

UNIVERSIDADE FEDERAL DE PERNAMBUCO



CENTRO DE INFORMÁTICA
TRABALHO DE GRADUAÇÃO



Classificação de qualidade da água em
reservatórios utilizando um algoritmo de
aprendizagem competitiva e não supervisionada

Autor: Vilmar Santos Nepomuceno

Orientadora: Teresa Bernarda Ludermir

Co-Orientador: Francisco de Assis Tenório de Carvalho

RECIFE
2006

UNIVERSIDADE FEDERAL DE PERNAMBUCO

VILMAR SANTOS NEPOMUCENO

Classificação de qualidade da água em reservatórios
utilizando um algoritmo de aprendizagem competitiva e
não supervisionada

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: Teresa Bernarda Ludermir

Co-Orientador: Francisco de Assis Tenório de Carvalho

Monografia apresentada em 22 de fevereiro de 2006.

Avaliada pela banca examinadora composta por:

Teresa Bernarda Ludermir (orientadora)

Renata Maria Cardoso Rodrigues de Souza (avaliadora)

Classificação de qualidade da água em reservatórios utilizando um algoritmo de aprendizagem competitiva e não supervisionada.

Vilmar Santos Nepomuceno

Teresa Bernarda Ludermir

(Orientadora) Centro de Informática

Francisco de Assis Tenório de Carvalho

(Co-orientador) Centro de Informática

EXAMINADORA

Renata Maria Cardoso Rodrigues de Souza

(Avaliadora) Centro de Informática

Dedico este trabalho aos meus pais, irmão e namorada. Em especial, a minha mãe, por estar presente em todos momentos da minha vida, sendo eles bons ou ruins.

AGRADECIMENTOS

Agradeço principalmente a minha mãe Maria de Lourdes, por tornar possível todo o meu estudo e desenvolvimento, agradeço ao meu falecido pai Walmir que mesmo com sua ausência foi responsável pela criação do meu caráter, ao meu irmão Victor por sempre compreender que fazer meus trabalhos de faculdade era mais importante do que ele jogar no computador e também a toda a família por entender minhas ausências nas reuniões familiares.

Um agradecimento especial a minha namorada Mirele por ter suportado a ausência do namorado em vários momentos, quantas festas perdidas pelas conhecidas noites no centro de informática, quantas noites mal dormidas sempre me esperando chegar em casa e nem sempre aceitando, mas compreendendo o esforço para terminar minha graduação.

Agradecer a professora Teresa e o professor Francisco pela oportunidade de realizar este trabalho.

Não poderia esquecer de agradecer todos os meus companheiros de noites mal dormidas, ou não dormidas, “agora durma com uma bronca dessa”, no centro de informática, Aécio (cabelo), Afonso (alonso), André (batoré), Túlio (gente que faz), Vicente (anaconda) e tantos outros com os quais dividi trabalhos, Fausto, Alexandre, Andrei, Paulo, etc. Também lembrar de Ives (slot) que apesar de não ser da minha turma, pegou muita carona e dividiu muito do seu conhecimento em computação gráfica e inteligência artificial comigo.

Agradeço aos professores Aluizio por despertar em mim o gosto por inteligência artificial e Valdir, do Centro de Ciências Biológica, por ter me ajudado, orientado e se tornado um grande amigo durante a minha iniciação científica.

Agradecer ao centro de informática por estar sempre 24h aberto, fazendo com que os projetos pudessem ser entregues no prazo, também ao DA de computação por comprar a mesa de ping-pong proporcionando vários momentos de lazer e descontração.

LISTA DE FIGURAS

Figura 1AB	A) Resultado MLP. B) Resultado SOM.....	14
Figura 2	Interface do RBMS.....	14
Figura 3	Representação do mapa de Kohonen.....	15
Figura 4	Diagrama de relacionamento entre as classes.....	26
Figura 5	Mapa inicial 9X9.....	26
Figura 6	Resultado do treinamento da melhor configuração.....	28
Figura 7	Resultado do treinamento da segunda melhor configuração.....	28
Figura 8	Resultado do treinamento da terceira melhor configuração.....	29
Figura 9	Sobreposição das classes.....	30

LISTA DE TABELAS

Tabela 1	Parâmetros do Índice de Qualidade da Água - IQA e respectivos pesos.....	12
Tabela 2	Valores do IQA.....	18
Tabela 3	Divisão inicial dos dados.....	19
Tabela 4	Valores ausentes.....	21
Tabela 5	Valores mal classificados.....	22
Tabela 6	Valores não-atômicos.....	22
Tabela 7	Estratificação dos dados.....	23
Tabela 8	Matriz de confusão, dados estratificados.....	23
Tabela 9	Replicação dos dados.....	24
Tabela 10	Quatro melhores configurações de treinamento da rede...	27
Tabela 11	Três melhores configurações de treinamento da rede usadas para demonstrar resultados.....	27
Tabela 12	Matriz de confusão encontrada após os testes.....	30
Tabela 13	Benchmark.....	31
Tabela 14	Resultados FERREIRA [5].....	31

LISTA DE ABREVIações

Agência Nacional de Águas	ANA
Back-Propagation Neural Network	BPNN
Centre for Intelligent Environmental Studies	CIES
Companhia de Tecnologia de Saneamento Ambiental	CETESB
Companhia Hidroelétrica do São Francisco	CHESF
Conselho Nacional do Meio Ambiente	CONAMA
Demanda Bioquímica de Oxigênio	DBO_{5,20}
Índice de Qualidade da Água	IQA
Learning Vector Quantization	LVQ1
Multi-layer Perceptron	MLP
National Sanitation Foundation	NSF
Nitrogênio Kjeldahl Total	NKT
Oxigênio Dissolvido	OD
Potencial Hidrogeniônico	pH
Principal Component Analysis	PCA
River Biology Monitoring System	RBMS
Self-organizing Maps	SOM

RESUMO

A informação sobre a qualidade da água dos rios brasileiros é pouca e esparsa. O uso de técnicas de aprendizagem de máquina visa um melhor aproveitamento dessa informação para realizar um monitoramento mais eficaz das águas. Para este trabalho foi usada uma técnica de rede neural.

Apresenta-se então nesse trabalho uma solução baseada no modelo de mapas auto-organizáveis de Kohonen e a comparação deste modelo com outras técnicas de aprendizagem propostas por outros autores.

ÍNDICE

DEDICATÓRIA.....	04
AGRADECIMENTOS.....	05
LISTA DE FIGURAS.....	06
LISTA DE TABELAS.....	07
LISTA DE ABREVIACÕES.....	08
RESUMO.....	09
1. INTRODUÇÃO.....	11
2. ESTADO DA ARTE.....	13
2.1 TRABALHOS RELACIONADOS.....	13
2.2 SELF-ORGANIZING MAPS (SOM)	15
3. PROBLEMA PROPOSTO E ANÁLISE DOS DADOS.....	16
3.1 PROBLEMA PROPOSTO.....	16
3.2 ANÁLISE DOS DADOS.....	18
4. METODOLOGIA	21
4.1 PRÉ-PROCESSAMENTO DOS DADOS.....	21
4.1.1 VALORES AUSENTES (MISSING VALUES)	21
4.1.2 VALORES MAL CLASSIFICADOS.....	22
4.1.3 VALORES NÃO-ATÔMICOS.....	22
4.1.4 DESBALANCEAMENTO DAS CLASSES.....	23
4.2 ASPECTOS OBSERVADOS.....	24
5. RESULTADOS.....	25
5.1 PROTÓTIPO.....	25
5.2 EXPERIMENTOS.....	26
5.3 ANÁLISE DOS RESULTADOS.....	30
6. CONCLUSÃO E TRABALHOS FUTUROS.....	33
REFERÊNCIAS BIBLIOGRÁFICAS.....	34
APÊNDICE A	36

1. INTRODUÇÃO

A qualidade da água está entre os mais relevantes e emblemáticos problemas do país. A informação esparsa ou inexistente em várias bacias, sem redes de monitoramento adequadas em termos de frequência, parâmetros e número de pontos de amostragem para o território que cobre a área dos reservatórios, dificultam uma visão totalmente fidedigna da condição dos corpos d'água destes reservatórios.

Com a inclusão do Sistema Nacional de Gerenciamento de Recursos Hídricos na Constituição de 1988, a aprovação da Lei 9.433, em 1997, estabelecendo a Política e o Sistema Nacional de Gerenciamento de Recursos Hídricos e a criação da Agência Nacional de Águas (ANA) em 2000, a água é, definitivamente, incorporada à agenda política brasileira [9].

A importância da qualidade da água está bem conceituada na Política Nacional de Recursos Hídricos [1] que define, dentre seus objetivos, “assegurar à atual e às futuras gerações a necessária disponibilidade de água, em padrões de qualidade adequados aos respectivos usos” [9].

A Política Nacional de Recursos Hídricos também determina, como uma das diretrizes de ação do Sistema Nacional de Gerenciamento de Recursos Hídricos, “A gestão sistemática dos recursos hídricos, sem dissociação dos aspectos de quantidade e qualidade e a integração da gestão dos recursos hídricos com a gestão ambiental” [9].

Para a gestão de qualidade da água foi usado, inicialmente, o índice de qualidade da água o IQA [1], elaborado em 1970, pelo National Sanitation Foundation (NSF), dos Estados Unidos, a partir de uma pesquisa de opinião realizada junto a especialistas em qualidade de água. Nessa pesquisa cada especialista indicou os parâmetros a serem avaliados, o peso relativo dos mesmos e a condição com que se apresenta cada parâmetro (Tabela 1). Uma explicação sobre cada parâmetro pode ser encontrada no Apêndice A deste trabalho.

Tabela 1: Parâmetros do Índice de Qualidade da Água - IQA e respectivos pesos.

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

No entanto o IQA tem o intuito, apenas, de comunicação com o público leigo, mas em termos computacionais existe uma grande perda de informação individual e de interação entre as variáveis [1].

O uso de métodos computacionais, tais como redes neurais [4], para reconhecimento de padrões da qualidade da água e análise das variáveis de forma automática, utilizando os indicadores físicos, químicos e biológicos, é de fundamental importância para este monitoramento, pois classificam de forma rápida e eficiente as amostras de água coletadas e analisadas.

Uma técnica de rede neural não supervisionada, mais especificamente o modelo baseado em aprendizado competitivo proposto por KOHONEN [8], será utilizado para resolver o problema de classificação da qualidade da água proposto neste trabalho. No modelo de mapas auto-organizáveis (SOM) de KOHONEN [8], “os neurônios são colocados em nós de um mapa normalmente com uma ou duas dimensões, eles sofrem ajustes seletivos através de vários padrões de entrada durante o processo de aprendizagem competitiva. Os neurônios ajustados tendem a ficar mais próximos de maneira que um sistema de coordenadas significativo seja montado para as características de entrada”. De acordo com HAYKIN [6] “Um mapa auto-organizável é caracterizado pela formação de um mapa topográfico dos padrões de entrada, em que a localização espacial dos neurônios corresponde intrinsecamente às características dos padrões de entrada”.

2. ESTADO DA ARTE

2.1 TRABALHOS RELACIONADOS

Utilizando métodos de aprendizagem de máquina é possível obter uma classificação sobre a qualidade da água utilizando indicadores físicos, químicos e biológicos. No trabalho de BARUAH [2] que descreveu estudos realizados no Lago Kasumigaura (Japão), utilizando alguns parâmetros característicos da água, tais como concentração de Clorofila e de Sedimentos Suspensos, já foi possível monitorar, satisfatoriamente, a qualidade da água. Nesse caso utilizando a rede neural do tipo Multi layer Feed-Forward empregando o algoritmo de aprendizado back-propagation, conhecida como Back-Propagation Neural Network (BPNN) [6].

No trabalho de TRAUTMANN [14] ele afirma que “para entradas de poucas dimensões é possível fazer uma avaliação subjetiva dos dados através de mapas gráficos, mas com entradas de muitas dimensões pode haver contra-indicações. Essa abordagem não constitui uma abordagem sistemática e objetiva para se montar uma arquitetura ótima ou fazer a comparação entre vários mapas a partir de diferentes condições de início”. Ele também menciona que “pode existir problemas para novos padrões de entrada, já que adaptações em aplicações de monitoramento são necessárias, para quando aparecer um conjunto de treinamento que seja muito dissimilar aos protótipos já existentes”. No caso dele ele investiu não apenas no modelo SOM clássico, mas também em um modelo incremental que aceitava o aumento do número de nós do mapa para contemplar os novos conjuntos de treinamentos.

Outro esforço realizado foi o de YONG [15] onde em seu trabalho ele avaliou o uso de dois modelos, o perceptron multicamadas (MLP) e o SOM, chegando a conclusão que “ambos os modelos possuíam uma classificação eficiente para o problema proposto”, mas nesse trabalho, assim como nos descritos acima, foram usados poucos parâmetros, nesse trabalho foram usados três parâmetros de entrada: o potencial hidrogeniônico (pH), oxigênio dissolvido e temperatura. Os resultados obtidos por Yong podem ser vistos na Figura 1AB, a Figura 1A mostra os resultados obtidos com o uso da MLP, enquanto que na Figura 1B é possível visualizar os resultados da classificação obtidos com o algoritmo de Kohonen.

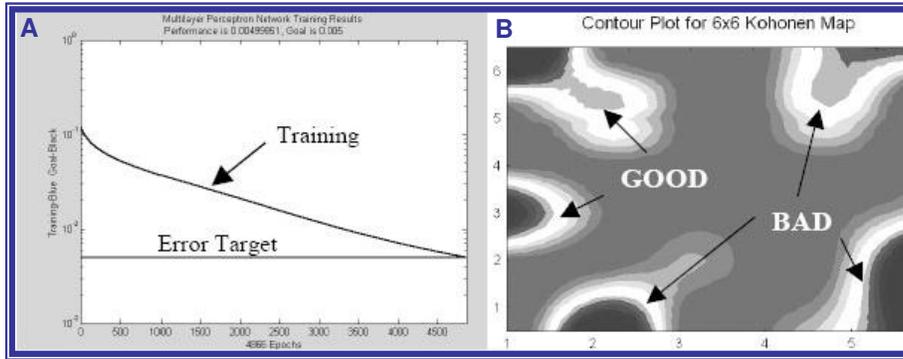


Figura 1AB: A) Resultado MLP. B) Resultado SOM.

O projeto desenvolvido no Centre for Intelligent Environmental Studies (CIES) por MARTIN [10] foi realizado com o intuito de analisar a composição biológica das águas. A ferramenta desenvolvida nesse trabalho, o River Biology Monitoring System (RBMS), possui uma interface amigável e apresenta três tipos de algoritmos, MLP, SOM e Bayes ingênuo, mas, diferentemente dos trabalhos apresentados acima, ela trabalha, apenas, com fatores biológicos. É possível visualizar na Figura 2 um exemplo da interface do RBMS ao se aplicar o SOM.

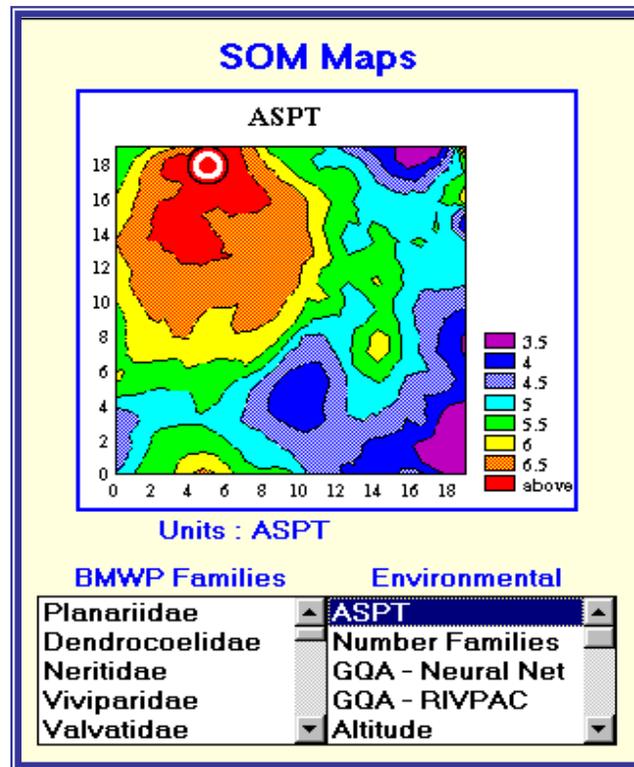


Figura 2: Interface do RBMS.

2.2 SELF-ORGANIZING MAPS (SOM)

O modelo de rede neural de Kohonen é formado somente por duas camadas: a de entrada e a de saída. Cada um dos neurônios de entrada excita cada um dos neurônios de saída e estes últimos estão dispostos em um mapa, denominado de SOM - self organizing map.

No modelo de rede neural proposto originalmente por Kohonen [8], cada padrão de entrada irá, através dos pesos entre a camada de entrada e a de saída, excitar de forma distinta cada um dos neurônios de saída, de acordo com a similaridade da entrada com os pesos das conexões aos respectivos neurônios de saída; assim o neurônio com maior excitação será ativado e deverá inibir a ativação dos demais, esta abordagem é denominada de winner-take-all.

O mapa se auto-organiza de forma que vetores de entrada muito similares ativam o mesmo neurônio, definindo conjuntos de vetores de entrada associados a cada neurônio, de mesma forma, conjuntos de vetores que possuam similaridades entre si devem ativar neurônios próximos no mapa de saída. Assim, comparando-se a localização de cada saída no mapa tem-se uma noção de quanto similares os vetores de entrada são entre si e cada região do mapa resalta determinadas similaridades e funções relacionadas à aplicação em questão. Uma representação do mapa de Kohonen é mostrada na Figura 3.

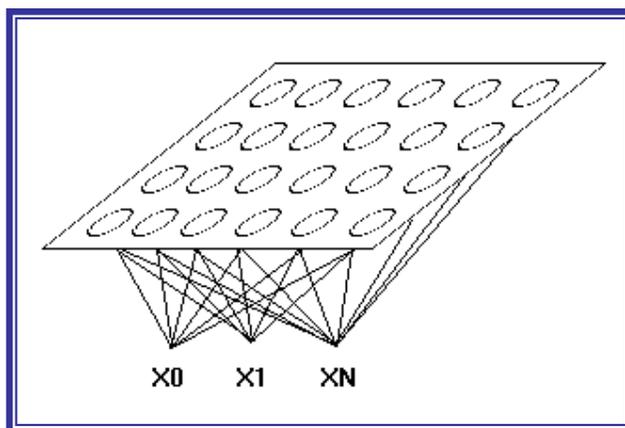


Figura 3: Representação do mapa de Kohonen.

3. PROBLEMA PROPOSTO E ANÁLISE DOS DADOS

O modelo SOM foi escolhido para servir de base para este trabalho, principalmente, por ser uma técnica que permite uma representação visual dos dados permitindo um melhor entendimento de especialistas de outras áreas e o uso dos nove parâmetros foi feito para se ter como base a classificação do IQA para definição das classes.

3.1 PROBLEMA PROPOSTO

Neste trabalho foi desenvolvido um classificador para a qualidade da água utilizando o modelo SOM clássico. Criou-se uma região topográfica onde a partir dos padrões de entrada foram geradas regiões de agrupamento, e utilizando o padrão de cálculo do IQA proposto pela ANA [1], os grupos foram classificados. O algoritmo do classificador realiza os seguintes passos:

I. Inicialização dos Pesos:

Os pesos, $w(t)$, dos nós de entrada conectados aos nós do mapa são escolhidos randomicamente. Foram gerados valores entre zero e um.

II. Apresentar Padrão:

A partir do conjunto de padrões de entrada é escolhido, randomicamente, um padrão $x(t)$ para ser apresentado a rede.

III. Calcular distância entre padrão e nós do mapa e escolher nó vencedor:

Calcula-se a distância, $d(t)$, do padrão de entrada para todos os nós, é então escolhido o nó vencedor. O vencedor é o nó que apresentar a menor distância quadrática para o padrão apresentado.

$$d(t) = E (x (t) - w(t))^2$$

IV. Atualização dos Pesos:

O nó vencedor tem seu peso atualizado, pela seguinte equação:

- $w(t+1) = w(t) + \eta(t)[x(t) - w(t)]$

Onde $\eta(t)$ é a taxa de aprendizagem naquele momento. Assim como o nó vencedor, os nós vizinhos também são atualizados seguindo a mesma equação, os nós que não estão na vizinhança mantêm o peso inalterado.

- $w(t+1) = w(t)$

A vizinhança diminui com o passar do tempo, assim como a taxa de aprendizagem.

V. Retornar ao passo II:

O algoritmo se repete até que todos os padrões de entrada sejam apresentados.

VI. Aplicar algoritmo LVQ1 [6]:

O algoritmo LVQ1 foi utilizado para delimitar melhor a borda das regiões geradas. O LVQ1 é uma técnica supervisionada. Foi usada, então, a informação das classes geradas pelo uso do IQA. O LVQ1 utiliza as seguintes equações para atualização dos pesos:

$w(t+1) = w(t) + \eta(t) (x(t) - w(t))$, caso x tenha sido classificado corretamente;

$w(t+1) = w(t) - \eta(t) (x(t) - w(t))$, caso x tenha sido classificado incorretamente;

$w(t+1) = w(t)$, para neurônios diferentes do vencedor.

Onde x é o padrão de entrada apresentado.

VII. Escolher nós vencedores de cada classe:

Selecionou-se então o nó mais representativo para cada classe, e criou-se uma região de influência para cada nó. Depois de selecionados, os nós da área de influência têm suas classes comparadas com a classe do nó vencedor.

Para apresentar os padrões à rede é utilizada a técnica de ten-fold cross-validation onde o conjunto de padrões de entrada é dividido em dez partes, onde nove servem para treino e um para teste. Todos os padrões são usados para treinamento e para teste.

3.2 ANALISE DOS DADOS

De acordo com a ANA existem cinco níveis de IQA, divididos em uma faixa de 0 – 100 (Tabela 2).

Tabela 2: Valores do IQA.

Valores do IQA	Qualidade da água
Varição 80 – 100 (Classe 0)	Ótima
Varição 52 – 79 (Classe 1)	Boa
Varição 37 – 51 (Classe 2)	Aceitável
Varição 20 – 36 (Classe 3)	Ruim
Varição 0 – 19 (Classe 4)	Péssima

Esses valores são calculados usando a fórmula abaixo, onde q_i é a qualidade do i -ésimo parâmetro e w_i é o peso correspondente ao i -ésimo parâmetro:

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

, onde

Os dados foram retirados do site da CETESB [3], já que não possuíamos uma base de dados própria. A idéia inicial era utilizar dados dos reservatórios da CHESF, mas tais dados não foram disponibilizados até agora.

Sendo assim foi necessária a busca de outra fonte de dados. Na CETESB [3] foi onde obtivemos dados das variáveis de água de vários rios de São Paulo. Essa foi a fonte encontrada com mais dados disponíveis e que possuía medição de todas as nove variáveis que são usadas para o cálculo do IQA.

Ainda assim, os dados disponíveis nos relatórios da CETESB [3] são escassos, não foram encontrados rios com a qualidade “Ruim” e poucos exemplos com a qualidade “Péssima” e “Boa”.

Os dados foram coletados em rios paulistanos, de onze de fevereiro de 2004 a oito de dezembro de 2004, nesse intervalo foram realizadas, em média, seis coletas por rio. Em CETESB [3] pode ser encontrada a rede de monitoramento usada para captação dos dados.

A base de dados da CETESB [3] apresentou a seguinte divisão inicial dos dados (Tabela 3).

Tabela 3: Divisão inicial dos dados.

Classe	Amostras
Classe 0	19
Classe 2	148
Classe 3	24
Classe 4	42

No trabalho de CONRADS [4] foi visto que “foram usados dados de três anos de coleta que eram medidas a cada 30 minutos”. Assim como no CONRADS [4] outros trabalhos usaram muitos dados para os treinamentos dos algoritmos desenvolvidos. No caso de YONG [15], eles produziram um sistema que possuía um sensor, o qual fazia a medição de parâmetros da água e enviava as informações obtidas, para um servidor de dados através de um

modem de telefone celular, o que tornava a base de dados deles atualizável em tempo real.

De um modo geral a divisão inicial dos dados mostrou um grande desbalanceamento entre as classes. Em aprendizagem de máquina, classes desbalanceadas podem provocar distorções nos resultados. Para o modelo SOM, essa divisão da massa de dados acarretaria problemas, como especialização das classes zero e três. O Classificador não conseguiria generalizar essas amostras nem especializar as amostras da classe dois, suficientemente, o que deixaria o classificador pobre. Para serem usadas no treinamento, essas amostras foram submetidas a vários tipos de pré-processamentos.

4. METODOLOGIA

Após a obtenção dos dados, foi necessário tratá-los, para que pudessem ser usados como padrões de entrada da rede neural desenvolvida neste trabalho. Para observação e análise dos resultados foi utilizado matriz de confusão que indica quantas amostras foram classificadas corretamente em relação ao espaço de amostras total.

4.1 PRÉ-PROCESSAMENTO DOS DADOS

Durante a análise dos dados foram encontrados problemas de ausência de valores para alguns atributos. Encontraram-se também problemas com valores mal classificados e valores que não eram atômicos, ou seja, representavam uma faixa de valores. Os dados possuíam também um grande desbalanceamento entre as classes o que também teve que ser resolvido.

4.1.1 VALORES AUSENTES (MISSING VALUES)

O problema dos dados ausentes foi resolvido calculando-se o valor médio para cada atributo de cada classe e substituindo os valores ausentes por esse valor médio. Essa técnica foi utilizada pela pequena quantidade de amostras afetadas. Na Tabela 4 são apresentadas essas amostras.

Tabela 4: Valores ausentes.

Característica	Amostras afetadas
Fósforo	12
Nitrogênio	12
Resíduo	27
Coliformes	1
DBO	4
pH	1

4.1.2 VALORES MAL CLASSIFICADOS (RUÍDO)

O problema dos dados mal classificados foi resolvido substituindo esses valores pelo valor máximo para o atributo, caso o valor seja maior, e pelo mínimo caso contrário. Esses dados poderiam causar over-fitting, acarretando um aumento na complexidade do modelo, além de que esses dados, se em maior quantidade, pode tornar a extração de informação complicada. Na tabela 5 são apresentadas as amostras afetadas.

Tabela 5: Valores mal classificados.

Característica	Amostras afetadas
pH	2
Temperatura	3
DBO	2
Fósforo total	4
NKT	3
OD	2
Turbidez	10
Coliforme fecal	4

4.1.3 VALORES NÃO-ATÔMICOS

O problema dos valores não-atômicos foi resolvido substituindo-se os intervalos que eles estavam representando pelo maior valor que quantificava esse intervalo, ou seja, discretizando os valores, por exemplo, o valor (<2) foi substituído por dois. A transformação desses dados foi realizada, pois a forma original deles impossibilitava o uso no protótipo criado. Na tabela 6 são apresentadas as amostras afetadas.

Tabela 6: Valores não-atômicos.

Característica	Amostras afetadas
DBO	57
Fósforo total	7
NKT	4
OD	2
Coliforme fecal	2

4.1.4 DESBALANCEAMENTO DAS CLASSES

Os dados retirados da CETESB [3] estavam inicialmente classificados pela resolução CONAMA, que é uma alternativa da resolução da ANA sendo que com o uso de mais parâmetros como alumínio, manganês, ferro total, etc. O número de parâmetros do CONAMA tornou inviável o uso de SOM. Foi necessário então converter esta classificação para a ANA, e para isso foi usado o cálculo do IQA. Após a reclassificação dos dados apareceram os problemas de desbalanceamento das classes. Para resolver esse problema foi utilizada inicialmente a técnica de estratificação dos dados, os dados foram estratificados randomicamente. A Tabela 7 mostra como ficou a divisão dos dados.

Tabela 7: Estratificação dos dados.

Classe	Amostras
Classe 0	19
Classe 2	29
Classe 3	24
Classe 4	21

Lembrando-se que para treinamento e teste, com os dados estratificados, foi utilizado ten-fold cross-validation. Foi observado que os resultados obtidos não foram satisfatórios. A tabela 8 mostra a matriz de confusão dessa abordagem. O erro de teste ficou em, aproximadamente, 38%, o que significa que mais de um terço das amostras de teste foram classificadas erradas.

Tabela 8: Matriz de confusão, dados estratificados.

Classe	0	2	3	4	Total	Recall	1 - Precision
0	12	7	3	0	22	54,55%	57,14%
2	9	11	5	3	28	39,29%	56,00%
3	7	4	13	0	24	54,17%	38,10%
4	0	3	0	16	19	84,21%	15,79%
Soma	28	25	21	19	93	-	-

Após verificar o mal resultado dessa abordagem foi utilizada a técnica de replicação, onde as amostras das classes que possuem poucos dados sofreram ajustes para ficar com uma quantidade equiparável de amostras. A Tabela 9 mostra como ficou a divisão dos dados.

Tabela 9: Replicação dos dados.

Classe	Amostras
Classe 0	152
Classe 2	154
Classe 3	144
Classe 4	168

Após o processo de limpeza dos dados, estes, para servirem de entrada para rede, sofreram uma normalização, onde todos os valores ficaram entre zero e um.

4.2 ASPECTOS OBSERVADOS

Durante os experimentos foram geradas matrizes de confusão, cada conjunto de parâmetros adotado foi analisado. Para montar a matriz de confusão era gerado como saída à classificação dos dados de teste.

A partir desta classificação fez-se a contagem de quantos falsos verdadeiros, que são, por exemplo, dados que pertenciam a classe 2 e foram classificados como classe 3, são encontrados para cada classe e de quantas amostras foram classificadas corretamente.

Foi calculado então o “recall” de cada classe, que é a porcentagem de amostras classificadas corretamente de cada classe, dividido pelo total de amostras classificadas como desta classe. Para uma melhor análise dos falsos verdadeiros foi calculado o “1-precision”, que é a divisão dos falsos verdadeiros de cada classe pelo total de amostras classificadas como desta classe.

Foi calculado também o erro geral do teste que é a divisão de todas as amostras de teste classificadas erradas, pelo total de amostras de teste apresentada.

5. RESULTADOS

Utilizando-se de um protótipo próprio, foram realizados vários experimentos com os padrões que possuíamos. Obtivemos resultados interessantes quanto ao agrupamento dos dados. O protótipo, experimentos e resultados serão os assuntos deste tópico.

5.1 PROTÓTIPO

Foi produzido um protótipo de um classificador, esse protótipo foi desenvolvido em linguagem Java [7], utilizando o J2SE 1.4.2 SDK. O diagrama de relacionamento entre as classes do classificador está na Figura 4, a classe principal do sistema é a Kohonen.class, onde foi feito o treinamento da rede e guardado toda a configuração do experimento, a classe KohonenPanel.class é responsável pela parte gráfica do sistema.

A classe Signals.class tem como objetivo gerar os sinais de entrada para serem apresentados a rede. As classes Node.class, Connection.class e Tuple.class representam a rede neural, com seus neurônios, sinapses e vetor de pesos, respectivamente. A classe LVQ1.class é responsável por aplicar o algoritmo LVQ1 [6].

Os resultados dos experimentos foram persistidos em arquivos textuais, assim como a configuração de cada experimento realizado. Foram gerados dois arquivos, o primeiro com os pesos obtidos como resultado no treinamento e o segundo com a saída comparada entre os padrões de entrada, do conjunto de teste. Os pesos de cada nó da rede, dando a classificação de cada padrão do conjunto de teste, também foi persistido.

O mapa inicial gerado é visto na Figura 5. Para uma melhor representação visual, o mapa 9X9 foi escolhido.

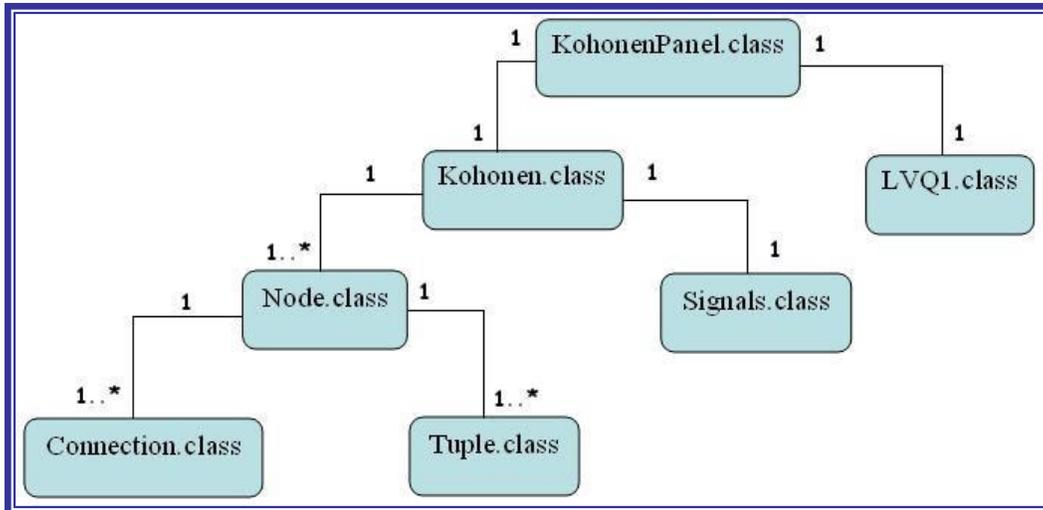


Figura 4: Diagrama de relacionamento entre as classes.

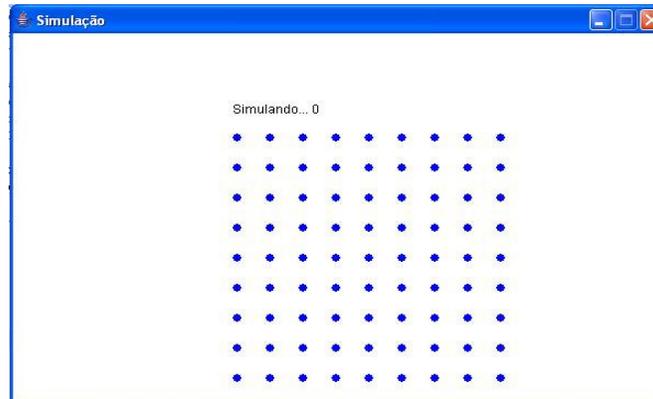


Figura 5: Mapa inicial 9X9.

5.2 EXPERIMENTOS

Foram realizados vários experimentos utilizando várias configurações para rede. Na Tabela 10 estão descritos as melhores configurações encontradas, levando-se em consideração, limites superior e inferior na inicialização dos pesos, taxa de aprendizagem ($\eta(n)$), número de iterações e tamanho do mapa.

Tabela 10: Quatro melhores configurações de treinamento da rede.

Limites dos pesos	$\eta(n)$	Iterações	Mapa
0 – 1	0.05	20	9X9
0 – 1	0.05	15	9X9
0 – 1	0.1	15	9X9
0 – 1	0.04	25	9X9

Foram realizados, também, experimentos com outros valores para cada parâmetro, por exemplo, foram utilizados valores entre 0 -100 para limite dos pesos, utilizando-se taxas de aprendizagem maiores, entre 10 e 20, com cinco iterações e mapas entre os tamanhos 7X7 e 9X9, mas não foram obtidos bons resultados e a abordagem seguida foi a de parâmetros apresentados na Tabela 10.

As Figuras 6, 7 e 8 mostram resultados gráficos de experimentos para as seguintes configurações de parâmetros (Tabela 11):

Tabela 11: Três melhores configurações de treinamento da rede usadas para demonstrar resultados.

Figura	Limites	$\eta(n)$	Iterações	Mapa
Figura 6	0 – 1	0.05	20	9X9
Figura 7	0 – 1	0.05	15	9X9
Figura 8	0 – 1	0.1	15	9X9

Na Figura 6, é possível visualizar a formação de quatro grupos de dados, de acordo com a classificação dada, no desenvolvimento do protótipo, a cor preta, é classificada como classe zero, a cor azul para classe dois, a cor vermelha para classe três e a cor amarela para classe quatro. A figura 6 apresenta uma disposição de pontos, onde se pode verificar um maior agrupamento dos pontos da classe três. Isso pode ser explicado, pelo maior número de padrões dessa classe, pois pela replicação dos dados, os padrões da classe três serão apresentados um maior número de vezes do que os das classes dois e quatro, algo semelhante, também, pode ser visto na classe zero. O fato fica mais evidente com o aumento da taxa de aprendizagem, como pode ser visto na Figura 8.

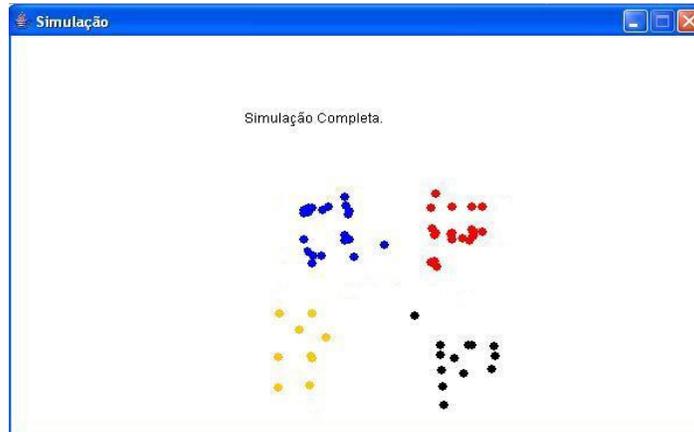


Figura 6: Resultado do treinamento da melhor configuração.

Na Figura 7 há uma dispersão maior dos pontos, isso se deve pelo menor número de iterações. Fica evidente também um menor agrupamento dos pontos que pertencem às classes com menos padrões, pois poucos pontos são ativados inicialmente para estas classes, por terem muitos padrões replicados, o que é revertido posteriormente com o ajuste dos pesos dos nós vizinhos para se aproximarem destas classes.

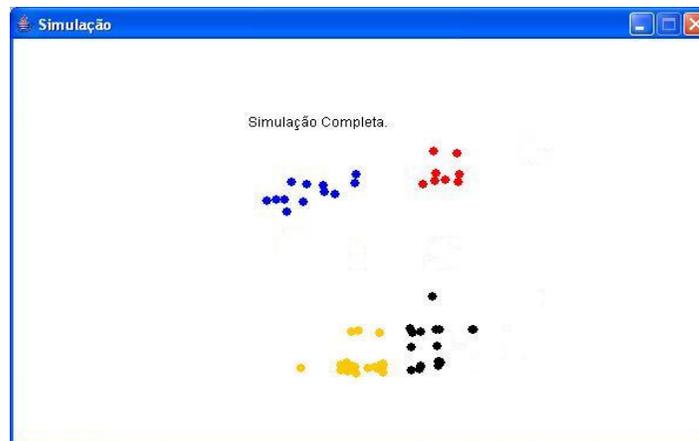


Figura 7: Resultado do treinamento da segunda melhor configuração.

Na Figura 8, os pontos são poucos, mas bem agrupados. Isso se deve a taxa de aprendizagem ser maior, ou seja, os nós se aproximam mais rapidamente, e o número menor de iterações faz com que outros nós mais próximos não tenham tido a oportunidade de entrar na área de influência do nó vencedor.

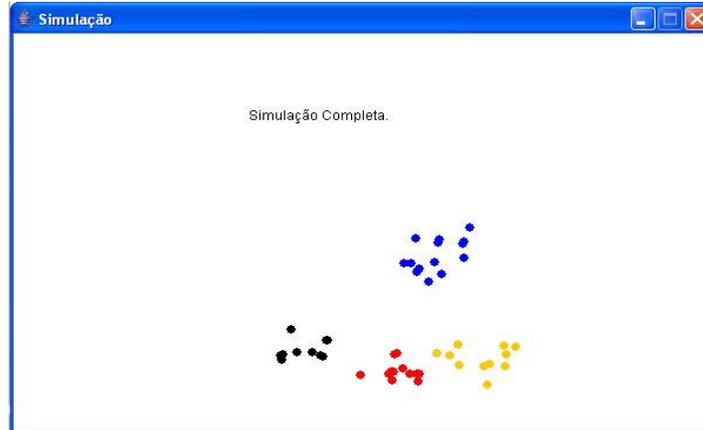


Figura 8: Resultado do treinamento da terceira melhor configuração.

Mesmo observando que esses casos, tem suas peculiaridades, foi visto que pelos parâmetros de configuração testados, esses foram o que melhor classificaram as classes.

Com algumas escolhas de parâmetros aconteceu a congruência das classes, ou seja, uma classe sobrepôs a outra, como pode ser visto na Figura 9, na implementação a uma ordem de pintura dos nós, primeiro os da classe zero (preto), depois da classe dois (azul), seguida da classe três (vermelho) e por último a classe quatro (amarelo). Os parâmetros de treinamento foram: limites entre 0 – 1, $\eta(n) = 0.01$ e 5 iterações. É possível ver que existe apenas um nó da cor azul, que é o nó vencedor da classe que foi sobreposta pelos nós vermelhos, e três da cor preta, que foram sobrepostas pelos nós da cor amarela. Isso pode ser explicado pela iniciação randômica dos pesos, que pode gerar pesos de classes diferentes em nós vizinhos, como também pela baixa taxa de aprendizagem e pequeno número de iterações, o que pode não ter permitido uma divergência dos nós de classes diferentes.

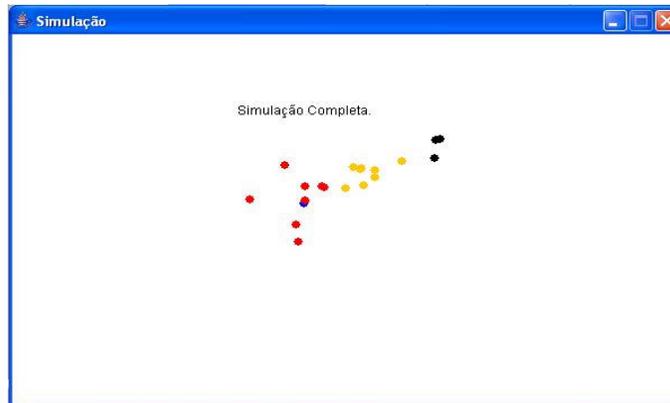


Figura 9: Sobreposição das classes.

5.3 ANÁLISE DOS RESULTADOS

Após várias execuções da melhor configuração de parâmetros observada foi encontrado um resultado satisfatório, levando-se em consideração o volume de dados que possuíamos. Foi realizada também a comparação entre o método proposto neste trabalho e o de outros autores, além de um benchmark, construído na ferramenta TANAGRA [13].

Na Tabela 12 encontra-se a matriz de confusão encontrada aplicando-se o classificador desenvolvido nesse trabalho, utilizando-se a melhor configuração encontrada. Para construir essa matriz foram realizados 50 testes, com a mesma configuração, gerando essa matriz que bem representa os resultados obtidos.

Tabela 12: Matriz de confusão encontrada após os testes.

Classe	0	2	3	4	Total	Recall	1 - Precision
0	132	9	12	0	153	86,27%	18,01%
2	23	96	32	4	155	61,94%	24,41%
3	6	18	119	0	143	83,21%	26,99%
4	0	4	0	163	167	97,06%	2,40%
Soma	161	127	163	167	618	-	-

O recall, ou porcentagem de acerto, relativo a classe dois, foi inferior aos outros recalls, por possuir uma variedade maior de dados, quase não tendo sido replicado o erro total dos testes ficou em 17,47%, bem abaixo do que foi encontrado quando usada à técnica de estratificação dos dados. Em comparação com a técnica de benchmark os resultados podem ser

considerados satisfatórios. O benchmark utilizado foi o algoritmo K-NN com cinco vizinhos, com divisão do conjunto de dados de 50% para treinamento e 50% para teste, os dados estavam com sua configuração inicial (Tabela 3), foram realizadas 100 repetições do algoritmo e a matriz que o melhor representa está na Tabela 13, o erro total de testes encontrado foi de 28,84%, bem acima do encontrado pela técnica utilizada neste trabalho.

Tabela 13: Benchmark.

Classe	0	2	3	4	Soma	Recall	1 - Precision
0	123	807	34	0	964	12,76%	67,72%
2	244	6901	312	267	7724	89,34%	26,49%
3	13	1023	116	54	1206	9,62%	77,30%
4	1	657	49	1399	2106	66,43%	18,66%
Soma	381	9388	511	1720	12000		

FERREIRA [5], em seu trabalho de conclusão, usou a abordagem de criação de hipercubos para classificação de água, onde os padrões de entrada geravam volumes que representavam as classes do projeto, a técnica é supervisionada, tentando classificar cada padrão de entrada, se o padrão não se encaixa em nenhum hipercubo é calculado a distância do padrão para cada um dos hipercubos o que estivesse menos distante tinha seu volume alterado para conter esse novo padrão. FERREIRA [5] realizou testes com divisão dos grupos do conjunto de dados de 50% para treinamento e 50% para teste, foram realizadas 100 repetições e como desempate, no caso de padrões provocarem a mesma alteração de volume, o padrão será classificado como a classe que tiver o menor volume. A Tabela 14 mostra a matriz que melhor representa os experimentos o erro total de testes encontrado foi de 29,63%, bem acima do encontrado pela técnica utilizada neste trabalho.

Tabela 14: Resultados FERREIRA [5].

Classe	0	2	3	4	Soma	Recall	1 - Precision
0	243	605	0	52	900	27,00%	76,86%
2	696	5647	974	383	7700	73,34%	17,73%
3	111	569	427	93	1200	35,58%	69,52%
4	0	43	0	2057	2100	97,95%	20,43%
Soma	1050	6864	1401	2585	11900	-	-

Os resultados mostrados na Tabela 14 se aproximaram melhor dos resultados encontrados na Tabela 12, apesar do erro total da rede ter sido maior, sendo até melhor no caso da classe dois, isso pode acontecer, pelo fato de que uma maior amostragem de dados geram um volume mais bem delimitado, classificando de uma forma mais precisa as amostras do que quando existem poucas amostras, que devem gerar volumes não muito bem definidos, isso explica também o grande número de falsos verdadeiros encontrados e, portanto, um significativo erro total de teste.

Enquanto que no método abordado neste trabalho, o fato de uma classe ter poucas amostragens pode fazer com que a rede decore aqueles resultados tendo, portanto, um melhor aproveitamento de classificação para estes dados, e no caso de ter-se uma variedade de amostras maior, a rede pode não obter uma convergência tendo um pior aproveitamento de classificação. A técnica adotada por FERREIRA [5] pode ser uma boa abordagem no caso de ter-se uma quantidade maior de dados diversificados.

Em MELO FILHO [11] foi utilizado um algoritmo de aprendizagem supervisionada, a MLP para a classificação de qualidade de água, o resultado encontrado por ele é caracterizado pela taxa de erro do sistema que ficou em 15,48%, considerada, por ele, satisfatória, a taxa de erro encontrada nesse trabalho foi de 17,47%. Apesar dos resultados estarem próximos, a visualização que o modelo SOM proporciona deve ser levado em consideração, podendo ser considerado uma melhor abordagem.

6. CONCLUSÃO E TRABALHOS FUTUROS

Os resultados obtidos nesse trabalho podem ser considerados satisfatórios, diante do fato de que dispomos de poucos dados. O ideal seria que o conjunto de dados tivesse um maior volume. Para contornar esse problema, existe a perspectiva da utilização dos dados que venham a ser disponibilizados pela CHESF.

A comparação dos resultados obtidos com o benchmark, originado de um software muito utilizado no mundo acadêmico, foi um ponto importante do trabalho, onde pôde ser visto que o caminho adotado para seguir, ou seja, o uso de técnicas de redes neurais especialmente o uso de SOM, foi bem escolhido.

O aprimoramento da técnica usada, ou a hibridização desta técnica pode ser feito, uma abordagem que pode ser seguida é a proposta por TRAUTMANN [14] onde ele investiu não apenas no modelo SOM clássico, mas também em um modelo incremental que aceitava o aumento do número de nós do mapa para contemplar novos padrões, que antes não apareciam.

O uso de técnicas estatísticas também pode ser feito, para melhorar o uso dos dados, técnicas, como o pca [12], podem ser abordadas para se realizar o estudo dos parâmetros de qualidade de água, já que uma das grandes dificuldades desse trabalho foi mapear os nove nós de entrada, que representavam os parâmetros, nos nós do mapa, e com a utilização dessa técnica podem ser excluídos alguns parâmetros que não acrescentam informações na rede.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ANA. Panorama da qualidade das águas superficiais no Brasil. / Agência Nacional de Águas --- Brasília: ANA, 2005. 175 p
- [2] BARUAH, P. J.; OKI, K. e NISHIMURA, H. A Neural Network Model for estimating Surface Chlorophyll and Sediment Content at the Lake Kasumi Gaura of Japan. Proceedings of Asian Conference on Remote Sensing, Taipei, Taiwan, 2000, p. 419-424. Disponível em: <<http://yasulab.lis.u-tokyo.ac.jp/~pjbaruah/publications/AJGpaper.pdf>>. Acesso em: Janeiro, 2006.
- [3] CETESB. Padrões de Qualidade para os Parâmetros Monitorados na Rede de Monitoramento, segundo a Resolução CONAMA 20/86. Companhia de Tecnologia de Saneamento Ambiental, São Paulo. Disponível em: <<http://www.cetesb.sp.gov.br/>>. Acesso em: Janeiro 2006.
- [4] CONRADS, P. A. ROEHL JR, E. A., MARTELLO, W. P., Estimating Point-Source Impacts on the Beaufort River Using Artificial Neural Network Models Disponível: http://smig.usgs.gov/SMIG/features_0302/beaufort.html. Acesso em: Janeiro 2006.
- [5] FERREIRA, A. G. A., 2006. Monitoração de Qualidade de Água com o Uso de Dados Simbólicos.
- [6] HAYKIN, S. Neural networks: a comprehensive foundation. 2ª ed, Prentice: Hall, Inc. , 1999.
- [7] JAVA. Java Technology. Disponível em: <<http://java.sun.com/>>. Acesso em: Janeiro 2006.
- [8] KOHONEN, T., 1990. "The self-organizing map" Proceedings of the IEEE78,1464-1480.

[9] LEI 9.433, em 1997, estabelecendo a Política e o Sistema Nacional de Gerenciamento de Recursos Hídricos e a criação da Agência Nacional de Águas. Disponível em: <http://www.ana.gov.br/Institucional/Legislacao/leis/lei9433.pdf>. Acesso em: Janeiro, 2006.

[10] MARTIN, R. W. e WALLEY W. J. A River Biology Monitoring System (RBMS) for English and Welsh Rivers. School of Computing, Staffordshire University, 1995. Disponível em: <http://www.cies.staffs.ac.uk/rbms.htm> Acesso em: Janeiro, 2006.

[11] MELO FILHO, A. C. , 2006. Sistema de Análise de Qualidade de Água Utilizando Rede Neural Artificial Multicamadas.

[12] MORRISON D. F.. Multivariate Statistical Methods, Second Edition [International Student Edition]. McGraw Hill International, Auckland, 1976.

[13] 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 2006.

[14] TRAUTMANN, T. e DENOEU, T. Comparison of dynamic feature map models for environmental Monitoring. Proc. of the International Conference on Artificial Neural Network, 1995, p.73-78.

[15] YONG, C. K. e LIM, C. M. An Integrated Water Quality Monitoring System using Artificial Neural Networks. Ngee Ann Polytechnic - Electronic and Computer Engineering Department, Singapore, 2001. Disponível em: http://www.np.edu.sg/~yck/nn_waterquality.pdf. Acesso em: Janeiro, 2006.

APÊNDICE A

A seguir é apresentado o significado ambiental dos parâmetros do IQA, segundo CETESB [3].

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 refração 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.

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 gran-negativas manchadas, de hastes não esporuladas que estão associadas com as fezes de animais de sangue quente e com o solo. As bactérias coliformes termotolerantes reproduzem-se ativamente a 44,5°C e são capazes de fermentar o açúcar.

O uso das bactérias coliformes termotolerantes 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 paratifóide, 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 canaveiras, 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, a lavagem 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 macronutrientes, 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 a forma 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 macro-nutrientes 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 de 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.

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 refração 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.

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.