



Universidade Estadual Paulista  
"Júlio de Mesquita Filho"

Programa Interunidades

unesp

Mestrado

---

## Engenharia Civil e Ambiental

**ANTONIO MARCOS GALVEZ SERRA**

**MODELAGEM EMPÍRICA PARA AJUSTE, CORREÇÃO E ANÁLISE DE  
CONFIABILIDADE DE PARÂMETROS DE REATORES UASB**

Bauru

2019

**ANTONIO MARCOS GALVEZ SERRA**

**MODELAGEM EMPÍRICA PARA AJUSTE, CORREÇÃO E ANÁLISE DE  
CONFIABILIDADE DE PARÂMETROS DE REATORES UASB**

Dissertação apresentada como requisito para a obtenção do título de Mestre em Engenharia Civil e Ambiental da Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Área de Concentração: Saneamento Ambiental.

Orientador: Prof. Dr. Rodrigo Braga Moruzzi.

Coorientador: Prof. Dr. Gustavo Henrique Ribeiro da Silva.

Bauru

2019

Serra, Antonio Marcos Galvez.

Modelagem empírica para ajuste, correção e análise de confiabilidade de parâmetros de reatores UASB.

- Bauru - SP, 2019  
97 f.

Orientador: Rodrigo Braga Moruzzi  
Coorientador: Gustavo Henrique Ribeiro da Silva.

Dissertação (Mestrado)-Universidade Estadual Paulista. Faculdade de Engenharia, Bauru, 2019

1. Estação de tratamento de esgoto. 2. Modelo matemático.  
3. Reatores UASB. I. Universidade Estadual Paulista. Faculdade de Engenharia. II. Título.

**ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE Mestrado de ANTONIO MARCOS GALVEZ SERRA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL E AMBIENTAL, DA FACULDADE DE ENGENHARIA - CÂMPUS DE BAURU.**

Aos 01 dias do mês de novembro do ano de 2019, às 14:00 horas, no(a) Instituto de Geociências e Ciências Exatas - Câmpus de Rio Claro, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. RODRIGO BRAGA MORUZZI - Orientador(a) do(a) Departamento de Planejamento Territorial e Geoprocessamento - DEPLAN / Instituto de Geociências e Ciências Exatas (IGCE) de Rio Claro – UNESP, Prof. Dr. ARNALDO SARTI do(a) Departamento de Bioquímica e Tecnologia Química / Instituto de Química - UNESP - Araraquara, Prof. Dr. SAMUEL CONCEIÇÃO DE OLIVEIRA do(a) Departamento de Bioprocessos e Biotecnologia / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE Mestrado de ANTONIO MARCOS GALVEZ SERRA, intitulada **DESENVOLVIMENTO DE UMA ESTRUTURA DE MODELAGEM EMPÍRICA PARA AJUSTE E CORREÇÃO DE PARÂMETROS VOLTADOS A REATORES UASB E ANÁLISE DA CONFIABILIDADE DO SISTEMA**. Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: Aprovado. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. RODRIGO BRAGA MORUZZI

Prof. Dr. ARNALDO SARTI

Prof. Dr. SAMUEL CONCEIÇÃO DE OLIVEIRA

PROPOSTA DE ALTERAÇÃO DO TÍTULO

A COMISSÃO EXAMINADORA PROPÕE A ALTERAÇÃO DO TÍTULO DO TRABALHO DO ALUNO:  
**ANTONIO MARCOS GALVEZ SERRA**

DE: "DESENVOLVIMENTO DE UMA ESTRUTURA DE MODELAGEM EMPÍRICA PARA AJUSTE E CORREÇÃO DE PARÂMETROS VOLTADOS A REATORES UASB E ANÁLISE DA CONFIABILIDADE DO SISTEMA"

PARA:

Modelagem empírica para  
ajuste, correção e análise de confiabilidade de parâmetros de reatores UASB.

Bauru, 01 de novembro de 2019.

  
Prof. Dr. Rodrigo Braga Moruzzi  
Orientador

Ao meu orientador, professor Rodrigo Braga Moruzzi, um pessoa admirável pelo seu conhecimento; mesmo a distância, compartilhou suas ideias e experiências, me orientando nos passos a seguir, para a construção desta dissertação.

Ao meu coorientador, professor Gustavo Henrique Ribeiro que propôs o estudo da Estação de tratamento de Esgoto de Candeia, me auxiliando no contato com a responsável técnica para dar início à dissertação, e que também me acompanhou em todas as etapas até a minha qualificação.

A Giselda Passos Giaferis, técnica Química do DAE, responsável pela estação de tratamento ETE Candeia na qual me auxiliou com os trâmites para ser autorizado e ter acesso aos resultados das análises laboratoriais dos parâmetros envolvidos nesta pesquisa; por me ter apresentado a literatura química e as informações já disponíveis da ETE de Candeia para a construção deste trabalho.

Aos técnicos do Laboratório de águas residuárias da UNESP- BAURU, pelo fornecimento solícito das tabelas com as medições laboratoriais dos parâmetros físico-químicos e biológicos da ETE Candeia.

## AGRADECIMENTOS

Principal e inicialmente a Deus.

Aos meus pais pelo apoio e custeio dos estudos que hoje me fizeram chegar até aqui, em especial ao meu pai, que mesmo contraindo um AVC durante esta dissertação, continuou me dando força para que eu levasse esse trabalho adiante.

A minha família, pela paciência e colaboração durante o decorrer desta pesquisa, que por muitas vezes possa ter me omitido em situações cotidianas em prol desta dissertação.

Ao meu amigo Marcos Henrique de Paula Dias da Silva pela ajuda na elaboração e redação, por seus conhecimentos matemáticos, que propiciou auxílio em softwares específicos estatísticos, que me permitiram construir as modelagens apresentadas neste trabalho.

## Resumo

A falha humana é um evento indesejado para os engenheiros, e a população que se vê refém destes empreendimentos, (visto os diversos incidentes que ocorrem no Brasil) por fatores como omissão de informações, medições equivocadas ou mesmo erros técnicos, passam despercebidos durante as vistorias. Na área de Saneamento, em se tratando de esgoto sanitário, os riscos não são menores e também devem ser evitados considerando o número de doenças que são proliferadas por falhas no saneamento de esgoto. A partir deste pressuposto, a presente pesquisa visou desenvolver um modelo preditor, usando ferramentas matemáticas que viabilizassem identificar e corrigir erros técnicos no registro e medições realizadas em parâmetros físicos, químicos e biológicos de dois reatores UASB da Estação Candeia da cidade de Bauru- SP. Para isto, após uma revisão sobre a tecnologia de reatores UASB e de estudos probabilísticos baseados no conceito de confiabilidade aplicado a ETEs (Estações de Tratamento de Esgotos), definiu-se, por meio de análise de regressão, um modelo multilinear, acoplado a um algoritmo de otimização, capaz de prever e corrigir parâmetros de entrada e saídas (inconsistentes ou ausentes) no registro do monitoramento da estação referentes ao tratamento de reatores UASB. Com este procedimento matemático, através de uma análise estatística conceitual de confiabilidade do sistema, foi possível corrigir dados e estimar informações de parâmetros ausentes, em 22% das coletas de medidas e registros nos relatórios técnicos que foram disponibilizados para este estudo.

Palavras-chaves: Esgoto Sanitário. Reator UASB. Análise de Confiabilidade. Regressão Linear Múltipla. Modelagem Matemática.



## **Abstract**

Human failure is an unwanted event for engineers, and the population that is hostage to these endeavors, (due to the various incidents that occur in Brazil) due to factors such as omission of information, wrong measurements or even technical errors, go unnoticed during surveys. . In the Sanitation area, when it comes to sanitary sewage, the risks are no less and should also be avoided considering the number of diseases that are proliferated by sewage sanitation failures. From this assumption, the present research aimed to develop a predictor model, using mathematical tools that could identify and correct technical errors in the recording and measurements performed on physical, chemical and biological parameters of two UASB reactors at Candeia Station of Bauru- SP. For this, after a review of UASB reactor technology and probabilistic studies based on the concept of reliability applied to sewage treatment plants (SEP), a multilinear model coupled to a optimization algorithm, capable of predicting and correcting input and output parameters (inconsistent or absent) in the station monitoring register for the treatment of UASB reactors. With this mathematical procedure, through a conceptual statistical analysis of system reliability, it was possible to correct data and estimate missing parameter information in 22% of the measurement collections and records in the technical reports that were made available for this study.

Keywords: Sanitary Sewer. UASB reactor. Reliability Analysis. Multiple Linear Regression.

Mathematical modeling.

## Lista de abreviaturas e siglas

BF	Biofiltro
CETESB	Companhia Ambiental do Estado de São Paulo
COT	Carbono Orgânico Total
CONAMA	Conselho Nacional do Meio Ambiente
Condut.	Condutividade
DAE	Departamento de Água e Esgoto
DBO	Demanda Bioquímica de Oxigênio
DQO	Demanda Química de Oxigênio
DS	Decantador Secundário
ETE	Estação de Tratamento de Esgoto
LAR	Laboratório de Águas Residuárias
NMP	Número Mais Provável
NTK	Nitrogênio Total Kjeldahl
OD	Oxigênio Dissolvido
PNMA	Política Nacional do Meio Ambiente
RALF	Reator Anaeróbio
SC	Sistema de Cloração
SS	Sólidos em Suspensão
SSed.	Sólidos Sedimentáveis
SST	Sólidos Suspensos Totais
STD	Sólidos Totais Dissolvidos
TDH	Tempo de detenção hidráulica
Temp.	Temperatura
Turb.	Turbidez
UASB	(Upflow Anaerobic Sludge Blanket Reactors)
UWWTD	(European Union Urban Wastewater Treatment Directive)

## LISTA DE ILUSTRAÇÕES

1.	Modelo de reator UASB.....	19
2.	Eficiências de remoção de DQO esperadas para reatores UASB.....	22
3.	Eficiências de remoção de DBO esperadas para reatores UASB.....	23
4.	Gráfico da FDP de uma distribuição Normal.....	30
5.	Medida de graus de achatamento da curva FDP.....	34
6.	Gráfico p-p plot de resíduos ajustado pela normalidade.....	55
7.	Temperatura de um composto em uma reação .....	56
8.	Teste de Kolmogorov – Smirnov para Normalidade.....	57
9.	Vista aérea da ETE Candeia.....	58
10.	Unidades/Componentes do tratamento na ETE Candeia .....	60
11.	Fluxograma das etapas da ETE Candeia .....	61
12.	Etapas do sistema de tratamento da ETE Candeia em planta.....	61
13.	Estação elevatória da ETE Candeia.....	62
14.	Etapa do Gradeamento da estação.....	62
15.	Detalhe da Calha Parshall .....	63
16.	Detalhe da caixa de areia.....	63
17.	Detalhe do Reator UASB de Candeia .....	64
18.	Detalhe do biofiltro.....	64
19.	Detalhe do decantador secundário .....	65
20.	Detalhe do Sistema de cloração.....	65
21.	Destino dos Efluentes .....	66
22.	Gráfico de distribuição de probabilidade para o parâmetro Sólidos Fixos.....	72
23.	Equação de regressão com variáveis dependentes e independentes.....	73
24.	Observação de um <i>outlier</i> no período do crescimento da DQO na etapa de entrada (dia x mgO <sub>2</sub> /L).....	78

## LISTA DE QUADROS

Quadro 1- Tempo de detenção hidráulica para projeto de reatores UASB.....	21
Quadro 2-Relação das ETEs ativas no Brasil que utilizam a tecnologia UASB.....	25
Quadro 3 - Correlação entre as variáveis X e Y.....	40
Quadro 4 – Teste de hipótese para a variável nula ( $H_0$ ).....	46
Quadro 5 - Organização dos dados coletados no laboratório de Águas Residuárias .....	67
Quadro 6 – Organização dos dados pelos Pontos de coleta.....	69
Quadro 7 – Parâmetros desconsiderados inicialmente.....	76
Quadro 8 - Quantidades de dados para cada parâmetro.....	77
Quadro 9- Parâmetros com 145 ou mais dados coletados.....	77
Quadro 10- Parâmetros de acordo com sua distribuição de probabilidade.....	79
Quadro 11 - Parâmetros classificados por G ou R.....	82
Quadro 12 - Modelo de Otimização Linear para o Oxigênio Dissolvido.....	87
Quadro13–Resultados de análises de amostras de esgoto tratado efluente.....	88

## LISTA DE TABELA

Tabela 1 - Determinação da CDF da V.A. normalizada.....	31
Tabela 2 - Teste F- ANOVA.....	48
Tabela 3 - Teste F da Análise de variância.....	49
Tabela 4 – Distribuição F central com seus graus de liberdade.....	52
Tabela 5 - Coeficientes de correlação entre os parâmetros no gradeamento.....	81
Tabela 6 - Concentração média de DBO <sub>5</sub> de operação para padrão ( $\alpha= 95\%$ ).....	85
Tabela 7 – Confiabilidade entre os parâmetros de saída do reator na regressão.....	86
Tabela 8- Diferenças das previsões para a saída do reator.....	87
Tabela 9- Lista de parâmetros corrigidos pelo método.....	89
Tabela 10– Confiabilidade da regressão com os dados modificados.....	90

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>14</b>
<b>2. OBJETIVOS.....</b>	<b>16</b>
<b>3. JUSTIFICATIVA E FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>17</b>
3.1 <i>Processos anaeróbios.....</i>	<i>17</i>
3.2 <i>Tecnologia UASB.....</i>	<i>18</i>
3.3 <i>Relação dos Municípios que utilizam a tecnologia de reatores UASB.....</i>	<i>24</i>
<b>4. REVISÃO DA LITERATURA.....</b>	<b>27</b>
4.1 <i>Critérios para avaliação do desempenho em estações de tratamento de esgoto.....</i>	<i>27</i>
4.1.1 <i>Distribuições Analíticas de Probabilidade de uma Variável Contínua.....</i>	<i>29</i>
4.2 <i>Confiabilidade de um Sistema.....</i>	<i>34</i>
4.2.1 <i>Cálculo do coeficiente de Confiabilidade.....</i>	<i>35</i>
4.3 <i>Técnicas para modelagem de Sistemas.....</i>	<i>37</i>
4.3.1 <i>Simulação de Monte Carlo.....</i>	<i>37</i>
4.3.2 <i>Análise por Árvore de Falhas.....</i>	<i>37</i>
4.3.3 <i>Aplicações da Simulação de Monte Carlo e da Análise por árvore de falhas.....</i>	<i>38</i>
4.4 <i>Correlação amostral.....</i>	<i>39</i>
4.5 <i>A análise de Regressão.....</i>	<i>41</i>
4.5.1 <i>Regressão linear simples.....</i>	<i>43</i>
4.5.1.1 <i>Coeficiente de Determinação ou Correlação.....</i>	<i>44</i>
4.5.1.2 <i>Testes de hipótese na regressão linear simples.....</i>	<i>45</i>
4.5.1.3 <i>Nível descritivo (p- valor).....</i>	<i>47</i>
4.5.2 <i>Regressão linear múltipla.....</i>	<i>50</i>
4.5.2.1 <i>Modelo Linear de 2º Grau.....</i>	<i>51</i>
4.5.2.2 <i>Análise da variância para a regressão linear múltipla.....</i>	<i>51</i>
4.5.2.3 <i>Detecção de outliers e sua Influência na regressão linear múltipla.....</i>	<i>53</i>
4.5.2.4 <i>Diagnóstico de Normalidade dos dados.....</i>	<i>55</i>

<b>5 MATERIAIS E MÉTODOS.....</b>	<b>58</b>
5.1 Área de estudos.....	58
5.2 Unidades do Sistema de Tratamento do Estudo.....	59
5.2.1 Descrição de cada componente das unidades .....	62
5.3 Lançamentos dos efluentes tratados da ETE.....	66
5.4 Metodologias da Pesquisa.....	66
5.5 Ensaios e tratamentos de dados.....	69
5.5.1 Organização dos parâmetros coletados .....	69
5.5.2 Redução de parâmetros .....	70
5.5.3 Eliminação de outliers .....	70
5.5.4 Normalidade ou Log-normalidade .....	71
5.5.5 Teste de correlação.....	72
5.5.6 Cálculo das equações de regressão linear múltipla .....	73
5.5.7 Correção dos parâmetros de saída ausentes .....	73
5.5.8 Correção dos parâmetros de entrada ausentes .....	74
5.5.9 Limitações do método .....	75
<b>6. RESULTADOS E DISCUSSÕES.....</b>	<b>76</b>
6.1 Cálculos do percentual esperado de atendimento às diversas metas de lançamento.....	85
<b>7. CONCLUSÃO.....</b>	<b>92</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>94</b>

## 1. INTRODUÇÃO

O erro humano no trabalho e em diversos setores da Engenharia deve ser uma preocupação constante, visto que, prejuízos materiais, desperdício de dinheiro, recursos e até mesmo o comprometimento de vidas humanas, causados por sua imprecisão devem ser investigados.

A preocupação ora apresentada, também se aplica a esta pesquisa, a qual aborda uma temática importante da Engenharia, que é o Saneamento Ambiental, mais especificamente as ETEs (Estações de Tratamento de Esgoto). Como destacado por Taheriyoun e Moradinejad (2014), em um estudo de confiabilidade de uma ETE, no qual os resultados mostram que os fatores humanos têm o maior impacto no fracasso do sistema de tratamento, tanto na fase de projeto, implementação ou operação.

Porém, em estudos sobre o planejamento de melhorias em ETEs, destaca-se a preferência por ajustes de parâmetros de controle em comparação ao desenvolvimento de atualizações físicas estruturais (por exemplo: aumento das dimensões do tanque; inserção de novos equipamentos), dado o custo e o tempo para implementação que estas requerem.

Um trabalho que merece destaque neste contexto, é o de Benedetti et al. (2010), que estudaram os benefícios e malefícios de 13 possíveis atualizações em uma ETE. Neste sentido, os autores pontuam os custos totais, as difíceis atualizações que envolvem principalmente a intervenção de construção, como sendo claramente mais caras do que as atualizações dos parâmetros de controle.

No contexto brasileiro, Campos (1999) afirma, como benefício da otimização na rotina de controle dos parâmetros, uma redução no tempo de detenção hidráulica (TDH) dos reatores, melhorando a eficiência da ETE a partir do aumento do volume tratado de esgoto, levando a uma menor perda de sólidos no efluente final, o que proporciona uma melhor qualidade do efluente em termos de sólidos suspensos, e DBO (Demanda Bioquímica de Oxigênio) parâmetros estes fiscalizados pela legislação brasileira ambiental, por meio da Resolução do Conselho Nacional do Meio Ambiente, CONAMA 430/2011 e ao Decreto Estadual 8468/1976.

A modelagem estatística está presente nas investigações anteriormente citadas (Taheriyoun, Moradinejad, 2014; Benedetti et al., 2010) como uma ferramenta para tratar o problema de ajuste de parâmetros de ETEs, ligado principalmente ao conceito de confiabilidade estatística. Para Benedetti et al. (2010), confiabilidade estatística diz respeito à porcentagem de tempo para que as concentrações esperadas de efluentes



atendam aos padrões de descarte especificados ou aos alvos de tratamento. Para Taheriyoun e Moradinejad (2014), refere-se à probabilidade de sucesso no bom desempenho da ETE, que deve atender aos padrões de qualidade do efluente.

Como importância deste tema, Messaoud et al (2013) realizaram um estudo sobre a avaliação da performance diária de uma ETE utilizando um modelo probabilístico, no qual ressaltam que a análise baseada na confiabilidade das ETEs permite ao engenheiro explorar a estrutura estatística dos dados do afluente e efluente, a fim de prever a probabilidade de eventos indesejáveis.

Dessa forma, a partir do que foi contextualizado, a hipótese deste estudo é que os principais parâmetros físico – químico e biológico de controle da ETE registrados pelo seu corpo técnico em análise laboratorial, possam ser acompanhados e assistidos indiretamente a partir de um modelo matemático de natureza estatística que permita a identificação e a correção de erros oriundos de vários fatores, dentre eles, o fator humano. Assim, busca-se nesta pesquisa, construir um modelo matemático que permita o monitoramento e controle dos principais parâmetros físico-químicos e biológicos de entrada e saída de reatores UASB presentes em uma ETE.

Um benefício direto deste estudo é seu potencial na produção instrumental de baixo custo e sua rápida implementação para o ajuste destes parâmetros que visam à correção de erros normativos (que se distam do contexto real) de coleta laboratorial em uma estrutura de ETE que seja bastante usada em território brasileiro, proporcionando um menor custo e um uso amplo da ferramenta em território nacional.

## **2. OBJETIVOS**

Tomando-se como hipótese que os principais parâmetros físicos, químicos e biológicos de uma ETE possam ser relacionados, foi definido o seguinte objetivo geral: Desenvolver uma estrutura de modelagem empírica aplicável ao ajuste e correção de parâmetros operacionais registrados na entrada e saída de reatores UASB. Com base neste objetivo geral, são propostos os seguintes objetivos específicos:

1. Avaliar a utilização da regressão linear múltipla como ferramenta para a modelagem destes parâmetros operacionais, para tratamento de esgoto sanitário por meio de sistema com reator UASB.
2. Definir os critérios para estabelecimento dos parâmetros adequados a esta modelagem.
3. Realizar análise de confiabilidade dos parâmetros obtidos a partir do modelo empírico elaborado.

### 3. JUSTIFICATIVA E FUNDAMENTAÇÃO TEÓRICA

Antes de tratar da modelagem matemática propriamente dita dos parâmetros de uma ETE, cabe abordar os processos anaeróbios, especificamente o sistema em estudo neste trabalho, que é o *Upflow Anaerobic Sludge Blanket* (UASB – traduzido como Sistema de Reator Anaeróbio de Manta de Lodo).

#### 3.1 Processos anaeróbios

Segundo Chernicharo (2016), diversos dos compostos orgânicos podem ser degradados pela via anaeróbia, sendo que o processo se mostra mais eficiente e mais econômico quando os dejetos são facilmente biodegradáveis. Também, os digestores anaeróbios têm sido largamente aplicados para o tratamento de resíduos sólidos, incluindo culturas agrícolas, dejetos de animais, lodos de ETE e lixo urbano, estimando-se que milhões de digestores anaeróbios tenham sido construídos em todo o mundo com esse propósito.

A digestão anaeróbia tem sido muito aplicada para o tratamento de efluentes de indústrias agrícolas, alimentícias e de bebidas, tanto em países desenvolvidos como em desenvolvimento. Também em relação ao tratamento de esgotos domésticos, tem-se verificado um enorme incremento na utilização da tecnologia anaeróbia, notadamente por meio dos reatores tipo UASB. Naturalmente que, nesse caso, a aplicabilidade da tecnologia anaeróbia depende de forma muito mais significativa da temperatura dos esgotos, devido à baixa atividade dos microrganismos anaeróbios em temperaturas abaixo de 20° C e à inviabilidade de aquecimento dos reatores. Isso porque os esgotos domésticos são bem mais diluídos que os efluentes industriais, resultando em baixas taxas de produção volumétrica de gás metano, o que torna antieconômica a sua utilização como fonte de energia para aquecimento (Campos, 1999).

As diversas características favoráveis da tecnologia anaeróbia, como baixa produção de sólidos; baixo consumo de energia; baixos custos de implantação e operação; tolerância a elevadas cargas orgânicas; possibilitam a operação com elevados tempos de retenção de sólidos e baixos tempos de detenção hidráulica no tratamento de águas residuais concentradas e diluídas. O uso dessa tecnologia para o tratamento de esgotos domésticos torna-se ainda mais atrativo no Brasil devido às condições climáticas, e, de fato, o país assume posição de vanguarda no cenário mundial, em relação à aceitação e à

disseminação de reatores anaeróbios, em especial dos reatores de fluxo ascendente e manta de lodo (UASB) (Lettinga, 1992).

As principais desvantagens da tecnologia anaeróbia são relacionadas à remoção insatisfatória de nutrientes (nitrogênio e fósforo) e patógenos; maior instabilidade dos reatores anaeróbios, devido a choques de carga orgânica e hidráulica, presença de compostos tóxicos ou ausência de nutrientes (Campos, 1999).

A aceitação e a disseminação da tecnologia anaeróbia para o tratamento de esgotos domésticos, notadamente dos reatores tipo UASB, colocam o Brasil em uma posição de vanguarda em nível mundial. Estima-se que exista atualmente mais de 400 reatores anaeróbios tipos UASB tratando esgotos domésticos no Brasil, a maioria deles localizada nos estados de São Paulo, Paraná, Bahia, Minas Gerais e no Distrito Federal. Existem ainda diversos reatores anaeróbios em operação em outros estados, particularmente no Espírito Santo, Mato Grosso do Sul, Pará, Pernambuco, Rio de Janeiro, Rio Grande do Norte e Rio Grande do Sul (Chernicharo, 2016).

### **3.2. Tecnologia UASB**

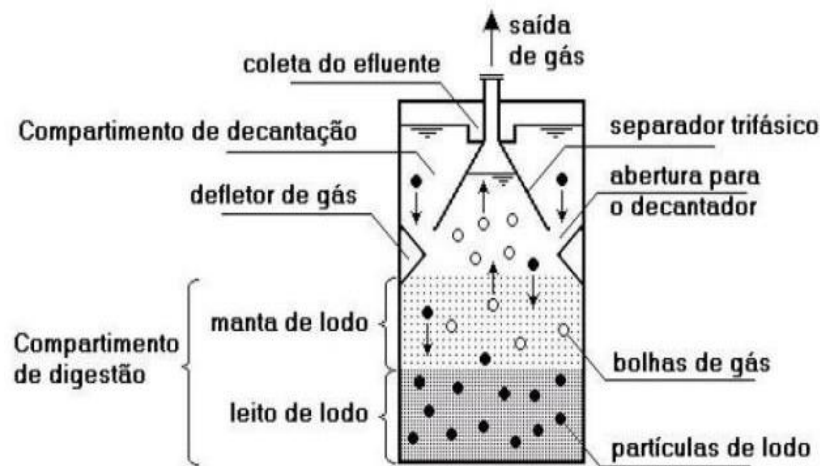
O Reator de fluxo ascendente anaeróbico (RAFA ou a sigla em inglês UASB) é um reator bioquímico, cujo tratamento biológico ocorre por um processo anaeróbio, isto é, na ausência de oxigênio. A decomposição da matéria orgânica é feita por microrganismos presentes em uma manta de lodo. A água residuária sai do fundo do reator e passa pela camada de lodo que age como uma biomassa ativa na conversão de matéria orgânica. O objetivo de um reator UASB é de promover uma remoção média de matéria orgânica DBO<sub>5</sub> da ordem de 70 %, em sendo assim, um tratamento complementar já é previsto como necessário (Von Sperling, 1995).

De acordo com Campos (1985), o reator UASB ou também conhecido como reator de manta de lodo, é uma unidade de fluxo ascendente que permite transportar as águas residuárias por uma região de elevada concentração de microrganismos anaeróbios. A principal diferença entre este tipo de reator com os demais, é que o UASB não possui qualquer material de enchimento para servir de suporte para a biomassa. “A imobilização dos microrganismos ocorre por meio de auto adesão, formando flocos ou grânulos densos suspensos, que se dispõe em camadas de lodo a partir do fundo do reator” (Campos, 1999).

Lettinga G. (1992) ainda reforça que os sistemas de tratamento de águas residuárias de leito anaeróbico de fluxo ascendente (UASB) representam uma tecnologia sustentável comprovada para uma gama de efluentes industriais muito diferentes, incluindo aqueles que contêm compostos tóxicos / inibidores. Cita ainda que o processo também é viável para o tratamento de águas residuais domésticas com temperaturas de 14 a 16° C e, provavelmente, ainda mais baixas e que em comparação com os sistemas convencionais de tratamento aeróbico, este processo anaeróbico oferece apenas vantagens.

Na Figura 1, está representado o esquema geral de um reator UASB. Seu processo consiste em um fluxo ascendente de esgotos por meio de um leito de lodo denso e de elevada atividade. O perfil de sólidos no reator varia de muito denso e com partículas granulares de alta capacidade de sedimentação próximos ao fundo (leito de lama) a um lodo mais disperso e leve próximo ao topo do reator (manta de lodo). A estabilização da matéria orgânica ocorre em todas as zonas de reação (leito e manta de lodo), com a mistura do sistema promovida pelo crescente fluxo de drenagem e bolhas de gás (Chernicharo, 2007). A Figura 01 mostra um reator UASB.

**Figura 1- Modelo de reator UASB**



Fonte: Chernicharo, 2007

A água residuária entra no fundo e o efluente sai do reator por meio de um decantador interno localizado na sua parte superior. Um dispositivo de separação trifásico de gases, sólidos e líquidos, localizado abaixo do decantador, garante condições ótimas para a sedimentação das partículas que circundam a camada de lodo, permitindo que sejam novamente digeridas em vez de serem removidas do sistema.

Embora algumas das partículas mais leves sejam perdidas com o efluente, o tempo médio de permanência dos sólidos no reator permanece alto, o suficiente, para manter o crescimento de uma massa densa de microrganismos formadores de metano, apesar do reduzido tempo de retenção hidráulica (Chernicharo, 1997).

De acordo com Campos (1999), entre os principais parâmetros que devem ser verificados durante a fase operacional de um reator UASB, tem-se:

1. Vazões afluentes ao sistema;
2. Características físico-químicas do lodo do esgoto afluente;
3. Eficiência e problemas operacionais das unidades de tratamento preliminar;
4. Produção e características do material retido nas grades e na caixa de areia;
5. Eficiência e problemas operacionais do reator anaeróbio;
6. Quantidade e características do biogás produzido no reator anaeróbio;
7. Quantidade e características do lodo produzido no reator anaeróbio.

Além destes, entende-se como necessária a verificação das características físico-químicas do lodo do esgoto efluente.

Observando que neste trabalho tem-se como um dos objetivos específicos a modelagem de parâmetros operacionais registrados na entrada e saída de reatores UASB em uma ETE que trata de esgoto sanitário, para a determinação da eficiência destes reatores, alguns parâmetros de projeto são importantes, entre os quais, temos: a carga hidráulica volumétrica ( CHV) dada pela Equação (1) e o tempo de detenção hidráulica (t) dado pela Equação (2) que são definidos a seguir :

- a) A carga hidráulica volumétrica (CHV)

É a quantidade ( volume) de esgotos aplicados diretamente no reator, por unidade de volume do mesmo.

$$CHV = \frac{Q}{V} \quad (1)$$

Onde:

CHV: carga hidráulica volumétrica ( m<sup>3</sup>/m<sup>3</sup>.d);

Q: Vazão ( m<sup>3</sup>/d);

V: Volume total do reator ( m<sup>3</sup>).

b) Tempo de detenção hidráulica ( t )

É o inverso da carga hidráulica volumétrica .

$$t = \frac{1}{CHV} \quad (2)$$

Onde :

t: tempo de detenção hidráulica (d)

Segundo (Van Haandel e Catunda, 1998), para esgotos domésticos sendo tratados na faixa de temperatura em torno de 20° C, tem-se adotado, tempo de detenção hidráulica pode variar de 6 a 16 horas, dependendo o tipo de despejo. Estudos em escala piloto com reatores a uma temperatura média de 25°C, alimentados com esgoto doméstico com alcalinidade relativamente elevada, mostraram que na ordem de 4 horas não afetam o desempenho desses reatores, nem sua estabilidade operacional.

No Quadro 1 são apresentadas algumas diretrizes para o estabelecimento dos tempos de detenção hidráulica em projetos de reatores de manta de lodo, tratando esgotos domésticos .

**Quadro 1 – Tempo de detenção hidráulica para projeto de reatores UASB**

Temperatura do esgoto (°C)	Tempo de detecção hidráulica (h)	
	Para Q média	Para Q máxima
15 a 18	≥ 10,0	≥ 7,0
18 a 22	≥ 8,0	≥ 5,5
22 a 25	≥ 7,0	≥ 4,5
> 25	≥ 6,0	≥ 4,0

Fonte : ABNT- NBR 12209

### **3.2.1 Cálculo da eficiência dos reatores UASB**

A modelagem matemática aplicada ao projeto e à operação de sistemas anaeróbios, particularmente para sistemas tratando substratos complexos como esgoto doméstico, vem sendo feita por meio de relações empíricas a partir de resultados experimentais em reatores em operação. Para determinação da eficiência dos reatores

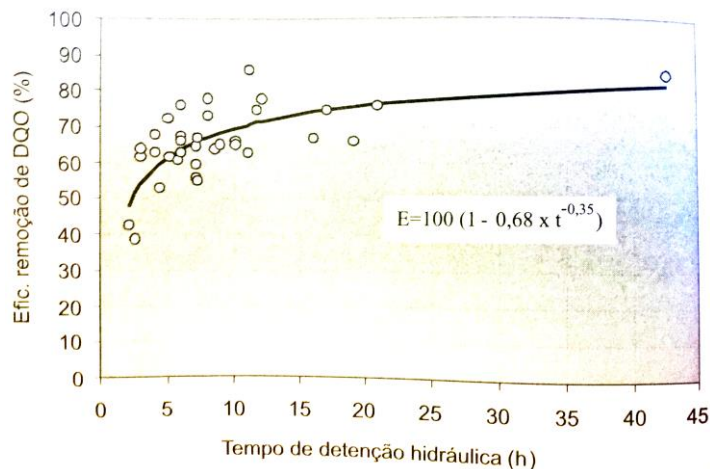
UASB, serão tomados como base o modelo de digestão anaeróbia N° 1, desenvolvido pelo grupo de trabalho da International Water Association (IWA), para modelagem matemática de processos de digestão anaeróbia (Batstone et al., 2002).

O resultado deste trabalho refere-se à operação de dezesseis reatores UASB em plena escala operando nas seguintes condições:

1. Temperatura de esgoto: entre 20 e 27° C.
2. Concentração de DQO do esgoto bruto: entre 300 e 1.400 mg/L.
3. Concentração de DBO do esgoto bruto: entre 150 e 850 mg/L.

Observando-se que as eficiências de remoção de DQO e DBO são afetadas, de forma significativa, pelo tempo de detenção hidráulica do sistema, tendo variado de 40% a 80% para a remoção de DQO, e de 40 a 90% para a remoção de DBO, a partir do ajuste dos resultados operacionais destes dezesseis reatores, foram obtidas curvas de eficiência para os parâmetros DQO e DBO do esgoto bruto, representadas pelas Figuras 2 e 3 respectivamente.

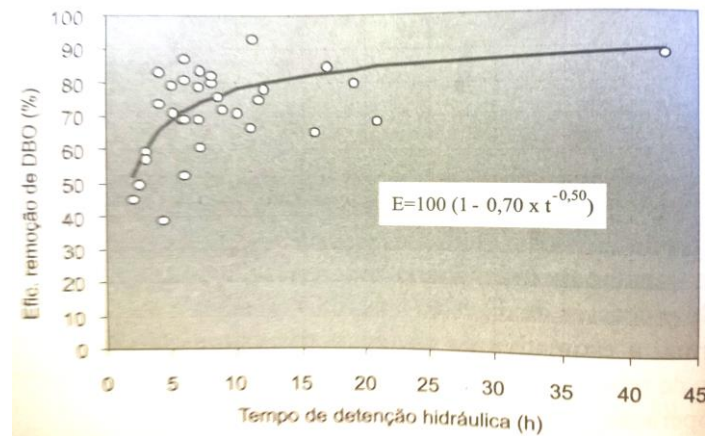
**Figura 2 : Eficiências de remoção de DQO esperadas para reatores UASB**



Fonte: Chernicharo, 1997, p. 236.



**Figura 3: Eficiências de remoção de DBO esperadas para reatores UASB**



Fonte: Chernicharo, 1997, p. 236..

As equações que descrevem a eficiência do resultado operacional destes dezesseis reatores UASB, em função do tempo de detenção hidráulica para os parâmetros DQO e DBO, são dadas respectivamente pelas equações 3 e 4 . Deve-se ressaltar, no entanto, a limitação das mesmas, devido ao número bastante limitado de dados utilizados para a determinação das constantes empíricas, que entre si exibiam grandes desvios.

$$E_{DQO} = 100 (1 - 0,68 \times t^{-0,35}) \quad (3)$$

Na qual:

$E_{DQO}$ : eficiência do reator UASB, em termos de remoção de DQO ( %)

t: tempo de detenção hidráulica (h);

0,68 : constante empírica;

0,35 : constante empírica

$$E_{DBO} = 100 (1 - 0,70 \times t^{-0,50}) \quad (4)$$

Na qual:

$E_{DQO}$ : eficiência do reator UASB, em termos de remoção de DBO ( %)

t: tempo de detenção hidráulica (h);

0,70 : constante empírica;

0,50 : constante empírica

À partir da eficiência esperada pelo sistema, pode-se estimar a concentração de DQO e de DBO no efluente final, conforme a Equação 5:

$$S = S_0 - \frac{E \times S_0}{100} \quad (5)$$

Na qual:

$S$ : concentração de DQO ou de DBO efluente ( mg/L);

$S_0$  : concentração de DQO ou de DBO afluente ( mg/L);

$E$  : eficiência de remoção de DQO ou de DBO ( %).

### ***3.3 Relação dos Municípios que utilizam a tecnologia de reatores UASB***

Devido à natureza prática das ETEs, em buscas nos repositórios da CAPES em fevereiro de 2018, não se localizou uma lista sistemática que apresentasse quais ETEs brasileiras utilizam reatores UASB. Embora existam muitas pesquisas relacionadas a uma ou outra ETE, não foram encontradas pesquisas de levantamento. Dessa forma, em março de 2018 realizou-se um levantamento sistemático por meio do site de buscas Google, para em língua portuguesa, identificar anúncios, notícias, sites próprios com as palavras-chave “ETE” AND “UASB”.

Assim, por uma questão de tempo e qualidade dos resultados, considerou-se verificar os sites dos 100 primeiros resultados da busca. Destes, procurou-se referências às ETEs que utilizam UASB, assim como seu nome, ou o nome do bairro em que se encontra (visto que muitas delas são conhecidas e tratadas apenas pelo nome do bairro), a empresa que gerencia e uma forma de contato oficial.

Foi possível nesta busca reunir 50 ETEs ativas até o momento, que utilizam reatores UASB, duas ETEs projetadas para utilizar reatores UASB, porém não construídas, e duas em processo de construção. Por conta de outras pesquisas que possam se beneficiar desta informação, apresenta-se o Quadro 2.

**Quadro 2 - Relação das ETEs no Brasil que utilizam da tecnologia UASB**

<b>Cidade</b>	<b>Estado</b>	<b>Situação</b>	<b>Nome da ETE/Bairro</b>
Almirante Tamandaré	Paraná	Ativa	São Jorge
Angra dos Reis	Rio de Janeiro	Desenvolvimento	Praia da Chácara
Araguari	Minas Gerais	Ativa	Novo Horizonte
Araras	São Paulo	Desenvolvimento	Horto Florestal
Arroio Grande	Rio Grande do Sul	Projeto	Ecoville
Bauru	São Paulo	Ativa	Candeia
Belém	Pará	Ativa	Vila da Barca
Belo Horizonte	Minas Gerais	Ativa	Onça
Betim	Minas Gerais	Ativa	Central
Bom Jesus do Amparo	Minas Gerais	Ativa	Arrudas
Brasília	Distrito Federal	Ativa	Paranoá
Campinas	São Paulo	Ativa	Piçarrão
Cascavel	Paraná	Ativa	Rio das Antas/Norte
Aracruz	Espírito Santo	Ativa	Guaraná
Cuiabá	Mato Grosso	Ativa	Hospital do Câncer
Curitiba	Paraná	Projeto	Ecoville
Divinópolis	Minas Gerais	Ativa	Itapecerica
Florianópolis	Santa Catarina	Ativa	Lagoa da Conceição
Florianópolis	Santa Catarina	Ativa	Canasvieiras
Florianópolis	Santa Catarina	Ativa	Barra da Lagoa
Fortaleza	Ceará	Ativa	HGWA
Foz do Iguaçu	Paraná	Ativa	Ouro Verde
Irupi	Espírito Santo	Ativa	Centro
Itabira	Minas Gerais	Ativa	Laboreaux
Itapema	Santa Catarina	Ativa	Águas de Itapema
Juazeiro do Norte	Ceará	Ativa	Malvas
Juiz de Fora	Minas Gerais	Ativa	Barreira do Triunfo
Juiz de Fora	Minas Gerais	Ativa	União Indústria

Continua

Continuação

Juiz de Fora	Minas Gerais	Ativa	Barbosa Lage
Juiz de Fora	Minas Gerais	Ativa	Santa Luzia
Lagoa de Prata	Minas Gerais	Ativa	Lagoa de Prata
Lauro de Freitas	Bahia	Ativa	VogVille
Lençóis	Bahia	Ativa	Lençóis
Londrina	Paraná	Ativa	Norte
Luzerna	Santa Catarina	Ativa	Luzerna
Natal	Rio Grande do Norte	Ativa	Baldo
Novo Hamburgo	Rio Grande do Sul	Ativa	Roselândia
Piracicaba	São Paulo	Ativa	Piracicamirim
Pomerode	Santa Catarina	Ativa	Corupá
Porto Alegre	Rio Grande do Sul	Ativa	Sarandi
Quirinópolis	Goiás	Ativa	Quirinópolis
Recife	Pernambuco	Ativa	Mangueira
Ribeirão Pires	São Paulo	Ativa	ETE ABC
Rio Claro	São Paulo	Ativa	Jardim das Flores
Rio de Janeiro	Rio de Janeiro	Ativa	Ilha do Fundão
Santo Cristo	Rio Grande do Sul	Ativa	Monjolo
São Carlos	São Paulo	Ativa	Monjolinho
São José do Rio Preto	São Paulo	Ativa	São José do Rio Preto
Saudades	Santa Catarina	Ativa	Saudades
Taquaritinga	São Paulo	Ativa	Taquaritinga
Uberlândia	Minas Gerais	Ativa	Uberabinha
Valinhos	São Paulo	Ativa	Capuava
Vila Velha	Espírito Santo	Ativa	Vale Encantado

Fonte: Elaborado pelo autor

## 4. REVISÃO DA LITERATURA

### *4.1 Critérios para avaliação do desempenho em estações de tratamento de esgoto*

Diversos trabalhos têm sido desenvolvidos sobre o projeto e a operação de estações de tratamento de esgoto, envolvendo operações físicas unitárias e processos químicos e biológicos unitários, uma vez que seu bom desempenho (como comentado) envolve os padrões de lançamento preconizados pelos órgãos e legislações ambientais.

Metcalf e Eddy (2003) citam que novas considerações são importantes para se operar e projetar estações de tratamento de esgotos, entre eles:

- a) necessidade de otimizar o desempenho das estações;
- b) confiabilidade nos processos de tratamento e seleção adequada dos parâmetros de projeto;
- c) controle de odor;
- d) estratégias de controle de processo;
- e) expansão da capacidade de tratamento e
- f) eficiência energéticas nos processos de tratamentos de esgoto.

As análises estatísticas de desempenho de processos de tratamento e desenvolvimento de métodos e procedimentos para a introdução de conceitos de confiabilidade e estabilidade no projeto têm sido por décadas efetuadas em diversos países, sendo que alguns deles versam sobre o comportamento das variáveis dos efluentes gerados em ETEs.

Dessas análises estatísticas, Dean e Forsythe (1976a), citam que a função de distribuição da concentração do efluente é importantíssima para se determinar a probabilidade de falha. Conhecida esta distribuição, uma expressão pode ser utilizada para definir a fração do tempo em que uma dada concentração foi excedida no passado e, desta forma, predizer o comportamento futuro de uma ETE, desde que as variáveis do processo continuem as mesmas.

Neste contexto, Niku e Schroeder (1981) reforçam ainda que as análises estatísticas por meio da teoria de probabilidades são essenciais para se conhecer a distribuição de frequência dos dados dos efluentes, onde umas séries de estudos efetuados partiram da determinação de Dean e Forsythe (1976a, 1976b); Niku et al., (1979, 1981a, 1981b, 1981c e 1982), Berthouex e Hunter (1981, 1983), Charles et al. (2005).

Ainda nesses estudos citados por Niku e Schroeder (1981), os diversos autores comentam da dificuldade de se utilizar distribuições tradicionalmente adotadas como a normal, uma vez que a análise realizada por esse modelo pressupõe que os dados sejam aleatórios, independentes e simetricamente distribuídos e na realidade, geralmente violam essas condições por apresentarem:

- a) presença de valores atípicos ;
- b) vários erros de medições;
- c) variância não constante;
- d) período com ausência de dados;
- e) distribuição assimétrica;
- f) correlação serial ( dados não independentes);
- g) relações complexas de causa e efeito;
- h) variáveis não medidas;
- i) mudança nos métodos de medição, causando eventual heterogeneidade nos dados amostrais;
- j) alteração nos pontos de monitoramento;
- k) alteração nos procedimentos de amostragem.

Berthouex e Hunter (1983) e Berthouex e Fan (1986) ressaltam, porém, que diversas análises de variáveis efluentes de estações de tratamento de esgoto, têm mostrado que, para a maioria dos processos, grande parte dos constituintes pode ser modelada pela distribuição log-normal, já que os dados geralmente apresentam assimetria positiva.

Metcalf e Eddy (2003) comentam que, em geral, o desempenho de processos de tratamento biológicos convencionais, considerando variáveis efluentes de esgotos que podem ser modificadas pelo tratamento, como DBO, carbono orgânico total, etc. são descritos mais frequentemente pela distribuição log-normal.

Messaoud et al. (2013) apresentaram um estudo sobre a avaliação do desempenho diário de uma ETE utilizando um modelo probabilístico. Neste trabalho, ressalta-se que a análise baseada na confiabilidade das ETE permite ao engenheiro explorar a estrutura estatística dos dados do afluente e efluente, a fim de prever a probabilidade de eventos indesejáveis. Este estudo também apresenta o contexto matemático proposto por Niku et al. (1979) para a avaliação da confiabilidade do tratamento da cidade de Khenchela na Argélia Oriental e para uma primeira avaliação dos

componentes críticos do processo de tratamento de águas residuais. Para este método, o primeiro passo foi determinar as leis reais de distribuição de probabilidade dos parâmetros de qualidade do efluente tratado selecionados para a cidade de Khenchela.

A qualidade dos efluentes tratados descarregados pela ETE foi estudada usando três parâmetros amplamente utilizados para descargas de ETE: Demanda Bioquímica de Oxigênio de cinco dias (DBO<sub>5</sub>), Demanda Química de Oxigênio (DQO) e Sólidos Suspensos Totais (SST). Os resultados deste estudo mostraram que a distribuição log-normal é a mais representativa do comportamento dos parâmetros considerados do efluente (DBO<sub>5</sub>, DQO e SST). Os resultados obtidos foram consistentes com a observação feita por Dean e Forsythe (1976), Niku e Schroeder (1981) e Oliveira e Sperling (2008). As variáveis de qualidades dos efluentes (DBO<sub>5</sub>, DQO e SST) não são distribuídas simetricamente e geralmente sua distribuição é inclinada para a direita. Com isso, a lei de distribuição de probabilidade para as concentrações de DBO<sub>5</sub>, DQO e SST de um processo de lodo ativado foi examinada, e mostrou que a distribuição log-normal fornece consistentemente um bom ajuste geral para observar os parâmetros do efluente. Neste contexto, segue as principais distribuições de probabilidade normalizadas.

#### 4.1.1 Distribuições Analíticas de Probabilidade de uma Variável Contínua

Segundo Thomopoulos e Jonhson (2003) as distribuições de uma variável contínua podem ser:

1) Distribuição Normal ou Gaussiana

- Distribuição mais conhecida e usada na prática;
- Sua função Densidade probabilidade (FDP) é dada pela Equação 6:

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ \frac{-1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right] \text{ para } -\infty < x < +\infty \quad (6)$$

Onde:  $\mu$  e  $\sigma$ , são respectivamente, a média e o desvio padrão da variável aleatória em estudo.

- Sua nomenclatura para uma variável aleatória Normal é:  $X \sim N(\mu; \sigma)$
- A distribuição normal é simétrica em torno da média (portanto, média e mediana são iguais);

- O ponto máximo da Função Densidade Probabilidade (FDP) ou MODA ocorre na média;
- O ponto de inflexão da FDP ocorre em:  $x = \mu \pm \sigma$
- Para  $n$  observações independentes da variável aleatória (V.A.)  $X$ , a média e a variância podem ser estimadas pelas Equações 7 e 8 respectivamente:

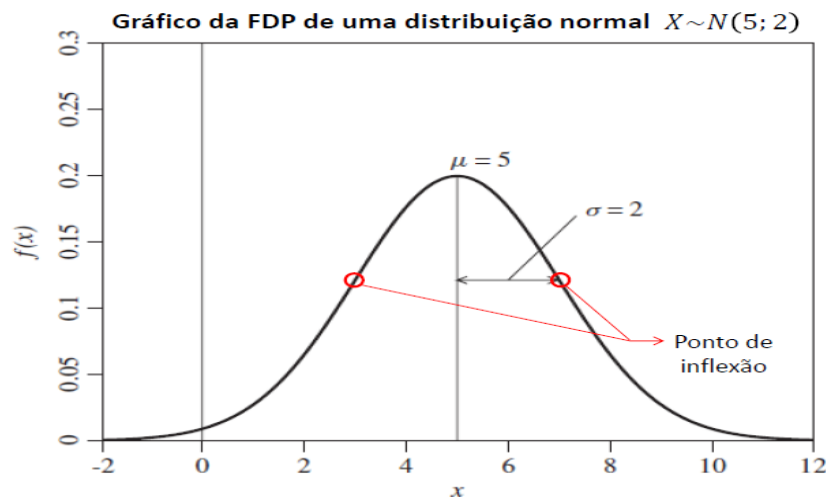
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (7)$$

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (8)$$

Demais medidas estatísticas também podem ser determinadas como: média geométrica, mediana, moda, quartis, covariância, coeficientes de assimetria, etc.

A Figura 4 ilustra a Função Densidade Probabilidade (FDP) de uma distribuição normal.

**Figura 4 - Gráfico da FDP de uma distribuição Normal**



Fonte: Elaborado pelo autor

Segundo Bolfarine e Sandoval (2010), sua desvantagem, é que não existe forma analítica para a função distribuição acumulada de probabilidades (CDF integral de  $f_x$ ) sua probabilidade deve ser calculada por integração numérica ou pela:

- Função distribuição acumulada de probabilidades (CDF) da V.A. normalizada, obtida numericamente e condensada conforme mostrada na Tabela 1 a seguir:



**Tabela 1 –Determinação da CDF da V.A. normalizada Z:**

A.1 Distribuição Normal  $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50398	.50797	.51196	.51595	.51993	.52392	.52790	.53188	.53585
0.1	.53982	.54379	.54775	.55171	.55567	.55961	.56355	.56749	.57142	.57534
0.2	.57925	.58316	.58706	.59095	.59483	.59870	.60256	.60641	.61026	.61409
0.3	.61791	.62171	.62551	.62930	.63307	.63683	.64057	.64430	.64802	.65173
0.4	.65542	.65909	.66275	.66640	.67003	.67364	.67724	.68082	.68438	.68793
0.5	.69146	.69497	.69846	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72574	.72906	.73237	.73565	.73891	.74215	.74537	.74857	.75174	.75490
0.7	.75803	.76114	.76423	.76730	.77035	.77337	.77637	.77935	.78230	.78523
0.8	.78814	.79102	.79389	.79673	.79954	.80233	.80510	.80784	.81057	.81326
0.9	.81593	.81858	.82121	.82381	.82639	.82894	.83147	.83397	.83645	.83891
1.0	.84134	.84375	.84613	.84849	.85083	.85314	.85542	.85769	.85992	.86214
1.1	.86433	.86650	.86864	.87076	.87285	.87492	.87697	.87899	.88099	.88297
1.2	.88493	.88686	.88876	.89065	.89251	.89435	.89616	.89795	.89972	.90147
1.3	.90319	.90490	.90658	.90824	.90987	.91149	.91308	.91465	.91620	.91773
1.4	.91924	.92073	.92219	.92364	.92506	.92647	.92785	.92921	.93056	.93188
1.5	.93319	.93447	.93574	.93699	.93821	.93942	.94062	.94179	.94294	.94408
1.6	.94520	.94630	.94738	.94844	.94949	.95052	.95154	.95254	.95352	.95448
1.7	.95543	.95636	.95728	.95818	.95907	.95994	.96079	.96163	.96246	.96327
1.8	.96406	.96485	.96562	.96637	.96711	.96784	.96855	.96925	.96994	.97062
1.9	.97128	.97193	.97257	.97319	.97381	.97441	.97500	.97558	.97614	.97670
2.0	.97724	.97778	.97830	.97882	.97932	.97981	.98030	.98077	.98123	.98169

Fonte : MORETTIN (2011)

No entanto, a distribuição normal permite a ocorrência de valores negativos para as variáveis aleatórias, no sentido de que existe uma probabilidade não nula de ocorrer tais valores.

Observa-se também que esse tipo de distribuição não é adequado para simular variáveis aleatórias do tipo carga, propriedade dos materiais e concentrações que em geral não possuem valores negativos.

Uma maneira simples de se evitar esse inconveniente é utilizar a distribuição log-normal nessa representação, uma vez que desse modo, os valores negativos não podem aparecer.

## 2) Distribuição Logarítmica Normal ou Log-Normal

Thomopoulos e Jonhson (2003) citam também que a distribuição log-normal é obtida através de uma transformação não linear sobre a distribuição normal. Seja uma variável aleatória G normalmente distribuída com média e desvio padrão, onde  $-\infty < g < +\infty$ , então:  $X = e^G$  será sempre positivo com variação de  $0 \leq x < +\infty$ . Assumimos, portanto que a variável aleatória X é do tipo Log-normal, pois seu logaritmo natural  $\ln(x)$  é normalmente distribuído.

- Sua função Densidade probabilidade (FDP) é dada pela Equação 9:

$$f_x(x) = \frac{1}{x\sigma_{\ln x}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu_{\ln x}}{\sigma_{\ln x}}\right)^2\right] \text{ para } -\infty < x < +\infty \quad (9)$$

Os dois parâmetros de distribuição são dados pelas Equações (10) e (11) respectivamente:

$$\mu_{\ln X} = E[\ln X] \quad (10)$$

$$\sigma_{\ln x}^2 = \text{Var}[\ln x] \quad (11)$$

Para o cálculo de probabilidades, temos a Inequação 12:

$$\begin{aligned} P[X < a] &= P[\ln(X) - \ln(a)] \\ P\left[\frac{\ln X - \mu_{\ln x}}{\sigma_{\ln x}} < \frac{\ln a - \mu_{\ln x}}{\sigma_{\ln x}}\right] \\ P\left[Z < \frac{\ln a - \mu_{\ln x}}{\sigma_{\ln x}}\right] \end{aligned} \quad (12)$$

Onde:  $\left(\frac{\ln a - \mu_{\ln x}}{\sigma_{\ln x}}\right)$  é encontrado na tabela de Z normalizado sendo chamado de  $(Z_{1-\alpha})$ .

A média e variância de x podem ser obtidas a partir da transformação dos dois parâmetros de distribuição log-normal através das Equações 13 e 14 respectivamente:

$$\mu_X = E[X] = \bar{X} = e^{\mu_{\ln X} + \frac{1}{2}\sigma_{\ln X}^2} \quad (13)$$

$$\sigma_X^2 = \text{Var}[X] = \mu_X^2 \left(e^{\sigma_{\ln X}^2} - 1\right) \quad (14)$$

Ou, caso conheçamos a média e a variância de X (o que ocorre usualmente), podemos obter os parâmetros internos da distribuição log-normal conforme as Equações 15 e 16:

$$\mu_{\ln X} = \ln \mu_X - \frac{1}{2}\sigma_{\ln X}^2 \quad (15) \quad e \quad \sigma_{\ln X}^2 = \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \quad (16)$$

Demais momentos de X respectivamente são dados pelas Equações 17 e 18:

$$\text{Moda: } Mo_x = e^{\mu_{\ln X} - \sigma_{\ln X}^2} \quad (17)$$

$$\text{Mediana: } Mde^x = e^{\mu_{\ln X}} \quad (18)$$

É importante salientar, que a expressão genérica  $E[X^n]$  de esperança, o enésimo momento ou momento de ordem n, são muito importantes para caracterizar as distribuições de probabilidade. A distribuição normal é caracterizada apenas pelo primeiro e segundo momentos. O primeiro, segundo, terceiro e quartos momentos

caracterizam a tendência central, dispersão, assimetria e curtose respectivamente, que são muito utilizados para caracterizar as funções densidade de probabilidade (FDP).

A seguir, através da Equação 19, segue a equação que rege o cálculo do momento de uma ordem genérica K.

$$\text{Momento de ordem } k: E[X^k] = e^{k\mu + \frac{1}{2}k^2\sigma^2} \quad (19)$$

Na qual se obtém os principais momentos através das Equações 20 e 21.

$$E[k^3] = \text{skewness} = e^{3\mu + \frac{9}{2}\sigma^2} \quad (20)$$

$$E[k^4] = \text{Curtose} = e^{4\mu + 8\sigma^2} \quad (21)$$

Obs: A obliquidade ou Skewness mede a assimetria das caudas na distribuição. As distribuições assimétricas, são aquelas que têm um dos lados da distribuição (abaixo ou acima da média) com mais elementos que o outro, enquanto as distribuições simétricas têm obliquidade igual a zero.

O critério para se conhecer o tipo de assimetria é:

a)  $Mo < Md < \bar{X}$  assimetria positiva (possui o lado esquerdo com valores abaixo da média com mais elementos) ;

b)  $\bar{X} < Md < Mo$  , assimetria negativa (possui o lado direito com valores abaixo da média com mais elementos);

Já a Curtose é uma medida adimensional que caracteriza o achatamento da curva da função (FDP). Ela pode ser:

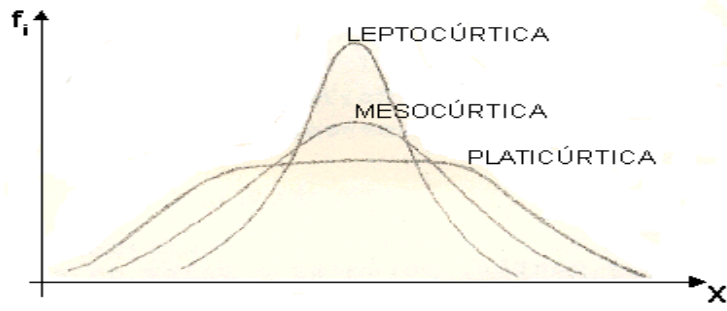
a)  $E[k^4] = 3$  , então o Grau de "achatamento" ou Grau de Curtose é igual ao da Curva Normal (simétrica) de mesma área e denomina-se a este tipo de Curva MESOCÚRTICA.

b)  $E[k^4] < 3$ , a distribuição é mais achatada do que a Curva Normal de mesma área e diz-se que é uma Curva PLATICÚRTICA.

c)  $E[k^4] > 3$  a distribuição é menos achatada (mais afilada) do que a Curva Normal de mesma área, denominando-se de Curva LEPTOCÚRTICA.

A Figura 5 esclarece cada um desses tipos:

**Figura 5 – Medida de graus de achatamento da curva FDP**



Fonte :Elaborado pelo autor

Cabem salientar que há diversas outras distribuições analíticas de probabilidade tais como: distribuição exponencial, uniforme, gamma, etc. que não fizeram parte do escopo desse trabalho.

#### **4.2 Confiabilidade de um Sistema.**

Define-se confiabilidade de um sistema, como a probabilidade de se alcançar um desempenho adequado por um determinado período de tempo, sob determinadas condições (OLIVEIRA, 2008).

Em se tratando de desempenho de uma estação de tratamento de esgoto, a confiabilidade pode ser entendida como a porcentagem de tempo na qual as concentrações que são almejadas para o efluente, em termos de cumprimento dos padrões de lançamento exigido pelos órgãos ambientais sejam cumpridas. Benedetti et al, (2010). Assim, uma ETE será completamente confiável se não houver violação dos limites de cada parâmetro físico, químico e biológico do efluente final, limites estes preconizados pelas legislações ambientais.

Diante do exposto, a confiabilidade estabelece a exigência mínima para se determinar a magnitude da probabilidade de falha que poderá ser aceita. Neste contexto, utilizando dessa definição, estudos realizados por Niku et al. (1979, 1981b, 1981c), , avaliaram a confiabilidade de processos de lodos ativados, analisando 43 estações de tratamento de esgotos em operação nos EUA. No primeiro trabalho, foi desenvolvido um coeficiente de confiabilidade, onde a concentração média do constituinte (valor de projeto) se relaciona aos valores limites a serem cumpridos em uma análise de probabilidade. A partir desse modelo de confiabilidade obtido, os autores concluíram que é possível a utilização da distribuição log-normal para prever, tanto a qualidade do

efluente em termos de concentrações de DBO e SST, quanto à confiabilidade e ao desempenho de estações de tratamento de esgotos.

#### 4.2.1 Cálculo do coeficiente de Confiabilidade

A caracterização das distribuições de probabilidade dos indicadores, efetuada na Revisão da literatura, mostrou que a distribuição log-normal pode ser adotada para descrever o comportamento da maioria dos dados disponibilizados pelas ETEs. As equações prescritas da média e variância dos dados log-normais, foram utilizadas para desenvolver a equação de confiabilidade segundo a metodologia proposta por Niku et al (1979). O coeficiente de confiabilidade (CDC), desenvolvido pelos autores, pode ser utilizado para estimar a confiabilidade das estações de tratamento, ou seja, a fração de tempo que a concentração efluente não exceda certo valor padrão de referência previamente fixado. (Para isto, utilizam-se os coeficientes de variação que é dado pela razão entre o desvio padrão e a média aritmética).

Niku et al (1979) citam ainda que este coeficiente de confiabilidade relaciona valores das concentrações médias (valores de projeto) ao padrão a ser alcançado em uma base probabilística. No Decreto Estadual de São Paulo N° 8.468, de 8 de setembro de 1976 (capítulo II, seção II, artigo 18, tópico 5, página 6) estabelece-se o padrão de emissão de efluentes para qualquer fonte poluidora a serem lançadas, direta ou indiretamente, nas coleções de água. Nele define que a DBO<sub>5</sub> deve estar a no máximo de 60 mg/L. Este limite somente poderá ser ultrapassado no caso de efluente de sistema de tratamento de águas residuárias que reduza a carga poluidora em termos de DBO<sub>5</sub> do despejo em no mínimo 80%. Assim, supondo que estejamos seguindo o padrão de 60 mg/L para DBO<sub>5</sub> e definimos que ele deva ser cumprido em 90% do tempo. O processo deve ser projetado e operado para alcançar uma concentração média efluente tal que, mesmo com a variabilidade presente, este percentual seja alcançado. O valor médio do constituinte,  $m_x$ , poderá ser obtido a partir da Equação 22:

$$m_x = (CDC) \cdot X_s \quad (22)$$

Onde:

$m_x$  = concentração média do constituinte;

$X_s$  = meta de qualidade ou padrão fixado pela legislação ou norma;

CDC = coeficiente de confiabilidade

Supondo que, para uma determinada probabilidade de falha  $\alpha$  entre 0 e 1, mais perto de 0, deseja-se projetar um processo para cada variável log-normal  $x$  observada, com a propriedade descrita pela Equação 23:

$$P(x \leq X_s) = 1 - \alpha \quad (23)$$

Assim,  $(1-\alpha)$  é a probabilidade de que a variável  $x$  não exceda o padrão  $X_s$ .

A partir de então, Niku et al (1979) utilizaram algumas propriedades da distribuição log-normal padronizada. Desta forma, utilizando-se das tabelas elaboradas para a distribuição normal central reduzida, puderam ser utilizadas, após a transformação dos parâmetros, através da Equação 24:

$$P\left[Z < \frac{\ln X_s - \mu_y}{\sigma_y}\right] = 1 - \alpha \quad (24)$$

Onde:  $Z$  é uma variável normal padronizada e o termo  $\frac{\ln X_s - \mu_y}{\sigma_y}$  corresponde ao valor padrão normal variado e será chamado de  $(Z_{1-\alpha})$ , de modo que  $P(Z \leq Z_{1-\alpha}) = 1 - \alpha$ . Com isto, foi possível desenvolver a expressão do cálculo da probabilidade, a partir do valor médio necessário para que o processo garanta um nível de confiabilidade  $(1-\alpha)$ , dado um conhecido padrão  $X_s$  e um coeficiente de variação  $CV$  (que representa a razão entre o desvio padrão e a média dos dados amostrais), dado pela Equação 25.

$$CV = (\sigma_x / \mu_x). \quad (25)$$

Rearranjando as equações por meio de algumas manipulações algébricas (que não cabem nesse contexto), chega-se ao coeficiente de confiabilidade dado pela Equação 26.

$$CDC = \sqrt{CV^2 + 1} \cdot \exp[-Z_{1-\alpha} \cdot \ln(CV^2 + 1)] \quad (26)$$

Importante: O CDC relaciona o valor médio do constituinte,  $\mu_x$ , ao padrão  $X_s$  para um nível de confiabilidade de  $1 - \alpha$ . Ressalta-se que o CDC é expresso com base nas propriedades dos dados originais e não dos logaritmos dos dados. Observa-se que os valores da média e variância de  $x$ , que seguem uma distribuição log-normal, devem ser transformados, conforme as Equações 27 e 28.

$$\mu_y = \ln\left(\frac{\mu^2}{\sqrt{\mu^2 + \sigma^2}}\right) \quad (27)$$

$$\sigma_y^2 = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right) \rightarrow \sigma_y = \sqrt{\ln\left(1 + \frac{\sigma^2}{\mu^2}\right)} \quad (28)$$

### ***4.3 Técnicas para modelagem de Sistemas***

#### **4.3.1 Simulação de Monte Carlo**

A simulação de Monte Carlo é uma técnica matemática desenvolvida originalmente para modelar a difusão aleatória de nêutrons. É usada em uma ampla variedade de aplicativos, incluindo física, finanças e confiabilidade do sistema. Ela utiliza ferramentas estatísticas para modelar matematicamente um sistema ou processo da vida real e, então, estima a probabilidade de obter um resultado bem-sucedido. A distribuição estatística do processo a ser modelado deve ser determinada primeiro antes que a simulação de Monte Carlo possa ser aplicada. Ela envolve o uso de números aleatórios como uma ferramenta para calcular algo que não é aleatório, assim se torna uma ferramenta versátil para analisar e avaliar medições complexas usando um modelo de sistema e para experimentar o modelo no sentido de inferir o comportamento do sistema (ADERIBIGBE, 2014; BONATE, 2001; ROBERT, CASELLA, 1999).

#### **4.3.2 Análise por Árvore de Falhas**

É uma técnica de identificação de perigos e análise de riscos, que parte de um evento topo escolhida para estudo e estabelece combinações de falhas e condições que poderiam causar a ocorrência desse evento. A técnica é dedutiva e pode ser qualitativa e quantitativa. O objeto da Análise por Árvore de Falhas são os sistemas. Os focos da Análise por Árvore de Falhas são o evento topo e as sequências de eventos que o produzem. Tendo por objetivo selecionar um evento topo identificado por qualquer técnica de identificação de perigos. Construir os níveis subsequentes ou ramos, identificando falhas que podem causar a ocorrência do evento topo. Podem ser falhas aleatórias de componentes, falhas de modo comum, falhas humanas ou indisponibilidade de equipamentos. A Análise por Árvore de Falhas possibilita o cálculo da frequência e de probabilidade de ocorrência de eventos básicos (evento cuja frequência é conhecida e geralmente obtida de banco de dados ou outro registro) (SEIXAS, 2001).

### 4.3.3 Aplicações da Simulação de Monte Carlo e da Análise por árvore de falhas

Benedetti et al. (2010) realizaram um estudo sobre a modelagem probabilística e a avaliação de atualizações em ETE em um contexto de avaliação baseado na qualidade do efluente. No trabalho, os autores definiram confiabilidade como a porcentagem de tempo para que as concentrações esperadas de efluentes atendessem aos padrões de descarga especificados ou aos alvos de tratamento. Assim, o foco do trabalho foi estudar os benefícios e malefícios de 13 possíveis atualizações na ETE para melhoria dos sistemas de tratamento urbanos de águas residuárias. Para este estudo, basearam-se em um trabalho do próprio autor (BENEDETTI, 2006), apoiando-se nos seguintes passos:

- 1) Seja fornecida uma série temporal suficientemente longa e representativa ao modelo da ETE, a fim de considerar as perturbações do influente do processo em diferentes escalas de tempo (de minutos no efeito “primeiro flush<sup>1</sup>”) meses em infiltração;
- 2) As atualizações da ETE são modeladas;
- 3) O descarte do efluente no rio é modelado;
- 4) Os modelos de ETE e rio estão integrados;
- 5) As incertezas do modelo são caracterizadas e propagadas para os resultados do modelo por meio de simulações de Monte Carlo (uma vez que a incerteza nas previsões do modelo de ETE é considerada grande, deve sempre ser quantificada) e os resultados probabilísticos de simulação são avaliados a partir de fatores econômicos e ambientais;

No trabalho de Taheriyoun e Moradinejad (2014), os mesmos fizeram uso do método de análise de árvore de falhas e simulação de Monte Carlo para investigar as condições que proporcionam o desempenho de uma ETE, utilizando o conceito de confiabilidade e afirmam que normalmente, a simulação de Monte Carlo é usada quando o valor exato dos resultados não pode ser calculado com algoritmos determinísticos, sendo o princípio por trás da metodologia de Monte Carlo, simular as ocorrências dos eventos primários usando um gerador de números aleatórios. Dessa forma os autores estudaram o caso de uma ETE, que atende 85.000 pessoas, equivalente a uma vazão média de 30.240 m<sup>3</sup>/dia. Seus resultados mostraram que os fatores humanos têm o maior impacto

---

1 O “primeiro flush” é o escoamento inicial de uma forte chuva. Nesta fase, a água proveniente da chuva nos centros urbanos, onde existe uma alta proporção de superfície impermeabilizada, carrega uma quantidade maior de poluentes que são descarregados nos esgotos destinados para escoamento de águas.



no fracasso do sistema. Na ordem subsequente, há fatores climáticos, fatores mecânicos e sistema de esgoto. Entre os eventos básicos, o erro do operador, a alta carga de influente, o problema de design e a falha do equipamento são de maior impacto na ocorrência do evento principal.

#### ***4.4 Correlação amostral***

Muitos problemas em engenharia e ciências envolvem explorar as relações entre duas ou mais variáveis. A ideia de correlação, remonta a Sir Francis Galton (1822-1911), que foi o primeiro a usá-lo em um estudo das alturas (X) e filhos (Y). Galton ajustou uma linha de mínimos quadrados e a usou para prever a altura dos filhos a partir da altura dos pais. Ele determinou que se a altura dos pais fosse acima da média, a altura dos filhos seria também acima da média, mas ainda assim abaixo da altura dos pais. Um efeito similar foi observado para alturas abaixo da média. Ou seja, a altura dos filhos “regrediu” em direção à média. Conseqüentemente, Galton se referiu a essa correlação entre as variáveis a uma linha dos mínimos quadrados chamada de linha de regressão (BARROS, 2001).

Uma vez que tenhamos decidido ajustar uma linha reta, a correlação amostral serve para estudar o comportamento conjunto de duas variáveis quantitativas distintas. Ou, em outras palavras, mede o grau de associação entre duas variáveis aleatórias X e Y. Para o estudo do comportamento conjunto de duas variáveis poderiam ser usados:

a) O Diagrama de dispersão

Representação gráfica do conjunto de dados. Nada mais é do que a representação dos pares de valores num sistema cartesiano.

Em síntese três situações marcantes poderiam acontecer:

1) Se, quando uma das variáveis “cresce”, a outra, em média, também “cresce”, dizemos que entre as duas variáveis existem correlação positiva, tanto mais forte quanto mais perto de uma reta imaginária que os pontos estiverem;

2) Se, quando uma das variáveis “cresce”, a outra, em média, também “decrece”, dizemos que entre as duas variáveis existem correlação negativa, tanto mais forte quanto mais perto de uma reta imaginária os pontos estiverem;

3) Se, os pontos estiverem dispersos, sem definição de direção, dizemos que a correlação é muito baixa, ou mesmo nula. As variáveis nesse caso são ditas não correlacionadas. O Quadro 3 apresenta a correlação entre duas variáveis.

**Quadro 3 - Correlação entre as variáveis X e Y**

$r_{xy} = -1$	Correlação perfeita negativa
$-1 < r_{xy} \leq 0$	Correlação negativa
$r_{xy} = 0$	Correlação nula
$0 < r_{xy} < 1$	Correlação positiva
$r_{xy} = 1$	Correlação perfeita positiva
$0,2 < r_{xy} \leq 0,4$	Correlação fraca
$0,4 < r_{xy} \leq 0,7$	Correlação moderada
$0,7 < r_{xy} \leq 0,9$	Correlação forte

Fonte: <http://leg.ufpr.br>

#### b) Coeficiente de correlação

É um valor numérico, uma medida, para o grau de associação entre duas variáveis. Um dos métodos mais conhecidos para se determinar essa correlação é o método de correlação linear de Pearson, que é dado pela Equação (29):

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} \quad (29)$$

Onde:

$r_{xy}$  = Coeficiente de Correlação de Pearson ( $-1 < r_{xy} \leq 1$ ) e os desvios padrões entre as variáveis, são dados pelas Equações de 30 a 32.

$$S_{xx} = \sum x^2 - \frac{1}{n} (\sum x)^2 \quad (30)$$

$$S_{yy} = \sum y^2 - \frac{1}{n} (\sum y)^2 \quad (31)$$

$$S_{xy} = \sum xy - \frac{1}{n} (\sum x)(\sum y) \quad (32)$$

Uma observação importante na análise correlação, é que quando calculamos um coeficiente de correlação com base em dados amostrais, o valor de r obtido é apenas uma estimativa de um parâmetro correspondente, o coeficiente de correlação da população, que denotamos por  $\rho$  (letra grega rô). O que r mede em uma amostra,  $\rho$ , mede em uma população.

Para fazer inferências sobre  $\rho$  com base em r, devemos fazer várias suposições sobre as distribuições das variáveis aleatórias cujos valores observamos com exceção de que os X's não são constantes, e sim valores de uma variável aleatória com distribuição normal. .

Como a distribuição de r é bastante complexa, sob tais suposições, é comum fazermos as inferências sobre  $\rho$  na transformação Z de Fischer, uma mudança de escala de r para Z, dada pela Equação 33:

$$Z = \frac{1}{2} \cdot \ln \frac{1+r}{1-r} \quad (33)$$

Aqui, a abreviatura  $\ln$  denota o “logaritmo natural” ou, o logaritmo na base  $e=2,71828\dots$  que a transformação de Fischer exige. Neste escopo, Fischer mostrou, que, sob as suposições da análise de correlação normal e para qualquer valor de  $\rho$ , a distribuição de Z é aproximadamente normal, cujos valores da média e do desvio padrão, são dados pelas Equações 34 e 35, respectivamente.

$$\mu_z = \frac{1}{2} \cdot \ln \frac{1+\rho}{1-\rho} \quad (34) \quad \text{e} \quad \sigma_z = \frac{1}{\sqrt{n-3}}, \quad (35) \quad \text{logo:}$$

A equação então normalizada é fornecida pela Equação 36.

$$z = \frac{Z - \mu_z}{\sigma_z} = (Z - \mu_z) \sqrt{n-3} \quad (36)$$

Esta equação tem aproximadamente a distribuição normal padronizada. A aplicação dessa teoria é extremamente facilitada pela tabela de distribuição normal, que nos fornece os valores de Z correspondentes a  $r = 0,00, 0,01, 0,02, \dots, 0,99$ .

#### **4.5 Análise de Regressão**

A análise de regressão consiste na realização de uma análise estatística com o objetivo de verificar a existência de uma relação funcional entre uma variável dependente com uma ou mais variáveis independentes. Em outras palavras, consiste na obtenção de uma equação que tenta explicar a variação da variável dependente pela variação dos níveis das variáveis independentes. (BARROS, 2001).

Para se tentar estabelecer qual equação melhor representa o fenômeno em estudo, pode-se fazer um gráfico chamado de diagrama de dispersão e verificar como se comporta a variável dependente ou variável de resposta (Y) em função da variação do(s) níveis das variáveis independentes ou preditoras (X). Este comportamento pode-se apresentar de diversas maneiras: linear, quadrático, cúbico, exponencial, logaritmo, etc. Para isto, deve-se verificar qual tipo de curva e equação de um modelo matemático, que mais se aproxime dos pontos representados no diagrama de dispersão. Fox, J. (1997).

Fox, J. (1997), recomenda ainda que, o modelo escolhido deve ser coerente com o que acontece na prática. Para isto, devem-se levar em conta, as seguintes considerações no momento de se escolher o modelo:

- o modelo escolhido deve ser condizente tanto no grau como no aspecto da curva, para representar em termos práticos o fenômeno em estudo;
- o modelo deve conter apenas as variáveis que são relevantes para explicar o fenômeno.

Hosmer e Lemeshow (2000) citam que a análise de regressão é um dos métodos estatísticos mais usados na prática. Aplicações de análise de regressão podem ser encontradas em muitas ciências, campos, incluindo medicina, biologia, agricultura, economia, engenharia, sociologia, geologia etc. Os objetivos da análise de regressão são triplos:

- 1) Estabelecer uma relação casual entre a variável dependente (Y) e das variáveis independentes  $x_1, x_2, x_3, \dots, x_n$ .
- 2) Prever a variável dependente (Y) com base nos valores de  $x_1, x_2, x_3, \dots, x_n$ .
- 3) Identificar quais variáveis dependentes (X) são mais importantes para a previsão da variável (Y).

De um modo geral, Fox (1997) recomenda que sejam analisados os seguintes procedimentos para a regressão:

1) Verificar se o problema real frequentemente recai em um campo científico e se esta questão se enquadra no escopo da análise de regressão.

2) Definir o modelo de regressão que será escrito, na qual minha variável dependente (Y) de resposta é escrita em função das variáveis independentes (X), juntamente com um erro aleatório denominado  $\varepsilon$ . Matematicamente indicamos pela Equação 37:

$$Y = f(x_1, x_2, x_3, \dots, x_n) + \varepsilon. \quad (37)$$

3) Realizar suposições distributivas sobre o erro aleatório na regressão do modelo. Essas premissas precisam ser verificadas ou testadas pelos dados.

4) Coletar os dados com uma quantidade suficiente de amostras, mostrando o modelo de banco de dados com valores que sejam os mais reais possíveis, significativos para se desenvolver a regressão.

5) De acordo com o software usado na análise, criar conjuntos de dados através de um formato apropriado e fácil de ser lido em um software escolhido para a realização da regressão.

6) Avaliar cuidadosamente se o modelo selecionado é sólido e que responde de uma maneira satisfatória a variável dependente (Y).

7) Se o modelo não for adequado, estará sujeito a aprimoramento ou modificações.

#### 4.5.1 Regressão linear simples

Cook e Weisberg (1982) comentam que a regressão linear simples, considera um único regressor ou preditor X e uma variável dependente ou variável de resposta Y. Supõe que a relação verdadeira entre Y e X seja uma linha reta e que a observação Y em cada nível de X seja uma variável aleatória. Como notado previamente, o valor esperado de Y para cada valor de X é indicado pela Equação (38):

$$E(Y|X) = Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (38)$$

Onde:

$Y_i$  = valor observado para a variável dependente Y no i-ésimo nível da variável independente X;

$\beta_0$  = constante de regressão. Representa o intercepto da reta com o eixo dos Y;

$\beta_1$  = coeficiente de regressão. Representa a variação de Y em função da variação de uma unidade da variável X;

$X_i$  = i-ésimo nível da variável independente X (  $i=1,2,\dots,n$ ),

$\varepsilon_i$  = é o erro que está associado à distância entre o valor esperado  $Y_i$  e o correspondente ponto na curva, do modelo proposto, para o mesmo nível i de X.

Para obter-se a equação estimada, utiliza-se o método dos mínimos quadrados (MMQ), visando à minimização dos erros. Assim, isolando o erro aleatório temos a Equação (39):

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i \quad (39)$$

Elevando ambos os membros ao quadrado, obtém-se a Equação (40)

$$\varepsilon^2_i = [Y_i - \beta_0 - \beta_1 X_i]^2 \quad (40)$$

Aplicando o somatório, obtém-se a Equação (41)

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [Y_i - \beta_0 - \beta_1 X_i]^2 \quad (41)$$

Por meio da obtenção de estimadores de  $\beta_0$  e  $\beta_1$ , que minimizem o valor obtido, na Equação (41), é possível alcançar a minimização da soma dos quadrados dos erros.

Para se encontrar o mínimo de uma equação, deve-se derivá-la em relação à variável de interesse e igualá-la a zero. Derivando então a equação (41) em relação à  $\beta_0$  e  $\beta_1$ , igualando-as a zero, obtém-se duas Equações (42) e (43) que, juntas, vão compor o chamado sistemas de equações normais. A solução desse sistema fornecerá:

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{S_{xy}}{S_{xx}} \quad (42) \quad \text{e} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (43)$$

Uma vez obtidas essas estimativas, é obtida a Equação (44) estimada:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (44)$$

#### 4.5.1.1. Coeficiente de Determinação ou Correlação

Segundo Barros (2001), coeficiente de determinação representado por  $R^2$ , é uma medida de ajustamento de um modelo estatístico linear generalizado, em relação aos valores observados. O  $R^2$  varia entre 0 e 1, indicando, em percentagem, o quanto o modelo consegue explicar os valores observados, sendo que, quanto maior o  $R^2$ , mais explicativo é o modelo, melhor ele se ajusta à amostra. Valores próximos de 1 indicam que o modelo proposto é adequado para descrever os dados experimentais.

Para a obtenção do  $R^2$ , define-se as Equações de (45) a (49) como segue:

$$SQ_{total} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (45)$$

Onde:

$SQ_{total} = SQD_y =$  Soma total dos quadrados;

$n =$  número de observações;

$y_i =$  valor experimental observado na condição  $i$ ,

$\bar{y} =$  valor experimental médio das observações.

Partindo de  $y_i$  (o valor observado) e  $\bar{y}$  (a média das observações), essa equação (45) dá-nos a soma total dos quadrados, ou seja, a soma dos quadrados das diferenças entre cada valor observado e a média.

A diferença dos quadrados dos resíduos, que representa a parte não explicada pelo modelo, é calculada pela Equação (46).

$$SQ_{res} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (46)$$

Onde:

$SQ_{res}$  = diferença dos quadrados dos resíduos;

$\hat{y}_i$  = é o valor estimado (previsto) da variável  $y$  na condição  $i$ .

A soma dos quadrados dos resíduos obtida pela regressão, é calculada pela Equação (47).

$$SQ_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (47)$$

Onde:

$SQ_{reg}$  = Soma dos quadrados da regressão.

A Equação (47) então indica-nos a diferença entre o valor estimado para cada observação e a média das observações, somando os respectivos quadrados dessas diferenças. Quanto menor for a diferença, maior poder explicativo terá o modelo.

Em todos os casos temos a Equação (48):

$$SQ_{total} = SQ_{reg} + SQ_{res}. \quad (48)$$

Normalizando a equação acima, tem-se a Equação (49):

$$R^2 = \frac{SQ_{reg}}{SQ_{total}} = 1 - \frac{SQ_{res}}{SQ_{total}} \quad (49)$$

#### 4.5.1.2. Testes de hipótese na regressão linear simples

Barros (2001) relata que admitindo um valor hipotético para um parâmetro populacional, com base nas informações de uma amostra, pode-se realizar um teste estatístico para aceitar ou rejeitar esse valor hipotético. Como a decisão será tomada de acordo com os elementos da amostra, fica evidente que a decisão estará sujeita a erros. Entretanto, pode-se dimensionar a probabilidade (risco) da decisão de se aceitar ou rejeitar tal hipótese.

Para testar um parâmetro populacional, deve-se cuidadosamente formular um par de hipóteses: uma que represente a afirmação e outra que represente seu complemento. Quando uma dessas hipóteses for falsa, a outra deve ser verdadeira. Essas duas hipóteses são chamadas de hipótese nula ( $H_0$ ) e hipótese alternativa ( $H_a$ ).

A hipótese nula ( $H_0$ ) é uma hipótese estatística que contém uma afirmação de igualdade, tal como: menor igual ( $\leq$ ), igual ( $=$ ) ou maior igual ( $\geq$ ).

A hipótese alternativa ( $H_a$ ) é geralmente, o complemento da hipótese nula. É a afirmação que deve ser verdadeira se  $H_0$  for falsa e contém uma afirmação de desigualdade tal como demonstrado a seguir:

$$\begin{matrix} \{H_0: \mu = k & \{H_0: \mu = k & \{H_0: \mu = k & \{H_0: \mu = k & \{H_0: \mu \geq k \\ \{H_a: \mu \neq k & \{H_a: \mu > k & \{H_a: \mu < k & \{H_a: \mu > k & \{H_a: \mu < k \end{matrix}$$

Em um teste de hipótese, sempre se pressupõe que a hipótese nula ( $H_0$ ) é verdadeira, decorrendo naturalmente as seguintes decisões:

- 1) Aceitar  $H_0$ , rejeitando  $H_a$  ou,
- 2) Rejeitar  $H_0$ , aceitando  $H_a$ .

Pelo fato da decisão ser baseada em uma amostra ao invés de ser na população, há sempre a possibilidade de tomar a decisão errada.

Neste caso, há dois possíveis erros, quando se realiza um teste estatístico, para aceitar ou rejeitar  $H_0$ . Pode-se rejeitar a hipótese  $H_0$  quando ela é verdadeira ou aceitar  $H_0$  quando ela for falsa.

O erro de rejeitar  $H_0$ , sendo  $H_0$  verdadeira, é denominado Erro Tipo I e a probabilidade de se obter esse erro é designada por  $\alpha$ .

O erro de aceitar  $H_0$ , sendo  $H_0$  falsa, é denominado Erro tipo II e a probabilidade de se obter esse erro é designada por  $\beta$ .

O Quadro 4 a seguir sintetiza os possíveis erros e acertos em uma decisão de teste de hipótese.

**Quadro 4 – Teste de hipótese para a variável nula ( $H_0$ )**

REALIDADE	
$H_0$ verdadeira	$H_0$ falsa
Decisão correta ( $1-\alpha$ )	Erro Tipo II ( $\beta$ )
Erro Tipo I ( $\alpha$ )	Decisão correta ( $1-\alpha$ )

Fonte: BARROS, 2001



O tomador da decisão deseja, obviamente, reduzir ao mínimo as probabilidades dos dois tipos de erros. A redução simultânea dos erros poderá ser alcançada com aumento do tamanho da amostra, evidentemente, com aumento dos custos. Para um mesmo tamanho de amostras, a probabilidade de incorrer em um erro Tipo II aumenta à medida que diminui a probabilidade do erro Tipo I, e vice-versa.

O nível de significância de um teste é definido como sendo a probabilidade máxima permissível para cometer o erro Tipo I, ou seja, o nível de significância é igual ao valor de  $\alpha$ .

A probabilidade de um Erro Tipo II é o valor de  $\beta$ . O valor de  $(1 - \beta)$  é chamado de poder do teste. Ele representa a probabilidade de rejeitar a hipótese nula quando a hipótese alternativa for verdadeira, sendo o valor do poder é difícil (e às vezes impossível) de determiná-lo.

#### 4.5.1.3 Nível descritivo (p-valor)

Ao realizar um teste de hipótese, partimos de um dado valor de  $\alpha$ , pré-fixado para a regra de decisão. Uma alternativa é deixar a cargo de quem vai utilizar as conclusões.

A ideia consiste em calcular, supondo que a hipótese nula seja verdadeira, a probabilidade de se obter estimativas mais desfavoráveis ou extremas do que está sendo fornecida pela amostra. Essa probabilidade será o nível descritivo, denotado por p-valor. Valores pequenos do p-valor evidenciam que a hipótese nula é falsa, pois, sendo a amostra nossa ferramenta de inferência estatística sobre a população, ela fornece uma estimativa que teria probabilidade muito pequena de acontecer, se  $H_0$  fosse verdadeira. O conceito de “pequeno” fica a cargo do usuário, que assim decide qual usar para comparar com o valor obtido p-valor.

Quanto menor for o p-teste, mais evidências há para se rejeitar a hipótese nula. Um p-valor muito pequeno indica um evento incomum.

Para usar o p-valor na decisão de um teste de hipótese, basta compararmos o p-valor com:

- 1) Se  $p\text{-valor} \leq \alpha$ , rejeitamos  $H_0$  o que acarreta aceitar  $H_a$
- 2) Se  $p\text{-valor} > \alpha$ , aceitamos  $H_0$  o que acarreta rejeitar  $H_a$ .

De maneira alternativa poderíamos testar a significância da regressão pelo método da Análise de Variância (ANOVA- Teste F).

Segundo Barros (2001), a análise de variância (ANOVA- Analysis of Variance) é uma ferramenta para a comparação de vários grupos ou estratos de interesse. Consiste em fazer uma partição da variabilidade total da variável resposta Y em componentes de acordo com o modelo e o teste a ser feito, A ANOVA apresenta um nível de confiança  $\alpha$  determinado pelo analista (por exemplo um nível de confiança de 95%), assim a seguinte identidade pode ser verificada:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

Ou, em outras palavras, através da equação (50):

$$SQ_{total} = SQ_{reg} + SQ_{res}. \quad (50)$$

Onde:

$$SQ_{total} = \text{Variação total em Y} = SQ_{Dy}.$$

$SQ_{reg}$  = variação em Y explicada pela regressão ajustada =  $\hat{\beta}_1 \cdot SPD_{xy}$ , de modo que:

$$SQ_{reg} = SQ_{res} = \text{variação não explicada pela regressão} = SQ_{Dy} - \hat{\beta}_1 \cdot SPD_{xy}$$

Baseado nessa identidade, a Tabela 2 pode ser elaborada.

**Tabela 2 - Teste F- ANOVA**

FONTES DE VARIACÃO	GRAUS DE LIBERDADE	SOMA DOS QUADRADOS	QUADRADO MÉDIO	TESTE F
FV	GL	SQ	QM	F
Regressão	1	$SQ_{reg}$	$QM_{reg} = SQ_{reg}$	$QM_{reg} / QM_{res}$
Resíduo	n-2	$SQ_{res}$	$QM_{res} = SQ_{res} / (n-2)$	
Total	n-1	$SQ_{total}$		

Fonte: BARROS, 2001

A estatística F obtida no quadro acima serve para testar a significância da regressão, ou seja, testar  $H_0: \beta_1 = 0$  versus  $H_a: \beta_1 \neq 0$ .

- Regra de decisão: Se  $F_{calc} \geq F_{(\alpha, 1, n-2)} \Rightarrow$  rejeita  $H_0$

OBS.: Para  $H_0: \beta_1 = 0$ , temos que  $(t_{calc})^2 = F_{calc}$

A equação estimada obtida, apenas estabelece uma relação funcional, entre a variável dependente e a variável independente, para representar o fenômeno em estudo.

Portanto a simples obtenção da equação estimada não responde ao pesquisador se a variação da variável independente influencia significativamente na variação da variável dependente.

Para se responder a esta pergunta, é necessário realizar um teste estatístico para as estimativas dos coeficientes da equação de regressão estimada. Um teste que pode ser realizado para verificar tal fato é o teste F da análise de variância. Portanto, é necessário realizar uma análise de variância dos dados observados, em função do modelo proposto.

A Tabela 3 a seguir exemplifica o teste F para a análise de variância.

**Tabela 3- Teste F da Análise de variância**

FONTE DE VARIAÇÃO	GRAUS DE LIBERDADE	SOMA DOS QUADRADOS	QUADRADO MÉDIO	TESTE F
FV	GL	SQ	QM	F
Regressão	p-1	$SQ_{reg}$	$SQ_{reg}/p$	$QM_{reg}/QM_{res}$
Resíduo	n-p	$SQ_{ind}$	$SQ_{ind}/(n - 1 - p)$	
Total	n-1	$SQ_{total}$		

Fonte: BARROS, 2001

Onde:

p = nº de coeficientes de regressão (não inclui o  $\beta_0$ )

n = nº de observações.

GL= número do grau de liberdade

$SQ_{ind}$ = soma dos quadrados do termo independente

As fórmulas para a obtenção das somas dos quadrados totais e da soma de quadrados do termo independente da regressão são as mesmas, tanto para o modelo linear de 1º grau quanto para o de 2º grau, as quais são dadas pelas Equações 51 e 52 respectivamente:

$$SQ_{total} = \sum Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \quad (51)$$

$$SQ_{\text{Independente da Regressão}} = SQ_{ind} = SQ_{Total} - SQ_{reg} \quad (52)$$

A soma de quadrados para a regressão varia de acordo com o modelo em teste. Assim tem-se que, para o modelo linear de 1º grau, a soma de quadrados da regressão, é dado pela Equação 53.

$$SQ_{reg} = \hat{\beta}_0 \sum_{i=1}^n Y_i + \hat{\beta}_1 \sum_{i=1}^n Y_i X_i - \frac{(\sum_{i=1}^n Y_i)^2}{n} \quad (53)$$

No modelo do 2º grau a soma de quadrados da regressão temos a Equação 54.

$$SQ_{reg} = \hat{\beta}_0 \sum_{i=1}^n Y_i + \hat{\beta}_1 \sum_{i=1}^n Y_i X_i + \hat{\beta}_2 \sum_{i=1}^n Y_i X_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \quad (54)$$

As hipóteses estatísticas para o teste F são as seguintes:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ , o que significa dizer que as  $p$  variáveis independentes não exercem influência na variável dependente, segundo o modelo proposto,

$H_a: \beta_i \neq 0$ , para pelo menos um  $i$ , o que significa dizer que pelo menos uma das  $p$  variáveis independentes exercem influência na variável dependente, segundo o modelo proposto.

O valor de  $F$  da análise de variância deve ser comparado, com o valor de  $F$  tabelado ( $F_{tab}$ ), o qual se obtém na tabela da distribuição  $F$  de acordo com o nível de significância ( $\alpha$ ) do teste, e o número de graus de liberdade ( $n$ ) para a regressão e independente da regressão e é representado pela Equação 55:

$$F_{tab} = F_{\alpha}(p; n-1-p) \quad (55)$$

A regra decisória para o teste  $F$  é:

Se  $F \geq F_{tab} \Rightarrow$  Rejeita-se  $H_0$  ao nível de significância que foi realizado o teste.  
Se  $F < F_{tab} \Rightarrow$  não se rejeita  $H_0$  ao nível de significância que foi realizado o teste. Pode-se inferir que o modelo proposto não é adequado para descrever o fenômeno.

#### 4.5.2 Regressão linear múltipla

Segundo Fox, J. (1997) a regressão múltipla envolve três ou mais variáveis, ou seja, uma única variável dependente ( $Y$ ) e duas ou mais variáveis independentes ou explanatórias ou covariáveis ou regressoras ( $X_i, i = 1, 2, \dots$ ). A teoria é uma extensão da análise de regressão linear simples. De modo similar a análise tem por objetivo estabelecer uma equação que possa ser usada para prever valores de  $Y$  para valores dados das diversas variáveis independentes. A finalidade das variáveis independentes adicionais é melhorar a capacidade de predição em confronto com a regressão linear simples. A técnica de cálculo é bastante complicada e pode ser facilitada com o auxílio de álgebra de matrizes, cujas equações lineares são dadas conforme a Equação 56.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (56)$$

é chamado de modelo de regressão linear múltipla com k variáveis regressoras. Os parâmetros  $\beta_i$  (i=1 a k) são chamados de regressões parciais.

#### 4.5.2.1 Modelo Linear de 2º Grau

Fox, J. (1997) cita o modelo estatístico para a situação, através da Equação 57:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon \quad (57)$$

em que:

$Y_i$  = valor observado para a variável independente Y no i-ésimo nível da variável independente X.

$\beta_0$  = constante de regressão;

$\beta_1$  = coeficiente de regressão;

$\beta_2$  = coeficiente de regressão;

$X_i$  = i-ésimo valor da variável independente X (i=1,2,3,...n)

$X_i^2$  = i-ésimo valor da variável independente X, elevada ao quadrado;

$\varepsilon_i$  = é o erro que está associado à distância entre o valor observado  $Y_i$  e o correspondente ponto na curva para o mesmo nível i de X.

Utilizando o método dos mínimos quadrados, no modelo do 2º grau, chegar-se-á ao sistema de equações normais, para obter as estimativas  $\beta_0, \beta_1, \beta_2$ .

$$\begin{cases} \sum_{i=1}^n Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i + \hat{\beta}_2 \sum_{i=1}^n X_i^2 \\ \sum_{i=1}^n Y_i X_i = \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 + \hat{\beta}_2 \sum_{i=1}^n X_i^3 \\ \sum_{i=1}^n Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^n X_i^2 + \hat{\beta}_1 \sum_{i=1}^n X_i^3 + \hat{\beta}_2 \sum_{i=1}^n X_i^4 \end{cases}$$

Uma vez obtidas essas estimativas, podemos escrever a Equação 58 estimada:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 \quad (58)$$

#### 4.5.2.2 Análise da variância para a regressão linear múltipla

Segundo Fox (1997), na análise de regressão múltipla, o teste F produz um teste mais geral. Por meio da sua utilização determina-se se qualquer das variáveis independentes no modelo possui poder de explicação. Cada variável pode então ser testada individualmente com o teste t para determinar se é uma das variáveis significativas.

O mesmo autor comenta que a análise de variância baseia-se na decomposição da soma dos quadrados total, SQT, (que corresponde à variação da variável resposta), na soma dos quadrados explicada, SQR, (que corresponde à variação da variável resposta que é explicada pelo modelo) e na soma dos quadrados dos resíduos, SQE, (que corresponde à variação da variável resposta que não é explicada pelo modelo).

Desta forma, pode-se escrever a Equação 59:

$$SQ_{total} = SQ_{res} + SQE \quad (59)$$

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

Ou, em outras palavras, através da Equação 60:

$$SQ_{total} = SQ_{reg} + SQ_{res} \quad (60)$$

Assim, no conceito de regressão linear múltipla, as hipóteses a testar serão:

$$\begin{cases} H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1: \exists j: \beta_j \neq 0, j = 1, 2, \dots, p \end{cases}$$

Para testar a hipótese  $H_0$ , utiliza-se a estatística do teste dada pela Equação 61.

$$F = \frac{\frac{SQ_{res}}{p}}{\frac{SQE}{n-p-1}} = \frac{QM_{res}}{QM_E} \sim F_{p, n-p-1} \quad (61)$$

Com  $\frac{SQ_{res}}{\sigma^2} \sim \chi_p^2$ ,  $\frac{SQE}{\sigma^2} \sim \chi_{n-p-1}^2$  e  $SQ_{res}$  e  $SQE$  são independentes.

Assim, sob  $H_0$ , a estatística de F segue uma distribuição F central com p e n-(p+1) graus de liberdade,  $F_{p, n-p-1}$ .

Portanto, se  $F_{obs} > F_{(1-\alpha, p, n-p-1)}$ , o quartil 1- $\alpha$  de distribuição F central com p e n-p-1 graus de liberdade. Ao rejeitarmos  $H_0$ , concluímos que pelo menos uma das variáveis explicativas contribui significativa para o modelo. Estas somas podem ser apresentadas pela Tabela 4 a seguir.

**Tabela 4 – Distribuição F central com seus graus de liberdade**

FONTE DE VARIAÇÃO GRAUS DE LIBERDADE SOMA DOS QUADRADOS QUADRADO MÉDIO				
TESTE F				
FV	GL	SQ	QM	F
Regressão	p-1	$SQ_{reg}$	$QM_{reg} = SQ_{reg}/p$	
Resíduo	n-p	$SQ_{res}$	$QM_{res} = SQ_{res}/(n-p-1)$	$q = QM_{reg}/QM_{res}$
Total	n-1	$SQ_{total}$		

Fonte: BARROS, 2001

Como visto anteriormente, temos o valor da regressão através da Equação 62:

$$R^2 = \frac{SQ_{reg}}{SQ_{total}} = 1 - \frac{SQ_{res}}{SQ_{total}} \quad (62)$$

Este coeficiente é usado para quantificar a capacidade explicativa do modelo, ou seja, segundo Esteves e Sousa (2007), é uma medida da proporção da variação da variável resposta que é explicada pela equação de regressão quando estão envolvidas as variáveis independentes  $x_1, x_2, \dots, x_p$ .

#### 4.5.2.3 Detecção de *outliers* e sua Influência na regressão linear múltipla

Depois que os dados são coletados e limpos, o próximo passo é colocar os dados em um modelo estatístico para que a relação entre a variável resposta e as variáveis independentes possa ser estabelecida. Escolher um modelo estatístico depende em grande parte da natureza dos dados e do experimento e para isto, se o modelo selecionado for de regressão linear múltipla, a determinação do modelo é resolver os parâmetros de regressão. Depois, é necessário realizar testes de hipótese para realizar o diagnóstico do modelo.

Como já mencionado, para o modelo de regressão clássico, utilizamos o método dos mínimos quadrados. Se o erro do modelo da regressão for normalmente distribuído, os acertos dos parâmetros da regressão são iguais à probabilidade máxima estimada. Existem dois aspectos do diagnóstico da regressão:

1) O modelo escolhido deve ser suficiente para refletir a relação entre a variável resposta e as variáveis independentes.

2) Se há algum ponto que se desvie significativamente do modelo escolhido. Neste caso, temos o que chamamos de um *outlier* e assim, o mesmo deve ser descartado.

A seguir serão explicados alguns critérios básicos que são comumente usados para o diagnóstico do sistema de regressão.

1) Coeficiente de correlação  $R^2 = 1 - \frac{SQ_{res}}{SQ_{total}}$ , próximo a 1;

2) Estimativa do erro da variância  $S^2$ . Um modelo de regressão preferido deve ser aquele que possui um valor menor que  $S^2$ .

3)  $R^2$  ajustado =  $1 - \left( \frac{n-1}{n-(p-1)} \right) (1 - R^2)$

Alguns estatísticos preferem utilizar o  $R^2$  ajustado. O  $R^2$  ajustado não necessariamente aumenta com a inserção de parâmetros no modelo. Aliás, se adicionarmos variáveis com pouco poder explicativo este tende a decrescer. Pelo que, quando existe uma diferença significativa entre  $R^2$  e  $R^2$  ajustado, estamos perante uma situação em que provavelmente tenham sido incluídas no modelo variáveis estatisticamente não significativas (Seber, 1977). Consequentemente, um critério para a seleção do modelo é escolher o que tem o  $R^2$  ajustado máximo.

De acordo com Pires e Branco (2007), *outliers* são observações extremas que se encontram de tal forma afastadas da maioria dos dados que surgem dúvidas sobre se elas poderão ou não ter sido gerada pelo modelo proposto para explicar a maioria dos dados.

Segundo Pires e Branco (2007), para se verificar se um determinado dado é um *outlier*, emprega-se uma regra empírica largamente utilizada no meio científico sendo que todos os conjuntos de dados dessa pesquisa foram tratados por esse processo.

Regra de exclusão da observação:

- Se observação  $>$  Quartil superior  $+1,5$  (Quartil superior – Quartil Inferior): exclui-se o dado ou
- Se Observação  $<$  Quartil inferior  $-1,5$ : exclui-se também o dado.

Onde:

Quartil superior: Percentil 75% ou Q3 (ou seja, que 75% dos dados estejam abaixo desse valor).

Quartil Inferior: Percentil 25% ou Q1 (ou seja, que 25% dos dados estejam abaixo desse valor).

Se um *outlier* for influente vai interferir sobre a função de regressão ajustada o que significa que a inclusão ou não desse ponto modifica substancialmente os valores ajustados. Assim, um ponto é influente se a sua exclusão na regressão ajustada provoca uma mudança substancial nos valores ajustados. Em relação aos resíduos esses são dados pela diferença entre os valores da variável resposta observada e a variável resposta estimada, isto é dado pela Equação 63:

$$\varepsilon_i = y_i - \hat{y}_i, i = 1, 2, \dots, n. \quad (63)$$

Ao realizarmos uma análise de resíduos pretendemos verificar se o modelo de regressão que está a ser utilizado é adequado. Para tal os resíduos devem verificar os pressupostos anteriormente impostos ao erro do modelo. Tais pressupostos são considerando o modelo  $Y = X\beta + \varepsilon$ , onde:



$Y = (y_1, y_2, \dots, y_n)'$ ,  $X$  é a matriz do modelo  $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$  e  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$

- a)  $\varepsilon_i$ ,  $i = 1, 2, \dots, n$  são normalmente distribuídos;
- b)  $Var(\varepsilon_i) = \sigma^2$ ,  $i = 1, 2, \dots, n$ , tem variância constante;
- c)  $\varepsilon_i$  e  $\varepsilon_j$ ,  $i \neq j$ , são independentes;

No caso da regressão linear múltipla, para além destes pressupostos, é preciso ainda verificar se existe colinearidade ou multicolinearidade entre as variáveis explicativas. A seguir serão apresentadas algumas “técnicas” por forma a verificar estes pressupostos.

#### 4.5.2.4 Diagnóstico de Normalidade dos dados

A normalidade dos resíduos pode ser analisada quer através de gráficos, quer usando alguns testes, nomeadamente através:

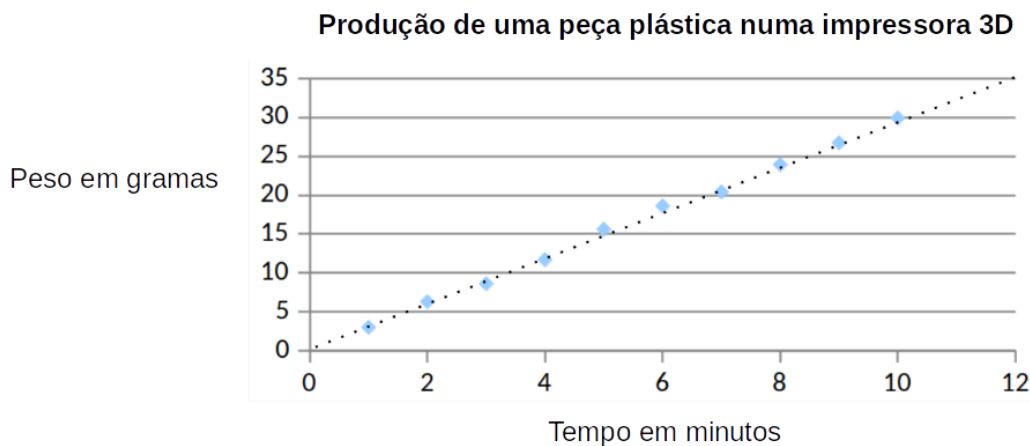
- a) gráfico P-P plot dos resíduos;
- b) histograma dos resíduos estandardizados;
- c) teste de Kolmogorov-Smirnov;
- d) teste de Shapiro-Wilk.

##### a) Gráfico P-P plot dos resíduos;

Neste gráfico, a distribuição de probabilidades dos valores observados com os valores esperados, representada por uma diagonal, segundo uma distribuição normal. Caso a normalidade se verifique, as observações registadas aproximam-se dessa diagonal, sem nenhum afastamento significativo.

A figura 6 representa a normalidade, aproximada através de uma reta .

**Figura 6 – Gráfico p-p plot de resíduos ajustado pela normalidade**

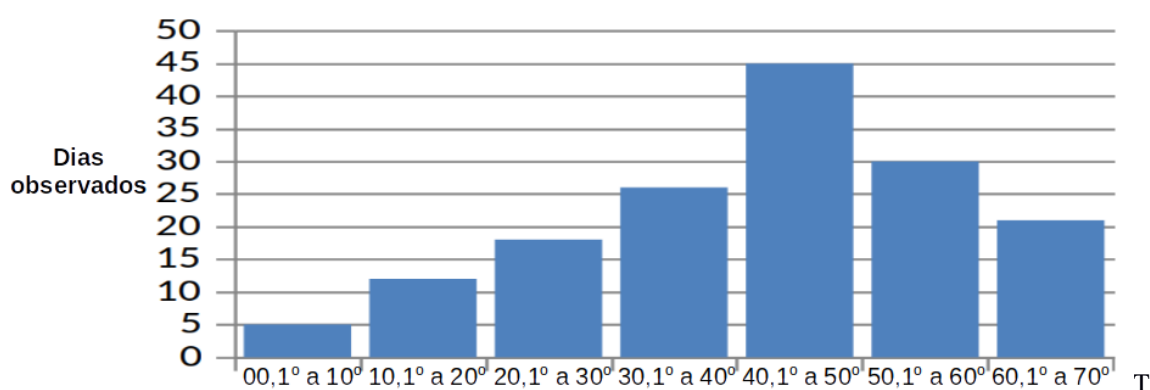


Fonte: Próprio autor

### b) Histograma dos resíduos estandardizados

Também se pode fazer um histograma dos resíduos no qual se procuram afastamentos evidentes em relação à forma simétrica e uni modal da distribuição normal. Este gráfico apenas deverá ser utilizado em amostras de dimensão elevada, já que quando se trabalha com amostras de dimensão reduzida o histograma não é muito conclusivo. A Figura 7 ilustra um histograma.

**Figura 7 – Temperatura de um composto em uma reação**  
**Temperatura do composto em reação**



Fonte: Próprio autor

### c) Teste de Kolmogorov-Smirnov (K-S)

Neste caso o teste de K-S é utilizado para testar as hipóteses:

$$\begin{cases} H_0: a \text{ distribuição é normal} \\ H_1: a \text{ distribuição não é normal} \end{cases}$$

A estatística de teste, segundo MAROCO (2003), é dada pela Equação 64:

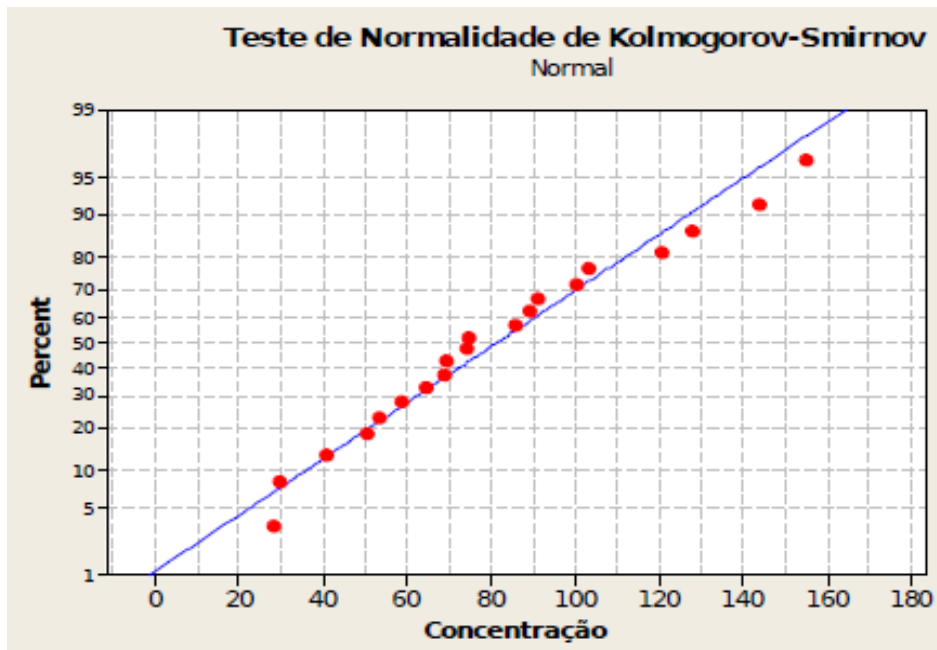
$$D = \max \left\{ \max \left( \left| F(x_i) - F_0(x_i) \right| \right) \right\} \quad (64)$$

em que  $F(x_i) - F_0(x_i)$  representa a diferença entre a frequência acumulada de cada uma das observações e a frequência acumulada que essa observação teria, sendo a sua distribuição normal.

Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada assumida pelos dados, neste caso da distribuição normal, e a função de distribuição empírica dos dados.

A Figura 8 ilustra o teste de Kolmogorov – Smirnov para Normalidade

**Figura 8 – Teste de Kolmogorov – Smirnov para Normalidade**



Fonte: Próprio Autor

#### **d) Teste de Shapiro-Wilk (S-W)**

Segundo Maroco (2003), este teste sugere-nos preferência em relação ao teste de K-S para amostras de pequenas dimensões ( $n < 30$ ). Neste caso, as hipóteses a serem testadas são as definidas anteriormente para o teste de K-S.

A estatística de teste é definida conforme a Equação 65:

$$W = \frac{\sum_{i=1}^n (a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (65)$$

Onde:

$a_i$ : são constantes geradas à partir da média, variância e covariância de  $n$  ordem.

## 5. MATERIAIS E MÉTODOS

### 5.1 Área de estudos

A ETE Candeia está localizada a cerca de 550 metros do bairro Núcleo Residencial Edison Bastos Gasparini no município de Bauru. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE) a população estimada no município de Bauru-SP no ano de 2017 foi de aproximadamente 371.690 habitantes, sendo das 21 cidades que compõem a macrorregião, a mais populosa (<<https://cidades.ibge.gov.br/>> acesso em 07 de fevereiro de 2018). Nesta cidade, a empresa que administra o saneamento básico é o DAE. A Figura 9 ilustra a Estação de Tratamento de Esgoto de Candeia.

**Figura 9 - Vista aérea da ETE Candeia.**



Fonte: Site do DAE

A ETE Candeia trata os esgotos sanitários ou domésticos lançados na bacia do córrego do Pau D'alho e do córrego Monte Belo, ambos afluentes do Ribeirão Água Parada, a noroeste do município, que recebe o esgoto gerado por diversos núcleos habitacionais e loteamentos da cidade (Gasparini, Índia Vanuíre, Pousada da Esperança I, Pousada da Esperança II, Jardim Helena, Vila São Paulo, Nova Bauru, Vitória Régia). Segundo o DAE, em comunicados à população, esta ETE possui uma capacidade volumétrica para tratar o esgoto de até 50.000 habitantes e na ocasião recebia cerca de 10% do esgoto doméstico do município (<<http://www.daebauru.sp.gov.br>> acesso em 17 de maio de 2018).

## ***5.2 Unidades do Sistema de Tratamento do Estudo***

Em relação à ETE Candeia, por conta do seu uso recente e de vários dos parâmetros das ETEs no Brasil terem sido baseadas em estudos estrangeiros (CAMPOS, 1999), o corpo técnico desta estação acredita que nas condições locais seria possível, apenas com reajustes não estruturais na unidade, melhorar o desempenho dos seus dois reatores e conseqüentemente reduzir a quantidade de resíduos não tratados despejados nos afluentes do Ribeirão Água Parada, o que proporcionaria benefícios ao abastecimento de cidades localizadas à jusante desta estação.

Conceitualmente o sistema de tratamento de esgoto doméstico é do tipo Reator Anaeróbico de Manta de Lodo e Fluxo Ascendente (UASB) + Biofiltro (BF) + Decantador Secundário (DS) + Desinfecção por Sistema de Cloração (SC), com capacidade para tratamento de uma vazão média de 60,0 L/s.

A ETE do tipo UASB + BF + DS + SC, constitui-se em um processo capaz de realizar o tratamento de esgoto pela associação em série de reatores anaeróbios, biofiltros, decantador secundário e sistema de cloração, atingindo eficiência de remoção de matéria orgânica superior a 90%, sendo suas principais vantagens:

- Compacidade;
- Simplicidade operacional;
- Baixo custo de implantação e operação;
- Baixo impacto em ambientes urbanos (ruído, odor, visual);
- Geração de 60% a menos de lodo que os processos convencionais.

Durante o processo de reconhecimento da estação, constatou-se que seu funcionamento, diversos parâmetros físicos químicos e biológicos não atingiam a regulamentação do órgão ambiental do CONAMA430/2011.

Por esta razão, o corpo técnico local inseriu duas novas etapas no protocolo previsto, que são: um Decantador no Primeiro Sistema de Pré-Tratamento; um Segundo Sistema de Pré-Tratamento após a passagem do esgoto pela Estação Elevatória. Considerando as modificações efetuadas pelo corpo técnico, tem-se o seguinte processo de tratamento orientado nos registros:

- a) *Primeiro Sistema de Pré-Tratamento*
  - Gradeamento
  - Caixa de areia

- b) *Decantador*
- c) *Estação elevatória de esgoto*
- d) *Segundo Sistema de Pré-Tratamento*
  - Gradeamento
  - Caixa de areia
  - Decantador
- e) *UASB*
  - Biofiltro
  - Sistema de aeração
  - Material filtrante
  - Decantador secundário
  - Desinfecção por sistema de cloração
  - Desidratação do lodo de descarte por centrífuga
  - Tratamento do biogás

A Figura 10 detalha as unidades e seus componentes.

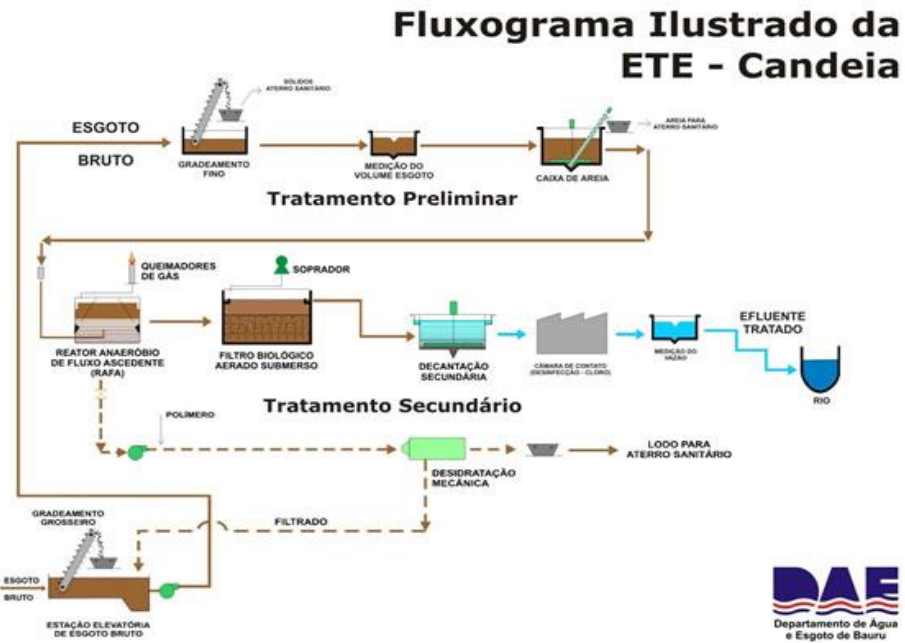
**Figura 10 – Unidades/Componentes do tratamento na ETE Candeia**

Unidade	Componentes
<b>Estação elevatória</b>	Poço e conjunto moto-bomba
<b>Pré- tratamento</b>	Gradeamento Caixa de areia
<b>Tratamento Primário</b>	Reatores Anaeróbios de Manta de lodo e fluxo ascendente UASB)
<b>Tratamento secundário</b>	Biofiltro ( BF) Decantadores Secundários (DS)
<b>Tratamento terciário</b>	Sistema de Cloração (SC)
<b>Tratamento do Biogás</b>	Queimador de gás
<b>Desidratação e Estabilização do Lodo</b>	Centrífuga

Fonte: Elaborado pelo autor

O Fluxograma completo do sistema é apresentado pela Figura 11 a seguir.

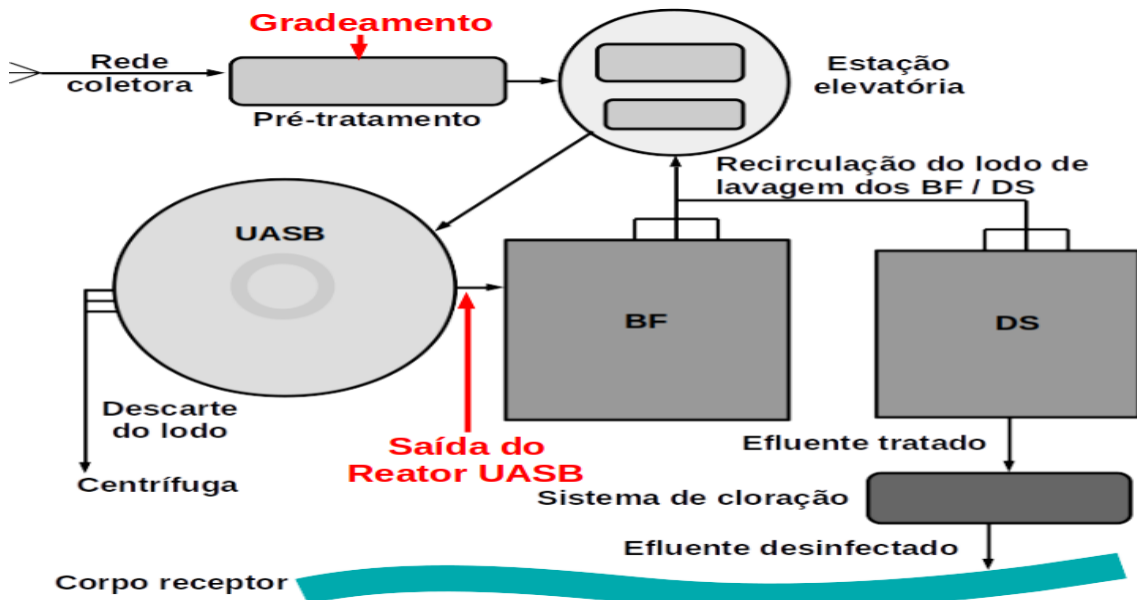
Figura 11 – Fluxograma das etapas da ETE Candeia



Fonte: DAE – BAURU

As etapas do sistema em vista superior, conforme Figura 12.

Figura 12 – Etapas do sistema de tratamento da ETE Candeia em planta.



Fonte: DAE- Bauru

### **5.2.1 Descrição de cada componente das unidades**

Estação Elevatória: Toda vez que não é possível o escoamento do esgoto pela ação da gravidade é necessário o uso de estações elevatórias para recalcar o esgoto de um ponto a outro com uma cota mais elevada. A Estação elevatória possui duas bombas que recalcam o esgoto até a caixa desarenadora. A Figura 13 ilustra a estação elevatória.

**Figura 13 – Estação elevatória da ETE Candeia**



Fonte: Elaborado pelo autor

Gradeamento: tem por objetivo remover sólidos grosseiros e proteger o conjunto moto-bomba que compõe a estação elevatória de esgoto bruto. É constituído por um cesto com grade média, com limpeza manual, onde o material retido é removido periodicamente e disposto em aterro sanitário. A Figura 14 ilustra a etapa do gradeamento.

**Figura 14 – Etapa do Gradeamento da estação**



Fonte: Elaborado pelo autor

Calha Parshall: É um trecho curto de canal, com geometria de fundo e paredes que acelera a velocidade do esgoto, e cria uma passagem por escoamento crítico. A



medição de nível é feita a montante, e pode ser relacionada diretamente à vazão. A Figura 15 ilustra a calha Parshall.

**Figura 15 – Detalhe da Calha Parshall**



Fonte: Elaborado pelo autor

Caixa de areia: A areia presente no esgoto deposita-se no fundo da unidade desarenadora, enquanto a matéria orgânica permanece em suspensão. A unidade desarenadora é do tipo canal, sendo que a areia é removida periodicamente com uma frequência quinzenal e com um auxílio de um caminhão fossa é acondicionada em bombonas para uma posterior disposição em aterro sanitário. A Figura 16 ilustra a caixa de areia

**Figura 16 – Detalhe da caixa de areia**



Fonte: Elaborado pelo autor

A Figura 17 mostra o Reator UASB, sendo que sua descrição e funcionamento já foram previamente discutidos no capítulo Justificativa e Fundamentação Teórica.

**Figura 17–Detalhe do Reator UASB de Candeia**

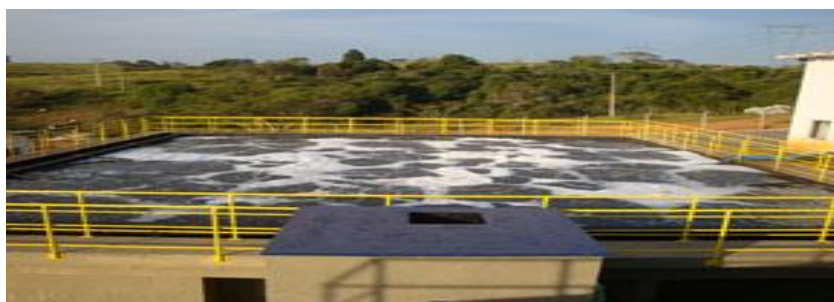


Fonte: Elaborado pelo autor

**Biofiltro (BF):** Os biofiltros são constituídos por meio de aeradores, provedores de ar via compressores e um tanque preenchido com material filtrante. O Leito filtrante tem a função de servir como meio suporte para a biomassa ativa. O esgoto e o ar fluem continuamente, ambos em fluxo ascendente.

Os biofiltros recebem o efluente anaeróbico do reator UASB. Nesta etapa, grande parte da matéria orgânica remanescente é metabolizada aerobicamente, ou seja, na presença de oxigênio. A principal função do biofiltro é a remoção de componentes orgânicos e nitrogênio na forma de nitrito, e segundo (Von Sperling, 2005), nesta fase a eficiência global de remoção média é na faixa de 83 a 95%. A Figura 18 detalha o biofiltro.

**Figura 18 – Detalhe do biofiltro**



Fonte: Elaborado pelo autor

**Decantador secundário (DS):** É a unidade que realiza o polimento final do efluente tratado, propiciando a redução de partículas de DQO, DBO<sub>5</sub>, sólidos em suspensão e alta taxa especialmente de fosfato e nitrato, a teores muito baixos, superiores

a 90%. O esgoto decantado sai pela parte de cima do decantador, após ser escoado pelas lâminas e é coletado por calhas coletoras. A Figura 19 mostra o decantador secundário.

**Figura 19–Detalhe do decantador secundário**



Fonte: Elaborado pelo autor

Sistema de cloração: Após o tratamento de efluentes, o cloro não apresenta boa eficiência na remoção de protozoários, devido ao seu maior tamanho, devendo haver um processo auxiliar de filtração a fim de removê-los. Para isto é realizado um sistema de desinfecção do efluente. Este sistema consiste em um gerador contínuo de hipoclorito de sódio produzido no mesmo local de aplicação, permanecendo o efluente por meia hora na câmara de contato por chicanas para a eliminação dos coliformes e posteriormente lançado na escada de aeração.

A Figura 20 mostra o sistema de cloração.

**Figura 20 – Detalhe do Sistema de cloração**



Fonte: Elaborado pelo autor

Tratamento do Biogás: Tem por finalidade, queimar o gás coletado pela câmara de gás, quando existir uma quantidade suficiente do mesmo, reduzindo o odor gerado pelo processo.

Descarte de Lodo: Diariamente, por meio da tomada de amostra no reator UASB, deve-se monitorar a altura da manta de lodo para não ultrapassar a altura de 3m. Quando a manta alcançar esta altura, deve ser feito a desidratação do lodo biológico utilizando-se centrifugas e a adição de polímeros catiônicos. O lodo é desaguado, para manter teor de sólidos entre 20 e 30% e acondicionando-o em caçambas para disposição final no aterro sanitário.

### ***5.3 Lançamentos dos efluentes tratados da ETE***

O lodo biológico da ETE Candeia é classificado como Resíduo Classe II-A (não perigoso e não inerte pela Norma Técnica NBR 10.004/2004 Classificação dos Resíduos Sólidos). O DAE tem CADRI (Certificado de Movimentação de Resíduos de Interesse Ambiental) expedido pela CETESB para a correta disposição do lodo biológico. O lançamento dos efluentes tratados da ETE Candeia é realizado no córrego Pau D'Alho, afluente do Ribeirão Água Parada, considerado Classe II, segundo a resolução do CONAMA 430/2011 e o Decreto Estadual nº 8.468/76. A Figura 21 mostra o local do lançamento dos efluentes tratados.

**Figura 21 - Destino dos Efluentes**



Fonte: Elaborado pelo autor

### ***5.4 Metodologias da Pesquisa***

Neste trabalho, investigou-se como alguns dos parâmetros coletados na etapa inicial (que precede a entrada do esgoto no reator UASB) influenciam os parâmetros de saída do reator. A investigação foi baseada nos dados disponibilizados pelo Laboratório de Águas Residuárias – LAR, os quais foram utilizados tais dados para fins desta pesquisa

sobre modelagem matemática dos dois Reatores UASB em uso na ETE Candeia, localizada no bairro Núcleo Residencial Edison Bastos Gasparini, na cidade de Bauru – SP. Foram disponibilizados quatro arquivos no formato. xlsx (planilhas eletrônicas, do “Excel”) referentes aos registros dos pontos de coletas e amostras das águas residuárias da ETE estudada. Os dados estavam divididos em grupos relativos aos seguintes períodos:

1. 06/01/2015 – 14/12/2015
2. 04/01/2016 – 19/12/2016
3. 03/01/2017 – 27/12/2017
4. 03/01/2018 – 25/06/2018

Os parâmetros cujos dados foram disponibilizados nessas planilhas estão listados no Quadro 5 a seguir:

**Quadro 5 - Organização dos dados coletados no laboratório de Águas Residuárias**

<b>Informações por linha</b>
Nome do coletor;
Pontos de coleta;
Contagem das análises realizadas;
Data;
Ocorrência de chuvas nas últimas 24h;
Horário da coleta;
Vazão da ETE em L/s;
Temperatura amostra em °C;
Cor Verdadeira em PtCo;
Condutividade em $\mu\text{S}/\text{cm } 25^\circ\text{C}$ ;
pH;
Turbidez em NTU;
DBO <sub>5</sub> em mgO <sub>2</sub> /L;
DQO em mgO <sub>2</sub> /L;
Oxigênio Dissolvido em mgO <sub>2</sub> /L;
Amônia em mg NH <sub>3</sub> /L;

Continua

Nitrogênio Amoniacal em mgNH <sub>3</sub> -N/L;
Nitrogênio Total em mg NTK /L;
Nitrogênio Nitrato em mgNO <sub>3</sub> -N /L;
Nitrogênio Nitrito em mgNO <sub>2</sub> -N /L;
Sólidos Totais em mg/L;
Sólidos Fixos em mg/L;
Sólidos Voláteis em mg/L;
Sólidos Sedimentáveis em mL/L;
Sólidos Totais Dissolvidos em mg/L;
Fósforo Total em mgP/L;
Óleos e Graxas em mg/L;
Cloro Total em mg/L;
Sólidos Suspensos Totais em mg/L;
Sólidos Suspensos Fixos em mg/L;
Sólidos Suspensos Voláteis em mg/L;
Coliformes Totais em NMP/100ml;
Coliformes Fecais em NMP/100ml;
Fosfato em mg(PO <sub>4</sub> ) <sup>-3</sup> /L;
Cor aparente em uC-PtCo;
Considerações.

Fonte: Elaborado pelo autor

Neste intervalo de 1.267 dias (entre 06/01/2015 – 25/06/2018), a partir das informações disponíveis dos arquivos, pôde-se contabilizar 164 dias de coleta. Para todos os constituintes de interesse (afluente e efluente) que possuíam todos os dados disponibilizados, nesses arquivos, realizaram-se as estatísticas bases relativas ao número de dados, tais como: média aritmética e geométrica, mediana, moda, valores máximos e mínimos, amplitude (diferença entre o maior e o menor valor de cada parâmetro), quartis inferior e superior, percentis de 10% e de 90%, variância, covariância, desvio padrão, assimetria e curtose.

As mesmas planilhas estão organizadas por colunas a partir das seguintes informações a partir da linha “Pontos de Coleta”, conforme o Quadro 6.

**Quadro 6 – Organização dos dados pelos Pontos de coleta**

<b>Informações por Colunas</b>
Gradeamento;
Saída do reator UASB;
Saída do biofiltro;
Saída do decantador;
Saída da câmara de contato;
Escada;
Jusante;
Manancial Pau d’alho;
Manancial Monte Belo.

Fonte: Elaborado pelo autor

### **5.5 Ensaaios e tratamentos de dados**

Conhecidas as análises estatísticas de cada parâmetro, condições nas quais o modelo matemático deve ser aplicado e as necessidades a suprir, sendo de caráter prático as atividades do corpo-técnico na ação de prevenção e identificação de erros, é apresentado a seguir passo o passo deste processo, com as justificativas lógicas e práticas para cada ação adotada.

#### **5.5.1 Organização dos parâmetros coletados**

Como procedimento inicial, os parâmetros coletados foram organizados em grupos de acordo com os processos de ação e reação estudados. Como por exemplo, a entrada do material no reator UASB e a saída do material após seu tratamento. Estes dados são agrupados em uma única folha de planilha eletrônica distribuídos em linhas pelos dias em que as coletas foram realizadas, e em colunas de acordo com os parâmetros coletados, dos quais, por uma questão de simplicidade linguística, serão tratados neste capítulo como parâmetros de entrada e parâmetros de saída. Nas ocasiões em que um mesmo parâmetro de entrada for analisado na saída, estes serão mencionados como parâmetros pareados.

### 5.5.2 Redução de parâmetros

Sem perda de generalidade, os parâmetros cujos dados coletados possuem pouca frequência, devem ser eliminados, pois em se tratando de construir um modelo de regressão linear múltipla dos parâmetros de entrada para prever os parâmetros de saída, é necessário um número considerável de dados para análise. Porém, este é um processo que se mostra dependente das quantidades gerais de dados, cabendo ao pesquisador uma análise e observação sobre as quantidades de parâmetros disponíveis para cada critério de frequência mínima adotada, enfatizando que o objetivo da regressão linear múltipla, é prever o comportamento de um parâmetro supostamente dependente de outros parâmetros independentes conhecidos.

### 5.5.3 Eliminação de *outliers*

Erros de digitação por vezes podem ser percebidos na forma de *outliers*, como parâmetros registrados com uma ordem de grandeza a mais ou a menos do que o padrão observado, ou até mesmo valores negativos para quantidades de massa medidas.

Na medida em que os menores e maiores valores medidos nas análises laboratoriais (chamados parâmetros de borda) de cada parâmetro possam estar cometidos de erros de digitação e sabendo que estes não interferirem na curva média das funções, podemos admitir que fossem potenciais *outliers* e sendo assim foram eliminados. Contudo, outras tantas vezes a presença dos *outliers* é uma parte natural do sistema, que em ocorrências atípicas, atingem parâmetros irregulares. A eliminação de *outliers* não prejudica o sistema a ser calculado pelas regressões, pois o próprio sistema previsto no modelo tem por objetivo responder à maioria dos casos, não a sua totalidade. Dessa forma, resultados atípicos, são pela sua própria classificação, casos incomuns.

Para a eliminação dos *outliers*, é preciso se embasar na quantidade total de dados coletados dos parâmetros do sistema, e na observação dos casos atípicos, sejam naturais ou por erros de medição. Neste processo, a eliminação dos potenciais *outliers* seguiram as recomendações prescritas por Pires e Branco (2007), no item 4.5.2.3 da Revisão da Literatura.



#### 5.5.4 Normalidade ou Log-normalidade

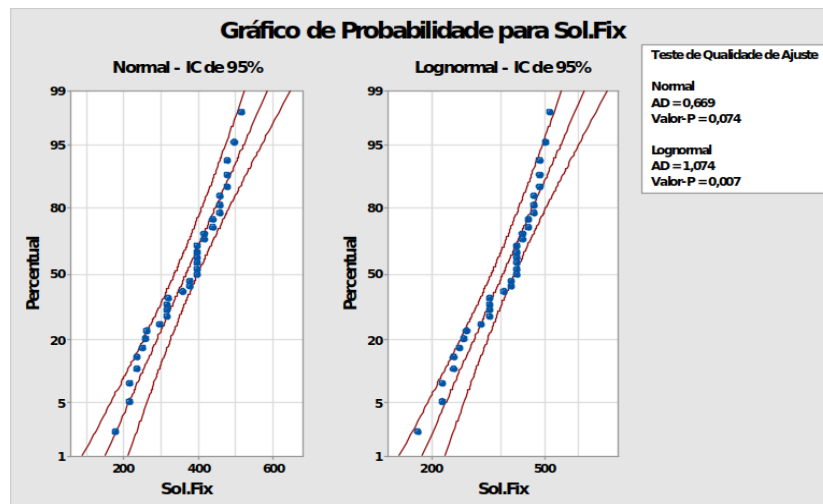
Mas antes do procedimento de regressão ser aplicado, se faz necessário o cálculo de correlações, visando eliminar os parâmetros de entrada que sejam fortemente correlacionados. Neste quesito, é preciso primeiramente considerar qual teste de correlação usar, no caso, se é um teste paramétrico (na qual se supõe que as populações das quais provêm às amostras tenham aproximadamente a forma de distribuição normal) ou não paramétrico (na qual não se saiba ou se tem dúvidas se as suposições de normalidade são atendidas).

Por fim, a partir deste resultado, o pesquisador deve optar pela eliminação dos parâmetros de entrada que não possuem comportamento normal nem log-normal, e seguir o procedimento de modelagem com os parâmetros restantes. Outra possibilidade é permanecer com todos os parâmetros e aderir a um teste não paramétrico para calcular suas correlações.

Uma característica particular para as ETE, que foi observada em diversas pesquisas relacionadas à confiabilidade destes sistemas, é abordar o comportamento de seus parâmetros como log-normais. Dessa forma, cabe também determinar quais dos parâmetros possuem uma distribuição log-normal. Nesta pesquisa, diversos parâmetros foram determinados (com softwares específicos como excel, Gnumeric ou LibreofficeCalc) como normais ou log-normais, em sendo assim, adotou-se os testes paramétricos, e nesse caso, as condições de normalidade dos parâmetros de entrada devem ser atendidas.

Caso ambas as hipóteses, de normalidade ou log-normalidade para um  $p$  estipulado chamado de ( $p$ -valor) sejam aceitas, é preciso tomar uma para a continuidade do processo (ou seja, o uso dela no cálculo da correlação de Pearson), assim, seria tomada aquela que possui a melhor qualidade de ajuste com a distribuição, que foi calculada numericamente através da correlação de Pearson e demonstrada para melhor lustração na Figura 22 (através do software Gnumeric) na qual se adotou a normalidade para os parâmetros dos sólidos fixos, pois foi a que possuiu o menor coeficiente de aderência (AD).

**Figura 22 - Gráfico de distribuição de probabilidade para o parâmetro Sólidos Fixos**



Fonte: Elaborado pelo Autor

Para os parâmetros de saída, pode-se utilizar o teste não paramétrico de Kolmogorov-Smirnov, cuja hipótese estatística pode ser aplicada em dados que possuam ou não distribuição normal. De forma simplória, isto pode ser calculado convertendo os valores dos dados coletados para logaritmo neperiano (conforme Revisão da literatura – item 4.1.1) e calculando a partir destes novos valores, sua normalidade por testes como Kolmogorov-Smirnov.

### 5.5.5 Teste de correlação

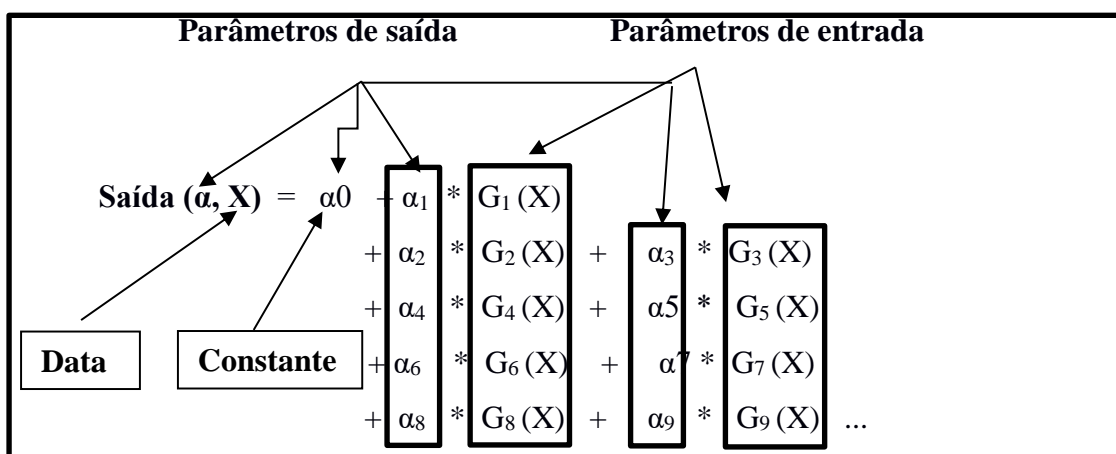
O teste de correlação escolhido (seja paramétrico ou não paramétrico) será aplicado aos parâmetros de entrada deste sistema, pois, como condição para o cálculo da regressão linear múltipla, os parâmetros independentes não podem ser fortemente correlacionados. Os parâmetros dependentes do sistema serão modelados cada qual, de forma individual, não havendo interações entre eles.

Uma vez calculada a correlação entre os parâmetros de entrada, define-se um valor para se considerar fortemente correlacionada. Os parâmetros com tais valores passarão por uma seleção para determinar quais destes serão eliminados e quais permanecerão no modelo.

### 5.5.6 Cálculo das equações de regressão linear múltipla

Como tratado anteriormente, a equação de regressão linear múltipla, se baseia em uma série de parâmetros considerados independentes para prever o resultado de um parâmetro dependente destes. Segue a Figura 23 que exemplifica o método.

Figura 23 – Equação de regressão com variáveis dependentes e independentes



Fonte: Elaborado pelo autor

Onde :

Saída (  $\alpha$ , X ) : representa o registro de saída a ser calculado de um parâmetro  $\alpha$  em um dia X específico com sua confiabilidade .

$\alpha_0$ : constante a ser obtida pela equação de regressão ;

$\alpha_1, \alpha_2, \dots \alpha_n$ : parâmetros de saída obtidos pela regressão ;

$G_1, G_2, \dots G_n$ : parâmetros de entrada.

São diversos os softwares de planilha eletrônica que realizam tal operação (a título de exemplo, Gnumeric, Excel, Libre office Calc).

Com estas equações calculadas, é preciso observar aquelas que possuem uma melhor confiabilidade, sendo estas as que melhor representam o sistema, ou de outro modo, os parâmetros de saída que se comportam mais linearmente baseado nos seus parâmetros de entrada.

### 5.5.7 Correção dos parâmetros de saída ausentes

Dispondo destas equações de regressão, é possível corrigir os parâmetros de saída ausentes. Dado que a equação de regressão prevê o comportamento dos parâmetros de saída, tomando para isto o conhecimento dos valores dos parâmetros de entrada. Nesta

etapa, observando na folha de dados coletados, os dias dos quais os parâmetros de entrada estão completos, porém alguns ou todos, os parâmetros de saída estão ausentes. Podem-se aplicar de maneira direta, as equações de regressão lineares múltiplas calculadas, para prever estes valores e assim completar os dados ausentes nos registros do corpo-técnico.

Apenas para reforçar a confiabilidade das previsões, quando calculado um parâmetro de saída ausente, sugere-se calcular também os outros parâmetros de saída não ausentes. A fim de comparar naqueles conhecidos, o quanto o modelo de regressão se afastou da realidade.

### **5.5.8 Correção dos parâmetros de entrada ausentes**

O aspecto principal desta etapa é completar os parâmetros de entrada ausentes, visto que estes são independentes, logo afetam todos os parâmetros de saída para aquele dia. Uma estratégia inicial para isto seria aplicar a regressão linear múltipla, isolando o parâmetro de entrada, e calculando-o diretamente. Contudo, sua solução deve satisfazer ao mesmo tempo, todas as equações de regressão lineares múltiplas consideradas. Assim, não vem a ser mais um problema de equação de 1º grau, e sim um sistema de equações de 1º grau, e a solução única obtida para este parâmetro ausente, deve completar todas as equações deste sistema.

Para um parâmetro de entrada ausente é possível reescrevê-lo em função deste, de modo que seja possível garantir se existe solução tal que satisfaça todas as equações. Mas na prática, isto se mostrou impossível, dado que as próprias equações de regressão lineares múltiplas calculadas, já possuem graus de confiabilidade menores que 100%, ao exigir do sistema uma única solução para resolver todas as equações de saída para aquele dia, isto se tornava incalculável, ou nas ocasiões que era possível, o resultado sugerido era absurdo (um nítido *outlier*).

Uma estratégia adotada neste procedimento foi o relaxamento do sistema. No qual, sugere-se que todos os parâmetros conhecidos possam ter um pequeno equívoco na sua coleta. Seja este equívoco de natureza humana, como falha na medição ou uma aproximação decimal que desencadeia sempre um erro. Mas também, estes equívocos podem ser de natureza do próprio sistema, como a não homogeneidade do material analisado, levando a uma medição diferente da média real.

Assumindo que estes equívocos possam existir, se aplica a partir do método de otimização linear SIMPLEX (também disponível em diversos softwares de planilha

eletrônica como Excel, Gnumeric e Libre Office Calc), uma busca por soluções daquele sistema de equações lineares, que minimizem as diferenças entre os dados coletados conhecidos e aqueles necessários para que o sistema tenha uma solução. De forma geral, calcula-se o mínimo necessário a se modificar nos dados conhecidos para que o sistema possa ter solução única.

Isto gera uma solução que estima o valor do parâmetro ausente, mas que também propõem quais devam ser os parâmetros até então conhecidos para aquele dia, baseados na sua mínima modificação. A vantagem deste método é seu potencial de aproximar até mesmo os dados coletados com precisão pelo corpo-técnico da ETE, daqueles que verdadeiramente solucionam o sistema de equações lineares. Seu resultado propõe um ajuste em toda a linha no qual este parâmetro ausente esta localizado. Porém assumindo uma precisão de muitas casas decimais a cada parâmetro.

#### **5.5.9 Limitações do método**

O método acima apresentado para completar parâmetros de entrada ausentes possui duas limitações. A primeira é a impossibilidade de que estes novos dados, sejam usados para calcular novas equações de regressão linear múltipla, visto que o resultado obtido é tendencioso às equações de regressão múltiplas originalmente utilizadas, levando assim a equações mais confiáveis para os parâmetros que embasaram o cálculo dos parâmetros ausentes.

A segunda diz respeito à correção de dias nos quais dois ou mais parâmetros de entrada estejam ausentes, pois, com dois graus de liberdade, o sistema buscará reduzir a condição de mínimo das diferenças dos parâmetros conhecidos, variando os dois ou mais parâmetros desconhecidos, avançando assim por uma direção que foge ao interesse de completar o sistema com valores prováveis, ainda que sejam propostos valores possíveis que completem este sistema, não há garantias de que estes tenham sido de fato, valores próximos daqueles que realmente ocorreram, dado que com a liberdade de dois ou mais parâmetros, nada restringe à condição de minimização que o sistema impõe.

## 6. RESULTADOS E DISCUSSÕES

Baseado no estudo-piloto utilizando o método descrito, os principais resultados são apresentados a seguir.

No caso dos parâmetros reunidos, foram eliminados aqueles cuja relação com o processo de tratamento não era relevante ou que possuíam uma distribuição constante (por exemplo, as dimensões do tanque), não afetando assim o estudo sobre a regressão dos demais parâmetros a partir de sua modificação.

Os dados referentes a estas duas colunas foram organizados em uma única página de planilha eletrônica, associando para cada linha as informações sobre os parâmetros físico-químicos e biológicos das coletas realizadas. Percebeu-se inicialmente, que três destes parâmetros (para ambas as etapas) apareciam ocasionalmente, porém em todos os 164 dias de coletas, não havia sequer um registro destas. Também foram encontrados dois parâmetros presentes em todas as coletas, mas cujos dados referentes à saída dos reatores UASB eram inexistentes, enquanto os dados referentes à etapa de gradeamento eram escassos (apenas uma e duas informações registradas).

Dessa forma, optou-se por eliminar do estudo estes cinco parâmetros (referente a ambas as etapas), listados no Quadro 7.

**Quadro 7- Parâmetros desconsiderados inicialmente**

<b>Nº</b>	<b>Parâmetro</b>
1.	Cloro Total
2.	Cloro Livre
3.	DBO Filtrada
4.	Coliformes Totais
5.	Coliformes Fecais

Fonte: Elaborado pelo autor

Dos 52 parâmetros restantes (considerando os parâmetros de cada etapa como diferentes), tem-se as seguintes distribuições de frequências de acordo com o Quadro 8.

**Quadro 8- Quantidades de dados para cada parâmetro**

<b>Parâmetros</b>	<b>Dados coletados</b>
15 parâmetros	164 dados coletados
1 parâmetro	162 dados coletados
2 parâmetros	155 dados coletados
1 parâmetro	147 dados coletados
5 parâmetros	146 dados coletados
2 parâmetros	145 dados coletados
1 parâmetro	117 dados coletados
1 parâmetro	100 dados coletados
2 parâmetros	40 dados coletados
19 parâmetros	39 dados coletados
1 parâmetro	38 dados coletados
2 parâmetros	24 dados coletados

Fonte: Elaborado pelo autor

Prosseguindo com esta escolha, observa-se a existência de vários parâmetros com 145 ou mais dados coletados, e o efeito disto na confiabilidade de um modelo capaz de prever parcialmente o comportamento do Reator UASB, conhecendo os dados na etapa de gradeamento. Assim, foram eliminados do estudo, os 26 parâmetros contendo de 24 a 117 dados coletados, restando para estudo 26 parâmetros, distribuídos nos seguintes 13 pares (parâmetro de gradeamento e parâmetro de saída do reator UASB), conforme o Quadro 9.

**Quadro 9 – Parâmetros (de entrada e saída) com 145 ou mais dados coletados**

#	Parâmetro
<b>1.</b>	Temperatura amostra
<b>2.</b>	Cor Verdadeira
<b>3.</b>	Condutividade
<b>4.</b>	pH
<b>5.</b>	Turbidez

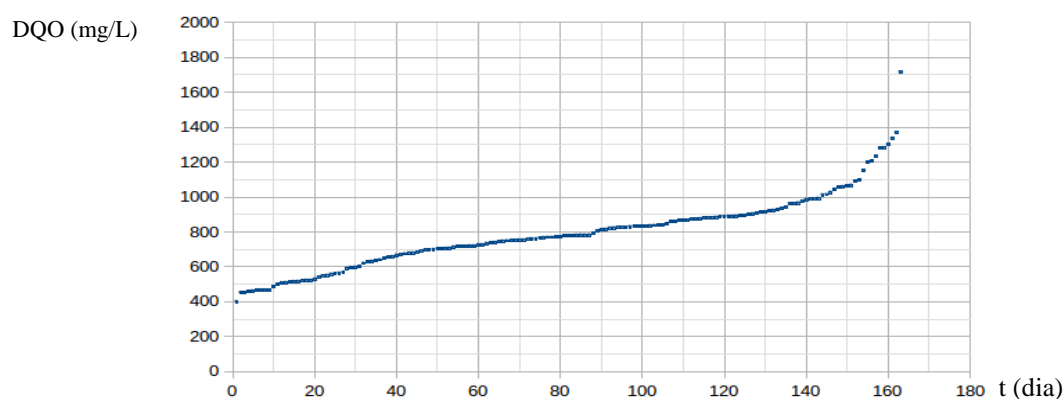
Continua

6.	DBO5
7.	DQO
8.	Oxigênio Dissolvido
9.	Amônia
10.	Nitrogênio Amoniacal
11.	Sólidos sedimentáveis
12.	Sólidos Totais Dissolvidos
13.	Cor aparente

Fonte: Elaborado pelo autor

A título de exemplo, na Figura 24, encontra-se um gráfico com a distribuição pontual do parâmetro DQO ( $\text{mgO}_2/\text{L}$ ) ordenado do menor para o maior. Observa-se que o primeiro registro, próximo de  $400 \text{ mg/L O}_2$ , está um pouco desconexo dos demais. Na perspectiva tomada, ele seria um potencial outlier. Por outro lado, a partir do registro do dia 150, é possível observar pequenos saltos que variam de  $1100 \text{ mg/L O}_2$  a quase  $1400 \text{ mg/L O}_2$ , permitindo a interpretação de que estes fossem dias atípicos, porém não tão distantes de ocorrerem algumas vezes ao longo dos anos que este registro cobre. Mas o último registro, que ultrapassa  $1700 \text{ mg/L O}_2$ , é um *outlier*, ou um fenômeno atípico extremamente raro para este sistema, justificando assim, a remoção dele assim como dos casos atípicos, para o benefício da compreensão do sistema de interesse que neste mesmo gráfico pode ser entendido como o segmento visualmente contínuo da curva.

**Figura 24 – Observação de um *outlier* no período do crescimento da DQO na etapa de entrada (dia x  $\text{mgO}_2/\text{L}$ )**



Fonte: Elaborado pelo autor



Identificou-se que os registros de algumas coletas de parâmetros foram digitados erroneamente pelos funcionários responsáveis, a título de exemplo tem-se os parâmetros: Sólidos Suspensos Totais, coletado da saída do reator UASB no dia 28/03/2017 com a medida de -3276 mg/L; Sólidos Suspensos Voláteis, coletado na etapa de gradeamento no dia 26/02/2018 com a medida de -25 mg/L. Também se identificam alguns possíveis *outliers*.

Considerando a quantidade de possíveis *outliers* identificados, adotou-se como estratégia para prevenir estes erros, e também garantir uma maior integralidade dos dados, a eliminação para todos os parâmetros, dos 3 maiores e 3 menores valores encontrados. Assim, para os casos de possíveis *outliers* identificados, têm-se estes erros eliminados, enquanto outros erros não identificados também podem ser eliminados. Esta redução dos maiores e menores valores, possui um efeito positivo até mesmo para os parâmetros que não apresentam erros, dado que naturalmente possam existir *outliers* que afetem a qualidade das inferências estatísticas.

Estes procedimentos ocorreram com o auxílio do software gratuito Gnumeric, chegando assim nos seguintes resultados para os parâmetros das etapas de gradeamento e da saída dos reatores UASB, no qual os parâmetros marcados com X representam um possível comportamento normal ou Log - normal conforme o Quadro 10.

**Quadro 10 - Parâmetros de acordo com sua distribuição de probabilidade**

Parâmetro	GRADEAMENTO		SAÍDA DO UASB	
	Normal	Log normal	Normal	Log normal
Temperatura				
Cor verdadeira				
Condutividade	X	X		X
pH	X	X		
Turbidez			X	X
DBO				X
DQO		X		X
OD				

Continua

Amônia		X		X
Nitrogênio Amoniacal		X		X
Sólidos Sedimentáveis				
Sólidos Totais Dissolvidos	X	X	X	X
Cor Aparente	X	X	X	

Fonte: Elaborado pelo Autor

Observam-se nesta Tabela os parâmetros cuja distribuição de probabilidade normal ou log-normal se mantiveram em pelo menos uma das duas etapas. Com isto, pode-se perceber que com exceção dos parâmetros pH, Turbidez e DBO, todos os demais possuem pelo menos um candidato à distribuição (Normal ou Log normal) que se mantêm após o processo de tratamento que ocorre nos reatores UASB. Ou seja, aqueles parâmetros sem um desses comportamentos, não adquire um comportamento após o tratamento, enquanto que todos aqueles com estes comportamentos, preservam pelo menos um deles após o tratamento.

A exceção à regra ocorre com o pH, turbidez e DBO, pois o pH na etapa de gradeamento possuiu uma distribuição de probabilidade Normal ou Log normal, enquanto na sua saída do reator UASB, estes comportamentos se alteram, transformando sua distribuição em algo diferente destas duas. Já a turbidez e o DBO, não possuíram nem comportamento normal ou Log normal na entrada (etapa de gradeamento), porém na saída do reator, estes parâmetros se encontram obedientes, possuindo o parâmetro turbidez comportamentos normal e log normal e o parâmetro DBO, comportamento log normal.

Observa-se nessas constatações que as distribuições dos efluentes que têm comportamento log normal, seguem as observações realizadas por Berthouex e Hunter (1983) e Berthouex e Fan (1986) e Metcalf e Eddy (2003) apresentados na Revisão da Literatura.

No caso da prática realizada na ETE CANDEIA, optou-se pelos parâmetros com correlação ( $x$ ), sendo  $|x| \geq 0,7$ , porém este é um valor a ser analisado de maneira particular. A título de exemplo, segue a Tabela 5, no qual se encontra como fortemente correlacionados os parâmetros Nitrogênio Amoniacal com Amônia, e Sólidos Totais Dissolvidos com Condutividade (ambos com correlações mínimas de 0,99).

Segue a Tabela 5 com os valores correlacionados.

Tabela 5 - Coeficientes de correlação entre os parâmetros no gradeamento

Parâmetro	Temperatura	Cor Verdadeira	Condutividade	pH	Turbidez	DBO <sub>5</sub>	DQO	OD	Amônia	Nitrôgenio Amoniacal	Sólido Sedimentável	Sólido Total Dissolvido
Cor Verdadeira	-0,0061											
Condutividade	-0,52	-0,073										
pH	-0,32	-0,47	0,42									
Turbidez	-0,24	0,14	0,021	-0,18								
DBO <sub>5</sub>	0,035	-0,090	0,090	0,0059	0,47							
DQO	-0,16	0,14	0,035	-0,26	0,67	0,50						
Oxigênio Dissolvido	0,11	-0,30	-0,25	0,20	-0,27	-0,21	-0,33					
Amônia	0,036	-0,073	0,0039	0,38	-0,12	0,10	-0,14	0,28				
Nitrogênio Amoniacal	0,029	-0,07	0,045	0,41	-0,12	0,10	-0,14	0,27	0,99			
Sólido Sedimentável	-0,13	0,0066	0,15	-0,058	0,54	0,55	0,46	-0,21	-0,064	-0,081		
Sólido Total Dissolvido	-0,47	-0,08	0,99	0,40	0,0230	0,11	0,04	-0,25	0,0033	0,04	0,12	
Cor Aparente	0,13	-0,033	0,055	0,11	-0,49	-0,23	-0,33	-0,017	-0,10	-0,10	-0,26	0,034

Fonte: Elaborado pelo autor

Uma vez identificados os parâmetros fortemente correlacionados, determinou-se em termos de seu comportamento físico-químico e biológico, qual representa maior importância e independência no sistema.

Para determinar quais parâmetros seriam descartados, analisou-se a quantidade de dados coletados para cada um deles e quais índices são mais importantes para se caracterizar a qualidade de uma água.

Observando que os índices químicos são os mais importantes, e que esta qualidade relaciona-se em função dos íons dissolvidos que são estimados pelo teor de sólidos totais dissolvidos (MACÊDO, 2002), optou-se pelo sólido dissolvido em relação ao parâmetro físico (condutividade).

Nos esgotos domésticos brutos, as formas predominantes são o nitrogênio orgânico e a amônia. Conjuntamente, estes dois parâmetros determinados em laboratórios, constituem o assim denominado Nitrogênio Total Kjeldahl (NTK). As demais formas de nitrogênio são usualmente de menor importância nos esgotos efluentes, a uma estação de tratamento (Macêdo, 2002). Assim, optou-se pela amônia. Neste caso, o estudo ficou com 11 parâmetros na etapa de gradeamento e 13 parâmetros na etapa de saída do reator UASB conforme se pode observar no Quadro 11, dos quais se trata pela nomenclatura G referindo-se ao Gradeamento e R referindo-se à saída do Reator UASB, e permanecendo na numeração abaixo apresentada.

O quadro 11 mostra a classificação desses parâmetros.

**Quadro 11 - Parâmetros classificados por G ou R**

<b>Parâmetros do gradeamento</b>	<b>Parâmetros de saída do reator UASB</b>
G01 - Temperatura amostra	R01 - Temperatura amostra
G02 - Cor Verdadeira	R02 - Cor Verdadeira
G03 - PH	R03 - Condutividade
G04 - Turbidez	R04 - pH
G05 - DBO <sub>5</sub>	R05 - Turbidez
G06 - DQO	R06 - DBO <sub>5</sub>
G07 - Oxigênio Dissolvido	R07 - DQO
G08 - Amônia	R08 - Oxigênio Dissolvido
G 09 - Sólidos sedimentáveis	R09- Amônia

Continua

G10–Sólidos Totais Dissolvidos	R 10 – Nitrogênio Amoniacal
G 11 – Cor Aparente	R11 – Sólidos Sedimentáveis
	R12 – Sólidos Totais Dissolvidos
	R 13 – Cor Aparente

Fonte: Elaborado pelo autor

A partir deste ponto da pesquisa, tomando-se como base o parâmetro DBO<sub>5</sub> cujo valor médio de entrada  $S_0$  (afluente) fornecido foi de 574,12mg/L, seu valor médio de saída  $S$  (efluente) de 148,26 e com desvio padrão de 56,63 mg/L, será determinada a eficiência média esperada pelo reator UASB deste parâmetro através da equação 5 (pág. 24).

$$S = S_0 - \frac{E \times S_0}{100}$$

$$148,26 = 574,12 - \frac{E \times 574,12}{100}$$

Assim, chega-se a uma eficiência estimada de  $E = 74,18\%$ .

Outra maneira de se obter a eficiência de remoção de DBO<sub>5</sub> do reator UASB, é em função do tempo de detenção hidráulica pela equação 4 (pág. 23).

Observando que a temperatura média do esgoto foi de aproximadamente 25 ° C, adota-se conforme o Quadro 1 (pág. 21) o tempo de detenção hidráulica 7 h.

Assim:

$$E_{\text{DBO}_5} = 100 (1 - 0,70 \times t^{-0,50})$$

$$E_{\text{DBO}_5} = 100 (1 - 0,70 \times 7^{-0,50}) = 73,54\%$$

Sendo o erro obtido entre estas equações menor que 1%, se observa uma convergência pela eficiência, na qual será adotada o valor de eficiência de 74% como padrão do efluente na saída do reator UASB.

Este número reflete a eficiência de remoção média de DBO<sub>5</sub> na ordem de 70% preconizada por Von Sperling (1995). Assim, o valor efluente médio corrigido será de:

$$S = S_0 - \frac{E \times S_0}{100}$$

$$S = 574,12 - \frac{0,74 \times 574,12}{100} = 149,27 \text{ mg/L}$$

Admitindo-se que o sistema de tratamento da ETE Candeia pós reator UASB, e da etapa dos biofiltros, segue conforme detalhado em seu processo de tratamento uma eficiência global média de 90% , temos :

$$S_{DBO_5} (global) = S_0 - \frac{E \times S_0}{100}$$

$$S_{DBO_5} (global) = 574,12 - \frac{90 \times 574,12}{100} = 57,41 \text{ mg/L}$$

O valor de remoção global de  $DBO_5$  é inferior ao preconizado pelo decreto nº 8468/76 em seu capítulo 2 (dos padrões), seção II (dos padrões de emissão) em seu artigo 18, que dispõe que “  $DBO_5$  5 dias, 20°C no máximo de 60 mg/l (sessenta miligramas por litro). Este limite somente poderá ser ultrapassado no caso de efluente de sistema de tratamento de águas residuárias que reduza a carga poluidora em termos de  $DBO_5$  5 dias, 20°C do despejo em no mínimo 80% (oitenta por cento)”.

Porém, mesmo com o atendimento normativo de sua eficiência, para melhor compreensão do cálculo da confiabilidade para este parâmetro de saída do reator UASB, será necessário tomar como base, o desvio padrão de 56,63 mg/L do parâmetro  $DBO_5$ , para a determinação do coeficiente de variação e do coeficiente de confiabilidade (CDC).

Assim, para um nível de confiabilidade  $\alpha$  de 95%, foram determinados, os coeficientes de variação (CV), e o coeficiente de confiabilidade (CDC) para o parâmetro de saída  $DBO_5$  (Vide Equações 25 e 26 do item 4.2.1 da pág. 36).

$$CV = \frac{\sigma}{\mu} = \frac{56,63}{149,27} = 0,38$$

$$CDC = (\sqrt{CV^2 + 1}) \cdot e^{(-Z_{1-\alpha} \cdot \sqrt{\ln(CV^2 + 1)})} = 0,58$$

Dessa forma, definida a meta para o padrão de  $DBO_5$  efluente de 149,27 mg/L, sua concentração de projeto será de:

$$\text{Concentração de projeto (CP)} = \text{Meta ou padrão (mg/L)} \times \text{CDC}$$

$$CP = 149,27 \times 0,58 = 87,18 \text{ mg/L.}$$

A tabela 6 detalha a concentração média do efluente  $DBO_5$  de saída do reator UASB para um nível de confiabilidade de 95%.

**Tabela 6- Concentração média efluente de DBO<sub>5</sub> na saída do reator**

Média (mg/L)	CV	1 - $\alpha$	$\alpha$	Z	CDC	CP	Meta na saída DBO <sub>5</sub> (mg/L)
148,26	0,38	95	5	-1,65	0,58	87,18	149,27

Fonte: próprio autor

Como exemplo de interpretação, da Tabela 6, a operação do reator UASB na ETE deveria conduzir ao valor médio de 87,18 mg/L de DBO<sub>5</sub> no efluente do reator UASB, para que, com a variabilidade dos dados, 95% dos dados estivessem abaixo de 149,27 mg/L. Finalmente, resta obter o valor da confiabilidade, ou seja, o percentual dos dados que atendem a esta especificação média de 149,27 mg/L, com 95% de confiança.

### **6.1 Cálculo do percentual esperado de atendimento à meta de lançamento**

Assim, conhecido os valores da concentração efluente e do coeficiente de variação do parâmetro DBO<sub>5</sub> de saída do reator UASB, foi calculado o percentual esperado para atendimento às metas de lançamento.

Observando que o parâmetro DBO<sub>5</sub> segue uma distribuição log-normal, os novos parâmetros para a média e o desvio padrão, foram determinados conforme a equações 27 e 28 da pág.36.

$$\mu_y = \ln\left(\frac{\mu^2}{\sqrt{\mu^2 + \sigma^2}}\right) = \ln\left(\frac{148,26^2}{\sqrt{148,26^2 + 56,63^2}}\right) = 4,94$$

$$\sigma_y = \sqrt{\ln\left(1 + \frac{\sigma^2}{\mu^2}\right)} = \sqrt{\ln\left(1 + \frac{56,63^2}{148,26^2}\right)} = 0,37$$

Desta forma, através da função DIST.LOGNORMAL do Excel, foi possível determinar o percentual de atendimento alcançado, que corresponde a área abaixo da curva fornecida pela função densidade probabilidade (FDP) da distribuição log-normal.

Esta distribuição demonstrada para a meta de 149,27 mg/L é dada por:

$$\text{DIST.LOGNORMAL}(149,27;4,94;0,37) = 57,27\%$$

Neste caso, a interpretação para a DBO<sub>5</sub> é de que apenas 57,27% dos resultados atenderão ao padrão médio de 149,27 mg/L com um nível de confiança de 95%.

A Tabela 7 a seguir demonstra a confiabilidade dos treze parâmetros de saída listados no Quadro 11 de acordo com cada meta ou padrão estabelecido por parâmetro.

**Tabela 7 – Confiabilidade entre os parâmetros de saída do reator na regressão**

	<b>Regressão obtida para</b>	<b>Confiabilidade</b>
<b>R01</b>	<b>Temperatura</b>	<b>95,91%</b>
R02	Cor Verdadeira	77,76%
<b>R03</b>	<b>Condutividade</b>	<b>89,46%</b>
R04	pH	69,79%
R05	Turbidez	40,52%
<u>R06</u>	<i>DBO5</i>	<u>57,27%</u>
R07	DQO	55,72%
R08	Oxigênio Dissolvido	80,01%
R09	Amônia	69,81%
R10	Nitrogênio Amoniacal	69,86%
R11	Sólidos Sedimentáveis	36,22%
<b>R12</b>	<b>Sólidos Totais Dissolvidos</b>	<b>89,36%</b>
R13	Cor aparente	61,31%

Fonte: Elaborado pelo autor

Apesar de 8 dos 13 parâmetros apresentarem confiabilidades próximas ou acima de 70%, como a intenção desta pesquisa é o processo de modelagem dos parâmetros ausentes baseados naqueles conhecidos, foram considerados nesta análise os parâmetros com confiabilidades próximas ou acima de 90%. Dessa forma, foram selecionados 3 dos 13 parâmetros de saída calculados (Temperatura, Condutividade e Sólidos Totais Dissolvidos).

No caso, para os parâmetros de Temperatura (R1), Condutividade (R3), foram calculados, os resultados para os dias em que faltavam estes dados. Com o intuito de comparar os resultados, também se calculou a Temperatura, a Condutividade e os Sólidos Totais Dissolvidos nestes dias, mesmo quando estes não estavam ausentes. Assim, na Tabela 8, estão presentes as diferenças obtidas em porcentagem, dos resultados reais e daqueles gerados pelas previsões da regressão linear múltipla (R1', R3', R12').



**Tabela 8- Diferenças das previsões para a saída do reator**

Data	Temperatura Medida	Temperatura Estimada	Diferença	Condutiv. Medida	Condutiv. Estimada	Diferença	STD Medido	STD Estimado	Diferença
	<b>R1</b>	<b>R1'</b>	<i>Dif %</i>	<b>R3</b>	<b>R3'</b>	<i>Dif %</i>	<b>R12</b>	<b>R12'</b>	<i>Dif %</i>
<b>26/03/18</b>	28,77	28,65	0,42%		<b>712,17</b>		303	<b>348,95</b>	-15,17%
<b>06/03/17</b>		<b>28,32</b>		948	942,17	0,61%	465	<b>464,23</b>	0,17%
<b>09/09/15</b>		<b>23,56</b>		906	954,64	-5,37%	444	<b>466,27</b>	-5,02%
<b>29/05/17</b>	24,4	24,02	1,56%		<b>1141,24</b>		535	<b>554,05</b>	-3,56%

Fonte: Elaborado pelo autor

Assim, tomando-se como base a Tabela 8, construíram-se as Equações (66), (67) e (68) de regressão dos parâmetros R1, R3 e R12 respectivamente.

$$8,79 + (0,93)*G1 + (-0,00)*G2 + (-0,65)*G3 + (-0,00)*G4 + (0,00)*G5 + (-0,00)*G6 + (-0,04)*G7 + (-0,00)*G8 + (-0,03)*G9 + (0,00)*G10 + (-0,00)*G11 = R1 \quad (66)$$

$$873,54 + (1,75)*G1 + (-0,27)*G2 + (-84,42)*G3 + (0,79)*G4 + (-0,11)*G5 + (0,09)*G6 + (-30,73)*G7 + (-0,38)*G8 + (-5,59)*G9 + (1,10)*G10 + (0,01)*G11 = R3 \quad (67)$$

$$480,59 + (0,30)*G1 + (-0,14)*G2 + (-47,80)*G3 + (0,39)*G4 + (-0,05)*G5 + (0,04)*G6 + (-14,20)*G7 + (-0,11)*G8 + (-2,00)*G9 + (0,52)*G10 + (0,05)*G11 = R12 \quad (68)$$

Para deixar mais claro o processo de otimização deste resultado, observa-se no Quadro 12, que no dia 03/01/2017, na qual faltava uma informação referente à variável Oxigênio Dissolvido na etapa de Gradeamento.

**Quadro 12 - Modelo de Otimização Linear para o Oxigênio Dissolvido**

	Parâmetro	Valor	Mínimo		Nova Estimativa		Máximo
<b>Entrada</b>	<b>Oxigênio Dissolvido</b>	-	<b>0,03</b>	≤		≤	<b>2,58</b>
	Temperatura amostra	26,30	0,00	≤		≤	0,00
	Cor Verdadeira	189,00	0,00	≤		≤	0,00
	pH	7,40	0,00	≤		≤	0,00
	Turbidez	276,00	0,00	≤		≤	0,00
	DBO5	884,00	0,00	≤		≤	0,00
	DQO	1281,00	0,00	≤		≤	0,00
	Amônia	75,00	0,00	≤		≤	0,00
	Sól. Sedimentáveis	12,00	0,00	≤		≤	0,00
	Sól. Totais Dissolv.	455,00	0,00	≤		≤	0,00
	Cor Aparente	211,00	0,00	≤		≤	0,00
<b>Saída</b>	Temperatura	27,20	0,00	≤		≤	0,00
	Condutividade	1032,00	0,00	≤		≤	0,00
	Sól. Tot. Dissolv	506,00	0,00	≤		≤	0,00

Fonte: Elaborado pelo autor

Tomando as saídas referentes à Temperatura, Sólidos Totais Dissolvidos e Condutividade, para descobrirmos qual o mínimo de erro pode-se cometer ao assumirmos um valor dentro dos intervalos válidos para Oxigênio Dissolvido.

Como resultado do Programa Simplex, obteve-se que a mínima diferença nas saídas ficou entre 96,01% e 100,32%, dado que 100% correspondem à realidade conhecida para aquele dia.

Assim, o melhor valor que se escolheu para Oxigênio Dissolvido, não satisfará exatamente as 3 equações (Temperatura, Condutividade e STD), chamados respectivamente pelas equações (61), (62) e (63) ao mesmo tempo, mas as satisfará dentro de um erro de até 4,33%. Dessa forma, pode-se dizer que o valor que melhor se encaixa para Oxigênio Dissolvido neste dia é 0,03. Qualquer outro valor para esta variável neste dia, implicaria em um erro maior ou igual à 4,33% para todas as medições deste sistema, assim, como foi considerado somente o erro que é necessário assumir nas medições para o sistema linear ter uma solução, basta encontrarmos um valor que satisfaça a condição de erro mínimo.

O Quadro 13, mostra o resultado da otimização do modelo de otimização linear proposto para este dia.

**Quadro 13 – Resultado de Otimização linear para o parâmetro Oxigênio Dissolvido.**

	<b>Parâmetro</b>	<b>Valor</b>	<b>Mínimo</b>		<b>Nova Estimativa</b>		<b>Máximo</b>
<b>Entrada</b>	Oxigênio Dissolvido	-	0,03	≤	<b>0,03</b>	≤	2,58
	temperatura amostra	26,30	24,50	≤	<b>26,30</b>	≤	26,30
	Cor Verdadeira	189,00	175,77	≤	<b>175,77</b>	≤	189,00
	pH	7,40	6,84	≤	<b>6,84</b>	≤	7,40
	Turbidez	276,00	256,68	≤	<b>276,00</b>	≤	276,00
	DBO5	884,00	822,13	≤	<b>822,13</b>	≤	884,00
	DQO	1281,00	1191,34	≤	<b>1281,00</b>	≤	1281,00
	Amônia	75,00	69,75	≤	<b>69,75</b>	≤	75,00
	Sól. Sedimentáveis	12,00	11,16	≤	<b>11,16</b>	≤	12,00
	Sól. Totais Dissolv.	455,00	423,15	≤	<b>455,00</b>	≤	455,00
	Cor Aparente	211,00	196,23	≤	<b>211,00</b>	≤	211,00
					0,93		1,00
				≤		≤	
<b>Saída</b>	Temperatura	27,20	25,30	≤	<b>26,30</b>	≤	27,20
	Condutividade	1032,00	959,77	≤	<b>959,77</b>	≤	1032,00
	Sól. Tot. Dissolv	506,00	470,58	≤	<b>473,25</b>	≤	506,00

Fonte: Elaborado pelo autor

Este procedimento foi realizado com os 32 dias nos quais apenas um parâmetro de gradeamento era ausente. Na Tabela 9, apresentam-se os dias corrigidos, o parâmetro ausente em cada dia, e o máximo de erro necessário de assumir aos demais parâmetros para esta correção.

**Tabela 9- Lista de parâmetros corrigidos pelo método**

Parâmetro corrigido	Data	Erro
<b>Amônia</b>	02/05/17	2,54%
	08/05/17	1,01%
	22/05/17	9,68%
	12/06/17	1,73%
	03/07/17	5,65%
	24/07/17	2,99%
	02/08/17	2,23%
	23/10/17	7,11%
<b>Cor Aparente</b>	22/06/15	6,79%
	30/06/15	1,62%
	19/02/18	2,11%
<b>DBO<sub>5</sub></b>	15/05/17	7,24%
	27/11/17	1,22%
<b>DQO</b>	19/09/16	2,71%
	28/05/18	10,34%
<b>Oxigênio Dissolvido</b>	11/07/16	5,47%
	03/01/17	2,19%
	30/01/17	1,18%
	17/04/17	7,00%
	24/04/17	2,05%
	28/08/17	3,01%
	18/09/17	7,16%
<b>pH</b>	07/12/15	2,06%
	31/05/16	14,37%
<b>Sólidos Sedimentáveis</b>	24/10/16	0,61%

Continua

Continuação

	15/01/18	3,01%
<b>Sólidos Totais Dissolvidos</b>	02/05/16	1,04%
	12/03/18	3,31%
	02/05/18	0,49%
<b>Temperatura da amostra</b>	05/03/18	5,48%
<b>Turbidez</b>	03/11/15	1,98%
	23/01/17	8,42%

Fonte: Elaborado pelo autor

Para estes erros máximos, obtive uma amplitude de erros em 13,88%, variando do menor dos erros em 0,49% para o maior dos erros em 14,37%. O erro mediano ficou em 2,85% enquanto o erro médio em 4,18%. Com estes 36 novos dias para estudo (32 nos quais faltava um parâmetro de gradeamento, e quatro que faltavam parâmetro de Saída do Reator UASB), calcularam-se novamente as equações de regressão linear múltipla para os 13 parâmetros de saída do Reator UASB. Assim, chegou-se ao resultado da Tabela 10.

**Tabela 10– Confiabilidade da regressão com os dados modificados**

<b>Regressão obtida para</b>	<b>Confiabilidade baseado apenas nos dados originais</b>	<b>Confiabilidade com os novos dados</b>	<b>Diferença</b>
<b>Temperatura</b>	<b>95,91%</b>	<b>95,96%</b>	0,01%
Cor Verdadeira	77,76%	75,59%	-2,80%
<b>Condutividade</b>	<b>89,46%</b>	<b>93,81%</b>	4,35%
pH	69,79%	64,27%	-7,90%
Turbidez	40,52%	32,02%	-8,50%
DBO <sub>5</sub>	58,04%	57,38%	-1,14%
DQO	55,72%	52,85%	-5,15%
Oxigênio Dissolvido	80,01%	78,81%	-1,50%
Amônia	69,81%	66,59%	-4,60%

Nitrogênio amoniacal	69,86%	66,50%	-4,81%
Sólidos Sedimentáveis	36,22%	39,37%	3,15%
<b>Sól. Totais Dissolvidos</b>	<b>89,36%</b>	<b>93,58%</b>	4,22%
Cor aparente	61,31%	59,39%	3,46%

Fonte: Elaborado pelo autor

Como visto na Tabela 10, a confiabilidade se ajusta em conformidade com aqueles utilizados nas equações de regressão linear múltipla (temperatura, condutividade, sólidos totais dissolvidos), melhorando até este índice. A exceção ocorreu como parâmetro Sólido Sedimentável, mas mesmo com a melhora no seu novo ajuste, é pouco interessante em termos de confiabilidade.

Aos demais parâmetros (até por não possuírem uma boa confiabilidade), a nova confiabilidade diminuiu.

## 7. CONCLUSÃO

Este estudo mostrou como conceitos de Matemática aliada a alguns testes estatísticos, podem ser combinados no desenvolvimento de um modelo preditor para o comportamento de alguns parâmetros de ETEs. O estudo-piloto, condensado na forma de relatos de prática ao longo do texto, serviu para validar e amparar este estudo nas necessidades e condições reais em que se dispõe dos dados em relatórios de ETEs.

Reconhecendo que qualquer procedimento de coleta está sujeito a erros, sejam de natureza humana ou mesmo oriundos de suas condições ambientais, este modelo aqui desenvolvido, se caracteriza como capaz de prever resultados pontuais nos dados quando estes são dependentes do restante do sistema e propor modificações gerais a todos os dados quando esses são dependentes do restante do sistema, e expor modificações gerais a todos os dados, quando estes se apresentarem como variáveis independentes na formação do sistema de equações.

Procedimentos como este podem ser aplicados em todas as etapas de tratamento, (ainda que neste estudo tenham sido disponibilizados apenas dados dos parâmetros afluentes e efluentes do reator UASB), desde que se desenvolva para esta, um aprofundamento e validação dentro das condições conhecidas e amparada a partir de um breve referencial teórico sobre a tecnologia de tratamento anaeróbio utilizado. Contudo, para pesquisas futuras, objetiva-se ampliar o domínio deste modelo preditor e corretor, visando à modelagem de todas as etapas de uma ETE e assim, proporcionar um recurso simplificado por meio do qual as equipes responsáveis por instalações desta natureza, possam utilizar no gerenciamento e controle do esgoto tratado.

Destaca-se também que, apesar dos custos reais envolvidos nas etapas de coletas, as quais envolvem desde instrumentos físicos à equipe técnica especializada, a presença de erros é inevitável e incorrigível com a realização de novas coletas. Dessa forma, este modelo preditor e corretor, ainda que limitado a poucas variáveis, porém, foram variáveis que tiveram em sua análise de confiabilidade um percentual bem elevado, ou seja, que estes novos valores corrijam os dados pré-existentes e calculem o valor da variável faltante do dia específico, com um menor erro possível. Importante salientar que este modelo estendido a diferentes etapas do tratamento, aumenta assim os ganhos e velocidade no reconhecimento de falhas e nas suas correções. Visto que no exemplo apresentado na prática, a aplicação deste modelo proporcionou a correção de dados referentes a 22% do total disponibilizado para esta pesquisa.

Esta investigação pode ser diretamente associada à prevenção de tratamentos de esgoto insuficientes, permitindo que as instituições redirecionem seus recursos para os aspectos mais precários e dos quais infiram o maior risco à saúde quando este esgoto não for suficientemente tratado. Assim, se terá economia de tempo, energia, capital investido e recursos na coleta destas informações, que podem ser preditas apenas em fator das outras. Permitindo que o corpo técnico da ETE foque seus esforços em aspectos mais importantes e representativos ao sistema.

Dessa forma, com a presente pesquisa, que desenvolveu a partir de um recurso acessível à maioria das pessoas e viável através de computadores pessoais ou mesmo em alguns aparelhos smartphones, com potencial auxílio para o gerenciamento de ETE e por basear-se em uma estrutura bastante comum no território brasileiro e em ascensão no contexto mundial, que são os reatores do tipo UASB, através dos processos de tratamento anaeróbios, este estudo possui um potencial para extrapolação e reuso em diferentes contextos e locais, proporcionando uma ferramenta auxiliar aos corpos técnicos ou dos órgãos fiscalizadores de ETE, e assim contribuindo para a comunidade na melhora do gerenciamento e consequente qualidade das águas residuais tratadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

ADERIBIGBE, A.A **Term Paper on Monte Carlo Analysis / Simulation**.Department of Electrical and Electronic Engineering.Faculty of Technology.University of Ibadan, 2014.

Ana M. Pires e João A. Branco; **Introdução aos Métodos Estatísticos Robustos**; XV Congresso Anual da Sociedade Portuguesa de Estatística, Lisboa, 19 a 21 de Agosto de 2007; Edições SPE. - www.ine.pt

ABNT – NBR 12209 “**Elaboração de projetos Hidráulico - sanitários de Sistema de Tratamento de Esgotos Sanitários**”, 2009.

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS; R.E. **Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria**. Campinas, SP: Editora da Unicamp, 2001.

BENEDETTI, L.; et al. **Probabilistic modelling and evaluation of wastewater treatment plant upgrades in a water quality based evaluation context**. 2010.

BERTHOUEX, P.M., FAN, R. **Evaluation of treatment plant performance: causes, frequency, and duration of upsets**.*Journal of Water Pollution Control Federation*, v. 58, n. 5, p.368 – 375, 1986.

BERTHOUEX, P.M., HUNTER, W.G. **How to construct reference distributions to evaluate treatment plant effluent quality**.*Journal of Water Pollution Control Federation*, v. 55, n. 12, p. 1417-1424,1983.

BOLFARINE, H. & SANDOVAL, M. C. (2010). **Introdução à inferência estatística**. Coleção Matemática Aplicada, SBM.

BONATE, P. L. **A Brief Introduction to Monte Carlo Simulation**. Clinical Pharmacokinetics, Quintiles, Kansas City, Missouri, 2001.

BRASIL. Lei no 6.938, de 31 de agosto de 1981. **Dispõe sobre a Política Nacional do Meio Ambiente, seus fins e mecanismos de formulação e aplicação, e dá outras providências**.

CAMPOS, J. R. **Tratamento de esgotos sanitários por processo anaeróbio e disposição controlada no solo**. Rio de Janeiro, ABES, 1999.



CAMPOS, J. R. **Uso de outros tipos de reatores anaeróbios para efluentes líquidos.** In: Encontro de Especialistas Em Digestão Anaeróbia. São Paulo, 1985.

CHERNICHARO, C. A. DE L. Reatores anaeróbios. **Princípios do tratamento Biológico de águas residuárias.** DESA/UFMG. Belo Horizonte – MG, v. 5, 1997.

CHERNICHARO, C. A. DE L. Reatores anaeróbios. **Princípios do tratamento Biológico de águas residuárias.** DESA/UFMG. Belo Horizonte – MG, 2007.

CHERNICHARO, C. A. de L. Reatores anaeróbios. **Princípios do tratamento Biológico de águas residuárias.** 2ª ed. ampliada e atualizada. Belo Horizonte: DESA – UFMG, v. 5, 2016.

CONSELHO NACIONAL DO MEIO AMBIENTE - CONAMA. Resolução n.º 430, de 13 de Maio de 2011 - **Dispõe sobre as condições e padrões de lançamento de efluentes.** 2011.

COOK, R. D. AND WEISBERG, S. (1982) ***Residuals and Influence in Regression.*** New York: Chapman and Hall.

DEAN, R. B.; FORSYTHE, S. L. **Estimating the reliability of advanced waste treatment. Part 1 and Part 2, Water & Sewage Works.** 1976.

Decreto 8.468, de 08 de setembro de 1976. **Aprova o Regulamento da Lei nº 997, de 31 de maio de 1976, que dispõe sobre a Prevenção e o Controle da Poluição do Meio Ambiente.** São Paulo, 1976.

FOX, J. ***Applied Regression Analysis, Linear Models, and Related Methods.*** Newbury Park, CA: Sage, 1997.

Galton, F. **Hereditary talents and character** *Macmillan's Magazine.* 12, 157- 166, 318-327, 1985.

Haandel van, A.C., Catunda Paula, F.C., Araújo, Luiz de Souza, 1998. **Biological sludge stabilisation Part 1: kinetics of aerobic sludge digestion.** *Water SA* 24 (3).

Hosmer, D. W. and Lemeshow, S. ***Applied Logistic Regression.*** New York: John Wiley & Sons, Inc. 2000.

LETTINGA G. **Treatment of Sewage under Tropical Conditions in Design of Anaerobic Processes for the Treatment of industrial and Municipal Wastes**, editors J. F. Malina, F.G. Pohland, Technomic Publishers Co., p. 147- 166. 1992.

MACÊDO, J.A.B., **Introdução à Química Ambiental**. (Química & Meio Ambiente & Sociedade) /Jorge Antônio Barros de Macêdo – Juiz de Fora- MG: JORGE MACÊDO, 1ª Edição, 487p, 2002.

MAROCO, J.; **Análise Estatística** – Com utilização do SPSS, 2ª edição; Edições Sílabo; 2003.

MESSAOUD, D.; BACHIR, A.; MAURICE, M. **Daily Performances Evaluation of Municipal Wastewater Treatment Plant using Probabilistic Model**. 2013.

METCALF & EDDY. *Wastewater engineering: treatment, and reuse*. New York: Metcalf & Eddy, Inc., 4 th.Ed, 1819p, 2003.

MONTGOMERY, D. **Estatística aplicada e probabilidade para engenheiros**. 4. ed. Rio de Janeiro, LTC, 2009.

MORETTIN, P. A. **Estatística Básica**. 7. ed. São Paulo: Saraiva, 2011.

NIKU, S., SCHROEDER, E.D., SAMANIEGO F.J. **Performance of activated sludge process and reliability-based design**. *Journal Water Pollution Control Association*, v. 51, n. 12, p. 2841 - 2857, Dec., 1979.

NIKU, S., SCHROEDER, E.D., TCHOBANOGLOUS, G., SAMANIEGO F.J. **Performance of activated sludge process: reliability, stability and variability**. Environmental Protection Agency, EPA Grant No R805097-01, pp. 1 – 124, 1981b.

NIKU, S., SCHROEDER, E.D., TCHOBANOGLOUS, G., SAMANIEGO F.J. **Performance of activated sludge process: reliability, stability and variability**. Environmental Protection Agency, EPA – 600/S2 – 81 – 227., pp. 1 – 11, 1981c.

NIKU, S.; SCHROEDER, E. D. **Stability of Activated Sludge Processes Based on Statistical Measures**. *Journal Water Pollution Control Federation*, Vol. 53, No. 4, pp.: 129-143. 1981.

OLIVEIRA, S. C., VON SPERLING, M. **Reliability analysis of wastewater treatment plants**. *Water Research* 42, pp.: 1182-1194. 2008.

ROBERT C. P.; CASELLA, G. **Monte Carlo Statistical Methods**. [S.l.]: Springer, 1999.

SEIXAS, E. S. **Construção de Árvores de Falhas**. Qualytek - Qualidade, Tecnologia e Sistemas LTDA, 2001.

TAHERIYOUN, M.; MORADINEJAD, S. **Reliability analysis of a wastewater treatment plant using fault tree analysis and Monte Carlo simulation**. 2014.

THOMOPOULOS, N. AND JOHNSON, A, **Tables and Characteristics of the Standardized Lognormal Distribution**. *Proceedings of the Decision Sciences Institute*, n. 103:1-6, 2003.

VON SPERLING, M. **Introdução à qualidade das águas e ao tratamento de esgotos**. Departamento de engenharia Sanitária e Ambiental da UFMG. 1995.

VON SPERLING, M. **Princípios do tratamento biológico de águas residuárias – Volume 1: Introdução à qualidade das águas e ao tratamento de esgotos**. Departamento de Engenharia Sanitária e ambiental – UFMG, 3ªEd. 2005.