

UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA



ANÁLISE DE QUALIDADE DE ÁGUA UTILIZANDO
REDE NEURAL ARTIFICIAL MULTICAMADAS

AÉRCIO CAVALCANTI DE MELO FILHO

Recife – PE
2006

Aécio Cavalcanti de Melo Filho

Análise de qualidade de água utilizando rede neural artificial multicamadas.

Monografia apresentada ao curso de Bacharelado em Ciência da Computação da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do grau de Bacharel em Ciência da Computação.

Orientadora: Prof^a. Dr^a. Teresa Bernarda Ludermir

Co-Orientador: Prof. Dr. Francisco de Assis Tenório de Carvalho

Recife – PE
2006

Aécio Cavalcanti de Melo Filho

**Análise de qualidade de água utilizando rede neural
artificial multicamadas.**

Monografia apresentada ao curso de Bacharelado em
Ciência da Computação da Universidade Federal de
Pernambuco, como parte dos requisitos para obtenção
do grau de Bacharel em Ciência da Computação.
Orientadora: Prof^a. Dr^a. Teresa Bernarda Ludermir
Co-Orientador: Prof. Dr. Francisco de Assis Tenório de
Carvalho

Aprovado em ___/___/___

BANCA EXAMINADORA

Prof^a. Dr^a. Teresa Bernarda Ludermir

Prof. Dr. Aluízio Fausto Ribeiro Araújo

Este trabalho é dedicado a minha família:
pai, mãe e irmãs e a minha noiva. Pois,
além de me darem todas as oportunidades
sempre me apoiaram e aceitaram minhas
escolhas mesmo quando não as
entendiam.

Agradecimentos

Agradeço aos meus pais, Aécio e Mércia, por sempre estarem presentes me apoiando nos momentos difíceis e por terem me dado todo o apoio que precisei para chegar até aqui, respeitando minhas decisões mesmo quando não as entendiam. Gostaria de agradecer também a minhas irmãs queridas, Lulu e Ró, que sempre cuidaram de mim e por sempre incentivar meus estudos.

Agradecer a Rafaela, minha noiva querida que tanto quero bem, por estar ao meu lado há mais de três anos, participando de decisões importantes da minha vida. Por compreender a minha dedicação para com o curso de Ciência da Computação, mesmo que, às vezes, chateada pela minha ausência.

Agradeço aos meus amigos de infância que tiveram sua participação na formação de meu caráter. Foram tantas brincadeiras, partidas de futebol, passeios de bicicleta, brigas, pazes, companheirismo, etc. Entre eles Filipe (peixe), Pedro (peu), Tiago (Zé Dumba), Rhand (com esse nome não precisa de apelido) e João Paulo (testa). A lista é vasta, por tanto um agradecimento especial a Filipe (peixe) e Lula, companheiros, nos momentos bons e ruins. Muito obrigado a vocês meus amigos.

Como esquecer dos amigos feitos na faculdade? Foram cinco anos de convivência com as mais diferentes figuras. Agradeço a André, Afonso, Túlio, Vicente, Vilmar, Ives, Paulo e também à Léo e Renan que na fase final da faculdade se mostraram bons amigos. Foram noites e noites acordados desenvolvendo os “super projetos”. Agradecer também a Alexandre, Andrey, Fausto, Igor e a todos os outros de minha sala e do centro de informática.

Um agradecimento aos professores do centro de informática e em especial para Geber, Chico, Teresa e Aluizio que despertaram em mim o gosto por inteligência artificial.

Para finalizar, gostaria de agradecer aos meus avós: Lucy, Lita e Luis que sempre foram fontes de inspiração pelas suas belas histórias de vida.

Obrigado a todos vocês, por fazerem parte da minha vida.

Resumo

O uso de métodos computacionais, tais como redes neurais artificiais (RNA), para reconhecimento de padrões da qualidade da água em ecossistemas têm sido bastante utilizados, pois classificam de forma eficiente e rápida as amostras de águas coletadas. Este trabalho propõe uma solução baseada em uma RNA multicamadas para classificação da qualidade da água. Os resultados obtidos por este trabalho serão comparados com os resultados de outros autores.

Sumário

1	INTRODUÇÃO.....	10
2	ESTADO DA ARTE	12
2.1	REDES NEURAIS ARTIFICIAIS	12
3	PROBLEMA PROPOSTO	15
4	ANÁLISE DOS DADOS.....	17
4.1	PRÉ-PROCESSAMENTO DOS DADOS	18
4.2	BALANCEAMENTO DAS CLASSES	20
5	METODOLOGIA	22
5.1	PROTÓTIPO.....	22
5.2	EXPERIMENTOS	23
5.3	ANALISE DOS EXPERIMENTOS	24
6	CONCLUSÕES.....	29
	REFERÊNCIAS BIBLIOGRÁFICAS	30
	ANEXOS	33
	ANEXO I: SIGNIFICADO AMBIENTAL DOS PARÂMETROS DO IQA.....	33
	ANEXO II: GRÁFICOS DE ÍNDICE DE QUALIDADE DOS PARÂMETROS DO IQA.....	39

Índice de Tabelas

TABELA 1: PESOS DOS PARÂMETROS DE QUALIDADE DE ÁGUA SEGUNDO NSF.	17
TABELA 2: CLASSIFICAÇÃO DA ÁGUA DE ACORDO COM O IQA.	18
TABELA 3: DIVISÃO DE DADOS EM CLASSES DE QUALIDADE DE ÁGUA.....	18
TABELA 4: QUANTIDADE DE AMOSTRAS AFETADAS POR VALORES AUSENTES.	19
TABELA 5: QUANTIDADE DE AMOSTRAS AFETADAS POR VALORES ANORMAIS.	19
TABELA 6: QUANTIDADE DE AMOSTRAS AFETADAS POR VALORES NÃO-ATÔMICOS.....	20
TABELA 7: REPLICAÇÃO DOS DADOS DE QUALIDADE DE ÁGUA.	20
TABELA 8: ESTRATIFICAÇÃO DOS DADOS DE QUALIDADE DE ÁGUA.	21
TABELA 9: DESCRIÇÃO DA TOPOLOGIA DA REDE.	25
TABELA 10: RESULTADOS DO TRABALHO PROPOSTO.....	25
TABELA 11: RESULTADOS DE FERREIRA.	27
TABELA 12: RESULTADOS DE NEPOMUCENO.	27
TABELA 13: RESULTADOS DO <i>BENCHMARK</i>	28

1 Introdução

Neste início de século, uma preocupação recorrente e justificada é a de que a água, tão abundante, em nosso planeta torne-se cada vez mais escassa para uso humano. Sem contar com a má distribuição geográfica, o maior problema encontrado hoje é a poluição dos rios e lagos. Esta ínfima porção de água doce – se comparada com porção salgada, mais abundante – corresponde apenas a 1% de toda água no mundo e está sendo destruída pela má gestão dos recursos hídricos. Executar a análise da água utilizando métodos computacionais é uma forma rápida e eficiente para obter informação sobre a condição da água em um determinado ponto.

As redes neurais, com sua habilidade de derivar o significado de dados complexos ou imprecisos, podem ser usadas para extrair padrões e detectar tendências que são muito complexas para serem observadas por humanos ou por outras técnicas computacionais. Uma rede neural treinada pode ser comparada a um especialista na categoria de informação que a rede recebeu para analisar. Outras vantagens da aprendizagem computacional, em especial redes neurais, são:

1. Aprendizagem adaptável: Uma habilidade de aprender novos padrões e tarefas baseadas nos dados disponibilizados para treinamento ou a partir de experiência inicial.
2. Auto-organização: Uma rede neural artificial é capaz de criar sua própria organização ou representação da informação que recebe durante o tempo da aprendizagem.

O uso de métodos computacionais, tais como redes neurais (RNA), para reconhecimento de padrões da qualidade da água em ecossistemas de forma automática, utilizando indicadores físicos, químicos e biológicos, é de fundamental importância para o monitoramento da qualidade da água, pois classificam de forma rápida e eficiente as amostras de água coletadas e analisadas. Esta automação é de fundamental importância, pois além de facilitar todo o trabalho de

análise permitirá que esta classificação seja feita pelo próprio pessoal envolvido no acompanhamento desta tarefa sem a necessidade constante de um especialista na área, acarretando uma diminuição dos custos associados a essa tarefa.

Na literatura[2],[5], [12],[14], [18] e [19], os indicadores mais relevantes passíveis de análise a serem utilizados são as concentrações de: clorofila, sedimentos suspensos, amoníaco, cádmio, cromo-total, níquel, chumbo, cobre e zinco; temperatura; condutância específica; oxigênio dissolvido e pH, além da concentração de determinadas espécies de microorganismos nas amostras de água. De posse desses indicadores, ou de um subconjunto dos mesmos, é possível classificar, de forma automática, as amostras de água coletadas nos rios de São Paulo.

A proposta deste trabalho é a criação de um sistema para classificação de água utilizando uma rede neural supervisionada de tipo multicamadas com algoritmo de retro propagação do erro. Os dados utilizados para treinamento e testes da rede são extraídos a partir de medições regulares realizadas pela CETESB em rios do estado de São Paulo. O resultado desse treinamento será a classificação da água nos vários pontos de coleta como: excelente, boa, ruim e muito ruim. Os resultados obtidos serão comparados com os resultados de outras abordagens para resolução do problema de classificação de água, como é o caso do trabalho de FERREIRA[7] e o de NEPOMUCENO[13]. Será feita, também, a comparação deste trabalho com uma técnica de *benchmark*, obtido a partir de um *software* de mineração de dados[17].

2 Estado da arte

Métodos computacionais, tais como redes neurais, já foram aplicados para monitorar a qualidade da água, como por exemplo, a análise de sedimentos suspensos e de clorofila no Lago *Kasumi Gaura*[2], ou utilizando o nível da água, temperatura e concentração de oxigênio nos rios *Cooper* e *Wando*[5]. Nesses casos foram utilizadas redes neurais artificiais com retro propagação do erro (*backpropagation*).

Outras metodologias foram aplicadas ao monitoramento da qualidade da água, utilizando redes neurais artificiais auto-organizáveis (SOM). Nos rios do Reino Unido[12], em dois rios em Singapura[19], no rio *Seine*[18] e ainda em pesquisas oceanográficas da FURG sobre a qualidade da água em nosso ecossistema costeiro[14].

Os parâmetros utilizados para testar os sistemas citados acima foram: pH, oxigênio dissolvido e temperatura[19], temperatura, pH e a concentração de NH_4 e NO_2 [13],[6] e ainda sedimentos suspensos[2] e temperatura e concentração de oxigênio[5].

Nos trabalhos acima citados os modelos neurais escolhidos foram capazes de classificar, satisfatoriamente, a qualidade de água. Em estudo comparativo YONG[19] avaliou a utilização de dois modelos, uma rede multicamadas (MLP) e uma rede do tipo SOM (*self organizing maps*) e concluiu que os dois modelos sugeridos possuem classificação eficiente para o problema atacado.

2.1 Redes Neurais Artificiais

Redes neurais artificiais (RNAs) experimentaram um renascimento nos últimos quinze anos, desde a descoberta e popularização do algoritmo de retropropagação do erro (*error backpropagation*) [9] em 1986. Antes desse algoritmo, o processo de treinamento de uma rede neural era muito complexo e geralmente eles eram limitados a resolver problemas linearmente separáveis.

Redes neurais artificiais imitam o comportamento de sistemas neurais biológicos de forma simplificada. Elas são representadas por um número de neurônios (similar às células cerebrais) conectados entre si através da rede por um conjunto de ligações (similar aos axônios). O ajuste dos pesos dessas ligações entre neurônios (similar às sinapses) é capaz de treinar uma rede para se comportar de várias maneiras diferentes.

A partir dessa representação simplista, é possível implementar um número incontável de modelos de redes neurais. Esses modelos podem ser construídos a partir da organização de neurônios em camadas ou grupos, por exemplo. As ligações entre os neurônios podem ser configuradas para carregar dados em diferentes direções através da rede. Diferentes tipos de neurônios, com diferentes tipos de ativação podem ser usados na rede, bem como diferentes tipos de algoritmos de treinamento podem ser aplicados em uma tentativa de melhora no desempenho da rede.

Entretanto existem algumas desvantagens associadas às redes MLP utilizando o algoritmo *backpropagation*. Embora esse algoritmo represente uma poderosa ferramenta para treinamento, eles apresentam uma baixa velocidade de convergência. O objetivo desses algoritmos é minimizar uma função custo, atualizando os fatores de ponderação (pesos) na direção de erro mínimo. Em problemas que podem ocorrer mínimos locais, o algoritmo de retro propagação pode convergir e ficar “encalhado”. Além disso, segundo Hecht-Nielsen[10], um outro problema associado à retro propagação é a existência de regiões muito planas na superfície de erro onde o gradiente é pequeno. Considerando uma taxa de aprendizado fixa, como é o caso, isto significará necessariamente uma convergência lenta. Outra desvantagem é a definição da topologia da rede que geralmente é feita por tentativa e erro, acarretando grande esforço envolvido para a definição de uma estrutura satisfatória.

Apesar de existirem diversas estruturas de redes neurais, a mais popular é a rede *feedforward* mostrada na Figura 1. Nesta configuração, um conjunto de

entradas é propagado pela rede a partir da camada de entrada, passando pelas camadas escondidas até a camada de saída. Esses dados passados para rede são modificados de acordo com os pesos em cada conexão entre os neurônios. Em cada neurônio os valores de suas entradas são combinados e modificados por uma determinada função, por exemplo, sigmóide. É calculada então a saída do neurônio que é passada para próxima camada da rede. Ao término desses cálculos à rede fornece uma saída (resposta).

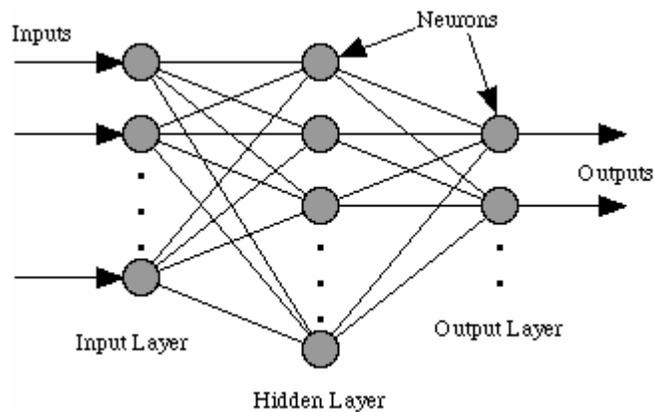


Figura 1: Estrutura de uma rede neural.

Rede Multicamadas (MLP)

A MLP seja, talvez, a arquitetura de rede neural mais utilizada. É uma rede do tipo *feedforward* composta por neurônios interconectados entre si (ver Figura 1). Geralmente essas redes são treinadas com o algoritmo *backpropagation*[9]. Este algoritmo funciona iterativamente alterando os pesos das conexões entre neurônios (proporcionalmente a taxa de aprendizagem da rede) até que o erro da saída da rede seja minimizado. Este algoritmo pode, ainda, ser adaptado de várias formas. Por exemplo, um termo de *momentum* pode ser introduzido na tentativa de acelerar a convergência para um erro mínimo.

Maiores detalhes sobre este algoritmo e sobre MLP estão fora do escopo deste trabalho, já que os mesmos possuem vasta documentação[9] e [20].

3 Problema Proposto

Para a solução do problema de classificação abordado por este trabalho foi escolhida uma rede neural do tipo multicamadas. A MLP foi escolhida devido à sua fácil implementação e também pelo fato de bons resultados terem sido alcançados na literatura [5] e [19] utilizando essa abordagem. Sua estrutura é composta por uma camada de entrada, n camadas escondidas e uma camada de saída. Sua alimentação é do tipo *feedforward* e o algoritmo de treinamento escolhido é o *backpropagation*. Abaixo segue uma descrição desse algoritmo:

Dado um conjunto de padrões de treinamento $P = \{(z_1, d_1), (z_2, d_2), \dots, (z_p, d_p)\}$, onde z_i é $(I \times 1)$, d_i é $(K \times 1)$ e $i = 1, 2, \dots, P$. Temos ainda as camadas escondidas, onde as saídas de seus neurônios são y e y é $(J \times 1)$ e o é $(K \times 1)$.

1. Inicialização da rede

Os parâmetros da rede são inicializados:

- Quantidade de iterações (ciclos)
- Taxa de aprendizagem – η

2. Inicialização dos pesos

Pesos W e V são inicializados randomicamente com valores pequenos (entre 0 e 1). O tamanho de W é $(K \times J)$ e V é $(J \times I)$.

3. Treinamento é inicializado

Um padrão de entrada p é apresentado à camada de entrada e as saídas das camadas da rede são computadas utilizando a função de ativação sigmóide:

$$y_j = f(v_j^t z), \text{ para } j = 1, 2, \dots, J$$

Onde v_j é a j -ésima linha de V e y é a saída dos neurônios das camadas escondidas.

$$o_k = f(v_k^t y), \text{ para } k = 1, 2, \dots, K$$

Onde v_k é a k -ésima linha de W e o é a saída da rede.

4. Erro é calculado

O erro da rede é calculado para cada neurônio da camada de saída.

$$E = \frac{1}{2}(d_k - o_k)^2 + E, \text{ para } k = 1, 2, \dots, K$$

Onde d_k é o valor esperado e o é a saída calculada para o neurônio k .

5. Retro propagação do erro

Sinal de erro é calculado para camada de saída δ_o ($K \times 1$) e para as camadas escondidas δ_y ($J \times 1$).

Para camada de saída, temos:

$$\delta_{ok} = \frac{1}{2}(d_k - o_k)(1 - o_k^2), \text{ para } k = 1, 2, \dots, K$$

E para as camadas escondidas, temos:

$$\delta_{yj} = \frac{1}{2}(1 - y_j^2) \sum \delta_{ok} w_{kj}, \text{ para } j = 1, 2, \dots, J$$

Com os sinais de erro calculados devemos reajustar os pesos da rede.

6. Ajuste de pesos

O ajuste dos pesos é feito em dois passos. O primeiro passo é ajustar os pesos da camada de saída:

$$w_{kj} = w_{kj} + \eta \delta_{ok} y_j, \text{ para } k = 1, 2, \dots, K \text{ e } j = 1, 2, \dots, J$$

Com os pesos da camada de saída calculados, reajustamos os pesos das camadas anteriores, da seguinte maneira:

$$v_{ji} = v_{ji} + \eta \delta_{yj} z_i, \text{ para } j = 1, 2, \dots, J \text{ e } i = 1, 2, \dots, I$$

7. Verifica conjunto de padrões

Se existir algum padrão, retorna ao passo 3. Caso contrário passa-se ao passo 8.

8. Finaliza treinamento

Se o erro $E < E_{max}$, finalize o treinamento. Caso contrário inicia-se um novo ciclo de treinamento retornando ao passo 3.

4 Análise dos Dados

Os dados utilizados para avaliação do sistema proposto foram retirados de uma pesquisa[3][3] realizada entre fevereiro e dezembro de 2004 nos rios de São Paulo. Esses dados foram escolhidos por possuírem a medição das nove variáveis utilizadas para o cálculo do IQA.

O IQA é composto por nove parâmetros, com seus respectivos pesos (w), que foram fixados em função da sua importância para a classificação da qualidade da água (ver Tabela 1). Além de seu peso (w) cada parâmetro possui um valor de qualidade (q) obtido do respectivo gráfico de qualidade, em função de sua concentração ou medida (ver Anexo II).

PARÂMETROS	PESOS
Oxigênio Dissolvido	$w = 0,17$
Coliformes Fecais	$w = 0,15$
Potencial Hidrogeniônico (pH)	$w = 0,12$
Demanda Bioquímica de Oxigênio ($DBO_{5,20}$)	$w = 0,10$
Temperatura	$w = 0,10$
Nitrogênio Total	$w = 0,10$
Fósforo Total	$w = 0,10$
Turbidez	$w = 0,08$
Resíduo Total	$w = 0,08$

Tabela 1: Pesos dos parâmetros de qualidade de água segundo NSF.

O IQA é então calculado, utilizando a fórmula abaixo, onde q_i é a qualidade do i -ésimo parâmetro e w_i é o peso correspondente ao i -ésimo parâmetro:

$$IQA = \prod_{i=1}^n q_i^{w_i}$$

A partir do valor do IQA, define-se a classificação da água, como descrito na Tabela 2:

VALOR DO IQA	CLASSIFICAÇÃO
Entre 80 - 100	0 – Ótima
Entre 51 - 79	1 – Boa

Entre 37 - 51	2 – Aceitável
Entre 20 - 36	3 – Ruim
Entre 0 - 19	4 – Péssima

Tabela 2: Classificação da água de acordo com o IQA.

Apesar dos dados sobre qualidade de água em rios, encontrados no *site* da CETESB[3] terem sido a fonte mais completa, não foi encontrado nenhum rio da “classe 1” e ainda havia poucos rios com das classes “0, 3 e 4”. Outro problema encontrado foi à presença de *missing values* (ausência de valor para determinado parâmetro) e *outliers* (valor muito distorcido em relação à média de um determinado parâmetro). A base de dados apresentou, então, as seguintes características:

CLASSE	QUANTIDADE
0 – Ótima	19
1 – Boa	0
2 – Aceitável	148
3 – Ruim	24
4 – Péssima	42

Tabela 3: Divisão de dados em classes de qualidade de água.

Como podemos observar na Tabela 3, existe uma grande desproporção na distribuição dos dados disponíveis. Essa desproporção pode provocar distorções nos resultados da rede neural, privilegiando as classes de padrões que contém mais amostras.

4.1 Pré-processamento dos dados

Na tentativa de resolução dos problemas acima descritos, foi aplicado um processamento nos dados. Esse processamento consistiu, basicamente, na resolução de valores ausentes (*missing values*) e de valores anormais (*outliers*). Ao final dessa etapa de processamento, algumas tentativas de balanceamento das classes foram feitas.

MISSING VALUES

Os *missing values* ou simplesmente valores ausentes, são caracterizados pela inexistência de informação sobre um parâmetro em um determinado padrão da base de dados. Para resolver esse problema, foi calculada a média de cada parâmetro da classe, substituindo os valores ausentes por esse valor. A Tabela 4 apresenta a quantidade de amostras afetadas:

PARAMETRO	QUANTIDADE DE AMOSTRAS
Fósforo Total	12
NKT	12
Resíduo	27
Coliformes	1
DBO	4
pH	1

Tabela 4: Quantidade de amostras afetadas por valores ausentes.

OUTLIERS

Os *outliers* são valores que fogem do padrão apresentado por um atributo da base de dados. A melhor forma encontrada para resolver esse problema foi substituir o valor anormal pelo máximo ou mínimo valor encontrado para um determinado atributo. Ex: o máximo valor de Coliforme fecal para uma água classificada como “Ótima” é de 500 NMP/100mL, quando encontrado um valor anormal muito maior que 500 (2700, por exemplo) deve-se substituir o valor anormal por 500. No caso de um valor mínimo o problema é atacado da mesma forma que a descrita no exemplo.

PARAMETRO	QUANTIDADE
Fósforo Total	4
NKT	3
Turbidez	10
Coliformes	4
DBO	2
OD	2
pH	2
Temperatura	3

Tabela 5: Quantidade de amostras afetadas por valores anormais.

VALORES NÃO-ATÔMICOS

O problema mais simples de resolver foi o de valores não-atômicos. Na base de dados esses valores foram encontrados em intervalos (ex: $< 0,01$), então para resolvê-los foi aplicada uma discretização dos valores, substituindo o valor não-atômico pelo maior valor possível para o intervalo.

PARAMETRO	QUANTIDADE
Fósforo Total	57
NKT	7
DBO	4
OD	2
Coliformes	2

Tabela 6: Quantidade de amostras afetadas por valores não-atômicos.

4.2 *Balanceamento das classes*

Após o processamento aplicado aos dados, conforme descrito na seção 4.1, foram utilizadas algumas abordagens na tentativa de balancear a quantidade de amostras das classes.

A primeira abordagem utilizada foi a de replicação dos dados, para que a quantidade das amostras com menos padrões pudessem ser equiparadas às amostras de maior expressividade. A tabela 7 mostra a nova disposição dos dados (comparar com Tabela 3).

CLASSIFICAÇÃO	QUANTIDADE DE AMOSTRAS
0 – Ótima	152
2 – Aceitável	154
3 – Ruim	144
4 – Péssima	168

Tabela 7: Replicação dos dados de qualidade de água.

A replicação dos dados acarretou um problema para a rede neural. Com a configuração mostrada acima, foi observado que a rede foi super treinada (*overfitting*), isto é, a rede decorou os padrões das classes que tiveram maior

número de replicações (classes 0 e 3) e não foi capaz de generalizar de forma satisfatória as classes com mais amostras (classes 2 e 4).

Por causa dos maus resultados obtidos, uma segunda abordagem foi utilizada. Nessa abordagem a quantidade de amostras foi definida pela menor quantidade de amostras encontradas dentre as classes. A partir daí os dados foram selecionados randomicamente para compor a nova base de dados. A distribuição da base dá-se como descrito abaixo:

CLASSIFICAÇÃO	QUANTIDADE DE AMOSTRAS
0 – Ótima	19
2 – Aceitável	19
3 – Ruim	19
4 – Péssima	19

Tabela 8: Estratificação dos dados de qualidade de água.

Após a execução de todos os passos acima os dados foram normalizados entre 0 e 1.

5 Metodologia

Nesta seção será discutida a criação de um protótipo de uma MLP, a definição de parâmetros utilizados pela rede e sua topologia. Será discutida ainda a distribuição da base de dados em treinamento e testes. E por fim serão apresentados os resultados do treinamento da rede.

5.1 Protótipo

O protótipo da MLP foi desenvolvido na linguagem JAVA™[11]. A modelagem do sistema é mostrada na figura abaixo:

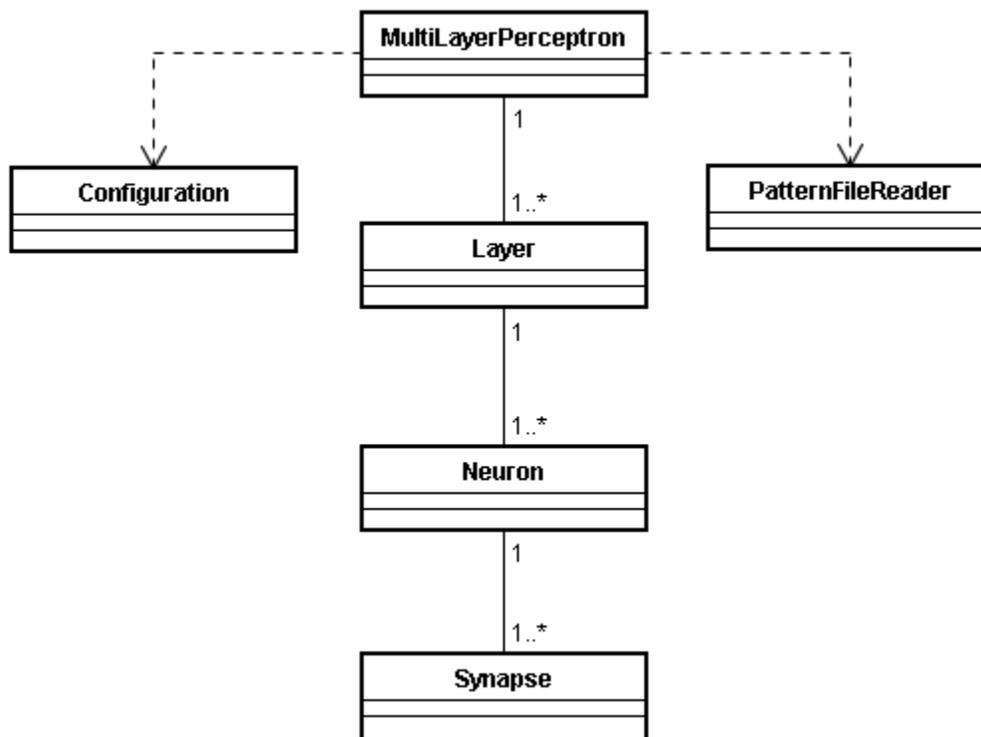


Figura 2: Modelagem do protótipo.

Conforme mostrado na Figura 1, uma rede neural do tipo MLP possui um conjunto de camadas (entrada, escondidas e saída), cada camada possui um conjunto de Neurônios. E por fim cada neurônio se conecta aos outros através de

um conjunto de sinapses. Assim foi modelado o sistema. Segue abaixo a descrição de cada entidade do sistema.

- *MultiLayerPerceptron* – controlador do sistema. Essa entidade realiza a construção da topologia da rede, o treinamento e os testes. É composto por um conjunto de camadas (*Layer*).
- *Layer* – representa uma camada de uma MLP, é composta por um conjunto de neurônios (*Neuron*).
- *Neuron* – possui um conjunto de sinapses de entrada e outro de saída. Essa classe é responsável pelo cálculo da saída do neurônio, bem como da regra de propagação, a regra delta.
- *Synapses* – representa a ligação entre dois neurônios. É nessa entidade que o peso da conexão é inicializado.
- *Configuration* – responsável pela inicialização dos parâmetros configuráveis da rede, tais como: taxa de aprendizagem, *momentum*, *threshold*, número de iterações, número de neurônios na camada de entrada, número de neurônios na camada de saída e, por fim, número de neurônios na camada escondida.
- *PatternFileReader* – entidade auxiliar do sistema. Ela carrega um conjunto de padrões, a partir de um determinado arquivo (treinamento, por exemplo).

5.2 Experimentos

Após a construção deste protótipo, iniciaram-se os experimentos. O primeiro passo foi definir a topologia da rede, isso foi realizado por tentativa e erro. Após várias tentativas a topologia vencedora foi a seguinte configuração:

- 2 camadas escondidas;
- 16 neurônios na camada escondida;
- Taxa de aprendizagem = 0,075
- *Momentum* = 0,9
- *Threshold* = 0,3

A próxima etapa, então, foi o desenvolvimento dos experimentos. Por possuir uma base de dados muito pequena, foi utilizada a técnica *Leave-one-out Cross Validation*[14] uma especialização do *k-Fold Cross Validation*, onde k (número de experimentos) é igual ao número total de padrões de exemplo. Nesta técnica, a cada experimento rodado é escolhido um padrão para testes e o resto é utilizado para treinamento. A Figura 3 descreve o funcionamento dessa técnica.

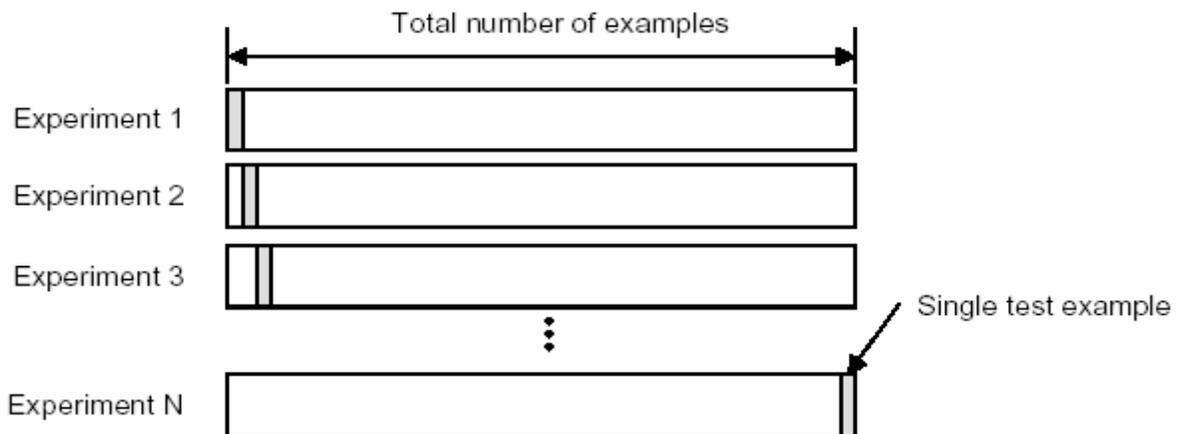


Figura 3: Funcionamento do *Leave-one-out Cross Validation*.

Dessa forma, para cada experimento, a rede era treinada com 75 padrões e testada com um padrão, já que após o tratamento dos dados descrito na seção 4, apenas 76 amostras são disponíveis para realização de experimentos (ver Tabela 8). Vale salientar que a configuração inicial da rede era armazenada (pesos das conexões), para que os experimentos fossem todos baseados na mesma rede neural.

5.3 Análise dos experimentos

Com a configuração descrita até o momento a rede é treinada em menos de seis segundos. Lembrando que para cada experimento a rede é treinada 76 vezes então o tempo médio de execução de um experimento é de oito minutos. Para obter uma quantidade razoável de testes, os experimentos foram repetidos 50 vezes. O tempo total de execução dos testes era de, aproximadamente, seis

horas. Esse tempo pode ser considerado razoável para treinamento e teste de uma rede neural MLP com algoritmo *backpropagation*.

Após o termino dos experimentos descritos acima, foi observado um resultado bastante satisfatório, se levado em conta à pequena quantidade de dados disponíveis. Utilizou-se ainda uma ferramenta para pesquisa em mineração de dados TANAGRA[17] para comparar o modelo proposto neste trabalho. Essa ferramenta possui uma vasta gama de métodos para mineração, como aprendizagem estatística e aprendizagem de máquina, por exemplo.

A Tabela 9 apresenta a configuração da rede utilizada na execução dos experimentos.

PARÂMETRO	VALOR
Nº de iterações	1000
Nº de neurônios na camada de entrada	9
Nº de neurônios na camada escondida	16
Nº de neurônios na camada de saída	4
Nº de camadas escondidas	2
Taxa de aprendizagem	0,075
<i>Threshold</i>	0,3
<i>Momentum</i>	0,9

Tabela 9: Descrição da topologia da rede.

Com a topologia descrita acima o melhor resultado encontrado foi:

CLASSE	AMOSTRAS*	ACERTOS	RECALL
0	950	732	77,05%
2	950	809	85,15%
3	950	768	80,84%
4	950	903	95,06%
TOTAL	3800	3212	84,52%

Tabela 10: Resultados do trabalho proposto.

* A quantidade de amostras consideradas nesta tabela refere-se à quantidade original de amostras (76) multiplicada pelo número de execuções dos experimentos (50).

A classificação dos dados utilizados estava de acordo com o CONAMA[4], que leva em consideração mais de 15 variáveis para classificar a água, entretanto a base disponível possuía muitos valores ausentes para os dados não contemplados por esta pesquisa. Por este motivo foi adotada a medida de classificação proposta pelo ANA[1], utilizando apenas nove variáveis. Isso acarretou uma perda muito grande de informação, desfavorecendo a capacidade de aprendizado da rede. Considerando estes problemas citados e a escassez de dados para treinamento e testes, a taxa de acerto (*Recall*) da rede foi bastante satisfatória.

Pode-se observar uma diferença muito grande na taxa de acerto entre as classes 4 e 0. Essa diferença foi atribuída aos dados, visto que elementos da classe 4 possuem uma quantidade de coliformes fecais bem superior à média das outras classes, esse atributo caracterizou muito bem a classe 4. Já a classe 0 não possuía nenhuma característica que se destaca em relação as outras classes, pois apesar de possuir uma taxa coliformes fecais baixa, outras classes também se comportavam da mesma maneira em algumas amostras.

Os resultados deste trabalho foram comparados com o trabalho de FERREIRA[7], que trata da análise qualidade de água utilizando dados simbólicos. No trabalho dele foi utilizada uma técnica de criação de hipercubos, gerados a partir dos padrões de entrada. Os hipercubos representam as classes de água (mais de um hipercubo pode representar à mesma classe). Essa técnica é supervisionada e funciona da seguinte forma: a cada padrão apresentado, tenta-se encaixar o padrão em algum hipercubo, se nenhum hipercubo suportar esse padrão, é calculada a distância desse padrão para os hipercubos. O hipercubo que estiver mais próximo do padrão deve alterar seu volume para conter esse padrão.

Nos testes de FERREIRA[7] o conjunto de dados foi dividido em 50% para treinamento e 50% para testes, em seguida foram realizadas 100 repetições. Para critério de desempate, no caso de algum padrão provocar a mesma alteração de

volume em dois ou mais hipercubos, o padrão será classificado pelo hipercubo de menor volume. Os resultados de FERREIRA[7] são mostrados na Tabela 11.

CLASSE	AMOSTRAS	ACERTOS	RECALL
0	900	243	27,00%
2	7700	5647	73,34%
3	1200	427	35,58%
4	2100	2057	97,95%
TOTAL	11900	8374	70,36%

Tabela 11: Resultados de FERREIRA.

Como podemos observar, para o conjunto de dados utilizado, a técnica adotada neste trabalho produziu melhores resultados do que a técnica utilizada por FERREIRA[7]. Enquanto que neste trabalho os dados foram estratificados, igualando a quantidade de amostras de todas as classes de padrões, em FERREIRA[7], os dados foram replicados. Essa replicação de dados gerou o mesmo problema encontrado no pré-processamento dos dados deste trabalho, isto é, as classes com mais amostras tenderam a aumentar muito os seus respectivos hipercubos, que acarretou na interseção entre classes, caracterizando um *overfitting*.

No trabalho de NEPOMUCENO[13], foi desenvolvida uma solução para análise de qualidade de água baseada no modelo de mapas auto-organizáveis de *Kohonen*[9]. Os resultados de NEPOMUCENO[13] foram bastante próximos dos resultados obtidos neste trabalho, como podemos ver abaixo, na Tabela 12.

CLASSE	AMOSTRAS	ACERTOS	RECALL
0	153	132	86,27%
2	155	96	61,94%
3	143	119	83,21%
4	167	163	97,06%
TOTAL	618	510	82,53%

Tabela 12: Resultados de NEPOMUCENO.

Diferente da técnica de processamento de dados utilizada neste trabalho, NEPOMUCENO[13] realizou uma replicação nos dados. De acordo com ele: “O *recall*, ou porcentagem de acerto, relativo à classe dois, foi inferior aos outros *recalls*, por possuir uma variedade maior de dados, quase não tendo sido replicado.”. Comparando as Tabelas 10 e 12, podemos observar um comportamento similar das duas redes, exceto a classe 2. Essa discrepância é reflexo das diferentes técnicas de processamento abordadas pelos trabalhos.

Os resultados ainda foram comparados com uma técnica de *benchmark* do TANAGRA[17]. Nesse caso a técnica utilizada foi uma rede neural do tipo MLP, com as mesmas configurações iniciais da rede proposta por esse trabalho. A única diferença entre as abordagens foi à quantidade de vezes que os experimentos foram executados, 50 vezes neste trabalho e apenas uma vez para o *benchmark*. Em ambas as abordagens foi utilizado *Leave-one-out Cross Validation*. A Tabela 13 apresenta os resultados do *benchmark*.

CLASSE	AMOSTRAS	ACERTOS	RECALL
0	19	14	73,68%
2	19	16	84,21%
3	19	15	78,94%
4	19	18	94,73%
TOTAL	76	63	82,89%

Tabela 13: Resultados do *benchmark*.

Comparando-se as Tabelas 10 e 13, podemos observar que nas duas abordagens os resultados obtidos são similares. Por pouco mais de 1% a taxa de acerto da rede proposta neste trabalho sobrepujou a taxa de acerto do *benchmark*. Nas duas abordagens a escassez de dados foi o maior desafio encontrado, prejudicando o resultado da rede como um todo.

6 Conclusões

Os resultados obtidos neste trabalho podem ser considerados satisfatórios, se consideramos a quantidade de dados disponíveis. A partir da comparação deste trabalho com o trabalho de FERREIRA[7], NEPOMUCENO[13] e o de *benchmark*, foi possível notar que a escolha de redes neurais, tanto MLP, quanto *Kohonen*, foi decisiva para apresentação de bons resultados para o problema atacado.

A aquisição de maior volume de dados é essencial para uma melhor classificação da rede. Como por exemplo, no trabalho de CONRADS[5], onde foram utilizados dados coletados durante três anos em intervalos de 30 minutos, ou aproximadamente 52560 amostras. Ao passo que este trabalho utilizou apenas 19 amostras de cada classe. Uma outra solução foi proposta por YONG[19], que desenvolveu um sistema para medição de parâmetros de água e as enviava para um servidor de dados via ondas de rádio. Desta forma a base se torna atualizável em tempo real.

Alguns ajustes podem ser feitos no sistema proposto por esse trabalho. Por exemplo, a utilização de algoritmos genéticos para definição da topologia da rede, como foi proposto por STEPNIEWSKI e KEANE[16]. Com a utilização desta técnica poderia se atingir a melhor configuração de rede para cada problema atacado, o que melhoraria consideravelmente o desempenho da rede.

Referências Bibliográficas

- [1] ANA. Agência Nacional de Águas; Panorama da Qualidade das Águas Superficiais no Brasil. Brasil, 2005. Disponível em: <http://www.ana.gov.br>. Acessada em novembro de 2005.
- [2] BARUAH, TAMURA, M., OKI, K., NISHIMURA, H.; *Neural Network Modeling of Lake Surface Chlorophyll and Sediment Content from Landsat TM Imagery*. Japão, 2000. Disponível em: <http://yasulab.lis.u-tokyo.ac.jp/~pjbaruah/publications/AJGpaper.pdf>. Acessada em dezembro de 2005.
- [3] CETESB. Companhia de Tecnologia de Saneamento Ambiental; Relatório de Qualidade das Águas Interiores do Estado de São Paulo. Brasil, 2004. Disponível em <http://www.cetesb.sp.gov.br>. Acesso em janeiro de 2006.
- [4] CONAMA. Conselho Nacional do Meio Ambiente. Disponível em: <http://www.mma.gov.br/port/conama/index.cfm>. Acesso em janeiro de 2006.
- [5] CONRADS, P. A. e ROEHL Jr., E. A.; *Comparing Physics-Based and Neural Network Models for Simulating Salinity, Temperature, and Dissolved Oxygen in a Complex, Tidally Affected River Basin*. Estados Unidos, 1999. Disponível em: http://smig.usgs.gov/SMIG/features_0302/beaufort.html. Acessada em dezembro de 2005.
- [6] DAWSON, C. W., WILBY, R. L., HARPAM, C., BROWN, M. R.; *Modelling Ranunculus Presence in the Rivers Test and Itchen Using Artificial Neural Networks*; Estados Unidos, 2000. Disponível em: <http://www.geocomputation.org/2000/GC016/Gc016.htm>. Acessado em janeiro de 2006.
- [7] FERREIRA, A. G. A., 2006. Monitoração de Qualidade de Água com o Uso de Dados Simbólicos. Brasil, 2006.
- [8] GALLAGHER, M. R.; *Multi-layer Perceptron Error Surfaces: Visualization, Structure and Modelling*. Austrália, 2000. Disponível em: <http://www.itee.uq.edu.au/~marcusg/thesis.html>. Acessado em janeiro de 2006.
- [9] HAYKIN, S.; *Neural networks: a comprehensive foundation*. 2ª Ed, Prentice: Hall, Inc., 1999.

- [10] HECHT-NIELSEN R. *Neurocomputing*. Addison Wesley Publ. Co., New York, 1990.
- [11] JAVA. *Java Technology*. Disponível em: <http://java.sun.com/>. Acesso em janeiro de 2006.
- [12] MARTIN, R. W. e WALLEY W. J.; *A River Biology Monitoring System (RBMS) for English and Welsh Rivers*. Inglaterra, 1995. Disponível em: <http://www.cies.staffs.ac.uk/rbms.htm>. Acessada em dezembro de 2005.
- [13] NEPOMUCENO, V. S.; *Classificação de água em reservatórios utilizando um algoritmo de aprendizagem competitiva e não supervisionada*. Brasil, 2006.
- [14] SANTOS, Leandro A. P., SANTOS, Sonia M., BOTELHO, Sílvia S. C.; *Reconhecimento da Qualidade da Água de Ecossistemas por Redes Neurais Artificiais; I Workshop de Tecnologia da Informação aplicada ao Meio Ambiente – Brasil, 2003*. Disponível em: www.cbcomp.univali.br/anais/pdf/2003/wma083.pdf. Acessado em dezembro de 2005.
- [15] SCHNEIDER, J.; *Cross Validation*; Disponível em: <http://www.anc.ed.ac.uk/rbf/intro/node43.html> . Acessada em fevereiro de 2006.
- [16] STEPNIEWSKI, S. W; KEANE, A. J. *Topology design of feedforward neural networks by genetic algorithms*. Berlim, 1996. Disponível em: <http://citeseer.ist.psu.edu/11208.html>. Acesso em fevereiro de 2006.
- [17] TANAGRA. *TANAGRA - A free DATA MINING software for teaching and research*. Disponível em: <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>. Acesso em janeiro de 2006.
- [18] TRAUTMANN, T. e DENOEU, T.; *Comparison of dynamic feature map models for environmental Monitoring*. International Conference on Artificial Neural Network, Cambridge/Reino Unido, 1995. Disponível em: <http://www.hds.utc.fr/~tdenoeux/congres/annie99.pdf>. Acessado em janeiro de 2006.
- [19] YONG, C. K. e LIM, C. M.; *An Integrated Water Quality Monitoring System using Artificial Neural Networks*. Singapura, 2001. Disponível em: http://www.np.edu.sg/~yck/nn_waterquality.pdf. Acessado em janeiro de 2006.

[20] ZURADA, J. M.; *Introduction to Artificial Neural Systems*. 1^a Ed, PWS Publishing Company, 1995.

Anexos

Anexo I: Significado Ambiental dos Parâmetros do IQA.

A seguir é apresentado o significado ambiental dos parâmetros do IQA, segundo CETESB (2003).

Oxigênio Dissolvido

Uma adequada provisão de oxigênio dissolvido é essencial para a manutenção de vida aquática em sistemas aquáticos. Através de medição do teor de oxigênio dissolvido, os efeitos de resíduos oxidáveis sobre águas receptoras e a eficiência do tratamento dos esgotos durante a oxidação bioquímica, podem ser avaliados.

A taxa de reintrodução de oxigênio dissolvido em águas naturais através da superfície depende das características hidráulicas e é proporcional à velocidade, sendo que a taxa de renovação do ar superficial em uma cascata é maior do que a de um rio de velocidade normal, que por sua vez apresenta taxa superior à de uma represa, onde a velocidade normalmente é bastante baixa.

Outra fonte importante de oxigênio nas águas é a fotossíntese de algas. Este fenômeno ocorre em águas poluídas ou mais propriamente, em águas eutrofizadas, ou seja, aquelas em que a decomposição dos compostos orgânicos lançados levou à liberação de sais minerais no meio, especialmente os de nitrogênio e fósforo que são utilizados como nutrientes pelas algas. Esta fonte não é muito significativa nos trechos iniciais de rios a jusante de fontes de lançamento de esgotos. A turbidez elevada dificulta a penetração dos raios solares e apenas poucas espécies resistentes às condições severas de poluição conseguem sobreviver. A contribuição fotossintética de oxigênio só é expressiva após grande parte da atividade bacteriana na decomposição de matéria orgânica ter ocorrido, bem como após terem se desenvolvido também os protozoários que, além de decompositores, consomem bactérias clarificando as águas e permitindo a penetração de luz.

Este efeito pode “mascarar” a avaliação do grau de poluição de uma água quando se toma por base apenas a concentração de oxigênio dissolvido. Sob este aspecto, águas poluídas são aquelas que apresentam baixa concentração de oxigênio dissolvido (devido ao seu consumo na decomposição de compostos orgânicos), enquanto que as águas limpas apresentam concentrações de oxigênio dissolvido elevadas, chegando até a um pouco abaixo da concentração de saturação. No entanto, uma água eutrofizada pode apresentar concentrações de oxigênio bem superiores a 10 mg/L, mesmo em temperaturas superiores a 20°C, caracterizando uma situação de supersaturação. Isto ocorre principalmente em lagos de baixa velocidade aonde chegam a se formar crostas verdes de algas na superfície.

Coliformes fecais

As bactérias do grupo coliforme são consideradas os principais indicadores de contaminação fecal. O grupo coliforme é formado por um número de bactérias que inclui os gêneros *Klebsiella*, *Escherichia*, *Serratia*, *Erwenia* e *Enterobactéria*. Todas as bactérias coliformes são gram-negativas manchadas, de hastes não esporuladas que estão associadas com as fezes de animais de sangue quente e com o solo.

As bactérias coliformes termo tolerantes reproduzem-se ativamente a 44,5°C e são capazes de fermentar o açúcar. O uso das bactérias coliformes termo tolerantes para indicar poluição sanitária mostra-se mais significativo que o uso da bactéria coliforme “total”, porque as bactérias fecais estão restritas ao trato intestinal de animais de sangue quente.

A determinação da concentração dos coliformes assume importância como parâmetro indicador da possibilidade da existência de microorganismos patogênicos responsáveis pela transmissão de doenças de veiculação hídrica, tais como febre tifóide, febre paratífóide, desintéria bacilar e cólera.

Potencial Hidrogeniônico (pH)

A influência do pH sobre os ecossistemas aquáticos naturais dá-se diretamente devido a seus efeitos sobre a fisiologia das diversas espécies. Também o efeito indireto é muito importante podendo determinadas condições de pH contribuir para a precipitação de elementos químicos tóxicos como metais pesados; outras condições podem exercer efeitos sobre as solubilidades de nutrientes. Desta forma, as restrições de faixas de pH são estabelecidas para as diversas classes de águas naturais, tanto de acordo com a legislação federal (Resolução nº. 20 do CONAMA, de junho de 1986), como pela legislação do Estado de São Paulo (Decreto no 8468/76). Os critérios de proteção à vida aquática fixam o pH entre 6 e 9.

Demanda Bioquímica de Oxigênio (DBO_{5,20})

A Demanda Bioquímica de Oxigênio de uma água é a quantidade de oxigênio necessária para oxidar a matéria orgânica por decomposição microbiana aeróbia para uma forma inorgânica estável. A DBO_{5,20} é normalmente considerada como a quantidade de oxigênio consumido durante 5 dias, em uma temperatura de incubação de 20°C. Pelo fato de a DBO_{5,20} somente medir a quantidade de oxigênio consumido num teste padronizado, não indica a presença de matéria não biodegradável, nem leva em consideração o efeito tóxico ou inibidor de materiais sobre a atividade microbiana. Os maiores aumentos em termos de DBO_{5,20} num corpo d'água são provocados por despejos de origem predominantemente orgânica. A presença de um alto teor de matéria orgânica pode induzir à completa

extinção do oxigênio na água, provocando o desaparecimento de peixes e outras formas de vida aquática.

Temperatura

Variações de temperatura são parte do regime climático normal e corpos d'água naturais apresentam variações sazonais e diurnas bem como estratificação vertical. A temperatura superficial é influenciada por fatores tais como latitude, altitude, estação do ano, período do dia, taxa de fluxo e profundidade. A elevação da temperatura em um corpo d'água geralmente é provocada por despejos industriais (indústrias canavieiras por exemplo) e usinas termoelétricas.

A temperatura desempenha um papel principal de controle no meio aquático, condicionando as influências de uma série de parâmetros físico-químicos. Em geral, à medida que a temperatura aumenta de 0 a 30°C, a viscosidade, tensão superficial, compressibilidade, calor específico, constante de ionização e calor latente de vaporização diminuem, enquanto a condutividade térmica e a pressão de vapor aumentam as solubilidades com a elevação da temperatura. Organismos aquáticos possuem limites de tolerância térmica superior e inferior, temperaturas ótimas para crescimento, temperatura preferida em gradientes térmicos e limitações de temperatura para migração, desova e incubação do ovo.

Nitrogênio Total

São diversas as fontes de nitrogênio nas águas naturais. Os esgotos sanitários constituem em geral a principal fonte, lançando nas águas nitrogênio orgânico devido à presença de proteínas e nitrogênio amoniacal devido à hidrólise sofrida pela uréia na água. Alguns efluentes industriais também concorrem para as descargas de nitrogênio orgânico e amoniacal nas águas, como algumas indústrias químicas, petroquímicas, siderúrgicas, farmacêuticas, de conservas alimentícias, matadouros, frigoríficos e curtumes.

A atmosfera é outra fonte importante devido a diversos mecanismos: fixação biológica desempenhada por bactérias e algas que incorporam o nitrogênio atmosférico em seus tecidos contribuindo para a presença de nitrogênio orgânico nas águas a fixação química, reação que depende da presença de luz e concorre para as presenças de amônia e nitratos nas águas, as lavagens da atmosfera poluída pelas águas pluviais concorrem para as presenças de partículas contendo nitrogênio orgânico bem como para a dissolução de amônia e nitratos.

Nas áreas agrícolas, o escoamento das águas pluviais pelos solos fertilizados também contribui para a presença de diversas formas de nitrogênio. Também nas áreas urbanas, as drenagens de águas pluviais associadas às deficiências do sistema de limpeza pública, constituem fonte difusa de difícil

caracterização. Como visto, o nitrogênio pode ser encontrado nas águas nas formas de nitrogênio orgânico, amoniacal, nitrito e nitrato. As duas primeiras chamam-se formas reduzidas e as duas últimas, formas oxidadas.

Os compostos de nitrogênio são nutrientes para processos biológicos. São tidos como macro nutrientes, pois, depois do carbono, o nitrogênio é o elemento exigido em maior quantidade pelas células vivas. Quando descarregados nas águas naturais conjuntamente com o fósforo e outros nutrientes presentes nos despejos, provocam o enriquecimento do meio tornando-o mais fértil e possibilitam o crescimento em maior extensão dos seres vivos que os utilizam, especialmente as algas, o que é chamado de eutrofização.

Quando as descargas de nutrientes são muito fortes, dá-se o florescimento muito intenso de gêneros que predominam em cada situação em particular. Estas grandes concentrações de algas podem trazer prejuízos aos usos que se possam fazer dessas águas, afetando seriamente o abastecimento público ou causando poluição por morte e decomposição. O controle da eutrofização através da redução do aporte de nitrogênio é comprometido pela multiplicidade de fontes, algumas muito difíceis de serem controladas como a fixação do nitrogênio atmosférico por parte de alguns gêneros de algas. Por isso, deve-se investir preferencialmente no controle das fontes de fósforo.

Os nitratos são tóxicos causando uma doença chamada metahemoglobinemia infantil, que é letal para crianças (o nitrato se reduz a nitrito na corrente sanguínea, competindo com o oxigênio livre, tornando o sangue azul).

Fósforo Total

O fósforo aparece em águas naturais devido, principalmente, às descargas de esgotos sanitários. Nestes, os detergentes superfosfatados empregados em larga escala domesticamente constituem a principal fonte, além da própria matéria fecal que é rica em proteínas. Alguns efluentes industriais, como os de indústrias de fertilizantes, pesticidas, químicas em geral, conservas alimentícias, abatedouros, frigoríficos e laticínios, apresentam fósforo em quantidades excessivas. As águas drenadas em áreas agrícolas e urbanas também podem provocar a presença excessiva de fósforo em águas naturais.

O fósforo pode se apresentar nas águas sob três formas diferentes. Os fosfatos orgânicos são as formas em que o fósforo compõe moléculas orgânicas, como a de um detergente, por exemplo. Os ortofosfatos, por outro lado, são representados pelos radicais que se combinam com cátions formando sais inorgânicos nas águas. Os polifosfatos ou fosfatos condensados são polímeros de ortofosfatos. No entanto, esta terceira forma não é muito importante nos estudos de controle de qualidade das águas, porque os polifosfatos sofrem hidrólise convertendo-se rapidamente em ortofosfatos nas águas naturais.

Assim como o nitrogênio, o fósforo constitui-se em um dos principais nutrientes para os processos biológicos, ou seja, é um dos chamados macronutrientes por ser exigido também em grandes quantidades pelas células. Por ser nutriente para processos biológicos, o excesso de fósforo em esgotos sanitários e efluentes industriais conduz a processos de eutrofização das águas naturais.

Turbidez

A turbidez de uma amostra de água é o grau de atenuação e intensidade que um feixe de luz sofre ao atravessá-la (e esta redução se dá por absorção e espalhamento, uma vez que as partículas que provocam turbidez nas águas são maiores que o comprimento de onda da luz branca), devido à presença de sólidos em suspensão, tais como partículas inorgânicas (areia, silte, argila) e de detritos orgânicos, algas e bactérias, plâncton em geral, etc.

A erosão das margens dos rios em estações chuvosas é um exemplo de fenômeno que resulta em aumento da turbidez das águas e que exige manobras operacionais como alterações nas dosagens de coagulantes e auxiliares nas estações de tratamento de águas. A erosão pode decorrer do mau uso do solo em que se impede a fixação da vegetação. Este exemplo mostra também o caráter sistêmico da poluição ocorrendo inter-relações ou transferência de problemas de um ambiente (água, ar ou solo) para outro.

Os esgotos sanitários e diversos efluentes industriais também provocam elevações na turbidez das águas. Um exemplo típico deste fato ocorre em consequência das atividades de mineração, onde os aumentos excessivos de turbidez têm provocado formação de grandes bancos de lodo em rios e alterações no ecossistema aquático.

Alta turbidez reduz a fotossíntese de vegetação enraizada submersa e algas. Esse desenvolvimento reduzido de plantas pode, por sua vez, suprimir a produtividade de peixes. Logo, a turbidez pode influenciar nas comunidades biológicas aquáticas. Além disso, afeta adversamente os usos doméstico, industrial e recreacional de uma água.

Resíduo Total

Em saneamento, sólidos nas águas correspondem a toda matéria que permanece como resíduo, após evaporação, secagem ou calcinação da amostra a uma temperatura pré-estabelecida durante um tempo fixado.

Para o recurso hídrico, os sólidos podem causar danos aos peixes e à vida aquática. Eles podem se sedimentar no leito dos rios destruindo organismos que fornecem alimentos ou também danificar os leitos de desova de peixes. Os sólidos podem reter bactérias e resíduos orgânicos no fundo dos rios promovendo

decomposição anaeróbia. Altos teores de sais minerais, particularmente sulfato e cloreto, estão associados à tendência de corrosão em sistemas de distribuição além de conferir sabor às águas.

Anexo II: Gráficos de Índice de Qualidade dos Parâmetros do IQA.

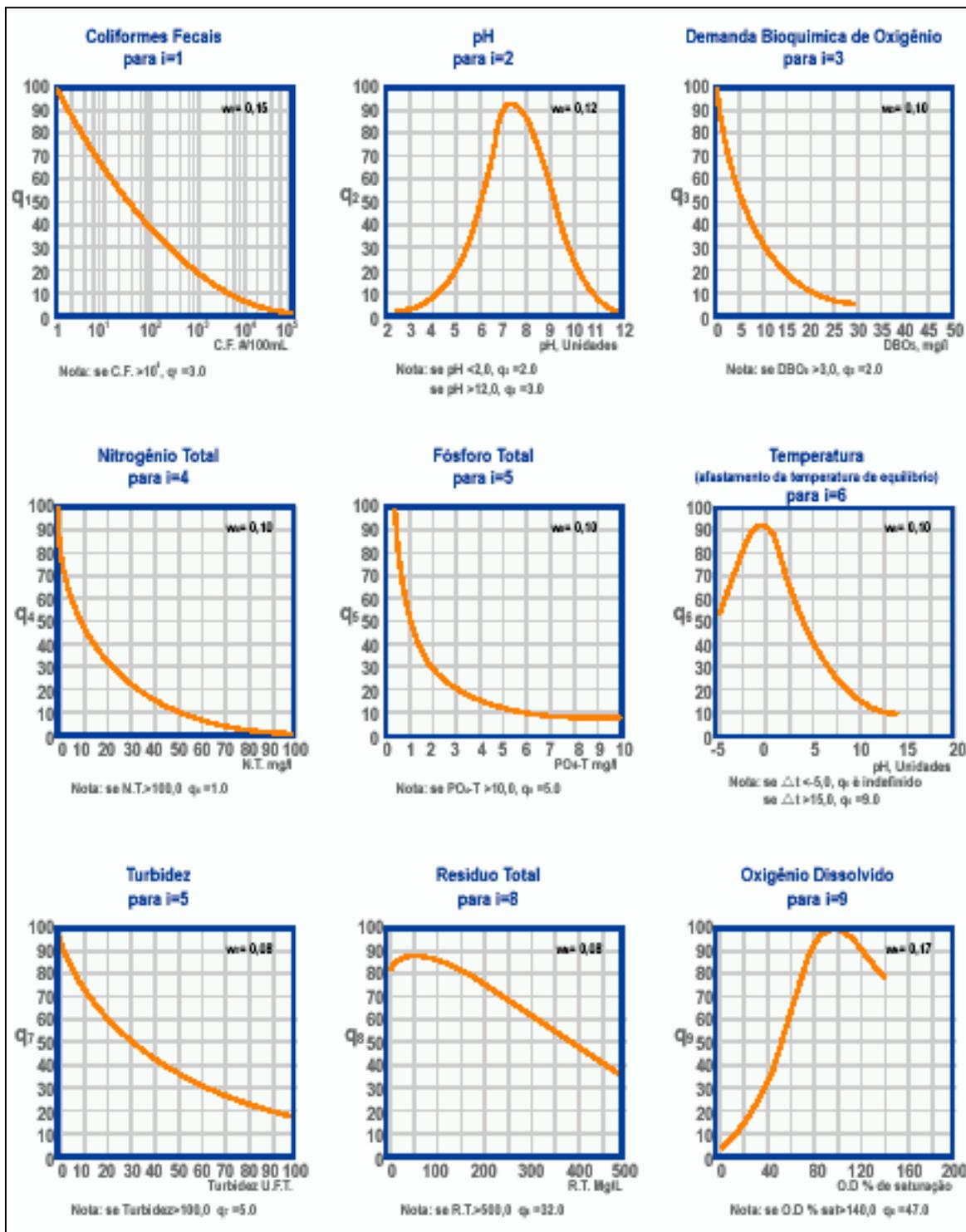


Figura: Gráficos de índice de qualidade dos parâmetros.