
TÉCNICA DE IDENTIFICAÇÃO DE MODELOS LINEARES E NÃO-LINEARES DE SÉRIES TEMPORAIS

Ivette Luna*

iluna@cose.fee.unicamp.br

Rosângela Ballini†

ballini@eco.unicamp.br

Secundino Soares*

dino@cose.fee.unicamp.br

*Universidade Estadual de Campinas – UNICAMP
Departamento de Engenharia de Sistemas – DENSIS
Faculdade de Engenharia Elétrica e de Computação – FEEC
13081-970 Campinas, SP, Brasil

†Universidade Estadual de Campinas – UNICAMP
Departamento de Teoria Econômica – DTE
Instituto de Economia – IE
13083-857 Campinas, SP, Brasil

ABSTRACT

In this work, an algorithm for identifying time series models is proposed. The strategy is based on Partial Mutual Information Criterion (PMI), which considers not only linear but also non-linear relations between variables under study. For calculating the PMI criterion, it is necessary to approximate marginal and joint probability densities, as well as conditional expected values. In this work, these operators are estimated using the city-block distance function and product multivariate kernel estimators. The algorithm is applied for identifying time series linear models and for selecting inputs for a non-linear model based on neural networks. The neural model is used for modelling monthly average streamflow series of a Brazilian river. Experimental results show good performance of the proposed approach.

KEYWORDS: Partial mutual information, input selection, Kernel functions, non-parametric regression, temporal series.

Artigo submetido em 19/12/2005

1a. Revisão em 28/08/2006

2a. Revisão em 20/09/2006

Aceito sob recomendação do Editor Associado

Prof. José Roberto Castilho Piqueira

RESUMO

Este trabalho apresenta uma proposta de identificação de modelos de séries temporais baseada no algoritmo de Informação Mútua Parcial (PMI). Este critério leva em conta tanto as relações lineares, como as relações não-lineares existentes entre as variáveis consideradas na análise. Para o cálculo do PMI é necessário estimar as funções de probabilidades marginais e conjunta e o valor esperado. Neste trabalho, essas funções são aproximadas através da função de diferença absoluta em combinação com as funções de *kernel*. O algoritmo é aplicado na identificação de modelos lineares de séries temporais, assim como na seleção de entradas de um modelo não-linear utilizando redes neurais artificiais. Estes modelos neurais são utilizados na modelagem de séries de vazões médias mensais do Brasil. Os resultados mostram a eficiência do algoritmo, tanto para identificação de modelos lineares como para não-lineares.

PALAVRAS-CHAVE: Informação mútua parcial, seleção de entradas, funções de Kernel, regressão não paramétrica, séries temporais.

1 INTRODUÇÃO

A área de identificação de sistemas é tratada, muitas vezes, como um problema de otimização que envolve algumas medidas para adequação de modelos matemáticos candidatos a representar um processo real, sendo que a seleção de modelos e o ajuste dos parâmetros são influenciados por diversos fatores, tais como: (i) conhecimento *a priori* do sistema (linearidade, grau de não-linearidade, atrasos); (ii) propriedades do modelo (complexidade); (iii) seleção da medida de erro a ser minimizada; (iv) presença de ruídos (Johansson, 1993), (Coelho and dos Santos Coelho, 2004).

De modo geral, a identificação de sistemas consiste de três etapas: determinação da estrutura, estimação dos parâmetros e validação do modelo (Aguirre, 2004). Este trabalho se concentra, basicamente, na determinação da estrutura do modelo matemático para previsão de séries temporais. Com este propósito, o processo de identificação das variáveis que definem um sistema é uma das etapas mais importantes na construção de um modelo, pois este deve representar de maneira eficiente a dinâmica do sistema e, no caso específico de previsão de séries temporais, encontrar um eficiente modelo de previsão, considerando sempre os objetivos da análise da série.

As variáveis de entrada de um modelo são escolhidas através de informação conhecida *a priori* ou de forma empírica, via tentativa e erro, acarretando na escolha de variáveis redundantes ou de pouca importância. Outra forma de se determinar as variáveis de entrada é a utilização de critérios de informação tais como de informação *Bayesiana* (Schwarz, 1978), de *Akaike* (Akaike, 1969), ou *Minimum Description Length*, os quais combinam a variância residual e a ordem do modelo, para estabelecer a necessidade de adequação do modelo com *princípio da parcimônia* (Haber and Unbehauen, 1990), (Ljung, 1999).

Para casos de relações não-lineares entre as variáveis, sugere-se o uso de testes de causalidade, tal como o teste de causalidade de Granger (Granger, 1989). Esta aplicação consiste em analisar se as defasagens de um dos regressores incluídos no modelo econométrico possuem conteúdo preditivo útil, muito superior ao dos outros regressores no modelo. A afirmação de que uma variável não possui conteúdo preditivo corresponde à hipótese nula de que os coeficientes de todas as defasagens daquela variável são iguais a zero. Entretanto, a utilização desse tipo de teste supõe que a série seja estacionária e que os resíduos seguem uma distribuição normal, o que não é válido para um grande número de séries temporais e, em especial para este trabalho, para as séries de vazões.

No caso de modelos paramétricos não-lineares, como redes neurais artificiais, a escolha das entradas é ainda mais importante, pois o número de entradas define parcialmente a

estrutura da rede, conduzindo a modelos mais complexos à medida que o número de entradas aumenta (Narendra and Parthasarathy, 1990). Além disso, quanto maior o número de entradas, maior será o número de parâmetros a serem ajustados, aumentando a complexidade computacional, o período de treinamento, e diminuindo o desempenho do modelo, devido a inclusão de entradas pouco relevantes e ao aumento do número de mínimos locais na superfície de erro (Zheng and Billings, 1995).

Se as relações ou dependências entre as variáveis do sistema são lineares, tem-se que, a análise destas através de medidas de relações lineares, como critérios de informação ou coeficiente de correlação serão suficientes (Sharma, 2000). Se o sistema envolve relações não-lineares, como acontece na maioria dos problemas reais, uma aproximação linear pode fornecer como resultado um modelo pouco eficiente. Nestes casos, é necessário utilizar medidas que considerem estas características na escolha das variáveis, para definir os estados associados, ou pelo menos, a maior parte destes.

Este trabalho apresenta uma metodologia de escolha de entradas para modelos de previsão de séries temporais, baseada no Critério de Informação Mútua Parcial (PMI - *Partial Mutual Information*), proposto inicialmente em (Sharma, 2000). Este critério determina os atrasos relevantes e, não necessariamente consecutivos, que forneçam a maior quantidade de informação ao sistema a partir de um número mínimo de entradas ou atrasos. A quantidade de informação que cada entrada armazena é representada pelo valor da informação mútua parcial associada. As variáveis ou entradas do modelo são selecionadas a partir de um conjunto inicial de possíveis entradas.

O método descrito neste trabalho tem como característica fundamental, a capacidade de captar relações não-lineares a partir da série de dados do sistema, sendo esta a diferença básica, quando comparado aos métodos de seleção de entradas tradicionais, baseados no critério de correlação. Sharma (2000) propôs o algoritmo PMI, assumindo que os dados tenham uma distribuição normal, independente do sistema. No trabalho de Bowden (2003) foi proposta a utilização da distância *city-block* ou diferença absoluta para a aproximação das funções de probabilidades marginais e conjuntas. Além dessa aproximação, em (Bowden, 2003) é utilizada uma rede neural de regressão (Specht, 1991), para aproximar os valores condicionais esperados, o que tornou o tempo de execução do algoritmo menor quando comparado com a metodologia proposta em (Sharma, 2000).

Este trabalho propõe a utilização da função de distância *city-block* ou diferença absoluta, tanto para a aproximação das funções de probabilidades marginais e conjuntas, como na aproximação dos valores esperados condicionais, através de

regressores de kernel (Scott, 1992), necessários no cálculo do critério PMI. A vantagem de utilizar a distância *city-block* em combinação com as funções de *kernel*, é que não é necessário assumir algum tipo de distribuição dos dados. Além disso, utiliza-se a distância absoluta na aproximação dos valores esperados condicionais, diminuindo assim a complexidade das aproximações feitas em (Sharma, 2000), assim como a necessidade de ajustar um modelo de rede neural de regressão a cada iteração, como feito em (Bowden et al., 2005).

Desta forma, as entradas finais serão aquelas que possuem uma dependência significativa com a saída, sendo esta dependência determinada por uma medida de significância, a qual representa um limite superior (γ -th percentil) aos valores PMI. Este limite superior é determinado por *bootstrapping* ou reordenamento dos dados, considerando o limiar de corte (γ th percentil) previamente definido.

Os resultados são comparados aos obtidos pelo modelo de rede neural, com entradas definidas pelo critério de Akaike (Akaike, 1969). A proposta é aplicada na identificação de modelos lineares de séries temporais sintéticas, assim como na solução de um problema real, que é a identificação de modelos de previsão de séries de vazões médias mensais do posto de Furnas, localizado na bacia do rio Grande, região sudeste do Brasil.

O objetivo do modelo de planejamento energético a cargo do Operador Nacional do Sistema Elétrico - ONS, é minimizar o custo total de operação através de decisões de geração (térmica ou hidráulica), intercâmbio entre regiões e cortes de carga (déficit), sendo este custo dependente das vazões no futuro. Por este motivo, se faz necessário o estudo das vazões, assim como o estudo e desenvolvimento de modelos capazes de capturar a natureza não linear das séries, associada à sazonalidade e elevado nível de incerteza apresentadas nas aflúências dos rios brasileiros.

Assim, a seleção de entradas para a série de vazão da usina de Furnas, é feita com o objetivo de construir um modelo de previsão não-linear utilizando redes neurais, sendo este modelo validado através de uma análise dos resíduos, através dos testes de autocorrelação residual, de correlação cruzada (Morettin and Tolo, 2004), e as estatísticas Q de Ljung-Box (Ljung and Box, 1978).

2 CRITÉRIO DE INFORMAÇÃO MÚTUA PARCIAL (PMI)

O critério de informação mútua surge da teoria de informação, embora seja aplicado em outras áreas tais como reconhecimento de padrões, processamento de imagens, seleção e identificação de variáveis de entrada para modelos não-paramétricos (Bonnlander and Weigend, 1994), (Zaffalon

and Hutter, 2002), (Martins, 2004).

A informação mútua é uma medida bastante usada para análise de dependência estocástica de variáveis aleatórias discretas (Cover and Thomas, 1991), (Soofi, 2000). Este critério fornece uma medida do grau de dependência entre variáveis, sendo um indicador importante na análise de séries temporais. Assim, se duas variáveis são independentes, o critério de informação mútua será zero; se duas variáveis são fortemente dependentes, a informação mútua terá um valor alto.

A informação mútua pode também ser considerada como uma medida da quantidade de informação armazenada em uma variável com relação a outra, sendo por este motivo, interessante na escolha de entradas para modelos de sistemas não-lineares, principalmente de redes neurais, onde o processo de aprendizado pode ser considerado como um mecanismo de extração de conhecimento (Bonnlander and Weigend, 1994), (Zheng and Billings, 1995).

O critério de informação mútua (MI - *Mutual Information*) entre duas variáveis discretas x e y é definido como:

$$MI = \frac{1}{N} \sum_{i=1}^N \log_e \left[\frac{f_{xy}(x_i, y_i)}{f_x(x_i)f_y(y_i)} \right] \quad (1)$$

sendo (x_i, y_i) o i -ésimo par de dados do conjunto de amostras, com $i = 1, \dots, N$, $f_{xy}(x, y)$ a função de probabilidade conjunta (*Probability Density Function* - PDF) de x e y e $f_x(x)$ e $f_y(y)$ são as funções de probabilidade marginais de x e y , respectivamente. Se x e y são independentes, a função de probabilidade conjunta será igual ao produto das probabilidades marginais, sendo o valor do MI igual a 0. Caso x e y sejam dependentes, o valor do MI será maior que 0.

Existem diversas maneiras de aproximar as funções de probabilidades marginais e conjuntas. Histogramas de frequência como médias da estimativa das probabilidades são utilizados em (Fraser and Swinney, 1986). Neste trabalho, foram utilizadas funções de *kernel* para aproximar tanto as funções de probabilidades marginais como as funções de probabilidade conjunta, por ser uma maneira mais eficiente e robusta como mostrado em (Sharma, 2000). A aproximação das funções de probabilidade é apresentada a seguir.

2.1 Aproximação das Funções de Probabilidades Marginal e Conjunta

Seja N o número de pares de dados entrada-saída $[\mathbf{x}(k), y(k)]$, com $k = 1, \dots, N$. A função de probabilidade marginal de uma variável x unidimensional pode ser aproximada através de estimadores não paramétricos de *kernel*,

utilizando a seguinte equação:

$$\hat{f}(\mathbf{x}) = \frac{1}{N\lambda} \sum_{i=1}^N K\left(\frac{x-x_i}{\lambda}\right) = \frac{1}{N} \sum_{i=1}^N K_\lambda(x-x_i) \quad (2)$$

sendo que $K_\lambda(t)$ é denominada de função de *kernel*, e λ é a largura de banda ou parâmetro de dispersão.

Existem diversas funções de *kernel* (Silverman, 1986). Dentre estas, (Bowden et al., 2005) utiliza a função de distância absoluta ou *city-block*. A vantagem de utilizar este tipo de função está no fato de ser simples em termos computacionais, assim como não precisar assumir nenhum tipo de distribuição dos dados como é feito em (Sharma, 2000) ou em (Akaho, 2002). A função de probabilidade marginal de \mathbf{x} utilizando a distância absoluta, é definida por:

$$\hat{f}_x(\mathbf{x}) = \frac{1}{N(2\lambda)^p} \sum_{i=1}^N \prod_{j=1}^p \exp^{-|\mathbf{x}_j - \mathbf{x}_{ij}|/\lambda}$$

isto é,

$$\hat{f}_x(\mathbf{x}) = \frac{1}{N(2\lambda)^p} \sum_{i=1}^N \exp\left[-\frac{1}{\lambda} \sum_{j=1}^p |\mathbf{x}_j - \mathbf{x}_{ij}|\right]$$

As Equações (3) e (4) surgem a partir da Equação (2), e uma adaptação para o caso de \mathbf{x} ser multidimensional e zando como função de *kernel* a função *city-block*. Na Equação (4), p é a dimensão do vetor \mathbf{x} e $\hat{f}_x(\mathbf{x})$ a aproximação função de probabilidade marginal de \mathbf{x} .

O parâmetro λ é calculado mediante:

$$\lambda = \left(\frac{4}{p+2}\right)^{1/(p+4)} N^{-1/(p+4)}$$

Embora a definição de λ dada na Equação (5) tenha sido derivada para dados com distribuição normal, esta equação é amplamente utilizada na literatura por ser uma definição simples e eficiente como mostrado em (Bowden et al., 2005).

Da mesma forma, a função de probabilidade conjunta de \mathbf{x} dado y , com y sendo a saída unidimensional, é definida por (Akaho, 2002):

$$\hat{f}_{xy}(\mathbf{x}, y) = \frac{1}{N(\lambda)^{p+1}} \sum_{i=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}_i}{\lambda}\right) K\left(\frac{y - y_i}{\lambda}\right) \quad (6)$$

ou seja:

$$\hat{f}_{xy}(\mathbf{x}, y) = \frac{1}{N(2\lambda)^{p+1}} \sum_{i=1}^N \exp\left[-\frac{1}{\lambda} s_i\right] \quad (7)$$

sendo s_i igual a:

$$s_i = \sum_{j=1}^p |\mathbf{x}_j - \mathbf{x}_{ij}| + |y - y_i| \quad (8)$$

Para ilustrar a utilização da aproximação das funções de probabilidades utilizando a função de *kernel* no cálculo do MI, considera-se o seguinte modelo:

$$y = x^4 + e_1 \quad x = \text{sen}(2\pi t/T) + e_2 \quad (9)$$

sendo $T = 20$, $t = 1, \dots, 200$ e e_1 e e_2 os sinais de ruído com distribuição normal de média zero e desvio padrão igual a 0,1. A Figura 1 e a Equação (9) ilustram de forma clara uma forte relação não-linear entre a variável independente x e a variável dependente y .

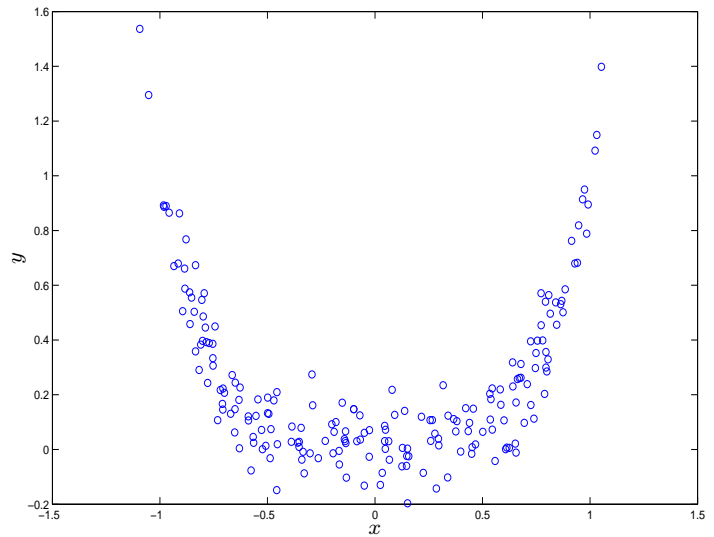


Figura 1: Modelo da Equação (9).

O coeficiente de correlação entre estas duas variáveis é de 0,0032, indicando que praticamente não existe relação entre ambas as variáveis, isto devido a relação entre x e y ser não-linear. Aplicando o critério de informação mútua (Equação (1)) nas amostras do modelo da Equação (9), obtém-se um valor de MI igual a 0,4199. A Tabela 1 mostra os valores obtidos para os critérios de correlação e para o MI, assim como o limiar para cada um. Este limiar foi calculado tendo como base o limiar de confiança ou percentil de 95%, indicando que a dependência entre as duas variáveis é forte, se o valor obtido for maior a este limiar, com uma probabilidade de erro de 5%.

Observa-se na Tabela 1 que, enquanto o coeficiente de correlação indica independência entre as variáveis, o critério de

Tabela 1: Comparação da informação mútua e o coeficiente de correlação para os dados da Figura 1.

Critério	Valor	Percentil (95%)
MI	0,4199	0,1154
Correlação	0,0032	0,0965

informação mútua indica dependência entre x e y , obtendo um valor de 0,4199 superior ao limiar igual a 0,1154.

Em (Sharma, 2000) é proposto o critério de informação mútua parcial (PMI) o qual é uma medida da redução da incerteza em y devido ao conhecimento de x , ou seja, o critério PMI mede a informação mútua entre a variável independente x e a variável dependente y , condicional a um conjunto de entradas \mathbf{z} selecionado inicialmente. Dado que \mathbf{z} existe, é necessário extrair a influência desse conjunto em relação aos atrasos ainda não selecionados, para calcular a real contribuição de x . Assim, a Equação (1) pode ser reformulada da seguinte maneira:

$$PMI = \frac{1}{N} \sum_{i=1}^N \log_e \left[\frac{f(x'_i, y'_i)}{f_{x'}(x'_i) f_{y'}(y'_i)} \right] \quad (10)$$

na qual,

$$x'_i = x_i - E(x_i | \mathbf{z}) \quad y'_i = y_i - E(y_i | \mathbf{z}) \quad (11)$$

sendo x'_i e y'_i os valores residuais correspondentes ao i -ésimo par de dados e $E(\cdot)$ o valor esperado condicional associado. As definições dadas na Equação (11) garantem que as variáveis x' e y' representem a informação restante, uma vez que o efeito das entradas escolhidas presentes em \mathbf{z} tem sido considerado.

Como pode ser observado na Equação (11), para o cálculo do PMI, é necessário ter uma boa estimativa do valor esperado $E(\cdot)$, para assim, estimar de forma precisa x' e y' .

Em (Sharma, 2000) foi assumido que os dados tenham uma distribuição normal, independente do sistema. Já no trabalho de Bowden (2003) foi proposto a utilização de uma rede neural de regressão (Specht, 1991), para aproximar os valores condicionais esperados.

Este trabalho propõe a utilização da função de distância *city-block* ou diferença absoluta, tanto para a aproximação das funções de probabilidades marginais e conjuntas, como na aproximação dos valores esperados condicionais, através de regressores de *kernel* (Scott, 1992), necessários no cálculo do critério. Assim, o valor esperado é aproximado através de um modelo de regressão não paramétrica de *kernel*, conhecido como Estimador de Nadaraya-Watson (Scott, 1992).

2.2 Aproximação do Valor Esperado Condicional

Seja $\hat{r}(x)$ a aproximação do valor esperado condicional $E(Y|X = x)$, definida da seguinte forma:

$$\hat{r}(x) = \sum_{i=1}^N w_{\lambda_x}(x, x_i) y_i \quad (12)$$

sendo,

$$w_{\lambda_x}(x, x_i) = \frac{K_{\lambda_x}(x - x_i)}{\sum_{j=1}^N K_{\lambda_x}(x - x_j)} \quad (13)$$

na qual, $K_{\lambda_x}(x - x_i)$ é a função de *kernel* para x . Nesta aproximação também foi utilizada a função *city-block*. Este critério foi adotado por ser mais simples de implementar, quando comparado com a aproximação feita em (Sharma, 2000) e em (Bowden, 2003).

2.3 Determinação do limiar de confiança para o PMI

Existem duas maneiras de finalizar o processo de seleção de entradas. A primeira consiste na construção de um teste de hipótese utilizando p seqüências diferentes de x em relação a y , através de um reordenamento ou *bootstrapping* da variável independente, e obtendo o respectivo vetor de PMIs. Ou seja, o limiar de confiança para determinar se uma entrada é relevante para modelar o sistema é calculado assumindo independência entre a variável de entrada e a variável de saída.

Os PMIs obtidos são reordenados de forma ascendente e o valor correspondente ao γ -th percentil será o limiar para rejeitar ou não rejeitar a hipótese nula. Dessa forma, a hipótese de independência entre x e y será rejeitada caso a entrada avaliada possua um PMI superior ao nível de significância ($\alpha = 1 - \gamma$) obtido, ou seja, a entrada será considerada relevante com uma possibilidade de $\alpha\%$ destas serem realmente independentes. Resultados do algoritmo proposto na modelagem de séries temporais utilizando redes neurais recorrentes encontram-se em (Luna et al., 2006), onde adotou-se $p = 100$ seqüências, um percentil $\gamma = 95\%$ e, portanto, um nível de significância $\alpha = 5\%$. Este mesmo limiar de confiança foi utilizado na identificação das séries sintéticas apresentadas na seção 3.1.

O segundo critério de parada consiste em estabelecer um mínimo de PMI para não rejeitar a entrada avaliada. Caso o valor do PMI seja menor a este valor mínimo estabelecido, a entrada é rejeitada. Neste caso, uma visualização gráfica facilita a análise, como será visto na seção 3.2.1. Este critério de parada é factível devido à natureza monotonicamente descendente do PMI a medida que as entradas são selecionadas. Neste trabalho, para o problema da série de vazões o limiar para o PMI foi de 0,05.

Em termos gerais, tanto o limiar de confiança como o valor mínimo pré-definido para o PMI refletem o rigor com o qual as entradas serão selecionadas. Maior o nível de significância na escolha de entradas, mais rigorosa será a seleção das mesmas, já que estas terão que possuir um valor mais alto de PMI para serem escolhidas. Ou seja, se o valor de α é baixo, corre-se o risco de escolher entradas pouco relevantes para a modelagem no final do processo. Por outro lado, se α é muito alto, entradas importantes podem estar sendo rejeitadas, prejudicando a modelagem da série. Os níveis de significância adotados neste trabalho foram definidos após uma seqüência de testes para valores diferentes de α .

2.4 Algoritmo PMI

O algoritmo para identificar as entradas para modelar um sistema, utilizando o algoritmo PMI proposto, pode ser resumido nos seguintes passos:

1. Construir um conjunto de possíveis entradas ao sistema a ser modelado, denotado como um vetor \mathbf{z}^* . Definir um vetor que armazenará as entradas selecionadas como \mathbf{z} . Inicialmente, o vetor \mathbf{z} é igual ao vetor nulo;
2. Calcular o PMI entre cada uma das variáveis em \mathbf{z}^* e a variável dependente y , condicional ao conjunto de entradas pertencentes ao vetor \mathbf{z} , utilizando a Equação (10);
3. Identificar a variável de entrada em \mathbf{z}^* que apresenta o maior PMI;
4. Se o PMI da entrada for superior ao limiar definido para o PMI, incluir a entrada em \mathbf{z} e retirá-la de \mathbf{z}^* . Caso contrário, fim do algoritmo.
5. Se o valor PMI da variável de entrada identificada for superior ao limiar definido, incluir a entrada em \mathbf{z} e retirá-la de \mathbf{z}^* . Se a dependência não é significativa, ir para o passo 7;
6. Repetir os passos 2-5 tantas vezes quanto necessário;
7. Este passo será alcançado somente quando todas as variáveis de entrada forem identificadas.

3 APLICAÇÕES

Nesta seção é apresentada a aplicação do algoritmo PMI proposto, na identificação de modelos lineares e não-lineares. Primeiramente, o algoritmo PMI é aplicado a duas séries sintéticas geradas por modelos lineares. O critério PMI também é aplicado no processo de identificação de uma série de vazão. Nesse caso, será utilizado um modelo de rede neural para previsão das vazões e, a seguir, é feita uma análise dos resíduos com o objetivo de avaliar o modelo identificado.

3.1 Séries sintéticas

O algoritmo proposto foi aplicado na identificação de dois modelos lineares propostos em (Sharma, 2000) como em (Bowden et al., 2005). Estes modelos são apresentados como uma maneira de verificar o desempenho do algoritmo descrito. Os modelos são auto-regressivos (AR) de ordem 1 e 9, respectivamente, sendo definidos pelas seguintes equações:

AR(1):

$$x_t = 0,9x_{t-1} + 0,866e_t$$

AR(9):

$$x_t = 0,3x_{t-1} - 0,6x_{t-4} - 0,5x_{t-9} + e_t$$

sendo e_t uma variável aleatória com distribuição normal de média zero e desvio padrão unitário.

A partir dos dois modelos auto-regressivos acima, foram gerados um total de 420 pontos para cada série sintética, sendo que, os primeiros 20 pontos foram desconsiderados com a finalidade de reduzir a influência da inicialização. O algoritmo PMI foi aplicado a cada série, onde os primeiros 15 atrasos para cada modelo AR foram avaliados, como candidatos a possíveis variáveis de entrada dos modelos de séries temporais. As Tabelas 2 e 3 apresentam os resultados obtidos para os modelos AR(1) e AR(9), respectivamente.

Tabela 2: Resultados do algoritmo PMI aplicado ao modelo AR(1).

Atraso	PMI	Percentil (95%)
1	0,5920	0,2034
2	0,2118	0,2186

Tabela 3: Resultados do algoritmo PMI aplicado ao modelo AR(9).

Atraso	PMI	Percentil (95%)
4	0.3146	0.2166
9	0.2174	0.1477
1	0.1310	0.1006
5	0.0413	0.0429

Como pode ser visto nas Tabelas 2 e 3, o algoritmo é capaz de reconhecer os atrasos corretos que influenciam cada série. Para o modelo AR(1), o atraso 1 foi o escolhido pelo algoritmo, pois para o atraso 2, o valor de PMI foi inferior ao valor do percentil ou limiar adotado. Isto significa que, todos

os outros atrasos terão um valor de PMI ainda menor e portanto, a única variável a ser considerada é o atraso 1 (Tabela 2). Já para o modelo AR(9), os atrasos selecionados foram 4, 9 e 1, pois para o atraso 5, o valor de PMI foi inferior ao limiar (Tabela 3). Portanto, os três primeiros atrasos com maior valor de PMI serão suficientes para construir o modelo de séries temporais.

Além disso, observando os valores dos coeficientes de ambos os modelos auto-regressivos e a ordem na qual os atrasos foram selecionados pelo algoritmo, pode-se verificar que, além de selecionar as variáveis corretas, o critério PMI fornece a informação da importância de cada uma destas variáveis para o modelo de série temporal, pois os atrasos com maior influência no modelo (coeficiente mais alto em valor absoluto no modelo AR), são os que possuem um maior valor de PMI.

3.2 Série de vazão

Redes neurais artificiais são sistemas não-lineares de processamento de informação formados pela interconexão de unidades simples de processamento denominadas neurônios artificiais (Von Zuben, 2003). Estes modelos têm sido amplamente utilizados na construção de modelos de séries temporais (Zealand et al., 1999), sendo um dos passos fundamentais é a escolha das variáveis de entrada para o modelo neural. Muitas entradas produzem um modelo mais complexo e nem sempre mais eficiente. Poucas entradas podem fornecer informação insuficiente para o modelo neural e gerar um modelo de previsão com um desempenho pobre.

Nesta seção, modelos neurais para previsão de séries temporais serão construídos através da seleção das entradas usando o critério PMI e os parâmetros do modelo são ajustados. A seguir, uma análise dos resíduos será feita com o objetivo de validar o modelo identificado.

Os dados utilizados consistem em uma série histórica de vazão afluenta média mensal do posto de Furnas, localizado na bacia do Rio Grande, região sudeste do Brasil, que abrange o período de janeiro de 1931 a dezembro de 1998, colhida em postos de medição de vazão, onde existe aproveitamento hidroelétrico. Esse conjunto de dados compõe um banco de vazões históricas mantido pelo Operador Nacional do Sistema Elétrico (ONS).

Como foi mencionado na seção 1, o planejamento e operação de sistemas de energia e recursos hídricos requer muitas variáveis de entrada, tendo entre estas, a previsão de vazões. Por este motivo, é necessário o estudo das vazões, assim como o estudo e desenvolvimento de modelos capazes de capturar a natureza não linear das séries, tanto para fins de simulação, como de otimização e planejamento.

Neste trabalho, as entradas para o modelo neural de previsão para a série de vazões foram selecionadas utilizando o algoritmo PMI proposto. Nesse caso, foi considerado um limiar de 0,05 para o PMI para determinar se uma entrada é relevante ou não ao sistema.

Uma vez identificadas as entradas, a próxima etapa consiste em ajustar os parâmetros do modelo neural de tal forma que, no final do processo, a rede neural forneça um modelo eficiente de previsão de série temporal. Para tal, foram utilizados dados correspondentes ao período de 1931 a 1990 para a identificação das variáveis e para o ajuste do modelo neural. Para este período, foi feita uma análise dos resíduos, com o objetivo de avaliar o modelo. Para a validação do modelo identificado foi utilizado o período de 1991 a 1998. A rede neural multicamada (MLP) foi utilizada por ser um modelo de previsão já conhecido na literatura (Haykin, 2001).

A Figura 2 ilustra o histórico de vazões em estudo. A Tabela 4 mostra os valores de média e desvio padrão para cada mês da série. Como pode ser visto, a série apresenta um comportamento periódico, onde os valores oscilam entre um valor máximo e um valor mínimo ao longo do ano. Além disso, na Figura 2 pode-se observar uma variação sazonal, com período aproximado de doze meses, resultando em uma série não estacionária.

Como o algoritmo de seleção de entradas requer que as séries em estudo sejam estacionárias, os dados de vazões são transformados com o objetivo de eliminar o componente sazonal. Desta forma, é feita a padronização dos dados da seguinte forma:

$$z_{t,m} = \frac{x_{t,m} - \mu_m}{\sigma_m} \quad (14)$$

sendo que μ_m é a estimativa da média e σ_m é a estimativa do desvio padrão do mês m ; t é o ano; $x_{t,m}$ é o dado original para o ano t e mês m . A série padronizada $z_{t,m}$ apresenta média zero e desvio padrão unitário. Assim, o valor esperado da série $\mu_t = E[z_{t,m}]$ e a variância $\sigma_t^2 = E[z_{t,m} - \mu_{t,m}]^2$ são invariantes no tempo, sendo essas as características de um processo estocástico estritamente estacionário (Box et al., 1994). A seguir, são apresentados os resultados para a série de vazão.

3.2.1 Modelo neural para a série de Furnas

Para os dados do posto de Furnas, um total de quinze atrasos compuseram o conjunto inicial de possíveis variáveis de entradas. A Tabela 5 ilustra a ordem na qual os atrasos foram escolhidos, assim como os correspondentes valores de PMI e

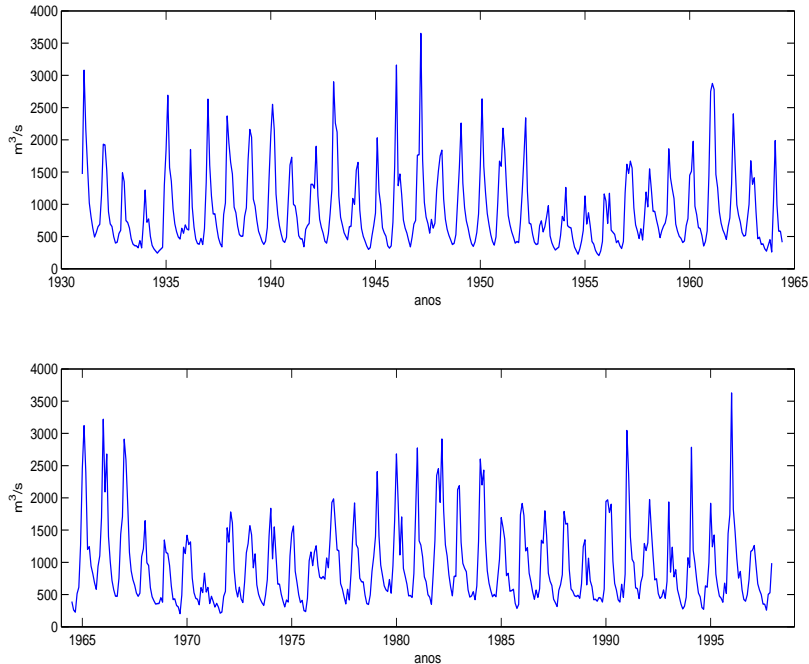


Figura 2: Série de vazões afluentes médias mensais para a Usina Hidroelétrica de Furnas: Jan/1931-Dez/1998.

Tabela 4: Médias e Desvios padrão para cada mês do histórico de Furnas.

Mês	Média	Desvio Padrão
1	1726,3	688,7
2	1661,8	639,3
3	1481,7	585,3
4	1010,1	319,5
5	744,5	201,0
6	606,6	151,6
7	508,9	117,9
8	422,8	101,5
9	428,7	153,9
10	519,0	157,3
11	730,6	284,1
12	1221,9	429,0

Tabela 5: Resultados do algoritmo PMI aplicado ao histórico de vazões do posto de Furnas: Jan/1931-Dez/1991.

Atraso	PMI	Percentil (95%)
1	0,4390	0,0709
11	0,2037	0,0530
14	0,0795	0,0441
12	0,0387	0,0159
10	0,0298	0,0116
4	0,0215	0,0115
9	0,0155	0,0055
6	0,0113	0,0043
8	0,0054	0,0021
3	0,0031	0,0009
5	0,0020	0,0005
7	0,0011	0,0004
15	0,0006	0,0001
2	0,0003	0,0001
13	0,0002	0,0000

percentil.

Como foi dito, nesta aplicação, a escolha das entradas é feita utilizando um nível de significância de 5% para determinar quais entradas são relevantes no modelo. Assim, para valores de PMI maiores que 0,05, a variável de entrada não é rejeitada, isto é, essa variável é significativa na identificação do modelo de previsão. Estes resultados podem ser visuali-

zados na Figura 3, onde observa-se que os atrasos 1, 11 e 14 possuem um valor de PMI maior que o nível de significância de 5% (representado pela linha pontilhada). Como o componente sazonal foi retirado previamente utilizando a Equação (14), o atraso 12 não foi escolhido como pode ser observado

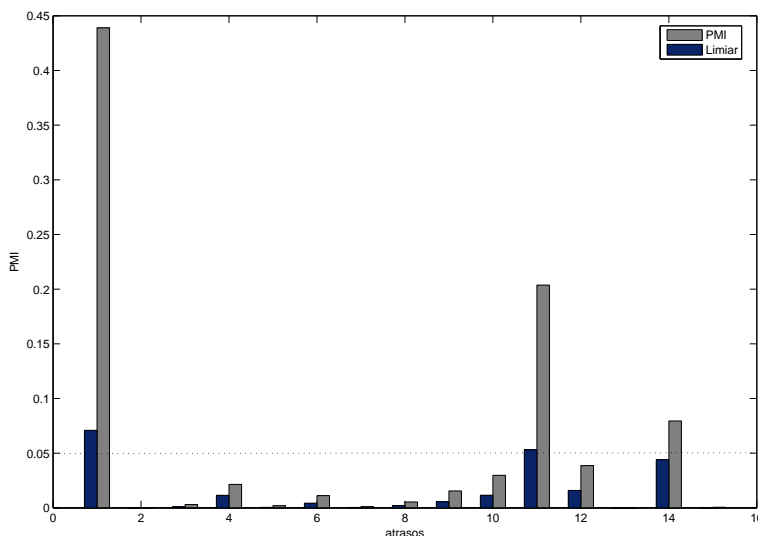


Figura 3: Seleção de entradas para a série de vazões afluentes mensais de Furnas utilizando o algoritmo PMI.

na Tabela 5 e Figura 3.

Após a seleção das variáveis de entrada, o próximo passo é a estimação dos parâmetros do modelo. A rede neural MLP foi ajustada utilizando o algoritmo de retropropagação do erro (Rumelhart and McClelland, 1986), com uma taxa de aprendizado igual a 0,15. O número de neurônios intermediários na rede neural MLP foi igual a 3 e após 1000 épocas, o processo de aprendizado ou de ajuste foi encerrado, pois verificou-se que a variação do erro quadrático médio de uma época para outra assumiu um valor menor que 10^{-3} .

A seguir, a rede neural estimada foi avaliada, com o intuito de verificar se o modelo ajustado é adequado ou não. Neste trabalho, a verificação foi feita analisando os resíduos estimados \hat{a}_t . Assim, se o modelo for adequado, os resíduos estimados deverão ser aproximadamente não-correlacionados. Para essa análise, dois testes de adequação foram usados: o teste de autocorrelação residual e o teste de correlação cruzada.

Indicando por $\hat{\rho}_k$ as autocorrelações estimadas e por \hat{s}_k as correlações cruzadas estimadas, sendo k os atrasos, os resíduos estimados \hat{a}_t são não-correlacionados se $\hat{\rho}_k \sim 0$ e $\hat{s}_k \sim 0$. As autocorrelações e correlações cruzadas são dadas por:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2}$$

$$\hat{s}_k = \frac{\sum_{t=k+1}^n \hat{a}_t (y_{t-k} - \bar{y})}{[\sum_{t=1}^n \hat{a}_t^2 \sum_{t=1}^n (y_t - \bar{y})^2]^{1/2}}, \quad k = 1, 2, \dots$$

sendo y o valor da série, \bar{y} o valor médio e n o número de dados. Assim, se os valores de $\hat{\rho}_k$ e \hat{s}_k são significativamente diferentes de zero, para um dado nível de significância α , os resíduos estimados são correlacionados. Caso contrário, os resíduos estimados são não-correlacionados. Para mais detalhes ver (Morettin and Tolo, 2004) e (de Holanda Sales et al., 1998).

A Tabela 6 mostra os valores das autocorrelações e correlações cruzadas estimadas. Além disso, esta tabela apresenta as informações sobre as estatísticas Q de Ljung-Box e seus respectivos valores-p (Ljung and Box, 1978). A estatística Q na defasagem k é um teste estatístico da hipótese nula de ausência de autocorrelação até a ordem k , dada por:

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{n-j} \quad (15)$$

sendo $\hat{\rho}$ a j -ésima autocorrelação estimada e n o número de observações. A estatística (15) segue uma distribuição χ^2 com $(k-p)$ graus de liberdade, sendo p o número de parâmetros no modelo ajustado.

A probabilidade (Prob) na Tabela 6 representa o valor-p do teste. Assim, se as referidas probabilidades são aproximadamente iguais a zero, ou menores que um dado nível de significância, rejeita-se a hipótese nula de ausência de autocorrelação, ou seja, as correlações são diferentes de zero. Vale notar que, se um processo estocástico for puramente aleatório, sua autocorrelação a qualquer defasagem $k > 0$ será nula. Assim, observa-se que, para todo valor de k , a probabilidade é significativamente diferente de zero, indicando que a hipótese nula deve ser rejeitada para níveis de significância acima

Tabela 6: Autocorrelações e correlações cruzadas dos resíduos.

k	$\hat{\rho}_k$	\hat{s}_k	$Q(k)$	Prob
1	0,000	0,000	0,0001	0,991
2	-0,010	-0,010	0,0657	0,968
3	0,010	0,010	0,1374	0,987
4	-0,014	-0,014	0,2788	0,991
5	0,071	0,071	3,8050	0,578
6	-0,044	-0,044	5,1472	0,525
7	-0,011	-0,009	5,2248	0,633
8	0,000	-0,003	5,2249	0,733
9	0,040	0,043	6,3563	0,704
10	0,025	0,019	6,7914	0,745
11	0,018	0,025	7,0333	0,796
12	0,001	-0,001	7,0336	0,855
13	0,000	0,000	7,0337	0,900
14	0,002	-0,004	7,0359	0,933
15	-0,030	-0,029	7,6579	0,937
16	0,022	0,022	7,9987	0,949
17	-0,024	-0,022	8,3932	0,957
18	0,009	0,009	8,4535	0,971
19	-0,038	-0,042	9,4921	0,964
20	-0,044	-0,041	10,869	0,950

de $\alpha = 1\%$. Logo, o modelo ajustado é adequado, pois os resíduos estimados são não correlacionados. A Figura 4 ilustra as autocorrelações estimadas dos resíduos \hat{a}_t do modelo neural, assim como as correlações cruzadas entre os resíduos estimados e a série temporal. As linhas pontilhadas na representação gráfica das autocorrelações e correlações cruzadas indicam os limites $\pm 2/\sqrt{n}$, os quais fornecem uma indicação geral de possível quebra do comportamento de ruído branco.

Após estimado o modelo de rede neural, este modelo foi validado usando o período de 1991 a 1998. Assim foram geradas as previsões um passo a frente para o período de Janeiro de 1991 a Dezembro de 1998. Os resultados de simulação do modelo neural para o posto de Furnas são ilustrados na Figura 5.

4 CONCLUSÕES

Neste trabalho foi proposto um algoritmo de identificação de sistemas lineares, e de seleção de entradas para modelos não-lineares de séries temporais. Este algoritmo é baseado no critério de informação mútua, que além de considerar as relações não-lineares existentes entre as variáveis de entrada e saída dos modelos, fornece a informação da importância

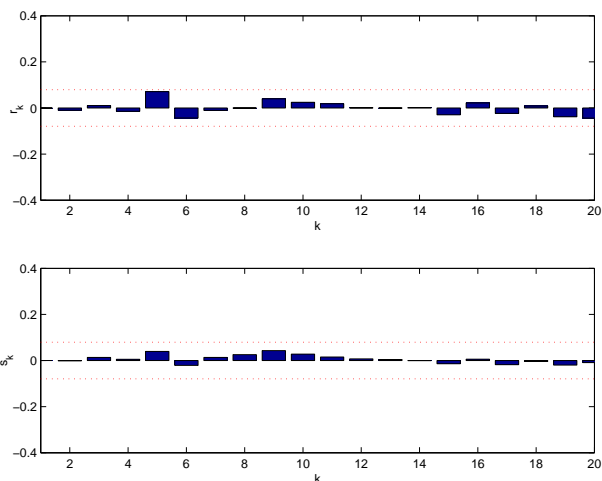


Figura 4: Funções de autocorrelações estimadas dos resíduos e função de correlação cruzada estimada para a série de vazões de Furnas.

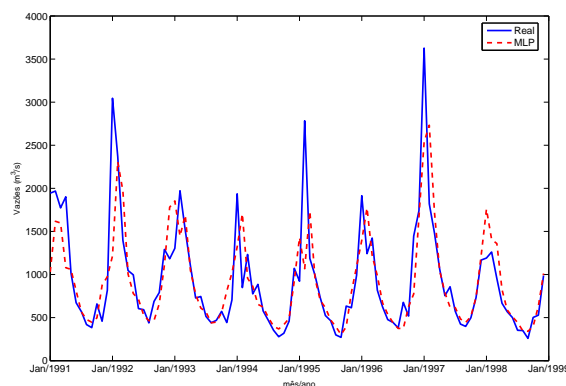


Figura 5: Série de vazões afluentes médias mensais do posto de Furnas: Jan/1991 - Dez/1998.

que cada variável possui para a construção do modelo. Para a construção do critério de informação mútua, neste trabalho foi proposto a utilização da função da distância absoluta em combinação da função de *kernel* para a aproximação das funções de probabilidades marginais e conjunta, assim como do valor esperado condicional.

O algoritmo é aplicado na identificação de duas séries sintéticas geradas por modelos lineares, assim como na seleção das entradas para um modelo neural para séries de vazões afluentes médias mensais. O modelo de série temporal não linear obtido é validado através da análise dos resíduos via testes de correlações cruzadas estimadas e autocorrelações parciais estimadas, assim como através dos coeficiente Q de Ljung-Box, indicando que o modelo neural é adequado para a série de vazões em estudo.

Em termos gerais, os resultados ilustram a eficiência do algoritmo, tanto para modelos lineares, como para modelos não-lineares de séries temporais.

AGRADECIMENTOS

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Financiadora de Estudos e Projetos (FINEP), pelo auxílio.

REFERÊNCIAS

- Aguirre, L. A. (2004). *Introdução à Identificação de Sistemas: Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*, 2 edn, Editora UFMG, Belo Horizonte, BH.
- Akaho, S. (2002). Conditionally independent component analysis for supervised feature extraction, *Neurocomputing* **49**: 139–150.
- Akaike, H. (1969). Fitting autoregressive models for prediction, *Annals of Institute of Statistical Mathematics* (21): 243–247.
- Bonnlander, B. V. and Weigend, A. S. (1994). Selecting input variables using mutual information and nonparametric density estimation, *Proc. of the 1994 International Symposium on Artificial Neural Networks*, Tainan, Taiwan, pp. 42–50.
- Bowden, G. J. (2003). *Forecasting Water Resources Variables Using Artificial Neural Networks*, Tese de Doutorado, University of Adelaide, Australia.
- Bowden, G. J., Maier, H. R. and Dandy, G. C. (2005). Input determination for neural network models in water resources applications. Part 1—background and methodology, *Journal of Hydrology* (301): 75–92.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3th edn, Holden Day, Oakland, California.
- Coelho, A. A. R. and dos Santos Coelho, L. (2004). *Identificação de Sistemas Dinâmicos Lineares*, 1 edn, Editora da UFSC, Florianópolis, SC.
- Cover, T. and Thomas, J. (1991). *Elements of Information Theory*, John Wiley & Sons.
- de Holanda Sales, P. R., de Bragança Pereira, B. and de Mesquita Vieira, A. (1998). Procedimentos Lineares para Identificação e Estimação dos Parâmetros de Modelos, *Revista Brasileira de Estatística* **59**(212): 25–51.
- Fraser, A. M. and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information, *Physical Review A* **33**(2): 1134–1140.
- Granger, C. W. J. (1989). Investigating causal relations by econometric models and cross-spectral methods, *Econometria* (3): 24–36.
- Haber, R. and Unbehauen, H. (1990). Structure identification of nonlinear dynamic systems – a survey on input/output approaches, *Automatica* **26**: 651–677.
- Haykin, S. (2001). *Redes Neurais, Princípios e Prática*, 2 edn, Bookman.
- Johansson, R. (1993). *System Modeling and Identification*, Prentice-Hall, Upper Saddle River, NY.
- Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika* **65**: 297–303.
- Ljung, L. (1999). *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, NJ.
- Luna, I., Soares, S. and Ballini, R. (2006). Partial Mutual Information Criterion For Modelling Time Series Via Neural Networks, *Proc. of the 11th Information Processing and Management of Uncertainty International Conference*, Vol. 1, pp. 2012–2019.
- Martins, D. C. (2004). *Redução de dimensionalidade utilizando entropia condicional média aplicada a problemas de bioinformática e de processamento de imagens*, Dissertação de Mestrado, IME-USP, Brasil.
- Morettin, P. A. and Toloi, C. M. C. (2004). *Análise de Séries Temporais*, 1 edn, Edgard Blücher Ltda.
- Narendra, K. and Parthasarathy, K. (1990). Identification and Control of Dynamical Systems using Neural Networks, *IEEE Transactions on Neural Networks* **1**(1): 4–27.
- Rumelhart, D. and McClelland, J. (1986). *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge.
- Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annual of Statistics* **6**(2): 461–464.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*, 1 edn, John Wiley & Sons Inc.
- Sharma, A. (2000). Seasonal to internannual rainfall probabilistic forecasts for improved water supply management: Part 1 – A strategy for system predictor identification, *Journal of Hydrology* (239): 232–239.

- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, 1 edn, Chapman and Hall.
- Soofi, E. S. (2000). Principal information theoretic approaches, *Journal of the American Statistical Association* pp. 1349–1353.
- Specht, D. F. (1991). A General Regression Neural Network, *IEEE Transactions on Neural Networks* **2**(6): 568–576.
- Von Zuben, F. J. (2003). Uma Caricatura Funcional de Redes Neurais Artificiais, *Learning and Nonlinear Models* **1**(2): 77–87.
- Zaffalon, M. and Hutter, M. (2002). Robust features selection by mutual information distributions, *18th International Conference on Uncertainty in Artificial Intelligence*, pp. 577–584.
- Zealand, C. M., Burn, D. H. and Simonovic, S. P. (1999). Short term streamflow forecasting using artificial neural networks, *Journal of Hydrology* (214): 32–48.
- Zheng, G. L. and Billings, S. A. (1995). Radial Basis Function Network Configuration Using Mutual Information and the Orthogonal Least Squares Algorithm, *Neural Networks* **9**(9): 1619–1637.