

# **PREVISÃO DE SÉRIES DE VAZÃO DO RIO SÃO FRANCISCO UTILIZANDO DADOS HISTÓRICOS DA TEMPERATURA DA SUPERFÍCIE DO MAR**

**Trabalho de Conclusão de Curso**

**Engenharia da Computação**

**Robson David Montenegro**  
**Orientador: Prof. Mêuser Jorge Valença**

**ROBSON DAVID MONTENEGRO**

**PREVISÃO DE SÉRIES DE VAZÃO  
DO RIO SÃO FRANCISCO UTILIZANDO  
DADOS HISTÓRICOS DA  
TEMPERATURA DA SUPERFÍCIE DO  
MAR**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia da Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

**Recife, novembro de 2009**

*À minha família.*

# Agradecimentos

À todos.

*(...) Quando dizemos que alguma coisa é infinita, queremos apenas dizer que não somos capazes de conceber os limites e fronteiras da coisa designada, não tendo concepção da coisa, mas de nossa própria incapacidade (...)*

- THOMAS HOBBS, 1588-1679, *Leviatã*

# Resumo

A principal fonte de energia do Brasil é a força das águas. As hidroelétricas executam o trabalho de transformar a força motriz hídrica em energia elétrica. Contudo, o planejamento da produção de energia nas usinas tem grande dependência da vazão afluyente e da quantidade de água presente em seus reservatórios. Na procura de um melhor planejamento, órgãos que formam o sistema energético brasileiro implantaram um conjunto de modelos estatísticos para previsão de vazão. Entretanto, modelos estatísticos não são técnicas mais adequadas para tratar dados com características não lineares como as séries de vazão. Portanto, este trabalho propõe a utilização de Redes Neurais Artificiais, modelos não lineares, utilizando séries históricas das Temperaturas da Superfície do Mar, como potencial melhora no desempenho de previsão. Visto que, a Temperatura da Superfície do Mar controla a evaporação das águas oceânicas à atmosfera, relacionando-a com o volume de água futuro proveniente das chuvas.

# Abstract

The main source of energy in Brazil is the power of water. The Hydroelectrics do the work of transforming the driving force water into electrical energy. However, the planning of energy production in hydroelectrics is strongly dependent on the inflow and the amount of water in their reservoirs. For a better planning, Brazilian's energy system responsables implemented a set of statistical models to forecast flow. Nevertheless, statistical models are most appropriate techniques to handle data with non-linear characteristics such as flow time series. Therefore, this monograph proposes the use of Artificial Neural Networks, which are nonlinear models, using historical series of sea surface temperatures, as a potential improvement in performance of forecast flow. Since the sea surface temperature controls the evaporation of ocean water to the atmosphere, relating to the future water volume from the rain.

# Sumário

<b>Resumo</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>Sumário</b>	<b>viii</b>
<b>Índice de Figuras</b>	<b>x</b>
<b>Índice de Tabelas</b>	<b>xi</b>
<b>Tabela de Símbolos e Siglas</b>	<b>xii</b>
<b>Capítulo 1 Introdução</b>	<b>13</b>
3.1 Caracterização do Problema	14
3.2 Objetivos	14
3.2.1 Objetivos Específicos	15
3.3 Estrutura da monografia	15
<b>Capítulo 2 Fundamentação Teórica</b>	<b>17</b>
4.1 Sistema Conexionista	17
4.1.1 <i>MultiLayer Perceptron</i>	18
4.1.2 <i>Backpropagation</i>	19
4.2 Coeficiente de correlação de postos de Spearman	20
4.3 Teste t	21
<b>Capítulo 3 Metodologia</b>	<b>22</b>
3.1 Aquisição dos dados	22
3.1.1 Vazão	22
3.1.1 Temperatura da Superfície do Mar	23
	viii



3.2	Definição de Regiões Relevantes	25
3.2.1	Cálculo do Coeficiente de Correlação de Postos de Spearman	26
3.2.2	Definição dos pontos importantes	26
3.2.3	Cálculo das Defasagens	28
3.3	Previsão de Vazão	28
3.3.1	Topologia MLP	28
3.3.2	Definição dos dados de entrada	29
3.3.3	Normalização dos Dados de Entrada	30
3.3.4	Execução MLP	30
<b>Capítulo 4 Resultados e Discussão</b>		<b>32</b>
4.1	Previsão de Vazão	32
4.1.1	Sobradinho	32
4.1.2	Três Marias	34
<b>Capítulo 5 Conclusão</b>		<b>37</b>
5.1	Contribuições	37
<b>Bibliografia</b>		<b>38</b>
<b>Apêndice A Quantidade de pontos importantes pelo ‘Teste t’</b>		<b>40</b>
<b>Anexo A Tabela de Valores Seleccionados para o Teste t</b>		<b>42</b>

# Índice de Figuras

<b>Figura 1.</b>	Arquitetura MLP com três camadas: entrada, intermediária e saída.....	18
<b>Figura 2.</b>	Função com mínimo local e mínimo global em destaque.....	19
<b>Figura 3.</b>	Fluxo do <i>backpropagation</i> : cálculo da saída ( <i>forward</i> ) e atualização dos pesos das conexões ( <i>backward</i> ).....	19
<b>Figura 4.</b>	Superfície terrestre dividida em células 2º por 2º.....	23
<b>Figura 5.</b>	Anomalias anuais do período 1880-2008 com intervalo de confiança de 95% na área azul (NOAA , 2009).....	25
<b>Figura 6.</b>	Séries temporais ocorrida e previstas para usina de Sobradinho. ....	34
<b>Figura 7.</b>	Séries temporais ocorrida e previstas para usina de Três Marias.....	36

# Índice de Tabelas

<b>Tabela 1.</b>	Divisão do ERSST em 16 períodos.....	24
<b>Tabela 2.</b>	Pontos importantes após aplicação do ‘Teste t’.....	27
<b>Tabela 3.</b>	Número de unidades escondidas para cada Usina.....	29
<b>Tabela 4.</b>	Parâmetro da MLP.....	30
<b>Tabela 5.</b>	Resultado de Previsão Vazão-Vazão para Usina de Sobradinho.....	33
<b>Tabela 6.</b>	Resultado de Previsão TSM-Vazão para Usina de Sobradinho.....	33
<b>Tabela 7.</b>	Resultado de Previsão do Tipo Vazão-Vazão para Usina de Três Marias. 34	
<b>Tabela 8.</b>	Resultado de Previsão do Tipo TSM-Vazão para Usina de Três Marias. 35	

# Tabela de Símbolos e Siglas

IPH – Instituto de Pesquisas Hidráulicas

UFRGS – Universidade Federal do Rio Grande do Sul

TSM – Temperatura da Superfície do Mar

RNA – Rede Neural Artificial

MLP – MultiLayer Perceptron

ONS – Operador Nacional do Sistema Elétrico

ERSST – Extended Reconstruction of Historical Sea Surface Temperature

PODAAC – Physical Oceanography Distributed Active Archive Center

ASCII – American Standard Code for Information Interchange

AVHRR – Advanced Very High Resolution Radiometer

NOAA – National Oceanic and Atmospheric Administration

ICOADS – International Comprehensive Ocean-Atmosphere

EPMA – Erro Percentual Médio Absoluto

EMA – Erro Médio Absoluto

EP – Erro Padrão

SSE – Soma dos Erros Quadráticos

# Capítulo 1

## Introdução

O Brasil tem 77,2% de sua energia elétrica proveniente de hidroelétricas (OECD/IEA, 2006). Esta energia depende fortemente da regularização da vazão e na disponibilidade hídrica dos reservatórios de usinas hidroelétricas instaladas no país. O sistema de produção hídrica brasileira sofreu grandes melhoras nos últimos anos, mas depende do conhecimento prévio do volume de água disponível para realizar o planejamento das usinas.

Portanto, a possibilidade de prever a vazão afluente significa obter um planejamento energético para que se possam atingir os melhores níveis de produção de energia, além de previsão de inundações, da umidade do solo, dos níveis de navegação das vias fluviais e da irrigação (Tucci, Clarke, *et al.*, 2002). Para tanto, foi desenvolvido, pelos órgãos que formam o sistema energético brasileiro, um conjunto de modelos para: previsão, operação e planejamento energético. Estes modelos de previsão pertencem à classe de modelos estocásticos que utilizam técnicas estatísticas e empíricas para explorar o conhecimento passado para prever o futuro (Tucci, Clarke, *et al.*, 2004).

Um estudo realizado pelo IPH-UFRGS (Tucci, Clarke, *et al.*, 2004) utilizou modelos de regressão linear múltipla para estimar, a partir de valores de variáveis climatológicas em meses anteriores, a vazão futura da bacia do Rio São Francisco. Uma das variáveis usadas foi a temperatura da superfície do mar (TSM). Esse estudo sugere a existência de relação dessa variável macro-climática e os valores de vazão da bacia. Esta hipótese parte do princípio de que parte da precipitação da bacia do São Francisco é de origem oceânica, relacionando a TSM, que controla a evaporação do oceano à atmosfera, ao volume de água futuro.

Entretanto, modelos estatísticos não são técnicas mais adequadas para tratar dados com características inerentemente não lineares como as séries de vazão (Chatfield, 1991). Contudo, Redes Neurais Artificiais são modelos formados por

elementos não lineares (neurônios artificiais) capazes de aprender a partir de experiências passadas. Estudos feitos utilizando Redes Neurais para previsão de séries temporais têm obtido resultados expressivos (Neto, Coelho, *et al.*, 2005).

A proposta da utilização de variáveis macro-climáticas se mostra bastante relevante (Matsuo, Cardoso, *et al.*, 2002). É possível vislumbrar resultados positivos no uso de técnicas de Redes Neurais para previsão de vazão da bacia do Rio São Francisco utilizando dados de TSM.

### **3.1 Caracterização do Problema**

Em períodos úmidos, quando a afluência natural é maior que a demanda, as usinas hidroelétricas aproveitam para aumentar suas reservas de água. Para que, em períodos secos, quando a demanda é maior que a afluência, essa escassez seja compensada (Valença, Ludermir e Vasconcelos, 2000). Por isso, é muito importante que se saiba com antecedência a quantidade de volume de água a ser retirado do reservatório, a cada instante de operação. Além disso, a concentração de gases e o aquecimento global têm mudado os cenários e o planejamento de abastecimento de água, que dependem da quantidade de água disponível nos rios.

O período seco compreende os meses de maio a novembro, e o período úmido de dezembro a abril. Saber com antecedência se o ano será chuvoso ou não, é conseguir prever qual serão as vazões para o período úmido. Isto é, possibilitar a melhora da previsão de vazão visando melhorar o processo decisório de operação das usinas hidroelétricas.

### **3.2 Objetivos**

Verificar a influência da Temperatura da Superfície do Mar na previsão de vazão do rio São Francisco em médio e longo prazo.

### 3.2.1 Objetivos Específicos

- Verificar quais áreas do oceano têm maior influência na vazão do rio São Francisco a partir das séries de TSM;
- Comparar a previsão de vazão no período úmido que utilizam TSM e que não utilizam TSM.

## 3.3 Estrutura da monografia

Este trabalho é dividido em cinco capítulos, resumidos à seguir:

### **Capítulo 1: Introdução**

Contém o texto introdutório do trabalho, abordando a problemática e os objetivos.

### **Capítulo 2: Fundamentação Teórica**

Contempla o fundamento das técnicas utilizadas no processo metodológico do trabalho. Para tal, aborda os Sistemas Conexionistas, detalhando a técnica MultiLayer Perceptron, e os métodos estatísticos utilizados.

### **Capítulo 3: Metodologia**

Explica a maneira na qual foram obtidos os dados e descreve, de maneira detalhada, os métodos para o processamento dos mesmos. Mostra como foi realizada a escolha dos parâmetros da rede MLP assim como sua utilização para a execução da previsão.

### **Capítulo 4: Resultados e Discussão**

Apresenta os resultados obtidos pela metodologia detalhada no **Capítulo 3**. Contém a discussão e análise do impacto do uso dos dados de temperatura da superfície do mar na previsão de vazão.

## **Capítulo 5: Conclusão**

Contém as principais contribuições desta monografia e as conclusões obtidas dos resultados.



## Capítulo 2

# Fundamentação Teórica

Este trabalho se destina a avaliar o desempenho de técnicas de computação inteligente na previsão de séries temporais utilizando dados globais de TSM. Os fundamentos que integram esse estudo são abordados à seguir. Na seção **4.1** os sistemas conexionistas são introduzidos e a não-linear escolhida, MultiLayer Perceptron, é detalhada. As seções **4.2** e **4.3** abordam os métodos estatísticos utilizados para selecionar as séries de TSM mais relevantes: coeficiente de correlação de Spearman e Teste t, respectivamente.

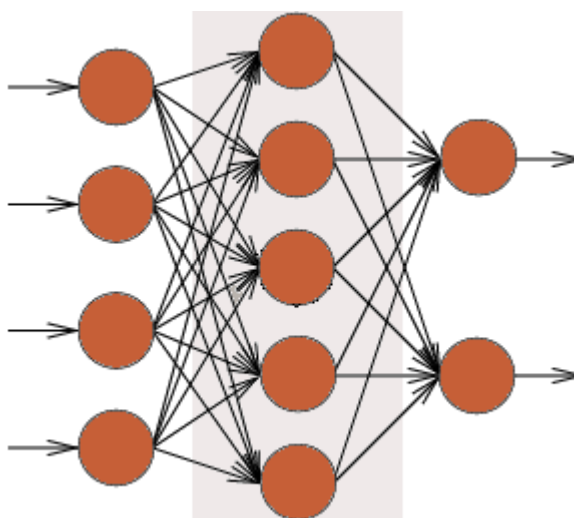
### 4.1 Sistema Conexionista

Rede Neural Artificial (RNA), ou sistema conexionista, é um sistema computacional baseado na estrutura física do cérebro humano. Uma RNA é formada por unidades de processamento (nodos) que são responsáveis por calcular determinadas funções matemáticas (normalmente não-lineares). Essas unidades podem estar organizadas em uma ou mais camadas e são interligadas por um grande número de conexões. Na maioria dos modelos, o conhecimento está associado aos pesos de suas conexões, que são ponderados a entrada recebida por cada neurônio da rede. O maior atrativo das RNAs é sua capacidade de aprender através de exemplos e generalizar a informação aprendida (Braga, Ludermir, & Carvalho, 2000).

Em aplicações de Redes Neurais artificiais, o modelo mais recorrente é o *MultiLayer Perceptron* (MLP) treinado com o algoritmo *backpropagation*. Sua popularidade pode ser explicada devido sua capacidade de aproximar uma função arbitrária não-linear em muitas variáveis. Neste trabalho, o modelo de Rede Neural escolhido para realizar a previsão de vazão foi o MLP, abordado a seguir.

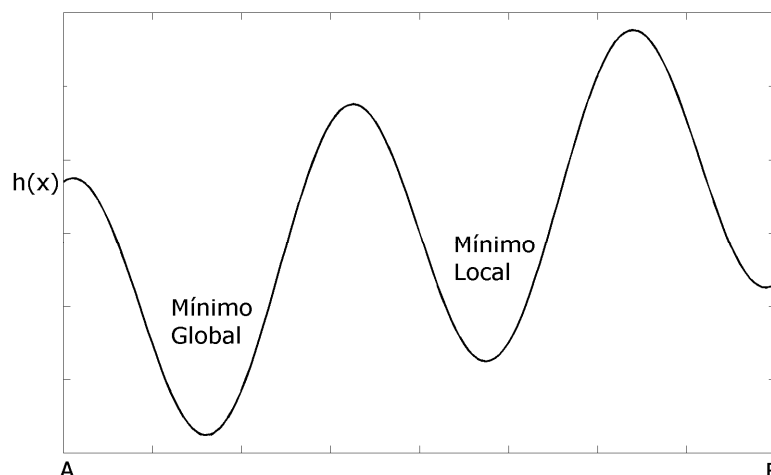
#### 4.1.1 *MultiLayer Perceptron*

A rede MLP é baseada no modelo perceptron proposto por Frank Rosenblatt em 1958 (Rosenblatt, 1958). Por apresentar apenas uma camada, a utilização do perceptron é viável apenas para problemas linearmente separáveis. Entretanto, o uso de pelo menos uma camada intermediária (escondida), permite a uma rede neural a aproximação de qualquer função (Cybenko, 1989). A **Figura 1** mostra um exemplo de uma rede perceptron multicamadas, tipicamente formada por uma camada de entrada, uma camada intermediária e uma camada de saída.



**Figura 1.** Arquitetura MLP com três camadas: entrada, intermediária e saída.

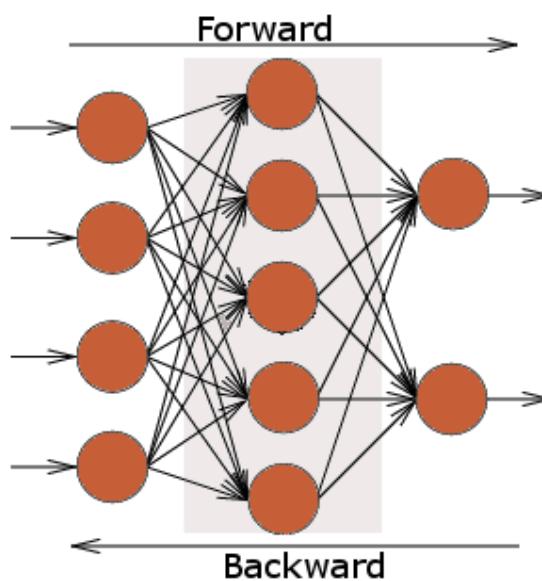
Apesar de capacidade de aproximação a qualquer função, uma rede MLP não garante a implementação da função. Dependendo da distribuição dos dados, a rede pode convergir a um mínimo local ou pode demorar mais para convergir (Braga, Ludermir, & Carvalho, 2000). A **Figura 2** mostra a função  $h(x)$  representando graficamente um mínimo local e um mínimo global.



**Figura 2.** Função com mínimo local e mínimo global em destaque.

#### 4.1.2 *Backpropagation*

*Backpropagation* é o algoritmo de aprendizado mais conhecido para redes MLP. Esta é uma técnica de treinamento supervisionado, ou seja, o par entrada-saída é conhecido, e pode ser dividida em duas fases: *forward* e *backward*. A fase *forward* calcula o valor de saída da rede para o padrão de entrada e a fase *backward* atualiza os pesos das conexões dos nodos da rede a partir do erro de saída (ver **Figura 3**).



**Figura 3.** Fluxo do *backpropagation*: cálculo da saída (*forward*) e atualização dos pesos das conexões (*backward*).

O algoritmo *backpropagation* é a generalização da regra delta de Widrow e Hoff (Widrow & Hoff, 1960). A proposta do método é definir os erros dos nodos das camadas intermediárias para que seja realizado o ajuste dos pesos. O ajuste dos pesos utiliza a técnica do gradiente descendente e a função de custo a ser minimizada é:

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (1)$$

onde  $E$  é o erro total,  $p$  é o número de padrões,  $k$  é o número de unidades de saída,  $d_i^p$  é a  $i$ -ésima saída desejadas e  $y_i$  é a  $i$ -ésima saída gerada pela rede. O erro resultante caracteriza a diferença entre a saída gerada pela rede e as saídas reais.

## 4.2 Coeficiente de correlação de postos de Spearman

O coeficiente de correlação de postos de Spearman é uma medida não-paramétrica de correlação proposto por Charles Spearman. O cálculo de correlação não faz suposição sobre as distribuições de freqüência das variáveis analisadas. O valor do coeficiente varia no intervalo  $[-1,1]$ , portanto, quanto mais próximo o coeficiente estiver desses extremos, maior a associação entre as variáveis. O coeficiente  $\rho$ , é dado pela fórmula:

$$\rho = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)} \quad (2)$$

onde  $x_i - y_i$  representa a diferença entre os postos, correspondentes a  $X_i$  e  $Y_i$ , respectivamente, e  $n$  é o número de valores de cada distribuição, que deve ter o mesmo valor para as duas distribuições. Os postos, também conhecidos como *ranks*, são as posições relativas aos valores da distribuição quando ordenados.

## 4.3 Teste t

Teste t é um teste de hipótese cuja avaliação é baseada na distribuição 't de Student' quando a Hipótese Nula é assumida. A distribuição 't de Student' é uma distribuição de probabilidade teórica, simétrica, campaniforme e semelhante a curva normal, porém com caudas mais largas e usada para um pequeno número de amostras. Em estatística, a Hipótese Nula é utilizada quando se quer confrontar fatos: uma hipótese é assumida até que provas estatísticas mostrem o contrário.

Este trabalho faz uso do 'Teste t' como forma de avaliação da importância do coeficiente dos postos de Spearman. O teste toma por base o valor da estatística 't de Student' dado pelas fórmulas:

$$t = \frac{r\sqrt{gl}}{\sqrt{1-r^2}} \quad (3)$$

$$gl = N - 2 \quad (4)$$

onde  $r$  é o Coeficiente de Correlação dos Postos de Spearman (2) e  $N$  é o número de amostras na distribuição. O  $gl$  (grau de liberdade) indica a proximidade da distribuição à distribuição normal, portanto, quanto maior o número de amostras, maior proximidade.

O teste utiliza a Tabela de Valores Seleccionados (*Anexo A*) como base de análise da importância do valor 't de Student' calculado por (3). O valor  $t_c$  da tabela é escolhido, de acordo com nível de confiabilidade e número de amostras. O valor  $t$  calculado é então submetido ao teste:

Se  $|t| > |t_c|$  então a hipótese assumida é rejeitada.

Com este teste é possível inferir, considerando o grau de confiabilidade escolhido, se as distribuições utilizadas no cálculo do coeficiente dos postos de Spearman difere significativamente de 0 (zero).

# Capítulo 3

## Metodologia

Este trabalho se destina a avaliar o impacto do uso da Temperatura da Superfície do mar no desempenho de previsão de vazão para o modelo de rede *MultiLayer Perceptron*. Os dados de vazão são fornecidos pela ONS (Operador Nacional do Sistema Elétrico) e são referentes as barragens Três Marias e Sobradinho. A partir da aquisição dos dados de TSM, é possível estabelecer as séries históricas para cada região do globo terrestre. Devido ao grande volume de dados, faz-se conveniente a filtragem das regiões que apresentam as séries históricas mais relevantes, ou seja, com maior correlação com as séries de vazão em questão. As séries relevantes de TSM também são definidas para 12 tipos de defasagens (1 à 12 meses).

Após a definição dos pontos relevantes para cada defasagem de cada mês, é possível construir grande banco de arquivos que servirão de entrada a rede MLP. Esta tem sua topologia definida experimentalmente, baseada no melhor desempenho de previsão do tipo Vazão-Vazão. A metodologia usada neste trabalho é detalhada nas seções subseqüentes.

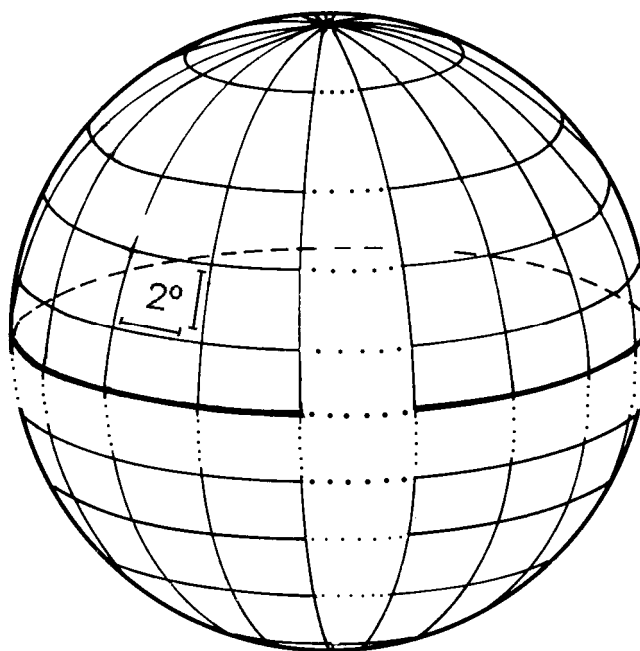
### 3.1 Aquisição dos dados

#### 3.1.0 Vazão

Os dados de Vazão são fornecidos pelo ONS e contemplam as vazões mensais, em  $m^3 / s$ , do período de janeiro de 1931 à dezembro de 2003. Neste trabalho são usadas as séries históricas das barragens de Três Marias e Sobradinho localizadas na bacia do Rio São Francisco no nordeste brasileiro.

### 3.1.1 Temperatura da Superfície do Mar

Os dados de TSM são extraídos do ERSST (Extended Reconstruction of Historical Sea Surface Temperature), um produto fornecido pelo PO.DAAC (Physical Oceanography Distributed Active Archive Center), e está disponível online em formato ASCII (PO.DAAC, 2009). Este produto consiste na temperatura média mensal de vários pontos da superfície terrestre. Cada ponto do globo é representado por uma célula de  $2^\circ$  por  $2^\circ$  de um grid esférico, como ilustra a **Figura 4**.



**Figura 4.** Superfície terrestre dividida em células  $2^\circ$  por  $2^\circ$ .

O grid é uma matriz com 180 linhas e 89 colunas, portanto, intuitivamente uma esfera com 180 meridianos e 89 paralelos formada por 16.020 células e suas respectivas temperaturas. Cada temperatura diz respeito a uma coordenada do tipo (latitude, longitude) com seus valores em graus Celsius. Temperaturas com valores abaixo dos  $-1,8^\circ$  são arredondadas para  $-1,8^\circ$  e, a regiões de terra, o valor  $-99,99^\circ$  é atribuído.

O ERSST é dividido em 16 arquivos, um arquivo por década (ver **Tabela 1**), referentes ao período de janeiro de 1854 a setembro de 2009. Cada mês desse período é representado por um grid, onde cada célula contém a temperatura média da superfície do mar.

**Tabela 1.** Divisão do ERSST em 16 períodos.

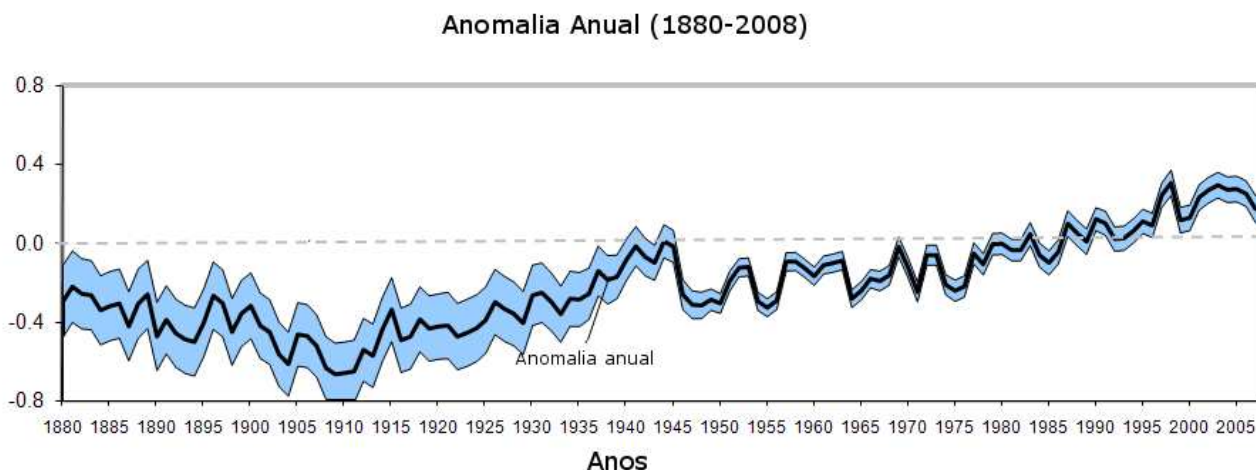
#	Período	Anos	Meses
1	Janeiro de 1854 – Dezembro de 1859	6	72
2	Janeiro de 1860 – Dezembro de 1869	10	120
3	Janeiro de 1870 – Dezembro de 1879	10	120
4	Janeiro de 1880 – Dezembro de 1889	10	120
5	Janeiro de 1890 – Dezembro de 1899	10	120
6	Janeiro de 1900 – Dezembro de 1909	10	120
7	Janeiro de 1910 – Dezembro de 1919	10	120
8	Janeiro de 1920 – Dezembro de 1929	10	120
9	Janeiro de 1930 – Dezembro de 1939	10	120
10	Janeiro de 1940 – Dezembro de 1949	10	120
11	Janeiro de 1950 – Dezembro de 1959	10	120
12	Janeiro de 1960 – Dezembro de 1969	10	120
13	Janeiro de 1970 – Dezembro de 1979	10	120
14	Janeiro de 1980 – Dezembro de 1989	10	120
15	Janeiro de 1990 – Dezembro de 1999	10	120
16	Janeiro de 2000 – Setembro de 2009	10	117

As temperaturas presentes no ERSST são derivadas de análises *in-situ* e de coletas feitas pelo detector de radiação AVHRR (Advanced Very High Resolution Radiometer) localizados nos satélites NOAA (National Oceanic and Atmospheric Administration). Os três canais do AVHRR operam em banda infra-vermelha e são capazes de detectar a radiação de calor e, conseqüentemente, as temperaturas em terra, na superfície do mar, e as nuvem sob a superfície.

Smith e Reynolds (Smith e Reynolds, 2004) utilizaram dados de anomalias de temperatura extraídos do ICOADS (International Comprehensive Ocean-Atmosphere) para reconstrução dos dados históricos da temperatura da superfície



do mar. As anomalias de temperatura oferecem uma visão geral da variação da temperatura média global comparadas a um valor de referência. A **Figura 5** mostra os valores de anomalias de temperaturas estipulados ao longo do tempo.



**Figura 5.** Anomalias anuais do período 1880-2008 com intervalo de confiança de 95% na área azul (NOAA , 2009).

Os valores de anomalias de temperatura são derivados de observações de elementos marítimos como navios e bóias, no decorrer do tempo. Devido a pequena quantidade de dados disponíveis para o período anterior a 1880, os dados desse período não são aproveitados, pelo seu baixo nível de confiança. A partir de 1950, a incerteza dos valores das medições diminui na maioria das regiões (Smith e Reynolds, 2004) apresentando limites de intervalos de confiança menores. Por isso, os valores de TSM utilizados neste trabalho são do período de janeiro de 1950 à dezembro de 2003.

## 3.2 Definição de Regiões Relevantes

Após a aquisição dos dados de TSM, é possível definir as séries históricas para cada um dos 16.020 pontos do globo nos 54 anos considerados. Este grande volume de dados torna sua utilização inviável como entrada para rede MLP. Nesse sentido, é necessário a filtragem dos pontos mais relevantes para a previsão, ou seja, identificação dos pontos no globo que apresentam maior correlação com as

séries de vazão em questão. Neste trabalho são construídas séries para cada um dos 12 meses do ano separadamente.

O método de correlação de postos de Spearman é utilizado para calcular os coeficientes de correlação entre os 16.020 pontos e as duas séries de vazão consideradas neste trabalho, para cada mês. O critério de aceitação do ponto é realizado pelo 'Teste t'. Este teste de hipótese é capaz de avaliar a importância do coeficiente de Spearman, inferindo se aquele ponto será descartado ou não. A seguir, as subseções **3.2.1** e **3.2.2** mostram como os cálculos de correlação e os testes estatísticos são realizados.

### **3.2.1 Cálculo do Coeficiente de Correlação de Postos de Spearman**

O primeiro passo para definição dos coeficientes de correlação é a separação das séries que serão objeto de estudo. Este trabalho se atem ao estudo da influencia da TSM nas vazões de barragens do Rio São Francisco, portanto, pontos de terra são descartados. Esta separação pode ser facilmente realizada pois os pontos de terra, no ERSST, tem valores de temperatura iguais a -99,99°. Após a separação, restam 11.074 pontos, dos 16.020 pontos iniciais.

Após a determinação das séries referentes aos pontos do mar, é feito o cálculo dos coeficientes de correlação entre os 11.074 pontos e cada uma das séries de vazão. Assim como realizado em (Tucci, Clarke, *et al.*, 2004), as correlações são calculadas para cada mês, separadamente. Por exemplo, para o mês de janeiro, foi construída uma série de 54 anos de todas as vazões de janeiro para o período 1950-2003, e para cada um dos 11.074 pontos foi construída uma série semelhante, de 54 valores, das TSM de janeiro. As condições necessárias para o uso de testes rigorosos de significância não são satisfeitas neste contexto, mas como uma indicação da importância (ou não) de um dado valor do coeficiente de correlação de Spearman, pode-se usar o 'Teste t'.

### **3.2.2 Definição dos pontos importantes**

O 'Teste t' foi utilizado para identificar as correlações mais importantes. Para uma confiabilidade de 95% e com, aproximadamente, 50 amostras, temos o valor do

parâmetro  $t_c = 2,009$ . Este valor é adquirido da Tabela de Valores Seleccionados (Anexo A). A utilização do 'Teste t' exige que o coeficiente de correlação utilizado, obedeça a distribuição 't de Student', para tanto, a seguinte transformação é efetuada:

$$t = \frac{s\sqrt{N-2}}{\sqrt{1-s^2}} \quad (5)$$

onde  $N$  é o número de amostras das séries e  $s$  é o coeficiente de correlação de Spearman. Portanto, a correlação é mais importante se  $|t| > |t_c|$ , ou seja, se  $|t| > 2,009$ . A **Tabela 2** mostra a quantidade de pontos que restam após a aplicação do 'Teste t'.

**Tabela 2.** Pontos importantes após aplicação do 'Teste t'.

Mês	Quantidade de Pontos Relevantes	
	Sobradinho	Três Marias
Janeiro	1.714	1.348
Fevereiro	1.180	944
Março	795	291
Maio	1.259	1.291
Abril	1.722	1.154
Junho	2.227	1.686
Julho	2.348	1.277
Agosto	2.444	1.641
Setembro	2.774	1.532
Outubro	2.554	1.520

Novembro	1.628	537
Dezembro	1.017	746

Como é possível perceber, o número de pontos relevantes continua alto mesmo após a aplicação do 'Teste t'. Em virtude disso, o mês que obteve o menor número de pontos relevantes foi selecionado e considerado como limite superior. Isto é, o número máximo de pontos considerados para previsão será o menor número de pontos relevantes achado.

### 3.2.3 Cálculo das Defasagens

O cálculo de correlação e 'Teste t' também é realizado para 12 defasagens (1 à 12) para cada um dos 12 meses. Neste trabalho, as séries de vazão foram adiantadas em relação as séries de TSM mês a mês, e os coeficientes de correlação são coletados. Portanto, o cálculo de correlação e o 'Teste t' são realizados para cada mês e para cada uma das suas 12 defasagens. No **Apêndice A** é possível encontrar todos os resultados para todas as defasagens do 'Teste t'.

## 3.3 Previsão de Vazão

Após identificar todos os pontos relevantes para as 12 defasagens de todos os meses, é possível selecionar os parâmetros de entrada para a rede MLP. Antes disso, porém, é necessário definir qual topologia de rede será usada (subseção 3.3.1). Após a definição da topologia e dos arquivos de entrada (subseção 3.3.2) com dados normalizados (subseção 3.3.3) inicia-se a fase de treinamento, validação e testes da Rede Neural (subseção 3.3.4).

### 3.3.1 Topologia MLP

O número de unidades escondidas da rede MLP é definido experimentalmente. São realizadas previsões utilizando apenas dados de vazão, para que se possa coletar a topologia que obtêm o menor erro. A média aritmética de três erros, Erro Percentual

Médio Absoluto (Eq. 6), Erro Médio Absoluto (Eq. 7) e Erro Padrão (Eq. 8), é calculada para redes MLP com 1 à 100 unidades na camada escondida.

$$EPMA = \frac{100}{N} \left( \sum_{p=1}^N \frac{|Z_p - Z_o|}{Z_o} \right) \quad (6)$$

$$EMA = \frac{1}{N} \left( \sum_{p=1}^N |Z_p - Z_o| \right) \quad (7)$$

$$EP = \sqrt{\frac{1}{N} \left( \sum_{p=1}^N (Z_p - Z_o)^2 \right)} \quad (8)$$

onde  $N$  é o número de previsões,  $Z_p$  é a vazão prevista pela MLP e  $Z_o$  é a vazão ocorrida. Obtêm-se, assim, as topologias de menor média de erro: 40 unidades escondidas para Sobradinho e 53 unidades escondidas para Três Marias, como mostra a tabela abaixo.

**Tabela 3.** Número de unidades escondidas para cada Usina

Barragem	Unidades Escondidas	EPMA(%)	EMA	EP
Sobradinho	53	30,5	6,07	1051,55
Três Marias	40	38,84	4,89	435,31

### 3.3.2 Definição dos dados de entrada

O período objeto de estudo deste trabalho é Novembro, período de transição, e Dezembro à Abril período úmido, portanto, toma-se como entrada as vazões dos seis meses anteriores: Maio à Outubro. Dentre todas as defasagens de todos os meses, o menor número de pontos relevantes que se obtêm é para março com defasagem igual a 1 (ver **Apêndice A**), apenas 55 pontos obtiveram êxito no 'Teste t'. Portanto, para cada um dos meses que se quer prever, são tomados de 1 à 55 pontos, em ordem de 'importância'.

### 3.3.3 Normalização dos Dados de Entrada

Antes de iniciar a fase de treinamento da rede, os dados precisam estar normalizados, pois TSM e Vazão estão em grandezas diferentes. Para tanto, todos os valores de cada um dos parâmetros são divididos pelo maior valor daquele parâmetro no conjunto de dados, com segue:

$$v'_{p,i} = \frac{v_{p,i}}{V_p} \quad (9)$$

onde  $v'_{p,i}$  é valor normalizado do  $i$ -ésimo padrão para o parâmetro  $p$ ,  $v_{p,i}$  é o valor do  $i$ -ésimo padrão para o parâmetro  $p$ , e  $V_p$  é o maior valor do parâmetro  $p$  no conjunto de dados. Após a normalização todos os dados terão valores no intervalo  $[-1, 1]$ .

### 3.3.4 Execução MLP

A rede MLP tem sua topologia definida experimentalmente (3.3.2). Os demais parâmetros como: número de épocas, taxa de aprendizado, quantidade máxima de falhas na validação e a função de Erro, foram definidos baseado em trabalhos relacionados a previsão de séries temporais utilizando a rede MLP (Battagello, Cordeiro e Dal Pino). Os dados foram normalizados no intervalo  $[-1,1]$ , por isso, a função Tangente foi escolhida como função de ativação.

**Tabela 4.** Parâmetro da MLP.

<b>Épocas</b>	200
<b>Taxa de Aprendizado</b>	1%
<b>Quantidade máxima de falhas na validação</b>	5
<b>Função de Ativação</b>	Tangente
<b>Função de Erro</b>	SSE (Soma dos Erros Quadráticos)

Para cada um dos 1300 arquivos, 650 para Três Marias e 650 para Sobradinho, é utilizado a configuração descrita na **Tabela 4**. Os resultados são mostrados e discutidos no **Capítulo 4**.

# Capítulo 4

## Resultados e Discussão

Este trabalho tem como objetivo principal avaliar o desempenho de previsão de vazão utilizando Redes Neurais treinadas com dados de TSM. Para isso, foram definidos os pontos com maior correlação para a vazão de cada mês, para cada uma das usinas. Esses dados foram organizados em vários arquivos e usados para treinamento, validação e testes de uma Rede MLP. Neste capítulo são expostos os resultados da metodologia detalhada no **Capítulo 3**.

### 4.1 Previsão de Vazão

Para cada usina, são obtidos 650 conjuntos de erros resultantes das previsões efetuadas pela rede MLP. Estes erros são os objetos da análise comparativa entre as previsões que utilizam apenas dados de vazão para previsão e as que utilizam TSM. Todos os resultados apresentados nesta seção são derivados de uma mesma configuração de rede, definida nas subseções 3.3.4 e 3.3.1. Os resultados mais relevantes para as usinas de Sobradinho e Três Marias são apresentados nas subseções seguintes.

#### 4.1.1 Sobradinho

A **Tabela 5** mostra os erros obtidos utilizando apenas os dados de vazão para previsão, chamada aqui, de previsão Vazão-Vazão. Na **Tabela 6**, encontram-se os cinco melhores resultados para previsão utilizando as TSM como parâmetro de entrada, as previsões TSM-Vazão. É possível afirmar que o desempenho da previsão TSM-Vazão é melhor que a previsão Vazão-Vazão. O EPMA é menor para as cinco simulações, enquanto o EMA é menor para três. O EP não obteve melhora para nenhuma das cinco simulações.



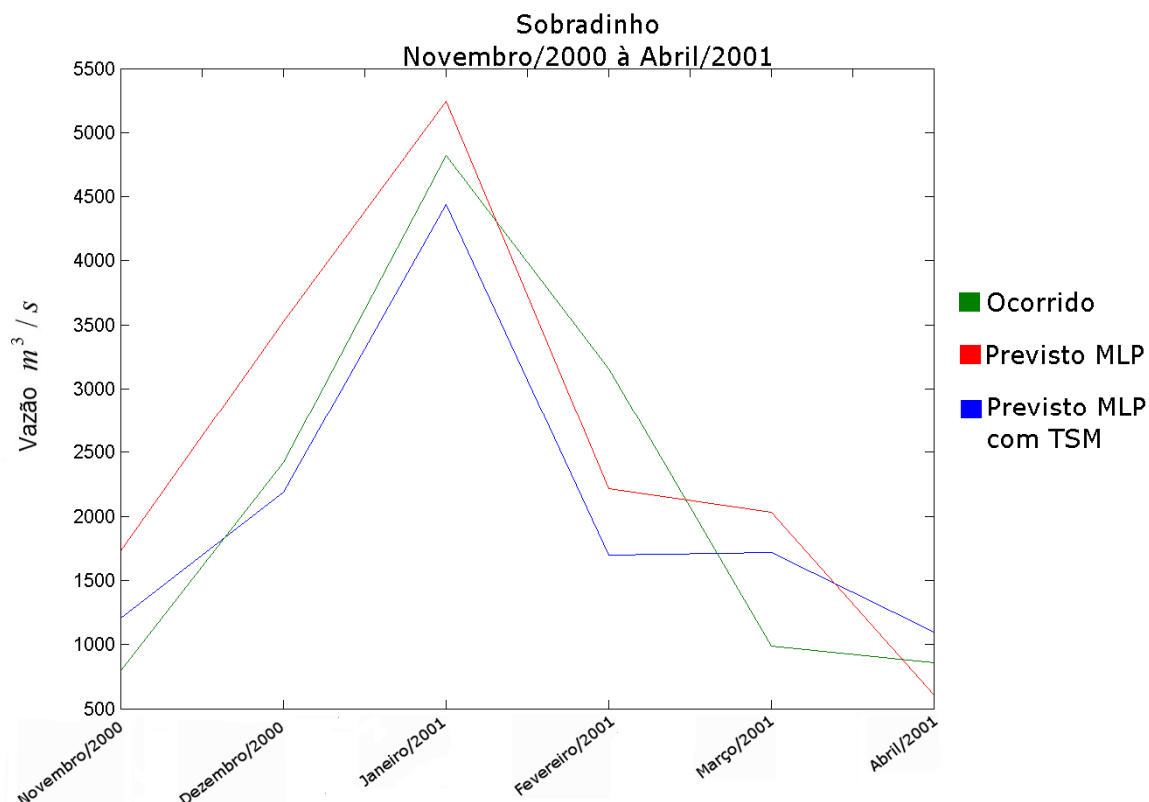
**Tabela 5.** Resultado de Previsão Vazão-Vazão para Usina de Sobradinho.

EPMA (%)	EMA	EP
30,5	6,07	1051,55

**Tabela 6.** Resultado de Previsão TSM-Vazão para Usina de Sobradinho.

Quantidade de Pontos Relevantes	Defasagem (meses)	EPMA (%)	EMA	EP
1	11	24,49	5,93	1153,95
4	5	26,34	4,66	1674,16
2	1	26,84	8,09	1239,17
1	2	26,94	8,27	1648,61
4	5	27,13	4,62	1173,12

É importante observar a quantidade de pontos relevantes para as melhores previsões (**Tabela 6**). Os bons resultados para a pequena quantidade de pontos relevantes mostram que informação em demasia pode causar uma degradação no desempenho da rede, possivelmente, porque existem informações conflitantes. A seguir, a **Figura 6** mostra as séries da vazão ocorrida, da vazão prevista pela rede treinada apenas com dados de vazão e prevista pela rede treinada com TSM, para o período úmido de 2000 à 2001.



**Figura 6.** Séries temporais ocorrida e previstas para usina de Sobradinho.

As três séries ilustradas acima mostram que há uma aproximação maior da previsão realizada pela rede treinada com as TSM ao ocorrido que a previsão realizada pela rede treinada apenas com dados de vazão. Fevereiro foi o único mês em que a rede treinada com TSM não obteve melhora.

#### 4.1.2 Três Marias

Os resultados para a previsão da usina de Três Marias são semelhantes as de Sobradinho. As **Tabelas 7** e **8** mostram, respectivamente, os resultados para previsão sem a utilização e com a utilização de TSM. O EPMA e o EMA são melhores para as cinco simulações, enquanto o EP é em apenas uma.

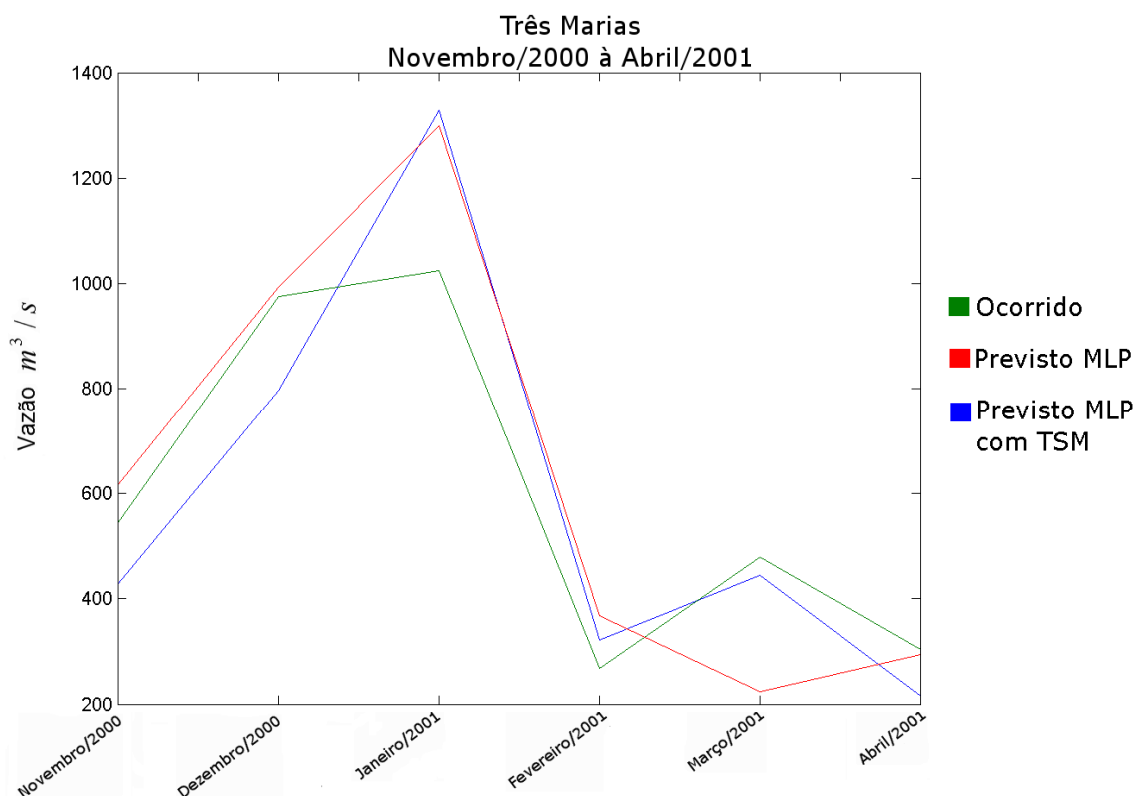
**Tabela 7.** Resultado de Previsão do Tipo Vazão-Vazão para Usina de Três Marias.

EPMA (%)	EMA	EP
38,84	4,89	435,31

**Tabela 8.** Resultado de Previsão do Tipo TSM-Vazão para Usina de Três Marias.

Quantidade de Pontos Relevantes	Defasagem (meses)	EPMA (%)	EMA	EP
2	1	32,83	3,92	497,25
3	11	35,85	4,21	759,93
5	12	36,3	0,67	399,33
3	6	37,3	4,75	594,3
20	3	37,47	1,37	668,67

Assim como visto em Sobradinho, o número de pontos relevantes a obter bons resultados é baixo para a usina de Três Marias. A seguir, a **Figura 7** ilustra as séries temporais previstas pela rede MLP e a observada em Três Marias no período de Novembro de 2000 à abril de 2001.



**Figura 7.** Séries temporais ocorrida e previstas para usina de Três Marias.

A **Figura 7** mostra que no período de Novembro à Janeiro a rede sem os dados de TSM obteve melhor desempenho na previsão. Por outro lado, no período de fevereiro à março a rede treinada com dados de TSM conseguiu uma melhor aproximação ao ocorrido.

# Capítulo 5

## Conclusão

Este trabalho analisa o impacto da Temperatura da Superfície do Mar na previsão de vazão em hidroelétrica da Bacia do Rio São Francisco.

A maior parte da energia brasileira tem origem em usinas hidroelétricas que tem seu planejamento fortemente dependente da disponibilidade dos reservatórios de água. A racionalização da água, visando compensar a produção de energia em períodos secos, leva em consideração, entre outros fatores, a vazão prevista pelo sistema implantado nas usinas, que utilizam técnicas estatísticas. Contudo, modelos estatísticos não são modelos mais viáveis para o tratamento de dados não lineares, como são as séries de vazão.

Este trabalho propôs o uso de Redes Neurais Artificiais para realizar a previsão de vazão utilizando dados históricos de Temperatura da Superfície do mar. Para tanto, utilizou-se as séries das usinas de Três Marias e Sobradinho, pertencentes a bacia do Rio São Francisco, como estudo de caso. As previsões foram realizadas por redes MultiLayer Perceptron treinadas, parte com dados de vazão, e parte com TSM e vazão.

### 5.1 Contribuições

Os resultados obtidos neste trabalho demonstram que, ao se considerar as TSM, há ganho de desempenho para a previsão de vazão. Isto indica que esta variável tem significância como parâmetro de entrada para o modelo de previsão de médio e longo prazo. Portanto, a utilização de modelos não lineares que consideram as TSM podem diminuir a incerteza e melhorar o processo decisório de planejamento nas usinas hidroelétricas no processo de racionalização.

# Bibliografia

BATTAGELLO, V. A.; CORDEIRO, M. X.; DAL PINO, A. **Redes Neurais Aplicadas à Previsão de Séries Temporais**. ITA. São José dos Campos.

BRAGA, A. D. P.; LUDERMIR, T. B.; CARVALHO, A. C. **Redes Neurais Artificiais Teoria e aplicações**. Rio de Janeiro: LTC, 2000.

CHATFIELD, C. **The Analysis of Time Series**. 4ª Edição. ed. New York: [s.n.], 1991.

CYBENKO, G. Approximation by superposition of a sigmoid function. **Mathematics of Control, Signals and Systems**, p. 303-314, 1989.

MATSUO, P. et al. **Previsão estatística de vazão trimestral na bacia em anomalias de TSM**. Anais do XII Congresso Brasileiro de Meteorologia. Foz de Iguaçu: [s.n.]. 2002. p. 1790-1797.

NETO, L. B. et al. Previsão da Vazão Afluente Mensal do Rio São Francisco usando-se Redes de Elman. **Engevista**, v. Vol. 7, n. 1, p. 63-72, 2005.

NOAA. Extended Reconstruction Sea Surface Temperature (ERSST.v3b). **NOAA Satellite and Information Service**, 2009. Disponível em: <<http://lwf.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php>>. Acesso em: 21 Novembro 2009.

OECD/IEA. The Energy Situation in Brazil: an overview. **International Energy Agency**, 2006. Disponível em: <<http://www.iea.org/publications>>. Acesso em: 28 Novembro 2009.

PO.DAAC. Sea Surface Temperature: NCEP Reynolds Historical Reconstructed Sea Surface Temperature Data Set. **Physical Oceanography DAAC**, 2009. Disponível em: <<http://podaac.jpl.nasa.gov/PRODUCTS/p118.html>>. Acesso em: 17 Novembro 2009.

ROSENBLATT, F. The perceptron: probabilistic model for information storage and organization in the brain. **Psychological Review**, p. 386–408, 1958.

SMITH, T. M.; REYNOLDS, R. W. Improved extended reconstruction of SST (1854-1997). **J. Climate**, 2004. 2466-2477.

TUCCI, C. E. **Previsão de médio prazo da afluência de reservatórios com base na previsão climática**. IPH-UFRGS. Porto Alegre. 2002.

TUCCI, C. E. **Previsão de Vazões na Bacia do Rio São Francisco com Base na Previsão Climática**. IPH-UFRGS. Porto Alegre. 2004.

VALENÇA, M. J.; LUDERMIR, T. B.; VASCONCELOS, A. **Sistema de apoio à decisão para a operação hidráulica de Sobradinho incorporando tendências macro-climáticas utilizando redes neurais**. Natal: [s.n.]. 2000. p. 07-17.

WIDROW, B.; HOFF, M. E. Adaptative switching circuits. **Institute of Radio Engineers**, p. Western Electronics Show and Convention, 1960.

# Apêndice A

## Quantidade de pontos importantes pelo 'Teste t'

Mês	Defasagem = 0		Defasagem = 1		Defasagem = 2		Defasagem = 3		Defasagem = 4		Defasagem = 5		Defasagem = 6	
	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias
Janeiro	1714	1348	980	961	623	815	491	454	432	596	788	502	382	335
Fevereiro	1180	944	537	335	503	703	248	330	698	288	275	334	510	193
Março	795	291	303	55	794	647	396	205	241	600	221	518	796	716
Maiο	1259	1291	584	224	799	168	893	347	610	385	438	262	2137	1287
Abril	1722	1154	1190	516	1137	286	743	348	950	397	662	396	2350	396
Junho	2227	1686	1618	777	1241	300	1129	188	846	291	1011	303	2334	447
Julho	2348	1277	1938	857	1355	364	1571	875	1113	155	1173	258	1815	261
Agosto	2444	1641	2164	1500	1603	1104	1178	790	1006	567	1143	290	1502	376
Setembro	2774	1532	1227	820	1204	666	1135	679	940	480	678	532	1024	609
Outubro	2554	1520	2067	1763	1170	1013	1229	824	1033	792	837	808	1184	848
Novembro	1628	537	1075	574	412	344	401	208	575	435	563	537	716	767
Dezembro	1017	746	982	692	195	301	483	197	467	273	559	488	730	551



Mês	Defasagem = 7		Defasagem = 8		Defasagem = 9		Defasagem = 10		Defasagem = 11		Defasagem = 12	
	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias	Sobradinho	Três Marias
Janeiro	528	257	891	265	586	473	381	750	222	636	481	300
Fevereiro	803	104	997	743	694	198	922	455	1275	1474	650	254
Março	1006	496	562	251	642	270	586	387	819	1058	668	554
Maio	964	302	687	668	1111	581	1004	481	1186	1725	590	362
Abril	1999	389	936	517	1080	560	1385	400	1721	1166	893	424
Junho	2257	536	1656	448	1311	665	1719	612	1713	993	1351	920
Julho	1772	314	1984	637	1095	502	1562	457	1479	839	1444	947
Agosto	1786	781	2171	867	1159	669	1498	823	1599	1219	1281	945
Setembro	716	178	1557	1239	993	1033	1234	426	1350	716	1365	1474
Outubro	1068	1018	1712	2581	1520	2090	1182	1401	971	1061	1284	1496
Novembro	618	330	1294	715	1130	812	1444	1188	901	1186	1202	537
Dezembro	311	669	315	140	592	443	1287	517	326	382	504	1505

# Anexo A

## Tabela de Valores Selecionados para o Teste t

Amostras	Percentual de Confiança										
	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646

<b>40</b>	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
<b>50</b>	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
<b>60</b>	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
<b>80</b>	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
<b>100</b>	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
<b>120</b>	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
$\infty$	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291