

MÉTODOS ESTOCÁSTICOS E NÃO-ESTOCÁSTICOS PARA A PREVISÃO DA VAZÃO EM RIOS

Guilherme Batista Cervilha - Uni-FACEF

Dr. Antônio Carlos da Silva Filho - Uni-FACEF

INTRODUÇÃO

Atualmente, mais de 85% da produção de energia elétrica, no Brasil, é hidrelétrica, mesmo com a diversificação devido à criação de termelétricas a gás, nucleares, etc. A capacidade brasileira de geração baseia-se em dois elementos gratuitos: a água das chuvas e a força da gravidade. O Brasil possui bacias hidrográficas compostas de centenas de rios permanentes e caudalosos e que se distribuem pelas regiões Norte, Nordeste, Sul, Sudeste e Centro-Oeste apresentando regimes de chuvas bem diferenciados. O sistema hidroelétrico brasileiro apresenta aspectos peculiares que o tornam diferente dos outros sistemas de base hidroelétrica existentes no mundo. Primeiro, pelas características das afluências nos rios brasileiros, que apresentam acentuada sazonalidade e elevado nível de incerteza, ao contrário do sistema do hemisfério norte, no qual o regime hidrológico é proveniente, basicamente, do degelo. Segundo, pelo isolamento dos sistemas, que, ao contrário dos demais sistemas hidroelétricos, não possuem interligação com sistemas vizinhos de base termoelétrica. Terceiro, porque apresenta forte acoplamento hidráulico entre suas unidades.

Posto isto, o planejamento da operação dessas usinas depende do conhecimento prévio do volume de água disponível nesses reservatórios. Dito de outra forma, é preciso conhecer a quantidade de água que estará disponível, com certa antecedência, para que se possa, finalmente, estimar a quantidade máxima de energia a ser gerada pela usina (SOARES, 2001). Assim, em função dos bons resultados da previsão das vazões, é possível realizar o planejamento energético de modo a atingir níveis ótimos de geração de energia.

Pode-se prever a vazão a curto, médio e longo prazo (TUCCI, 2002). A previsão de curto prazo é usada para estimar a vazão num trecho da bacia com antecedência de horas ou poucos dias (TUCCI, 2002). A previsão de médio prazo envolve a previsão de vazão com antecedência de um a alguns meses e depende, fortemente, das condições conhecidas de clima e dos oceanos que podem influenciar as vazões futuras na bacia hidrográfica (TUCCI, 2002). A previsão de longo prazo é a estimativa, normalmente estatística, dos riscos de ocorrência de vazões num determinado local, como por exemplo: o risco de inundação numa seção de um rio, a probabilidade de períodos úmidos e secos, dentre outros (TUCCI, 2002).

O setor elétrico usa, normalmente, o método de Box-Jenkins (BOX, 1976), (HOFF, 1983) para prever a vazão afluente. Este método pressupõe uma relação linear entre os valores presente e passado da vazão. Podem ser aliados aos modelos lineares autoregressivos (AR), de média móvel (MA) e o autoregressivo de média móvel (ARMA), que igualmente não são os mais adequados para tratar um conjunto de dados com características não lineares e não estacionárias (CHATFIELD, 1991) inerentes às séries de vazão. Mas existem alternativas. Assim, por exemplo, as Redes Neurais Artificiais (RNAs) (FOG, 1995; LACHTERMACHER, 1995; SARLE, 1995) são modelos compostos de um certo número de elementos não lineares (neurônio artificial) operando em paralelo e dispostos em camadas similarmente às redes neurais biológicas. Através de algoritmos específicos, as RNAs podem aprender certo conhecimento (HAYKIN, 1994; EVANS, 1991; SIQUEIRA, 2002). As RNAs podem ser grupadas em duas categorias: redes sem realimentação (*feedforward networks*) e com realimentação (*recurrent networks*). As redes sem realimentação são estáticas, isto é, uma dada entrada pode produzir apenas um conjunto de saída; logo, não produzem nenhum tipo de memória. Por outro lado as redes recorrentes permitem que as características temporais da informação sejam memorizadas.

O objetivo deste trabalho é investigar a previsão da vazão afluente, num período de alguns anos, mês a mês, para vários rios. Pretende-se fornecer subsídios para auxiliar a decisão dos especialistas do setor elétrico envolvidos em planejamento energético.

MATERIAIS E MÉTODOS

Desenvolvemos uma investigação para um melhor conhecimento a respeito do processo responsável pela coleta dos dados fluviométricos históricos dos rios de grande importância social e econômica do estado de São Paulo. De forma genérica, o processo de medição é bem simples, concretizado conforme as seguintes etapas:

1. Elege-se uma seção do rio, onde não há turbulência e com seção transversal bem definida. Nesta seção, fixam-se réguas graduadas, onde se medem os níveis diários d'água;
2. A partir de então, calculam-se as áreas para cada nível de água.
3. Finalmente, obtém-se a vazão multiplicando a área encontrada pela velocidade da água.

Hoje, as leituras das vazões dos rios foram automatizadas e são informadas em tempo real via satélite, com possibilidades reduzidas de erro.

Os dados fluviométricos históricos foram coletados no Sistema de Informação para o Gerenciamento de Recursos Hídricos do Estado de São Paulo (SIGRH) através do site: <http://www.sigrh.sp.gov.br/cgi-bin/bdhm.exe/flu>. Assim, os rios selecionados foram: Rio Pardo, Rio Piracicaba, Rio Paraíba do Sul e Rio Jacaré Guaçu. Enfocamos o estudo também nas séries temporais do Rio São Francisco (de grande importância nacional) e os dados fluviométricos históricos do mesmo foram captados no site da Agência Nacional das Águas através do endereço eletrônico: <http://200.140.135.139/Usuario/DadoPesquisar.aspx>

Exemplo de séries temporais históricas:

Nome:	Rio Pardo
Município:	Ribeirão Preto/SP
Prefixo:	4C-001
Latitude:	21°06'13"
Longitude:	47°45'26"
Área (KM²):	10679

Tabela 1. Vazões para o Rio Pardo

DADOS MENSAIS DE VAZÃO (m³/s) - ESTAÇÃO FLUVIOMÉTRICA

ANO	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	MEDIA
1944	181	201	289	174	111	94	77	60	49	58	121	94	126
1945	131	314	211	130	91	89	81	60	49	51	101	264	131
1946	442	260	275	180	117	100	103	70	57	78	81	92	155
1947	248	260	536	240	162	121	105	83	100	132	133	229	196
1948	293	323	333	187	130	100	81	70	55	71	120	155	160
1949	226	321	231	220	135	113	84	61	55	43	83	164	145
1950	283	537	343	242	157	116	95	74	59	83	134	278	200
1951	296	396	334	251	149	117	100	86	66	68	73	112	171
1952	148	222	320	183	108	108	84	65	57	68	93	70	127
1953	73	85	88	120	62	60	52	42	46	52	69	100	71
1954	110	228	112	93	120	89	61	44	34	41	42	111	90
1955	171	98	144	127	73	69	46	40	43	41	62	182	91
1956	168	116	139	87	95	99	70	86	62	60	67	148	100
1957	216	254	288	212	153	105	100	79	109	81	123	160	157
1958	181	259	248	164	177	153	109	79	84	117	127	136	153
1959	333	249	212	172	111	87	65	71	53	49	90	116	134
1960	266	286	297	150	132	103	82	60	48	54	102	237	151
1961	318	416	422	229	191	123	99	82	62	57	111	127	186
1962	203	368	343	178	141	118	88	73	79	135	120	244	174
1963	435	307	202	124	91	79	70	62	50	50	74	45	132
1964	91	201	104	77	103	65	68	49	40	94	100	232	102
1965	336	424	430	203	138	91	92	65	60	126	157	271	199
1966	379	297	440	234	173	128	98	86	94	81	171	262	204
1967	413	448	282	195	151	140	106	97	97	93	128	220	197
1968	220	164	178	120	98	88	82	87	92	101	83	100	118
1969	79	85	86	77	79	69	48	47	36	46	98	118	72
1970	214	365	294	188	122	115	105	114	106	129	130	93	165
1971	70	68	83	83	89	104	79	79	66	92	83	167	89
1972	191	397	307	187	133	106	125	104	98	153	217	195	184
1973	250	268	176	221	156	131	105	95	81	84	119	203	157
1974	368	220	298	219	135	122	98	86	94	93	120	265	176
1975	206	286	230	187	147	124	117	77	66	79	167	202	157
1976	189	330	424	327	208	205	199	185	209	187	240	325	252
1977	514	378	235	294	185	150	114	90	113	118	143	224	213
1978	291	190	167	118	109	114	91	74	75	86	162	208	140
1979	202	269	199	117	133	120	95	182	194	110	117	213	163
1980	301	346	218	264	175	146	127	126	130	128	154	280	200
1981	463	303	179	118	110	123	125	112	88	165	262	430	207
1982	386	329	429	281	205	175	152	161	153	206	155	365	250
1983	542	660	473	331	273	383	248	204	334	367	313	461	382
1984	387	240	167	137	164	110	117	133	144	126	105	199	169
1985	332	303	346	202	155	158	133	113	101	110	118	121	183
1986	196	236	203	186	175	116	107	114	89	107	130	266	160
1987	351	333	215	184	200	165	136	84	103	143	118	141	181

1988	178	256	245	167	131	132	122	91	76	127	118	124	147
1989	255	310	253	155	121	103	102	99	122	121	127	189	163
1990	349	149	247	161	135	90	79	85	81	99	82	92	137
1991	168	413	317	381	226	166	122	100	102	172	96	174	203
1992	235	272	226	176	165	113	104	93	128	178	264	179	178
1993	161	346	269	227	153	141	105	104	110	111	115	124	164
1994	218	148	205	157	141	107	98	86	75	87	98	157	131
1995	179	642	326	252	159	113	100	86	80	130	130	178	198
1996	355	234	297	189	157	114	98	85	123	142	246	279	193
1997	539	400	298	191	155	225	156	115	98	103	160	209	221
1998	169	276	208	158	137	139	109	94	73	104	89	198	146
1999	314	248	338	175	124	111	101	91	93	86	90	119	157
2000	346	403	265	169	117	102	93	91	106	80	140	180	174
2001	170	131	109	98	89	77	71	64	73	86	120	182	106
2002	266	349	264	145	110	91	89	85	110	68	157	160	158
2003	322	352	216	174	133	91	87	87	83	75	115	157	158
MÉDIA	265	293	260	183	140	120	101	89	90	103	128	189	163
MÍNIMA	70	68	83	77	62	60	46	40	34	41	42	45	71
MÁXIMA	542	660	536	381	273	383	248	204	334	367	313	461	382

Iniciamos o estudo com algo mais tradicional, como a aplicação do método polinômios de regressão. Assim, as variáveis trabalhadas foram as seguintes: Y é a variável dos dados fluviométricos históricos e X é o conjunto dos números inteiros positivos.

Em uma primeira etapa, utilizamos os primeiros doze pontos da série para calcularmos a regressão e, posteriormente, prever para os próximos pontos. Variamos o grau do polinômio de regressão de um (polinômio de primeiro grau ou regressão linear) até o grau de polinômio cinco (o polinômio de regressão de mais alto grau). Como estamos trabalhando com séries mensais, previmos até doze pontos à frente. Uma vez feita a previsão, comparamos o resultado ponto a ponto, com os reais (que estão na série) da seguinte forma:

1. Calculamos o módulo do desvio, que é o módulo da diferença entre os valores reais e os previstos;
2. Com o resultado do item 1 em mãos, encontramos o desvio absoluto em porcentagem, dividindo o resultado do item 1 pelo valor real correspondente e multiplicamos por 100 (cem).

3. Ao final do processo, fizemos uma média para o cálculo do desvio absoluto percentual médio, conhecido como MAPE. A interpretação do MAPE é que ele é o erro percentual médio da previsão, ou seja, uma medida da precisão (ou da incerteza) do método. Quanto menor o valor do MAPE, melhor será a nossa previsão.

Posteriormente, utilizamos do segundo ponto ao décimo-terceiro ponto da série e repetimos todo o processo. Esta metodologia se repetiu até a utilização de todos os pontos possíveis da série.

Após isto, repetimos o mesmo processo utilizando vinte e quatro, trinta e seis, quarenta e oito e sessenta pontos para calcularmos a regressão, mas sempre prevendo doze pontos à frente.

Em uma segunda etapa do método dos polinômios de regressão, criamos como variação doze séries mensais: uma para o mês de janeiro, outra para o mês de fevereiro e assim por diante. Como neste caso não houve mais ciclos de doze meses, utilizamos os vinte primeiros pontos (que equivalem por vinte anos) para calcularmos a regressão e fizemos a previsão para os cinco pontos à frente. Como na primeira etapa, os resultados das previsão foram comparados com os valores reais, com a perspectiva de encontrarmos o menor valor para o desvio absoluto percentual médio (MAPE). Após isto, utilizamos do segundo ponto até o vigésimo-primeiro e aplicamos novamente o método. Repetimos esta metodologia até a utilização total de todos os pontos possíveis da série.

Em um segundo instante, aplicamos o método das médias móveis de 3, 5, 7, 9, 11, 23, 35, 47 e 59 meses, nas séries temporais das vazões dos rios. Assim como no método do polinômios de regressão, também utilizamos o valor do MAPE para avaliação dos resultados em todas as etapas do estudo das médias móveis. Na primeira etapa, consideramos que o pontos originados nas médias móveis corresponderia à previsão do próximo ponto da série. Exemplo:

$$X = [X1 , X2 , X3 , X4 , X5 , X6]$$

$$\text{Média Móvel 3 meses} = M1$$

$$M1 = \frac{X1 + X2 + X3}{3}$$

3

$$M1 = Y4$$

Y4 = Valor Previsto

Na segunda etapa, aplicamos o método das médias móveis de 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 35, 37, 39, 41 e 43 meses nas séries mensais.

Na terceira etapa, consideramos que o primeiro ponto originado das médias móveis seria igual o segundo, o segundo ponto seria igual ao terceiro e assim por diante. Aplicamos nesta etapa médias móveis de 3, 5, 7, 9, 11, 23, 35, 47 e 59. Com isso, obteríamos o valor da previsão. Exemplo:

$$X = [X1 , X2 , X3 , X4 , X5 , X6]$$

$$\text{Média Móvel 3 meses} = M1$$

$$M1 = \frac{X1 + X2 + X3}{3}$$

$$M1 = M2$$

$$M2 = \frac{X2 + X3 + Y4}{3}$$

$$Y4 = 3M2 - X2 - X3$$

$$Y4 = \text{Valor Previsto}$$

Na quarta etapa, aplicamos as médias móveis de 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 35, 37, 39, 41 e 43 meses com a mesma metodologia da terceira etapa nas séries mensais.

Na quinta etapa fizemos um tratamento nos dados brutos das séries temporais mensais encontrando o valor máximo e mínimo da mesma e calculando a diferença entre eles. Depois, subtraímos todos os valores das séries mensais pelo valor mínimo e dividimos resultado pela diferença entre o valor máximo e mínimo. Com isto, criamos uma nova série com dados variando entre 0 e 1 e, posteriormente, aplicamos o método do polinômio de regressão nestes pontos, utilizando 5, 10, 15 e 18 pontos para o cálculo da regressão, variamos o grau do polinômio de um até cinco e previmos até 5 pontos à frente.

Todas as análises, foram realizadas no MATLAB (HANSELMAN, 2003),

RESULTADOS

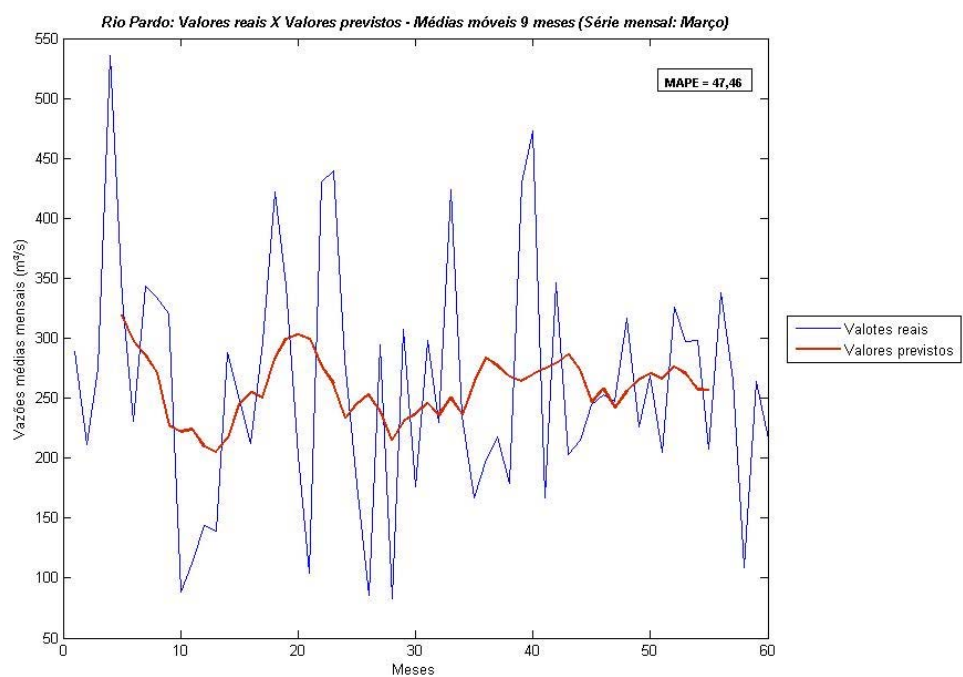
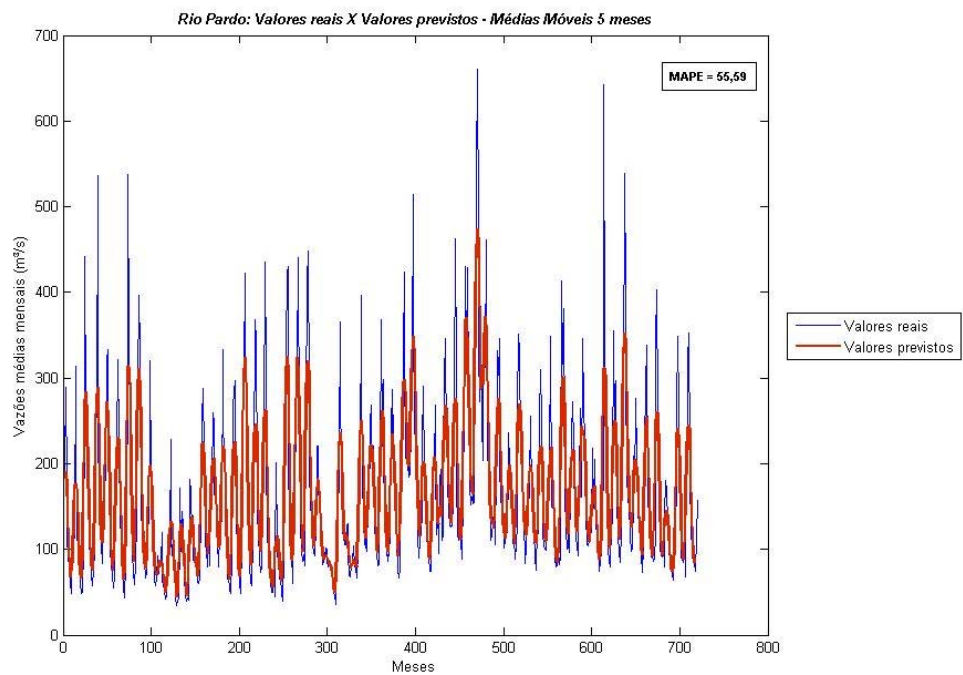
Alguns dos resultados dos métodos que obtiveram MAPE < 10 podem ser vistos na tabela abaixo, onde os casos com tratamento referem-se àqueles casos onde os valores originais das séries foram substituídos por valores “alisados”, segundo o método das diferenças normalizadas:

Tabela 2. Alguns casos com MAPE < 10 para o Rio Pardo

Método	Mês das séries mensais	Quantidade de pontos a serem utilizados na regressão (np)	Quantidade de pontos a serem previstos (npr)	Grau máximo dos polinômios a serem construídos (gm)	Com tratamento / Sem tratamento	MAPE
Polinômios de regressão	Agosto	18	5	5	Sem tratamento	8,24
Polinômios de regressão	Setembro	18	5	5	Sem tratamento	9,64
Polinômios de regressão	Outubro	18	5	5	Sem tratamento	1,25
Polinômios de regressão	Outubro	18	5	5	Com tratamento	3,12

ANÁLISE

Como em todo trabalho de pesquisa científica obtivemos resultados relevantes que superaram as expectativas, mas por outro lado, também obtivemos resultados irrelevantes. É importante salientar que o objetivo de nosso estudo é a comparação de vários métodos para descobrir qual método é melhor para cada caso e, os resultados irrelevantes merecem uma análise especial, pois os mesmos podem fornecer informações importantes para ajustamento dos métodos quantitativos. Alguns destes resultados com MAPE alto foram obtidos pelo método de médias móveis e são exibidos nos gráficos abaixo:



As dificuldades apresentadas durante o trabalho se referem ao elevado grau de dispersão dos dados fluviométricos históricos.

CONCLUSÃO

Os resultados alcançados através da metodologia proposta parecem ser adequados para a previsão da vazões de rios. Para o planejamento estratégico e de investimento no setor energético erro médio menor que 10% pode ser considerado irrelevante devido ao alto grau de dispersão dos dados analisados.

Observa-se nitidamente que com o refinamento dos dados fluviométricos históricos, dos métodos aplicados e com a combinação entre os métodos cresce a expectativa de obtermos melhores resultados, já que há uma rica fonte de alternativas ainda não exploradas.

BIBLIOGRAFIA

AQUINO, R., B. de, CARVALHO, M., A., Jr., SOUZA, B., A., 1999. Redes Neurais Artificiais: Uma Aplicação ao Planejamento da Operação de Sistemas Hidrotérmicos de Geração. In *Proceedings of the IV Brazilian ENGEVISTA*, v. 7, n. 1, p. 63-72, abril 2005 *72 Conference on Neural Networks – IV Congresso Brasileiro de Redes Neurais*, pp. 164-169.

BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. California, USA, San Francisco: Holden Day, 2. ed., 1976.

CHAKRABORTY, K., MEHROTA, K., CHILUKURI, K. M., and RANKA, S., 1992. Forecasting the Behaviour of Multivariate Time Series Using Neural Networks. In *Neural Networks*, v5, 961-970.

CHATFIELD E., 1991. *The Analysis of Time Series*, New York, USA, Chapman and Hall, fourth edition.

ELMAN J. L., Finding Structure in Time, 1990. In *Cognitive Science*, vol. 14, pp. 179-211.

EVANS, R. M. and S. ALVIN, J., 1991. Relating Numbers of Processing Elements in a Sparse Distributed Memory Model to Learning Rate and Generalization, In *ACM APL Quote Quad*, v21(4), 166-173.

FOG, T.L. et al, 1995. Training and Evaluation of Neural Networks for Multi-Variate Time Series Processing. In *Proceedings of IEEE International Conference on Neural Networks*. IEEE Press.

HANSELMAN, Duane; LITTLEFIELD, Bruce. *Matlab 6 – Curso Completo*. São Paulo: Prentice Hall, 2003. 676 p.

HAYKIN S. , 1994. *Neural Networks, a comprehensive Foundation*; London, Macmillan College Publishing Co.

HILL, T., O'CONNOR, M. e REMUS, W. Neural network models for time series forecasts. *Management Science*. v. 42, n: 7, p. 1082 – 1092, jul. 1996.

HOFF C. J. *A Practical Guide to Box-Jenkins Forecasting*, Belmont, CA., USA: Lifetime Learning Publications. 1983.

LACHTERMACHER, G. and FULLER, J.D., 1995. Backpropagation in Time series Forecasting. In *Journal of Forecasting*. Vol 14, 381-393.

MAKRIDAKIS, S., WHEELWRIGHT, S. C., McGEE, V. E. *Forecasting: methods and applications*. 2 .ed. John Wiley e Sons, 1983.

MILIONI, A., Z., e ACIOLI A., O., 2001. Modelos Econométricos para a Previsão de Cheias Fluviais . In *Simpósio Brasileiro de Pesquisa Operacional*.

OLIVO, Acioli Antonio de; MILIONI, Armando Zeferino . Modelos Econométricos para a Previsão de Cheias Fluviais. In: XXXIII SBPO - Simpósio Brasileiro de Pesquisa Operacional, 2001, Campos do Jordão, SP. Anais do XXXIII SBPO, 2001. v. 1. p. 534-542.

SARLE, W.S., 1995. Stopped Training and other remedies for Overfitting. In *Proceedings of the 27th Symposium on the Interface*. Eletropaulo, 1995. Rios / Reservatórios / Enchentes. In *Departamento de Patrimônio Histórico. História e Energia*, v. 5.

SIQUEIRA, T. G., SOARES F., S., 2002. Aplicação de Redes Neurais com função de ativação de Base Radial na predição de séries temporais não estacionárias, in *XIV Congresso Brasileiro de Automação*.

SOARES F., S., BESSA, M. R., 2001. Otimização da Operação de Reservatórios de superfície: Um estudo comparativo entre diferentes metodologias aplicadas ao sistema de Furnas, In *XVI SNPTEE*.

TUCCI, C. E. M., BRUN, G., 2001. Previsão em tempo real do volume afluente ao reservatório de Ernestina. In *Revista Brasileira de Recursos Hídricos*. V.6, n.2, p.73 – 79.

TUCCI, C. E. M., ROBIN T. C., DIAS P. L. Da S., COLLISCHONN W., 2002. Previsão de médio prazo da afluência de reservatórios com base na previsão climática. In Relatório Final Projeto: BRA/00/029, Instituto de Astronomia, Geofísica e Ciências Atmosféricas Universidade de São Paulo e Instituto de Pesquisas Hidráulicas Universidade Federal do Rio Grande do Sul.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*. Kent(Ohio) 14, p. 35 – 62, 1998.

ZURADA J. M. and CHOLEWO T., J., Sequential Network Construction for Time Series Prediction. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp 2034–2039, 1997.