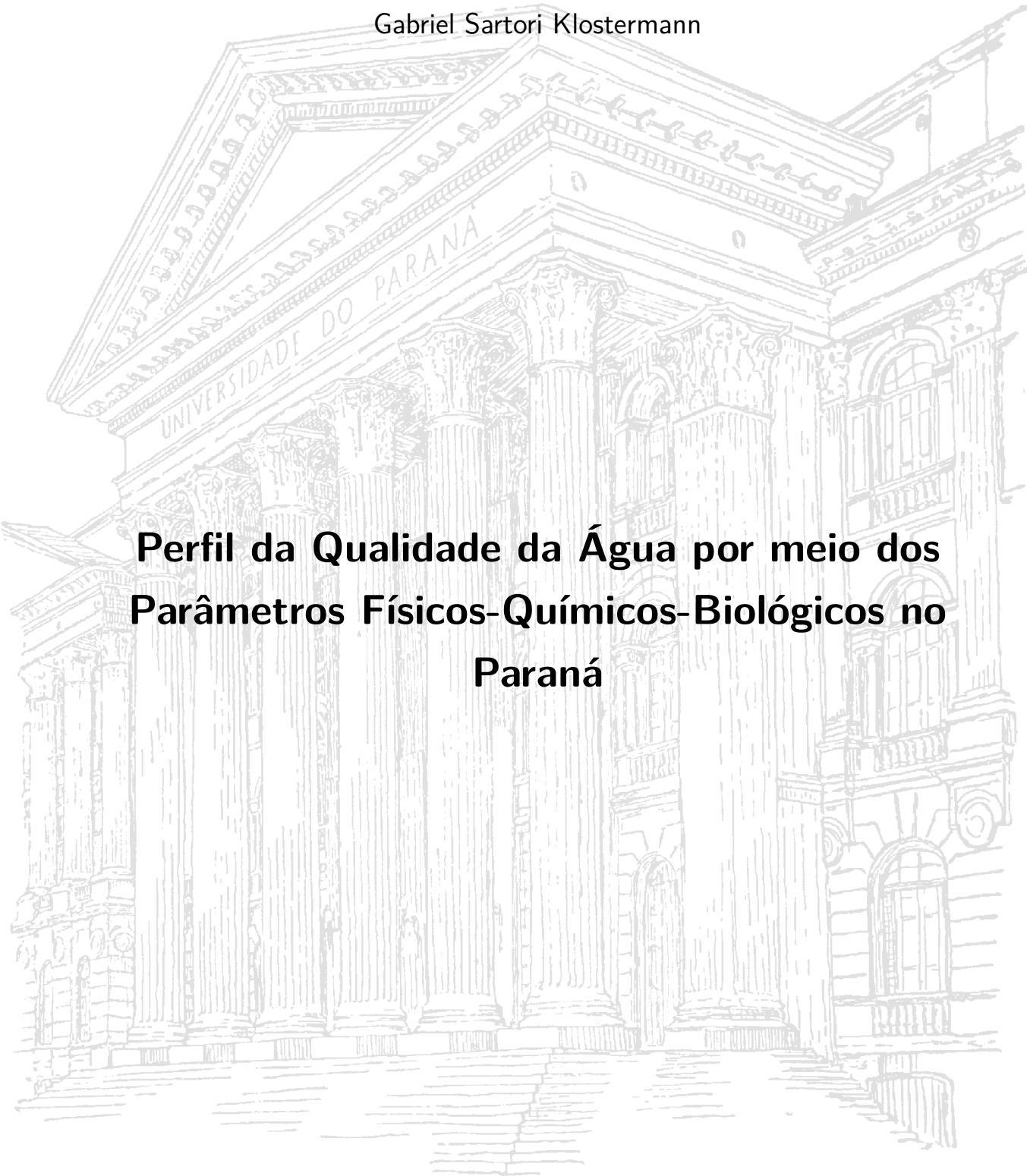


Universidade Federal do Paraná

Gabriel Sartori Klostermann



**Perfil da Qualidade da Água por meio dos  
Parâmetros Físicos-Químicos-Biológicos no  
Paraná**

Curitiba

2018



Gabriel Sartori Klostermann

# **Perfil da Qualidade da Água por meio dos Parâmetros Físicos-Químicos-Biológicos no Paraná**

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório B do Curso de Graduação em Estatística da Universidade Federal do Paraná

Universidade Federal do Paraná  
Setor de Ciências Exatas  
Departamento de Estatística

Orientador: Prof. Dr. Paulo Justiniano Ribeiro Junior

Curitiba  
2018



# Agradecimentos

Agradeço a meus pais pelo apoio à escolha da profissão, sempre me motivando, mesmo com a distância entre Curitiba e Joinville. Agradeço ao meu tio Dietter que sempre quando eu voltava para Joinville discutíamos as análise da qualidade da água e os resultados preliminares, compartilhando toda a sua experiência no estudo do Rio Cachoeira. Agradeço aos meus amigos, Angela, Daniel, Eduardo e Jhenifer, por estarem comigo nos momentos mais divertidos e complicados durante a graduação e principalmente por tornarem o PET entre 2014 a 2015 o período mais divertido e desafiador da graduação, realizando projetos, pesquisas e reuniões, sempre com a orientação da incrível tutora, Silvia Emiko Shimakura. Agradeço ao meu orientador por toda experiência passada, todas as valiosas discussões que tivemos desde os projetos de pesquisa, o próprio TCC, o perfil do profissional de estatística, e sempre com a dose das mais irreverentes analogias e filosofias de análise de dados. Agradeço a todos os meus colegas que me acompanharam na graduação, desde da época do PSE, passando pelas excursões, semana acadêmica, e durante o convívio no campus. Agradeço as conversas com o professor Cristovão do DHS, a Camila Bittencourt, e todo grupo do projeto Integra, que me orientaram para o apelo prático da análise, as indicações de matérias, dúvidas em parâmetros da qualidade da água e compartilharem o panorama das reuniões de recurso hídrico e enquadramento dos corpos d' água. Agradeço ao professores de Estatísticas, ao professor Walmes por ter me ajudado com as dúvidas mais triviais em R e no desenho do framework do MCGLM, ao professor Wagner, por ter me ajudado na construção dos modelos multivariados e dúvidas neste processo se mostrando sempre solícito e ao professor Cesar que sempre se mostrou solícito em ajudar em qualquer análises de dados que eu mostrava. Além disso, todos os professores, especialmente os que estão no Laboratório de Estatística e Geoinformação (LEG). Um ambiente fantástico que a graduação proporciona, com palestras, dúvidas da aula, brincadeiras, e o tão aguardado café das 16h00. Agradeço aos colegas de trabalho no Ministério Público Estadual em especial ao Paulo, meu orientador no meu estágio, por ter me ajudado na correção deste texto e a Maurielle, geógrafa na qual tirou muitas dúvidas sobre o fenômeno. Agradeço à equipe do AGUASPARANÁ por ter se mostrado muito receptivo e interessado no estudo panorâmico da qualidade da água, na apresentação parcial dos resultados, no direcionamento do estudo e dúvidas no banco de dados.



*“As verdades são simples e as mentiras são essenciais”*

— Daniel Goleman

*“A universidade não transforma ninguém num profissional, mas sim, nos oportuniza à se transformar em um profissional”* — Autor Desconhecido

*“O mercado filtra os profissionais”*

— Autor Desconhecido

*“Seus canalhas...”*

— Autor Desconhecido



# Resumo

Este trabalho apresenta um estudo da qualidade da água no Estado do Paraná, por meio de análises de parâmetros físicos-químicos-biológicos. O monitoramento dos parâmetros é realizado pelo instituto AGUASPARANÁ com a abrangência em todo o estado e pelo grupo Integra DHS/UFPR que realiza coletas na região do Alto Iguaçu. Ao todo foram obtidas 8500 coletas em doze bacias hidrográficas em 293 pontos de monitoramentos no período de 1982 a 2016. Além dos parâmetros foi informado na base de dado o índice IQA que é calculado a partir de nove parâmetros. O IQA foi recalculado para comparação com o repassado pelo AGUAS PARANÁ . O resultado motivou a revisão dos parâmetros, que acabou levando à remoção de cerca de 1200 valores e o próprio calculado na amostra. Obtido um banco de dados consistente, foi retomada a análise do IQA no sentido de avaliar quais parâmetros estão contribuindo para a oscilação do índice. Os parâmetros coliformes fecais e turbidez apresentam as maiores variabilidades. Além disto, foram feitas análises descritivas para compreensão das coletas segmentadas por bacia hidrográfica, a fim de verificar padrões ao longo do tempo, como crescimento, decaimento ou constante. Ainda com o intuito de identificar diferenças e similaridades entre as bacias hidrográficas foram aplicadas as técnicas de Análise de Componentes Principais (ACP) e *t-Distributed Stochastic Neighbor Embedding* (t-SNE). Ambas as técnicas foram aplicadas nas diversas escalas dos parâmetros, concentração, carga e transformação IQA. Nas escalas concentração e transformação IQA os dois primeiros componentes não explicam 70% da variabilidade para a maioria das bacias, diferentemente, na escala carga houve explicação de quase 100%. A técnica t-SNE não apresentou resultados satisfatórios em nenhuma das escalas. Uma possível explicação é o fato de que as técnicas não incorporam as dependências entre rios, pontos monitorados e a própria assimetria da variável coletada. Além disso, o trabalho também contribuiu com a discussão da autodepuração do rio. Modelos foram aplicados no contexto as técnicas do *framework Multivariate Covariance Generalized Linear Model* (MCGLM) que incorpora múltiplas respostas e possibilita combinar matrizes de dependência espacial, temporal e trajetória de coletas. O rio Barigui foi utilizado como estudo de caso pela existência de cinco pontos de monitoramento e razoáveis coletas ao longo do período. Como resultado, um modelo foi selecionado por critérios estatísticos e apontou que não é mantida a mesma qualidade da água ao longo dos pontos de monitoramento. Para finalizar todas as análises foram feitas com softwares livres e gratuitos, portanto, os resultados podem ser reproduzidos em outras bacias hidrográfica e rios.

**Palavras-chave:** Autodepuração; Bacias Hidrográficas; Índice de Qualidade da Água; *Multivariate Covariance Generalized Linear Model*; Paraná; Qualidade da Água; Rio Barigui;

*t-Distributed Stochastic Neighbor Embedding*

# **Lista de ilustrações**

Figura 1 – Curvas empíricas da qualidade da água . . . . .	28
Figura 2 – Processo de Autodepuração . . . . .	32
Figura 3 – Diferentes características de coletas ao longo do tempo . . . . .	35
Figura 4 – Estrutura do MCGLM . . . . .	39
Figura 5 – Dependência espacial entre monitoramento . . . . .	40
Figura 6 – Dependência temporal por monitoramento . . . . .	40
Figura 7 – Dependência da ordem das coletas ao longo do tempo . . . . .	41
Figura 8 – Matrizes de dependência . . . . .	42
Figura 9 – Frequência de coleta por ano de cada equipe . . . . .	45
Figura 10 – Quantidade de coletas nas estações de monitoramento por bacias . . . . .	46
Figura 11 – Comparação do IQA calculado pelo AGUASPARANÁ e pelo autor . . . . .	47
Figura 12 – IQA ao longo do tempo por Bacia . . . . .	49
Figura 13 – Boxplot dos Pesos dos Parâmetros por Bacia . . . . .	49
Figura 14 – IQA calculado sem CF e TURB . . . . .	50
Figura 15 – Relação IQA e enquadramento por Bacia . . . . .	51
Figura 16 – Relação OD por classe de Rio . . . . .	51
Figura 17 – Concentração do OD ao longo do tempo por Bacia . . . . .	52
Figura 18 – Tonelas por dias de OD por Bacia . . . . .	53
Figura 19 – Boxplot das épocas do ano por bacia . . . . .	53
Figura 20 – Parâmetros da qualidade da água ao longo do tempo . . . . .	55
Figura 21 – Comparação das fontes da qualidade da água . . . . .	55
Figura 22 – PCA nas transformações do IQA na bacia do Iguaçu e Litorânea . . . . .	56
Figura 23 – Scores PCA . . . . .	57
Figura 24 – Score - t-SNE . . . . .	58
Figura 25 – Análise Descritiva no Parâmetro OD . . . . .	60
Figura 26 – Convergência Modelo . . . . .	63
Figura 27 – Avaliação dos Pressupostos do Modelo . . . . .	63
Figura 28 – Bandas de Predição para OD . . . . .	64
Figura 29 – Bandas de Predição para DBO . . . . .	65
Figura 30 – Correlação das respostas . . . . .	65
Figura 31 – Curvas médias de variação dos parâmetros de qualidade das águas para o cálculo do IQA . . . . .	78



# **Lista de tabelas**

Tabela 1 – Exemplo Cálculo IQA . . . . .	29
Tabela 2 – Quantidades de dados por bacia . . . . .	46
Tabela 3 – Quantidade de Dados Inválidos . . . . .	48
Tabela 4 – Sumarização Alto Iguaçu por Fonte . . . . .	54
Tabela 5 – Nº de Coletas por Monitoramento . . . . .	59
Tabela 6 – Funções de transformação para o IQA . . . . .	76



# **Lista de quadros**

Quadro 1 – Pesos dos parâmetros do IQA . . . . .	28
Quadro 2 – Relação entre IQA, qualidade da água e classe do rio . . . . .	29
Quadro 3 – Limites definidos por parâmetros em cada classe do rio . . . . .	31
Quadro 4 – Parâmetros da Água e das condições da coleta . . . . .	34
Quadro 5 – Matrizes Significativas por Resposta . . . . .	62



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>2</b>	<b>OBJETIVOS</b>	<b>21</b>
<b>2.1</b>	<b>Objetivo Geral</b>	<b>21</b>
<b>2.2</b>	<b>Objetivos Específicos</b>	<b>21</b>
<b>3</b>	<b>REVISÃO DA LITERATURA</b>	<b>23</b>
<b>3.1</b>	<b>Parâmetros da Qualidade da Água</b>	<b>23</b>
3.1.1	Oxigênio Dissolvido (OD)	23
3.1.2	Coliformes Totais (CT)	23
3.1.3	Coliformes Fecais (CF)	23
3.1.4	Potencial Hidrogeônico (pH)	24
3.1.5	Demanda Bioquímica de Oxigênio (DBO)	24
3.1.6	Fósforo Total (FT)	24
3.1.7	Nitrogênio Total	25
3.1.8	Sólidos Totais (ST)	25
3.1.9	Temperatura da Água (TA)	25
3.1.10	Turbidez (Turb)	26
3.1.11	Condutividade (Condut)	26
3.1.12	Demanda Química de Oxigênio (DQO)	26
3.1.13	Vazão	26
<b>3.2</b>	<b>Índice de Qualidade da Água (IQA)</b>	<b>27</b>
<b>3.3</b>	<b>Enquadramento de Corpos d'Água</b>	<b>29</b>
<b>3.4</b>	<b>Autodepuração do Rio</b>	<b>31</b>
<b>4</b>	<b>MATERIAL E MÉTODOS</b>	<b>33</b>
<b>4.1</b>	<b>Procedimento Metodológico</b>	<b>33</b>
<b>4.2</b>	<b>Banco de Dados</b>	<b>34</b>
<b>4.3</b>	<b>Métodos</b>	<b>36</b>
4.3.1	Análise Descritiva	36
4.3.2	Redução de Dimensionalidades	36
4.3.2.1	ACP	37
4.3.2.2	t-SNE	37
4.3.3	Modelo Multivariado	38
<b>4.4</b>	<b>Recursos Computacionais</b>	<b>43</b>

<b>5</b>	<b>RESULTADOS E DISCUSSÕES . . . . .</b>	<b>45</b>
<b>5.1</b>	<b>Análise Descritiva . . . . .</b>	<b>45</b>
5.1.1	Característica dos Dados . . . . .	45
5.1.2	Montagem do Banco de Dados . . . . .	47
5.1.3	Contribuição dos Parâmetros . . . . .	48
5.1.4	Relação Enquadramento com IQA e Limites CONAMA . . . . .	50
5.1.5	Parâmetros por Bacia Hidrográfica . . . . .	52
5.1.6	Comparação entre Bacias . . . . .	52
5.1.7	Alto Iguaçu . . . . .	54
<b>5.2</b>	<b>Redução de Dimensionalidade . . . . .</b>	<b>56</b>
5.2.1	Identificação de diferenças e similaridades entre bacias e intra estações . . .	56
<b>5.3</b>	<b>Modelo Multivariado . . . . .</b>	<b>59</b>
5.3.1	Rio Autodepurável . . . . .	59
<b>6</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>67</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>69</b>
	<b>APÊNDICES</b>	<b>73</b>
	<b>APÊNDICE A – EQUAÇÕES IQA . . . . .</b>	<b>75</b>
	<b>APÊNDICE B – CURVAS DE DESEJABILIDADE DO IQA . . . . .</b>	<b>77</b>

# 1 Introdução

A Lei nº 9.433 ([BRASIL, 1997](#)) instituiu a Política Nacional de Recursos Hídricos, criando o Sistema Nacional de Gerenciamento de Recursos Hídricos (SINGREH). O primeiro artigo estabelece os fundamentos desta política, determinando que a água é um recurso natural limitado, um bem de domínio público, dotado de valor econômico.

A referida lei também estabelece que a gestão da bacia hidrográfica deve ser descentralizada, visando o uso múltiplo das águas. Ao mesmo tempo, é o espaço geográfico de atuação onde os Comitês de Bacias Hidrográficas (CBHs) buscam promover o planejamento regional, controlar os usos da água na região, proteger e conservar as fontes de captação da bacia.

Uma bacia hidrográfica pode ser caracterizada por um conjunto de terras drenadas por um rio e seus afluentes, formada nas regiões mais altas do relevo por divisores de água, onde as águas das chuvas escoam superficialmente formando os riachos e rios. As águas de superfície escoam nas regiões inferiores do terreno, sendo que as cabeceiras são formadas por riachos que brotam em terrenos íngremes das serras e montanhas. À medida que as águas dos riachos descem, juntam-se a outros riachos, aumentando o volume e formando os primeiros rios. Esses pequenos rios continuam seus trajetos recebendo água de outros tributários, formando rios maiores até desembocar no oceano ([TEODORO et al., 2015](#)).

As atividades humanas requerem usos múltiplos dos recursos hídricos. Essas atividades podem gerar impactos e deterioração da qualidade da água, assim como interferir na quantidade de água disponível ([VEIGA, 2011](#)). A contaminação dos mananciais de abastecimento público, por rejeitos oriundos das atividades humanas tem sido um dos maiores fatores de risco para a saúde humana, especialmente nas regiões com condições inadequadas de saneamento básico ([GRUNITZKI et al., 2013](#)).

De acordo [Guedes et al. \(2009\)](#), um recurso hídrico que tenha recebido uma descarga de material orgânico passa por um processo natural de recuperação, conhecido como autodepuração, podendo ocorrer através de processos físicos, químicos e bioquímicos. Conforme a carga de poluentes aumenta, de modo que faça com que a capacidade de autodepuração do corpo hídrico seja ultrapassada, o mesmo tende a tornar-se poluído.

O comprometimento da qualidade da água pela contaminação por esgotos domésticos, muitas vezes lançados no ambiente sem tratamento prévio, implica, entre outras consequências, o aumento da incidência de doenças de veiculação hídrica, como cólera, diarreia, amebíase e esquistossomose. Essa situação assume proporções mais

graves em países ou regiões onde é maior a pobreza. Nos países em desenvolvimento, 90% das doenças infecciosas são transmitidas através da água ([FREITAS; SANTOS, 1999](#)).

Desta forma, se tornam importantes ações de monitoramento da qualidade das águas como ferramentas para o manejo adequado dos recursos hídricos. Além disso, o monitoramento de qualidade de água constitui um alerta para as instituições de ensino, pesquisa e extensão, autoridades públicas e a população sobre significativos índices de poluição dos recursos naturais ([GONÇALVES et al., 2005](#)).

O monitoramento por meio de parâmetros de qualidade da água ou indicadores de qualidade de água é uma tentativa que todo programa de monitoramento prevê como forma de acompanhar, por meio dessas informações a possível deterioração ou poluição de um corpo hídrico ([GILBERTO et al., 2002](#)).

Para tanto, se fez, neste trabalho, uso do Índice de Qualidade das Águas (IQA) desenvolvido pela National Sanitation Foundation (NSF) em 1970 no Estados Unidos e utilizado no Brasil pela Companhia de Tecnologia de Saneamento Ambiental (CETESB), desde 1975, em uma versão adaptada da versão original da NSF.

No Paraná, a coleta dos parâmetros utilizados na composição do IQA e de outros parâmetros na água, é realizada por equipes da Superintendência de Desenvolvimento de Recursos Hídricos e Saneamento Ambiental (SUDERHSA), orgão vinculado ao instituto AGUASPARANÁ, e as análises laboratoriais são realizadas em um laboratório do Instituto Ambiental do Paraná (IAP).

Os dados coletados e analisados no laboratório são armazenados no Sistema de Informações Hidrológicas e o AGUASPARANA utiliza para calcular o IQA. Na UFPR, também são feitas coletas da qualidade da águas e analisadas laboratorialmente pelo grupo Integra na região do Alto Iguaçu desde 2005.

Dessa forma, a aquisição do banco de dados do estudo foi viabilizado pela união dois programas de monitoramento, passando por validações, transformações na escala nos parâmetros, categorização de variáveis contínuas e recálculo do IQA com a contribuição de cada parâmetro.

Neste trabalho, objetiva-se estudar a evolução dos parâmetros coletados e o próprio IQA por meio de séries históricas, métodos multivariados e modelo multivariado.

Dessa forma, as técnicas elencadas vão permitir auxiliar na discussão da qualidade da água no Paraná, quais parâmetros variaram ao longo do tempo e quais estão contribuindo mais acentuadamente no IQA. Além disso, se os rios enquadrados estão de acordo com a qualidade indicada pela resolução 357 do Conselho Nacional do Meio Ambiente (CONAMA) de 2005. Investiga-se ainda, se há diferenças e similaridades nas estações de monitoramento de mesma bacia hidrográfica e contribuir com a discussão

da autodepuração de rios.

As análises de dados são extensas e optou-se por apresentar aqui no texto apenas uma parte dos resultados. Diversas análises complementares são disponibilizadas em <http://www.leg.ufpr.br/~paulojus/iqa/> e ao longo do texto que se segue são indicados os links relacionados com o tópico em discussão.

Espera-se que os resultados apresentados e discutidos sejam capazes de fornecer subsídios técnicos consistentes, a ponto de reavaliar as discussões vigentes nos comitês de bacias hidrográficas, e também produzir relatórios expondo a situação atual e histórica da qualidade da água.



## 2 Objetivos

### 2.1 Objetivo Geral

O objetivo geral do trabalho é qualificar a discussão da qualidade da água no Estado do Paraná, via série histórica, nas bacias e rios monitorados e apontar métodos estatísticos que lidem com as características do fenômeno.

### 2.2 Objetivos Específicos

- Montar um banco de dados em qualidade da água;
- analisar a contribuição dos parâmetros que compõem o IQA;
- verificar por bacia o atendimento às normas da CONAMA e a relação do IQA com a classe do rio;
- explorar descritivamente os parâmetros da qualidade da água, nas diferentes escalas, comparando as bacias e na região do Alto Iguaçu;
- identificar diferenças e similaridades por meio dos parâmetro entre as bacias e internamente nas estações de monitoramento;
- analisar a variação dos parâmetros ao longo do monitoramento do rio Barigui acondendo estruturas de dependência espacial, temporal e ordenação das coletas.



## 3 Revisão da Literatura

### 3.1 Parâmetros da Qualidade da Água

#### 3.1.1 Oxigênio Dissolvido (OD)

A concentração de OD na água é essencial para os ciclos de vida de peixes, organismos aquáticos e para o funcionamento adequado dos ecossistemas. Os níveis de OD indicam a saúde desses ecossistemas, uma vez que o oxigênio está envolvido em praticamente todos os processos químicos e biológicos.

O déficit extremo de OD pode levar ao que ocorrerá uma mortandade de peixes nos rios, por asfixia, em condições anaeróbicas, caracterizadas por uma coloração escura e gases odoríferos, que ocasionam incômodos às populações ribeirinhas ([ÁGUAS-ANA, 2017](#)).

As principais fontes de OD, em água, são a renovação natural proveniente da atmosfera e a fotossíntese das plantas aquáticas, inclusive dos fitoplânctons. Já as principais causas da diminuição do OD são, a respiração das plantas, a Demanda Bioquímica de Oxigênio (DBO), originada da decomposição de matérias orgânicas e sedimentos e remoção de águas saturadas de oxigênio e íons inorgânicos.

#### 3.1.2 Coliformes Totais (CT)

O grupo coliformes consiste de vários gêneros de bactérias pertencentes à família Enterobacteriaceae. A determinação de coliformes fecais inclui também as espécies que não vivem necessariamente nos intestinos humanos, podendo ser encontrados no solo e nos vegetais([AGUASPARANA, 2011](#)).

#### 3.1.3 Coliformes Fecais (CF)

Coliformes fecais, ou coliformes termotolerantes, são bactérias que estão presentes em grande quantidade no intestino de animais de sangue quente. Sua presença num curso d'água pode estar relacionada com o lançamento de esgotos domésticos. Um grande número de microrganismos patogênicos – que transmitem doenças como hepatite, cólera e disenterias, entre outras – também pode estar presente nos esgotos. A *Escherichia coli*, é a principal bactéria do grupo de CF, sendo abundante nas fezes humanas e de animais, sendo bastante importante a inspeção sanitária da bacias, para estabelecer a origem da contaminação ([SPERLING, 2007](#)).

### 3.1.4 Potencial Hidrogeônico (pH)

O pH é um parâmetro definido no intervalo entre 0 a 14. O pH neutro (igual a 7) não indica necessariamente uma pureza da água, mas um equilíbrio entre substâncias ácidas e alcalinas, e alterações no pH da água podem afetar a fauna e flora aquática. Valores reduzidos, podem ser indicativos de presença de efluentes industriais assim como valores elevados, podem estar associados à proliferação de algas ([SPERLING, 2007](#)).

### 3.1.5 Demanda Bioquímica de Oxigênio (DBO)

A DBO indica a quantidade de oxigênio consumido nos processos biológicos de degradação da matéria orgânica no meio aquático. É, portanto, um indicador das cargas orgânicas nos corpos hídricos. Cargas orgânicas provenientes de esgotos domésticos não tratados têm forte influência no aumento da DBO, sobretudo em rios e córregos de pequeno porte e com capacidade limitada de autodepuração ([ÁGUAS-\(ANA\), 2017](#)).

Os esgotos domésticos e certos despejos industriais (matadouros, laticínios, curtumes, usinas de açúcar e álcool, etc.) são ricos em matéria orgânica e seu lançamento nos cursos d'água pode causar problemas à fauna aquática, devido à queda do OD e como consequência mortandades de peixes ([AGUASPARANA, 2011](#)).

### 3.1.6 Fósforo Total (FT)

O fósforo é um nutriente essencial para todas as formas de vida, pois é parte de suas estruturas celulares. Alguns efluentes industriais como os de indústrias de fertilizantes, pesticidas, químicas em geral, conservas alimentícias, abatedouros, frigoríficos e laticínios apresentam fósforo em quantidades excessivas, assim como as águas drenadas de áreas agrícolas e urbanas.

Nos ambientes urbanos, a concentração de fósforo na água indica principalmente a poluição por efluentes domésticos e industriais. No campo, as concentrações de fósforo estão geralmente associadas à entrada de sedimentos e nutrientes com origem em processos erosivos pelo manejo inadequado do solo e fertilizantes.

O fósforo tende a se acumular e causar problemas em corpos hídricos lênticos, tais como lagos e reservatórios. O crescimento de plantas e algas é normalmente limitado pelas concentrações de fósforo em condições naturais. O excesso deste nutriente pode ocasionar o crescimento excessivo da flora aquática e o desequilíbrio dos ecossistemas, por meio da eutrofização ([SPERLING, 2007](#)).

Denomina-se eutrofização o fenômeno causado pelo excesso de nutrientes (nitrogênio e, principalmente, fósforo) num corpo d'água, o que leva à proliferação excessiva

de algas e outras plantas aquáticas, que, ao entrarem em decomposição, ocasionam o aumento do número de microrganismos e a consequente diminuição de OD, degradando a qualidade da água. A eutrofização ocorre com mais intensidade em lagos e reservatórios do que em rios, pelas condições ambientais para o crescimento de algas serem mais favoráveis, como turbidez e velocidades baixas ([AGUASPARANA, 2011](#)).

### 3.1.7 Nitrogênio Total

O nitrogênio total é a soma de quatro espécies, que constituem a chamada série nitrogenada: o amoniacal, orgânico, nitrito e nitrato. A série nitrogenada permite identificar a idade da poluição.

A ocorrência de níveis elevados de nitrato é atribuída à falta de esgotamento sanitário em áreas urbanas, e à aplicação de fertilizantes nitrogenados, em áreas rurais. Os nitratos apresentam alta mobilidade na água subterrânea e podem contaminar extensas áreas ([SPERLING, 2007](#)).

### 3.1.8 Sólidos Totais (ST)

Os ST correspondem ao material que permanece como resíduo após a evaporação e secagem da coleta a uma temperatura escolhida até obter uma massa constante. A classificação dos sólidos pode ser química ou física. Fisicamente eles são classificados segundo suas dimensões: sólidos dissolvidos, com dimensões inferiores a 2,0, e sólidos em suspensão, com dimensões superiores a esta.

Do ponto de vista químico, os sólidos são classificados em fixos ou voláteis. Sólidos voláteis são os que volatilizam a temperaturas inferiores a 550 °C, tanto substâncias orgânicas quanto sais minerais. Os sólidos fixos são aqueles que permanecem após a completa evaporação da água, geralmente os sais ([SÃO, 2009](#)).

O excesso de sólidos dissolvidos na água pode causar alterações no sabor e problemas de corrosão. Já os sólidos em suspensão provocam a turbidez da água gerando problemas estéticos e prejudicando a atividade fotossintética. O teor de sólidos dissolvidos representa a quantidade de substâncias dissolvidas na água, geralmente sais minerais. Os componentes primários que contribuem para a contagem de sólidos totais dissolvidos são sais de cálcio, magnésio, sódio, cloro, bicarbonato e enxofre ([SÃO, 2009](#)).

### 3.1.9 Temperatura da Água (TA)

Temperatura é a medida de intensidade de calor. Ela influencia todos os processos físico-químicos e biológicos da água. O lançamento de efluentes industriais com temperatura elevada, como água de refrigeração, provoca uma poluição térmica dos

rios, podendo afetar a fauna e flora aquáticas, pois provoca a diminuição da concentração de OD e pode aumentar a toxicidade de muitas substâncias ([AGUASPARANA, 2011](#)).

### 3.1.10 Turbidez (Turb)

A turbidez é causada pela presença de partículas em suspensão na água. Ela é medida pela dificuldade da luz em atravessar a água. A erosão das margens dos rios, despejos de esgotos domésticos e industriais provocam elevações na turbidez das águas. A alta turbidez dificulta a fotossíntese, portanto a produção de oxigênio, da vegetação e das algas presentes na água, prejudicando a vida dos organismos aquáticos. Além disso, afeta negativamente os usos doméstico, industrial e recreacional dos corpos d'água ([AGUASPARANA, 2011](#)).

A turbidez reflete a interferência de materiais em suspensão na passagem da luz através da água. É, portanto, um bom indicador da quantidade de sólidos em suspensão e, consequentemente, de processos erosivos na bacia hidrográfica. Nas cidades o aumento da turbidez pode refletir despejos domésticos e industriais de fontes pontuais, muitas vezes associadas à infraestrutura de drenagem, e também à poluição de origem difusa ([AGUASPARANA, 2011](#)).

### 3.1.11 Condutividade (Condut)

A condutividade representa a capacidade de condução de energia elétrica pela água a 25°C. É originada da presença de sais dissolvidos na água na forma de ions. Os ions podem ter origem antropogênica (descargas industriais e de esgoto doméstico) ou geônicas (decomposição de rochas). Altas cargas de sais na água têm efeitos negativos, podendo afetar as comunidades aquáticas ([AGUASPARANA, 2011](#)).

### 3.1.12 Demanda Química de Oxigênio (DQO)

A DQO é uma medida de quantidade de oxigênio necessário para a oxidação da matéria orgânica através de um oxidante químico (dicromato de potássio) tanto a matéria biodegradável quanto a não biodegradável por isso os valores da DQO são maiores que a DBO. O aumento da concentração de DQO num corpo d'água se deve principalmente a despejos de origem industrial.

### 3.1.13 Vazão

Para uma interpretação adequada das condições de um corpo hídrico os aspectos qualitativo e quantitivos da água devem ser trabalhados juntos. Variações na concentração de um poluente pontual (quando os efluentes atingem o corpo d'água de

forma concentrada no espaço), como também de poluição difusa (quando os poluentes adentram o corpo d'água distribuídos ao longo do tempo).

Somente quando se conhece a vazão é possível diferenciar o tipo de poluição. A vazão permite ainda calcular o fluxo em massa de poluentes numa seção de rio. As medidas de vazões coletadas são vazões instantâneas, e na falta destas, vazões médias diárias determinadas com base nos dados fluviométricos das estações de monitoramento([AGUASPARANA, 2011](#)).

## 3.2 Índice de Qualidade da Água (IQA)

A operacionalização de uma rede de monitoramento de qualidade das águas origina uma grande quantidade de dados analíticos, sendo difícil sua compilação e apresentação de uma maneira sintética e de fácil compreensão para o público geral. Isso levou ao desenvolvimento, em vários países, de índices que permitem expressar a qualidade da água de uma maneira global, sintetizando os resultados dos diversos parâmetros analisados.

Um desses índices é o IQA, desenvolvido pela National Sanitation Foundation do Estados Unidos da América. A definição dos parâmetros que compõem o IQA se baseou numa pesquisa de opinião feita entre 142 especialistas, os quais indicaram os parâmetros que deveriam ser medidos, bem como sua importância. Os parâmetros finais que foram incorporados são (OD, CF, DBO, pH, Temp, NT, FT, Turb e ST) ([AGUASPARANA, 2011](#)).

O IQA é determinado pelo produtório ponderado das qualidades estabelecidas para os nove parâmetros que o compõem, conforme a equação 3.1 podendo variar de 0 a 100 pontos:

$$IQA = \prod_{i=1}^9 q_i^{w_i} \quad (3.1)$$

- $i = 1, 2, 3, \dots, 9$  identifica o i-ésimo parâmetro;
- $q_i$  = valor da transformação do i-ésimo parâmetro;
- $w_i$  = peso relativo do i-ésimo parâmetro (em que  $\sum_{i=1}^9 w_i = 1$ ).

O quadro 1 apresenta os pesos ( $w_i$ ) para cada parâmetro, conforme definição do IQA.

Quadro 1 – Pesos dos parâmetros do IQA

Parâmetros IQA	Pesos ( $w_i$ )
% OD	0.17
CF	0.15
pH	0.12
DBO	0.10
FT	0.10
NT	0.10
$\Delta$ Temp	0.10
ST	0.08
Turb	0.08

Fonte: Portal da Qualidade das Águas e AGUASPARANÁ

A Figura 1 mostra para cada parâmetro, o comportamento na respectiva transformação ( $q_i$ ) para o IQA. Para calcular estas curvas foram utilizadas as funções descritas no Apêndice A, obtidos na referência obtidos na referência [Sperling \(2007\)](#).

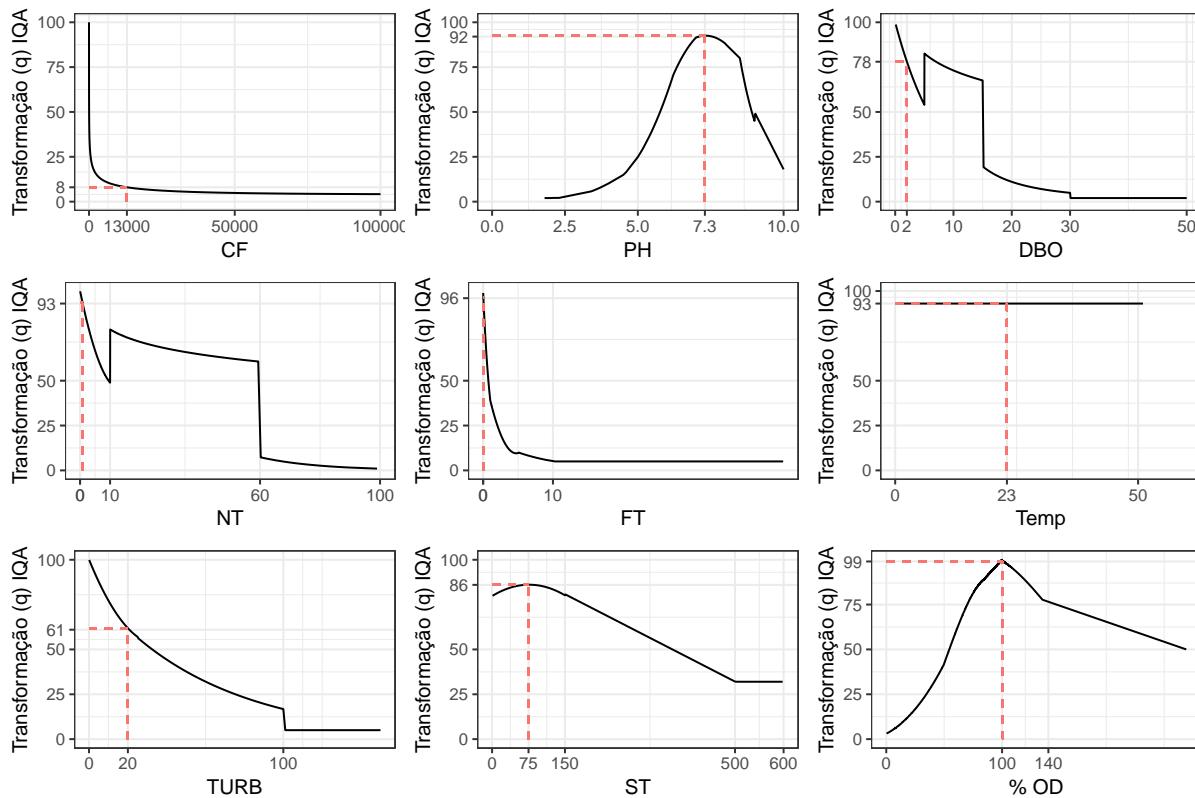


Figura 1 – Curvas empíricas da qualidade da água

A fim de comparação, as curvas teóricas do IQA estão no Apêndice B apresentam o comportamento ideal da conversão do parâmetro para a escala do IQA. Nota-se divergência com as curvas da Figura 31, principalmente para os parâmetros NT e DBO.

Os demais não apresentam o mesmo crescimento ou decaimento suave. De qualquer forma optou-se por recalcular o IQA.

O cálculo está exemplificado para uma amostra a partir das linhas pontilhadas que representam os valores indicado na coluna "Concentração" da Tabela 1.

Tabela 1 – Exemplo Cálculo IQA

Parâmetro	Concentração	qi	wi	$q_i^{wi}$
% OD	100.80	99.17	0.17	2.18
CF	13000.00	8.02	0.15	1.37
pH	7.30	92.52	0.12	1.72
DBO	2.00	78.12	0.10	1.55
NT	0.80	93.66	0.10	1.57
FT	0.00	96.40	0.10	1.58
Temp	23.00	93.00	0.10	1.57
Turb	20.00	61.87	0.08	1.39
ST	75.00	86.08	0.08	1.43
IQA				61.00

O resultado da amostra pode ser categorizado conforme o quadro 2 e ainda mais, pode ser relacionado com o nível de qualidade de água dado a classe do rio. No caso, os parâmetros coletados foram de um rio enquadrado de classe 2, e portanto era de se esperar o IQA entre 70 a 90.

Quadro 2 – Relação entre IQA, qualidade da água e classe do rio

Valor IQA	Qualidade da água	Classe do Rio
91 - 100	Ótima	1
71 - 90	Boa	2
51 - 70	Regular	3
26 - 50	Ruim	4
0 - 25	Péssima	Fora de Classe

Fonte: Portal da Qualidade das Águas e AGUASPARANÁ

### 3.3 Enquadramento de Corpos d'Água

O enquadramento de águas doces, enfoque do presente estudo, é citado de acordo com o Artigo 4º da Resolução CONAMA 357/2005:

As águas doces são classificadas em:

I - classe especial: águas destinadas:

- a) ao abastecimento para consumo humano, com desinfecção;
- b) à preservação do equilíbrio natural das comunidades aquáticas; e,

c) à preservação dos ambientes aquáticos em unidades de conservação de proteção integral.

II - classe 1: águas que podem ser destinadas:

a) ao abastecimento para consumo humano, após tratamento simplificado;

b) à proteção das comunidades aquáticas;

c) à recreação de contato primário, tais como natação, esqui aquático e mergulho, conforme Resolução CONAMA no 274, de 2000;

d) à irrigação de hortaliças que são consumidas cruas e de frutas que se desenvolvam rentes ao solo e que sejam ingeridas cruas sem remoção de película; e

e) à proteção das comunidades aquáticas em Terras Indígenas.

III - classe 2: águas que podem ser destinadas:

a) ao abastecimento para consumo humano, após tratamento convencional;

b) à proteção das comunidades aquáticas;

c) à recreação de contato primário, tais como natação, esqui aquático e mergulho, conforme Resolução CONAMA no 274, de 2000;

d) à irrigação de hortaliças, plantas frutíferas e de parques, jardins, campos de esporte e lazer, com os quais o público possa vir a ter contato direto; e

e) à aquicultura e à atividade de pesca.

IV - classe 3: águas que podem ser destinadas:

a) ao abastecimento para consumo humano, após tratamento convencional ou avançado;

b) à irrigação de culturas arbóreas, cerealíferas e forrageiras;

c) à pesca amadora;

d) à recreação de contato secundário; e

e) à dessedentação de animais.

V - classe 4: águas que podem ser destinadas:

a) à navegação; e

b) à harmonia paisagística.

Na lei 9.433 de 1997, no artigo 9º estabelece os objetivos para enquadramento:

I - assegurar às águas qualidade compatível com os usos mais exigentes a que forem destinadas;

II - diminuir os custos de combate à poluição das águas, mediante ações preventivas permanentes.

A referida resolução da CONAMA dispõe dos limites esperados para os parâmetros da qualidade da água de acordo com suas respectivas classes de rios, conforme exposto no Quadro 3.

Quadro 3 – Limites definidos por parâmetros em cada classe do rio

Parâmetro	UM	Classe 1	Classe 2	Classe 3
OD	mg/L	>6	>5	>4
CF	NMP/100ml	<200	<400	<1000
DBO	mg/L	<3	<5	<10
ST	mg/L	<500	<500	<500
CT	NMP/100ml	<1000	<5000	<20000
pH		[6 - 9]	[6 - 9]	[6- 9]

### 3.4 Autodepuração do Rio

Um dos objetivos do trabalho é aplicar uma metodologia que contribua com o debate da autodepuração do rio. Sendo assim, a temática é de suma importância conforme disposto no trabalho de ([ANDRADE, 2018](#)):

A proteção dos recursos hídricos envolve o monitoramento e a avaliação de sua qualidade ao longo dos rios, estabelecendo metas e controlando as descargas de poluentes, de forma que um nível aceitável de qualidade da água seja mantido. A compreensão da autodepuração dos corpos d' água constitui em elemento básico para a adoção destas medidas e ações.

Do texto de ([ZSCHORNACK, 2016 apud SPERLING, 2005](#)) o processo de autodepuração pode ser caracterizado na citação e na Figura 2.

O processo de autodepuração se desenvolve ao longo do tempo e da direção longitudinal do curso d'água, por isso, segundo von Sperling (2005), é possível identificar as fases das zonas de autodepuração, divididas em: zona de degradação, zona de decomposição ativa, zona de recuperação e zona de águas limpas.

- a) Zona de águas limpas - localizada em região à montante do lançamento do efluente (caso não exista poluição anterior) e também após a zona de recuperação. Essa região é caracterizada pela elevada concentração de oxigênio dissolvido e vida aquática superior;
- b) Zona de degradação - localizada à jusante do ponto de lançamento, sendo caracterizada por uma diminuição inicial na concentração de oxigênio dissolvido e presença de organismos mais resistentes;
- c) Zona de decomposição ativa - região onde a concentração de oxigênio dissolvido atinge o valor mínimo e a vida aquática é predominada por bactérias e fungos (anaeróbicos);
- d) Zona de recuperação - região onde se inicia a etapa de restabelecimento do equilíbrio anterior à poluição, com presença de vida aquática superior.

# Processo de Autodepuração

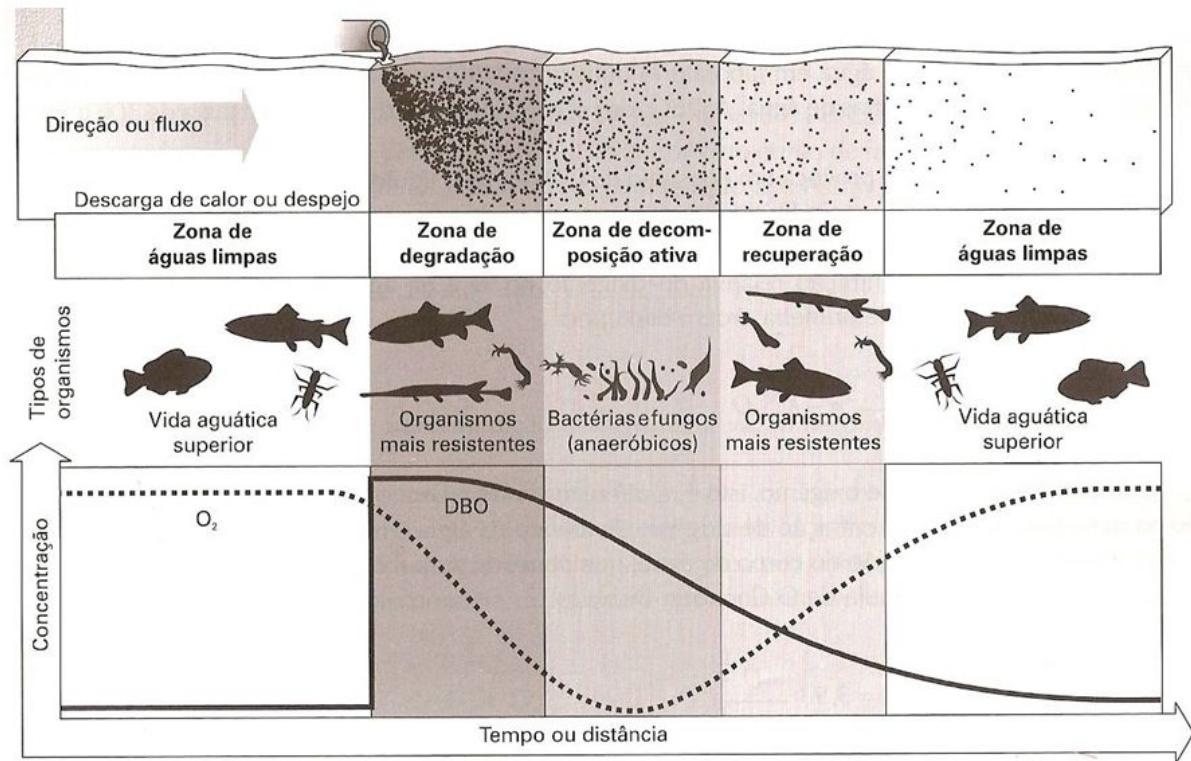


Figura 2 – Processo de Autodepuração

## 4 Material e Métodos

### 4.1 Procedimento Metodológico

O contato inicial com os dados foi durante o estágio no Ministério Público do Paraná em que foi produzido um estudo descritivo da qualidade da água, focada na visualização temporal dos parâmetros para cada estação de monitoramento.

Por se tratar de um estudo de monitoramento da qualidade da água, optou-se por primeiramente analisar o IQA, já que resumia nove parâmetros e sua interpretação é mais simples e direta. O IQA recebido, foi calculado pelo instituto que enviou os dados, porém sem os devidos pesos por parâmetros. Portanto, iniciou-se um processo de recalcular o IQA, a fim de comparar as fontes.

O resultado desta comparação levou à revisão dos parâmetros da qualidade da água e consequentemente, ocorreu a remoção das amostras para quais os valores dos parâmetros estavam fora dos limites de aceitabilidade e o descarte do IQA.

Obtido um banco de dados consistente, foi retomada a análise de verificação de qual parâmetro que compõe o IQA que mais estava contribuindo para o seu decréscimo ou crescimento.

Além disso, o procedimento de análise de dados, também foi desenvolvido descritivamente para comparações dos parâmetros na escala de concentração, carga e peso do IQA entre bacias e diferenças entre as coletas do AGUASPARANÁ e do DHS/UFPR.

Na literatura da área, além das análises descritivas, é muito comum utilizar técnicas como componentes principais para reduzir a dimensionalidade dos parâmetros e investigar a associação dos componentes criados conforme trabalhos de ([CANDIDO, 2015; TRINDADE, 2013; ARRUDA, 2014](#)).

A aplicação desta técnica no presente estudo, não se mostrou eficiente na maioria das bacias, pois a variação explicada não atingiu 70% utilizando os dois primeiros componentes. Como alternativa foi aplicada a técnica *t-SNE*, já que é um método não linear, mas que também não distinguiu as variáveis categóricas com os parâmetros de qualidade da água.

Outra discussão da qualidade da água, é saber o grau de autodepuração do rio, nesse sentido, foi criado um modelo multivariado para o Rio Barigui, dos parâmetros que compõem o IQA, para ajudar a discutir a capacidade de restauração do rio.

## 4.2 Banco de Dados

Os dados dos parâmetros da qualidade da água, foram disponibilizados pelo instituto AGUASPARANÁ ao Ministério Público Estadual no setor de Núcleo de Inteligência, via correio eletrônico em setembro de 2016. A partir da organização dos dados, foram identificadas 7961 coletas em 269 estações de monitoramento no estado do Paraná, datada entre março de 1982 a setembro de 2013. Ressalta-se, que há dados mais recentes a partir de outubro de 2016, no [portal de informações hidrológicas](#), mas não foram incluídas.

A ausência de dados entre 2013 a 2016 reflete a deficiência tecnológica e financeira da instituição, segundo os servidores consultados. O banco de dados original contém as variáveis identificadoras: nome da bacia, rio, classe do rio, município, estação de monitoramento, data, condições do tempo (chuva, nublado e bom), latitude e longitude. No Quadro 4 estão os parâmetros de qualidade da água. Adotou-se o símbolo  $\infty$  para representar que não há um limite superior definido.

Quadro 4 – Parâmetros da Água e das condições da coleta

Variáveis	Unidade de Medida	Amplitude
OD	mg/L	[0 - 9.2]
CF	NMP/100mL	[0 - $\infty$ )
pH		[0-14]
DBO	mg/L	[0 - $\infty$ )
NT	mg/L	[0 - $\infty$ )
FT	mg/L P	[0 - $\infty$ )
Turb	NTU	[0 - $\infty$ )
ST	mg/L	[0 - $\infty$ )
Temp	°C	[0 - $\infty$ )
IQA		[0 - 100]
CT	NMP/100mL	[0 - $\infty$ )
DQO	mg/L	[0 - $\infty$ )
Cond	$\mu$ /cm	[0 - $\infty$ )
NK	mg/L N	[0 - $\infty$ )
TA	°C	[0 - $\infty$ )
Vazões	m <sup>3</sup> /s	[0 - $\infty$ )

A partir dessas variáveis, foram adicionadas novas variáveis relativos à época (verão, outono, inverno, primavera) e unidades de gestão, que reconfigura, as bacias hidrológicas naturais, obtido em [AGUASPARANÁ](#). O uso do solo nas estações de monitoramento foram atribuídas por meio do estudo realizado em 2002, disponível em [ITCG](#) e repetido para todos os anos.

Além do mais, foram criadas novas variáveis a partir dos parâmetros originais, de acordo com os limites da resolução CONAMA que delimita as concentrações dos

parâmetros para os respectivos enquadramentos dos rios. Também foram criadas, novas variáveis, após conversão das concentrações que estão na unidade de medida (mg/L) para Toneladas por dia (T/dia), recebendo a terminologia de carga. A conversão desta variável está exemplificada abaixo:

$$Carga = \frac{\left( \left( \frac{\text{parâmetro}}{1000} * \text{vazão} \right) * 60 * 60 * 24 \right)}{1000} \quad (4.1)$$

Além disso, foi recalculado o IQA utilizando a metodologia descrita na seção 3.2. Foram extraídos respectivos pesos nas curvas de referências fazendo com que todas os parâmetros que compõem o IQA assumam valores entre 0 a 100.

Por outro lado, foram adicionadas novas coletas com as amostra do DHS que realizam monitoramento independente ao AGUASPARANÁ na região do Alto Iguaçu desde 2005, com periodicidade recorrente. Desta forma, somando os dados do AGUASPARANÁ e do DHS, o banco de dados possui ao todo 8506 coletas.

Em um primeiro momento, verificou-se que a quantidades de coletas varia por estações de monitoramento. Identificou-se que o número de coletas por estação de monitoramento é diferente. Na Figura 3 exemplifica-se quatro estações de monitoramento distintas em número de coletas.

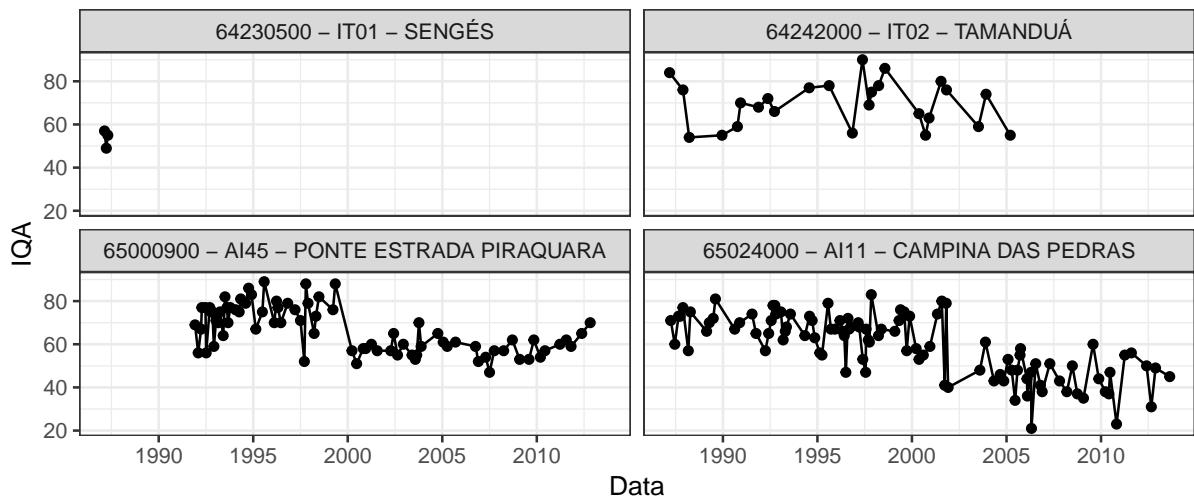


Figura 3 – Diferentes características de coletas ao longo do tempo

A Figura 3 apresenta estações que foram desativadas com poucas coletas ou num período prolongado e das estações com registros recentes, foram instaladas em épocas diferentes. Entretanto, no banco de dados não há registro da data de desativação das estações de monitoramento.

## 4.3 Métodos

O trabalho desenvolvido preocupou-se em passar as ideias principais dos métodos utilizados desde a análise descritiva, polinômios locais, redução de dimensionalidade e modelos multivariados. Nas subseções subsequentes serão indicados ao leitor os artigos originais para o interesse do detalhamento das metodologias. Desta forma, o foco do texto é apresentar os resultados obtidos com tais metodologias.

### 4.3.1 Análise Descritiva

Para investigar as relações de interesse, num primeiro momento, foram feitas análises exploratórias por meio de gráficos temporais dos parâmetros, em escala da concentração, carga e convertido pela transformação do IQA. Ainda mais, por meio de gráficos de barras, foi visualizado as classificações dos parâmetros, de acordo com a resolução da CONAMA para cada classe de rio.

No que se refere ao cruzamento de dados contínuos com categóricos, foram utilizados *boxplots* para os resultados da época do ano (verão, outono, inverno e primavera) e gráficos de dispersão para visualizar a transição dos parâmetros ao longo do tempo, conforme orientações [Bussab e Morettin \(2010\)](#). Para o objetivo do atendimento dos limites da CONAMA e a relação do IQA com a classe de rio foi usado o gráfico de barras.

Nos gráficos em que as abscissas se referem à datas, foram adicionadas linhas suaves que capturaram a variação dos dados. Esta técnica baseia-se em ajustes não paramétricos, já que não requisitam nenhum pressuposto prévio, caracterizando como métodos exploratórios. De forma geral, essas técnicas são utilizados para se ter um ideia inicial do comportamento e possíveis padrões nos dados segundo [\(JAMES et al., 2013\)](#).

Para construção dessas linhas suaves foi usada a técnica de polinômios locais, em que ajusta regressões locais em janelas, e é robusta a *outliers*. Para detalhes do método está descrito em [Faraway \(2016\)](#).

### 4.3.2 Redução de Dimensionalidades

Técnicas de redução de dimensionalidade permitem converterem uma alta dimensão de dados em um plano de visualização 2D ou 3D que mantenha as principais características dos dados originais. Além de reduzir o conjunto de dados é possível atribuir legendas por meio de variáveis categóricas para identificação de padrões.

#### 4.3.2.1 ACP

A ACP permite explicar a estrutura de variância e covariância de p-variáveis aleatórias originais, em menos de p componentes, para se ter uma redução daquelas variáveis. Estes componentes principais são não correlacionados entre si, assim a informação contida nas variáveis originais pode ser substituída por estes componentes (MINGOTI, 2013).

A qualidade da análise, ou dos  $q < p$  componentes na explicação da estrutura de variância e covariância pode ser verificada por meio da avaliação da proporção de variância total explicada. Os parâmetros da qualidade da água foram transformados multivariadamente, para que o conjunto de dados adequasse na distribuição normal multivariada, conforme (BOX; COX, 1964). A segunda transformação consistiu na padronização da variável, pelas respectivas médias e desvios padrões, gerando novas variáveis, centradas em 0 e com variância 1 (MINGOTI, 2013).

Uma vez determinados os componentes é possível construir *scores*, que podem ser calculados para cada coleta. Estes scores podem ser reproduzidos em gráficos 2D para atribuição de legendas a fim de encontrar padrões.

Em contrapartida, o método de componentes principais apresenta resultados mais condizentes com o fenômeno, quando os dados possuem normalidade e são independentes (MANLY, 2008).

Foi aplicado o ACP por cada bacia em todas as escalas de qualidade da água: concentração, carga e transformação IQA.

#### 4.3.2.2 t-SNE

O t-SNE é um algoritmo de redução de dimensionalidade não linear, que herda as ideias de redução de dimensionalidade do ACP porém pretende ser mais flexível. No estudo foram aplicados em todos os dados sem divisão por bacia hidrográfica. Diferentemente do ACP o único pressuposto é que as variáveis devem ser numérica para utilização.

A metodologia do cálculo do *t-SNE* se encontra no artigo original de Maaten e Hinton (2008), mas de maneira sintética e exemplificada neste [link](#).

O algoritmo é controlado pelo parâmetro de perplexidade, que acomoda o comportamento global do hiperespaço para o local, que na prática é a quantidade de vizinhos mais próximos entre os pontos usados para calcular distância. Segundo o próprio autor, o parâmetro deve ser fixada entre 5 a 50 e não pode ser maior ou próximo da quantidade de dados no estudo. Ao passo que, a projeção de um novo plano, são atribuídas legendas na nuvem de pontos para identificação de similaridades ou diferenças.

Como desvantagem, a distância das nuvens de pontos entre os grupos não implica na intensidade da diferença.

#### 4.3.3 Modelo Multivariado

Com o objetivo de discutir a sustentabilidade do rio Barigui, foi utilizado a classe modelos desenvolvido por Bonat e Jørgensen (2016), denominado de *multivariate covariance generalized linear models* (MCGLM). Segundo Bonat e Jørgensen (2016), cuja a principal característica do modelo é permitir acomodar respostas multivariadas, tanto para respostas de contagem, positivos reais, assimétricos, com diferentes funções de ligação, e com diferentes estruturas de covariância, como medidas repetidas, longitudinal, temporal, espacial e espaço temporal.

Neste modelo os efeitos  $\beta$  das covariáveis são estimados pela função *quasi-score* e os parâmetros de dispersão  $\tau$  e  $\rho$  por meio da função de estimação de Pearson, que são descritos em detalhes na seção 3.2 do artigo original. A inferência baseia-se na distribuição assintótica dos estimadores.

A Figura 4 apresenta a formulação do modelo MCGLM. O texto em azul refere-se à dimensão das matrizes para cada componente da fórmula. No tópico (Parâmetros - Estimação)  $\beta$  e  $\tau$  estão em negrito para diferenciar os vetores dos escalares das respostas. O texto em vermelho é o indexador dos conteúdos para auxiliar na explicação do texto abaixo.

A seguir os conceitos serão indicados pelas partes da Figura 4. A construção do modelo para os efeito de médias é composta por modelos para cada resposta (1), permitindo especificar as mesmas funções de ligação usualmente utilizadas nos modelos lineares generalizados (MLG) (2). A especificação do modelo se completa com a definição da estrutura de covariância para todas as respostas que é especificada pela fórmula (3). A estrutura é composta pela matriz bloco diagonal das covariâncias de cada resposta e a matriz de correlação entre as respostas, indicado na expressão (4).

A construção da matriz de covariância para cada resposta é feita conforme a expressão (5), dada pela função de variância e a função de covariância.

A função de variância, expressa a relação entre a média e variância sendo determinada conforme a natureza do dado e relacionado no conjunto (6). Como caso particular na fórmula  $u^p$  assume-se distribuições indicadas no conjunto(7).

A função de covariância dos dados é determinada pela expressão (8). Esta função é uma combinação linear de matrizes que representam a dependência entre os dados para cada resposta.

Em cada matriz (8) há o escalar  $\tau$  que quantifica se a respectiva matriz é signi-

ficativa para a explicação da variabilidade. Essas matrizes (9) podem ser criadas pelo usuário ou já pré estabelecidas na biblioteca mcglm disponível [CRAN](#). Neste estudo as matrizes de dependência ( $Z_D$ ) foram criadas conforme as características do fenômeno.

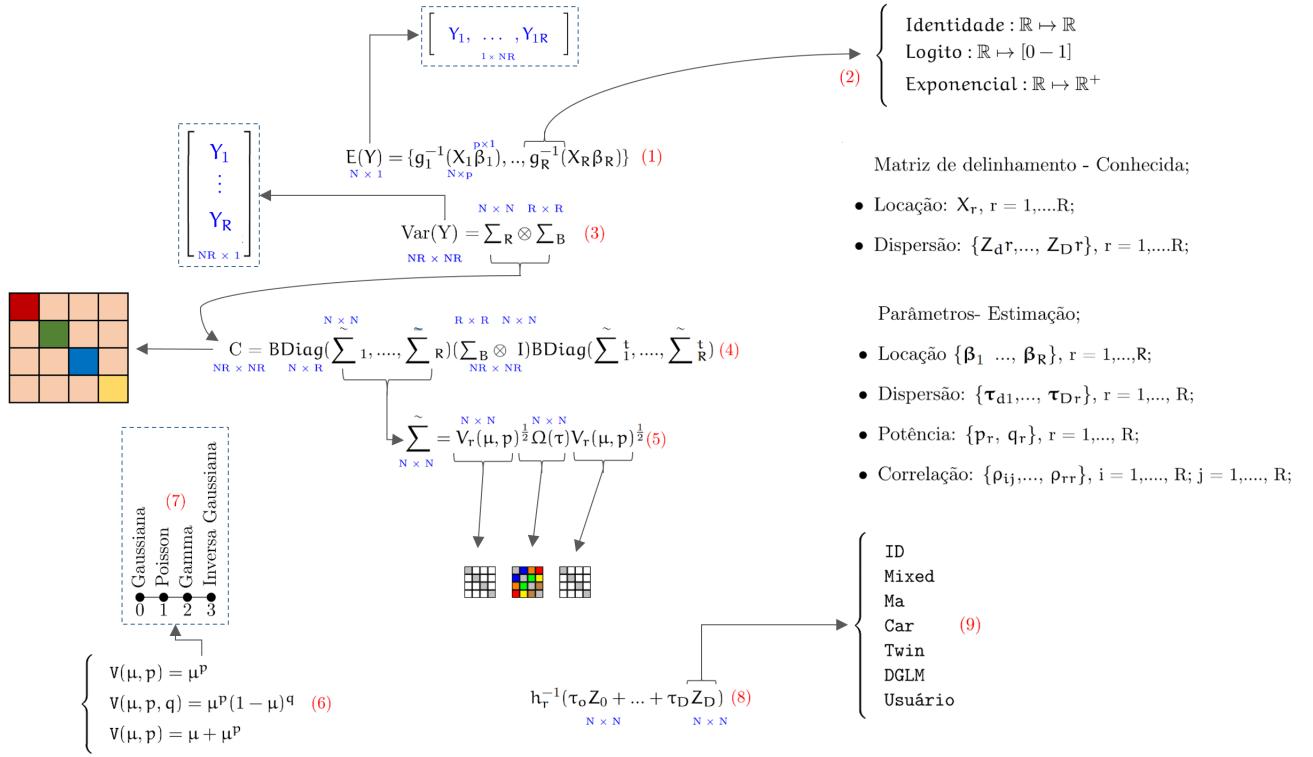


Figura 4 – Estrutura do MCGLM

O framework de MCGLM é flexível a ponto de englobar diferentes modelos, com diferentes combinações de covariáveis para efeitos de médias e com diferentes estruturas de dependência entre os dados. Para que se possa ajustar o modelo é necessário: os dados, indicando respostas e covariáveis, matrizes de dependência para cada resposta, com as respectivas as funções de ligação, para variância e covariância.

Em contrapartida, a única pressuposição do modelo é estabelecer adequadamente a relação entre média e variância. Por outro lado, esta relação pode ser estimada pelo modelo. Para verificar a adequação do modelo, somente é necessário verificar a relação da média com a variância por meio dos resíduos de *pearson* com os valores ajustados.

Considerando o fenômeno em estudo pode-se pensar que há dependência espacial, temporal e de trajetória ao longo do rio. Neste sentido, a dependência desta verificação pode ser ilustrada conforme as Figuras 5, 6 e 7 que exemplificam as ideias fundamentais em duas estações de monitoramento.

Inicialmente, é natural pensar que existe influência espacial entre as estações de

monitoramento conforme indicado na Figura 5. Por se tratar de uma série temporal, pode-se pensar que cada estação de monitoramento à medida que acontecem as coletas são dependentes da anterior, dessa forma ilustradas na Figura 6. Ao mesmo tempo, a dependência temporal pode estar associada pela diferença em dias entre a coleta anterior.

Pode-se pensar que a dependência está associada ao rio como um todo, independente da estações de monitoramento. Deste modo, a Figura 7, representa a vizinhança das coletas conforme a trajetória ao longo do tempo independente da estação de monitoramento. Também será representado pela diferença em dias.

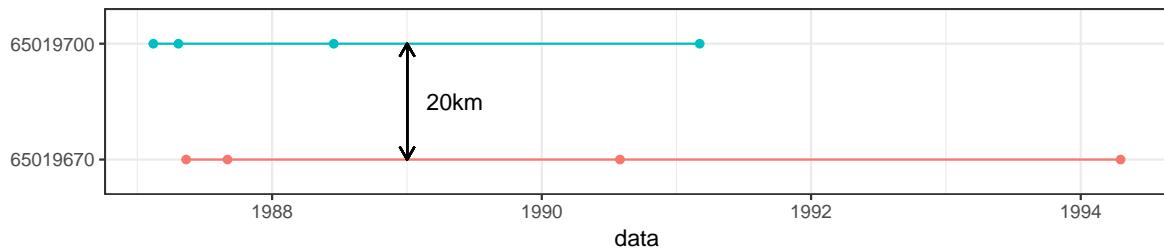


Figura 5 – Dependência espacial entre monitoramento

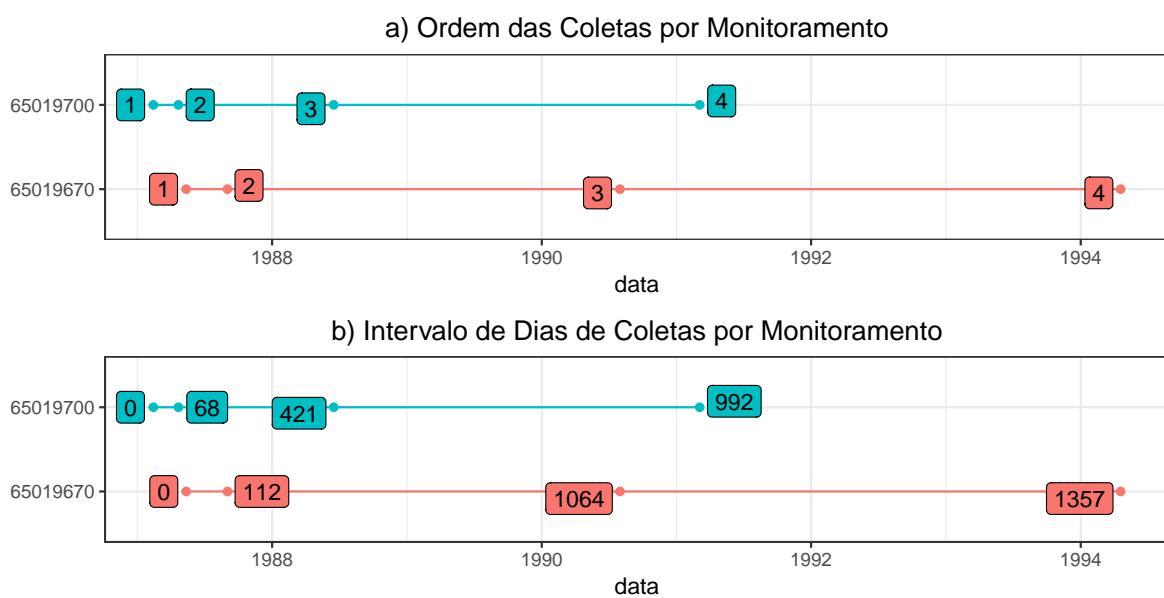


Figura 6 – Dependência temporal por monitoramento

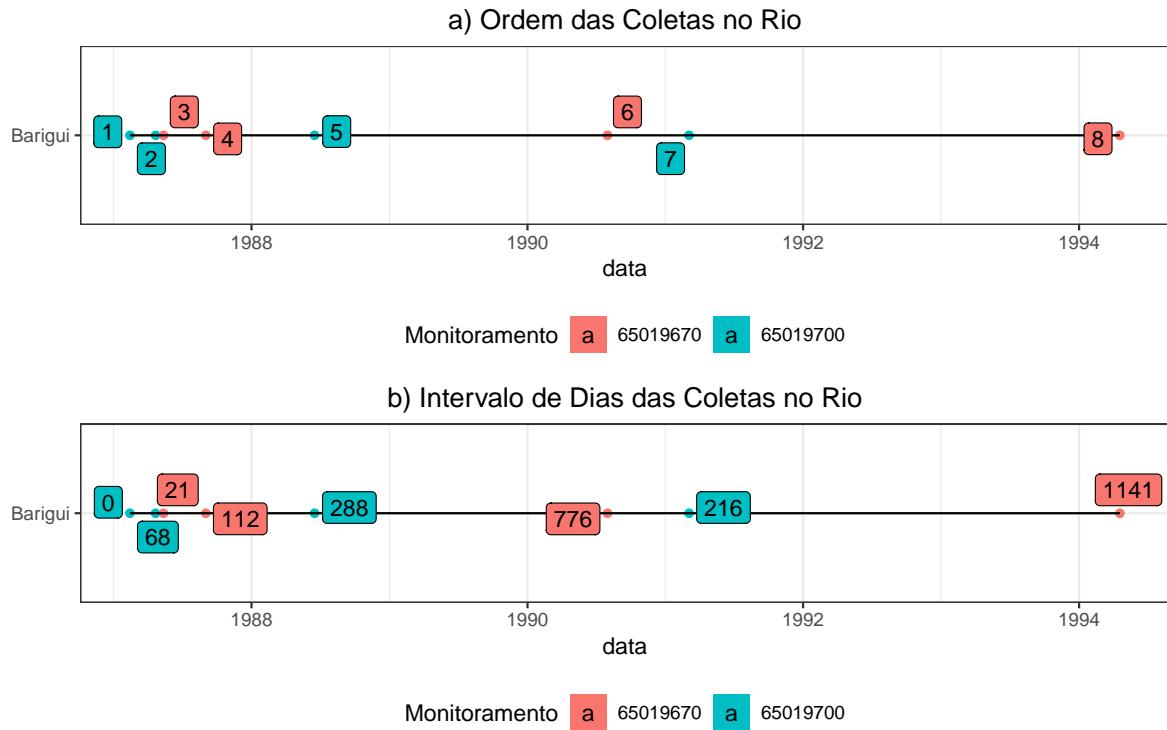


Figura 7 – Dependência da ordem das coletas ao longo do tempo

O intuito da construção dessas Matrizes é verificar se a variabilidade na resposta pode estar associada, pela dependência espacial entre as estações de monitoramento, ou pela ordem das coletas na própria estação de monitoramento ou considerando toda a trajetória das coletas no rio como um todo, independente da estação de monitoramento. A ponderação pela diferença em dias representa se a correlação da série é linear, tanto intra monitoramento ou levando em consideração todo o rio.

Em outras palavras, daquilo que não é captado pelo efeito das covariáveis na média da resposta, há alguma variabilidade restrita às próprias estações de monitoramento ao longo do tempo, ou algum efeito ao longo do tempo que afetou todas as estações de monitoramento durante as coletas no rio como um todo.

Além dessas matrizes foram construídas matrizes diagonais, considerando os dados sem dependência e a matriz para dados heterocedásticos considerando variabilidade por monitoramento.

Ou seja, na prática, se nenhuma das matrizes anteriores combinadas forem significativas, para uma específica resposta, os dados são independentes ao longo do tempo. Caso apenas a matriz de heterocedasticidade seja significativa, há alguma característica inerentes nas estações que não foi identificada. As especificações de todas as matrizes citadas estão exemplificadas na Figura 8.

Na Figura 8 são especificadas as matrizes de dependência que possibilitam

capturar a variabilidade dos dados. Em todas as matrizes há oito linhas por oito colunas, indicando oito coletas ilustrativas nos gráficos das Figuras 5, 6 e 7.

<p>(a) Z0</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>1</th><th>2</th><th>3</th><th>4</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td></tr> <tr><td>3</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td></tr> <tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr> </tbody> </table>		1	2	3	4	1	2	3	4	1	1								2		1							3			1						4				1					1					1				2						1			3							1		4								1	<p>(b) Z1 – Monitoramento 1</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>1</th><th>2</th><th>3</th><th>4</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td></tr> <tr><td>3</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td></tr> <tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr> </tbody> </table>		1	2	3	4	1	2	3	4	1	1								2		1							3			1						4				1					1					1				2						1			3							1		4								1		
	1	2	3	4	1	2	3	4																																																																																																																																																													
1	1																																																																																																																																																																				
2		1																																																																																																																																																																			
3			1																																																																																																																																																																		
4				1																																																																																																																																																																	
1					1																																																																																																																																																																
2						1																																																																																																																																																															
3							1																																																																																																																																																														
4								1																																																																																																																																																													
	1	2	3	4	1	2	3	4																																																																																																																																																													
1	1																																																																																																																																																																				
2		1																																																																																																																																																																			
3			1																																																																																																																																																																		
4				1																																																																																																																																																																	
1					1																																																																																																																																																																
2						1																																																																																																																																																															
3							1																																																																																																																																																														
4								1																																																																																																																																																													
<p>(c) Z1 – Monitoramento 2</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>1</th><th>2</th><th>3</th><th>4</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td></tr> </tbody> </table>		1	2	3	4	1	2	3	4	1	1								2									3									4									1				1					2					1				3						1			4							1		<p>(d) Inverso da Distância(KM)</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>1</th><th>2</th><th>3</th><th>4</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td></td><td></td><td></td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td></tr> <tr><td>2</td><td></td><td>1</td><td></td><td></td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td></tr> <tr><td>3</td><td></td><td></td><td>1</td><td></td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td></tr> <tr><td>4</td><td></td><td></td><td></td><td>1</td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td></tr> <tr><td>1</td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td><td>1</td><td></td><td></td><td></td></tr> <tr><td>2</td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td><td></td><td>1</td><td></td><td></td></tr> <tr><td>3</td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td><td></td><td></td><td>1</td><td></td></tr> <tr><td>4</td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td><td></td><td></td><td></td><td>1</td></tr> </tbody> </table>		1	2	3	4	1	2	3	4	1	1				0.05	0.05	0.05	0.05	2		1			0.05	0.05	0.05	0.05	3			1		0.05	0.05	0.05	0.05	4				1	0.05	0.05	0.05	0.05	1	0.05	0.05	0.05	0.05	1				2	0.05	0.05	0.05	0.05		1			3	0.05	0.05	0.05	0.05			1		4	0.05	0.05	0.05	0.05				1		
	1	2	3	4	1	2	3	4																																																																																																																																																													
1	1																																																																																																																																																																				
2																																																																																																																																																																					
3																																																																																																																																																																					
4																																																																																																																																																																					
1				1																																																																																																																																																																	
2					1																																																																																																																																																																
3						1																																																																																																																																																															
4							1																																																																																																																																																														
	1	2	3	4	1	2	3	4																																																																																																																																																													
1	1				0.05	0.05	0.05	0.05																																																																																																																																																													
2		1			0.05	0.05	0.05	0.05																																																																																																																																																													
3			1		0.05	0.05	0.05	0.05																																																																																																																																																													
4				1	0.05	0.05	0.05	0.05																																																																																																																																																													
1	0.05	0.05	0.05	0.05	1																																																																																																																																																																
2	0.05	0.05	0.05	0.05		1																																																																																																																																																															
3	0.05	0.05	0.05	0.05			1																																																																																																																																																														
4	0.05	0.05	0.05	0.05				1																																																																																																																																																													
<p>(e) Monitoramento Vizinhança de 1º Ordem</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>1</th><th>2</th><th>3</th><th>4</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td><td>1</td><td>1</td><td></td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr> <tr><td>3</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>1</td></tr> </tbody> </table>		1	2	3	4	1	2	3	4	1	1	1							2	1	1	1						3		1	1	1					4			1	1					1				1	1				2					1	1	1		3						1	1	1	4							1	1	<p>(f) Monitoramento Inverso do Intervalo em Dias</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>1</th><th>2</th><th>3</th><th>4</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>0.009</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td>0.009</td><td>1</td><td>0.001</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td>0.001</td><td>1</td><td>0.001</td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td>0.001</td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td><td></td><td>1</td><td>0.015</td><td></td><td></td></tr> <tr><td>2</td><td></td><td></td><td></td><td></td><td></td><td>0.015</td><td>1</td><td>0.002</td></tr> <tr><td>3</td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.002</td><td>1</td><td>0.001</td></tr> <tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.001</td><td>1</td></tr> </tbody> </table>		1	2	3	4	1	2	3	4	1	1	0.009							2	0.009	1	0.001						3		0.001	1	0.001					4			0.001	1					1					1	0.015			2						0.015	1	0.002	3							0.002	1	0.001	4								0.001	1
	1	2	3	4	1	2	3	4																																																																																																																																																													
1	1	1																																																																																																																																																																			
2	1	1	1																																																																																																																																																																		
3		1	1	1																																																																																																																																																																	
4			1	1																																																																																																																																																																	
1				1	1																																																																																																																																																																
2					1	1	1																																																																																																																																																														
3						1	1	1																																																																																																																																																													
4							1	1																																																																																																																																																													
	1	2	3	4	1	2	3	4																																																																																																																																																													
1	1	0.009																																																																																																																																																																			
2	0.009	1	0.001																																																																																																																																																																		
3		0.001	1	0.001																																																																																																																																																																	
4			0.001	1																																																																																																																																																																	
1					1	0.015																																																																																																																																																															
2						0.015	1	0.002																																																																																																																																																													
3							0.002	1	0.001																																																																																																																																																												
4								0.001	1																																																																																																																																																												
<p>Trajetória Coleta Vizinhança Ordem 1</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td></tr> <tr><td>5</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>6</td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr> <tr><td>7</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td></tr> <tr><td>8</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>1</td></tr> </tbody> </table>		1	2	3	4	5	6	7	8	1	1	1							2	1	1	1						3		1	1	1					4			1	1	1				5				1	1	1			6					1	1	1		7						1	1	1	8							1	1	<p>Trajetória Coleta Inverso da Diferença em Dias</p> <table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>0.015</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>2</td><td>0.015</td><td>1</td><td>0.048</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>3</td><td></td><td>0.048</td><td>1</td><td>0.009</td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td></td><td></td><td>0.009</td><td>1</td><td>0.003</td><td></td><td></td><td></td></tr> <tr><td>5</td><td></td><td></td><td></td><td>0.003</td><td>1</td><td>0.001</td><td></td><td></td></tr> <tr><td>6</td><td></td><td></td><td></td><td></td><td>0.001</td><td>1</td><td>0.005</td><td></td></tr> <tr><td>7</td><td></td><td></td><td></td><td></td><td></td><td>0.005</td><td>1</td><td>0.001</td></tr> <tr><td>8</td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.001</td><td>1</td></tr> </tbody> </table>		1	2	3	4	5	6	7	8	1	1	0.015							2	0.015	1	0.048						3		0.048	1	0.009					4			0.009	1	0.003				5				0.003	1	0.001			6					0.001	1	0.005		7						0.005	1	0.001	8							0.001	1		
	1	2	3	4	5	6	7	8																																																																																																																																																													
1	1	1																																																																																																																																																																			
2	1	1	1																																																																																																																																																																		
3		1	1	1																																																																																																																																																																	
4			1	1	1																																																																																																																																																																
5				1	1	1																																																																																																																																																															
6					1	1	1																																																																																																																																																														
7						1	1	1																																																																																																																																																													
8							1	1																																																																																																																																																													
	1	2	3	4	5	6	7	8																																																																																																																																																													
1	1	0.015																																																																																																																																																																			
2	0.015	1	0.048																																																																																																																																																																		
3		0.048	1	0.009																																																																																																																																																																	
4			0.009	1	0.003																																																																																																																																																																
5				0.003	1	0.001																																																																																																																																																															
6					0.001	1	0.005																																																																																																																																																														
7						0.005	1	0.001																																																																																																																																																													
8							0.001	1																																																																																																																																																													

Figura 8 – Matrizes de dependência

Na matriz (a) assume-se que a variabilidade é única para todas as coletas, já que

só há valores na diagonal principal. Para as Matrizes (b) e (c) assume-se que cada estação de monitoramento possui sua própria variabilidade. Na matriz (d) a variabilidade é expressa pela inversa da distância entre as estações de monitoramento. Nas matrizes (e) e (f) é a dependência da vizinhança de primeira ordem das coletas na mesma estação de monitoramento, porém na (f) é ponderado pelo inverso da diferença em dias.

Considerando o rio como um todo, as matrizes (g) e (h) relacionam a vizinhança de primeira ordem das coletas, independente da estação e a matriz (h) pondera pelo inverso da diferença em dias de uma coleta à outra.

## 4.4 Recursos Computacionais

Foi usado o *software R* na versão 3.4 ([R Core Team, 2017](#)), para criar novas variáveis, validar informações, gerar gráficos, aplicar técnicas multivariadas, como componentes principais, *t-SNE* e modelos com resposta multivariada do conjunto MCGLM. Principais pacotes utilizados R ([WICKHAM, 2016](#)), ([WICKHAM et al., 2018](#)) e ([CHENG; KARAMBELKAR; XIE, .](#)).

Também foi usado o *software QGIS* na versão 2.18.16 ([QGIS Development Team, 2009](#)), para operações de geoprocessamento, como a intersecção da camada do uso do solo na respectiva coordenada da estação de monitoramento e identificação da altitude do local de monitoramento.



# 5 Resultados e Discussões

## 5.1 Análise Descritiva

### 5.1.1 Característica dos Dados

Primeiramente é apresentado a quantidade de dados, por ano, separado por origem (AGUASPARANÁ e DHS/UFPR), posteriormente a frequência de coletas em cada bacia, a distribuição de coletas dentro das bacias.

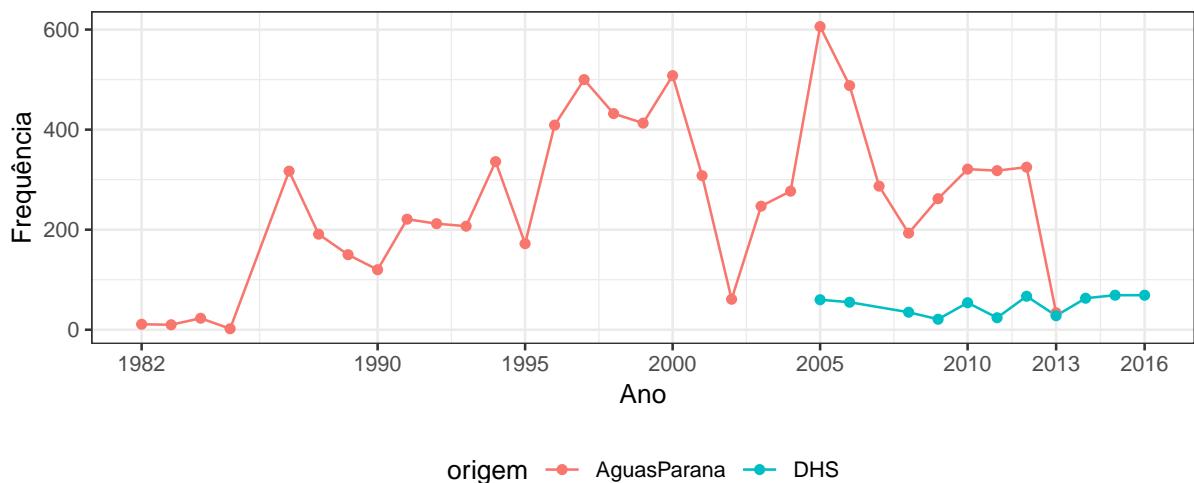


Figura 9 – Frequência de coleta por ano de cada equipe

A Figura 9 mostra a oscilação na quantidade de coletas realizadas, pelo instituto do AGUASPARANÁ, alternando quantidades expressivas e escassa de coletas, ao longo dos anos. Segundo os servidores consultados do órgão, o motivo desta oscilação é por motivos técnicos, logísticos, e estruturais, e que o cenário ideal seriam coletas trimestrais. Por outro lado, as coletas da equipe do DHS, apresentam menor variabilidade entre os anos de coleta iniciada em 2005.

A Tabela 2 a seguir mostra por bacia hidrográfica, a quantidade de coletas, juntamente com o número de diferentes estações, municípios, rios e o ano mais recente de coleta.

Tabela 2 – Quantidades de dados por bacia

Bacia	Coletas	Estações	Municípios	Rios	Ano
Iguaçu	5666	160	56	76	2016
Tibagi	534	29	18	17	2013
Ivaí	523	22	19	16	2013
Litorânea	416	18	6	14	2013
Piquiri	397	16	13	8	2013
Ribeira	354	11	7	7	2013
Cinzas	213	10	10	6	2012
Paraná 3	163	8	5	4	2012
Pirapó	109	10	9	4	2012
Itararé	62	5	3	4	2012
Paranapanema 3	42	2	2	2	2007
Paraná 1	25	1	1	1	2000
Paranapanema 1	2	1	1	1	1984
Total	8506	293	148	153	2016

A tabela 2 está ordenada pela quantidade de coletas. A bacia Iguaçu domina praticamente as coletas em relação às demais, consequentemente possui a maioria das estações de monitoramento, identificadas nos diferentes rios e cidades. A coluna ano, apresenta o último ano em que houve coleta, as bacias Paranapanema 3, Paraná 1 e Paranapanema 1, não tiveram coletas recentes, e a última apresenta apenas duas coletas, portanto, foi removida das análises.

A Figura 10 apresenta comparativamente por meio dos diagramas *boxplots* a quantidade de coletas nas estações de monitoramento por bacias e o ponto representa o número de coletas.

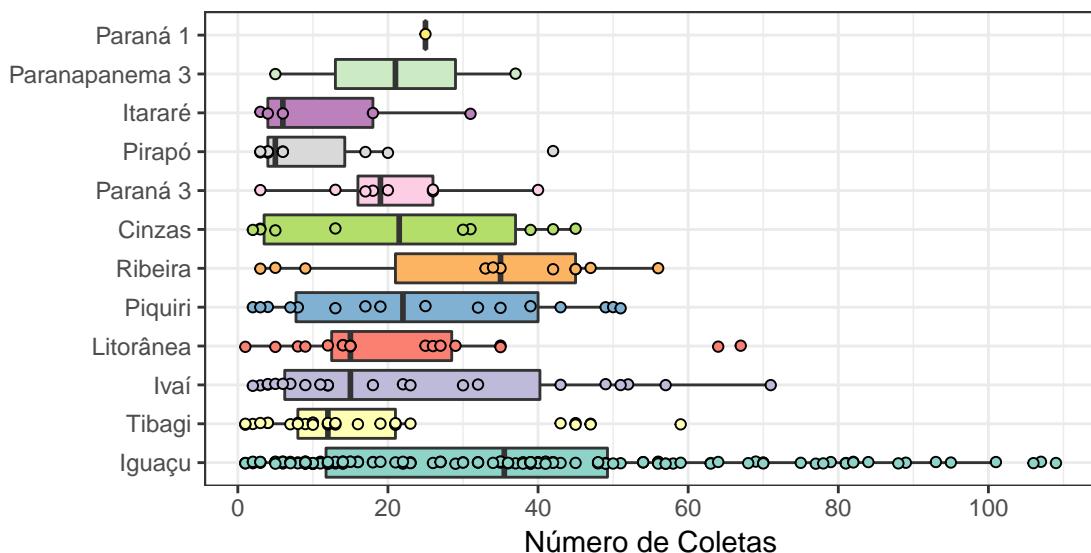


Figura 10 – Quantidade de coletas nas estações de monitoramento por bacias

A Figura 10 evidencia que, em todas as bacias, não há uniformidade na quantidade de coletas. Em todas há uma estação de monitoramento com menos de 10 coletas. Além disso, há estações de monitoramento da bacia do Iguaçu, com mais coletas do que todas as estações da Paraná 1 e Paranapanema 3 juntas.

No *link* (Característica) é apresentado a distribuição espacial das estações, sendo predominantes nas cidades com maiores populações e principalmente na região metropolitana de curitiba. No mesmo *link*, também é mostrado a quantidade de dados por época e enquadramento para cada bacia e a quantidade de dados faltantes por variável.

### 5.1.2 Montagem do Banco de Dados

Nesta seção destaco as incosistências registradas, nos parâmetros da qualidade, e sem passar por esta inspeção, poderia haver análises equivocadas. A identificação das inconsistências passa por comparar o IQA calculado pelo instituto AguasParaná, e pelo autor utilizando as fórmulas no apêndice.

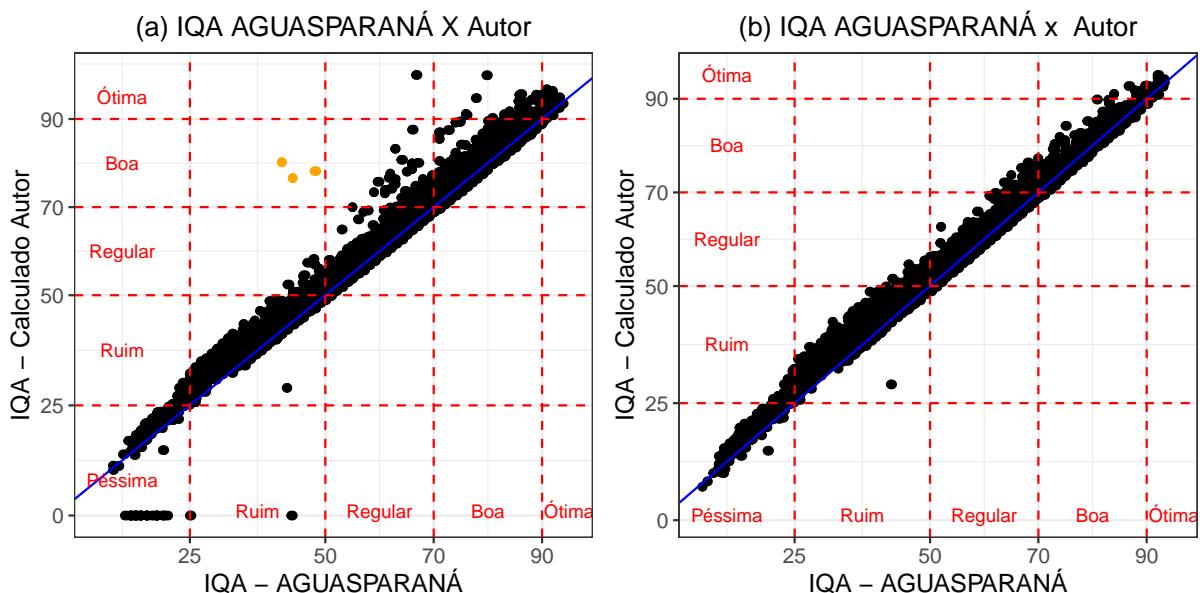


Figura 11 – Comparação do IQA calculado pelo AGUASPARANÁ e pelo autor

Ambos os gráficos da Figura 11 apresentam o valor do IQA calculado pelo instituto e aqui recalculados. No gráfico à esquerda, a maioria dos valores está acima da linha azul que representa a reta  $x=y$ , indicando que o IQA calculado pelo autor é maior que o fornecido.

Ainda que a maioria dos valores esteja no quadrante da reta azul, há valores bem discrepantes, como os pontos em laranja, e que implicam em diferentes categorizações da qualidade da água conforme o Quadro 2.

O gráfico à direita mostra novamente a comparação dos cálculos, mas levando em consideração a remoção dos dados incongruentes por meio dos especialistas em parâmetros da qualidade da água conformes as verificações indicadas na Tabela 3.

Como não tive acesso na metodologia do AGUASPARANÁ, não posso inferir sobre o procedimento de readequação do IQA com valores inválidos.

Tabela 3 – Quantidade de Dados Inválidos

Limites	Nº de Inválidos
$0 < OD < 9.2$	1115
$DQO > DBO$	187
$NT > NK$	25
$CT > CF$	40
$0 < PH < 14$	6

### 5.1.3 Contribuição dos Parâmetros

Nesta sub seção são destacadas as contribuições dos parâmetros que compõem o IQA está na oscilação do índice em todas as bacias conforme mostra a Figura 12.

No gráfico da Figura 12 a linha vermelha representa a média na bacia e a linha azul é um ajuste suavizado, por meio dos polinômios locais. A interpretação das linhas é da seguinte forma, caso a linha média esteja englobada no ajuste suave, então o IQA não variou ao longo do tempo. No decorrer do estudo, gráficos semelhantes à Figura 12, para cada parâmetros serão apresentados, e mantêm as mesmas propriedades e interpretações.

Em todas as bacias não há intersecção da linha média com o ajuste suavizado ao longo do tempo. Nota-se que a partir dos anos 2000, houve um decaimento e logo após uma crescimento, o que diferiu nas bacias foi a intensidade deste comportamento, já que as amplitudes dos eixos são iguais.

Com o intuito de descobrir qual parâmetro está causando esta oscilação é apresentado a Figura 13 que converte os parâmetros nas escalas de concentração, descritos no Apêndice A para o peso do IQA, com amplitude fechada de 0 a 100.

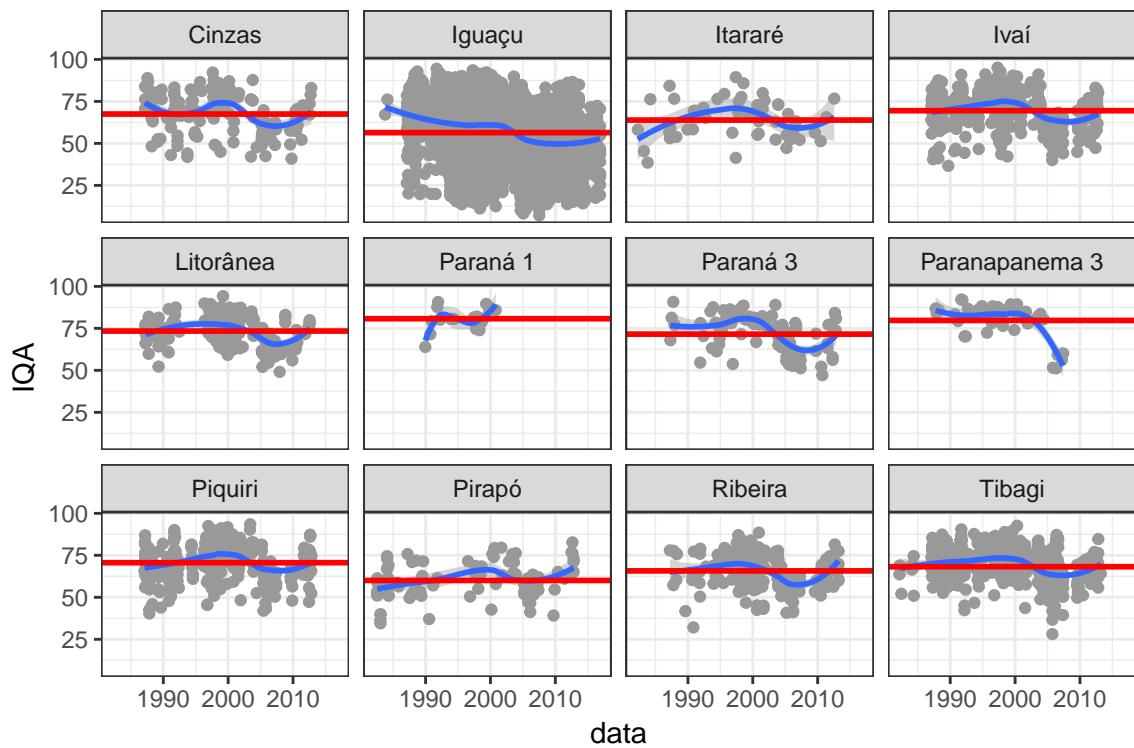


Figura 12 – IQA ao longo do tempo por Bacia

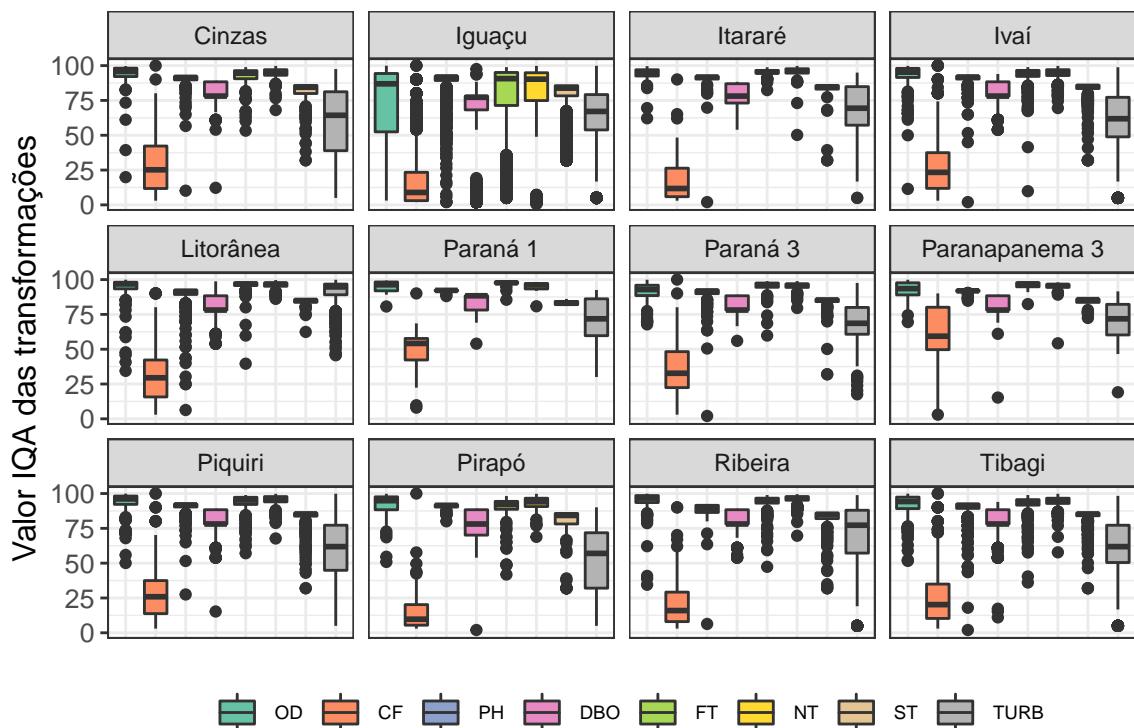


Figura 13 – Boxplot dos Pesos dos Parâmetros por Bacia

Na Figura 13 observa-se que o CF e a Turb, possuem as menores medianas em todas as bacias enquanto que para outros parâmetros, a mediana está bem próxima do valor máximo, e apenas há *outliers* inferiores.

Desta forma a Figura 14 apresenta o IQA sem os parâmetros CF e TURB, redistribuindo os pesos igualmente para os outros parâmetros conforme o Quadro 1 que elenca os pesos originais.

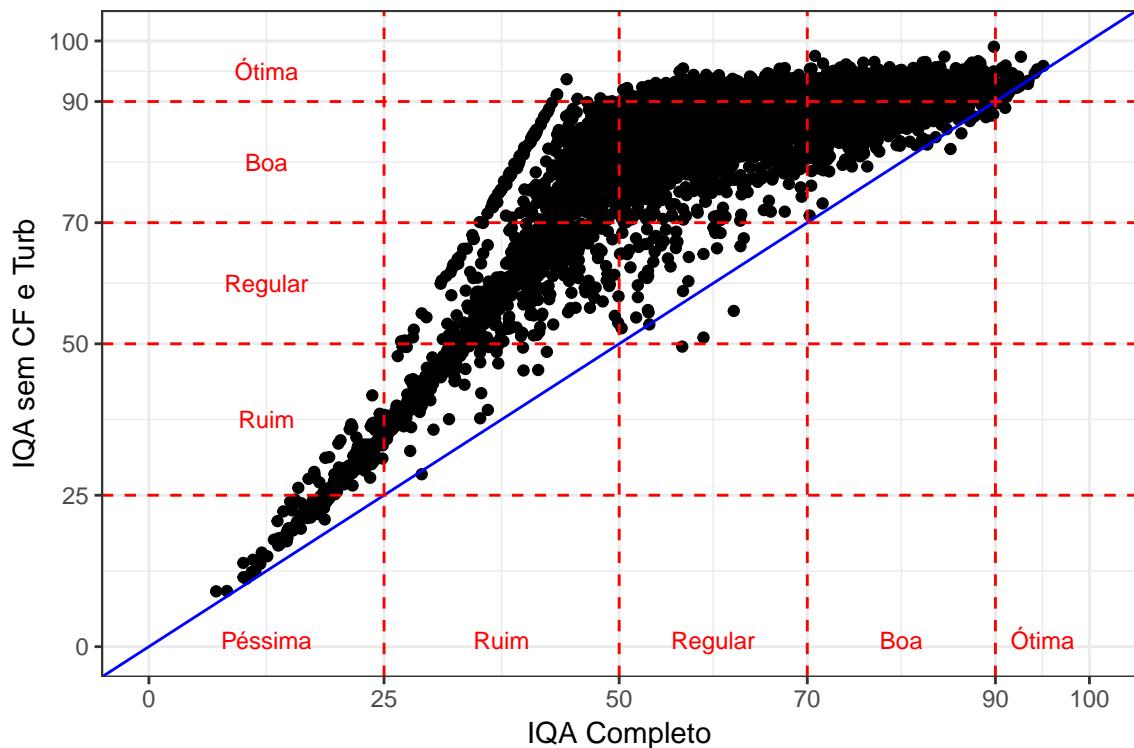


Figura 14 – IQA calculado sem CF e TURB

Mais análise da identificação dos parâmetro que contribui para o IQA está disponível no *link* (Valida), assim como, foram adicionadas análises para a variável Turb, comparando temporalmente por bacia na escala do IQA e também recalculando o IQA sem a variável Turb.

#### 5.1.4 Relação Enquadramento com IQA e Limites CONAMA

A proposta da seção é visualizar a relação da qualidade da água com o enquadramento descrito no Quadro 2 e se os limites de vigências aplicados pelas resolução da CONAMA estão sendo atendidos conforme o enquadramento dos corpos hídricos.

A Figura 15 apresenta o enquadramento dos corpos hídricos relacionado com as categorias do IQA, sendo destacado a quantidade de coletas em cada categoria e no eixo x a respectiva porcentagem.

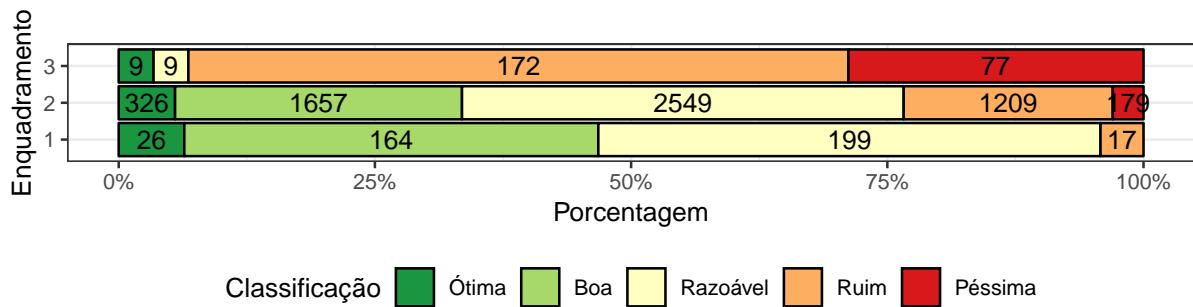


Figura 15 – Relação IQA e enquadramento por Bacia

Observa-se que a qualidade da água está inferior à esperada, em todas as classes. Era de se esperar que a classe 1 tivesse a maioria das coletas fosse ótima, na classe 2 a maioria das coletas fosse boa e para o enquadramento de classe 3 que a maioria das coletas estivesse na categoria regular.

Quanto ao atendimento dos limites estabelecidos pela CONAMA, o gráfico de barras na figura 16 para o parâmetro OD, relaciona que a maioria das coletas em todas as bacias estão sendo atendidas.

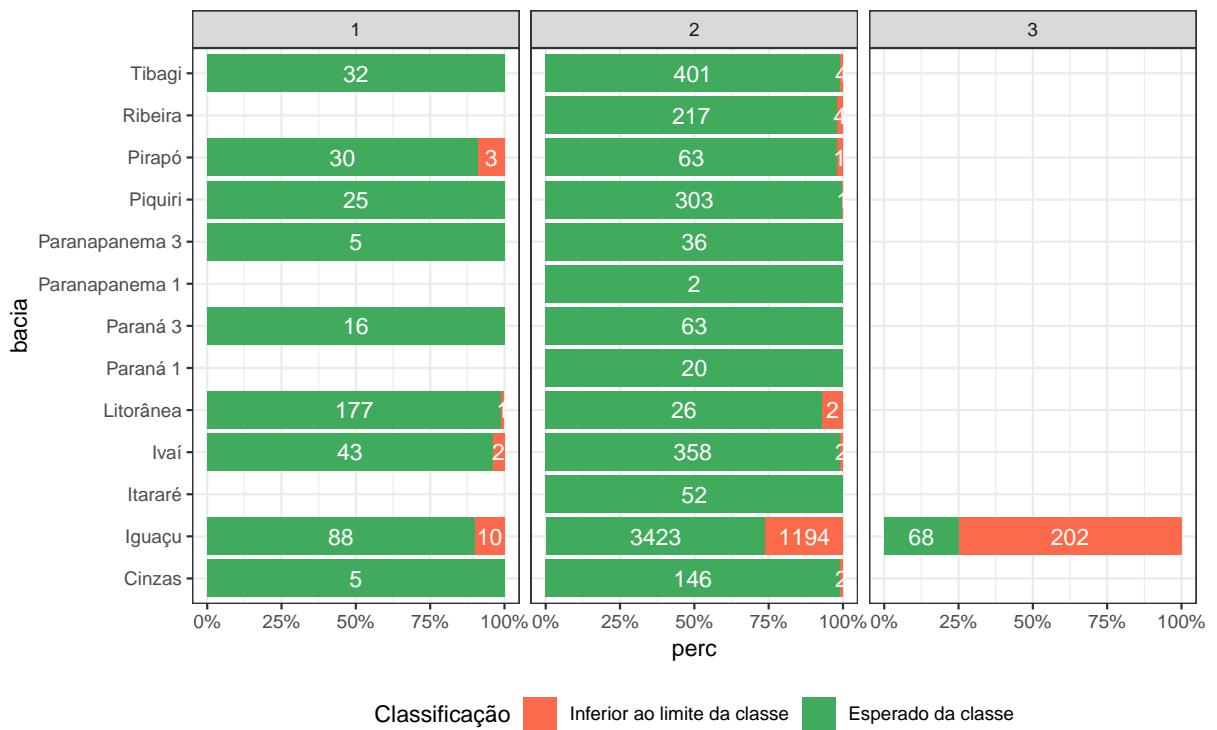


Figura 16 – Relação OD por classe de Rio

No link (CONAMA) são apresentadas as porcentagens atendida dos parâmetros estabelecidas pela CONAMA por bacia. Os parâmetros CF e CT são os mais preocupantes levando em consideração os limites da CONAMA em todos os enquadramentos.

Também é colocado a relação do enquadramento e qualidade da água para todos os dados e segmentado por bacia. Os resultados mostram que na maioria das bacias, a qualidade está inferior a esperada. Na classe 1 predomina a qualidade Boa, na classe 2 predomina a qualidade Regular e na classe 3, predomina a qualidade Ruim.

### 5.1.5 Parâmetros por Bacia Hidrográfica

### 5.1.6 Comparação entre Bacias

As comparações por bacia foram feitas da seguinte forma, gerou-se gráficos temporais dos parâmetros na escala da concentração, carga e peso do IQA. A visualização completa de todos os parâmetros está no *link* (Escalas). No presente texto, é exemplificado para o parâmetro OD.

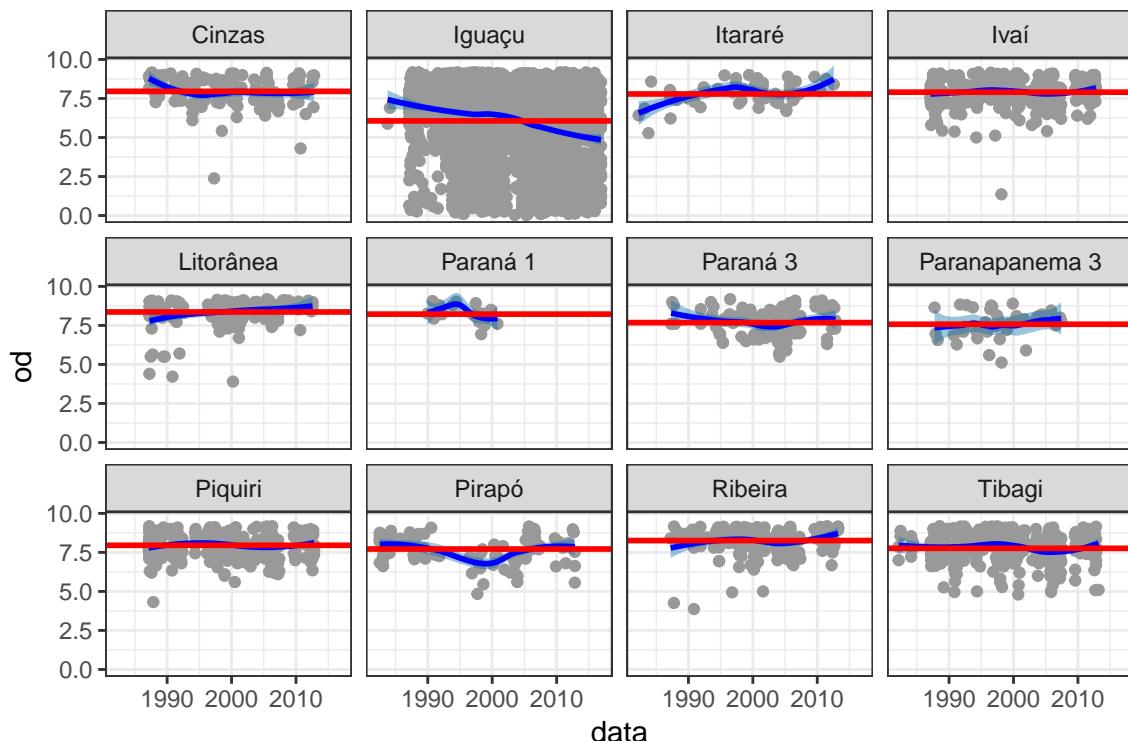


Figura 17 – Concentração do OD ao longo do tempo por Bacia

Na Figura 17 são apresentadas as concentrações medidas de OD em diferentes bacias. Observa-se que as bacias do Iguaçu, Itararé e Pirapó apresentam oscilação ao longo do tempo e no caso a Iguaçu está decrescendo enquanto para as demais os valores se mantêm sem grandes alterações.

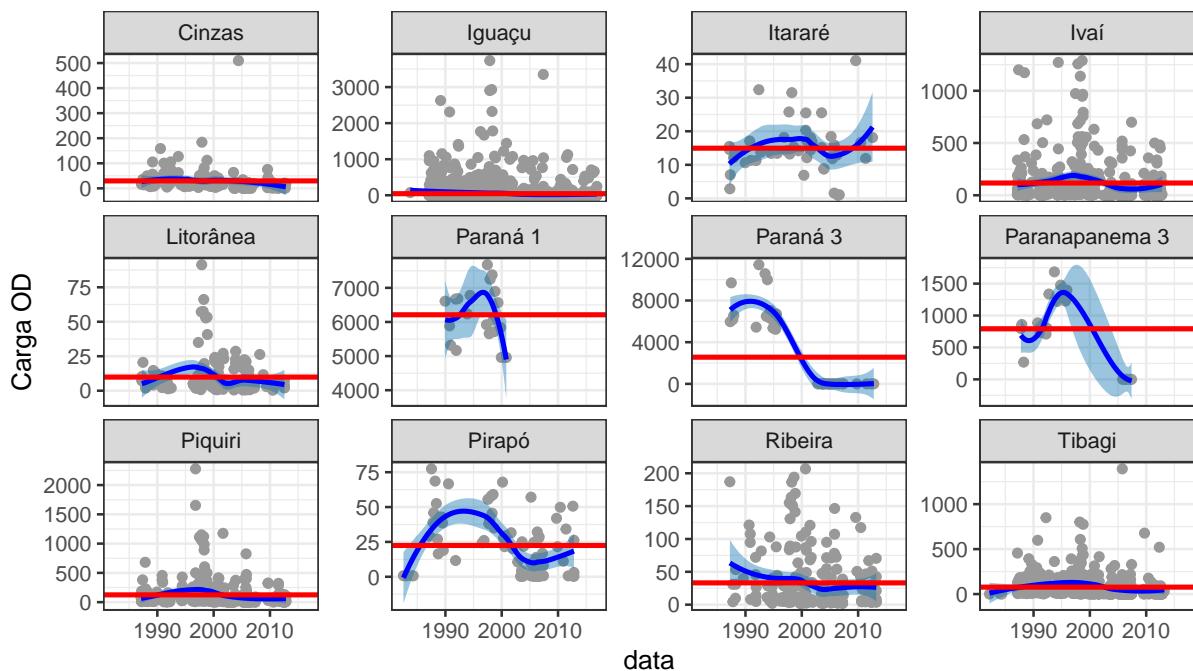


Figura 18 – Tonelas por dias de OD por Bacia

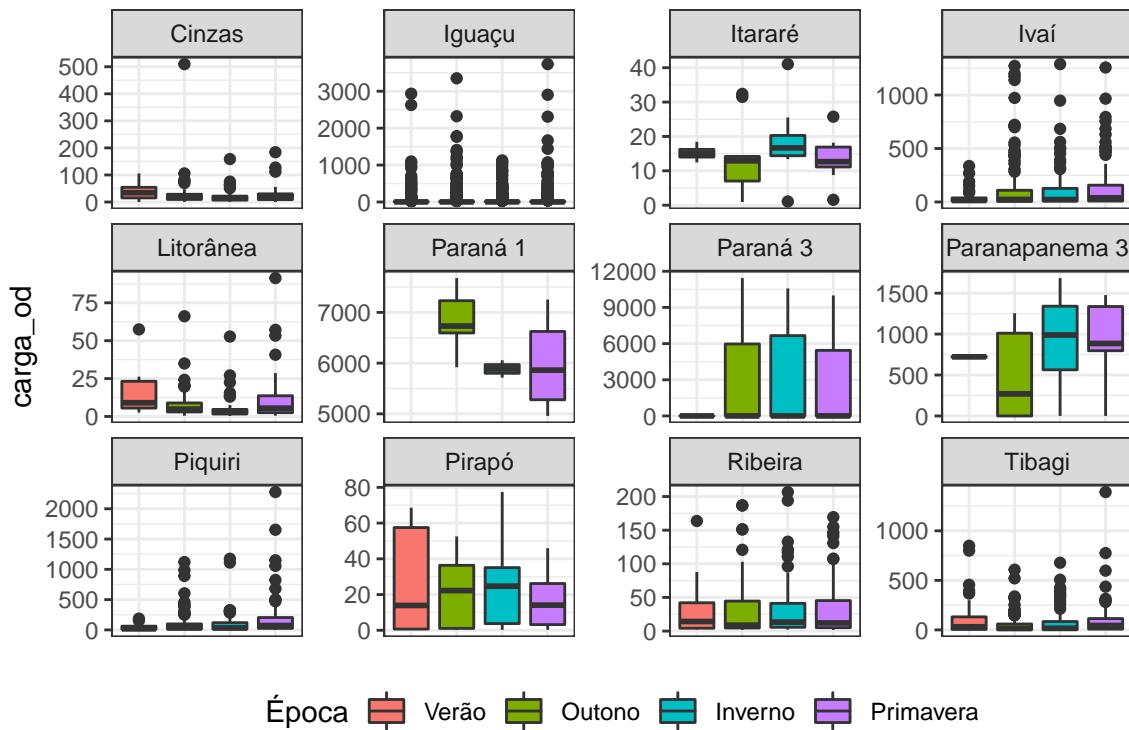


Figura 19 – Boxplot das épocas do ano por bacia

Entretanto ao considerar a carga do OD na unidade (kg/dias), visivelmente todas as bacias apresentam um decréscimo após os anos 2000, ou seja está diminuindo a quantidade de OD transportado em todas as bacias conforme mostra a Figura 18.

Por sua vez, ao identificar a época da coleta, foi construído o *boxplot* na tentativa de identificar diferenças. Para as bacias Litorânea e Cinzas o OD aumenta na época do Verão, na bacia de Itararé e Paranapanema 3 o OD diminui no Outono e na Paraná 1 aumenta conforme apresentado na Figura 19.

### 5.1.7 Alto Iguaçu

A região do Alto Iguaçu concentra praticamente metade das estações de monitoramento do estado conforme indica a Tabela 4. A unidade está em sua maior parte na região metropolitana de Curitiba e apresenta valores preocupantes de qualidade da água. Esta unidade também é objeto de estudo do DHS/UFPR o que permite que sejam feitas comparações e combinações dos dados de ambas as fontes.

Tabela 4 – Sumarização Alto Iguaçu por Fonte

Fonte	Coletas	Estações	Municípios	Rios	Ano
AguasParana	4033	80	18	39	2013
DHS	545	24	11	7	2016
Total	4578	104	21	40	2016

Por se tratar da mesma bacia, a Figura 20 apresenta os parâmetros da qualidade ao longo do tempo. Os parâmetros CF, Turb, ST e CT foram recalculados na escala log na base 10 devido às assimetrias verificadas nos dados.

Praticamente em todos os gráficos o parâmetro se mantém ao longo do tempo, exceto o OD, já identificado nos gráficos anteriormente. Os parâmetros de CF e CT apresentam comportamentos diferentes. À medida que passa o tempo, o CT aumenta enquanto que o CF a partir de 2010 voltou para a mesma resposta do ínicio do monitoramento.

A relação entre os CT e CF é explicada, por meio da Figura 21, em que os valores das coletas do DHS/UFPR estão aumentando ao longo do tempo mas em compensação está diminuindo nas coletas do AGUASPARANÁ.

Além disso em ambas as fontes, nota-se que praticamente todos os parâmetros, possuem a mesma característica na curva ao longo dos anos, exceto o OD, em que a escala da curva é diferente. Portanto o recorte da origem pode propiciar resultados diferentes. No link (Alto Iguaçu) também é feito esta comparação a nível de carga e peso do IQA.

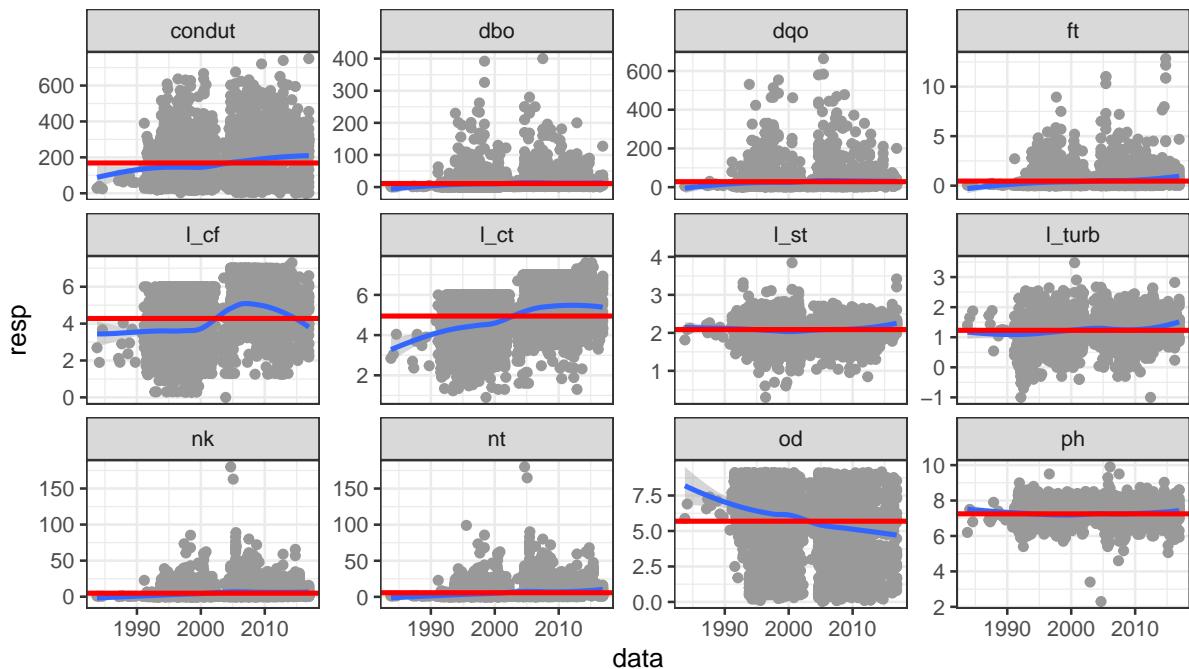


Figura 20 – Parâmetros da qualidade da água ao longo do tempo

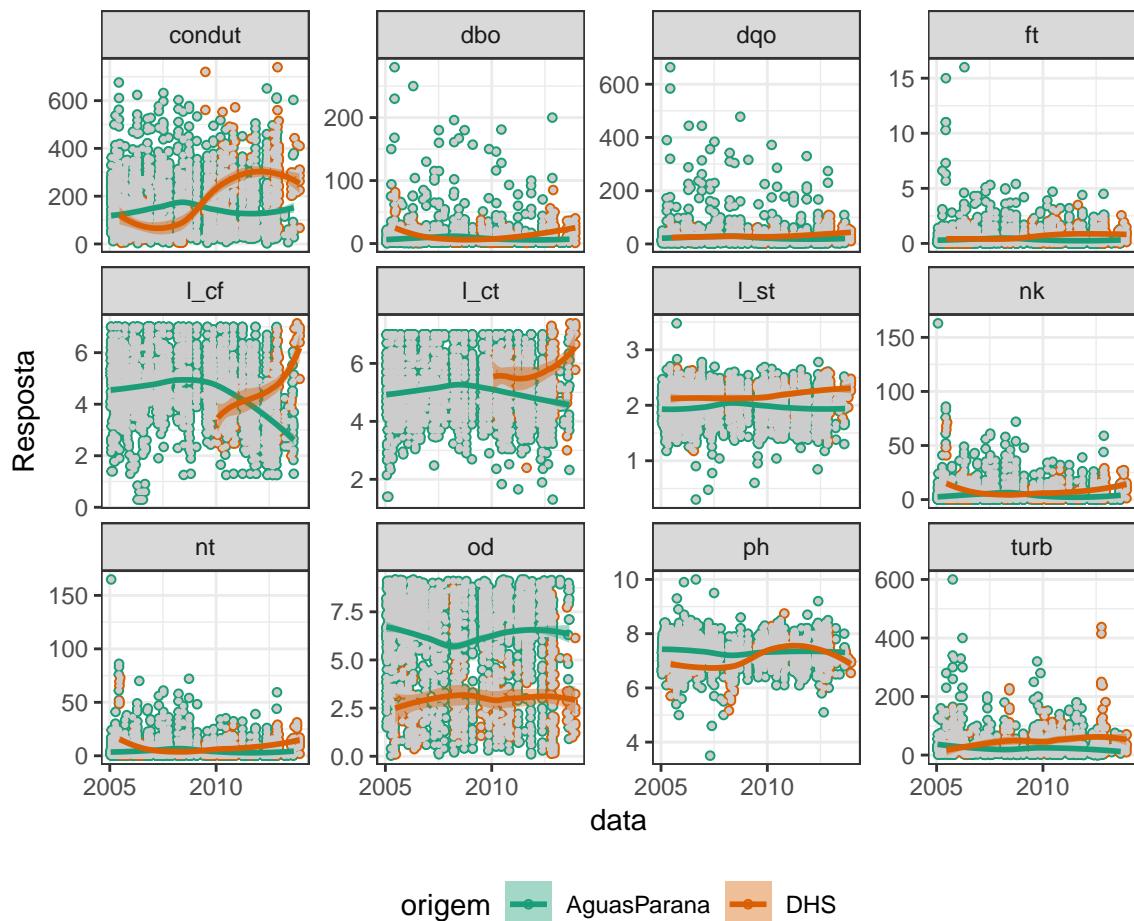


Figura 21 – Comparação das fontes da qualidade da água

## 5.2 Redução de Dimensionalidade

### 5.2.1 Identificação de diferenças e similaridades entre bacias e intra estações

Com o intuito de identificar a variabilidade por bacia considerando os parâmetros em escala de carga, concentração e peso do IQA, é aplicada a técnica PCA. Como desdobramento do resultado foram construídos novos *scores* a partir dos componentes principais. O no plano projeto foi atribuído possíveis legendas como classe do rio, época do ano, rio e período que pudessem explicar a variabilidade.

Os resultados estão exemplificados na Figura 22, ambos apresentam divergência na associação entre CF e ST, assim como a Turb e CF. Na Litorânea OD e PH estão inversamente associados, já na Iguaçu estão diretamente associados. Por outro lado, parâmetros como FT, NT, DBO estão positivamente correlacionados em ambas as bacias.

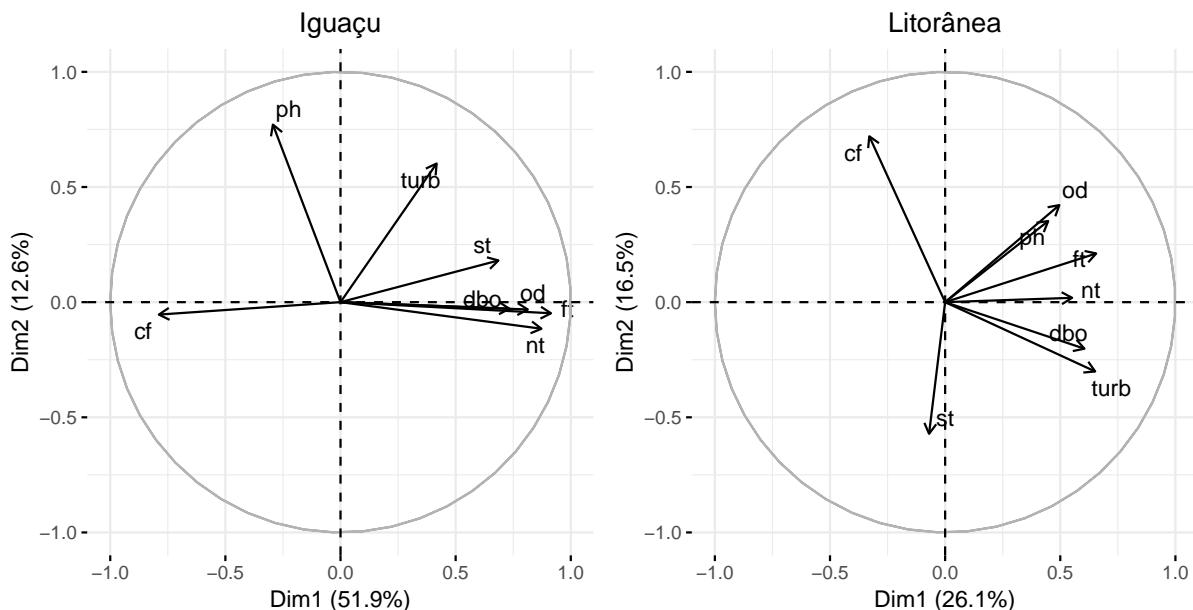


Figura 22 – PCA nas transformações do IQA na bacia do Iguaçu e Litorânea

Levando em consideração o somatório dos dois componentes do ACP, a explicação da variabilidade na maioria dos gráficos não passa de 70%. Mesmo assim, optou-se por produzir scores e atribuir variáveis categóricas, no intuito de identificar diferenças conforme mostra a Figura 23, mas não foram encontradas diferenças visuais.

No gráfico da estações de monitoramento, a elipse é o intervalo de confiança do grupo, e como estão todas sobrepostas não há diferenças evidentes.

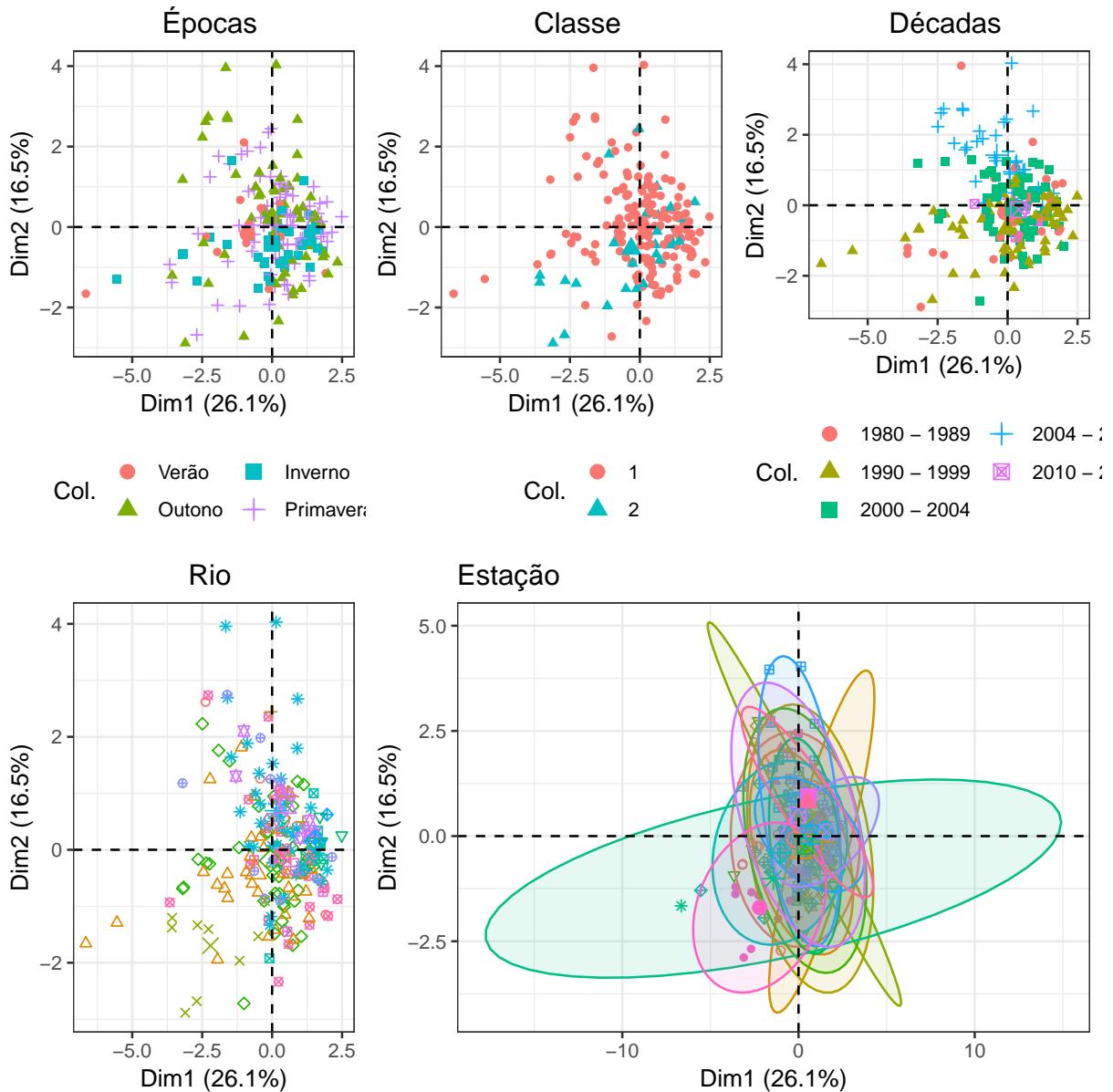


Figura 23 – Scores PCA

No *link* (ACP) são apresentadas para todas as bacias em diferentes escala, concentração, carga e transformação IQA, as coordenadas do ACP e os *scores*. Na escala da concentração e transformação IQA, em nenhuma das bacias foram verificadas diferenças visuais ao identificar os grupos nos gráficos de *score*. Por outro lado na escala carga, na maioria das bacias, a explicação da variabilidade por dois componentes atinge quase 100 %, e todas as variáveis apontam na mesma direção.

A fim de comparar resultados no métodos de redução de dimensionalidade, foi aplicado o *t-SNE*. O conjunto de dados aplicado não foi separado por bacia, sem transformação multivariada, sem centralizar e padronizar as variáveis. No presente texto é apresentado para a escala do IQA, mais resultados são disponibilizados no *link* (*t-SNE*) nas escalas de concentração e carga.

Os resultados são apresentados na Figura 24. Como adição, é identificado o grupo da qualidade da água conforme o Quadro 2. Verifica-se diferença neste grupo, já que os próprios dados usados são das transformações do IQA. Entretanto nos demais grupos, a nuvem de pontos não é discernida nas legendas atribuídas como: classe de rio, época e bacia. Exceto na atribuição da classificação da qualidade da água, pois os dados usados eram justamente das variáveis transformadas na escala IQA.

Em todas as escalas, de maneira geral, a redução de dimensionalidade nos planos 2D não identificou diferenças entre as legendas atribuídas.

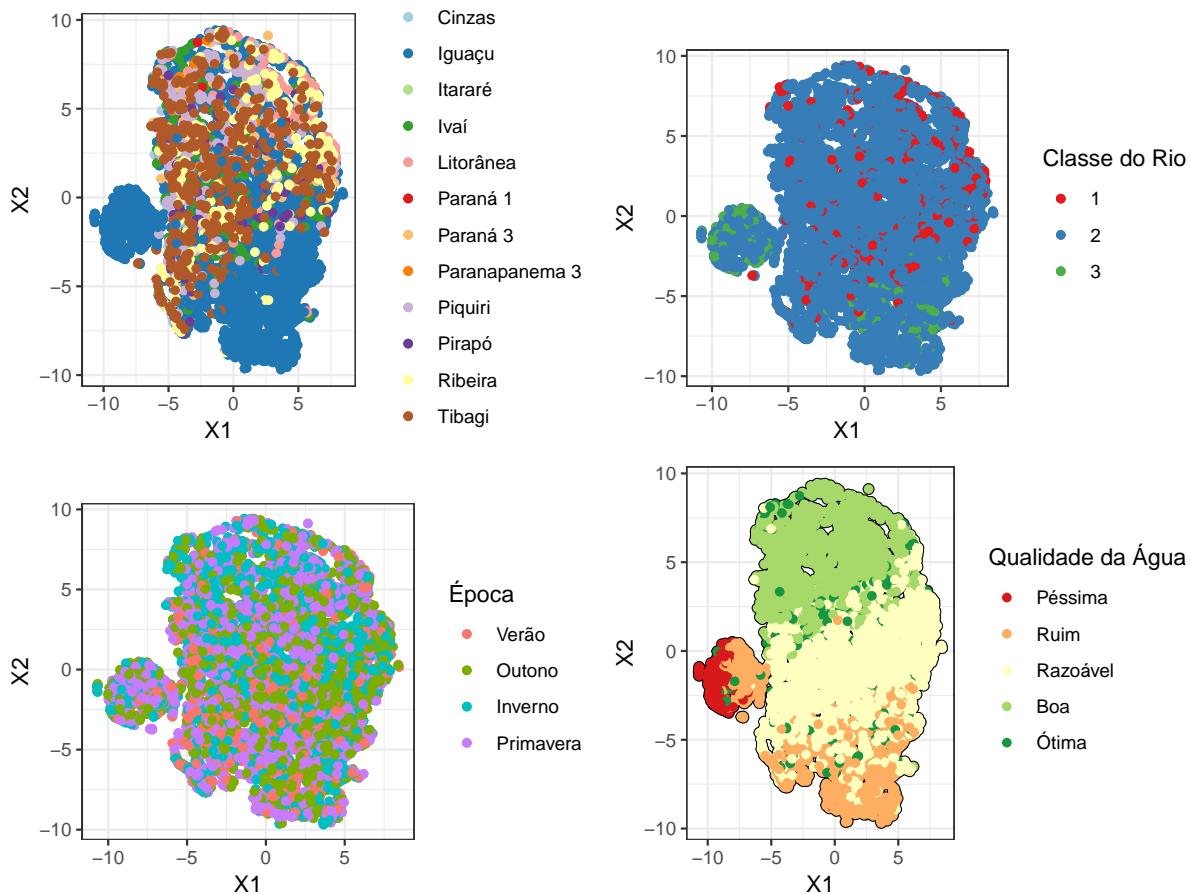


Figura 24 – Score - t-SNE

O uso dessas técnicas não mostrou diferenças, muito possivelmente por não considerar a dependência temporal entre as observações, a hierarquia da bacia, para o rio, para estação de monitoramento e por não incluir a proximidade das estações de monitoramento.

## 5.3 Modelo Multivariado

### 5.3.1 Rio Autodepurável

Nesta seção são apresentados os resultados da análise do Rio Barigui. A escolha deste rio, foi motivada pela existência de cinco estações de monitoramento e razoáveis número de coletas que compreendem todo o período de tempo, conforme Tabela 5. No caso, ressalta-se que os pontos monitorados não cobrem toda a extensão do rio, mas a maior parte da extensão, sendo assim, as inferências valem para os trechos avalidos.

Tabela 5 – Nº de Coletas por Monitoramento

Monitoramentos	65019520	65019650	65019670	65019685	65019700	Total
Nº Coletas	37	39	42	39	79	236

A análise tem como foco contribuir com a discussão da sustentabilidade do Rio, e para isso foi verificado se o parâmetro da qualidade da água é estatisticamente diferente entre as estações, já que podem ser consideradas fotografias do rio. A comparação é feita levando em consideração a ordem das estações no rio, já que pode ser indicada pela altitude no ponto de monitoramento. Dessa maneira a ordenação das estações segue a trajetória do rio.

Os parâmetros avaliados foram os que compõem o IQA e todos na escala da concentração. Para a temperatura não houve convergência nos algoritmos numéricos e portanto não serão reportados os resultados. Para as respostas CF, Turb e ST foi passado o log na base 10, para diminuir a assimetria dos dados e na variável OD, por ser tratar de uma variável limitada foi transformada na escala de 0 a 1, utilizando a seguinte expressão.

$$Intod = \frac{od - min_{od}}{max_{od} - min_{od}} \quad (5.1)$$

Além de usar as estações de monitoramento como covariável categórica, foi considerado como categórica a época do ano (Verão, Outono, Inverno e Primavera) e o ano como numérico. A construção do modelo multivariado passou primeiramente por análises descritivas, construindo *boxplots* das respostas por época, pontos de monitoramento e a série temporal de cada resposta, na ordem do rio. No trabalho será exemplificado para a resposta OD.

Na Figura 25, no gráfico (a), visualmente não existe diferença por época. No gráfico (b), visualmente há diferença de medianas e variação entre as estações, principalmente entre as duas primeiras com as duas últimas. No gráfico (c), é apresentado a

série temporal das estações conforme a ordem do rio. Observando as linhas suavizadas percebe-se que há pouca oscilação dos valores em cada uma das estações.

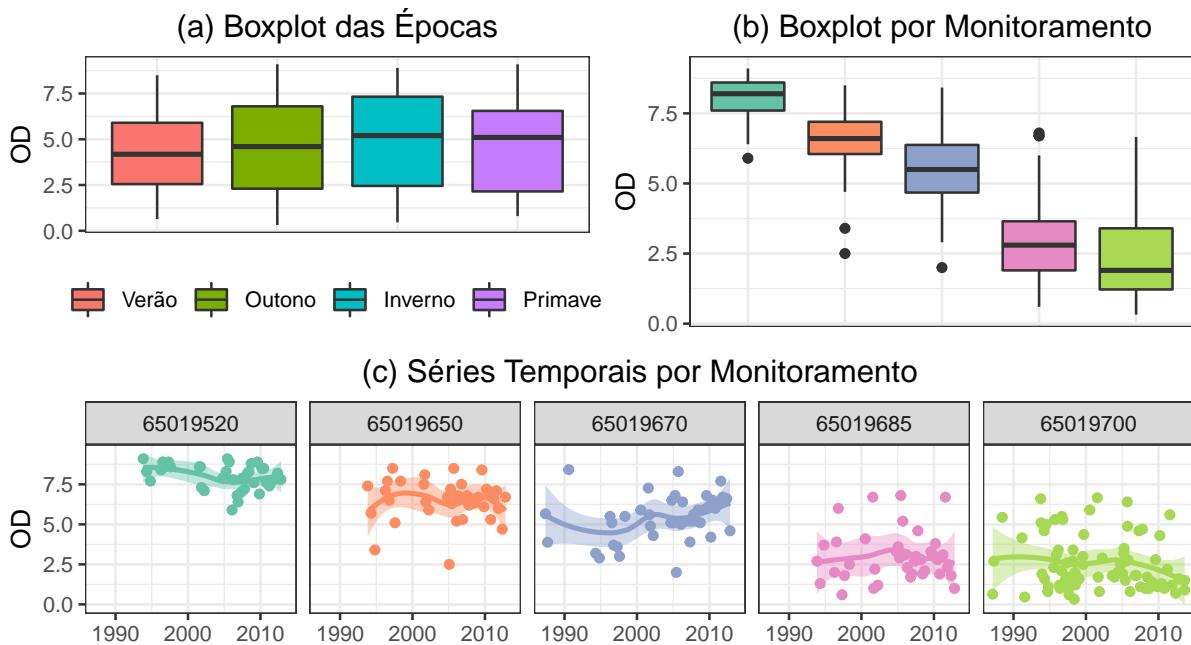


Figura 25 – Análise Descritiva no Parâmetro OD

Após realizar a análise descritiva, o procedimento de modelagem foi realizado da seguinte forma:

Primeiramente ajustou-se para cada resposta (parâmetros que compõem o IQA) as covariáveis mais significativas. Levou-se em consideração o procedimento backward, em que se começa com todas as covariáveis com interação e retirando as menos significativas, a nível de 5%. Neste procedimento controlou-se a matriz ( $\tau$ ) de dependência do modelo pela matriz Z0 (variabilidade única para os dados), indicado na Figura 8 no gráfico (a). Foi adotado fixar z0 pois deseja-se maximizar a explicação pelas covariáveis do que pela variabilidade, já que as covariáveis possibilitam uma interpretação mais informativa.

No segundo passo, após definir as covariáveis mais significativos para cada resposta, definiu-se a matriz inicial de dependência ( $\tau$ ) do modelo, se a variabilidade é única em todos os monitoramentos (Z0) ou se possui variabilidade própria (Z1), conforme indicado pela Figura 8 nos gráficos (a), (b) e (c).

No terceiro passo, após definir a matriz inicial de variabilidade (Z0 ou Z1), adicionou-se uma de cada vez as matrizes de dependência espacial (Inverso da Distância), temporal e trajetória. As matrizes estão exemplificadas na Figura 8, indicadas nos gráficos das letras (d) até (h).

Caso significativo mais de uma matriz de dependência, cria-se um novo modelo combinado com a matrix de variabilidade inicial e as demais significativas. O novo modelo é testado com as matrizes combinadas para verificação da significância conjunta, sempre a nível de 5%.

Até então foi estabelecido para cada resposta as covariáveis e as matrizes de dependência, sendo necessário modelar conjuntamente. Com este objetivo, foi necessário estabelecer um valor inicial nas estimativas deste modelo.

As estimativas iniciais dos ( $\beta$ ), ( $\tau$ ) e ( $\rho$ ) para o modelo final foram definidas pelo modelo multivariado combinado 2 a 2 entre todas as respostas, totalizando 28 combinações. Dessa forma, foi calculado a média dos ( $\beta$ ) e ( $\tau$ ) para cada resposta e colocados como *input* inicial. A correlação foi apenas extraída de cada modelo.

Os modelos resultantes para cada resposta são mostradas abaixo nas especificações de entrada na linguagem R:

- int od ~ monit \* ano + epoca
- l cf ~ monit \* ano
- ph ~ monit + epoca
- dbo ~ monit + epoca
- ft ~ monit
- nt ~ monit \* ano
- l st ~ monit
- l turb ~ monit + ano \* epoca

Para todas as respostas, há efeito da estação de monitoramento, ou seja, em pelo menos uma das estações de monitoramento é estatisticamente significativo. Em outras palavras, a qualidade do parâmetro não se mantém ao longo do rio.

Para as respostas FT, NT e ST, somente há efeito de monitoramento. Para PH e DBO, há efeito aditivo de época do ano. Nas respostas CF e OD, há efeito de interação com o ano e a estação de monitoramento, ou seja, dependendo do ano com a estação de monitoramento a resposta aumenta ou diminui e a Turb tem interação com o ano e a época mais efeito aditivo da estação de monitoramento. Além do mais cada uma das resposta também é apresentado as matrizes de significativas de variabilidade por cada resposta no referido Quadro 5.

Nos modelos DBO, FT e NT, somente a matriz Z1 foi significativa, na prática representa que cada estação de monitoramento possui sua própria variabilidade, isto

é, um indicativo que existe alguma variável não identificada, que está causando esta oscilação, ou naturalmente existe uma variabilidade intrínseca de cada estação de monitoramento.

Nos modelos para ST e Turb foi detectado que a variabilidade é referente a uma variabilidade única ( $z_0$ ) mais uma variabilidade ao decorrer na trajetória das coletas ao longo do tempo, independente da estação de monitoramento, em outras palavras, está refletindo alguma variabilidade ambiental ou industrial que afeta todo o rio.

Nos modelos de OD, CF e PH, também há efeito das trajetória das coletas, mas ponderado pelo inverso da diferença em dias, ou seja, a coleta no tempo  $i$  é linearmente dependente no tempo anterior.

As matrizes de monitoramento vizinhança e inverso dias, representados nos gráficos 8 não foram significativas, mas se fossem, significaria que existe alguma variável não identificada está causando a variabilidade ao longo do tempo em cada estação de monitoramento. A matriz de dependência espacial, foi especificada simétrica, mas considerando o fluxo do rio é mais natural pensar que a influência é uma triangular superior.

Quadro 5 – Matrizes Significativas por Resposta

	OD	CF	PH	DBO	FT	NT	ST	Turb
Z0		X	X				X	X
Z1	X			X	X	X		
Espacial								
Monitoramento Vizinhança								
Monitoramento Inverso Dias								
Trajetória Coleta							X	X
Trajetória Inverso Dias	X	X	X					

Como desdobramento dos resultados do modelo, primeiramente verifica-se a convergência do modelo na Figura 26. Nas duas janelas gráficas inferiores os  $\beta$  e  $\tau$ , na respectivo lado direito e esquerdo, convergem para o 0, significando que o modelo encontrou as estimativas mais adequadas.

Para validação do ajuste do modelo produziu-se o Gráfico 27 para todas as respostas, comparando os resíduos de *pearson* padronizados com os valores preditos, e a linha vermelha representa um ajuste suave de tendência.

Apesar dos resíduos estarem assimétricos, na maioria dos parâmetros, como no caso da DBO, em que a faixa de variação do eixo y é de -2 a 6, não há nenhum tendência de queda ou subida na linha vermelha, se mantendo constante no valor 0, em que o valor predito e o observado são iguais.

Antes de mostrar os resultados finais da predição, optou-se por adicionar efeitos

não paramétrico como *splines* na covariável ano e também tentou-se adicionar termos da harmonia de *Fourier* que captura a variação sazonal, mas em ambos os casos não houve ganho informativo.

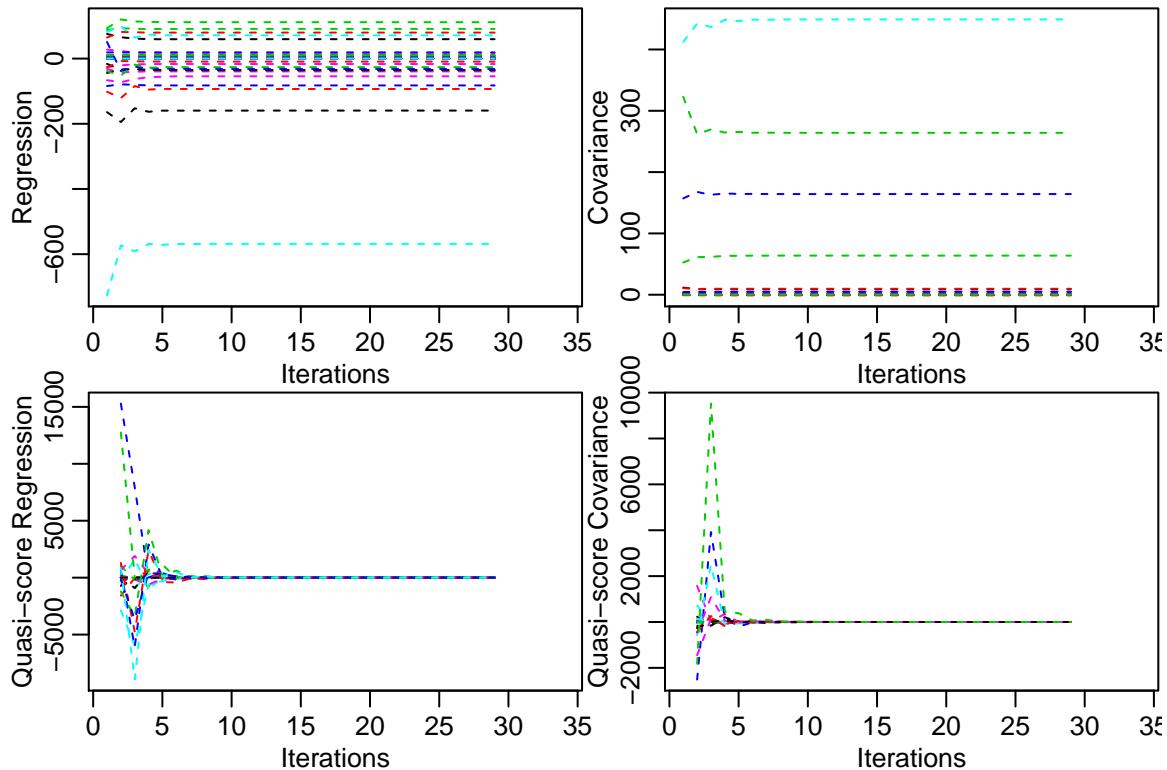


Figura 26 – Convergência Modelo

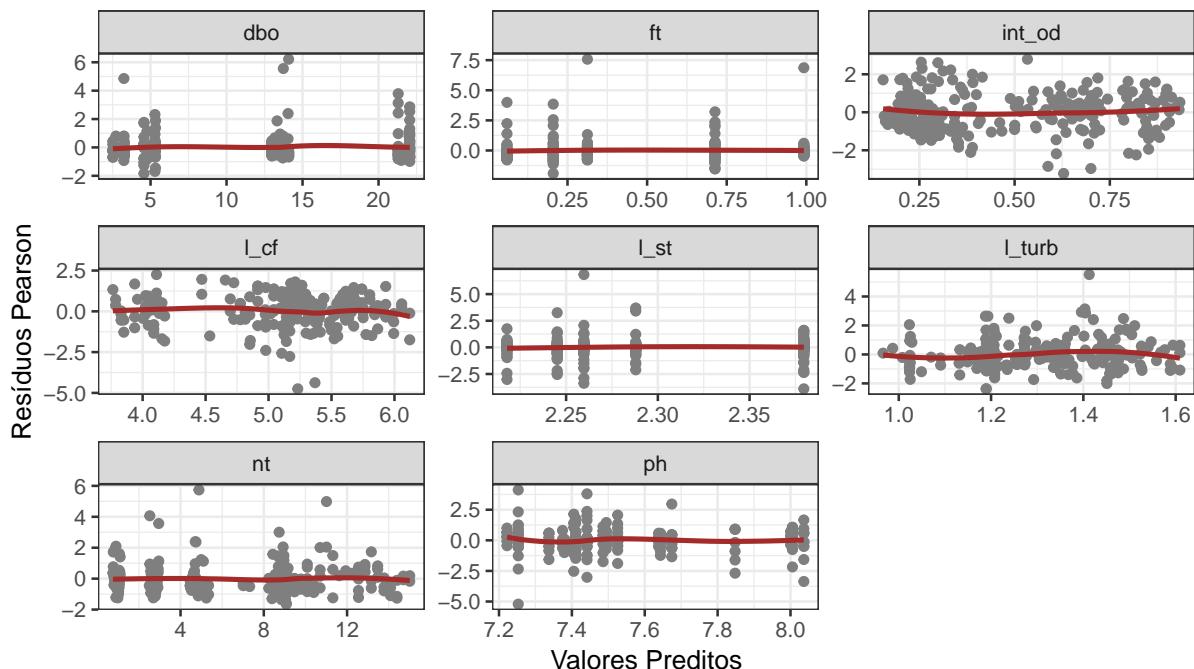


Figura 27 – Avaliação dos Pressupostos do Modelo

Para exposição dos resultados do modelo, optou-se por realizar bandas de previsões para cada uma das respostas na escala da transformação. No presente trabalho está exemplificado na resposta OD (IntOD), e DBO presente nas Figuras 28, 29. O gráfico do DBO foi limitado na escala de 0 a 60, para destaque das bandas de previsões.

A discussão das duas respostas será comparada com Figura 2, que apresenta o comportamento das respostas conforme a influência do efluente não tratado no rio.

Levando em consideração a comparação, o DBO aumenta à medida que prossegue o monitoramento no Rio para todas as épocas. O OD em contrapartida, só diminui à medida que prossegue o monitoramento. Portanto, pode-se dizer que o rio, considerando especificadamente a figura de referência, não está sendo capaz de se restabelecer.

Além disso com o modelo multivariado é estimado a correlação entre as respostas, descontando os efeitos das covariáveis para cada resposta e as matrizes de variabilidade, também pode ser nomeada como a matriz de correlação residual das respostas. No Gráfico 30, é destacado as correlações significativas que são positivas em azul e negativas em vermelho. Nota-se que a ST está altamente correlacionado com a Turb, e também com os nutrientes (FT e NT) altamente correlacionados. O OD é o parâmetro em que a maioria das correlações são inversas, como o DBO e os nutrientes. No *link* (MCGLM) é realizado a análise para todas as respostas, desde as análises descritivas até as bandas de previsão.

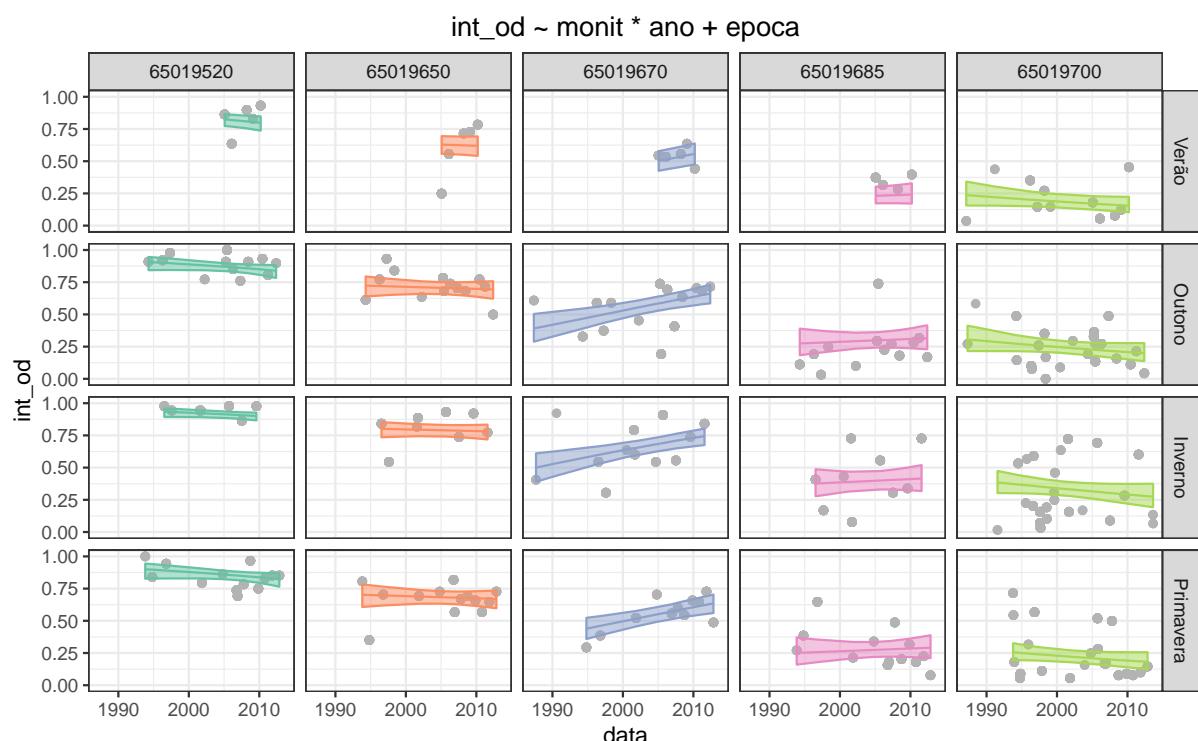


Figura 28 – Bandas de Previsão para OD

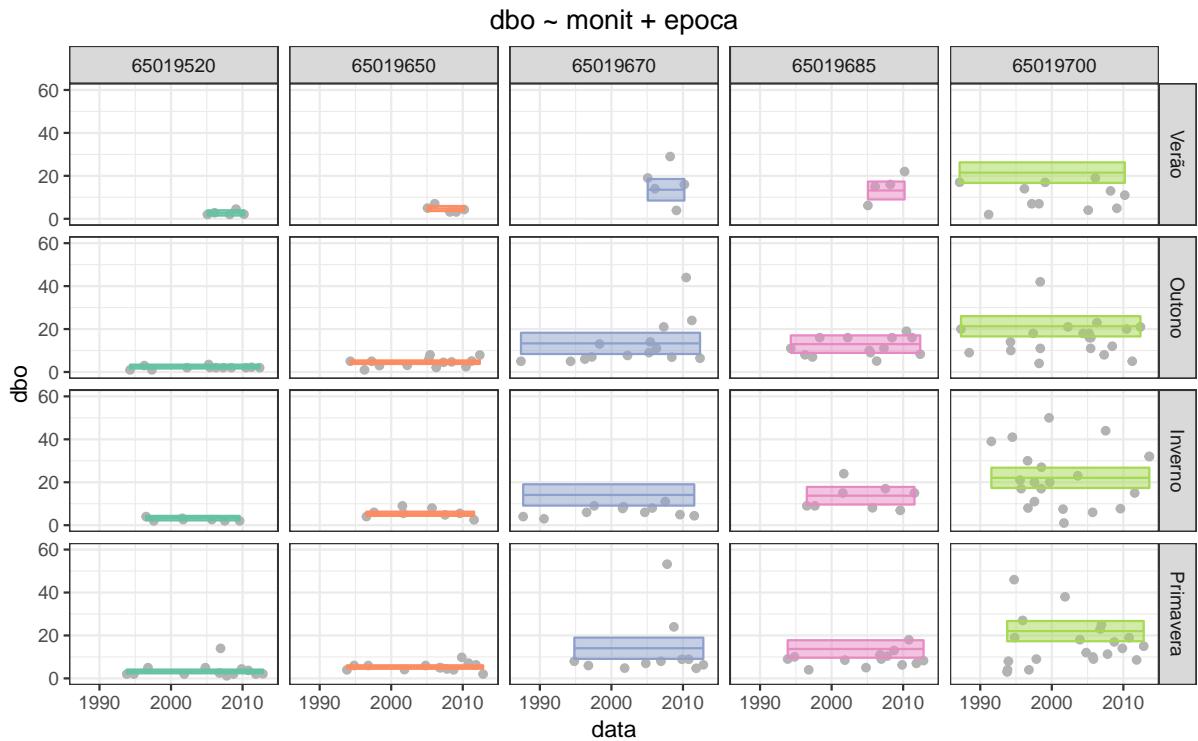


Figura 29 – Bandas de Predição para DBO

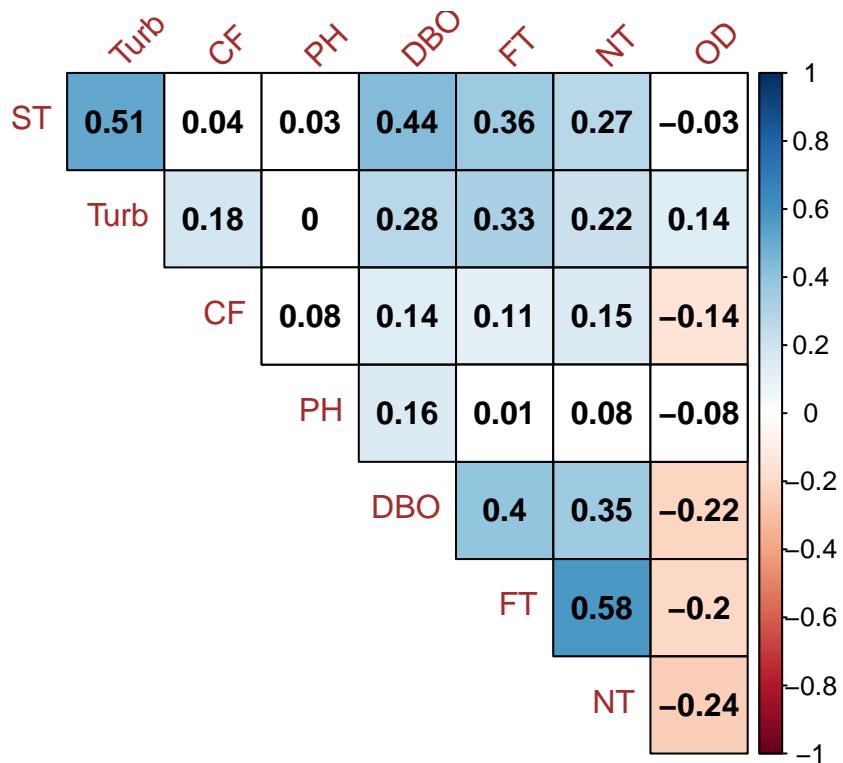


Figura 30 – Correlação das respostas



## 6 Considerações Finais

A consolidação deste trabalho passou por etapas fundamentais, sendo a primeira delas a mais importante, assegurar validade nos dados. Antes mesmo de discutir a qualidade da água, se fez necessário e vital, validar as informações coletadas, por meio dos limites de aceitabilidade de cada parâmetro, construindo um banco de dados consistente.

Havendo o banco de dados consistente, foi inicialmente analisado o IQA para compreensão da variabilidade de cada parâmetro que o compõe e identificado a Turb e CF como determinantes no resultado. O IQA também necessita ser revisto nas equações que convertem o parâmetro para a escala de 0 a 100, já que muitos não condizem com as curvas descritas na teoria.

De outro modo a visualização de cada parâmetro, em diferentes escalas (concentração, carga, transformação do IQA) ao longo do tempo por bacia, acrescido da média e do polinômio local *loess* permite comparar se o parâmetros mudou ao longo do tempo ou não.

Na tentativa de agrupar parâmetros que maximizem a variabilidade de cada bacia e o agrupamento por características, a utilização dos métodos de PCA e *t-SNE*, não produziram resultados discriminativos em todas as escalas. Uma possível explicação é o fato de que não está, nas análises incorporando o efeito de temporal, a hierarquia entre rios e bacia e a dependência espacial dos pontos de monitoramento.

A metodologia MCGLM se mostrou muito flexível e totalmente informativa em problemas de qualidade da água, no sentido de contribuir com a discussão da sustentabilidade do rio, já que é possível testar dependência temporal por estação de monitoramento, ao longo de todo o rio, dependência espacial e coletas não periódicas. Inclusive a metodologia pode ser aplicada no rio Passaúna nas oito estações de monitoramento, pela existência de razoáveis coletas que compreendem o período totalizando 800 dados.

Como potenciais trabalhos futuros, os dados disponíveis também podem ser usados, para cruzamento de informações de saúde por municípios, como mortalidade infantil, sendo necessário alguma operação de geoprocessamento, para atribuir um valor de qualidade da água no município. Outro estudo seria estimar do preço de um rio, por meio das outorgas realizadas.

Outro cruzamento de interesse é comparar as atribuições do enquadramento para cada rio e o impacto em cada parâmetro da qualidade da água. Outro estudo

de interesse é atribuir o uso de solo para cada coleta, por meio de imagens digitais de satélite, a fim de verificar se a transição do espaço rural para o urbano afetou a qualidade da água. Como possível prosseguimento inicial, seria sistematizar as análises à medida que um novo dado é coletado.

Uma discussão para trabalhos futuros é qual a escala mais adequada para analisar dados da qualidade da água: concentração, carga ou transformação IQA. Neste trabalho foram geradas as análise em diferentes escalas para promover esta discussão com os especialistas.

Por fim acredito que este material sirva para direcionar o estudante de Estatística no contexto de recursos hídricos, parâmetros de qualidade da água e métodos que lidem com a característica do dado. Ao mesmo tempo, direcionar o estudante de recursos hídricos, com diferentes abordagens de técnicas estatísticas, em que não foram encontradas na literatura pelo autor.

# REFERÊNCIAS

ÁGUAS-(ANA), A. N. D. *Conjuntura dos recursos hídricos no Brasil*. [S.l.]: ANA Brasília, 2017. Citado 2 vezes nas páginas [23](#) e [24](#).

AGUASPARANA. Qualidade das Águas da bacia hidrográfica do alto iguaçu. v. 1, 2011. Citado 5 vezes nas páginas [23](#), [24](#), [25](#), [26](#) e [27](#).

ANDRADE, L. N. de. Autodepuração dos corpos d'água. *Revista da Biologia*, v. 5, p. 16–19, 2018. Citado na página [31](#).

ARRUDA, N. M. B. d. Avaliação de variáveis de qualidade de água dos reservatórios das usinas hidrelétricas de foz do areia, segredo e caxias, como instrumento de gestão de bacias hidrográficas. UFPR, 2014. Citado na página [33](#).

BONAT, W. H.; JØRGENSEN, B. Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 65, n. 5, p. 649–675, 2016. Citado na página [38](#).

BOX, G. E.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 211–252, 1964. Citado na página [37](#).

BRASIL. Lei no 9.433, de 8 de janeiro de 1997. *Diário Oficial da União*, jan 1997. Citado na página [17](#).

BUSSAB, W. d. O.; MORETTIN, P. A. *Estatística básica*. [S.l.]: Saraiva, 2010. Citado na página [36](#).

CANDIDO, L. d. O. *Avaliação do índice de qualidade das águas na área urbana da bacia hidrográfica do Rio Mourão, Paraná*. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2015. Citado na página [33](#).

CHENG, J.; KARAMBELKAR, B.; XIE, Y. *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. [S.l.]. R package version 2.0.0.9000. Disponível em: <<http://rstudio.github.io/leaflet/>>. Citado na página [43](#).

FARAWAY, J. J. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. [S.l.]: CRC press, 2016. v. 124. Citado na página [36](#).

FREITAS, M. d.; SANTOS, A. H. M. Importância da água e da informação hidrológica. *O estado das águas no Brasil*, ANEELSIRMMASRH Brasília, v. 2, 1999. Citado na página [18](#).

GILBERTO, N. et al. Índice de qualidade de água em microbacia sob uso agrícola e urbano. *Scientia Agricola*, Escola Superior de Agricultura, 2002. Citado na página [18](#).

GONÇALVES, C. S. et al. Water quality in a watershed located in a tobacco producing area. *Revista Brasileira de Engenharia Agrícola e Ambiental*, SciELO Brasil, v. 9, n. 3, p. 391–399, 2005. Citado na página [18](#).

GRUNITZKI, R. et al. Ferramenta web para determinação do índice de qualidade de água a partir da reestruturação das equações que descrevem as curvas dos indicadores de qualidade. *XX Simpósio Brasileiro de Recursos Hídricos, Bento Gonçalves, RS*, 2013. Citado na página 17.

GUEDES, H. A. S. et al. Modelagem da qualidade da água e da capacidade de autodepuração do rio pomba. Universidade Federal de Viçosa, 2009. Citado na página 17.

JAMES, G. et al. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112. Citado na página 36.

MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. Nov, p. 2579–2605, 2008. Citado na página 37.

MANLY, B. F. M. *Métodos estatísticos multivariados: uma introdução*. [S.l.]: Bookman, 2008. Citado na página 37.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. [S.l.]: Editora UFMG, 2013. Citado na página 37.

QGIS Development Team. *QGIS Geographic Information System*. [S.l.], 2009. Disponível em: <<http://qgis.osgeo.org>>. Citado na página 43.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. Citado na página 43.

SÃO, C. A. d. E. D. Paulo-cetesb. *Relatório de qualidade das águas superficiais no estado de São Paulo*, 2009. Citado na página 25.

SPERLING, M. V. *Introdução à qualidade das águas e ao tratamento de esgotos: princípio do tratamento biológico de águas residuárias*. [S.l.]: Editora UFMG, 2005. v. 3. Citado na página 31.

SPERLING, M. V. *Estudos e modelagem da qualidade da água de rios*. [S.l.]: DESA/UFMG, 2007. Citado 4 vezes nas páginas 23, 24, 25 e 28.

TEODORO, V. L. I. et al. O conceito de bacia hidrográfica e a importância da caracterização morfométrica para o entendimento da dinâmica ambiental local. *Revista ReBrAM*, v. 11, n. 1, p. 137–156, 2015. Citado na página 17.

TRINDADE, A. L. C. Aplicação de técnicas estatísticas para avaliação de dados de monitoramento de qualidade das águas superficiais da porção mineira do rio são francisco. UFMG, 2013. Citado na página 33.

VEIGA, N. Cianobactéria tóxica na água para consumo humano. *Artigo Científico*, 2011. Citado na página 17.

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<http://ggplot2.org>>. Citado na página 43.

WICKHAM, H. et al. *dplyr: A Grammar of Data Manipulation*. [S.l.], 2018. R package version 0.7.5. Disponível em: <<https://CRAN.R-project.org/package=dplyr>>. Citado na página 43.

ZSCHORNACK, T. *Avaliação do impacto da implantação do sistema de esgotamento sanitário na qualidade da água da bacia hidrográfica do rio Cachoeira sob a ótica da saúde ambiental.* [S.l.]: Joinville, Univille, 2016. Citado na página 31.



# Apêndices



## APÊNDICE A – Equações IQA

Tabela 6 – Funções de transformação para o IQA

Parâmetros	Limite Mínimo (>)	Limite Máximo ( $\leq$ )	Equação $q_i$
OD	Concentração de Saturação de OD		$C_s = (14.62 - 0.3898 * \text{Temp} + 0.006969 * \text{Temp}^2 - 0.00005896 * \text{Temp}^3) * (1 - 0.00000228675 * \text{altitude})^{5.167}$
	Percentagem de Saturação		$\text{sat} = 100 * (\text{OD}/C_s)$
	0	50	$3 + (0.34 * \text{sat}) + 0.008095 * (\text{sat}^2) + 1.35252 * 0.00001 * (\text{sat}^3)$
%OD	50	85	$3 - 1.166 * (\text{sat}) + 0.058 * (\text{sat}^2) - 3.803435 * 0.0001 * (\text{sat}^3)$
	85	100	$3 + 3.7745 * (\text{sat})^{0.704889}$
	100	140	$3 + 2.9 * (\text{sat}) - 0.02496 * (\text{sat}^2) + 5.60919 * 0.00001 * (\text{sat}^3)$
	140		$3 + 47$
$\log_{10}(CF)$	0	1	$100 - 33 * \log_{10}(CF)$
	1	5	$100 - 37.2 * \log_{10}(CF) + 3.60743 * (\log_{10}(CF)^2) / 3$
	5		
PH	0	2	$2$
	2	4	$13.6 - 10.6 * \text{PH} + 2.4364 * (\text{PH}^2)$
	4	6.2	$155.5 - 77.36 * \text{PH} + 10.2481 * (\text{PH}^2)$
	6.2	7	$-657.2 + 197.38 * \text{PH} - 12.9167 * (\text{PH}^2)$
	7	8	$-427.8 + 142.05 * \text{PH} - 9.695 * (\text{PH}^2)$
	8	8.5	$216 - 16 * \text{PH}$
	8.5	9	$1415823 * \text{EXP}(-1.1507 * \text{PH})$
	9	10	$50 - 32 * (\text{PH}-9)$
	10	12	$633 - 106.5 * \text{PH} + 4.5 * (\text{PH}^2) / 3$
	12	14	
DBO	0	5	$99.96 * \text{EXP}(-0.1232728 * \text{DBO})$
	5	15	$104.67 - 31.5463 * \log_{10}(\text{DBO})$
	15	30	$4394.91 * C^{-1.99809} / 2$
	30		
NT	0	10	$100 - 8.169 * \text{NT} + 0.3059 * \text{NT}^2$
	10	60	$101.9 - 23.1023 * \log_{10}(\text{NT})$
	60	100	$159.3148 * \text{EXP}(-0.0512842 * \text{NT}) / 2$
	100		
FT	0	1	$99 * \text{EXP}(-0.91629 * \text{FT})$
	1	5	$57.6 - 20.178 * \text{FT} + 2.1326 * (\text{FT}^2)$
	5	10	$19.8 * \text{EXP}(-0.13544 * \text{FT}) / 5$
	10		
$\Delta \text{Temp}$	94 (assumido o valor constante de 94 pela CETESB, por se considerar que, nas condições brasileiras, a temperatura dos corpos d'água não se afasta da temperatura de equilíbrio.		
Turb	0	25	$100.17 - 2.67 * \text{Turb} + 0.03775 * \text{Turb}^2$
	25	100	$84.76 * \text{EXP}(-0.016206 * \text{Turb}) / 5$
	100		
ST	0	150	$79.75 + 0.166 * \text{ST} - 0.001088 * \text{ST}^2$
	150	500	$101.67 - 0.13917 * \text{ST} / 32$
	500		

# APÊNDICE B – Curvas de Desejabilidade do IQA

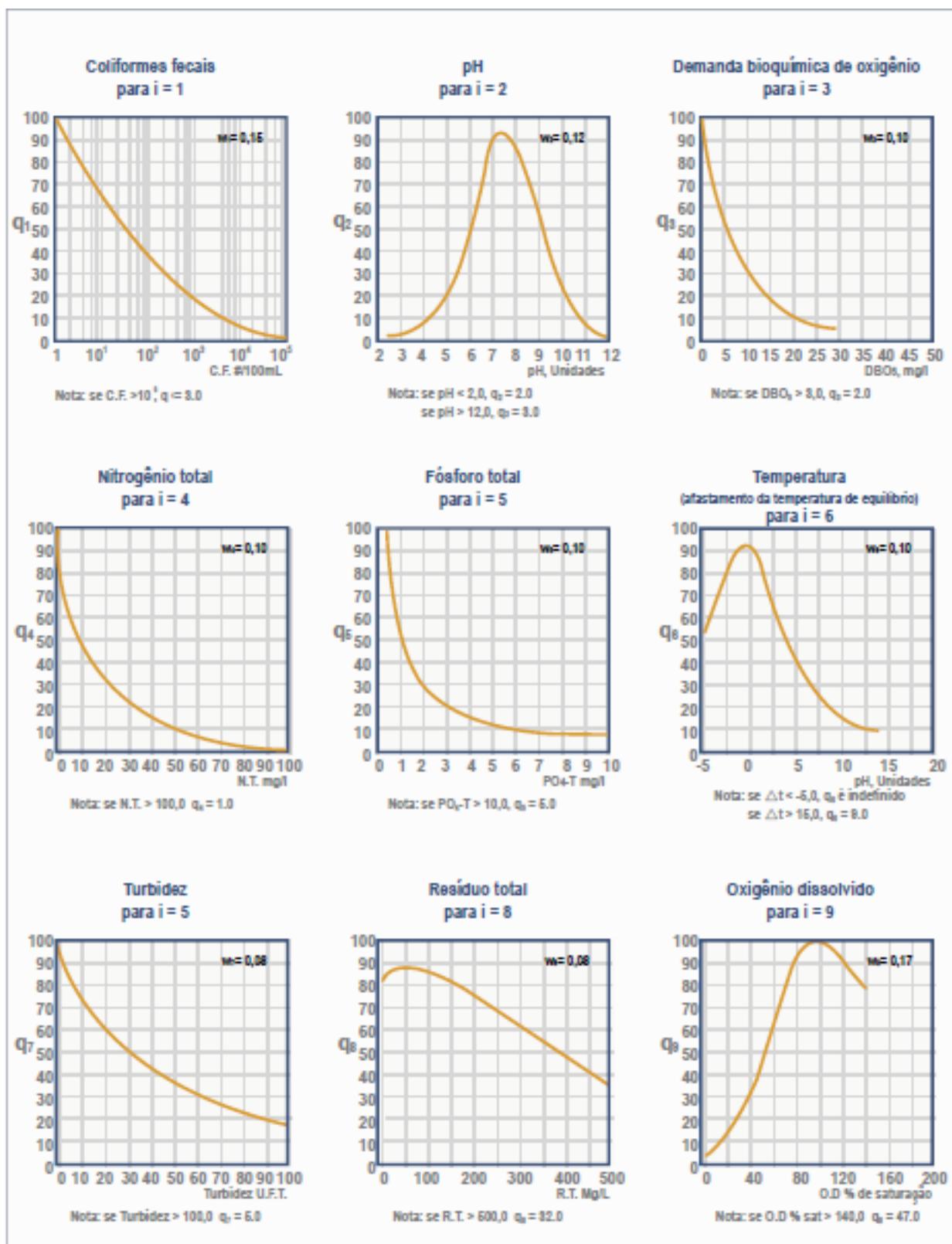


Figura 31 – Curvas médias de variação dos parâmetros de qualidade das águas para o cálculo do IQA