

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA**

**“Implementação de Ponderação Baseada em IQAs na
Análise Exploratória (HCA e PCA) para Estudos de
Qualidade de Água”**

Por:

Glauciene Paula de Souza Marcone

**Dissertação submetida ao Departamento de Química como
parte dos requisitos para obtenção do título de Mestre em
Ciências**

1º Orientador: Mário César Ugulino de Araújo

2º Orientador: Wallace Duarte Fragoso

João Pessoa – Agosto/ 2006

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA

**“Implementação de Ponderação Baseada em IQAs em
Análise Exploratória (HCA e PCA) para Estudos de
Qualidade de Água”**

Por

Glauciene Paula de Souza Marccone

Dissertação submetida ao programa de pós-graduação em química, da Universidade Federal da Paraíba, como requisito parcial à obtenção do título de Mestre em Química, área de concentração em “Química Analítica”.

1º Orientador: Prof. Dr. Mário César Ugulino de Araújo

2º Orientador: Prof. Dr. Wallace Duarte Fragoso

João Pessoa – Agosto/ 2006

M 321 i

Marcone, Glauciene Paula de Souza

Implementação de ponderação baseada em IQAs na análise exploratória (HCA e PCA) para estudos de qualidade de água/ Glauciene Paula de Souza Marcone. – João Pessoa, 2006.

47p.: il.

Orientadores: Mário César Ugulino de Araújo e Wallace Duarte Fragoso

Dissertação (mestrado) UFPB/CCEN

1. Água-Qualidade 2. HCA 3. PCA 4. IQA

UFPB/BC.

CDU:556 (043)

**“Implementação de Ponderação Baseada em
IQAs na Análise Exploratória (HCA e PCA)
para Estudos de Qualidade de Água”**

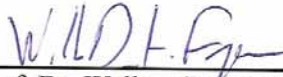
Glauciene Paula de Souza Marcone

Aprovada em 31 de agosto de 2006

BANCA EXAMINADORA:



Prof. Dr. Mário César Ugulino de Araújo
Orientador



Prof. Dr. Wallace Duarte Fragoso
2º. Orientador



Profa. Dra. Marta Maria Menezes Bezerra Duarte
Examinadora



Profa. Dra. Teresa Cristina Bezerra Saldanha
Examinadora

*Ao Deus Criador e conhecedor de todas as
nossas limitações, mas que pelo amor infinito que
sente por nós, nos conduz sempre em frente como
um Pai bondoso.*

*Aos meus irmãos, ao meu sobrinho Talles e aos
meus pais, presentes de Deus em minha vida.*

Eu dedico.

*" Que o senhor te abençoe e te guarde;
que o senhor faça resplandecer seu rosto sobre ti
e tenha misericórdia de ti
que o senhor sobre ti levante o seu rosto
e te dê a paz ".*

Números 6, 24-26

AGRADECIMENTOS

- *Agradeço a Deus, pelas graças concedidas e amor incondicional.*
- *Ao Prof. Dr. Mário César, pela orientação deste trabalho, pela confiança e valorização de minhas potencialidades e principalmente pela amizade construída.*
- *Ao Prof. Dr. Wallace Fragoso pela indispensável co-orientação, pela amizade, confiança e paciência.*
- *Às Profas. Dras. Tereza Saldanha e Vânia Medeiros, pela amizade, delicadeza e apoio indispensáveis.*
- *À Profa. Dra. Valdinete Lins Silva pela disponibilidade e apoio sempre solícitos.*
- *À CAPES, pela concessão da bolsa.*
- *À coordenação de Pós-Graduação em Química e em especial à Profa. Dra. Regiane de Cássia, pela atenção e amizade.*
- *Especialmente aos meus amigos tão pacientes e necessários na fase final deste trabalho, Alessandra Félix, Edilene Oliveira, Sueny Freitas e Simone Simões, pela ajuda indispensável.*
- *A todos os amigos da família LAQA, pela convivência amorosa e amizades cultivadas e de forma especial à Amália Gama, Sófacles Soares, Francisco Silva e Antônio Cícero pelo desprendimento na fase inicial deste trabalho e a Ricardo Alexandre e Sérgio Ricardo pela amizade e co-orientação no período de iniciação científica.*
- *Aos meus irmãos, Netinho, Siele, Leninha, Márcia e Yure, ao meu tio, Marcos e à minha preciosa mãe Cássia, por estarem sempre ao meu lado. E ao meu amado sobrinho Talles, “coisinha fofa da titia”.*
- *Ao meu pai, José Paulo (in memória) pela lembrança sempre amável.*
- *À família catequese, pelo amor que me ofereceram, pelos ensinamentos preciosos e pela compreensão da ausência muitas vezes sentida.*

SUMÁRIO

ÍNDICE DE FIGURAS	ix
ÍNDICE DE TABELAS	x
LISTA DE SIGLAS	xi
RESUMO	xii
ABSTRACT	xiii
1 – INTRODUÇÃO	1
1.1 – Controle de qualidade dos sistemas hídricos	3
1.2 – Principais parâmetros de qualidade de água	5
1.2.1 – Parâmetros físicos	5
1.2.2 – Parâmetros químicos	6
1.2.3 – Parâmetro biológico	8
1.3 – Índices de qualidade de água (IQAs)	8
1.3.1 – Formulações de Índices de qualidade de água	10
1.3.1.1 – Escolha dos parâmetros (variáveis) que representem a qualidade da água	10
1.3.1.2 – Geração da função de sub-índice de qualidade (qi): transformação dos parâmetros de dimensões e de unidades diferentes para uma mesma escala	10
1.3.1.3 – Atribuição de pesos a cada parâmetro	11
1.3.1.4 – Agregação dos sub-índices para a formulação do IQA final	12
1.3.2 – Aplicações dos índices de qualidade de água	12
1.3.2.1 – IQAs para avaliação de águas para consumo humano	13
1.3.2.2 – IQAs para a proteção da vida aquática	15
1.4 – Quimiometria	17
1.4.1 – Tratamento quimiométrico e estatístico dos dados	17
1.4.1.1 – Pré-processamento dos dados	17
1.4.1.2 – Análise hierárquica de agrupamentos – HCA	18
1.5.1.3 – Análise de componentes principais – PCA	19
1.5 – Análise exploratória em qualidade de água e a necessidade de ponderação das variáveis	20
1.6 – OBJETIVOS	21
1.6.1 – Objetivo geral	21
1.6.2 – Objetivos específicos	21

2 – MATERIAIS E MÉTODO -----	23
2.1 – Banco de dados -----	23
2.1.1 – Parâmetros que compõem os dados -----	23
2.2 – Softwares -----	24
2.3 – Metodologia proposta -----	24
2.3.1 – Análise exploratória: HCA e PCA ponderadas -----	24
3 – RESULTADOS E DISCUSSÃO -----	26
3.1 – Exame preliminar dos dados -----	26
3.2 – Análise quimiométrica e estatística dos dados -----	27
3.2.1 – Análise hierárquica de agrupamentos ponderada (HCA ponderada) -----	27
3.2.2 – Análise de componentes principais ponderada (PCA ponderada) -----	31
4 - CONCLUSÕES E PROPOSTAS FUTURAS -----	39
5 - REFERÊNCIAS BIBLIOGRÁFICAS -----	41
ANEXOS -----	46

ÍNDICE DE FIGURAS

Figura 1.1 – Função de qualidade para o parâmetro turbidez -----	11
Figura 1.2 – Representação das duas primeiras PCs em que os círculos representam as amostras-----	19
Figura 1.3 – Representação dos escores e as projeções na direção da componente principal (a) e os pesos dados pelos co-senos dos ângulos entre cada variável e a componente principal (b) -----	20
Figura 2.1 – Esquema do pré-processamento aplicado a matriz de dados X com m linhas e n colunas. Após o autoescalonamento é gerada a matriz E, que é ponderada multiplicando-a pela matriz de pesos (matriz diagonal), resultando na matriz ponderada, E_p a ser usada na HCA e PCA -----	25
Figura 3.1 – Dendrogramas dos dados sem rotulagem usando ponderação dada pelo $IQA_{BÁSCARON}$ (a) e ponderação dada pelo IQA_{CETESB} (b) -----	28
Figura 3.2 – Dendrogramas dos dados com rotulagem usando ponderação dada pelo $IQA_{BÁSCARON}$ (a) e ponderação dada pelo IQA_{CETESB} (b) -----	29
Figura 3.3 – Gráfico dos escores para as duas primeiras PCs sem ponderação (■ bom; ■ razoável; ■ ruim; ■ péssima) -----	32
Figura 3.4 – Gráfico dos escores para as duas primeiras PCs ponderadas segundo o $IQA_{BÁSCARON}$ (■ bom; ■ razoável; ■ ruim; ■ péssima) -----	33
Figura 3.5 – Gráfico da variância explicada <i>versus</i> o número de componentes principais ponderadas segundo $IQA_{BÁSCARON}$ -----	34
Figura 3.6 – Gráfico dos pesos das variáveis nas componentes principais PC1 <i>versus</i> PC2, ponderadas segundo o $IQA_{BÁSCARON}$, onde: OD (Oxigênio Dissolvido); CF (Coliforme Fecal); pH; DBO (Demanda Bioquímica de Oxigênio); NT (Nitrogênio Total); PT (Fósforo Total); Tu (Turbidez); ST (Sólidos Totais); T (Temperatura) -----	35
Figura 3.7 – Gráfico dos escores para as duas primeiras PCs ponderadas segundo o IQA_{CETESB} (■ bom; ■ razoável; ■ ruim; ■ péssima) -----	36
Figura 3.8 – Gráfico da variância explicada <i>versus</i> o número de componentes principais ponderadas segundo IQA_{CETESB} -----	37
Figura 3.9 – Gráfico dos pesos das variáveis nas componentes principais PC1 <i>versus</i> PC2, ponderadas segundo o IQA_{CETESB} , onde: OD (Oxigênio Dissolvido); CF (Coliforme Fecal); pH; DBO (Demanda Bioquímica de Oxigênio); NT (Nitrogênio Total); PT (Fósforo Total); Tu (Turbidez); ST (Sólidos Totais); T (Temperatura) -----	37

ÍNDICE DE TABELAS

Tabela 2.1 – Parâmetros medidos no monitoramento -----	23
Tabela 2.2 - Parâmetros com seus respectivos pesos. Os pesos do IQA _{BÁSCARON} variam de 1-4, enquanto os pesos do IQA _{CETESB} têm somatório igual a 1 -----	25
Tabela 3.1 - Classificação das amostras do banco de dados segundo o IAP -----	26
Tabela 3.2 - Incertezas associadas a variabilidade dos dados do IAP -----	27
Tabela 3.3 - Classificação das amostras referentes ao dendrograma rotulado -----	30

LISTA DE ABREVIATURAS E SIGLAS

IQA – Índice de Qualidade de Água

PCA – Análise de Componente Principal (*Principal Component analysis*)

HCA – Análise Hierárquica de Agrupamentos (*Hierarchical Cluster Analysis*)

WQI – Water Quality Index

CETESB – Companhia Estadual de Tecnologia de Saneamento Básico do Estado de São Paulo

FNS/US – Fundação Nacional de Saúde (*National Sanitation Foundation – Units States/NSF-US*)

UNICEF – Fundo das Nações Unidas para a Infância

CONAMA – Conselho Nacional do Meio Ambiente

IAP – Instituto Ambiental do Paraná

OMS – Organização Mundial da Saúde

ONU – Organização das Nações Unidas

PNRH – Política Nacional de Recursos Hídricos

OD – Oxigênio Dissolvido

DBO₅ ou DBO – Demanda Bioquímica de Oxigênio a 20°C incubada por 5 dias

RESUMO

Técnicas quimiométricas de análise exploratória são aplicadas com frequência ao estudo de sistemas aquáticos, tendo em vista o comportamento multivariado dos mesmos. Os Índices de Qualidade de Água (IQAs) são uma forma simples de indicar a qualidade de sistemas hídricos. Neste trabalho é proposta a implementação da ponderação proveniente dos IQAs na metodologia de avaliação de qualidade de águas por meio da análise hierárquica de agrupamentos (*Hierarchical Cluster Analysis* – HCA) e da análise de componentes principais (*Principal Component Analysis* – PCA). Ambas, HCA e PCA foram ponderadas com o intuito de refletirem a importância relativa que os parâmetros monitorados têm no conceito de qualidade. Para a ponderação das variáveis que compõem os IQAs foram utilizados os pesos de dois IQAs distintos: um baseado em somatório de termos relativos às variáveis que o compõe, cujos pesos variam de 1 a 4, e outro baseado em produtório relativo a estas mesmas variáveis, que tem seus pesos variando de 0 a 1. O conjunto de dados é composto por 568 amostras de águas de vinte rios das bacias de Ribeira, Iguaçu, Cinzas, Irai e Piquiri, localizadas no Estado do Paraná. As amostras foram coletadas no período de 1987 a 2001 e tiveram a sua qualidade classificada, segundo os critérios do Instituto Ambiental do Paraná (IAP), como: boa, razoável, ruim e péssima. A HCA ponderada foi utilizada para se estudar grupos de amostras conforme suas similaridades. Foram observados dois grandes grupos: G1 que agrega águas de melhor qualidade e G2, contendo as águas de pior qualidade. Essa separação foi evidenciada melhor quando os pesos do IQA_{BASCARON} foram usados. Com a PCA ponderada foi constatado que os escores da primeira componente principal (PC1) concordam com a variação de qualidade de água dada para o sistema em análise, segundo a classificação dada pelo IAP. Melhores resultados foram observados usando os pesos do IQA_{BASCARON}, cuja PC1 porta 62% da variância explicada. Foi visto também que as relações entre os pesos das variáveis originais em PC1 e PC2 correspondem às relações esperadas nos sistemas hídricos.

Palavras chave: HCA, PCA, ponderação, IQA, qualidade de água

ABSTRACT

Quimiometric techniques of exploratory analysis are frequently applied on the study of aquatic systems having in mind the multivariate behavior of the same. The Water Quality Indices-WQIs (índice de qualidade de águas – IQAs, in Portuguese) represent a simple form of indicating the hidric systems quality. In this work the weighting from the Water Quality Indices is implemented in the methodology of waters quality evaluation using Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA). Both, HCA and PCA, were weighted in order to point out the relative importance of the parameters for the quality concept. For the weighting of the variables that composes the IQAs, It was used the weights of two distinct IQAs. The IQA_{CETESB} was based on the sum of terms relatives to these variables, whose weights are in the range from 1 to 4; the $IQA_{BÁSCARON}$ was based on product, whose weights varies from 0 to 1. The data set used includes 568 water samples from twenty rivers distributed among the basins of Ribeira, Iguaçu, Cinzas, Iraí and Piquiri, located in the State of Paraná. The samples were collected in the period from 1987 to 2001 and its quality was classified according to the criteria of the *Instituto Ambiental of Paraná* (IAP) as good, reasonable, bad and very bad. First, the weighted HCA was used in order to study sample groups according to their similarities. Two great groups were observed: G1, including the best quality waters and G2, with the worst quality waters. This separation was more evident when the weights of the $IQA_{BÁSCARON}$ were used. The weighted PCA showed that the scores of the first principal component agree with the water quality variation for the analyzed aquatic system according to the classification by the IQA_{CETESB} . The best results of the weighted PCA were observed using the weights of the $IQA_{BÁSCARON}$ whose PC1 carries 62% of the explained variance. In addition the relation between the loadings of PC1 and PC2 agree with the expected behavior of the aquatic system.

Key works: HCA, PCA, weighting, WQI, water quality

1 - INTRODUÇÃO

Atualmente fala-se muito sobre a preservação ambiental, para que as futuras gerações possam viver de forma sustentável e em harmonia com os ecossistemas da Terra: sistemas aquáticos, ar, solo e biota. Na prática, são constatados altos índices de poluição dos compartimentos ambientais causados pelos mais variados aportes, como por exemplo, poluição causada pela ação antrópica.

Desde a Antigüidade, o homem busca sua evolução sócio-econômica e tecnológica. Os tempos atuais são marcados por uma crescente industrialização, devido à procura por bens manufaturados e por um modo de vida mais confortável. Neste contexto, são gerados, todos os dias, milhares de toneladas de resíduos de produção ou de consumo, provenientes das indústrias tais como, os gases tóxicos que são lançados na atmosfera, os efluentes contaminados por metais pesados, compostos orgânicos biodegradáveis ou não, dentre outros. Em geral, essas substâncias são lançadas nos compartimentos ambientais sem nenhum tratamento prévio.

Estes fatores são os principais responsáveis por problemas gravíssimos de saúde pública, como epidemias que assolam a humanidade desde a Idade Média (400 a. 1400 d.C). Como exemplo, pode-se citar a peste bubônica, que matou cerca de 25 milhões de pessoas na Europa, e foi causada por falta de saneamento nas cidades ou pelo despejo, nos rios europeus, de efluentes de esgotos domésticos^[1].

Além desses e de outros aspectos qualitativos, a problemática da água envolve fatores quantitativos.

Analisando a quantidade de água disponível no planeta, é sabido que dois terços da superfície da Terra ou 71% são cobertos por água, sendo 97% deste total constituído de água salgada, e apenas 3% de água doce. Desta porção de água doce 2% está disponível nos rios, lagos (reservatórios superficiais) e aquíferos (reservatórios subterrâneos), 1% fica armazenado nas geleiras e uma quantidade mínima (0,0005%) está na forma de vapor d'água espalhada na atmosfera^[2].

Além de ser um bem econômico, a água é indispensável às atividades humanas, no que diz respeito à produção de alimentos (agricultura e pecuária), aos processos industriais em geral e ao abastecimento público^[3].

O problema da escassez de água doce de qualidade para consumo humano é global. Existem políticas mundiais para melhorar a distribuição e diminuir a poluição das fontes ainda disponíveis deste bem indispensável à vida e ao desenvolvimento sócio-econômico das nações. A Organização Mundial de Saúde (OMS) tem como meta reduzir pela metade, até 2015, os números atuais de pessoas que não dispõem de água potável e de saneamento, estimados respectivamente em 1 bilhão e 2,4 bilhões de habitantes. Estes problemas são também fonte de preocupação da Organização das Nações Unidas (ONU) cujo secretário geral, Kofi Annan, tem afirmado que “no ritmo que vai, a água se tornará cada vez mais uma fonte de tensão e fruto de feroz competição entre as nações”^[1].

Atualmente, estes problemas ainda permanecem, porque falta saneamento básico e a distribuição de água potável é precária para boa parte da população mundial. Isto acarreta o aumento do número de doenças de veiculação hídrica, como por exemplo, febre tifóide, cólera, salmonelose, shigelose e outras gastroenterites, poliomielite, hepatite A, verminoses, amebíase e giardíase, que causam surtos epidêmicos, mortalidade infantil e o agravamento dos vários problemas de saúde pública^[4].

No ano de 2003, o Fundo das Nações Unidas para a Infância (UNICEF, do inglês, *United Nations International Children's Emergency Fund*) afirmou que, cerca de 10.000.000 de pessoas morrem por ano em todo o mundo em virtude de doenças causadas pela contaminação das águas. Estas mortes acontecem porque menos da metade da população mundial (35%) tem acesso à água potável, ou seja, um bilhão e 200 milhões de pessoas. Também, segundo o UNICEF, cerca de 43% (um bilhão e 800 milhões) dos habitantes de todo o mundo não dispõem de esgotamento sanitário^[3].

Conforme censo realizado em 2000 pelo Instituto Brasileiro de Geografia e Estatística (IBGE), 76,1% da população brasileira é, em média, atendida por rede de abastecimento de água. Todavia, a rede de esgoto atende, apenas, cerca de 40% da população e desta parcela de esgoto coletado, somente 34,5% são tratados, o que tem provocado a grave crise de saúde pública verificada há décadas no país^[5]
[4]

A heterogeneidade da distribuição de água no Brasil (80% está na região amazônica e os outros 20% nas demais regiões) e o aporte de esgotos domésticos e industriais nos sistemas hídricos são fatores que levam à escassez de água potável

o que aumenta a preocupação com os problemas ambientais relativos qualidade de águas.

Neste sentido, é imprescindível a implementação de políticas públicas para a gestão de bacias hidrográficas, com a finalidade de preservar as nascentes dos rios, bem como a qualidade da água durante o seu curso, o que tem tornado importante em todo o mundo a implementação de programas de monitoramento de recursos hídricos^[6].

Particularmente no Brasil, com o intuito de proteger as reservas de águas superficiais e subterrâneas de cada região do território nacional, bem como regulamentar a gestão e estabelecer o controle de qualidade dos sistemas hídricos pode-se destacar:

- a Política Nacional de Recursos Hídricos (PNRH), pela Lei Nº 9433 de 8 de janeiro de 1997, desenvolvida como ação política e administrativa, fundamentada em estudos sobre riscos químicos à saúde pública e ambiental, que favorecem o aperfeiçoamento da legislação e indicam a prática da gestão dos recursos hídricos^[7].
- a Resolução proposta pelo Conselho Nacional do Meio Ambiente (CONAMA) de nº 357 de 17 de março de 2005^[8], que classifica e enquadra os corpos d'água^a.
- a Portaria de nº 518 de 25 de março de 2004 do Ministério da Saúde (MS)^[9], que visa promover o controle e vigilância dos padrões de potabilidade da água para consumo humano.

1.1 - Controle de qualidade dos sistemas hídricos

A qualidade dos sistemas aquáticos está associada às suas propriedades físicas, químicas e biológicas. A alteração da mesma pode ser causada por fatores naturais, climáticos e/ou hidrogeológicos, bem como, antrópicos (efluentes domésticos e industriais ou chorume de aterros sanitários). Estes contaminantes podem atingir reservatórios hídricos superficiais e subterrâneos, causando impactos

^a Segundo a resolução CONAMA N°357/2005 as águas do território nacional são classificadas conforme sua salinidade, como doce, salobra e salina. As águas doces podem pertencer às classes especial, I, II, III e IV, as quais destinam-se basicamente ao abastecimento público. O enquadramento nas respectivas classes dependerá do tratamento prévio que terá a água para se tornar potável.

negativos em suas características normais e em seu estado de equilíbrio. Deste modo, diz-se que o ambiente aquático sofreu um processo de poluição^[4].

Alterações físicas, químicas e biológicas de um ecossistema aquático que caracterizem a poluição dos mesmos, além de causar prejuízos ambientais, acabam refletindo na economia, seja pela diminuição da oferta de peixes, seja pelo aumento do custo de tratamento da água para abastecimento público, e ainda pelos prejuízos causados às atividades de turismo e lazer.

Por isso, é de fundamental importância o gerenciamento efetivo e eficiente dos recursos hídricos, porque água de boa qualidade é um bem que, a cada dia se torna mais escasso, devido ao alto grau de contaminações e ao desperdício. Neste contexto, vale salientar que cerca de 69% da água consumida no país é destinada à irrigação^[10] e que ocorrem constantes vazamentos nas redes distribuidoras de água e mau uso por parte da população.

A qualidade de água de um determinado manancial é avaliada através do seu monitoramento, que assegura os diferentes usos das águas. Esta avaliação se dá por meio de análises físicas e químicas, que identificam e quantificam a substância poluente, e de análises biológicas que mostram os efeitos que tais substâncias causam ao meio ambiente. Vale salientar que a população que consome água de má qualidade está sujeita não apenas às doenças diarréicas de veiculação hídrica, já citadas anteriormente, mas também a doenças degenerativas cerebrais, como o mal de Alzheimer (que tem como fator relevante o excesso de alumínio em águas superficiais e subterrâneas) e a formação de compostos carcinogênicos no estômago, causados por excesso de nitrato e nitrito em águas profundas^[11].

Para garantir a distribuição de água de boa qualidade à população, bem como proteger os mananciais contra a deposição de compostos orgânicos tóxicos, excesso de produtos biodegradáveis e de pesticidas é que são implantadas medidas para a fiscalização dos sistemas aquáticos, que através de diretrizes gerais, estabelecem os limites aceitáveis de certas substâncias nos sistemas hídricos.

A dinâmica dos sistemas aquáticos pode ser avaliada através de programas de monitoramento de longo prazo que englobem a determinação de parâmetros físicos, químicos e biológicos, respeitando os aspectos temporais e espaciais da amostragem, ou seja, fazendo coleta de amostras em diferentes pontos ao longo da bacia hidrográfica e em épocas distintas, de modo que seja observada a sazonalidade^[12].

1.2 – Principais parâmetros de qualidade de água

Dentre os vários parâmetros utilizados nos programas de monitoramento alguns como os mostrados a seguir, são fundamentais e a determinação dos mesmos comumente é exigida pelos órgãos responsáveis pela gestão dos recursos hídricos.

1.2.1 - Parâmetros físicos

➤ Turbidez

A turbidez^[13], como o próprio nome indica, está relacionada à aparência turva da água, causada por material coloidal ou em suspensão, como por exemplo, partículas constituídas por plâncton, bactérias, argilas e/ou por fontes de poluição que lançam material fino nos corpos hídricos. A turbidez é medida pelo espalhamento da luz (normalmente, luz azul) causado por esses materiais. Algumas vezes, a turbidez pode também estar associada à absorção de luz por partículas coloidais ou em suspensão.

O aumento da turbidez reduz a zona de luz prejudicando a fotossíntese onde ela ainda pode ocorrer (zona eufótica), interferindo nos equilíbrios naturais entre os compartimentos ambientais^[14].

➤ Temperatura

É uma medida da intensidade de calor, transferido por radiação, condução ou convecção (entre o corpo d'água, a atmosfera e o solo). A elevação da temperatura acima da faixa usual tem uma relação direta com os equilíbrios químicos e biológicos entre eles com a diminuição da solubilidade dos gases e conseqüente aumento da taxa de transferência desses gases para a atmosfera^[14,15].

1.2.2 - Parâmetros químicos

➤ pH

É uma medida da quantidade de íons hidrogênio, H^+ , em solução, que pode indicar acidez, neutralidade ou basicidade do meio, causada por sólidos e gases dissolvidos, provenientes da dissolução de rochas, absorção de gases da atmosfera, oxidação da matéria orgânica e fotossíntese. Dependendo do valor de pH os sistemas aquáticos podem apresentar características corrosivas ou incrustantes, bem como interferir na vida aquática e influenciar na taxa de crescimento de microorganismos^[14,15].

➤ Nitrogênio total

A medida de nitrogênio total corresponde à soma das quantidades de nutrientes nas formas: amônia, nitrato, nitrito e nitrogênio orgânico (dissolvido ou em suspensão). Ela pode indicar poluição por efluentes domésticos e fertilizantes, bem como a presença de constituintes em sua forma natural (proteínas, clorofila e vários outros compostos biológicos)^[15].

A presença de nitrogênio em grande quantidade pode causar eutrofização de ambientes aquáticos e a morte de peixes que são sensíveis a certos níveis de amônia livre^[16].

➤ Fósforo total

O fósforo presente no sistema aquático também se insere na gama de nutrientes que são indispensáveis para o crescimento dos microorganismos responsáveis pela estabilização da matéria orgânica^[13]. Em excesso, o fósforo causa a eutrofização do meio. A determinação de fósforo total é feita pela soma das parcelas de fósforo inorgânico (ortofosfatos e polifosfatos) e fósforo orgânico. As principais fontes de fósforo são os esgotos domésticos e os detergentes biodegradáveis^[16].

➤ **Sólidos ou resíduos totais**

Corresponde às impurezas da água (exceto os gases dissolvidos) e, de acordo com o tamanho das partículas, podem ser classificados como suspensos ou dissolvidos. Dependendo de suas características químicas, estes se classificam em voláteis (fração orgânica) e fixos (fração inorgânica). A soma destas quantidades (dissolvidos voláteis e fixos e suspensos voláteis e fixos) constitui os sólidos ou resíduos totais, cuja medida é importante na detecção de fonte de poluição por esgoto doméstico^[17].

➤ **Demanda bioquímica de oxigênio (DBO₅)**

Representa de forma indireta a quantidade de matéria orgânica biodegradável que foi consumida pelos microorganismos aeróbios presentes no sistema aquático e constitui um importante parâmetro na caracterização do seu grau de poluição^[15].

Medidas baixas de DBO indicam que há muita matéria orgânica e baixa concentração de oxigênio dissolvido na água e vice-versa^[17]. O termo DBO₅ indica a quantidade de oxigênio resultante após incubação (estabilização da matéria orgânica carbonácea) da amostra analisada por um período de cinco dias, a 20°C^[18].

➤ **Oxigênio dissolvido (OD)**

O oxigênio dissolvido é imprescindível para a manutenção da vida aquática e pode ser proveniente do oxigênio atmosférico ou de organismos fotossintéticos. A concentração de OD na água é influenciada pela temperatura, pois a quantidade de gás dissolvido diminui com o aumento da mesma^[1,13,15]. Este é o principal parâmetro para caracterizar efeitos de poluição causada por despejos orgânicos^[16].

➤ **Condutividade**

A condutividade está associada à quantidade de ânions e cátions dissolvidos na água que conferem à mesma capacidade de conduzir corrente elétrica. Este

parâmetro depende dos estados de oxidação dos íons e da temperatura e tem uma relação direta com as concentrações relativas destes^[14,18].

1.2.3 - Parâmetro biológico

➤ Coliforme fecal (coliformes termotolerantes)

As bactérias do grupo coliforme são organismos indicadores de contaminação fecal associados potencialmente à presença de agentes patogênicos que são responsáveis pela transmissão de doenças por veiculação hídrica, tais como: cólera, febre tifóide, leptospirose, dentre outras^[16].

Os coliformes fecais ou termotolerantes são um subgrupo de bactérias do grupo coliforme (principalmente, a *Escherichia Coli*) que fermentam a lactose a $44,5 \pm 0,2$ °C em 24 horas. Estas bactérias são indicadoras de organismos originados no trato intestinal de humanos e de outros animais de sangue quente. Por esta razão, para indicar poluição sanitária de forma significativa, determina-se a concentração de bactérias coliformes termotolerantes ao invés de bactérias do tipo coliformes totais^[9].

1.3 – Índices de qualidade de água (IQAs)

De acordo com a Resolução CONAMA n° 357 de 17 de março de 2005^[8] nos seus artigos 15° e 16°, que indicam os padrões da água para o abastecimento humano após tratamento convencional, bem como a Portaria do MS n° 518 de 25 de março de 2004^[9] que regulamenta o controle e vigilância da qualidade de água para consumo humano e seu padrão de potabilidade, faz-se necessário a determinação de mais de 60 parâmetros, dentre eles: óleos e graxas, corantes artificiais, DBO₅, OD, turbidez, cor, pH, temperatura e de substâncias potencialmente prejudiciais, como por exemplo, alumínio, nitrato, nitrito e metais pesados, além de testes ecotoxicológicos para o monitoramento de agrotóxicos.

Visto o grande número de análises necessárias para o monitoramento da qualidade da água^[14], é necessário um alto investimento financeiro para que seja

organizada uma estrutura física e técnica, especializada para este fim. Contudo, a realidade da maioria das agências reguladoras e administradoras que examinam a qualidade da água no país é precária, o que limita muito o atendimento a esta demanda.

Diante da gama de informações geradas pelo monitoramento hídrico, cuja interpretação pode ser complicada, surgiu a necessidade de uma abordagem mais simples sobre a qualidade da água, para se chegar a um resultado compreensível para a sociedade em geral (população, prefeituras, órgãos de controle ambiental, comitês de bacias hidrográficas, organizações não-governamentais, entre outros). Essa abordagem deve ser baseada nas variáveis ambientais que são críticas para a qualidade da água e deve servir como ferramenta para examinar tendências, indicar condições ambientais específicas (controle de fontes de poluição pontuais e difusas), bem como para auxiliar órgãos ambientais na avaliação de programas de gerenciamento de recursos hídricos.

Neste contexto, uma forma simples de se mostrar a qualidade da água é através do cálculo do Índice de Qualidade de Água (IQA). O IQA é expresso através da combinação ponderada de um número essencial de parâmetros de qualidade de águas, resultando em um único valor que pode indicar de modo classificatório a qualidade de um determinado curso de água. Entretanto, ao se construir tais IQAs, alguns problemas podem ocorrer, como por exemplo: perda de informação na definição de qualidade da água. Além do mais, os IQAs em geral são de baixa aplicabilidade para outros diferentes sistemas hídricos^[19], uma vez que a universalização deste índice não é válida, pois cada região tem suas características específicas.

O IQA mais utilizado no país como ferramenta de gestão dos recursos hídricos é aquele adotado pela Companhia Estadual de Tecnologia de Saneamento Básico do Estado de São Paulo (CETESB). O IQA-CETESB baseia-se no IQA elaborado pela *National Sanitation Foundation – Units States (NSF-US)*^[1], que considera a determinação de nove parâmetros de qualidade de águas: sólidos totais, turbidez, temperatura, coliformes fecais, pH, Oxigênio Dissolvido (OD), Demanda Bioquímica de Oxigênio a 20°C e 5 dias (DBO₅), nitrato e fosfato.

Vale destacar que, os IQAs não podem ser usados como única fonte de avaliação da qualidade da água, mas servem de orientação para a utilização e o

tratamento da água, a distribuição de recursos, para a informação pública e o planejamento ambiental.

1.3.1 - Formulações de índices de qualidade de água

O cálculo de um IQA compreende quatro etapas básicas^[20]:

- Escolha dos parâmetros (variáveis) que representem a qualidade da água;
- Geração dos sub-índices (qi): transformação dos parâmetros de dimensões e de unidades diferentes para uma mesma escala;
- Atribuição de pesos a cada parâmetro;
- Agregação dos sub-índices para a formulação do IQA final.

1.3.1.1 - Escolha dos parâmetros (variáveis) que representem a qualidade da água

Devido ao grande número de constituintes encontrados na água nas formas dissolvida, coloidal ou em suspensão (substâncias neutras, iônicas, orgânicas e radioativas), e também de outros parâmetros de qualidade como cor, odor, número de bactérias dissolvidas, de fungos etc. é necessário uma escolha criteriosa e a menos subjetiva possível dos parâmetros mais representativos para a avaliação da qualidade da água de determinada região. Essa escolha pode ser feita por métodos estatísticos, que levem em conta a frequência de ocorrência e a correlação entre os parâmetros, a fim de diminuir a incerteza e a subjetividade inerente à avaliação complexa da qualidade da água. Os parâmetros a serem escolhidos para construção do IQA dependerão da finalidade do uso da água, da caracterização sócio-econômica da região em questão e dos critérios legais de aceitabilidade.

1.3.1.2 - Geração da função de sub-índice de qualidade (qi)

A geração da função de sub-índice de qualidade (qi) consiste na transformação dos parâmetros de dimensões e de unidades diferentes para uma mesma escala.

As variáveis que compõem os IQAs têm unidades e ordens de grandeza diferentes bem como provocam impactos diferentes no corpo hídrico. Desta maneira, os dados são transformados em uma única escala, comumente chamada de escala de qualidade, que varia de 0 a 1 ou, a mais comum, de 0 a 100, indo dos valores inaceitáveis para os aceitáveis. Esta transformação é comumente chamada de normalização.

A função de sub-índice de qualidade ou simplesmente função de qualidade, pode ser obtida na maioria das vezes a partir de dados históricos que demonstram o comportamento de cada variável ao longo de um certo período. Dependendo do comportamento do parâmetro, esta função pode ser linear, não linear, segmentada linear ou segmentada não linear. Deste modo, é construído o gráfico da função de sub-índice de qualidade para um dado parâmetro, com os valores de qualidade que variam de 0 a 100 *versus* os valores históricos do parâmetro. A **Figura 1.1** mostra um exemplo de função de sub-índice de qualidade obtida para o parâmetro turbidez.

Os sub-índices obtidos de todos os parâmetros são agrupados em um único número que corresponderá ao IQA final.

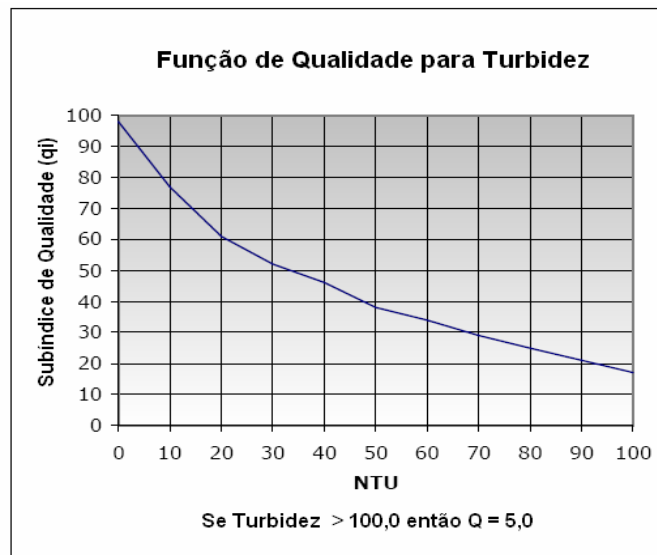


Figura 1.1 Função de qualidade para o parâmetro turbidez. (Fonte: adaptada da NSF-US^[12]).

1.3.1.3 - Atribuição de pesos a cada parâmetro

Assim como a escolha dos parâmetros, a atribuição dos respectivos pesos é uma tarefa subjetiva que deve ter critérios bem estabelecidos, para refletir a real importância de cada um deles

Alguns índices utilizam o mesmo peso para todas as variáveis, mas a maioria as distribuem numa escala de 0 a 1, fazendo uma análise através de estudo Delphi^[21,22]. Este estudo consiste numa pesquisa criteriosa entre especialistas em um determinado assunto para definir características sobre um tema em questão. Por exemplo, BROWN et alii, 1970^[23] utilizaram o estudo Delphi para escolher os parâmetros, bem como os seus respectivos pesos. Quando a formulação de um IQA é feita atribuindo diferentes pesos às variáveis, ocorre na maioria das vezes o chamado efeito eclipse^[24,25] que será discutido na próxima seção.

1.3.1.4 - Agregação dos sub-índices para a formulação do IQA final

A formulação matemática do IQA final pode ser baseada em produtório, somatório, análise fatorial, operador mínimo e operador máximo, dentre outros^[20]. Ao aplicar uma determinada formulação matemática deve-se observar o efeito eclipse^[25]. Este efeito consiste em mascarar um efeito ambiental devido a uma formulação inadequada do IQA que por sua vez ocorre quando um determinado parâmetro que compõe o referido IQA é ponderado de forma indevida. Desse modo, o IQA não refletirá as condições ambientais observadas, ou seja, o efeito negativo de uma determinada variável pode ser mascarado, acarretando uma avaliação equivocada do estado de um corpo hídrico^[24-27].

Pelo exposto acima, conclui-se que na formulação do IQA é imprescindível fazer uma escolha criteriosa dos parâmetros a serem incorporados, levando em conta os limites estabelecidos pela legislação de acordo com a finalidade a que se destina o sistema hídrico em análise^[28-30], e a atribuição adequada dos pesos aos parâmetros, quando esta formulação envolve uma ponderação.

O cálculo final de um IQA segue geralmente uma escala crescente de 0 a 100 no sentido do melhoramento de qualidade da água.

1.3.2 - Aplicações dos índices de qualidade de água

Dentre os inúmeros IQAs encontrados na literatura, os IQAs da NSF-US^[23] e da CETESB^[31] são específicos para indicação da qualidade da água para abastecimento público. Para avaliar a qualidade da água no tocante à proteção da

vida aquática são usados, por exemplo, o IQA de Bascarán^[32] e o IQA utilizando operador mínimo^[24], que serão descritos mais adiante.

1.3.2.1 – IQAs para a avaliação de águas para consumo humano

BROWN et alii^[23], apoiados pela NSF-US, propuseram em 1970 o Índice de Qualidade de Água – IQA_{NSF}. Este índice foi formulado baseando-se numa pesquisa que envolveu 142 profissionais, que trabalhavam com a temática de qualidade da água. Foram indicados 35 parâmetros. Contudo, apenas 9 foram escolhidos (temperatura, pH, OD, DBO₅, coliformes fecais, nitrato, fósforo total, sólidos totais ou resíduo total e turbidez), como forma de padronizar a avaliação da qualidade da água. O cálculo do IQA_{NSF}, de acordo com a **Equação (1)**, é feito através da somatório dos produtos entre os sub-índices de qualidade de cada parâmetro, q_i e o seu respectivo peso, w_i .

$$IQA_{FNS} = \sum_{i=1}^n q_i w_i \quad (1)$$

Os pesos foram atribuídos às variáveis escolhidas considerando sua importância para o consumo humano e os riscos para a saúde pública^[33]. O valor q_i de cada parâmetro é obtido usando a função de sub-índice construída conforme variação dos mesmos no tempo e espaço, através de dados históricos locais.

O resultado do cálculo de IQA_{NSF} leva a valores que podem variar de 0 a 100% (máxima qualidade), constituindo-se numa ferramenta muito simples para diagnosticar a qualidade de um determinado manancial, bem para acompanhar a variabilidade do mesmo no decorrer de sua extensão e do tempo.

O IQA_{NSF} é bastante aplicado como ferramenta necessária para auxiliar a gestão dos recursos hídricos, principalmente nos Estados Unidos^[12], sendo bastante difundido na Escócia e na Polónia^[33].

No Brasil^[34], o uso de IQA como indicador de poluição de determinado sistema aquático foi iniciado pela CETESB^[31,35], que adaptou o IQA_{NSF} da seguinte forma: ao invés de somatório, foi aplicado um produtório, no qual os sub-índices de qualidade dos parâmetros, q_i , são elevadas aos seus respectivos pesos, w_i , conforme a **Equação (2)**:

$$IQA_{CETESB} = \prod_{i=1}^n q_i^{w_i} \quad (2)$$

Os parâmetros considerados relevantes para indicar a qualidade da água destinada ao abastecimento público são basicamente os mesmos usados no IQA_{NSF} , (exceto pelo uso de nitrogênio total e fósforo total ao invés de nitrato e fosfato, respectivamente), bem como os pesos (w_i) e as funções de sub-índices usadas na obtenção dos valores de (q_i) de cada um dos parâmetros.

A metodologia de indicação de qualidade de água baseada no IQA_{CETESB} é bastante comum, sendo utilizada no país por agências de gerenciamento dos recursos hídricos e por estudos científicos que têm como objetivo avaliar impactos na dinâmica aquática, causados por fatores antrópicos, como despejo de esgotos industriais, domésticos ou por atividades agropecuárias.

O Centro de Recursos Ambientais da Bahia, usando o IQA_{CETESB} , realizou em 2001 a avaliação de 13 bacias hidrográficas monitoradas naquele Estado com o intuito de se tomar decisões para implementação de ações preventivas e corretivas para manter e melhorar a qualidade ambiental dessas bacias^[36]. Este mesmo índice também foi utilizado por MACEDO Jr. et alii (2003)^[37] para avaliar a poluição causada por atividades urbanas num trecho do rio Açu, no Rio Grande do Norte. Estes conseguiram classificar entre boa e ótima a qualidade da água no trecho observado.

CARVALHO et alii (2000)^[33] investigaram os impactos de atividades agropecuárias adotando o IQA_{CETESB} como uma das ferramentas para inferir sobre a poluição em fontes de água para o abastecimento público da cidade de São Carlos, SP e constataram a correlação entre os parâmetros medidos e a sazonalidade, bem como com a cobertura vegetal ao longo das margens do sistema hídrico.

O Instituto Ambiental do Paraná (IAP)^[38], através do IQA_{CETESB} avaliou, no período de monitoramento de 1987 a 2001, a qualidade da água em cerca de vinte rios distribuídos nas bacias hidrográficas de Ribeira, Iguaçu, Cinzas, Iraí e Piquiri, localizadas no Estado do Paraná e constataram a variação de qualidade entre os estados classificados como péssimo, ruim, razoável, bom e ótimo.

1.3.2.2 – IQAs para proteção da vida aquática

CONESA FDEZ-VITORIA^[32] (1995), introduziu o Índice de Qualidade de Bascarón (IQA_{Bascarón}) tendo em vista a sobrevivência das espécies aquáticas associadas com o conceito de qualidade de água. O IQA_{Bascarón} pode ser de dois tipos, IQA de Bascarón subjetivo (IQA_{sub}) e objetivo (IQA_{obj}).

O IQA_{sub} baseia-se num somatório, conforme **Equação (3)**:

$$IQA_{sub} = k \frac{\sum_i C_i P_i}{\sum_i P_i} \quad (3)$$

onde C_i e P_i são, respectivamente, as medidas dos parâmetros normalizados (sub-índice de qualidade) e os pesos (que variam de 1 a 4) de cada parâmetro medido. O k é uma constante de subjetividade que depende do aspecto visual do rio e varia de 0,25 a 1,0 do pior para o melhor aspecto.

Quando o IQA_{Bascarón} possui k igual a 1, este é chamado de IQA_{obj}, e é descrito pela **Equação (4)**:

$$IQA_{obj} = \frac{\sum_i C_i P_i}{\sum_i P_i} \quad (4)$$

PINHEIRO e LOCATELLI^[7] (2003), empregaram uma metodologia baseada no IQA_{Bascarón} para observar a qualidade da água dos mananciais da bacia hidrográfica do rio Itajaí. Ao usar o IQA_{obj} a qualidade da água melhorou ao longo dos anos, enquanto que o IQA_{sub} mostrou a tendência inversa, mas não na mesma intensidade. Estas tendências antagônicas observadas revelam que o crescimento do IQA_{obj} se deu pelo aumento da cobertura vegetal nos pequenos mananciais, bem como pela precipitação ocorrida durante as coletas das amostras. O decréscimo no IQA_{sub} foi resultado do despejo de efluentes domésticos e industriais ao longo dos mananciais.

Tanto o IQA_{sub} quanto o IQA_{obj} se baseiam em 20 parâmetros^b de qualidade de água, inviabilizando o cálculo em muitos casos de monitoramento. Com o intuito de diminuir o custo da grande quantidade de análises, bem como o tempo gasto, PESCE e WUNDERLIN^[39] (2002) propuseram um IQA mínimo, IQA_{min} , calculado segundo a **Equação (5)**, na qual são utilizados apenas três parâmetros: OD, condutividade, turbidez ou sólidos totais.

$$IQA_{min} = \frac{C_{OD} + C_{Cond} + C_{tur\ ou\ ST}}{3} \quad (5)$$

onde, C_{OD} , C_{Cond} e $C_{tur\ ou\ ST}$ são, respectivamente, os valores medidos de oxigênio dissolvido, condutividade e turbidez ou sólidos totais após normalização dos parâmetros.

Perfis similares foram obtidos (PESCE e WUNDERLIN, 2002)^[39] para a evolução do IQA_{min} e dos IQA_{sub} e IQA_{obj} do sistema analisado. Como o IQA_{sub} pode superestimar a poluição, devido à constante k , a comparação entre o IQA_{min} e o IQA_{sub} pode fornecer um resultado não confiável. Porém, como o perfil do IQA_{min} e o IQA_{obj} foram também similares pode-se confiar no resultado obtido do IQA_{min} , obtendo-se com o seu uso uma grande economia no custo total do monitoramento dos recursos hídricos.

O Instituto Ambiental do Paraná (IAP)^[34,38,40-41] utilizou a mesma equação do $IQA_{Bascarón}$, a **Equação (4)** para verificar a qualidade da água da região. Com base nos dados históricos determinaram-se as variáveis relevantes^c e seus respectivos pesos, mantendo a soma dos mesmos iguais a 100. O índice foi chamado de IQAR, Índice de Qualidade de Água de Reservatório.

Outro IQA para a proteção da vida aquática, usando a metodologia de operador mínimo foi aplicado por SILVA E JARDIM (2006)^[24] aos dados da bacia do rio Atibaia, São Paulo. Ao considerarem apenas as variáveis mais críticas para a proteção da vida aquática que são amônia total e OD obtiveram valores de IQA

^b Os parâmetros de qualidade que compõem o $IQA_{Bascarón}$ são: amônia, DBO_5 , cálcio, cloreto, DQO (Demanda Química de Oxigênio), OD, dureza, magnésio, nitrato, nitrito, óleos e graxas, pH, fósforo (ortofosfatos), sólidos dissolvidos, sólidos totais dissolvidos, sulfatos, surfactantes, temperatura, coliformes totais e turbidez.

^c Para o cálculo do IQAR, o IAP utiliza os parâmetros: déficit de OD, fósforo total, nitrogênio inorgânico total, clorofila "a", profundidade Secchi, DQO, Fitoplâncton, tempo de residência e profundidade média.

semelhantes aos obtidos pela CETESB (que utiliza nove parâmetros) no mesmo sistema observado.

1.4 – Quimiometria

1.4.1 - Tratamento quimiométrico e estatístico dos dados

Segundo Kowalski ^[42], “quimiometria é uma disciplina da química que usa métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para fornecer o máximo de informação química na análise de dados de natureza multivariada”.

Métodos de reconhecimento de padrão não supervisionados são utilizados para se fazer análises qualitativas de dados de natureza multivariada ^[43], como é o caso da análise hierárquica de agrupamentos (do inglês: *Hierarchical Cluster Analysis* - HCA) e a análise de componentes principais (do inglês: *Principal Component Analysis*- PCA) ^[44].

1.4.1.1 – Pré-processamento dos dados

Antes da aplicação de HCA e PCA é necessário observar se os dados necessitam de algum tratamento prévio, chamado de pré-processamento dos dados. Técnicas de pré-processamento são empregadas para corrigir variações sistemáticas ou aleatórias que mascarem uma variação de interesse. Contudo, se for usada uma técnica inadequada de pré-processamento, informação útil pode ser removida, e para evitar isso; os dados devem ser reexaminados e comparados com os originais.

Para dados de natureza discreta que apresentem unidades e magnitudes diferentes, a técnica de pré-processamento mais adequada é o autoescalamento ^[45], que consiste em centrar os dados na média, e em seguida fazer a normalização da covariância. O autoescalamento é calculado segundo as **Equações (6) e (7)**, mostradas a seguir.

A matriz de covariância é dada pela expressão:

$$Cov(x, y) = \frac{1}{N-1} \sum_i (x_i - x_{med})(y_i - y_{med}) \quad (6)$$

onde x, x_{med} e y, y_{med} são as variáveis e as respectivas médias.

A normalização da covariância é dada por:

$$Cor(x, y) = \frac{1}{N-1} \sum_i \frac{(x_i - x_{med})}{S_x} \frac{(y_i - y_{med})}{S_y} \quad (7)$$

onde S_x e S_y representam os desvios padrão da cada variável.

1.4.1.2- Análise hierárquica de agrupamentos – HCA

A análise hierárquica de agrupamentos é utilizada na análise de dados de natureza multivariada com a finalidade de agrupar as amostras que tenham similaridade entre si e heterogeneidade em relação a outros agrupamentos em termos de uma determinada distância (por exemplo: distância Euclideana ou de Mahalanobis)^[44,46]. As diferenças e similaridades entre os agrupamentos são analisadas através de um gráfico conhecido como dendrograma.

Para construir um dendrograma é necessário escolher em primeiro lugar o tipo de distância utilizada para medir a similaridade entre as amostras. Em segundo lugar, deve ser escolhido o método adequado para a estruturação dos agrupamentos.

A distância mais comumente utilizada em HCA é a distância Euclideana, calculada segundo a **Equação (8)** e o método para estruturação de agrupamentos empregado a dados discretos, é chamado método *Ward*^[47].

$$D = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (8)$$

onde x e y representam as amostras referentes às variáveis i .

O método *Ward* agrupa os pontos de modo a minimizar o resíduo interno do grupo^[48]. Desta maneira, são formados grupos segundo a ordem de similaridade, e

devido à minimização da variação interna, são construídos agrupamentos mais ou menos do mesmo tamanho.

1.4.1.3 - Análise de componentes principais – PCA

A análise de componentes principais^[42,49] baseia-se na combinação linear^d das variáveis originais correlacionadas que compõem o conjunto de dados, a fim de que sejam gerados novos eixos de coordenadas, denominados de componentes principais. No espaço descrito por estes novos eixos as amostras são dispostas de tal modo que a estrutura dos dados e as classes que são formadas podem ser visualizadas com maior facilidade.

As componentes principais são projetadas na direção da maior variância dos dados. A primeira componente principal (PC1) segue na direção da variância máxima, enquanto a segunda componente principal (PC2) que deve ser ortogonal a PC1, segue a segunda maior variância e assim por diante (**Figura 1.2**).

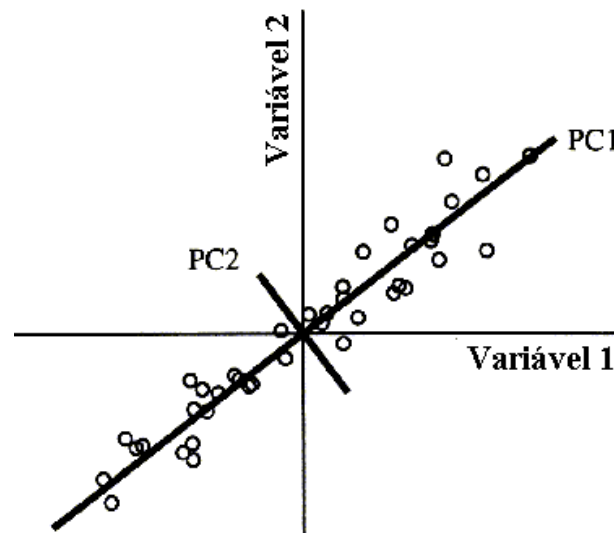


Figura 1.2 - Representação das duas primeiras PCs em que os círculos representam as amostras. (Fonte: adaptado de BEEBE et alii, 1998^[44])

^d Estatisticamente a combinação entre as variáveis depende da correlação entre as mesmas. De forma geral, a combinação linear de p variáveis é dada por: $y = a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum_i a_i x_i$, que pode ser escrita em termos da variância das variáveis correlacionadas pela expressão: $S_y^2 = \sum_i a_i^2 S_i^2$, onde S_y^2 é a variância, a_i é o coeficiente linear e S_i é o desvio padrão de cada variável.

As coordenadas das amostras nas componentes principais (PCs) são dadas pelos valores de escores (**Figura 1.3.a**) e os coeficientes da combinação linear são dados pelos valores dos pesos (**Figura 1.3.b**), que são os co-senos dos ângulos entre os eixos originais e os novos eixos dados pelas PCs. Os valores dos pesos nas PCs indicam literalmente o peso que cada variável original tem neste novo sistema de eixos.

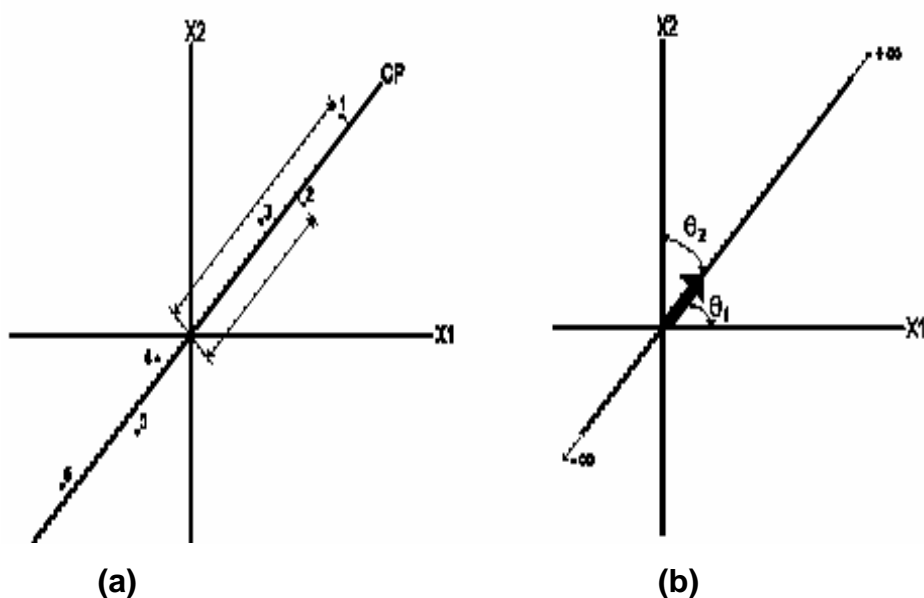


Figura 1.3 - Gráfico dos escores e as projeções na direção da componente principal (a) e os pesos dados pelos co-senos dos ângulos entre cada variável e a componente principal (b). (Fonte: adaptado de BEEBE et alii, 1998^[44])

1.5 – Análise exploratória em qualidade de água e a necessidade de ponderação das variáveis

O grande número de parâmetros medidos para avaliar a qualidade da água pode gerar confusão quando analisados simultaneamente, uma vez que deve ser levada em consideração a interação entre os mesmos e os equilíbrios aquáticos, dados muitas vezes por reações de complexação, precipitação, oxi-redução e ácido base e ainda os fatores termodinâmicos e cinéticos complexos^[50]. Deste modo, o uso de ferramentas quimiométricas para a análise exploratória de dados ambientais de natureza multivariada tem sido o foco de vários estudos, como os citados a seguir.

VEGA et alii (2003)^[51], BENGRAINE et alii (1995)^[52] e MARENGO et alii (1994)^[53], utilizaram HCA e PCA para investigar a variabilidade hidroquímica dos sistemas aquáticos, quanto aos fatores sazonais, temporais e em relação à poluição antrópica.

Como a análise exploratória é feita a partir dos dados originais, a PCA é construída conforme o ambiente geofísico e estrutural em questão e gera modelos específicos que podem ser aplicados no monitoramento ambiental.

Neste contexto, técnicas quimiométricas podem auxiliar em planos de amostragem com menor número de parâmetros analisados; por exemplo, ANDRADE et alii (2004)^[54], propuseram a redução do número de variáveis utilizadas no programa de monitoramento de águas profundas. Em 2004, PARINET et alii^[55] verificaram as variáveis mais importantes para a avaliação do estado trófico em lagos.

É importante ressaltar que nos dos trabalhos citados acima e na pesquisa bibliográfica realizada, não foram encontrados estudos fazendo ponderações das variáveis originais, para que fosse refletida a importância qualitativa das variáveis ambientais no conceito de qualidade de água.

1.6 – Objetivos

1.6.1 - Objetivo Geral:

Fazer a análise exploratória de dados ambientais provenientes de sistemas aquáticos levando-se em consideração a importância que cada variável possui no conceito de qualidade de água.

1.6.2 - Objetivos Específicos:

1) Inserir na análise exploratória de dados ambientais a mesma atribuição da importância qualitativa dos parâmetros que compõem os IQAs usando ponderação no pré-processamento.

2) Avaliar se análise exploratória ponderada expressa a qualidade de água do sistema em análise, tendo como referência a metodologia de construção de IQAs em geral e a classificação dada pelo IQA_{CETESB} .

3) Verificar se a análise exploratória permite inferir sobre a qualidade da água para fins de abastecimento público, observando a relação entre o comportamento multivariado dos dados e os fenômenos hidroquímicos em geral, inerentes ao sistema em análise.

2 – MATERIAIS E MÉTODOS

2.1 - Banco de dados

Para a aplicação da metodologia proposta foi utilizado um banco de dados proveniente do monitoramento realizado pelo IAP no período de 1987 a 2001 em cerca de vinte rios das bacias de Ribeira, Iguaçu, Cinzas, Iraí e Piquiri, localizadas no Estado do Paraná. O banco de dados é composto de 568 amostras e foi adquirido via *internet*^[38].

2.1.1 - Parâmetros que compõem os dados

O monitoramento foi realizado com a finalidade de avaliar a qualidade da água destinada ao abastecimento público, seja ela bruta ou tratada. Para isso, foram feitas medidas das seguintes variáveis físicas, químicas e biológicas, apresentadas na **Tabela 2.1**.

Tabela 2.1 – Parâmetros medidos no monitoramento

Característica	Parâmetros	Unidade
Parâmetros físicos	Turbidez	UNT
	Temperatura	°C
Parâmetros químicos	pH	-
	N-Total	mg l ⁻¹ - N
	P-Total	mg l ⁻¹ - P
	OD	mg l ⁻¹ - O ₂
	DBO ₅	mg l ⁻¹ - O ₂
	Sólidos Totais	mg l ⁻¹ - O ₂
	Condutividade	µS cm ⁻¹
Parâmetro biológico	Coliforme fecal	NMP(100 mL) ⁻¹

2.2 - Softwares

Para pré-tratamento dos dados e realização dos cálculos quimiométricos utilizou-se o MICROSOFT EXCEL XP, os pacotes UNSCRUMBLER^[56] da CAMO A/S e o STATISTICA versão 6.0, bem como o ambiente operacional MATLAB for Windows, versão 6.1, da Mathworks, 2001.

2.3 - Metodologia proposta

2.3.1 – Análise exploratória: HCA e PCA ponderadas

Utilizar HCA e PCA para fazer a indicação da qualidade da água e comparar esses resultados com a classificação dada pelo IQA pode levar a considerações interessantes. Contudo, vale a pena destacar que o IQA é dado como um simples número que pode encobrir observações importantes acerca do sistema hídrico em análise. Entretanto, através de HCA e PCA essas variações podem ser vistas também de maneira simples, fornecendo, contudo uma informação mais completa, sem necessidade de usar modelos pré-estabelecidos, como por exemplo, as funções de qualidade usadas pelo IQA_{CETESB}.

Nesta metodologia, os dados originais são previamente autoescalonados^[45] e ponderados antes da realização tanto da HCA, como da PCA, pois como nos IQAs em geral, cada variável tem a sua importância relativa definida no contexto da qualidade da água. Por exemplo, alguns parâmetros como DBO₅, sólidos totais e OD são qualitativamente mais importantes que temperatura e pH. Esta é a idéia levada em conta quando pesos são atribuídos aos parâmetros de qualidade na formulação dos IQAs em geral, logo, a metodologia aqui proposta apenas transpõe esta idéia de ponderação.

Para construção de HCA e PCA foram utilizados dois conjuntos de pesos diferentes: os pesos provenientes do IQA_{BASCARÓN} e os pesos do IQA_{CETESB}. A **Tabela 2.2** mostra os respectivos pesos utilizados para a ponderação da PCA.

Tabela 2.2 Parâmetros com seus respectivos pesos. Os pesos do IQA_{BASCARÓN} variam de 1 a 4 enquanto os pesos do IQA_{CETESB} têm somatório igual a 1.

Parâmetros	Unidade	Pesos IQA _{CETESB}	Pesos IQA _{BASCARÓN}
OD	mg l ⁻¹ O ₂	0,17	4
Coliforme fecal	NMP (100 mL) ⁻¹	0,15	3
pH	-	0,12	1
DBO ₅	mg l ⁻¹ - O ₂	0,10	3
N-Total	mg l ⁻¹ - N	0,10	2
P-Total	mg l ⁻¹ - O ₂	0,10	1
Turbidez	UNT	0,08	2
Sólidos Totais	mg l ⁻¹	0,08	4
Temperatura	°C	0,10	1

Fonte: Adaptado de PESCE E WUNDERLIN, 2002^[39] e de CETESB^[56]

A ponderação dos dados faz parte da etapa de pré-processamento e é feita logo após o autoescalamento antes de ser realizado o cálculo para obtenção dos dendrogramas e das PCs. A **Figura 2.1** mostra o esquema de pré-processamento, que tem como resultado uma matriz ponderada a ser usada nos cálculos da HCA e da PCA.

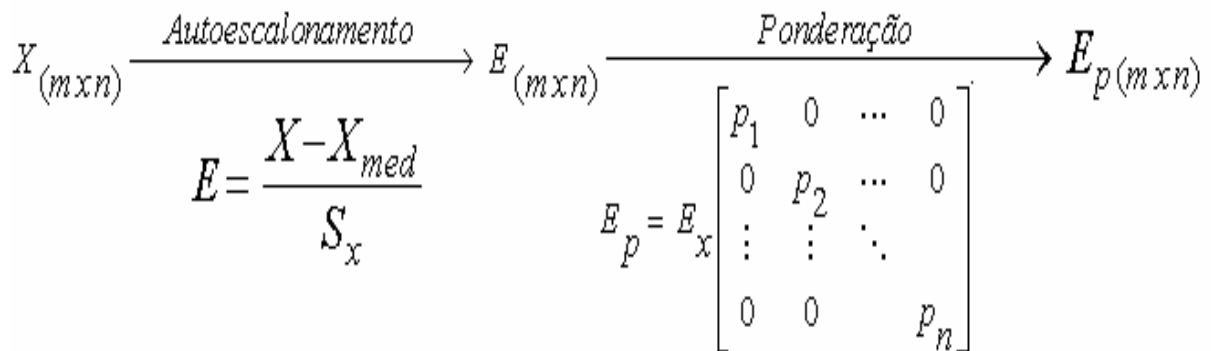


Figura 2.1 – Esquema de pré-processamento aplicado à matriz de dados X com m linhas e n colunas. Após o autoescalamento é gerada a matriz E, que é ponderada multiplicando-a pela matriz de pesos (matriz diagonal), resultando na matriz ponderada, E_p a ser usada em HCA e PCA.

3 – RESULTADOS E DISCUSSÃO

3.1 - Exame preliminar dos dados

O banco de dados do IAP é composto por 568 amostras que apresentam valores de IQA que variam segundo a classificação dada pelo próprio IAP, como: bom, razoável, ruim e péssima.

A classificação das amostras do conjunto de dados foi feita com base nos seus respectivos valores de IQA. Os valores que delimitam cada classe e o número de amostras de cada uma são mostrados na **Tabela 3.1**.

Tabela 3.1- Classificação das amostras do banco de dados segundo o IAP

Classe	Variação no Valor IQA _{IAP}	Número de Amostras
PÉSSIMO	0-25	71
RUIM	26-50	131
RAZOÁVEL	51-70	201
BOM	71-90	165
ÓTIMO	91-100	-

Fonte: Adaptado de IAP^[38]

Inicialmente foram calculadas as médias amostrais e os desvios padrão das amostras referentes a cada classe da **Tabela 3.1**. Em seguida, foram calculados os intervalos correspondentes à média, mais ou menos três desvios padrão para cada classe. Assim podemos avaliar a sobreposição entre os intervalos que correspondem às regiões de incerteza na classificação das amostras. Essas incertezas vêm do fato dos limites das classes em termos de valor de IQA serem subjetivos, e não absolutos, não significando que exista uma diferença marcante entre as amostras com IQA 70 (maior IQA classificado como razoável) e 71 (menor IQA classificado com bom). O limite de três desvios padrão foi escolhido porque ele cobre praticamente toda a extensão da distribuição normal: 99,7%. Os intervalos estão apresentados na **Tabela 3.2**.

Tabela 3.2- Incertezas associadas à variabilidade dos dados do IAP^[38]

Classe	Média Amostral	Desvio Padrão	Intervalo de confiança
PÉSSIMO	19,0	3,04	19,0 ± 9,12
RUIM	39,2	7,27	39,2 ± 21,81
RAZOÁVEL	61,3	6,11	61,3 ± 18,13
BOM	75,5	4,58	75,5 ± 13,74

Observa-se na **Tabela 3.2** que os intervalos de confiança para cada classe sobrepõem-se, como por exemplo, as amostras da classe ruim têm o limite inferior do seu intervalo de confiança para média populacional iniciando-se em 17,4; que pela definição do IQA corresponderia a uma amostra classificada como péssima. Da mesma forma, o limite superior do intervalo com 99,74% de confiança para as amostras da classe péssima é de 28,1; que corresponderia pela definição a uma amostra da classe ruim. Isso ocorre porque as fronteiras entre as classes dadas pelo IQA_{CETESB} são arbitrárias e não rigorosamente definidas, uma vez que se trata apenas de um índice para avaliação e diagnóstico qualitativo de determinado corpo hídrico.

3.2 - Análise quimiométrica e estatística dos dados

3.2.1 - Análise hierárquica de agrupamentos ponderada (HCA ponderada)

Para a análise hierárquica de agrupamentos, os dados foram autoescalados e ponderados usando tanto os pesos do IQA_{BASCARÓN} quanto os pesos do IQA_{CETESB}.

A HCA foi realizada com o objetivo de identificar padrões de qualidade de águas ao longo dos agrupamentos formados. Foi utilizada a distância Euclideana para medir a similaridade entre as amostras e o método *Ward* para a construção dos agrupamentos.

Primeiramente foram construídos dendrogramas sem rotulagem das classes com o intuito de observar a possível formação de agrupamentos. Em seguida, as amostras nos dendrogramas foram rotuladas segundo o IQA_{CETESB} , usado como referência para confirmar a ligação entre a qualidade de água e os agrupamentos formados.

➤ **HCA ponderada sem rotulagem das classes**

Na **Figura 3.1**, observam-se os dendrogramas referentes aos dados autoescalados e ponderados com os pesos do $IQA_{BASCARÓN}$ e com os pesos do IQA_{CETESB} . A numeração observada corresponde apenas à seqüência das amostras na matriz de dados (568 amostras analisadas e nove parâmetros determinados).

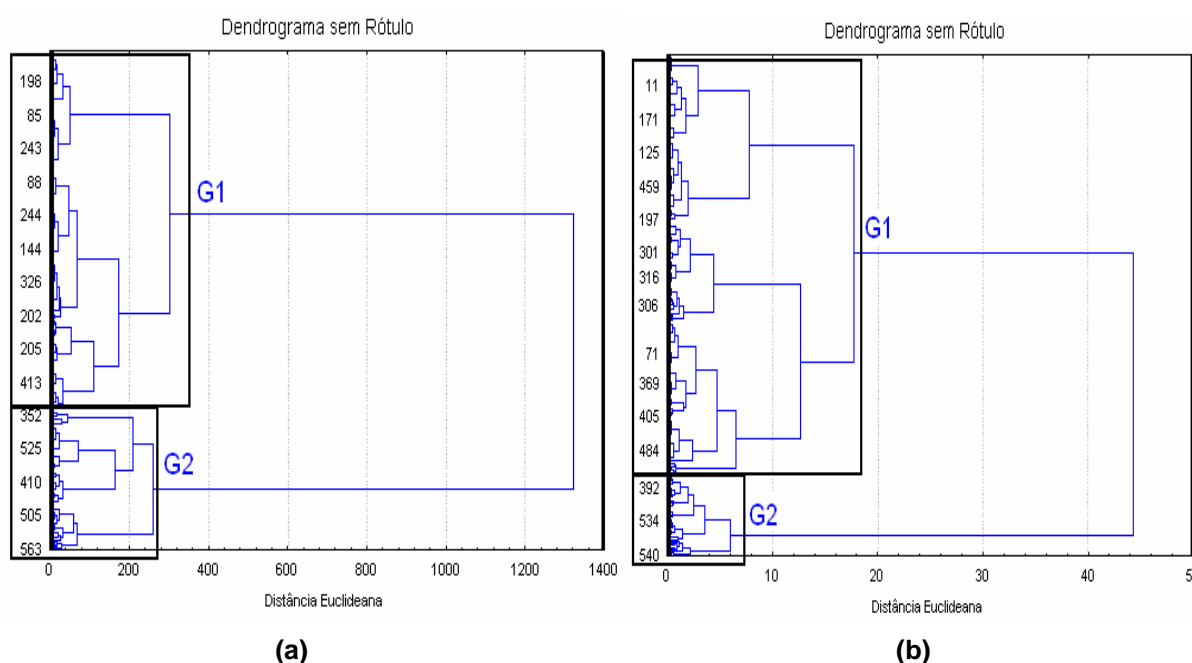


Figura 3.1- Dendrogramas dos dados sem rotulagem usando ponderação dada pelo $IQA_{BASCARÓN}$ (a) e ponderação dada pelo IQA_{CETESB} (b).

Cada agrupamento observado na **Figura 3.1** é composto por amostras que têm homogeneidade dentro do grupo e heterogeneidade em relação às demais classes. Quanto menor a distância entre as classes ou amostras, maior é a similaridade entre elas. Para ambos os conjuntos de pesos observam-se a formação de dois grandes grupos denominados G1 e G2, que se ligam a uma distância de 1350 quando os pesos do $IQA_{BASCARÓN}$ são usados (**Figura 3.1a**) e a

aproximadamente 44,5 quando os pesos do IQA_{CETESB} são usados (**Figura 3.1b**). Nota-se a presença de três subgrupos dentro do agrupamento denominado G1. Esses subgrupos são mais similares entre si e estão ligados a uma distância de 280 (**Figura 3.1a**) e 18 (**Figura 3.1b**). Observa-se, também, a presença de dois subgrupos dentro do agrupamento G2 que se ligam a uma distância de 240 (**Figura 3.1a**).

➤ HCA ponderada com rotulagem das classes

Os dendrogramas da seção anterior foram rotulados usando a classificação do IAP, como já foi dito anteriormente. O dendrograma obtido após a rotulagem é mostrado na **Figura 3.2**.

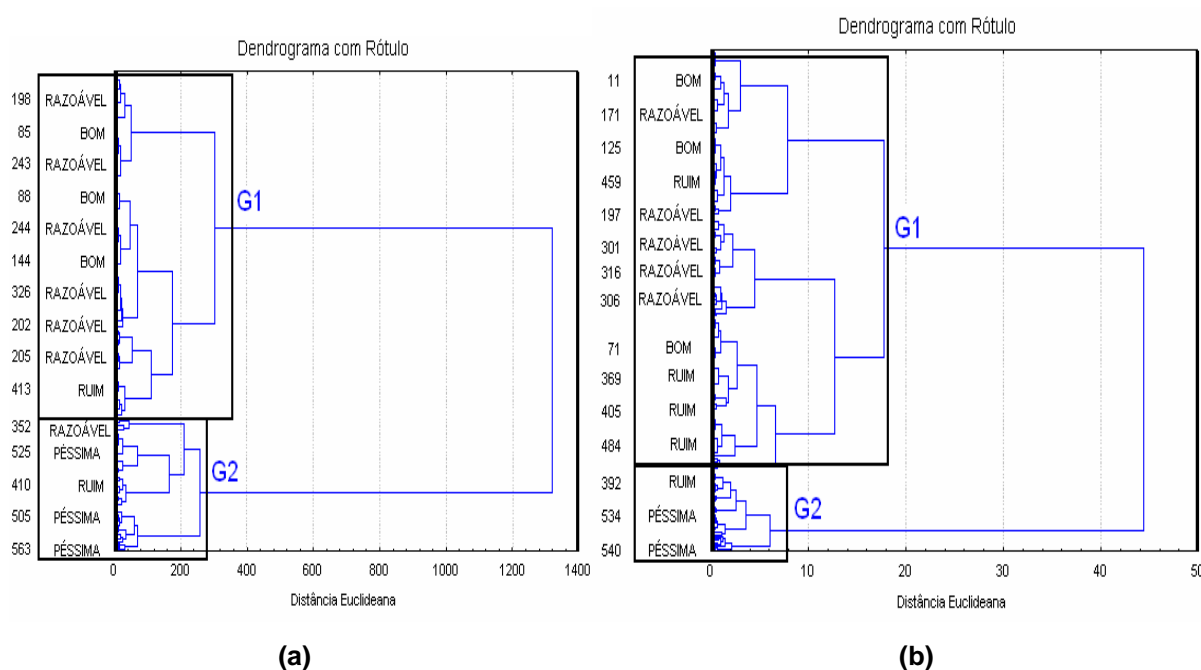


Figura 3.2 - Dendrogramas dos dados com rotulagem usando ponderação dada pelo IQA_{BASCARÓN} (a) e ponderação dada pelo IQA_{CETESB} (b).

Ao comparar os agrupamentos de amostras com sua respectiva classificação (bom, razoável, ruim e péssimo), observa-se (**Figura 3.2**) que os grupos G1 e G2 estão separados de acordo com a qualidade da água das amostras desses grupos. Nota-se, também, que as águas de “melhor” qualidade (boa e razoável) estão contidas, de forma geral, no agrupamento G1, enquanto que as águas de “pior” qualidade (ruim e péssima) estão contidas no agrupamento G2.

Como mostra a **Figura 3.2.a**, onde as amostras foram ponderadas com os pesos do $IQA_{BASCARÓN}$, as amostras 413 e 352 (respectivamente ruim e razoável na **Tabela 3.3**) não estão classificadas nos grupos esperados, ou seja, a 413 está em G1 e a 352 está em G2. A amostra 352 é considerada de qualidade razoável por apresentar seu IQA_{CETESB} igual a 51, enquanto a 413 é dita ruim por apresentar seu IQA_{CETESB} igual a 43 (**Tabela 3.1**).

Tabela 3.3- Classificação das amostras referentes ao dendrograma rotulado

Amostra com pesos $IQA_{BASCARÓN}$	Valor de qualidade	Qualidade IQA_{IAP}	Amostra com pesos IQA_{CETESB}	Valor de qualidade	Qualidade IQA_{IAP}
198	65	Razoável	11	81	Bom
85	72	Bom	171	67	Razoável
243	65	Razoável	125	85	Bom
88	87	Bom	459	49	Ruim
244	56	Razoável	197	62	Razoável
144	71	Bom	301	69	Razoável
326	70	Razoável	316	66	Razoável
202	67	Razoável	306	68	Razoável
205	63	Razoável	71	80	Bom
413	43	Ruim	369	42	Ruim
352	51	Razoável	405	41	Ruim
525	19	Péssima	484	29	Ruim
410	30	Ruim	392	27	Ruim
505	15	Péssima	534	24	Péssima
563	21	Péssima	540	12	Péssima

No entanto, como vimos na **Tabela 3.2**, o intervalo de confiança para nossas amostras da classe razoável é de $61,3 \pm 18,13$ e da classe ruim é de $39,2 \pm 21,81$. Portanto, ambas as amostras estão localizadas na faixa de 43-61, onde os intervalos de confiança para essas duas classes se sobrepõem e não podemos classificar com exatidão essas amostras.

Na **Figura 3.2.b**, onde as amostras foram ponderadas com os pesos do IQA_{CETESB} , nota-se que o grupo G1 não inclui majoritariamente amostras boas e razoáveis, visto que as amostras como 369 ($IQA_{IAP} = 42$), 405 ($IQA_{IAP} = 41$) e 484 ($IQA_{IAP} = 29$), todas classificadas como ruins (**Tabela 3.3**), pertencem a esse grupo. Ainda assim, o dendrograma separa as amostras por qualidade, já que essas amostras ruins estão agrupadas em um mesmo subgrupo de G1.

Uma outra observação é que a amostra 451 ($IQA_{IAP} = 49$) da classe ruim (**Tabela 3.3**) está em um outro subgrupo de G1 (**Figura 3.2.b**), ligando-se a amostras das classes razoável e boa. Contudo, esta amostra não pode ser classificada com exatidão como ruim, pois seu valor de IQA_{IAP} cai na faixa de 43-61, que corresponde à sobreposição das classes razoável e ruim segundo seus respectivos intervalos de confiança ($61,3 \pm 18,13$ e $39,2 \pm 21,81$).

Conclui-se, então, que a HCA ponderada utilizando os pesos do $IQA_{BASCARÓN}$ se mostrou mais adequada para análise geral do comportamento da qualidade de água, pois forneceu agrupamentos bem mais distintos. A partir desta HCA, pode ser deduzida a existência de dois agrupamentos gerais de qualidade de água: G1 para as águas de “melhor” qualidade e G2 para as águas de “pior” qualidade. Ao contrário, a HCA que usou os pesos do IQA_{CETESB} agrupou amostras de qualidade ruim no agrupamento G1, formado por águas de melhor qualidade.

3.2.2 - Análise de componentes principais ponderada (PCA ponderada)

Foi realizada a análise de componentes principais com os dados previamente autoescalados e ponderados usando os pesos de $IQA_{BASCARÓN}$ e de IQA_{CETESB} , a fim de ser observada a variação da qualidade da água.

Primeiramente foi feita a PCA sem ponderação e em seguida a PCA ponderada a fim de compará-las e então investigar a influência do uso de pesos na indicação de qualidade de água.

➤ PCA sem ponderação

Considerando os dados sem ponderação, apenas autoescalonados, observamos na **Figura 3.3**, um certo perfil de qualidade, no qual as amostras variam de boa a péssima indo do sentido negativo para o sentido positivo da primeira componente principal (PC1), que porta apenas 42% da informação total dos dados (variância explicada).

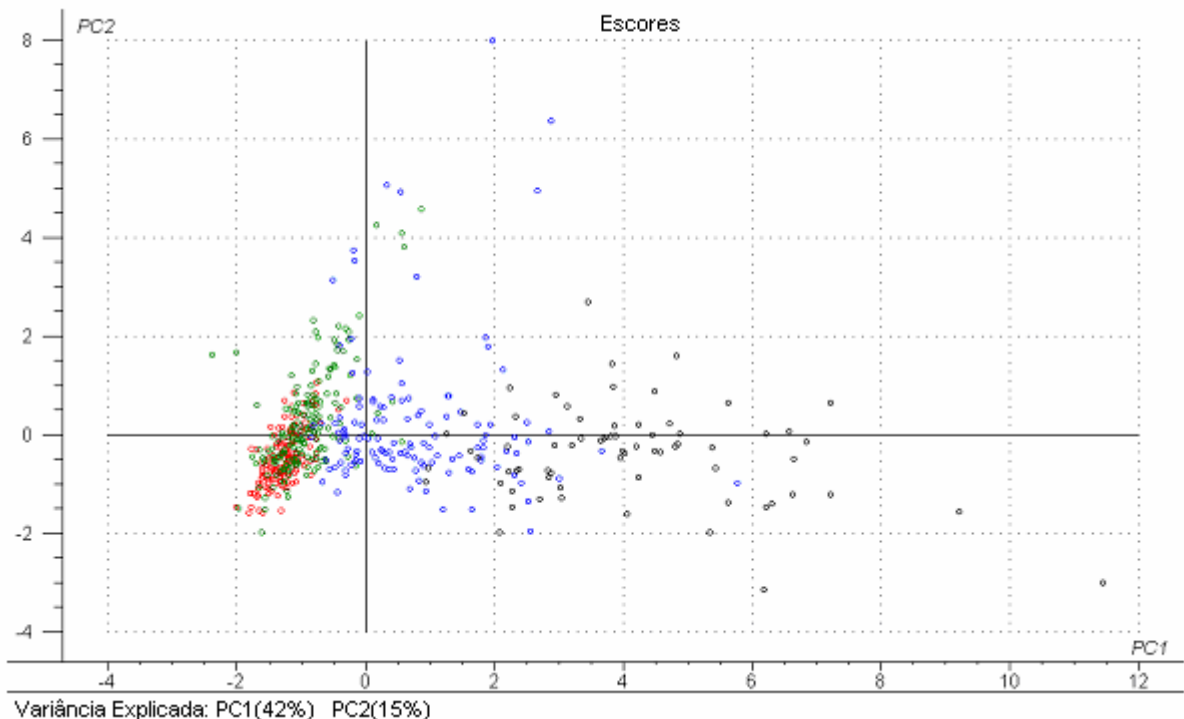


Figura 3.3 – Gráfico dos escores para as duas primeiras PCs sem ponderação (■ bom; ■ razoável; ■ ruim; ■ péssima)

As amostras dispostas ao longo de PC1 referem-se às classes de qualidade destacadas pelo IQA_{IAP} e indicam a mesma tendência observada na HCA, no que se refere às fronteiras pouco rigorosas entre as classes, observada pela mistura das mesmas. Entretanto, observa-se que os escores negativos de PC1 correspondem às águas de “melhor” qualidade enquanto os escores positivos correspondem às águas de “pior” qualidade.

➤ **PCA ponderada usando os pesos do IQA_{BASCARÓN}**

A **Figura 3.4** mostra o gráfico dos escores das duas primeiras componentes principais (PC1 X PC2), obtido para as amostras que têm qualidades variando de péssima a boa. Observa-se uma maior dispersão dos dados no sentido negativo de PC1 e no sentido positivo de PC2.

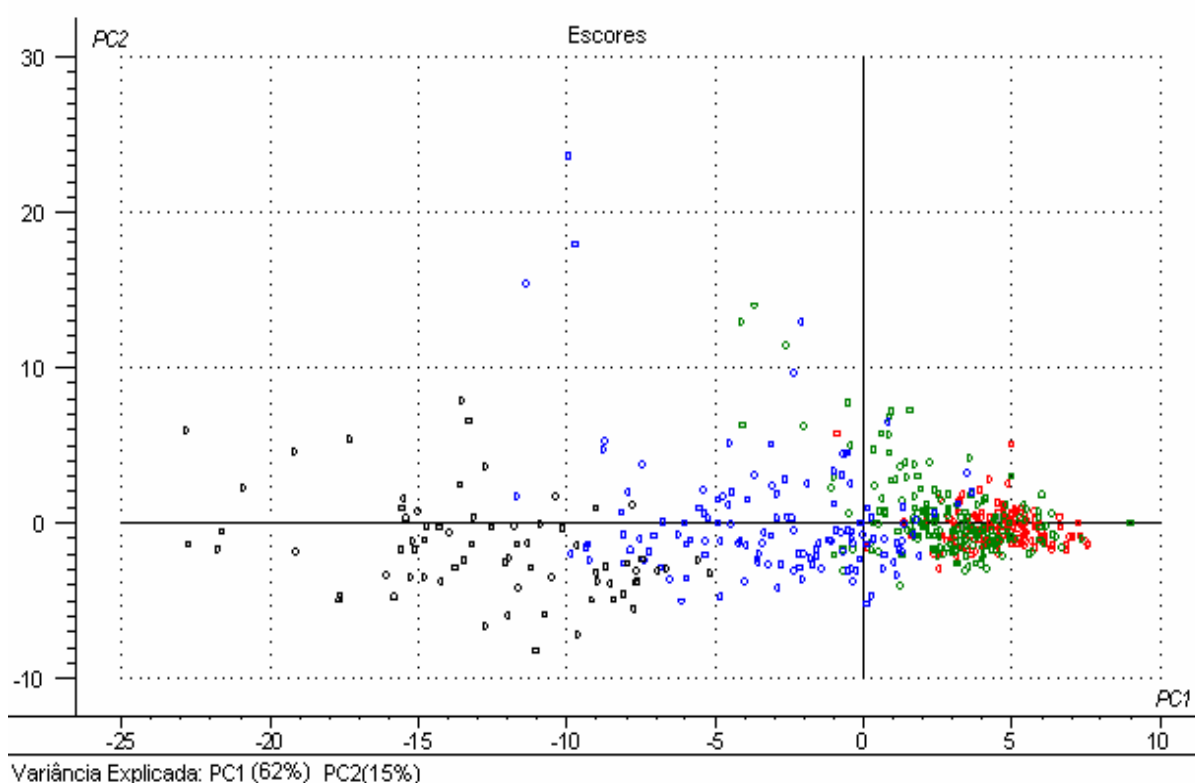


Figura 3.4 – Gráfico dos escores para as duas primeiras PCs, segundo o IQA_{BASCARÓN} (■ bom; ■ razoável; ■ ruim; ■ péssima).

Verifica-se também na **Figura 3.4**, uma relação entre os escores de PC1 e qualidade da água, uma vez que, caminhando do sentido positivo para o sentido negativo dessa componente, a qualidade das amostras de água varia na seqüência: boa, razoável, ruim e péssima (segundo a classificação do IQA_{IAP}), à medida que os valores dos escores tornam-se mais negativos.

A primeira componente principal (PC1) descreve 62% da variância explicada, ou seja, modela 20% a mais de informação quando comparada com a PCA não ponderada (**Figura 3.3**). Observamos também que na PCA ponderada com os pesos do IQA_{BASCARÓN} (**Figura 3.4**), as classes estão mais dispersas em PC1, numa faixa

de escores mais ampla (de -23 a + 9) que na PCA não ponderada (-2 a 12), conduzindo a uma interpretação de qualidade mais resolvida.

É interessante notar que em ambas as PCAs as amostras de melhor qualidade, classificadas como boas, estão menos dispersas que as demais, o que indica que esta categoria de qualidade está bem definida.

Vale ressaltar que a diferença do sentido em que as amostras apresentam “melhor” qualidade (**Figura 3.4**), quando se comparam estes resultados com os resultados da PCA não ponderada (**Figura 3.3**) é arbitrária e depende apenas da execução do algoritmo empregado no cálculo. A interpretação baseia-se apenas na disposição relativa entre os valores dos escores e/ou pesos da PCA.

A segunda componente principal (PC2) explica 15% da variância. A **Figura 3.5** mostra o gráfico da variância explicada versus o número de componentes principais.

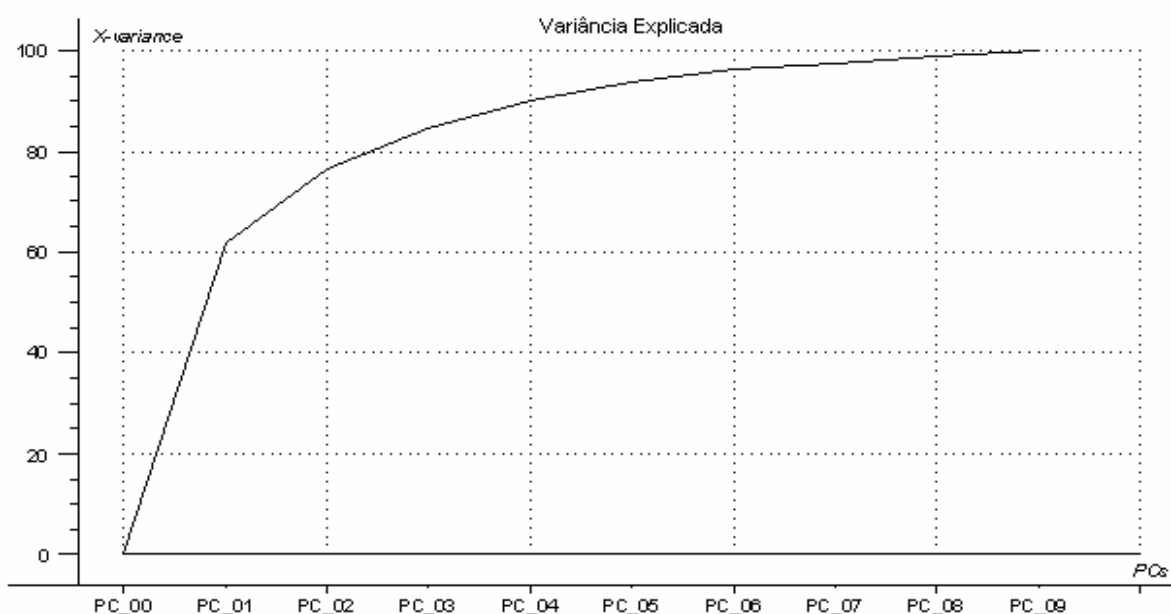
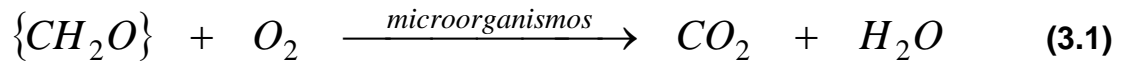


Figura 3.5 – Gráfico da variância explicada *versus* o número de componentes principais ponderadas segundo o IQA_{BASCARÓN}.

Ao observar o gráfico dos pesos de PC1 *versus* PC2 para cada variável (**Figura 3.6**) é importante destacar algumas relações:

- DBO e OD apresentam correlação inversa na primeira e na segunda componente principal, uma vez que seus pesos aparecem em quadrantes opostos para ambas as componentes. Essa relação está de acordo com a dinâmica química dos sistemas aquáticos envolvendo o processo de consumo

de matéria orgânica nas fontes hídricas. Este processo é representado pela reação química genérica dada pela **Equação 3.1**, abaixo:



De acordo com a **Equação 3.1**, com o aumento da matéria orgânica (CH_2O) a quantidade de oxigênio necessária para degradá-la será cada vez maior, o que implica em uma redução na quantidade de oxigênio dissolvido presente no corpo hídrico.

➤ Os pesos da PCA dos parâmetros relacionados à presença de nutrientes (fósforo total, nitrogênio total e DBO) e coliformes fecais em relação a PC1 e PC2 (que explicam 77% de variância dos dados) têm valores negativos, em contraste com os valores positivos assumidos por OD. Esta correlação inversa é também corroborada pelo processo descrito pela **Equação 3.1**, onde a quantidade de matéria orgânica é aumentada devido ao aumento de nutrientes, o que se reflete na diminuição de oxigênio dissolvido no sistema aquático.

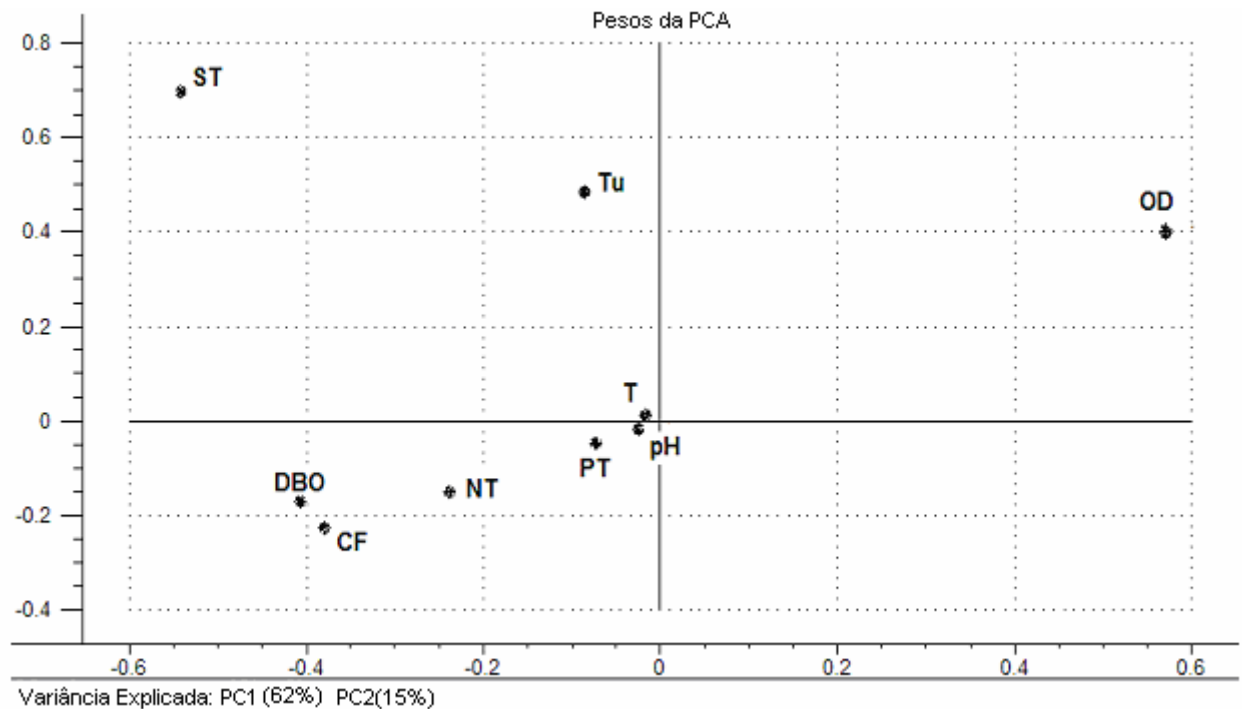


Figura 3.6 – Gráfico dos pesos das variáveis nas componentes principais PC1 versus PC2, ponderadas segundo o $IQA_{BASCARÓN}$, onde: OD (Oxigênio Dissolvido); CF (Coliforme Fecal); pH; DBO (Demanda Bioquímica de Oxigênio); NT (Nitrogênio Total); PT (Fósforo Total); Tu (Turbidez); ST (Sólidos Totais); T (Temperatura).

➤ Ao levar em consideração a variabilidade dos dados em questão, os parâmetros temperatura e pH (**Figura 3.6**) têm pouca influência para a definição das componentes principais PC1 e PC2, uma vez que seus pesos se aproximam de zero em ambas as PCs. Contudo, esse comportamento não pode ser generalizado para outros bancos de dados, já que esses são parâmetros importantes para os equilíbrios que regulam a qualidade da água, e juntos com os demais parâmetros complementam a informação sobre a dinâmica química dos sistemas hídricos em geral.

➤ **PCA ponderada usando os pesos do IQA_{CETESB}**

A **Figura 3.7** mostra o gráfico dos escores de PC1 (50% da variância explicada) *versus* PC2 (12 % da variância explicada).

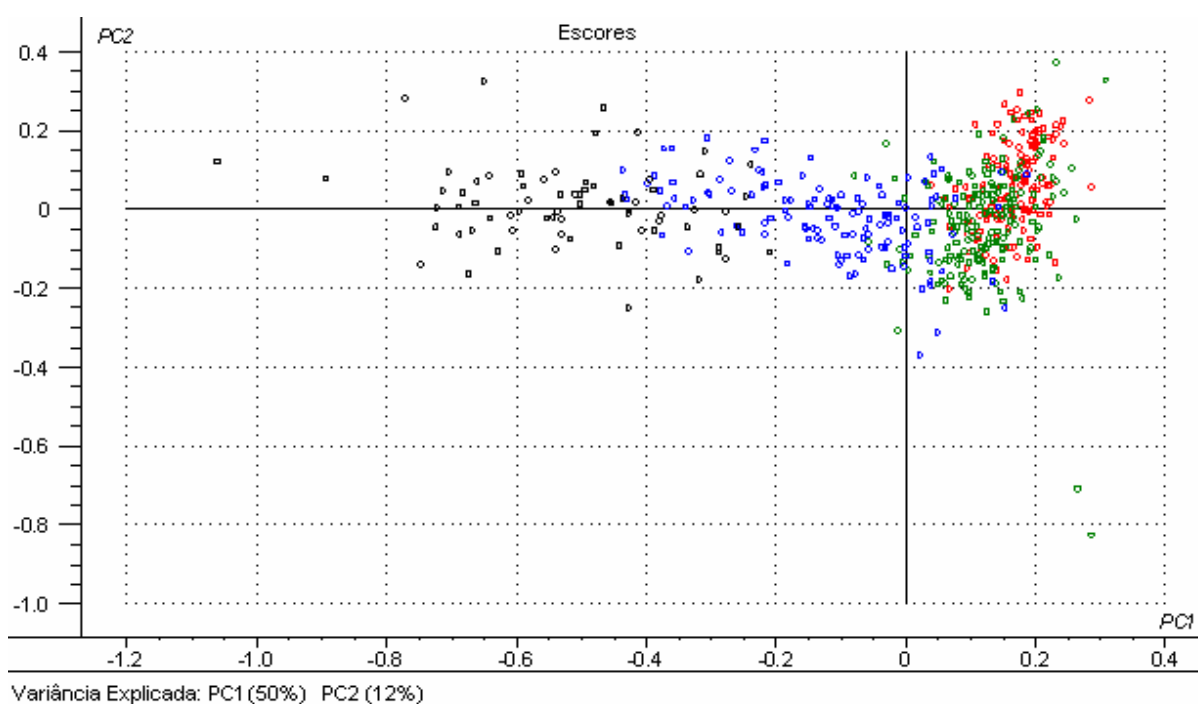


Figura 3.7 – Gráfico dos escores para as duas primeiras PCs, ponderadas segundo o IQA_{CETESB} (■ bom; ■ razoável; ■ ruim; ■ péssima).

Constata-se também a mesma relação entre a qualidade da água (**Figura 3.7**) e os escores ao longo de PC1, embora a quantidade de variância explicada pelas duas primeiras componentes (**Figura 3.8**), nesse caso, seja menor (62%) que no caso da PCA anterior (77%). Notamos ainda que a variação nos valores de escores de PC1 neste caso é bem mais estreita variando de – 1,1 a 0,4.

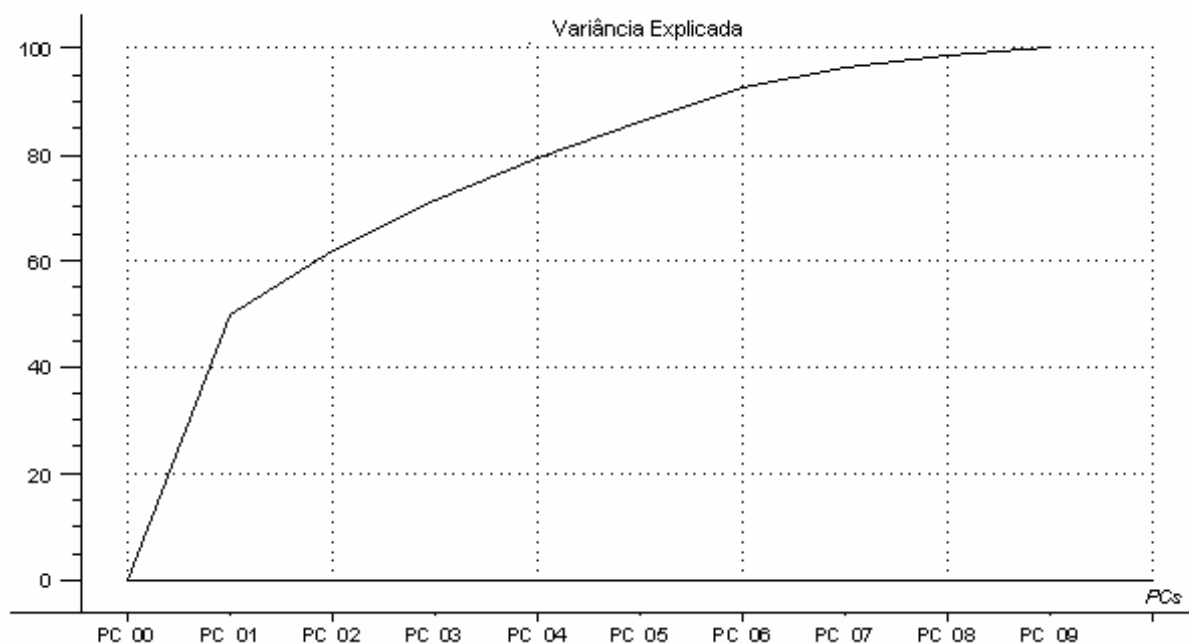


Figura 3.8 – Gráfico da variância explicada versus o número de componentes principais ponderadas segundo IQA_{CETESB}

Em relação aos pesos da PC1 (**Figura 3.9**), constatou-se que os nutrientes, DBO e coliformes fecais apresentam também valores negativos de pesos em PCA, em contraste com os valores positivos de OD, como na PCA ponderada com pesos do $IQA_{BASCARÓN}$, compatível com o esperado para o sistema aquático.

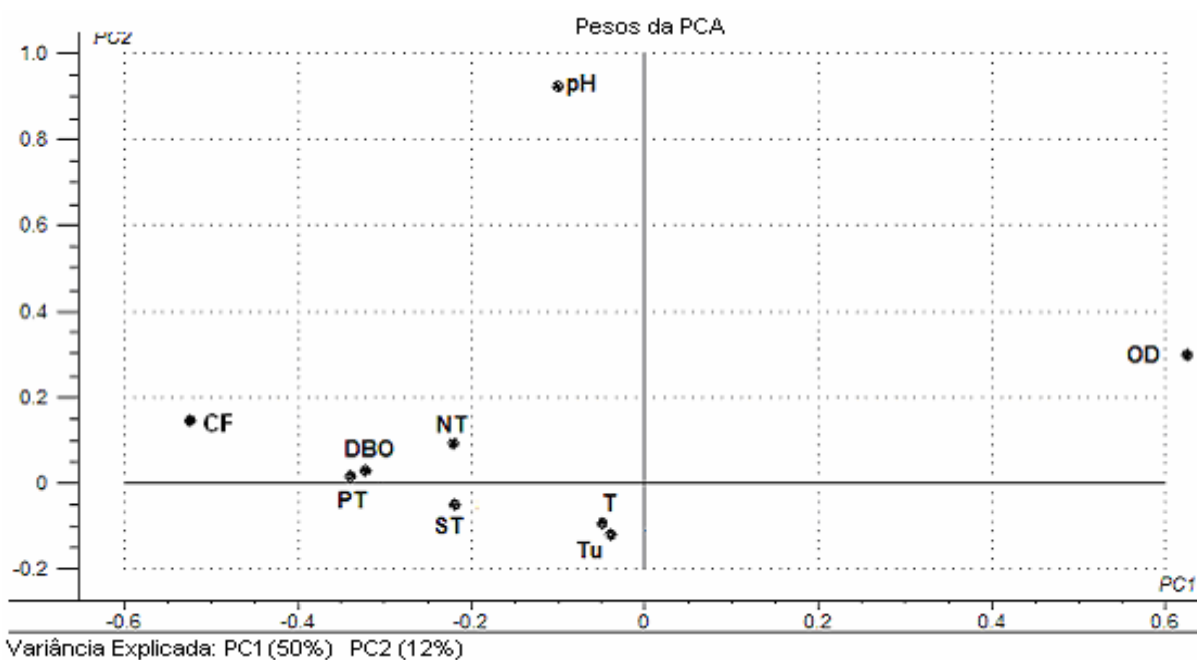


Figura 3.9 – Gráfico dos pesos das variáveis nas componentes principais PC1 versus PC2, ponderadas segundo o IQA_{CETESB} , onde: OD (Oxigênio Dissolvido); CF (Coliforme Fecal); pH; DBO (Demanda Bioquímica de Oxigênio); NT (Nitrogênio Total); PT (Fósforo Total); Tu (Turbidez); ST (Sólidos Totais); T (Temperatura).

Temperatura e pH se relacionam de maneira oposta em PC2 (**Figura 3.9**) provavelmente porque os pesos do IQA_{CETESB} (**Tabela 2.2**) dados a estas variáveis são diferentes, embora que, inferir sobre a segunda componente principal não seja tão relevante para esta discussão em específico, uma vez que a mesma não porta a informação que caracteriza a qualidade da água.

Ao comparar os resultados das PCAs ponderadas usando os pesos de $IQA_{BASCARÓN}$ e de IQA_{CETESB} , observamos semelhanças acerca da informação qualitativa da qualidade da água. Todavia, como a PCA ponderada com os pesos do $IQA_{BASCARÓN}$ portam mais informação, ou seja, apresenta uma maior variância explicada (77%) nas duas primeiras componentes, PC1 e PC2. Ela explica melhor a relação entre os parâmetros e a sua importância dentro do contexto de qualidade de água.

4 – CONCLUSÕES E PROPOSTAS FUTURAS

A proposta deste trabalho foi implementar nas metodologias de análise multivariada: HCA e PCA, quando aplicadas a parâmetros de qualidade de águas, a mesma atribuição de importância relativa dos parâmetros de qualidade usadas nos IQAs. Para tanto foram utilizados dois conjuntos de pesos referentes a IQAs diferentes, ou seja, os pesos do IQA_{CETESB} e os pesos do IQA_{BASCARÓN}.

A metodologia utilizada neste trabalho para a indicação da qualidade da água se assemelha à metodologia dos IQAs em geral, com a vantagem de levar em consideração toda a variabilidade dos dados e a correlação entre as variáveis envolvidas.

Na HCA ponderada observou-se a formação de agrupamentos (G1 e G2) que indicaram de forma geral o comportamento da água, diagnosticando sua qualidade como “melhor” (G1) e “pior” (G2). Utilizando a HCA ponderada com os pesos do IQA_{BASCARÓN} esta separação foi vista de forma mais clara.

Com os resultados da PCA ponderada foi possível estabelecer uma clara relação entre a variação de qualidade e os escores de PC1. Quando os pesos do IQA_{BASCARÓN} foram usados PC1 apresentou 62% da variância explicada, enquanto que quando os pesos do IQA_{CETESB} foram usados PC1 apresentou 50%. A PC1 obtida sem ponderação apresenta apenas 42% da variância explicada e discrimina menos as amostras de classes de qualidade diferentes.

Os pesos da PCA ponderada apresentaram concordância com o comportamento esperado, tendo em conta a dinâmica dos sistemas aquáticos. Por exemplo, evidencia-se a correlação inversa entre o parâmetro OD e os parâmetros relacionados à presença de nutrientes (fósforo total, nitrogênio total e DBO) e coliformes fecais.

A proposta de implementação de pesos às metodologias de análise exploratória leva em consideração as características específicas da região em estudo, já que utiliza dados históricos que portam essas informações. Não é razoável utilizar curvas de qualidade obtidas com dados de regiões diferentes como é o caso das curvas utilizadas no IQA_{CETESB}, que se referem a dados históricos provenientes da NSF-US.

Finalmente, pode-se afirmar que proposta apresentada é uma ferramenta útil para indicar a qualidade da água e para contribuir para idéia de usar ponderação em análise multivariada de dados ambientais de natureza hídrica. Contudo, ela ainda precisa ser avaliada utilizando outros conjuntos de dados, de preferência de regiões distintas, para testar sua reprodutibilidade.

Como propostas futuras espera-se:

Complementar as informações geradas pela HCA e PCA ponderadas através da construção de cartas de controle estatístico para que sejam detectadas anomalias quanto aos programas de controle de qualidade de sistemas hídricos.

Otimizar os conjuntos de pesos utilizados, de modo a tornar menos subjetiva possível sua atribuição.

5 - REFERÊNCIAS BIBLIOGRÁFICAS

1. ROCHA, J. C.; ROSA, A. H.; CARDOSO, A. A. *Introdução a química ambiental*. 1ª ed. Porto Alegre: Bookman, 2004.
2. WEBER, R., Sistemas costeiros e oceânicos, *Química Nova*, **15(12): 137, 1992**.
3. GUIMARÃES, J.R.; NOUR, E. D. A., Tratando nossos esgotos: processos que imitam a natureza, *Caderno Temáticos de Química Nova Escola*, **01: 19, 2001**.
4. LESER, W. S. et alii. *Elementos de epidemiologia geral*. São Paulo: Atheneu, 1985.
5. Atlas de saneamento/IBGE. Rio de Janeiro, 2004.
6. SILVA, A. C. DA COSTA., Novos Rumos para o Saneamento - É Hora de Sermos Responsáveis, *Revista Brasileira de Saneamento e Meio Ambiente*, **1(18): 5, 2001**.
7. LOCATELLI, N. D.; PINHEIRO, A., Avaliação da qualidade das águas dos mananciais da bacia do rio Itajaí. Anais do XV simpósio brasileiro de recursos hídricos, Curitiba-PR, 2003, v. 01, p. 289-302.
8. Resolução CONAMA Nº 357, de 17 de março de 2005. CONSELHO NACIONAL DO MEIO AMBIENTE. Disponível em: <<http://www.mma.gov.br/conama/>>. Acesso em: 28 de julho de 2006.
9. Portaria N.º 518, de 25 de março de 2004. MINISTÉRIO DA SAÚDE. Disponível em: <<http://www.uniagua.org.br/website/images/destaque/portaria518.pdf>>. Acesso em: 28 de julho de 2006.
10. GRASSI, M. T., As águas do planeta Terra, *Cadernos Temáticos de Química Nova na Escola (Meio Ambiente)*, **1(1): 31, 2001**.
11. FREITAS, M. B.; BRILHANTE, O. M. & ALMEIDA, L. M., Importância da Análise de Água para a Saúde Pública em Duas Regiões do Estado do Rio de Janeiro: Enfoque para Coliformes Fecais, Nitrato e Alumínio, *Cad. Saúde Pública*, **17 (3): 651, 2001**.
12. <<http://bcn.boulder.co.us/basin/watershed/index.html>>. Acesso em 12 de junho de 2006.

13. APHA. Standard methods for the examination of water and wastewater. 20. ed. Washington: American Public Health Association, 1998.
14. MACEDO, J. A. *Introdução à química ambiental – química e meio ambiente e sociedade*, 1ª ed. Juiz de Fora: CRQ-MG, 2002.
15. BAIRD, C. *Química ambiental*, tradução da 2. ed. Porto Alegre: Bookman, 2002.
16. VON SPERLING, M. *Introdução à qualidade das águas e ao tratamento de esgotos*, 2. ed. Belo horizonte: DESA / UFMG, 1996.
17. MANAHAN, S. E. *Fundamentals of environmental chemistry*, 2. ed. Boca Raton: CRC Press LLC, 1994.
18. MACEDO, J. A. *Métodos laboratoriais de análises físico-químicas e bacteriológicas*. 2. ed. Belo Horizonte: CRQ-MG, 2003.
19. KHAN, F.; HUSAIN, T.; LUMB, A., Water quality evaluation and trend analysis in selected watersheds of the Atlantic region of Canada, *Environmental Monitoring and Assessment*, **88: 221, 2003**.
20. ABBASI, S. A., Water quality indices state-of-the-art, Scientific Contribution No.-INCOH/SAR-25/2002, Published by National Institute of Hydrology, 2002. Disponível em: <http://scholar.google.com/scholar?hl=en&lr=&q=cache:xY3Yu5Qe4DUJ:www.nih.ernet.in/general/Water%2520Quality%2520Indices.doc+water+quality+index+++ABBASI,+S.+A.>>. Acesso em 20 de março de 2006.
21. http://www3.dsi.uminho.pt/gavea/delphi/ver_informacao_login.htm>. Acesso em 14 de junho de 2006.
22. ABBASI, S. A.; AYRA, D. S. *Environmental impact assessment: available techniques, emerging trends*. Delhi: Discovery, 2000.
23. OWN, R.M. et alii., A water quality index—do we dare?, *Water Sewage Works*, **11: 339, 1970**.
24. SILVA, G. S., JARDIM, W.F., Um novo índice de qualidade de águas para a proteção da vida aquática aplicada ao rio Atibaia, região de Campinas/ Paulínia-SP, *Quim. Nova*, **29(4): 689, 2006**.
25. Flores, J. C., Comments to the use of water quality indices to verify the impact of cordoba city (argentina) on suquia river, *Water Res.*, **36: 4664, 2002**.
26. Smith, D. G., A better water quality indexing system for river and streams, *Water Res.*, **24: 1237, 1990**.

27. SMITH, D. G. A., New form of water quality index for rivers and streams, *Water Sci. Technol*, 21(2): 123, 1989.
28. OTT, W. R. *Environmental indices: theory and practice*. Michigan: Ann Arbor Sc., 1978.
29. WALSKI, T.M., PARKER, F.L., Consumers water quality index, *J. Environ. Eng. Div.*, 1: **593, 1974**.
30. BERNARDES, R. S.; ABREU, L. M.; DORNAS, M., O índice da qualidade da água distribuída à população do Distrito Federal. Anais do XX Congresso nacional de engenharia sanitária e ambiental, Salvador-BA, 2004, p. 1099-1107.
31. PENNA, Y. M. A. D. *Aplicação do índice de qualidade de água multiplicativo*. São Paulo: CETESB, 1977.
32. CONESA Fdez-Vitora, V. *Guia metodológico para la evolucion del impacto ambiental*. 2 ed. Madri: Mundi-Prensa, 1995.
33. CARVALHO, R. A.; SCHLITTLER, M. H. F.; TORNISIELO, V. L., Relações da atividade agropecuária com parâmetros físico-químicos da água, *Quím. Nova*, **23(5): 618, 2000**.
34. Seleção de índice e indicadores da qualidade de água, Recife: CPRH, 2003. Disponível em: <http://www.cprh.pe.gov.br/frme-index-secao.asp?idconteudo=460>>. Acesso em 14 de junho de 2006.
35. Companhia de Tecnologia de Saneamento Ambiental (CETESB). *Watershed management 2000–Brasil/ Canadá*, São Paulo: CETESB, 2000.
36. Avaliação da qualidade das águas: relatório técnico/avaliação ambiental, Bahia: CRA, 2001. Disponível em: <http://www.seia.ba.gov.br/SGDIA/transarg/arquivos/Bacia%20Hidrogr%C3%A1fica/AVALIA%C3%87%C3%83O%20DA%20QUALIDADE%20DAS%20%C3%81GUAS%20-%202001/ARQUIVO/sumario2001.pdf>>. Acesso em: 14 de junho de 2006.
37. MACEDO, JR. G. M.; COSTA, C. E. F. S.; NETO, I. C., Avaliação hidroquímica e da qualidade das águas de um trecho do rio açu, rio grande do norte, *Revista de Geologia*, **16 (2): 27, 2003**.
38. Monitoramento da qualidade de água. Paraná: SUDERHSA. Disponível em: http://www.pr.gov.br/meioambiente/suderhsa/monit_quali_igq.shtml>. Acesso em: 15 de junho de 2006.

39. PESCE, S. F.; WUNDERLIN, D. A., Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquia River, *Water Res.*, **34(11): 2915, 2000**.
40. Banco de dados, CRIA. Disponível em: <<http://hydro.cria.org.br/iqa>>. Acesso em: 18 de junho de 2006.
41. Monitoramento da qualidade das águas dos rios da região metropolitana de Curitiba, no período de 1992 a 2005. Curitiba: 2005. Disponível em: <http://www.pr.gov.br/meioambiente/iap/pdf/rel_monit_qual_aguas_rios_rmc_9205.pdf>. Acesso em: 19 de maio de 2006.
42. FERREIRA, M. M. C., et alii., Quimiometria I: calibração multivariada, um tutorial, *Quím. Nova*, **22 (5): 724, 1999**.
43. BRUNS, R. E.; FAIGLE, J. F. G., Quimiometria, *Quím. Nova*, **8: 84, 1985**.
44. BEEBE K. R.; PELL, R. J.; SEASHOLTZ, M. B., *Chemometric: a practical guide*. New York: John Wiley & Sons, 1998.
45. NETO, B. B.; SCARMINIO, I. S.; BRUNS, R.E. *Como fazer experimentos*. 1ª ed. Campinas: UNICAMP, 2001.
46. VANDEGINSTE, B. G. M. et alii. *Handbook of chemometrics and qualimetrics: parte B*. Amsterdam: Elsevier, 1998.
47. STATSOFT, Manual do Usuário, Statistica For Windows versão 6.0, N. York, 1998.
48. HAIR, JR. et alii. *Análise multivariada de dados*, 5. ed. Porto Alegre: Bookman, 2005.
49. WOLD, S.; ESBENSEN, K.; GELADI, P., Principal Component Analysis, *Chemom. Intell. Lab. Syst.*, **2: 37, 1987**.
50. MEDEIROS, V. M., et alii., Screening analysis of river seston downstream of na effluent discharge point using near-infrared reflectance spectrometry and wavelet-based spectral region selection, *Water Research*, **39: 3089, 2005**.
51. VEGA, M. et. alii., Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis, *Water. Research*, **32 (12): 1998**.
52. BENGRAINE, K.; MARHABA, T.F., Using principal component analysis to monitor spacial and temporal changes in water quality. *Journal of Hazardous Materials*, **100: 179, 2003**.
53. MARENGO, E. et. alii., How Chemometrics can helpfully assist in evaluating environmental data. Lagoon water, *Analytica Chimica Acta*, **317: 53, 1995**.

54. ANDRADE, J. M. et. alii., Selection of analytical variables to optimize laboratory efforts in future groundwater studies, *Analytica Chimica Acta*, **292: 253, 1994**.
55. PARINET, B.; LHOTE, A.; LEGUBE, B., Principal component analysis: an appropriate tool for water quality evaluation and management application to a tropical lake system, *Ecological Modelling*, **178: 295, 2004**.
56. Índice de qualidade de água. Disponível em: <http://www.cetesb.sp.gov.br/Agua/rios/indice_iap_iqa.asp>. Acesso em: 29 de julho de 2006.

ANEXOS

1. Curvas dos sub-índices de qualidade dos parâmetros utilizados no IQA_{NSF}:

