



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE TECNOLOGIA E RECURSOS NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL E AMBIENTAL
CAMPUS I - CAMPINA GRANDE**

**EXTRAÇÃO DE CONHECIMENTO EM
BASE DE DADOS NA GESTÃO DA QUALIDADE DA ÁGUA DO SISTEMA
DE ABASTECIMENTO DE CAMPINA GRANDE - PB**

CLARISSA CÂMARA DE FREITAS

**CAMPINA GRANDE - PB
2015**

CLARISSA CÂMARA DE FREITAS

**EXTRAÇÃO DE CONHECIMENTO EM
BASE DE DADOS NA GESTÃO DA QUALIDADE DA ÁGUA DO SISTEMA
DE ABASTECIMENTO DE CAMPINA GRANDE - PB**

Dissertação apresentada no Curso de Mestrado em Engenharia Civil e Ambiental da Universidade Federal de Campina Grande - UFCG, em cumprimento às exigências para obtenção do grau de Mestre em Engenharia Civil e Ambiental.

Área: Engenharia de Recursos Hídricos e Sanitária

Orientador: Prof. Dr. Rui de Oliveira

CAMPINA GRANDE - PB

2015

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

F866e Freitas, Clarissa Câmara de.
Extração de conhecimento em base de dados na gestão da qualidade da água do sistema de abastecimento de Campina Grande - PB / Clarissa Câmara de Freitas. – Campina Grande, 2015.
66f. : il. color.

Dissertação (Mestrado em Engenharia Civil e Ambiental) – Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais, 2015.

"Orientação: Prof. Dr. Rui de Oliveira".
Referências.

1. Qualidade da água tratada. 2. Sistemas inteligentes. 3. Extração de conhecimento. 4. Mapas de Kohonen. 5. Lógica fuzzy. I. Oliveira, Rui de. II. Título.

CDU 628.16(043)

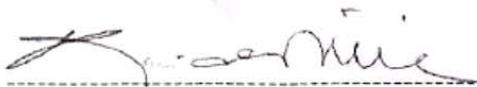
**EXTRAÇÃO DE CONHECIMENTO EM
BASE DE DADOS NA GESTÃO DA QUALIDADE DA ÁGUA DO SISTEMA
DE ABASTECIMENTO DE CAMPINA GRANDE - PB**

Data da aprovação: 2 de Fevereiro de 2015,

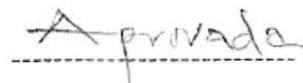
COMISSÃO EXAMINADORA

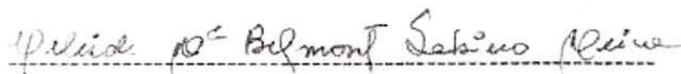
COMISSÃO EXAMINADORA

PARECER

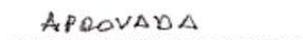


Prof. Rui de Oliveira – Dr.



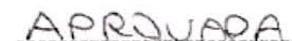


Profa. Celeide Maria Belmont Sabino Meira – Dra.





Prof. Fernando Fernandes Vieira – Dr.





Prof. Peter Batista Cheung – Dr.



Campina Grande, 2 de fevereiro de 2015

Ao meu DEUS e à minha família.

AGRADECIMENTOS

Agradeço a Deus pela Sua bondade e amor para comigo, que me fez voar como águia, me supriu em tudo e me fortaleceu até o dia de hoje.

À minha família: Silene Brandão, Miguel Ângelo, Salete Leite, Luiz Alberto Leite Maria José Araújo (Hona) e Nelice Medeiros, que de forma única e indestrutível me geraram em suas almas e me motivaram a ser firme em todos os passos.

À Júlia Câmara, Joaquim Câmara e Marieliza Braga, que foram compreensíveis e participaram ativamente dessa conquista.

A Emanuel Campos, Igor Ogata, Cayo Farias, Juscelino Henriques e George Ferreira pelo apoio, companheirismo, carinho e confiança que perdurarão por muitas décadas.

A Josete, Isabel Cristina, Savana e todos os que compõem o Laboratório de Saneamento da Universidade Federal de Campina Grande.

Ao professor Rui de Oliveira pelo apoio e confiança em mim investidas.

Aos colegas e amigos que fazem parte da turma de mestrado em Engenharia Civil e Ambiental.

A todos que fazem parte do Grupo de Pesquisa Saúde Ambiental, por me ensinarem que além de técnicas e experimentos, um verdadeiro pesquisador é feito da construção a passos curtos do conhecimento compartilhado.

A todos os professores do curso de Pós-Graduação em Engenharia Civil e Ambiental da UFCG, em especial aos da área de Engenharia de Recursos Hídricos e Sanitária, pela excelência no ensino e na transmissão de toda experiência.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro durante o curso.

A todos que direta ou indiretamente cooperaram para o enriquecimento deste trabalho, registro aqui minha gratidão.

RESUMO

Os sistemas inteligentes têm sido aplicados a uma grande variedade de estudos que não se limitam apenas à área computacional, na qual originalmente adquiriu o termo Inteligência Artificial, mas possuem utilidades diretas nas diversas áreas do conhecimento, entre elas a Engenharia Ambiental. O presente estudo contempla a modelagem matemática por Redes Neurais Artificiais (RNA's) e Lógica Fuzzy da qualidade da água do sistema de abastecimento da cidade de Campina Grande, estado da Paraíba. O modelo apresenta importantes contribuições para os estudos de pesquisa operacional, de modo que para a modelagem foram feitos monitoramentos de indicadores da qualidade da água com base na Diretriz Nacional do Plano de Amostragem da Vigilância Ambiental em Saúde. O processo de aprendizagem adotado nas RNA's foi a Aprendizagem Competitiva ou *Clustering*, fornecendo resultados satisfatórios para todas as variáveis selecionadas. A aplicação de lógica fuzzy objetivou a extração de conhecimentos através de regras difusas, com a meta de interpretação dos dados processados pela rede neural. Foram simulados cenários supondo situações enfrentadas por um operador do sistema de abastecimento. Pontos específicos da cidade apresentaram problemas quanto à concentração do cloro residual livre, por não possuírem manutenção há um longo tempo. Pontos cuja pressão é menor apresentaram concentrações elevadas de bactérias heterotróficas. Ajustes operacionais foram feitos diretamente no modelo, levando em consideração a variação das concentrações de cloro residual livre e combinado, pH e turbidez, verificando o comportamento da concentração de bactérias heterotróficas em todos os pontos do sistema de abastecimento. O modelo constitui uma ferramenta estratégica de gestão da qualidade da água.

Palavras-Chave: Qualidade da água tratada. Sistemas inteligentes. Extração de conhecimento. Mapas de Kohonen. Lógica fuzzy.

ABSTRACT

Intelligent systems have been applied to a wide variety of studies that not limited to the computational area where originally was named with the term Artificial Intelligence, but have direct utility in various areas of knowledge, including Environmental Engineering. This study considers the mathematical modeling by Artificial Neural Networks (ANN) and Fuzzy Logic of water quality of Campina Grande supply system, Paraiba state, northeast Brazil. The model has important contributions to operational research studies, so for the modeling was made monitoring of water quality indicators based on the brazilian National Guideline for the Sampling Plan in Environmental Health Surveillance. The learning process adopted in the ANN was Competitive or Clustering Learning, providing satisfactory results for all selected variables. The application of fuzzy logic aimed at the extraction of knowledge by fuzzy rules, with the goal of interpreting the data processed by the neural network. Scenarios simulating situations faced by a water supply system operator were constructed and modeled. Specifics points of the city presented problems as low concentrations of free residual chlorine, due to long term lack of maintenance. Points whose water pressure is lower, showed higher concentrations of heterotrophic bacteria. Operational adjustments were made directly on the model, taking into account variations of free and combined residual chlorine concentrations, pH and turbidity, checking the behavior of the concentration of heterotrophic bacteria in all points of the supply system. The model is a strategic tool for managing water quality.

Keywords: Quality of treated water. Intelligent systems. Extraction of knowledge. Kohonen maps. Fuzzy logic.

SUMÁRIO

LISTA DE FIGURAS	i
LISTA DE TABELAS.....	ii
LISTA DE ABREVIATURAS.....	iv
INTRODUÇÃO	1
1.1 Objetivos	7
1.1.1. Objetivo geral.....	7
1.1.2 Objetivos específicos	7
FUNDAMENTAÇÃO TEÓRICA.....	8
2.1 Sistemas de abastecimento de água.....	9
2.1.2 Realidade brasileira quanto à operação, manutenção e qualidade da água de abastecimento público.	12
2.3 Redes auto-organizáveis	13
2.3.1 Mapas auto-organizáveis	14
2.3.1 Formas de visualização.....	15
2.3.2 Algoritmo de treinamento.....	16
2.3.3 Interpretação do mapa SOM	18
2.4 Lógica fuzzy	20
2.4.1 Teoria dos conjuntos difusos.....	20
2.4.2 Operações em conjuntos difusos	22
2.4.3 Regras fuzzy	24
2.4.4 Funções de pertinência.....	24
2.4.5 Inferência fuzzy.....	26
2.4.6 Sistemas de inferência Mamdani.....	26
2.5 Extração de conhecimento em base de dados	28
2.5.1 Estado da arte de extração de conhecimento de mapas auto-organizáveis a partir de regras difusas	28

MATERIAIS E MÉTODOS.....	32
3.1 Descrição do sistema	33
3.2 Metodologias de aquisição de variáveis	40
3.2.1 Cloro residual livre	41
3.4.2 Cloro residual combinado	41
3.2.3 pH	42
3.4.4 Turbidez	42
3.4.5 Bactérias heterotróficas facultativas (BHF).....	42
3.3 Ferramentas utilizadas	43
3.4 Definição do problema.....	44
3.5 Construção da rede neuro-fuzzy	45
RESULTADOS	47
DISCUSSÃO	56
5.2 Mapas de Kohonen	58
5.3 Simulação de cenários e tomada de decisão.....	59
CONCLUSÕES.....	61
REFERÊNCIAS.....	63

LISTA DE FIGURAS

Figura 2.1 - Domicílios abastecidos de água por rede geral, segundo as grandes regiões do Brasil - 2000/2008	11
Figura 2.2 - Tipos de arranjos da rede SOM.	15
Figura 2.3 - Matriz-U e sua escala de cores.	16
Figura 2.4 - Treinamento de uma rede SOM. Fase de ajuste dos pesos (1), processo de cooperação dos neurônios vizinhos (2) e processo de adaptação (3) em que a rede não sofre mais mudanças em suas iterações.	18
Figura 2.5 - Conjunto difuso convexo e não-convexo.....	22
Figura 2.6 - Funções de inferência e suas representações gráficas	25
Figura 3.1 - Localização geográfica do município de Campina Grande.	33
Figura 3.2 - Zonas de pressão de Campina Grande.	34
Figura 3.3 - Representação esquemática do sistema de abastecimento de água de Campina Grande, entre a captação e o reservatório R9.....	38
Figura 3.4 - Rede de distribuição de água de Campina Grande: localização dos reservatórios de distribuição de água potável e área de influência (AI) das unidades de reservação.	39
Figura 3.5 - Representação esquemática do modelo neuro-fuzzy.	46
Figura 4.1 – Dispersão e histogramas das variáveis analisadas.	49
Figura 4.2 - Esforço computacional do processo de aprendizagem da rede neural.....	50
Figura 4.3 - Mapas auto-organizáveis dos indicadores de qualidade da água.	51
Figura 4.4 - Mapas BMU para de pontos de monitoramento (a) e indicadores da qualidade da água (b).	51
Figura 4.5 - Superfícies de resposta para os pontos de monitoramento P1, P2 e P8 relacionadas ao comportamento das variáveis CRL e BHF.....	54
Figura 4.6 - Superfícies de resposta para os pontos de monitoramento P1, P6 e P8 relacionadas ao comportamento das variáveis CRL e BHF.....	54
Figura 4.7 - Superfícies de resposta para os pontos de monitoramento P3, P6 e P8 relacionadas ao comportamento das variáveis CRL e BHF.....	55

LISTA DE TABELAS

Tabela 2.1 - Percentual de municípios do País, total de domicílios, economias residenciais e economias abastecidas em relação ao número de domicílios.	11
Tabela 3.2 - Pontos de coleta e monitoramento no sistema de abastecimento.....	41
Tabela 4.3 - Resultados para Extração de Regras.....	52
Tabela 4.4 - Resultados para extração de regras em relação aos pontos de monitoramento...	52
Tabela 4.5 - Valores dos indicadores por pontos de monitoramento simulados pela rede neuro-fuzzy.	53

LISTA DE QUADROS

Quadro 2.1 - Quadro representativo das operações propriedades dos conjuntos difusos.....	23
Quadro 3.1 - Bairros inseridos em suas respectivas zonas de pressão.	34
Quadro 3.2 - Quadro resumo do sistema de reservação de Campina Grande.....	36

LISTA DE ABREVIATURAS

RNA	<i>Rede Neural Artificial</i>
SOM	<i>Self-Organizing Map (Mapa Auto-Organizável)</i>
BMU	<i>Best Matching Unit (unidade mais semelhante)</i>
FIS	<i>Fuzzy Inference System (Sistema de Inferência Difusa)</i>
CRL	<i>Cloro Residual Livre</i>
CRC	<i>Cloro Residual Combinado</i>
BHF	<i>Bactérias Heterotróficas Facultativas</i>
ETA	<i>Estação de Tratamento de Água</i>

*I*NTRODUÇÃO

Redes Neurais Artificiais constituem um tipo de modelo matemático da grande área da Inteligência Artificial, cujo objetivo principal é reproduzir matematicamente a atividade sináptica cerebral no processo de aprendizagem das informações, adquirindo comportamento inteligente. Ao cumprir tal objetivo, o modelo é capaz de prever cenários futuros, reconhecer padrões, auxiliar na tomada de decisões, reproduzir ambientes complexos e compor planos de monitoramento em diversas áreas, como economia, agricultura e eletricidade, bem como a área de saneamento, foco do presente trabalho. Aliada a uma rede neural, a lógica fuzzy possui a capacidade de explicar as informações contidas na rede e auxiliar em processo de controle e gestão. Em estudos de saneamento, um modelo neuro-fuzzy tem a função de auxiliar na tomada de decisão e, através de sua capacidade de aprendizagem, construir cenários abrangendo todas as características pertinentes do objeto de estudo.

Nesse contexto, Saneamento básico é definido como o conjunto das seguintes ações: abastecimento de água, esgotamento sanitário, limpeza pública, drenagem pluvial e controle de vetores. O conceito de abastecimento de água insere-se no conceito mais amplo de saneamento definido pela Organização Mundial da Saúde como sendo o controle de todos os fatores do meio físico do homem, que exercem ou podem exercer efeitos deletérios sobre seu bem-estar físico, mental ou social. Logo o saneamento compreende um conjunto de ações sobre o meio ambiente no qual vivem as populações, visando garantir a elas condições de salubridade, que protejam a sua saúde (HELLER *et al.*, 2006).

A principal base que promove a execução de tais projetos é a garantia da promoção da saúde pública como conceito estabelecido pela OMS - Organização Mundial de Saúde, quando define a saúde como um processo de bem estar pleno do ser humano. Ainda que seja um conceito considerado inatingível, a engenharia sanitária compreende que favorecer um estado salutar à população, se inicia com serviços de prevenção de estados ou doenças que comprometam seu bem estar. Por isso, além da realização de projetos que se utilizem desta visão para seu plano de execução, é importante também que haja planejamento e monitoramento adequados dos mesmos.

No Brasil, há uma crescente preocupação com a qualidade da água distribuída à população humana. A cada revisão do padrão de potabilidade, ao lado dos indicadores mais tradicionais de qualidade higiênica, novos indicadores são introduzidos com a preocupação de levar a um maior aperfeiçoamento do controle e da vigilância da qualidade da água. Desde a Portaria nº 36 de 19 de janeiro de 1990, até a Portaria nº 518 de 25 de março de 2004, atualizada em 12 de dezembro de 2011 pela Portaria nº 2914, do Ministério da Saúde, ocorreram mudanças sensíveis no padrão de potabilidade, sendo notável a recomendação da

monitoração dos indicadores mais específicos *Escherichia coli* e cianobactérias. Também, a monitoração de cianotoxinas e a recomendação da obtenção de valores de turbidez inferiores a 0,5 UT, na saída da unidade de tratamento, com vistas a assegurar a eficiência de remoção de enterovírus, cistos de *Giardia spp.* e oocistos de *Cryptosporidium sp.*, foram mudanças que objetivaram um controle de qualidade mais exigente por parte do operador do sistema de abastecimento. Há dois conjuntos de atividades estreitamente relacionados à oferta de água potável à população; o primeiro, exercido pelo operador do sistema de abastecimento, é o controle da qualidade, e o segundo, exercido pela autoridade de saúde pública, é a vigilância da qualidade da água para consumo humano (BRASIL, 2011).

De acordo com a “Diretriz nacional do plano de amostragem da vigilância em saúde ambiental relacionada à qualidade da água para consumo humano” (BRASIL, 2007), a monitoração da qualidade da água pode ser entendida como atividade de vigilância ou de investigação. Enquanto a monitoração de investigação consiste em buscar informações sobre a qualidade da água em casos de acidentes ou eventos de surto/epidemia, a monitoração de vigilância consiste em avaliar, continuamente, a qualidade da água consumida pela população, permitindo a identificação de fatores de riscos e a definição de estratégias de melhoria da situação existente, além do acompanhamento dos impactos resultantes das medidas implementadas. A monitoração de vigilância da qualidade da água deve conciliar os princípios de amostragem (abrangência e representatividade) com outros aspectos como infraestrutura e capacidade instaladas e a complementaridade com o controle de qualidade levado a efeito pelo serviço de abastecimento.

A monitoração de vigilância da qualidade da água deve ser baseada num plano de amostragem, implementado pelos municípios em articulação com os estados devendo ser observado que todos os municípios devem implantar os indicadores sentinelas, independentemente de seu porte, bem como realizar análise dos dados de qualidade da água para consumo humano em conexão com as informações constantes no Sistema de Monitorização das Doenças Diarreicas Agudas - MDDA.

Comumente, há uma imediata associação da vigilância da qualidade da água à determinação de indicadores microbiológicos, embora seja reconhecido que não existem organismos que indiquem a presença ou ausência da ampla variedade de patógenos possíveis de serem inativados ou resistirem aos diversos processos de tratamento de água (BASTOS, 2005). Mas, variáveis físico-químicas revestem-se de grande valor em vista de serem determinadas através de métodos mais simples, rápidos e de menor custo, exigindo nível de especialização bem menor que aqueles para as determinações microbiológicas, podendo ser

usadas como indicadores sentinelas. O termo sentinelas (BRASIL, 2011) é utilizado em analogia às chamadas fontes sentinelas e aos Sistemas de Vigilância Sentinela bastante utilizados na Epidemiologia. Utilizado para os indicadores sanitários, o termo pretende conferir aos mesmos a condição de instrumentos de identificação precoce de situações de riscos, em relação à água consumida pela população, que podem resultar em doenças de transmissão hídrica passíveis de prevenção e controle com medidas de saneamento básico.

No âmbito da vigilância da qualidade da água, os indicadores sentinelas referidos na “Diretriz nacional do plano de amostragem da vigilância em saúde ambiental relacionada à qualidade da água para consumo humano” (BRASIL, 2011) são a turbidez e o cloro residual livre (CRL). De fato, esses indicadores estão relacionados mais diretamente à qualidade higiênica, mas outros indicadores como o pH e o cloro residual combinado (CRC) devem fazer parte do elenco de indicadores auxiliares, em vista de sua relação com a desinfecção da água.

A turbidez que na rede de distribuição deve ser sempre inferior a 5,0 UT (BRASIL, 2011) e é causada pela presença de material em suspensão e coloidal, guarda uma relação inversa com a transparência da água, além de poder comprometer o desempenho da desinfecção. O cloro residual livre (CRL) exprime a soma das concentrações de ácido hipocloroso (HClO) e do radical hipoclorito (ClO^-), sendo o primeiro o real agente desinfetante e o segundo uma espécie de desinfetante potencial que, ao reagir com íons hidrogênio (H^+), forma ácido hipocloroso. O padrão de potabilidade brasileiro estabelece a faixa entre 0,2 e 2,0 mgCRL/L, em qualquer ponto da rede de distribuição de água, devendo a desinfecção ser realizada num pH inferior a 8,0, com um tempo de contato mínimo de 30 minutos.

O pH é uma das variáveis químicas de maior influência sobre a qualidade da água, por conferir agressividade quando inferior a 6,0 e propiciar a precipitação de carbonato de cálcio, quando superior a 9,5. O desempenho da desinfecção com cloro é altamente dependente do pH, sendo melhor em águas mais ácidas. O padrão de potabilidade (BRASIL, 2011), ora em vigor no Brasil, recomenda que o pH da água de abastecimento seja mantido na faixa entre 6,0 e 9,5.

Monocloramina (NH_2Cl), resultante da reação de ácido hipocloroso com sais de amônio, e dicloramina (NHCl_2) e tricloramina (NCl_3), formadas, respectivamente, pela oxidação de mono e dicloramina pelo ácido hipocloroso, constituem o cloro residual combinado (CRC) o qual pode coexistir com o CRL em sistemas de abastecimento de água,

particularmente em locais da rede de distribuição, como reservatórios e extremidades da rede, que favoreçam a acumulação de matéria orgânica nitrogenada passível de amonificação. Quantidades significativas de CRC indicariam manutenção deficiente e sua pesquisa pode contribuir para o estabelecimento de um programa mais eficiente de manutenção dos sistemas de abastecimento de água apresentando certa complementaridade com a contagem de bactérias heterótrofas facultativas (BFH). CRC, como mono e dicloramina têm poder desinfetante reduzido em comparação com o CRL, além de ser responsável pela ocorrência de algum odor e por uma significativa demanda de cloro livre.

As bactérias heterotróficas podem ser identificadas a partir do gênero as quais pertencem, podendo ser: *Pseudomonas*, *Clostridium*, *Desulfovibrio*, *Serratia*, *Mycobacterium*, *Acinetobacter*, *Alcaligenes*, *Klebsiella*, *Flavobacterium*, *Micrococcus*, *Bacillus*; além das espécies: *Escherichia coli*, *Aerobacter aerogenes*, *Citrobacter freundii* e *Klebsiella sp* (HELLER E PÁDUA, 2006).

As bactérias mesófilas aeróbias heterotróficas estão presentes em águas superficiais em grande número e são indicativas de insalubridade (COELHO et al., 2010). Segundo Farache Filho e Dias (2008), mesmo que a maioria das bactérias heterotróficas da flora natural da água não seja considerada patogênica, é importante que sua densidade seja mantida sob controle, pois densidades muito elevadas dessas bactérias na água podem causar riscos à saúde do consumidor. Tais indicadores são essenciais para o monitoramento da qualidade da água por caracterizarem a relevância do sistema operacional e a segurança da qualidade da água para uso humano. Entretanto, utilizá-los num processo de tomada de decisão implicaria na necessidade de análises profundas e que fossem capazes de abranger todos os perfis possíveis diante do sistema operacional.

Diante disto, os sistemas inteligentes são importantes ferramentas no planejamento, monitoramento e aperfeiçoamento eficazes de ações de saneamento. Isto se deve à grande capacidade de modelos inteligentes generalizarem, ou seja, de convergirem para qualquer sistema a ser simulado. Não são necessárias leis físicas, químicas ou de origem fenomenológica, pois o principal mecanismo que torna as redes neurais aplicáveis é o aprendizado.

Os mapas auto-organizáveis são especialmente inspirados no córtex cerebral dos mamíferos, cujas evidências biológicas demonstram que a sua organização é altamente estruturada, resultando em regiões do cérebro capacitadas em funções específicas de processamento sensorial de sinais como visão, audição, controle motor, entre outros (VAN HULLE, 2000; KOHONEN, 1997). Os neurônios tornam-se sensíveis a determinados estímulos em particular e a outros não, especializando-se no processamento de um determinado sinal, que

pode ser explicado pela separação dos canais nervosos que ligam os órgãos sensoriais ao cérebro.

As células corticais são organizadas de um modo complexo e abstrato: elas tornam-se sensíveis aos estímulos de acordo com uma relação específica de similaridade, onde seus parâmetros são ajustados para que tal comportamento ocorra.

Desta forma, o aprendizado competitivo, sob a ótica de uma rede neural artificial, tem o sentido de quantificação vetorial.

O que diferencia tal algoritmo dos demais é que apenas um neurônio de saída está ativo em um determinado instante, o que torna o modelo adequado para cenários em que as características são estatisticamente acentuadas, assim como um sistema de abastecimento de água (ETA), em que os indicadores de qualidade da água tratada são não lineares e tornam um modelo matemático mecanicista muito robusto. A aplicação da rede SOM (*self organizing maps*) resulta em simulações semelhantes ao cenário real, com menor esforço computacional.

1.1 Objetivos

1.1.1. Objetivo geral

Desenvolver um modelo neuro-fuzzy para o diagnóstico da qualidade da água do sistema de abastecimento de Campina Grande - PB, fornecendo aos operadores do sistema uma ferramenta de tomada de decisão.

1.1.2 Objetivos específicos

- Gerar mapas SOM a partir da base de dados obtidos pelo monitoramento em pontos estratégicos da cidade de Campina Grande, utilizando indicadores preconizados pela “Diretriz Nacional do Plano de Amostragem da Vigilância em Saúde Ambiental Relacionada à Qualidade da Água para Consumo Humano” (BRASIL, 2007);
- Gerar mapas SOM considerando os pontos de monitoramento como variáveis do modelo na identificação de problemas relativos aos indicadores da qualidade da água;
- Identificar os grupos formados no processo de aprendizagem da rede SOM, a correlação existente entre as variáveis e as justificativas para cada perfil identificado;
- Comparar os resultados gerados pela rede neural auto-organizável e os dados de treinamento;
- Extrair regras difusas do mapa SOM, com base nos pesos de maior potencial, que caracterizem o sistema;
- Simular cenários que possibilitem a tomada de decisão, considerando as possíveis situações enfrentadas por um operador do sistema de abastecimento.

FUNDAMENTAÇÃO TEÓRICA

2.1 Sistemas de abastecimento de água

A água é um elemento essencial ao ser humano; quando contaminada, influencia diretamente a saúde e a qualidade de vida da população. De acordo com a Organização Mundial de Saúde (OMS), “todas as pessoas em quaisquer estágios de desenvolvimento e condições socioeconômicas têm o direito de ter acesso a um suprimento adequado de água potável e segura”. Para assegurar esse direito, os sistemas de abastecimento de água desempenham um importante papel, uma vez que, nesses sistemas a água bruta passa por diferentes processos até que se torne potável e não ofereça risco à saúde da população.

A Portaria do Ministério da Saúde nº2914, de 12 de dezembro de 2011, define sistema de abastecimento de água para consumo humano como sendo:

Uma instalação composta por um conjunto de obras civis, materiais e equipamentos, destinada à produção e à distribuição canalizada de água potável para populações, sob a responsabilidade do poder público, mesmo que administrada em regime de concessão ou permissão, assegurando que a população receba um produto que não ofereça risco à saúde.

A implantação de sistemas de abastecimento de água se constitui no maior investimento para a saúde pública, visto que há uma diminuição da incidência de doenças de veiculação hídrica que acometem elevado número de pessoas, principalmente crianças menores de cinco anos.

Segundo Heller et al. (2006), os sistemas de abastecimento de água são compostos pelas seguintes unidades:

- Unidade de produção: inclui captação no manancial, adução de água bruta, tratamento e adução de água tratada;
- Unidade de distribuição: inclui os reservatórios e a rede de distribuição;
- Unidade de transporte: corresponde ao conjunto composto pela estação elevatória e a adutora correspondente.

De acordo com Barbosa et al. (1999), os sistemas de distribuição de água devem ser projetados e operados com o intuito de atingir vários objetivos:

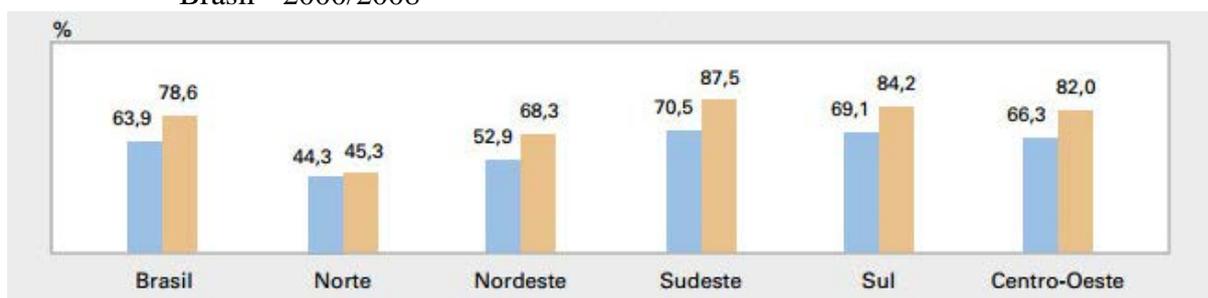
- i. Objetivos técnicos: ligados ao desempenho hidráulico, como por exemplo, garantia de pressões mínimas e máximas, garantia de água suficiente para proteção contra incêndio, confiabilidade operacional, etc.;
- ii. Objetivos econômicos: traduzem-se pela minimização de custos associados aos componentes dos sistemas e aos custos operacionais.

Antes da implantação de qualquer sistema de abastecimento de água, é necessário considerar os elementos condicionantes que irão auxiliar na escolha do tipo de sistema a ser instalado e a tecnologia de tratamento a ser utilizada, tais como porte da localidade, densidade demográfica, mananciais, características topográficas, instalações existentes, características geológicas e geotécnicas, energia elétrica, recursos humanos, tecnologia de tratamento e alcance do projeto no tempo (HELLER et al., 2006). Segundo os mesmos autores, pode-se destacar, dentre os demais elementos, a importância do porte da localidade, visto que comunidades de pequeno porte podem proporcionar uma simplificação do sistema, em contrapartida as localidades de maior porte podem requerer sistemas mais complexos, em termos de sofisticação tecnológica e operacional. Na escolha do manancial deve-se realizar uma análise conjunta da quantidade e qualidade da água e com relação ao alcance do projeto no tempo é necessário fazer estimativas de crescimento populacional para que o projeto atenda à demanda presente e não se torne insuficiente para as futuras gerações. A tecnologia de tratamento a ser adotada no sistema deve conduzir a um menor custo sem, contudo, deixar de lado a segurança na produção de água potável (PARSEKIAN, 1998).

A implantação de um sistema de abastecimento requer um embasamento em várias normas da Associação Brasileira de Normas Técnicas, como por exemplo, a NBR 12.211 (ABNT, 1989) que trata dos estudos de concepção de sistemas públicos de abastecimento de água, ou seja, trata dos estudos de arranjos, sob o ponto de vista qualitativo e quantitativo, das diferentes partes de um sistema, organizados de modo a formarem um todo integrado, para a escolha da melhor solução sob os pontos de vista técnico, econômico, financeiro e social. Além das normas da ABNT deve-se levar em consideração a Portaria do Ministério da Saúde nº2914/2011, referente à qualidade da água para consumo humano.

De acordo com a Pesquisa Nacional de Saneamento Básico (PNSB) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2008), 33 municípios do País não dispunham de rede geral de distribuição de água em nenhum de seus distritos, valendo-se de soluções alternativas. Dentre essas, destacam-se o abastecimento efetuado através de carros-pipas (14 municípios), poços particulares (13) e chafarizes, bicas ou minas (3). Vale ressaltar que essa situação vem diminuindo sistematicamente no País: em 1989, a PNSB identificou 180 municípios sem o referido serviço, enquanto em 2000, 116 (Figura 2.1).

Figura 2.1 - Domicílios abastecidos de água por rede geral, segundo as grandes regiões do Brasil - 2000/2008



Fonte: Pesquisa Nacional de Saneamento Básico (IBGE, 2008).

Ainda segundo o PNSB, observa-se que, do primeiro PNSB feito em 1989, houve um crescimento de 3,5% nessa cobertura que, em 2008, alcançou a marca de 99,4% dos municípios do País. O maior avanço nesses 19 anos ocorreu na Região Norte, que aumentou de 86,9% dos municípios com o referido serviço para 98,4%, representando um acréscimo de quase 12 pontos percentuais. Ressalta-se que a Região Sudeste foi a única que apresentou a totalidade dos municípios que a integram abastecidos por rede geral de distribuição de água, em pelo menos um distrito ou parte dele, fato este já identificado na PNSB 2000. (Tabela 2.1).

Tabela 2.1 - Percentual de municípios do País, total de domicílios, economias residenciais e economias abastecidas em relação ao número de domicílios.

Grandes Regiões	Percentual de municípios do País	Domicílios		Economias abastecidas	
		Total (em 1 000 domicílios)	Distribuição percentual (%)	Total (em 1 000 economias)	Percentual em relação ao número de
Brasil	100,0	57 656	100,0	45 343	78,6
Norte	8,1	4 035	7,0	1 829	45,3
Nordeste	32,2	15 011	26,0	10 254	68,3
Sudeste	30,0	25 341	43,9	22 183	87,5
Sul	21,3	9 004	15,6	7 582	84,2
Centro-Oeste	8,4	4 264	7,4	3 495	82,0

Fonte: Pesquisa Nacional de Saneamento Básico (IBGE, 2008).

2.1.2 Realidade brasileira quanto à operação, manutenção e qualidade da água de abastecimento público.

A operação de um sistema de abastecimento de água em tempo real é uma tarefa muito complexa, devido à necessidade de se garantir confiabilidade no atendimento dos serviços (CARRIJO e REIS, 2004), e manter o fornecimento de água com qualidade à crescente população.

Uma das maiores dificuldades enfrentadas atualmente pelo setor de saneamento é o aumento populacional e, conseqüentemente, o aumento da demanda por água potável, porém, muitos dos sistemas de abastecimento foram projetados e implantados há muito tempo pelas empresas de saneamento. De acordo com Bággio (1998), essas empresas foram criadas em uma época de enormes demandas por projetos e construções de sistemas de abastecimento de água a fim de superar os déficits da cobertura de serviços, tendo sido negligenciados os aspectos de operação dos ditos sistemas. As conseqüências são sentidas hoje, constatando-se sérios comprometimentos na rotina dessas prestadoras de serviços que se deparam com inúmeras reclamações dos consumidores, a respeito da má qualidade dos serviços, chegando, em alguns casos, a conviverem com racionamentos e/ou desabastecimentos, desgastando seriamente a imagem das empresas. Esse cenário é visualizado frequentemente nas cidades com sistemas de abastecimento de pequeno porte, onde, rotineiramente, ocorrem interrupções nos serviços de abastecimento.

Outro aspecto importante que tem sido observado é que nas cidades de pequenos e médios portes, os sistemas de abastecimento são operados por funcionários (operadores) que, na maioria das vezes, possuem baixo grau de escolaridade dificultando o acesso a cursos de capacitação sobre operação e manutenção das ETAs. Conseqüentemente, a maior parte desses sistemas é operada de forma “intuitiva”, ou seja, com base na experiência dos operadores. Estudos realizados por Parsekian (1998) comprovaram que o nível de escolaridade dos operadores de ETAs é relativamente baixo, cerca de 32% cursaram o 1º grau incompleto e 30% o 1º grau completo, ou seja, mais da metade do pessoal que trabalha nas estações de tratamento possui no máximo o primeiro grau, não percebendo, muitas vezes, o tamanho de sua responsabilidade.

2.3 Redes auto-organizáveis

Em 1962, os neurofisiologistas Hubel e Wiesel, este sueco e aquele canadense, realizaram um experimento que visava a reprodução computacional de estímulos visuais do córtex cerebral de gatos. Tal experimento foi a principal motivação para a construção de modelos de redes neurais cujas características se baseiam na cognição e transmissão de sinais do córtex cerebral.

Em 1973, von der Malsburg apresentou o primeiro modelo de redes neurais auto-organizáveis para a modelagem do córtex visual das vértebras posteriores, em mapas característicos. Sucessivamente, Teuvo Kohonen, em 1982, desenvolveu um algoritmo de codificação vetorial, cujo objetivo é a produção de um mapa topológico que localiza vetores e um espaço de dimensão elevada, para aumentar a compreensão destes dados. Os fundamentos para este modelo derivam da auto organização, adquirida de considerações neurobiológicas, bem como de bases adquiridas da teoria das comunicações, já que o modelo utiliza um codificador e um decodificador, cujas compreensões serão facilitadas nos tópicos adiante.

Em 1987, surgia um modelo com forte inspiração biológica e cognitiva, o qual fundamenta sua construção de processamento *online*, ou seja, proporciona uma solução para o dilema estabilidade-plasticidade, denominado de ART - *Adaptive Resonance Theory*. Tal modelo foi desenvolvido pelo professor Gail Carpenter, originando uma série de modelos da família ART. Basicamente, um modelo ART busca solucionar um problema muito comum em redes neurais, que é a capacidade de adaptação da mesma. Suponha-se uma rede cuja expectativa é de que ela forneça a resposta correta para uma determinada situação. Com o passar do tempo, espera-se que a rede seja capaz de executar sua tarefa de forma satisfatória para quaisquer padrões apresentados. Entretanto, ao longo do tempo o conhecimento sobre as informações contidas no domínio pode ser modificado, bem como os padrões inseridos na rede. Isso significa dizer que o desempenho da rede pode decair em relação à fase de treinamento. A proposta é ter um algoritmo plástico, ou seja, que se adapte a diferentes situações indefinidamente. Além da plasticidade, a rede neural deve ser estável de modo que o conhecimento previamente adquirido seja preservado, bem como que este não se sobreponha a conhecimentos futuros (HAYKIN, 1993).

Uma rede auto-organizável tem a peculiaridade de agrupar e classificar dados independentemente da presença de um supervisor ou crítico externo. Para tal, os dados devem apresentar similaridades e diferenças, que os distingam de ruídos aleatórios.

Quanto ao algoritmo de treinamento, ele pode ser classificado em aprendizado competitivo e aprendizado hebbiano. O algoritmo competitivo é mais usado para problemas de classificação, reconhecimento de padrões, extração de características e agrupamento de dados. Já o aprendizado hebbiano, proposto pelo neurofisiologista Donald Hebb, em 1949, é baseado no princípio de atualização dos pesos, que estabelece que se houver sincronismo entre os níveis de atividade das entradas e saídas, os pesos de uma conexão sináptica devem ser ajustados (HEBB, 1949). Então, conclui-se que uma rede SOM (*self organizing maps* ou mapas auto-organizáveis) pode ser usada para treinamento e mapeamento, implicando na classificação automática do vetor de entrada, e este na formação dos mapas usando exemplos, podendo ser um conjunto de dados de treinamento.

2.3.1 Mapas auto-organizáveis

Um mapa auto-organizável é caracterizado por uma grade neural, baseada na aprendizagem competitiva. Nesta grade, os neurônios competem entre si, a fim de se manterem ativados, de modo que apenas um neurônio ficará ativo por um determinado tempo. Este neurônio é denominado de “neurônio vencedor” e se mantém ativo por conexões laterais inibitórias.

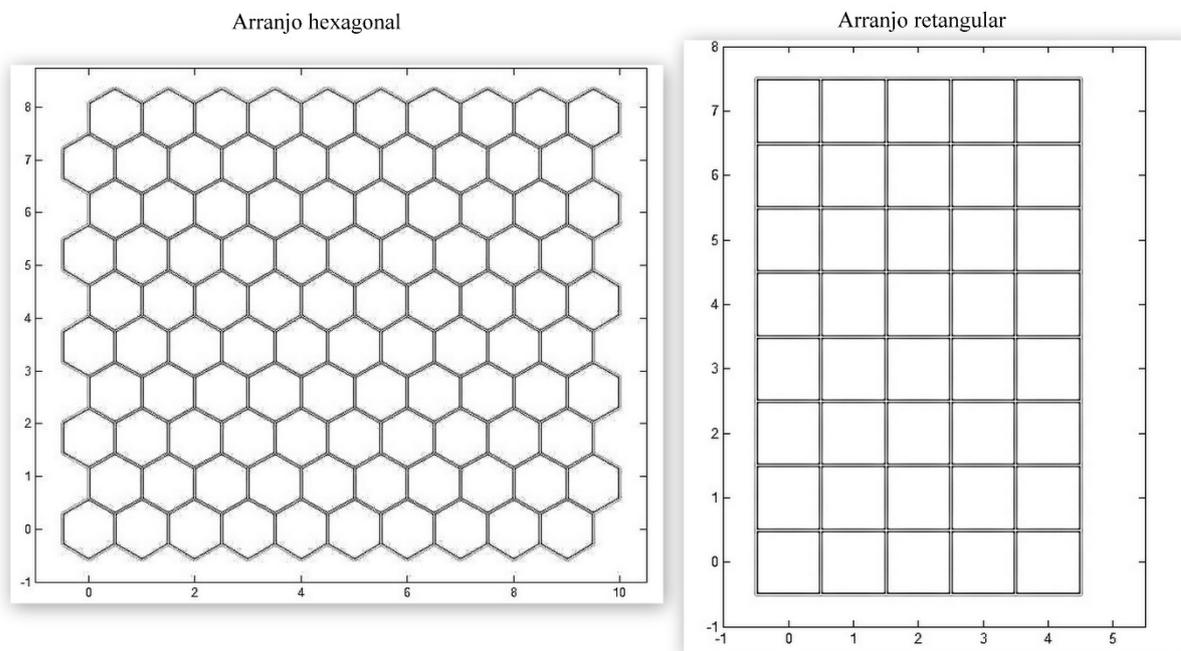
A organização dos neurônios é feita uni ou bidimensional, podendo ser também de dimensões maiores. Os neurônios se tornam seletivos em relação aos padrões, de modo que a localização ou coordenadas dos mesmos indicam características intrínsecas aos dados. Assim, as localizações dos neurônios sintonizados aos vários padrões e estímulos se tornam peculiares por causa de sua ordem, para diversas características.

A principal inspiração de um mapa auto-organizável é a estrutura do córtex cerebral, caracterizada pela setorização das funções, onde funções como a tátil e motora são mapeadas em áreas diferentes do cérebro. Então, pode-se descrever um mapa computacional como sendo processadores que operam paralelamente a respeito dos sinais a eles transmitidos. Cada sinal de entrada recebe uma distribuição de valores calculados de parâmetros por valor de atividade máxima, de forma que a informação possa ser facilmente localizada (KOHONEN, 2001).

O arranjo da rede SOM é feita através da apresentação dos nós de entrada, ou camada de entrada, ou ainda denominada de modelo de referência, que está totalmente conectada aos neurônios da camada de saída. Tais conexões ou pesos possuem valores correspondentes aos dos vetores da camada de entrada. Os neurônios da camada de saída podem se organizar entre

si por arranjos retangulares, em que cada neurônio apresenta oito neurônios vizinhos, e de forma hexagonal, com seis neurônios vizinhos para cada neurônio (Figura 2.2).

Figura 2.2 - Tipos de arranjos da rede SOM.



Fonte: Próprio Autor

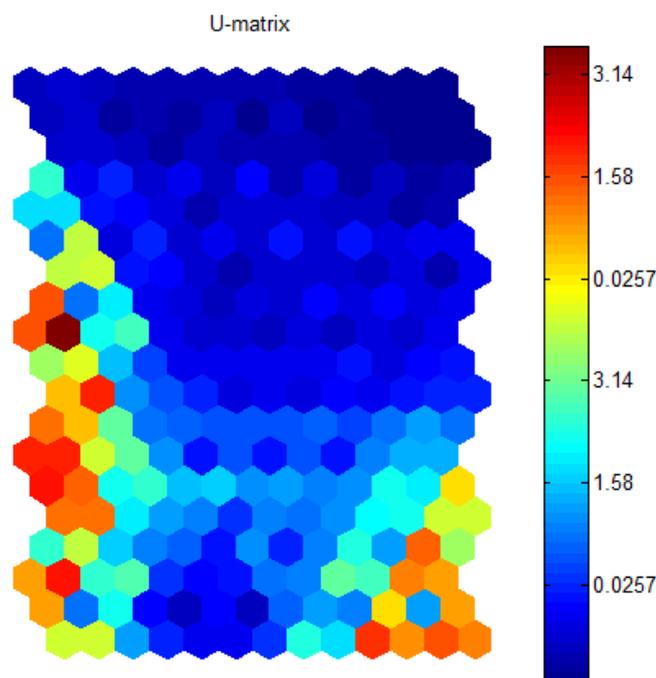
2.3.1 Formas de visualização

O modo mais comum de visualização dos resultados de um mapa auto-organizável é denominada de U-Matriz, matriz de distância unificada ou Matriz-U. Esse método apresenta as abstrações do SOM como agrupamentos.

O cálculo se baseia na distância entre matrizes, estimando as distâncias entre os pesos sinápticos dos neurônios em relação a cada um de seus vizinhos, interpretando estes valores por meio de uma escala de cores.

A Figura 2.3 apresenta um exemplo de Matriz-U, sendo possível verificar uma escala de cores que varia no sistema RGB (*red green blue*). Uma Matriz-U é gerada após o processo de aprendizagem, indicando as maiores distâncias entre os neurônios através da cor vermelha e as menores distâncias através da cor azul. As distâncias menores representam os grupos formados no processo de aprendizagem e seu comportamento no modelo (SILVA, 2013) As cores em tons de azul representam menores distâncias entre os neurônios; cores em tons de vermelho representam maiores distâncias.

Figura 2.3 - Matriz-U e sua escala de cores.



Fonte: Próprio Autor

2.3.2 Algoritmo de treinamento

O algoritmo de treinamento do SOM visa ajustar os pesos sinápticos de modo que estes sejam capazes de representar os domínios de entrada, submetidos à rede. Os dados poderão sofrer ou não normalização, a fim de que eles sejam colocados na mesma escala. A rede é alimentada com um conjunto de dados extenso, de modo que a mesma possa se aproximar o máximo possível do cenário a ser modelado. Este processo de alimentação é caracterizado pela iteratividade (VESANTO et al., 2000).

a) Inicialização dos pesos e neurônio vencedor

Os vetores de pesos podem ser inicializados de forma aleatória, em que os valores para os componentes dos vetores de referência são gerados randomicamente. Os pesos também podem ser inicializados de forma linear, o que se dá com a determinação de dois autovetores da matriz de autocorrelação de x com os maiores autovalores (SILVA, 2013).

O treinamento utiliza aprendizado competitivo. Quando um conjunto de dados é apresentado para a rede, a distância euclidiana entre o vetor de dados e os vetores peso é

calculada. O neurônio cujo vetor peso é mais similar à entrada é chamado de “a melhor unidade de correspondência”, do inglês “best match unit” - BMU. Os pesos do BMU e neurônios próximos a ele na estrutura SOM são ajustados ao vetor de entrada (Figura 2.4). A magnitude da mudança diminui com o tempo e com a distância (dentro da rede) a partir da BMU. A fórmula de atualização para um neurônio v com vetor de pesos $W_v(s)$ é dada pela Equação 2.1.

$$W_v(s+1) = W_v(s) + \Theta(u, v, s) \alpha(s) (D(t) - W_v(s)) \quad Eq (2.1)$$

em que s é o índice do passo, t um índice para a amostra do treino, u é o índice do BMU em relação à $D(t)$, $\alpha(s)$ é um coeficiente de aprendizagem e $D(t)$ é o vector de entrada; $\Theta(u, v, s)$ é a função de vizinhança, que indica a distância entre os neurônios u e v no passo s .

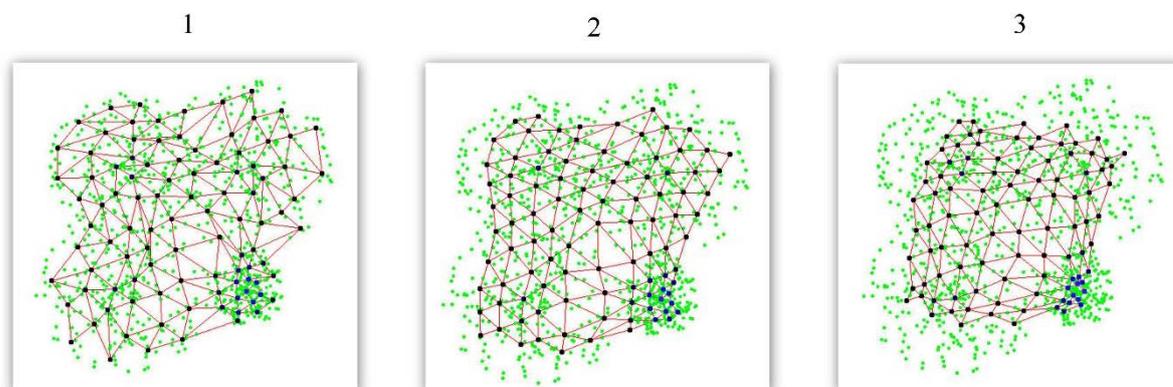
Com o término do treinamento e os grupos definidos, o procedimento para seleção do neurônio com maior potencial para representação de cada grupo deve ser realizado. O cálculo do potencial de um neurônio para representar uma determinada classe é dado por (CHIU, 1994):

$$P_{ji} = \sum_{l=1}^{N_i} \exp(-\|w_j - x(l)\|^2) \quad Eq (2.2)$$

sendo P_{ji} a medida do potencial do neurônio j que pertence ao grupo i , N_i o número de padrões do vetor de entrada alocados para o grupo i , w_j o vetor de pesos do neurônio j , $x(l)$ o padrão l do vetor de entradas e $\|\cdot\|$ a distância Euclidiana.

Assim, a medida do potencial de um neurônio é uma função da distância do seu vetor de pesos em relação aos N padrões de entrada alocados ao grupo, ao qual este neurônio pertence, concluindo-se que quanto menor a distância calculada, maior será o potencial do neurônio (VESANTO et al., 2000).

Figura 2.4 - Treinamento de uma rede SOM. Fase de ajuste dos pesos (1), processo de cooperação dos neurônios vizinhos (2) e processo de adaptação (3) em que a rede não sofre mais mudanças em suas iterações.



Fonte: Próprio Autor

A função de vizinhança $\Theta(u, v, s)$ depende da distância entre a estrutura BMU (neurônio u) e neurônio v . Na forma mais simples, o valor 1 é atribuído para todos os neurônios próximos o suficiente do BMU e 0 para os outros. A função de Gauss é uma escolha comum, entretanto, independentemente da forma funcional, a função de vizinhança diminui com o tempo. Isto significa que no início, quando a os neurônios estão afastados entre si, a auto-organização ocorre em escala global. Quando a distância é reduzida para apenas um par de neurônios, os pesos estão convergindo para as estimativas locais.

O processo é repetido para cada vetor de entrada para um (geralmente grande) número de ciclos λ . A rede associa nós de saída com grupos ou padrões do conjunto de dados de entrada. Se estes padrões podem ser nomeados ou rotulados, os nomes podem ser ligados aos nós associados na rede treinada.

Durante o mapeamento, haverá um único neurônio vencedor: o neurônio cujo vetor peso encontra-se mais próximo do vetor de entrada

2.3.3 Interpretação do mapa SOM

Há duas maneiras de interpretar um SOM: uma seria no processo de treinamento, onde os pesos de toda a vizinhança são movidos na mesma direção do neurônio vencedor. Desta forma, pode-se dizer que itens semelhantes tendem a excitar neurônios adjacentes. Portanto, uma rede SOM forma um mapa semântico onde amostras semelhantes são mapeadas juntas e vetores desiguais em intervalos desiguais. Isto pode ser visualizado por uma U-Matrix.

A outra maneira é pensar em pesos neuronais como direcionadores do espaço de entrada. Eles formam uma aproximação discreta da distribuição de amostras de treino. Mais neurônios apontam para regiões com alta concentração da amostra de treinamento e menos onde as amostras são escassas.

Uma rede SOM pode ser considerada uma generalização não-linear de análise de componentes principais (PCA). Tem sido demonstrado, usando dados geofísicos artificiais e reais, que a rede SOM tem muitas vantagens sobre a extração de conhecimento e outros métodos convencionais tais como Funções Ortogonais Empíricas (EOF) ou APC (KIANG, 2001).

2.4 Lógica fuzzy

Para lidar com sistemas imprecisos, Lotfali Askar-Zadeh, matemático, engenheiro eletrônico e cientista da computação estadunidense, nascido no Azerbaijão, e professor de Ciência da Computação na Universidade da Califórnia em Berkeley, redescobriu e defendeu em 1965 a Lógica Nebulosa, ou Lógica Fuzzy. Esta havia sido introduzida em 1930 pelo filósofo e lógico polonês Jan Lukasiewicz, na qual através do estudo de termos do tipo alto, velho e quente, ele propôs a utilização de um intervalo de valores [0,1] que indicaria a possibilidade que uma declaração fosse verdadeira ou falsa. O primeiro conjunto Fuzzy, no entanto, foi proposto pelo filósofo Max Black, que descreveu também algumas ideias e operações preliminares. A lógica Fuzzy lida com sistemas vagos, em que os conjuntos de regras que definem o sistema avaliado podem ser expressos por meio de variáveis linguísticas, que são variáveis cujos valores são palavras ou sentenças (ZADEH, 1975).

Assim, para uma situação em que se pretende avaliar níveis de temperatura, cujos valores são expressos na forma de variáveis linguísticas (como alto, médio e baixo), em que os valores de temperatura podem ser agrupados nestas variáveis linguísticas. O modelo de lógica Fuzzy objetiva modelar computacionalmente situações subjetivas, vagas ou ambíguas.

2.4.1 Teoria dos conjuntos difusos

Zadeh (1975) propôs na Teoria dos Conjuntos Difusos uma generalização da Teoria Clássica dos Conjuntos, que diz que cada elemento tem uma pertinência μ , com valor 0 ou 1, que indicam pertinência ou não pertinência. Isso significa dizer que dado um conjunto, um determinado elemento é pertinente ou não pertinente se pertence ou não ao conjunto. Dessa forma, a lógica clássica denota pertinência limitada. Entretanto, um conjunto difuso permite vários graus de pertinência ou não pertinência, contidos num intervalo [0,1].

Então, pode-se concluir que, para a teoria clássica, um conjunto A definido sobre o conjunto universo U é denotado pela função característica $\mu_A(x)$:

$$\mu_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases} \quad \text{Eq. (2.3)}$$

Ao transformar o conjunto clássico da Equação 2.3 para um conjunto difuso, faz-se necessário o uso da teoria da possibilidade, que permite múltiplas possibilidades de soluções

definidas através de uma função de distribuição. A função de distribuição, também chamada de função de pertinência, indica a adequação de uma amostra a um conceito ou possibilidade do evento ocorrer. Vale salientar que esta adequação não se enquadra em um conceito estatístico, ou seja, não se trata da chance do evento ocorrer, mas sim da opinião de especialista sobre a ocorrência do evento (ZADEH, 1975).

Sabendo que os conjuntos difusos são capazes de lidar com verdades vagas ou imparciais, as funções de pertinência se distinguem por mapear um valor entre 0 e 1, indicando seu grau de pertinência.

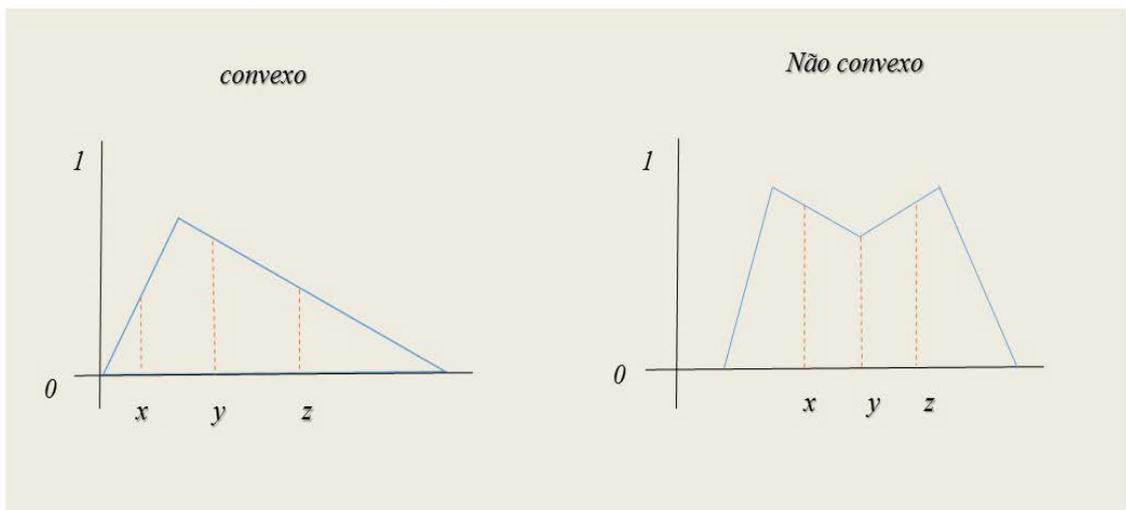
$$\mu_A(x): x \rightarrow [0 \ 1] \quad \text{Eq. (2.4)}$$

Para a representação de conjuntos Fuzzy, é necessária a escolha da função de pertinência. Para tanto, é necessário avaliar qual problema deverá ser modelado e também a capacidade computacional disponível para o processamento das respostas. Funções não lineares são mais usadas em problemas de maior complexidade, em contrapartida, demandam maior esforço computacional.

As funções de pertinência podem ser descritas pelas seguintes propriedades:

- I. O suporte da função de pertinência de um conjunto difuso A é a região do universo em que $\mu_A(x) > 0$.
- II. Um conjunto difuso normal é aquele, cuja função de pertinência possui ao menos um elemento x no universo com pertinência igual a 1. Caso nenhum elemento x apresente tal valor de pertinência, o conjunto é dito subnormal.
- III. Um conjunto difuso convexo é descrito por uma função de pertinência, cujos valores de pertinência comportam-se de três formas: estritamente monotonicamente crescentes ou estritamente monotonicamente decrescentes ou, estritamente monotonicamente crescentes em seguida estritamente monotonicamente decrescentes, depois estritamente crescente de novo. Portanto, se A é um conjunto difuso e a relação $x < y < z$ existe em A , então tem-se a Figura 2.5 (ROSS, 2004).

Figura 2.5 - Conjunto difuso convexo e não-convexo.



Fonte: Adaptado de ROSS (2004).

A forma geométrica da função de pertinência é determinada pela imprecisão na variável difusa correlata. A função mais utilizada é a triangular, por ser considerada mais prática, mas também são usadas com frequência a função trapezoidal, função-s, função- π e função-z.

2.4.2 Operações em conjuntos difusos

Tal qual a teoria clássica dos conjuntos, os conjuntos difusos obedecem a certas propriedades e podem ser operados de diversas maneiras. As operações básicas da teoria dos conjuntos difusos são: complemento, interseção (t-norma) e união (t-conorma) e são baseadas no conceito de pertinência ou não de um elemento a um conjunto difuso, entretanto outras operações e propriedades são apresentadas na Quadro 2.1 considerando A , B e C conjuntos fuzzy num universo U . As operações entre conjuntos são importantes pois os cálculos que regem o modelo são baseados nas mesmas.

Quadro 2.1 - Quadro representativo das operações propriedades dos conjuntos difusos.

Operações e Propriedades	Informações
<i>Igualdade</i>	$A=B \leftrightarrow (\mu_A(x) = \mu_B(x)), \forall x \in U$ Dois conjuntos são iguais quando possuem os mesmos elementos.
<i>Inclusão</i>	$A \subseteq B \leftrightarrow (\mu_A(x) \leq \mu_B(x)), \forall x \in U$ A é subconjunto de B.
<i>União</i>	$A \text{ OR } B = A \cup B = \{(x, \max(\mu_A(x); \mu_B(x))) \mid x \in U\}$ Zadeh (1960) define que um conjunto união deve obedecer as propriedades: Comutativa, Associatividade, Monotonicidade e Coerência nos contornos.
<i>Interseção</i>	$A \text{ AND } B = A \cap B = \{(x, \min(\mu_A(x); \mu_B(x))) \mid x \in U\}$ Zadeh (1960) define que um conjunto interseção deve obedecer às normas-T caracterizadas pelas seguintes propriedades: Comutativa, Associatividade, Monotonicidade e Coerência nos contornos.
<i>Complemento</i>	$\text{NOT } A = \neg A = \{(x, \mu_{\neg A}(x)) \mid x \in U \text{ e } \mu_{\neg A}(x) = 1 - \mu_A(x)\}$ O complemento de um conjunto fuzzy A pode ser denotado por $\neg A$.
<i>Diferença</i>	$A - B = (x, (\mu_A \cap_{\neg B}(x))) \mid x \in U$ A diferença entre dois conjuntos é o que sobra da interseção desses conjuntos.
<i>Idempotência</i>	$A \cup A = A$ $A \cap A = A$ Não há diferenciação para operações com conjuntos fuzzy idênticos.
<i>Identidade</i>	$A \cap \emptyset = A$ $A \cup \emptyset = A$ $A \cap U = \emptyset$ $A \cup U = U$ Sejam \emptyset o conjunto fuzzy vazio e U o conjunto que representa o universo e considere os graus de pertinência do conjunto vazio 0 e do conjunto-universo.
<i>Absorção</i>	$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$ Propriedade com caráter <i>booleano</i> .
<i>Comutatividade</i>	$A \cap B = B \cap A$ $A \cup B = B \cup A$ A ordem dos operandos não altera o resultado final.
<i>Associatividade</i>	$(A \cap B) \cap C = A \cap (B \cap C)$ $(A \cup B) \cup C = A \cup (B \cup C)$ A ordem dos operandos não altera o resultado, para o caso de três conjuntos.
<i>Distributividade</i>	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ A ordem em que as operações são efetuadas podem, de certa forma, alterar o resultado.
<i>Complemento duplo</i>	$\neg(\neg A) = A$ Complemento duplo de um conjunto é reversível.
<i>Lei Transitiva</i>	Se $A \overset{TM}{\sim} B$ e $B \overset{TM}{\sim} C \rightarrow A \overset{TM}{\sim} C$ Se A tem relação com B, e B tem relação com C, então A deve ter relação com C.
<i>Lei de Morgan</i>	$\neg(A \cup B) = \neg A \cap \neg B$ $\neg(A \cap B) = \neg A \cup \neg B$ O complementar da reunião de uma coleção finita de conjuntos é a interseção dos complementares desses conjuntos.

Fonte: Próprio Autor

2.4.3 Regras fuzzy

Para obter os resultados de um modelo fuzzy é necessária a implementação de regras que serão responsáveis por operar os conjuntos difusos. É preciso um raciocínio coerente, dividido em duas etapas: (1) avaliar o antecedente à regra e (2) avaliar o resultado consequente à regra. Uma base de regras difusas é composta basicamente pelos elementos SE-ENTÃO, especificando a relação linguística que uma variável possui em relação às saídas do sistema.

Conclui-se que a parte SE da regra é a antecedente da regra e a parte ENTÃO é a consequente da regra.

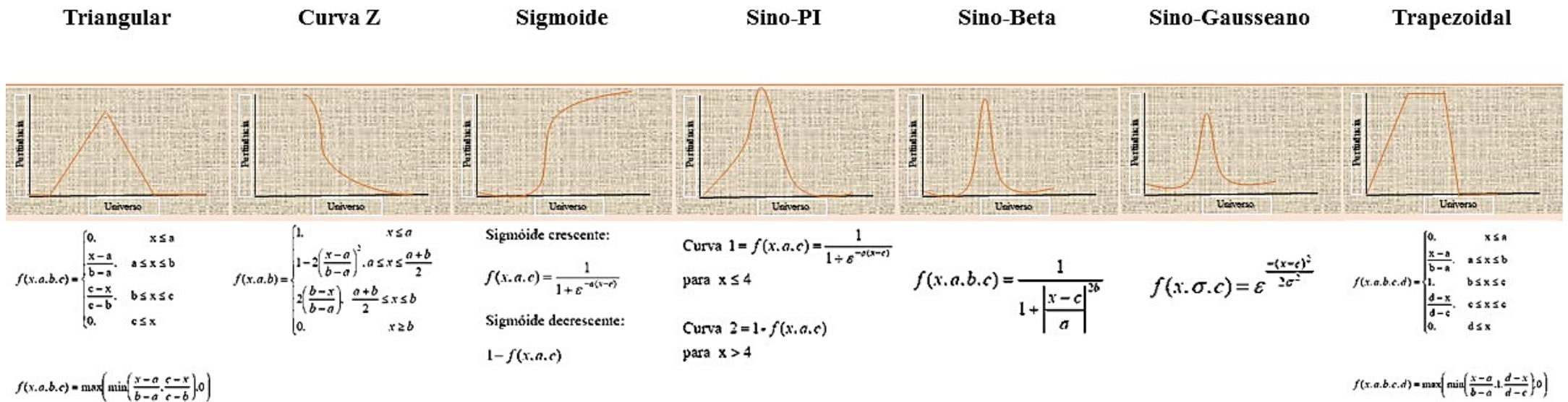
Para casos em que existem vários antecedentes, é necessário encontrar um grau de pertinência resultante de todos os antecedentes. Há ainda a possibilidade de se utilizar o conectivo E. Para este caso, métodos de combinação devem ser usados, com a condição de não ultrapassar o valor de menor pertinência.

2.4.4 Funções de pertinência

A função de pertinência apresenta o significado ou a interpretação para o modelo. Dessa forma, pode-se afirmar que uma função de pertinência expõe as propriedades semânticas conceituais. Então, ela apresenta a coerência de uma determinada proposição ou valor específico de uma variável, em relação a um grau de pertinência pré-estabelecido, comparando a entrada com o grau de pertinência.

Não há uma metodologia comum para a escolha das funções de pertinência. A atribuição da melhor função dependerá da complexidade do sistema a ser modelado e também do esforço computacional. A Figura 2.6 apresenta as principais funções de pertinência e suas representações gráficas.

Figura 2.6 - Funções de inferência e suas representações gráficas



Fonte: Próprio Autor

2.4.5 Inferência fuzzy

A inferência fuzzy é o processo de avaliação das entradas, com base nas regras previamente definidas e na teoria dos conjuntos. Esse processo é baseado na capacidade lógica do ser humano de tomar decisões, analisando as possibilidades existentes e os limites para cada situação.

A inferência é feita de acordo com o método utilizado, que, por sua vez, deve se adequar ao tipo de problema a ser resolvido. Existem vários métodos de inferência, entretanto, na presente pesquisa iremos abordar o estilo Mamdani, usado em sua composição, pois possui uma estrutura mais simples e ao mesmo tempo eficaz.

2.4.6 Sistemas de inferência Mamdani

O método de inferência Mamdani foi criado pelo professor Ebrahim Mamdani, da Universidade de Londres (Reino Unido), em 1975, baseando-se em regras difusas no intuito de representar experiências da vida real. O processo de raciocínio nesse sistema é dividido em quatro passos: fuzzyficação, avaliação das regras fuzzy, agregação das regras fuzzy e defuzzyficação.

a) Fuzzyficação

Nesta etapa, para a obtenção do grau de pertinência de uma determinada entrada, é necessário buscar esse valor na base de conhecimento do sistema fuzzy.

b) Avaliação das regras fuzzy

Depois de fuzzyficadas, as entradas são aplicadas aos antecedentes a fim de serem obtidos os consequentes para cada regra. Os operadores *e* e *ou* são usados para um único antecedente composto, obtendo assim um único resultado; então, para o caso do operador *ou* é usada a operação da união, e no caso do operador *e* é usado o operador de interseção.

Depois da obtenção de um único valor para o antecedente, é necessária a obtenção do valor do consequente através do método de correlação entre os mesmos. O método mais usado é denominado de Método de Correlação *Clipped* (“cortado” em inglês) ou Método da Correlação Mínima, que limita ou “corta” o valor do consequente para ao nível de verdade do

antecedente, ou seja, o conseqüente deve ser correspondente ao antecedente. Este método é mais usado por possuir conjunto de cálculos menos complexos e mais rápidos, gerando uma superfície de saída mais fácil de ser interpretada na etapa seguinte de defuzzyficação.

c) Agregação das regras fuzzy

Nesta etapa as funções membro dos conseqüentes de cada regra em um único conjunto fuzzy.

d) Defuzzyficação

O processo de defuzzyficação é caracterizado por ser a etapa que transforma a saída em valor. Para isto, é mais utilizada a técnica do centroide, que obtém o ponto onde uma linha vertical divide ao meio o conjunto agregado. A equação matemática que expressa esse ponto é:

$$COG = \frac{\sum_{x=a}^b \mu(x)x}{\sum_{x=a}^b \mu(x)} \quad Eq. (2.5)$$

A precisão do método depende do intervalo escolhido; quanto maior o intervalo, mais preciso, entretanto a velocidade do cálculo aumenta.

Pode-se concluir que o valor resultante do centroide é a interpretação numérica das variáveis de entrada de acordo com o conjunto de regras a elas associado.

2.5 Extração de conhecimento em base de dados

Extração de conhecimento (também conhecido como processo KDD, do inglês knowledge-discovery in databases), é um processo de extração de informações de base de dados, que cria relações de interesse que não são observadas pelo especialista no assunto. Recentemente, foram encontradas evidências em pedaços de argila de que há 5500 anos já se faziam registros de dados relativos a cultivo, colheita e outros negócios da época (PYLE, 1999).

A partir de então, o homem desenvolveu técnicas para entender o modo como os dados se comportam, e quais informações poderiam ser extraídas desses perfis. Em 1823, um matemático chamado Charles Babbage iniciou o desenvolvimento de uma máquina para tratamento de funções matemáticas, ferramenta que ficou conhecida por se aproximar do propósito de analisar dados e atribuir soluções. Em 1871, uma ferramenta se estendeu a esse propósito, denominada de *Calculating Engine*, que apesar de representar uma ideia brilhante, a ferramenta possuía falhas excessivas no processamento de informações (PEAT, 1985).

O invento de Babbage representou um elo para o avanço de análise de dados e processamento de informações. Os estudos do mesmo fizeram com que houvesse avanços importantes para o desenvolvimento do primeiro computador, chamado *Electronic Numerical Integrator and Calculator-Eniac* -, construído em 1944 por J. P. Eckert e J. W. Mauchy, com a ajuda de John von Neumann durante a fase de projeto. Conforme os avanços computacionais aconteceram nos anos seguintes, a quantidade de departamentos de análises de dados criados, principalmente no Estados Unidos, foram aumentando, de modo que um número cada vez maior de dados pôde ser processado (NEUMANN, 1998).

Assim, atualmente uma grande quantidade de informações é gerada, bem como é possível extrair informações importantes sobre um determinado assunto, e com isso decidir as diretrizes de um cenário, de modo mais acertado. A principal meta de técnicas de extração de conhecimento é a obtenção de conhecimento escondido a partir de uma realidade registrada em um determinado intervalo de tempo.

2.5.1 Estado da arte de extração de conhecimento de mapas auto-organizáveis a partir de regras difusas

Segundo Silva (2013), as redes neurais seriam mais populares se a capacidade de explicação dos seus resultados fosse parte integrante de sua funcionalidade. Nesta perspectiva,

diversas pesquisas têm sido desenvolvidas para atribuir interpretação às informações processadas.

A utilização dos sistemas difusos tem grande importância no campo de extração de conhecimento por ser um modelo capaz de explicar o comportamento de variáveis linguísticas. Entretanto, um modelo difuso não apresenta a capacidade de generalização, descrevendo apenas o que lhe foi delimitado; o modelo também apresenta total dependência de um especialista para a descrição de suas regras. A partir destas limitações, a atribuição de um modelo neural apresenta-se como uma alternativa eficaz que une a capacidade de aprendizagem a partir de dados das redes neurais, com a capacidade de explanação do conhecimento da lógica fuzzy.

A seguir, são apresentados importantes trabalhos já desenvolvidos com abordagem em extração de conhecimento de mapas auto-organizáveis através de regras difusas.

- No trabalho realizado por Alatrística-Salas et al. (2014), apresenta-se um processo de extração de conhecimento aplicado a dados hidrológicos. Para tanto, foram combinados métodos de redes neurais e aprendizagem de máquina para extrair conhecimento em dados coletados em estações localizadas ao longo de vários rios. Em primeiro lugar, os dados foram pré-processados, a fim de obter diferentes aproximações espaciais. Mais tarde, foram aplicados a um algoritmo para extrair padrões sequenciais. Finalmente propõe-se uma combinação de duas técnicas: uma para realizar uma filtragem com base em padrões de medida de interesse e outra para apresentá-los graficamente, no intuito de ajudar aos especialistas. Tais elementos podem ser utilizados para analisar os indicadores especializados a fim de ajudar a interpretação dos dados ecológicos e de monitoramento do rio.

- O trabalho de Santos et al. (2005) apresenta um quadro de apoio a decisão da seleção de parâmetros bioquímicos que foram considerados em sensoriamento remoto de águas contidas em barragens. A estrutura permite a comparação da eficácia de dois tipos de modelos, utilizando o método de árvores de decisão. O primeiro usa todos os indicadores de qualidade da água, incluindo as variáveis de tempo e consumo, enquanto o segundo modelo é baseado apenas em parâmetros remotos adquiridos em tempo real. Ao comparar as duas estratégias sob diversos critérios (por exemplo, custo, tempo e confiança), este último método mostrou-se a melhor alternativa.

- Waters et al. (2003) analisam os efeitos químicos em sistemas biológicos (CEBS) utilizando uma base de conhecimento, que é projetada para atender às necessidades sobre informações de "sistemas toxicológicos", envolvendo o estudo da perturbação por produtos químicos na água e em seus ecossistemas, o acompanhamento da evolução molecular e parâmetros toxicológicos convencionais, e a integração de dados de resposta biológica para

descrever o funcionamento do organismo. A base de conhecimento foi sendo desenvolvido a partir de informações de múltiplos bancos de dados de domínio público, cujos indicadores podem ser a assimilação toxicológica, biológica e química da proteína. Com conteúdo de dados de alta qualidade, CEBS acabará por tornar-se um recurso para apoio de tomada de decisão conduzindo a pesquisa que contribui efetivamente para a segurança de diversas substâncias e para a melhoria das avaliações de risco de produtos químicos no ambiente.

- Mounce (2012) demonstra como mapas SOM podem ser usados para a mineração de fontes de dados microbiológicos a partir de sistemas urbanos de água. Dados de diferentes da composição microbiológica dos sistemas de drenagem de diferentes partes do mundo foram obtidas a partir de amostras e pós processadas pela ferramenta de software T-Align antes da redução de dimensionalidade com Análise de Componentes Principais (PCA). Estes conjuntos de dados foram então analisados através de redes SOM e características adicionais foram utilizadas na rotulagem do mapa. Os resultados iniciais mostram que a produção visual da análise SOM fornece um meio rápido com intuito de explorar hipóteses para o aumento da compreensão e interpretação de ecologia microbiana.

- Body et al. (2002) apresentam uma nova categoria de ferramentas de apoio à decisão que se baseia nos Sistemas de Informação Geográfica de hoje (GIS) e tecnologias de Processamento Analítico On-Line (OLAP) para facilitar a extração de conhecimento geográfico (GKD). Esta nova categoria, denominada Spatial OLAP (solap), tem sido um tema de pesquisa nos últimos 5 anos em poucos laboratórios da universidade e agora está sendo implementado por pioneiros em diferentes áreas, incluindo a saúde pública na qual oferece inúmeras vantagens. Neste artigo, apresenta-se um exemplo de uma aplicação SOLAP no campo da saúde ambiental: o projeto ICEM-SE. Após ter apresentado este exemplo, descreve-se a concepção deste sistema e explica-se como ele fornece acesso fácil e rápido aos dados pormenorizados e agregados, que são necessários para a tomada de decisões GKD e na saúde pública. O conceito SOLAP também é descrito e é feita uma comparação com aplicações GIS tradicionais.

- Em Dzeroski et al. (1997), foi descrita uma metodologia particular de extração de conhecimento de dados em vários domínios ecológicos. Estes incluem a classificação biológica dos rios britânicos, baseados em dados de bioindicadores, a análise da influência dos parâmetros físicos e químicos em organismos bioindicadores selecionados em rios eslovenos e a classificação biológica dos rios eslovenos com base nos parâmetros físicos e químicos, bem como dados bioindicadores. Em todos os três casos, os modelos de extração de conhecimento na forma de regras foram extraídos de dados obtidos através de monitoramento ambiental e /

ou interpretação técnica das amostras adquiridas. Foram verificadas evidências positivas para a utilização da aprendizagem de máquina em modelação ecológica.

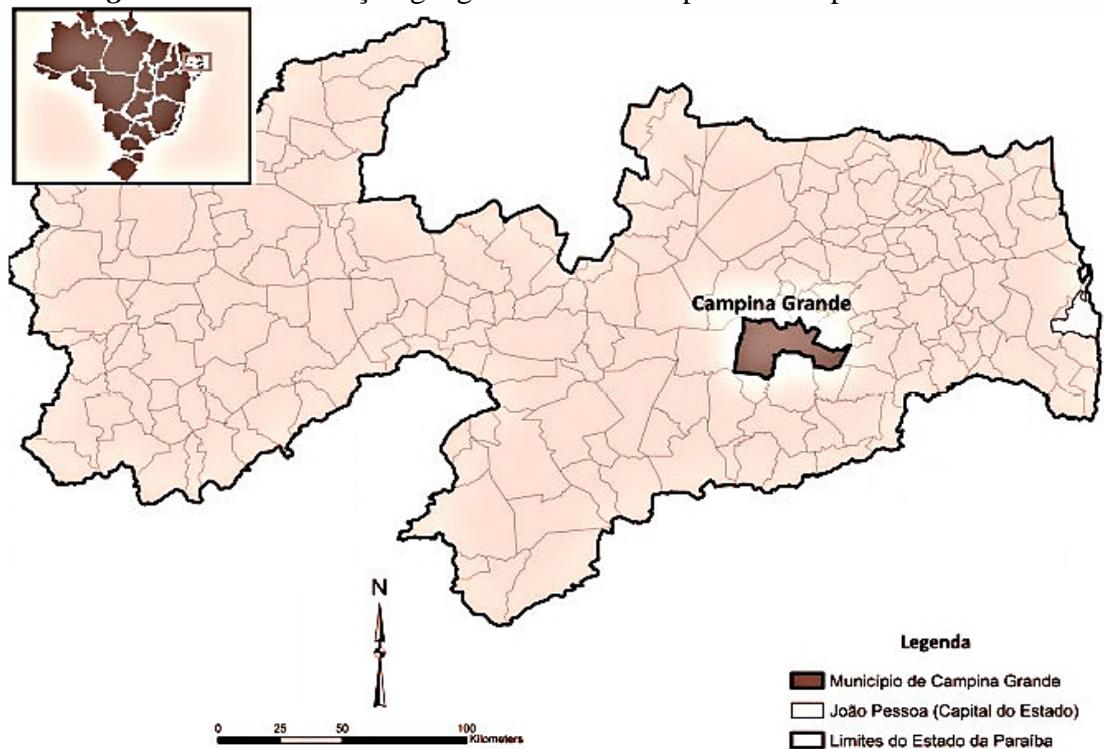
- Gur - Ali et al. (1993) utilizaram uma abordagem híbrida, pelo método Probabilístico Indutivo de Aprendizagem (Pril), que fornece uma medida probabilística de eficácia para uma regra individual, permitindo que o usuário defina os níveis de classificação incorreta máximos, ou mínimos de confiabilidade, com confiança pré-determinado o qual cada regra irá satisfazer este critério. O usuário é capaz de quantificar a confiabilidade do processo de decisão, ou seja, o cumprimento das regras, que é de importância crucial nos processos de decisão automatizadas. O Pril é descrito com o seu algoritmo associado. Também é apresentado um exemplo ilustrativo com base no processo de reivindicações e de compensação dos trabalhadores.

MATERIAIS E MÉTODOS

3.1 Descrição do sistema

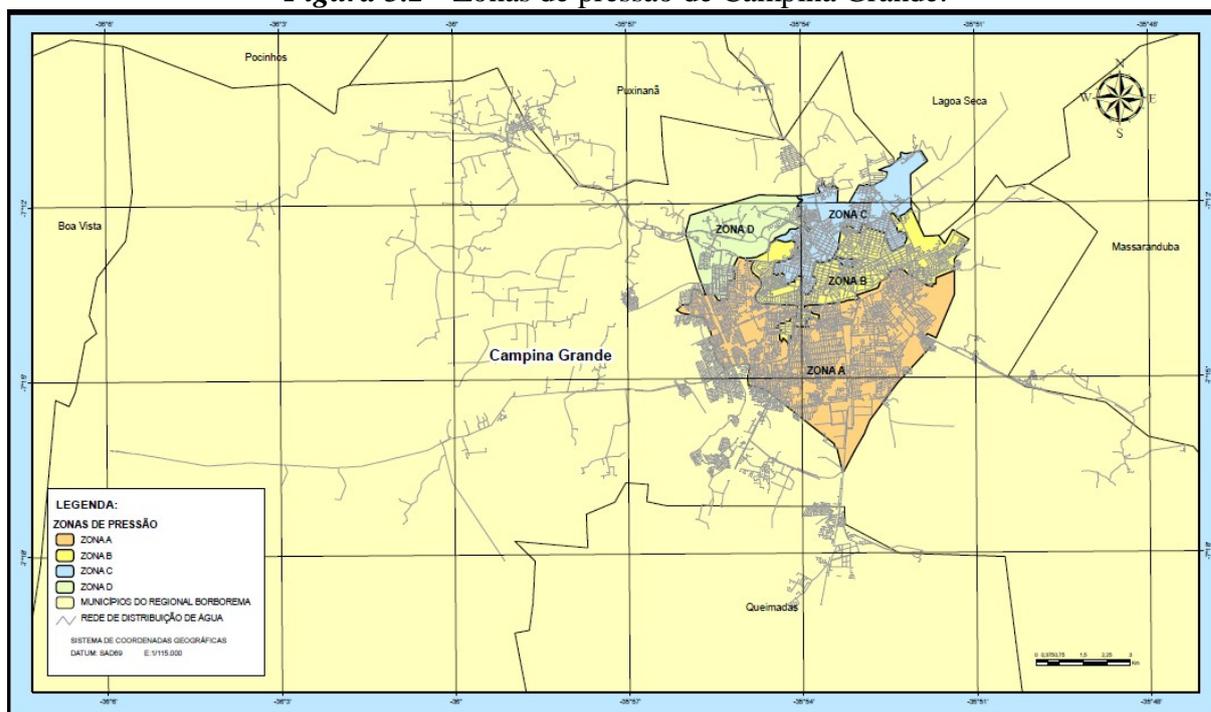
A área de estudo compreende o sistema de abastecimento de água da cidade de Campina Grande, no estado da Paraíba. A cidade de Campina Grande é a sede do município homônimo, localizado na mesorregião do Agreste Paraibano, na parte oriental do Planalto da Borborema, a 7°13'11" de latitude Sul e a 35°52'31" de latitude Oeste (Figura 3.1). O município de Campina Grande apresenta uma área de aproximadamente 621 km², com uma zona urbana de quase 96 km², em seus limites.

Figura 3.1 - Localização geográfica do município de Campina Grande.



Fonte: *Próprio Autor.*

O sistema de abastecimento é operado pela Companhia de Água e Esgotos da Paraíba (CAGEPA), empresa com atribuições para captar, tratar e distribuir água ao consumidor, sendo que, devido as variações de altitude, a rede de distribuição foi dividida em quatro zonas de pressão A, B, C e D (Figura 3.2).

Figura 3.2 - Zonas de pressão de Campina Grande.

Fonte: CAGEPA, 2011.

Os bairros inseridos em cada zona de pressão são listados no Quadro 3.1. O R-5 é um reservatório do tipo semi-enterrado com capacidade de armazenamento de 8.000 m³, localizado no bairro da Prata, cota 567 m. Sua função é a distribuição de água, bem como de poço de sucção, e é alimentado pela adutora de diâmetro nominal (DN) 700 mm.

Quadro 3.1 - Bairros inseridos em suas respectivas zonas de pressão.

Zonas	Principais Bairros
A e D	Quarenta, Liberdade, Cruzeiro, Santa Rosa, Jardim Paulistano, Catolé, Mirante, José Pinheiro, Malvinas, Presidente Médice, Distrito Industrial, Jeremias, Bodocongó, São Januário, Severino Cabral, Lagoa de Dentro, São José da Mata e a cidade de Pocinhos.
B e C	Centro, Santo Antônio, São José, Prata, Centenário, Alto Branco, Juracy Palhano, Nações, Cuités, Palmeira, Prata, Monte Santo, Conjunto dos Professores, Distrito de Jenipapo e as cidades de Lagoa Seca, Alagoa Nova, São Sebastião de Lagoa de Roça e Matinhas.

Fonte: LEAL, 2012.

A rede de distribuição é alimentada por um conjunto de três adutoras de água tratada as quais se estendem ao longo de um percurso um pouco superior a 20 km. Tal sistema de distribuição compreende um número significativo de reservatórios de distribuição, sendo os principais o R-9, e o R5, situado na cota tal e tal. São reservatórios importantes porque recebem água tratada, aduzida da Estação de Tratamento, e a distribuem diretamente para a rede e para outros reservatórios. No Quadro 3.2 encontra-se um quadro resumo do conjunto de reservatórios do sistema de abastecimento de Campina Grande, com base no estudo de Meneses (2011).

Quadro 3.2 - Quadro resumo do sistema de reservação de Campina Grande.

Unidade	Descrição e condições de operação
R-1	É abastecido a partir da rede de distribuição que abastece o bairro do Alto Branco. Serve de poço de sucção da elevatória EEAT-11 que atende ao R-07, bairro das Nações. Seu nível mínimo operacional é de 1,60 m e o máximo é de 2,50 m. Há leitura de nível.
R-2	É abastecido por subadutora a partir do R-5. Seus níveis operacionais máximo e mínimo são 2,50 m e 0,50 m. Há leitura de nível.
R-4	É abastecido pela EEAT-8 (R-5). Serve de poço de sucção para a EEAT-09, que atende ao sistema do Brejo e não funciona no horário de ponta. Há o controle de pressão para a rede de distribuição, cujo valor máximo na saída do reservatório é de 3,0 m.c.a. O nível mínimo para atendimento é de 2,5 m e o máximo de operação é de 5,50 m. Há leitura de nível.
R-5	É abastecido diretamente da adutora DN 700, cujo controle de vazão é realizado no R-9. Serve de poço de sucção para a EEAT-8 e para parte do Centro. Atende também por gravidade, à parte do Bairro de Bodocongó e do Centro, cujo controle de vazão de distribuição é realizado por válvulas borboletas. O nível mínimo de operação é de 1,30 m. Já o máximo é de 3,60 m. Há leitura de nível.
R-6	É abastecido através da adutora de DN 700. Seu nível é conhecido através da medida de pressão. O nível máximo é 35,50 m.c.a e o mínimo de 15,00 m.c.a.
R-7	Esse reservatório é abastecido através da EEAT-8. O nível mínimo de operação é de 2,00 m e o máximo é de 4,00 m.
R-9	É o principal reservatório. É composto por 4 câmaras, sendo atendido pelas adutoras de DN 500, DN 700 e DN 800. Porém, a adutora de DN 700 abastece principalmente o R-5 e a entrada de água no R-9 é por duas válvulas borboletas que regulam a entrada de água para as câmaras 2 e 3. O nível mínimo de abastecimento é de 2,5 m e o máximo operacional é de 5,10 m. Após as 21:00 h, término do horário de ponta, inicia-se o acionamento dos conjuntos na EEAT-Gravatá. Caso o R-9 esteja com nível de 3,0 m faz-se o seguinte acionamento: 1 bomba na adutora de DN 700 e 1 bomba na adutora de DN 800. Às 0:00 h, se o nível do R-9 estiver em 3,50 m, aciona-se o 2º conjunto na adutora de DN 800.
R-10	Reservatório alimentado pela EE-3. Não há leitura de nível. O bombeamento para esta unidade é de aproximadamente 10 h/dia, normalmente das 0:00 h às 10 h e das 13 h às 17:00 h.
R-11	Também é acionado pela EE-3 e não dispõe de escala de nível. A alimentação se dá pela mesma adutora que atende o R-12.
R-12	É alimentado pela EE-3. O nível máximo é de 4,00 m e o mínimo é de 2,00 m.
R-13	Esse reservatório é abastecido por até 3 conjuntos da EEAT-3. O nível do reservatório é conhecido através da medida de pressão. O valor mínimo operacional é de 16,00 m.c.a e o máximo é de 23,00 m.c.a.
R-14	Alimentado pela EE-4. O nível mínimo para o abastecimento é de 3,0 metros. Já o nível máximo operacional é de 4,20 m.
R-18	É abastecido pela EEAT-7, abastecendo o bairro do Mutirão. O nível também é conhecido através da medição de pressão. O valor para atendimento é de 5,00 m.c.a. e o máximo, 13,50 m.c.a.
R-29	Alimentado a partir da EEAT-9. O nível também é conhecido através da medição de pressão. O valor máximo para atendimento é 21,00 m.c.a.

Fonte: *Meneses, 2011.*

O R-0, conforme representado esquematicamente na Figura 3.3 (ALBUQUERQUE, 2007), é o reservatório inicial, localizado na estação de tratamento, em Gravatá (7°23'7.4" Sul e 35°58'41" Oeste), município de Queimadas, distante 21 km de Campina Grande. A estação de tratamento é do tipo convencional, com capacidade para tratar 1,5 m³/s, composta por calha Parshall, floculadores mecânicos, decantadores de fluxo horizontal, filtros de gravidade, tanque de contato de cloro e casa de química completa (MENEZES, 2011).

Ao chegar a estação de tratamento a água bruta passa por um tanque de tranquilização, cujo objetivo é reduzir a velocidade. Em seguida, a água recebe a adição do coagulante sulfato de alumínio e da cal na calha Parshall. A água passa por 24 floculadores mecânicos de eixo vertical tipo paletas, com volume total de 1.341,20 m³.

A água floculada percorre decantadores do tipo convencional com escoamento horizontal, com área total de 6.730 m². A água segue para a filtração, que é realizada por 11 filtros rápidos de seção quadrada com meio filtrante de dupla camada, com uma área total de 475 m² e taxa de filtração máxima de 300 m³/m²/dia (ALBUQUERQUE, 2007; MENEZES, 2011).

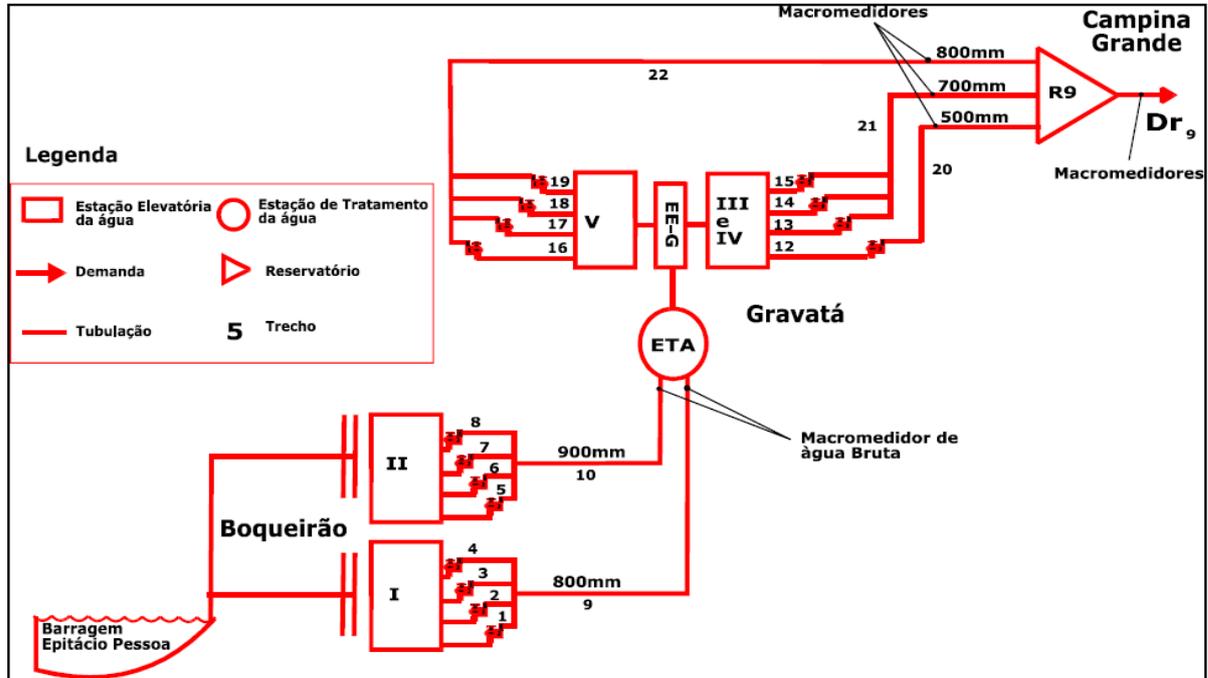
Por fim, a água é submetida à etapa de desinfecção, utilizando cloro gasoso. Após a adição do cloro, a água passa pelas câmaras de contato de volume total de 4.000 m³. O objetivo deste processo é garantir a eliminação dos microrganismos patogênicos e assegurar uma concentração residual que atenda ao padrão de potabilidade e garanta a segurança da água para consumo humano. (LEAL, 2012).

A água tratada é recalçada até a cidade de Campina Grande por meio de três estações elevatórias em três linhas de adução. Da primeira elevatória partem duas adutoras: a DN 500 mm, a mais antiga em operação, desde 1957, que abastece o R-9 e a DN 700 mm, implantada em 1972, que abastece o R-5, localizado na cota 567 m e possui uma derivação para o R-9. Da segunda elevatória parte a adutora de DN 800 mm, construída em 1994. Outros 30 reservatórios espalhados por toda a área da cidade são atendidos por estas canalizações, estando três deles (R8, R16 e o R21) desativados de acordo com informação da empresa (CAGEPA, 2009).

Ao longo do percurso até os reservatórios principais (R-9 e R-5), as adutoras sofrem derivações. A adutora de DN 500 mm sofre derivação antes do R-9, a fim de atender aos bairros da Catingueira, das Cidades, Catolé de Zé Ferreira e Conjunto Nenzinha Cunha Lima. A adutora DN 700 mm possui derivação para atender ao Distrito Industrial de Campina Grande e ao Distrito de Galante. A adutora de DN 800 mm abastece o Distrito Industrial de Queimadas, o bairro do Ligeiro e as indústrias ao longo da Alça Sudoeste (Figura 3.3). O R-9 é um reservatório do tipo apoiado localizado no bairro de Santa Rosa, cota 547m, com a capacidade

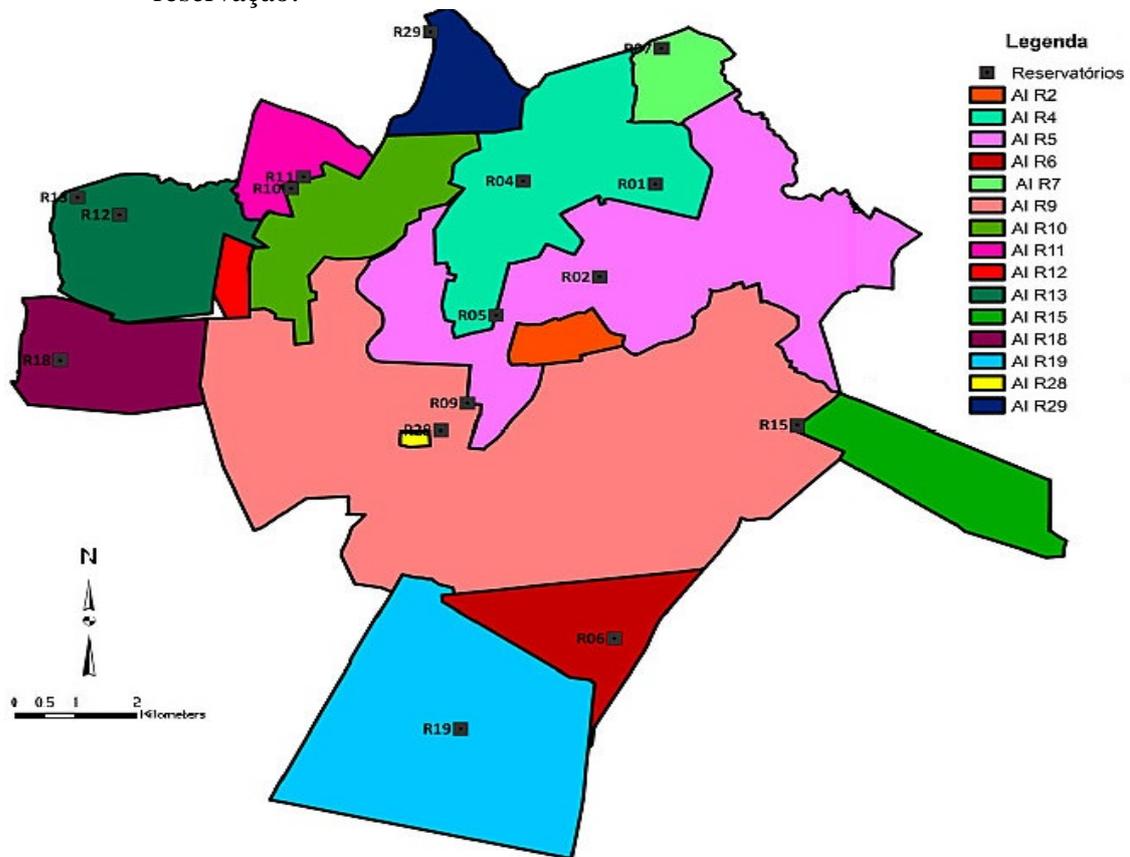
de armazenamento de 29.000 m³. Ele é responsável pelo abastecimento das zonas de pressão A e D (Figura 3.4). Já as zonas de pressão B e C são alimentadas pelo reservatório R-5.

Figura 3.3 - Representação esquemática do sistema de abastecimento de água de Campina Grande, entre a captação e o reservatório R9.



Fonte: Albuquerque, 2007.

Figura 3.4 - Rede de distribuição de água de Campina Grande: localização dos reservatórios de distribuição de água potável e área de influência (AI) das unidades de reservação.



Fonte: Próprio Autor.

De acordo com dados fornecidos pela Companhia, o controle da operação é centralizado - CCO. As decisões operacionais são tomadas a partir do CCO, mediante o recebimento de informações de cada unidade operacional em que há operador. O contato do CCO com as unidades é realizado por meio de rádio ou telefone.

Todas as interferências nas unidades são realizadas de forma manual pelos operadores: acionamento e desligamento de bombas, abertura e fechamento de válvulas.

As válvulas de controle na rede de distribuição são manuseadas pelas equipes de manobristas (disponíveis 24 horas/dia) que são sempre acionadas pelo CCO. Na maioria das vezes, as ocorrências são devido a vazamentos e rompimentos das tubulações. Há sempre equipes de plantão nos fins de semana, feriados e no período noturno. Durante a noite, a equipe plantonista atua principalmente na retirada de vazamentos em ramais prediais; caso haja vazamentos em redes, as válvulas são fechadas e os serviços são realizados na manhã seguinte.

3.2 Metodologias de aquisição de variáveis

Desde o ano de 2007, o Grupo de Pesquisa Saúde Ambiental, que reúne professores e estudantes das Universidades Estadual da Paraíba e Federal de Campina Grande, desenvolveu trabalho de monitoração de vigilância da qualidade da água do sistema de abastecimento de Campina Grande. Trabalhos anteriores (SANTOS, 2011; LEAL, 2012, VIEIRA, 2012) analisaram os resultados com a utilização de métodos convencionais (modelagem matemática e estatística), usando sistemas de informação geográfica (SANTOS, 2011) e de análise de risco (VIEIRA, 2012).

A escolha dos pontos de monitoramento seguiu critérios de distribuição geográfica e locais estratégicos para identificação de áreas vulneráveis e que abasteçam um maior número de pessoas. Desse modo, as amostras de água foram coletadas em pontos da rede de distribuição em conformidade com a “Diretriz nacional do plano de amostragem da vigilância em saúde ambiental relacionada à qualidade da água para consumo humano” para representar uma melhor amostragem adotando, normalmente, escolas públicas como pontos de coleta de amostras.

Para a construção do modelo neuro-fuzzy foram utilizados dados referentes ao período de coleta de março a outubro de 2011, para as variáveis Cloro Residual Livre, Monocloramina, Dicloramina, Tricloramina, pH, Turbidez e Bactérias Heterotróficas Facultativas.

Os dados foram coletados nos reservatórios R-9, R-5 e R-0, nas adutoras de DN 500, 700 e 800 mm, bem como em locais de alta vulnerabilidade como os apresentados na Tabela 3.2, perfazendo um total de 241 dados coletados para cada uma das variáveis analisadas. As coletas foram realizadas manualmente de uma torneira diretamente ligada ao sistema público de distribuição de água, após uma descarga de dois minutos, e as amostras mantidas em um frasco PET recoberto por uma capa escura e com capacidade de 1 litro, de acordo com as recomendações de APHA, AWWA, WPCF (1999). As amostras foram analisadas em triplicata e as variáveis cloro residual (livre e combinado), pH e turbidez foram determinadas *in loco* devido à instabilidade do cloro e da inadequação de qualquer procedimento de preservação das amostras coletadas. Os métodos analíticos utilizados na pesquisa seguiram os procedimentos padrões descritos no Standard Methods for the Examination of Water and Wastewater (APHA, AWWA, WPCF, 1999).

Tabela 3.2 - Pontos de coleta e monitoramento no sistema de abastecimento.*Pontos de Coleta e Monitoramento*

	P1	P2	P4	P5	P6	P7	P8	P9
Localização	R0	AD ¹ 500	AD 700	AD 800	E.M. ² Melo Leitão	R-05	EEEFM ³ Solon de Lucena	EEEFM Clementino Procópio
	7°29'47.43" S 36°09'16.35 "O	7°14'4.76"S 35°54'19.95"O		7°13'59.13"S 35°53'57.25"O		7°13'9.84"S 35°53'41.19"O	7°12'56.22"S 35°52'54.18"O	7°13'30.26"S 35°53'25.96"O
Área de Abrangência	Gravatá, Queimadas e Caturité.	Santa Rosa, Quarenta, Cruzeiro, Santa Cruz, Dinamérica, Centenário e Bodocongó.		Santa Rosa, Quarenta, Cruzeiro, Santa Cruz, Dinamérica, Centenário e Bodocongó.		Prata, Bela Vista, Monte Santo e Centro	Centro, Alto Branco, Santo Antônio e Jardim Tavares	São José e Estação Velha.

¹ AD – Adutoras. As coletas em adutoras foram feitas no R9, como explica a Descrição do Sistema, na pág 47.

¹ EM - Escola Municipal

¹ EEEFM – Escola Estadual de Ensino Fundamental e Médio

3.2.1 Cloro residual livre

Para a determinação do cloro residual foi utilizado o método titulométrico DPD - SFA. Neste método as espécies de cloro residual são determinadas por análise volumétrica de oxidação-redução, utilizando o sulfato ferroso amoniacal (SFA) como titulante e o oxalato ou sulfato de N,N - dietil - p - fenilenediamina (DPD) como indicador. O cloro livre reage instantaneamente com DPD, na ausência de íons iodeto, com produção de coloração vermelha cuja intensidade está relacionada à quantidade de cloro livre disponível. O cloro livre pode ser imobilizado com glicina (ácido aminoacético) permitindo a determinação de frações interferentes de bromo e iodo sendo este resultado usado para subtrair do resultado de uma prova na qual o cloro livre não tenha sido imobilizado.

3.4.2 Cloro residual combinado

À mistura remanescente do teste de CRL é adicionada, primeiramente, uma pequena porção de cristais de iodeto de potássio (KI), seguida de titulação com solução padrão de sulfato ferroso amoniacal, para a determinação de monocloramina. Em seguida, é adicionada uma porção maior de cristais de KI e procedida a titulação com solução padrão de sulfato ferroso

amoniacal para a determinação de dicloramina. Tricloramina é determinada por titulação com solução padrão de SFA na presença de DPD e KI, numa alíquota de amostra tomada à parte.

3.2.3 pH

O pH foi determinado pelo método potenciométrico, com o auxílio do pHmetro portátil TECNOPON modelo PA210 P, calibrado com soluções tampões de pH 4 e 7. As leituras de pH foram feitas de forma direta em alíquotas das réplicas (três) de amostras coletadas.

3.4.4 Turbidez

A turbidez foi determinada pelo método nefelométrico com a utilização de um turbidímetro portátil provido de fonte de luz de tungstênio. Este método é baseado na comparação da intensidade de luz desviada pela amostra, com a intensidade da luz desviada por uma suspensão padrão de referência. A quantificação da turbidez é diretamente relacionada à intensidade da luz desviada.

3.4.5 Bactérias heterotróficas facultativas (BHF)

Para determinação das bactérias heterotróficas foi utilizado o método de plaqueamento em profundidade (“Pour Plate”), utilizando meio Agar Padrão para Contagem (PCA - Plate Count Agar) e os resultados foram expressos em Unidades Formadoras de Colônias por mililitro (UFC/mL). Todas as análises foram realizadas em triplicata.

3.3 Ferramentas utilizadas

Para a realização do presente trabalho foi utilizada uma biblioteca de funções *somtoolbox* desenvolvida pela Universidade de Helsinki, com apoio de Teuvo Kohonen, projetada para a linguagem avançada de programação MATLAB® (VESANTO et al., 2000). Em seguida, os pesos dos mapas auto-organizáveis foram exportados para o *toolbox* Fuzzy Logic, também do MATLAB®, cuja funcionalidade está em extrair as regras difusas baseadas na rede SOM.

Diversos cenários foram simulados, após o modelo ter sido validado. Esses cenários possuem o propósito de simular a atuação de um operador em momentos de conflito ou falhas no sistema de abastecimento, baseados nos indicadores selecionados, para que seja demonstrada a importância do modelo no processo de tomada de decisão.

3.4 Definição do problema

De acordo com estudos já realizados no âmbito do sistema de abastecimento de Campina Grande, o controle da operação é centralizado, as decisões operacionais são tomadas a partir do CCO, mediante o recebimento de informações de cada unidade operacional em que há operador. O contato do CCO com as unidades é realizado por meio de rádio ou telefone. Todas as interferências nas unidades são realizadas de forma manual pelos operadores: acionamento e desligamento de bombas, abertura e fechamento de válvulas. Portanto, o sistema não é automatizado.

Não há pontos de recloração nos reservatórios localizados na cidade e estes não passam por manutenção periódica com finalidade de limpeza e revisão de seus acessórios. O banco de dados do Serviço de Abastecimento é atualizado manualmente e não contém dados relativos à qualidade da água em todos os reservatórios de distribuição (ALBUQUERQUE, 2007).

Diante dos problemas percebidos através de consultas a estudos, relatórios, bem como ao acesso a dados da própria companhia, a construção de um modelo automático baseado em rede neural integrada com a lógica fuzzy, constitui-se em uma solução estratégica para gerenciamento do sistema de abastecimento, em termos de qualidade da água para consumo humano.

3.5 Construção da rede neuro-fuzzy

Após a coleta dos dados, foi feito o tratamento estatístico removendo os *outliers* dos conjuntos amostrais, através do método de Grubbs (SOKAL e ROHLF, 1981). Em seguida, o vetor de entrada foi apresentado ao algoritmo, de modo que a ativação inicial de cada unidade de saída foi estimada. Os dados foram normalizados com base em sua variância, a fim de que todos os vetores de entrada fossem reduzidos a uma mesma escala, evitando maior impacto de certos valores elevados, sobre as distâncias dos neurônios.

O modelo foi construído com 70 neurônios na camada oculta, cujos pesos foram divididos de tal forma que a soma destes sempre fosse 1. O número de neurônios foi definido com base em testes para a verificação do desempenho da rede no processo de aprendizagem. Este procedimento evita que um neurônio possa se tornar dominante e permanecer ativo em todo o tempo.

As variáveis selecionadas foram atribuídas como neurônios de entrada, e são: cloro residual livre, monocloramina, dicloramina, tricloramina, bactérias heterotróficas mesófilas, pH e turbidez. O procedimento estatístico adotado pela rede foi o erro topográfico, que quantifica a capacidade do mapa de representar a topologia dos dados de entrada. A função de aproximação foi a Gaussiana, por apresentar menor erro topográfico e erro de quantização média, que representa a média das distâncias entre cada vetor de dados e o correspondente vetor de pesos do neurônio vencedor (BMU). Outro modelo utilizando os pontos de coleta como variável foi construído, com o objetivo de verificar os pontos que influenciam mais sobre a qualidade da água, os indicadores que predominam nesses pontos e a correlação entre eles.

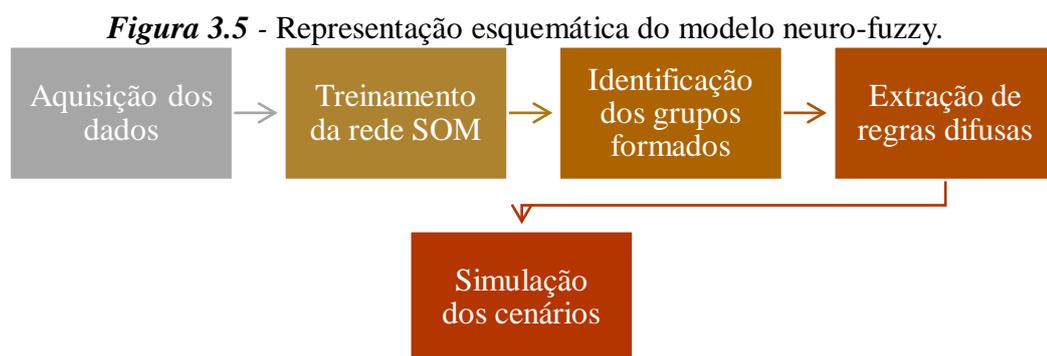
Para estas condições foram feitas 200 iterações, com uma matriz de 10x7 neurônios, utilizando o algoritmo de atualização dos pesos sinápticos dos neurônios em que, a cada vez que um item é apresentado à rede, ele sofre o processo de incremento, onde o modelo é postergado somente após a apresentação de todos os elementos (ULTSCH e SIEMON, 1990). A distribuição dos dados no modelo foi de 70% para o processo de treinamento e 30% para o processo de teste da rede.

Após treinamento do SOM e a identificação dos grupos, o processo para extração das regras difusas foi iniciado. O primeiro passo foi selecionar entre todos os neurônios, que pertencem a um grupo específico, aquele que tem o maior potencial para representar o centro do grupo. A seleção foi feita através do uso da Equação 2.2 que permite estimar o potencial de um neurônio. A partir das informações de cada grupo foram extraídas 40 regras. Foi possível extrair dos pesos dos neurônios 64 características do sistema de abastecimento de água de

Campina Grande. Essas características demonstraram também a relação existente entre algumas variáveis e o comportamento de outras. A conversão dos mapas em forma de valores numéricos (pesos) ocorreu com auxílio de comandos específicos da linguagem de programação usada. Esta metodologia foi baseada em um estudo sobre a aplicação de extração de conhecimento na detecção de falhas em transformadores elétricos, realizado por Silva (2013). A partir das regras extraídas, foi possível verificar sua conformidade com os valores obtidos no processo de monitoração. Através das regras extraídas do SOM, o especialista poderá verificar porque, para determinados valores de entrada dos indicadores, o sistema difuso apresenta como resposta determinado tipo de valor em termos de BHF, por exemplo, e com seu conhecimento pode validar ou não os resultados apresentados pelo sistema difuso. O uso de BHF como resposta se deve ao fato deste indicador representar um risco à segurança da qualidade da água.

As regras difusas auxiliaram também na percepção do comportamento das variáveis ao passo em que os valores de outras foram modificados, no sistema como um todo. Os cenários foram simulados com base na alteração das concentrações de CRL (cenário I) próximo ao limite mínimo recomendado pela Portaria vigente (0,2 mg/L) no ponto P3, alteração dos níveis de turbidez acima do recomendado pela Portaria vigente (5,0 UT) no ponto P0 e concentrações de CRL (cenário III) acima do recomendado pela Portaria vigente (2 mg/L embora seja admitido até 5 mg/L), verificando o que ocorre com as demais variáveis ao longo do sistema de abastecimento.

Enfim, pode-se resumir o modelo construído utilizando o organograma da Figura 3.5.



RESULTADOS

4

A Figura 4.1 apresenta gráficos de dispersão e histogramas de todas as variáveis analisadas para indicadores da qualidade da água. A dispersão dos dados originais está no triângulo superior, os valores do mapa auto-organizável sobre o triângulo inferior, e os histogramas na diagonal estão decodificados em preto para o conjunto de dados e em vermelho para os valores do mapa SOM.

A partir deste processo de visualização pode-se notar uma conformidade da rede SOM com os dados de treinamento, em que há pelo menos um grupo característico para cada par de variáveis analisadas. Cada variável exerce forte influência sobre a outra, o que reforça o uso de extração de regras difusas para a modelagem do sistema.

Para o processo de treinamento, foram obtidos os valores de erro topográfico e erro de quantização vetorial de 0,076 e 0,755, respectivamente. Isto significa que, tanto em relação ao processo de aprendizagem quanto em relação à representação dos mapas, a rede neural foi capaz de simular os dados de entrada de modo fidedigno ao cenário real. Ley et al. (2011) realizaram um estudo sobre a similaridade hidrológica de captações num rio do estado da Renânia Palatinado, Alemanha, para o período de 1993 a 2008, abrangendo uma enorme variabilidade das condições meteorológicas e de escoamento, através da aplicação de uma rede SOM. Neste modelo, Ley et al. (2011) obtiveram erro de quantização de 1,64, em uma matriz 10x7. Park et al. (2014) realizaram um estudo sobre o padrão de qualidade da água em reservatórios da Coreia, usando a rede SOM como classificador. No processo de treinamento, obtiveram respectivamente um erro de quantização e erro topográfico de 0,142 e 0,013, para uma matriz 4x8. Voutilainen et al. (2012), ao realizar um estudo correlacionando a presença de plânctons com a qualidade da água em rios da Finlândia, alcançaram erro topográfico de 0,129 e um erro de quantização vetorial de 0,339, em uma matriz neural de 8x8. O que se pode perceber é que o erro obtido neste trabalho se mantém em um intervalo adequado, para uma matriz 10x7, cujo esforço computacional é reduzido, como é possível perceber com a Figura 4.2, o que significa dizer que para cada variável modelada a rede terá um tempo de processamento diferente. Portanto, Quanto maior o número de clusters, maior será o tempo de processamento.

Em comparação aos trabalhos supracitados, o presente modelo apresentou erros topográficos e de quantização baixos em relação a topografia adotada.

Figura 4.1 – Dispersão e histogramas das variáveis analisadas.

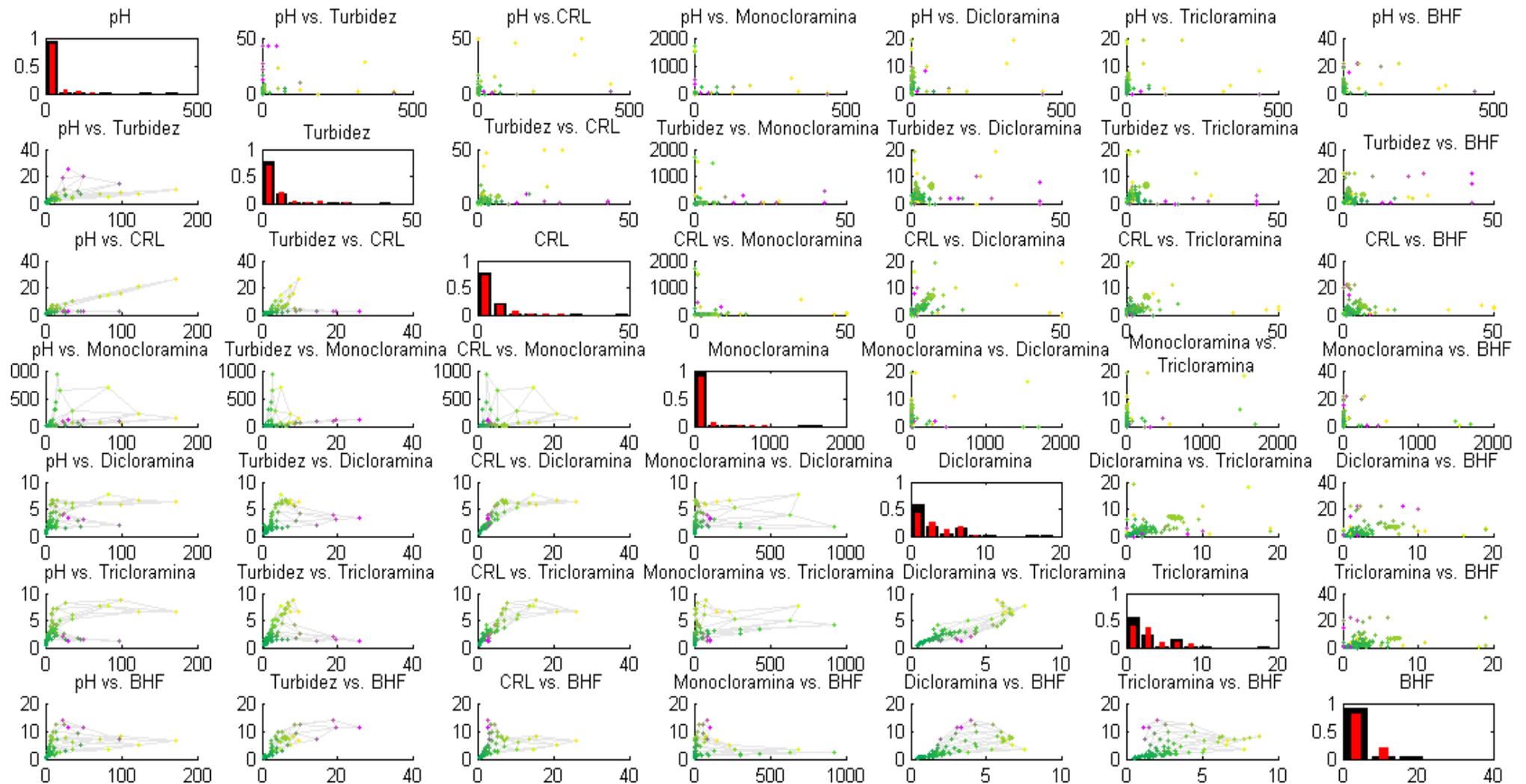


Figura 4.2 - Esforço computacional do processo de aprendizagem da rede neural.



A Figura 4.3 apresenta os mapas auto-organizáveis para os indicadores da qualidade da água.

Os pontos de monitoramento foram inseridos na rede neural, a fim de que informações também fossem extraídas através de regras difusas. Cada ponto de monitoramento foi interpretado como variável, e todos os indicadores foram inseridos como rótulos, cujo erro de quantização vetorial foi de 0,446. Este erro diz que o SOM preserva a topologia dos dados de entrada. Assim, foi possível verificar onde e como cada indicador se comporta na rede de abastecimento (Figura 4.4).

Figura 4.3 - Mapas auto-organizáveis dos indicadores de qualidade da água.

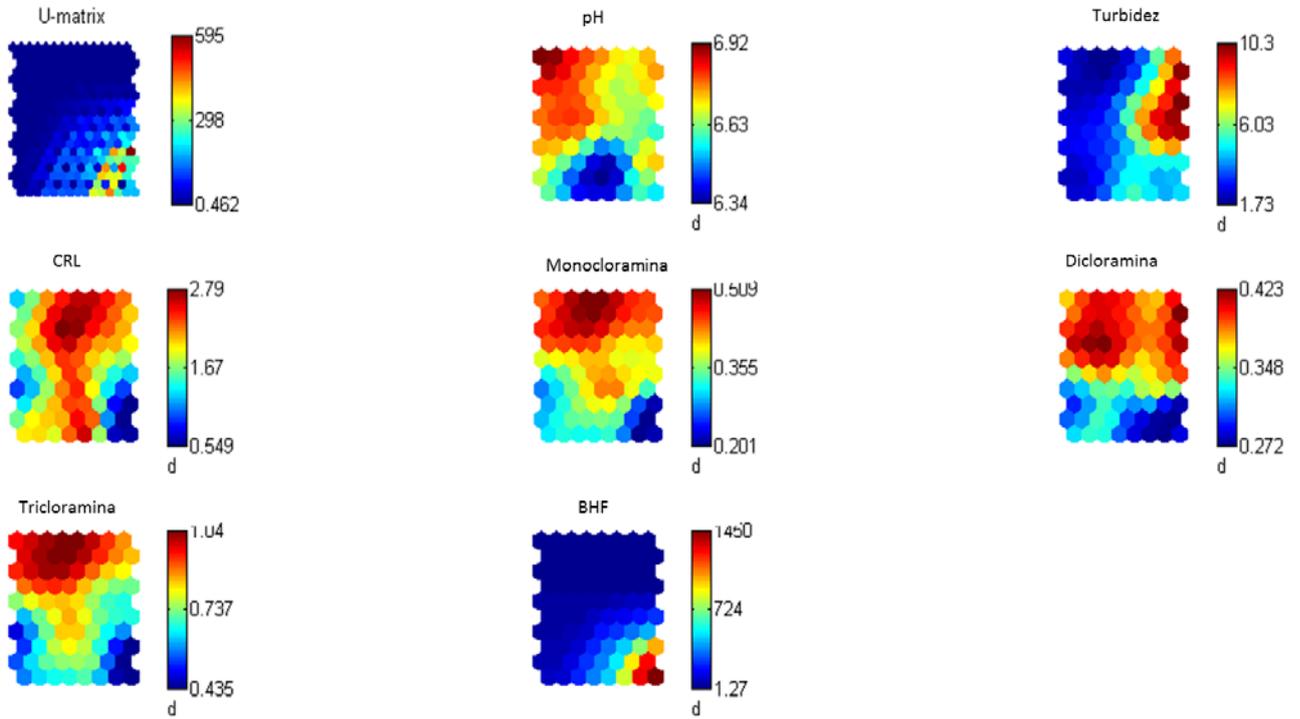
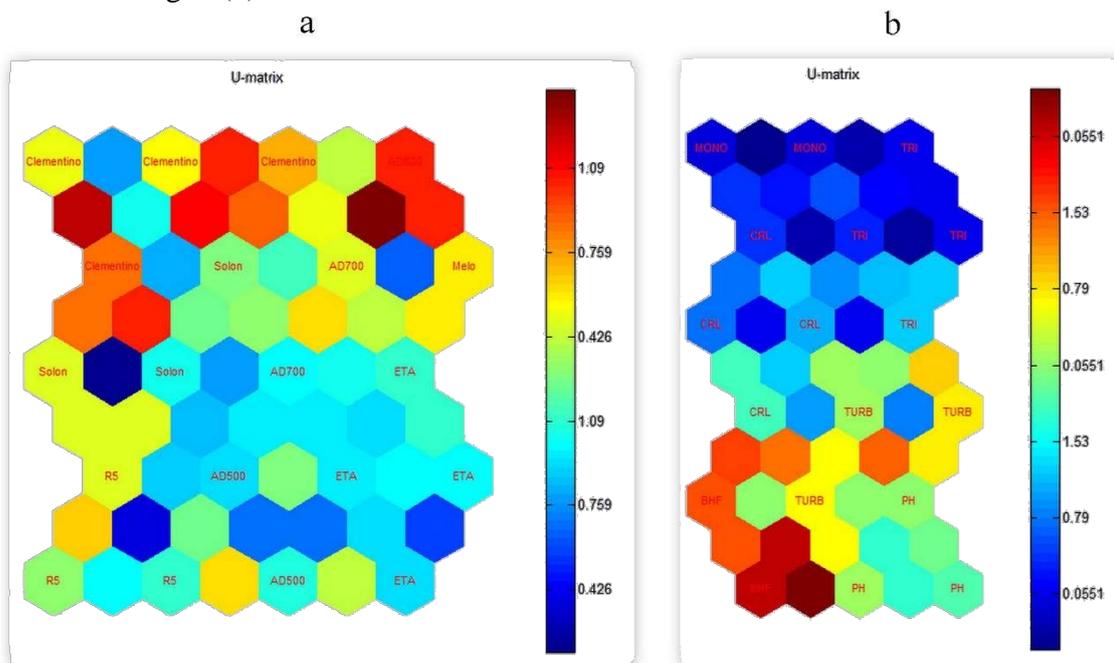


Figura 4.4 - Mapas BMU para de pontos de monitoramento (a) e indicadores da qualidade da água (b).



Para cada grupo processado (Figura 4.5), tem-se os valores dos neurônios com maior potencial (Tabela 4.3). Isso significa que foi possível verificar, através dos valores dos pesos dos neurônios de maior potencial, qual a influência de uma variável em relação a outra, e seu comportamento nos pontos de monitoramento. Esse processo resultou nas regras difusas, como as apresentadas abaixo:

- Se pH é 6,65 e Turbidez é 5,83 UT e CRL é 1,79 mg/L e Monocloramina é 0,365 mg/L e Dicloramina é 0,332 mg/L e Tricloramina é 0,763 mg/L, então BHF será 123 UFC/mL.
- Se pH é 6,62 e Turbidez é 6,74 UT, CRL é 1,98 mg/L e Monocloramina é 0,401 mg/L e Dicloramina é 0,332 mg/L e Tricloramina é 0,824 mg/L, então BHF será 393 UFC/mL.

Tabela 4.3 - Resultados para Extração de Regras.

	<i>Pesos dos Neurônios</i>						
	pH	Turbidez (UT)	CRL (mg/L)	Monocloramina (mg/L)	Dicloramina (mg/L)	Tricloramina (mg/L)	BHF (UFC/mL)
<i>Grupo 1</i>	6,47	6,01	2,32	0,32	0,24	0,66	337,42
<i>Grupo 2</i>	6,47	5,95	2,18	0,31	0,25	0,67	463,05
<i>Grupo 3</i>	6,50	4,69	2,49	0,33	0,26	0,75	491,79
<i>Grupo 4</i>	6,60	5,43	1,43	0,25	0,25	0,54	832,12
<i>Grupo 5</i>	6,63	5,18	1,60	0,31	0,29	0,62	84,56

Para as variáveis pH, Turbidez, cloro residual livre (CRL) e bactérias heterotróficas (BHF) também foram feitas as extrações de regras com base nos mapas SOM, para cada ponto. A escolha destas variáveis se deve à sua importância para a manutenção da qualidade da água do sistema de abastecimento, e por representarem indicadores de segurança da água.

Tabela 4.4 - Resultados para extração de regras em relação aos pontos de monitoramento.

	<i>Pesos dos Neurônios</i>								
	P1	P2	P3	P4	P5	P6	P7	P8	
<i>CRL (mg/L)</i>	1,87	3,32	1,39	2,25	2,00	2,25	1,60	0,13	<i>Grupo 1</i>
<i>pH</i>	6,78	6,58	7,42	7,14	6,76	6,78	6,62	6,67	<i>Grupo 2</i>
<i>Turbidez (UT)</i>	6,62	4,20	7,67	6,30	9,47	6,02	6,76	4,14	<i>Grupo 3</i>
<i>BHF</i>	77,00	4,00	1,00	1,00	1,00	2,00	0,00	42,00	<i>Grupo 4</i>
<i>Monocloramina (mg/L)</i>	1,85	3,63	1,84	2,13	1,61	1,20	0,06	0,05	<i>Grupo 3</i>
<i>Dicloramina (mg/L)</i>	1,10	2,71	1,66	1,66	1,57	1,16	0,08	0,02	<i>Grupo 4</i>
<i>Tricloramina (mg/L)</i>	0,20	0,72	0,72	0,48	0,72	1,00	0,02	0,03	<i>Grupo 5</i>

Os cenários foram construídos com base nos limites estabelecidos pela Portaria 2914/2011 do Ministério da Saúde, sendo os resultados são apresentados na Tabela 4.5. Mediante a suposição de que um operador esteja diante de quadros críticos em relação às concentrações de cada indicador estudado, nos devidos pontos de monitoramento, foram realizadas simulações com o propósito de verificar qual o comportamento de determinadas variáveis diante da alteração de uma delas, ao longo do percurso da água no sistema. Esta

avaliação ocorreu com base no exemplo a seguir: se o CRL ultrapassar o valor determinado na Portaria 2914/2011 MS em P1, quais serão os seus efeitos nos pontos sucessivos? Quais serão os valores dos outros indicadores? Estas perguntas foram feitas ao passo em que novos valores eram inseridos para cada variável na rede, originando a Tabela 4.5.

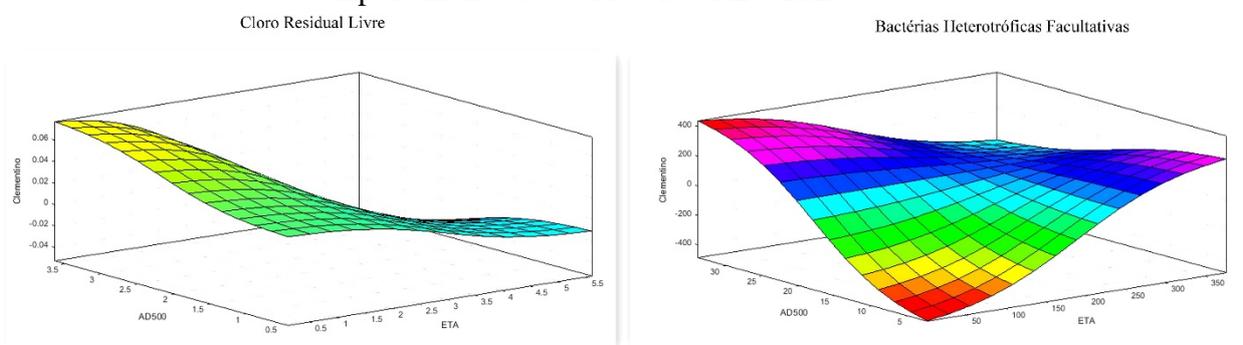
Para os indicadores de segurança da qualidade da água, CRL e BHF, foram feitas superfícies de resposta, que mostram o perfil desses indicadores nos diversos pontos monitorados.

Tabela 4.5 - Valores dos indicadores por pontos de monitoramento simulados pela rede neuro-fuzzy.

	pH	Turbidez (UT)	CRL (mg/L)	Monocloramina (mg/L)	Dicloramina (mg/L)	Tricloramina (mg/L)	BHF (UFC/ml)	Pontos
Cenário I	6,4755	6,0134	2,3282	0,32051	0,24884	0,66887	337,4238	P1
	6,6065	5,4355	1,5000	0,25263	0,25351	0,54244	832,1257	P2
	6,5563	4,6066	0,98591	0,22652	0,26413	0,49522	1232,7137	P3
	6,567	5,6388	1,8536	0,31431	0,2824	0,62862	275,658	P4
	6,5231	5,3302	2,1033	0,34479	0,30148	0,71708	377,3983	P5
	6,5026	3,6313	2,4363	0,36224	0,30624	0,77393	415,9645	P6
	6,6182	4,1735	1,3898	0,26471	0,26727	0,5409	765,3897	P7
	6,5091	5,2555	1,2657	0,26065	0,26731	0,51111	875,3201	P8
Cenário II	pH	Turbidez (UT)	CRL (mg/L)	Monocloramina (mg/L)	Dicloramina (mg/L)	Tricloramina (mg/L)	BHF (UFC/ml)	Pontos
	6,7651	9,3281	1,8721	0,42917	0,41455	0,79717	10,577	P1
	6,7373	3,638	1,7073	0,36937	0,38458	0,71445	26,8715	P2
	6,7334	4,697	1,6768	0,41166	0,39804	0,76037	20,3649	P3
	6,7734	6,082	1,817	0,43628	0,39226	0,84927	9,3558	P4
	6,7045	7,2944	2,0924	0,43923	0,39049	0,86156	3,5471	P5
	6,7256	8,3854	1,5	0,44322	0,40084	0,86074	2,1691	P6
	6,7455	3,0168	2,1002	0,39427	0,39785	0,81675	21,3994	P7
6,7579	3,0366	1,9995	0,43213	0,40514	0,87865	18,4408	P8	
Cenário III	pH	Turbidez (UT)	CRL (mg/L)	Monocloramina (mg/L)	Dicloramina (mg/L)	Tricloramina (mg/L)	BHF (UFC/ml)	Pontos
	6,8461	3,3945	1,9824	0,45625	0,40117	0,9371	11,2552	P1
	6,7175	3,4274	2,3991	0,47334	0,39387	0,97796	3,8771	P2
	6,7226	4,1613	2,4666	0,47948	0,38841	0,97816	1,5565	P3
	6,7185	3,1255	2,1902	0,42411	0,41176	0,91669	18,1881	P4
	6,8261	2,9895	2,1518	0,45514	0,40085	0,98054	12,9332	P5
	6,8045	2,383	2,4064	0,48327	0,40765	1,0192	5,7285	P6
	6,7347	2,4063	2,5889	0,49792	0,39918	1,0258	1,9977	P7
6,7759	2,3248	2,5875	0,50289	0,38936	1,0316	1,055	P8	

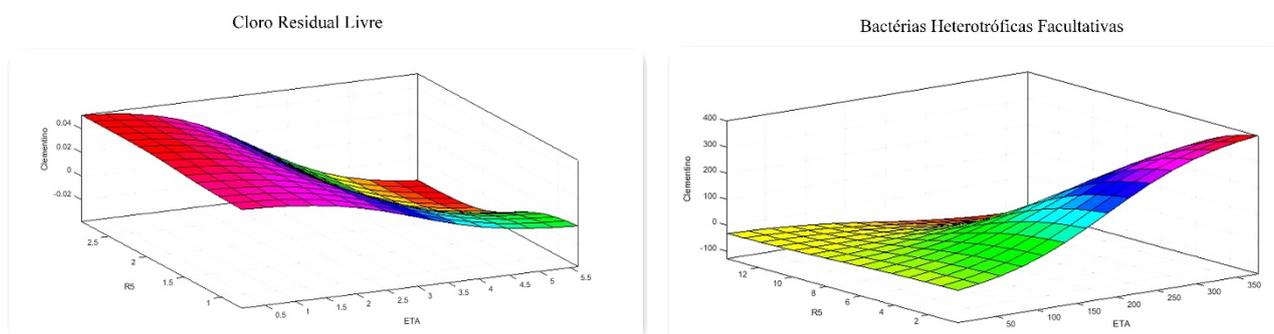
A Figura 4.5 representa a superfície de resposta da rede neuro-fuzzy para os pontos de monitoramento em relação aos indicadores de qualidade da água nos pontos P1, P2 e P8. Vale salientar que, para facilitar a visualização de todas as variáveis na superfície, os dados foram normalizados. A escolha dos pontos das superfícies de respostas foi feita com base nos resultados mais relevantes apresentados na Tabela 4.5. Estes dados foram interpretados com base em valores acima ou abaixo do recomendado pela Portaria 2014/2011 para as variáveis estudadas. Gradações de cor próximas ao vermelho representam concentrações mais elevadas, que tendem a níveis intermediários para cores próximas ao violeta e verde, e concentrações baixas em cores próximas ao azul escuro.

Figura 4.5 - Superfícies de resposta para os pontos de monitoramento P1, P2 e P8 relacionadas ao comportamento das variáveis CRL e BHF.



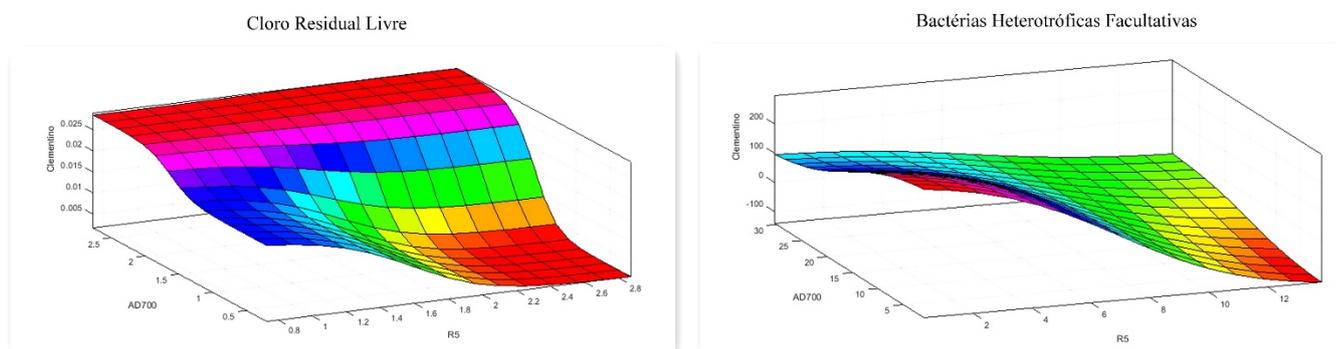
A Figura 4.6 representa a superfície de resposta da rede neuro-fuzzy para os pontos de monitoramento em relação aos indicadores de qualidade da água nos pontos P1, P6 e P8.

Figura 4.6 - Superfícies de resposta para os pontos de monitoramento P1, P6 e P8 relacionadas ao comportamento das variáveis CRL e BHF.



A Figura 4.7 representa as superfícies de resposta da rede neuro-fuzzy para os pontos de monitoramento em relação aos indicadores de qualidade da água nos pontos P3, P6 e P8.

Figura 4.7 - Superfícies de resposta para os pontos de monitoramento P3, P6 e P8 relacionadas ao comportamento das variáveis CRL e BHF.



DISCUSSÃO

5

Os histogramas e dispersões da Figura 4.1 mostram que a rede neural, a partir do processo de aprendizagem, reproduziu com grande aproximação as características da qualidade da água do sistema de abastecimento de Campina Grande. É possível observar a correlação existente entre CRL e pH, previsível quimicamente, já que os valores de pH acima do recomendável pela Portaria vigente resulta em decaimento das concentrações de ácido hipocloroso, que possui alto poder desinfetante. A elevação do pH pode ser explicada por falhas em algumas operações unitárias.

A dicloramina, que se caracteriza como bom desinfetante, predomina em pH baixo; a monoclорamina, desinfetante pouco ativo, predomina em pH acima de 7,5.

Grupos especiais foram detectados pela rede no sistema de abastecimento: em alguns pontos da cidade o pH e a dicloramina formam dois grupos distintos caracterizados por uma distância euclidiana maior e outra menor, evidenciando que, nestes locais, existem uma maior concentração de dicloramina à medida em que o pH se mantém baixo, garantindo a segurança da população abastecida nestas regiões, devido a seu potencial desinfetante; em contraposição, há agrupamento que mostra uma maior distância euclidiana, ou uma menor concentração em termos de dicloramina; nestas mesmas condições há um crescimento nas concentrações de tricloramina, que pode ser explicado pela presença de nitrogênio amoniacal e pela falta de manutenção da rede, podendo resultar em odor e sabor característicos e em irritação das mucosas; outro grupo identificado pela rede neural é em relação ao CRL e à turbidez, pois nos pontos P7 e P8 a concentração de CRL diminui na mesma proporção em que a turbidez aumenta, resultando não apenas na intervenção da aparência da água, mas também afetando no processo de desinfecção; é possível afirmar que no ponto P8 a desinfecção esteja ocorrendo com uma menor eficiência, como é possível perceber no agrupamento turbidez vs. BHF, que possui uma distância euclidiana menor, caracterizando elevados valores de turbidez à medida em que a concentração de bactérias se eleva; à medida em que o pH se eleva, a concentração de bactérias se eleva também, já que nesta mesma região há uma elevada concentração de CRC, cujo poder desinfetante é menor. Este cenário é perceptível nos pontos P3, P6, P7 e P9 podendo ser explicado por falhas nos processos operacionais na ETA. O CRL é apontado pela rede neural como um cluster de maior distância euclidiana, sugerindo a interpretação de que nos pontos P3, P6, P7 e P9 sua concentração é baixa, podendo ser abaixo do recomendado pela Portaria. A falta de manutenção na rede e até mesmo nos reservatórios de distribuição explica esse cenário.

5.2 Mapas de Kohonen

Os mapas auto-organizáveis apresentados na Figura 4.2 reafirmam os resultados exibidos nas dispersões, com destaque para o CRL que possui uma gradação próxima aos tons de amarelo e vermelho, significando uma distância euclidiana maior, ou seja, uma menor concentração da variável na rede de distribuição. Este perfil se assemelha ao mapa de dicloramina, que também possui um decaimento visível: a presença de microrganismos, aliada à presença de nitrogênio amoniacal, resultantes da falta de manutenção bem como falhas no processo de dosagem de cloro na ETA são fatores que levam a estes resultados.

Os mapas mostram ainda a formação de três grupos para bactérias: um apresenta características de baixas concentrações, outro de concentrações intermediárias e o último de concentrações elevadas que implicam violação da Portaria 2914/2011. Estas características contrapõem os perfis de CRL e dicloramina, e se assemelham ao perfil de turbidez e pH, que sugere que ao passo em que a água é distribuída, o CRL sofre um rápido decaimento pelo aumento do pH. Parte do CRL é transformado em CRC; entretanto, das frações de CRC, a que possui menor concentração é a dicloramina, que rapidamente é oxidada para tricloramina. Ainda sobre o CRL, parte de seu potencial é inibido pelos altos valores de turbidez, aumentando a concentração de BHF.

Para entender o comportamento dos indicadores, particularmente na sua relação com os pontos de monitoração da cidade, a Figura 4.3 apresenta um mapa BMU rotulado para as duas análises citadas anteriormente. O CRL é presente em pontos como a ETA, AD 500 e AD 700; como já foi descrito, as adutoras supracitadas se encontram no R-9, cuja importância é distinta para o sistema de abastecimento da cidade. Este mesmo reservatório é apontado pela rede neural como um ponto onde as concentrações de monocloramina e tricloramina são mais elevadas. Isso é justificado pelo R-9 ser considerado um dos reservatórios mais antigos da cidade, e por não passar por manutenções periódicas. É presumível, portanto, a presença de nitrogênio amoniacal no R-9 e nas suas tubulações, que é oxidada até tricloramina, variável apontada em maior concentração pela rede neural. A matéria nitrogenada é decorrente da presença de resíduos de origem animal ou vegetal, que podem existir na rede de abastecimento através do processo de carreamento dos mananciais até as tubulações e que persistem na falta de manutenção e limpeza de reservatórios de distribuição.

5.3 Simulação de cenários e tomada de decisão

Após a formulação das regras difusas através dos mapas auto-organizáveis, o Cenário I, apresentado na Tabela 4.5 não sofreu alterações, o que significa dizer os valores obtidos da lógica Fuzzy foram diretamente apresentados na sua forma numérica. Cada uma das análises realizadas através dos mapas auto-organizáveis se tornam claras a partir de sua apresentação numérica, bem como na forma de superfícies de resposta. O Cenário II apresenta modificações no R-5 para que o valor mínimo em termos de CRL, exigido pela Portaria 2914/2011 MS, seja mantido. O Cenário III é caracterizado com uma possível manutenção nos reservatórios e também em manutenção dos trechos. Esta simulação foi feita realizando uma diminuição nos valores de variáveis como turbidez e CRC que representam indicadores específicos de operação do sistema.

É possível perceber que o ponto mais crítico da cidade é o P8, sendo caracterizado por estar numa área cuja rede é a mais antiga da cidade, ainda formada por cimento amianto. Apesar disso, no Cenário II, com a aplicação de um ponto de recloração no R-5, há garantias de que a água permanecerá com um mínimo de 0,2 mg/L de CRL. Modificações em termos do acréscimo de CRL nos pontos adjacentes ao P1 não mostraram importância, ultrapassando inclusive, o que preconiza a Portaria do Ministério da Saúde sobre o limite máximo recomendado de CRL de 2,0 mg/L em qualquer ponto da rede. As adutoras possuem concentrações diferenciadas de bactérias, especialmente as adutoras de DN 700 mm e DN 800 mm, pois no horário de ponta (adotado pela concessionária como sendo das 17h às 21h), ocorre a paralisação de alguns conjuntos moto-bomba. Durante o horário de ponta, no sistema de adução de água bruta, dos cinco conjuntos disponíveis (três conjuntos na DN800 mm e dois na DN900 mm) apenas os correspondentes à adutora de DN900 mm funcionam e no sistema de adução de água tratada só o conjunto da adutora de DN 500 mm e um conjunto da adutora de DN 700 mm. Após o horário, o funcionamento do sistema adutor de água tratada ocorre pelo acionamento paralelo de duas bombas para a adutora de DN 700 mm, duas também para a adutora de DN 800 mm e o conjunto da adutora de DN500 mm (LEAL, 2011). O desligamento destes conjuntos durante o horário de ponta resulta no acúmulo de bactérias nas águas de reservação do sistema adutor.

Apesar da grande quantidade de dados o tempo de processamento dos dados para cada variável analisada, no tocante aos indicadores de qualidade, foi baixo, demonstrando que a arquitetura é adequada para o estudo realizado.

A degradação da qualidade da água é avaliada através de grande número de indicadores físico-químicos e microbiológicos. Entretanto, em 2007, o Ministério da Saúde,

através da “Diretriz Nacional do Plano de Amostragem da Vigilância em Saúde Ambiental Relacionada à Qualidade da Água para Consumo Humano” (BRASIL, 2007) recomendou a utilização dos indicadores sentinelas CRL e turbidez, para o controle e, principalmente, a vigilância da qualidade da água, em vista de sua capacidade de indicar precocemente perigos associados à ingestão da água. Além disso, bactérias heterotróficas, microrganismos que têm matéria orgânica como fonte de carbono e crescimento em biofilmes, assumem papel relevante como indicadores da degradação da qualidade da água como este aqui avaliado.

Desta forma, foi possível verificar que os indicadores são suficientes para validar o modelo proposto, de modo a torná-lo uma ferramenta de auxílio à tomada de decisão. O modelo pode ser adequado a qualquer situação corriqueira no tocante a operacionalização do sistema de abastecimento, e ajuda a evitar problemas futuros como a contaminação em pontos específicos da cidade.

A rede neural auto-organizável, apesar de não se utilizar de coeficientes ou leis químicas e físicas, através do processo de aprendizagem reproduziu os fenômenos ocorridos na rede de abastecimento, e caracterizou cada ponto de monitoramento descrevendo as variáveis pertinentes de acordo com a sua correlação.

A lógica fuzzy auxiliou no processo de interpretação da rede auto-organizável, e cooperou para o levantamento de estratégias no que tange à gestão da qualidade da água.

Os principais problemas encontrados sempre estão relacionados com a gestão e operacionalização do sistema.

CONCLUSÕES



Conclui-se que para o presente trabalho foi possível a geração dos mapas SOM a partir da base de dados obtidos pelo monitoramento em pontos estratégicos da cidade de Campina Grande, utilizando indicadores preconizados pela “Diretriz Nacional do Plano de Amostragem da Vigilância em Saúde Ambiental Relacionada à Qualidade da Água para Consumo Humano” (BRASIL, 2007). Os pontos de monitoramento foram suficientes para o treinamento da rede auto-organizável.

Os mapas SOM construídos para os pontos de monitoramento como variáveis do modelo foram importantes na identificação de problemas relativos aos indicadores da qualidade da água.

Foi possível identificar os grupos formados no processo de aprendizagem da rede SOM, a correlação existente entre as variáveis e as justificativas para cada perfil identificado.

Os resultados obtidos pela rede neural auto-organizável são compatíveis com os dados de treinamento.

A extração das regras difusas do mapa SOM, com base nos pesos de maior potencial, que caracterizem o sistema, foi importante no processo de interpretação dos mapas e auxiliou na construção de cenários que possibilitem a tomada de decisão, considerando as possíveis situações enfrentadas por um operador do sistema de abastecimento.

Uma das recomendações de trabalhos futuros é a aplicação do modelo em um conjunto de dados da qualidade da água atualizado diariamente, de modo que o modelo neuro-fuzzy se adeque a metodologias de previsão em relação a um determinado intervalo de tempo futuro. Recomenda-se também a inserção de outros reservatórios no modelo, como por exemplo o R-2, que abastece a região identificada pelo P8 e que pode influenciar na melhoria da qualidade da água a partir do instante em que suas características forem assimiladas pelo modelo.

REFERÊNCIAS

- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12.211 - Estudos de concepção de sistemas públicos de abastecimento de água**, Rio de Janeiro: ABNT, 1989.
- ALATRISTA-SALAS, H., AZÉ, J., BRINGAY, S., CERNESSON, F., SELMAOUI-FOLCHER, N., TEISSEIRE, MA.. Knowledge discovery process for spatiotemporal data: Application to river water quality monitoring. **Ecological Informatics**, 2014.
- ALBUQUERQUE, A. A. **Análises e Métodos de Otimização para Eficiência Energética de Estações Elevatórias em Sistemas de Abastecimento de Água**. Tese de Doutorado. Programa Institucional de Doutorado Temático em Recursos Naturais da UFCG. Campina Grande, 2007.
- APHA, AWWA, WPCF. **Standard Methods for the Examination of Water and Wastewater**. 20th ed., Washington, D.C: American Public Health Association/American Water Works Association/Water Environment Federation, 1999. 1220p.1
- BÁGGIO, M. A. **Gerenciamento da operação de sistemas de abastecimento de água: nova filosofia**. Londrina: SANEPAR, 1998. 21p.
- BARBOSA, P. S. F.; COSTA, A. A.; SANTOS JÚNIOR, J. B. S. Um algoritmo de programação linear para análise de redes hidráulicas. **Revista Brasileira de recursos Hídricos**, v.4, n.4. p.75-84, 1999.
- BÉDARD, Y. . Integrating GIS components with knowledge discovery technology for environmental health decision support. **International journal of medical informatics**, v. 70, n. 1, p. 79-94, 2003.
- BODY, M., MIQUEL, M., BÉDARD, Y., TCHOUNIKINE, A.. A multidimensional and multiversion structure for OLAP applications. In: **Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP**. ACM, 2002. p. 1-6.
- BORGES, J.T; SPARRAPAN, R.; GUIMARÃES, J. R.; EBERLIN, M. N. **A influência da amônia no potencial de formação de trihalometanos**. In: CONGRESSO INTERAMERICANO DE INGENIERIA SANITÁRIA Y AMBIENTAL. 28, Canun, México, octubre, 2002.
- BRASIL. Ministério da Saúde (MS), Secretaria de Vigilância em Saúde (SVS), Coordenação Geral de Vigilância em Saúde Ambiental (CGVAM), **Vigilância da Qualidade de Água para Consumo Humano (Vigiagua)**. Relatório das atividades vigiagua 1998 a 2005. 2005.
- BRASIL. MINISTÉRIO DA SAÚDE. **Portaria nº2914**. Disponível na Internet: http://bvsmms.saude.gov.br/bvs/publicacoes/portaria_2914_2011.pdf. 2011. Acesso em 18/08/2013.

BRASIL. Ministério da Saúde. **Portaria n.º 518**, de 25 de março de 2004. Estabelece os procedimentos e responsabilidades relativos ao controle e vigilância da qualidade da água para consumo humano e seu padrão de potabilidade, e dá outras providências, 2004.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Diretriz nacional do plano de amostragem da vigilância em saúde ambiental relacionada à qualidade da água para consumo humano**. Brasília: Ministério da Saúde, 2007. 60p.

CAGEPA – Companhia de Água e Esgotos da Paraíba. **Relatório de Informações Cadastrais. Campina Grande, 2008**. in MENESES, R.A, **Diagnostico Operacional de Sistemas de Abastecimento de Água: o caso de Campina Grande. Campina Grande**, (2011). 144p. Dissertação (Mestrado) – Universidade Federal de Campina Grande, Campina Grande.

CALIJURI, M. L.; SANTIAGO, A. da F.; CAMARGO R. de A.; MOREIRA NETO R. F. Estudo de indicadores de saúde ambiental e de saneamento em cidade do Norte do Brasil. **Revista Engenharia Sanitária e Ambiental**, v.14 n.1, 2009.

CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. **Computer vision, graphics, and image processing**, v. 37, n. 1, p. 54-115, 1987.

CARRIJO, I. B.; REIS, L. F. R. **Extração de regras operacionais ótimas de sistemas de distribuição de água através de algoritmos genéticos e aprendizado de máquina**. In: Seminário Hispano-Brasileiro sobre Sistemas de Abastecimento Urbano de Água João Pessoa (Brasil), 2004.

COELHO, M. I. S.; MENDES, E. S.; CRUZ, M. C. S.; BEZERRA, S. S.; SILVA, R. P. P. **Avaliação da qualidade microbiológica de águas minerais consumidas na região metropolitana de Recife, Estado de Pernambuco**. Biblioteca virtual em saúde. Pesquisa em base de Dados. DOI: 10.4025/actascihealthsci.v32i1.3837. 2010.

CORDÃO, M. J. de S. Modelagem e otimização da disposição espacial de unidades de reservação em redes de distribuição de água utilizando geotecnologias. Dissertação de Mestrado. Universidade Federal de Campina Grande. PPGECA/UFPG. Campina Grande, 2009. 115P.

DI BERNARDO, L. **Tecnologias de tratamento de Água com filtração rápida**. In: Seminario Taller Selección de Tecnología para el Mejoramiento de la Calidad de Agua. Santiago de Cali, 2000.

DUARTE, A..de S.; CUNHA, A.C. da; BRITO, D.C. de. **Monitoramento da qualidade da água utilizada para consumo humano no bairro Santa Rita, Macapá, Amapá, Brasil**, 2009. In: CONGRESSO BRASILEIRO DE ENGENHARIA SANITÁRIA E AMBIENTAL, 25, 2009, Recife, **Anais...** Rio de Janeiro: ABES 2009.

DŽEROSKI, S., GRBOVIĆ, J., WALLEY, W. J., KOMPARE, B.. Using machine learning techniques in the construction of models. II. Data analysis with rule induction. **Ecological Modelling**, v. 95, n. 1, p. 95-111, 1997.

FARACHE FILHO, A.; DIAS, M. F. F. Qualidade microbiológica de águas minerais em galões de 20 litros. **Alimentação e Nutrição**, v. 19, n. 3, p. 243-248, 2008.

GUR-ALI, OZDEN; WALLACE, WILLIAM A.. . Induction of rules subject to a quality constraint: Probabilistic inductive learning. **Knowledge and Data Engineering, IEEE Transactions on**, v. 5, n. 6, p. 979-984, 1993.

GRUBBS, F. E. Sample criteria for testing outlying observations. **The Annals of Mathematical Statistics**, p. 27-58, 1950.

HAYKIN, S.. **Neural Networks: A Comprehensive Foundation**. 3ª edição. Prentice-Hall, 2001.

HEBB, D. O. **The organization of behavior: A neuropsychological theory**. Psychology Press, 1949.

HELLER, L.; PÁDUA, V. L. (Org.). **Abastecimento de água para consumo humano**. Belo Horizonte: UFMG, 2006.859p.

HUBEL, D. H.; WIESEL, T.; N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **The Journal of physiology**, v. 160, n. 1, p. 106, 1962.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE) (Brasil). **Censo 2000**. Disponível na Internet:<http://www.ibge.gov.br/home/estatistica/populacao/censo2000/default.shtm>. 2000. Acesso em 18/08/2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE) (Brasil). **Pesquisa Nacional de Saneamento Básico: 2008**. Brasília, DF: Ministério das Cidades; Ministério do Orçamento e Gestão, 2008.

KIANG, M. Y. Extending the Kohonen Self-Organizing Map Networks for Clustering Analysis, **Computational Statistics & Data Analysis**, v.38, i. 2, p. 161-180, 2001. DOI: 10.1016/S0167-9473(01)00040-8

KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological cybernetics**, v. 43, n. 1, p. 59-69, 1982.

KOHONEN, T. **Self-Organizing Maps**. 3ª ed., New York: Springer, 2001.

KOHONEN, T., KASKI, S., LAPPALAINEN..**Self-Organization Formation of Various Invariant-Feature Filters in the Adaptive-Subspace SOM**. **Neural Computation**, vol 9, nº6, pg. 1321-1344, 1997.

KOVÁCS, Zsolt L. **Redes neurais artificiais**. Editora Livraria da Física, 2006.

LEAL, E. S. Modelagem da degradação de cloro residual livre em sistemas de adução de água de abastecimento de porte médio. 2012. 106p. **Dissertação** (Mestrado em Engenharia Civil e Ambiental) – Universidade Federal de Campina Grande, Paraíba, 2012.

LEY, R., CASPER, M. C., HELLEBRAND, H., MERZ, R.. Catchment classification by runoff behaviour with self-organizing maps (SOM). *Hydrology and Earth System Sciences*, 15(9), 2947-2962, 2011.

MAMDANI, E. H.; ASSILIAN, S. An experiment in linguistic synthesis with a fuzzy logic controller. **International journal of man-machine studies**, v. 7, n. 1, p. 1-13, 1975.

MENESES, R. A, Diagnóstico operacional de sistemas de abastecimento de água: o caso de Campina Grande. 2011. 144p. **Dissertação** (Mestrado em Engenharia Civil e Ambiental) – Universidade Federal de Campina Grande, Paraíba, 2011.

MOUNCE, S. R. . Data mining T-RFLP profiles from urban water system sampling using self-organizing maps. In: **Natural Computation (ICNC), 2012 Eighth International Conference on**. IEEE, 2012. p. 212-216.

NEUMANN, J. **Classification and Evaluation of Algorithm for Rule Extraction from Artificial Neural Networks**, PhD Summer Project, ICCS Division of Informatics, University of Edinburgh, August, 1998.

PARK, Y. S., KWON, Y. S., HWANG, S. J., PARK, S.. Characterizing effects of landscape and morphometric factors on water quality of reservoirs using a self-organizing map. **Environmental Modelling & Software**, v. 55, p. 214-221, 2014.

PARSEKIAN, M. P. S. Análise e proposta de formas de gerenciamento de estações de tratamento de águas de abastecimento completo em cidades de porte médio do estado de São Paulo. 1998. 194f. **Dissertação** (Mestrado em Hidráulica e Saneamento). Universidade de São Paulo, SP, 1998.

PAZ, L. P. S. Modelo conceitual de tecnologias de tratamento de comunidades de pequeno porte. 2007. 398f. **Tese** (Doutorado em Hidráulica e Saneamento), Escola de Engenharia de São Carlos, São Carlos, 2007.

PEAT, M. M. **Train dispatching simulation model: capabilities and description**. USA, 1985.

PYLE, D. **Data preparation for data mining**. Morgan Kaufmann, 1999.

ROSS, T. J. **Fuzzy Logic with Engineering Applications**, 2a ed., England: John Wiley and Sons Ltd, 2004.

SANCHES, S. M.; SILVA, C. H. T. de P.; VIEIRA, E. M. Agentes desinfetantes alternativos para o tratamento de água. **Revista Química Nova na Escola**; n. 17; 2003.

SANTOS, M. F., CORTEZ, P., QUINTELA, H., NEVES, J., VICENTE, H. Ecological mining: a case study on dam water quality. 2005.

SANTOS, S. G. Distribuição espacial de bactérias heterotróficas na rede de distribuição de água de Campina Grande-PB / Sarah Golzio dos Santos. — **Dissertação** (Mestrado em Engenharia Civil e Ambiental) – Universidade Federal de Campina Grande. Campina Grande, 2011.

SILVA, A.C.M.. Extração de conhecimento em forma de regras difusas a partir de mapas auto-organizáveis de Kohonen: aplicação em diagnóstico de faltas incipientes em transformadores. 2013.

SILVA; S. A.; OLIVEIRA, R, **Manual de análise físico-químicas de águas de abastecimento e residuárias**. Campina Grande, PB: O Próprio Autor, 2001. 266p.

SOKAL, R. R. Rohlf; ROHLF, F. J. Biometry (2nd edn). **New York: WH Feeman and Company**, v. 668, 1981.

ULTISCH, A.; SIEMON, H. **Kohonen's Self Organization Feature Maps for Exploratory Data Analysis**. In: **International Neural Network Conference**. Kluwer Academic Press, Paris, pg. 305-308, 1990.

VAN HULLE, M.M. **Faithful Representation and Topographic Maps: From Distortion to Information-Based Self-Organization**. John Wiley & Sons, 2000.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. **SOM Toolbox for Matlab 5**, Technical Report A57. Finland: Helsinki University of Technology, 2000.

VIEIRA, B.G.A.. Análise de risco aplicada à qualidade da água do sistema de abastecimento de Campina Grande (PB) . **Dissertação** (Mestrado em Engenharia Civil e Ambiental) – Universidade Federal de Campina Grande. Campina Grande, 2012.

VON DER MALSBURG, C. Self-organization of orientation sensitive cells in the striate cortex. **Kybernetik** 14.2 (1973): 85-100.

VOUTILAINEN, A., RAHKOLA-SORSA, M., PARVIAINEN, J., HUTTUNEN, M. J., VILJANEN, M. . Analysing a large dataset on long-term monitoring of water quality and plankton with the SOM clustering. **Knowledge and Management of Aquatic Ecosystems**, n. 406, p. 04, 2012.

WATERS, M. Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. **Environmental Health Perspectives**, v. 111, n. 6, p. 811, 2003.

ZADEH, L. The Concept of a Linguistic Variable The Concept of a Linguistic Variable – I, **Information Sciences** , v. 8, i. 3, p. 199-249, 1975.