*

Os coliformes são formados por um número de bactérias que inclui os gêneros *Klebsiella*, *Escherichia*, *Serratia*, *Erwenia* e *Enterobactéria*. Essas bactérias são associadas com as fezes de animais de sangue quente e são medidas em águas residuárias para identificar possíveis infecções intestinais que possam estar ocorrendo na população que dispõe do tratamento de esgoto. (CEBALLOS, 2017)

Para determinação de coliformes termotolerantes, o método utilizado é o de tubos múltiplos e a técnica se baseia na detecção dos subprodutos do metabolismo da fermentação da lactose em meio de cultura lactosado. Os coliformes, por sua habilidade de fermentar a lactose presente no meio, são detectados diretamente. A densidade bacteriana, ou seja, quantidade de bactérias presentes no tubo é estimada a partir de uma tabela de Número Mais Provável (NMP) que contém resultados de variadas combinações de tubos positivos e negativos para coliformes termotolerantes conforme descrito em CETESB (2017).

Para a determinação específica de *E. coli*, o método utilizado é do substrato enzimático. Quando os coliformes totais metabolizam o indicador ONPG contido no substrato enzimático, a amostra adquire uma coloração amarela. Quando a *E. coli* metaboliza o indicador MUG contido no substrato enzimático, a amostra fluoresce quando exposta a luz ultravioleta de 365 nm. Os resultados, também, são dados em Número Mais Provável por 100 mL (NMP/100mL) de acordo com tabela. O procedimento e a tabela estão descritos em CETESB (2017).

4.2 Linguagem R e método PCA

A etapa da análise de dados foi feita por meio do programa estatístico RStudio que é um software livre e utiliza algumas linguagens de programação, entre elas a que foi utilizada neste trabalho: o R. Esse software permite a realização de cálculos matemáticos e estatísticos avançados e plotagem de gráficos. Além disso, foi desenvolvido no intuito de auxiliar alunos do mundo todo a desenvolverem trabalhos e pesquisas com grandes conjuntos de dados sem que os mesmos precisem adquirir um software pago.

Para se obter os resultados de PCA, foram seguidos os seguintes passos:

- Organização do conjunto de dados;
- Exclusão de variáveis de periodicidade muito restrita;
- Exclusão de datas que não possuíam análise nenhuma;
- Interpolação linear de valores não numéricos,
- Autoescalamamento dos dados;
- Realização da análise PCA e obtenção dos gráficos.

O primeiro passo foi a organização do conjunto de dados. As variáveis em colunas e as amostras em linhas e ordenadas por suas respectivas datas de coleta.

Após a organização do conjunto de dados, as variáveis que apresentavam periodicidade muito restrita e, portanto, não possuíam preenchimento completo (*E. coli*, N-NH₃, DBO e CT), em um primeiro momento, foram desconsideradas por exigirem uma exclusão de um número significativo de datas das demais variáveis. Os valores vazão e volume têm uma relação linear extremamente forte, ou seja, seu coeficiente de correlação de Pearson é igual a 1. Assim sendo apenas uma dessas variáveis foi mantida, no caso, a variável vazão.

Além disso, como as análises possuem periodicidades diferentes entre si, algumas datas tinham o resultado de uma determinada análise, sem o resultado das outras. Gerando para o conjunto de dados valores sem preenchimento naquela data específica. O R, nesses casos, identifica esse valor como não numérico e o descreve na tabela de dados como NaN (sigla de *not a number*). Estes valores identificados como NaN foram interpolados linearmente, utilizando uma interpolação polinomial, que nada mais é do que uma boa aproximação para gerar preenchimento em dias de não realização das análises e ter um resultado mais significativo. Contudo, variáveis que continham um valor excessivo de NaN (>10%) tiveram que ser excluídas da PCA.

O último passo foi o autoescalamamento. Mesmo os dados estando na mesma unidade de medida (mg.L⁻¹), os intervalos de valores são muito diferentes entre as variáveis. Portanto, antes

da decomposição dos valores em componentes principais, é importante que seja realizado o autoescalamamento desses dados. Ao fim do autoescalamamento, o desvio padrão de todas as colunas é igual. (SMITH, 2002).

Para obtenção dos gráficos, inicialmente, a decomposição em componentes principais foi dividida em afluentes e efluentes das ETEs. Os resultados da porcentagem de variância explicada (que é o quanto cada componente principal explica aquele conjunto de dados), o gráfico de pesos (contribuição das variáveis por dimensão) e gráfico de escores (relação entre as amostras) são apresentados para as duas primeiras componentes principais

Após a obtenção desses gráficos acima para efluente e afluente, PCA foi calculado para cada ETE individualmente, de forma a identificar os processos internos e suas possíveis correlações e seguindo os mesmos passos descritos anteriormente.

Por fim, a PCA foi recalculada, dessa vez especificamente para o ano de 2020, dividindo o conjunto de dados em semanas para trazer uma maior quantidade de variáveis, desconsiderando apenas o valor de *E-Coli* por ser um procedimento mensal. Ou seja, variáveis que possuam mais do que um resultado por semana, utilizou-se o valor médio; variáveis que possuam apenas um valor, este foi considerado e em semanas que o conjunto de dados não possuía preenchimento, o valor era interpolado.

5. RESULTADOS E DISCUSSÃO

Ao todo, o conjunto de dados possui 4380 observações, sendo que metade (2190 observações) é referente aos afluentes das estações e a outra metade referente aos efluentes das estações. As variáveis inicialmente consideradas para a análise foram DQO, PT, SS e vazão por serem variáveis que possuem resultados em maior número de dias.

5.1 Afluentes

Os resultados da decomposição em componentes principais para afluentes das estações estão expressos nas figuras abaixo.

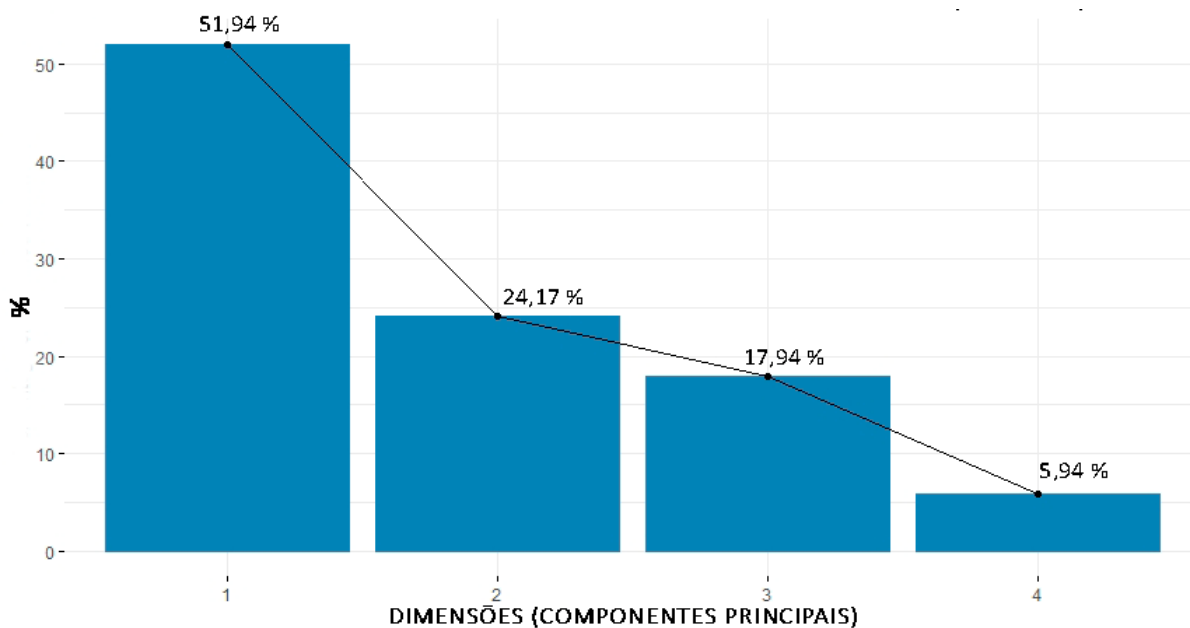


Figura 8 – Porcentagem de variâncias explicadas para os afluentes das estações.

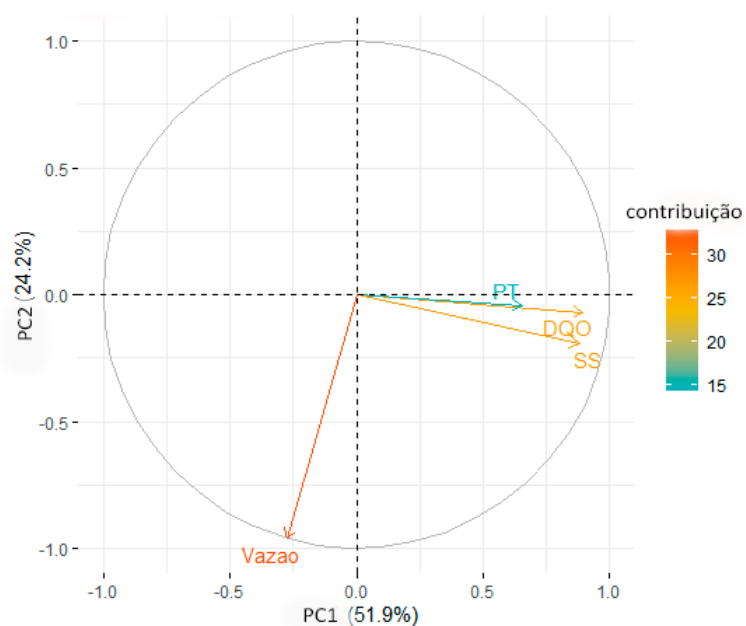


Figura 9 – Gráfico de pesos dos afluentes das estações.

Tabela 1 – Matriz de correlação das variáveis utilizadas na análise dos afluentes da estação.

	DQO	SS	PT	Vazão
DQO	1.000	0.756	0.385	-0.173
SS	0.756	1.000	0.380	-0.078
PT	0.385	0.380	1.000	-0.107
Vazão	-0.173	-0.078	-0.107	1.000

Pelo gráfico de proporção das variâncias explicadas, nota-se que a componente principal um (PC1) carrega a maior parte da informação dos dados de afluente. De acordo com o gráfico de pesos, a componente contém predominantemente a variância presente nas variáveis fósforo total (PT), sólidos suspensos (SS) e DQO. Além disso, de acordo com a matriz de correlação identificada logo abaixo, é possível identificar forte correlação entre DQO e SS e correlação moderada entre essas duas variáveis e PT. O que é um resultado razoável, visto que DQO e SS dizem respeito à matéria orgânica e inorgânica biodegradável e em suspensão, respectivamente. Já PT possui uma correlação menor por ser proveniente majoritariamente por origem fisiológica e uma pequena parcela por detergentes e sabões (MOTA, 2009).

A contribuição total de cada variável individualmente para cada componente principal (identificada ao lado direito do gráfico de pesos) é calculada por uma média aritmética entre as contribuições individuais de determinada variável em cada componente principal e os respectivos autovalores das componentes principais, sendo C_1 e C_2 as respectivas contribuições individuais de determinada variável em PC-1 e PC-2.

Para se calcular essas contribuições (C_1 e C_2), é necessário inicialmente calcular a correlação de cada variável com cada PC individualmente. Essa correlação é calculada utilizando-se a equação de correlação de Pearson abaixo.

$$r_{xy} = \frac{Cov(X, Y)}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Equação 5 – Cálculo da correlação de Pearson de cada variável isoladamente com cada componente principal.

A partir desse valor de correlação, é possível calcular o valor de \cos^2 que diz respeito à qualidade da representação para cada variável. Isso porque os gráficos de pesos são representados matematicamente pelo cosseno dos ângulos entre as variáveis originais e as componentes principais. O valor de \cos^2 nada mais é do que os valores de correlação, obtidos pela Equação 5, ao quadrado. Então, a contribuição individual de cada variável em cada PC (C_1 e C_2) é a divisão do seu valor de \cos^2 pelo valor total de \cos^2 de determinada componente. A partir desses valores de C_1 e C_2 e considerando que AV_1 e AV_2 são os respectivos autovalores de PC1 e PC-2, calcula-se a contribuição total de cada variável a partir da equação abaixo (HOLLAND, 2008).

$$\text{contribuição} = \frac{(C_1 * AV_1) + (C_2 * AV_2)}{AV_1 + AV_2}$$

Equação 6 – Cálculo da contribuição nos gráficos de pesos.

Dessa forma, é possível identificar se uma determinada variável possui alta contribuição em seu gráfico de pesos em determinada componente principal pela coloração atrelada à sua seta. Caso a seta possua coloração próxima do laranja é uma contribuição elevada, caso a seta possua coloração azul/azul claro é uma contribuição mediana para baixa.

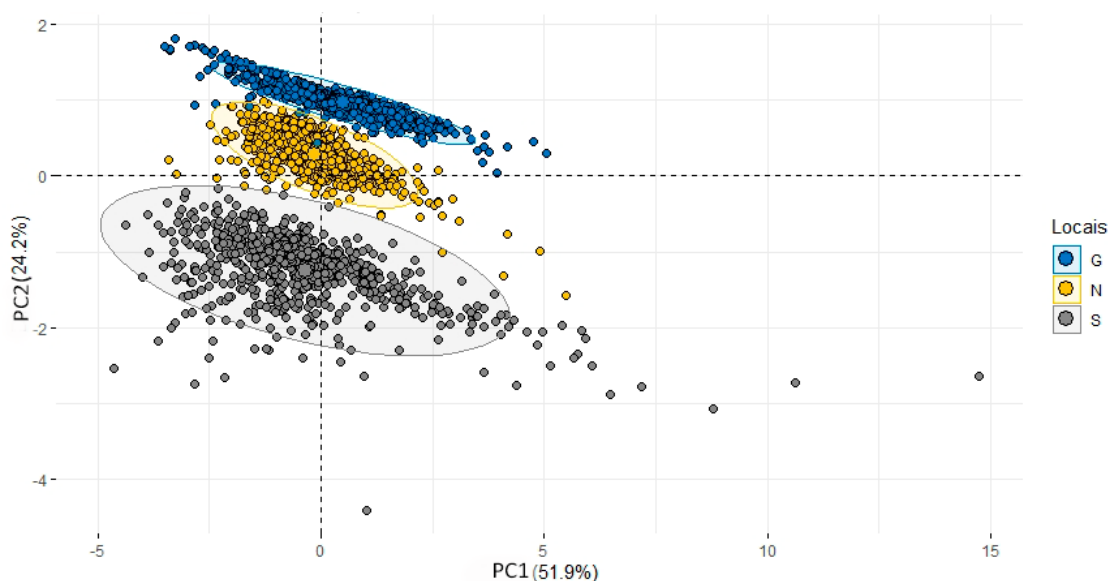


Figura 10 – Gráfico de escores dos afluentes das estações. ETE Sul (S ●), ETE Norte (N ●) e ETE Gama (G ●).

De acordo com o gráfico de escores para os afluentes das estações, é possível identificar uma separação bastante nítida em relação a componente principal de número dois (PC-2) para todas as ETEs (N, S e G). Essa componente, de acordo com o gráfico de pesos, diz respeito basicamente a vazão, o que está de acordo, visto que as vazões de cada estação são bem diferentes entre si. A ETE Brasília Norte possui uma vazão média de aproximadamente 500 L/s, enquanto que a ETE Brasília Sul possui uma vazão média de aproximadamente 1200 L/s e a ETE Gama possui uma vazão média de aproximadamente 200 L/s. É interessante notar que os *clusters* formados entre as ETEs Norte e Gama estão mais próximos do que o *cluster* formado pela ETE Sul, isso pelas maiores proximidades das vazões das ETEs Gama e Norte do que da vazão da ETE Sul.

Em relação a componente principal de número um (PC-1) é possível observar uma discrepância em relação aos dados da ETE Sul para com os demais, no caso, os dados possuem um maior intervalo entre os valores de SS, DQO e PT. Como é a maior ETE, possui uma maior variação do material orgânico que chega à estação. Em comparação, a ETE Gama, possui seus dados com uma variação consideravelmente inferior, o que se explica pelo seu menor tamanho se comparada com as demais.

5.2 Efluentes

Os resultados da decomposição em componentes principais para efluentes das estações está expresso nas figuras abaixo.

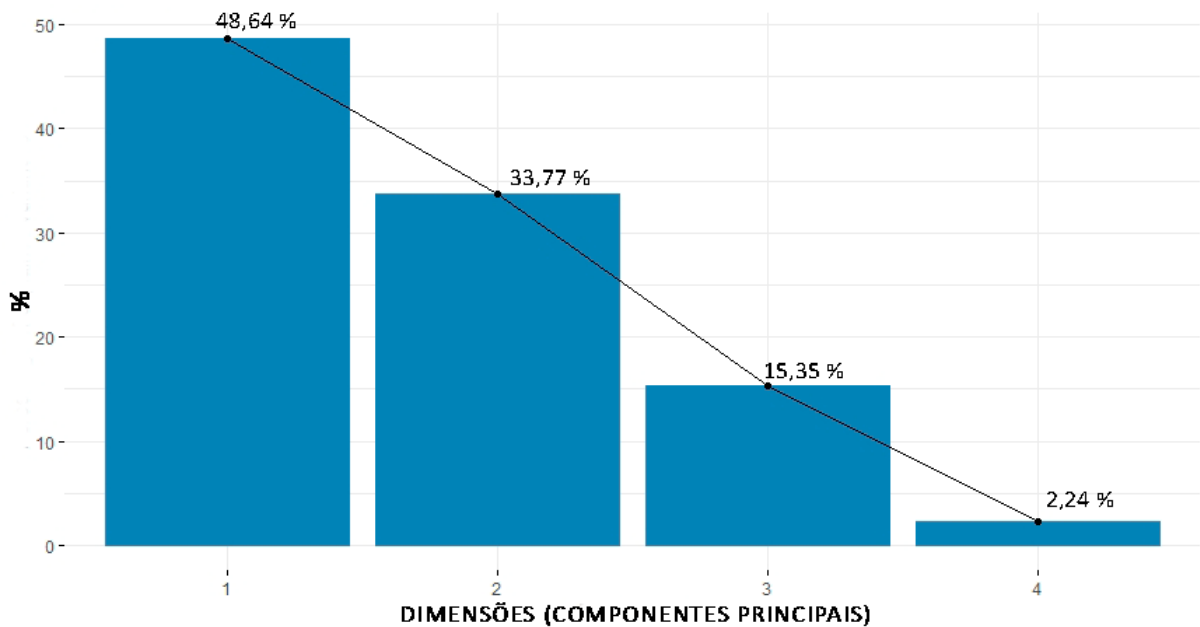


Figura 11 – Gráfico da porcentagem de variâncias explicadas para os efluentes das estações com os valores anômalos.

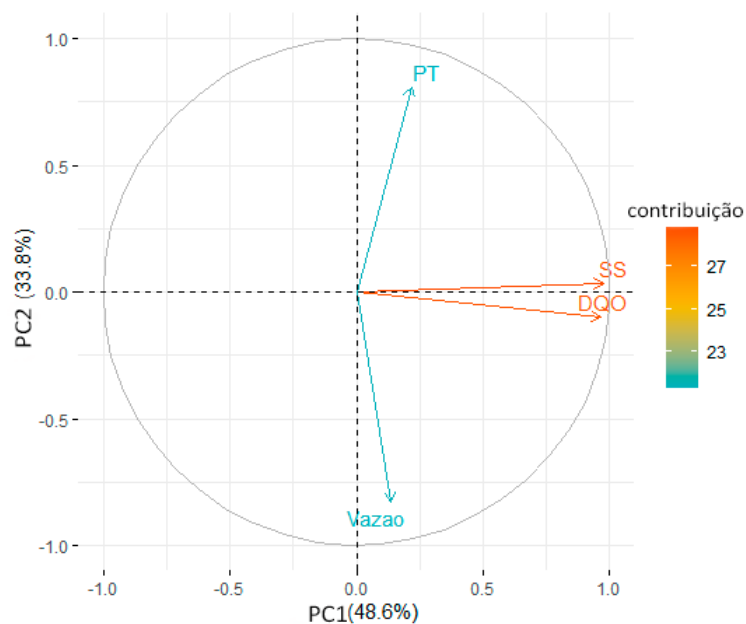


Figura 12 – Gráfico de pesos dos efluentes das estações com os valores anômalos.

Tabela 2 – Matriz de correlação dos efluentes da estação com os valores anômalos.

	DQO	SS	PT	Vazão
DQO	1.000	0.901	0.070	0.150
SS	0.901	1.000	0.190	0.062
PT	0.070	0.190	1.000	-0.343
Vazão	0.150	0.062	-0.343	1.000

Para os gráficos de porcentagem das variâncias explicadas, novamente a componente principal um (PC-1) carrega a maior parte das informações. Essa, entretanto, é composta somente e majoritariamente pelas variáveis DQO e sólidos suspensos (SS), sendo que diferentemente dos afluentes, a variável de fósforo total (PT) possui uma correlação quase inexistente com as variáveis DQO e SS de acordo com a matriz de correlação.

Entretanto, ao se avaliar o gráfico de escores dos efluentes logo abaixo, é possível identificar dois pontos completamente anômalos ao comportamento dos demais.

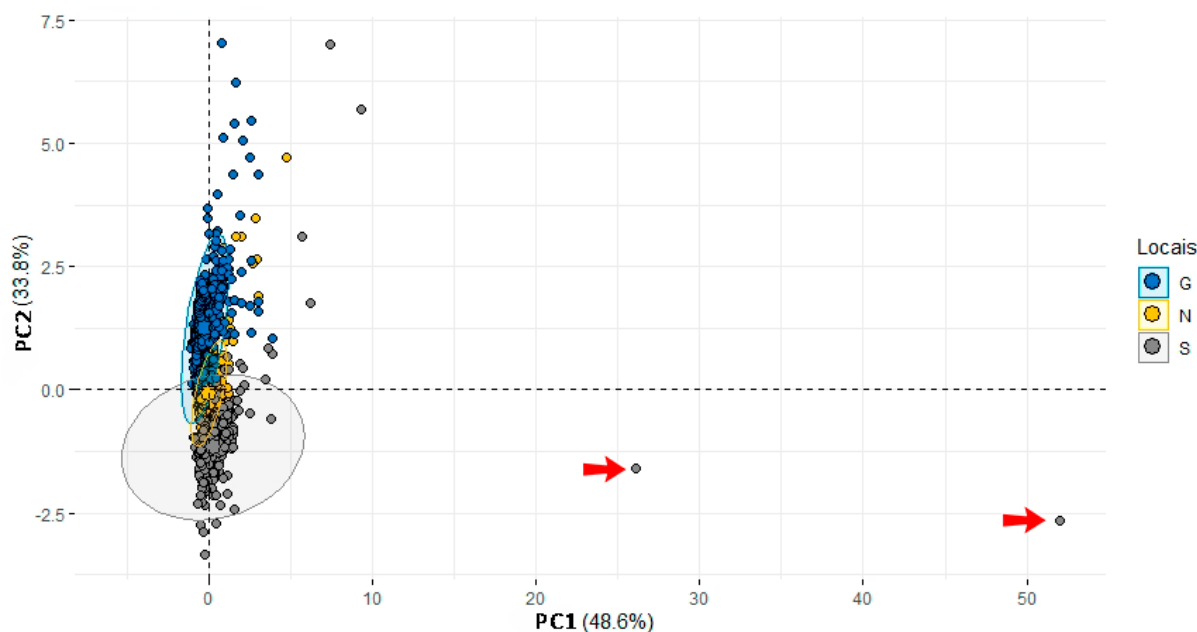


Figura 13 – Gráfico de escores dos efluentes das estações. Amostras com comportamento anômalo identificadas por setas vermelhas.

É possível notar a dificuldade de se agrupar bem os valores dos efluentes de cada estação de tratamento. Para o método aplicado nesse conjunto de dados, a PC1 está sendo decomposta basicamente em função dos valores anômalos, comprimindo assim o restante do conjunto de dados.

Valores anômalos ou *outliers*, quando ocorrem, são informados e notificados pela CAESB. Esses valores em específico ocorreram pela descarga de efluentes não domésticos por caminhões fossa, com alta carga de gordura. Efluentes como este possuem alto teor de carga orgânica e aumentam a suspensão das partículas no afluente que chega à estação, entretanto eles não aumentam os valores de vazão e de fósforo. Pela alta carga orgânica recebida pela estação e por ser uma carga orgânica altamente estabilizada, o tratamento pode não ser tão eficiente, tendo o efluente final retornado às águas pluviais ainda com muita matéria orgânica (MOTA, 2009).

Assim sendo, para melhor aplicabilidade do método e considerando que os valores são completamente irregulares, os mesmos foram retirados do conjunto de dados e a PCA foi recalculada. Os gráficos obtidos estão dispostos abaixo.

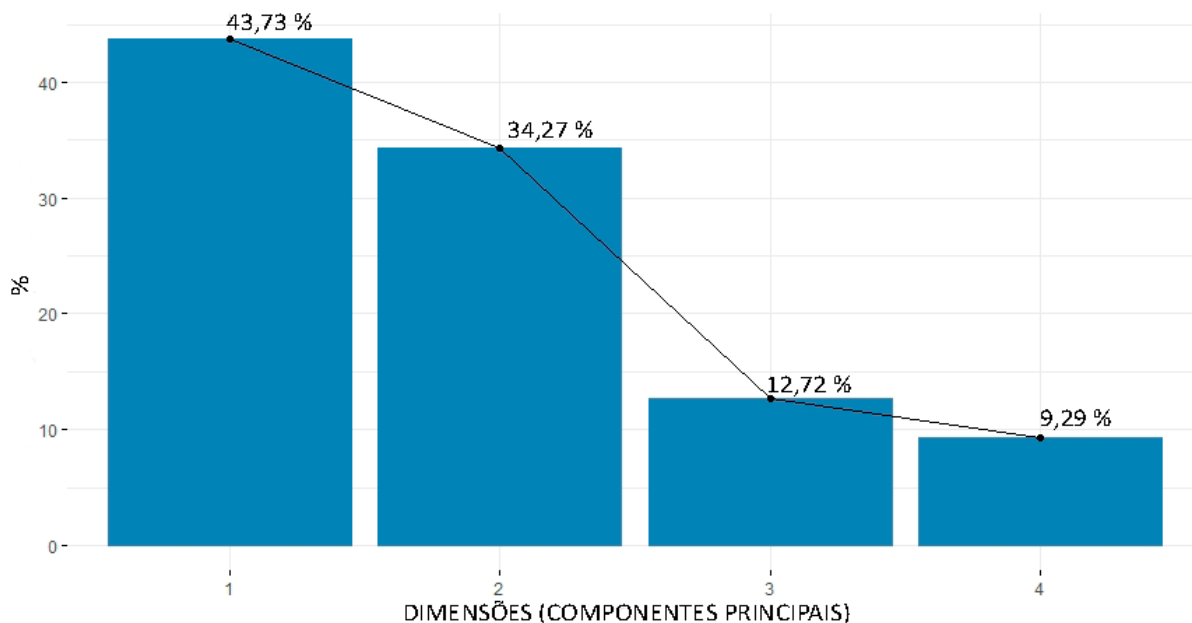


Figura 14 – Gráfico da porcentagem de variâncias explicadas para os efluentes das estações sem os valores anômalos.

Observa-se que a componente principal um (PC-1) teve uma diminuição de variância explicada em relação ao gráfico com os valores anômalos. Entretanto, como as duas componentes principais antes da retirada dos valores anômalos somavam 82,41% dos dados e após a retirada dos valores anômalos, somavam 78%, a diferença foi de apenas 4,41%, o que não acarreta em perda na representatividade do modelo.

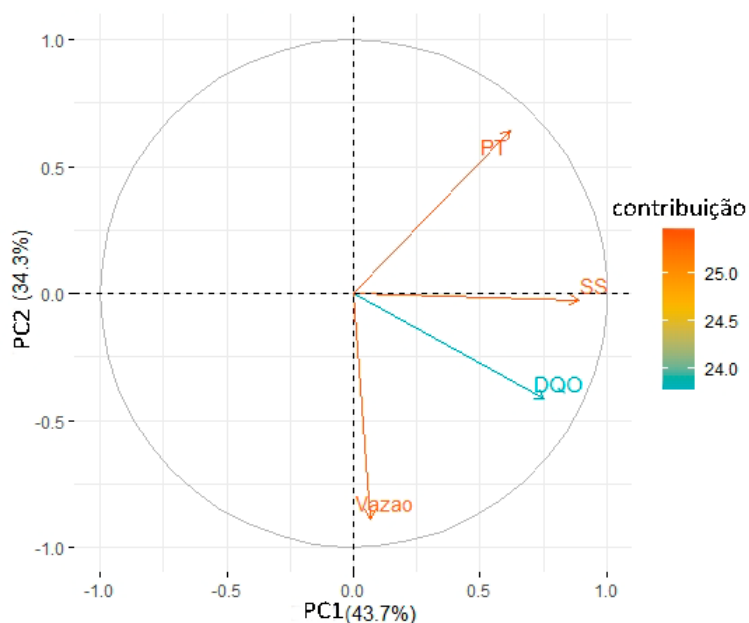


Figura 15 - Gráfico de pesos dos efluentes das estações sem os valores anômalos.

Tabela 3 – Matriz de correlação dos efluentes sem os valores anômalos

	DQO	SS	PT	Vazão
DQO	1.000	0.521	0.135	0.257
SS	0.521	1.000	0.433	0.079
PT	0.135	0.433	1.000	-0.344
Vazão	0.257	0.079	-0.344	1.000

Com a retirada dos valores anômalos, nota-se que o gráfico de pesos muda em relação ao anterior e as correlações também, de acordo com a matriz. A correlação inicialmente forte entre DQO e SS se altera para uma correlação moderada, de forma que os escores inicialmente se apresentavam de forma altamente comprimida e agora não mais, de acordo com o novo gráfico apresentado a seguir, onde identifica-se a melhor separação das amostras de cada ETE em função de PC-2.

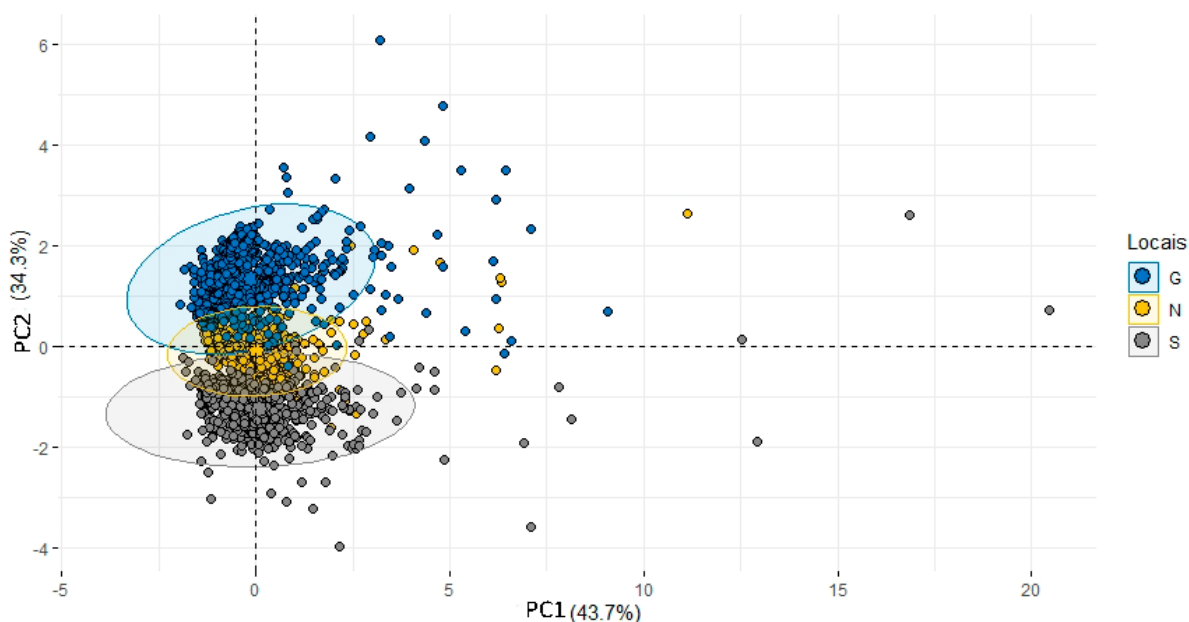


Figura 16 – Gráfico de escores dos efluentes das estações sem os valores anômalos.

Comparando com os dados de afluentes observa-se que os escores das ETEs G, N e S encontram-se mais próximos, entretanto este é um bom resultado, visto que se os três efluentes estão com características muito parecidas, significa basicamente que o tratamento foi similarmente eficiente nas três estações de tratamento. Os valores se distanciam entre si, porém basicamente em função da componente principal dois (PC-2) que, novamente, diz respeito quase que majoritariamente à vazão.

5.3 PCA para cada ETE individualmente

Em seguida, o método PCA foi aplicado individualmente para cada estação de tratamento, de forma a identificar as correlações entre afluente e efluente dentro da própria estação. Os resultados obtidos foram dispostos abaixo.

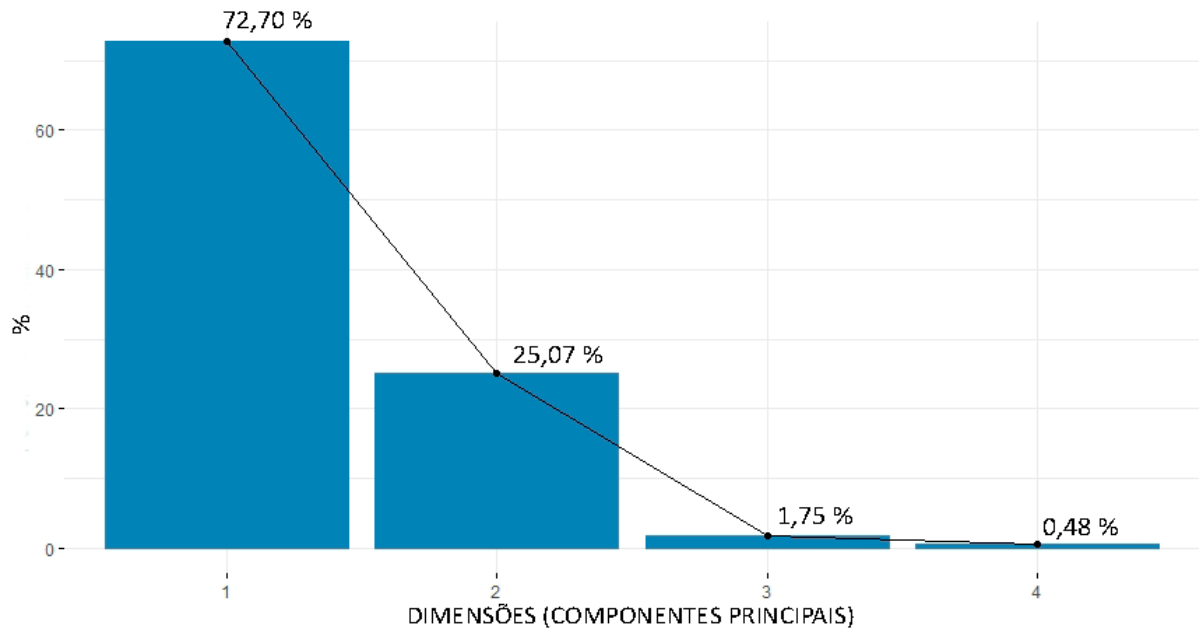


Figura 17 - Gráfico da porcentagem de variâncias explicadas para a ETE Brasília Norte.

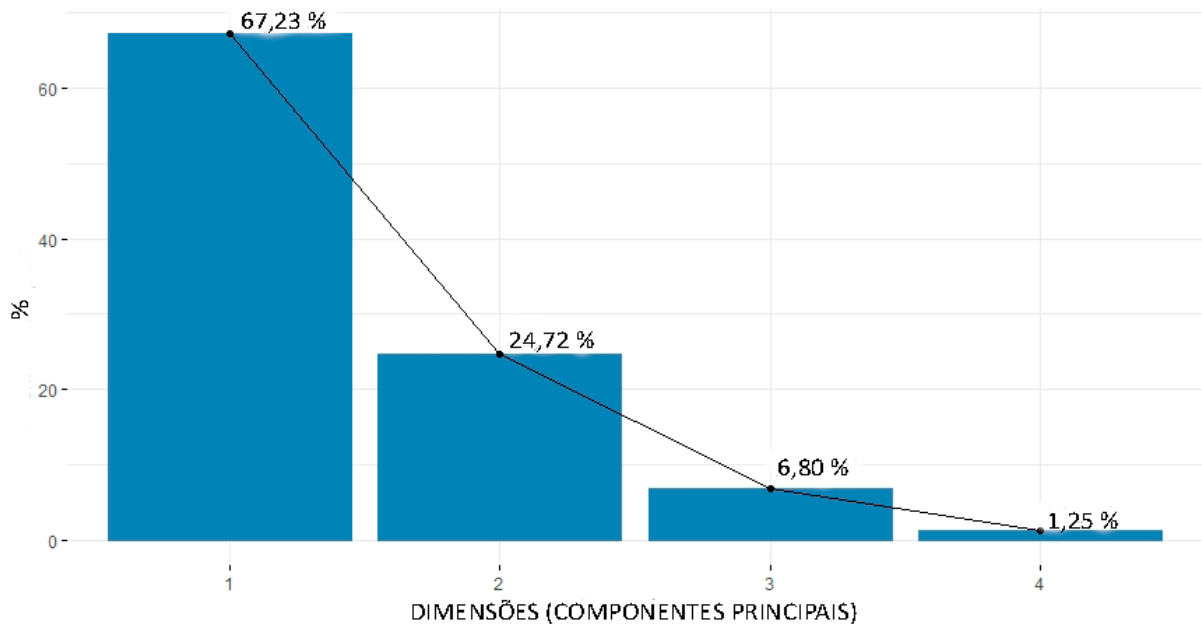


Figura 18- Gráfico da porcentagem de variâncias explicadas para a ETE Brasília Sul.

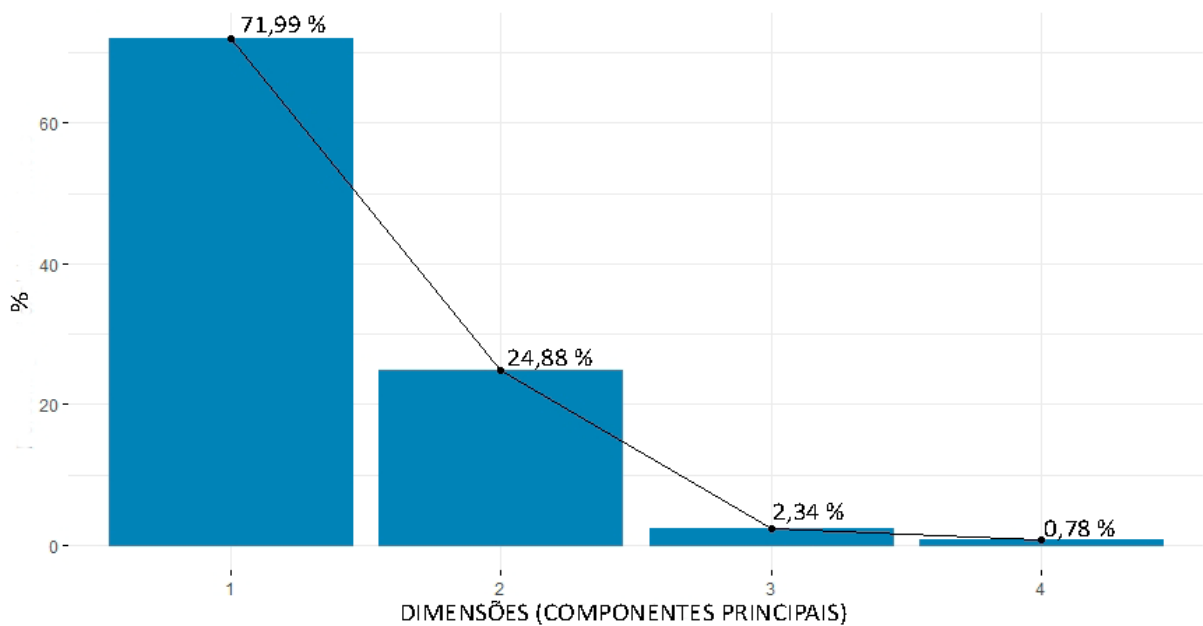


Figura 19 – Gráfico da porcentagem de variâncias explicadas para a ETE Gama.

Ao se avaliar os gráficos de porcentagem de variâncias explicadas, um padrão é bastante explicitado: a PC-1 carrega a maior parte da informação. Como a mesma é composta pelos dados de DQO, sólidos suspensos (SS) e fósforo total (PT), de acordo com os gráficos de pesos, dispostos em uma mesma imagem logo abaixo, há um forte indicativo que tanto em afluente, quanto em efluente, em qualquer uma das três estações, as três variáveis estão diretamente correlacionadas; enquanto que a vazão se mantém quase que ortogonal à componente principal dois (PC-2). O que faz sentido, como anteriormente discutido.

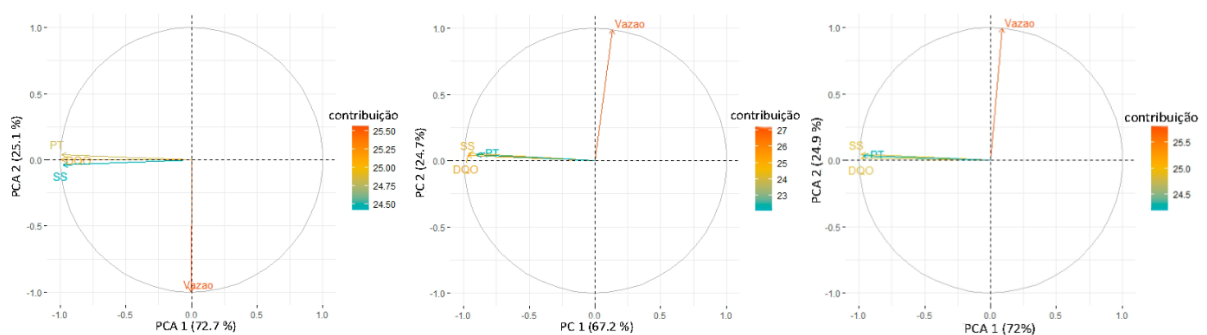


Figura 20 – Gráficos de pesos para as três estações de tratamento de esgoto. Resultados, da esquerda para a direita, respectivamente, da ETE Norte*, ETE Sul e ETE Gama.

**A inversão da seta de vazão na ETE Norte em relação às demais se trata de uma ambiguidade matemática do método PCA, de forma que não foram observadas nenhuma diferença significativa entre os resultados das três ETEs.*

Se tratando dos gráficos de escores de cada estação de tratamento, plotados logo abaixo, é possível identificar de maneira bem clara a separação em relação ao afluente e ao efluente.

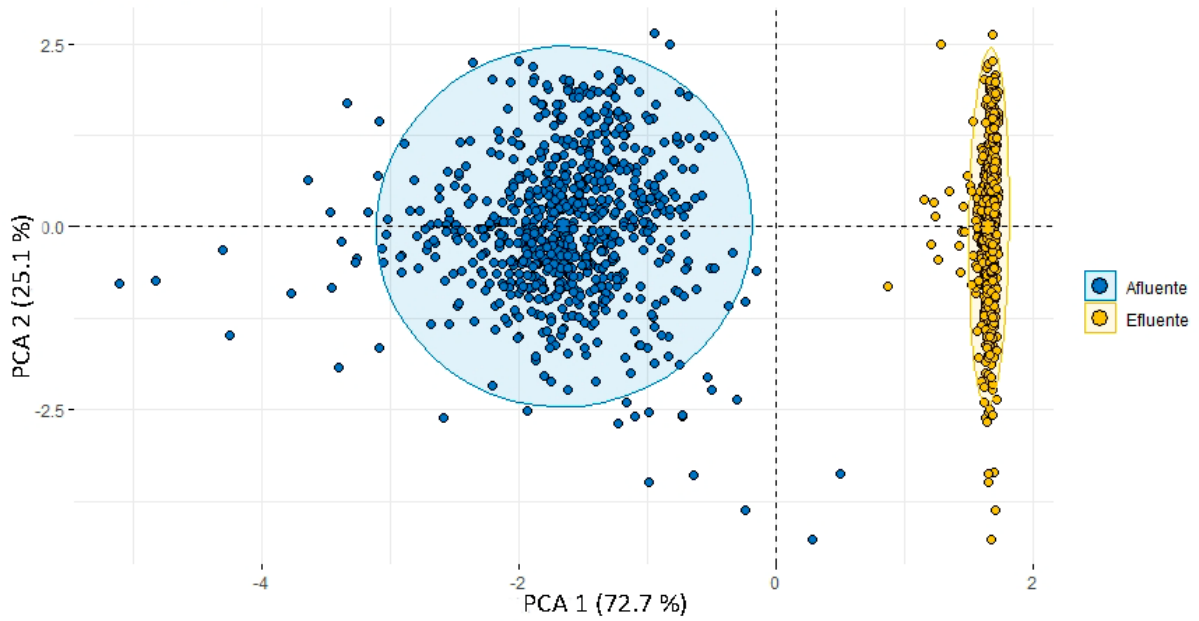


Figura 21 – Gráfico de escores da ETE Brasília Norte. Afluente (●) e Efluente (●)

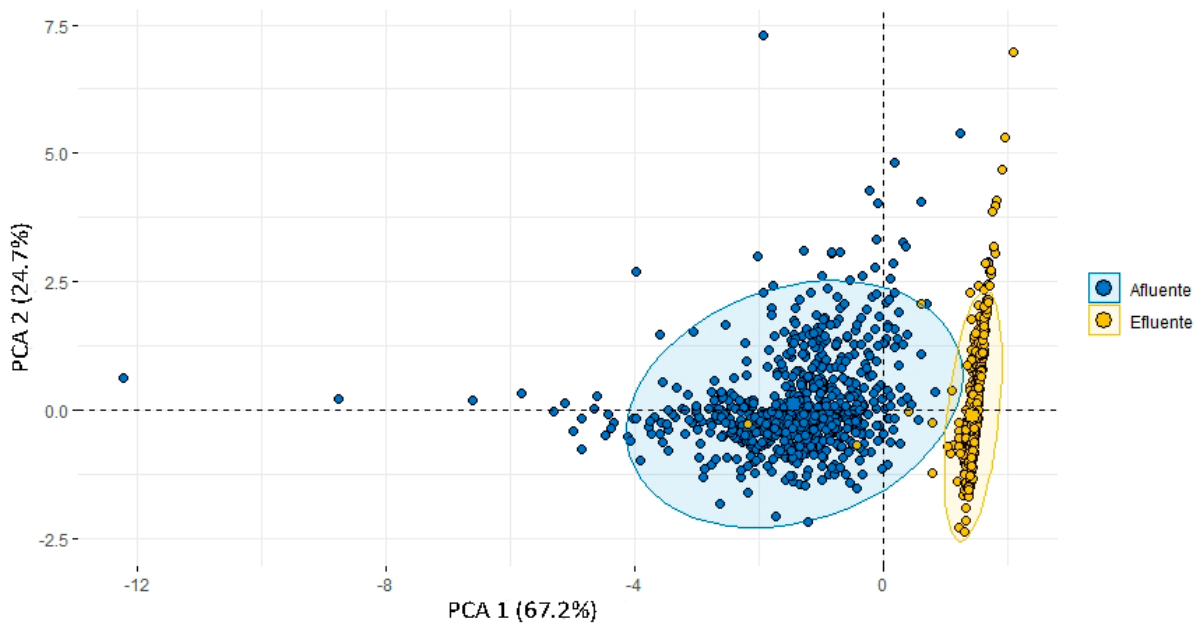


Figura 22 – Gráfico de escores da ETE Brasília Sul. Afluente (●) e Efluente (●)

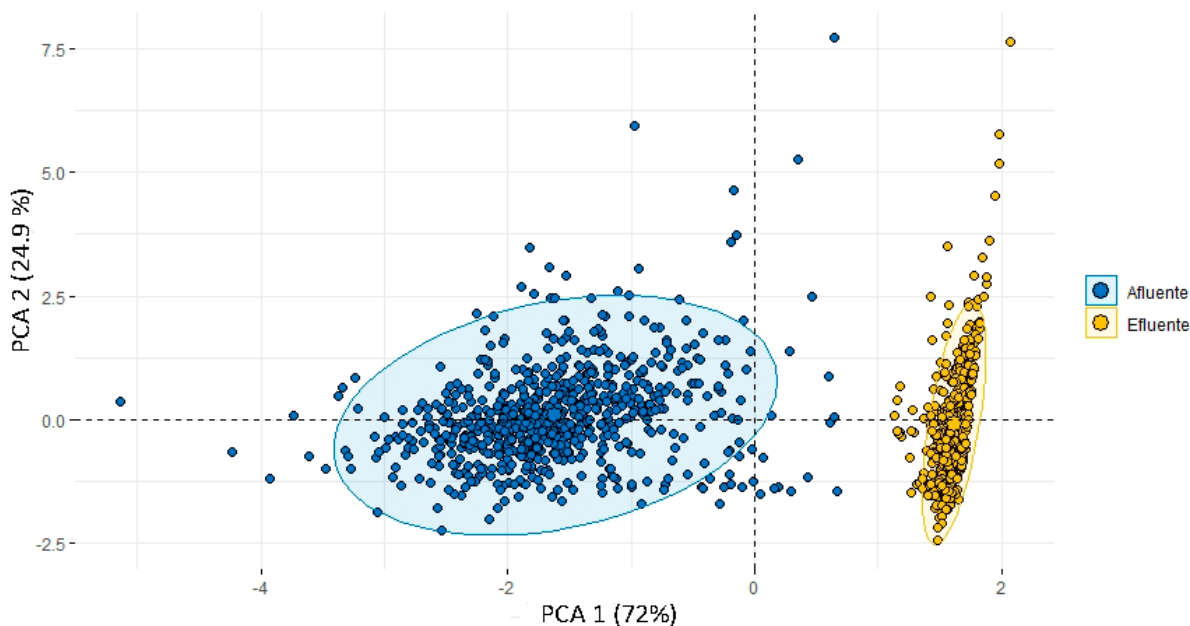


Figura 23 – Gráfico de escores da ETE Gama. Afluente (●) e Efluente (●)

Essa separação é um resultado bastante positivo e esperado, considerando que o afluente chega à estação com valores de variáveis quase que aleatórios e após o tratamento é possível identificar um efluente uniforme, demonstrando alta eficiência no tratamento. Além disso, vale ressaltar que entre as três estações de tratamento, a que possui um tratamento mais eficiente e que deixa o efluente mais uniforme, de acordo com os gráficos de escores, é a ETE Sul, visto que os resultados nessa ETE formam um *cluster* altamente unificado em comparação às outras duas.

5.4 PCA realizada com apenas os dados referentes ao ano de 2020 acrescentando variáveis

Como discutido anteriormente, as variáveis NT, DBO e N-NH₃, foram inicialmente descartadas por possuírem baixa periodicidade de análise e conseqüentemente não possuírem dados suficientes em relação à amostragem quase diária das variáveis consideradas na seção anterior. Então, de forma a tentar correlacioná-las, o conjunto de dados foi dividido semanalmente, apenas no ano de 2020, onde possuía maior preenchimento, e apenas um valor foi considerado por semana. Em análises que possuíam mais de um valor durante a semana, considerou-se a média. Os resultados são apresentados a seguir.

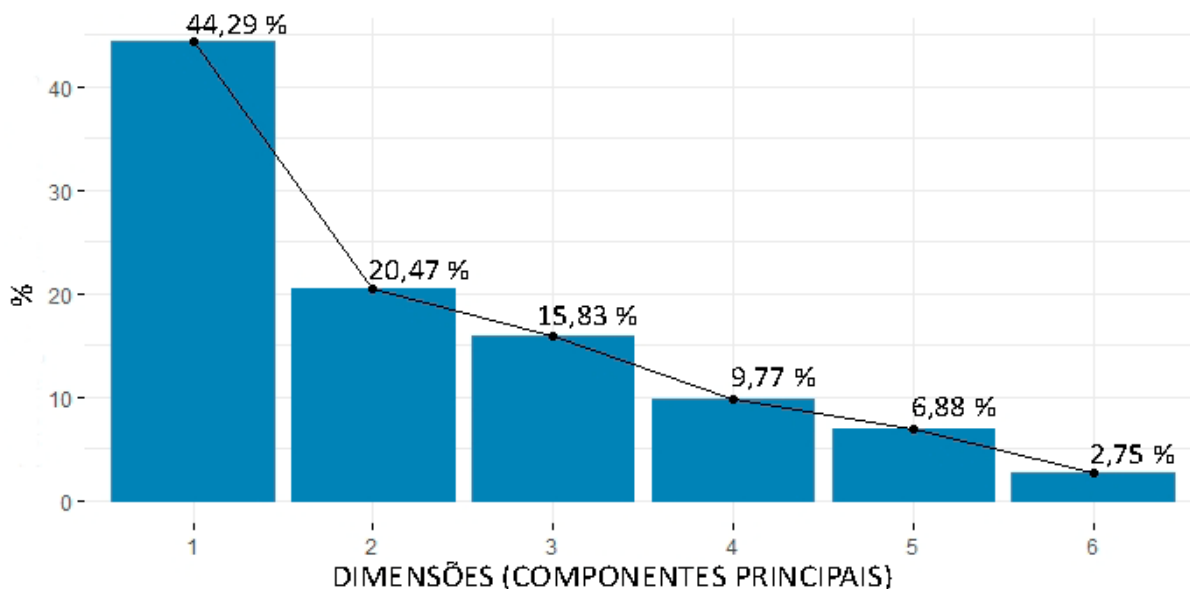


Figura 24 – Gráfico da porcentagem de variâncias explicadas dos afluentes das estações acrescentando mais variáveis.

Ao se analisar o gráfico de porcentagem das variâncias explicadas, observa-se que a componente principal 1 (PC-1) explica cerca de 44% das informações do conjunto de dados e ela é composta majoritariamente por DBO, sólidos suspensos (SS), DQO e fósforo total (PT). Sendo que a componente principal 2 (PC-2) é composta quase que inteiramente por vazão e nitrogênio total (NT).

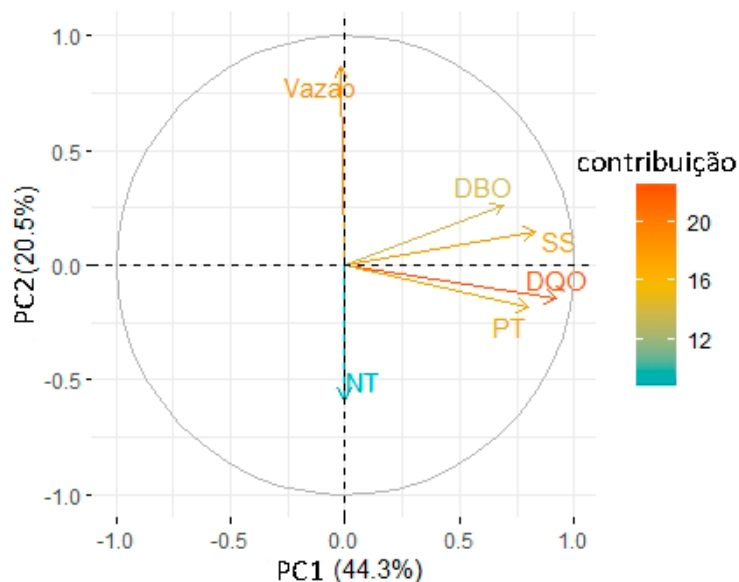


Figura 25 – Gráfico de pesos dos afluentes das estações acrescentando mais variáveis.

Tabela 4 – Matriz de correlação em análise de afluentes com mais variáveis.

	DQO	SS	PT	DBO	NT	Vazão
DQO	1.000	0.708	0.651	0.547	0.076	-0.131
SS	0.708	1.000	0.451	0.353	-0.085	0.071
PT	0.651	0.451	1.000	0.386	-0.029	-0.157

DBO	0.547	0.353	0.387	1.000	-0.006	0.204
NT	0.076	-0.085	-0.028	-0.006	1.000	-0.128
Vazão	-0.131	0.071	-0.157	0.204	-0.128	1.000

A componente principal um (PC-1), como mostrada pelo gráfico de pesos se refere basicamente às variáveis DBO, SS, DQO e PT; sendo que a componente principal dois (PC-2) se correlaciona de maneira inversa aos valores de NT e vazão. Para compreender a não relação entre NT e as demais variáveis, é necessário identificar a origem do nitrogênio em uma estação de tratamento de esgoto. O nitrogênio que chega à estação é proveniente, majoritariamente, de urina e degradação de aminoácidos (MOTA, 2009). Dessa forma, é viável relacioná-lo inversamente com a vazão, visto que quanto maior a vazão, mais diluída será a urina que é despejada no esgoto e que carrega esses compostos. Além disso, as variáveis restantes dizem respeito majoritariamente à matéria orgânica carbonácea, como discutido por BERTOLAMI, (2017) e VALENTE (1997), não tendo relação necessariamente direta com a concentração de NT.

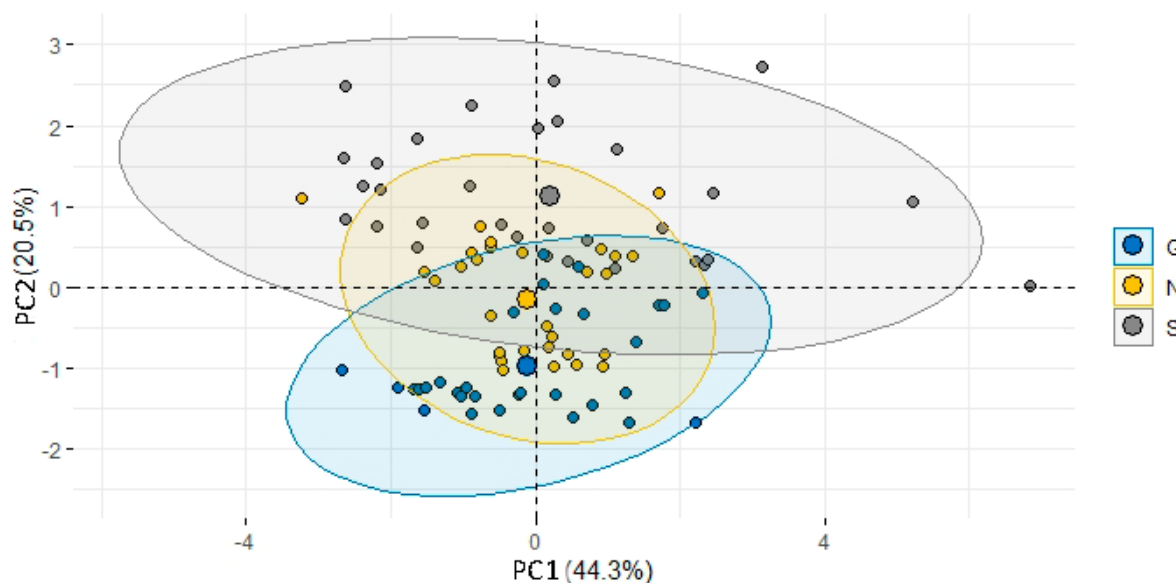


Figura 26 – Gráfico de escores dos afluentes das estações, acrescentando mais variáveis.

Ao se analisar o gráfico de escores dos afluentes da estação (Figura 31), é esperado que o afluente não apresente uma clara divisão em grupos, visto que não se controla o que chega à estação. Entretanto, pode-se observar uma separação discreta entre as ETEs em PC-2, sendo que a ETE Gama apresenta escores mais negativos, seguida da ETE Norte e depois ETE Sul, que apresenta valores positivos.

Se tratando dos efluentes da estação, a análise foi repetida e os resultados estão dispostos abaixo.

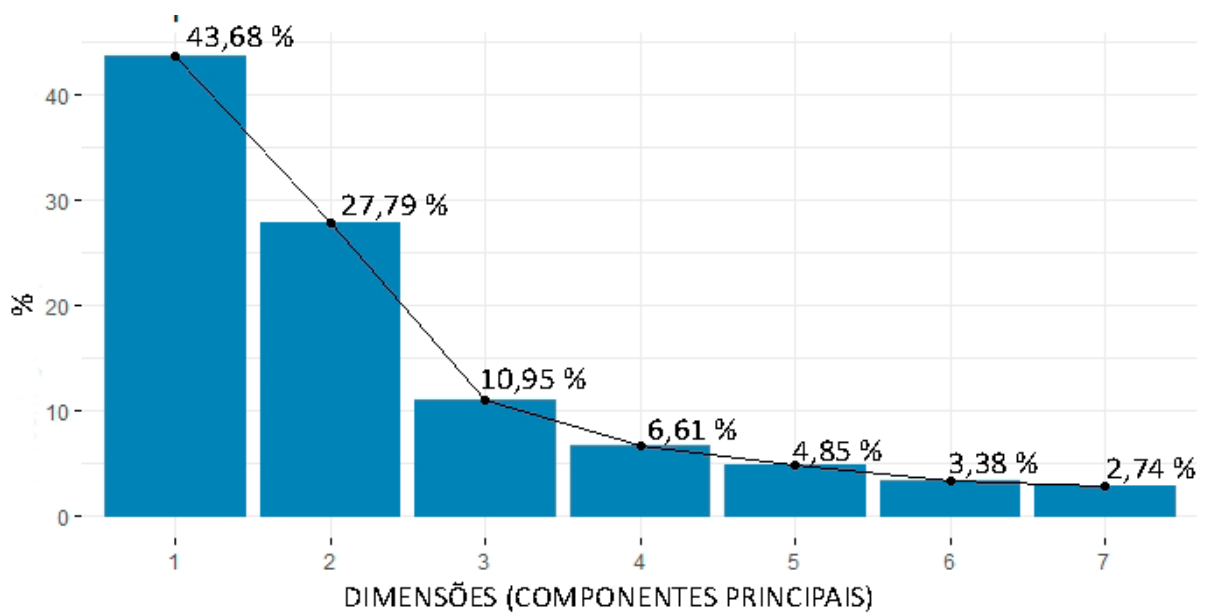


Figura 27 – Gráfico da porcentagem de variâncias explicadas dos efluentes das estações acrescentando mais variáveis.

Aqui é importante notar que o gráfico de porcentagem de variâncias explicadas possui uma variável a mais, em relação à análise de afluentes, que se refere ao nitrogênio amoniacal (N-NH₃), pois este é medido somente no efluente.

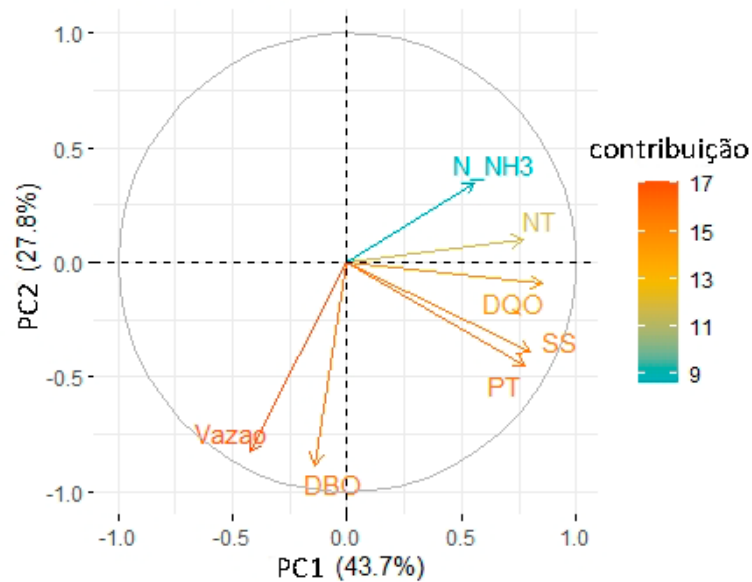


Figura 28 - Gráfico de pesos dos efluentes das estações acrescentando variáveis de menor periodicidade.

Tabela 5 – Matriz de correlação em análise de efluentes com mais variáveis.

	DQO	SS	PT	DBO	NT	N-NH ₃	Vazão
DQO	1.000	0.645	0.576	-0.038	0.512	0.407	-0.251
SS	0.645	1.000	0.777	0.181	0.427	0.244	-0.028
PT	0.576	0.777	1.000	0.235	0.536	0.233	0.022

DBO	-0.038	0.181	0.235	1.000	-0.159	-0.218	0.716
NT	0.512	0.427	0.536	-0.159	1.000	0.251	-0.349
N_NH3	0.407	0.244	0.233	-0.218	0.251	1.000	-0.430
Vazão	-0.251	-0.028	0.022	0.716	-0.349	-0.430	1.000

Avaliando-se o gráfico de pesos e a matriz de correlação, é possível aferir que a DBO está se correlacionando de maneira significativa com a vazão na saída da estação pela componente principal dois (PC-2). Enquanto as outras variáveis possuem correlações medianas entre si pela componente principal um (PC-1), o que seria naturalmente esperado de um efluente mineralizado ao término de seu tratamento.

Observando-se exclusivamente os dados, seria possível afirmar que como a DBO não se correlaciona diretamente de maneira relevante com as outras variáveis, o tratamento nessas estações seria mais eficaz na remoção de micronutrientes como fósforo e nitrogênio e menos eficiente na remoção de matéria orgânica carbonácea. Entretanto, ao se comparar a carga média (variável multiplicada pela vazão média) de saída e de entrada de todas as variáveis das estações aqui presentes, nota-se as seguintes porcentagens de remoção:

Tabela 6 – Remoção em porcentagem de cada parâmetro nas estações de tratamento.

	DQO	SS	PT	DBO	NT	Estação
Remoção (%)	94,74	90,77	87,55	94,25	38,18	ETE Brasília Norte
	95,52	97,44	95,37	97,14	83,92	ETE Brasília Sul
	94,60	97,00	95,78	98,28	53,45	ETE Gama

A remoção de DBO em todas as estações de tratamento de esgoto é altamente eficiente como mostra a Tabela 5, sendo em todas as estações superior a 94%. E em comparação com a remoção de NT, por exemplo, é altamente superior, podendo-se afirmar que nitrogênio é mais dificilmente eliminado como micronutriente em comparação a materiais carbonáceos.

Dessa forma, a explicação mais viável para essa variável ter se deslocado das demais em relação a PC-1 é o baixo preenchimento dessa análise no conjunto de dados em relação às demais, podendo o método não ter sido altamente eficiente para explicação dessa variável em específico.

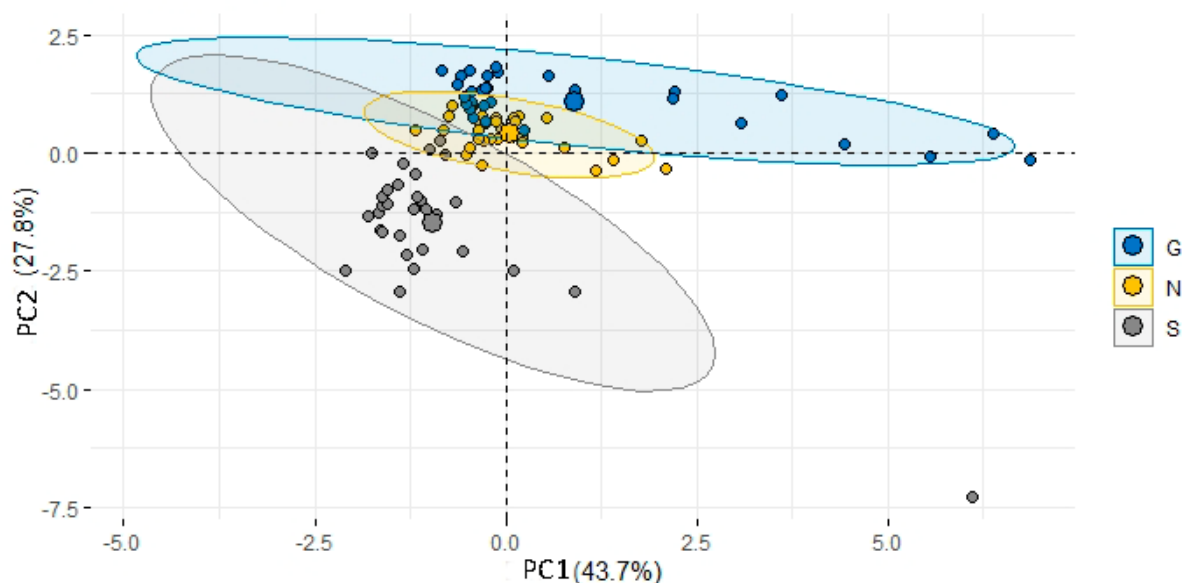


Figura 29 - Gráfico de escores dos efluentes das estações, acrescentando mais variáveis.

Por fim, ao se avaliar o gráfico de escores, por mais que o conjunto de dados seja pequeno, é possível identificar novamente uma separação razoável entre as três ETEs. A ETE com resultados de menor variação é a ETE Norte e isso se dá, principalmente, pela etapa a mais de equalização a mais que a ETE Norte possui. Os tanques de equalização conseguem reter o líquido que chega, evitando passagem desnecessária de efluente em momentos de baixa movimentação.

6. CONCLUSÃO

Os resultados mostraram que os parâmetros DQO, SS e PT estão diretamente correlacionados, tanto no efluente, quanto no afluente das estações. Além disso, foi possível identificar, ao se avaliar as estações individualmente, que a ETE Brasília Sul possui um tratamento mais eficiente por produzir um efluente mais uniforme.

A PCA se mostrou eficiente na evidenciação de similaridades entre as ETES. Além disso, a partir dessas análises, foi possível identificar valores anômalos, buscar a causa desses valores anômalos, retirá-los do banco de dados e repetir a análise obtendo resultados mais expressivos. O método, entretanto, não se mostrou eficiente para a análise de dados com grandes quantidades de valores ausentes.

Por meio do gráfico de escores, foi possível identificar que após o tratamento os efluentes das três ETEs apresentam maior similaridade do que o afluente, o que contribui pra afirmar que a empresa possui uma alta qualidade de eficiência no seu tratamento. Além de que foi possível observar que em caso de resultados anômalos, como os identificados neste trabalho,

a empresa possui um controle total do ocorrido, notificando em seu sistema interno a causa e acionando as equipes responsáveis para resolução do problema.

Este trabalho também possibilitou um primeiro contato com a ferramenta de Análise de Componentes Principais (PCA), que é um modelo amplamente utilizado para a análise exploratória de dados multivariados nas mais diferentes áreas da Ciência.

7. PERSPECTIVAS

- Colaboração com a CAESB em futuros trabalhos;
- Modelagem estatística de dados para anos irregulares;
- Identificação de possíveis falhas de operação por análise do banco de dados;

8. BIBLIOGRAFIA

- ANGELAKIS, A.N.; SNYDER, S.A. “Wastewater Treatment and Reuse: Past, Present, and Future”. Revista Water 2015, 7a ed.; p. 4887-4895.
- ANJUM, M.; RASHEED, H.; AHMED, W. (2016). “Impact of waste water treatment on quality of influent & effluent water.” Jornal Imperial de Pesquisas Interdisciplinares (IJIR). 2a ed. p. 844-849.
- APHA, A. (2005). *Standard methods for the examination of water and wastewater*. New York. American Public Health Association, American Water Works Association, Water Environment Federation.
- APU, A. (2018). “Introduction to sewage treatment.” Dhaka University of Engineering & Technology (DUET), Gazipur.
- BERTOLAMI, G. K. (2017). A técnica de Demanda Bioquímica de Oxigênio e o kit rápido da MachereyNagel NANOCOLOR® BOD5-TT. São Paulo: Carvalhaes.
- CEBALLOS, B.S.O. de; DINIZ, C.R. Técnicas de microbiologia sanitária e ambiental. Campina Grande: EDUEPB, 2017. 11000kb. 324 p.: il. Disponível em <www.uepb.edu.br/ebooks/> Acesso em 8 nov de 2020.
- CETESB (2017). Qualidade das águas subterrâneas no estado de São Paulo – Boletim de 2017. SP - Brasil. Disponível em <<http://www.cetesb.sp.gov.br>> Acesso em 11 nov 2020.
- COMPANHIA DE SANEAMENTO AMBIENTAL DO DISTRITO FEDERAL (CAESB) (2015). Brasília, DF. Sistemas de Esgotamento. Disponível em: <<https://www.caesb.df.gov.br/esgoto/sistemas-de-esgotamento.html>> Acesso em 11 nov. 2020.
- DIRETORIA COLEGIADA (2014). Relatório anual da administração (PDF). CAESB. Consultado em 12 de novembro de 2020.
- FOGLER, H. S.; EATON, A.D.; RICE, E.W. “Elements of chemical reaction engineering”. 4. ed. Nova Jersey: Prentice Hall, 2005. 1120 p.
- GRANGER, C.W.J.; HYUNG, N. Correlação espúria em modelos de painel: uma análise do desmatamento da Amazônia*. Pesquisa e Planejamento Econômico (PPE) - Artigos, Instituto de Pesquisa Econômica Aplicada, ano 1997, v. 27, n. 3, 8 dez. 1997. Meio Ambiente. Recursos Naturais: Artigos, p. 462-490.
- HALLS, S.; YAMAZAKI, K. Qualidade da água: O impacto da eutrofização. 1. ed. Brasília: Instituto Internacional de Ecologia, 2002. v. 3, cap. De onde vêm os nutrientes e como eles provocam a eutrofização, p. 13-17. Disponível em: <<http://www.blogdocancado.com/wp-content/uploads/2011/02/eutrofizacao1.pdf>> Acesso em: 8 nov. 2020.
- HOLLAND, S.M. “Principal components analysis (PCA)”. Universidade da Geórgia (2008), Atenas.
- KONZEN, G.B.; FIGUEIREDO, J.A.S.; QUEVEDO, D.M. História da qualidade dos parâmetros de água. – um estudo no rio Sinos/Brasil. Brasil. J. Biol., São Carlos, v. 75, n. 2, supl. p. 1-10, Maio, 2015. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-69842015000300003&lng=en&nrm=iso>. Acesso em 12 Nov. 2020. <http://dx.doi.org/10.1590/1519-6984.0213>.
- LOUGON, M. S., ROCHA, S. A., GUIMARÃES, H. F., LOUZADA, F. L., & GARCIA, G. D. (2011). Caracterização dos sólidos totais, fixos e voláteis nas águas. Jerônimo Monteiro: Universidade do Vale do Paraíba.
- MENDONÇA, H. V.; RIBEIRO, C. B. M.; BORGES, A. C.; BASTOS, R. R. Remoção de nitrogênio e fósforo de águas residuárias de laticínios por sistemas alagados construídos

- operando em bateladas. *Ambi-Água*, Taubaté, v. 7, n. 2, p. 75-87, 2012. (<http://dx.doi.org/10.4136/ambi-agua.805>)
- MOTA, F.S.B.; SPERLING, M.V. (2009) Nutrientes de esgoto sanitário: utilização e remoção. Rio de Janeiro: ABES, 2009 428p. Projeto PROSAB. Acesso em 07 nov. de 2020.
- MURTHA, N.A.; CASTRO, J.E.; HELLER, L. Uma perspectiva histórica das primeiras políticas públicas de saneamento e de recursos hídricos no Brasil. *Ambiente e soc.*, São Paulo, v.18, n.3, p.193-210, set.2015. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2015000300012&lng=pt&nrm=iso>. Acesso em 12 nov. 2020. <https://doi.org/10.1590/1809-4422ASOC1047V1832015>.
- NATHANSON, J.A.; AMBULKAR, A. “*Wastewater treatment*”. Oxford: Enciclopédia Britânica, 2020. 17 p.
- URBANIZADORA PARANOAZINHO (2012). ETE de sobradinho passa por ampliação para atender condomínios da região. UPSA, 2012. Disponível em: <<http://www.upsa.com.br/2012/11/28/ete-de-sobradinho-passa-por-ampliacao-para-atender-condominios-da-regiao/>>. Acesso em: 18, de nov. de 2020.
- RIBEIRO, A. Partes de um rio; Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/geografia/partes-um-rio.htm>. Acesso em 12 de novembro de 2020.
- SERVIÇO AUTÔNOMO DE ÁGUA E ESGOTO (SAAE). Tratamento de Esgoto. Cerquilha, São Paulo, 15 jan. 2020. Disponível em: <https://www.saaec.com.br/esgoto/tratamento-de-esogoto/#:~:text=Assim%2C%20podemos%20ent%C3%A3o%20separar%20tratamento,tratamento%20terci%C3%A1rio%20ou%20p%C3%B3s%20tratamento.> Acesso em: 10 nov. 2020.
- COMPANHIA DE SANEAMENTO BÁSICO DO ESTADO DE SÃO PAULO (SABESP). Tipos de Tratamento. Disponível em: <http://site.sabesp.com.br/site/interna/Default.aspx?secaoId=61>. Acesso em: 10 nov. 2020.
- SHLENS, J. “*A Tutorial on Principal Component Analysis*”. Instituto Salk para Estudos Biológicos e Universidade da Califórnia, 2ª edição, Dezembro de 2005.
- SMITH, L. I. “*A tutorial on Principal Components Analysis*”, Fevereiro de 2002.
- SOUZA, A.M.d.; POPPI, R.J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: Um tutorial, Parte i. *Química Nova*, Campinas, v. 35, n. 1, p. 223-229, 22 jul. 2011. Semanal.
- TCHOBANOGLIOUS, G.; BURTON, F.L.; STENSEL, H. D. (2003) “*Wastewater engineering: treatment and reuse*”. 4ª ed.; República Popular da China; Companhia McGraw. 1819 p.
- VALENTE, J. S., PADILHA, P. M., & SILVA, A. M. (1997). Oxigênio dissolvido (OD), demanda bioquímica de oxigênio (DBO) e demanda química de oxigênio (DQO) como parâmetros de poluição no ribeirão Lavapés/Botucatu - SP. Universidade Estadual Paulista, Botucatu.
- VALDERRAMA, L. *et al.* Proposta experimental didática para o ensino de análise de componentes principais. *Quím. Nova*, Brasília, v. 39, n. 2, p. 245-249, 26 out. 2015. Semanal.