



SPOLM 2007

ISSN 2175-6295

Rio de Janeiro- Brasil, 08 e 09 novembro de 2007.

## **APOIO A DECISÃO COM APLICAÇÃO DE REDE NEURAL ARTIFICIAL CONSTRUTIVA E SISTEMA DE INFORMAÇÕES NA AVALIAÇÃO DE QUALIDADE DE ÁGUA DE RESERVATÓRIOS**

**Ivna Valença, Jorge Cruz, Jarley Nóbrega, Meuser Jorge Silva Valença**

Universidade Salgado de Oliveira

Departamento de Sistemas de Informação, Av. Mascarenhas de Moraes, 1919 - Imbiribeira  
Recife - PE 51150-001

[ivnavalenca@yahoo.com](mailto:ivnavalenca@yahoo.com), [jpn@jarley.com](mailto:jpn@jarley.com), [meuserj@yahoo.com.br](mailto:meuserj@yahoo.com.br)

**Domingos Vanderlei Filho, Sheila Melo**

Companhia Hidro Elétrica do São Francisco

Departamento de Meio Ambiente, Rua Delmiro Gouveia, 333 - Bongi Recife – Pernambuco  
50761-901

[dvanderlei@yahoo.com.br](mailto:dvanderlei@yahoo.com.br), [sheilab@chesf.gov.br](mailto:sheilab@chesf.gov.br)

**Teresa B. Ludermir**

Universidade Federal de Pernambuco

Centro de Informática, Rua Acadêmico Hélio Ramos, s/n - Cidade Universitária Recife –  
Pernambuco 50740-530

[tbl@cin.ufpe.br](mailto:tbl@cin.ufpe.br)

### **Resumo**

No monitoramento de qualidade da água de reservatórios podem ser avaliados diversos parâmetros. Dentre as variáveis coletadas pode-se destacar: temperatura, pH, salinidade, clorofila, oxigênio dissolvido entre outras. No caso da Companhia Hidro Elétrica do São Francisco (CHESF) essas medições são realizadas atualmente com periodicidade trimestral, em diversos pontos de coleta dos reservatórios. No entanto, tais dados coletados não estão sendo ainda utilizados para o controle preventivo de monitoramento da qualidade da água, nem estão sendo armazenados em um banco de dados unificado. Assim, propõe-se a introdução de um sistema de informações ambientais para qualidade de água (SIAQUA) com o objetivo de reunir todos os dados relevantes em um único banco de dados, e que gere informações concisas, confiáveis e que permita sua utilização de forma fácil por meio de um sistema baseado em redes neurais artificiais (RNA) que faça interpolação e/ou previsão. A utilização deste sistema vem facilitar e apoiar a tomada de decisão, tanto no que diz respeito à quantidade de pontos que devem ser monitorados, quanto às condições da qualidade da água dos reservatórios que tem o impacto direto nos demais usuários da água.

**Palavras-Chaves:** Apoio à Decisão, Redes Neurais Artificiais; Sistema de Informações;

### Abstract

When monitoring the reservoir's water quality it's possible evaluate several different kind of parameters. Among the measured variables it's possible to highlight: temperature, pH, salinity, chlorophyll, dissolved oxygen and others. Nowadays the Companhia Hidroelétrica do São Francisco (CHESF) makes the monitor of diverse parameters of quality of the water on theirs reservoirs every three months. These measurements are performed in many sampling points of the reservoirs. However, such measurements are still neither being used for the preventive control and monitor of the water quality and nor stored in a centralized data base. The Environmental Information System for Water Quality Control (SIAQUA) has the objective to congregate and concentrate the most important data in a only one data base, and that It could provide reliable information. Additionally, It is expected that the SIAQUA make the things easier for the users because the system offers a interpolation and/or forecast based in artificial neural network technology. The use of this system contributes to make easy the decision making in terms of the amount of sampling points that must be monitored, and about the water quality of the reservoirs and its consequences on other users of the water.

**Keywords:** Decision Support; Artificial Neural Networks; Information Systems; Water Quality.

## 1. INTRODUÇÃO

Todos os usuários que utilizam as águas dos reservatórios da CHESF localizados no rio São Francisco podem ser beneficiados com a análise das informações que podem ser geradas a partir da coleta de dados de qualidade das águas. Hoje, apesar de a CHESF realizar esta coleta, tais informações não são totalmente aproveitadas pela falta de disponibilidade de um sistema que permita a consulta e análise destas informações haja vista, não se dispor de um banco de dados unificado que permita uma análise integrada através de ferramentas de previsão e de interpolação.

Como exigência legal, a CHESF, faz o monitoramento de diversos parâmetros de qualidade da água dos seus reservatórios. Dentre estes dados coletados pode-se destacar: temperatura, pH, salinidade, clorofila, oxigênio dissolvido, entre outros. Essas medições são realizadas atualmente com periodicidade trimestral em diversos pontos de coleta dos reservatórios. No entanto, tais dados coletados não estão sendo ainda utilizados para o controle preventivo de monitoramento da qualidade da água, nem estão sendo armazenados em um banco de dados unificado.

O SIAQUA tem o objetivo de reunir todos os dados relevantes em um único banco de dados, e assim gerar informações concisas, confiáveis e que permita sua utilização de forma fácil através de um sistema que faça interpolação e/ou previsão. A utilização deste sistema vem facilitar a tomada de decisão tanto do que diz respeito à quantidade de pontos que devem ser monitorados quanto às condições da qualidade da água dos reservatórios que tem o impacto direto nos demais usuários da água tais como: agricultores, piscicultores, fazendeiros, ou seja, pessoas e animais que estejam em contato direto com a água do rio ou próximas a ela.

Como exemplo de resultados interessantes que podem ser obtidos a partir deste sistema, pode-se citar: a interpolação e/ou a previsão de diversos parâmetros de qualidade da água, os quais estão diretamente relacionados com as atividades de empreendimentos de piscicultura em tanques-rede. Com este objetivo utiliza-se a técnica de redes neurais construtivas as quais facilitam o uso desta ferramenta, pois é possível obter uma arquitetura ótima com uma estrutura que evolui durante o processo de treinamento.

Salienta-se que a interpolação de parâmetros de qualidade da água busca minimizar custos, por meio da otimização da quantidade dos pontos de coleta que permitam uma interpolação com desempenho satisfatório, isto vai contribuir para um redimensionamento da

capacidade de produção em tanques-rede dos reservatórios.

Relacionado a este tipo de atividade econômica é possível destacar dois pontos, onde a previsão pode ser bastante útil:

- previsão do total de gases dissolvidos - aviso preventivo aos piscicultores durante uma cheia dos parâmetros de qualidade da água relacionados à mortalidade de peixes em função da supersaturação gasosa;
- previsão de clorofila - embora essa atividade seja apenas uma das atividades antrópicas que podem causar degradação na qualidade da água, recomenda-se monitoramento periódico da água em seus aspectos químicos físicos e biológicos, pois um aumento significativo de nutrientes (nitrogênio e fósforo), provenientes de transformações dos restos de alimento como a ração e resíduos metabólicos de peixes, podem levar a eutrofização desses reservatórios a qual pode ser traduzida em um aumento percentual de clorofila.

Este artigo está organizado em 5 seções. A seguir na seção 2, apresenta-se o sistema proposto e sua forma de implementação. Na seção 3 é apresentada a rede neural construtiva NSRBN. Os resultados obtidos com exemplos de utilização do sistema são descritos na seção 4. A seção 5 apresenta as conclusões e recomendações, bem como os possíveis desdobramentos futuros deste trabalho.

## **2. O SISTEMA DE INFORMAÇÕES SIAQUA**

O sistema de informações proposto neste artigo está descrito e apresentado em suas características nas subseções a seguir.

### **2.1. VISÃO GERAL**

O Quadro 1 mostra uma visão geral deste sistema.

QUADRO 1 – Visão geral do SIAQUA.

<b>O problema de</b>	A falta de um banco unificado.
<b>Afeta</b>	A CHESF e demais órgãos interessados na utilização dos dados.
<b>O impacto disso é</b>	A coleta dos dados hoje é realizada a partir de empresas terceirizadas que seguem um padrão mínimo estabelecido pelo IBAMA (em caso de rios federais), tais dados são transformados em informações através de relatórios e guardados em papel ou em CD. Como não existe um padrão para os relatórios e sim um conjunto mínimo de informações que necessita ser apresentada, cada empresa confecciona seus relatórios seguindo os seus próprios padrões. Com a falta de padronização é impossível realizar uma análise detalhada dos dados, tendo em vista que uma empresa “X” que realiza as coletas pode não realizar estas nos mesmos pontos coletados pelas empresas anteriores.
<b>Uma solução de sucesso seria</b>	A padronização de todos os relatórios voltados às coletas de dados sobre a água poderia possibilitar um melhor aproveitamento futuro das informações através de uma série histórica.
<b>O problema de</b>	Falta de um sistema de interpolação dos dados: Não existe hoje um critério claro que defina quantos pontos devem ser coletados em cada reservatório.
<b>Afeta</b>	A CHESF, demais órgãos interessados na utilização dos dados e usuários das águas dos reservatórios.
<b>O impacto disso é</b>	A não existência dos dados em todos os locais desejados.
<b>Uma solução de sucesso seria</b>	A criação de um sistema de interpolação que permita ao usuário obter o parâmetro de qualidade da água no local desejado.
<b>O problema de</b>	Falta de um sistema de previsão: O sistema de previsão tem por objetivo antecipar possíveis variações nos parâmetros de qualidade da água.
<b>Afeta</b>	A CHESF e demais usuários das águas dos reservatórios.
<b>O impacto disso é</b>	Não permitir antecipar medidas que possam minimizar possíveis impactos ambientais.
<b>Uma solução de sucesso seria</b>	A criação de um modelo de previsão para diversos parâmetros de qualidade da água.

### 2.1.1. Resumo da Capacidade do SIAQUA

Todos os usuários que utilizam as águas dos reservatórios da CHESF, localizados no rio São Francisco, podem ser beneficiados com a análise das informações que podem ser geradas a partir da coleta de dados de qualidade das águas. Hoje, apesar de a CHESF realizar esta coleta, tais informações não são totalmente aproveitadas pela falta de disponibilidade de um sistema que permita a consulta e análise destas informações, considerando-se que não existe um banco de dados unificado.

O SIAQUA tem o objetivo de reunir todos os dados relevantes em um único banco de dados, e assim gerar informações concisas, confiáveis e que permitam sua utilização de forma fácil por meio de um sistema que incorpore ferramentas que permitam fazer interpolação e/ou previsão.

Portanto, a utilização deste sistema vem facilitar a tomada de decisão, tanto no que diz respeito à quantidade de pontos que devem ser monitorados, quanto às condições da qualidade da água dos reservatórios, que tem o impacto direto nos demais usuários da água tais como: agricultores, piscicultores, fazendeiros, ou seja, pessoas e animais que estejam em contato direto com a água do rio ou próximas a ela.

O usuário do SIAQUA poderá utilizar a aplicação inicialmente instalada no seu computador pessoal. Devido às restrições de segurança nos ambientes de trabalho, a aplicação deverá prover também acesso ao sistema através da web, sincronizando as informações entre cliente e servidor. Como característica mais avançada, o sistema proposto deverá também estar disponível em ambientes móveis como aparelhos telefônicos celulares e Assistentes Pessoais Digitais (em inglês - PDA), permitindo ao usuário mobilidade no uso da aplicação.

## 2.2. CARACTERÍSTICAS DO SIAQUA

A seguir são mostradas algumas características do sistema.

- Registro das coletas realizadas, dos usuários, dos parâmetros de coleta, das bacias hidrográficas, das estações, dos reservatórios e dos pontos de coleta: Permite o cadastramento na base de dados do sistema.
- Interpolação: Decidir quais são os melhores pontos de coleta, interpolar a variável desejada nas coordenadas desejadas e traçar os caminhos que deveram ser seguidos para a realização da tarefa. O processo de interpolação utiliza a técnica de redes neurais construtivas [1-2].
- Previsão: Permite a previsão de diversos parâmetros de qualidade da água com base nas informações coletadas. O processo de interpolação utiliza a técnica de redes neurais construtivas [1-2].
- Seleção de variáveis: Permite a seleção das variáveis coletadas.
- Geração de relatórios de usuários, coletas, bacias, estações, reservatórios, pontos de coleta e solicitação de análises estatísticas: Gera um relatório, podendo ser impresso posteriormente.
- Solicitações: Permite a solicitação dos dados por pesquisadores.

## 2.3. OUTROS REQUISITOS DO SIAQUA

Segue a descrição de alguns requisitos adicionais propostos para o sistema. Portanto, no tocante aos padrões aplicáveis a solução proposta deverá estar disponível na plataforma Java 2 e fará utilização do TCP/IP como protocolo de comunicação. O acesso ao servidor deverá ocorrer usando o Protocolo de Transferência de Hipertexto HTTP. A utilização de *Web Services* deverá ser estudada pela equipe de desenvolvimento.

No que se refere aos requisitos de ambientes a aplicação deverá rodar na plataforma Microsoft Windows e ser portada para Linux posteriormente. O acesso à base remota não deverá ser acessível para todos os usuários, tendo em vista políticas internas da empresa quanto à proteção dos dados, apenas usuários com permissões específicas terão acesso.

As Figuras de 1 a 8 a seguir fornecem uma visão de algumas telas do sistema com relação a suas diversas funcionalidades.

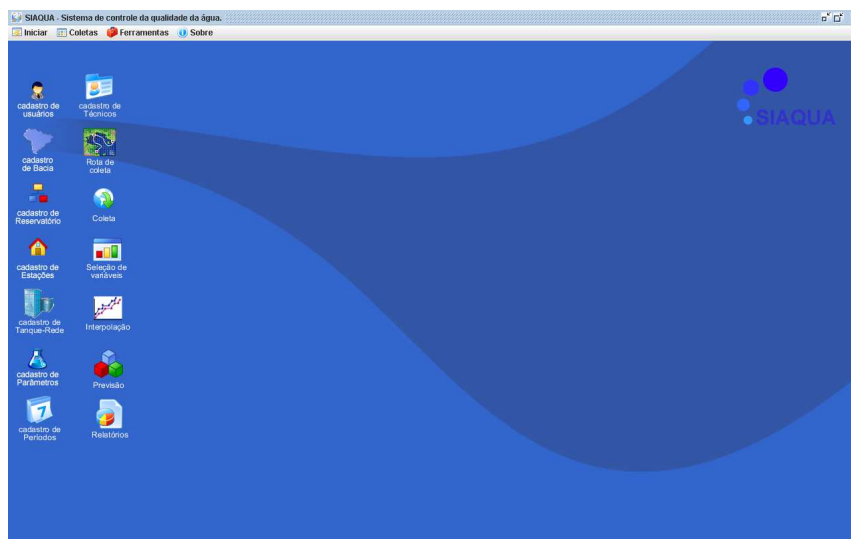


FIGURA 1: Tela de Abertura

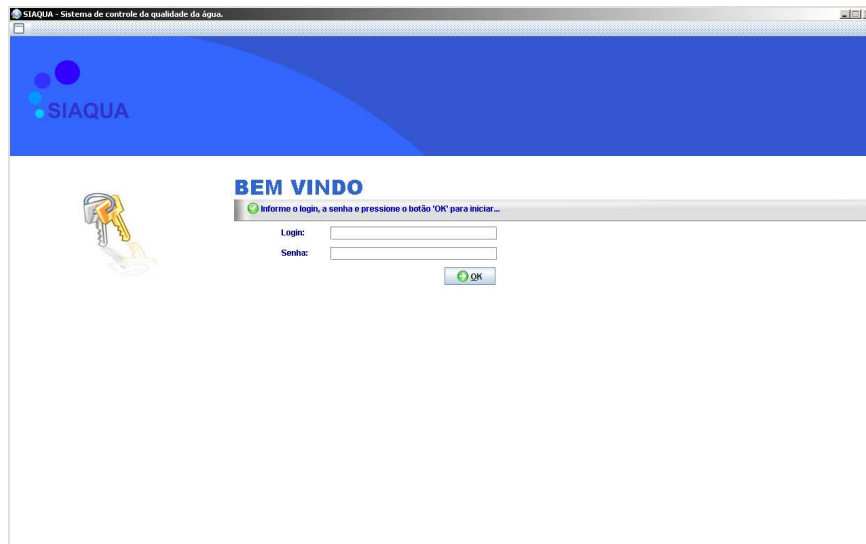


FIGURA 2: Tela de *Login*

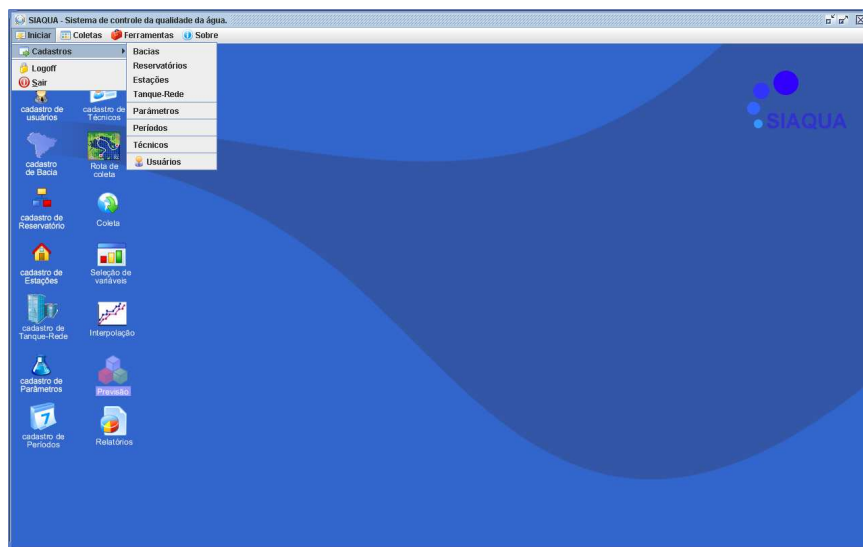


FIGURA 3: Tela com Menu Ferramentas

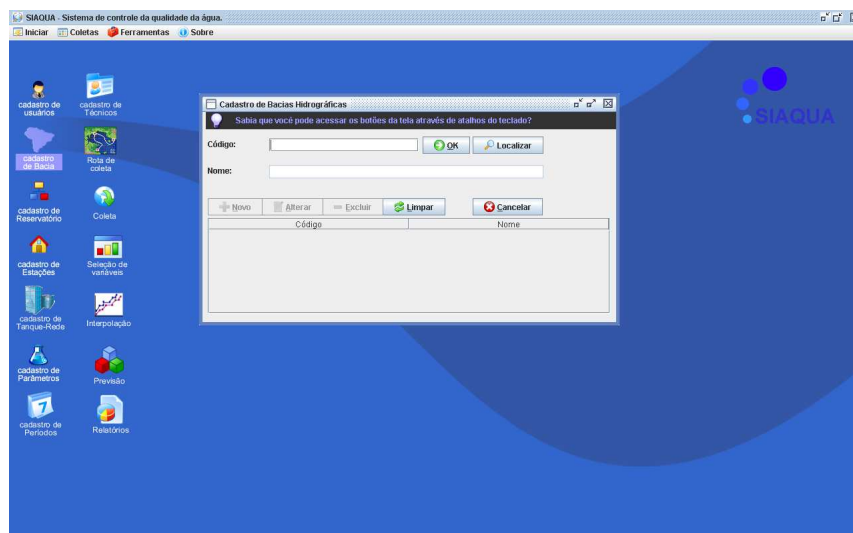


FIGURA 4: Tela de Cadastro de Bacia

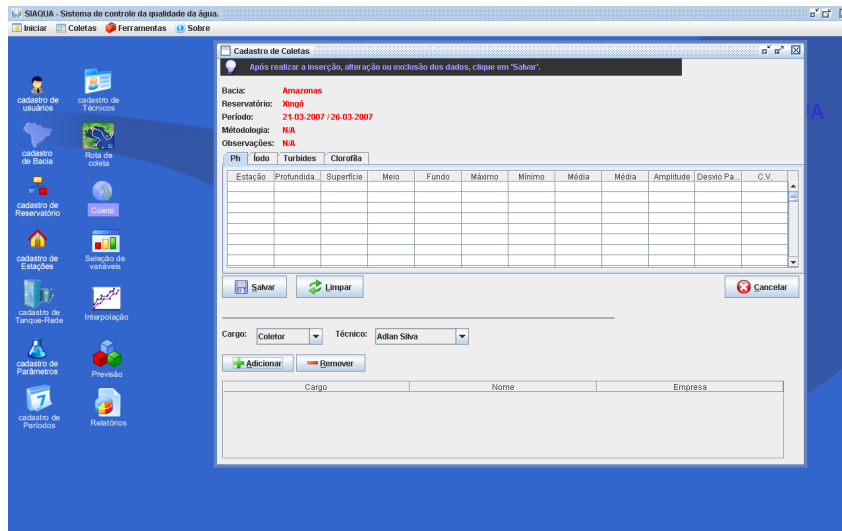


FIGURA 5: Tela de Cadastro de Coleta

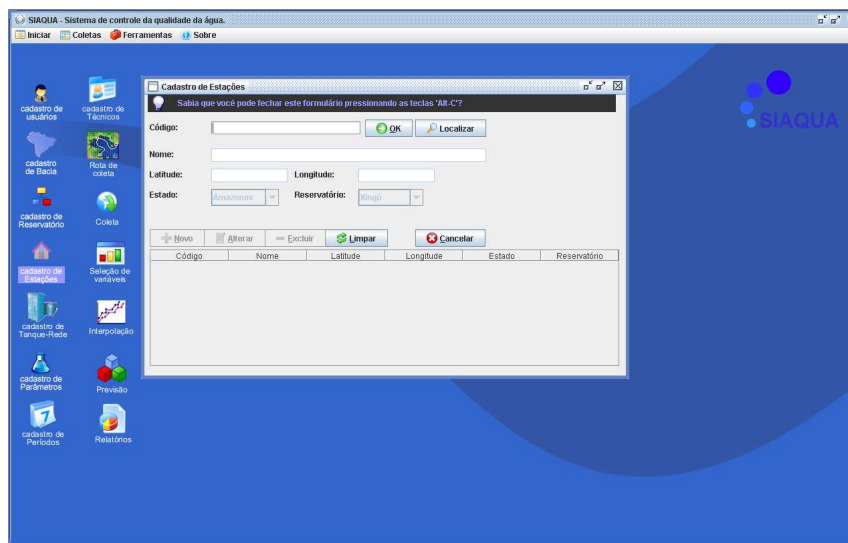


FIGURA 6: Tela de Cadastro de Estações

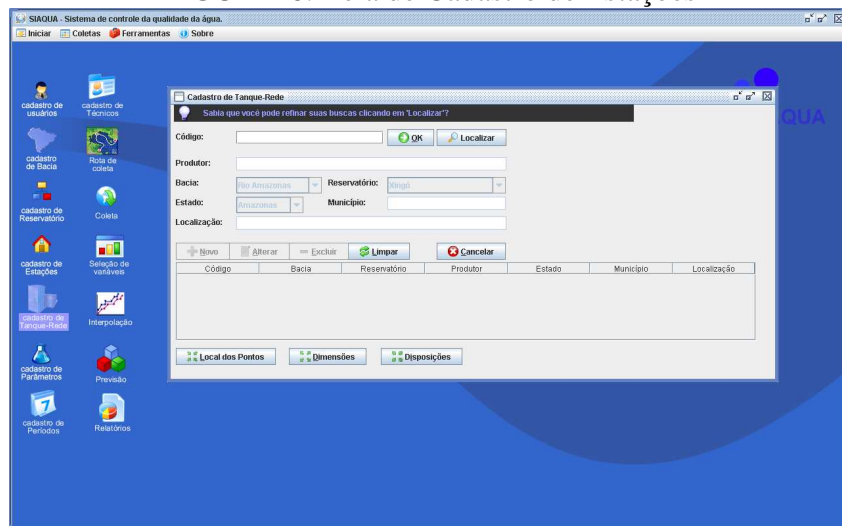


FIGURA 7: Tela de Cadastro de Tanque-rede

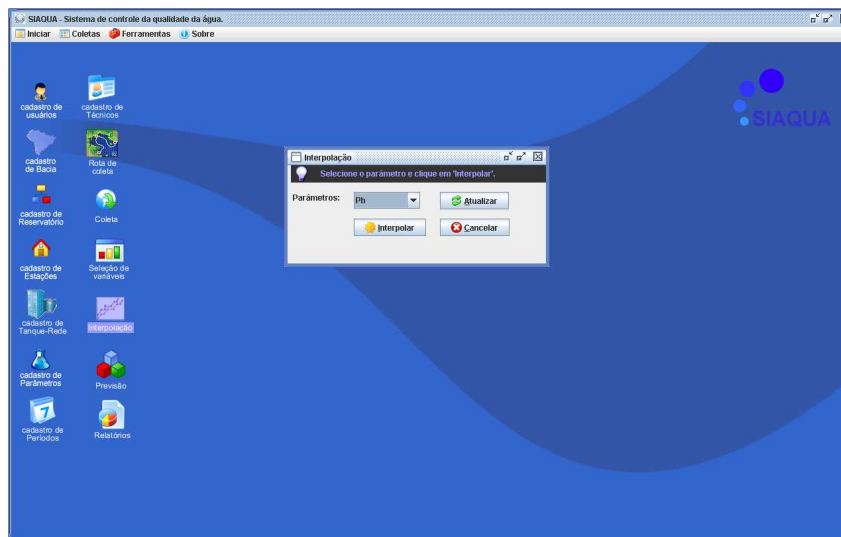


FIGURA 8: Tela para acesso a ferramenta de Interpolação

### 3. REDE NEURAL CONSTRUTIVA

Neste trabalho foram utilizadas RNAs auto-organizáveis que se caracterizam por utilizarem uma metodologia híbrida, redes neurais e conceitos estatísticos, com o objetivo de estimar RNAs de arquitetura ótima com uma estrutura que evolui durante o processo de treinamento. A classe de RNAs *feedforward* de alta ordem, chamadas de redes neurais por combinação de blocos de regressões sigmóides não-lineares (NSRBN), são capazes de aproximar qualquer função contínua definida sobre um conjunto compacto [1]. Essas RNAs têm como fundamento a auto-organização, além de utilizar um aprendizado incremental baseado em um algoritmo construtivo. Este algoritmo construtivo é responsável pela escolha do modelo de ótima complexidade, ou seja, pela definição da arquitetura da rede. As NSRBN podem ser vistas como aproximações polinomiais multivariadas, uma vez que elas têm como elemento de construção básico polinômios homogêneos de grau  $k$ , nos quais  $k$  varia de 1 a  $N$  (complexidade máxima, ou seja, máximo grau polinomial).

A arquitetura de uma RNA, de acordo com a possibilidade de mudanças no seu tamanho durante o processo de treinamento, pode ser classificada em estática ou dinâmica. Uma rede é dita de arquitetura estática quando o número de camadas e unidades de processamento permanece constante desde a concepção da RNA até a finalização do seu treinamento. Por outro lado, uma RNA de arquitetura dinâmica tem como característica principal um processo de mudanças no tamanho durante o treinamento. Nas NSRBN um algoritmo construtivo é utilizado para incrementar a estrutura da rede, partindo-se de uma pequena estrutura, até que se atinja uma estrutura de tamanho ótimo, segundo critérios estatísticos de seleção adequados.

O modelo NSRBN a rede é construída por blocos de polinômios homogêneos, utilizando uma função de ativação adequada (tangente hiperbólica, por exemplo), para os neurônios escondidos. Quanto aos neurônios de saída, estes têm uma ativação mais complexa, uma vez que realizam uma regressão logística não-linear com relação à saída dos neurônios escondidos. Desta forma, essas redes NSRBN podem aproximar uniformemente qualquer função contínua definida em um conjunto compacto limitado e também lidar com a realização de funções booleanas, o que não é possível com os algoritmos tradicionais do grupo GMDH – *Group Method Of Data Handling* (combinatorial e *multilayer*) [1].

O primeiro passo para a construção de um modelo combinatorial consiste em dividir a função  $f(x)$  em blocos compostos de termos de igual grau, como apresentado na Figura 9.

$$f(x) = f_1(x) + f_2(x) + \dots + f_d(x) \quad (1)$$



O bloco de grau  $p$  ( $p=1, 2, \dots, d$ ), de acordo com o polinômio de Kolmogorov-Gabor (teorema de Weierstrass), pode ser expresso como uma soma de produtos das entradas, isto é:

$$f_p(x) = \sum_{i_1=1}^N \sum_{i_2=i_1}^N \dots \sum_{i_N=i_1}^N a_p(i_1, i_2, i_3, \dots, i_N) \cdot x_1^{q_{p1}(i_1)} \cdot x_2^{q_{p2}(i_2)} \dots x_N^{q_{pN}(i_N)} \quad (2)$$

em que,  $q_{p1}(i_1) + q_{p2}(i_2) + q_{p3}(i_3) + \dots + q_{pN}(i_N) = p$  e  $p \geq q_{pm}(i_m) \geq 0$  para  $m=1, 2, 3, \dots, N$ .

Tem-se como objetivo realizar o mapeamento de cada um dos polinômios homogêneos de grau  $p$ .

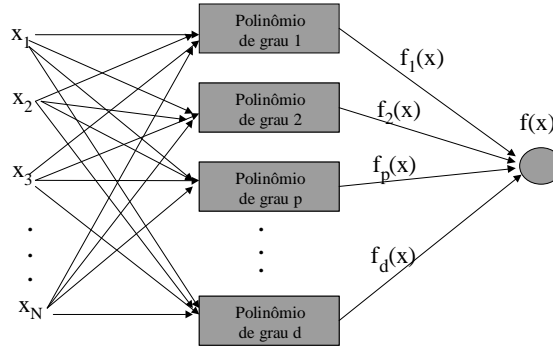


Figura 9: Arquitetura combinatorial

Uma forma simples seria realizar todos os produtos das variáveis de entrada procurando aproximar diretamente a hypersuperfície não-linear, entretanto isto não é prático, a menos que  $d$  e  $N$  sejam relativamente pequenos (modelo combinatorial). Logo, para se ter um modelo eficiente, faz-se necessário que ele possua funções de ativação que sejam capazes de realizar vários termos dos polinômios homogêneos de forma otimizada, isto é, sem que ocorra um crescimento explosivo no número de parâmetros necessários para ajustar a função desejada. Para esse objetivo, as NSRBN têm por base o algoritmo combinatorial (composta pela soma de blocos) com uma estrutura de blocos similar a uma rede *Multilayer Perceptron* (MLP), sendo que, diferentemente destas, os neurônios de saída realizam a técnica estatística de regressão logística não-linear [3], com relação aos neurônios escondidos.

Definição 1: uma rede NSRBN é uma rede combinatorial composta pela soma de  $p$  blocos ( $p=1, \dots, d$ ) com estrutura similar a uma MLP, na qual as unidades escondidas desses blocos realizam uma regressão sigmóide linear das entradas (tal qual uma MLP), e as unidades de saída realizam uma regressão do tipo sigmóide não-linear das unidades escondidas (tal qual uma regressão logística não-linear). Essa nova estrutura formada é um aproximador universal com forte capacidade de mapeamento não-linear. Utiliza-se aqui o termo sigmóide, significando qualquer função sigmoidal, como uma generalização da regressão logística clássica.

A estrutura de cada bloco é composta de apenas uma camada escondida, tendo  $Nh$  unidades escondidas em cada bloco com função de ativação do tipo [3]:

$$f(x) = (\sigma_{net(h)} + \theta_h) \quad (3)$$

em que  $net(h)$  é a entrada líquida para cada neurônio escondido,  $h$  é um número inteiro que representa as unidades escondidas em cada bloco (variando de 1 até a ordem  $p$  do bloco),  $\theta_h$  é a tendência e  $\sigma_{net(h)}$  é, por exemplo, a função tangente hiperbólica dadas por:

$$\sigma_{net(h)} = \frac{e^{net(h)} - e^{-net(h)}}{e^{net(h)} + e^{-net(h)}} \quad (4)$$

Poder-se-ia utilizar outras funções, como é o caso da função recíproca representada na Equação 5, isto é,

$$f_r(x) = \frac{net(h)}{1 + |net(h)|} \quad (5)$$

O valor de  $p$  correspondente ao grau do bloco é quem vai definir o número de neurônios em cada bloco, isto é: o bloco de grau 1 ( $p=1$ ) tem  $h=1$ , logo um neurônio

escondido; o bloco de grau 2 ( $p=2$ ) terá 2 neurônios escondidos ( $h=1$  e  $h=2$ ) e assim sucessivamente, de tal forma que o bloco de grau  $p$  terá  $h=p$  e portanto  $p$  neurônios escondidos.

A estrutura proposta para cada bloco é similar à da Figura 10, na qual apresenta-se o mapeamento para um bloco de grau  $p$  (a ordem dos neurônios escondidos  $h$  variará de 1 até  $p$ ), para  $i$  variáveis de entrada (onde  $i = 1, \dots, N$ ) e uma única unidade de saída.

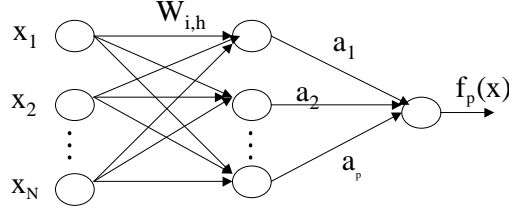


Figura 10: Arquitetura de um bloco polinomial

Esse bloco é uma rede *feedforward* completamente conectada com uma única camada escondida. A diferença básica desse bloco de grau  $p$  para uma rede MLP é que os neurônios de saída são expressos como uma regressão sigmóide não-linear das unidades escondidas.

Seja  $\mathbf{x} = [1, x_1, x_2, x_3, \dots, x_N]^T$  o vetor coluna das variáveis de entrada, no qual o elemento  $x_i$  denota um elemento qualquer dele. As entradas são ponderadas pelos vetores de pesos  $\mathbf{w} = [w_{0h}, w_{1h}, w_{2h}, \dots, w_{Nh}]^T$ ,  $h=1, \dots, p$  nos quais  $p$  é a ordem do bloco. Portanto, utilizando-se função de ativação estabelecida na Equação 3, a saída de cada unidade escondida,  $f(\text{net}(h))$ , desse bloco de grau  $p$  pode ser escrito como:

$$f(\text{net}(h)) = (\sigma_{\text{net}(h)} + \theta_h) \quad (6)$$

Em que:  $\text{net}(h) = \mathbf{w}_h^T \mathbf{x}$ , ou seja,  $\text{net}(h) = \sum_{i=1}^N w_{ih} x_i + w_{0h}$ ,  $h=1, 2, \dots, p$ .

Logo, a unidade de saída (que realiza uma regressão logística não-linear) desse bloco pode ser representada pela seguinte equação:

$$f_p(x) = a_1 (f(\text{net}(1)))^1 + a_2 (f(\text{net}(2)))^2 + a_3 (f(\text{net}(3)))^3 + \dots + a_p (f(\text{net}(p)))^p \quad (7)$$

sendo  $\mathbf{a} = [a_1, a_2, a_3, \dots, a_p]^T$  o vetor dos pesos das unidades escondidas para a unidade de saída.

Fazendo-se uso da Equação 6, a função  $f_p(x)$  pode ser escrita da seguinte forma:

$$f_p(x) = a_1 (\sigma_{\text{net}(1)} + \theta_1) + a_2 (\sigma_{\text{net}(2)} + \theta_2)^2 + \dots + a_p (\sigma_{\text{net}(p)} + \theta_p)^p \quad (8)$$

ou

$$f_p(x) = \sum_{h=1}^p a_h (\sigma_{\text{net}(h)} + \theta_h)^h \quad (9)$$

com  $h$  representando a ordem dos neurônios na camada escondida, ( $h=1, \dots, p$ ), do bloco de grau  $p$ . A arquitetura de uma rede NSRBN (Figura 9) é composta pela soma de blocos com grau variando de 1 a  $d$ . Aqui, está sendo utilizado o bloco  $p$  (Figura 10) para representar qualquer um desses blocos. Portanto, usando-se esta arquitetura, uma função desconhecida  $f$  em  $R^N$  pode ser aproximada por uma rede NSRBN, através da seguinte expressão:

$$f(x) = \sigma_{\text{net}(o)} \left( \sum_{p=1}^d f_p(x) \right) \quad (10)$$

em que:  $\sigma_{\text{net}(o)}$  uma função de ativação adequada para a unidade de saída.

Por outro lado, a construção das redes NSRBN é feita de forma incremental, de maneira que pode-se representa-la em uma forma similar a de um bloco (Equação 8), conforme apresenta-se a seguir. Considere o passo do algoritmo para o qual se se adiciona à rede o bloco de ordem  $d$ ,  $f_d$ , tendo-se anteriormente adicionado os blocos  $f_0, f_1, \dots, f_{d-1}$ . Fazendo-se  $f_0 \equiv 0$ , uma função  $f$  pode ser representada por:

$$f(x) = \sigma_{\text{net}(o)} \left( \sum_{p=0}^{d-1} f_p(x) + f_d(x) \right) \quad (11)$$

na qual os pesos dos blocos  $f_0, f_1, \dots, f_{d-1}$  são congelados ao adicionar-se o bloco de ordem  $d$ ,  $f_d$ .

Portanto, pode-se escrever  $f_x$  da seguinte forma:

$$f(x) = \sigma_{net(o)} \{ c_0 + c_1(\sigma_{net(i)} + \theta_1) + c_2(\sigma_{net(2)} + \theta_2)^2 + \dots + c_d(\sigma_{net(d)} + \theta_d)^d \} \quad (12)$$

na qual o vetor  $\mathbf{c} = [c_1, c_2, \dots, c_d]$  representa os pesos das unidades escondidas para as unidades de saída e a constante  $c_0$  está representando todos os blocos congelados até a ordem  $d-1$ . A Equação 9 é, portanto, idêntica a uma regressão logística não-linear quando a função de ativação utilizada é a sigmóide logística tradicional.

A potencialidade dessas redes está fundamentada nas características impostas pela regressão sigmóide não-linear na unidade de saída. Esse algoritmo construtivo tem como base um método de aprendizagem construtiva (também chamado de auto-organização na teoria das redes GMDH). O princípio construtivo pode ser formulado como segue: quando um modelo incrementa gradualmente sua complexidade, certos critérios, que são chamados de critérios de seleção ou função objetivo e que têm a propriedade de complemento externo, passam através de um valor mínimo global. O encontro deste mínimo indicará a existência de um modelo de complexidade ótima.

Portanto, usando-se essa arquitetura proposta uma função qualquer  $f(x)$  pode ser aproximada pelo uso direto de uma rede NSRBN por dividir  $f(x)$  em blocos homogêneos de ordem 1 até o bloco de ordem  $d$ . Dessa forma, a função total  $f(x)$  (polinômio completo) pode ser aproximada pela soma das funções parciais  $f_p(x)$ ,  $p=1, \dots, d$  (polinômios homogêneos), de maneira incremental podendo ser representada matematicamente por:

$$f(x) = g(\hat{f}_1(x) + \hat{f}_2(x) + \dots + \hat{f}_p(x) + \dots + f_d(x)) \quad (13)$$

O modelo proposto tem como princípio manter congelados os pesos de todos os blocos anteriores ao atualmente otimizado, de tal forma que o número de pesos a ser otimizado a cada grau de complexidade não aumente explosivamente, como no algoritmo combinatorial tradicional. O algoritmo de abordagem construtiva procede como se descreve a seguir. Seja  $p$  um passo do algoritmo para qual o bloco  $f_p(x)$  é adicionado à rede, então uma função qualquer  $f(x)$  é sucessivamente aproximada por:

$$f(x) = g(f_1(x)) \quad (14)$$

$$f(x) = g(\hat{f}_1(x) + f_2(x)) \quad (15)$$

...

...

...

$$f(x) = g(\hat{f}_1(x) + \hat{f}_2(x) + \dots + f_p(x)) \quad (16)$$

em que os pesos dos blocos  $\hat{f}_{p-1}(x)$  são congelados quando o bloco de ordem  $p$  é adicionado à rede e  $g$  representa uma função de ativação adequada para a unidade de saída.

## 4. EXEMPLOS DE APLICAÇÃO

Para ilustrar a utilização do sistema, duas aplicações foram realizadas, uma referente à previsão e outra a interpolação.

### 4.1. PREVISÃO

Os dados armazenados no sistema SIAQUA podem ter diversas utilidades no que diz respeito à área de previsão e em especial na área de ecologia. Diversos trabalhos têm sido publicados com a utilização de redes neurais na área de ecologia de tal modo que se decidiu incorporar esta ferramenta ao sistema, [4-8].

Nos reservatórios sob a operação da Chesf, a atividade de piscicultura intensiva em tanques-rede, foi implantada inicialmente na região de Paulo Afonso (BA), sobretudo nos reservatórios de Xingó e tem-se expandido para os demais reservatórios do rio São Francisco, [9]. Entretanto, embora essa atividade seja apenas uma das atividades antrópicas que podem

causar degradação na qualidade da água, recomenda-se monitoramento periódico da água em seus aspectos químicos físicos e biológicos, pois um aumento significativo de nutrientes (nitrogênio e fósforo), provenientes de transformações dos restos de alimento (ração) e resíduos metabólicos de peixes, podem levar a eutrofização desses reservatórios (que pode ser traduzida em um aumento percentual de clorofila). No exemplo apresentado consiste em realizar a previsão da clorofila-a com os dados disponíveis no SIAQUA para o reservatório de Xingo, situado no rio São Francisco, Figura 11. A clorofila-a serve como um indicador inicial das condições de eutrofização de um lago.

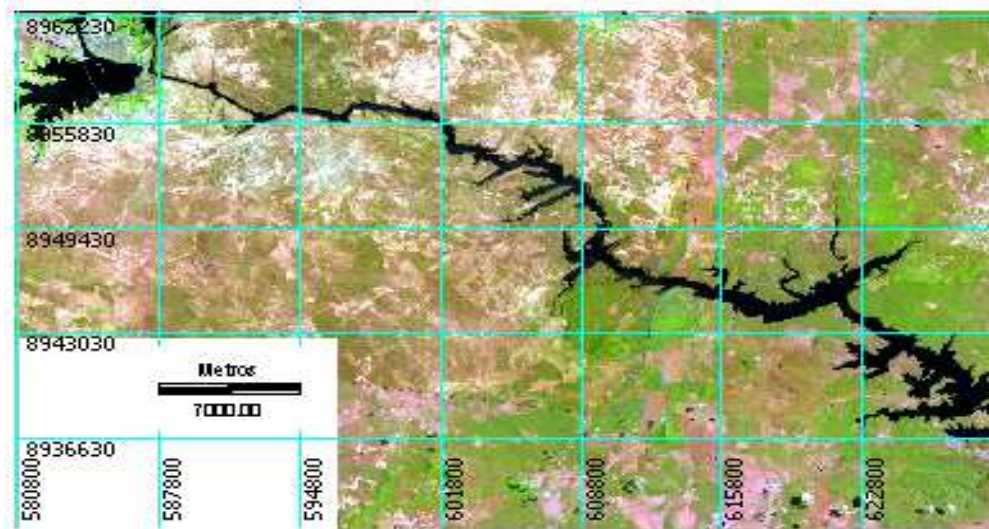


FIGURA 11: Imagem Landsat – TM da região do reservatório de Xingo

A coleta de dados ambientais primários foi efetuada no período de 11 a 17 de junho de 2002. Foram selecionados 79 (setenta e nove) pontos de amostragem, distribuídos nos diferentes trechos ao longo do reservatório, incluindo o corpo central e meandros. As Figuras 12, 13 e 14 apresentam a localização destes pontos de amostragem, [9].

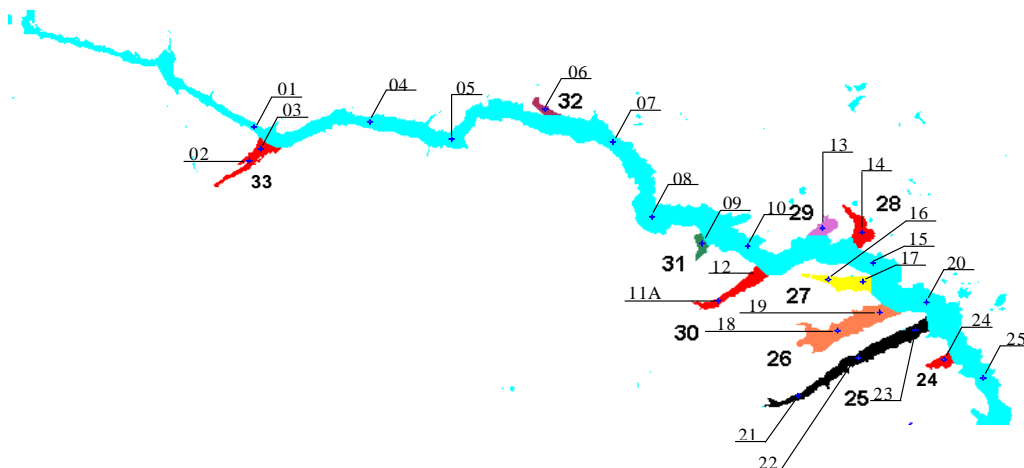


FIGURA 12: Mapa do reservatório de Xingo com indicação das estações de amostragem (1 a 25)

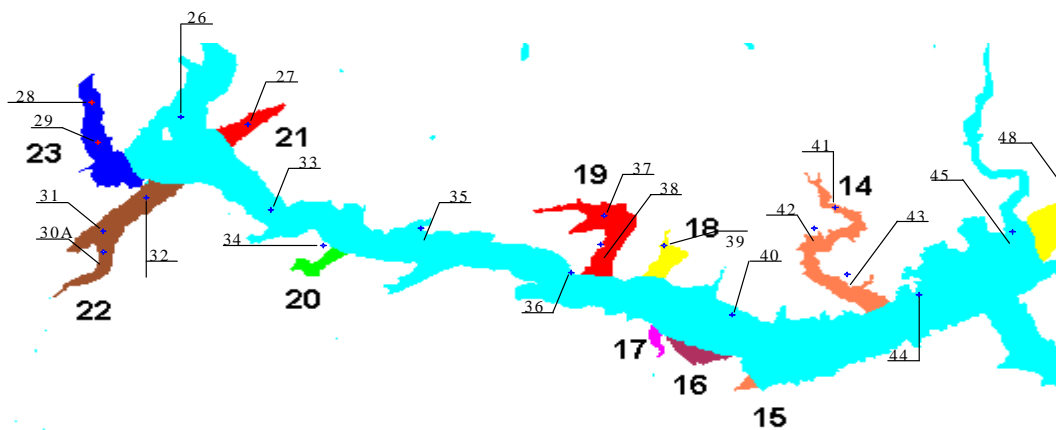


FIGURA 13: Mapa do reservatório de Xingo com indicação das estações de amostragem (26 a 49)

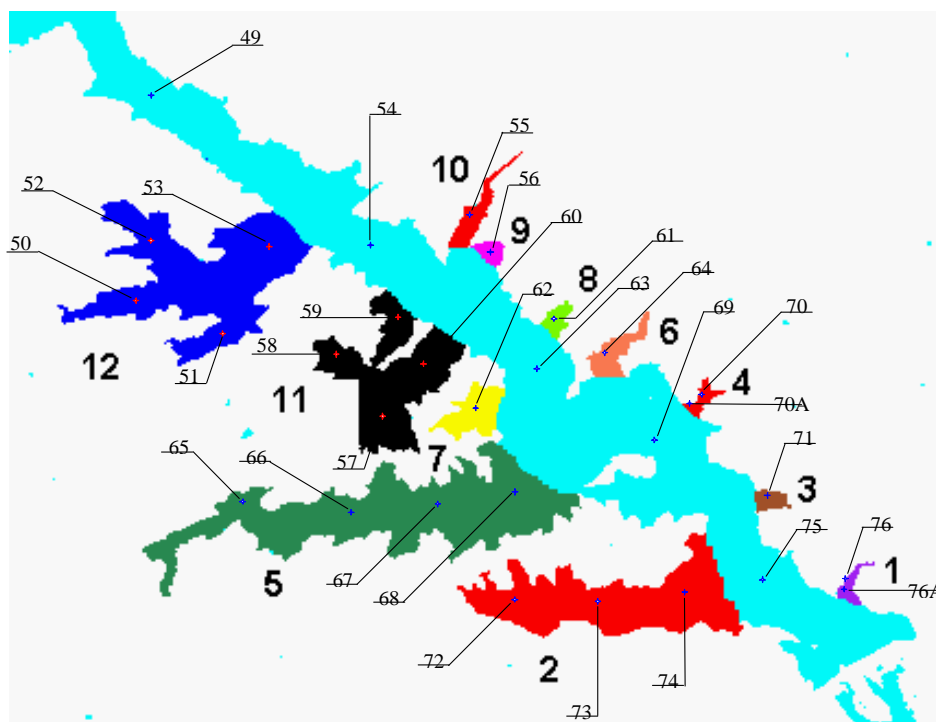


FIGURA 14: Mapa do reservatório de Xingo com indicação das estações de amostragem (49 a 76A)

Os dados armazenados no SIAQUA e aqui utilizados foram obtidos em uma campanha realizada para caracterização física, química e biológica da água, sendo efetuadas medidas das seguintes variáveis: temperatura, pH, turbidez, condutividade elétrica, radiação solar subaquática, transparência e oxigênio dissolvido. Amostras de água para determinação do material em suspensão, clorofila-a, nitrogênio e fósforo foram coletadas com garrafas de Van Dorn na superfície, no limite da zona eufótica e no fundo, em diferentes estações de amostragem. Estas foram distribuídas ao longo do corpo central do reservatório e seus principais meandros, incluindo todos aqueles anteriormente identificados como potencialmente exploráveis para a implantação de áreas aquícolas, totalizando 79 pontos amostrados.

As variáveis utilizadas como entradas para a rede neural para a previsão de clorofila-a, foram: pH, condutividade elétrica, temperatura, alcalinidade, nitrato, nitrito amônia, orto-

fosfato, fósforo total e fosfato total, totalizando 10 entradas para a rede. A saída da rede foi o valor medido da clorofila-a. O treinamento e validação cruzada da rede neural foram realizados com a utilização de 71 postos, deixando-se 8 postos para realização da avaliação de desempenho da rede neural artificial. Os coeficientes de correlação obtidos para treinamento e previsão, avaliação do desempenho, foram respectivamente 0,75 e 0,74.

#### **4.2. INTERPOLAÇÃO**

O objetivo da interpolação consiste em obter informações dos parâmetros de qualidade da água em locais onde não foram realizadas medições, e que sejam áreas em potencial para a atividade de tanques-rede.

O exemplo apresentado neste trabalho consiste em utilizar os dados dos postos amostrados e armazenados no SIAQUA, com informação de entrada para a rede neural de tal forma que a rede seja capaz depois de treinada fornecer informações em outros pontos desejados. Para a etapa de treinamento da rede neural e para a validação cruzada foram utilizados dados de 71 postos e os 8 postos restantes ficaram para a etapa de avaliação do desempenho da rede.

Como entrada para a rede neural foram utilizados as seguintes variáveis: coordenadas planas (UTM) dos pontos de amostragem do reservatório de Xingo, profundidade e a temperatura dos 3 postos mais próximos, o que totalizou 6 entradas para a rede neural. A saída desejada no caso foi a temperatura média no local de observação desejado. O treinamento e validação cruzada da rede neural foram realizados com a utilização de 71 postos, deixando-se 8 postos para a realização da avaliação de desempenho da rede neural. Os coeficientes de correlação obtidos para treinamento e previsão, avaliação do desempenho, foram respectivamente 0,71 e 0,70.

### **5. CONCLUSÕES E RECOMENDAÇÕES**

O Sistema de Informações Ambientais sobre Qualidade da Água vem no sentido de atender o objetivo de reunir todos os dados relevantes em um único banco de dados, e assim gerar informações concisas, confiáveis e que permitam sua utilização de forma fácil por meio de um sistema que faça interpolação e/ou previsão dos parâmetros de qualidade da água.

A utilização deste sistema vem facilitar a tomada de decisão tanto do que diz respeito à quantidade de postos que devem ser monitorados quanto às condições da qualidade da água dos reservatórios que tem o impacto direto nos demais usuários da água.

Este sistema uma vez desenvolvido ficará disponível na Internet de tal forma que os dados possam ser consultados e utilizados por outros usuários que necessitem de tais informações.

Neste trabalho foram apresentados dois exemplos de utilização deste sistema, os quais podem fazer um diferencial para a gestão ambiental dos reservatórios da Chesf. A técnica de redes neurais construtivas apresentou-se como uma ferramenta poderosa, haja visto sua facilidade de utilização por não especialistas, bem como pelos resultados obtidos com as variáveis trabalhadas.

O próximo passo consistirá na realização de um projeto piloto onde será realizada a instalação de equipamentos para coleta automática dos parâmetros de qualidade da água para um dos reservatórios da Chesf, onde a atividade de piscicultura em tanques-rede já seja realizada. Com os dados coletados espera-se realizar previsão com redes neurais, com uma antecedência de até 15 dias à frente, que permitam ao gestor ambiental fazer uma comunicação aos piscicultores para que eles tomem as providências necessárias no sentido de minimizar a mortalidade de peixes, como aquelas ocorridas em fevereiro de 2004 e de 2005.

Um dos reservatórios candidatos é o reservatório de Xingo, que possui 13 (treze) empreendimentos de piscicultura em tanques-rede, totalizando 1.593 unidades, com um volume total cultivável de 6.372 m<sup>3</sup>, povoado por, aproximadamente, 1.723.000 peixes.

A principal espécie cultivada é a tilápia nilótica (*Oreochromis niloticus*), com uma produção total estimada de 862 toneladas por ciclo médio de produção de 6 meses, ou cerca de 1.722 toneladas/ano [9].

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] VALENÇA, M.J.S. (1999). Análise e Projeto de Redes Neurais com Abordagem Construtiva para Modelagem de Sistemas Complexos. Tese de doutorado, UFPE, Brasil, 1999, 223 p.
- [2] VALENÇA, M.J.S. (2005). Aplicando Redes Neurais: um guia completo. Livro Rápido, Olinda – PE, 284 p.
- [3] M.J.S. Valença, T.B. Ludermir, “Selforganization neurons blocks networks”, ICCIMA, IEEE, pages 60-64, New Delhi, Índia, Sep. 1999.
- [4] RECKNAGEL, F.; FRENCH, M.; HARKONEN, P.; YABUNAKA, K.I. (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.* 96, 11–28.
- [5] WALTER M.; RECKNAGEL F.; CARPENTER C.; BORMANS M. (2001) Predicting eutrophication effects in the Burrinjuck Reservoir (Australia) by means of the deterministic model.
- [6] CHEN, Q.; MYNETT, A.E. (2003). Integration of data mining techniques with heuristic knowledge in a fuzzy logic modelling of eutrophication in Taihu Lake. *Ecological Modelling*, 162: 55-67.
- [7] JEONG K-S.; RECKNAGEL F.; JOO G-J. (2005). Prediction and elucidation of population dynamics of the bluegreen algae *Microcystis aeruginosa* and the diatom *Stephanodiscus hantzschii* in the Nakdong riverreservoir system (South Korea) by a recurrent Artificial Neural Network. In: Recknagel F (ed) *Ecological Informatics. Scope, Techniques and Applications*. 2nd Edition. Springer-Verlag, Heidelberg, New York, 255-273.
- [8] RECKNAGEL, F.; KIM, B.; WELK, A. (2005). Artificial Neural Network Approach to Unravel and Forecast Algal Population Dynamics of Two Lakes Different in Morphometry and Eutrophication. In: Recknagel, F. (2005). *Ecological Informatics*. 2nd Edition. Springer-Verlag. New York , 1-485.
- [9] COMPANHIA HIDRO ELÉTRICA DO SÃO FRANCISCO – Chesf. Relatório Técnico sobre o Redimensionamento da Capacidade de Produção em Tanques-rede do Reservatório de Xingo, realizado pela Fundação Apolônio Salles de Desenvolvimento Educacional – FADURPE, Recife, 2002.